

# Populationsgenetik eisenzeitlicher Reiternomaden der Eurasischen Steppe

*Next Generation Sequencing* alter DNA aus archäologischen humanen Skeletten  
des 1. Jahrtausends vor Christus

Dissertation  
zur Erlangung des Grades  
Doktor der Naturwissenschaften

Am Fachbereich Biologie  
der Johannes Gutenberg-Universität Mainz

Martina Unterländer  
geb. am 04.08.1981 in Speyer

Mainz 2014



Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 28.05.2014



# Vorbemerkung

Die vorliegende Dissertation wurde im Rahmen des vom BMBF geförderten interdisziplinären Verbundprojektes „Palaeogenetische Untersuchungen zu wirtschaftlichen Innovationen und sozialer Mobilität in der eurasischen Steppe 3.500–300 v. Chr.“ angefertigt. Die palaeogenetischen Untersuchungen wurden in zwei Teilprojekte gegliedert, wobei Frau S. W. die kupfer- und bronzezeitlichen Populationen des nordpontischen Raumes untersuchte und sich die vorliegende Arbeit den eisenzeitlichen Reiternomaden Zentralasiens widmet.

Neben der populationsgenetischen Untersuchung des Probenmaterials und der Interpretation der Ergebnisse im Hinblick auf die demographischen Prozesse in der Eurasischen Steppe während der Eisenzeit, sollte eine Methode zur Maximierung des Datenvolumens bei gleichzeitiger Einsparung wertvollen Knochen- und Zahnmaterials etabliert werden.

Um der Darstellung beider Ansätze gerecht zu werden, wurde die vorliegende Arbeit in einen populationsgenetischen Teil A und einen methodischen Teil B untergliedert.



# Inhaltsverzeichnis

A Populationsgenetik .....	1
A1 Einleitung .....	1
A1.1 Die Palaeopopulationsgenetik des Menschen .....	1
A1.2 Die Skythen .....	8
A1.3 Die Palaeopopulationsgenetik der Skythen .....	10
A2 Die populationshistorische Fragestellung .....	14
A3 Das Untersuchungsmaterial .....	16
A4 Methodenüberblick .....	19
A4.1 Die untersuchten Marker .....	19
A4.2 Die selektive Anreicherung der Marker aus alter DNA .....	25
A4.3 DNA-Sequenzierung und Genotypisierung .....	26
A4.4 Biostatistik .....	27
A5 Ergebnisse .....	28
A5.1 Datensatz skythischer Individuen bzw. Populationen .....	28
A5.2 Haplogruppenzusammensetzung der skythischen Populationen .....	34
A5.3 Populationsgenetik der skythischen Kulturgruppen .....	36
A5.4 Die genetische Struktur der Pazyryk-Population .....	43
A5.5 Das Ursprungsgebiet der Skythen .....	45
A5.6 Die phänotypischen Marker .....	47
A6 Diskussion .....	54
A6.1 Anmerkungen zur verwendeten Stichprobe und den Markern .....	54
A6.2 Die mitochondriale Variabilität skythischer Populationen .....	57
A6.3 Die Populationsgenetik der skythischen Kulturgruppen .....	60
A6.4 Die Struktur der Pazyryk-Population .....	64
A6.5 Das Ursprungsgebiet der Skythen .....	65

A6.6 Die phänotypischen Charakteristika der Skythen.....	71
A7 Zusammenfassung .....	79
A8 Ausblick.....	86
B Methoden.....	87
B 1 Methodischer Stand der palaeogenetischen Forschung.....	87
B 2 Methodische Zielsetzung .....	92
B 3 Methoden zur Datengenerierung .....	93
B 3.1 Untersuchungsmaterial .....	93
B 3.2 Die Auswahl der mitochondrialen Marker .....	94
B 3.3 Die nukleären Marker .....	96
B3.4 DNA-Anreicherungsverfahren .....	98
B3.5 Markierung der DNA-Fragmente zur parallelen Sequenzanalyse.....	108
B3.6 Sanger Sequenzierung und <i>Next Generation Sequencing</i> .....	114
B3.7 Kontaminationskontrolle .....	128
B4 Verarbeitung der Rohdaten.....	129
B5 Methoden der Datenanalyse .....	131
B6 Ergebnisse.....	137
B6.1 DNA-Erhalt .....	137
B6.2 Amplifikationserfolg .....	138
B6.3 Sequenziererfolg.....	140
B6.4 Kontamination .....	144
B6.5 Haplogruppenzuordnung .....	150
B6.6 Genotypen.....	152
B7 Diskussion .....	154
B7.1 DNA-Erhalt .....	154
B7.2 Kontaminationen .....	155
B7.3 Auswahl der mitochondrialen Marker.....	160



B7.4 Anreicherungsmethoden im Vergleich .....	162
B7.5 Sequenziermethoden und –erfolg .....	165
B7.6 Authentizität der Daten.....	168
B7.7 Erreichung der gesetzten Ziele .....	170
B8 Zusammenfassung der methodischen Ergebnisse .....	172
Referenzen.....	173
Anhang .....	191
Abbildungsverzeichnis .....	191
Tabellenverzeichnis.....	192
Alignment der HVR1-Konsensussequenzen .....	197
SNPs der kodierenden Region des mitochondrialen Genoms.....	202
Haplogruppenzuordnung anhand von SNPs der kodierenden Region und der HVR	207
Paarweise $F_{ST}$ -Werte mit zugehörigen <i>p-values</i> .....	212
Reynold’s distances.....	216
Frequenzen der abgeleiteten Allele für alle untersuchten Marker .....	217

## Inhalt der Daten-DVD

Daten der 454-Sequenzierungen mit Zusatzinformationen

Sanger-Sequenzen

Informationen zum *Capture* und der Illumina-Sequenzierung

Sequenzdaten und Zusatzinformationen zu den Referenzpopulationen

Tabelle mit den Polymorphismen der mitochondrialen DNA



# A Populationsgenetik

## A1 Einleitung

### A1.1 Die Palaeopopulationsgenetik des Menschen

Die Erforschung prähistorischer Populationen, ihrer Lebensweise, Herkunft und Entwicklung erfolgte lange Zeit nur anhand von Bestattungsritualen, Siedlungsresten und materiellen Hinterlassenschaften des täglichen Lebens, wie beispielsweise Keramik und Steingeräten. Die Palaeopopulationsgenetik befasst sich hingegen mit der biologischen Verwandtschaft von Bevölkerungen. Der Genpool einer Bevölkerung ist das Resultat ihrer evolutionären und demographischen Entwicklung. So kann die Zunahme an neuen Mutationen ein Anzeichen für Bevölkerungsexpansion sein (FU 1997). Ein Heterozygotendefizit kann auf eine Populationsstruktur, oder auch Inzucht hinweisen (ROUSSET & RAYMOND 1995). Durch die Berechnung der genetischen Distanz kann die Verwandtschaft zwischen verschiedenen Bevölkerungsgruppen evaluiert werden. Der in der Vergangenheit am häufigsten verwendete genetische Marker der Palaeogenetik ist die mitochondriale DNA. Ihr Vorteil liegt in der erhöhten Kopienzahl im Vergleich zur DNA des Zellkerns – der nukleären DNA – und in ihrem Vererbungsmodus. Anders als die nukleäre DNA, bei der jeder Marker in zweifacher Ausführung, den sog. Allelen, von denen eines vom Vater und eines von der Mutter ererbt wurde, vorliegt, wird die mitochondriale DNA lediglich uniparental von der Mutter an ihre Kinder weitergegeben. Dadurch ist es möglich einen maternalen Stammbaum über mehrere Generationen nachzuvollziehen. Die erhöhte Kopienzahl erlaubt es selbst in stark degradiertem archäologischem Skelettmaterial noch mitochondriale DNA nachzuweisen und zu untersuchen. So konnte durch die Analyse mitochondrialer DNA palaeo-, meso- und neolithischer Populationen gezeigt werden, dass es keine kontinuierliche Bevölkerungsentwicklung vom Palaeo- zum Neolithikum in Zentraleuropa gab (BRAMANTI *ET AL.* 2009) und dass auch 2.000 Jahre nach Ankunft der ersten Ackerbauern in Zentraleuropa um 5.500 v. Chr. sich diese, zumindest in einigen Gebieten, nicht mit den Jäger-Sammlern vermischt haben, sondern in Parallelgesellschaften nebeneinander existierten (BOLLONGINO *ET AL.* 2013). Eine umfangreiche Studie an 364 frühneolithischen bis bronzzeitlichen Individuen des Mittel-Elbe-Saale Gebietes zeigt die Bevölkerungsentwicklung in dieser Zeit anhand

mitochondrialer Linien auf. Die Ergebnisse lassen auf eine zunächst kontinuierliche Entwicklung nach Einführung des Ackerbaus durch die Ankunft der Linienbandkeramik (LBK) schließen. Gegen Ende des Frühneolithikums und im Mittelneolithikum weisen die Daten auf einen bidirektionalen Genfluss mit Süd-Skandinavien hin, durch den der Anteil an Jäger-Sammler-Linien wieder steigt, nachdem diese in der LBK nicht nachgewiesen werden konnten. Mit dem Auftreten der Schnurkeramik im Spätneolithikum tauchen neue mitochondriale Linien auf, die von den Autoren mit einem osteuropäischen Einfluss assoziiert werden, wohingegen die etwa 300 Jahre später eintreffende Glockenbecherkultur als Verbindung zur Iberischen Halbinsel interpretiert wird. All diese Einflüsse zusammengekommen führen zu einer Zusammensetzung des mitochondrialen Genpools wie sie auch im modernen Zentraleuropa anzutreffen ist (BRANDT *ET AL.* 2013). In einem weiteren Fall konnte durch die Untersuchung an mitochondrialer DNA nachgewiesen werden, dass die Begründer der minoischen Kultur Kretas nicht wie von Sir Arthur Evans angenommen aus Afrika stammen (EVANS 1921). Die molekularen Analysen ergaben, dass sich diese bronzezeitliche Zivilisation wahrscheinlich aus den Nachfahren der neolithischen Erstbesiedelung der Insel Kreta entwickelte (HUGHEY *ET AL.* 2013). Die Palaeopopulationsgenetik ermöglicht es, den Blick auf die Menschen vergangener Jahrtausende zu richten und ihre demographische Entwicklung nachzuvollziehen. Durch die palaeogenetische Forschung ist es weiterhin möglich die Verbindung zu anderen *Hominini*, wie dem Neandertaler, näher zu beleuchten. So zeigten erste Untersuchungen an Fragmenten mitochondrialer DNA des Neandertalers, dass diese außerhalb des Variationsbereiches moderner Menschen liegt (KRINGS *ET AL.* 1997), was durch die spätere Analyse ganzer mitochondrialer Genome bestätigt wurde (BRIGGS *ET AL.* 2009A). Die Untersuchung der mitochondrialen DNA eines Fingerknochens, der in der Denisova-Höhle in Sibirien gefunden wurde, offenbarte eine maternale Linie, die sich stark von denen modernen Menschen und Neandertalern unterscheidet und nicht mehr als Schwestertaxon zu diesen eingeordnet werden kann (KRAUSE *ET AL.* 2010B).

Weiterhin konnte ein fast komplettes mitochondriales Genom eines Individuums aus dem Mittleren Pleistozän untersucht werden. Die Überreste stammen von der Fundstelle *Sima de los Huesos* in einer Höhle des Gebirgszuges *Sierra de Atapuerca* in Nord-Spanien. Sie zeigen morphologische Ähnlichkeit zu Fossilien, die als *Homo heidelbergensis* klassifiziert wurden und weisen zusätzlich Neandertaler-Merkmale auf. Eine phylogenetische Analyse mit modernen Populationen, Denisova-Individuen und Neandertalern ergab, dass die mtDNA des *Sima de los Huesos* Individuums eine größere Ähnlichkeit zu Denisovans als zu

Neandertalern zeigt und sich einen gemeinsamen Vorfahren mit den Denisova-Individuen teilt (MEYER *ET AL.* 2014).

Bislang gibt es zahlreiche Studien prähistorischer und moderner Populationen, die anhand mitochondrialer DNA durchgeführt wurden. Die Untersuchung dieses Markers bleibt deshalb besonders attraktiv, obwohl er nur die mütterliche Seite einer Populationshistorie widerspiegelt, da große Mengen an Vergleichsdaten zur Verfügung stehen. Die Untersuchung nukleärer Marker ist aufwendiger, wurde jedoch auch für prähistorisches Probenmaterial bereits erfolgreich eingesetzt. Sie ermöglicht es beispielsweise Selektionsereignisse einzelner genetischer Marker zu detektieren und ihren Ursprung mit besonderen Begebenheiten der Vergangenheit zu korrelieren. So konnte die Selektion des Markers -13.910\*T für Laktasepersistenz mit der Ausbreitung des Neolithikums und mit ihm der Vieh- und Milchwirtschaft in Zentraleuropa in Zusammenhang gebracht werden (BURGER *ET AL.* 2007, ITAN *ET AL.* 2009). Durch die Analyse des *FOXP2* Gens wurde gezeigt, dass Neandertaler und der moderne Mensch die gleiche Genvariante aufweisen, die angeblich mit der Fähigkeit zu Sprechen in Verbindung steht. Diese Studie offenbarte nicht nur, dass der Neandertaler die gleichen Möglichkeiten zur Sprachentwicklung hatte wie der moderne Mensch, sondern auch, dass der Selektionszeitpunkt dieser Mutation vor dem letzten gemeinsamen Vorfahren gelegen haben muss (KRAUSE *ET AL.* 2007).

Die Entwicklung neuer DNA-Anreicherungsverfahren und Sequenzieretechniken ermöglichte die Analyse einer größeren Menge nukleärer DNA als bisher – bis hin zu ganzen Genomen. Die Anwendung auf alte DNA beschränkt sich bislang jedoch auf die Analyse einzelner, bzw. weniger Individuen. Konnte durch die vergleichende Analyse mitochondrialer Genome von modernen Menschen und Neandertalern keine genetische Vermischung festgestellt werden, zeigt der Vergleich eines autosomalen Genoms, das aus den Sequenzen von drei verschiedenen Neandertalern zusammengesetzt wurde, mit fünf Genomen moderner Menschen aus unterschiedlichen Regionen der Welt, Hinweise für Genfluss vom Neandertaler zu Nicht-Afrikanern (GREEN *ET AL.* 2010). Die Analyse des Denisova-Genoms zeigte, dass im Vergleich zu den Ergebnissen aus der Analyse der mitochondrialen DNA, eine phylogenetische Einordnung aufgrund autosomaler Daten, das Denisova-Individuum als Schwestergruppe des Neandertalers platziert (REICH *ET AL.* 2010). Durch diese Studien wurde verdeutlicht, dass uniparental vererbte Marker nur einen begrenzten Ausschnitt der genealogischen Information unseres Erbguts repräsentieren und erst die gleichzeitige Analyse multipler *loci* eine genauere Auflösung der Struktur einer Population erlaubt (VEERAMAH &

HAMMER 2014). Durch den Einsatz einer neuen Methode zur Sequenzier Vorbereitung gelang es, eine höhere Sequenzabdeckung (*coverage*) des Denisova-Genoms zu erreichen. Die größere Anzahl an Sequenzen ermöglicht es zum einen besser zwischen zufälligen Sequenzierfehlern und echten Genotypen zu unterscheiden und zum anderen wird die Gefahr reduziert einen Homozygotenüberschuss aufgrund von *allelic dropout* zu detektieren. Durch die höhere *coverage* war es somit möglich den Heterozygotenanteil des Genoms zu ermitteln, der wesentlich geringer als der Anteil in heutigen Populationen ist. Es konnten allerdings keine ungewöhnlich langen homozygoten Abschnitte im Genom entdeckt werden, wodurch Inzucht unter den unmittelbaren Vorfahren des untersuchten Individuums ausgeschlossen werden kann. Die Autoren schließen daraus auf eine sehr geringe genetische Diversität der Denisova-Population im Vergleich mit modernen Populationen (MEYER *ET AL.* 2012). Aus einem weiteren Knochenfragment aus der Denisova-Höhle im Altai-Gebirge, gelang es ebenfalls ein hervorragend abgedecktes Genom zu erhalten. Im Vergleich mit dem Denisova-Genom, dem Neandertaler-*draft*-Genom, einem Neandertaler-Genom aus dem Kaukasus und 25 Genomen moderner Populationen bildet das neue Genom eine phylogenetische Gruppe mit den Neandertaler Individuen. Anders als im Denisova-Genom konnten bei diesem Individuum lange homozygote Regionen im Genom nachgewiesen werden, die auf eine enge Verwandtschaft der Eltern hinweisen. Diese Regionen ausgenommen zeigte das Genom auch insgesamt eine sehr geringe Heterozygotität. Weiterhin konnte im Zuge dieser Studie der Anteil an Neandertaler-DNA in Populationen außerhalb Afrikas auf 1,5–2% bestimmt werden. Die Neandertaler Sequenzen in den modernen Populationen teilen mehr abgeleitete Allele mit dem Neandertaler aus dem Kaukasus als dem Altai-Neandertaler was dafür spricht, dass der Zeitpunkt des Genflusses nach der Trennung der verschiedenen Neandertaler Populationen gelegen haben muss (PRÜFER *ET AL.* 2014).

Das älteste Genom eines anatomisch modernen Menschen, das bisher sequenziert wurde stammt von einem 24.000 Jahre alten Individuum aus dem südsibirischen Mal'ta. Die genomische Analyse dieses Individuum zeigt im Vergleich mit 147 nicht-afrikanischen Populationen eine genetische Verbindung zu westeurasischen Populationen und Hinweise auf Genfluss der sibirischen Mal'ta Population zu den ersten amerikanischen Ureinwohnern. Die Autoren sehen diesen Befund als mögliche Erklärung für die cranialen Charakteristika der ersten Amerikaner, die nicht in die ostasiatische Variabilität fallen und für die Präsenz der mitochondrialen Haplogruppe X, die sich eher modernen westeurasischen als ostasiatischen Populationen zuordnen lässt (RAGHAVAN *ET AL.* 2014). Es wird allgemein angenommen, dass

die amerikanischen Ureinwohner über die Beringbrücke auf den amerikanischen Kontinent gelangten und asiatischen Ursprungs sind. Aus genetischer Sicht begründet sich die These auf die Analyse mitochondrialer Linien amerikanischer Ureinwohner, die zum größten Teil den Haplogruppen A, B, C und D zugeordnet werden können und sich in heutigen asiatischen Populationen wiederfinden. Zu einem geringen Anteil wurde jedoch auch die Haplogruppe X detektiert (BROWN *ET AL.* 1998, TORRONI *ET AL.* 1993). Die Präsenz dieser Haplogruppe und die fehlenden ostasiatischen Charakteristika in Schädeln amerikanischer Ureinwohner wurden als Indizien für die Solutréen-Hypothese verwendet, die einen europäischen Ursprung dieser Bevölkerung vermutet (BRADLEY & STANFORD 2004). Genfluss von der sibirischen Mal'ta Population des Jungpalaeolithikums zu den Vorfahren der amerikanischen Ureinwohner konnte auch Anhand genomischer Analysen eines ca. 12.600 Jahre alten Individuums aus West-Montana nachgewiesen werden, dessen Begräbnis Artefakte der Clovis-Kultur enthielt. Die Analysen zeigten eine größere genetische Affinität des Clovis-Individuums zu allen getesteten Populationen amerikanischer Ureinwohner als zu eurasischen Populationen. Unter den amerikanischen Populationen zeigt das Individuum eine größere genetische Affinität zu den indigenen Populationen Südamerikas (RASMUSSEN *ET AL.* 2014). Durch statistische Analysen anhand eines genomischen SNP-Set wurde die genetische Verwandtschaft des Clovis Individuums mit dem Mal'ta Individuum, einem 4.000 Jahre alten Palaeo-Eskimo und modernen Populationen untersucht. Für das Genom des Palaeo-Eskimos wurde bereits in einer früheren Studie eine größere genetische Affinität zu sibirischen Populationen aus dem Bereich der Beringstraße als zu amerikanischen Ureinwohnern und Grönland-Inuits nachgewiesen. Die Autoren schließen daraus, dass die Palaeo-Eskimo-Population, die einst die Beringstraße überquerte dies unabhängig von den Vorfahren der heutigen amerikanischen Ureinwohner und Inuits tat (RASMUSSEN *ET AL.* 2010). Die Ergebnisse der statistischen Analysen ergaben eine engere genetische Verbindung zwischen dem Individuum aus Mal'ta zu amerikanischen Ureinwohnern als zwischen diesen und dem Palaeo-Eskimo, was die vorherigen Studien bestätigt. Weiterhin zeigt das Clovis-Individuum eine engere Verbindung zu ostasiatischen und indigenen amerikanischen Populationen als das Mal'ta Individuum. Die Autoren sehen in den Ergebnissen die Bestätigung für ein Modell, das Genfluss von Ost-Asien und der Mal'ta Population zu den amerikanischen Ureinwohnern annimmt und somit die Solutréen-Hypothese negiert (RASMUSSEN *ET AL.* 2014).

Die ersten genomischen Daten zu mesolithischen Individuen aus Spanien stammen von zwei 7.000 Jahre alten Jäger-Sammlern des Fundplatzes La Braña-Arintero in León. Eine

Hauptkomponentenanalyse mit weltweiten Populationen platziert die La Braña Individuen in die Nähe der Variabilität moderner europäischer Populationen – werden nur europäische Populationen integriert, zeigen die spanischen Mesolithiker eine größere genetische Affinität zu nordeuropäischen als zu südeuropäischen Populationen (SÁNCHEZ-QUINTO *ET AL.* 2012). In einer weiterführenden Studie wurde das gesamte Genom des Individuums La Braña 1 bis zu einer 3,4fachen *coverage* sequenziert. Die Hauptkomponentenanalysen anhand eines umfangreicheren SNP-Sets bestätigt die Ergebnisse der ersten Studie und somit die Affinität zu nordeuropäischen Populationen. Weiterhin wurde die genetische Verbindung zu dem jungpalaeolithischen Individuum aus dem südsibirischen Mal'ta untersucht. Verschiedene statistische Tests auf gemeinsame genetische Abstammung legen nahe, dass das Individuum aus Mal'ta eine engere genetische Verwandtschaft zu dem Individuum aus La Braña zeigt als zu modernen europäischen und asiatischen Populationen. Die Autoren schließen daraus eine genetische Kontinuität im prähistorischen Eurasien. Die genomischen Analysen ergaben weiterhin, dass das Individuum aus La Braña die ancestralen Allele an verschiedenen Genen zeigt, die mit einer Pigmentierung der Haut assoziiert werden. Der Befund deutet darauf hin, dass die helle Hautfarbe moderner Europäer im Mesolithikum noch nicht allgegenwärtig war (OLALDE *ET AL.* 2014).

Im Februar 2012 wurde das Genom einer 5.300 Jahre alten Mumie veröffentlicht, die im September 1991 in den Ötztaler Alpen entdeckt wurde. Basierend auf den Resultaten von Hauptkomponentenanalysen, die anhand unterschiedlicher SNP-Sets durchgeführt wurden, schließen die Autoren auf eine genetische Verwandtschaft dieses Individuums zu modernen Populationen aus Sardinien. Durch die Untersuchung qualitativ unterschiedlicher SNP-Sets konnte gezeigt werden, dass die Zuordnung eines Individuums zu einer Population zuverlässiger ist, je höher die *coverage* der verwendeten SNP-Positionen ist, also je mehr Sequenzen zur Bestätigung eines Genotyps vorliegen. Im Zuge der metagenomischen Analyse aller verfügbaren Sequenzdaten dieses Individuums wurde ein geringer Teil der Sequenzen als bakterielle DNA identifiziert. 0,16% der bakteriellen Sequenzen konnten dem Bakterium *Borrelia burgdorferi* zugeordnet werden – ausreichend um 60% des Bakteriengenoms abzudecken. Dieser Befund deutet darauf hin, dass es sich hier um den ersten Nachweis der Infektion eines Menschen mit dem Erreger der Lyme-Borreliose handelt. Durch die Untersuchung verschiedener genomischer Marker wurde neben einer Prädisposition für eine Erkrankung der Herzkranzgefäße festgestellt, dass das Individuum laktoseintolerant war und braune Augen hatte (KELLER *ET AL.* 2012).



Aus einer 100 Jahre alten Haarsträhne konnte das Genom eines australischen Aborigines sequenziert werden. Die Analyse ausgewählter SNPs dieses Genoms im Vergleich mit SNP-Daten von 1.220 Individuen aus 79 Populationen ergab, dass kein *admixture* mit europäischen Bevölkerungen stattgefunden hat. Die Autoren sehen in den Ergebnissen ihrer Studie weiterhin, dass die australischen Aborigines Nachfahren der ersten Menschen sind, die den australischen Kontinent besiedelten und somit wahrscheinlich eine der ältesten bestehenden Populationen außerhalb Afrikas repräsentieren (RASMUSSEN *ET AL.* 2011).

Die vorgestellten Studien zeigen die Vielfalt an Informationen die anhand genomischer Analysen gewonnen werden kann. Für die statistischen Analysen, die verwendet wurden um genetische Verwandtschaft zwischen verschiedenen Populationen nachzuweisen, wurde nie die gesamte Information der Genome verwendet sondern lediglich eine Auswahl variabler Positionen. Dass für den Vergleich verschiedener Populationen oder Individuen nicht unbedingt die gleichen variablen Positionen untersucht werden müssen, konnte anhand einer Studie 5.000 Jahre alter Überreste von drei Jäger-Sammlern und einem Ackerbauern aus Skandinavien gezeigt werden. Die Sequenzierung dieser Individuen resultierte in insgesamt 249 Millionen Basenpaaren genomischer DNA, wobei die Verteilung der erhaltenen Sequenzdaten pro Individuum jedoch wenige Überschneidungen an variablen Positionen ergab. Die Anwendung einer *Procrustes* Analyse ermöglicht in einem solchen Fall die gemeinsame Analyse unterschiedlicher Daten verschiedener Individuen. Für jedes Individuum wird dazu gemeinsam mit den Referenzpopulationen eine eigene Hauptkomponentenanalyse durchgeführt, die Ergebnisse werden dann durch Translation, Rotation und Skalierung auf Deckung gebracht und können so gemeinsam ausgewertet werden. Die Studie ergab eine große genetische Ähnlichkeit zwischen dem Ackerbauern und Populationen aus Süd-Europa. Im Gegensatz dazu zeigten die Jäger-Sammler eine größere genetische Ähnlichkeit zu Nord-Europäern (SKOGLUND *ET AL.* 2012).

Die palaeogenetische Forschung ermöglicht einen Einblick in die genetische Verwandtschaft und die demographischen Prozesse prähistorischer Populationen, der durch die alleinige Analyse moderner Populationen oder kultureller Hinterlassenschaften nicht gegeben wäre. Sie bietet die Möglichkeit, durch die Archäologie indizierte Bevölkerungszusammenhänge und Migrationshypothesen aufgrund von Ähnlichkeiten in Kultur und Fundgut zu überprüfen.

## A1.2 Die Skythen

Das Bild der eurasischen Steppe ist im 1. Jahrtausend v. Chr. durch Gruppen reiter-nomadischer Stämme geprägt, deren Verbreitungsgebiet sich vom Altai-Gebirge im Osten bis ins Karpatenbecken im Westen erstreckt (PARZINGER 2004). Es sind keine unmittelbaren schriftlichen Hinterlassenschaften dieser Populationen überliefert. Als einzige Zeugnisse existieren eindrucksvolle Grabhügel, sogenannte Kurgane, die sich im gesamten Steppenraum aus der Landschaft erheben, vereinzelte Anzeichen ehemaliger Siedlungen sowie die historischen Berichte antiker Geschichtsschreiber, die Einblick in die kulturellen Gepflogenheiten gewähren und Hinweise auf die Verbreitung dieser nomadisch und seminomadisch lebenden Gruppen geben. Die Skythen – ein Volk, das im 7. Jahrhundert v. Chr. das Gebiet nördlich des Schwarzen Meeres besiedelte – sind, aufgrund der detaillierten Berichte in den *Historien* des Herodot (etwa 480 – 420 v. Chr.), die bekannteste dieser Gruppen. Doch auch Schilderungen über weitere Bevölkerungsgruppen dieser Zeit liegen vor: Sie stellen Sauromaten als die östlichen Nachbarn der Schwarzmeer-Skythen dar (HERODOTUS 4.21); zwischen Kaspischem Meer und Aralsee hingegen erstreckte sich das Verbreitungsgebiet der Massageten und von der Nordgrenze des Iran bis ins kasachische Steppengebiet jenes der Saken. Allerdings verwendeten die Perser den Begriff Saken für das Volk der Skythen im Allgemeinen (HERODOTUS 7.64). Die Sauromaten, die von Hippokrates (460–377 v. Chr.) als ein skythisches Volk beschrieben werden (HIPPOCRATES XVII), werden aus den antiken Schriften auch mit Sarmaten übersetzt, wodurch lange Zeit Uneinigkeit herrschte, ob es sich um das gleiche Volk handelt (MORDVINTSEVA 2013). Meist werden die Sarmaten jedoch als Nachfolger der Sauromaten beschrieben (GRAKOV 1978). Über die Massageten wird berichtet, dass sie die gleiche Kleidung trugen wie die Skythen und auch die gleiche Lebensweise an den Tag legten (HERODOTUS 1.215).

In den historischen Berichten finden sich keine Beschreibungen zu der Situation im heutigen Süd-Sibirien. Herodot erwähnt zwar Beziehungen der Skythen zu ihnen verwandten Gruppen im Osten, doch der Osten wird nicht näher spezifiziert. Erst seit Beginn des 20. Jahrhunderts, nach der Entdeckung der ersten Kurgane in Tuva und dem Altai, deren Fundmaterial, insbesondere Waffen, Pferdegeschirr und eine charakteristische Kunstform, der sogenannte Tierstil, auffällige Ähnlichkeiten mit den Artefakten aus der Nordschwarzmeer-Steppe aufweist, lässt sich diese Verbindung bestätigen. Aufgrund der Ähnlichkeit im Fundmaterial der eisenzeitlichen Populationen aus dem Bereich des Altai-Gebirges zu den Skythen des

Nordschwarzmeerraumes werden die östlichen Gruppen auch als Altai-Skythen bezeichnet (ROLLE 1992) und die Kunstobjekte einem zusammenfassenden skytho-sibirischer Tierstil zugeordnet (SAMASHEV 2002). Die unterschiedlichen Gruppen im Osten und Westen der Eurasischen Steppe der skythischen Kultur zuzuordnen, bzw. unter dem Begriff Skythen zu subsumieren, bleibt für einige Forscher jedoch umstritten, da diese von einer unabhängigen Entwicklung der eisenzeitlichen Populationen ausgehen (YABLONSKY 2000).

Das Ursprungsgebiet der skythischen Kultur konnte bis heute nicht eindeutig lokalisiert werden. Nach den ersten archäologischen Entdeckungen in den Steppen nördlich des Schwarzen Meeres im 18. Jahrhundert, die als skythisch identifiziert werden konnten, wurde davon ausgegangen, dass hier Ursprung und Zentrum ihrer Kultur zu finden seien (YABLONSKY 2000). Bei Ausgrabungen an eisenzeitlichen Monumenten der Transural-Steppe Zentralasiens und Südsibiriens zu Beginn des 20. Jahrhunderts kamen Artefakte ans Tageslicht, die große Ähnlichkeit mit den Funden der Nordschwarzmeer-Region aufwiesen. Terenozhkin und seine Anhänger formulierten daraufhin die Hypothese eines zentralasiatischen Ursprungs der skythischen Kultur (YABLONSKY 2000). Archäologische Indizien legen jedoch nahe, dass der Ausgangspunkt im Osten der eurasischen Steppe zu suchen ist. So gilt der Kurgan Aržan 1, der von Grjaznov in Tuva ausgegraben wurde (GRJAZNOV 1984) als der älteste skythische Kurgan und eine großangelegte Datierung eisenzeitlicher Monumente bestätigte, dass die Kurgane im Westen jüngeren Datums als die Grabhügel im Osten der eurasischen Steppe sind (ALEKSEEV *ET AL.* 2001, ALEKSEEV *ET AL.* 2002). Des Weiteren wurden Artefakte mit Elementen der charakteristischen Tierstil-Kunst im Umkreis des Flusses Jenissei und im Gebiet des heutigen China bereits im 10. Jh. v. Chr. entdeckt (BASHILOV & YABLONSKY 2000, PARZINGER 2004).

Unabhängig von ihrem Ursprungsort lässt sich die Zeit der skythischen Kultur für Ost und West in drei Phasen gliedern: die vor- bzw. initial-skythische Phase vom 9. bis zur Mitte des 7. Jh. v. Chr., eine früh-skythische Phase vom 7. bis zum 6. Jh. v. Chr. und die klassische Skythen-Epoche vom 5. bis zum 4. Jh. v. Chr. (ALEKSEEV *ET AL.* 2001).

### A1.3 Die Palaeopopulationsgenetik der Skythen

Mehrere palaeogenetische Studien haben sich bereits mit der Untersuchung skythischer Reiternomaden beschäftigt. Die Mehrheit dieser Analysen umfassten Individuen der Pazyryk-Kultur, die im 4.–3. Jahrhundert v. Chr. im gesamten Bereich des Altai-Gebirges verbreitet war. Ein Großteil der Bestattungen wurde im Bereich von Permafrostböden entdeckt, wodurch sich die meisten Individuen durch eine hervorragende Erhaltung auszeichnen und besonders gut für palaeogenetische Studien eignen. Auch die Kleidung der Toten, Grabausschmückungen und –beigaben aus organischem Material sind oft noch sehr gut erhalten und können in ihrer ursprünglichen Form rekonstruiert werden. Die ersten Analysen wurden an drei Individuen des russischen Altai-Gebirges durchgeführt (VOEVODA *ET AL.* 2000). Die Ergebnisse dieser Studie legen nahe, dass sich der mitochondriale Genpool der Pazyryk Population aus europäischen und asiatischen Linien zusammensetzt<sup>1</sup>. Die Autoren sehen darin allerdings nicht das Resultat einer Vermischung europäischer und ostasiatischer Populationen, sondern halten eine zentralasiatische Protopopulation, die westeurasische und osteurasische Merkmale in sich vereint, als Ursprung für wahrscheinlicher. Das Vorkommen ost- und westeurasischer Linien konnte von CLISSON *ET AL.* 2002 bestätigt werden, die zwei Individuen aus einem Kurgan des Fundplatzes Berel' untersuchten. Bei dem ersten Individuum handelt es sich nach anthropologischen Untersuchungen um einen 30–50 Jahre alten Mann, das zweite, weniger gut erhaltene Skelett gehört vermutlich zu einer alten Frau – beide wurden gemeinsam mit 13 Pferden in einem Kurgan bestattet. Die Ergebnisse der Ausgrabung haben gezeigt, dass die Frau erst einige Jahre nach dem Mann bestattet wurde. Aufgrund dieses ungewöhnlichen Befundes sollte durch die genetische Studie geklärt werden, ob die beiden Individuen in einem verwandtschaftlichen Verhältnis zueinander standen. Dadurch, dass die Individuen unterschiedliche mitochondriale Linien zeigen, konnte eine Verwandtschaft auf der mütterlichen Seite bereits ausgeschlossen werden. Die Untersuchung nukleärer *loci* sollte klären, ob es sich um Vater und Tochter handeln kann. Obwohl trotz des hervorragenden Erhaltungszustandes der Bestatteten kein vollständiges DNA Profil erhalten wurde, konnte eine enge Verwandtschaft auf der väterlichen Seite ebenfalls ausgeschlossen werden. Im Gegensatz zur Annahme einer Protopopulation, halten die Autoren den Einfluss

---

<sup>1</sup> Eine Erklärung zu der Bezeichnung europäische und asiatische Linien und eine kritische Betrachtung zur Verwendung dieser Begriffe findet sich in Kapitel A4.1 Die mitochondrialen Marker.

weiter entfernter Regionen auf den Genpool der Skythen für plausibel. Sie begründen diese Annahme durch die Entdeckung chinesisch geprägter Objekte in einem untersuchten Kurgan.

Eine Verbindung zu China sehen auch Ricaut und Kollegen in einem von zwei Individuen, die vor ca. 2.500 Jahren im Sebystei-Tal im russischen Teil des Altai-Gebirges beigesetzt wurden. Die mitochondriale Linie des zweiten Individuums interpretieren die Autoren als Teil eines Genpools palaeo-asiatischer Gruppen, da diese heute nur noch in einigen Eskimo Populationen Nordostasiens, Nordamerikas und Grönlands zu finden ist (RICAUT *ET AL.* 2004B). Einen eher europäischen Einfluss sprechen Ricaut und Kollegen einem weiteren Individuum aus dem russischen Teil des Altai-Gebirges zu, das ins 4.–2. Jahrhundert v. Chr. datiert (RICAUT *ET AL.* 2004A). Die mitochondriale Linie dieses Individuums lässt sich der Haplogruppe N1a zuordnen, die in modernen Populationen Zentral-, Nord- und Ostasiens bisher nicht detektiert werden konnte und nur zu geringen Anteilen in Populationen Indiens, des Nahen Ostens und Europas nachgewiesen wurde. Eine Verbindung zu Indien oder zum Nahen Osten anzunehmen, scheint demnach ebenfalls plausibel. Die Funde persisch-archaischer Kunstobjekte in pazyrykzeitlichen Grabhügeln sprechen zumindest für einen kulturellen Einfluss aus diesen Gebieten (POLOSMAK *ET AL.* 2006). Neben Grabhügeln des kasachischen und russischen Bereichs des Altai wurden auch Kurgane aus der Mongolei untersucht. Die genetische Analyse eines Mannes und einer Frau mit Kind aus Olon-Kurin-Gol ergab, dass die Frau und das Kind wahrscheinlich in einem direkten verwandtschaftlichen Verhältnis zueinander standen, da sie die gleiche mitochondriale Linie tragen. Die Linie lässt sich der Haplogruppe HV2 zuordnen, was wiederum ebenfalls als Verbindung zum Iran bzw. Südasien gesehen werden könnte, da dort heute das Hauptverbreitungsgebiet dieser Linie liegt (QUINTANA-MURCI *ET AL.* 2004, ROOTSFORREAL.COM ). Der Mann trägt die Haplogruppe U5a1, die zwar generell den europäischen oder westeurasischen Linien zugeordnet wird, jedoch bereits im Neolithikum in Südsibirien, an der Westküste des Baikalsees gefunden wurde (MOODER *ET AL.* 2006, NASAB 2011). Aufgrund der kulturellen Ähnlichkeiten der Pazyryk Gruppen in den verschiedenen Bereichen des Altai und der Tatsache, dass eine ähnliche U5a1 Linie bereits in Verch Kal'džin 2 auf dem Ukok-Plateau entdeckt wurde (VOEVODA *ET AL.* 2000), gehen die Autoren von der Möglichkeit aus, dass es eine genetische Verwandtschaft zwischen den verschiedenen Pazyryk Populationen geben könnte (PILIPENKO *ET AL.* 2010). Eine etwas umfangreichere Studie zur mongolischen Eisenzeit unterstützt die These einer homogenen Pazyryk Population (GONZÁLEZ-RUIZ *ET AL.* 2012). Untersucht wurden 13 Individuen aus Baga Turgen Gol und drei aus Tsengel Khairkhan von denen die

eine Hälfte westeurasische und die andere Hälfte osteurasische Linien aufweist. Somit wird auch die Vermutung eines zusammengesetzten Genpools der Pazyryk Population aus europäischen und asiatischen Linien bestätigt. Den Ursprung dieser Variabilität sehen die Autoren eher in Migrationsbewegungen europäischer und asiatischer Populationen als im lokalen Phänomen der Protopopulation Zentralasiens. Ihre Annahme stützen sie auf bisherige Ergebnisse zur Bronzezeit in Zentralasien. Neben den 16 Pazyryk Individuen haben sie drei bronzezeitliche Individuen untersucht, die jeweils osteurasische Linien tragen. Ergebnisse aus vorherigen Studien zu bronzezeitlichen Individuen aus Chakassien (KEYSER *ET AL.* 2009) und Kasachstan (LALUEZA-FOX *ET AL.* 2004) zeigen fast ausschließlich westeurasische Linien. Daraus folgern Gonzales-Ruiz und Kollegen, dass es erst zu Beginn der Eisenzeit zu einer Vermischung der Linien auf der jeweils anderen Seite des Altai-Gebirges gekommen ist. Es wurde jedoch mittlerweile in verschiedenen Studien gezeigt, dass osteurasische Linien bereits in der Bronzezeit und im Neolithikum in der westsibirischen Waldsteppe, dem nördlichen Schwarzmeergebiet und in Nordwest Russland verbreitet waren (DER SARKISSIAN *ET AL.* 2013, MOLODIN *ET AL.* 2012A, NIKITIN *ET AL.* 2012). Auch finden sich westeurasische Linien bereits in der Bronzezeit im Tarim Becken in China (LI *ET AL.* 2010). Die zugrundeliegenden populationsgenetischen Dynamiken der Vorläuferkulturen in den verschiedenen Gebieten zu kennen, ist von entscheidender Bedeutung, um den Ursprung der skythischen Populationen zu klären und ihre Zusammensetzung zu verstehen. Verschiedene Arbeiten haben sich mit den Veränderungen von der Bronze zur Eisenzeit im Bereich der Eurasischen Steppe beschäftigt. So wurden in der Studie von Keyser 2009 in Chakassien nicht nur bronzezeitliche Individuen analysiert, sondern speziell Gruppen, von der frühen Bronzezeit bis zur späten Eisenzeit, die sich mit einer Kurgan-Kultur assoziieren lassen. Darunter befinden sich auch zehn Individuen, die von den Autoren der Tagar-Kultur zugeordnet werden und vom 8. Jh. v. Chr. bis zum 1. Jh. n. Chr. datieren. Zeigen die bronzezeitlichen Individuen noch fast ausschließlich europäische Linien (10 von 11), steigt der Anteil asiatischer Linien in der Eisenzeit dagegen auf 40% an. Die Tagar- und die Pazyryk-Kultur weisen also ähnliche Anteile west- und osteurasischer Linien auf (KEYSER *ET AL.* 2009). Eine weitere Studie, die sich jedoch nicht direkt mit der skythischen Kultur assoziieren lässt, beschäftigt sich mit den Veränderungen der Bronze- zur Eisenzeit in Kasachstan (LALUEZA-FOX *ET AL.* 2004). Hier zeigt sich ein ähnliches Bild wie in Chakassien: Keine ostasiatischen Linien in der Bronzezeit im Gegensatz zu einer ungefähr 50:50 Verteilung ost- und westeurasischer Linien in der Eisenzeit. In dem hohen Anteil westeurasischer Linien in der Bronzezeit sehen die Autoren das Resultat einer Expansion europäisch geprägter Bevölkerungen bis hin nach China, auf die

dann ab dem 6. Jahrhundert vor Christus eine westwärts gerichtete Expansion vermutlich aus Sibirien und der Mongolei folgt. Die bisher einzige veröffentlichte aDNA Studie zu westlichen Skythen behandelt 16 Individuen aus dem Rostov-Don Gebiet, die vom 6.–3. Jahrhundert v. Chr. datieren. Die untersuchte Gruppe zeigt zu 69% westeurasische und zu 31% osteurasische Linien (DER SARKISSIAN 2011). In der Ukraine wurden zwar bereits ostasiatische Linien im Neolithikum detektiert, dabei handelte es sich jedoch lediglich um Linien der Haplogruppe C. Die Rostov-Don Gruppe zeigt dagegen fünf verschiedene Linien, die sich den vier Haplogruppen A, C, D und F zuordnen lassen und damit eine wesentlich höhere Vielfalt ostasiatischer Linien im Vergleich zu den vorhergehenden Epochen in diesem Bereich. Aufgrund eines Vergleiches zwischen den Haplotypen der Rostov-Don Individuen und der Haplotypen-Verteilung in modernen Populationen sieht die Autorin in ihren Daten Einflüsse aus Sibirien und dem Zentralasiatischen Korridor womit sie den Irak, Iran, Pakistan und Indien meint.

Der Konsens der genannten Studien liegt in der Feststellung, dass sich die eisenzeitlichen Populationen aus west- und osteurasischen Linien zusammensetzen und das es eine Verbindung zwischen den einzelnen Gruppen zu geben scheint. Die Aussagekraft der bisherigen Studien war jedoch begrenzt, da entweder nur vereinzelte Individuen untersucht wurden (RICAUT *ET AL.* 2004A, RICAUT *ET AL.* 2004B, VOEVODA *ET AL.* 2000) oder weil sie sich nur auf eine bestimmte geographische Region beschränkten (GONZÁLEZ-RUIZ *ET AL.* 2012, KEYSER *ET AL.* 2009), was wiederum auf einen Mangel an Vergleichsdaten zurückzuführen ist.

## A2 Die populationshistorische Fragestellung

Wie bereits im vorherigen Kapitel beschrieben, waren bisherige Studien zu skythenzeitlichen Populationen entweder auf eine geringe Individuenanzahl, oder eine geographische Region beschränkt. Übergeordnete Vergleiche waren selbst bei Studien mit einer größeren Probenanzahl aufgrund mangelnden Vergleichsmaterials nur eingeschränkt möglich. Weiterhin beschränken sich die Studien im Osten Zentralasiens auf das 4.–3. Jahrhundert v. Chr. und umfassen nur Individuen der Pazyryk-Kultur. Ziel dieser Arbeit war es, durch die Anwendung effizienterer DNA-Anreicherungsverfahren und den Einsatz neuer Sequenziertechnologien den bestehenden Datensatz zu skythischen Populationen zu erweitern. Bei der Probenauswahl wurde zum einen darauf geachtet, dass die Fundkomplexe mit der skythischen Kultur in Verbindung gebracht werden können und zum anderen, dass eine möglichst große Zeitspanne und ein möglichst großes Verbreitungsgebiet abgedeckt wird, um chronologische Veränderungen und räumliche Unterschiede nachvollziehen zu können. Mit dem erweiterten Datensatz sollen die populationsgenetischen Zusammenhänge der eisenzeitlichen Gruppen des eurasischen Steppengürtels genauer untersucht und auf der Grundlage von archäologischen Forschungen bislang angenommene Beziehungen eingehender überprüft werden.

Im Speziellen wurden folgende Gesichtspunkte analysiert:

### 1) Die diachrone Bevölkerungsentwicklung im Bereich des Altai-Gebirges

Durch die Analyse der Individuen aus dem Bereich des Altai-Gebirges, die sich zeitlich verschiedenen Kulturstufen zuordnen lassen, soll geklärt werden, ob es in dieser Region eine kontinuierliche Bevölkerungsentwicklung im Verlauf des 1. Jahrtausend vor Christus gab und gleichzeitig, ob die unterschiedlichen Kulturstufen aus genetisch unterschiedlichen Populationen hervorgegangen sind oder ob eine gemeinsame Ursprungspopulation zugrunde liegt.

### 2) Die Struktur der Pazyryk-Kultur

Die Populationen des 4.–3. Jahrhunderts v. Chr. sind eine gesonderte Betrachtung wert, denn die zahlreichen Funde aus dem russischen, kasachischen und mongolischen Teil des Altai-Gebirges werden alle der Pazyryk-Kultur zugeschrieben. Das umfangreiche Probenmaterial ermöglicht die Überprüfung auf eine Populationsstruktur kulturell homogener Funde unterschiedlicher geographischer Lage.



### 3) Das Verhältnis der Skythen zu den Sarmaten

Weiterhin soll das Verhältnis der Skythen aus dem Nordschwarzmeerraum und der frühen Sarmaten aus Pokrovka zueinander geprüft werden. So finden sich auf der einen Seite charakteristische Elemente des skythischen Tierstils in den früh-sarmatischen Kurganen aus Pokrovka (MALASHEV & YABLONSKY 2004), die eine Verbindung zu den Schwarzmeerskythen nahelegen – auf der anderen Seite wird von einer Verdrängung der Skythen durch sarmatische Verbände im 2. Jahrhundert v. Chr. ausgegangen (SMIRNOV 1980). Von daher stellt sich die Frage, ob sich die Sarmaten genetisch von den Skythen unterscheiden und somit tatsächlich von einem Bevölkerungswechsel im nordpontischen Raum ausgegangen werden kann.

### 4) Das Verhältnis zwischen westlichen und östlichen skythischen Populationen

Die gemeinsame Betrachtung der westlichen Populationen mit skythischer Sachkultur und deren Bezug zu den Populationen des Altai-Gebirges soll zur Klärung der übergeordneten Frage beitragen, ob eine generelle Verbindung zwischen den reiternomadischen Stämmen des 1. Jahrtausends v. Chr. besteht, die deren Zusammenfassung unter dem Begriff Skythen rechtfertigt.

### 5) Das Ursprungsgebiet der Skythen

Ein weiterer Punkt ist die Frage nach dem Ursprung der skythischen Populationen. Zur Überprüfung der unterschiedlichen Hypothesen, wurden die eisenzeitlichen Gruppen mit verfügbaren Daten aus dem Neolithikum und der Bronzezeit der westlichen und östlichen Eurasischen Steppe und der westsibirischen Waldsteppe verglichen.

### 6) Die phänotypischen Charakteristika der Skythen

Es wurden zusätzlich nukleäre Marker untersucht, die mit einer phänotypischen Ausprägung assoziiert sind und von denen vermutet wird, dass sie entweder in europäischen oder asiatischen Populationen unter positiver Selektion stehen. Die Analyse ermöglicht einen Einblick in das Erscheinungsbild der untersuchten Individuen und kann weiterhin Hinweise auf den geographischen Ursprung der Skythen geben.

## A3 Das Untersuchungsmaterial

Für die vorliegende Arbeit wurden 126 Proben von 22 Fundplätzen ausgewählt. Die Datierungen reichten vom 9. bis zum 3. Jh. v. Chr. und die Fundplätze erstreckten sich vom Minusinsker Becken über Tuva, das Altai Gebirge und Kasachstan bis nach West-Russland zum Bereich nördlich des Schwarzen Meeres. Die Individuen des Fundplatzes Pokrovka wurden teilweise bereits für eine frühere Studie in Mainz analysiert.

**Tab. A 1 Probenliste nach Fundplatz und Region.**

Koordinaten sind ungefähre Angaben. Lab\_ID bezeichnet die interne Laborbenennung der Probe; K=Kurgan; Ind.=Individuum.

Region	Fundplatz	Koordinaten	Datierung	Kultur	Lab_ID	Kurgan	Grab	Ind.
Ostkasachstan	Tar Asu	49°17'N 86°19'O	4.–3.Jh.v.Chr.	Pazyryk	Ta1	23		
Ostkasachstan	Berel'	49°20'N 86°12'O	4.–3.Jh.v.Chr.	Pazyryk	Be1	16		
					Be2	16		
					Be3	16		
					Be4	32		
					Be6	10		
					Be8	34		
					Be9	73	20	
					Be11	31		
					Be12	36		
	Be14	71						
Ostkasachstan	Ismailovo	50°18'N 81°26'O	9.–7.Jh.v.Chr.	Zevakino-Čilikta	Is1	6		
					Is2	10	1	
					Is4	27		
Ostkasachstan	Zevakino	50°12'N 81°49'O	9.–7.Jh.v.Chr.	Zevakino-	Ze1	25		
					Ze2	99b		
					Ze3	224		
					Ze4	10	1	
					Ze5	10	2	
					Ze6	46	1	
					Ze7	33		
					Ze8	10	4	
					Ze9	336		
Ostkasachstan	Majemir	49°10'N 84°57'O	unbekannt	unbekannt	MaI_1			
					MaII_2	1		
					MaII_3	2		
Südkasachstan	Cirikrabort	44°05'N 62°54'O	4.–2.Jh.v.Chr.	Saken	CR1	1		1
					CR2	1		3
					CR3	1		4
Südsibirien Tuva	Aržan 2	52°05'N 93°42'O	7.–6.Jh.v.Chr.	Aldy Bel	A1	2	2	
					A2	2	5	1
					A3	2	5	2
					A4	2	7	

Region	Fundplatz	Koordinaten	Datierung	Kultur	Lab_ID	Kurgan	Grab	Ind.
Südsibirien Tuva	Aržan 2	52°05'N 93°42'O	7.–6.Jh.v.Chr.	Aldy Bel	A5	2	8	
					A6	2	12	
					A7	2	13a	1
					A8	2	13a	2
					A9	2	13b	
					A10	2	14	1
					A11	2	14	2
					A12	2	15	
					A13	2	18	
					A14	2	20	1
					A15	2	20	2
					A17	2	22	
					A18	2	23	
					A19	2	24	
A20	2	25						
A21	2	26						
A22	2	29						
A23	2	ADE	1					
A24	2	GNDK						
Südsibirien russisches Altai-Gebirge	Ak-Alacha 1	49°17'N 87°32'O	4.–3.Jh.v.Chr.	Pazyryk	Ak1_1	1	1	
					Ak1_2	1	1	
	Ak-Alacha 4	49°17'N 87°32'O	4.–3.Jh.v.Chr.	Pazyryk	Ak4_1	1		
					Ak-Alacha 5	49°17'N 87°32'O	4.–3.Jh.v.Chr.	Pazyryk
	Ak5_2	1						
	Ak5_3	3	1					
	Ak5_4	3	1					
	Ak5_5	3	2					
Ak5_6	4	1						
Ak5_7	4	2						
Ak5_8	5							
Südsibirien russisches Altai-Gebirge	Alagail 2	50°12'N 87°42'O	4.–3.Jh.v.Chr.	Pazyryk	Ala1	8	12	
					Ala2	8	15	
					Ala4	10		
	Barburgazy 1	49°49'N 89°08'O	4.–3.Jh.v.Chr.	Pazyryk	B1_1	4(A)		
					B1_2	7	1	
	Barburgazy 3	49°49'N 89°08'O	4.–3.Jh.v.Chr.	Pazyryk	B3_1	5	2	
	Balik-Sook	50°48'N 86°00'O	4.–3.Jh.v.Chr.	Pazyryk	BS1	1		
					BS2	27		
	Borotal 2	50°12'N 87°42'O	4.–3.Jh.v.Chr.	Pazyryk	Bt1	2		
					Bt2	3	2	
Dcholin 2	49°48'N 89°22'O	4.–3.Jh.v.Chr.	Pazyryk	D1	1			
Südsibirien russisches Altai-Gebirge	Justyd 12	49°46'N 89°18'O	4.–3.Jh.v.Chr.	Pazyryk	J12_1	1		
					J12_2	3A		
					J12_3	6		
					J12_6	10		
					J12_7	17	2	
					J12_8	17	3	
	J12_9	18						
	Justyd 22	49°46'N 89°18'O	4.–3.Jh.v.Chr.	Pazyryk	J22_1	1		
	Kuturguntas 1	49°25'N 87°35'O	4.–3.Jh.v.Chr.	Pazyryk	K1	1		
					K2	1		

Untersuchungsmaterial A

Region	Fundplatz	Koordinaten	Datierung	Kultur	Lab_ID	Kurgan	Grab	Ind.
Südsibirien russisches Altai-Gebirge	Moinak 2	49°19'N 87°35'O	4.–3.Jh.v.Chr.	Pazyryk	Mo1	1	1	
					Mo2	2	1	
	Ulandryk 1	49°41'N 89°06'O	4.–3.Jh.v.Chr.	Pazyryk	U1_1	9		
					U1_2	11		
	Ulandryk 2	49°41'N 89°06'O	4.–3.Jh.v.Chr.	Pazyryk	U2_1	1		
					U2_2	2		
Ulandryk 4	49°41'N 89°06'O	4.–3.Jh.v.Chr.	Pazyryk	U4_1	2			
				U4_4	3			
Verch Kal'džin 1	49°23'N 87°34'O	4.–3.Jh.v.Chr.	Pazyryk	VK1_1	1			
Minusinsker Becken	Barsučij Log	54°00'N 91°11'O	5.Jh.v.Chr.	Tagar	BL1		1	1
					BL2		1	2
					BL3		1	3
					BL4		1	4
					BL5		1	5
					BL6		2	1
SW-Russland Voronezh Oblast	Kolbino	51°37'N 39°11'O	3.Jh.v.Chr.	Skythen	Kol1	8	1	
					Kol2	10		
					Kol3	10_I		
					Kol4	44	2	
					Kol5	44	4	
SW-Russland Nord- Kaukasus	Novozavedennoe 2	44°16'N 43°38'O	8.–6.Jh.v.Chr.	Initiale Skythen	Nov1	10		
					Nov2	13	2	
					Nov3	1		
					Nov4	9	2	
					Nov5	3	1	
					Nov6	9	1	
					Nov7	2		
					Nov8	6		
					Nov9	12		
					Nov10	14		
SW-Russland Orenburg Oblast	Pokrovka	52°04'N 53°56'O	5.–2.Jh.v.Chr.	Frühe Sarmaten	PR1	2		
					PR2	2		
					PR3	2		
					PR4	2		
					PR5	2		
					PR6	2		
					PR7	2		
					PR8	2		
					PR9	2		
					PR10	2		
					PR11	2		
					PR13	2		

## A4 Methodenüberblick

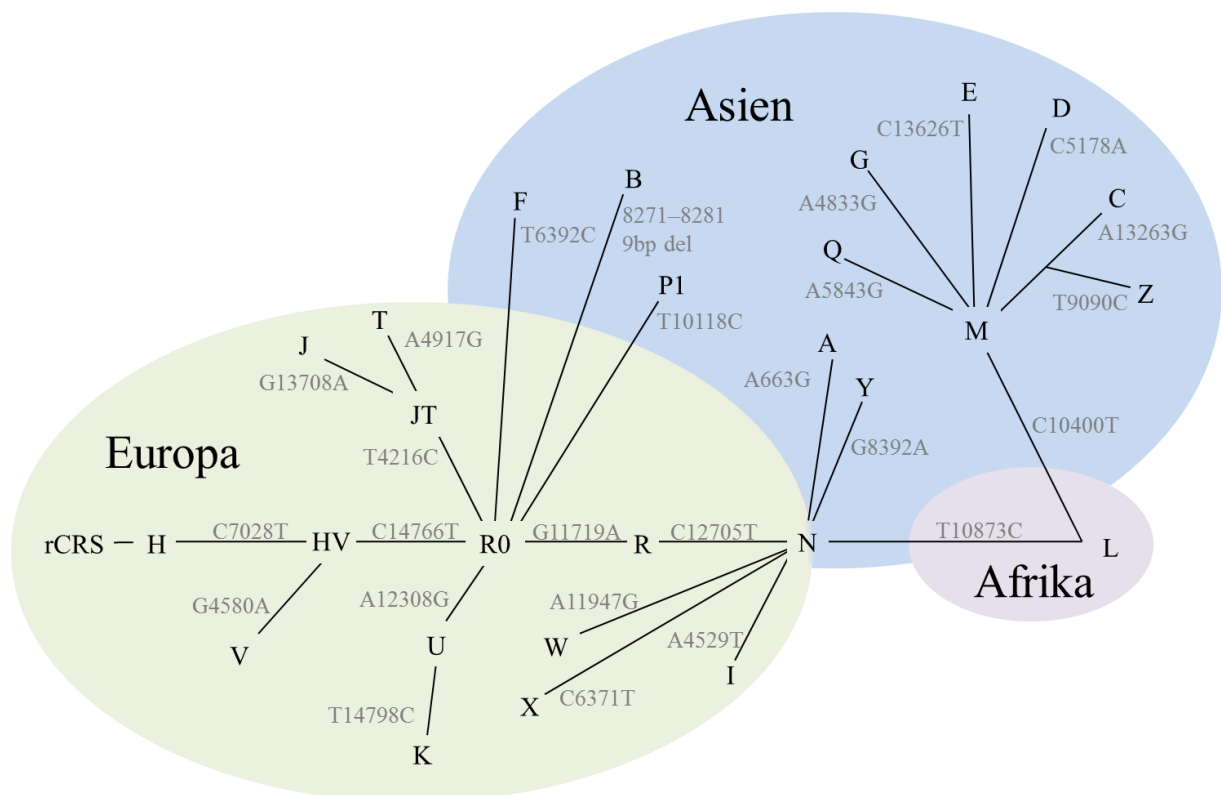
Eine detaillierte Beschreibung der verwendeten Methoden findet sich in Teil B

### A4.1 Die untersuchten Marker

#### Die mitochondrialen Marker

Die menschliche Erbinformation ist in Form von 22 Autosomen und zwei Geschlechtschromosomen im Zellkern gespeichert. Neben diesem sogenannten Kerngenom findet sich ein weiterer Informationsspeicher in den Mitochondrien, den Energieproduzenten der Zelle. Das menschliche mitochondriale Genom besteht aus einem ringförmigen DNA Molekül von 16.569 Basenpaaren (bp). Im Gegensatz zum Kerngenom, welches pro Zelle in einem diploiden Chromosomensatz vorkommt, befinden sich in jedem Mitochondrium 2–10 DNA-Moleküle, und jede Zelle besitzt mehrere tausend Mitochondrien. Aufgrund ihrer hohen Kopienzahl ist die mitochondriale DNA (mtDNA) besonders prädestiniert für palaeogenetische Studien, da die DNA in archäologischem Zahn- oder Skelettmaterial je nach Alter und Lagerung bereits mehr oder weniger degradiert vorliegt. Die mtDNA wird rein maternal vererbt, anhand ihrer kann also die mütterliche Vererbungslinie nachvollzogen werden. Mitochondrialen Linien, oder Haplotypen, definieren sich anhand spezifischer SNPs (*single nucleotide polymorphisms*), womit im vorliegenden Fall Basensubstitutionen relativ zur verwendeten Referenzsequenz (hier: *revised Cambridge Referenz Sequence – rCRS*; ANDREWS ET AL. 1999; *Accession number: NC\_012920*;) gemeint sind. Phylogenetisch nah verwandte Haplotypen werden zu übergeordneten Haplogruppen zusammengefasst. Für die vorliegende Arbeit wurden 37 Fragmente des mitochondrialen Genoms ausgewählt, die Haplogruppen, bzw. Haplotypen spezifische SNPs tragen. Die SNPs verteilen sich auf den kodierenden und den nichtkodierenden Bereich des mitochondrialen Genoms, also den Bereich der Gene trägt und für Proteine kodiert und den Bereich der keine Gene trägt und dessen Basenabfolge keine Aminosäuren und somit Proteine kodiert. Die Mutationen des kodierenden Bereiches erlauben eher eine grobe Eingliederung in die Haupthaplogruppen und die Mutationen des nichtkodierenden Bereiches, in vorliegendem Fall der Hypervariablen Region 1 (HVR1), ermöglichen eine feinere Unterteilung in Subhaplogruppen.

Seit Beginn der neunziger Jahre wird mitochondriale DNA systematisch für populationsgenetische Untersuchungen eingesetzt, wofür neben RFLPs (*restriction fragment length polymorphism*) meist die HVR1 verwendet wurde (COMAS *ET AL.* 1998, DERENKO *ET AL.* 2007, HERRNSTADT *ET AL.* 2002, HORAI *ET AL.* 1993, TORRONI *ET AL.* 1992, TORRONI *ET AL.* 1993, TORRONI *ET AL.* 1998, TORRONI *ET AL.* 2000). Aufgrund ihrer erhöhten Mutationsrate im Vergleich zum restlichen Genom eignet sie sich hervorragend auch relativ rezente Evolutionsereignisse zu untersuchen (INGMAN *ET AL.* 2000, PARSONS *ET AL.* 1997). Bis dato wurden so enorme Mengen an mitochondrialen Daten gesammelt und es zeigt sich, dass die Verteilung der Haplogruppen einer geographischen Struktur folgt (s. Abb. A1), die sich auf die Ausbreitung des anatomisch modernen Menschen aus Afrika zurückführen lässt (CANN *ET AL.* 1987). So findet sich die Haplogruppe L hauptsächlich in Afrika, die Haplogruppen A, B, C, D, E, F, G, M finden sich in asiatischen und H, HV, I, J, K, T, U, W, X und V in europäischen Populationen wieder (QUINTANA-MURCI *ET AL.* 2004).



**Abb. A 1 Geographische Verteilung der Haplogruppen in modernen Bevölkerungen.**

Für jede Haplogruppe ist die Mutation angegeben, die zu ihrer Detektion untersucht wurde, wobei vor der entsprechenden Position die Base der Referenzsequenz steht und danach die anzeigende Mutation (rCRS als Ursprung für die Leserichtung) (Quelle: modifiziert nach MITOMAP).

Bei der geographischen Einteilung der Haplogruppen ist zu beachten, dass die Bezeichnung gewisser Linien als europäisch oder asiatisch darauf gründet, dass in Zentraleuropa ausschließlich europäische Linien detektiert wurden und in Asien, oder genauer in Ostasien nur asiatische. Dabei gilt zu bedenken, dass es sich hierbei um die Haplogruppenverteilung in modernen Populationen handelt und keine Aussage getroffen werden kann, wie die Variation in diesen Gebieten in vorhergehenden Epochen aussah. Weiterhin existieren weiträumige Überschneidungsbereiche, in denen die Haplogruppen sich entlang eines Gradienten ausdünnen bzw. zunehmen. Um die Grenzen der Verbreitungsgebiete nicht ganz so scharf zu fassen, werden in vielen Publikationen statt europäische und asiatische Linien, die Begriffe west- und osteurasisch verwendet. Unterschiede gibt es auch in der Zuteilung mitochondrialer Linien deren Hauptverbreitungsgebiet Südasien ist, womit meist Indien, Pakistan und der Iran gemeint sind. Einige Autoren bilden hier eine eigenständige Gruppe südasiatischer Linien, viele teilen die betroffenen Haplotypen ihren Haupthaplogruppen entsprechend in die beiden Großregionen Ost und West auf. Aufgrund der heutigen Haplogruppenverteilung in den Bereichen Zentralasiens (COMAS *ET AL.* 1998, COMAS *ET AL.* 2004, QUINTANA-MURCI *ET AL.* 2004) wurde angenommen, dass die skythischen Individuen europäische und ostasiatische Linien tragen können, was auch bereits in früheren Studien zu eisenzeitlichen Populationen im Bereich des Altai-Gebirges gezeigt wurde (CLISSON *ET AL.* 2002, VOEVODA *ET AL.* 2000). Bei der Auswahl der zu untersuchenden SNPs wurde deshalb darauf geachtet, dass die gesamte Variabilität Zentralasiens erfasst werden kann.

## Die nukleären Marker

Neben der mitochondrialen DNA wurden zudem 21 Fragmente des Kerngenoms untersucht. Die Zusammenstellung der Marker und das Analyseverfahren wurden von S. W. im Zuge ihres Teilprojektes vorgenommen. Die ausgewählten Fragmente enthalten Mutationen, bei denen ein starker Zusammenhang mit einer bestimmten phänotypischen Eigenschaft festgestellt wurde und von denen vermutet wird, dass sie in einer oder mehreren Populationen unter Selektion stehen. Für die vorliegende Studie sind diejenigen Marker von besonderem Interesse, die entweder in Asien oder Europa selektiert sind. Die Allelfrequenzen entsprechender *loci* unterscheiden sich stark in den beiden Populationen wodurch zugeteilt werden kann, ob die in dieser Studie untersuchten Individuen eher die europäische oder die asiatische Ausprägung teilen. Zusätzlich wird es durch die Korrelation von Genotyp und Phänotyp möglich, sich eine Vorstellung vom Erscheinungsbild prähistorischer Populationen zu machen. Die bisher stärkste Korrelation zwischen Genotyp und Augenfarbe wurde für eine Mutation im Intron 86 des *HERC2* Gens nachgewiesen. Individuen, die an dieser Stelle den A/A-Genotyp aufweisen, haben mit einer 80%igen Wahrscheinlichkeit braune Augen. Im Gegenzug weisen Träger des G/G-Genotyps nur mit einer Wahrscheinlichkeit von 1% braune Augen auf (STURM *ET AL.* 2008). In den meisten Fällen sind mehrere Gene, bzw. mehrere Mutationen, für die Ausprägung eines bestimmten Merkmals verantwortlich, oder der Zusammenhang zwischen Genotyp und Phänotyp trifft nur für eine bestimmte Population zu. So wurde klar nachgewiesen, dass eine Mutation des Laktase Gens positiv mit der Fähigkeit korreliert auch im Erwachsenenalter Milch verdauen zu können (ITAN *ET AL.* 2010) – die Milchverträglichkeit in Afrikanern und Europäern wird dabei jedoch von unterschiedlichen Mutationen beeinflusst (MULCARE *ET AL.* 2004). Auch scheint es so, dass die hellere Hautfarbe in Europäern und Asiaten auf unterschiedliche Mutationen zurückzuführen ist (NORTON *ET AL.* 2007). Die hier untersuchten Marker können entsprechend ihrer Funktion in verschiedenen Gruppen eingeteilt werden: Pigmentierung, Calcium Aufnahme und Vitamin D Synthese, Regulierung von Körpertemperatur und Wasserhaushalt und die Verarbeitung körperfremder Substanzen, sog. Xenobiotika. Zusätzlich wurde der Amelogenin *locus* untersucht, der in unterschiedlich langen Varianten auf dem X und dem Y-Chromosom vorkommt, wodurch er sich zur genetischen Determination des Geschlechts eignet (s. Tab. A2). Die rs-Nummern und die chromosomale Position der untersuchten Marker sind in Tabelle B4 aufgeführt.



**Tab. A 2 phänotypische Eigenschaften der untersuchten nukleären loci**

Fett hervorgehoben sind die Marker, deren Frequenzen sich zwischen Europäern und Ostasiaten unterscheiden und von denen Selektion in einer der beiden Gruppen angenommen wird.

	Name	SNP	Phänotyp / Selektionsfaktor (hypothetisch)	Info/Referenz
Pigmentierung	<b>HERC2</b>	A>G	homozygot GG = zu 99% blaue, graue oder grüne Augenfarbe	starke Korrelation zwischen Genotyp und Phänotyp (STURM <i>ET AL.</i> 2008); aDNA Daten für Chakassien: BOUAKAZE <i>ET AL.</i> 2009: gesamt Bronze - Eisenzeit: 65,2% helle Augen, 34,8% braune Augen; im Vergleich dazu: modern Zentralasien: DONNELLY <i>ET AL.</i> 2012: 17% G-Allel (helle Augen) in Kasachen (Ethnie), 0-4% in Mongolen, bis zu 42% in den westsibirischen Khanty, 9% in Yakut; moderne HCB: 100% braune Augen
	<b>SLC24A5</b>	G>A	hellere Hautpigmentierung	Für alle drei Gene wurde eine Korrelation mit der Pigmentierung der Haut nachgewiesen (LAMASON <i>ET AL.</i> 2005 für <i>SLC24A5</i> , GRAF <i>ET AL.</i> 2005 für <i>SLC45A2</i> und SHRIVER <i>ET AL.</i> 2003 für <i>TYR</i> ); Alle drei haben großen Anteil an der natürlichen Variation der Hautpigmentierung in Südasien (STOKOWSKI <i>ET AL.</i> 2007); <i>SLC24A5</i> und <i>TYR</i> spielen eine wichtige Rolle bei der Evolution hellerer Haut in Europäern aber nicht in Ostasiaten; die relativ hohe Frequenz des abgeleiteten A-Allels von <i>SLC24A5</i> in Zentralasien ist wahrscheinlich auf Genfluss mit europäischen Populationen zurückzuführen (NORTON <i>ET AL.</i> 2007).
	<b>SLC45A2 (MATP)</b>	C>G	hellere Pigmentierung	
	<b>TYR</b>	C>A	hellere Pigmentierung	
Calcium Aufnahme und Vitamin D Synthese	<b>LCTa (MCM6, -13910T)</b>	C>T	Laktasepersistenz / Milchwirtschaft	
	<b>LCTb (MCM6, -22018A)</b>	G>A (C>T)	Laktasepersistenz / Milchwirtschaft	
	CYP3A4	G>A	verringerte Genexpression / Schutz vor Rachitis	Außerhalb Afrikas selektiert, Frequenz korreliert mit dem Breitengrad (THOMPSON <i>ET AL.</i> 2004); Da <i>CYP3A4</i> entscheidend an der Vitamin D-Synthese beteiligt ist wird der Schutz vor Rachitis als Selektionsfaktor vermutet (SCHIRMER <i>ET AL.</i> 2006).
	<b>SLC12A3</b>	G>C	niedrige Enzymaktivität / Stärkung des Calciumhaushalts	Der Test auf erweiterte Haplotypen Homozygotität (EHH) lässt auf Selektion des C-Allels schließen, welches allerdings nur in geringer Frequenz und nur in Europäern vorkommt. Die niedrigere Sonneneinstrahlung wird hier als Selektionsfaktor angenommen. (VORMFELDE <i>ET AL.</i> 2007).
	TRPV6b	G>A	ungeklärt -- Immunsystem?	Die abgeleiteten Allele des <i>TRPV6 locus</i> bilden außerhalb Afrikas einen selektiven Vorteil (AKEY <i>ET AL.</i> 2006); kein einzelnes Selektionsereignis sondern parallele Selektion in verschiedenen Populationen, eventuell im Zusammenhang mit der Einführung von Agrikultur (HUGHES <i>ET AL.</i> 2008).
	TRPV6c	C>T	ungeklärt -- Immunsystem?	

	Name	SNP	Phänotyp / Selektionsfaktor (hypothetisch)	Info/Referenz
Temperatur und Wasser	<b>EDAR</b>	T>C	C = dickeres Haar /Schutz vor Kälte	selektiert in asiatischen Populationen (FUJIMOTO <i>ET AL.</i> 2008)
	<b>ABCC11</b>	G>A (C>T)	AA= Funktionsverlust des ABCC11 Proteins → trockenes Cerumen, weniger Achselschweiß / Kälteanpassung, geruchsarmer Achselschweiß	Zusammenhang zwischen Allelfrequenz und Breitengrad, weshalb Klimaanpassung vermutet wird (OHASHI <i>ET AL.</i> 2011, YOSHIURA <i>ET AL.</i> 2006).
	CYP3A5	A>G	erhöhter Anteil an nicht-funktionellem Enzym / Anpassung an höhere Wasservorkommen	KUEHL <i>ET AL.</i> 2001; Das funktionelle Enzym führt zu einer vermehrten Salz- und somit Wasserrückführung in der Niere, was vorteilhaft in trockenen Gebieten ist. Es wird vermutet, dass ein Zusammenhang zwischen dem aktiven Enzym und Bluthochdruck besteht. Die Frequenz des nicht-funktionellen Enzyms steigt mit der Entfernung zum Äquator. Es wird angenommen, dass die Wasserrückführung in Gebieten mit erhöhtem Wasservorkommen irrelevant wird und der negative Effekt Bluthochdruck zur Selektion des nicht-funktionellen Enzyms beiträgt. Zumindest sprechen die extremen Frequenzunterschiede zwischen Populationen für einen populationspezifischen Selektionsdruck (THOMPSON <i>ET AL.</i> 2004).
Xenobiotika	<b>ADH1Ba</b>	C>T	erhöhte Expression / Schutz vor Alkoholismus	Eine erhöhte Enzymmenge bei gleichzeitiger Aktivitätssteigerung führt zu einer Akkumulation von Acetaldehyd, was Symptome wie Erröten des Gesichts, Herzklopfen, Schwindel und Übelkeit hervorruft. Es wird vermutet, dass diese negativen Effekte zu einer Aversion gegen übermäßigen Alkoholgenuß führen und die Träger dieser Mutationen deshalb weniger suchtanfällig sind. (LI <i>ET AL.</i> 2008, WALL <i>ET AL.</i> 2005). Selektiert in ostasiatischen Populationen; vermuteter Zusammenhang zwischen Selektion und Reis-Domestikation bzw. der Verarbeitung durch Fermentation (PENG <i>ET AL.</i> 2010)
	<b>ADH1Bb</b>	G>A	erhöhte Enzymaktivität -> Beschleunigung des ersten Schrittes beim Alkoholabbau -> Akkumulation von Acetaldehyd / Schutz vor Alkoholismus	
	NAT2 (NAT2*5)	T>C	langsam arbeitendes Enzym / Ernährungsumstellung durch Agrikultur	höhere Frequenz in ackerbaubetreibenden Gesellschaften im Vergleich zu Jäger-Sammlern (SABBAGH <i>ET AL.</i> 2011)
	ABCB1a	C>T	erhöhter Schutz vor xenobiotischen Angriffen an der Blut-Hirn-Schranke / Adaption an neue Umwelt	Adaption an körperfremde Stoffe aus der Umwelt (WANG <i>ET AL.</i> 2007)
	ABCB1b	G>T		
	ABCB1c	C>T		
	CASP12	C>T	nicht-funktionelles Protein / geringeres Sepsisrisiko	SALEH <i>ET AL.</i> 2004 ; Beginn der Selektion mit steigender Populationsgröße (XUE <i>ET AL.</i> 2006)
♀♂	AMEL	6bp Indel	Der Amelogenin <i>locus</i> ist auf dem Y-Chromosom 112bp lang und auf dem X-Chromosom nur 106bp. (MANNUCCI <i>ET AL.</i> 1994)	

Bei Untersuchungen der nukleären DNA, ist die Diploidie des Chromosomensatzes zu beachten. Eines der Chromosomen wurde vom Vater und eines von der Mutter geerbt, was dazu führt, dass die zu untersuchenden Marker je zweimal existieren und daher in verschiedenen Ausprägungen (Allelen) vorliegen können.

## A4.2 Die selektive Anreicherung der Marker aus alter DNA

Es wurden zwei verschiedene Methoden angewendet, um die ausgewählten DNA-Fragmente soweit zu vervielfältigen, dass eine Untersuchung ihrer Basenabfolge möglich wird: Die erste Methode folgt dem klassischen Ansatz der Polymerase-Kettenreaktion (PCR – *polymerase chain reaction*). Hierbei wurden ausgewählte Positionen des mitochondrialen Genoms und des Kerngenoms mithilfe sogenannter Multiplex-PCR-Ansätze amplifiziert. Der Multiplex-Ansatz erlaubt die gleichzeitige Amplifikation mehrerer *loci*, ohne dass zusätzliches DNA-Extrakt eingesetzt werden muss. So wurden die 21 nukleären *loci* in einem Multiplex-Ansatz und die 37 Fragmente des mitochondrialen Genoms, aufgeteilt auf drei Ansätze, jeweils simultan vervielfältigt.

Die zweite Methode beruht auf einer relativen Anreicherung der gesuchten *loci*, indem diese von dem DNA-Extrakt separiert werden. Hierzu werden vorab sogenannte *sequence baits* entworfen, einzelsträngige RNA-Fragmente, deren Basenabfolge komplementär zu den gesuchten *loci* ist und die zusätzlich eine Biotin-Markierung tragen. Die *baits* werden mit dem DNA-Extrakt gemischt und hybridisieren an die ihnen komplementären, gewünschten DNA-Fragmente. Zu diesem Hybridisierungsansatz werden dann magnetische Kügelchen (*magnetic beads*) gegeben, die einen Streptavidin-Überzug tragen, an den das Biotin der *baits* gebunden wird. So können, mit Hilfe eines Magneten, die gewünschten Fragmente aus einem DNA Gemisch herausgezogen werden. Im zweiten Schritt werden auch diese *loci* durch eine PCR vervielfältigt. Diese Methode wird als *Capture*, oder auch *Hybrid Capture* bezeichnet und wurde angewendet, um ganze mitochondriale Genome zu erlangen.

### A4.3 DNA-Sequenzierung und Genotypisierung

Zur Bestimmung der Basenabfolge der vervielfältigten DNA-Fragmente wurden verschiedene Sequenzieretechniken angewandt, zum einen die Sanger-Sequenzierung und weiterhin zwei NGS-Methoden (*Next Generation Sequencing*). Der grundlegende Unterschied zwischen der Sanger- und den NGS-Methoden liegt in der Menge an DNA-Fragmenten, die gleichzeitig sequenziert werden kann. Bei beiden NGS-Methoden erfolgt die Sequenzierung außerdem klonal, also jedes Startmolekül wird separat erfasst, wohingegen das Ergebnis der Sanger-Sequenzierung eine Konsensussequenz aller Moleküle der Reaktion darstellt und für die Erfassung der Einzelmoleküle eine bakterielle Klonierung vorgeschoben werden muss.

Die durch die verschiedenen Sequenziermethoden ermittelten DNA-Sequenzen des mitochondrialen Genoms wurden mit der rCRS (ANDREWS *ET AL.* 1999; *Accession number: NC\_012920*) verglichen und die Unterschiede (SNPs) zu dieser notiert. Anhand der Unterschiede zur Referenzsequenz wurde mit Hilfe des Programms *HaploGrep* (KLOSS-BRANDSTÄTTER *ET AL.* 2011), jedem Individuum die entsprechende Haplogruppe zugeordnet. Die Haplogruppenzuordnung des Programms basiert auf dem *PhyloTree (Built 15)*, einem phylogenetischen Baum, der die globale Variation mitochondrialer DNA erfasst und ständig aktualisiert wird (VAN OVEN & KAYSER 2009).

Für die Sequenzauswertung der nukleären *loci* ist zu beachten, dass in der Regel je drei mögliche Ausprägungen existieren: die homozygot aneztrale Form (Bsp.: AA), die heterozygote Form (AB) und die homozygot abgeleitete Form (BB). So wurde zuerst für jedes Individuum und jeden phänotypischen Marker der vorliegende Genotyp bestimmt und anschließend die Frequenzen für die aneztralen und die abgeleiteten – hier die selektierten – Allele berechnet. Die für die skythischen Populationen ermittelten Allelhäufigkeiten der unterschiedlichen Marker wurden mit Daten der *Allele Frequency Database ALFRED* (KIDD 2012, RAJEEVAN *ET AL.* 2012) und der NCBI SNP-Datenbank (DBSNP, SHERRY *ET AL.* 2001) abgeglichen. Die NCBI SNP-Datenbank stellt Daten aus verschiedenen Projekten zur Verfügung, darunter Allelfrequenzen aus dem *HapMap Project* (HAPMAPCONSORTIUM 2003), dem *Environmental Genome Project* (EGP) (NIEHS\_SNPs) und dem *1000 Genomes Project* (THE1000GENOMESCONSORTIUM *ET AL.* 2012). Verwendet wurden zentralasiatische und fernöstliche Daten der ALFRED Datenbank und Han Chinesen aus Beijing und – als Vertreter europäischer Populationen – Daten des CEPH (*Centre d'Etude du Polymorphisme*

*Humain*), gewonnen von Bewohner Utahs mit nord- und westeuropäischen Wurzeln, aus der NCBI SNP-Datenbank. Die Populationen, die für die zentralasiatischen Daten verwendet wurden umfassen Kasachen, Mongolen, Uiguren und Tsaatan – ausgewählt nach dem ungefähren Ursprungsgebiet der skythischen Populationen. Die fernöstlichen Daten stammen von Jakuten und Tschuktschen.

Um den schwankenden Populationsgrößen für die untersuchten Marker Rechnung zu tragen, wurden für die ermittelten Frequenzen binomiale Konfidenzintervalle gebildet (CLOPPER & PEARSON 1934). Das Konfidenzintervall ist abhängig von der Stichprobengröße und gibt den Bereich an, der mit einer 95%igen Wahrscheinlichkeit die wahre Verteilung der Parameter umfasst.

## A4.4 Biostatistik

Zur statistischen Auswertung wurden mit Hilfe des Programmpakets Arlequin 3.5.1.2 (EXCOFFIER & LISCHER 2010), die Haplotypen- und Nukleotiddiversität, paarweise  $F_{ST}$ -Werte und  $F_u$ 's  $F_s$  Werte berechnet, wobei die Haplotypendiversität  $\hat{H}$  und die Nukleotiddiversität  $\hat{\pi}$  als Maß für die Variabilität innerhalb einer Population dienen und der  $F_{ST}$ -Wert die genetische Distanz zwischen zwei Populationen angibt.  $F_u$ 's  $F_s$ -Werte werden normalerweise als Test benutzt um festzustellen ob sich eine Population im Gleichgewicht befindet und Selektions-neutralität herrscht.  $F_u$  hat allerdings festgestellt, dass die  $F_s$  Statistik sensitiv auf demographische Expansion reagiert, was sich in negativen  $F_s$ -Werte widerspiegelt (FU 1997). Ob die untersuchten Populationen eine genetische Struktur aufweisen wurde mit Hilfe einer AMOVA (*Analysis of Molecular Variance*) getestet, die auch in das Arlequin Paket integriert ist. Das Programm SAMOVA (*Spatial Analysis of Molecular Variance*) definiert Gruppen von Populationen, die eine geographische Homogenität bei maximaler genetischer Differenzierung aufweisen (DUPANLOUP ET AL. 2002). Für all diese Analysen wurden jeweils die HVR1 Sequenzen der untersuchten Individuen von Position 16040 bis 16400 verwendet.

## A5 Ergebnisse

### A5.1 Datensatz skythischer Individuen bzw. Populationen

Von den ursprünglich zur Analyse vorgesehenen 126 Proben mussten zehn Individuen aufgrund ihrer Datierung, bzw. unsicherer Datierung ausgeschlossen werden. Drei Individuen wurden doppelt beprobt, eins stammte aus einem Massengrab und eines konnte keinem existierenden Grab zugeordnet werden. Die verbleibenden 111 Proben wurden auf ihren DNA-Gehalt hin getestet. Fünf der Proben enthielten keine DNA, bei acht Proben war der Erhalt zu schlecht um ein eindeutiges mtDNA Profil zu erhalten und eine Probe lieferte inkonsistente Sequenz-Ergebnisse. Für die Analysen verblieben somit Proben von 97 Individuen, für die es gelang, mitochondriale DNA zu vervielfältigen und anhand der vorliegenden Mutationen entsprechende Haplogruppen zuzuordnen (s. Tab. A3). Eine Übersicht über die Polymorphen Positionen auf denen die Haplogruppenzuordnung basiert findet sich im Anhang in den Tabellen 10–15.

#### Tab. A 3 Haplogruppen

Für einige Proben wurden keine Haplogruppen angegeben, die Begründung ist jeweils in der letzten Spalte gegeben: a: keine DNA, b: zu schlechter DNA Erhalt, c: Datierung, d: Probenzuordnung, e: nicht genug Material für eine Validierung der Ergebnisse. K. = Kurgan; Ind. = Individuum.

<i>Region</i>	<i>Fundplatz</i>	<i>K.</i>	<i>Grab</i>	<i>Ind.</i>	<i>Lab_ID</i>	<i>Haplogruppe</i>	<i>*</i>
Minusinsker	Barsučij Log		1	1	BL1	C5+16093	
Becken	Barsučij Log		1	2	BL2	U5a1a2 (U5a2c3)	
	Barsučij Log		1	3	BL3	C5+16093	
	Barsučij Log		1	4	BL4	C5+16093	
	Barsučij Log		1	5	BL5	U2e2	
	Barsučij Log		2	1	BL6	A4	
	Tuva	Aržan 2	2	M2		A1	-
	Aržan 2	2	M5	1	A2	A8	
	Aržan 2	2	M5	2	A3	G	
	Aržan 2	2	M7		A4	D4b1a2a	
	Aržan 2	2	M8		A5	C5b1	
	Aržan 2	2	M12		A6	U5a1f1	
	Aržan 2	2	M13a	1	A7	Y1	
	Aržan 2	2	M13a	2	A8	U4a3	
	Aržan 2	2	M13b		A9	U5a1d2b	
	Aržan 2	2	M14	1	A10	A11 (A8)	
	Aržan 2	2	M14	2	A11	G	
	Aržan 2	2	M15		A12	-	c
	Aržan 2	2	M18		A13	-	c
	Aržan 2	2	M20	1	A14	T1a	
	Aržan 2	2	M20	2	A15	-	d
	Aržan 2	2	M22		A17	H (H1+16189)	

<i>Region</i>	<i>Fundplatz</i>	<i>K.</i>	<i>Grab</i>	<i>Ind.</i>	<i>Lab_ID</i>	<i>Haplogruppe</i>	<i>*</i>
	Aržan 2	2	M23		A18	-	c
	Aržan 2	2	M24		A19	C4	
	Aržan 2	2	M25		A20	G2a	
	Aržan 2	2	M26		A21	A4	
	Aržan 2	2	M29		A22	-	c
	Aržan 2	2	ADE		A23	-	c
	Aržan 2		GndK		A24	-	d
Altai	Balik Sook	1			BS1	X2b	
	Balik Sook	27			BS2	U4b1a4	
Chuya-Tal	Alagail 2	8	12		Ala1	C4a1+16129	
	Alagail 2	8	15		Ala2	C4	
	Alagail 2	10			Ala4	T	
	Barburgazi 1	4			B1_1	G	
	Barburgazi 1	7		1	B1_2	D4m2	
	Barburgazi 3	5		2	B3_1	T1a1b	
	Baratal 2	2			Bt1	T2b+@16296	
	Baratal 2	3		2	Bt2	U2e2a	
	Dcholin 2	1			D1	T1a	
	Justyd 12	1			J12_1	T2b+@16296	
	Justyd 12	3A			J12_2	-	b
	Justyd 12	6			J12_3	U7	
	Justyd 12	10			J12_6	Z	
	Justyd 12	17		2	J12_7	T2b+@16296	
	Justyd 12	17		3	J12_8	A8	
	Justyd 12	18			J12_9	G1a1	
	Justyd 22	1			J22_1	T2b+@16296	
	Ulandryk 1	9			U1_1	W4a	
	Ulandryk 1	11			U1_2	Z1a	
	Ulandryk 2	1			U2_1	F2a	
	Ulandryk 2	2			U2_2	F2a	
	Ulandryk 4	2			U4_1	U2e2	
	Ulandryk 4	3			U4_4	K	
Ukok-Plateau	Ak Alakha 1	1	1		Ak1_1	C4a1+16129	
	Ak Alakha 1	1	1		Ak1_2	-	d
	Ak Alakha 4	1			Ak4_1	A11 (A8)	
	Ak Alakha 5	1			Ak5_1	C4	
	Ak Alakha 5	1			Ak5_2	-	b
	Ak Alakha 5	3	1		Ak5_3	-	b
	Ak Alakha 5	3	1		Ak5_4	D4b1	
	Ak Alakha 5	3	2		Ak5_5	U2e1a	
	Ak Alakha 5	4	1		Ak5_6	U2e1a	
	Ak Alakha 5	4	2		Ak5_7	C4a1+16129	
	Ak Alakha 5	5			Ak5_8	A11 (A8)	
	Kuturguntas 1	1			K1	CRS	
	Kuturguntas 1	1			K2	-	d
	Moinak 2	1		1	Mo1	D4b1a2a1	
	Moinak 2	2		1	Mo2	D4h4a	
	Verch Kal'džin1	1			VK1_1	K1 (K1b2/K1c2)	
kasachisches	Berel'	16			Be1	-	d
Altai-Gebirge	Berel'	16			Be2	D4	
	Berel'	16			Be3	CRS	
	Berel'	32			Be4	A11 (A8)	

<i>Region</i>	<i>Fundplatz</i>	<i>K.</i>	<i>Grab</i>	<i>Ind.</i>	<i>Lab_ID</i>	<i>Haplogruppe</i>	<i>*</i>
	Berel'	10			Be6	A6	
	Berel'	34			Be8	D4	
	Berel'	73			Be9	A4f	
	Berel'	31			Be11	C4a1+16129	
	Berel'	36			Be12	HV2	
	Berel'	71			Be14	A6	
	Tar Asu	23			Ta1	K2b1a	
Ostkasachstan	Ismailovo	6			Is1	D4h1	
	Ismailovo	10		1	Is2	HV	
	Ismailovo	27			Is4	H2a1	
	Majemir				MaI_1	-	c
	Majemir	1			MaII_2	-	c
	Majemir	2			MaII_3	-	c
	Zevakino	25			Ze1	-	c
	Zevakino	99b			Ze2	T2b	
	Zevakino	224			Ze3	K1a1 (K1b1b)	
	Zevakino	10	1		Ze4	D4h4a (D4j+16311)	
	Zevakino	10	2		Ze5	I	
	Zevakino	46	1		Ze6	C4 (C4a1/C4b8/C4+152+16093)	
	Zevakino	33			Ze7	U4a1	
	Zevakino	10	4		Ze8	D4	
	Zevakino	336			Ze9	D4j3	
Südkasachstan	Čirikrabort	1		1	CR1	-	a
	Čirikrabort	1		3	CR2	-	b
	Čirikrabort	1		4	CR3	-	a
Orenburg Region	Pokrovka	2			Pr1	U3	
	Pokrovka	2			Pr2	-	b
	Pokrovka	2			Pr3	M* (M5/M7b'd/M13b)	
	Pokrovka	2			Pr4	U1a'c	
	Pokrovka	2			Pr5	T	
	Pokrovka	2			Pr6	F1b	
	Pokrovka	2			Pr7	N1a1a1a1a	
	Pokrovka	2			Pr8	T2	
	Pokrovka	2			Pr9	U2e2	
	Pokrovka	2			Pr10	H2a1f	
	Pokrovka	2			Pr11	T1a	
	Pokrovka	2			Pr13	U5a1d2b	
Woronesch	Kolbino	8	1		Kol1	X4	
Region	Kolbino	10			Kol2	H8c	
	Kolbino	10_I			Kol3	U4	
	Kolbino	44	2		Kol4	-	b
	Kolbino	44	4		Kol5	J2b1a	
Nord-Kaukasus	Novozavedennoe	10			Nov1	-	a
	Novozavedennoe	13	2		Nov2	-	a
	Novozavedennoe	1			Nov3	-	b
	Novozavedennoe	9	2		Nov4	-	a
	Novozavedennoe	3	1		Nov5	CRS	
	Novozavedennoe	9	1		Nov6	-	e
	Novozavedennoe	2			Nov7	H1c	
	Novozavedennoe	6			Nov8	-	b
	Novozavedennoe	12			Nov9	T2g1	
	Novozavedennoe	14			Nov10	X	



Um ein möglichst vollständiges Bild der genetischen Zusammensetzung skythischer Populationen zu erhalten wurden alle verfügbaren Literaturdaten in die Analysen einbezogen (s. Tab. A 4). Weiterhin wurden die Individuen, inklusive der Literaturdaten, für die weiteren Analysen in chronologisch und geographisch verschiedene kulturelle Gruppen aufgeteilt. Probenanzahl und Fundplätze der jeweiligen Populationen, inklusive der verwendeten Literaturdaten sind Tabelle A 5 zu entnehmen.

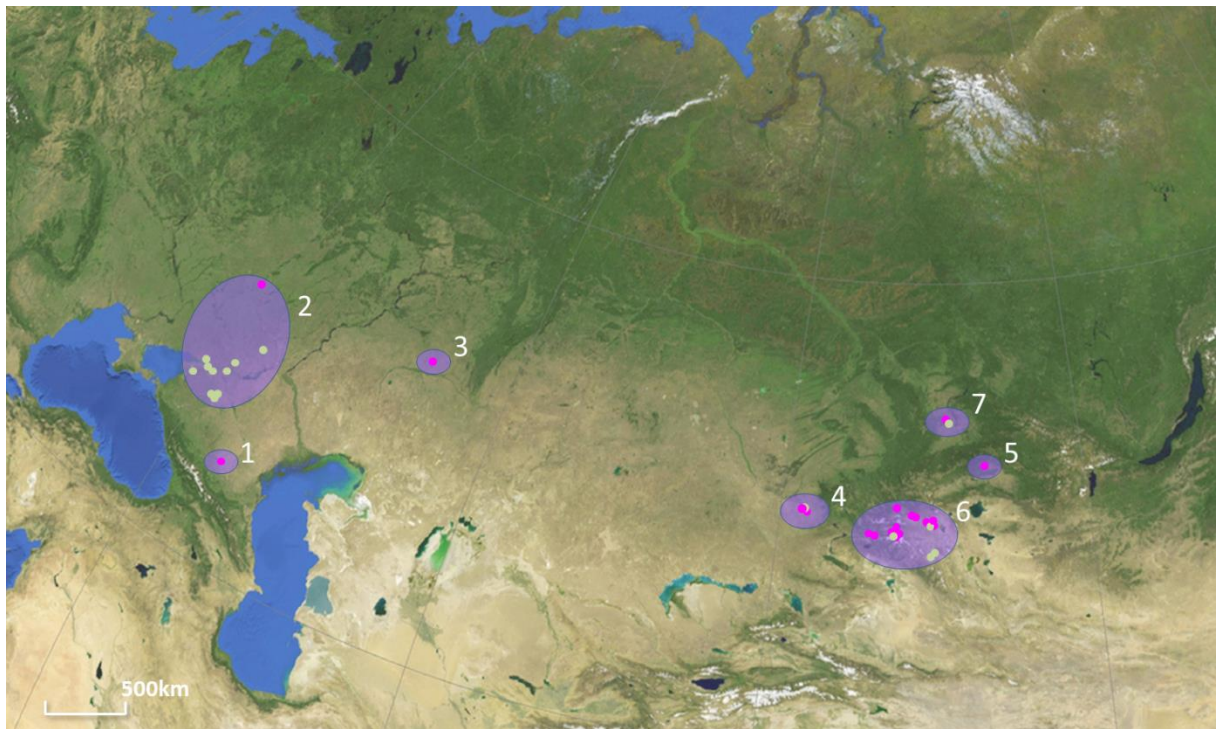
**Tab. A 4 Literaturdaten**

<i>Region</i>	<i>Fundplatz/Region</i>	<i>Probe</i>	<i>Haplogruppe</i>	<i>Quelle</i>
Chakassien	Anach	S21	T2b2b	KEYSER <i>ET AL.</i> 2009
	Anach	S22	T2b2b	
	Chernogorsk	S23	T2c1	
	Chernogorsk	S24	M25	
	Oust-Abakabsty	S25	G2a	
	Beysky region	S26	C5b1	
	Bogratsky region	S27	CRS	
	Bogratsky region	S28	F1b	
	Bogratsky region	S29	CRS	
	Bogratsky region	S32	H5	
Altai	Verch Kal'džin	VK2_1	C	VOEVODA <i>ET AL.</i> 2000
	Verch Kal'džin	VK2_3	U5a1d2b	
	Kizil	95KBI52	N1a1a1a1a	RICAUT <i>ET AL.</i> 2004A
	Sebystei	SEB96K1	F2a	RICAUT <i>ET AL.</i> 2004B
	Sebystei	SEB96K2	D4	
kasachisches Altai	Berel'	BeK11S1	CRS	CLISSON <i>ET AL.</i> 2002
	Berel'	BeK11S2	D4g1	
NW-Mongolei	Olon Kurin Gol	OKG6	HV2	PILIPENKO <i>ET AL.</i> 2010
	Olon Kurin Gol	OKG10	U5a1	
	Baga Turgen Gol	BTG05.T1	D4b1	GONZÁLEZ-RUIZ <i>ET AL.</i>
	Baga Turgen Gol	BTG05.T2	K	
	Baga Turgen Gol	BTG05.T8.1	U5a1	
	Baga Turgen Gol	BTG05.T8.2	C	
	Baga Turgen Gol	BTG05.T8.3	D	
	Baga Turgen Gol	BTG06.T3	J	
	Baga Turgen Gol	BTG06.T8	D	
	Baga Turgen Gol	BTG06.T10A	HV6	
	Baga Turgen Gol	BTG06.T10B	D	
	Baga Turgen Gol	BTG06.T11A	K	
	Baga Turgen Gol	BTG06.T11B	K	
	Baga Turgen Gol	BTG06.T12	U5a1	
	Baga Turgen Gol	BTG06.T13	C4a1+16129	
	Tsengel Khairkhan	TSK07.T1	A	
	Tsengel Khairkhan	TSK07.T2A	G2a	
	Tsengel Khairkhan	TSK07.T2B	T1a	

<i>Region</i>	<i>Fundplatz/Region</i>	<i>Probe</i>	<i>Haplogruppe</i>	<i>Quelle</i>
SW-Russland	Rostov-Don Area	RD1	F1b	DER SARKISSIAN 2011
	Rostov-Don Area	RD2	C	
	Rostov-Don Area	RD5	U5a1	
	Rostov-Don Area	RD6	T1a	
	Rostov-Don Area	RD7	T2	
	Rostov-Don Area	RD8	A4	
	Rostov-Don Area	RD9	CRS	
	Rostov-Don Area	RD10	H2a1	
	Rostov-Don Area	RD11	T1a	
	Rostov-Don Area	RD12	U2e	
	Rostov-Don Area	RD13	D	
	Rostov-Don Area	RD14	I3a	
	Rostov-Don Area	RD15	I3a	
	Rostov-Don Area	RD16	U5a	
	Rostov-Don Area	RD17	D4b1	

**Tab. A 5 Probenanzahl pro Fundplatz nach Kultur inklusive Literaturdaten**

Region	Kultur	Datierung	Fundplatz/Region	N
Ost	Zevakino-Čilikta	9.–7. Jh. v. Chr.	Ismailovo	3
			Zevakino	8
	Aldy Bel	7.–6. Jh. v. Chr.	Aržan 2	15
	Tagar	5. Jh. v. Chr.	Barsučij Log	6
	Tagar/Tes	8. Jh. v. Chr. – 1. Jh. n. Chr.	Chakassien	10
	Pazyryk	4.–3. Jh. v. Chr.	Berel'	11
			Tarasu	1
			Ak Alacha 1, 4 und 5	8
			Kuturguntas 1	1
			Moinak 2	2
			Verch Kal'džin 1 und 2	3
			Alagail 2	3
			Baburgazy 1 und 3	3
			Balik Sook	2
			Baratal 2	2
			Dcholin 2	1
			Justyd 12 und 22	7
			Ulandryk 1, 2 und 4	6
			Sebystei	2
	Kizil	1		
Olon-Kurin-Gol	2			
Bajan-Ölgii	16			
West	initial Skythen	8.–6. Jh. v. Chr.	Novozavedennoe	4
	Skythen	6.–2. Jh. v. Chr.	Rostov-Don	15
			Kolbinho	4
	Sarmaten	5.–2. Jh. v. Chr.	Pokrovka	11



**Abb. A 2 Verteilung der Fundplätze über die Eurasische Steppe**

Dargestellt sind Proben die im Zuge dieser Studie bearbeitet wurden (pink) und Daten aus der Literatur (grün). Die Nummern beziehen sich auf die vorab definierten Gruppen: **1** initiale Skythenzeit (N=4); **2** klassische Skythenzeit (N=19); **3** frühe Sarmaten (N=11); **4** Zevakino-Čilikta-Kultur (N=11); **5** Aldy Bel-Kultur (N=15); **6** Pazyryk-Kultur (N=71); **7** Tagar/Tes-Stufe (N=16). (Quelle der Karte: GINKOMAPS)

## A5.2 Haplogruppenzusammensetzung der skythischen Populationen

Die ermittelten Haplotypen finden sich auch in modernen europäischen, bzw. westeurasischen und ostasiatischen, bzw. osteurasischen Populationen wieder (COMAS *ET AL.* 1998, COMAS *ET AL.* 2004, QUINTANA-MURCI *ET AL.* 2004). Von 78 Individuen aus dem östlichen Skythen-Territorium zeigen 34 west-eurasische und 44 osteurasische Linien. Während bei den Proben aus Ostkasachstan, die zu den ältesten hier untersuchten gehören und der Zevakino-Čilikta-Stufe zugeordnet werden, die westeurasischen Linien mit einem Anteil von 54,5% leicht überwiegen, dominieren in den darauf folgenden Kulturstufen eher osteurasische Linien (54,3 – 66,7%). Die 19 erfolgreich analysierten Individuen aus dem Westen der eurasischen Steppe weisen dagegen fast ausschließlich europäische Linien auf. Unter den elf Individuen des frühsarmatischen Fundplatzes Pokrovka befanden sich lediglich zwei osteurasische Linien und die Haplogruppen der acht erfolgreich analysierten Proben westlicher Skythen der Fundplätze Kolbino und Novozavedennoe lassen sich alle unter die westeurasischen Linien subsumieren.

Unter Einbeziehung der Literaturdaten konnte der Datensatz für die Altai Region um 25 Individuen der Pazyryk erweitert werden (CLISSON *ET AL.* 2002, GONZÁLEZ-RUIZ *ET AL.* 2012, PILIPENKO *ET AL.* 2010, RICAUT *ET AL.* 2004A, RICAUT *ET AL.* 2004B, VOEVODA *ET AL.* 2000), wodurch sich die Anteile europäischer und ostasiatischer Linien jedoch nur um 2,2 % ändern (s. Tab. A 6). Anders sieht es in der südsibirischen Republik Chakassien und der westeurasischen Steppe aus: Für die vorliegende Arbeit wurden sechs Individuen der Tagar Kultur aus dem 5. Jh. v. Chr. untersucht, die alle von dem Fundplatz Barsučij Log im Bereich des Minusinsker Beckens stammen. Der Anteil an ostasiatischen Linien liegt für diesen Fundplatz bei 66,7%. Um den Datensatz zu erweitern wurden zehn eisenzeitlichen Proben aus Chakassien (KEYSER *ET AL.* 2009), die aus dem 8. Jh. v. Chr. bis zum 1. Jh. n. Chr. stammen und somit der Tagar-Kultur bzw. der späteren Tes-Stufe zugeordnet werden, in die Berechnungen mit einbezogen. Der Anteil an ostasiatischen Linien liegt bei den Tagar/Tes Proben von Keyser bei lediglich 40%. Gemeinsam mit den hier analysierten Barsučij Log Individuen liegt der Anteil ostasiatischer Linien bei 50%.

Im Westen der eurasischen Steppe konnten nur vier Individuen aus der klassischen Skythenzeit aus Kolbino, nördlich des Schwarzen Meeres, erfolgreich analysiert werden. Zur Erweiterung des Datensatzes bot sich hier die Arbeit von Der Sarkissian an, die in ihrer

Dissertation 16 Individuen aus dem Rostov-Don Gebiet untersuchte (DER SARKISSIAN 2011). Ein Individuum ordnet sie der sarmatischen Kultur zu, weshalb es nicht mit den Kolbino Daten zusammen analysiert wurde, die restlichen 15 der frühen- bzw. klassischen Skythenzeit. Die Individuen aus Kolbino zeigen keine ostasiatischen Linien, sondern zu 100% europäische, wohingegen die Rostov-Don Skythen 31% ostasiatische Linien aufweisen. Die 15 skythischen Individuen des Rostov-Don Gebietes und die 4 aus Kolbino wurden, als Vertreter der westlichen Skythen, gemeinsam betrachtet und analysiert.

**Tab. A 6 Anteile west- und osteurasischer Linien**

Verteilung nach Kulturgruppen; inkl. Lit. = Proben dieser Studie + Literaturdaten, N = Probenanzahl

	Kultur	Quelle	Datierung [v.Chr.]	N	mtDNA Linien [%]	
					europäisch	ostasiatisch
Ost	Zevakino-Čilikta	diese Studie	900–600	11	54.5	45.5
	Aldy Bel	diese Studie	700–500	15	33.3	66.7
	Tagar	diese Studie	500–400	6	33.3	66.7
	Tagar/Tes	inkl. Lit.	800–100 n.Chr.	16	50.0	50.0
	Pazyryk	diese Studie	400–200	46	45.7	54.3
	Pazyryk	inkl. Lit.	400–200	71	47.9	52.1
West	Initial Skythen	diese Studie	800–600	4	100	0
	Skythen	diese Studie	300–200	4	100	0
	Skythen	inkl. Lit.	600–200	19	73.7	26.3
	Sarmaten	diese Studie	500–200	11	81.8	18.2

### A5.3 Populationsgenetik der skythischen Kulturgruppen

Die Variabilität der mitochondrialen Daten innerhalb von und zwischen Populationen gibt Aufschluss über den Grad der genetischen Ähnlichkeit von Individuen und Gruppen. Zur Klärung der Frage, ob sich die skythische Kultur mit ihren Trägern über den gesamten Steppengürtel verbreitete oder eben nur von jeweils benachbarten Nomadenstämmen übernommen und weitergetragen wurde, wurden die vorab definierten chronologisch und geographisch verschiedenen kulturellen Gruppen miteinander verglichen. Die Gruppen umfassen dabei jeweils die in dieser Studie analysierten Daten und die Literaturdaten. Die Tagar-Population des Fundplatzes Barsučij Log ist zusätzlich gesondert aufgeführt.

#### Haplotypen- und Nukleotiddiversität

Als Maß für Variabilität innerhalb einer Population dienen die Haplotypen- und die Nukleotiddiversität (s. Tab. A 7). Die niedrigste Haplotypendiversität zeigt die Tagar-Gruppe aus Barsučij Log mit  $0,8000 \pm 0,1721$ . Die restlichen Populationen weisen Haplotypendiversität zwischen 0,9583 und 1,000 auf. Die Nukleotiddiversitäten reichen von 0,009234 für die initialen Skythen bis zu 0,021657 für die Sarmaten.

**Tab. A 7 Haplotypen- und Nukleotiddiversität**

N = Anzahl der Individuen, K = Anzahl an unterschiedlichen Haplotypen

Kultur	Quelle	N	K	$\hat{H} \pm sd$	$\pi \pm sd$
Sarmaten	diese Studie	11	11	$1.0000 \pm 0.0388$	$0.021657 \pm 0.012325$
Zevakino-Čilikta	diese Studie	11	11	$1.0000 \pm 0.0388$	$0.013095 \pm 0.007828$
Skythen	diese Studie	4	4	$1.0000 \pm 0.1768$	$0.018467 \pm 0.013179$
Initial Skythen	diese Studie	3	3	$1.0000 \pm 0.2722$	$0.009234 \pm 0.008026$
Aldy Bel	diese Studie	15	14	$0.9905 \pm 0.0281$	$0.017781 \pm 0.010003$
Skythen	alle	19	17	$0.9883 \pm 0.0210$	$0.018402 \pm 0.010157$
Pazyryk	alle	71	47	$0.9855 \pm 0.0049$	$0.018466 \pm 0.009774$
Pazyryk	diese Studie	45	31	$0.9788 \pm 0.0098$	$0.017972 \pm 0.009617$
Tagar/Tes	alle	16	12	$0.9583 \pm 0.0363$	$0.019621 \pm 0.010889$
Tagar	diese Studie	6	4	$0.8000 \pm 0.1721$	$0.019391 \pm 0.012264$

## Genetische Distanzen

Mit dem Programmpaket Arlequin 3.5.1.2 wurden paarweise  $F_{ST}$ -Werte bestimmt (*Distance method: pairwise difference; Gamma a value = 0*), wobei ein höherer  $F_{ST}$ -Wert eine größere genetische Distanz bedeutet. Negative  $F_{ST}$ -Werte können mit dem Wert Null gleichgesetzt werden und zeigen keine genetische Distanz an. Ein Vergleich zwischen 13 modernen Populationen aus Europa, dem Mittleren Osten, Indien, China, Papua Neu Guinea und Australien ergab einen maximalen  $F_{ST}$ -Wert von 0,133 (BRAMANTI ET AL. 2009). Ausgehend von diesem Wert als Maximalwert, wurde folgende Einteilung vorgenommen:  $F_{ST}$ -Werte unter 0,01 wurden als zu gering eingestuft um eine genetische Distanz anzuzeigen, die Werte 0,01–0,03 werden als geringe genetische Distanz gewertet, Werte von 0,03 bis 0,06 zeigen eine mittlere, 0,06–0,09 eine erhöhte und Werte über 0,09 eine große genetische Distanz an.

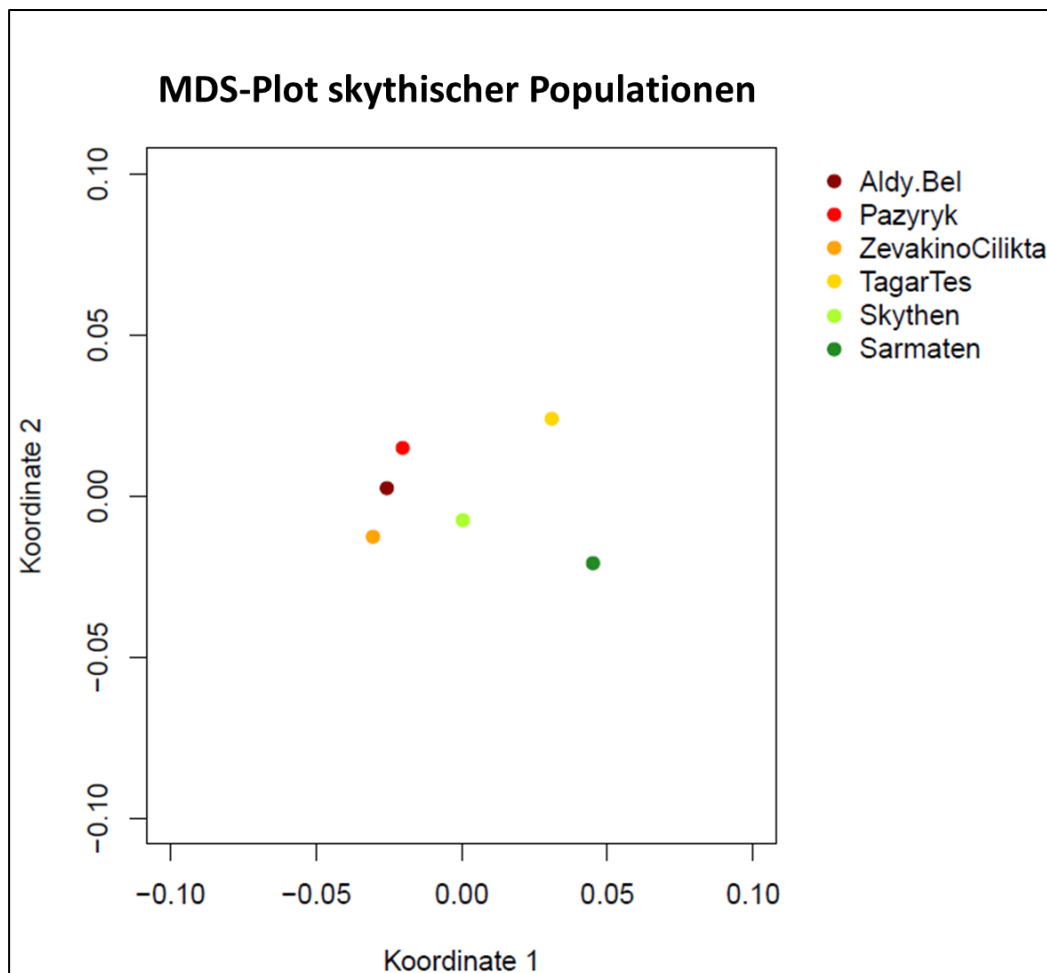
Weiterhin wurden linearisierte  $F_{ST}$ -Werte nach Reynolds bestimmt (REYNOLDS ET AL. 1983). Anhand der Reynolds Distanzen (s. Anhang Tab.19–21) wurde eine nicht metrische Multidimensionale Skalierung (*nicht metrische MDS*) mit der isoMDS Funktion des R Paketes „MASS“ durchgeführt und anschließend ein zweidimensionaler Plot erstellt (R\_CORE\_TEAM 2013, VENABLES & RIPLEY 2002).

Es wurden zuerst alle westlichen mit den östlichen Individuen verglichen, wobei die westliche Gruppe einen signifikant geringen  $F_{ST}$ -Wert von 0,01733 (*p-value: 0,01855±0,0036*) zur östlichen Gruppe zeigt. Im nächsten Schritt wurden die Unterschiede der zeitlich und räumlich verschiedenen Gruppen zueinander untersucht (s. Tab. A 8). Die Ergebnisse der  $F_{ST}$ -Wert Kalkulationen zeigen keine genetische Distanz zwischen der Zevakino-Čilikta-, der Aldy Bel- und der Pazyryk-Kultur und eine sehr geringe genetische Distanz zwischen diesen Kulturen und der Tagar/Tes-Gruppe. Dagegen bestehen zwischen der Tagar-Kultur und den restlichen Gruppen mittlere bis große genetische Distanzen ( $F_{ST}$ -Werte: 0,03351–0,09589).

Die westlichen Gruppen, die mit der skythischen Kultur in Verbindung gebracht werden, zeigen, wie die Populationen des Altai-Gebirges, keine genetische Distanz zueinander. Die initialen Skythen und die Individuen aus der frühen bis klassischen Skythenzeit zeigen außerdem keine genetische Distanz zu den östlichen Gruppen, abgesehen von der Tagar-Kultur. Die frühen Sarmaten zeigen keine Unterschiede zu den anderen Gruppen im Westen, offenbaren jedoch eine größere genetische Distanz zu den östlichen Gruppen – abgesehen von der Tagar/Tes Population, zu der sie nur eine geringe genetische Distanz aufweisen.

**Tab. A 8  $F_{ST}$ -Werte zwischen den skythischen Kulturgruppen**Signifikante  $F_{ST}$ -Werte ( $p$ -value < 0,05) sind fett hervorgehoben. Die  $p$ -values finden sich im Anhang Tab. 16.

	1	2	3	4	5	6	7	8
1 TagarTes								
2 Tagar								
3 Aldy Bel	0,01119	0,04039						
4 Pazyryk	0,00504	0,04204	-0,00959					
5 Zevakino-Čilikta	0,01507	0,08830	-0,00223	-0,01111				
6 initial-Skythen	-0,08808	0,03351	-0,05889	-0,07597	-0,02800			
7 Skythen	-0,00181	<b>0,07235</b>	-0,01405	-0,00055	-0,00660	-0,12205		
8 Sarmaten	0,00428	<b>0,09589</b>	0,04650	<b>0,04020</b>	<b>0,05125</b>	-0,10115	-0,01468	

**Abb. A 3 MDS-Plot der skythischen Populationen**Aufgrund der geringen Populationsgröße wurden die Individuen aus der initialen Skythenzeit im Westen nicht mit einbezogen. *Stress value* = 0

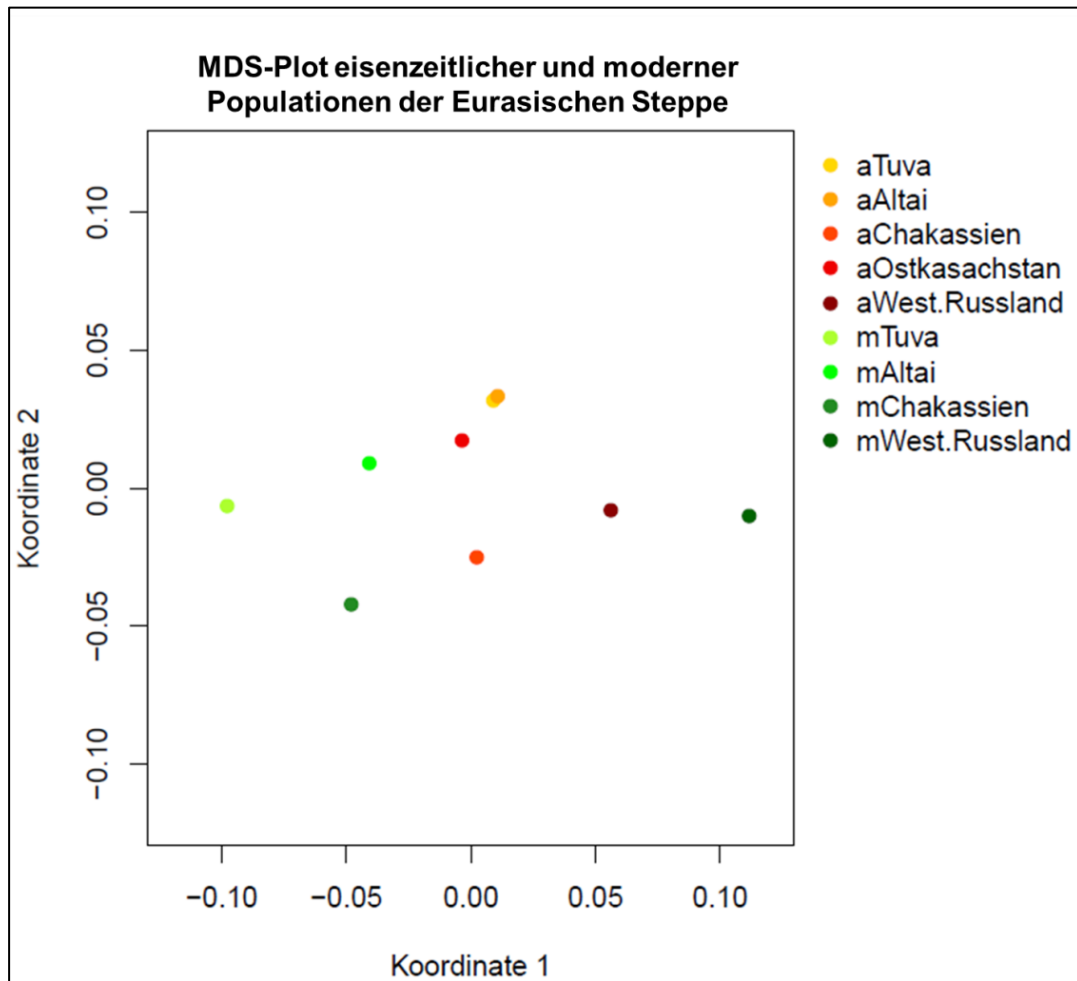


Die skythischen Populationen wurden zusätzlich mit modernen Populationen aus dem Bereich des Altai (N = 110), Tuva (N = 90), Chakassien (N = 53) (DERENKO *ET AL.* 2003) und West-Russland (N = 198) (MALYARCHUK *ET AL.* 2002) verglichen. Die skythische Gruppe der Aldy Bel-Stufe aus Tuva zeigt mit einem  $F_{ST}$ -Wert von 0,04729 eine mittlere genetische Distanz zu der modernen Population aus Tuva. Alte und moderne Populationen aus dem Altai und Chakassien zeigen jeweils geringe genetische Distanzen zueinander (0,01777 bzw. 0,01393), ebenso die Gruppe der West-Skythen und Sarmaten zur modernen Population aus Südwest-Russland (0,01279). Die modernen Populationen aus dem Osten der eurasischen Steppe zeigen geringe bis mittlere genetische Distanzen zueinander jedoch hohe bis sehr hohe zu der modernen Population aus Südwest-Russland.

**Tab. A 9  $F_{ST}$ -Werte zwischen skythischen Gruppen und modernen Populationen gleicher Regionen**

Signifikante  $F_{ST}$ -Werte ( $p$ -value <0,05) sind fett hervorgehoben. Die  $p$ -values finden sich im Anhang Tab. 17. a = eisenzeitliche Populationen (aTuva = Aldy Bel; aAltai = Pazyryk; aChakassien = Tagar/Tes; aOstkasachstan = Zevakino-Čilikta; aWest-Russland = Initiale Skythen, Skythen und Sarmaten); m = moderne Populationen

		1	2	3	4	5	6	7	8	9
1	aTuva									
2	aAltai	-0.00959								
3	aChakassien	0.01119	0.00504							
4	aOstkasachstan	-0.00223	-0.01111	0.01507						
5	aWest-Russland	0.01508	<b>0.01871</b>	0.00767	0.01582					
6	mTuva	<b>0.04729</b>	<b>0.05384</b>	<b>0.03679</b>	<b>0.03767</b>	<b>0.09892</b>				
7	mAltai	0.01401	<b>0.01777</b>	0.01281	-0.00286	<b>0.03437</b>	<b>0.02612</b>			
8	mChakassien	<b>0.03740</b>	<b>0.03619</b>	0.01393	0.01587	<b>0.05090</b>	<b>0.02065</b>	<b>0.00953</b>		
9	mWest-Russland	<b>0.07058</b>	<b>0.05786</b>	<b>0.05994</b>	<b>0.04839</b>	<b>0.01279</b>	<b>0.15671</b>	<b>0.08136</b>	<b>0.10217</b>	



**Abb. A 4 MDS-Plot eisenzeitlicher und moderner Populationen der Eurasischen Steppe**  
a = eisenzeitliche Population (aTuva = Aldy Bel; aAltai = Pazyryk; aChakassien = Tagar/Tes;  
aOstkasachstan = Zevakino-Čilikta; aWest-Russland = Skythen und Sarmaten); m = moderne Population;  
*stress-value* = 0,039

## Test auf genetische Barrieren

Die Ergebnisse der SAMOVA untermauern die größere genetische Distanz der frühen Sarmaten und auch der Tagar/Tes-Gruppe zu den restlichen Skythen, indem sie als eigene Gruppen separiert werden. Die Unterscheidung zwischen den frühen Sarmaten und den restlichen Populationen und auch die Trennung in drei Gruppen (frühe Sarmaten, Tagar/Tes und die restlichen skythischen Gruppen) liefern jedoch keine signifikanten Ergebnisse. Signifikante  $F_{CT}$ -Werte finden sich lediglich für die Unterteilung der Populationen in  $K=4$  und  $K=5$ . Der höchste  $F_{CT}$ -Wert wird bei  $K=6$  erreicht. Hierbei gilt jedoch zu beachten, dass sobald der  $F_{SC}$ -Wert negativ wird, die Variabilität innerhalb einer Population größer als die Variabilität innerhalb der Gruppe ist. Erreicht der  $F_{SC}$ -Wert also Null, ist eine Aufspaltung in weitere Gruppen nicht wünschenswert, da die Anzahl maximal differenzierter Populationen erreicht ist (RODRÍGUEZ-ROBLES *ET AL.* 2010). Da für die skythischen Populationen keine positiven Werte für  $F_{SC}$  erreicht werden (s. Tab. A 10), erscheint eine Aufspaltung in Untergruppen nicht sinnvoll.

### Tab. A 10 Ergebnisse der SAMOVA

$K$  = Anzahl der Gruppen;  $F_{CT}$ =Varianz zwischen den Gruppen,  $F_{SC}$ =Varianz zwischen Populationen innerhalb der Gruppen,  $p$ =Signifikanzniveau, Paz=Pazyryk, TaTe=Tagar/Tes, AB=Aldy Bel, ZC=Zevakino-Čilikta, Sky=Skythen, IniSky=initiale Skythen, FSar=frühe Sarmaten

K	$F_{CT}$	$F_{SC}$	P ( $F_{CT}$ )	P ( $F_{SC}$ )	Struktur
2	0,03884	-0,00781	0,13490+-0,01096	0,78299+-0,01603	(Paz, TaTe, AB, ZC, Sky, IniSky) (FSar)
3	0,02927	-0,01337	0,05767+-0,00649	0,89541+-0,01041	(Paz, AB, ZC, Sky, IniSky) (FSar) (TaTe)
4	0,02992	-0,02235	0,00684+-0,00305	0,96774+-0,00561	(Paz, AB, ZC) (FSar) (TaTe) (Sky, IniSky)
5	0,04253	-0,04302	0,00880+-0,00288	0,92082+-0,00828	(Paz) (AB, ZC) (FSar) (TaTe) (Sky, IniSky)
6	0,11302	-0,12636	0,05083+-0,00762	0,98925+-0,00411	(Paz) (AB) (ZC) (FSar) (TaTe) (Sky, IniSky)

## Test auf Bevölkerungsexpansion

Für die Populationen wurden Fu's  $F_S$ -Werte bestimmt (s. Tab. A 11). Negative  $F_S$ -Werte können als Hinweis auf eine Bevölkerungsexpansion gewertet werden. Hierbei ist zu beachten, dass das Signifikanzniveau bei einem *p-value* von 0,02, statt wie üblich 0,05 liegt (FU 1997). Abgesehen von der Tagar-Kultur zeigen alle Populationen negative  $F_S$ -Werte, wobei jedoch die Tagar-Kultur, die Tagar/Tes-Gruppe und die initialen Skythen den Signifikanzwert von 0,02 nicht erreichen.

**Tab. A 11** Fu's  $F_S$ -Werte

Kultur	N	$F_S$	<i>p-value</i>
Aldy Bel	15	-7,34392	0,00000
Pazyryk	71	-25,09447	0,00000
Tagar	6	1,83665	0,81100
Tagar/Tes	16	-2,70141	0,08100
Zevakino-Čilikta	11	-7,29093	0,00000
Initial Skythen	3	-0,07696	0,24600
Skythen	19	-9,13951	0,00000
Sarmaten	11	-5,07773	0,00700

## A5.4 Die genetische Struktur der Pazyryk-Population

Die Funde der Pazyryk-Kultur verteilen sich auf den kasachischen, den russischen und den mongolischen Teil des Altai-Gebirges, wobei sie sich innerhalb des russischen Teils weiterhin auf zwei Regionen konzentrieren: das Ukok-Plateau und das Chuya-Tal (s. Tab. A12). Im Folgenden wurden die Pazyryk-Gruppen verschiedener geographischer Regionen auf ihre genetische Homogenität hin untersucht.

**Tab. A 12 Pazyryk-Population nach Region**

Gruppe	N	Fundplatz
kasachisches Altai	12	Berel'
		Tarasu
Ukok-Plateau	14	Ak Alacha 1, 4 und 5
		Kuturguntas 1
		Moinak 2
		Verch Kal'džin 1 und 2
Chuya-Tal	24	Alagail 2
		Baburgazy 1 und 3
		Baratal 2
		Dcholin 2
		Justyd 12 und 22
		Ulandryk 1, 2 und 4
		Sebystei
mongolisches Altai	18	Olon-Kurin-Gol
		Bajan-Ölgii

Zur Ermittlung der genetischen Distanzen wurden auch hier die  $F_{ST}$ -Werte bestimmt. Im ersten Schritt wurden die Unterschiede zwischen den Regionen kasachisches, russisches und mongolisches Altai-Gebirge berechnet und anschließend wurden die russischen Individuen in ihre geographischen Subregionen unterteilt. Vergleicht man die  $F_{ST}$ -Werte zwischen der kasachischen, der russischen und der mongolischen Pazyryk-Gruppe, zeigen sich insgesamt nur geringe genetische Distanzen (s. Tab. A 13). Es wird aber deutlich, dass sich die kasachische und die mongolische Gruppe näher stehen ( $F_{ST}=0,00439$ ), als beide zur russischen ( $F_{ST}=0,01184$  bzw.  $0,02187$ ). Durch die Unterteilung der russischen Gruppe in die beiden Regionen Ukok-Plateau und Chuya-Tal zeigt sich weiterhin, dass sich die Unterschiede zur russischen Pazyryk-Gruppe auf die Chuya Region zu konzentrieren scheinen. So gibt es keinen genetischen Unterschied der kasachischen sowie der mongolischen Gruppe zu den Individuen des Ukok-Plateaus; die Chuya-Gruppe weist hingegen eine mittlere genetische Distanz zur Ukok-Gruppe ( $F_{ST}=0,02979$ ) und zu den

mongolischen Pazyryk-Individuen ( $F_{ST}=0,03184$ ) auf und zeigt mit 0,05969 einen etwas höheren  $F_{ST}$ -Wert und somit eine größere genetische Distanz zur kasachischen Pazyryk-Gruppe (s. Tab. A 14).

**Tab. A 13  $F_{ST}$ -Werte zwischen den Pazyryk-Gruppen nach Land**

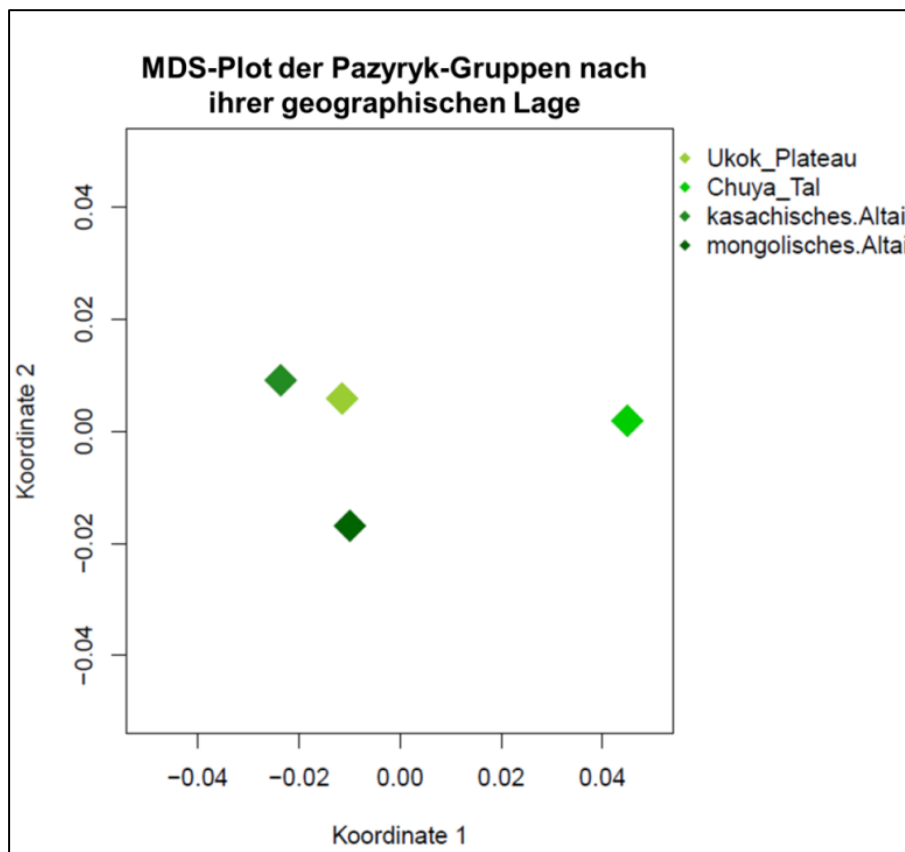
Die  $F_{ST}$ -Werte sind in der Tabelle links unten angegeben, die dazugehörigen *p-values* rechts oben

	kasachisches Altai	mongolisches Altai	russisches Altai
kasachisches Altai		0,37109+-0,0205	0,14160+-0,0088
mongolisches Altai	0,00439		0,18652+-0,0140
russisches Altai	0,02187	0,01184	

**Tab. A 14  $F_{ST}$ -Werte zwischen den Pazyryk-Gruppen nach geographischer Subregion**

Die  $F_{ST}$ -Werte sind in der Tabelle links unten angegeben, die dazugehörigen *p-values* rechts oben

	kasachisches Altai	mongolisches Altai	Ukok-Plateau	Chuya-Tal
kasachisches Altai		0,37402+-0,0135	0,48730+-0,0168	0,02637+-0,0046
mongolisches Altai	0,00439		0,49805+-0,0144	0,04883+-0,0092
Ukok-Plateau	-0,00395	-0,00257		0,09180+-0,0095
Chuya-Tal	0,05969	0,03184	0,02979	



**Abb. A 5 MDS-Plot der Pazyryk-Gruppen**

stress-value = 0

## A5.5 Das Ursprungsgebiet der Skythen

Eine  $F_{ST}$ -Wert Berechnung zur Bestimmung der genetischen Distanz wurden auch verwendet um weitere Hinweise auf ein mögliches Ursprungsgebiet der skythischen Populationen zu erhalten bzw. Verbindungen zwischen bronzezeitlichen und den eisenzeitlichen Individuen herzustellen. Dazu wurden neolithische und bronzezeitliche Populationen aus Zentralasien und den angrenzenden Gebieten mit den skythischen Gruppen in Ost und West verglichen. Hierzu wurden die skythischen Individuen in die Gruppen „Eisenzeit West“ (alle westlichen Individuen), „Frühe Eisenzeit Ost“ (Zevakino-Čilikta-Gruppe) und „Eisenzeit Ost“ (restliche Individuen Ost) aufgeteilt. Die allgemeine Bezeichnung „Zentralasien“ bezieht sich hier auf die Populationen aus der westsibirischen Waldsteppe. Die *p-values* der  $F_{ST}$ -Werte sind in der Tabelle ohne Standardabweichung angegeben, die Werte mit den dazugehörigen Standardabweichungen finden sich im Anhang in Tab. 18.

Den skythischen Populationen zeitlich am nächsten, nämlich am Übergang von der Bronze- zur Eisenzeit, zu Beginn des 1. Jahrtausends v. Chr., liegt die Gruppe aus der westsibirischen Waldsteppe Zentralasiens. Sie zeigt keine genetische Distanz zur „frühen Eisenzeit Ost“ ( $F_{ST}=0,00988$ ), eine sehr geringe genetische Distanz zur „Eisenzeit West“ ( $F_{ST}=0,01224$ ) und eine etwas höhere genetische Distanz zur „Eisenzeit Ost“ ( $F_{ST}=0,02778$ ). Alle drei skythischen Gruppen zeigen keine genetische Distanz zu den Individuen der späten Bronzezeit in Zentralasien und besonders große Unterschiede zur bronzezeitlichen Gruppe aus dem Tarim Becken in China ( $F_{ST}$ -Werte 0,12696–0,19477). Wohingegen die genetischen Distanzen zur chinesischen Population aus Lajia wesentlich geringer sind ( $F_{ST}$ -Werte 0,01449–0,03548). Die Gruppe „Eisenzeit West“ zeigt keine genetische Distanz zu den bronzezeitlichen Gruppen aus Chakassien und Kasachstan und zur Jamnaja-Kultur des nördlichen Schwarzmeerraums, die an den Übergang des Äneolithikums zur Bronzezeit datiert. Generell zeigt die Gruppe „Eisenzeit West“ jeweils die geringsten genetischen Distanzen zu den vorhergehenden Epochen im Westen, außer zur bronzezeitlichen Gruppe aus Nordwest Russland, hier zeigt sie von den eisenzeitlichen Populationen die größte genetische Distanz ( $F_{ST}=0,08048$ ).

**Tab. A 15  $F_{ST}$ -Werte zwischen den skythischen Gruppen, bronzezeitlicher und neolithischer Populationen**  
rechts unten  $F_{ST}$ -Werte, links oben: *p-values*;  $F_{ST}$ -Werte mit *p-values* <0,05 sind fett hervorgehoben

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		0.61523	0.18945	0.30566	0.35840	0.02637	0.07324	0.17480	0.02344	0.00293	0.25098	0.11328	0.07910	0.00586	0.02637	0.00391
2	-0,00928		0.01660	0.04492	0.62891	0.00000	0.00098	0.16895	0.02344	0.00000	0.14648	0.05664	0.00195	0.00684	0.00000	0.00000
3	0,01515	<b>0,01909</b>		0.19238	0.49902	0.00000	0.00000	0.65332	0.46289	0.00000	0.03613	0.11426	0.00000	0.12207	0.38770	0.00391
4	0,00988	<b>0,02778</b>	0,01224		0.04785	0.00000	0.00000	0.84766	0.02441	0.00000	0.00000	0.11426	0.00293	0.00684	0.13574	0.16504
5	0,00865	-0,01083	-0,00516	<b>0,05752</b>		0.49902	0.14941	0.45703	0.42285	0.01172	0.41504	0.36914	0.25586	0.12695	0.16699	0.00488
6	<b>0,06612</b>	<b>0,04747</b>	<b>0,07338</b>	<b>0,12438</b>	-0,00745		0.12012	0.00391	0.00781	0.00293	0.16797	0.01660	0.12695	0.00781	0.00000	0.00000
7	0,04194	<b>0,03978</b>	<b>0,08144</b>	<b>0,12233</b>	0,03368	0,01778		0.00098	0.00000	0.00098	0.42383	0.01660	0.18555	0.00000	0.00000	0.00000
8	0,02349	0,01748	-0,00981	-0,02497	-0,00115	<b>0,09254</b>	<b>0,10215</b>		0.69922	0.00000	0.00098	0.44629	0.03516	0.31152	0.63672	0.18262
9	<b>0,09385</b>	<b>0,05963</b>	-0,0036	<b>0,07582</b>	0,00680	<b>0,11584</b>	<b>0,17067</b>	-0,02828		0.00000	0.00000	0.02344	0.01562	0.41211	0.73633	0.00195
10	<b>0,19477</b>	<b>0,12696</b>	<b>0,15606</b>	<b>0,22578</b>	<b>0,13737</b>	<b>0,08013</b>	<b>0,13668</b>	<b>0,20484</b>	<b>0,27006</b>		0.01074	0.00391	0.00781	0.00000	0.00000	0.00000
11	0,01449	0,01544	<b>0,03548</b>	<b>0,14475</b>	0,00053	0,02073	-0,00008	<b>0,12393</b>	<b>0,20750</b>	<b>0,13804</b>		0.02051	0.34668	0.00000	0.00391	0.00000
12	0,05112	0,02800	0,0255	0,04350	0,00541	<b>0,06834</b>	<b>0,08113</b>	-0,00015	<b>0,10674</b>	<b>0,14423</b>	<b>0,10778</b>		0.04199	0.00684	0.00977	0.00195
13	0,05134	<b>0,04759</b>	<b>0,08048</b>	<b>0,11516</b>	0,01697	0,01553	0,01928	<b>0,06471</b>	<b>0,11759</b>	<b>0,09301</b>	0,00378	<b>0,07425</b>		0.02148	0.00000	0.00000
14	<b>0,10411</b>	<b>0,06710</b>	0,02533	<b>0,08167</b>	0,05590	<b>0,10669</b>	<b>0,16321</b>	0,01202	0,00141	<b>0,2672</b>	<b>0,17836</b>	<b>0,13823</b>	<b>0,10620</b>		0.59570	0.06934
15	<b>0,05323</b>	<b>0,04952</b>	0,00076	0,01905	0,02877	<b>0,10633</b>	<b>0,13419</b>	-0,01136	-0,02284	<b>0,20365</b>	<b>0,09651</b>	<b>0,06922</b>	<b>0,09476</b>	-0,01162		0.03516
16	<b>0,09379</b>	<b>0,07514</b>	<b>0,05015</b>	0,02021	<b>0,12911</b>	<b>0,15376</b>	<b>0,17149</b>	0,02139	<b>0,12750</b>	<b>0,27328</b>	<b>0,17902</b>	<b>0,10913</b>	<b>0,14919</b>	0,05812	<b>0,03340</b>	

**Tab. A 16 Populationen der  $F_{ST}$ -Tabelle**

1	Frühe Eisenzeit Ost (9.–7. Jh. v. Chr.); N=11	
2	Eisenzeit Ost (7.–3. Jh. v. Chr.); N=86	
3	Eisenzeit West (8.–3. Jh. v. Chr.); N=33	
4	Übergang Bronze- zu Eisenzeit Zentralasien (9.–8. Jh. v. Chr.); N=14	MOLODIN <i>ET AL.</i> 2012A
5	Späte Bronzezeit Zentralasien (Ende des 2. Jt. v. Chr.); N=8	MOLODIN <i>ET AL.</i> 2012A, 2012B
6	Mittlere Bronzezeit Zentralasien (3.–2. Jt. v. Chr.); N=46	MOLODIN <i>ET AL.</i> 2012A
7	Frühe Bronzezeit Zentralasien (4.–3. Jt. v. Chr.); N=27	MOLODIN <i>ET AL.</i> 2012A
8	Bronzezeit Chakassien (1800–800 v. Chr.); N=11	KEYSER <i>ET AL.</i> 2009
9	Bronzezeit Kasachstan (1400–1000 v. Chr.); N=8	LALUEZA-FOX <i>ET AL.</i> 2004
10	Bronzezeit Tarim Becken China (1980 v. Chr.); N=20	LI <i>ET AL.</i> 2010
11	Bronzezeit Lajia China (1900 v. Chr.); N=13	GAO <i>ET AL.</i> 2007
12	Neolithikum Nordwest Russland (5500 v. Chr.); N=11	DER SARKISSIAN <i>ET AL.</i> 2013
13	Bronzezeit Nordwest Russland (1500 v. Chr.); N=23	DER SARKISSIAN <i>ET AL.</i> 2013
14	Eneolithikum nördlicher Schwarzmeerraum (4500–3000 v. Chr.); N=10	WILDE <i>ET AL.</i> 2014
15	Yamnaya-Kultur; nördlicher Schwarzmeerraum (3000–2500 v. Chr.); N=28	WILDE <i>ET AL.</i> 2014
16	Katakombengrab-Kultur; nördlicher Schwarzmeerraum (2500–2000 v. Chr.); N=24	WILDE <i>ET AL.</i> 2014



## A5.6 Die phänotypischen Marker

Eine Übersichtstabelle mit den Ergebnissen erfolgreich analysierter Individuen für alle Marker findet sich in Teil B (s. Tab. B 49). Nachstehend sind die individuellen Ergebnisse für die phänotypischen Marker dargestellt, deren Frequenzen sich in europäischen und asiatischen Populationen unterscheiden (s. Tab. A17). Von den jeweils zwei *ADH* und *LCT* *loci* für Alkoholverträglichkeit und Laktasepersistenz wurde stellvertretend einer ausgewählt. Soweit vorhanden unterstützen sich die Ergebnisse der beiden *loci* und die Darstellung wäre somit redundant. Die Gene für die Hautpigmentierung sind hingegen alle in der Tabelle aufgeführt, da sich in diesem Fall der letztendliche Phänotyp, also das Erscheinungsbild des Individuums, aus der Summe der einzelnen Marker, die mit einer Pigmentierung der Haut assoziiert werden, zusammensetzt (STOKOWSKI *ET AL.* 2007).

**Tab. A 17 Ergebnisse zu phänotypischen Markern**

Blau = spezifisch für Asien; Grün = spezifisch für Europa; \*hell meint blau, grau oder grün; int. = intermediär

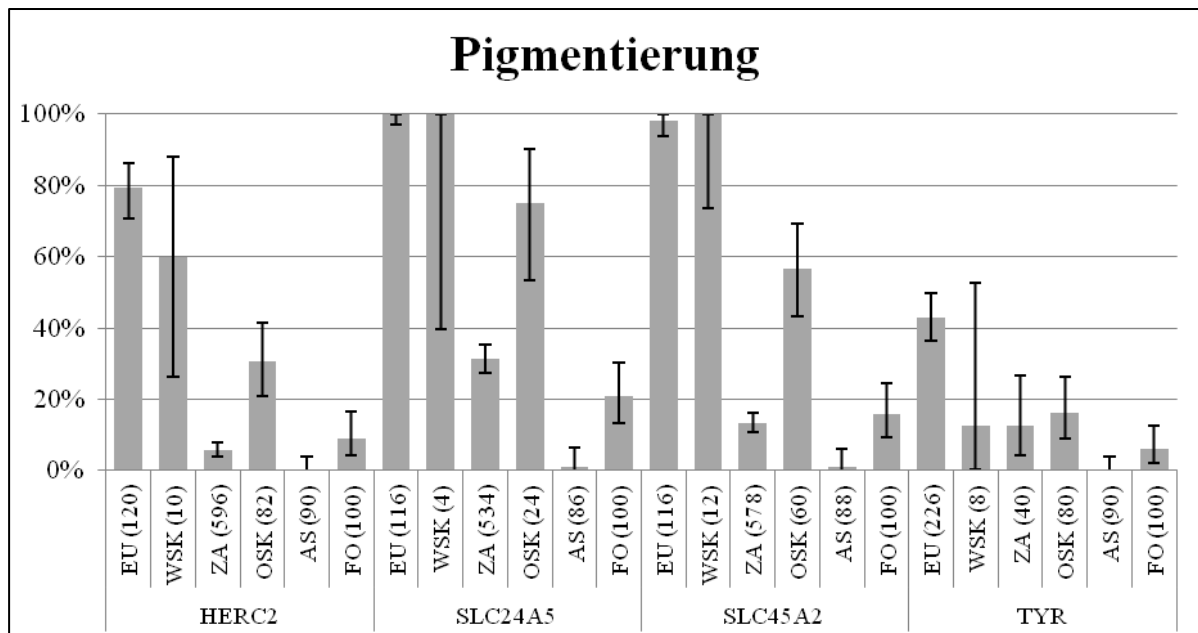
	Probe	ABCC11	Ohrschmalz	ADH1Ba	Schutz vor Alkoholismus	EDAR	Haardicke	HERC2	Augenfarbe	LCTa	verträgt Milch	SLC24A5	Pigmentierung	SLC45A2	Pigmentierung	TYR	Pigmentierung	AMELXY	Haplogruppe	
	anz.	C	feucht	C	nein	T	dick	A	braun	C	nein	G	dunkel	C	dunkel	C	dunkel			
	abg	T	trocken	T	ja	C	dünn	G	hell*	T	ja	A	hell	G	hell	A	hell			
Tuva – Aržan 2	A2							A	braun	C	nein									
	A4	T	trocken					A	braun	C	nein					C	dunkel	männlich		
	A6	C	feucht	C	nein					C	nein					C	dunkel	weiblich		
	A8									C	nein					C	dunkel			
	A9	C/T	feucht	C	nein	C	dick	A	braun	C/T	ja					C	dunkel	weiblich		
	A11	C/T	feucht	C	nein			A	braun	C	nein								männlich	
	A14	C/T	feucht	C	nein	T/C	int.	A/G	braun	C	nein	G	dunkel			C	dunkel	männlich		
	A21	T	trocken	C	nein	C	dick	A	braun	C	nein	G/A	int.			C	dunkel	männlich		
Barsučij Log	BL1	C	feucht	C	nein			A/G	braun	C/T	ja			G	hell	C	dunkel	männlich		
	BL2	C	feucht	C	nein					C/T	ja			G	hell	C/A	int.			
	BL3	C	feucht	C	nein			G	hell*	C	nein			G	hell	C/A	int.	männlich		
	BL4	C/T	feucht	C	nein			G	hell*	C/T	ja			G	hell	C/A	int.	männlich		
	BL5													G	hell	C	dunkel			
	BL6	C/T	feucht	C	nein			A/G	braun	C	nein			C/G	int.			männlich		
Chunya	U1_2	T	trocken					A	braun	C	nein					A	hell	weiblich		
	U2_2	C	feucht	C	nein			A	braun	C	nein			C/G	int.	C	dunkel			
	U4_1	C	feucht	C	nein			A/G	braun	C	nein			C/G	int.	C/A	int.			
	B1_1	T	trocken			C	dick	A	braun	C	nein			G	hell	C	dunkel			
	J12_8	C/T	feucht	C	nein	T	dünn	A	braun	C	nein	A	hell	C/G	int.	C	dunkel	weiblich		
	J22_1	C/T	feucht	C	nein	C	dick	A/G	braun	C	nein	A	hell	C/G	int.	C	dunkel	männlich		
	Ala2	C/T	feucht	C	nein	T/C	int.	A/G	braun	C	nein			G	hell	C/A	int.	weiblich		
BS1	C/T	feucht	C	nein			A	braun	C/T	ja	A	hell	C/G	int.	C/A	int.	männlich			
Ukok	Ak1_1	C/T	feucht	C	nein	T	dünn	A/G	braun	C	nein	A	hell	C	dunkel	C	dunkel	männlich		
	K1			C	nein			A	braun	C	nein			C/G	int.	C	dunkel			
	Mo1	C	feucht	C	nein					C	nein			G	hell	C	dunkel	weiblich		
VK1_1	C/T	feucht	C	nein			A	braun	C/T	ja			C	dunkel	C	dunkel	weiblich			

	Probe	ABCC11	Ohrenschmalz	ADH1Ba	Schutz vor Alkoholismus	EDAR	Haardicke	HERC2	Augenfarbe	LCTa	verträgt Milch	SLC24A5	Pigmentierung	SLC45A2	Pigmentierung	TYR	Pigmentierung	AMELXY	Haplogruppe	
	anz.	C	feucht	C	nein	T	dick	A	braun	C	nein	G	dunkel	C	dunkel	C	dunkel			
	abg	T	trocken	T	ja	C	dünn	G	hell*	T	ja	A	hell	G	hell	A	hell			
Kasachisches Altai	Be1	C/T	feucht	C	nein	T/C	int.	A	braun	C	nein	G/A	int.			C	dunkel			
	Be4	C/T	feucht	C	nein			A/G	braun							C	dunkel			
	Be6	C	feucht					A	braun	C	nein					C	dunkel			
	Be8							A	braun					C	dunkel	A	hell			
	Be9	T	trocken			T/C	int.	A/G	braun	C	nein	A	hell	C	dunkel	C/A	int.	weiblich		
	Be11	T	trocken							C	nein									
	Be12	C/T	feucht	C	nein	T	dünn	A/G	braun	C	nein	A	hell	C/G	int.	C	dunkel			
	Be14	C	feucht	C	nein			A	braun	C	nein			C/G	int.	C	dunkel	männlich		
Ostkasachstan	Is1	C	feucht	C	nein			A/G	braun	C	nein	A	hell	C/G	int.	C	dunkel	männlich		
	Is2	C	feucht	C	nein			A/G	braun	C	nein			G	hell	C	dunkel	männlich		
	Is4	C	feucht	C	nein			A/G	braun	C	nein			G	hell	C	dunkel	männlich		
	Ze2	C/T	feucht	C	nein			A	braun	C	nein			C/G	int.	C	dunkel			
	Ze3	C	feucht	C	nein			G	hell*	C	nein			G	hell	C	dunkel	männlich		
	Ze4			C	nein			A	braun	C	nein			C	dunkel	C	dunkel			
	Ze5	T	trocken					A	braun			G/A	int.	C	dunkel	C	dunkel	männlich		
	Ze6	C/T	feucht	C/T				A/G	braun					C/G	int.			weiblich		
	Ze7	C/T	feucht	C	nein			A/G	braun	C/T	ja			C	dunkel	C/A	int.	weiblich		
	Ze8	C/T	feucht	C	nein			A/G	braun	C	nein					C	dunkel	weiblich		
Ze9	C	feucht	C	nein			A/G	braun	C	nein							weiblich			
Westl. Skythen	Kol1	C	feucht	C	nein			A/G	braun	C	nein			G	hell	C/A	int.	weiblich		
	Kol4			C	nein			A/G	braun	C	nein			G	hell	C	dunkel			
	Kol5	C	feucht					G	hell*					G	hell					
	Nov5	C	feucht	C	nein			A/G	braun			A	hell	G	hell	C	dunkel	weiblich		
Nov7	C	feucht	C	nein			A/G	braun	C	nein	A	hell	G	hell	C	dunkel	männlich			
Nov8	C	feucht							C	nein			G	hell						

Für die skythischen Populationen ermittelte Frequenzen der abgeleiteten (selektierten) Allele ausgewählter *loci* sind in nachstehen Balkendiagrammen jeweils den Frequenzen der europäischen und asiatischen und soweit vorhanden zentralasiatischen und fernöstlichen Vergleichspopulationen (s. Kapitel A 4.3) gegenübergestellt. Die Frequenzen sind mit binomialen Konfidenzintervallen dargestellt. Für die Diagramme wurden die Marker, je nach ihrer Funktion, zu folgenden Gruppen zusammengefasst:

- 1) Pigmentierungs-*loci* (s. Abb. A 6),
- 2) Marker, die mit dem Calciumhaushalt und der Vitamin D Synthese assoziiert werden (s. Abb. A 7),
- 3) Gene, die für Temperaturanpassung und den Wasserhaushalt verantwortlich gemacht werden (s. Abb. A 8) und
- 4) Marker, die einen Einfluss auf die Verstoffwechslung von Xenobiotika haben (s. Abb. A 9).

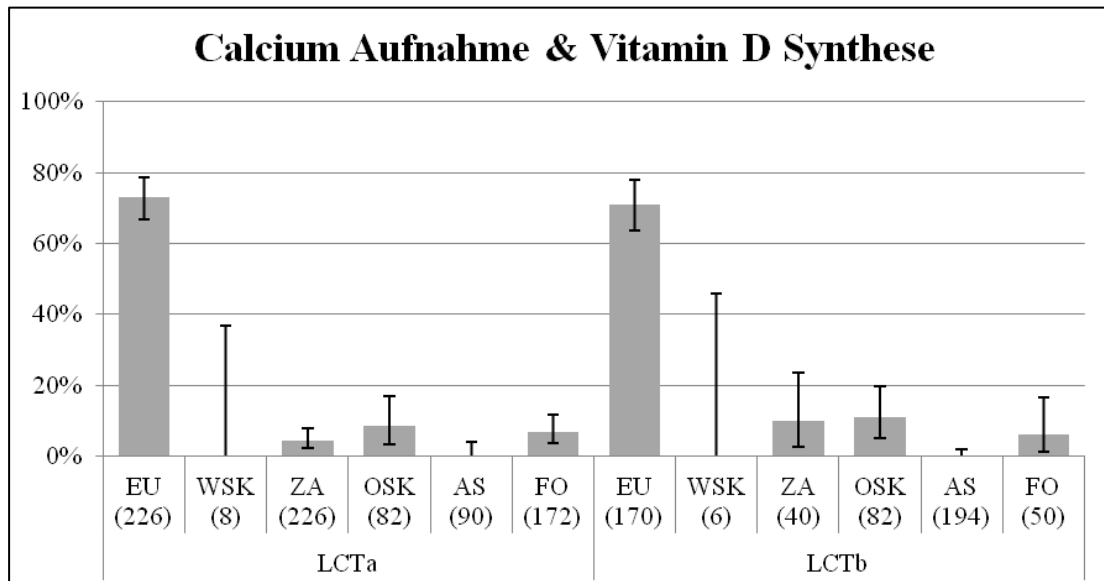
Die Ergebnisse für die Ost- und für die West-Skythen wurden getrennt aufgeführt, wobei WSK die Westskythen, OSK die östliche Skythengruppe, EU Populationen europäischen Ursprungs, ZA Zentralasiaten, AS Ostasiaten und FO Populationen aus Fernost bezeichnet. Die Nummer in Klammern gibt die Gesamtallelzahl wieder.



**Abb. A 6** Frequenzen der abgeleiteten Allele der Pigmentierungsgene

Das selektierte Allel des *HERC2* locus korreliert mit der Ausprägung einer hellen Augenfarbe und die selektierten Allele der drei weiteren Gene sind mit einer hellen Pigmentierung der Haut assoziiert. EU=Europäer, WSK=West-Skythen, ZA=Zentralasiaten, OSK=Ost-Skythen, AS=Han Chinesen, FO=fernöstliche Populationen.

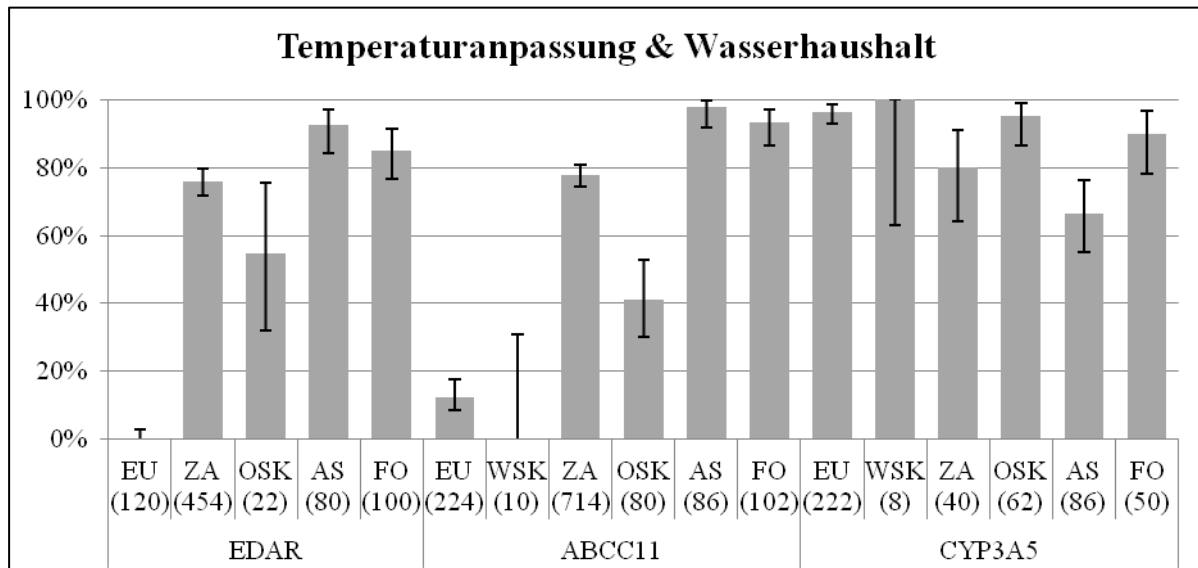
Die Ergebnisse für die Pigmentierungs-*loci* zeigen sehr anschaulich, dass die Allelfrequenzen der West-Skythen eher mit denen der europäischen Vergleichspopulation korrelieren und die östlichen Skythen den intermediären Status zwischen den Allelfrequenzen in modernen asiatischen und europäischen Populationen aufweisen.



**Abb. A 7 Frequenzen der abgeleiteten Allele zweier *loci*, die mit Laktasepersistenz assoziiert sind**

Die abgeleiteten Allele der *LCT loci* geben den Anteil an Individuen an, die auch im Erwachsenenalter in der Lage sind Milchzucker zu verarbeiten, womit die Möglichkeit einer erhöhten Calciumaufnahme einhergeht. EU=Europäer, WSK=West-Skythen, ZA=Zentralasiaten, OSK=Ost-Skythen, AS=Han Chinesen, FO=fernöstliche Populationen.

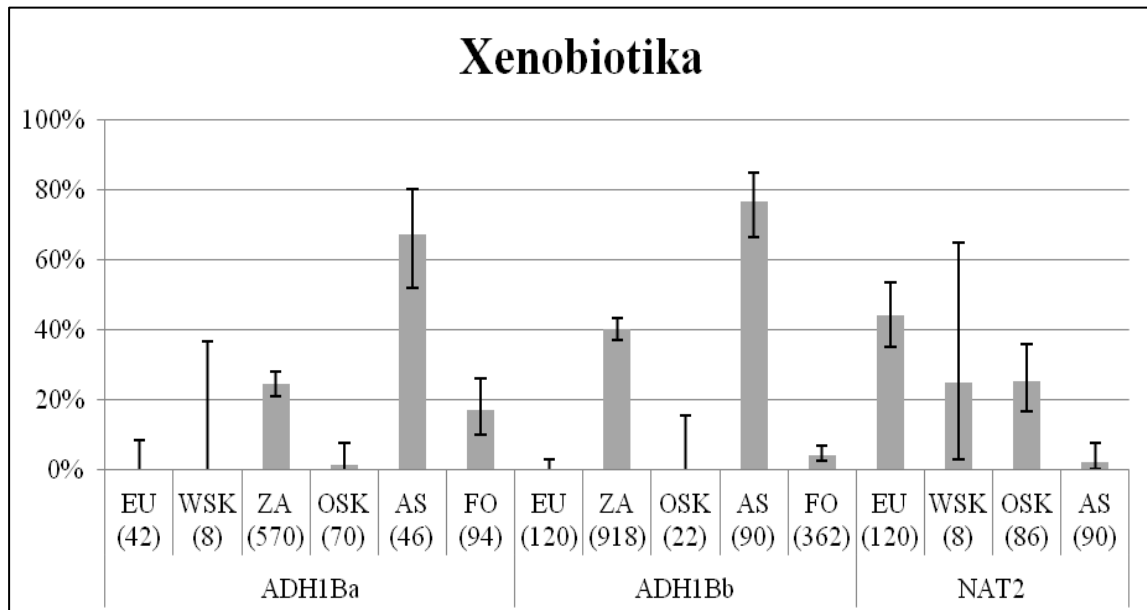
Bei den Laktase-*loci* tragen die West-Skythen jeweils ausschließlich die ancestralen C- bzw. G-Allele, die mit einer Milchunverträglichkeit assoziiert werden und in einer Frequenz von 100% in den ostasiatischen Populationen vorkommen. Für ungefähr 10% der Ostskythen kann eine Milchverträglichkeit angenommen werden, wodurch sie sich in der Variationsbreite der zentralasiatischen und fernöstlichen Populationen bewegen.



**Abb. A 8** Frequenzen der abgeleiteten Allele von Genen, die mit einer Anpassung an Temperatur & Wasservorkommen assoziiert werden

Die abgeleiteten Allele von *EDAR* und *ABCC11* stehen in Ostasien und die von *CYP3A5* außerhalb Afrikas unter Selektion. EU=Europäer, WSK=West-Skythen, ZA=Zentralasiaten, OSK=Ost-Skythen, AS=Han Chinesen, FO=fernöstliche Populationen.

Für den *locus EDAR*, der mit der Haardicke und schaufelförmigen Vorderzähnen assoziiert ist (FUJIMOTO *ET AL.* 2008, KIMURA *ET AL.* 2009) und *ABCC11*, dessen selektiertes Allel zu trockenem Ohrenschmalz und vermindertem Schweißfluss führt (TOYODA *ET AL.* 2009), liegen die Frequenzen der östlichen Skythenpopulation zwischen Europa und Ostasien. Für den Marker *CYP3A* zeigen die West- und die Ost-Skythen keinen Unterschied zu europäischen oder zentralasiatischen Populationen.



**Abb. A 9** Frequenzen der abgeleiteten Allele von Genen, die mit der Umsetzung von Xenobiotika assoziiert sind.

Von dem *locus NAT2* wird vermutet, dass er außerhalb Afrikas unter Selektion steht. *ADH* unterliegt einer positiven Selektion in ostasiatischen Bevölkerungen und wird mit einem Schutz vor Alkoholismus assoziiert. EU=Europäer, WSK=West-Skythen, ZA=Zentralasiaten, OSK=Ost-Skythen, AS=Han Chinesen, FO=fernöstliche Populationen.

Die Ostskythen weisen fast ausschließlich die ancestralen C- bzw. G-Allele an den *ADH-loci* auf und entsprechen somit eher dem europäischen als ostasiatischen Phänotyp. Für den Marker *NAT2* nehmen die Ostskythen die intermediäre Position zwischen Europa und Ostasien ein. Zentralasiatische Vergleichsdaten existieren zu diesem *locus* nicht.

## Hardy-Weinberg-Gleichgewicht

Zur Überprüfung ob sich die skythische Population für die untersuchten *loci* im Hardy-Weinberg-Gleichgewicht befindet, wurde anhand der beobachteten und erwarteten Anzahl der Genotypen ein *Chi-squared* ( $\chi^2$ ) Test durchgeführt. Der  $\chi^2$ -Test gibt die Wahrscheinlichkeit, dass die Differenz zwischen beobachteter und erwarteter Genotypenanzahl durch Zufall erreicht werden kann, wenn sich die Population im Hardy-Weinberg-Gleichgewicht befindet. Die ermittelten Werte für  $\chi^2$  werden mit der  $\chi^2$ -Verteilung des ersten Freiheitsgrades verglichen. Der Freiheitsgrad ist hier abhängig von der Anzahl möglicher Genotypen und der

Anzahl an Parametern, die berechnet werden. Je größer der Wert für  $\chi^2$  desto geringer die Wahrscheinlichkeit, dass die Differenz zwischen beobachteter und erwarteter Genotypenanzahl zufällig ist und dadurch desto unwahrscheinlicher, dass die Genotypen einer Population im Hardy-Weinberg-Gleichgewicht entstammen (HAMILTON 2009, S. 24–25). Für eine Stichprobenanzahl unter fünf Individuen pro Genotyp ist der  $\chi^2$ -Test nicht mehr exakt. Ist die erwartete Genotypenanzahl = 0 kann der  $\chi^2$ -Wert aufgrund einer Division durch 0 nicht berechnet werden (s. Tab. A 18). Zusätzlich wurde der HWExact-Test durchgeführt (Parameter alternative = „less“), der im HardyWeinberg-Paket für R implementiert ist (GRAFFELMAN 2013, R\_CORE\_TEAM 2013). Der Test überprüft, ob weniger Heterozygote vorliegen als erwartet und somit an den ausgewählten *loci* ein signifikanter Homozygoten-Überschuss vorliegt ( $p$ -value < 0,05).

**Tab. A 18** Ergebnisse des  $\chi^2$ -Test und des HWExact-Tests

	<i>N</i> (Ind)	<i>N</i> (hom. anz.)	<i>N</i> (het.)	<i>N</i> (hom. abg.)	$\chi^2$	Zufallswahrscheinlichkeit ( <i>df</i> =1)	HWExact-Test <i>p</i> -value
HERC2	41	19	19	3	0,3572	0,5501	0,8132593
SLC24A5	12	1	4	7	0,1481	0,7003	0,5606407
SLC45A2	30	7	12	11	1,0325	0,3096	0,2285627
TYR	40	29	9	2	1,2022	0,2729	0,2515027
LCTa	41	34	7	0	0,3572	0,5501	1,0000000
LCTb	41	32	9	0	0,6232	0,4299	1,0000000
CYP3A4	16	0	0	16	#DIV/0!	#DIV/0!	1,0000000
SLC12A3	21	21	0	0	#DIV/0!	#DIV/0!	1,0000000
TRPV6b	39	0	0	39	#DIV/0!	#DIV/0!	1,0000000
TRPV6c	27	0	1	26	0,0096	0,9219	1,0000000
EDAR	11	3	4	4	0,7822	0,3765	0,3293641
ABCC11	40	14	19	7	0,0160	0,8994	0,5464636
CYP3A5	31	0	3	28	0,0801	0,7771	1,0000000
ADH1Ba	35	34	1	0	0,0074	0,9317	1,0000000
ADH1Bb	11	11	0	0	#DIV/0!	#DIV/0!	1,0000000
NAT2	43	25	14	4	0,9027	0,3421	0,2567533
ABCB1a	36	15	15	6	0,4444	0,5050	0,3429059
ABCB1b	33	12	19	2	2,3693	0,1237	0,9721519
ABCB1c	39	13	14	12	3,0922	0,0787	0,0626307
CASP12	11	0	0	11	#DIV/0!	#DIV/0!	1,0000000

## A6 Diskussion

Eine detaillierte Diskussion der verwendeten Methoden findet sich in Teil B

### A6.1 Anmerkungen zur verwendeten Stichprobe und den Markern

Die Probenauswahl umfasste Individuen aus dem Minusinsker Becken, Tuva, dem Altai-Gebirge, Ostkasachstan, Südkasachstan, dem südlichen Uralgebirge, dem Oblast Voronezh in Südwest-Russland und dem nördlichen Kaukasus. Die Fundstellen decken somit einen großen Bereich des Verbreitungsgebietes der eisenzeitlichen Reiternomaden ab. Von den ausgewählten Proben gelang es, für insgesamt 97 skythenzeitliche Individuen von 20 verschiedenen Fundplätzen, genetische Daten zu produzieren. Mit 78 Individuen verteilt auf 17 Fundplätze stammt der Großteil der Proben aus dem Bereich des Altai-Gebirges. Die restlichen 19 Proben verteilen sich auf drei Fundplätze in der westeurasischen Steppe. Trotz des insgesamt guten Erhaltungszustandes der Proben mussten einige aufgrund zu geringen DNA-Gehaltes von den Analysen ausgeschlossen werden. Dazu zählen unter anderem die Proben des Fundplatzes Cirikrabad in Südkasachstan. Hier standen leider nur drei Proben zur Verfügung, mit denen, als Stellvertreter der Saken, eine zusätzliche Kulturgruppe für den Datensatz verlorengegangen ist. Es gilt jedoch zu bedenken, dass selbst bei einer positiven Analyse der Individuen die Anzahl zu gering gewesen wäre, um sie in die populationsgenetischen Studien mit einzubeziehen. Die Ergebnisse hätten jedoch einen ersten Hinweis auf mögliche Verbindungen zu den anderen Populationen liefern und als Basis für zusätzliche Analysen aus diesem Bereich dienen können.

Von dem Fundplatz Novozavedennoe aus dem nördlichen Kaukasus standen zehn Individuen für die Analysen zur Verfügung. Für vier dieser Individuen konnte eine Haplogruppe bestimmt und für drei die komplette HVR1 Sequenz reproduziert werden. Der Fundplatz datiert in die initiale Skythenzeit, aus der im Westen keine weiteren Proben vorlagen und keine Literaturdaten zur Verfügung standen um die Stichprobenanzahl zu erhöhen. Auch für den Fundplatz Kolbino aus dem Oblast Voronezh konnten nur vier Individuen reproduziert werden. Diese konnten jedoch mit den Literaturdaten aus dem Rostov-Don-Gebiet (DER SARKISSIAN 2011) zusammengefasst werden. Die westlichen Skythen der Fundplätze Novozavedennoe und Kolbino zeigen alle westeurasische Haplogruppen. Ohne die Einbeziehung der Literaturdaten, die zu 31% osteurasische Haplogruppen aufweisen wäre



eine Verbindung zu den östlichen Populationen weit weniger offensichtlich gewesen bzw. hätte aufgrund der Datenlage verneint werden müssen. Dies verdeutlicht, wie wichtig eine gewisse Mindestanzahl an Individuen für die Untersuchungen ist, um die tatsächliche Variabilität der Daten annähernd erfassen zu können. Für die initialen Skythen konnten also nur drei HVR1 Sequenzen für die Berechnungen der  $F_{ST}$ -Werte verwendet werden womit diese Stichprobe zu klein für aussagekräftige Interpretationen ist.

Die Tagar-Kultur wurde ebenfalls nur durch eine eingeschränkte Stichprobenanzahl repräsentiert. Für die vorliegende Arbeit wurden sechs Individuen der Tagar-Kultur aus dem 5. Jahrhundert v. Chr. untersucht, die alle von dem Fundplatz Barsučij Log im Bereich des Minusinsker Beckens stammen. Drei der sechs Individuen weisen den gleichen Haplotyp auf, was auf verwandtschaftliche Beziehungen hindeuten könnte. Aufgrund der Tatsache, dass es sich hier lediglich um sechs Vertreter einer Kulturstufe handelt, von denen die Hälfte womöglich verwandt ist, bilden diese Individuen keine repräsentative Stichprobe für die Tagar-Kultur und sollten mit Vorsicht interpretiert werden. Um den Datensatz zu erweitern wurden zehn eisenzeitlichen Proben aus Chakassien (KEYSER *ET AL.* 2009) in die Berechnungen mit einbezogen. Diese stammen aus dem 8. Jahrhundert v. Chr. bis zum 1. Jahrhundert n. Chr. und werden somit der Tagar-Kultur bzw. der späteren Tes-Phase zugeordnet. Der Anteil an ostasiatischen Linien liegt bei den Tagar/Tes Proben von Keyser bei lediglich 40%. Gemeinsam mit den hier analysierten Barsučij Log Individuen liegt der Anteil ostasiatischer Linien bei 50%. Die Schwankungen der Anteile ostasiatischer Linien könnten zum einen auf die geringe Stichprobenanzahl zurückzuführen sein, zum anderen könnten auch kulturelle Einflüsse eine Rolle spielen. Die chakassischen Proben decken eine Zeitspanne von 800 Jahren und zwei Kulturstufen ab, wobei leider nicht individuell zugeordnet werden kann, welches Individuum in welcher Epoche bestattet wurde. Somit erweitern die Literaturdaten den Probensatz zwar auf eine größere Stichprobenanzahl, aufgrund der zeitlichen und kulturellen Diversität ist jedoch auch der erweiterte Datensatz nur mit Vorbehalten zu interpretieren. Zum Vergleich sind in allen Analyseschritten die Tagar-Individuen des Fundplatzes Barsučij Log gesondert aufgeführt. Generell ist die Beprobung mehrerer Bestattungen und mehrerer Fundplätze gleicher Datierung sinnvoll, um eine Kulturstufe möglichst genau und objektiv darstellen zu können.

Die Untersuchung der nukleären Marker wurde nicht für alle Individuen einer Kulturstufe durchgeführt. Es konnten weiterhin nicht für jedes Individuum alle Marker erfolgreich typisiert oder ausreichend reproduziert werden. Die ermittelten Genotypen beruhen auf

mindestens zwei 454-Sequenzierungen aus unabhängigen Multiplex-PCRs, wobei nur diejenigen als gesichert angesehen werden, die in mindestens drei Sequenzierungen konsistente Ergebnisse zeigen. Die Allelfrequenzen sollten deshalb als vorläufige Ergebnisse angesehen und unter Vorbehalt interpretiert werden. Für welches Individuum welche Marker typisiert werden konnten und der Stand der Reproduktion ist Tab. B 49 zu entnehmen. Die Skythen, in ihrer Stellung als mobile zentralasiatische Gemeinschaft unbekannter Herkunft, sollten vor allem im Hinblick auf die Marker untersucht werden, deren Frequenzen sich zwischen europäischen und asiatischen Populationen stark unterscheiden. Aus diesem Grund sind einige der ausgewählten Marker irrelevant für diese Studie, da die Frequenzen zwischen Europäern und Asiaten kaum, oder gar nicht differieren. In den meisten Fällen ist dies auf ein Selektionsereignis vor der Trennung der beiden Populationen zurückzuführen, z.B. *CYP3A4* (THOMPSON *ET AL.* 2004), oder *CASP12* (XUE *ET AL.* 2006). Die Untersuchung dieser Marker hat gezeigt, dass sich die Frequenzen der skythischen Population, moderner Europäer und Ostasiaten nicht voneinander unterscheiden (s. Anhang Tab. 22), weshalb sich der Ergebnisteil und die Diskussion auf die Marker konzentrieren, deren Allelfrequenzen Unterschiede zwischen den untersuchten Populationen aufweisen.

Für die populationsgenetischen Analysen wurde die mitochondriale DNA verwendet. Dieser Marker wird über die mütterliche Linie vererbt und repräsentiert somit lediglich die weibliche Populationshistorie. Generell kann die Untersuchung eines Markers nur einen begrenzten Ausschnitt der genealogischen Information darstellen. Da es sich aufgrund des Vererbungsmodus bei mitochondrialen Haplogruppen um sehr alte Linien handeln kann, können diese in einem Individuum keine Information über rezentere demographische Prozesse beinhalten. Zuverlässige Aussagen über demographische Prozesse erfordern statistische Methoden, die mehrere *loci* gleichzeitig berücksichtigen (ROSENBERG & NORDBORG 2002). Die Beschränkung auf die mitochondrialen DNA ermöglicht es auf der anderen Seite eine große Menge an Individuen unterschiedlicher Populationen zu untersuchen. Die genomische DNA für die gleiche Anzahl an Individuen zu untersuchen, die in dieser Studie verwendet wurde, ist momentan noch nicht möglich. Die großen Datenmengen genomischer Analysen können weiterhin, mit der aktuell zur Verfügung stehenden Rechenleistungen moderner Computer, nicht für die Simulation demographischer Modelle verarbeitet werden. Die Untersuchung mitochondrialer DNA ermöglicht Einblicke in einen Teil der demographischen Prozesse einer Population.

## A6.2 Die mitochondriale Variabilität skythischer Populationen

Die Analyse der mitochondrialen DNA hat gezeigt, dass die skythischen Populationen Haplogruppen aufweisen, die in heutigen Populationen den Europäern bzw. Ostasiaten zugeordnet würden, jedoch auch im heutigen Zentralasien zu finden sind (COMAS *ET AL.* 1998, COMAS *ET AL.* 2004, QUINTANA-MURCI *ET AL.* 2004). Für die vergleichenden Betrachtungen und die biostatistischen Analysen wurden die skythischen Individuen nach chronologischen und geographischen Gesichtspunkten in Gruppen aufgeteilt, die jeweils mit einer distinkten Kulturstufe korrelieren. Weiterhin wurden in alle Analysen verfügbare Daten aus der Literatur miteinbezogen. Um die Verteilung west- und osteurasischer Linien in den Populationen besser vergleichen zu können sind diese farblich markiert in Abb. A6 wiedergegeben. Insgesamt tragen die östlichen Skythen ungefähr zu gleichen Teilen west- und osteurasische Linien. Die ältesten Individuen der Zevakino-Čilikta Stufe zeigen mit 54,5% etwas mehr westliche Linien als die nachfolgenden Gruppen. Den größten Anteil an osteurasischen Linien, mit jeweils 66,7%, zeigen die Individuen der Fundplätze Aržan 2 des 7.–6. Jahrhunderts v. Chr. und der Fundplatz Barsučij Log aus dem 5. Jahrhundert v. Chr. Die jüngste Gruppe bilden die Individuen der Pazyryk-Kultur des 4.–3. Jahrhunderts v. Chr. – hier zeigen mit 52,1% etwas mehr als die Hälfte der Individuen östliche Linien.

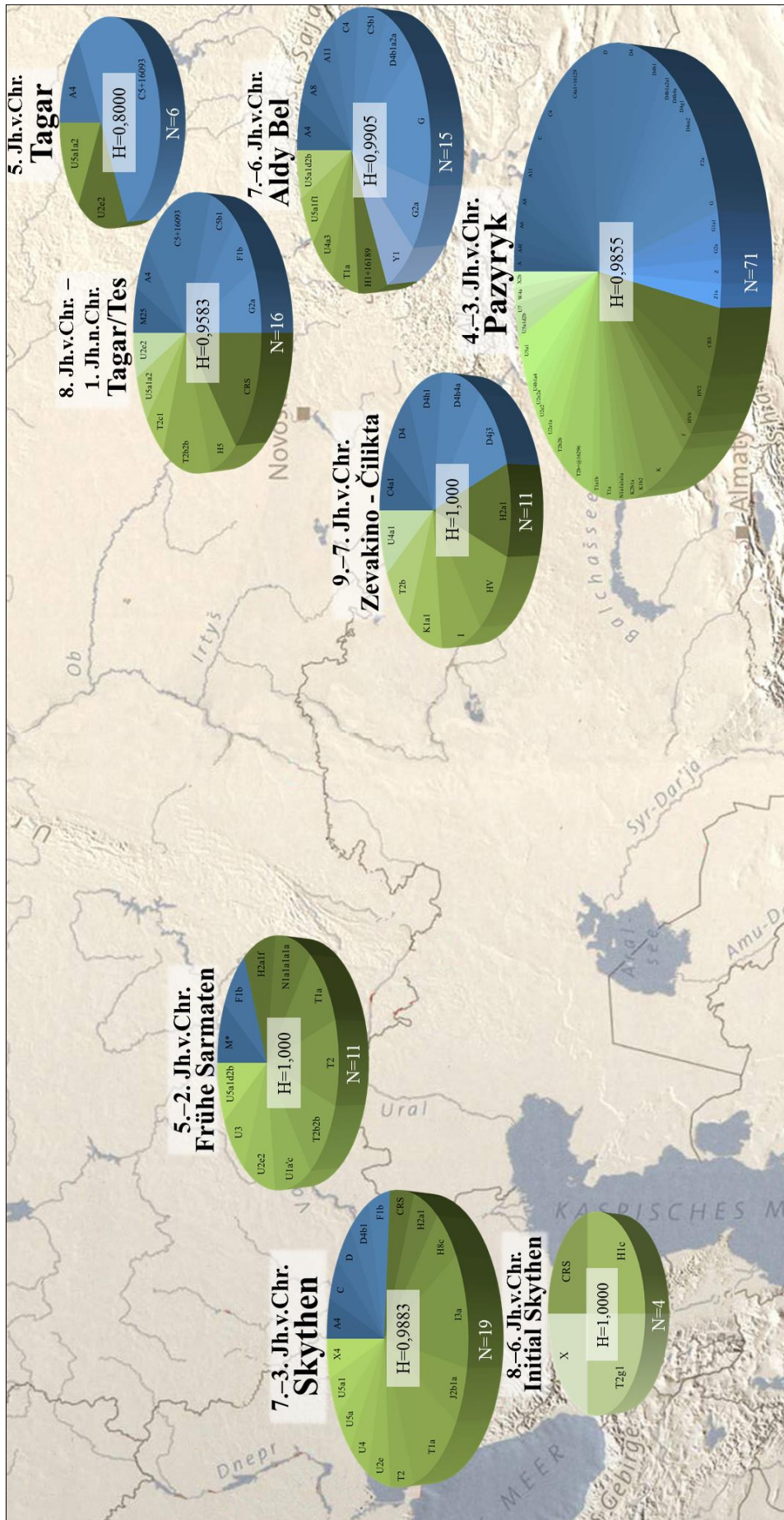
Der höhere Anteil an westeurasischen Linien der Zevakino-Čilikta Kultur könnte die Übergangsphase einer durch westliche Linien geprägten späten Bronzezeit markieren. Von der bronzezeitlichen Andronovo-Kultur wird ausgegangen, dass sie sich von West-Eurasien nach Süden und Osten, bis zum Bereich des Altai-Gebirges ausbreitete (KOHL 2007). Studien der Bronzezeit in Kasachstan und Chakassien zeigen nahezu ausschließlich europäische Haplogruppen (KEYSER *ET AL.* 2009, LALUEZA-FOX *ET AL.* 2004) und auch im Bereich der westsibirischen Waldsteppe nimmt der Anteil an westeurasischen Linien gegen Ende der Bronzezeit stark zu (MOLODIN *ET AL.* 2012A). Der sprunghafte Anstieg asiatischer Linien in Ostkasachstan von 0 auf 45,5% könnte dabei entweder auf ein Beprobungsartefakt, aufgrund der geringen Anzahl bronzezeitlicher Individuen, oder eine westwärts gerichtete Gegenbewegung zu Beginn der Eisenzeit zurückzuführen sein. Der höhere Anteil asiatischer Linien in Tuva könnte ebenfalls das Resultat einer westwärts gerichteten Bewegung östlicher oder südöstlicher Gruppen sein, oder jedoch die tatsächliche Zusammensetzung der indigenen Bevölkerung widerspiegeln. Das gleiche gilt für die Pazyryk-Population, deren Haplogruppenverteilung auf Einflüsse aus Ost und West zurückgeführt werden könnte.

Solange die Verhältnisse im Altai-Gebirge in den vorhergehenden Epochen nicht erforscht wurden, lässt sich über den Ursprung der Populationen jedoch nur spekulieren. Die geringen genetischen Distanzen zwischen den östlichen Skythen-Populationen, die im nachfolgenden Kapitel genauer besprochen werden, sprechen für eine gemeinsame Bevölkerungshistorie.

Im Zuge dieser Studie wurden weiterhin zehn Individuen des initial-skythischen Fundplatzes Novozavedennoe und fünf skythische Individuen aus Kolbino untersucht. Von jeweils vier Individuen der beiden Fundplätze konnten die Haplogruppen bestimmt werden, die sich alle den westeurasischen Linien zuordnen lassen. Bezieht man jedoch die Daten der Dissertation von Clio Der Sarkissian mit ein, ergibt sich für die West-Skythen ein Anteil asiatischer Linien von 26,3% (DER SARKISSIAN 2011). Etwas weiter östlich, im Bereich südlich des Ural, liegt der früh-sarmatische Fundplatz Pokrovka. Von den elf erfolgreich typisierten Individuen zeigen zwei (18,2%) asiatische Linien.

Die Individuen der initialen Skythenzeit weisen alle europäische Haplotypen auf. Geht man von einer repräsentativen Stichprobe aus, würde dies gegen einen Einfluss östlicher Populationen auf die Entwicklung der skythischen Kultur im Westen sprechen – zumindest in der Anfangsphase. Um eine zuverlässige Aussage zu treffen ist jedoch die Untersuchung weiterer Individuen dieser frühesten Phase der skythischen Kultur erforderlich.

Die Haplogruppenzusammensetzung einer Population kann Hinweise auf ihr Ursprungsgebiet geben. Es werden häufig moderne DNA Daten verwendet um Rückschlüsse auf prähistorische Bevölkerungen zu ziehen. Um konkrete Aussagen treffen zu können werden jedoch Daten rezenter und alter Populationen aus den entsprechenden Regionen benötigt. Rezente Populationen aus dem Altai, Tuva und Chakassien zeigen mit 65,5–88,9% im Schnitt etwas höhere Anteile osteurasischer Linien (DERENKO *ET AL.* 2003) als die eisenzeitlichen Gruppen dieser Regionen. Moderne Bevölkerung SW-Russlands zeigen nur ungefähr 8% osteurasische Linien (MALYARCHUK *ET AL.* 2002) im Vergleich zu 20% in der Eisenzeit. Dies könnte darauf hindeuten, dass der Austausch zwischen West und Ost in der Eisenzeit höher war als heute, was zu einer nomadisch, bzw. seminomadisch lebenden Bevölkerung passen würde. Werden die Menschen sesshaft gehen durch genetische Drift seltenere Linien verloren, oder das Verhältnis westlicher und östlicher Linien verschiebt sich durch den Zustrom neuer Linien aus dem Westen bzw. Osten.



**Abb. A 10 Haplogruppenverteilung der skythischen Populationen**

blau für ostasiatische und grün für europäische Linien. H = Haplotypendiversität (max.=1); N = Probenanzahl - jeweils mit Literaturdaten, außer bei der Tagar-Kultur.

### A6.3 Die Populationsgenetik der skythischen Kulturgruppen

Die meisten der Populationen, die im Zuge dieser Arbeit untersucht wurden, zeigten eine sehr hohe Haplotypendiversität ( $H$ ) zwischen 0,95 und 1,0 (s. Abb. A 10), wie sie auch in modernen Bevölkerungen Zentralasiens zu finden ist (COMAS *ET AL.* 1998, DERENKO *ET AL.* 2003). Eine hohe Haplotypendiversität kann auf eine hohe effektive Populationsgröße und/oder einen ständigen Zustrom neuer Linien hindeuten. Bemerkenswert ist, dass in einem Kurgan die gleiche Haplogruppe selten mehrfach auftaucht, wodurch es folglich nahezu auszuschließen ist, dass gemeinsam bestattete Individuen in einem direkten mütterlichen Verwandtschaftsverhältnis zueinander standen. Ausnahmen bilden der Fundplatz Barsučij Log im Minusinsker Becken ( $H=0,80$ ), wo von sechs Individuen eines Kurgans drei die gleiche Haplogruppe tragen und der Kurgan Aržan 2 aus Tuva. Hier zeigen zwei Individuen den gleichen Haplotypen, die Fürstin aus Grab 5 und ein männliches Individuum aus Grab 14. Den anthropologischen Untersuchungen nach ist das männliche Individuum 10–20 Jahre älter als die Fürstin und die Individuen wurden gleichzeitig bestattet (ČUGUNOV *ET AL.* 2010), es kann sich demnach nicht um Mutter und Sohn handeln. Es könnte sich um ein Geschwisterpaar handeln, oder eben eine weiter in der Vergangenheit liegende Verbindung darstellen. Dass zwei bis drei Individuen den gleichen Haplotypen tragen, findet sich auch an weiteren Fundplätzen, hier jedoch jeweils in unterschiedlichen Grabhügeln.

Die Untersuchung der genetischen Distanzen zwischen den verschiedenen skythenzeitlichen Populationen offenbart eine große Homogenität. Die Tatsache, dass die Zevakino-Čilikta-, die Aldy Bel- und die Pazyryk-Kultur keine genetische Distanz zueinander zeigen und alle drei Kulturgruppen nur sehr geringe  $F_{ST}$ -Werte (0,00504–0,01507) und somit geringe genetische Distanzen zur Tagar/Tes-Gruppe offenbaren, spricht für einen genetischen Austausch zwischen diesen Gruppen. Da die untersuchten Populationen zeitlich aufeinander folgen lässt sich weiterhin schließen, dass es sich um eine kontinuierliche Bevölkerungsentwicklung im 1. Jahrtausend v. Chr. im Altai-Gebirge handelt. Zu modernen Populationen vergleichbarer geographischer Lage zeigen die skythischen Gruppen aus dem Bereich des Altai-Gebirges geringe bis mittlere genetische Distanzen ( $F_{ST}$ -Werte: 0,01281–0,05384). Eine kontinuierliche Bevölkerungsentwicklung von der Eisenzeit bis heute wäre also denkbar. Die Analyse verschiedener demographischer Szenarien durch *Approximate Bayesian Computation* (ABC), die von Friso Palstra (CNRS, Paris) durchgeführt wurde, bestätigte die Hypothese, dass die skythischen Individuen aus dem Osten der eurasischen Steppe einer einzigen Population

entstammen. Die Simulation zeigte weiterhin Anzeichen dafür, dass die Ost-Skythen eine schwache, aber signifikante Expansion im Verlauf des 1. Jahrtausend v. Chr. durchmachten. In weiteren Simulationen hat sich gezeigt, dass eine kontinuierliche Bevölkerungsentwicklung von der Eisenzeit zu modernen zentralasiatischen Populationen, vor allem zu turksprachigen Stämmen angenommen werden kann (F. Palstra, persönliche Mitteilung).

Doch nicht nur die räumlich und zeitlich sehr diversen Gruppen des Altai-Gebirges und der angrenzenden Gebiete zeigen ein hohes Maß an genetischer Ähnlichkeit untereinander, sondern auch die skythischen Populationen des Westens. So zeigen die initialen Skythen, Skythen und frühen Sarmaten keine genetischen Distanzen zueinander, trotz der geographischen Distanz zwischen dem Rostov-Don-Gebiet und dem frühsarmatischen Fundplatz Pokrovka. Dies deutet darauf hin, dass die Skythen und die frühen Sarmaten einer gemeinsamen Population entstammen. Dass von einer Verdrängung der Skythen durch die Sarmaten im 2. Jahrtausend v. Chr. ausgegangen wird (SMIRNOV 1980) könnte mit den kulturellen Entwicklungen gegen Ende der Skythenzeit zusammenhängen. Eine Entwicklung, die in der nördlichen Peripherie Chinas ihren Anfang nahm und durch eine westwärts Bewegung hunnischer Verbände den Beginn der Hunno-Sarmatischen-Periode einläutete (PARZINGER 2006, S.711). Wie der Begriff schon nahelegt werden die Sarmaten als Teil dieser Bevölkerungsbewegung verstanden, wie auch die Wusun in Kasachstan und die Xiongnu (Hunnen) in Innerasien. Die frühen Sarmaten scheinen jedoch von diesen Entwicklungen noch wenig beeinflusst zu sein. Eine geringe genetische Distanz findet sich nicht nur zwischen den frühen Sarmaten und den Skythen, sondern auch zwischen diesen Gruppen und der modernen Vergleichspopulation aus Südwest-Russland ( $F_{ST}$ -Wert=0,01279). Die geringe genetische Distanz könnte auf eine kontinuierliche Bevölkerungsentwicklung in Südwest-Russland hindeuten. Die westlichen Skythen und die initiale Skythen-Gruppe zeigen zusätzlich keine genetischen Distanzen zu den östlichen Skythen-Populationen, was für einen Genfluss zwischen den Nomadengruppen im Osten und Westen der Eurasischen Steppe spricht. Im Gegensatz dazu weisen die modernen Populationen aus Tuva, dem Altai und Chakassien hohe genetische Distanzen zu der modernen Population aus Südwest-Russland auf ( $F_{ST}$ -Werte: 0,08136–1,5671). Die Ergebnisse deuten auf eine homogene Population durch Genfluss in der Eisenzeit hin und eine wesentlich stärkere Strukturierung der modernen Bevölkerungen in Ost und West (s. Abb. A 4).

Einzigste Außenseiter dieser genetisch einheitlich erscheinender Populationen sind zum einen die Individuen des tagarzeitlichen Fundplatzes Barsučij Log im Minusinsker Becken und zum

anderen die frühen Sarmaten aus Pokrovka im Westen. Die Individuen der Tagar-Kultur zeigen eine mittlere bis hohe genetische Distanz ( $F_{ST}$ -Werte von 0,03351–0,09589) zu allen anderen Skythen-Populationen, wobei nur die  $F_{ST}$ -Werte zwischen der Tagar-Population und den West-Skythen (0,07235) und der Tagar-Gruppe und den frühen Sarmaten (0,09589) das Signifikanzlevel von  $<0,05$  erreichen. Die Unterschiede der Individuen aus Barsučij Log zu den anderen Gruppen könnte zwar auf die geringe Stichprobenanzahl der Tagar-Population zurückzuführen sein; von Seiten der Archäologie wird jedoch angenommen, dass es sich bei der Tagar-Kultur um eine lokale Entwicklung aus der bronzezeitlichen Karasuk-Kultur handelt (LEONT'EV *ET AL.* 1996) und somit eine, vom Altai-Gebirge unabhängige Bevölkerungsentwicklung im Bereich des Minusinsker Beckens darstellt, was ebenfalls die genetische Distanz zu den anderen Populationen erklären könnte. Die frühen Sarmaten aus Pokrovka zeigen keine Unterschiede zu den anderen Gruppen im Westen und zur Tagar/Tes'-Population, offenbaren jedoch eine größere genetische Distanz als die Skythen zu den restlichen Gruppen im Osten ( $F_{ST}$ -Werte von 0,04020 bis 0,05125). Eine mögliche Erklärung wäre eingeschränkter Genfluss mit den skythischen Populationen des Ostens, ausgenommen die Tagar/Tes'-Population. Die geringe genetische Distanz der frühen Sarmaten zur Tagar/Tes'-Population könnte eventuell doch bereits auf die Einflüsse der Hunnen zurückzuführen sein, die sich gegen Ende der Skythenzeit von China nach Norden und Westen ausbreiten und sich gemeinsam mit Wusun und Sarmaten weiter Teile der Eurasischen Steppe bemächtigten (PARZINGER 2006, S.874). Vor allem die späte Tes'-Stufe zeigt Neuerungen, die auf einen Einfluss dieser Gruppen schließen lassen (PARZINGER 2006, S.818). Es könnte sich jedoch auch hier um ein Beprobungsartefakt handeln, oder das Resultat von Genfluss mit einer Population, die den West-Skythen näher steht als den Ost-Skythen.

Insgesamt überwiegt die genetische Ähnlichkeit aller untersuchten Gruppen im Vergleich zu den Unterschieden vereinzelter Populationen. Der Befund lässt darauf schließen, dass die kulturellen Ähnlichkeiten der skythischen Populationen das Resultat der Verbreitung einer hochmobilen Bevölkerung sind. Jede Population besitzt zusätzlich ihre kulturellen Eigenheiten, die sich nicht auf die Gesamtheit der hier als Skythen bezeichneten Bevölkerung übertragen lassen und ebenfalls nicht zwingend mit einer genetischen Verbindung einhergehen. So zeigen sich in der Pazyryk-Kultur deutliche Einflüsse aus der persisch-achämenidischen Kunst (POLOSMAN *ET AL.* 2006), die in den früheren Populationen noch nicht vorhanden sind. Die aufgrund der genetischen Homogenität postulierte Bevölkerungs-



kontinuität im Bereich des Altai-Gebirges lässt jedoch darauf schließen, dass die Verbindung der Pazyryk-Gruppe zum Bereich des heutigen Iran vorwiegend auf kulturellem Austausch basierte.

## A6.4 Die Struktur der Pazyryk-Population

Die genetische Homogenität der untersuchten Populationen ist überraschend, handelt es sich doch um klar definierte, räumlich und zeitlich getrennte Kulturgruppen. Obwohl alle diese Gruppen auch Gemeinsamkeiten aufweisen, wie die besondere Tierstil-Kunstform, Pferdegeschirr und Waffenbeigaben in den Begräbnissen, die als typisch für die Skythen des Nordschwarzmeerraumes angesehen wurden, wird die Existenz einer zugrundeliegenden skythischen Kultur kontrovers diskutiert. Die Verwendung von Skythen als Überbegriff für diese Gruppen ist umstritten (YABLONSKY 2000).

Zur Überprüfung, ob sich diese Homogenität auch innerhalb einer klar definierten Kulturgruppe findet, oder sich Substrukturen offenbaren, bietet sich eine Betrachtung der Individuen der Pazyryk-Kultur an. Von dieser Gruppe, die vom 4. bis zum 3. Jahrhundert im gesamten Altai-Gebirge verbreitet war, sind bisher die meisten genetischen Daten verfügbar. Durch die Einbeziehung der bereits veröffentlichten Sequenzen und der im Zuge dieser Arbeit analysierten Proben können 71 Individuen verteilt auf drei Länder untersucht werden. Die Berechnung der  $F_{ST}$ -Werte offenbart jedoch keine bzw. lediglich eine geringe genetische Distanz zwischen diesen geographischen Gruppen ( $F_{ST}$ -Werte: 0,00439–0,02187), wobei sich die kasachische und die mongolische Pazyryk-Gruppe jedoch deutlich näher stehen als beide zur russischen. Die russische Gruppe lässt sich, ebenfalls nach geographischen Gesichtspunkten, weiterhin in zwei Untergruppen aufteilen: das Ukok-Plateau und das Chuya-Tal. Hierbei zeigt sich, dass sich die erhöhte genetische Distanz auf das Chuya-Tal zu konzentrieren scheint. Die  $F_{ST}$ -Werte der Individuen des Chuya-Tals zu den anderen Bereichen des Altai reichen von einem geringen bis mittleren  $F_{ST}$ -Wert von 0,02979 zum Ukok-Plateau bis zu einem mittleren bis erhöhten  $F_{ST}$ -Wert von 0,05969 zur Pazyryk-Gruppe des kasachischen Altai. Zwischen dem Ukok-Plateau und den anderen Gruppen finden sich dagegen keine bzw. sehr geringe Unterschiede. Die wahrscheinlichste Erklärung ist, dass aufgrund der geographischen Abgeschiedenheit der Chuya-Gruppe die Mobilität und der Austausch zu den anderen Pazyryk-Gruppen eingeschränkt wurden.

## A6.5 Das Ursprungsgebiet der Skythen

Hinsichtlich des Ursprungsgebietes der Skythen bzw. der skythischen Kultur existieren verschiedene Hypothesen. Auf der Informationsbasis antiker Schriften und der fortschreitenden archäologischen Forschung, wurden zu verschiedenen Zeiten unterschiedliche Thesen als am wahrscheinlichsten erachtet. So wurde aufgrund der Berichte in den Historien des Herodot über das Volk der Skythen und den Anfängen der archäologischen Forschung im Bereich nördlich des Schwarzen Meeres, das Ursprungsgebiet der skythischen Kultur in diesem Bereich angesiedelt. Mit der Ausweitung der Grabungstätigkeiten in die Transuralsteppe und später in den Bereich des Altai-Gebirges, wurden auch ein zentralasiatischer Ursprung und die Entstehung der Kultur im Bereich des Altai-Gebirges diskutiert. Durch umfangreiche Radiokarbondatierungen wurde nachgewiesen, dass sich die ältesten Monumente, die sich mit der skythischen Kultur in Verbindung bringen lassen, im Osten der Eurasischen Steppe befinden (ALEKSEEV *ET AL.* 2001, ALEKSEEV *ET AL.* 2002). Des Weiteren wurden Artefakte mit Elementen der charakteristischen Tierstil-Kunst im Umkreis des Flusses Jenissei und im Gebiet des heutigen China bereits im 10. Jh. v. Chr. entdeckt (BASHILOV & YABLONSKY 2000, PARZINGER 2004). Die archäologischen Befunde deuten also auf einen östlichen Ursprung der skythischen Kultur hin. Es wurden jedoch nicht alle Monumente datiert und aufgrund möglicher Fundlücken oder kultureller Integration kann ein westlicher oder zentralasiatischer Ursprung nicht ausgeschlossen werden.

Die Frage, ob sich die kulturellen Artefakte, die dem skythischen Stil zugesprochen werden, durch Tauschhandel über die Eurasische Steppe verbreiteten oder von einer Migration der Träger dieser Kultur ausgegangen werden kann, wurde in den vorhergehenden Kapiteln beantwortet, bleibt die Frage nach der Richtung dieser Bevölkerungsbewegung.

Verbindungen lassen sich aufgrund kultureller Artefakte vom Altai-Gebirge zum Bereich des heutigen Iran ziehen, zeigen doch die Pazyryk Funde deutliche Einflüsse persisch-archämenidischer Kunst. Die Haplogruppe HV2 des Individuums Be12 aus Berel' findet sich heute fast ausschließlich im Bereich des Iran wieder (die Information wurde bereitgestellt von P. Forster aus der rootsforreal-Datenbank – [ROOTSFORREAL.COM](http://ROOTSFORREAL.COM) ). Die Präsenz der Haplogruppe HV2 im Iran soll hiermit erwähnt, jedoch nicht überinterpretiert werden, denn die Haplogruppenverteilung in modernen Populationen spiegelt nicht zwangsläufig die Verteilung in vorhergehenden Epochen wider. Da die Pazyryk-Kultur als Teil der

kontinuierlichen Bevölkerungsentwicklung im Bereich des Altai-Gebirges eng mit den vorhergehenden Epochen verknüpft ist, kann von einer persischen Ursprungspopulation abgesehen werden.

Um die genetische Zusammensetzung prähistorischer Bevölkerungen zu begreifen und ihren Ursprung zu lokalisieren ist die Betrachtung früherer Zeitstufen wesentlich wichtiger und informativer, als lediglich Rückschlüsse von modernen Bevölkerungen zu ziehen. Das Neolithikum vor allem aber die Bronzezeit in Zentralasien und auch den angrenzenden Gebieten spielt deshalb eine essentielle Rolle bei den Überlegungen zu möglichen Ursprungsgebieten eisenzeitlicher Nomadengruppen. Aus dem 2. Jahrtausend v. Chr. liegen zwei Studien aus China vor. Eine der Studien untersuchte Individuen des Dorfes Lajia am Gelben Fluss, die dort während einer Überschwemmung ums Leben kamen. Die mitochondrialen Linien dieser Individuen lassen sich alle den ostasiatischen Haplogruppen zuordnen, zeigen somit keinen europäischen Einfluss (GAO *ET AL.* 2007). Anders sieht es bei der Gruppe der zweiten Studie aus. Hier wurden Mumien untersucht die im Tarim Becken West-Chinas entdeckt wurde. Zwei von 20 untersuchten Individuen zeigen westeurasische Linien, womit der Fundplatz den frühesten Nachweis europäischer Linien in diesem Gebiet liefert (LI *ET AL.* 2010). Aus der Mongolei wurden drei bronzezeitliche Individuen untersucht, deren Haplotypen sich alle der ostasiatischen Haplogruppe D zuordnen lassen (GONZÁLEZ-RUIZ *ET AL.* 2012). Insgesamt weisen China und die Mongolei somit lediglich 5% europäische Linien in der Bronzezeit auf, womit diese Regionen als einzige Quelle skythischer Individuen ausscheiden. Lalueza-Fox und Kollegen haben in einer Studie von 2004 die mitochondriale DNA kasachischer Individuen aus der späten Bronzezeit untersucht. Es konnten acht Individuen erfolgreich analysiert werden, die sich auf ganz Kasachstan verteilen (LALUEZA-FOX *ET AL.* 2004). Bemerkenswert ist, dass hier nur europäische Linien gefunden wurden, liegt doch bereits in der frühen Eisenzeit Ostkasachstans der Anteil ostasiatischer Linien bei 45,5%. Abgesehen von einem möglichen *sampling bias* aufgrund der geringen Individuenanzahl wäre eine starke Migrationsbewegung von Ost nach West zu Beginn der Eisenzeit eine mögliche Erklärung für den plötzlichen Anstieg ostasiatischer Linien.

Eine detaillierte Studie zu den demographischen Prozessen in der westsibirischen Waldsteppe liefern die Studien von MOLODIN *ET AL.* 2012A und MOLODIN *ET AL.* 2012B. Hier wurden verschiedene Kulturgruppen von der frühen Bronzezeit bis zum Übergang der Bronze- zur Eisenzeit untersucht. Die Populationen der frühen und mittleren Bronzezeit weisen jeweils ca. 60% ostasiatische Linien auf, was den Vermutungen widerspricht, dass es erst mit Beginn der

Eisenzeit zu einer Zunahme osteurasischer Linien westlich des Altai-Gebirges kam (GONZÁLEZ-RUIZ *ET AL.* 2012). Bemerkenswert ist weiterhin, dass sich die europäischen Linien in der Zeit des 4. bis zum Ende des 3. Jahrtausend v. Chr. alle der Haplogruppe U zuordnen lassen. Im Speziellen finden sich Linien der Haplogruppen U2, U4 und U5a, die mittlerweile als charakteristisch für europäische Jäger-Sammler-Fischer Populationen gelten (BOLLONGINO *ET AL.* 2013, BRAMANTI *ET AL.* 2009). Die Haplogruppe U5a wurde auch in neolithischen bzw. bronzzeitlichen Jäger-Sammler- bzw. Fischer Populationen aus dem Bereich des Baikalsees entdeckt (MOODER *ET AL.* 2006, NASAB 2011). Die Analyse der mitochondrialen DNA eines frühen modernen Menschen aus dem russischen Kostenki ergab die Haplogruppe U2 (KRAUSE *ET AL.* 2010A). Die erwähnten Linien werden zwar verallgemeinernd den europäischen Haplogruppen zugeordnet, da sie in ostasiatischen Populationen heute nicht vorkommen. Diese Linien repräsentieren jedoch vielmehr einen alten eurasischen Genpool. Es ist anzunehmen, dass die bronzzeitliche Population Zentralasiens das genetische Substrat früher Jäger-Sammler Bevölkerungen in sich trägt.

Erst zu Beginn des 2. Jahrtausend v. Chr. treten Linien der Haplogruppe T hinzu. Weitere europäische Linien tauchen erst in der Übergangsphase von der Bronze zur Eisenzeit auf. Generell lässt sich ab der späten Bronzezeit eine Zunahme europäischer Linien feststellen. So zeigt der spätbronzezeitliche Fundplatz Stary Sad 50% westeurasische Linien. In der späten Irmen-Kultur, die in die Übergangsphase zur Eisenzeit datiert wird, tauchen zusätzlich neue Haplogruppen auf und es kommt zu einer direkt sprunghaften Zunahme europäischer Linien auf 93%. Die Autoren führen dies auf eine Zuwanderung aus dem Süden, also der kasachischen Steppe, bzw. dem Südwesten zurück. Gleichzeitig lässt sich am Fundplatz Chicha 1 jedoch auch der Einfluss nördlicher Kulturgruppen erkennen (s. auch PILIPENKO *ET AL.* 2008). Es wird vermutet, dass es aufgrund der klimatischen Veränderung zu Beginn der Eisenzeit zu einer Zuwanderung aus dem Norden kam, die sich jedoch nicht im Genpool der südsibirischen Bevölkerung nachweisen lässt.

Aufgrund der Anteile west- und osteurasischer Linien käme die spätbronzezeitliche Kultur der sibirischen Waldsteppe als Ursprung der östlichen Skythen in Frage – es finden sich allerdings europäische Haplogruppen im skythischen Genpool die sich in der späten Bronzezeit Westsibiriens noch nicht nachweisen lassen. Denkbar wäre, dass verschiedene Einflüsse zur Formation der skythischen Populationen beigetragen haben. So scheint es eine ostwärts gerichtete Bewegung europäisch geprägter Populationen in der späten Bronzezeit gegeben zu haben, die bis nach Chakassien im Norden (KEYSER *ET AL.* 2009) und Nordwest-

China im Süden reichte (LI *ET AL.* 2010). Diese Gruppen könnten sich mit der lokalen Bevölkerung des Altai-Gebirges und der angrenzenden Gebiete vermischt haben und so die vorliegende Haplogruppenzusammensetzung erklären. Denkbar wäre jedoch auch eine lokale Entwicklung im Bereich des Altai-Gebirges, denn bisher gibt es keine mitochondrialen Daten der bronzezeitlichen Bevölkerungen dieses Gebietes. Ein neolithisches Individuum aus dem 4. Jahrtausend v. Chr. trägt die Haplogruppe A4 (PILIPENKO 2012). Dies zeigt, dass bereits während des Neolithikums östliche Linien im russischen Bereich des Altai-Gebirges verbreitet waren. Die Haplogruppe A4 findet sich auch unter den Individuen der Aldy Bel und der Tagar-Kultur. Die zusätzliche Präsenz dieser Haplogruppe in der skythischen Gruppe des Rostov-Don-Gebietes im Westen, könnte für eine Verbindung der West-Skythen zum Altai-Gebirge sprechen.

In einer umfangreichen Studie zu eneolithischen und bronzezeitlichen Populationen des nordpontischen Raumes konnten ausschließlich westeurasische Linien detektiert werden (WILDE *ET AL.* 2014 *IN PRESS*). Auch frühere Studien an eneolithischen und bronzezeitlichen Individuen aus der Ukraine ergaben ausschließlich westeurasische mitochondriale Linien (NIKITIN *ET AL.* 2010, NIKITIN 2011). Im Gegensatz dazu wurde die osteurasische Haplogruppe C in neolithischen Individuen von dem Fundplatz Yasinovatka in der Ukraine detektiert (NIKITIN *ET AL.* 2012). Die Haplogruppe C wurde ebenfalls in neolithischen Individuen der Fundplätze Yuzhnyy Oleni Ostrov und Popovo aus Nordwest-Russland gefunden. Etwas weiter nordwestlich wurden auf dem bronzezeitlichen Fundplatz Bol'shoy Oleni Ostrov neben Linien der Haplogruppe C, ein D-Haplotyp und die Haplogruppe Z1a detektiert, die ebenfalls den osteurasischen Linien zugeordnet werden (DER SARKISSIAN *ET AL.* 2013). Die Präsenz osteurasischer Linien in neolithischen Individuen West-Eurasiens könnte mit einem östlichen Einfluss korreliert werden, der vor die eisenzeitlichen Migrationsbewegungen datiert. Welche demographischen Prozesse dazu führen, dass im nordpontischen Raum keine ostasiatischen Linien im Eneolithikum und der Bronzezeit zu finden sind ist unklar. Die größere Vielfalt ostasiatischer Linien in der Eisenzeit (DER SARKISSIAN 2011) im Vergleich zum Neolithikum legt eine Zuwanderung der Träger dieser Linien aus weiter östlich gelegenen Gebieten nahe. Neben der bereits erwähnten Haplogruppe A4, weist die mitochondriale Haplogruppe F1b, die in den Skythen des Rostov-Don-Gebietes und in den frühen Sarmaten detektiert wurde auf eine Verbindung zwischen Ost und West hin. In den veröffentlichten Studien prähistorischer Populationen wurde sie bisher nur in neolithischen Individuen aus dem Bereich des Baikalsees gefunden (MOODER *ET AL.* 2006, NASAB 2011). Das Vorhandensein einzelner

Haplogruppen stellt keine ausreichende Begründung für weiträumige Migrationsbewegungen dar, sondern kann nur Hinweise auf mögliche Verbindungen liefern. Als Zeichen erhöhter Mobilität können die Ergebnisse der  $F_S$ -Werte nach  $F_u$  gedeutet werden, denn die Tatsache, dass für die meisten Populationen signifikant negative  $F_S$ -Werte detektiert wurden, deutet auf eine expandierende Bevölkerung hin, könnte jedoch auch ein Hinweis auf Migration sein.

Um Hinweise auf eine eventuelle Ursprungspopulation der Skythen zu erhalten, wurden die genetischen Distanzen zwischen den eisenzeitlichen Skythen-Populationen und Kulturgruppen vorhergehender Epochen berechnet. Eine enge Verbindung scheint zwischen allen skythischen Gruppen und der spätbronzezeitlichen Gruppe aus Zentralasien zu bestehen, zumindest zeigen die  $F_{ST}$ -Wert-Berechnungen keine genetischen Distanzen an ( $F_{ST}$ -Werte: -0,01083–0,00865). Ob es sich hierbei tatsächlich um die Ursprungspopulation der Skythen handelt, oder einfach beide Gruppen gleichermaßen beeinflusst wurden lässt sich nicht mit Bestimmtheit sagen, vor allem da es sich bei der spätbronzezeitlichen Gruppe um lediglich acht Individuen handelt. Dennoch könnte es ein Hinweis darauf sein, dass die Entwicklungen der späten Bronzezeit in Zentralasien auch einen Einfluss auf die Gebiete westlich und östlich davon hatten. Die genetische Homogenität der skythischen Gruppen legt zwar eine gemeinsame Ursprungspopulation nahe, es ist jedoch wahrscheinlich, dass die skythischen Populationen in Ost und West ihre genetische Diversität aus mehr als einer Quelle schöpften. So zeigt die Gruppe „Eisenzeit West“ keine genetische Distanz zu den bronzezeitlichen Gruppen aus Chakassien und Kasachstan ( $F_{ST}$ -Werte: -0,00981 bzw. -0,00360), die beide auffällig hohe Anteile westeurasischer Linien tragen. Generell zeigt die westliche Skythengruppe inklusive der frühen Sarmaten jeweils geringere genetische Distanzen zu den westlichen Vergleichsgruppen als die östlichen Skythengruppen – ausgenommen hierbei ist die bronzezeitliche Gruppe aus Nordwest-Russland, die jedoch mit 61% wiederum einen sehr hohen Anteil asiatischer Linien aufweist.

Betrachtet man die östlichen Vergleichsgruppen, so findet sich eine geringe genetische Distanz zwischen der chinesischen Population aus Lajia und den östlichen Skythengruppen ( $F_{ST}$ -Werte: 0,01449–0,01544) und nur etwas größere Unterschiede zur eisenzeitlichen Gruppe West ( $F_{ST}$ -Wert: 0,03548). Dagegen zeigen jedoch alle drei eisenzeitlichen Gruppen die jeweils größte genetische Distanz zu der chinesischen Population des Tarim Beckens ( $F_{ST}$ -Werte: 0,12696–0,19477). Dieser Befund überrascht in Anbetracht der Tatsache, dass es sich hier um eine der ältesten, aus west- und osteurasischen Linien zusammengesetzten Populationen im Osten handelt, die bisher untersucht wurde wohingegen sich die Gruppe aus

Lajia nur aus osteurasischen Linien zusammensetzt. Vergleicht man jedoch die einzelnen mitochondrialen Linien, so zeigt sich, dass die Population des Tarim Beckens keinen einzigen Haplotypen mit den eisenzeitlichen Populationen gemein hat. Vielmehr noch sind die Haplotypen der Tarim Population in keiner der untersuchten Gruppen zu finden, was entweder auf eine andere Entstehungsgeschichte dieser Gruppe hindeutet, auf ein Beprobungsartefakt oder darauf, dass gemeinsame Linien aufgrund der abgeschiedenen Lage durch genetische Drift verloren gingen. Insgesamt zeigt sich ein komplexes Bild von genetischer Ähnlichkeit auf der einen Seite und extremen Unterschieden in teilweise benachbarten Gebieten auf der anderen Seite. Es scheint außer Frage, dass verschiedene Migrationswellen zur Diversität im gesamten zentralasiatischen Raum und der angrenzenden Gebiete beigetragen haben. Um die sich andeutenden Populationsdynamiken sicher entschlüsseln zu können, werden größere Datensätze aus der Bronzezeit benötigt.

In einem ersten Ansatz wurde von Friso Palstra (CNRS, Paris) versucht, durch die Simulation verschiedener demographischer Szenarien die komplexen Bevölkerungszusammenhänge anhand der eisenzeitlichen Daten skythischer Populationen und moderner Daten aus Europa und Ostasien zu entschlüsseln. Die bisherigen Ergebnisse deuten auf einen gemeinsamen Ursprung der skythischen Populationen aus Europa mit substantiellem Genfluss aus Ostasien hin (Friso Palstra, persönliche Mitteilung). Zu beachten ist hierbei, dass solche Simulationen sehr rechenaufwendig sind und deshalb auf einfachen Modellen beruhen, die mitunter die zugrundeliegende Komplexität nicht erfassen können.



## A6.6 Die phänotypischen Charakteristika der Skythen

Im Zuge dieser Studie wurden 21 Fragmente des Kerngenoms untersucht. Zwanzig dieser Fragmente enthalten jeweils eine Mutation, für die ein starker Zusammenhang mit einer phänotypischen Eigenschaft, wie z. B. Haut- oder Augenfarbe festgestellt wurde und von der vermutet wird, dass sie in einer oder mehreren Populationen unter positiver Selektion steht. Zusätzlich wurde der Amelogenin *locus* untersucht, der sich auf dem X- und dem Y-Chromosom in seiner Länge unterscheidet und dadurch eine molekulare Geschlechtsbestimmung ermöglicht. Nachfolgend werden die Frequenzen der selektierten Allele und damit die phänotypischen Charakteristika der Skythen diskutiert und den Frequenzen moderner Populationen gegenübergestellt. Das Hauptaugenmerk liegt hierbei auf den Markern, für die positive Selektion in europäischen oder asiatischen Populationen angenommen wird. Insgesamt wurden 51 Individuen untersucht, darunter 45 Ost-Skythen und sechs Individuen aus dem Bereich des Schwarzen Meeres. Von 32 Individuen konnte das Geschlecht bestimmt werden, wobei die Verteilung mit 14 weiblichen und 18 männlichen Individuen ausgeglichen ist. (N=Individuenanzahl)

### Laktasepersistenz

Unter anderem wurden zwei Marker untersucht, deren abgeleitete Allele in europäischen Populationen mit Laktasepersistenz, also der Fähigkeit auch im Erwachsenenalter Milchzucker verdauen zu können, korrelieren (ENATTAH *ET AL.* 2002). Die westlichen Individuen der Eisenzeit zeigen alle die homozygot anzestrale Form (N=4 für *LCTa* bzw. 3 für *LCTb*), sind also laktoseintolerant. Aufgrund des geringeren Datensatzes an skythischen Individuen aus dem Westen der Eurasischen Steppe ist der Fehlerbereich für die ermittelten Allelfrequenzen sehr hoch. Bezieht man die Fehlerspanne mit ein erreichen die Maximalwerte mit 37% bzw. 46% die Hälfte der Allelfrequenzen der modernen europäischen Vergleichspopulation und entsprechen ungefähr der Frequenz, die für moderne Populationen im Bereich nördlich des Schwarzen Meeres angenommen wird (ITAN *ET AL.* 2010). Im Vergleich zu den West-Skythen zeigen die östlichen Skythen mit 8,5% (N=41) bzw. 11% (N=41) höhere Anteile der abgeleiteten Allele der *LCT*-Marker, wobei keines der untersuchten Individuen Träger der homozygot abgeleiteten Variante ist. Mit 8,5% weisen die Ost-Skythen eine höhere Frequenz für das abgeleitete *-13.910\*T* Allel auf als die modernen

Vergleichspopulationen Zentralasiens (4,4%; N=113). Für den zweiten *LCT locus* entspricht die Allelfrequenz der Ost-Skythen (11%) in etwa dem Anteil des abgeleiteten Allels der modernen Vergleichspopulation Zentralasiens (10%; N=20)). Der Selektionsursprung für das *-13.910\*T* Allel wird aufgrund von Simulationen verschiedener Ausbreitungsmodelle in einer Population von Milchviehhaltern um 7.500 Jahre vor heute zwischen dem Balkan und Zentraleuropa angenommen (ITAN *ET AL.* 2009). In einer genetischen Studie acht früh- bzw. mittelneolithischer Individuen aus Europa, die in das 6. Jahrtausend v. Chr. datieren, wurden noch keine Anzeichen des abgeleiteten *-13.910\*T* Allels gefunden (BURGER *ET AL.* 2007). Gegen Ende des Neolithikums in Frankreich konnte in 26 Individuen ebenfalls kein Hinweis auf Laktasepersistenz detektiert werden (LACAN *ET AL.* 2011). Analysen, die im Zuge des gleichen Projekts wie auch die vorliegende Arbeit angefertigt wurden, ergaben keine abgeleiteten Allele und somit keine Laktasepersistenz in eneolithischen und bronzezeitlichen Individuen aus dem Nordpontikum (S. Wilde, persönliche Mitteilung). Ein Vergleich mittelalterlicher mit modernen Populationen Ungarns ergab einen Anstieg laktosetoleranter Individuen von 13% im Mittelalter (N=23) bis zu 61% im heutigen Ungarn (N=181) bzw. von 33% im Mittelalter, wenn man nur die als *commoners* bezeichneten Individuen miteinbezieht (N=9) (NAGY *ET AL.* 2011). Die vorgestellten Studien implizieren, dass das abgeleitete *-13.910\*T* Allels erst relativ spät die hohen Frequenzen erreicht, die in modernen Population Zentral- und Nordeuropas detektiert werden. In Relation zu diesen Studien sind die Frequenzen der abgeleiteten Allele im eisenzeitlichen Zentralasien relativ hoch, was auf einen europäischen Einfluss hindeutet.

### Pigmentierungs-*loci*

Es wurden weiterhin Marker untersucht, die mit einer Pigmentierung der Iris und der Haut assoziiert sind, darunter ein SNP im Intron 86 des *HERC2* Gens, der mit der Ausprägung der Augenfarbe assoziiert wird (STURM *ET AL.* 2008) und die Gene *SLC24A5*, *SLC45A2* und *TYR*, für die ein Zusammenhang mit der Pigmentierung der Haut nachgewiesen wurde (GRAF *ET AL.* 2005, LAMASON *ET AL.* 2005, SHRIVER *ET AL.* 2003). Die Marker *SLC24A5*, *SLC45A2* und *TYR* spielen eine wichtige Rolle bei der Evolution hellerer Haut in Europäern, nicht aber in Ostasiaten (NORTON *ET AL.* 2007), weshalb höhere Frequenzen der abgeleiteten Allele in europäischen, oder europäisch geprägten Populationen zu erwarten sind, als in nicht-europäischen Populationen. Die Frequenzen der abgeleiteten Allele der Pigmentierungs-*loci*

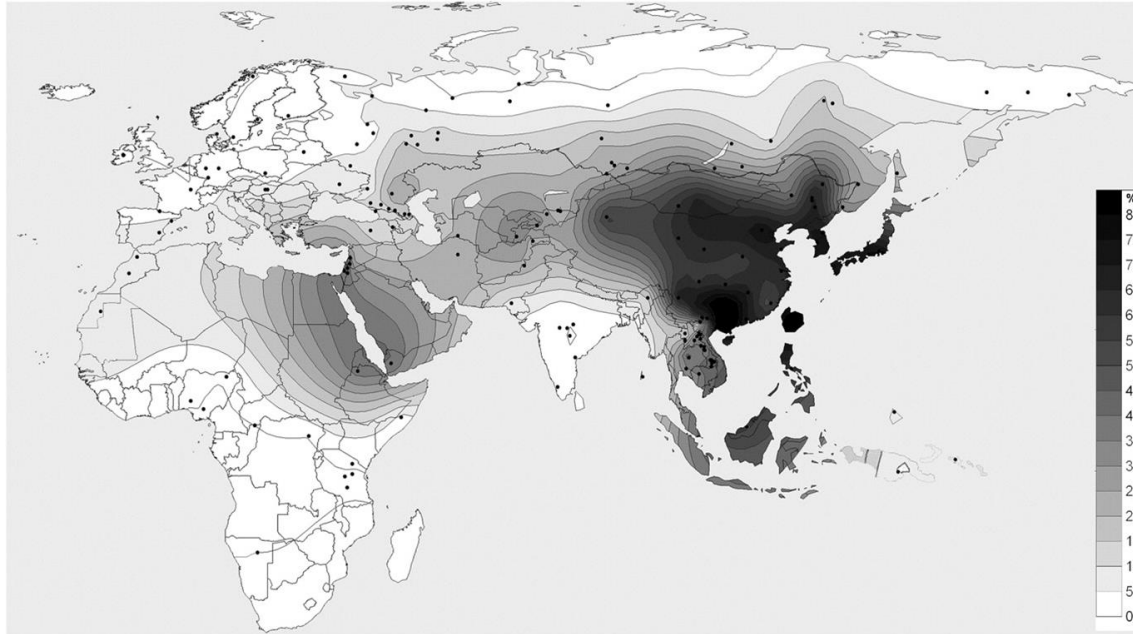
erreichen in modernen europäischen Populationen 79% für *HERC2* (N=60), 100% für *SLC24A5* (N=58), 98% für *SLC45A2* (N=58) und 43% für *TYR* (N=113). In der modernen asiatischen Vergleichspopulation erreichen die abgeleiteten Allele jeweils nur eine Frequenz von 0–1% (N=43–45). Zu den Markern *HERC2*, *SLC24A5* und *SLC45A2* liegen zusätzlich zu den modernen auch bronzezeitliche Vergleichsdaten aus dem 2. Jahrtausend v. Chr. aus Chakassien vor (BOUAKAZE ET AL. 2009). Die bronzezeitliche Population aus Chakassien zeigt für den *HERC2 locus* mit 72% (N=9) eine Frequenz des abgeleiteten Allels, die vergleichbar mit der in modernen europäischen Populationen ist. Die Ost-Skythen weisen an diesem *locus* mit 30% (N=41) eine geringere Frequenz des abgeleiteten Allels auf, die jedoch über der asiatischen (0%; N=45), zentralasiatischen (6%; N=298) und der fernöstlichen Vergleichspopulation (9%; N=50) liegt. Ein ähnliches Verhältnis zeigen die Frequenzen der beiden *loci SLC24A5* und *SCL45A2* für die erwähnten Populationen. Der hohe Anteil abgeleiteter Allele in der skythischen und vor allem der bronzezeitlichen Population im Vergleich zu modernen Populationen aus Zentralasien könnte auf einen starken europäischen Einfluss auf das östliche Zentralasien hindeuten, der sich mindestens bis ins 2. Jahrtausend vor Christus zurückdatieren lässt. Für die Allele des *TYR locus* liegen keine bronzezeitlichen Vergleichsdaten vor. Die Ost-Skythen weisen eine Frequenz von 16% für das abgeleitete Allel dieses *locus* auf (N=40), die leicht über der Frequenz von 13% moderner zentralasiatischer Populationen liegt (N=20). Das abgeleitete Allel des *TYR locus* findet sich nicht in der ostasiatischen Population (N=45) und nur zu 6% in der Vergleichspopulation aus Fernost (N=50). Für diesen Marker zeigen die Ost-Skythen also ebenfalls einen europäischen Einfluss. Für den *SLC24A5 locus* wurde bereits anhand von Studien an modernen Populationen vermutet, dass der relativ hohe Anteil des abgeleiteten Allels in zentralasiatischen Populationen im Vergleich zu Ostasiaten auf Genfluss mit europäischen Populationen zurückzuführen ist (NORTON ET AL. 2007). Der leichte Rückgang in den Frequenzen der abgeleiteten Allele von der Bronze- zur Eisenzeit könnte wiederum für einen Zustrom asiatischer Gruppen aus dem Osten oder Südosten zu Beginn der Eisenzeit sprechen. Bemerkenswert ist, dass die Frequenz des abgeleiteten *HERC2* Allels in eneolithischen und bronzezeitlichen Populationen aus dem nordpontischen Bereich lediglich 16% (N=47) beträgt (WILDE ET AL. 2014 IN PRESS). Die West-Skythen zeigen dagegen eine Frequenz des abgeleiteten Allels von 60% (N=5). Auch die Frequenz des abgeleiteten Allels des *SLC45A2 locus* bleibt im Nordpontikum mit 43,2% (N=22) unter dem Anteil in West-Skythen (100%; N=6), Ost-Skythen (57%; N=30) und bronzezeitlichen Gruppen der osteurasischen Steppe (70%; N=10).

Dieser Befund wirft die Frage auf, aus welchem Bereich genau – geographisch und zeitlich – dieser offensichtliche europäische Einfluss stammt, der sich in den bronze- und eisenzeitlichen Populationen aus dem Bereich des Altai-Gebirges detektieren lässt. Um diese Frage zu klären fehlen jedoch neolithische und bronzezeitliche Vergleichsdaten aus möglichen Ursprungsgebieten wie beispielsweise Ost- und Nordosteuropa. Die zusätzliche Untersuchung phänotypischer Marker zentralasiatischer Populationen aus der frühen Bronzezeit und wenn möglich auch vorhergehender Epochen ermöglicht die genauere Datierung eines putativen europäischen Einflusses bzw. die Überprüfung der Allelfrequenzen in einer zentralasiatischen Protopopulation um einen Einfluss aus anderen Gebieten zu bestätigen oder zu widerlegen. Die Untersuchung des jungpaläolithischen Individuums aus dem südsibirischen Fundplatz Mal'ta offenbarte eine dunkle Pigmentierung für Haare, Haut und Augen. Das abgeleitete Allel des *EDAR locus*, das mit Haardicke und schaufelförmigen Zähnen assoziiert ist und in ostasiatischen Populationen unter Selektion steht wurde nicht detektiert (RAGHAVAN *ET AL.* 2014).

## Verstoffwechslung körperfremder Stoffe

Eine weitere Gruppe von Genen, die untersucht wurde, sind solche, die in den Prozess der Verstoffwechslung körperfremder Stoffe involviert sind. Im Folgenden sollen die Ergebnisse für Typisierung der Mutationen im *ADH1B*-Gen näher betrachtet werden. Mutationen im *ADH1B*-Gen werden mit einer reduzierten Gefahr für eine Alkoholabhängigkeit assoziiert und stehen in ostasiatischen Populationen unter positiver Selektion (SHEN *ET AL.* 1997). In modernen Populationen zeigen die Frequenzen des abgeleiteten Allels des *ADH1Bb locus* (*ADH1B\*47His*) einen abfallenden Gradienten von Ostasien (85%) nach Europa (0%) (s. Abb. A11). Die moderne Vergleichspopulation aus Zentralasien zeigt für diesen *locus* eine Frequenz des abgeleiteten Allels von 40% (N=459). Für den *ADH1Ba locus* liegt die Frequenz des abgeleiteten Allels der modernen zentralasiatischen Vergleichspopulation bei 25% (N=275), in modernen europäischen Populationen ebenfalls bei 0% (N=21). Für das abgeleitete Allel des *ADH1Ba* Markers zeigen die West-Skythen eine Frequenz von 0% (N=4) und die Ost-Skythen eine Frequenz von 1% (N=35). Für das abgeleitete Allel des *ADH1Bb* Markers zeigen die Ost-Skythen eine Frequenz von 0% (N=11); Daten zu den West-Skythen liegen nicht vor. Die Skythen zeigen somit selbst im Vergleich mit Zentralasien sehr geringe Anteile der abgeleiteten Allele dieser *loci*. Die positive Selektion des *ADH1B\*47His*

Allels wurde mit der Einführung und Verbreitung der Reisdomestikation in Verbindung gebracht (PENG *ET AL.* 2010). Die extrem niedrige Frequenz für das abgeleitete Allel in den skythischen Populationen könnte eventuell auf einen Lebensstil zurückgeführt werden, der nicht mit dem Anbau von Reis in Verbindung steht.



**Abb. A 11** Heutige Frequenzen des ADH1B\*47His Allels (abgeleitetes Allel des ADH1Bb *locus*) nach BORINSKAYA *ET AL.* 2009

## Klimaanpassung

Wie für die beiden *ADH* Marker, so stehen auch die abgeleiteten Allele der *loci* *EDAR* und *ABCC11* unter Selektion in ostasiatischen Populationen (FUJIMOTO *ET AL.* 2008, OHASHI *ET AL.* 2011). Für beide wird vermutet, dass sie aufgrund einer Anpassung an ein kälteres Klima unter positiver Selektion stehen. Das abgeleitete Allel des *EDAR locus* ist mit der Ausprägung von dickerem Haar und schaufelförmigen Schneidezähnen assoziiert (FUJIMOTO *ET AL.* 2008). Das abgeleitete Allel des *locus* *ABCC11* führt zu einem Funktionsverlust des *ABCC11*-Proteins der zu trockenem Ohrenschmalz und einer verminderten Produktion von Achselschweiß führt (OHASHI *ET AL.* 2011, YOSHIURA *ET AL.* 2006). Das abgeleitete Allel des *EDAR locus* erreicht in ostasiatischen Populationen eine Frequenz von 93% (N=40), in der zentralasiatischen Vergleichspopulation 76% (N=227) und in modernen Europäern 0% (N=60). Die Ost-Skythen zeigen einen Anteil des selektierten Allels von 55% (N=11) und liegen somit zwischen dem heutigen Zentralasien und Europa. Für diesen *locus* konnten nur 11 Individuen erfolgreich typisiert werden, wodurch die ermittelte Allelfrequenz eine relativ

große Fehlerspanne von 32–75% umfasst und somit auch in den Variationsbereich der zentralasiatischen Populationen fallen könnte. Die Frequenz des abgeleiteten Allels für den Marker *ABCC11* (41%; N=40) liegt hingegen eindeutig zwischen den Allelfrequenzen der Vergleichspopulationen aus Europa (13%; N=112) und Zentralasien (78%; N=357). Die Ergebnisse weisen somit ebenfalls auf einen europäischen Einfluss auf das eisenzeitliche Zentralasien hin.

### Zusammenfassende Bewertung

Die zentralasiatische Vergleichspopulation wurde nach ihrer geographischen Lage, entsprechend des Verbreitungsgebietes der Ostskythen, ausgewählt. Die Frequenzen der abgeleiteten Allele dieser Gruppe liegen zwischen den Frequenzen europäischer und ostasiatischer Populationen. Die Allelfrequenzen der Ost-Skythen fallen für die *LCT loci* und die Marker *TYR*, *EDAR* und *CYP3A5* in etwa in die Variationsbreite der zentralasiatischen Vergleichspopulation. Die Frequenzen der restlichen Marker nehmen eine intermediäre Position zwischen der europäischen und zentralasiatischen Verteilung ein, abgesehen von den abgeleiteten Allelen der *ADH loci* deren Häufigkeiten in der Population der Ost-Skythen den in europäischen Populationen entsprechen. Für die meisten *loci* hat also eine Verschiebung der Allelfrequenzen in Richtung ostasiatische Variabilität von der Eisenzeit bis heute stattgefunden. Dies könnte darauf hindeuten, dass der europäische Einfluss auf Zentralasien, oder genauer das östliche Zentralasien in der Eisenzeit größer war als heute, oder der Einfluss aus östlichen Gebieten in den letzten 3.000 Jahren zugenommen hat. Positive Selektion könnte den Frequenzanstieg abgeleiteter Allele der Marker *EDAR*, *ABCC11* und der *ADH loci* erklären. Unter der Voraussetzung einer kontinuierlichen Bevölkerungsentwicklung müsste man jedoch im Gegenzug eine negative Selektion der Marker annehmen, die eine geringere Frequenz in modernen zentralasiatischen und asiatischen Populationen zeigen. Durch den Einsatz statistischer Methoden und Simulationen der demographischen Entwicklungen von der Eisenzeit bis heute könnte geklärt werden, ob Selektion, genetische Drift oder ein Bevölkerungswechsel für die Veränderung der Allelfrequenzen verantwortlich ist.

Die ermittelten Allelfrequenzen der skythischen Individuen aus dem Westen der Eurasischen Steppe sind aufgrund der geringen Individuenanzahl nicht sehr aussagekräftig. Als genereller

Trend kann jedoch festgehalten werden, dass die Allelfrequenzen der westlichen Skythen in den Variationsbereich europäischer Populationen fallen.

### Korrelation phänotypischer Ausprägung und mitochondrialer Haplogruppe

Die Individuen wurden dahingehend untersucht, ob eine Korrelation zwischen phänotypischer Ausprägung und der mitochondrialen Haplogruppe existiert. Es sollte überprüft werden, ob sich dadurch Hinweise auf eine Vermischung von Populationen aus unterschiedlichen geographischen Regionen detektieren lassen. Es konnte kein Zusammenhang zwischen dem Geschlecht eines Individuums und seiner Haplogruppe oder dem Genotyp festgestellt werden. Nur für drei Individuen konnte eine Korrelation zwischen phänotypischer Ausprägung und Haplogruppe festgestellt werden, wobei für alle drei Individuen jedoch maximal vier der 20 phänotypischen Marker erfolgreich typisiert werden konnten. Die Ergebnisse deuten darauf hin, dass die eisenzeitliche Population bereits eine Mischpopulation darstellt und dass der Zeitpunkt des *Admixtures* weiter in der Vergangenheit liegen muss. Generell setzt die Idee einer Korrelation zwei völlig getrennte Ursprungspopulationen voraus, wovon eine europäische Haplogruppen und phänotypische Ausprägungen aufweist und die andere zu 100% die asiatischen Genotypen und Haplogruppen trägt. Da die Zusammensetzung des mitochondrialen Genpools prähistorischer Populationen Zentralasiens nicht ausreichend aufgeklärt ist und nicht alle unter Selektion stehenden Marker selbst in den entsprechenden modernen Populationen fixiert sind lassen sich allein durch eine vergleichende Betrachtung keine sinnvollen Rückschlüsse auf eine eventuelle Vermischung zweier, oder mehrerer Ursprungspopulationen treffen.

## Hardy-Weinberg-Gleichgewicht

Zur Überprüfung ob sich die Populationen für die untersuchten *loci* im Hardy-Weinberg-Gleichgewicht befinden wurden ein  $\chi^2$ -Tests mit anschließender Berechnung der Zufallswahrscheinlichkeit und ein HWExact-Test durchgeführt. Die Analysen wurden nur für die Genotypen der Ost-Skythen durchgeführt, da die Individuenanzahl der West-Skythen zu gering ist um ein aussagekräftiges Ergebnis zu liefern. Die Ergebnisse des  $\chi^2$ -Tests zeigen keine signifikante Abweichung der Population vom Hardy-Weinberg-Gleichgewicht. Eine Abweichung würde in diesem Fall eine zugrundeliegende Populationsstruktur bedeuten, worauf es anhand des Datensatzes keine Hinweise gibt. Da der  $\chi^2$ -Tests nicht gut für kleine Stichproben geeignet ist, wurde zusätzlich der HWExact-Test des HardyWeinberg-Pakets für R durchgeführt. Anhand des HWExact-Tests konnte kein signifikanter Homozygotenüberschuss festgestellt werden. Ein Homozygotenüberschuss könnte auf Selektion oder eine zugrundeliegende Populationsstruktur hindeuten (DEVLIN *ET AL.* 1990), wofür auch mit diesem Test keine Hinweise gefunden wurden. Die Ergebnisse deuten also auf eine homogene Population hin, sollten jedoch aufgrund der Tatsache, dass auch für die Ost-Skythen teilweise nur sehr wenige Individuen pro *locus* erfolgreich getestet wurden mit Vorsicht betrachtet werden.



## A7 Zusammenfassung

Die vorliegende Arbeit widmet sich der populationsgenetischen Untersuchung eisenzeitlicher Reiternomaden der eurasischen Steppe, die aufgrund ihrer materiellen Hinterlassenschaften mit der skythischen Kultur in Verbindung gebracht werden. Durch den Einsatz von Multiplex-Amplifikationen und der 454-Sequenzierung gelang es 78 Individuen aus dem Bereich des Altai-Gebirges und 19 Individuen aus der westeurasischen Steppe zu analysieren und somit den bislang umfangreichsten Datensatz skythischer Populationen zu produzieren. Für die Analysen wurden zusätzlich bereits veröffentlichte Daten aus der Literatur hinzugezogen, wodurch insgesamt 147 eisenzeitliche Individuen zur Verfügung standen.

Anhand des Datensatzes sollten folgende Punkte untersucht werden:

- 1) Die demographischen Prozesse im Bereich des Altai-Gebirges
- 2) Die genetische Struktur der Pazyryk-Kultur
- 3) Das Verhältnis der Skythen zu den frühen Sarmaten
- 4) Das Verhältnis zwischen westlichen und östlichen skythischen Populationen
- 5) Das Ursprungsgebiet der Skythen
- 6) Die phänotypischen Charakteristika der Skythen

### 1) Die demographischen Prozesse im Bereich des Altai-Gebirges

Die verschiedenen Populationen aus dem Osten der Eurasischen Steppe werden aufgrund von Ähnlichkeiten kultureller Artefakte mit den Skythen des Nordschwarzmeerraumes, bzw. der skythischen Kultur assoziiert. Aufgrund regionaler Besonderheiten im kulturellen Fundgut wird jedoch jede Gruppe einer individuellen Kulturstufe zugeordnet. Es stellt sich die Frage, ob aufgrund der kulturellen Gemeinsamkeiten von einer einzigen Population ausgegangen werden kann, oder ob die Unterschiede im kulturellen Fundgut anzeigend für eine individuelle genetische Entwicklung jeder einzelnen Gruppe sind.

Die Analysen ergaben keine bzw. nur geringe genetische Distanzen zwischen den diversen Gruppen aus dem Altai-Gebirge und den angrenzenden Regionen, was auf Genfluss zwischen den untersuchten Populationen hindeutet. Einzige Ausnahme bildet die Tagar-Population

bestehend aus sechs Individuen des Fundplatzes Barsučij Log, die mittlere bis erhöhte genetische Distanzen zu den weiteren Populationen aus dem Bereich des Altai-Gebirges zeigt ( $F_{ST}$ -Werte = 0,04039–0,08830). Aufgrund der geringen Individuenanzahl handelt es sich hierbei wahrscheinlich um ein Beprobungsartefakt und erst die Untersuchung weiterer Individuen dieser Kulturgruppe würde eine präzisere Aussage über die genetische Verwandtschaft dieser Population zu den anderen Gruppen erlauben. Da die untersuchten Individuen aus zeitlich aufeinander folgenden Kulturstufen stammen, die vom 9. bis ins 3. Jahrhundert v. Chr. datieren, deutet die geringe genetische Distanz gleichzeitig auf eine kontinuierliche Bevölkerungsentwicklung im 1. Jahrtausend v. Chr. hin. Die Simulation verschiedener demographischer Szenarien für die östlichen Gruppen ergab, dass die skythischen Individuen einer einzigen expandierenden Population entstammen (F. Palstra, persönliche Mitteilung), wodurch die Befunde bestätigt werden.

## 2) Die Struktur der Pazyryk-Kultur

Zur Pazyryk-Kultur des 4.–3. Jahrhunderts v. Chr. liegen von allen skythischen Populationen die meisten genetischen Daten vor. Die zahlreichen Funde verteilen sich über den kasachischen, mongolischen und russischen Teil des Altai-Gebirges, wobei sie sich in Russland auf das Ukok-Plateau und das Chuya-Tal konzentrieren. Die Daten zur Pazyryk-Kultur wurden dahingehend untersucht, ob sich eine genetische Substruktur aufgrund der geographischen Verteilung einer kulturell einheitlichen Population feststellen lässt.

Ein Vergleich zwischen den Populationen des kasachischen, russischen und mongolischen Teils des Altai-Gebirges ergab nur geringe genetische Distanzen zwischen diesen Gruppen ( $F_{ST}$ -Werte = 0,00439–0,02187). Nach der Aufteilung der russischen Population in die Gruppen Ukok-Plateau und Chuya-Tal zeigte sich, dass die Chuya-Population etwas größere genetische Distanzen zu den anderen Gruppen zeigte ( $F_{ST}$ -Werte = 0,02979–0,05969) als diese unter sich ( $F_{ST}$ -Werte = -0,00395–0,00439). Die Ergebnisse weisen auf eine Substruktur innerhalb der Pazyryk-Population hin, die auf eingeschränkten Genfluss – womöglich aufgrund geographischer Isolation – zwischen der Population des Chuya-Tals und den restlichen Pazyryk-Gruppen hindeutet.

### 3) Das Verhältnis der Skythen zu den frühen Sarmaten

Neben den östlichen Gruppen wurde auch das Verhältnis der skythischen und sarmatischen Gruppen des Westens zueinander geprüft. Die Präsenz charakteristischer Elemente des skythischen Tierstils in den frühsarmatischen Kurganen aus Pokrovka (MALASHEV & YABLONSKY 2004) legt eine Verbindung zu den Schwarzmeerskythen nahe. Gleichzeitig wird jedoch von einer Verdrängung der Skythen durch sarmatische Verbände im 2. Jahrhundert v. Chr. ausgegangen (SMIRNOV 1980), was dafür sprechen würde, dass es sich hier um völlig getrennte Populationen handelt. Es wurde untersucht, ob sich die Sarmaten genetisch von den Skythen unterscheiden und somit tatsächlich von einem Bevölkerungswechsel im nordpontischen Raum ausgegangen werden kann.

Die populationsgenetischen Untersuchungen zeigen, dass es keine genetischen Distanzen zwischen den Frühen Sarmaten aus dem Orenburg-Gebiet und den West-Skythen gibt. Die Verdrängung der Skythen durch die Sarmaten im 2. Jahrtausend v. Chr. ist wahrscheinlich das Resultat einer Entwicklung, die in der nördlichen Peripherie Chinas ihren Anfang nahm und durch eine westwärts Bewegung hunnischer Verbände den Beginn der Hunno-Sarmatischen-Periode einläutete (PARZINGER 2006). Wie der Begriff schon nahelegt werden die Sarmaten als Teil dieser Bevölkerungsbewegung verstanden, wie auch die Wusun in Kasachstan und die Xiongnu (Hunnen) in Innerasien. Die frühen Sarmaten scheinen jedoch von diesen Entwicklungen noch unbeeinflusst zu sein.

### 4) Das Verhältnis zwischen westlichen und östlichen skythischen Populationen

Trotz kultureller Ähnlichkeiten wird eine Zusammenfassung der verschiedenen eisenzeitlichen Gruppen unter dem Begriff Skythen von Seiten der Archäologie stark kritisiert (YABLONSKY 2000). Um die Frage zu klären, ob sich nur kulturelle Artefakte über den weiten Bereich der Eurasischen Steppe verbreiteten oder eine genetische Verbindung zwischen den westlichen und östlichen Populationen besteht wurden zuerst alle westlichen mit allen östlichen Individuen verglichen und anschließend die einzelnen Populationen im Osten mit denen im Westen. Zusätzlich wurde das Verhältnis moderner Populationen aus dem Westen und Osten der Eurasischen Steppe zueinander untersucht und die genetische Distanz mit der zwischen den eisenzeitlichen Bevölkerungen verglichen.

Die Gruppe aller westlichen Individuen zeigt eine signifikant geringe genetische Distanz zur Gruppe der östlichen Individuen ( $F_{ST}$ -Wert = 0,01733), was bedeutet, dass die beiden Regionen genetisch nicht voneinander isoliert waren oder die Populationen einen relativ rezenten gemeinsamen Ursprung hatten. Die meisten der Gruppen in Ost und West zeigen ebenfalls geringe genetische Distanzen zueinander. Ausnahmen bilden die bereits erwähnte Tagar-Kultur, die mittlere bis hohe genetische Distanzen zu den Gruppen im Westen zeigt ( $F_{ST}$ -Werte = 0,03351–0,09589) und die frühen Sarmaten, die mittlere bis erhöhte genetische Distanzen zu den Individuen der Aldy Bel ( $F_{ST}$ -Wert = 0,04650), Pazyryk- ( $F_{ST}$ -Wert = 0,04020) und Zevakino-Čilikta-Kultur zeigt ( $F_{ST}$ -Wert = 0,05125). Diese Substruktur innerhalb der ansonsten sehr homogenen skythischen Gruppen könnte auf ein Beprobungsartefakt, eingeschränkten Genfluss, oder genetische Einflüsse aus anderen Bereichen der Eurasischen Steppe, die nicht alle Gruppen gleichermaßen beeinflussen, zurückzuführen sein. Die hohen Haplotypendiversitäten der untersuchten Populationen deuten auf eine große effektive Populationsgröße oder den ständigen Zustrom neuer mitochondrialer Linien hin. Die genetischen Substrukturen der skythischen Populationen könnten daher auf den Zustrom aus unterschiedlichen Regionen zurückzuführen sein.

Die modernen Populationen aus dem Altai, Tuva und Chakassien im Osten zeigen signifikant hohe genetische Distanzen zu der modernen Vergleichspopulation aus Südwest-Russland ( $F_{ST}$ -Werte = 0,08136–0,15671). Die eisenzeitlichen Individuen aus dem Westen der Eurasischen Steppe zeigen eine geringe genetische Distanz zur modernen Vergleichspopulation aus Südwest-Russland ( $F_{ST}$ -Wert = 0,01279). Gleichfalls zeigen die östlichen Populationen der Eisenzeit geringe bis mittlere genetische Distanzen zu den modernen Vergleichspopulationen des Ostens ( $F_{ST}$ -Werte = 0,01281–0,05384). Die Ergebnisse könnten auf eine kontinuierliche Bevölkerungsentwicklung im Osten und Westen der Eurasischen Steppe hindeuten, wobei von einem größeren Genfluss zwischen dem Osten und Westen der Eurasischen Steppe in der Eisenzeit im Vergleich zu heute auszugehen ist.

## 5) Das Ursprungsgebiet der Skythen

Die Ergebnisse von Radiokarbondatierungen an skythischen Monumenten und die Präsenz charakteristischer Tierstilelemente im spätbronzezeitlichen China deuten auf einen östlichen Ursprung der skythischen Kultur hin. Es wurden jedoch nicht alle Monumente datiert und aufgrund möglicher Fundlücken oder kultureller Integration kann ein westlicher oder zentralasiatischer Ursprung nicht ausgeschlossen werden. Die skythischen Populationen wurden deshalb mit neolithischen und bronzezeitlichen Gruppen West-, Ost- und Zentralasiens verglichen um dadurch Hinweise auf ihren Ursprung zu erhalten und folgende Hypothesen zu prüfen:

- a) Unabhängige Populationen in Ost und West
- b) Entstehung im Nordpontikum und anschließende Ausbreitung nach Osten
- c) Entstehung in Zentralasien und Ausbreitung nach West und Ost
- d) Entstehung im Bereich des Altai-Gebirges mit anschließender Ausbreitung nach Westen

Da die populationsgenetischen Analysen darauf hindeuten, dass die skythischen Populationen in Ost und West wahrscheinlich einer gemeinsamen Ursprungspopulation entstammen oder zumindest substantieller Genfluss zwischen den verschiedenen Gruppen existierte, kann der alleinige kulturelle Transfer und somit Szenario a) ausgeschlossen werden.

Anhand von Vergleichen der mitochondrialen Haplogruppenzusammensetzung bronze- und eisenzeitlicher Populationen lässt sich die Herkunft der genetischen Variabilität der skythischen Gruppen nicht eindeutig feststellen:

Die West-Skythen weisen ostasiatische Haplogruppen auf, die in den vorherigen Epochen im Westen der Eurasischen Steppe nicht vorhanden sind. Ein Ursprung im Westen ist also nur unter Einbeziehung substantiellen Genflusses aus dem Osten haltbar.

Aufgrund der Anteile west- und osteurasischer Linien käme die spätbronzezeitliche Kultur der westsibirischen Waldsteppe als Ursprung der östlichen Skythen in Frage – es finden sich allerdings europäische Haplogruppen im skythischen Genpool die sich in der späten Bronzezeit Westsibiriens noch nicht nachweisen lassen. Es muss also von einem zusätzlichen westeurasischen Einfluss ausgegangen werden.

Die Ostskythen weisen ungefähr zu gleichen Teilen west- und osteurasische Linien auf. Die bronzezeitlichen Gruppen aus China und der Mongolei zeigen hauptsächlich osteurasische Haplogruppen. Die bronzezeitlichen Gruppen aus Chakassien und Kasachstan, für die vermutet wird, dass sie sich vom Westen der Eurasischen Steppe nach Osten ausgebreitet haben, setzen sich fast ausschließlich aus westeurasischen Linien zusammen. Keine der verfügbaren bronzezeitlichen Populationen im Osten der Eurasischen Steppe zeigt eine Haplogruppenzusammensetzung, die vergleichbar mit der der Ost-Skythen wäre. Da bisher keine verlässlichen Daten aus der Bronzezeit des Altai-Gebirges vorliegen, ist zum jetzigen Zeitpunkt davon auszugehen, dass die Zusammensetzung des Genpools der Ost-Skythen das Resultat eines Zusammenschlusses bronzezeitlicher Gruppen westlicher und südöstlicher Herkunft darstellt.

Die  $F_{ST}$ -Wert Berechnungen ergeben keine genetischen Distanzen aller eisenzeitlichen Gruppen zur spätbronzezeitlichen Gruppe aus Sary Sad in der westsibirischen Waldsteppe ( $F_{ST}$ -Werte = -0,01083–0,00865). Zur nachfolgenden Gruppe der späten Irmen-Kultur dieser Region bestehen geringe genetische Distanzen ( $F_{ST}$ -Werte = 0,00988–0,02778); ebenso zur bronzezeitliche Gruppe aus Chakassien ( $F_{ST}$ -Werte = -0,00981–0,02349). Die Eisenzeit Gruppe West zeigte zusätzlich keine genetische Distanz zur Yamnaya Kultur des nördlichen Schwarzmeerraumes und zur bronzezeitlichen Gruppe aus Kasachstan. Die östlichen Gruppen der Eisenzeit weisen geringe genetische Distanzen zur bronzezeitlichen Gruppe aus Lajia, China auf ( $F_{ST}$ -Werte = 0,01449–0,01544) zu der auch die West-Skythen nur eine geringe bis mittlere genetische Distanz zeigen ( $F_{ST}$ -Wert = 0,03548). Die Daten weisen auf die Möglichkeit eines gemeinsamen zentralasiatischen Ursprungs hin; zeigen aber zusätzlich, dass die östlichen und westlichen Gruppen der Eisenzeit durchaus auch lokalen Einflüssen ausgesetzt waren.

Die Simulation der verschiedenen demographischen Szenarien deutet auf einen westeurasischen Ursprung mit substanziellem Genfluss aus Ostasien hin (Friso Palstra, pers. Mitteilung). Die bronzezeitlichen Populationen wurden jedoch noch nicht in die Simulationen integriert.

## 6) Die phänotypischen Charakteristika der Skythen

Es wurden nukleäre Marker untersucht, die mit einer phänotypischen Ausprägung assoziiert sind und von denen vermutet wird, dass sie entweder in europäischen oder asiatischen Populationen unter positiver Selektion stehen. Durch die Analyse sollten ein Einblick in das Erscheinungsbild der untersuchten Individuen und Hinweise auf den geographischen Ursprung der Skythen gewonnen werden. Die ermittelten Frequenzen der abgeleiteten Allele der skythischen Populationen wurden dazu mit denen moderner Bevölkerungen Europas, Ostasien, Zentralasiens und des Fernen Ostens verglichen.

Die Allelfrequenzen der Ost-Skythen fallen für die *LCT loci* und die Marker *TYR*, *EDAR* und *CYP3A5* in etwa in die Variationsbreite der zentralasiatischen Vergleichspopulation. Die Frequenzen der restlichen Marker nehmen eine intermediäre Position zwischen der europäischen und zentralasiatischen Verteilung in modernen Populationen ein, abgesehen von den abgeleiteten Allelen der *ADH loci* deren Häufigkeiten in der Population der Ost-Skythen den in modernen europäischen Populationen entsprechen. Die Ergebnisse deuten darauf hin, dass der europäische Einfluss auf Zentralasien, oder genauer das östliche Zentralasien in der Eisenzeit größer war als heute, oder der Einfluss aus östlichen Gebieten in den letzten 3.000 Jahren zugenommen hat. Durch den Einsatz statistischer Methoden und Simulationen der demographischen Entwicklungen von der Eisenzeit bis heute könnte geklärt werden, ob Selektion, genetische Drift oder ein Bevölkerungswechsel für die Veränderung der Allelfrequenzen verantwortlich ist.

Die Allelfrequenzen der westlichen Skythen fallen in den Variationsbereich moderner europäischer Populationen. Eine Ausnahme bilden die Frequenzen der abgeleiteten Allele der Laktase loci, die mit jeweils 0% der Frequenz moderner ostasiatischer Populationen entspricht. Die ermittelten Allelfrequenzen der skythischen Individuen aus dem Westen der Eurasischen Steppe sind aufgrund der geringen Individuenanzahl jedoch nicht sehr aussagekräftig.

## A8 Ausblick

Die vorliegende Studie liefert einen Einblick in die genetischen Zusammenhänge und demographischen Entwicklungen eisenzeitlicher Populationen der Eurasischen Steppe. Dabei bleiben die Analysen und somit die Schlussfolgerung jedoch beschränkt auf die mütterliche Seite der Bevölkerungshistorie, da nur die mitochondriale DNA untersucht wurde. Durch die Analyse Y-chromosomaler Marker könnte die Information zu den demographischen Prozessen der skythischen Populationen um die männliche Seite erweitert werden. Das Problem bleibt jedoch, dass durch die Untersuchung einzelner genetischer Marker nur ein Teil der genealogischen Information wiedergegeben wird. Die Untersuchung neutraler nukleärer Marker wäre der logisch nächste Schritt um die Bevölkerungsentwicklung der Skythen detaillierter nachvollziehen zu können.

Für weiterführende Analysen ist außerdem eine Erweiterung des Datensatzes sinnvoll, da einige Epochen bzw. Kulturgruppen lediglich durch singuläre Fundplätze repräsentiert sind. Somit könnte geklärt werden, ob die Tagar-Kultur eine eigenständige Entwicklung im Bereich des Minusinsker Beckens erfahren hat und welche Verbindung zwischen den frühen Sarmaten und der Tagar/Tes-Population besteht. Eine eingehendere Untersuchung der initialen Skythenzeit im Bereich nördlich des Schwarzen Meeres könnte außerdem Hinweise darauf geben, ob hier bereits ein ostasiatischer Einfluss zu spüren ist. Weiterhin könnten Computersimulationen dazu eingesetzt werden um zu überprüfen, ob auch im Westen der Eurasischen Steppe eine kontinuierliche Bevölkerungsentwicklung vorliegt, vor allem zwischen der Bronze- und der Eisenzeit. In einer Kooperation mit Friso Palstra (CNRS Paris) wurden, durch den Einsatz von Modellierungen bereits erste Schritte unternommen die komplexen demographischen Prozesse Zentralasiens vom Neolithikum zur Eisenzeit besser nachvollziehen zu können. Die Erstellung eines umfangreichen Datensatzes zu bronzezeitlichen und neolithischen Populationen Zentralasiens und auch der angrenzenden Gebiete und deren Einbeziehung in die simulierten Modelle können zur Lokalisation eines Ursprungsorts der skythischen Population beitragen.



# B Methoden

## B 1 Methodischer Stand der palaeogenetischen Forschung

Die Möglichkeit, Sequenzen alter DNA, in diesem Fall Sequenzen eines ausgestorbenen Wollhaarmammuts zu extrahieren, präsentierten Paul H. Johnson und seine Kollegen 1981 auf dem Treffen der Gesellschaft für Palaeopathologie in Detroit (JOHNSON *ET AL.* 1985). Wenige Jahre später wurden – durch unspezifische bakterielle Klonierung – die ersten mitochondrialen Sequenzen eines ausgestorbenen Quaggas veröffentlicht und damit der Grundstein der palaeogenetischen Forschung gelegt (HIGUCHI *ET AL.* 1984). Nur wenige Studien folgten auf diese ersten Erfolge (DORAN *ET AL.* 1986, ROGERS & BENDICH 1985), denn die in nahezu allen aDNA-Sequenzen vorhandenen Basenläsionen erschweren eine bakterielle Klonierung immens (PÄÄBO 1989). Des Weiteren ist bei einer ungerichteten, bakteriellen Klonierung die Reproduktion der Ergebnisse nur schwierig zu realisieren, weil die Vervielfältigung des gleichen Fragmentes nicht gewährleistet ist. Speziell bei alter DNA ist die Bestätigung der detektierten Mutationen jedoch unerlässlich, da Basenveränderungen, wie z.B. Deaminierungen, die nach dem Tod eines Organismus durch die Abbauprozesse entstehen können, zum Einbau falscher Nukleotide während der Replikation führen (GILBERT *ET AL.* 2003). Ein entscheidender Fortschritt wurde durch die Entwicklung der Polymerase Kettenreaktion (PCR) erzielt (MULLIS & FALOONA 1987). Mit ihrer Hilfe wurde es möglich, gezielt bestimmte Regionen des Genoms zu vervielfältigen und so die evolutionären Prozesse innerhalb und zwischen verschiedenen Spezies zu vergleichen. Die PCR war und ist auch heute noch die Methode der Wahl, wenn es um die gezielte Untersuchung einzelner Positionen oder Bereiche im Genom geht. Sie wurde erfolgreich in zahlreichen palaeogenetischen Studien eingesetzt (BOLLONGINO 2006, BRACE *ET AL.* 2012, BRAMANTI *ET AL.* 2009, BRANDT *ET AL.* 2013, BURGER *ET AL.* 2004, HAAK *ET AL.* 2005, HOFREITER *ET AL.* 2002, TROY *ET AL.* 2001). Für aDNA Studien wurden bevorzugt Fragmente des mitochondrialen Genoms untersucht, da die mitochondriale DNA in einer höheren Kopienzahl vorliegt als die Kern-DNA und so vor allem bei besonders schlecht erhaltenem Probenmaterial bessere Erfolgsaussichten verspricht. Für Untersuchungen menschlicher Populationen wurde vor allem die Hypervariable Region 1 (HVR1) der Kontrollregion des mitochondrialen Genoms verwendet, die im Vergleich zur kodierenden Region eine erhöhte

Mutationsrate aufweist (INGMAN *ET AL.* 2000, PARSONS *ET AL.* 1997) wodurch auch rezenterere Evolutionsereignisse nachvollzogen werden können. Mit wachsender Datenmenge zeigte sich jedoch, dass die alleinige Betrachtung der HVR1 eine zu geringe Auflösung der Variabilität bietet und weitere Marker, wie die Hypervariable Region 2 (HVR2) und Mutationen der kodierenden Region wurden in die Analysen einbezogen (FINNILÄ *ET AL.* 2001). Die Anzahl an Markern, die untersucht werden können wird jedoch, vor allem in der Palaeogenetik, durch das nur limitiert zur Verfügung stehende Probenmaterial beschränkt. Erste Fortschritte im Bestreben mehr Informationen aus Zahn- und Knochenmaterial zu gewinnen wurde durch den Einsatz modifizierter PCR-Methoden, wie dem Einsatz von Multiplex-PCRs oder *two-step*-PCRs, erreicht (KRAUSE *ET AL.* 2006). Jedoch können auch bei der Multiplex-PCR, aufgrund von Primer-Primer-Interaktionen, oder sterischen Hinderungen, nicht beliebig viele Positionen gleichzeitig amplifiziert werden. Eine Lösung für dieses Problem bietet die RainStorm-Plattform: in einer Weiterentwicklung der Emulsions-PCR (NAKANO *ET AL.* 2003) werden hier in einzelnen Tropfen die PCR-Chemikalien und je ein spezifisches Primerpaar mit dem DNA-Template vereint und somit Mikroreaktoren geschaffen, womit diese Technik in der Lage ist, bis zu 1,5 Mio. Amplifikationen parallel durchzuführen (TEWHEY *ET AL.* 2009). Ein weiterer Ansatz ist die selektive Anreicherung genetischer Marker durch *Hybrid Capture*. Hierbei werden Sequenzen komplementär zu den gesuchten DNA-Abschnitten hergestellt und entweder auf einer festen Oberfläche (Array) (ALBERT *ET AL.* 2007) oder in einer Lösung (GNIRKE *ET AL.* 2009) bereitgestellt. Nach Zugabe des DNA-Extraktes hybridisieren die gesuchten Fragmente an die künstlich hergestellten Gegenstücke und können so aus dem Sequenzgemisch separiert werden. Mit den wachsenden Möglichkeiten zur DNA Amplifikation und besseren Ausschöpfung des vorhandenen Probenmaterials steigen jedoch auch die Anforderungen an die Sequenzanalyse und Auswertung der generierten Daten.

Zur Sequenzanalyse hat sich die Strangabbruchmethode nach Sanger, mit Adaptionen zur automatisierten Sequenzanalyse, durchgesetzt (SANGER *ET AL.* 1977, SMITH *ET AL.* 1986), bis im Jahre 2005 die erste Sequenziertechnik der neuen Generation der Öffentlichkeit präsentiert wurde (MARGULIES *ET AL.* 2005). Diese erste Technik wurde von der Firma 454 Life Science entwickelt und beruht auf dem Prinzip des *Pyrosequencing* (RONAGHI *ET AL.* 1996) nach vorheriger klonaler Vervielfältigung der DNA-Fragmente in einer Emulsions-PCR. Wenig später folgten der *Genome Analyzer* der Firma Solexa, die 2007 von Illumina übernommen wurde und das SOLiD System von Applied Biosystems. Bei der Sequenzierung mit dem

*Genome Analyzer* werden die DNA-Fragmente durch eine *solid-phase-* oder *bridge-PCR* vervielfältigt und anschließend mittels *reversible terminator technology* sequenziert (BENTLEY *ET AL.* 2008). SOLiD bedeutet *Sequencing by Oligo Ligation Detektion*. Das Prinzip beruht auf dem Einbau von je zwei Basen, die in Kombination eine spezifische Fluoreszenzmarkierung tragen. Der Prozess muss für jede Sequenz mehrmals mit unterschiedlichem Startpunkt wiederholt werden, bis jeder Position die richtige Base zugeordnet werden kann. Diese neuen Techniken beruhen alle auf dem *sequencing by synthesis* Prinzip, was bedeutet, dass jedes Nukleotid direkt beim Einbau erfasst wird. Auf diese Weise erhält man die Basenabfolge jedes einzelnen Ausgangsmoleküls im Gegensatz zur Sanger-Sequenzierung, bei der die Sequenzanalyse nach der Synthese stattfindet und nur eine Konsensussequenz aller Ausgangsmoleküle erfasst wird. Anschauliche Darstellungen der bisher erwähnten Methoden finden sich in den ersten Reviews zum *Next-Generation Sequencing* (MARDIS 2008, METZKER 2010, SHENDURE & JI 2008). Alle bisher genannten Sequenziertechnologien beruhen auf der Detektion einer Lichtemission, die spezifischen Basen zugeordnet werden kann, anders funktioniert das Prinzip der *Personal Genome Machine* (PGM) von Ion Torrent. Hier werden die DNA-Fragmente ebenfalls durch eine Emulsions-PCR vervielfältigt und anschließend in Vertiefungen auf einer Halbleiterplatte eingebracht. Wird eine Base eingebaut, werden Ionen frei, die zu einer Spannungsänderung führen, die für jede Vertiefung individuell gemessen werden kann (ROTHBERG *ET AL.* 2011). Die Sequenzierung einzelner Moleküle ohne vorherige Amplifikation ermöglicht die Firma Pacific Biosciences mit dem PacBio RS II (Prinzip s. EID *ET AL.* 2009).

Der entscheidende Vorteil des sogenannten *Next Generation Sequencing* (NGS) gegenüber der herkömmlichen Sequenzierung ist die hochparallele Sequenzanalyse (s. Tab. B1). Bei gleicher DNA-Einsatzmenge kann ein Vielfaches an Daten produziert werden, wodurch die neuen Techniken schnellen Anklang in der Palaeogenetik fanden. So wurden bereits Ende 2005 die ersten Ergebnisse der Sequenzierung von Mammut-DNA veröffentlicht (POINAR *ET AL.* 2006) und wenig später zwei Studien zu Neandertaler DNA (GREEN *ET AL.* 2006, NOONAN *ET AL.* 2006). Um das Potential der neuen Sequenziermethoden ökonomisch nutzen zu können wurden verschiedene Methoden entwickelt zur parallelen Analyse mehrerer Proben. Das generelle Prinzip besteht in der Markierung der Proben durch eine individuelle Erkennungssequenz, die als Barcode, TAG oder Index bezeichnet wird. Beim sog. *tagging* werden die Erkennungssequenzen, oder TAGs an die DNA-Fragmente ligiert (MEYER *ET AL.* 2007). Sollen nur bestimmte Abschnitte der DNA untersucht werden, können die TAGS an

die Primersequenzen angehängt und somit während der Amplifikation in die DNA-Sequenzen integriert werden (BINLADEN *ET AL.* 2007). Bei der Sequenzmarkierung für die Illumina-Sequenzierung, werden in einem ersten Schritt Sequenzieradapter an die DNA-Fragmente ligiert und der Index durch eine nachträgliche Index-PCR in den Sequenzieradapter integriert (MEYER & KIRCHER 2010). Ein Nachteil der neuen Sequenziermethoden – ob mit oder ohne zusätzliche Sequenzmarkierung – liegt in der aufwendigen Probenvorbereitung, die ein erhöhtes Kontaminationsrisiko mit sich bringt. So zeigte sich, dass die Ergebnisse der Studie von GREEN *ET AL.* 2006 stark kontaminiert waren und die enorme Menge an Daten, die mit den neuen Techniken generiert werden kann sorgfältig geprüft werden muss (GREEN *ET AL.* 2009, WALL & KIM 2007). Ein weiterer Nachteil der NGS Methoden ist die höhere Fehlerrate im Vergleich zur Sanger Sequenzierung, wobei jede Plattform ihre spezifischen Fehlerquellen aufweist und somit, je nach Fragestellung mehr oder weniger geeignet ist (KIRCHER & KELSO 2010). Die Sequenzierer von Illumina/Solexa und 454 wurden bisher am häufigsten für palaeogenetische Studien verwendet (KNAPP & HOFREITER 2010). In der Anfangsphase wurden ausschließlich sog. *shotgun runs* durchgeführt (GILBERT *ET AL.* 2007, MILLER *ET AL.* 2008). Bei diesem Verfahren werden keine Marker vorselektiert sondern der gesamte Inhalt eines DNA-Extraktes wird sequenziert und je nach Fragestellung werden spezielle Fragmente des Genoms nachträglich herausgefiltert, wie z.B. die mitochondriale DNA (GILBERT *ET AL.* 2008) oder Mikrosatelliten (ALLENTOFT *ET AL.* 2009). Durch diese Studien wurde deutlich, dass die DNA-Fragmente aus prähistorischen Proben zu einem Großteil mikrobiellen Ursprungs sind und sich der Anteil an endogener DNA oftmals lediglich in einem Bereich von 1% –5% bewegt (BLOW *ET AL.* 2008, MEYER *ET AL.* 2012, RAMIREZ *ET AL.* 2009). Ein Ansatz die Menge an endogener DNA zu erhöhen ist die Behandlung des Extraktes mit Restriktionsenzymen, die bevorzugt exogene DNA schneiden. In einer Studie des Neandertaler Genoms konnte dadurch die Menge an endogener DNA von 3,1% auf 13,1% erhöht werden (GREEN *ET AL.* 2010). Für viele palaeogenetische Fragestellungen sind die mitochondriale DNA oder Teile des nukleären Genoms ausreichend, was *shotgun* Sequenzierungen, aufgrund der enormen Menge exogener DNA-Sequenzen und unerwünschter endogener DNA, unattraktiv macht. Die selektive Anreicherung bestimmter Marker vor der eigentlichen Sequenzierung, bietet eine gute Möglichkeit das Verhältnis von endogener zu exogener DNA positiv zu beeinflussen. Eine Methode besteht in der Voramplifikation relevanter Marker durch Multiplex-PCR (STILLER *ET AL.* 2009) und die in bisherigen Studien am häufigsten angewandte Technik beruht auf der relativen Anreicherung

von DNA-Fragmenten durch *Hybrid Capture* entweder *on array* (BURBANO *ET AL.* 2010), oder *in solution* (AVILA-ARCOS *ET AL.* 2011, BRIGGS *ET AL.* 2009B).

**Tab. B 1** Sequenzierplattformen im Vergleich

<i>Hersteller / Gerät</i>	<i>Spezifikation</i>	<i>Sequenzlänge [Basen]</i>	<i>Seq. / ~24Std.</i>	<i>Durchsatz</i>	<i>Quelle</i>
Applied Biosystems / 3730xl DNA Analyzer	96 Kapillaren / 16 Sequenzierplatten/ rapidseq	bis zu 900	3.840	~3,4 Mb	3730xl APPLIEDBIOSYSTEMS
454 / GS FLX+	GS FLX Titanium XL+	bis zu 1000	1 Mio.	~700 Mb	GS FLX 454
Illumina / HiSeq 2500	rapid run mode / dual flow cell	2 x 100	1,2 Mrd.	~120 Gb	HiSeq 2500 ILLUMINA
Applied Biosystems / 5500xl	nanobeads	bis zu 75	600 Mio.	~45 Gb	5500xl APPLIEDBIOSYSTEMS
Ion Torrent / PGM™	Ion 318™ Chip v2	bis zu 400	18 Mio.	~6 Gb	PGM IONTORRENT
Pacific Biosciences / PacBio RS II	pro SMRT® Cell / P4-C2 Chemie	~ 5500	400.000	~2,2 Gb	PacBio RS II PACIFICBIOSCIENCES

## B 2 Methodische Zielsetzung

Aufgrund des nur eingeschränkt zur Verfügung stehenden wertvollen archäologischen Probenmaterials und der Erkenntnis, dass zur Beantwortung populationsgenetischer Fragestellungen möglichst viele genetische Marker untersucht werden müssen, sollte eine Methode gefunden werden, die es erlaubt eine größere Anzahl an Markern zu untersuchen, ohne dabei zusätzliches Probenmaterial verwenden zu müssen. Die technischen Entwicklungen im Bereich der DNA Sequenzanalyse ermöglichen mittlerweile die parallele Analyse mehrerer Millionen Nukleotide. Um das Potential der neuen Sequenziertechnologien voll ausschöpfen zu können und die Probenbearbeitung effizienter zu gestalten, wurden Methoden gefunden, die es erlauben mehrere Proben gleichzeitig zu prozessieren. Ziel dieser Dissertation war es die neuen Sequenzier Techniken für die speziellen Anforderungen und Fragestellungen der Palaeopopulationsgenetik nutzbar zu machen. Hierzu sollten vorab informative Stellen der kodierenden Region des mitochondrialen Genoms, die Haplogruppen definierende SNPs enthalten, ausgewählt werden. Diese sollten gemeinsam mit der Hypervariablen Region 1 (HVR1) des mitochondrialen Genoms in einem Multiplex-PCR-Ansatz vervielfältigt werden. Gleichzeitig sollten im zweiten Teil dieses Projektes von Sandra Wilde nukleäre Marker ausgewählt werden, von denen vermutet wird, dass sie in einer oder mehreren Populationen unter positiver Selektion stehen. Die nukleären Marker sollten ebenfalls in einem Multiplex-PCR-Ansatz vervielfältigt werden. Die amplifizierten Fragmente sollten anschließend mit probenspezifischen Barcodes versehen werden, um eine parallele Sequenzanalyse auf der Roche GS FLX 454 Plattform zu ermöglichen.

## B 3 Methoden zur Datengenerierung

### B 3.1 Untersuchungsmaterial

Im Zuge dieser Arbeit wurden Knochen- und Zahnproben von 126 Individuen aus 22 Fundplätzen untersucht (s. Tab. B 2). Eine detaillierte Liste mit den Fundangaben zu jeder Probe, die bearbeitet wurde, findet sich in Teil A, Kapitel A3.

**Tab. B 2** Die Tabelle gibt die Anzahl an getesteten Proben aus den jeweiligen Fundregionen wieder

<i>Region</i>	<i>Fundplatz</i>	<i>Koordinaten</i>	<i>Datierung</i>	<i>Kultur</i>	<i>Anzahl</i>
Minusinsker Becken	Barsučij Log	~ 54°00'N 91°11'O	5.Jh.v.Chr.	Tagar	6
Tuva	Aržan 2	~ 52°05'N 93°42'O	7.–6.Jh.v.Chr.	Aldy Bel	23
russisches Altai	Ak Alacha 1, 4, 5	~ 49°17'N 87°32'O	4.–3.Jh.v.Chr.	Pazyryk	11
	Alagail 2	~ 50°12'N 87°42'O	4.–3.Jh.v.Chr.	Pazyryk	3
	Balik Sook	~ 50°48'N 86°00'O	4.–3.Jh.v.Chr.	Pazyryk	2
	Borotal 2	~ 50°12'N 87°42'O	4.–3.Jh.v.Chr.	Pazyryk	2
	Barburgazy 1, 3	~ 49°49'N 89°08'O	4.–3.Jh.v.Chr.	Pazyryk	3
	Dcholin 2	~ 49°48'N 89°22'O	4.–3.Jh.v.Chr.	Pazyryk	1
	Justyd 12, 22	~ 49°46'N 89°18'O	4.–3.Jh.v.Chr.	Pazyryk	8
	Moinak 2	~ 49°19'N 87°35'O	4.–3.Jh.v.Chr.	Pazyryk	2
	Kuturguntas 1	~ 49°25'N 87°35'O	4.–3.Jh.v.Chr.	Pazyryk	2
	Ulandryk 1, 2, 4	~ 49°41'N 89°06'O	4.–3.Jh.v.Chr.	Pazyryk	6
	Verch Kal'džin 1	~ 49°23'N 87°34'O	4.–3.Jh.v.Chr.	Pazyryk	1
kasachisches Altai	Berel'	~ 49°20'N 86°12'O	4.–3.Jh.v.Chr.	Pazyryk	10
	Tar Asu	~ 49°17'N 86°19'O	4.–3.Jh.v.Chr.	Pazyryk	1
	Majemir	~ 49°10'N 84°57'O	unbekannt	unbekannt	3
Ostkasachstan	Ismailovo	~ 50°18'N 81°26'O	9.–7.Jh.v.Chr.	Zevakino-Čilikta	3
	Zevakino	~ 50°12'N 81°49'O	9.–7.Jh.v.Chr.	Zevakino-Čilikta	9
Südkasachstan	Cirikrabort	~ 44°05'N 62°54'O	4.–2.Jh.v.Chr.	Saken	3
West Russland	Kolbinho	~ 51°37'N 39°11'O	3.Jh.v.Chr.	Skythen	5
	Novozavedennoe 2	~ 44°16'N 43°38'O	8.Jh.v.Chr.	initial Skythen	10
	Prokrovka	~ 52°04'N 53°56'O	5.–2.Jh.v.Chr.	Sarmaten	12

## B 3.2 Die Auswahl der mitochondrialen Marker

Ziel war es neben der Hypervariablen Region 1 (HVR1) der Kontrollregion ein Set von Fragmenten der kodierenden Region des mitochondrialen Genoms zu untersuchen, die haplogruppenspezifische Mutationen enthalten. Hintergrund ist die Tatsache, dass durch die alleinige Analyse der HVR1 aufgrund von auftretenden Parallel- oder Rückmutationen nicht immer eindeutig die Zugehörigkeit zu einer Haplogruppe bestimmt bzw. keine Aufschlüsselung in Sub-Haplogruppen vorgenommen werden kann (FINNILÄ *ET AL.* 2001). Für die zuverlässige Haplogruppen-Zuordnung ist es deshalb wichtig, die Daten der HVR1 und der kodierenden Region gemeinsam zu betrachten (VAN OVEN & KAYSER 2009). Da in der Eurasischen Steppe mit einer Mischpopulation aus west- und osteurasischen Linien zu rechnen ist (COMAS *ET AL.* 2004), sollten die untersuchten Marker in der Lage sein, die mögliche Variationsbreite an maternalen Linien dieses Gebietes zu erfassen. Die Auswahl der mitochondrialen SNPs musste deshalb so getroffen werden, dass aus einer möglichst geringen Anzahl zuverlässig alle Haupthaplogruppen bestimmt werden können. Es wurden SNPs der kodierenden Region des mitochondrialen Genoms ausgewählt, von denen in mehreren Publikationen gezeigt wurde, dass sie mit einer der Haupthaplogruppen korrelieren (s. Tab. B 3). Auf Vorschlag von Antonio Torroni wurden zusätzlich einige SNPs mit aufgenommen, die eine Einteilung in Subhaplogruppen ermöglichen. Bei den SNPs handelt es sich jeweils um Basenveränderungen relativ zur mitochondrialen Referenzsequenz, der rCRS (*revised Cambridge Reference Sequence* – ANDREWS *ET AL.* 1999). Das mitochondriale Referenzgenom lässt sich der Haplogruppe H zuordnen und wurde willkürlich als Referenz festgelegt, da es das erste vollständig sequenzierte mitochondriale Genom eines Menschen war (ANDERSON *ET AL.* 1981, ANDREWS *ET AL.* 1999). Im Jahr 2012 wurde vorgeschlagen eine neue Referenz zu verwenden, die an der, evolutionär gesehen, eigentlichen Wurzel des mitochondrialen Stammbaums liegt, die RSRS (*Reconstructed Sapiens Reference Sequence*) (BEHAR *ET AL.* 2012). Zurzeit werden beide Referenzen parallel verwendet, wobei für die vorliegende Arbeit alle Mutationen relativ zur rCRS angegeben wurden.



Tab. B 3 SNPs der kodierenden Region des mitochondrialen Genoms

<i>Position</i>	<i>SNP</i>	<i>Haplogruppe</i>	<i>Referenz</i>
456	C -> T	H5 <sup>1</sup>	GRIGNANI <i>ET AL.</i> 2006, BRANDSTÄTTER <i>ET AL.</i> 2008
663	A -> G	A	HERRNSTADT <i>ET AL.</i> 2002, TAMM <i>ET AL.</i> 2007, TORRONI <i>ET AL.</i> 1992
3010	G -> A	H1 <sup>1</sup> , J1 <sup>1</sup> , D4 <sup>1</sup>	FINNILÄ <i>ET AL.</i> 2001, KIVISILD <i>ET AL.</i> 2002, BRANDSTÄTTER <i>ET AL.</i> 2003, TAMM <i>ET AL.</i> 2007
4216	T -> C	T, J, R2 <sup>1</sup>	FINNILÄ <i>ET AL.</i> 2001, PALANICHAMY <i>ET AL.</i> 2004
4529	A -> T	I	TORRONI <i>ET AL.</i> 1994, FINNILÄ <i>ET AL.</i> 2001
4580	G -> A	V	FINNILÄ <i>ET AL.</i> 2001, HERRNSTADT <i>ET AL.</i> 2002
4833	A -> G	G	KIVISILD <i>ET AL.</i> 2002, KONG <i>ET AL.</i> 2006
4917	A -> G	T	FINNILÄ <i>ET AL.</i> 2001, PALANICHAMY <i>ET AL.</i> 2004
5178	C -> A	D	TORRONI <i>ET AL.</i> 1992, KIVISILD <i>ET AL.</i> 2002, TAMM <i>ET AL.</i> 2007
5843	A -> G	Q	FRIEDLAENDER <i>ET AL.</i> 2005, MERRIWETHER <i>ET AL.</i> 2005
6371	C -> T	X	FINNILÄ <i>ET AL.</i> 2001, TAMM <i>ET AL.</i> 2007
6392	T -> C	F, U8a <sup>1</sup>	FINNILÄ <i>ET AL.</i> 2001, TANAKA <i>ET AL.</i> 2004, KONG <i>ET AL.</i> 2006
6776	T -> C	H3 <sup>1</sup>	HERRNSTADT <i>ET AL.</i> 2002, BRANDSTÄTTER <i>ET AL.</i> 2003
7028	C -> T	alle außer H	TORRONI <i>ET AL.</i> 1994, BRANDSTÄTTER <i>ET AL.</i> 2003
8271-8281	9bp del	B	TORRONI <i>ET AL.</i> 1992, TAMM <i>ET AL.</i> 2007
8392	G -> A	Y	TANAKA <i>ET AL.</i> 2004, DERENKO <i>ET AL.</i> 2007
8994	G -> A	W, H21 <sup>1</sup>	FINNILÄ <i>ET AL.</i> 2001, HERRNSTADT <i>ET AL.</i> 2002, BRANDSTÄTTER <i>ET AL.</i> 2008
9090	T -> C	Z	FINNILÄ <i>ET AL.</i> 2001, KONG <i>ET AL.</i> 2006
10034	T -> C	I	FINNILÄ <i>ET AL.</i> 2001, PALANICHAMY <i>ET AL.</i> 2004
10118	T -> C	P1	FRIEDLAENDER <i>ET AL.</i> 2005
10400	C -> T	M	MACA-MEYER <i>ET AL.</i> 2001, HERRNSTADT <i>ET AL.</i> 2002, KIVISILD <i>ET AL.</i> 2002
10873	T -> C	L	MACA-MEYER <i>ET AL.</i> 2001, COMAS <i>ET AL.</i> 2004, TORRONI <i>ET AL.</i> 2006
11719	G -> A	R0 <sup>1</sup>	ACHILLI <i>ET AL.</i> 2004, BRANDSTÄTTER <i>ET AL.</i> 2008
11947	A -> G	W	HERRNSTADT <i>ET AL.</i> 2002, PALANICHAMY <i>ET AL.</i> 2004
12308	A -> G	K, U	FINNILÄ <i>ET AL.</i> 2001, HERRNSTADT <i>ET AL.</i> 2002
12705	C -> T	N	KIVISILD <i>ET AL.</i> 2002, COMAS <i>ET AL.</i> 2004
13263	A -> G	C	TORRONI <i>ET AL.</i> 1992, TAMM <i>ET AL.</i> 2007
13626	C -> T	E	FRIEDLAENDER <i>ET AL.</i> 2007
13708	G -> A	J, H18 <sup>1</sup>	TORRONI <i>ET AL.</i> 1994, FINNILÄ <i>ET AL.</i> 2001, BRANDSTÄTTER <i>ET AL.</i> 2003, 2008
14766	C -> T	alle außer H, HV und V	FINNILÄ <i>ET AL.</i> 2001, BRANDSTÄTTER <i>ET AL.</i> 2003
14798	T -> C	K, J1c <sup>1</sup>	FINNILÄ <i>ET AL.</i> 2001, BRANDSTÄTTER <i>ET AL.</i> 2003

<sup>1</sup> Diese zusätzlichen Marker wurden der ursprünglichen Auswahl von A. Torroni hinzugefügt.

### B 3.3 Die nukleären Marker

Die Auswahl der nukleären Marker und der Entwurf der zur Amplifikation benötigten Primer sowie des Reaktionsansatzes wurden von S. W. vorgenommen. Die Auswahl war auf Marker beschränkt, deren phänotypische Ausprägung wahrscheinlich positiv selektiert wurde. Zusätzlich wurde der Amelogenin *locus* untersucht, dessen Länge sich auf dem X- und dem Y-Chromosom unterscheidet und somit eine Geschlechtsbestimmung erlaubt. Insgesamt wurden 21 nukleäre *loci* untersucht (s. Tab. B 4). Ursprünglich wurden 23 Markern ausgewählt, zwei davon jedoch nachträglich von den Analysen ausgeschlossen, da sie unspezifische Sequenzergebnisse lieferten. In den Berechnungen zum Sequenziererfolg und den Kontaminationen sind diese Marker jedoch enthalten.

**Tab. B 4 Die nukleären Marker**

Die Marker sind nach ihrer Funktion sortiert. Zu jedem sind die Referenznummer, die entscheidende Mutation und die Positionsangabe im Genom angegeben

	Name	dbSNP	SNP	selektiertes Allel	Lage	Chromosom
Pigmentierung	HERC2	rs12913832	A>G	G	Intron 86 /Promotor OCA2	15q13.1
	SLC24A5	rs1426654	G>A	A	e3: Ala111Thr	15q21.1
	SLC45A2	rs16891982	C>G	G	e5: C1122G; Leu374Phe	5p13.3
	TYR	rs1042602	C>A	A	Ser192Tyr	11q14-21

Kalzium & Vitamin D	LCTa	rs4988235	C>T	T	C-13910T; Intron MCM6	2q21
	LCTb	rs182549	G>A	A	G-22018A; Intron MCM6	
	CYP3A4	rs2740574	A>G	A	Promotor	7q22.1
	SLC12A3	rs1529927	G>C	C	e6: Ala264Gly	16q13
	TRPV6b	rs4987667	G>A	A	e9: Val378Met	7q33-34
	TRPV6c	rs4987682	C>T	T	e15: Thr681Met	

Klima	EDAR	rs3827760	T>C	C	T1540C; Val370Ala	2q11-13
	ABCC11	rs17822931	C>T	T	G538A (C/T); Gly180Arg	16q12.1
	CYP3A5	rs776746	A>G	G	Intron	7q22.1

	Name	dbSNP	SNP	selektiertes Allel	Lage	Chromosom
Xenobiotika	ADH1Ba	rs3811801	C>T	T	regulatorische Region	4q22
	ADH1Bb	rs1229984	G>A	A	Arg47His	
	NAT2	rs1801280	T>C	C	T341C; Ile114Thr	8p23.1-21.3
	ABCB1a	rs1128503	C>T	T	e12: C1236T; synonym	7q21.1
	ABCB1b	rs2032582	G>A/T	T	e21: G2677T/A; Ala893Ser/Thr	
	ABCB1c	rs1045642	C>T	T	e26: C3435T; synonym	
	CASP12	rs497116	C>T	T	Arg125Stop	11q22.3
♀♂	AMEL		6bp Indel			Xp22.3-.1 Yp11

## B3.4 DNA-Anreicherungsverfahren

### Polymerase Kettenreaktion

Die am weitesten verbreitete Methode DNA anzureichern ist die von Kary B. Mullis entwickelte Polymerase Kettenreaktion (PCR - *polymerase chain reaction*) (MULLIS & FALOONA 1987). Hierzu werden Oligonukleotide (Primer) synthetisiert, die gesuchte DNA Regionen flankieren. In einem zyklischen Prozess werden alle DNA-Stränge denaturiert, die Primer hybridisiert und an ihren freien 3' OH-Enden mit Hilfe einer Polymerase und freien Nukleotidbausteinen komplementär zum gesuchten DNA Abschnitt verlängert. Somit wird der gewünschte DNA Abschnitt exponentiell vermehrt.

Die Standard-PCR wurde eingesetzt um die Proben auf ihren DNA-Gehalt hin zu untersuchen, oder um einzelne Fragmente zu amplifizieren, die in den Multiplex-Reaktionen ausgefallen sind.

**Tab. B 5 Standard-PCR-Ansatz**

<i>Für einen 50 µL Ansatz:</i>	
<i>Chemikalien</i>	<i>Menge [µL]</i>
HPLC-H <sub>2</sub> O	27,5-34,5
Puffer	5
MgCl <sub>2</sub>	5
dNTPs	1
BSA	1
upper Primer	0,5-1
lower Primer	0,5-1
Taq	0,5
Target	2-8

**Tab. B 6 Laufkonditionen eines Standard-PCR-Ansatzes**

<i>Programm</i>	<i>Temp</i>	<i>Zeit</i>
initial	94°C	6Min
Denaturierung	94°C	40s
Hybridisierung	56-58°C	40s
Elongation	72°C	40s
Zyklen	38-45	

## Multiplex-PCR

Die Multiplex-PCR ist eine Weiterentwicklung des Standard-PCR-Ansatzes, bei der statt einem Primerpaar mehrere parallel zum Einsatz kommen. Somit können bei gleicher DNA Einsatzmenge mehrere DNA-Fragmente parallel vervielfältigt werden.

### *Mitochondriale Multiplex-PCR*

Zur Amplifikation der ausgewählten SNPs der kodierenden Region wurden 30 Primerpaare entworfen. Des Weiteren wurden sieben überlappende Primerpaare für die HVR1 von Position 16011-16413 und alternativ vier Primerpaare, die den gleichen Bereich abdecken (16013-16410) hinzugefügt. Beim Entwerfen der Primer wurde darauf geachtet, dass diese später alle bei einer Hybridisierungstemperatur von 56°C eingesetzt werden können. Die Unterschiede zwischen den Schmelztemperaturen der einzelnen Primer sollten jeweils nicht mehr als 4°C betragen. Außerdem sollten die Primer an ihren 3'Enden eine Adenin oder Thymin Base tragen. Hintergrund ist die Tatsache, dass Adenin und Thymin nur zwei Wasserstoffbrücken zueinander aufbauen, statt wie Cytosin und Guanin drei, ihre Bindung zueinander also weniger stark ist. Dies führt dazu, dass das 3'-OH-Ende des Primers erst an die DNA bindet und somit verlängert werden kann, wenn der Rest des Primers bereits gebunden ist. Dadurch sollte eine spezifischere Bindung des Primers und somit eine spezifischere Amplifikation des gesuchten Fragmentes gewährleistet werden. Die für den Multiplex-PCR-Ansatz ausgewählten Primer (s. Tab. B 7) wurden auf drei Ansätze (s. Tab. B 8) verteilt um die überlappenden Fragmente zu trennen und so die Bildung von unerwünschten Nebenprodukten zu verringern.

Tab. B 7 Primer für die Multiplex-PCR

Primer	Sequenz 5' → 3'	Länge	$T_m$	$T_{Ann}$	$L_{Prod}$	SNP Position	SNP
423_U	AATTTTATCTTTTGGCGGTATGCACTT	27-mer	58,6	52,8	110	456	C->T
485_L	GATGGGCGGGGGTTGTATTG	20-mer	58,5				
654_U	CTCACATCACCCATAAACAAATAGG	26-mer	56,8	51,3	94	663	A->G
699_L	AACTCACTGGAACGGGGATGCT	22-mer	58,5				
2992_U	CAACAATAGGGTTTACGACCTCGAT	25-mer	56,6	53,6	116	3010	G->A
3057_L	CTCCGGTCTGAACTCAGATCACGTA	25-mer	58,4				
4155_U	TACCCCGATTCCGCTACGA	20-mer	58,9	52,9	113	4216	T->C
4221_L	ATGCTGGAGATTGTAATGGGTATGGA	26-mer	58,7				
4499_U	CTGGCCAACCCGTCATCTA	20-mer	57,0	53,5	103	4529	A->T
4554_L	GCATGTTTATTCTAGGCCTACTCAGG	27-mer	57,1				
4549_U	ACAGCGCTAAGCTCGCACTGAT	22-mer	58,0	51,9	113	4580	G->A
4617_L	ATGGCAGCTTCTGTGGAACGAG	22-mer	58,0				
4815_U	GAATAGCCCCCTTTCACCTCTGAGTC	26-mer	58,7	55,3	101	4833	A->G
4864_L	TGAGATGGGGGCTAGTTTTTGTTCAT	25-mer	58,9				
4871_U	GGCCTGCTTCTTCTCACATGACA	23-mer	58	53,4	120	4917	A->G
4940_L	ACTGCCTGCTATGATGGATAAGATTGA	27-mer	58,3				
5163_U	CCAGCACCACGACCCTACTACTATCT	26-mer	58	51,5	71	5178	C->A
5179_L	GGATGGAATTAAGGGTGTAGTCATGTT	28-mer	58				
5836_U	AAATCACCTCGGAGCTGGTAAAAAAG	25-mer	58,0	52,2	88	5843	A->G
5875_L	GGGGTGAGGTAATAATGGCTGAGT	23-mer	57,2				
6336_U	CACCCTGGAGCCTCCGTAGAC	21-mer	57,3	53,5	116	6371	C->T
6403_L	ATGGCAGGGGGTTTTATATTGATAATT	27-mer	57,5			6392	T->C
6764_U	CAATTGGCTCCTAGGGTTTATCGT	25-mer	57,8	51,8	101	6776	T->C
6814_L	GATGATTATGGTAGCGGAGGTGAAA	25-mer	57,4				
6975_U	GGTGGCCTGACTGGCATTGTA	21-mer	57,1	53,1	120	7028	C->T
7046_L	TATGATGGCAAATACAGCTCCTATTGA	27-mer	57,2				
8226_U	CATGCCCATCGTCTTAGAATTAA	23-mer	54,9	52,6	112	8271-8281	9bp del
8287_L	GCTAAGTTAGCTTTACAGTGGGCTCTA	27-mer	55,4				
8385_U	TACAGTGAAATGCCCAACTAAATACTA	28-mer	55,9	50,7	89	8392	G->A
8417_L	TTTAGTTGGGTGATGAGGAATAGTGTA	28-mer	55,7				
8932_U	ACTTCTTACCACAAGGCACACCTACA	26-mer	57,2	53,4	116	8994	G->A
8996_L	AGTAATGTTAGCGGTTAGGCGTACG	25-mer	57,0				
9072_U	GAAGCGCCACCCTAGCAATATC	22-mer	56,7	51,3	101	9090	T->C
9124_L	TAAGGCGACAGCGATTCTAGGATAG	26-mer	58,3				
10000_U	CATCTATTGATGAGGGTCTTACTCTTTTA	29-mer	54,3	47,2	107	10034	T->C
10048_L	AAATTAAGGCGAAGTTTATTACTCTTTTT	29-mer	54,3				
10105_U	TTAATAATCAACACCCTCCTAGCCTTAC	28-mer	56	51,1	109	10118	T->C
10166_L	GGTCGAAGCCGCACTCGTA	19-mer	56,1				
10387_U	TCTGGCCTATGAGTGACTACAAAAAG	26-mer	55,1	48,8	117	10400	C->T
10451_L	AGGGGCATTTGGTAAATATGATTATC	26-mer	55,1				
10865_U	CAACCACCCACAGCCTAATTATTAGC	26mer	58,2	49,9	83	10873	T->C
10895_L	TGGGGAACAGCTAAATAGGTTGTTGT	26mer	58,2				
11700_U	AGCTTACC GGCGCAGTCA	19-mer	59	53,2	89	11719	G->A
11743_L	GTGCGTTCGTAGTTTGAGTTTGCTAG	26-mer	57,6				
11935_U	ACCACGTTCTCCTGATCAAATATCAC	26-mer	56,5	51,6	101	11947	A->G
11983_L	CCCCATTGTGTTGTGGTAAATATGTA	26-mer	55,9				
12303_U	GATAACAGCTATCCATTGGTCTTAGGC	27-mer	58,2	51,1	103	12308	A->G
12352_L	GGAAGTCAGGGTTAGGGTGGTTATAG	26-mer	58				
12692_U	CAGACCCAAACATTAATCAGTTCTTCA	27-mer	56,8	51,0	110	12705	C->T
12754_L	GCCCTCTAGCCGATGAACA	20-mer	57,4				

Primer	Sequenz 5' → 3'	Länge	$T_m$	$T_{Ann}$	$L_{Prod}$	SNP Position	SNP
13231_U	GCGCCCTTACACAAAATGACATC	23-mer	57,3	51,1	94	13263	A->G
13275_L	GGTTGGTTGATGCCGATTGTAACAT	26-mer	58,3				
13620_U	AAGCGCCTATAGCACTCGAATAATTCT	27-mer	58,3	53,8	116	13626	C->T
13683_L	CCAGGCGTTTAATGGGGTTTAGTAG	25-mer	58,2				
13701_U	ACCCACCCCTACTAAACCCCATTA	25-mer	58,9	53,7	88	13708	G->A
13740_L	GATGCGGGGAAATGTTGTTAGT	23-mer	57,9				
14717_U	CAACCACGACCAATGATATGAAAAAC	26-mer	57,5	51,4	120	14766	C->T
14784_L	GGAGTTCGATGAATGAGTGGTTAATT	26-mer	57,6				
14783_U	ATACGCAAACTAACCCCTAATAAAA	27-mer	56,3	52,5	105	14798	T->C
14839_L	GCCAAGGAGTGAGCCGAAGTT	21-mer	57,1				

**HVR1 Primer**

Primer	Sequenz 5' → 3'	Länge	$T_m$	$T_{Ann}$	$L_{Prod}$
16011_U	AGCACCCAAAGCTAAGATTCTAATTT	26-mer	54,7	51,9	130
16088_L	GTGGCTGGCAGTAATGTACGAAATAC	26-mer	57,1		
16071_U	GGGTACCACCCAAGTATTGACTCA	24-mer	55,8	52,2	134
16153_L	TGATGTGGATTGGGTTTTATGTACTA	27-mer	55,0		
16119_U	GTACATTACTGCCAGCCACCATG	23-mer	55,9	52,5	138
16207_L	TGATAGTTGAGGGTTGATTGCTGTAC	26-mer	55,5		
16185_U	TACATAAAAACCCAATCCACATCAAAAAC	28-mer	57,6	54,1	139
16271_L	GGTGGGTAGGTTTGTGGTATCCT	24-mer	56,2		
16233_U	AGTACAGCAATCAACCCTCAACTATC	26-mer	54,1	52,4	127
16305_L	TGTACGGTAAATGGCTTTATGTACTATG	28-mer	54,4		
16274_U	AAAGCCACCCCTCACCCACTAG	22-mer	58,1	53,5	116
16345_L	TGGGGACGAGAAGGGATTGAC	22-mer	58,8		
16340_U	ACATAAAGCCATTTACCGTACATAGCAC	28-mer	57,2	54,7	127
16413_L	CACTCTTGTGCGGGATATTGATTTC	25-mer	57,8		

**Alternative HVR1 Primer**

16013_U	AGCACCCAAAGCTAAGATTCTAATTTAA	28-mer	56	52,9	196
16152_L	TGATGTGGATTGGGTTTTATGTACTAC	28-mer	55,8		
16122_U	ATTACTGCCAGCCACCATGAAT	22-mer	55	54,3	196
16271_L	GGTGGGTAGGTTTGTGGTATCCT	24-mer	56,2		
16235_U	AAGTACAGCAATCAACCCTCAACTATCAC	29-mer	58,3	54,5	163
16346_L	ATGGGGACGAGAAGGGATTGGA	22-mer	58,3		
16274_U	AAAGCCACCCCTCACCCACTAG	22-mer	58,1	55,9	182
16410_L	TTGTGCGGGATATTGATTTACG	23-mer	58,4		

**Tab. B 8 Verteilung der Primer auf die drei Multiplex-PCR-Ansätze**

<i>Ansatz A</i>		<i>Ansatz B</i>		<i>Ansatz C</i>	
<i>SNP</i>	<i>Primer</i>	<i>SNP</i>	<i>Primer</i>	<i>SNP</i>	<i>Primer</i>
456	423_U	663	654_U	3010	2992_U
	485_L		699_L		3057_L
4216	4155_U	4529	4499_U	4580	4549_U
	4221_L		4554_L		4617_L
4833	4815_U	4917	4871_U	5178	5163_U
	4864_L		4940_L		5179_L
5843	5836_U	6371	6336_U	6776	6764_U
	5875_L	6392	6403_L		6814_L
7028	6975_U	8271-8281	8226_U	8392	8385_U
	7046_L		8287_L		8417_L
8994	8932_U	9090	9072_U	10034	10000_U
	8996_L		9124_L		10048_L
10115	10105_U	10398	10387_U	10873	10865_U
	10166_L		10451_L		10895_L
11719	11700_U	11947	11935_U	12308	12303_U
	11743_L		11983_L		12352_L
12705	12692_U	13263	13231_U	13626	13620_U
	12754_L		13275_L		13683_L
13708	13701_U	14766	14717_U	14798	14783_U
	13740_L		14784_L		14839_L

<b>HVR1 Primer</b>	
I	16011_U
	16088_L
IV	16185_U
	16271_L
VII	16340_U
	16413_L

II	16071_U
	16153_L
V	16233_U
	16305_L

III	16119_U
	16207_L
VI	16274_U
	16345_L

**Tab. B 9 mitochondrialer Multiplex-PCR-Ansatz**

<i>Für einen 40µL Ansatz:</i>	
<i>Chemikalien</i>	<i>Menge [µL]</i>
Target	4-6
Qiagen-MM	20
Wasser	9-11
Primer:	
Ansatz A	5,2
Ansatz B	4,8
Ansatz C	4,8



**Tab. B 10 Laufkonditionen der mitochondrialen Multiplex-PCR**

<i>Programm</i>	<i>Temp</i>	<i>Zeit</i>
Initiale Denaturierung	95°C	15 Min
Denaturierung	94°C	40 s
Hybridisierung	56-57°C	90 s
Elongation	72°C	90 s
finale Elongation	72°C	10 Min
Zyklen	36-45	

*Nukleäre Multiplex-PCR*

Die Auswahl der nukleären Marker und das Design des Ansatzes wurden von Sandra Wilde im Zuge ihres Teilprojektes vorgenommen.

Die Primerpaare für die 21 nukleären Fragmente wurden nach dem gleichen Prinzip entworfen, das im vorherigen Abschnitt bereits beschrieben wurde. Die Hybridisierungstemperaturen wurden hier jedoch auf 58°C abgestimmt. Da die Fragmente nicht überlappen, können sie in einem einzigen Multiplex-Ansatz vervielfältigt werden.

**Tab. B 11 Primerpaare zur Amplifikation der nukleären Marker**

<i>Name</i>	<i>Primer</i>	<i>Sequenz 3' → 5'</i>
ABCB1a	S01U	TGTCTGTGAATTGCCTTGAAGTTTTT
	S01L	GCCACCGTCTGCCCACTCT
ABCB1b	S02U	GGACAAGCACTGAAAGATAAGAAAGAACT
	S02L	ATTAATCAATCATATTTAGTTTGACTCACCT
ABCB1c	S03U	AACAGCCGGGTGGTGTCA
	S03L	GACTCGATGAAGGCATGTATGTTGG
ABCC11	S04U	AGCCTAGAGTCCCCAAACCTCAC
	S04L	CACTTCTGGGCATCTGCTTCTGC
ADH1Ba	S05U	AAATGAAATTCAAGAACATCTCTGACC
	S05L	GTTTTAGGGCCATTCTCTTTTGT
ADH1Bb	S06U	TCTAGTTACAATCTTTTCTGAATCTGAACAG
	S06L	GGGTCACCAGGTTGCCACTAA
CASP12	S08U	AATCGAAATGGTTCTGAACTTGACCT
	S08L	GGAGCCATTACCTGAGCTGTGAGATT
CYP3A4	S09U	GTAGGTGTGGCTTGTGGGATGAA
	S09L	AAGTGGAGCCATTGGCATAAAATCTAT
CYP3A5	S10U	CCACCCAGCTTAACGAATGCTCTA
	S10L	GATGAAGGGTAATGTGGTCCAAACAG
EDAR	S11U	AATGCTCAGCTCCACGTACAACCTCTG
	S11L	GCCCCAATCTCATCCCTCTTC
HERC2	S12U	CCAAGAGGCGAGGCCAGTTTCA
	S12L	AAGCCTCGGCCCTGATGATG

Name	Primer	Sequenz 3' → 5'
LCTa	S13U	CTGCGCTGGCAATACAGATAAGA
	S13L	CAAATGCAACCTAAGGAGGAGAGTT
LCTb	S14U	AAGATGTCCTTAAAAACAGCATTCTCA
	S14L	CCAAAGTACTGGGACAAAGGTGTG
NAT2	S15U	CACTGGCATGGTTCACCTTCTCC
	S15L	CCAGACCCAGCATCGACAATGTAA
SLC12A3	S16U	CGTGGACCCCATTAACGACATC
	S16L	ATGGCCTCCTCACCTTGGACTC
SLC24A5	S17U	GTTCAGCCCTTGGATTGTCTCAGG
	S17L	GGAGCAATATTTACCTAGGAAAGCAGT
SLC45A2	S18U	AGAATAAAGTGAGGAAAACACGGAGTT
	S18L	GAAAGAGGAGTCGAGGTTGGA
TRPV6b	S20U	AGGCCTGAAAGCCCTGAGCAT
	S20L	ACCAGCCGGATATCGTCCTTAGG
TRPV6c	S21U	GGAAAACTAGAGCTGGGCTGTCC
	S21L	GGCACTGCTGCGGGAGGTA
TYR	S22U	TTTGTCTGGATGCATTATTATGTGTCA
	S22L	CTTCATGGGCAAAATCAATGTCTC
AMEL X/Y	S23U	CCCTGGGCTCTGTAAAGAATAGTG
	S23L	ATCAGAGCTTAAACTGGGAAGCTG

Tab. B 12 nukleärer Multiplex-PCR-Ansatz

Für einen 80 µL Ansatz	
Chemikalien	Menge [µL]
HPLC-H <sub>2</sub> O	23,6
MgCl <sub>2</sub>	20
Puffer	12
Primermix	9,2
dNTPs	3,2
BSA	2,4
Taq	1,6
Target	8

Tab. B 13 Laufkonditionen der nukleären Multiplex-PCR

Programm	Temp.	Zeit
Initiale Denaturierung	94°C	6 min
Denaturierung	94°C	40 s
Hybridisierung	58°C	60 s
Elongation	72°C	60 s
Zyklen	46	
final	72°C	10 min

## Agilent *SureSelect Target Enrichment*

Parallel zur DNA Anreicherung durch die oben beschriebenen PCR-Ansätze wurde das von Agilent entwickelte *SureSelect Target Enrichment XT* Protokoll getestet. Die Methode basiert auf einer relativen Anreicherung bestimmter Marker durch deren selektive Extraktion aus einem DNA Sequenzgemisch. Hierfür werden, komplementär zu den gesuchten DNA Sequenzabschnitten, 120 Basen lange RNA Sequenzen entworfen, sog. *baits*. Die *baits* und die gesuchten DNA-Fragmente werden in Lösung hybridisiert und über magnetische *beads*, die über eine Biotin/Streptavidin-Bindung an die *baits* gebunden werden, aus dem Sequenzgemisch isoliert (*Capture*). Für diese Studie wurde ein von S. K. erstellter *bait*-Katalog (KREUTZER 2011) verwendet, der das gesamte mitochondriale Genom abdeckt.

Abgesehen von nachstehenden Änderungen erfolgte die Durchführung des *Captures* nach Herstellerprotokoll:

Der erste Schritt, die mechanische Fragmentierung der DNA-Stränge, ist bei der Arbeit mit alter DNA überflüssig, da diese bereits fragmentiert vorliegt. Im nächsten Schritt wird die DNA, um sie auf die Sequenzierung vorzubereiten, in eine sog. *library* überführt. Dies erfolgte nach dem Protokoll von MEYER & KIRCHER 2010 anstelle des vorgesehenen Illumina Protokolls (s. Kap. B3.6 Illumina *Library Preparation*). Des Weiteren wurde direkt im Anschluss an die *library*-Präparation, anstatt nach dem *Capture*, eine Index-PCR durchgeführt um die Proben mit individuellen Barcodes zu versehen. Die so markierten DNA-Fragmente verschiedener Proben können nun als eine Probe gemeinsam weiter bearbeitet werden. Durch den Barcode ist es möglich einzelne DNA-Fragmente am Ende wieder der ursprünglichen Probe zuzuordnen.

Die benötigte Einsatzmenge für ein *Capture* beträgt 500ng DNA in einem maximalen Volumen von 3,4µL. Aufgrund der geringen DNA Konzentration in aDNA Proben, wurden mehrere Proben gemeinsam *gecaptured*. Dafür wurden von sechs skythenzeitlichen Proben *libraries* erstellt. Diese wurden äquimolar vereinigt, um die insgesamt benötigte DNA-Menge von 500ng zu erreichen. Hierzu wurden die Konzentrationen der Proben mit Hilfe des Agilent 2100 Bioanalyzers bestimmt. Für die Messung wurde der *High Sensitivity DNA Assay*-Chip verwendet. Die Chip- und Probenvorbereitung erfolgte entsprechend des Herstellerprotokolls. Zur Berechnung der benötigten Einsatzmenge je Probe wurde jeweils der Hauptpeak des Bioanalyser-Elektropherogramms verwendet. Die Konzentration wurde

dann dazu verwendet die DNA Menge der *library* zu errechnen und zu ermitteln wie viel  $\mu\text{L}$  eingesetzt werden müssen um die geforderten 500ng zu erreichen. Die berechneten Volumina werden vereinigt und müssen, falls sie die  $3,4\mu\text{L}$  überschreiten mit Hilfe einer Vakuumzentrifuge bei maximal  $45^\circ\text{C}$  eingengt werden, bis die gesamte Flüssigkeit verdampft ist. Das entstandene Pellet wird anschließend in  $3,4\mu\text{L}$  Elutionspuffer des Qiagen *MinElute PCR Purification Kits* gelöst.

Durch die Verwendung des *library* Protokolls nach Meyer & Kircher 2010 tragen die DNA-Fragmente Adapter an ihren Enden, die vom Illumina-Protokoll abweichen. Da die Adapter während des Hybridisierungsschrittes blockiert werden müssen, um eine Dimer-Bildung und somit das unspezifische Extrahieren unerwünschter DNA-Fragmente zu vermeiden, muss der im Protokoll erwähnte *SureSelect Indexing Block #3* durch die *Blocking-Oligos* von Meyer ersetzt werden.

**Tab. B 14 Blocking-Oligos des Meyer Protokolls.**

Es wird ein Mix aus allen Oligonukleotiden erstellt, so dass diese je in einer Konzentration von  $100\mu\text{M}$  vorliegen.

<i>Block</i>	<i>Sequenz 5' → 3'</i>
BO1.P5.F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-Pho
BO2.P5.R	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-Pho
BO3.P7.part1.F	AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-Pho
BO4.P7.part1.R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-Pho
BO5.P7.part2.F	ATCTCGTATGCCGTCTTCTGCTTG-Pho
BO6.P7.part2.R	CAAGCAGAAGACGGCATACGAGAT-Pho

**Tab. B 15 Ansatz der Blocking-Oligos & der vereinigten libraries**

<i>Reagenz</i>	<i>Menge</i>
Librarypool	$3,4\mu\text{L}$
SureSelect Indexing Block #1	$2,5\mu\text{L}$
SureSelect Block #2	$2,5\mu\text{L}$
Block Oligo Mix nach Meyer (Tab. B 14)	$0,6\mu\text{L}$

Das *Capture*-Produkt wird mit Hilfe des Qiagen *MinElute PCR Purification Kits* aufgereinigt und in  $33\mu\text{L}$  Elutionspuffer gelöst.

*Amplifikation nach der Hybridisierung*

Die DNA-Fragmente aus dem *Capture*-Prozess werden nun in drei parallelen PCR-Ansätzen vervielfältigt. Der PCR-Ansatz wird auf Eis pipettiert.

**Tab. B 16 PCR-Ansatz zur Amplifikation der DNA-Fragmente aus der Hybridisierungsreaktion**

<i>Reagenz</i>	<i>Volumen/Probe [<math>\mu\text{L}</math>]</i>
HPLC H <sub>2</sub> O	24
5x Herculase II Puffer	10
DMSO (100%)	2,5
dNTPs (25mM)	0,5
Primer IS5 (10 $\mu\text{M}$ )	1,25
Primer IS6 (10 $\mu\text{M}$ )	1,25
Herculase II Fusion Pol.	0,5
Template	10
Gesamt	50

**Tab. B 17 Laufkonditionen zur Amplifikation der *Capture*-Produkte**

<i>Programm:</i>	<i>Temp. [<math>^{\circ}\text{C}</math>]</i>	<i>Zeit [s]</i>
initiale Denaturierung	98	30
Denaturierung	98	10
Hybridisierung	60	30
Elongation	72	30
Zyklen	10	

Anschließend werden die PCR-Produkte wieder vereinigt, indem sie mit dem Qiagen *MinElute PCR Purification Kit* über einen Filter aufgereinigt und in 33 $\mu\text{L}$  EB gelöst werden.

*Konzentrationsbestimmung und –einstellung für die Sequenzierung*

Für die Illumina-Sequenzierung werden von der Firma GATC Biotec AG 100ng DNA in 5 $\mu\text{l}$  veranschlagt. Da drei verschiedene *Capture* gemeinsam sequenziert werden sollten, wurden die Konzentrationen auf 150ng DNA in 7,5 $\mu\text{l}$  eingestellt. Zur Berechnung des benötigten Probenvolumens für einen äquimolaren Probenpool wurden die DNA-Konzentrationen nach dem *Target Enrichment* erneut mit dem Agilent 2100 Bioanalyser bestimmt. Die errechneten Volumina wurden vereinigt und mit Hilfe der Vakuumzentrifuge bei maximal 45 $^{\circ}\text{C}$  eingeeengt. Das entstandene Pellet wurde anschließend in 7,5 $\mu\text{L}$  Elutionspuffer des Qiagen *MinElute PCR Purification Kits* gelöst.

### B3.5 Markierung der DNA-Fragmente zur parallelen Sequenzanalyse

Die parallele Analyse unterschiedlicher Proben auf den NGS Plattformen erfordert eine Markierung der DNA-Fragmente durch eine Erkennungssequenz, um sie nach der Analyse den jeweiligen Proben zuordnen zu können. Diese Erkennungssequenz, je nach Protokoll auch als Barcode, TAG oder Index bezeichnet, besteht aus sechs bis acht Basen, die auf unterschiedliche Weise in das DNA-Fragment integriert werden. So können diese Erkennungssequenzen bei der Sequenzanalyse mit ausgelesen werden.

Werden spezifische DNA-Fragmente vor der Sequenzierung vervielfältigt, besteht die Möglichkeit codierte Primer zu verwenden, deren Sequenzerweiterung während der Amplifikation in das zu sequenzierende DNA-Molekül integriert wird (BINLADEN *ET AL.* 2007). Der Nachteil dieser Methode besteht darin, dass für jeden Barcode ein eigenes Primerset benötigt wird. Bei der Illumina-Sequenzierung werden die Indexe ebenfalls durch eine PCR in die DNA-Fragmente integriert. Diese Index-PCR (s. Kapitel B3.6 Index-PCR) findet jedoch erst statt nachdem die Sequenzieradapter bereits an die DNA-Fragmente ligiert wurden. Da diese eine universelle Primerbindungsstelle enthalten, wird nur ein Set an Indexen, bzw. Index-Primern benötigt.

Um die Barcodes auch bei einer Voramplifikation ausgewählter DNA-Abschnitte universeller einsetzen zu können gibt es eine weitere Möglichkeit der Markierung. Diese besteht darin, die Barcodes nach der Amplifikation, an die DNA-Fragmente zu ligieren (MEYER *ET AL.* 2008). Die Barcodes, oder auch TAGs nach Meyer bestehen aus einer sechs bis acht Basen langen Erkennungssequenz an beiden Enden, getrennt durch eine Restriktionsschnittstelle in der Mitte. Der Aufbau ist palindromisch, sodass zwei einzelsträngige TAG Moleküle zu einem doppelsträngigen Adapter hybridisiert werden können (s. Abb. B 1).

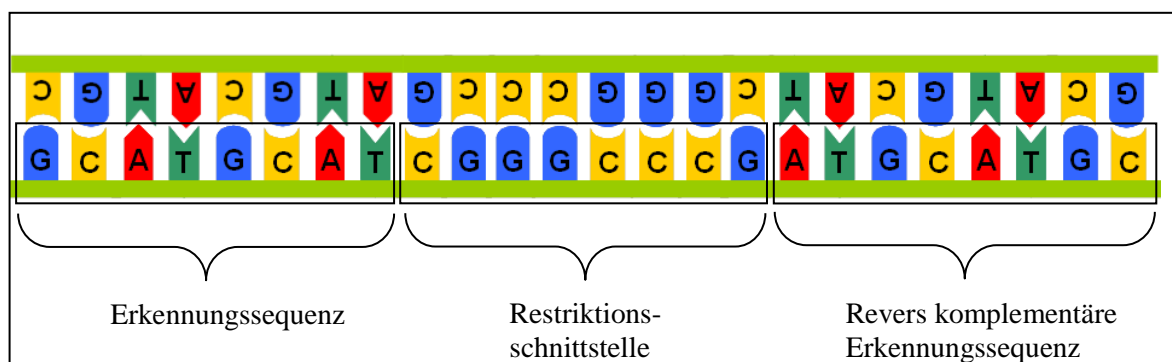


Abb. B 1 Aufbau eines Meyer-TAGs

Das Protokoll nach Meyer wurde mit leichten Veränderungen angewendet, die in den jeweiligen Sektionen erwähnt werden. Generell wurden für die Aufreinigungsschritte nach jedem enzymatischen Schritt keine *SPRI beads* sondern das Qiagen *MinElute PCR Purification Kit* verwendet. Das Kit ermöglicht die Aufreinigung kleinerer Fragmente mit einer geringeren Verlustrate. Außerdem wurden einzelne Tubes verwendet anstatt die Reaktionen in einer *96-well* Platte durchzuführen.

### Hybridisierung der TAGs

Für das *tagging* werden doppelsträngige Adapter benötigt, weshalb die palindromischen Oligonukleotide in einem ersten Schritt zu Doppelsträngen hybridisiert werden müssen. Hierfür werden je 10µL der TAG-Stammlösungen aliquotiert, in einem Cycler auf 95°C erhitzt und mit einer Rate von 0,1°C/s auf Raumtemperatur gekühlt. Anschließend können sie bis zu ihrer Verwendung bei -20°C gelagert werden.

### Quantifizierung der PCR-Produkte

Eine Quantifizierung der DNA wurde nur in einigen Fällen zur Kontrolle des relativen DNA Gehaltes nach den einzelnen enzymatischen Schritten durchgeführt. Auf eine generelle Kontrolle des DNA Gehaltes und dessen Anpassung für die Versuche wurde verzichtet, da die zur Verfügung stehenden Methoden der DNA-Quantifizierung nur den gesamt DNA-Gehalt wiedergeben. Somit werden auch mikrobielle Verunreinigungen und durch die vorgeschobene Multiplex-PCR verursachte Primerdimere gemessen. Eine Anpassung der Einsatzmenge aufgrund der gemessenen DNA-Konzentration würde mitunter zu einer Unterrepräsentation der gewünschten Amplifikate führen. Aus diesen Gründen wurde generell das gesamte PCR-Produkt in die Reaktionen eingesetzt.

## Vereinigen der PCR-Produkte

Alle PCR-Produkte, die durch den gleichen TAG markiert werden sollen, werden vor dem ersten Reaktionsschritt vereinigt. Dies betrifft in jedem Fall die drei mitochondrialen Multiplex-Ansätze jeder Probe und optional einen nukleären Multiplex-Ansatz, da dieser durch die entsprechenden Primer leicht von den mitochondrialen Fragmenten zu trennen ist. Die entsprechenden PCR-Ansätze werden zusammen aufgereinigt und mit 20µL Elutionspuffer (EB) eluiert und somit aufkonzentriert. In den nachfolgenden Reaktionsansätzen bezieht sich die Angabe „PCR-Produkt“ jeweils auf das Produkt des vorhergehenden Schrittes.

### *Blunt-end repair*

Die für die PCRs verwendete Taq-Polymerase hat die Eigenschaft an das Ende eines jeden Fragmentes ein dATP anzuhängen. Die nach der PCR vorliegenden DNA-Fragmente weisen somit keine glatten Enden (*blunt-ends*), sondern einen A-Überhang an jedem 3'-Ende auf. Zur Ligation der TAGs werden jedoch glatte Enden benötigt, weshalb ein *blunt-end repair* durchgeführt wird. Zwei Enzyme werden dazu eingesetzt, die T4 DNA Polymerase, die über eine 3'→5' Exonuklease Aktivität zum Schneiden überhängender 3'-Enden verfügt und eine zusätzliche 5'→3' Polymerase Aktivität zur Rekomentierung überhängender 5'-Enden aufweist und die T4 Polynukleotidkinase zur Phosphorylierung der 5'-Enden. Für den Reaktionsansatz wird anstatt dem im Meyer Protokoll verwendeten Tango Puffer der 5X T4 DNA Polymerase Puffer verwendet. Aus den benötigten Reagenzien wird hier, wie auch in den folgenden Schritten, für die eingesetzte Probenanzahl ein Mastermix erstellt. Der Mastermix wird durch vorsichtiges auf- und ab-pipettieren gemischt und anschließend je 20µL davon mit je 20µL des PCR-Produktes gemischt.



**Tab. B 18 Reaktionsansatz für das *blunt-end repair***

<i>Menge/Probe</i>	<i>Reagenz</i>
20 µl	PCR-Produkt
2.6 µl	ddH <sub>2</sub> O
8 µl	5X T4 DNA pol Puffer
1 µl	BSA (10mg ml <sup>-1</sup> )
1.6 µl	dNTPs (2.5mM)
4 µl	ATP (10mM)
2 µl	T4 PNK (10U µl <sup>-1</sup> )
0.8 µl	T4 DNA Polymerase (5U µl <sup>-1</sup> )

Die Reaktionsansätze werden für 15 Min bei 12°C und anschließend 15 Min bei 25°C in einem Thermocycler inkubiert, anschließend aufgereinigt und in 29µl EB eluiert.

### Ligation der TAGs

Die PCR-Produkte und TAGs müssen gut gemischt werden bevor alle weiteren Chemikalien zugegeben werden. Aus den restlichen Chemikalien wird ein Mastermix erstellt, der vor der Zugabe der Ligase 10s gevortext werden muss, da sich das PEG4000 aufgrund seiner hohen Viskosität schlecht homogenisieren lässt.

**Tab. B 19 Reaktionsansatz für die Ligation**

<i>Menge/Probe</i>	<i>Reagenz</i>
29 µl	phosphoryliertes PCR-Produkt
4 µl	10X T4 Ligase Puffer
4 µl	PEG 4000
1.6 µl	TAG (500µM)
0.4 µl	ddH <sub>2</sub> O
1 µl	T4 Ligase (5u/µL)

Die Ansätze werden über Nacht bei 22°C inkubiert, anschließend aufgereinigt und in 23µl EB eluieren.

### *Adapter fill-in*

Die T4-Ligase benötigt ein 5'-Phosphat für die Ligation zwischen zwei Molekülen. Fehlt die Phosphatgruppe an einem der Moleküle wird nur einer der beiden Doppelstränge verbunden und es bleibt eine Lücke zwischen TAG und PCR-Produkt, die vor den weiteren Schritten geschlossen werden muss. Hierfür wird die Bst-Polymerase eingesetzt, die in der Lücke ansetzt, den restlichen Strang nach der Lücke verdrängt und einen neuen durchgehenden Strang synthetisiert.

**Tab. B 20 Reaktionsansatz für das Adapter *fill-in***

<i>Menge/Probe</i>	<i>Reagenz</i>
23 µl	PCR-Produkt
3 µl	10x ThermoPol™ Puffer
3 µl	dNTPs (2,5mM)
0 µl	ddH <sub>2</sub> O
1 µl	Bst Polymerase (8 u/µl)

Der Ansatz wird 30 Min bei 37°C inkubiert.

Die PCR-Produkte tragen nun einen individuellen Barcode, den TAG und können während der Aufreinigung vereinigt werden. Davor wurde abweichend vom ursprünglichen Protokoll keine DNA-Quantifikation durchgeführt.

In den weiteren Schritten werden die DNA-Fragmente dephosphoryliert und anschließend an der Restriktionsschnittstelle des TAGs geschnitten. Auf diese Weise ist erstens keine *blunt-end repair* Reaktion zur Ligation der Sequenzieradapter für die 454 nötig, da das Restriktionsenzym beim Schneiden glatte Enden erzeugt und zweitens werden nur die Fragmente mit Sequenzieradaptern versehen, die an beiden Enden einen TAG tragen, denn nur der geschnittene Adapter besitzt eine Phosphatgruppe am 5'-Ende, die für die Ligation der Sequenzieradapter benötigt wird. Die Erstellung der *sequencing libraries* wurde jedoch von den Sequenzierfirmen vorgenommen, die routinemäßig ein *blunt-end repair* durchführen, wobei alle Fragmente ohne Phosphatgruppe erneut phosphoryliert werden. Die Dephosphorylierung und der anschließende Restriktionsverdau sind daher überflüssig. Die *getaggen* Proben wurden vereinigt, aufgereinigt und in der von der Sequenzierfirma geforderten Menge Elutionspuffer (100µL) eluiert.

Für die ersten 454-Sequenzierungen wurden die Dephosphorylierung und der Restriktionsverdau jedoch wie folgt durchgeführt:

Die *getaggt*en Proben werden vereinigt, aufgereinigt und in 20µL EB eluiert. Aufgrund der Filterkapazität wurden in diesem Schritt noch nicht alle Proben vereinigt, sondern jeweils bis zu sechs Stück.

## Dephosphorylierung

Bei der Dephosphorylierung wurden alle Phosphatgruppen am 5'-Ende abgeschnitten.

**Tab. B 21 Reaktionsansatz für die Dephosphorylierung**

Menge/Probe	Reagenz
20 µl	getaggtetes Produkt
3 µl	10x NEBuffer
5.5µl	ddH <sub>2</sub> O
1.5µl	Calf Intestine Phosphatase – CIP (10u/µl)

Die Reaktionsansätze werden für 30 Min bei 37°C inkubiert, anschließend aufgereinigt und in 25µL EB eluiert.

## Restriktionsverdau

Die TAGs werden an der eingebauten Erkennungssequenz von einem Restriktionsenzym geschnitten, wobei *blunt-ends* erzeugt werden.

**Tab. B 22 Ansatz für den Restriktionsverdau**

Menge/Probe	Reagenz
25 µl	getaggtetes Produkt
3 µl	Universalpuffer
1 µl	ddH <sub>2</sub> O
1 µl	SrfI (10u/µL)

Der Ansatz muss eine Stunde bei 37°C inkubieren, anschließend wird er aufgereinigt und in EB eluiert. Die Menge an EB beträgt 15µL wenn direkt die Ligation der Sequenzieradapter folgt, oder wie im vorliegenden Fall in 100µL für die Sequenzierfirma.

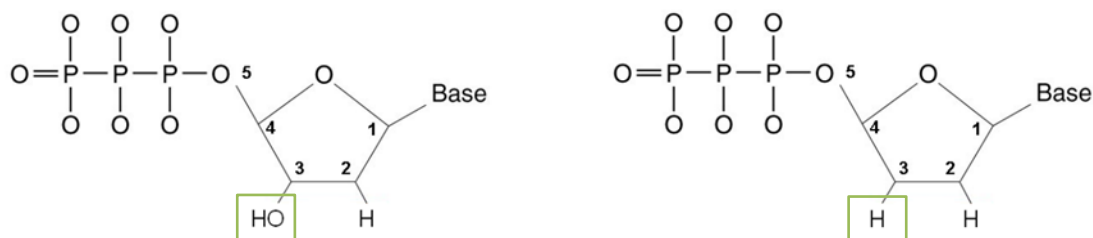
## B3.6 Sanger Sequenzierung und *Next Generation Sequencing*

Zur Analyse der DNA Sequenzen kamen verschiedene Sequenziertechniken und -plattformen zum Einsatz. Für die Analyse einzelner DNA Sequenzen wurde die Sanger Methode angewandt. Das *Next Generation Sequencing* (NGS) ermöglicht einen wesentlich höheren Probendurchsatz als die herkömmliche Sanger Methode und wurde deshalb zur Analyse der Multiplex-Ansätze, oder für ganze mitochondriale Genome verwendet. Für das NGS müssen die DNA-Fragmente je nach verwendeter Sequenzierplattform unterschiedlich auf die Sequenzanalyse vorbereitet werden, wobei plattformspezifische Sequenzieradapter an die DNA-Enden ligiert werden. Die DNA-Fragmente mit den entsprechenden Sequenzieradaptern bezeichnet man als *sequencing library*. Die Adapter enthalten zum einen die Erkennungssequenz für den Sequenzierprimer und zum anderen eine Bindungsstelle, die das DNA-Fragment fixiert – je nach Plattform an einem *bead* (454), oder einer Sequenzierplatte (Illumina).

Nachfolgend werden die verwendeten Plattformen mit den verschiedenen Sequenzier-techniken vorgestellt und die hierfür notwendige DNA-Präparation erläutert.

### Sanger Sequenzierung

Diese Methode zur DNA-Sequenzanalyse basiert auf der Strangabbruchmethode nach Sanger (SANGER *ET AL.* 1977). Ähnlich der PCR wird auch hier ein Reaktionsgemisch aus Puffer, dNTPs und Polymerase angesetzt. Der Unterschied zur PCR liegt darin, dass nur ein Primer und kein Primerpaar verwendet wird und dass zusätzlich zu den dNTPs auch ddNTPs eingesetzt werden, die nach ihrem Einbau zum Abbruch der Strangsynthese führen, da ihnen die 3'OH-Gruppe fehlt (s. Abb. B 2).



**Abb. B 2** dNTP (links) und ddNTP (rechts)

Im Gegensatz zur ursprünglichen Methode, bei der in vier getrennten Reaktionen jeweils ein Nukleotid radioaktiv markiert war, tragen die vier verschiedenen ddNTPs jeweils unterschiedliche Farbstoffmarkierungen, wodurch die Reaktion auf einen Ansatz beschränkt werden kann. Wie bei der PCR verläuft die Reaktion in Zyklen, wobei sich Denaturierung, Hybridisierung und Elongation stets wiederholen. Bei diesem sog. *Cyclesequencing* werden nun dNTPs und ddNTPs in einem bestimmten Verhältnis eingesetzt, was zur Folge hat, dass die Neusynthese in jedem Zyklus zufällig an unterschiedlichen Stellen abbricht und somit an jeder Position des DNA-Strang wenigstens einmal ein farbmarkiertes ddNTP eingebaut werden sollte. Für den Reaktionsansatz wurde das *BigDye® Terminator v1.1* Kit von Applied Biosystems verwendet.

**Tab. B 23 Cyclesequencing-Ansatz**

<i>Für einen 10µL Ansatz:</i>	
<i>Chemikalien</i>	<i>Menge/Probe [µL]</i>
HPLC H <sub>2</sub> O	5,5
BigDye® Terminator 5x Sequencing Buffer	1,5
2,5x Big Dye® Terminator v1.1 Ready Reaction Mix	1
Primer (10µM)	1
Target	1

**Tab. B 24 Laufkonditionen des Cyclesequencing**

<i>Programm</i>	<i>Temperatur</i>	<i>Zeit</i>
Denaturierung	92°C	30sec
Hybridisierung	56°C - 58°C	15sec
Elongation	60°C	2,5min
Zyklen	25	

Vor der eigentlichen Sequenzanalyse werden die Proben des *Cyclesequencing*-Ansatzes durch eine Filtration mit Sephadex G-50 Fine (GE Healthcare) aufgereinigt und mit je 10µL Formamid in einer 96er-Sequenzierplatte vorgelegt.

Die Ermittlung der DNA Sequenz erfolgt anschließend mit dem ABI PRISM 3130 *Genetic Analyzer* von Applied Biosystems, durch eine Kapillarelektrophorese. Das Gerät enthält vier Kapillaren, an die eine Spannung angelegt wird, wodurch die Sequenzierprodukte gemäß ihrer Ladung zur Anode wandern und dabei durch die Polymermatrix der Kapillaren ihrer Länge nach aufgetrennt werden. Die Trennschärfe liegt hierbei bei einem Nukleotid. Bei dem verwendeten Polymer handelt es sich um das *Performance Optimized Polymer 6* (POP-6™).

Die Injektionszeit (*injection time* = IT) richtet sich hierbei nach dem DNA Gehalt der Probe. Wie lange die Spannung angelegt wird (*run time* = RT), richtet sich nach der Länge der zu analysierenden Fragmente. Am Ende der Kapillaren ist ein Sichtfenster eingelassen, hinter dem ein Laser und eine CCD (*charge-coupled devise*) Kamera installiert sind. Der Laser regt die Farbmarkierungen der ddNTPs am Ende jeden Sequenzabschnitts an. Die anschließend emittierte Strahlung wird von der Kamera erfasst und mit Hilfe der ABI PRISM *Sequencing Analysis*<sup>TM</sup> 3.0 Software in Signale eines Elektropherogramms umgewandelt. Jeder Base wird hierbei ein farbiger Signalpeak zugeordnet: G = schwarz; A = grün; T = rot; C = blau.

## Roche 454 GS FLX

### *454 Library Preparation*

Die Vorbereitung der DNA-Fragmente für die Sequenzierung auf der 454 Plattform wurde von den jeweiligen Sequenzierfirmen, GATC Biotech AG und LGC Genomics GmbH durchgeführt und sei deshalb hier nur kurz umrissen. Eine ausführliche Beschreibung der 454 *library preparation* findet sich in den Anhängen der ersten Veröffentlichung zur 454 Sequenzierplattform (MARGULIES *ET AL.* 2005).

Proben moderner DNA werden im ersten Schritt der *library preparation* fragmentiert. Bei alter DNA ist dieser Schritt überflüssig, da die DNA bereits in mehr oder weniger fragmentierter Form vorliegt. Durch die Fragmentierung der DNA, artifiziell oder natürlich, sind die Enden der beiden DNA-Stränge in der Regel nicht gleich lang. Für die nachfolgenden Schritte sind jedoch gleich lange DNA-Stränge Voraussetzung, weshalb im nächsten Schritt ein *enzymatic end polishing* oder *blunt-end repair* durchgeführt wird. Hierfür werden drei Enzyme eingesetzt, die T4 DNA Polymerase, die über eine 3'→5' Exonuklease Aktivität zum Schneiden überhängender 3'-Enden und eine 5'→3' Polymerase Aktivität zur Rekomentierung überhängender 5'-Enden verfügt, das Klenow Fragment der *E.coli* DNA Polymerase, das die gleichen Eigenschaften besitzt und die T4 Polynukleotidkinase zur Phosphorylierung der 5'-Enden.

Für die 454 Sequenzierung werden zwei unterschiedliche Sequenzieradapter, A und B, verwendet. Beide weisen je Bindungsstellen für einen PCR Primer und für den Sequenzierprimer auf. Außerdem enthält jeder Adapter eine vier Basenpaar lange Erkennungssequenz. Die A- und B-Adapter weisen zum einen eine unterschiedliche Nukleotidsequenz auf und zum anderen trägt der B-Adapter eine Biotin Markierung am 5'-Ende. Die Adapter wurden entwickelt um eine gerichtete Ligation zu ermöglichen. Hierfür weisen beide Adapter nur ein glattes Ende auf, sodass nur dieses an das DNA-Fragment ligiert werden kann. Nach dem Protokoll von MARGULIES *ET AL.* 2005 wird hierfür die Quick Ligase verwendet. Im Idealfall trägt nun jedes DNA-Fragment auf der einen Seite einen A- und auf der anderen Seite einen B-Adapter.

Nach der Ligation wird die DNA auf ein Agarosegel aufgetragen und der Länge nach aufgetrennt. Der Bereich von 250 – 500bp wird mit einem Skalpell herausgeschnitten und die DNA aus dem Gel extrahiert. Bei den Sequenzierungen alter DNA wurde dieser

Größenbereich auf 90 – 150bp verkleinert, da aDNA generell in kürzeren Abschnitten vorliegt und in diesem speziellen Fall Amplikons sequenziert wurden, deren Länge 90 – 150bp beträgt.

Im Anschluss an die Größenselektion wird ein sog. *nick repair* durchgeführt, um die Lücke, die bei der Ligation zwischen dem 3'-Ende des DNA-Fragments und dem Adapter entsteht, zu schließen. Dies geschieht mit Hilfe der Bst-Polymerase, wobei sie den unterbrochenen DNA-Strang verdrängt und einen durchgängigen neu synthetisiert.

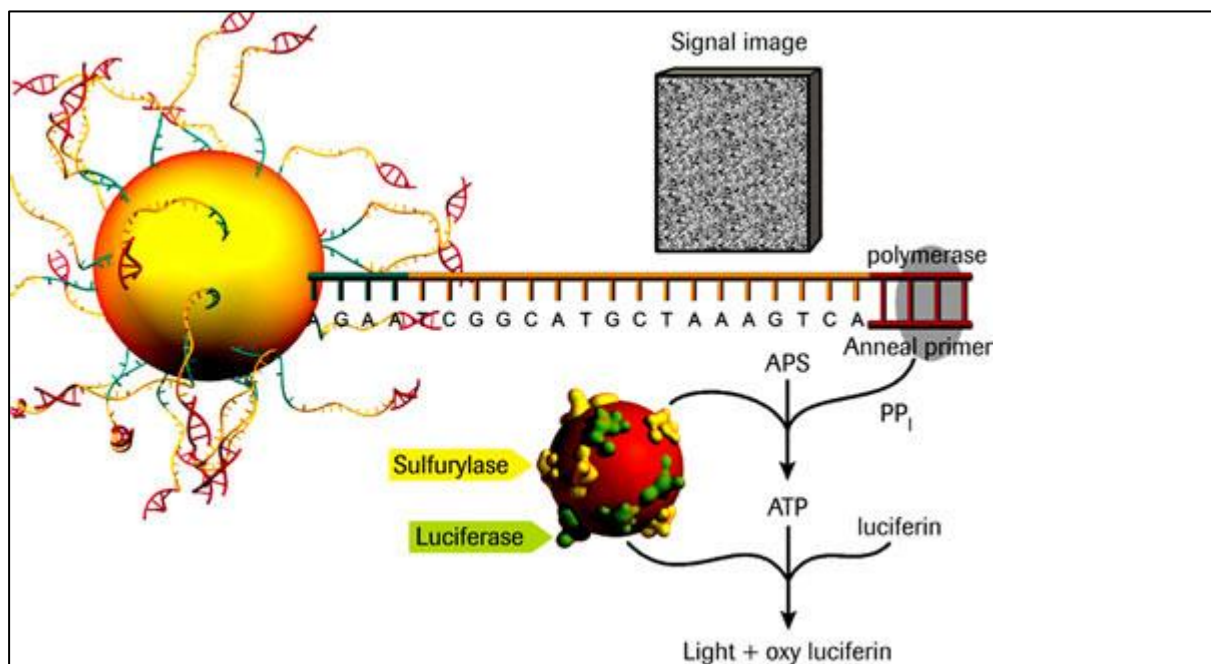
Im letzten Schritt der *library preparation* werden die Fragmente mit den korrekt ligierten A- und B-Adaptoren isoliert und einzelsträngig gemacht. Hierfür wird die fast fertige *library* mit magnetischen *beads*, die einen Streptavidinüberzug tragen gemischt. Die Biotin-Moleküle der B-Adapter binden an das Streptavidin. Durch einen Magneten, der die *beads* mit dem Streptavidinüberzug anzieht werden nun die Fragmente mit B-Adapter von denen ohne getrennt. Anschließend werden die Fragmente denaturiert, wodurch eine einzelsträngige *library* mit Fragmenten entsteht, die auf der einen Seite einen A-Adapter und auf der anderen einen B-Adapter tragen, denn Fragmente, die auf beiden Seiten einen B-Adapter tragen, bleiben am *bead* hängen.

#### *454-Sequenzierung*

Die Fragmente mit den korrekt ligierten Adaptoren werden über die B-Adapter erneut an *beads* gebunden, wobei das Verhältnis von Fragmenten zu *beads* so gewählt ist, dass ein Fragment pro *bead* bindet. Die DNA-Fragmente werden in einer Emulsions-PCR vervielfältigt, wobei die Emulsion so eingestellt wird, dass jeder Wassertropfen ein Reaktionsgefäß darstellt, das ein *bead* mit DNA-Fragment und die erforderlichen PCR Chemikalien enthält. Am Ende trägt jedes *bead* mehrere Millionen Kopien ein und desselben DNA-Strangs. Anschließend wird die Emulsion aufgebrochen, die DNA-Doppelstränge denaturiert und die *beads*, die nun jeweils einzelsträngige Klone des ursprünglichen DNA-Fragmentes tragen, jeweils in eine Vertiefung einer Picotiterplatte eingebracht. In die einzelnen Vertiefungen werden nun die Chemikalien für das *Pyrosequencing* zu den *beads* gegeben. Die Pyrosequenzierung erfolgt nach dem *sequencing by synthesis* Prinzip, d.h. jedes eingebaute Nukleotid wird direkt erfasst (RONAGHI *ET AL.* 1996). Wird ein Nukleotid eingebaut, setzt dies eine Enzymkaskade in Gang, an deren Ende Licht emittiert wird (s. Abb. B3), was durch eine CCD Kamera erfasst wird.



Die Kamera ist in der Lage, die Lichtinformation der richtigen Position auf der Picotiterplatte und somit der richtigen Ursprungssequenz zuzuordnen. Dadurch, dass die vier verschiedenen dNTPs nacheinander und nicht gleichzeitig zugegeben werden, kann somit festgestellt werden wann und wo welches Nukleotid eingebaut wurde und anhand der Stärke des Lichtsignals auch wie viele. Nach jedem Einbauzyklus werden nicht eingebaute dNTPs durch das Enzym Apyrase deaktiviert.



**Abb. B 3 Schematische Darstellung einer Pyrosequenzierungsreaktion.**

Sobald ein dNTP eingebaut wird, wird Diphosphat frei (PP), welches zusammen mit APS (Adenosin 5' Phosphosulfat), katalysiert durch Sulfurylase, zu ATP (Adenosintriphosphat) umgewandelt wird. Die Luciferase wandelt ATP und Luciferin zu Oxy-Luciferin um, wobei Energie in Form von Licht freigegeben wird. (Quelle: 454.com)

## Illumina-Sequenzierung

### *Illumina Library Preparation*

Zur Vorbereitung auf die Sequenzierung mit dem Illumina Sequenzierer müssen die DNA-Fragmente mit den entsprechenden Sequenzieradaptern versehen werden. Die Präparation der *sequencing library* erfolgte im Reinraumlabor nach dem Protokoll von MEYER & KIRCHER 2010.

**Tab. B 25 Adapter und Primer für die Präparation der Illumina Library**

<i>Adaptersequenzen 5' → 3'</i>	
IS1_adapter.P5	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCG*A*T*C*T
IS2_adapter.P7	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCG*A*T*C*T
IS3_adapter.P5+P7	A*G*A*T*CGGAA*G*A*G*C
<i>Primersequenzen (5' → 3')</i>	
IS4_indPCR.P5	AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTT
IS5_reamp.P5	AATGATACGGCGACCACCGA
IS6_reamp.P7	CAAGCAGAAGACGGCATAACGA
IS7_short_amp.P5	ACACTCTTCCCTACACGAC
IS8_short_amp.P7	GTGACTGGAGTTCAGACGTGT
Index (hier: Index1)	CAAGCAGAAGACGGCATAACGAGATcctgcgaGTGACTGGAGTTCAGACGTGT
<i>Sequenzierprimer (5' → 3')</i>	
Read 1	ACACTCTTCCCTACACGACGCTCTTCCGATCT
Index Read	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
Read 2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

### *Hybridisierung der Adapter und der Positivkontrolle*

Zu Beginn der Library Erstellung müssen die einzelsträngigen Teiladapter zu vollständigen Adaptern hybridisiert werden. Das gleiche gilt für die Positivkontrolle. Diese besteht aus den Roche 454 Sequenzierprimern und einer artifiziellen Sequenz (SANCHEZ *ET AL.* 2003), die nicht im menschlichen Genom zu finden ist.

**Tab. B 26 Positivkontrolle zur Illumina-Sequenzierung**

Zusammensetzung der Positivkontrolle aus den Roche Sequenzierprimern A und B und einer artifiziiellen Sequenz (Nonsense\_1). Nonsense\_2 besteht aus der revers komplementären Sequenz von Nonsense\_1 und wird mit dieser zur Positivkontrolle hybridisiert.

<i>Positivkontrolle (Nonsense) 5' → 3'</i>	
Roche Primer	CCATCTCATCCCTGCGTGTC
Artifizielle	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA
Roche Primer	CCTATCCCCTGTGTGCCTTG
Nonsense_1	CCATCTCATCCCTGCGTGTCAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTG ACAACCTATCCCCTGTGTGCCTTG
Nonsense_2	CAAGGCACACAGGGGATAGTTGTCAGACTTTTCACGACGTGGCACCTAGTTTAGTC AGTTGACACGCAGGGATGAGATGG

**Tab. B 27 Hybridisierungsansätze der Adapter und der Positivkontrolle**

<i>Adapter-MIX IS1/IS3</i>	<i>Volumen [µl]</i>
UV-HPLC	10
TE-Puffer	10
IS1_adapter_P5.F (500 µM)	40
IS3_adapter_P5+P7.R (500 µM)	40
<i>Adapter-Mix IS2/IS3</i>	<i>Volumen [µl]</i>
UV-HPLC	10
TE-Puffer	10
IS2_adapter_P7.F (500 µM)	40
IS3_adapter_P5+P7.R (500 µM)	40
<i>Positivkontrolle</i>	<i>Volumen [µL]</i>
TE-Puffer	46
Nonsense 1	2,05
Nonsense 2	2,01

**Tab. B 28 Programm zur Hybridisierung der Adapter und der Positivkontrolle**

<i>Programm</i>	<i>Temperatur</i>	<i>Zeit</i>
Initiale Denaturierung	95 °C	10 sec
graduelle Absenkung	0,1 °C/sec	14 min*
Hold	12°C	-
* Temperaturdifferenz = 83°C => Dauer = 830s = 14 min		

Nach der Hybridisierung werden die beiden Adapter IS1/IS3 und IS2/IS3 vereinigt.

*Konzentrationsbestimmung der Extrakte*

Die DNA-Konzentration wurde mit dem Qubit<sup>®</sup> Fluorometer von Invitrogen<sup>™</sup> bestimmt. Verwendet wurde hierfür das Qubit<sup>™</sup> dsDNA HS (High Sensitivity) Assay Kit mit einem Messbereich von 10pg/µL – 100ng/µL. Die Messung erfolgte nach Herstellerprotokoll.

### *Präparation der sequencing library*

In den nachfolgend beschriebenen Reaktionsansätzen wurden je eine Leer- und eine Positivkontrolle mitgeführt. Nach jedem Schritt wird eine Aufreinigung durchgeführt. Dies geschieht in allen Fällen mit dem Qiagen *MinElute PCR Purification Kit*.

### *Blunt-end repair*

Damit die Sequenzieradapter richtig an die DNA-Fragmente ligieren können, müssen diese glatte Enden aufweisen. Hierzu wird, wie auch beim *Tagging* (s. Kap. B3.5), ein *blunt-end repair* durchgeführt.

**Tab. B 29** Reaktionsansatz des *Blunt End Repairs*

<i>Reagenz</i>	<i>Volumen/Probe [μL]</i>
H <sub>2</sub> O	0,4
Tango Puffer (10x)	7
dNTPs (jedes 10mM)	0,7
ATP (10mM)	7
T4 PNK (10 U/μl)	3,5
T4 DNA Pol (5U/μl)	1,4
Probe	50
gesamt	70

Für das *blunt-end repair* werden jeweils 20μl des MM zu 50 μl DNA Extrakt gegeben. Die Ansätze werden in einem Thermoblock zuerst 15 Min. bei 25°C und anschließend 5 Min. bei 12°C inkubiert. Danach werden die Ansätze aufgereinigt und in 22μL Elutionspuffer (EB) eluiert.

### *Adapter Ligation*

Mit Hilfe der DNA-Ligase werden die zuvor hybridisierten Adapter an die glatten Enden der DNA-Fragmente ligiert.

**Tab. B 30 Reaktionsansatz zur Adapter Ligation**

<i>Reagenz</i>	<i>Volumen [<math>\mu</math>L]</i>
H <sub>2</sub> O	10
T4 DNA Ligase Puffer (10x)	4
PEG-4000 (50%)	4
Adapttermix (je 100 $\mu$ M)	1
T4 DNA Ligase (5U/ $\mu$ l)	1
Probe	20
gesamt	40

Weist der Ligase Puffer nach dem Auftauen ein Präzipitat auf, muss er vor dem Ansetzen auf 37°C erwärmen und gründlich gemischt werden. Des Weiteren sollte der MM vor Zugabe der T4 Ligase gevortext werden, da PEG sehr viskos ist und sich schlecht homogenisieren lässt. Von dem fertigen MM werden dann je 20 $\mu$ l zu 20 $\mu$ l des Eluats aus dem vorherigen Schritt gegeben. Der Ansatz wird für 30 Min. bei 22°C inkubiert, anschließend aufgereinigt und mit 22 $\mu$ l EB eluiert.

#### *Adapter Fill-In*

Wie bereits bei der 454-Library Präparation beschrieben entsteht bei der Adapter Ligation eine Lücke zwischen dem 3'-Ende des DNA-Fragments und dem Adapter, die durch nachfolgenden Schritt geschlossen wird. Dies geschieht mit Hilfe der Bst Polymerase, wobei sie den unterbrochenen DNA-Strang verdrängt und einen durchgängigen neu synthetisiert.

**Tab. B 31 Reaktionsansatz zum Adapter *Fill-In***

<i>Reagenz</i>	<i>Volumen [<math>\mu</math>L]</i>
H <sub>2</sub> O	13,5
ThermoPol Puffer (10x)	4
dNTPs (je 10mM)	1
<i>Bst</i> Polymerase, gr. Fragment (8U/ $\mu$ l)	1,5
Probe	20
gesamt	40

20 $\mu$ l des MM werden zu den 20 $\mu$ l Probe gegeben, gut vermischt und für 20 Min. bei 37°C inkubiert. Anschließend werden die Ansätze wieder aufgereinigt und diesmal in 37 $\mu$ L EB eluiert.

*Index-PCR*

Um die parallele Sequenzierung unterschiedlicher Proben zu ermöglichen, müssen diese mit einer individuellen Erkennungssequenz markiert werden. Diese Indexe werden für die Illumina-Sequenzierung nicht wie beim *Tagging* an die DNA-Fragmente ligiert, sondern durch eine PCR in die Adaptersequenz eingebaut. Jeder Index besteht aus sechs Nukleotiden und unterscheidet sich zu allen anderen Indexen durch mindestens zwei Nukleotide (s. Tab. B32). Die PCR wird in 3 Ansätzen pro Probe durchgeführt.

**Tab. B 32** Liste der verwendeten Indexe

<i>Index</i>	<i>Sequenz 5' → 3'</i>
1	TGACTG
2	ACACTG
3	AGTATG
4	TATATG
5	ATGATG
6	TCGATG
7	ACATCG
8	TACACG
9	AGACAG
10	TAGATC
11	ACTATC
12	AGACGC
13	TGATGC
14	ATGCAC
15	ATATAC
16	TATCAG
17	ACGACG
18	TGCATC
19	TACGAC
20	TATGCG

**Tab. B 33** Reaktionsansatz der Index-PCR

<i>Reagenz</i>	<i>Volumen [<math>\mu</math>L]</i>
H <sub>2</sub> O	25,5
10x Gold Buffer	5
MgCl <sub>2</sub> (25mM)	5
dNTPs (je 10mM)	1
BSA (20mg/mL)	1
Primer IS4 (10 $\mu$ M]	1
AmpliTaq (5U)	0,5
Index-Primer (10 $\mu$ M)	1
Probe	10

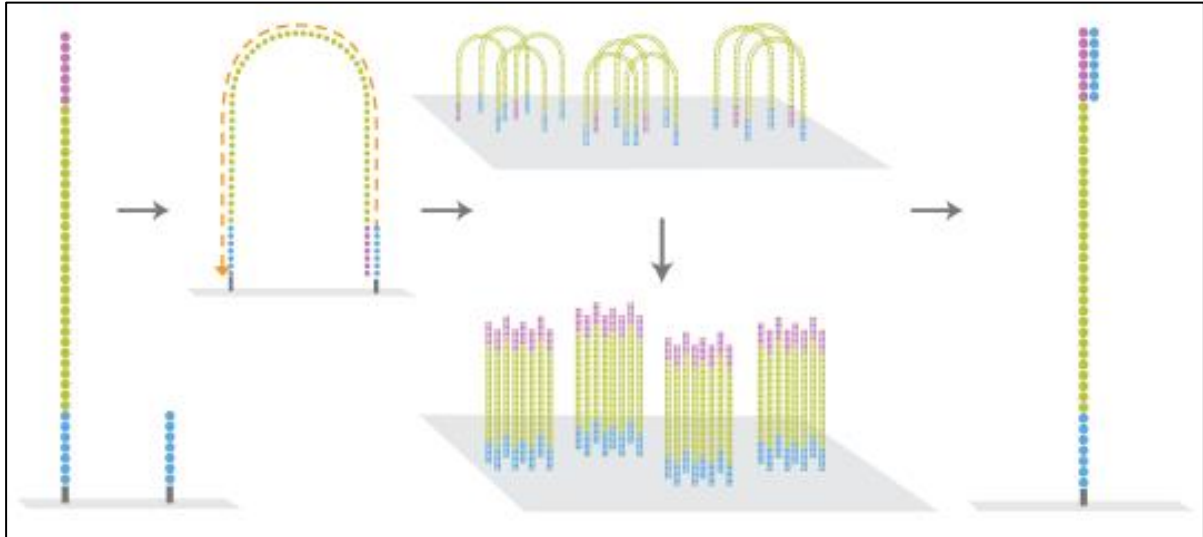
**Tab. B 34 Laufkonditionen der Index-PCR**

<i>Programm</i>	<i>Temp. [°C]</i>	<i>Zeit [s]</i>	<i>Zyklen</i>
Initiale Denaturierung	94	2 [min]	1
Denaturierung	94	40	12–24*
Annealing	60	40	
Elongation	72	40	
Finale Elongation	72	10 [min]	1
	8	hold	
*Die Zyklenzahl ist abhängig von der Ausgangskonzentration:			
> 100ng:	12 Zyklen		
> 10ng:	16 Zyklen		
> 1ng:	20 Zyklen		
> 100pg:	24 Zyklen		

Die drei Ansätze je Probe werden anschließend während der Aufreinigung wieder vereinigt und in 33 µl EB eluiert.

### *Illumina-Sequenzierung*

Die mit Adaptern versehenen DNA-Fragmente werden auf einen Glaträger, die sog. *flow-cell* gegeben, die mit kovalent an die Oberfläche gebundenen Primern versehen ist. Auf ihr erfolgen die Amplifikation der DNA-Fragmente und deren anschließende Sequenzierung. Die Amplifikation erfolgt durch die Hybridisierung der Adapter an die *flow-cell* Primer. Dadurch, dass die Primer fest an die Oberfläche gebunden sind, sind auch alle Amplifikate an der Oberfläche der *flow-cell* gebunden (ADESSI ET AL. 2000). Für jeden weiteren Vervielfältigungsschritt bilden die DNA-Fragmente eine Brücke zu benachbarten Primern auf der *flow-cell* aus, weswegen diese Art der PCR *Bridge-PCR* genannt wird. Durch diese Art der Amplifikation entstehen Anhäufungen (*Cluster*) der PCR-Produkte eines Ausgangsfragmentes. Da die Amplifikation in beide Richtungen erfolgt, bei der Sequenzierung jedoch nur eine Stranghälfte sequenziert werden kann, da es ansonsten zu Signalüberlagerungen kommt, wird am Ende der *Bridge-PCR* einer der beiden Primer auf der *flow-cell* Oberfläche durch ein Restriktionsenzym geschnitten. Die Brücken werden somit aufgelöst und die DNA-Doppelstränge richten sich auf. Anschließend werden die Doppelstränge denaturiert, wodurch ein *Cluster* nun aus gleichgerichteten, einzelsträngigen Kopien des ursprünglichen DNA-Fragmentes besteht (s. Abb. B4).

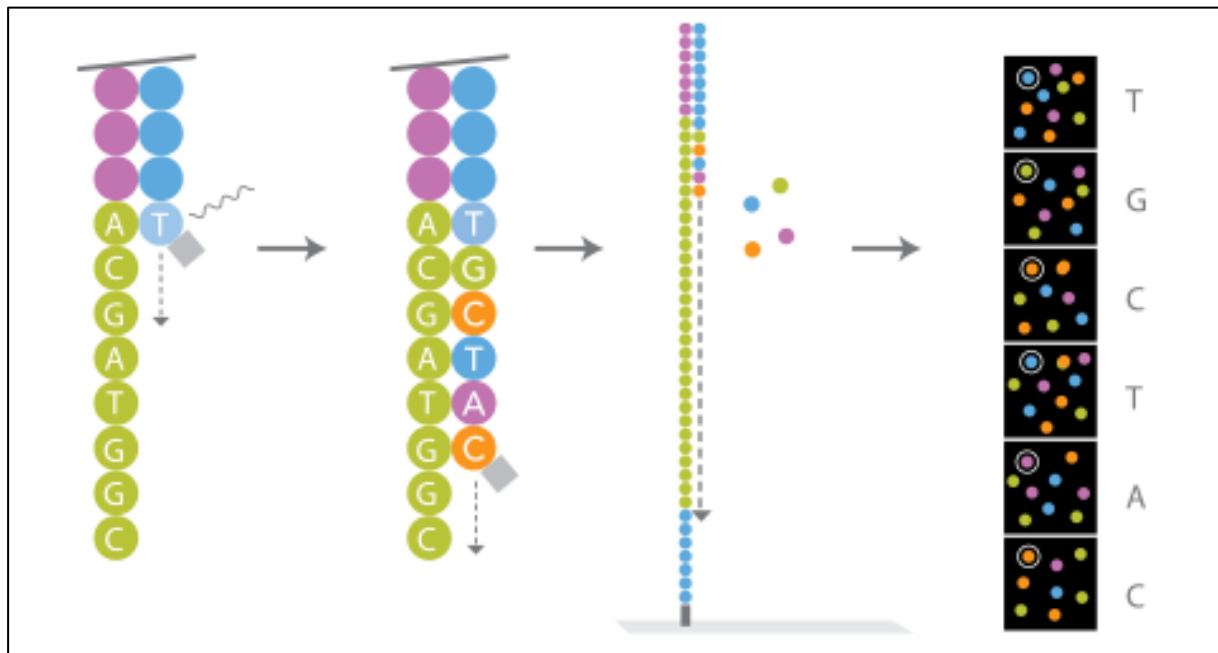


**Abb. B 4 Illumina Bridge-PCR**

Die Abbildung zeigt an die flow-cell gebundene DNA-Fragmente und Primer, die Amplifikation und die resultierenden Cluster. Der letzte Abschnitt zeigt die Anlagerung des Sequenzierprimers an ein DNA-Fragment (Quelle: [https://www.dkfz.de/gpcf/hiseq\\_technology.html](https://www.dkfz.de/gpcf/hiseq_technology.html)).

Im Anschluss an die Amplifikation lagert sich der Sequenzierprimer am freien Adapter-Ende des DNA-Fragmentes an. Zur *flow-cell* werden nun die Chemikalien für die Sequenzierung gegeben, darunter farbmarkierte, reversibel geblockte dNTPs. Nach erfolgreichem Einbau eines dNTPs wird die Farbmarkierung durch einen Laser angeregt und das emittierte Lichtsignal von einer CCD Kamera erfasst und den entsprechenden Koordinaten des *Clusters* zugeordnet (s. Abb. B5). Die überschüssigen dNTPs werden in einem Waschschrift entfernt und das eingebaute Nukleotid für den nächsten Syntheseschritt reaktiviert.





**Abb. B 5 Illumina sequencing by synthesis**

Die Abbildung zeigt den Ablauf der Sequenzierung: Einbau eines reversibel geblockten dNTPs, Lichtemission, Entfernung des Blockers, Einbau weiterer dNTPs komplementär zum analysierten DNA-Fragment. Der letzte Abschnitt zeigt das von der CCD-Kamera erfasste Bild, wobei jeder Punkt einem Sequenz-Cluster entspricht (Quelle: [https://www.dkfz.de/gpcf/hiseq\\_technology.html](https://www.dkfz.de/gpcf/hiseq_technology.html)).

Werden mehrere Proben gleichzeitig analysiert, tragen die Adapter einen probenspezifischen Index, der nach dem ersten Sequenzierschritt in einem separaten *Index-read* abgelesen werden kann. Soll das DNA-Fragment von beiden Seiten sequenziert, also eine *paired-end* Sequenzierung durchgeführt werden, findet die zweite Sequenzierrunde nach dem *Index-read* statt. Durch ein *paired-end-modul* wird es ermöglicht die Einzelstränge erneut zu amplifizieren und so neue Cluster zu bilden. Nach dieser zweiten Amplifikation wird nun der andere Primer geschnitten und die Sequenzierung für den komplementären Strang wiederholt.

## B3.7 Kontaminationskontrolle

Eines der großen Probleme bei der Arbeit mit alter DNA sind Kontaminationen. Dies betrifft Kontaminationen durch die Bearbeiter, mikrobielle Verunreinigungen bei Bodenlagerung, die Verwendung unsauberer Chemikalien und Kreuzkontaminationen der Proben während der unterschiedlichen Bearbeitungsschritte. Im besten Fall werden bereits während der Bergung von Skelettfunden entsprechende Sicherheitsvorkehrungen, wie das Tragen von Handschuhen, Mundschutz und Haube, getroffen und die Proben direkt vor Ort entnommen. Oft ist eine direkte Probennahme auf der Grabung jedoch nicht möglich und Knochen werden, wie auch im Fall der hier verwendeten Proben, teilweise bis zu Jahrzehnten in anthropologischen Sammlungen aufbewahrt bevor sie beprobt werden. Aus diesem Grund haben sich eine Reihe von Dekontaminationsverfahren und Kontaminationskontrollen etabliert, die eine zuverlässige Analyse endogener DNA ermöglichen sollen. Die Probenbearbeitung bis zur DNA-Amplifikation wird in einem speziellen Reinraumlabor durchgeführt – strikt getrennt von den Post-PCR-Laboren, um einen Übertrag bereits amplifizierter DNA (*carry over* Kontamination) zu vermeiden. Um einer Kontamination durch die Bearbeiter vorzubeugen, muss spezielle Schutzkleidung, inklusive Overall, Überschuhen, Haube, Mundschutz und drei Paar Handschuhe, getragen werden. Zur Dekontamination der Proben werden sie einer mechanischen Oberflächenentfernung unterzogen und zusätzlich mit UV-Licht bestrahlt, was zu Basenmodifikationen führt, die wiederum eine Amplifikation entsprechender DNA-Abschnitte verhindert (OU ET AL. 1991). Auch die Reaktionsgefäße und sonstige Labormaterialien werden, soweit dies möglich ist, mit UV-Licht bestrahlt. Chemikalien werden vor ihrer Verwendung auf Kontaminationen getestet. Während der Pulverisierung der Probe, der Extraktion der DNA und deren Amplifikation werden Kontrollen mitgeführt um die Sauberkeit der Arbeitsprozesse, Geräte und Chemikalien zu überprüfen. Des Weiteren werden wenn möglich zwei unabhängige Extraktionen einer Probe durchgeführt. Die DNA-Amplifikationen und Sequenzanalysen werden wiederholt um die ersten Ergebnisse zu bestätigen, oder sporadische Kontaminationen aufzudecken. Kontaminierte Kontrollen werden in der Regel so analysiert wie auch die Proben und die Ergebnisse werden mit denen der Proben verglichen, die gleichzeitig bearbeitet wurden.

## B4 Verarbeitung der Rohdaten

### Sanger Sequenzierung

Die Ausgabe des ABI PRISM 3130 *Genetic Analyzer* sind sog. abi-Dateien, die Sequenz- und Farbinformation verknüpfen. Mit Hilfe der ABI PRISM *Sequencing Analysis*<sup>TM</sup> 3.0 Software werden die ausgegebenen Chromatogramme auf die richtige Basenzuordnung hin kontrolliert und gegebenenfalls korrigiert.

### Roche 454 GS FLX

Die aus den 454-Sequenzierungen resultierten FASTA-Dateien, wurden mit Hilfe eines *Perl* Skripts, das von B. Rieger (Institut für Genetik, Johannes Gutenberg Universität Mainz) geschrieben wurde, sortiert. Als Eingabe dienen die FASTA-Datei aus der Sequenzierung, eine TXT-Datei mit den TAG Sequenzen und eine TXT-Datei mit den Primersequenzen. Das Skript sortiert die Sequenzdaten zuerst nach Meyer-TAGs und bildet in einem individuellen Projektordner, der zu Beginn des Skriptes kreiert wird, einen Unterordner für jeden gefundenen TAG. Anschließend werden die Sequenzen, die einem TAG zugeordnet werden konnten nach den verwendeten Primern sortiert, wobei in jedem TAG Ordner FASTA-Dateien für jeden gefundenen Primer gebildet werden. Aufgrund der Tatsache, dass das 454GS FLX Gerät Schwierigkeiten hat in Homopolymerregionen die exakte Basenanzahl wiederzugeben, kann man auswählen, dass vor der Zuordnung von TAGs und Primern mehrere gleiche Basen hintereinander auf eine reduziert werden. Optional können die Primersequenzen abgeschnitten werden. Die Sequenzen können somit eindeutig der ursprünglichen Probe zugeordnet werden und entweder für jedes Primerpaar getrennt oder Probenweise aligniert werden.

Da für diese Arbeit Amplikons sequenziert wurden, befinden sich in den Ausgabedateien große Mengen an redundanter Information durch die Vervielfältigungsschritte. Mit Hilfe des Programms cdHIT454 (NIU *ET AL.* 2010) werden identische Sequenzen auf eine reduziert.

## Sequenz-Alignierung und Bildung der Konsensussequenz

Die Sequenzen aus der Sanger- und der 454-Sequenzierung wurden mit Hilfe des Programm SeqMan Pro™ der DNASTAR Lasergene® Pakete 8, 9 und 10 an die jeweiligen Referenzsequenzen aligniert. Da hier die Ergebnisse aus verschiedenen Sequenzierungen zum ersten Mal gemeinsam betrachtet werden können, wurde in einem ersten Schritt verstärkt auf Widersprüche zwischen den einzelnen Sequenzen geachtet. Anschließend wurde das gleiche Programm verwendet, um für jede Probe die Konsensussequenzen der HVR1 und der Fragmente der kodierenden Region zu erstellen. Zur Darstellung der Alignments wurde das Programm MegAlign™ des DNASTAR Lasergene® Paketes 11 verwendet.

## Verarbeitung der Rohdaten aus der Illumina-Sequenzierung

Die Auswertung der 50bp *paired end* Illumina-Sequenzierung wurde von C. S. mit Hilfe verschiedener frei zugänglicher Programme durchgeführt. Die Programme und ihre Funktion sollen hier deshalb nur kurz genannt werden:

Die Sequenzen wurden von der Firma GATC bereits nach den Indexen sortiert.

**Tab. B 35 Analyse-Pipeline der Illumina Sequenzdaten**

	<i>Programm</i>	<i>Funktion</i>
1	KeyAdapterTrimFastQ_cc.py	Die Sequenzieradapter werden von den DNA-Fragmenten entfernt; Adapter-Dimere und –Chimere werden entfernt
2	QualityFilterFastQ.py	Qualitätsfilter
3	ea-utils: FastqJoin	Verbinden der <i>paired-end-reads</i>
4	BWA	Indexen des Referenzgenoms
5	BWA	Erstellen eines Burrows-Wheeler-Alignments
6	Samtools	Bearbeitung des Alignments
7	Picardtools: MarkDuplicates	Entfernung von Duplikaten
8	GATK	Realignierung an indels, Rekalibrierung der Basenqualität und SNP call
9	Samtools: mpileup	Erstellen einer Konsensussequenz im .FASTA-Format

## B5 Methoden der Datenanalyse

### Analyse der mitochondrialen Daten

#### Authentifizierung der mitochondrialen Daten

Die für die populationsgenetischen Analysen verwendeten Sequenzen der HVR1 wurden als authentisch eingestuft, wenn aus drei unterschiedlichen PCR-Ansätzen das gleiche Ergebnis ermittelt werden konnte. Für die SNPs der kodierenden Region wurde eine 454 Sequenzierung als ausreichend anerkannt, sofern sich zwischen den detektierten SNPs und zur HVR1 keine Widersprüche ergaben. Durch den hierarchischen Aufbau der Haplogruppenzuordnung anhand von SNPs der kodierenden Region des mitochondrialen Genoms lassen sich Widersprüche leicht erkennen und ausschließen. Aus dem vereinfachten mitochondrialen Stammbaum in Abb. A 1 lässt sich dieser hierarchische Aufbau erkennen. Beispielsweise wird eine Probe nur dann der Haplogruppe A zugeordnet, wenn sie neben dem SNP an Position 663 auch die Abweichungen zur rCRS an den Positionen zeigt, die im Stammbaum zwischen der rCRS und der Haplogruppe A liegen, also 7028T, 14766T, 11719A, 12705T und 663G.

#### Zuordnung der mitochondrialen Haplogruppen

Die ermittelten Konsensussequenzen der einzelnen Proben wurden mit der rCRS (*revised Cambridge Reference Sequence*) (ANDREWS *ET AL.* 1999) verglichen und die Unterschiede zu dieser notiert. Mit Hilfe des Programms *HaploGrep* (KLOSS-BRANDSTÄTTER *ET AL.* 2011) wurden alle detektierten SNPs, also der HVR1 und der kodierenden Region, jeder Probe, mit dem *PhyloTree* verglichen. *PhyloTree* ist eine Datenbank, in der die globale mtDNA Variation erfasst und ständig aktualisiert wird - die hier verwendete Version *mtDNA tree Build 15* wurde im Oktober 2012 veröffentlicht (VAN OVEN & KAYSER 2009).

## Biostatistik

Um eine Aussage über die populationsgenetischen Zusammenhänge der reiternomadischen Gruppen treffen zu können, wurden verschiedene biostatistische Verfahren angewendet.

Als Maß für die Variabilität innerhalb einer Population dienen die Haplotypendiversität  $\hat{H}$  und die Nukleotiddiversität  $\hat{\pi}$ . Erstere berechnet sich aus der Probenanzahl  $n$ , der Anzahl verschiedener Haplotypen  $k$  und der Frequenz des jeweiligen Haplotypen  $p_i$ .

$$\hat{H} = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2\right)$$

Die Nukleotiddiversität berechnet sich aus der Wahrscheinlichkeit, dass zwei zufällig ausgewählte Nukleotide an der gleichen Sequenzposition unterschiedlich sind. Sie ist das Äquivalent zur Haplotypendiversität auf Nukleotidebene.

$$\hat{\pi}_n = \frac{\sum_{i=1}^k \sum_{j<i} p_i p_j d_{ij}}{L}$$

Die Variabilität zwischen den räumlich und kulturell differenzierten Gruppen wird mithilfe des  $F_{ST}$ -Wertes ermittelt. Wobei ein niedriger  $F_{ST}$ -Wert eine geringe genetische Distanz und somit engere Verwandtschaft bedeutet. Bestimmt wurden paarweise  $F_{ST}$ -Werte zwischen den Populationen anhand einer Distanzmatrix paarweiser Sequenzunterschiede zwischen den Individuen einer Population.

$$F_{ST} = \frac{\sigma_a^2}{\sigma_T^2}$$

Aufschluss darüber, ob die genetischen Unterschiede der räumlich bzw. kulturell differenzierten Gruppen untereinander größer sind als innerhalb der einzelnen Gruppen und sie sich somit klar voneinander differenzieren lassen, gibt die AMOVA (*Analysis of Molecular Variance*) (EXCOFFIER ET AL. 1992). Die Populationen werden für diese Analyse zu Gruppen gefasst, damit beispielsweise die frühen mit den klassischen Skythen oder die östlichen mit den westlichen Skythen verglichen werden können. Die Kriterien der Gruppenzusammensetzung werden hierbei vom Bearbeiter bestimmt, um hypothetische Populationszusammenhäng oder -differenzierungen zu überprüfen. Bei der AMOVA werden drei Werte bestimmt:

$F_{CT}$  = Varianz zwischen Gruppen relativ zur Gesamtvarianz

$$F_{CT} = \frac{\sigma_a^2}{\sigma_T^2}$$

$F_{SC}$  = Varianz zwischen Populationen innerhalb von Gruppen

$$F_{SC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2}$$

$F_{ST}$  = Varianz zwischen Populationen relativ zur Gesamtvarianz

$$F_{ST} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_T^2}$$

Die Möglichkeit einer unvoreingenommenen Analyse bietet das Programm SAMOVA 1.0 (*Spatial Analysis of Molecular Variance*), denn es bildet selbsttätig Gruppen aus einem vorgegebenen Datensatz (DUPANLOUP ET AL. 2002). Lediglich die Anzahl der Gruppen wird vom Bearbeiter vorab festgelegt. Zusätzlich zu den molekularen Daten, also den Sequenzen, werden dem Programm auch die geographischen Informationen der Populationen gestellt. SAMOVA fasst die Populationen folgend zu immer neuen Gruppen zusammen und testet, bei welchen Kombinationen die innere Variabilität ihren Minimalwert bzw. die Variabilität zwischen den Gruppen ihren Maximalwert erreicht. Innerhalb des Analyseprozesses werden zunächst die geographisch benachbarten Populationen zusammengefasst, um die Analyse sukzessive auf Gruppierungen weiter entfernter Populationen auszuweiten. Mit diesem Vorgehen kann eruiert werden, ob genetische Differenzierungen aufgrund geographischer

Barrieren eine Rolle spielen. Da dem Programm SAMOVA eine veraltete Version von AMOVA zugrunde liegt und die Berechnung der Signifikanzen deshalb nicht zuverlässig ist (Dupanloup, persönliche Mitteilung), wurden die von SAMOVA gebildeten Gruppen anschließend einer AMOVA unterzogen um somit die Signifikanz zu prüfen.

Zusätzlich wurden für die Populationen  $F_u$ 's  $F_S$ -Werte bestimmt, die als Test benutzt werden um festzustellen ob sich eine Population im Gleichgewicht befindet und Selektionsneutralität herrscht.  $F_u$  hat allerdings festgestellt, dass die  $F_S$  Statistik sensitiv auf demographische Expansion reagiert, was sich in negativen  $F_S$ -Werte widerspiegelt (FU 1997). Berechnet wird die Wahrscheinlichkeit  $S'$ , eine zufällige neutrale Stichprobe mit einer bestimmten Anzahl an Allelen zu beobachten, die gleich groß oder größer als der tatsächlich beobachtete Wert ist, gegeben die beobachtete Zahl an paarweisen Sequenzunterschieden als Schätzwert für  $\theta$ .

$$F_S = \ln\left(\frac{S'}{1-S'}\right) \quad \text{Wobei } S' = \Pr(K \geq k_{obs} | \theta = \hat{\theta}_\pi)$$

Die Berechnungen der Haplotypen- und Nukleotiddiversität, des Fixationsindex  $F_{ST}$ , der  $F_u$ 's  $F_S$ -Werte sowie die AMOVA wurden mit Hilfe des Programmpakets Arlequin 3.5.1.2 (EXCOFFIER & LISCHER 2010) durchgeführt. SAMOVA 1.0 ist ein eigenständiges Programm, das, wie auch Arlequin, frei aus dem Internet heruntergeladen werden kann. Für alle statistischen Analysen wurden jeweils die HVR1 Sequenzen der untersuchten Individuen von Position 16040 bis 16400 verwendet.



## Analyse der nukleären Daten

### Authentifizierung der nukleären Daten

Bei der Auswertung nukleärer *loci* ist darauf zu achten, dass diese in drei möglichen Ausprägungen vorliegen können. Da für jeden *locus* zwei mögliche Allele existieren, ergeben sich daraus die Möglichkeiten der homozygot ancestralen Form, die heterozygote Ausprägung, oder die homozygot abgeleitete Form. Die Reproduktion der Ergebnisse ist hier von besonderer Bedeutung, da es speziell bei aDNA häufig zu einem sog. *allelic dropout* kommen kann. Das heißt, dass bei der Sequenzierung ein Allel unterrepräsentiert, oder gar nicht erfasst wird und der untersuchte Genort somit fälschlicherweise als homozygot eingestuft wird. Die Ergebnisse wurden als gesichert anerkannt wenn diese durch drei 454 Sequenzierungen aus unabhängigen PCRs bestätigt wurden. Unter Vorbehalt wurden auch die Ergebnisse mit einbezogen, die durch zwei 454 Sequenzierungen aus unabhängigen Amplifikationen bestätigt wurden. Vor allem bei heterozygoten *loci* liegt hier kein Zweifel an der Authentizität vor, lediglich die homozygoten *loci* sind nach nur zwei Sequenzierungen mit Vorsicht zu betrachten.

### Bestimmung des Genotyps und der Allelfrequenzen

Die nukleären Sequenzen wurden nach *locus* und Individuum getrennt, mit Hilfe des Programms SeqMan Pro™ des Lasergene® Pakets aligniert und aus den Alignments die entsprechenden Konsensussequenzen gebildet. Die Konsensussequenzen wurden mit den jeweiligen Referenzsequenzen abgeglichen und für die untersuchten SNPs die vorliegenden Allele notiert. Anschließend wurde die Frequenz  $p$  für das abgeleitete Allel  $A$  für jeden untersuchten Genort bestimmt, indem die Anzahl der abgeleiteten Allele durch die Gesamtzahl der Allele geteilt wurden.

$$p(A) = \frac{N_A}{N_{ges}}$$

Die ermittelten Allelhäufigkeiten im untersuchten Datensatz wurden mit der NCBI SNP-Datenbank abgeglichen. Diese enthält unter anderem Daten des *HapMap Projects*, des EGP (*Environmental Genome Project*) und des *1000Genomes Project*. Für die meisten loci wurden die Daten des *HapMap Projects* als Vergleich herangezogen. Das *HapMap Project* enthält Daten folgender Populationen: Yoruba aus Ibadan, Nigeria; Japaner aus Tokio; Han Chinesen aus Beijing und als Vertreter europäischer Populationen Daten des CEPH (*Centre d'Etude du Polymorphisme Humain*), gewonnen von Bewohner Utahs mit nord- und westeuropäischen Wurzeln. Für Allelvorkommen in Zentralasien wurden noch keine Daten erhoben, weswegen die im Zuge dieser Arbeit gewonnenen Daten in Relation zu den Utah-Europäern und den Han Chinesen als Vertreter Ostasiens betrachtet wurden.

Um den schwankenden Populationsgrößen für die untersuchten Marker Rechnung zu tragen, wurden für die ermittelten Frequenzen binomiale Konfidenzintervalle gebildet. Die Berechnung erfolgte nach einer Formel von CLOPPER & PEARSON 1934, die sich besonders für kleine Stichproben eignet und auch in der Lage ist adäquate Intervalle für Frequenzen von 100% oder 0% angeben zu können.

$$\sum_{j=k}^n \binom{n}{j} p_U^j \times (1 - p_U)^{n-j} = \frac{1 - \delta}{2}$$

$$\sum_{j=0}^k \binom{n}{j} p_O^j \times (1 - p_O)^{n-j} = \frac{1 - \delta}{2}$$

Sonderfall: Für  $k = 0$  setze  $p_U = 0$ , für  $k = n$  setze  $p_O = 1$ .

Wobei  $p_U$  und  $p_O$  die untere bzw. obere Intervallgrenze markieren;  $k$  ist im vorliegenden Fall die Menge abgeleiteter Allele und  $n$  dementsprechend die Menge aller Allele;  $\delta$  gibt das Vertrauensniveau, hier: 0,95.

## B6 Ergebnisse

### B6.1 DNA-Erhalt

Von den ursprünglich zur Analyse vorgesehenen 126 Proben mussten 15 aufgrund unsicherer Datierung oder Problemen bei der Zuordnung zu einem Individuum von den Analysen ausgeschlossen werden. Die restlichen 111 Proben wurden nach der Extraktion mit mitochondrialen Primern auf ihren DNA-Gehalt hin getestet. Eine Probe musste von den Analysen ausgeschlossen werden obwohl sie einen guten DNA-Erhalt zeigte, da das Probenmaterial nicht ausreichte, um ein vollständiges DNA Profil zu erhalten. Bei fünf (4,5%) der getesteten Proben konnte keine DNA nachgewiesen werden. Weitere acht Proben lieferten zwar sporadische Amplifikationsergebnisse. Der Erhalt war jedoch zu schlecht um ein eindeutiges mtDNA Profil zu erhalten. Aufgrund des mangelhaften DNA-Gehaltes mussten somit insgesamt 13 von 111 Proben von den Analysen ausgenommen werden, was 11,7% entspricht. Insgesamt konnten also 97 Proben für die Analysen verwendet werden.

**Tab. B 36 Probenanzahl nach Region**

Angegeben ist die Anzahl der Proben, die zur Untersuchung vorgesehen waren, die tatsächlich analysiert wurden und diejenigen, die eine ausreichende Menge an DNA enthielten.

<i>Region</i>	<i>Proben</i>			<i>% Erfolg der untersuchten Proben</i>
	<i>Gesamt</i>	<i>Untersucht</i>	<i>Verwendet</i>	
Minusinsker Becken	6	6	6	100%
Tuva	23	15	15	100%
Russisches Altai	41	39	36	92,3%
Kasachisches Altai	14	10	10	100%
Ostkasachstan	12	11	11	100%
Südkasachstan	3	3	0	0%
West-Russland	27	27	19	70,4%
Gesamt	126	111	97	87,4% Gesamt Ost: 96,3%

## B6.2 Amplifikationserfolg

Der allgemein gute DNA-Erhalt spiegelt sich auch im Amplifikationserfolg wieder. 96% der durchgeführten PCRs lieferten positive DNA Ergebnisse. Wobei eine Amplifikation als erfolgreich gewertet wurde, wenn eine sichtbare DNA-Bande nach der Gelelektrophorese verzeichnet werden konnte. In der Aufschlüsselung des Amplifikationserfolges nach Region zeigt sich ein besserer Erhalt in den östlichen bzw. nördlichen Gebieten im Vergleich zum Süden und Westen (s. Tab. B 37 und B 38). Da die mitochondriale Multiplex-PCR jeweils aufgeteilt auf zwei bis drei Ansätze amplifiziert wurde, beziehen sich die Angaben für den Amplifikationserfolg auf die Anzahl an Ansätzen. Für die Amplifikation nukleärer DNA wurden nur Individuen verwendet, bei denen die Amplifikation mitochondrialer DNA ein positives Ergebnis lieferte, wobei die nukleären *loci* ausschließlich im Multiplex-PCR-Ansatz vervielfältigt wurden. Alle 52 Individuen, die für die Analyse der nukleären *loci* ausgewählt wurden, zeigten positive DNA Ergebnisse, wobei jedoch bei einigen Proben der DNA-Gehalt zu gering war, um ein vollständiges nukleäres Profil zu erhalten, bzw. die Ergebnisse ausreichend zu reproduzieren.

**Tab. B 37 Amplifikationserfolg nach Region**

	<i>Gesamt</i>				
	<i>ges</i>	<i>pos</i>	%	<i>neg</i>	%
gesamt	1532	1470	96	62	4
Altai	646	632	97,8	14	2,2
Tuva	222	217	97,7	5	2,3
Minusinsk	79	79	100	0	0,0
Ostkasachstan	177	170	96	7	4,0
Südkasachstan	7	1	14,3	6	85,7
West-Russland	394	364	92,4	30	7,6

**Tab. B 38 Amplifikationserfolg nach PCR-Methode und Region**

	<i>mtPCR</i>					<i>mtMP</i>					<i>nuMP</i>				
	<i>ges</i>	<i>pos</i>	%	<i>neg</i>	%	<i>ges</i>	<i>pos</i>	%	<i>neg</i>	%	<i>ges</i>	<i>pos</i>	%	<i>neg</i>	%
gesamt	998	949	95,1	49	4,9	800	789	98,6	11	1,4	237	226	95,4	11	4,6
Altai	398	392	98,5	6	1,5	386	386	100	0	0	103	95	92,2	8	7,8
Tuva	129	124	96,1	5	3,9	157	152	96,8	5	3,2	39	39	100	0	0
Minusinsk	45	45	100	0	0	41	41	100	0	0	21	21	100	0	0
Ostkasachstan	87	82	94,3	5	5,7	93	93	100	0	0	48	46	95,8	2	4,2
Südkasachstan	5	1	20	4	80	6	0	0	6	100	0	0	0	0	0
West-Russland	329	300	91,2	29	8,8	117	117	100	0	0	26	25	96,2	1	3,8

Der Amplifikationserfolg der für diese Studie entworfenen Primerpaare kann nur anhand der Sequenzierergergebnisse beurteilt werden, da viele der Fragmente eine ähnliche Länge aufweisen und dadurch in der Gelelektrophorese nicht separiert werden können. Ein positives Sequenzierergergebnisse wurde deshalb als Amplifikationserfolg gewertet (s. Tab. B 39). Zum Vergleich wurden die Ergebnisse der 454-Sequenzierung den Ergebnissen aus den Sanger-Sequenzierungen gegenübergestellt.

**Tab. B 39 Amplifikationserfolg pro Fragment**

Evaluert anhand positiver Ergebnisse aus 454- und Sanger-Sequenzierung.

		<i>MP - 454</i>			<i>MP - Sanger</i>		
	<i>Primer</i>	<i>Fragmente</i>	<i>Sequenzen</i>	<i>% Erfolg</i>	<i>Fragmente</i>	<i>Sequenzen</i>	<i>% Erfolg</i>
1	423U/485L	211	145	68,7	9	4	44,4
2	654U/699L	211	195	92,4	9	8	88,9
3	2992U/3057L	211	175	82,9	9	5	55,6
4	4155U/4221L	211	181	85,8	9	7	77,8
5	4499U/4554L	211	200	94,8	9	6	66,7
6	4549U/4617L	211	158	74,9	9	6	66,7
7	4815U/4864L	211	194	91,9	9	9	100,0
8a	4871U/4940L	208	177	85,1	9	3	33,3
8b	4889U/4956L	3	3	100,0	6	5	83,3
9a	5163U/5179L	142	132	93,0	0	0	-
9b	5163U/5201L	69	68	98,6	9	8	88,9
10	5836U/5875L	211	197	93,4	9	6	66,7
11	6336U/6403L	211	183	86,7	9	8	88,9
12	6764U/6814L	211	170	80,6	9	6	66,7
13	6975U/7046L	211	160	75,8	9	6	66,7
14	8226U/8287L	211	178	84,4	9	6	66,7
15	8385U/8417L	211	138	65,4	9	7	77,8
16	8932U/8996L	211	159	75,4	9	8	88,9
17	9072U/9124L	211	174	82,5	9	8	88,9
18a	10000U/10048L	188	2	1,1	6	0	0,0
18b	9932U/10074L	23	13	56,5	3	3	100,0
19	10105U/10166L	211	158	74,9	9	8	88,9
20	10387U/10451L	211	171	81,0	9	8	88,9
21	10865U/10895L	211	200	94,8	9	7	77,8
22	11700U/11743L	211	195	92,4	9	6	66,7
23	11935U/11983L	211	189	89,6	9	6	66,7
24a	12303U/12352L	208	188	90,4	3	1	33,3
24b	12281U/12337L	3	1	33,3	6	0	0,0
25	12692U/12754L	211	179	84,8	9	6	66,7
26a	13231U/13275L	208	194	93,3	3	3	100,0
26b	13231U/13266L	3	3	100,0	6	5	83,3
27	13620U/13683L	211	174	82,5	9	6	66,7
28	13701U/13740L	211	191	90,5	9	6	66,7
29	14717U/14784L	211	177	83,9	9	5	55,6
30	14783U/14839L	211	172	81,5	9	6	66,7

		<i>MP - 454</i>			<i>MP - Sanger</i>		
	<i>Primer</i>	<i>Fragmente</i>	<i>Sequenzen</i>	<i>% Erfolg</i>	<i>Fragmente</i>	<i>Sequenzen</i>	<i>% Erfolg</i>
31	16011U/16088L	208	164	78,8	9	7	77,8
32	16071U/16153L	208	96	46,2	9	6	66,7
33	16119U/16207L	208	155	74,5	9	7	77,8
34	16185U/16271L	208	144	69,2	9	7	77,8
35	16233U/16305L	208	106	51,0	9	6	66,7
36	16274U/16345L	208	173	83,2	9	9	100,0
37	16340U/16413L	208	143	68,8	9	7	77,8
38	16013U/16152L	3	2	66,7			
39	16122U/16271L	3	1	33,3			
40	16235U/16346L	3	2	66,7			
41	16274U/16410L	3	2	66,7			

### B6.3 Sequenziererfolg

Für die vorliegende Arbeit wurden 13 454-Sequenzierungen (*runs*) durchgeführt. Durch die individuelle Markierung der DNA-Fragmente einer Probe mit einer Erkennungssequenz, dem TAG, war es möglich in jedem run mehrere Proben gleichzeitig zu analysieren. Insgesamt wurden 16 Probenpools erstellt: Amplikon 1–16. Abgesehen von der Markierung der Proben wurden die weitere Sequenzier Vorbereitung und die eigentliche Sequenzierung von zwei Firmen übernommen, wobei run 10 (Amplikon 11) bei der Firma LGC Genomics GmbH in Berlin und die restlichen Sequenzierungen bei der GATC Biotec AG in Konstanz durchgeführt wurden. Amplikon 9 + 10 und Amplikon 14 + 15 wurde jeweils gemeinsam sequenziert, wofür sie von der Firma GATC Biotec AG mit einem zusätzlichen Barcode versehen wurden. Jede der Amplikon-Proben besteht aus einem Pool an Proben, deren DNA zuvor in mitochondrialen und nukleären Multiplex-PCRs vervielfältigt wurden. Zusätzlich zu den 454-runs wurden einige mitochondriale Multiplex-Reaktionen mit der Sanger Methode sequenziert (s. Tab. B 40).

**Tab. B 40 Sequenziererfolg der Multiplex-PCRs**

Anzahl der Multiplex-PCRs, erfolgreich amplifizierten und sequenzierten Proben und die Anzahl an Fragmente, zu denen ein Sequenzergebnis erhalten wurde.

<i>Sequenziermethode / DNA</i>	<i>MP-PCR-Ansätze</i>	<i>Probenanzahl</i>	<i>amplifizierte Fragmente</i>	<i>erfolgreich sequenzierte Fragmente</i>	<i>Sequenziererfolg % pos</i>
454 / mitochondrial	33	211	7798	6182	79,3
454 / nukleär	10	199	4577	3076	67,2
Sanger / mitochondrial	7	9	311	241	77,5

Wie viele Sequenzen bei den einzelnen *runs* produziert wurden ist Tab. B 41 zu entnehmen. Bei der Darstellung wurden die unterschiedlichen Methoden zur Sequenzbewertung durch das Perl-Skript *sort* berücksichtigt. Da die Hauptfehlerquelle der 454-Sequenzierung in der Detektion der korrekten Basenzahl an Homopolymerstellen liegt, wurde das Skript so konzipiert, dass die Daten auf unterschiedliche Weise sortiert werden können. Eine Option trägt dem Homopolymerproblem Rechnung indem alle Basen, die mehrfach hintereinander auftauchen, auf eine reduziert werden und bei der zweiten Möglichkeit werden die Sequenzen unverändert der originalen TAG-Struktur zugeordnet.

**Tab. B 41 Erfolg der 454-Sequenzierungen**

Für jeden *run* ist die Anzahl an Sequenzen insgesamt aufgeführt, die Anzahl an Sequenzen, die erfolgreich einem TAG zugeordnet werden konnten und diejenigen, die einem TAG und einem der verwendeten Primer zugeordnet werden konnten. Aus Letzteren berechnet sich der Sequenziererfolg.

<i>run</i>	Name	Homopolymere	Gesamt	TAG	TAG & Primer	nicht verwertbar	% Erfolg
1	Amplikon1	reduziert	112768	66024	55440	57328	49,16
		unreduziert	112768	59024	43513	69255	38,59
2	Amplikon2	reduziert	552130	461023	423043	129087	76,62
		unreduziert	552130	439576	335332	216798	60,73
3	Amplikon3	reduziert	389483	348919	285725	103758	73,36
		unreduziert	389483	320653	245464	144019	63,02
4	Amplikon4	reduziert	208284	187695	146190	62094	70,19
		unreduziert	208284	180243	132004	76280	63,38
5	Amplikon5	reduziert	265615	201713	179656	85959	67,64
		unreduziert	265615	191521	148207	117408	55,80
6	Amplikon6	reduziert	189402	145503	132216	57186	69,81
		unreduziert	189402	140173	123038	66364	64,96
7	Amplikon7	reduziert	156426	114634	105365	51061	67,36
		unreduziert	156426	110878	94144	62282	60,18
8	Amplikon8	reduziert	227552	217702	189776	37776	83,40
		unreduziert	227552	210897	175452	52100	77,10
9	Amplikon9	reduziert	151908	108513	92599	59309	60,96
		unreduziert	151908	101580	84971	66937	55,94
10	Amplikon10	reduziert	207605	178742	152621	54984	73,52
		unreduziert	207605	169905	139650	67955	67,27
11	Amplikon11	reduziert	162196	134370	108717	53479	67,03
		unreduziert	162196	34296	26446	135750	16,30
12	Amplikon12	reduziert	119486	107222	90464	29022	75,71
		unreduziert	119486	102245	82983	36503	69,45
13	Amplikon14	reduziert	231870	171120	129168	102702	55,71
		unreduziert	231870	160648	118027	113843	50,90
14	Amplikon15	reduziert	180328	157256	121122	59206	67,17
		unreduziert	180328	147294	110020	70308	61,01
15	Amplikon16	reduziert	120702	114825	99728	20974	82,62
		unreduziert	120702	105817	85387	35315	70,74

Amplikon1–12 enthalten zusätzliche Proben anderer Bearbeiter. Amplikon14–16 umfassen nur skythische Proben, wobei für Amplikon16 lediglich nukleärer Multiplex-PCRs *gepoolt* wurden. Für *run* 12 und 13 können deshalb zusätzlich die tatsächlich alignierten Sequenzen dargestellt werden (s. Tab. B 42).

**Tab. B 42 Anteil erfolgreich alignierter Sequenzen**

mt = mitochondrial, nuk = nukleär

run	Name	Homopolymere	Gesamt	TAG	Anteil	TAG & Primer	aligniert	%
12	Amplikon14	reduziert	231870	171120	mt	14398	12948	89,9
					nuk	114770	20731	18,1
	Amplikon15	reduziert	180328	157256	mt	9922	9081	91,5
					nuk	111201	16167	14,5
13	Amplikon16	reduziert	120702	114825	nuk	99728	36985	37,1

Weiterhin wurde eine Sequenzierung, von zuvor durch Multiplex-PCRs amplifizierten Proben, mit dem Illumina HiSeq des Instituts für Genetik der Johannes Gutenberg-Universität Mainz durchgeführt. Die Daten waren jedoch nicht zu verwenden (s. Abb. B 6).

	+	+	+	+	+	+	+	+
	420	430	440	450	460	470	480	490
	+	+	+	+	+	+	+	+
NC_012920.seq	CTTTTAACAGTCACCCC-CCAAC TAACACATTATTTTCCCCTCCCCTCCCATACTACTAATCTCATCAATA							
5423	----	.A.	.AC.	.A.	.G.	.....	.....	.....
8203	----	.C.	.C.	.C.	.C.	.CC.	.CC.	.CC.
1678	----	.C.	.C.	.C.	.C.	.CC.	.C.	.C.
5303	----	.C.	.C.	.C.	.C.	.C.	.C.	.T.
2524	----	.C.	.C.	.C.	.C.	.TCT.	.C.	.T.
1299	----	.C.	.T.	.CC.	.TC.	.TC.	.C.	.C.
19542	----	.C.	.C.	.C.	.C.	.T.	.T.	.C.
2563	----	.....	.....	.....	.....	.....	.....	.....
6860	----	.G.	.....	.....	.....	.....	.....	.....
20078	----	.....	.....	.....	.....	.....	.C.	.....
2117	----	.C.	.....	.....	.....	.....	.C.	.C.
8494	----	.A.	.GA.	.AA.	.A.	.A.	.A.	.C.

**Abb. B 6 Ergebnis zur Illumina Amplikon Sequenzierung**

Alignment der Sequenzen aus der Amplikon-*library* für SNP-Position 456 (Primer: 423\_U/480\_L)



Eine zweite Illumina-Sequenzierung wurde nach einem mitochondrialen *Capture* von der GATC Biotech AG in Konstanz durchgeführt. Insgesamt wurden 18 Proben mit unterschiedlichen Indexen gemeinsam sequenziert. Jeweils sechs Proben wurden gemeinsam *gecaptured*. Dabei umfassen die Indexe 1 und 3–7 Proben des Eneolithikums und der Bronzezeit aus dem nördlichen Schwarzmeergebiet, Index 8–13 Proben des griechischen Neolithikums und die Indexe 14–19 skythenzeitliche Proben. Im Folgenden werden nur die Ergebnisse für die skythischen Proben dargestellt. Für Index 16 kann nur die Anzahl an *raw reads* angegeben werden, da nach der Zerstörung einer Festplatte die Rohdaten für diesen Index nicht wiederhergestellt und somit nicht ausgewertet werden konnten. Die *raw reads* einer Probe beziehen sich auf die Anzahl an Sequenzen die den probenspezifischen Index tragen. In Tab. B 43 ist die Anzahl an Sequenzen nach jedem Bearbeitungsschritt gegeben.

**Tab. B 43 Ergebnis der Illumina-Sequenzierung nach mitochondrialem Capture**  
trim = Adapter *trimming*; qual = Qualitätsfilter; Dup. = Duplikate

Index	Probe	raw reads: read 1+2	Seq. nach trimming	Seq. nach Qualitätsfilter	Sequenzpaare verbunden	an die rCRS aligniert	
						mit Dup.	ohne Dup.
14	A3	12547274	12430160	10109350	3697782	2152834	2022
15	A21	16502522	16436376	14143119	5269514	3901801	4376
16	Ak5_3	9978138					
17	Be9	45046496	45044423	42149406	9064994	4891305	28350
18	Bt1	20377164	20371977	18763196	4826242	2564251	8245
19	J12_9	10928212	10828225	9313790	3517819	1994123	9097

Nach Entfernung der Sequenzduplikate bleiben im Schnitt 0,09% der Rohdaten pro Index, die sich an das mitochondriale Genom alignieren lassen. Dieser Anteil reicht aus, um eine Abdeckung von 99,4–100% zu erreichen (Tab. B 44).

**Tab. B 44 Abdeckung des mitochondrialen Genoms**  
SD = Standardabweichung

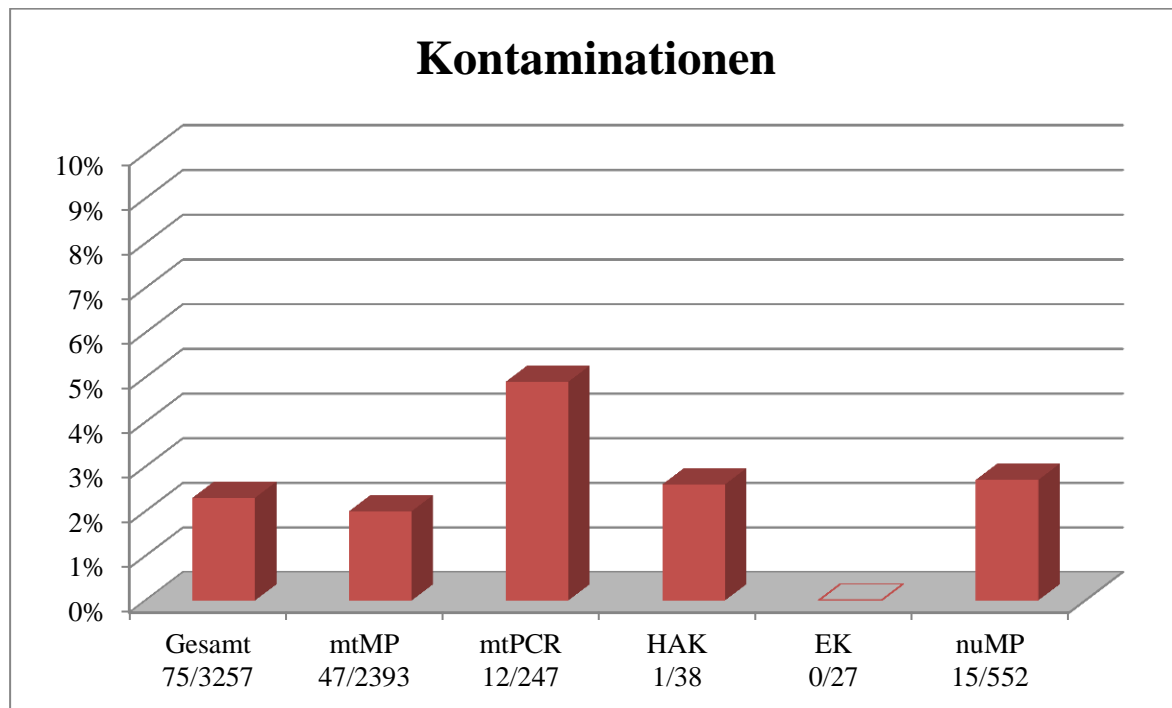
Index	Probe	Basen gesamt	nicht abgedeckte Basen	$\emptyset$ Abdeckung [Basen]	SD [Basen]	sequenzierte Basen	maximale Abdeckung	minimale Abdeckung
14	A3	16569	102	7,77	4,10	128667	27	0
15	A21	16569	3	16,70	6,64	276695	42	0
16	Ak5_3							
17	Be9	16569	0	143,39	23,98	2375783	175	6
18	Bt1	16569	2	34,36	11,04	569281	65	0
19	J12_9	16569	3	39,71	15,37	657912	93	0

## B6.4 Kontamination

Bei den durchgeführten Standard-PCRs wurden insgesamt 312 Kontrollen mitgeführt. Dabei handelt es sich um 247 reine PCR-Kontrollen, 38 Hydroxylapatitkontrollen (HAK), die seit dem Pulverisierungsschritt mitgeführt wurden und 27 Extraktionskontrollen (EK). Davon wurden 12 (4,9%) PCR-Kontrollen sowie eine (2,6%) der Hydroxylapatitkontrollen positiv auf DNA getestet. In den mtMultiplex-PCR-Ansätzen wurden insgesamt 187 Kontrollen mitgeführt. Von diesen zeigten 44 Gelbanden unterschiedlicher Stärke. Aufgrund der erhöhten Primermenge verglichen zu einer Standard-PCR ist es nicht immer ersichtlich, ob es sich um eine Kontamination, Primerdimere oder Sequenzartefakte handelt. 30 dieser Kontrollen wurden sequenziert um zu überprüfen, ob es sich um Kontaminationen handelt und welche Fragmente betroffen sind. Die 30 Kontrollen umfassen 385 Fragmente, von denen 33 (8,6%) lesbare Sequenzen ergaben und somit als echte Kontamination gewertet wurden. Diese Anzahl wurde auf die insgesamt 543 Fragmente aller 44 Kontrollen, die Gelbanden zeigten, hochgerechnet. So ergab sich eine theoretische Anzahl kontaminierter Fragmente von 47. Die 187 mitgeführten Kontrollen umfassen 2393 Fragmente. Die angegebene Kontaminationsrate ergibt sich aus dem Verhältnis von 47 Kontaminationen in 2393 Fragmenten (2%).

Im gleichen Verfahren wurde auch die Kontaminationsrate der nukleären Multiplex-PCRs berechnet. In jeder der 11 nukleären Multiplex-PCRs wurden mindestens 2 Leerkontrollen mitgeführt. Von insgesamt 24 Leerkontrollen zeigten 9 mehr oder weniger eindeutige DNA-Banden nach der Gelelektrophorese, die auf Kontaminationen hindeuten können. Sechs der neun putativ kontaminierten Kontrollen wurden bei verschiedenen 454 Sequenzierungen mitgeführt. Bei 23 Fragmenten je Probe ergeben sich für die 6 Leerkontrollen somit 138 mögliche kontaminierende Fragmente, von denen 10 lesbare Sequenzen ergaben. Hochgerechnet ergibt das 15 Fragmente für alle 9 Kontrollen mit insgesamt 207 Fragmenten. Insgesamt ergibt sich somit eine Kontaminationsrate von 15 Fragmenten aus 552 (24 Leerkontrollen mit je 23 Fragmenten), was einer Kontaminationsrate von 2,7% entspricht.

Insgesamt wurden 75 (2,3%) kontaminierende Fragmente in 3257 getesteten Fragmenten detektiert (s. Abb. B 7).



**Abb. B 7 Kontaminationsrate**

Angegeben für die Kontrollen insgesamt und aufgeteilt auf die verschiedenen Analyseschritte; mtMP = mitochondriale Multiplex-PCR, mtPCR = mitochondriale PCR, HAK = Hydroxylapatitkontrolle, EK = Extraktionskontrolle, nuMP = nukleäre Multiplex-PCR

### Kreuzkontamination während der 454-Sequenzierung

Zur Überprüfung auf Kreuzkontaminationen während der Probenvorbereitung für die 454-Sequenzierung wurden die Sequenzen jedes Probenpools (Amplikon1–16) jeweils nach allen TAGs (1–60) sortiert – egal ob verwendet oder nicht. So kann auf Überträge beim Ansetzen der TAGs oder bei deren Ligation an die Probenfragmente kontrolliert werden. Die Sortierung der DNA-Fragmente anhand der TAGs wurde jeweils nach zwei unterschiedlichen Methoden durchgeführt. Gesucht wurde immer nach dem TAG-Fragment ab der Restriktionsschnittstelle, wobei die Homopolymerregionen des TAGs einmal vor der Sortierung reduziert wurden und einmal unreduziert zugeordnet wurden. Die Analysen mit reduzierten Homopolymeren zeigen für die meisten *runs* TAGs, denen Sequenzen zugeordnet werden konnten, obwohl sie nicht verwendet wurden. Bei der Zuordnung von Sequenzen zu TAGs ohne die Homopolymerregionen zu reduzieren, werden überschüssigen TAGs nur noch bei den Probenpools Amplikon7, 9, 10, 14 und 15 gefunden (s. Tab. B 45).

**Tab. B 45 Kreuzkontamination 454**

Aufgelistet sind die 454-runs, die Amplikon-Probenpools, die im jeweiligen *run* sequenziert wurden, die Methode nach der sortiert wurde, die TAG-Nummer nach der gesucht und die im jeweiligen Ansatz verwendet wurde und die TAGs, die zusätzlich zu den verwendeten gefunden wurden.

<i>run</i>	Name	Homopolymere	sortiert	verwendet	zusätzlich vorhanden
run1	Amplikon1	reduziert	1-60	6-25	27,35,41,54,59
		unreduziert			keine zusätzlichen TAGs gefunden
run2	Amplikon2	reduziert	1-60	1-25, ohne 4	27,29,33,39,40,53,54,59
		unreduziert			keine zusätzlichen TAGs gefunden
run3	Amplikon3	reduziert	1-60	1-39	41,48,50,54,59
		unreduziert			keine zusätzlichen TAGs gefunden
run4	Amplikon4	reduziert	1-60	1-34	36,41,54,59
		unreduziert			keine zusätzlichen TAGs gefunden
run5	Amplikon5	reduziert	1-60	1-36	41,47,50,54,59
		unreduziert			keine zusätzlichen TAGs gefunden
run6	Amplikon6	reduziert	1-60	1-40	41,54,59
		unreduziert			keine zusätzlichen TAGs gefunden
run7	Amplikon7	reduziert	1-60	1-60 ohne 25	25
		unreduziert			25
run8	Amplikon8	reduziert	1-60	1-60, ohne 25	keine zusätzlichen TAGs gefunden
		unreduziert			keine zusätzlichen TAGs gefunden
run9	Amplikon9	reduziert	1-60	1-51, ohne 25	25,52,53,54,55,56,57,59
		unreduziert			25,52,53,54,55,56,57,59
	Amplikon10	reduziert	1-60	13-60	6,7,10
		unreduziert			6
run10	Amplikon11	reduziert	1-60	1-9, 12-48	10,54,59
		unreduziert			keine zusätzlichen TAGs gefunden
run11	Amplikon12	reduziert	1-60	1-40	41,44,47,54,59
		unreduziert			keine zusätzlichen TAGs gefunden
run12	Amplikon14	reduziert	1-60	1-38	39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60
		unreduziert			39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59
	Amplikon15	reduziert	1-60	21-59	7,15,16
		unreduziert			keine zusätzlichen TAGs gefunden
run13	Amplikon16	reduziert	1-60	1-58	keine zusätzlichen TAGs gefunden
		unreduziert			keine zusätzlichen TAGs gefunden

## Kreuzkontamination während der Illumina-Sequenzierung

Bei der Vorbereitung zur Illumina-Sequenzierung wurde eine Leerkontrolle mitgeführt. Der DNA-Gehalt der Leerkontrolle wurde nach der Erstellung der *sequencing library* mit dem Bioanalyzer gemessen. Abgesehen von Primer- und Adapterdimeren konnten keine verunreinigenden DNA-Fragmente detektiert werden.

Nach dem *Capture* und der anschließenden Illumina-Sequenzierung wurden die Ergebnisse für jede Probe miteinander verglichen. Als Referenz wurden die bereits bekannten SNPs aus

den Sanger- und 454-Sequenzierungen herangezogen. Für den SNP *call* aus den Illumina Daten und zur Erstellung der Konsensussequenzen wurden zwei unterschiedliche Programme verwendet. Das Programm GATK geht für den SNP *call* von einem haploiden Genom aus und gibt die Daten in einer *.vcf*-Datei aus. Zur Erstellung der Konsensussequenzen wurde *Samtools* verwendet, das von einem diploiden Genom ausgeht. Durch die Verwendung unterschiedlicher Parameter unterscheiden sich die variablen Positionen aus beiden Programmen, weshalb die Ergebnisse beider Analysen für den SNP Vergleich angegeben werden. Verglichen wurden zum einen die Ergebnisse der skythischen Proben, die zusammen gecaptured und sequenziert wurden (s. Tab. B 46) und zum anderen eine der skythischen Proben mit einer der griechischen Proben, die zusammen sequenziert, aber nicht zusammen gecaptured wurden (s. Tab. B 47).

**Tab. B 46 Kreuzkontamination bei Illumina-Sequenzierung & Capture**

Vergleich der Ergebnisse aus der Illumina-Sequenzierung (*vcf* + *fasta*) der Proben, die gemeinsam *gecaptured* wurden, mit den Resultaten aus den Sanger- und 454-Sequenzierungen (454). kD = keine Daten verfügbar

Pos	rCRS	A3 (Index 14)			A21 (Index 15)			Be9 (Index 17)			Bt1 (Index 18)			J12_9 (Index 19)		
		454	vcf	fasta	454	vcf	fasta	454	vcf	fasta	454	vcf	fasta	454	vcf	fasta
663	A				G	G	R	G	G	G					G	G
4216	T										C	C	C			
4824	A			R	G	G	G	G	G	G					G	G
4833	A	G												G		
4917	A										G	G	G			
7028	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
10398	A	G												G		
10400	C	T												T		
10873	T	C												C		
11719	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
12705	C	T			T	T	T	T	T	T				T	T	T
14766	C	kD		T	T	T	T	T	T	T	T	T	T	T	T	T
14783	C	C												C		
16126	T										C	C	Y			
16223	C	T		Y	T	T	T	T	T	T				T	T	T
16263	T													C		
16286	C				T		Y									
16290	C				T	T	Y	T	T	T					T	Y
16292	C						M	A	A	A					A	M
16294	C			T							T	T	T			
16304	T			C							C	C	C			
16319	G			R	A	A	A	A	A	A					A	R
16325	T													C		
16362	T	C		Y	C	C	C	C	C	C				C	C	C

korrekte Zuordnung  
 SNP wurde nicht erkannt  
 falsch zugeordneter SNP

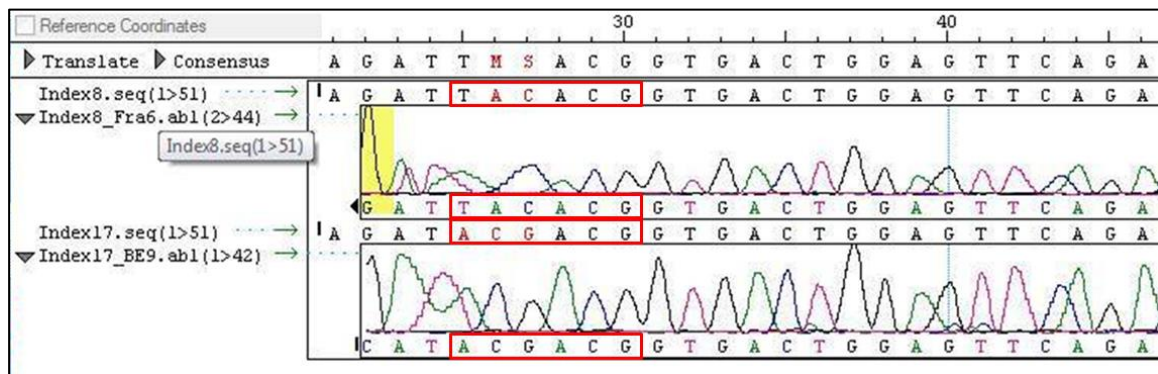
Der Vergleich zeigt, dass die SNPs der Proben Be9 und Bt1 korrekt und der Probe A21 nahezu korrekt zugeordnet wurden. Über 50% der variablen Positionen der Proben A3 und J12\_9 wurden entweder nicht oder falsch zugeordnet. SNPs, die nur in den Proben A3 und J12\_9 vorkommen, werden nicht detektiert, wohingegen in diesen Proben SNPs festgestellt werden, die in den anderen zu finden sind, was auf eine Kreuzkontamination zwischen den Proben hindeutet.

**Tab. B 47 Kreuzkontamination bei der Illumina-Sequenzierung**

Vergleich der Ergebnisse aus der Illumina-Sequenzierung (vcf + fasta) von zwei Proben, die gemeinsam sequenziert aber nicht zusammen *gecaptured* wurden, mit den Resultaten aus den Sanger- (San) und 454-Sequenzierungen (454). kD = keine Daten verfügbar

Pos	rCRS	Fra6 (Index8)			Be9 (Index17)		
		San	vcf	fasta	454	vcf	fasta
16223	C		T	T	T	T	T
16263	T						
16286	C						
16290	C		T	T	T	T	T
16292	C		A	A	A	A	A
16294	C						
16304	T						
16319	G		A	A	A	A	A
16325	T						
16362	T		C	Y	C	C	C
	korrekte Zuordnung						
	falsch zugeordneter SNP						

Zur Kontrolle auf eine Kreuzkontamination zwischen den Indexen, wurde mit dem Primer IS6 von den Libraries der Proben Fra6 (Index 8) und Be9 (Index 17) eine Sanger-Sequenzierung durchgeführt (s. Abb. B 8).



**Abb. B 8 Index Kontrolle durch Sanger-Sequenzierung**

Sequenziert wurden die Indexe 8 und 17 und die Libraries, die mit diesen Indexen markiert wurden.

Die Sequenzen zeigen keine Doppelpeaks, die das Resultat einer Kreuzkontamination zwischen den Indexen wären.

## B6.5 Haplogruppenzuordnung

Die Bestimmung der Haplogruppen wurde zum einen anhand der HVR1 Daten vorgenommen und zusätzlich anhand aller untersuchten Positionen, also der HVR1 und der SNPs der kodierenden Region. Die Übersichtstabellen mit allen detektierten SNPs finden sich im Anhang.

**Tab. B 48 Haplogruppen nach SNPs der *coding region* und der HVR1**

Zuordnung zum einen nach SNPs der HVR1 und weiterhin nach SNPs der HVR1 und der kodierenden Region

<i>Probe</i>	<i>Haplogruppe nach SNPs der HVR1</i>	<i>Haplogruppe nach allen SNPs</i>
BL1	C5+16093	C5+16093
BL2	U5a2c3/U5a2d1	U5a1a2
BL3	C5+16093	C5+16093
BL4	C5+16093	C5+16093
BL5	U2e2	U2e2
BL6	A4b	A4
A2	A8	A8
A3	M6/M9/G/M13'46'61+16362/M33+16362/D	G
A4	D4b1/D4d	D4b1a2a
A5	C5b1	C5b1
A6	U5a1f1	U5a1f1
A7	Y1	Y1
A8	U4a3/U4b1a1a1	U4a3/U4b1a1a1
A9	U5a1d2b	U5a1d2b
A10	A11	A11
A11	M6/M9/G/M13'46'61+16362/M33+16362/D	G
A14	T1a	T1a
A17	H1+16189/H1j9/H1ap1/H3+16189/R0+16189	H1+16189
A19	C	C4
A20	G2a	G2a
A21	A4	A4
Ak1_1	C4a1+16129	C4a1+16129
Ak4_1	A11	A11
Ak5_1	C	C4
Ak5_4	D4b1/D4d	D4b1/D4d
Ak5_5	U2e	U2e1a
Ak5_6	U2e	U2e1a
Ak5_7	C4a1+16129	C4a1+16129
Ak5_8	A11	A11
Ala1	C4a1+16129	C4a1+16129
Ala2	C	C4
Ala4	T	T
B1_1	M6/M9/G/M13'46'61+16362/M33+16362/D	G
B1_2	D4m2	D4m2
B3_1	T1a	T1a1b
BS1	X	X2b



<i>Probe</i>	<i>Haplogruppe nach SNPs der HVR1</i>	<i>Haplogruppe nach allen SNPs</i>
BS2	U4	U4b1a4
Bt1	T2b+!16296	T2b+!16296
Bt2	U2e2	U2e2a
D1	T1a	T1a1b
J12_1	T2b+!16296	T2b+!16296
J12_3	K1b1a2	U7
J12_6	Z	Z
J12_7	T2b+!16296	T2b+!16296
J12_8	A8	A8
J12_9	D4j4a	G1a1
J22_1	T2b+!16296	T2b+!16296
K1	CRS	H-CRS
Mo1	D4b1a2a1	D4b1a2a1
Mo2	L3a	D4h4a/D4j+16311
U1_1	M13a	W4a
U1_2	Z1a	Z1a
U2_1	F2a	F2a
U2_2	F2a	F2a
U4_1	U2e2	U2e2
U4_4	K	K
VK1_1	K1c2	K1b2
Be2	M6/M9/G/M13'46'61+16362/M33+16362/D	D4
Be3	CRS	H-CRS
Be4	A11	A11
Be6	A4	A6
Be8	M6/M9/G/M13'46'61+16362/M33+16362/D	D4
Be9	A4f	A4f
Be11	C4a1+16129	C4a1+16129
Be12	HV2	HV2
Be14	A4	A6
Ta_1	K2b1a	K2b1a
Is1	D4h1/D4j8	D4h1/D4j8
Is2	CRS	HV-CRS
Is4	H2a1	H2a1
Ze2	T2b	T2b
Ze3	K1a1/K1a3a+16093/K1a4d/K1a4f/K1a26/K1a21/K1b1b	K1a1/K1a3a+16093/K1a4d/K1a4f/K1a26/K1a21/K1b1b
Ze4	L3c/M74/D4h4a/D4j+16311/L4	D4h4a/D4j+16311
Ze5	I	I
Ze6	C4a1/C4b8/C4+152+16093	C4a1/C4b8/C4+152+16093
Ze7	U4a1	U4a1
Ze8	M6/M9/G/M13'46'61+16362/M33+16362/D	D4
Ze9	D4j3	D4j3
Pr1	U3	U3
Pr10	H2a1f	H2a1f
Pr11	T1a	T1a
Pr13	U5a1d2b	U5a1d2b
Pr3	M5/M7b'd/M13b	M5/M7b'd/M13b
Pr4	U1a'c	U1a'c
Pr5	T	T
Pr6	F1b	F1b
Pr7	L3e2b	N1a1a1a1a
Pr8	T2	T2

Probe	Haplogruppe nach SNPs der HVR1	Haplogruppe nach allen SNPs
Pr9	U2e2	U2e2
Kol1	X4	X4
Kol2	H8	H8c
Kol3	U4	U4
Kol5	J2b1a	J2b1a
Nov5	CRS	H-CRS
Nov7	U4b2	H1c
Nov9	keine HVR1 Daten	T2g1
Nov10	X	X

## B6.6 Genotypen

Die Genotypen der Proben wurden nur durch 454-Sequenzierungen bestimmt. Die Ergebnisse wurden als gesichert angesehen, wenn aus mindestens drei Sequenzierungen aus unabhängigen Multiplex-PCRs übereinstimmende Genotypen erhalten wurden. Der Großteil der Resultate basiert auf einem DNA-Extrakt (s. Tab. B 49).

**Tab. B 49 Genotypen**

Reproduzierte Genotypen sowie Proben, von denen zwei verschiedene DNA-Extrakte verwendet wurden sind fett gedruckt. Genotypen in Kleinbuchstaben sind nur durch zwei 454-Sequenzierungen aus unabhängigen PCRs belegt.

Marker	ABCB1a	ABCB1b	ABCB1c	ABCC11	ADH1Ba	ADH1Bb	CASP12	CYP3A4	CYP3A5	EDAR	HERC2	LCTa	LCTb	NAT2	SLC12A3	SLC24A5	SLC45A2	TRPV6b	TRPV6c	TYR	AMELXY															
Probe	C	T	G	T	A	C	T	G	A	C	T	G	A	T	C	A	G	C	T	G	A	T	C	G	C	G	A	C	G	A	C	T	C	A	X	Y
A2						<b>T/T</b>						a/a	c/c																							
A4	t/t	<b>G/T</b>	c/t	<b>T/T</b>	c/c			a/a	g/g		a/a	<b>C/C</b>	g/g	<b>T/T</b>				a/a	t/t	<b>C/C</b>	xy															
A6	c/t		t/t	c/c	c/c	g/g			g/g			<b>C/C</b>	g/g	<b>T/T</b>	g/g				t/t	c/c	xx															
A8	c/c	g/g	c/c									c/c		<b>T/T</b>				a/a	t/t	<b>C/C</b>																
A9	c/t		c/t	c/t	<b>C/C</b>	<b>G/G</b>		<b>A/A</b>	<b>G/G</b>	c/c	<b>A/A</b>	<b>C/T</b>	g/a	<b>T/T</b>	g/g			<b>A/A</b>		<b>C/C</b>	xx															
A11		g/t	<b>C/T</b>	<b>C/T</b>	c/c						a/a	<b>C/C</b>	g/g	<b>T/C</b>	<b>G/G</b>			<b>A/A</b>	t/t		xy															
A14		g/t	c/t	c/t	c/c	g/g	t/t		g/g	t/c	a/g	<b>C/C</b>	g/g	<b>T/T</b>	g/g	<b>G/G</b>		<b>A/A</b>	<b>T/T</b>	c/c	xy															
A21	<b>T/T</b>	<b>G/T</b>		<b>T/T</b>	c/c	g/g	<b>T/T</b>	a/a	<b>G/G</b>	c/c	<b>A/A</b>	<b>C/C</b>	g/g	<b>T/T</b>	g/g	g/a		<b>A/A</b>	t/t	c/c	xy															
BL1	<b>C/C</b>	<b>G/G</b>	<b>C/C</b>	<b>C/C</b>	<b>C/C</b>				g/g		<b>A/G</b>	<b>C/T</b>	g/a	<b>T/T</b>			g/g	<b>A/A</b>		<b>C/C</b>	xy															
BL2	<b>C/T</b>	g/t	<b>C/T</b>	<b>C/C</b>	<b>C/C</b>			a/a	g/g			c/t	g/a	<b>T/C</b>			<b>G/G</b>	a/a	t/t	c/a																
BL3	<b>C/C</b>	<b>G/T</b>	<b>C/T</b>	<b>C/C</b>	<b>C/C</b>	<b>G/G</b>	<b>T/T</b>		<b>G/G</b>		<b>G/G</b>	<b>C/C</b>	<b>G/G</b>	<b>T/C</b>			<b>G/G</b>	<b>A/A</b>	<b>T/T</b>	<b>C/A</b>	xy															
BL4	<b>C/C</b>	<b>G/G</b>	c/c	<b>C/T</b>	<b>C/C</b>	g/g	t/t	<b>A/A</b>	g/g		<b>G/G</b>	<b>C/T</b>	<b>G/A</b>	<b>T/C</b>	g/g		<b>G/G</b>	<b>A/A</b>		<b>C/A</b>	xy															
BL5			t/t					a/a			<b>A/G</b>		g/a	t/t			<b>G/G</b>	<b>A/A</b>		c/c																
BL6	c/t		t/t	<b>C/T</b>	<b>C/C</b>			<b>A/A</b>	g/g		<b>A/G</b>	<b>C/C</b>	<b>G/G</b>	<b>T/C</b>	g/g		c/g	a/a	t/t		xy															

Marker	ABCB1a	ABCB1b	ABCB1c	ABCC11	ADH1Ba	ADH1Bb	CASP12	CYP3A4	CYP3A5	EDAR	HERC2	LCTa	LCTb	NAT2	SLC12A3	SLC24A5	SLC45A2	TRPV6b	TRPV6c	TYR	AMELXY
Probe	C/T	G/T	A/C	T/C	C/T	G/A	C/T	G/A	A/G	T/C	A/G	C/T	G/A	T/C	G/C	G/A	C/G	G/A	C/T	C/A	X/Y
U1_2	c/c	g/t		t/t							a/a	C/C	G/G	T/T				a/a	t/t	a/a	xx
U2_2	t/t	T/T	T/T	C/C	C/C			a/a	G/G		A/A	C/C	G/G	T/T			c/g	A/A	T/T	C/C	
U4_1	C/C	G/G	T/T	C/C	c/c				g/g		a/g	C/C	G/G	T/T			c/g		c/t	C/A	
B1_1		G/T	C/C	T/T					g/g	c/c	a/a	C/C	g/g	C/C	g/g		g/g	A/A		C/C	
J12_8	C/C	G/A	C/C	C/T	C/C	G/G	T/T	A/A	G/G	t/t	A/A	C/C	G/G	T/C	G/G	A/A	c/g	A/A	T/T	C/C	XX
J22_1	C/T	G/T	C/T	C/T	C/C	g/g	T/T	A/A	G/G	c/c	A/G	C/C	G/G	T/C	g/g	A/A	C/G	A/A	T/T	C/C	xy
Ala2	C/C		C/T	C/T	c/c		t/t			t/c	A/G	C/C	g/g	T/T	g/g		g/g	A/A	t/t	c/a	XX
BS1	C/C		C/T	C/T	C/C		t/t		g/g		A/A	c/t	G/G	T/T	g/g	a/a	C/G	A/A		C/A	XY
Ak1_1	T/T	T/T	T/T	C/T	C/C	g/g	t/t	A/A	A/G	t/t	A/G	C/C	G/G	T/T	G/G	a/a	C/C	A/A	t/t	C/C	xy
K1	C/C	G/G	C/C		C/C				G/G		A/A	C/C	g/a	T/T			C/G			C/C	
Mo1	c/t	G/T	C/C	C/C	c/c							C/C	G/G	T/T			G/G	A/A		c/c	xx
VK1_1	T/T	G/T	C/T	C/T	C/C						a/a	C/T	G/A	T/T			C/C	A/A	t/t	C/C	XX
Be2	c/t	g/t	C/T	C/T	c/c					t/c	A/A	C/C	g/g	T/T	g/g	g/a		A/A		c/c	
Be4	c/c	G/G	C/C	C/T	C/C						A/G		G/G	T/C				a/a		C/C	
Be6	c/c		T/T	c/c							A/A	C/C	g/g	c/c	g/g			a/a		c/c	
Be8									g/g		A/A		g/g	t/t			c/c	A/A		a/a	
Be9	C/T	g/t	T/T	T/T					g/g	t/c	A/G	c/c	G/G	c/c	g/g	a/a	c/c	a/a		c/a	xx
Be11			T/T	t/t					G/G		a/g	c/c		T/C							
Be12	C/T	g/t	T/T	C/T	C/C				g/g	t/t	A/G	c/c	g/g	T/C		a/a	c/g	a/a		c/c	
Be14	C/C	G/G	T/T	C/C	C/C			A/A	A/G		A/A	C/C	G/G	C/C	g/g		C/G	A/A	T/T	C/C	xy
Is1	C/C	g/g	C/C	c/c	C/C			A/A	G/G		A/G	C/C	G/G	T/T	g/g	A/A	C/G	A/A	T/T	C/C	xy
Is2	C/T		C/C	c/c	C/C	g/g		A/A	a/g		A/G	C/C	G/A	T/T			g/g	A/A	T/T	C/C	xy
Is4	C/C	G/G	C/C	C/C	C/C	G/G	t/t	a/a	g/g		A/G	C/C	G/G	T/C	g/g		G/G	A/A	T/T	C/C	xy
Ze2	T/T	g/t	c/t	c/t	c/c		t/t		g/g		A/A	C/C	g/g	T/T			c/g	a/a	T/T	c/c	
Ze3		G/G		C/C	C/C				G/G		G/G	C/C		T/C	g/g		G/G	A/A	t/t	C/C	xy
Ze4	c/t			c/t	c/c						A/A	C/C	G/G		g/g		c/c	A/A	T/T	C/C	
Ze5	C/T	G/G	c/t	T/T					G/G		A/A		G/G	T/T		g/a	C/C	A/A	t/t	C/C	xy
Ze6	c/t	g/t	c/c	c/t	c/t			a/a	G/G		A/G	c/c	G/G	T/C			c/g	a/a	t/t	c/a	xx
Ze7	C/T	g/t	C/T	c/t	c/c						a/g	C/T	G/A	T/C		g/a	C/C	A/A	t/t	c/a	XX
Ze8	c/t	g/g	C/C	c/t	c/c				g/g		A/G	C/C	g/g	T/T				A/A		c/c	XX
Ze9		G/T		C/C	c/c						a/g	c/c	G/G	T/T							xx
Kol1		G/T	C/T	C/C	C/C				G/G		a/g	C/C	G/G				G/G	A/A	t/t	c/a	XX
Kol4			C/T		c/c			a/a	G/G		A/G	C/C		T/T			G/G	a/a		c/c	
Kol5		g/g	T/T	c/c							G/G			t/t			g/g	a/a			
Nov5	C/T	G/T		C/C	C/C			a/a	G/G		a/g		G/G			A/A	G/G	A/A	T/T	C/C	XX
Nov7	c/t	G/T	C/T	C/C	C/C				G/G		A/G	C/C	G/G	T/C		A/A	G/G	a/a	c/t	C/C	xy
Nov8		t/t		c/c								C/C		t/c			g/g				

## B7 Diskussion

### B7.1 DNA-Erhalt

Bei 4,5% der Proben konnte keine DNA nachgewiesen werden und bei weiteren 7,2% war der DNA-Erhalt zu schlecht um ein vollständiges HVR1-Profil für die populationsgenetischen Untersuchungen zu erhalten. Für die einzelnen Regionen zeigen sich erhebliche Unterschiede im DNA-Erhalt. Von 81 untersuchten Proben aus dem Osten der Eurasischen Steppe konnten 78 (96,3%) erfolgreich analysiert werden. Von den drei Proben aus Südkasachstan konnte keine für die Auswertung verwendet werden und aus West-Russland waren es 19 von 27 (70,4%). Die Ergebnisse sprechen jedoch für einen insgesamt guten und für die Proben aus dem östlichen Bereich der Eurasischen Steppe hervorragenden DNA-Erhalt. Der gute Erhaltungszustand des Probenmaterials spiegelt sich auch im Amplifikationserfolg von insgesamt 96% wider. Aufgeteilt nach Region zeigt der Osten einen Amplifikationserfolg von 97,7%, der Süden 14,3% und der Westen 92,4%. Im Vergleich dazu konnten bei einer Studie von 150 eneolithischen und bronzezeitlichen Proben aus dem Nordpontikum lediglich 42% erfolgreich analysiert werden (WILDE *ET AL.* 2014 *IN PRESS*). Weitere aDNA Studien verschiedener Epochen und Klimazonen zeigen einen DNA-Erhalt von 42% für frühneolithische Individuen aus Europa (HAAK *ET AL.* 2005), 55,6% für Etrusker aus Italien, die in das 1. Jahrtausend v. Chr. datieren (GHIROTTI *ET AL.* 2013), 73,45% für Individuen der Trichterbecher-Kultur aus Norddeutschland (RUDOLPH 2007) und 80% für Proben aus Peru, die auf 650–1400 Jahre nach Christus datieren (FEHREN-SCHMITZ *ET AL.* 2011). Die Lagerung des Knochen- und Zahnmaterials nimmt entscheidenden Einfluss auf den Erhaltungszustand, wobei eine konstant niedrige Temperatur den DNA-Zerfall verlangsamt und den Befall durch Mikroorganismen reduziert (BURGER *ET AL.* 1999). Die meisten Proben stammen aus dem Bereich des Altai-Gebirges und wurden aus Permafrostböden geborgen. Die frostkonservierten Gräber bieten ideale Bedingungen für den Erhalt biologischen Materials. Von den Proben, die keine nachweisbare Menge an DNA enthielten wurden drei in West-Russland und zwei in Südkasachstan geborgen. Fünf der acht Proben, für die zwar DNA nachgewiesen werden konnte, aber nicht in ausreichender Menge, stammten ebenfalls aus Südkasachstan bzw. West-Russland. Es spielen allerdings nicht nur die klimatischen Bedingungen des Fundplatzes eine entscheidende Rolle, sondern auch die Probenlagerung nach der Ausgrabung. Da die Proben aus Südkasachstan gemeinsam mit den Proben aus

Ostkasachstan aufbewahrt wurde, die alle erfolgreich analysiert werden konnten, kann der schlechtere Erhalt des südkasachischen Probenmaterial eindeutig auf die klimatischen Verhältnisse in dieser Region bezogen werden. Über die Lagerungsbedingungen der Proben aus dem Westen der Eurasischen Steppe liegen leider keine Informationen vor.

## B7.2 Kontaminationen

Kontaminationen lassen sich bei der Arbeit im DNA-Spurenbereich nicht vollständig vermeiden. Bereits geringe Mengen an Fremd-DNA in Chemikalien, Reaktionsgefäßen oder der Probe selbst können die Ergebnisse paläogenetischer Studien kompromittieren. Bei jedem Arbeitsschritt wurde deshalb darauf geachtet Kontaminationen zu entfernen, zu vermeiden oder gegebenenfalls zu detektieren. In 2,3% der mitgeführten Kontrollen konnte DNA nachgewiesen werden. Um festzustellen, ob es sich dabei um sporadische Kontaminationen oder systematische handelt, die auch die Probensequenz beeinflussen können, wurden die Kontaminationen analysiert und mit den Ergebnissen der Proben verglichen, die in der gleichen Reaktion, oder im gleichen Arbeitsschritt behandelt wurden. In einigen Fällen zeigte eine Kontamination die gleiche Sequenzabfolge wie eine der Proben. Am häufigsten ist dies der Fall, wenn weder Probensequenz noch Kontamination einen Basenaustausch aufweisen und das untersuchte Fragment somit der rCRS entspricht. Da stets nur ein Fragment der mitochondrialen DNA betrachtet wird ist dieser Umstand nicht weiter verwunderlich, denn gleich welchen Haplotyp eine Probe trägt, abgesehen von den variablen Positionen entsprechen die restlichen Abschnitte der Referenzsequenz. Durch die Wiederholung der Amplifikationen konnte in all diesen Fällen der Haplotyp bestätigt werden, was bedeutet, dass die Kontamination keinen Einfluss auf die Proben hatte. In den Fällen, wo die kontaminierende Sequenz Mutationen zeigte, konnte in keiner der mitgeführten Proben die gleiche Mutation gefunden werden, bzw. wenn die Mutation auch in einer der Probe gefunden wurde, dann stets in Kombination mit weiteren Basenaustauschen, die im Kontrollfragment nicht zu finden waren oder umgekehrt. Sporadische PCR-Kontaminationen können auch nur die Proben betreffen, ohne in den Leerkontrollen detektiert werden zu können. Die Kontamination einer Probe würde sich bei der Sanger-Sequenzierung als Doppelpeak im Chromatogramm bemerkbar machen, bei der 454- und der Illumina-Sequenzierung in der

Präsenz zwei oder auch mehrerer Sequenz-Populationen. Im Fall einer sporadischen Kontamination kann eine fehlerhafte Auswertung durch die Wiederholung des Versuches ausgeschlossen werden. Eine Kontamination der Probe selbst würde sich ebenfalls durch unterschiedliche Sequenzen für einen DNA-Abschnitt zeigen, wäre jedoch auch bei Wiederholung des Versuchs konsistent, was bei einer sporadischen Kontamination nicht der Fall wäre. Bei der Analyse alter DNA kann ein Doppelpack an einer Basenposition auch auf Deaminierungen hinweisen, die durch den Abbauprozess der DNA nach dem Tod der Zelle entstehen. Im Fall einer Deaminierung handelt es sich jedoch zum einen immer um einen Austausch von Cytosin zu Thymin, oder Guanin zu Adenin und zum anderen sollte sich der Austausch nicht reproduzieren lassen.

Wie oben erwähnt zeigt sich die Kontamination einer Probe darin, dass sich bei der Sequenzanalyse zwei überlagerte Haplotypen reproduzieren lassen. Ein Individuum des Fundplatzes Pokrovka zeigt dieses Phänomen in Form einer reproduzierbaren Signalüberlagerung an Position 16248. Die Referenzsequenz zeigt an dieser Position ein C, das Individuum Pr4 zeigt in Sequenzen aus zwei unabhängigen Extraktionen jeweils C und T. Alle weiteren Mutationen der HVR1 (16182C, 16183C, 16189C, 16249C) zeigen eindeutige Signale. Es ergeben sich drei Möglichkeiten: 16248C ist die authentische Base und 16248T die Kontamination, 16248T ist die Mutation der Probe und 16248C die kontaminierende Sequenz oder die Probe Pr4 zeigt an dieser Stelle eine Heteroplasmie und beide Varianten sind authentisch. Da nur eine Position inkonsistent ist, müsste im Falle einer Kontamination, diese den gleichen Haplotyp wie Pr4 zeigen, bis auf Position 16248, denn ansonsten müssten auch die anderen variablen Positionen Sequenzüberlagerungen zeigen. Zwei Haplotypen, die nur an Position 16248 unterschiedlich sind existieren laut *HaploGrep* nicht. Der Haplotyp mit der rCRS Variante an Position 16248 wird mit 100%iger Übereinstimmung der Haplogruppe U1a'c zugeordnet. Der gleiche Haplotyp mit 16248T wird ebenfalls der Haplogruppe U1a'c zugeordnet, allerdings wird der Basenaustausch als private Mutation dieser Probe gewertet, die sich sonst in dieser Kombination nicht im Referenzdatensatz findet. Als logische Schlussfolgerung bleibt die Heteroplasmie an dieser Stelle als wahrscheinlichste Erklärung anzunehmen. Die Untersuchung zusätzlicher SNPs bestätigt die Haplogruppe U. Aus diesen Gründen wurde die Probe in alle Analysen mit einbezogen.

Ein weiterer wichtiger Punkt um die Authentizität einer Probe nachzuweisen ist, dass die detektierten Mutationen einen phylogenetischen Sinn ergeben. So zeigte sich, dass ein 454 *run* der Probe D1 kontaminierende Sequenzen enthielt. Die Mutationen des kodierenden

Bereichs des mitochondrialen Genoms zeigten an einigen Stellen gemischte Signale und eine Mutation die auf die Haplogruppe N und ihre Abzweigungen hindeutet. Die HVR1 zeigte ebenfalls gemischte Signale, sodass keine Haplogruppe bestimmt werden konnte. In einem zweiten 454 *run* fielen einige der Fragmente im kodierenden Bereich aus, die HVR1 zeigt jedoch einen Haplotyp, der sich klar der Haplogruppe T1a zuordnen lässt. Durch zusätzliche Sanger Sequenzierungen ließ sich der T1a Haplotyp für die HVR1 reproduzieren und auch die Fragmente der kodierenden Region zeigten hierbei eindeutig Mutationen, die auf die Haplogruppe T hinweisen. Dadurch, dass eine Kombination unterschiedlicher SNPs der HVR1 und der kodierenden Region eine Haplogruppe bestimmt, sind Kontaminationen einzelner Fragmente leicht zu erkennen, oder auch eine Kontamination der Reaktion, durch vorhandene Mischsequenzen.

Im Gegensatz zu mitochondrialer DNA wird bei der Analyse nukleärer Marker jeweils nur eine Mutation untersucht, die in zwei Ausprägungen vorliegen kann, was unweigerlich zu Mischsignalen führt, wodurch die Kontamination einer Probe nur schwer zu detektieren ist. 2,7% der nukleären Kontrollen enthielten DNA, deren Sequenzabfolge zwangsläufig mit der einer oder mehrerer Proben übereinstimmte. Die Genotypen, die als gesichert angegeben wurden, sind durch mindestens drei 454-*runs* bestätigt, wodurch eine Fehlinterpretation aufgrund sporadischer Kontaminationen ausgeschlossen werden kann. Die Kontamination der Probe selber kann nicht vollständig ausgeschlossen werden. Aufgrund der guten Probenerhaltung ist es jedoch unwahrscheinlich, dass sich die Kontamination gegen den authentischen Genotyp durchsetzt.

Zur Analyse der Sequenzdaten aus den 454-Sequenzierungen müssen diese vorab nach den probenspezifischen TAGs sortiert werden. Dabei hat sich gezeigt, dass sich bei der Sortierung der Daten nach vorheriger Reduktion der Homopolymere einige TAGs finden, die nicht verwendet wurden. Dies könnte auf eine Kreuzkontamination zwischen den TAGs hinweisen. Sortiert man die Daten jedoch ohne die Homopolymere zu reduzieren verschwinden in den meisten Datensätzen alle überschüssigen TAGs. Ein Vergleich zwischen den reduzierten Sequenzen der TAGs, Primer und amplifizierten Regionen hat gezeigt, dass sich die kurzen TAG-Sequenzen in den unterschiedlichsten Sequenzabschnitten wiederfinden. So beinhaltet die Sequenz des mitochondrialen Primers 11743L die komplette Basenabfolge von TAG 15. TAG 41 findet sich in der Zusammensetzung aus dem Ende von TAG 4 und dem Anfang der mitochondrialen Primer 6336U und 16413L und der nukleären Primer S4L und S15U. TAG 59 findet sich in jedem run, da er sich aus dem Ende des Primers 14839L und dem

Anfang des sich anschließenden DNA Fragmentes zusammensetzt. Durch die Reduzierung der Homopolymerregionen entstehen Sequenzähnlichkeiten, die in den normalen Fragmenten nicht vorhanden sind, weshalb bei der normalen Sortierung auch kaum zusätzliche TAGs gefunden werden. Ausnahmen sind *run 9* und *run 12*. Bei diesen *runs* wurden je zwei Probenpools gleichzeitig sequenziert: in *run 9* Amplikon 9 und 10 und in *run 12* Amplikon 14 und 15. Um dies zu ermöglichen wurde jeder Probenpool von der Sequenzierfirma mit einer zusätzlichen Erkennungssequenz versehen um die Sequenzen am Ende wieder dem jeweiligen Amplikon-Pool zuordnen zu können. Es besteht die Möglichkeit, dass es dabei zu einer Kreuzkontamination zwischen den Erkennungssequenzen kam, oder ebenfalls ein Zuordnungsproblem besteht.

Amplikon 7 wurde komplett von den Analysen ausgeschlossen. Für diesen Probenpool wurden alle TAGs außer TAG 25 verwendet. Bei der Sortierung mit Reduzierung der Homopolymere wurden die TAGs 01–41, 54 und 59 gefunden, wobei die Ordner ab TAG 37 nur nukleäre Marker enthielten. Bei der unreduzierten Sortierung wurden nur den ersten 40 TAGs Sequenzen zugeordnet, wobei hier bereits ab TAG 36 keine mitochondrialen Marker mehr gefunden werden. TAG 25 wurde nach beiden Sortiermethoden gefunden, obwohl er nicht verwendet wurde. Durch vorhergehende Sequenzierungen wurde zusätzlich festgestellt, dass die Sequenzen nicht zu den Proben passen, denen sie laut TAG zugeordnet werden müssten. Ein Fehler bei der Sequenzierung ist nahezu auszuschließen, da die Sequenzverteilung in diesem Fall zufälliger alle TAGs betreffen müsste. Der Ausfall einiger TAGs oder Proben bietet keine Erklärung für die falschen Mutationen in den Proben. Kontaminationen treten eher sporadisch auf und sollten sie einen ganzen run betreffen würde es sich wahrscheinlich um eine Kontaminationsquelle handeln deren Sequenz in jeder Probe auftauchen müsste. Konsistent falsche Haplotypen würden Sinn ergeben, wenn alle Proben vertauscht worden wären, was jedoch nicht das Fehlen von 19 TAGs erklären würde. Eine Möglichkeit wäre, dass ein Fehler bei der Aliquotierung der TAGs dazu geführt hat, dass TAGs falsch beschriftet und somit versehentlich zwei Proben mit dem gleichen TAG markiert wurden. In diesem Fall müsste jede Probe dieses *runs* in sich Inkonsistenzen aufweisen, Unstimmigkeiten zeigen sich jedoch erst im Vergleich mit den Ergebnissen aus anderen Sequenzierungen. Als Folge dieses ungeklärten Phänomens wurden die Proben dieses *runs* nicht in die Auswertungen einbezogen.

Die Ergebnisse der Illumina-Sequenzierung nach vorherigem Capture legen eine Kreuzkontamination zwischen den Proben nahe. So wurden in den Proben A3, J12\_9 und



Fra6 Mutationen detektiert, die in diesen nicht vorkommen, aber in den Proben Be9 und Bt1, der gleichen Sequenzierung. Mögliche Ursachen für dieses Phänomen sind Kreuzkontaminationen der Indexe, Kreuzkontaminationen der Proben oder die falsche Zuordnung von Index und Probensequenz nach der Sequenzierung. Um die Indexe auf Kreuzkontaminationen zu prüfen, wurde mit dem Primer IS6, der vor der Indexsequenz liegt, in die *Libraries* sequenziert, bevor diese *gepoolt* wurden. Im Falle einer Kreuzkontamination zwischen den Indexen müssten diese Mischsignale aufweisen. Die Sequenzen waren jedoch eindeutig und entsprachen der jeweiligen Basenabfolge der Indexsequenz. Eine hypothetische Kreuzkontamination zwischen den Probenextrakten wäre vor oder während der *Library*-Erstellung am wahrscheinlichsten, da die Proben zu diesem Zeitpunkt noch keinen individuellen Barcode tragen. Die Kreuzkontaminationen wurden allerdings sowohl zwischen den skythischen Proben detektiert, die zusammen in eine *sequencing library* überführt und gleichzeitig *gecaptured* wurden, als auch zwischen einer skythischen (Be9) und einer griechischen Probe (Fra6), die gemeinsam sequenziert, aber getrennt *gecaptured* wurden. Eine Kreuzkontamination zwischen den Proben, bevor diese mit einem spezifischen Index markiert wurden kann nicht völlig ausgeschlossen werden, erklärt jedoch nicht die Kreuzkontamination zwischen den Proben Be9 und Fra6. Da eine Kreuzkontamination zwischen den Indexen bereits ausgeschlossen wurde, bleibt nur die fehlerhafte Zuordnung der Sequenzen zu den Indexen. Mittlerweile konnte gezeigt werden, dass die falsche Zuordnung von Sequenzen zu den entsprechenden Indexen durch die Bildung von Clustern auf der *flowcell* entsteht, die ein Sequenzgemisch statt einer Sequenzkolonie enthalten (KIRCHER ET AL. 2012).

### B7.3 Auswahl der mitochondrialen Marker

Die Fundplätze der untersuchten Individuen erstrecken sich vom Schwarzen Meer über Kasachstan bis zum Altai-Gebirge, Tuva und dem Minusinsker Becken. Sie verteilen sich somit über eine weite Strecke der Eurasischen Steppe, die einen Korridor von Europa nach Ost-Asien bildet. Studien an modernen Populationen haben gezeigt, dass sich im Bereich Zentralasiens sowohl mitochondriale Linien finden, die charakteristisch für europäische Populationen sind als auch solche, die als ost-asiatische Linien bezeichnet werden. Ziel war es demnach die Marker so auszuwählen, dass alle asiatischen und europäischen Haplogruppen zuverlässig bestimmt werden können. Meist wird nur eine einzige Basenveränderung angegeben um eine Haplogruppe zu definieren. Obwohl die Ergebnisse zeigen, dass oft bereits das Vorhandensein, oder die Abwesenheit dieser einen Mutation ausreicht, um eine Haplogruppe zu bestimmen, ist die sichere Zuordnung erst durch die Kombination verschiedener SNPs gewährleistet. So kann ein Basenaustausch von G zu A an Position 13708 auf die Haplogruppe J oder die Subhaplogruppe H18 hindeuten. Erst durch die zusätzliche Betrachtung von Position 7028, die die Haplogruppe H von allen anderen unterscheidet, kann eine eindeutige Zuordnung vorgenommen werden. Dadurch, dass für die Proben jeweils alle SNP-Positionen untersucht wurden, konnten die Basenveränderungen anhand des mitochondrialen Stammbaums bis zur tatsächlichen Haplogruppe nachverfolgt werden. Einige der Haplogruppen, die dem asiatischen Bereich zugeordnet werden, wie z.B. C, D, G und Z, entwickelten sich aus der Haplogruppe M (KONG *ET AL.* 2003) und sollten demzufolge zusätzliche zu ihrer spezifischen Mutation die anzeigende Mutation für die Haplogruppe M tragen.

Durch die alleinige Analyse der HVR1 des mitochondrialen Genoms ist die verlässliche Haplogruppenzuordnung nicht immer gewährleistet. Erst durch die zusätzliche Betrachtung der kodierenden Region wird die Einteilung bestätigt, oder gar erst möglich (VAN OVEN & KAYSER 2009). Um die Relevanz zusätzlicher Marker für den untersuchten Datensatz zu demonstrieren wurde eine parallele Haplogruppenbestimmung nach Basenveränderungen der HVR1 und anhand HVR1 und kodierender Region vorgenommen. Die Analysen zeigen, wie unterschiedlich die Ergebnisse mitunter ausfallen können. So wird der Probe J12\_3 nach Analyse der HVR1 die Haplogruppe K1b1a2 zugeordnet, die Betrachtung aller verfügbarer SNPs führt dagegen zur Haplogruppe U7. Die HVR1 Mutationen der Probe Nov7 weisen auf eine U-Linie hin, die kodierende Region zeigt eindeutig eine H-Linie. Dadurch, dass eine

beschränkte Anzahl an Mutationen untersucht wurde, konnten auch unter Einbeziehung der Mutationen der kodierenden Region manche Haplogruppen nicht eindeutig zugeordnet werden. Meistens handelt es sich dabei lediglich um die Unterscheidung zwischen zwei Subhaplogruppen, wie D4h1 oder D4j8 im Fall der Probe Is1. Durch die Betrachtung der HVR1 ist es teilweise nicht möglich eine Haupthaplogruppe zuzuordnen. Die Proben A3 und Be2 zeigen die gleichen SNPs in der HVR1, für die laut *HaploGrep* die Haplogruppen M6, M9, M13'46'61+16362, M33+16362, G und D gleich wahrscheinlich sind. Durch die Einbeziehung der Basensubstitutionen der kodierenden Region lässt sich der Probe A3 eindeutig die Haplogruppe G und der Probe Be2 die Haplogruppe D4 zuordnen. Betrachtet man im Gegenzug nur die Mutationen der kodierenden Region, erlauben diese zwar eine zuverlässige Zuordnung zu den Haupthaplogruppen, in vielen Fällen jedoch keine weitere Untergliederung in entsprechende Subhaplogruppen. So wird die Probe U1\_1 zwar eindeutig der Haplogruppe W zugeordnet, erst die zusätzliche Betrachtung der HVR1 erlaubt die genauere Bestimmung der mitochondrialen Linie W4a.

Die Ergebnisse dieser Arbeit zeigen, dass die ausgewählten Marker eine erfolgreiche und zuverlässige Zuordnung zu den Haupthaplogruppen ermöglichen, die unter Einbeziehung der HVR1 zusätzlich in entsprechende Sub-Haplogruppen unterteilt werden konnten. Einzige Ausnahme stellt das Fragment um die Position 10034 dar, das in den meisten 454 Sequenzierungen ausfiel. Die wahrscheinlichste Ursache ist der Ausfall des Primerpaares, das zur Amplifikation dieses speziellen *locus* entworfen wurde, obwohl der Primer-Test in einer einfachen PCR positiv verlief. Die T → C Mutation an Stelle 10034 ist anzeigend für die Haplogruppe I, die jedoch auch durch eine A → T Transversion an Position 4529 zugeordnet werden kann (HERRNSTADT *ET AL.* 2002), wodurch der Ausfall des Fragments keine Folgen hatte.

## B7.4 Anreicherungsverfahren im Vergleich

Die PCR ist eine simple und zuverlässige Methode, gewünschte DNA-Abschnitte zu vervielfältigen. Ihr Nachteil ist die erhebliche Menge an DNA Extrakt, die benötigt wird um eine größere Anzahl unterschiedlicher Fragmente zu untersuchen und die gewonnenen Daten zu reproduzieren. Neben der HVR1, die in sieben überlappenden Fragmenten amplifiziert wurde, sollten zusätzlich 30 Fragmente der kodierenden Region der mtDNA untersucht werden. Aus diesem Grund wurde die Multiplex-PCR Variante gewählt, die es ermöglicht, bei gleicher Einsatzmenge eines DNA-Extrakts, mehrere Fragmente simultan zu amplifizieren. Obwohl die entworfenen Primerpaare auf zwei bis drei Ansätze aufgeteilt werden mussten, da einige Paare überlappende Bereiche amplifizieren und ihr gemeinsamer Einsatz zur Bildung unerwünschter Nebenprodukte geführt hätte, wurde durch diese Methode wesentlich weniger Extrakt verbraucht, als bei einer einzelnen Amplifikation der gewünschten Marker. So wurden im Schnitt für die Standard-PCR und die Multiplex-PCR je 5µL DNA-Extrakt eingesetzt, in einer Multiplex-Reaktion jedoch 12–13 mal mehr Fragmente amplifiziert wie in einer Standard-PCR. Der Amplifikationserfolg der mitochondrialen Multiplex-PCR liegt mit 98,6% etwas höher als der Erfolg der Standard-PCR mit 95,1%. Hierbei muss jedoch berücksichtigt werden, dass die Standard-PCR zum Testen der Proben-Extrakte auf ihren generellen DNA-Gehalt verwendet wurde und nur Proben, die DNA enthielten für die Multiplex-PCR verwendet wurden. Weiterhin wurde der Amplifikationserfolg für die Multiplex-Reaktionen, wie auch für die Standard-PCRs, anhand der DNA-Banden bei der Gelelektrophorese beurteilt. Bei den Multiplex-Reaktionen bedeutet eine sichtbare Bande allerdings nur, dass eines der Fragmente im Ansatz erfolgreich amplifiziert wurde. Der tatsächliche Erfolg lässt sich erst nach der Sequenzierung beurteilen, wodurch er allerdings direkt an die Qualität dieser gekoppelt ist. Vergleicht man also den Sequenziererfolg der mitochondrialen Multiplex-PCR-Fragmente von 79,3% aus der 454-Sequenzierung bzw. 77,5% aus der Sanger-Sequenzierung mit dem Amplifikationserfolg der Standard-PCRs von 95,1%, ist die Amplifikation einzelner Fragmente deutlich erfolgreicher, jedoch nicht effizienter. Die umfangreiche Untersuchung aller haplogruppenrelevanter Marker, wäre mit dem alleinigen Einsatz einfacher PCRs nicht möglich gewesen, da die Extraktmenge nicht ausgereicht hätte, um alle Fragmente wenigstens einmal zu vervielfältigen. Das Gleiche gilt für die Amplifikation der nukleären Marker und vor allem, wenn beides untersucht werden soll, mitochondriale und nukleäre DNA. Der Vorteil an der

nukleären Multiplex-PCR ist, dass alle *loci* gemeinsam in einem Ansatz vervielfältigt werden können, da sich die gesuchten Fragmente nicht überlappen. Es hat sich allerdings gezeigt, dass es zur Bildung unspezifischer Nebenprodukte kommt, obwohl nur spezifische Primer verwendet wurden und die Ansätze so konzipiert waren, dass die Primer weit genug voneinander entfernt an die Target-Region hybridisieren. Dies zeigte sich durch sichtbare DNA-Banden bei der Gelelektrophorese in Leerkontrollen, die keine lesbaren Sequenzen ergaben und in der großen Anzahl nicht alignierter Fragmente nach der 454-Sequenzierung.

Zur DNA-Anreicherung durch *Hybrid Capture* wurde im Zuge dieser Arbeit nur ein Versuch durchgeführt, weshalb weder zu genereller Qualität noch Quantität dieser Methode eine solide Aussage getroffen, sondern lediglich ein erster Eindruck vermittelt werden kann. Im Gegensatz zu rein PCR basierten Methoden beruht das *Capture* – zumindest im ersten Schritt – auf einer relativen Anreicherung gewünschter Fragmente. In einem zweiten Schritt werden die isolierten Fragmente zwar durch eine PCR vervielfältigt, wobei jedoch so wenig Zyklen wie möglich durchlaufen werden, um auf der einen Seite eine ausreichende DNA Konzentration für die Sequenzierung zu erreichen und auf der anderen Seite eine ungleichmäßige Amplifikation der DNA-Fragmente zu vermeiden. Wie bei der Multiplex-PCR ist der Erfolg des *Captures* nur indirekt messbar. Hierfür wurde eine DNA-Quantifizierung mittels Bioanalyser durchgeführt. Der Bioanalyser gibt den DNA-Gehalt aufgetrennt nach Fragmentlänge an. Davon ausgehend, dass alte DNA in fragmentiertem Zustand mit einer durchschnittlichen Länge von 60–200bp vorliegt (NOONAN *ET AL.* 2006, POINAR *ET AL.* 2006, RAMIREZ *ET AL.* 2009) wird die Anzahl an DNA-Sequenzen dieser Länge oft als Anhaltspunkt für den authentischen DNA-Gehalt der Probe angenommen. Problematisch hierbei ist jedoch die Tatsache, dass auch Verunreinigungen durch bakterielle DNA, die durch die Bodenlagerung in die Probe eingebracht wurden, mit der Zeit degradieren und diese DNA somit das gleiche fragmentierte Muster wie die Probe aufweist. Das Gleiche gilt für moderne Kontaminationen, die durch die Dekontaminationsprozesse im Labor, wie die Behandlung mit Natriumhypochlorit, ebenfalls fragmentiert werden (GARCÍA-GARCERÀ *ET AL.* 2011). Eine Übersicht über das Verhältnis endogener DNA zu Kontaminationen unterschiedlichen Ursprungs und die effizienteste Vermeidung einer Detektion kontaminierender statt authentischer DNA Fragmente bietet der Artikel von KIRSANOW & BURGER 2012. Durch die *Capture* Methode soll zwar der exogene DNA-Gehalt verringert werden, die Effizienz dieses Schrittes lässt sich jedoch erst nach der Sequenzierung beurteilen und Konzentrationsangaben können sich neben exogener Kontamination auch auf

unerwünschte endogene DNA-Fragmente beziehen, die durch DNA-DNA Interaktionen mit aus dem Extrakt isoliert werden.

Die Ergebnisse der Illumina-Sequenzierung zeigen, dass trotz der Anreicherung mitochondrialer Fragmente durch das *Capture*, ein großer Anteil der analysierten DNA-Moleküle nicht zur gesuchten Region passt (s. Tab. B 43). Bei den skythischen Proben sind es im Schnitt 33% der ursprünglichen Sequenzpaare mit Duplikaten, die sich an das mitochondriale Genom alignieren lassen. Ohne Duplikate liegt der Anteil erfolgreich alignierter Sequenzen bei 0,09% der *raw read pairs*. Trotz des hohen Anteils redundanter Information war es durch einen *Capture*-Ansatz möglich, für die Probe Be9 ein komplettes und für die restlichen Proben ein fast komplettes mitochondriales Genom zu erhalten. Aufgrund der falschen Sequenzzuordnung bei der Sequenzierung sind die Ergebnisse mit äußerster Vorsicht zu betrachten. Nur durch die zusätzliche Information aus Sanger- und 454-Sequenzierung lässt sich schlussfolgern, dass die Ergebnisse der Proben Be9 und Bt1 authentisch sind. Ein Großteil der Sequenzen, die den Proben A3 und J12\_9 zugeordnet wurden, ist jedoch falsch. Obwohl auch für diese Proben, das mitochondriale Genom nahezu komplett abgedeckt wurde, lassen sich aus diesem Umstand keine direkten Schlussfolgerung über die Qualität des Captures treffen, da die Abdeckung durch falsch zugeordnete Sequenzen erreicht wurde. Würden die kontaminierenden Sequenzen dieser Proben entfernt, wäre wahrscheinlich nur noch ein kleiner Teil des Genoms abgedeckt. Die Probenqualität scheint dabei eine große Rolle zu spielen. Die skythischen Proben stammen alle aus Sibirien und datieren in das 1. Jahrtausend v. Chr. Insgesamt ist bei diesen Proben von einer besseren Erhaltung auszugehen als bei Proben aus dem griechischen Neolithikum. Von den sibirischen Proben zeigt die Probe A3 die schlechteste Erhaltung. Geht man davon aus, dass die Ergebnisse dieser ersten Versuche zumindest für die Proben Be9 und Bt1 repräsentativ sind, ist das *Capture* die Methode der Wahl für gut erhaltene Proben. Der Extrakt-Verbrauch für eine Reaktion ist zwar höher als bei den PCR basierten Methoden, die Ausbeute ist dafür jedoch wesentlich größer. Für weniger gut erhaltene Proben, oder Studien mit einer relativ großen Probenanzahl empfiehlt sich die Voramplifikation eines spezifischen Markersets durch Multiplex-PCR.

## B7.5 Sequenziermethoden und –erfolg

Für die vorliegende Arbeit wurden drei verschiedene Sequenziermethoden eingesetzt, die Sanger-Sequenzierung und zwei NGS-Methoden: die 454- und die Illumina-Sequenzierung. Die Sanger-Sequenzierung ist eine bewährte Methode zur Sequenzanalyse eines eingeschränkten Datensatzes. Die Probenvorbereitung und die Sequenzanalyse konnten im Haus durchgeführt werden, wodurch der Prozess kontrolliert und besonders für kurze aDNA-Fragmente optimiert werden konnte. Die Fehlerrate der Sanger-Sequenzierung ist sehr gering mit einer Genauigkeit pro Base von 99,999% (SHENDURE & JI 2008). Die Qualität der Sequenzen liegt somit höher als bei der 454- und der Illumina-Sequenzierung (KIRCHER & KELSO 2010). Das neueste Gerät zur Sanger-Sequenzierung (3730xl *DNA Analyzer*) erreicht in voll aufgerüstetem Zustand einen Durchsatz von 3.840 Sequenzen pro Tag (s. Tab. B 1), das hier verwendete Modell 3130 *Genetic Analyzer* von Applied Biosystems mit vier Kapillaren 96 Sequenzen. Im Zuge dieser Arbeit sollten 37 mitochondriale und 21 nukleäre Fragmente analysiert werden. Bei alleinigem Einsatz der Sanger-Sequenzierung wäre also die Bearbeitung von 1,7 Proben pro Tag möglich gewesen. Für die neuen Sequenziertechnologien liegt der maximale Durchsatz mit 1Mio. für die 454 Sequenzierung und 1,2 Mrd. Sequenzen pro Tag für die Illumina-Sequenzierung wesentlich höher. Berücksichtigt man den vergleichsweise geringen DNA-Gehalt prähistorischen Probenmaterials und die nun relativ gering erscheinende Menge von 60 Fragmenten pro Probe wäre der Einsatz dieser Methode für einzelne Proben unverhältnismäßig. Der methodische Teil vorliegender Studie beschäftigte sich deshalb mit der Kombination aus Multiplex-PCR, 454-Sequenzierung und der individuellen Probenmarkierung mit der *tagging*-Methode für eine parallele Sequenzierung unterschiedlicher Proben (MEYER *ET AL.* 2007) zur Analyse palaeopopulationsgenetisch relevanter Daten. Der erfolgreiche Einsatz der 454-Sequenzierung zur Analyse von alter DNA wurde bereits vor Beginn dieser Arbeit gezeigt (POINAR *ET AL.* 2006). Bis heute wurde jedoch keine der etablierten NGS-Methoden für palaeopopulationsgenetische Studien eingesetzt. Das erste prähistorische humane mitochondriale Genom wurde durch eine *shotgun* Sequenzierung erhalten, wobei der Anteil mitochondrialer DNA bei 0,2% lag. Einige Stellen dieses mitochondrialen Genoms waren nicht oder nicht ausreichend abgedeckt, weshalb die entsprechenden Fragmente in Multiplex-PCRs nachamplifiziert wurden (GILBERT *ET AL.* 2008). Für großangelegte Populationsstudien ist die Beschränkung auf phylogenetisch relevante Marker und deren Voramplifikation durch eine Multiplex-PCR deshalb sinnvoll.

Für die Analyse nach vorheriger Multiplex-PCR zeigen die Sanger Sequenzierung und die 454-Sequenzierung mit 77,5 bzw. 79,3% etwa den gleichen Erfolg. Allerdings ermöglichte die 454-Sequenzierung die Analyse von 60 Proben mit je 58 Fragmenten (37 mitochondriale und 21 nukleäre) gleichzeitig. Theoretisch wäre auch eine größere Probenanzahl denkbar, im Zuge dieser Arbeit wurden jedoch nur 60 verschiedene TAGs eingesetzt wodurch die maximale Anzahl limitiert wurde. In keinem der durchgeführten *runs* fiel eine Probe komplett aus (ausgenommen *run 7*), es wurden jedoch nicht immer Sequenzergebnisse zu allen untersuchten Markern erhalten. Der Ausfall einzelner Fragmente hängt zum einen mit der Probenqualität zusammen. Dies zeigt sich darin, dass für die meisten 454 *runs* einige Proben fast vollständig abgedeckt sind, wie z.B. bei *run 15* die mitochondrialen Fragmente der Proben J12\_8 und BL6, bei anderen wiederum viele Fragmente fehlen, wie z.B. bei *run 15* Probe A3 aus dem Kurgan Aržan 2, Grab 5, Skelett 2, deren schlechtere Erhalt im Vergleich zu anderen Individuen des gleichen Kurgans bereits im Grabungsbericht erwähnt und mit der speziellen Fundsituation begründet wurde (ČUGUNOV *ET AL.* 2010). In anderen Fällen zeigen die Ergebnisse eines kompletten *runs* eine schlechte Probenabdeckung, wie zum Beispiel *run 16*. Nach der Sortierung konnten zwar 82,6% der reduzierten Sequenzen einem TAG und einem Primer zugeordnet werden, davon ließ sich jedoch lediglich 37,1% an die gesuchten Fragmente alignieren. Bei den restlichen Daten handelte es sich um Dimere, oder Sequenzartefakte aus Primern und TAGs. Auch bei Amplikon 14 und 15 wurden jeweils nur 18,1% bzw. 14,5% der nukleären Fragmente korrekt aligniert, wohingegen 89,9% bzw. 91,5% der mitochondrialen Marker erfolgreich aligniert werden konnten. Die Ursache liegt dem-entsprechend wahrscheinlicher in der Voramplifikation als in der Sequenziertechnik. Dadurch, dass in einem nukleären Multiplex-PCR-Ansatz mehr Fragmente gleichzeitig amplifiziert werden, kommt zu vermehrter Dimer-Bildung zwischen den eingesetzten Primern. Verstärkt wird dieser Effekt durch die geringere Menge nukleärer DNA im Vergleich zum mitochondrialen DNA-Gehalt einer Probe.

Die Hauptfehlerquelle der 454-Sequenzierung liegt in der Detektion der korrekten Basenanzahl, vor allem in Homopolymerregionen, da die Signalstärke nicht immer korrekt in die tatsächliche Basenanzahl umgewandelt wird (QUINLAN *ET AL.* 2008). Da im vorliegenden Fall nur bekannte Basenabschnitte untersucht wurden, konnten die Sequenzierfehler leicht erkannt und von den tatsächlichen Mutationen unterschieden werden. Die Überprüfung auf Kontaminationen während der Sequenziervorbereitung bzw. der Sequenzierung ergab, dass



Probleme eher bei der Auswertung entstehen als bei der Durchführung der Experimente. Dies wurde durch den Vergleich der unterschiedlichen Methoden zur Datensortierung deutlich.

Es wurden lediglich zwei Illumina-Sequenzierungen durchgeführt. Im ersten Versuch sollten die Amplikons aus den Multiplex-PCRs mit Indexen versehen und als Probenpool sequenziert werden – ein direkter Vergleich zu den vorab durchgeführten 454-Sequenzierungen. Es war möglich, die Sequenzen nach Indexen zu sortieren und somit den Proben zuzuordnen, eine Auswertung war aufgrund der extremen Fehlerrate jedoch nicht möglich. Der Grund dafür liegt wahrscheinlich an der Datenjustierung des Illumina Sequenzierers. Die Messungen beim Einbau der ersten Basen werden zur Kalibrierung des Systems verwendet. Liegt eine unausgeglichene Basenverteilung in diesen ersten Zyklen vor, wie z.B. durch eine Erkennungssequenz zu Beginn des *reads*, oder sich wiederholende Primersequenzen, wie bei der Multiplex-PCR, verläuft die Einstellung des Gerätes inkorrekt. Die Folgen dieser Fehleinstellung sind zum einen der Verlust von Sequenzen, da Cluster, die mit gleichen Basen beginnen als redundante Information eingestuft und entfernt werden und zum anderen die Anhäufung von Sequenzierfehlern aufgrund von Problemen bei der Basenzuordnung in Sequenzen (KIRCHER *ET AL.* 2011). Aus diesem Grund wird die Indexdetektion bei Illumina als separater *read* durchgeführt, bei dem die Basenkalibrierung nicht auf den ersten Zyklen basiert.

Die zweite Illumina-Sequenzierung wurde nach vorherigem *Capture* des mitochondrialen Genoms durchgeführt. Es sollten wiederum mehrere Proben parallel sequenziert werden weshalb jede Probe mit einem individuellen Index versehen wurde. Nach der Auswertung der Mutationen wurde deutlich, dass es zu einem erheblichen Sequenzübertrag zwischen den Proben gekommen sein musste, was auf die falsche Zuordnung von Sequenzen zu den entsprechenden Indexen durch die Bildung von Clustern auf der *flowcell* zurückgeführt werden kann, die ein Sequenzgemisch statt einer Sequenzkolonie enthalten (KIRCHER *ET AL.* 2012). Zum Zeitpunkt der Sequenzierung, Ende 2011, war es nicht möglich die erhaltenen Ergebnisse so auszuwerten, dass eine zuverlässige Zuordnung zu der ursprünglichen Probe ermöglicht wird. Erst durch die Entwicklung eines Dekontaminationskriptes (C. Sell, s. BOLLONGINO *ET AL.* 2013 SOM: *removal of sequencing artifacts*) speziell für diesen Problemfall ist es möglich kontaminierende bzw. falsch zugeordnete Sequenzen zu identifizieren und aus dem Datensatz zu entfernen. Durch den Verlust der Sequenzdaten für Index 16 kann diese Methode auf den vorliegenden Datensatz jedoch nicht angewendet werden, da alle Proben miteinander verglichen werden müssen. Zumindest für die skythische

Probe Be9 kann jedoch von einem authentischen Ergebnis ausgegangen werden. Durch die Ergebnisse aus vorhergehenden Sanger- und 454-Sequenzierungen konnten die bekannten Sequenzpositionen geprüft und keine Diskrepanzen gefunden werden. Somit konnte gezeigt werden, dass es prinzipiell möglich ist mit einem *Capture* und anschließender Illumina-Sequenzierung ein vollständiges mitochondriales Genom zu erhalten. Durch die Markierung der Proben mit zwei verschiedenen Indexen (KIRCHER *ET AL.* 2012) und die bioinformatische Dekontamination der Sequenzierergebnisse können die Nachteile der parallelen Illumina-Sequenzierung überwunden werden.

## B7.6 Authentizität der Daten

Um sicherzustellen, dass sich die Schlussfolgerungen dieser Arbeit auf valide Daten stützen, wurden verschiedene Vorsichtsmaßnahmen getroffen. Dazu gehören die Durchführung der Experimente, bis zum Ansatz der DNA-Amplifikation, in einem physikalisch von allen Post-PCR-Bereichen abgetrennten Reinstraumlabor, das Tragen von Schutzkleidung, die Dekontamination von Proben, Verbrauchsmaterialien und Arbeitsflächen, die Mitführung von Kontaminationskontrollen bei allen Versuchsschritten und die Reproduktion der Ergebnisse. Die HVR1-Sequenzen wurden als gesichert anerkannt wenn sie sich aus drei unabhängigen Amplifikationen bestätigen ließen, bzw. in dreiviertel der Sequenzen die gleiche Mutation zu finden war. SNPs der kodierenden Region wurden als authentisch eingestuft, wenn sie phylogenetischen Sinn ergaben und in keinem Kontrast zu den Ergebnissen der HVR1 Sequenzen standen. Durch die Kontaminationskontrollen konnte ausgeschlossen werden, dass sich sporadische Verunreinigungen in den Proben manifestierten. Eine Kontamination der Probe selbst ist nie völlig auszuschließen, aufgrund der guten Erhaltung des eisenzeitlichen Probenmaterials jedoch unwahrscheinlich, dass der endogene Haplotyp der Probe völlig überdeckt wird.

Die Schwierigkeit bei der Beurteilung nukleärer Marker als authentisch liegt in der Tatsache, dass lediglich eine variable Position untersucht wird und die Ursache für Mischsignale eher als Heterozygotie denn als das Resultat einer Kontamination wahrgenommen wird. Ein weiteres Problem bei der Analyse nukleärer Marker ist das sog. *allelic dropout* – ein Phänomen, das zur Detektion homozygoter Allele führt, wo eigentlich Heterozygotie vorliegt. Die Reproduktion der Ergebnisse ist deshalb besonders wichtig, um einerseits Kontaminationen und andererseits Homozygotie aufgrund von *allelic dropout* ausschließen zu können. Je geringer der DNA-Gehalt desto wahrscheinlicher ist es, dass ein Allel nicht detektiert wird (GAGNEUX *ET AL.* 1997). Die skythischen Proben zeichnen sich durch eine sehr gute Erhaltung aus, weshalb Marker, die für eine Probe in mindestens drei Versuchen übereinstimmende Ergebnisse lieferten als authentisch eingestuft wurden. Weniger abgesicherte Ergebnisse werden unter Vorbehalt präsentiert und sollten ohne weitere Reproduktion nicht als endgültig betrachtet werden.

Aufgrund der hervorragenden Erhaltung der Proben, der Dekontaminationsverfahren und Kontaminationskontrollen im Labor und der Reproduktion der Ergebnisse, besteht keinerlei Zweifel an der Authentizität der hier präsentierten Daten.

## B7.7 Erreichung der gesetzten Ziele

Ziel des methodischen Teils dieser Arbeit war es eine effiziente Methode zur Datengenerierung aus prähistorischem Skelettmaterial für populationsgenetische Studien zu etablieren. Dafür wurde die Kombination aus einer parallelen Amplifikation populationsgenetisch relevanter Daten in einer Multiplex-PCR und anschließender 454-Sequenzierung gewählt. Zusätzlich wurde eine Methode getestet, die eine parallele Sequenzanalyse durch individuelle Probenmarkierung erlaubt. Die Auswertung der ausgewählten mitochondrialen Marker erlaubte eine zuverlässige Zuordnung zu den bekannten Haplogruppen. Aufgrund der Tatsache, dass viele Haplogruppen durch mehrere SNPs angezeigt werden und die generelle Zuordnung einem hierarchischen System folgt, indem die Kombination aus mehreren SNPs ausschlaggebend ist, fiel der Komplettausfall eines Markers nicht ins Gewicht. Durch die Reproduktion der Multiplex-PCRs oder einzelner Marker konnten die Proben, trotz sporadischer Ausfälle einzelner SNPs, stets einer Haplogruppe zugeordnet werden.

Obwohl die 454-Sequenzierungen nicht alle gleich erfolgreich verliefen und jeweils einige Marker ausfielen, war der Informationsgewinn wesentlich höher als es mit dem alleinigen Einsatz der Sanger-Sequenzierung möglich gewesen wäre. Ein Vorteil der 454-Sequenzierung, nämlich die Erfassung einzelner Ausgangsmoleküle im Vergleich zur Erfassung von Konsensussequenzen bei der Sanger-Sequenzierung kommt, nach vorheriger Amplifikation durch die Multiplex-PCR, leider kaum zum Tragen, da die sequenzierten Fragmente zum Großteil eben bereits Kopien darstellen. Das nachträgliche Kollabieren gleicher Sequenzen um die eigentliche Abdeckung der Targetregion zu bestimmen ist zwar mit dem Programm CD-HIT-454 prinzipiell möglich, durch fragmentarische Reste von Primer- und TAG-Sequenzen an den Enden der DNA-Fragmente, die nicht geschnitten werden konnten, jedoch praktisch nicht ausreichend. Hier müsste ein gesondertes Skript eingesetzt werden, das entweder die Fragmente erkennt und schneidet, oder das die Zielregion extrahiert und auswertet. Ist es nicht möglich die eigentliche Abdeckung der Zielregion zu bestimmen, müssen die 454-Sequenzierungen konsequenterweise wie einfache Sequenzierungen gewertet und die Versuche dementsprechend wiederholt werden. Doch selbst in diesem Fall bietet die 454-Sequenzierung 37,5x so viel Information wie eine Sanger-Sequenzierung.

Die Ligation probenspezifischer TAG-Sequenzen an DNA-Fragmente ermöglichte die parallele Sequenzierung von bis zu 60 Proben in einer Reaktion. Nach der Sequenzierung gelang es die DNA-Sequenzen nach ihren TAGs zu sortieren und somit den jeweiligen Proben zuzuordnen. Durch die Wiederholung der Versuche und dem Vergleich mit Fragmenten, die durch Sanger-Sequenzierungen erhalten wurden, zeigte sich, dass die Zuordnung korrekt war. Der Nachteil dieser Methode liegt in der Notwendigkeit zusätzlicher Reaktionsschritte, die zum einen eine erhöhte Kontaminationsgefahr der Proben darstellen und zum anderen zu einem Materialverlust durch zusätzliche Aufreinigungsschritte führen. Die Aufreinigungsschritte wären nicht nötig, wenn man bereits für die Amplifikation der Fragmente codierte Primer verwenden würde (BINLADEN *ET AL.* 2007). Man bräuchte jedoch für jeden TAG ein eigenes Primerset, was höhere Kosten und weniger Flexibilität bei der Fragmentauswahl bedeutet, denn statt einem neuen Primerpaar müssten 60 neue Primerpaare, für jeden TAG, angesetzt werden.

Trotz der aufgeführten Nachteile jeder einzelnen Methode überwiegen die Vorteile in Bezug auf Materialersparnis und Informationsgewinn. Im Zuge dieser Arbeit ist es somit gelungen, die neuen Sequenziertechniken auch für Palaeopopulationsgenetische Studien sinnvoll und erfolgreich einzusetzen.

## B8 Zusammenfassung der methodischen Ergebnisse

Ziel dieser Dissertation war es, eine Methode zu finden, die es erlaubt eine größere Anzahl an Markern zu untersuchen, ohne dabei zusätzliches Probenmaterial verwenden zu müssen. Hierzu sollten informative Stellen der kodierenden Region des mitochondrialen Genoms, die Haplogruppen definierende SNPs enthalten, ausgewählt und gemeinsam mit der HVR1 des mitochondrialen Genoms in einem Multiplex-PCR-Ansatz vervielfältigt werden. Die amplifizierten Fragmente sollten anschließend mit probenspezifischen Barcodes versehen werden, um eine parallele Sequenzanalyse auf der Roche GS FLX 454 Plattform zu ermöglichen und dadurch die neuen Sequenzieretechniken für die speziellen Anforderungen und Fragestellungen der Palaeopopulationsgenetik nutzbar zu machen.

In der vorliegenden Studie ist es gelungen, ein Set an mitochondrialen Markern der kodierenden Region B zusammenzustellen, das es ermöglicht die mitochondriale Variabilität europäischer, asiatischer und zentralasiatischer Populationen zu erfassen. Die individuelle Markierung der DNA-Fragmente, durch die *tagging*-Methode nach MEYER *ET AL.* 2007, ermöglichte die parallele Sequenzanalyse von 37 mitochondrialen und 21 (23) nukleären Fragmenten je Probe für bis zu 60 Proben in einer Sequenzierung. Die Anzahl der verwendeten TAGs stellt hierbei den limitierenden Faktor dar und eine höhere Parallelisierung wäre theoretisch möglich. Im Anschluss an die Sequenzierung gelang es die DNA-Fragmente anhand der TAGs den eingesetzten Proben zuzuordnen. Kreuzkontaminationen, die anhand überschüssiger TAGs nach der 454-Sequenzierung detektiert wurden, konnten auf bioinformatische Fehlzuordnungen zurückgeführt werden. Den gleichen Verfahrensansatz – *getaggte* Multiplex-PCR-Ansätze – mit dem Illumina HiSeq Gerät zu sequenzieren scheiterte, denn aufgrund der großen Anzahl an Molekülen, die nach der Amplifikation mit der gleichen Basensequenz starten, kommt es zu einer fehlerhaften Basenkalibrierung des Illumina-Gerätes. Die Sequenzen aus diesem Lauf enthielten zu viele Basensubstitutionen, die auf eine falsche Zuordnung durch die Analyse-Software zurückzuführen sind, um sie sinnvoll auswerten zu können. Eine zweite Illumina-Sequenzierung wurde nach vorheriger DNA-Anreicherung durch *Hybrid Capture* durchgeführt. Der individuelle Probenbarcode, oder auch Index, wurde hier in einem separaten *Index-read* erfasst, was die Kalibrierung für den *Base call* nicht beeinflusst und somit lesbare Sequenzen ergab. Die Sequenzen wurden jedoch aufgrund falscher Zuordnung von Probensequenz zu Index nicht in die Analysen mit einbezogen.

---

## Referenzen

454 454 / GS FLX. <http://454.com/products/gs-flx-system/>

[Zugriff: Feb. 2014]

Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R. *et al.* (2004). *The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool.* American Journal of Human Genetics 75, 910-8.

Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J.J., Mayer, P. & Kawashima, E. (2000). *Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms.* Nucleic Acids Res 28, E87.

Akey, J.M., Swanson, W.J., Madeoy, J., Eberle, M. & Shriver, M.D. (2006). *TRPV6 exhibits unusual patterns of polymorphism and divergence in worldwide populations.* Hum Mol Genet 15, 2106-13.

Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X. *et al.* (2007). *Direct selection of human genomic loci by microarray hybridization.* Nat Methods 4, 903-5.

Alekseev, A.Y., Bokovenko, N.A., Boltrik, Y., Chugunov, K.A., Cook, G., Dergachev, V.A. *et al.* (2001). *A chronology of the scythian antiquities of Eurasia based on new archaeological and 14C Data.* Radiocarbon 43, 1085-1107.

Alekseev, A.Y., Bokovenko, N.A., Boltrik, Y., Chugunov, K.A., Cook, G., Dergachev, V.A. *et al.* (2002). *Some problems in the study of the chronology of the ancient nomadic cultures in Eurasia (9<sup>th</sup> - 3<sup>rd</sup> centuries BC).* Geochronometria 21, 143-150.

Allentoft, M., Schuster, S.C., Holdaway, R., Hale, M., Mclay, E., Oskam, C. *et al.* (2009). *Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data.* Biotechniques 46, 195-200.

Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M.H., Coulson, A.R., Drouin, J. *et al.* (1981). *Sequence and organization of the human mitochondrial genome.* Nature 290, 457-465.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. & Howell, N. (1999). *Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA.* Nat Genet 23, 147.

Appliedbiosystems 5500xl / LifeTechnologies / AppliedBiosystems.

<http://www.lifetechnologies.com/order/catalog/product/4460730?CID=search-product>

[Zugriff: Feb. 2014]

Appliedbiosystems 3730xl / LifeTechnologies / AppliedBiosystems  
<http://www.lifetechnologies.com/order/catalog/product/3730XL?CID=search-3730xl>  
[Zugriff: Feb. 2014]

Avila-Arcos, M.C., Cappellini, E., Romero-Navarro, J.A., Wales, N., Moreno-Mayar, J.V., Rasmussen, M. *et al.* (2011). *Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA*. *Sci Rep* 1, 74.

Bashilov, V.A. & Yablonsky, L.T. (2000). *Some Current Problems Concerning the History of Early Iron Age Eurasian Nomadic Societies*. In: *Kurgans, Ritual Sites, and Settlements - Eurasian Bronze and Iron Age*, Davis-Kimball, J.M., E. M.; Koryakova, L.; Yablonsky, T., ed. Archaeopress, Oxford, 9-12.

Behar, D.M., Van Oven, M., Rosset, S., Metspalu, M., Loogvali, E.L., Silva, N.M. *et al.* (2012). *A "Copernican" reassessment of the human mitochondrial DNA tree from its root*. *Am J Hum Genet* 90, 675-84.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G. *et al.* (2008). *Accurate whole human genome sequencing using reversible terminator chemistry*. *Nature* 456, 53-9.

Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A. *et al.* (2004). *Genetic signatures of strong recent positive selection at the lactase gene*. *American Journal of Human Genetics* 74, 1111-1120.

Binladen, J., Gilbert, M.T., Bollback, J.P., Panitz, F., Bendixen, C., Nielsen, R. & Willerslev, E. (2007). *The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing*. *PLoS ONE* 2, e197.

Blow, M.J., Zhang, T., Woyke, T., Speller, C.F., Krivoschapkin, A., Yang, D.Y. *et al.* (2008). *Identification of ancient remains through genomic sequencing*. *Genome Research* 18, 1347-53.

Bollongino, R. (2006). *Die Herkunft der Hausrinder in Europa : eine aDNA-Studie an neolithischen Knochenfunden*. Thesis (doctoral), R. Habelt Universität Mainz., Bonn.

Bollongino, R., Nehlich, O., Richards, M.P., Orschiedt, J., Thomas, M.G., Sell, C. *et al.* (2013). *2000 years of parallel societies in Stone Age Central Europe*. *Science* 342, 479-81.

Borinskaya, S., Kal'ina, N., Marusin, A., Faskhutdinova, G., Morozova, I., Kutuev, I. *et al.* (2009). *Distribution of the alcohol dehydrogenase ADH1B\*47His allele in Eurasia*. *Am J Hum Genet* 84, 89-92; author reply 92-4.

Bouakaze, C., Keyser, C., Crubezy, E., Montagnon, D. & Ludes, B. (2009). *Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis*. *Int J Legal Med* 123, 315-25.



- Brace, S., Palkopoulou, E., Dalen, L., Lister, A.M., Miller, R., Otte, M. *et al.* (2012). *Serial population extinctions in a small mammal indicate Late Pleistocene ecosystem instability*. Proc Natl Acad Sci U S A 109, 20532-6.
- Bradley, B. & Stanford, D. (2004). *The North Atlantic ice-edge corridor: a possible Palaeolithic route to the New World*. World Archaeology 36, 459-478.
- Bramanti, B., Thomas, M.G., Haak, W., Unterländer, M., Jores, P., Tambets, K. *et al.* (2009). *Genetic discontinuity between local hunter-gatherers and central Europe's first farmers*. Science 326, 137-40.
- Brandstätter, A., Parsons, T.J. & Parson, W. (2003). *Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups*. Int J Legal Med 117, 291-8.
- Brandstätter, A., Zimmermann, B., Wagner, J., Gobel, T., Rock, A.W., Salas, A. *et al.* (2008). *Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East*. BMC Evol Biol 8, 191.
- Brandt, G., Haak, W., Adler, C.J., Roth, C., Szecsenyi-Nagy, A., Karimnia, S. *et al.* (2013). *Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity*. Science 342, 257-61.
- Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U. *et al.* (2009a). *Targeted retrieval and analysis of five Neandertal mtDNA genomes*. Science 325, 318-21.
- Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U. & Paabo, S. (2009b). *Primer extension capture: targeted sequence retrieval from heavily degraded DNA sources*. J Vis Exp, 1573.
- Brown, M.D., Hosseini, S.H., Torroni, A., Bandelt, H.J., Allen, J.C., Schurr, T.G. *et al.* (1998). *mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America?* American Journal of Human Genetics 63, 1852-61.
- Burbano, H.A., Hodges, E., Green, R.E., Briggs, A.W., Krause, J., Meyer, M. *et al.* (2010). *Targeted investigation of the Neandertal genome by array-based sequence capture*. Science 328, 723-5.
- Burger, J., Hummel, S., Hermann, B. & Henke, W. (1999). *DNA preservation: a microsatellite-DNA study on ancient skeletal remains*. Electrophoresis 20, 1722-1728.
- Burger, J., Rosendahl, W., Loreille, O., Hemmer, H., Eriksson, T., Gotherstrom, A. *et al.* (2004). *Molecular phylogeny of the extinct cave lion Panthera leo spelaea*. Molecular Phylogenetics and Evolution 30, 841-849.
- Burger, J., Kirchner, M., Bramanti, B., Haak, W. & Thomas, M.G. (2007). *Absence of the lactase-persistence-associated allele in early Neolithic Europeans*. Proc Natl Acad Sci U S A 104, 3736-41.

Cann, R.L., Stoneking, M. & Wilson, A.C. (1987). *Mitochondrial DNA and human evolution*. Nature 325, 31-6.

Clisson, I., Keyser, C., Francfort, H.P., Crubezy, E., Samashev, Z. & Ludes, B. (2002). *Genetic analysis of human remains from a double inhumation in a frozen kurgan in Kazakhstan (Berel site, Early 3rd Century BC)*. Int J Legal Med 116, 304-8.

Clopper, C.J. & Pearson, E.S. (1934). *The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial*. Biometrika 26, 404-413.

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Martinez-Arias, R. *et al.* (1998). *Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations*. American journal of human genetics 63, 1824-38.

Comas, D., Plaza, S., Wells, R.S., Yuldaseva, N., Lao, O., Calafell, F. & Bertranpetit, J. (2004). *Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages*. European Journal of Human Genetics 12, 495–504.

Čugunov, K.V., Parzinger, H. & Nagler, A. (2010). *Der skythenzeitliche Fürstenkurgan Aržan 2 in Tuva*. In. Mainz, Verlag Philipp von Zabern.

dbSNP Database of Single Nucleotide Polymorphisms <http://www.ncbi.nlm.nih.gov/SNP/> [Zugriff: Sept. 2013]

Der Sarkissian, C. (2011). *Mitochondrial DNA in Ancient Human Populations of Europe*. Dissertation, University of Adelaide, Adelaide.

Der Sarkissian, C., Balanovsky, O., Brandt, G., Khartanovich, V., Buzhilova, A., Koshel, S. *et al.* (2013). *Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of north East europe*. PLoS genetics 9, e1003296.

Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M. *et al.* (2007). *Phylogeographic analysis of mitochondrial DNA in northern Asian populations*. Am J Hum Genet 81, 1025-41.

Derenko, M.V., Grzybowski, T., Malyarchuk, B.A., Dambueva, I.K., Denisova, G.A., Czarny, J. *et al.* (2003). *Diversity of Mitochondrial DNA Lineages in South Siberia*. Annals of Human Genetics 67, 391-411.

Devlin, B., Risch, N. & Roeder, K. (1990). *No excess of homozygosity at loci used for DNA fingerprinting*. Science 249, 1416-20.

Donnelly, M.P., Paschou, P., Grigorenko, E., Gurwitz, D., Barta, C., Lu, R.B. *et al.* (2012). *A global view of the OCA2-HERC2 region and pigmentation*. Hum Genet 131, 683-96.

Doran, G.H., Dickel, D.N., Ballinger, W.E., Jr., Agee, O.F., Laipis, P.J. & Hauswirth, W.W. (1986). *Anatomical, cellular and molecular analysis of 8,000-yr-old human brain tissue from the Windover archaeological site*. Nature 323, 803-806.

- Dupanloup, I., Schneider, S. & Excoffier, L. (2002). *A simulated annealing approach to define the genetic structure of populations*. Mol Ecol 11, 2571-2581.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G. *et al.* (2009). *Real-time DNA sequencing from single polymerase molecules*. Science 323, 133-8.
- Enattah, N.S., Sahi, T., Savilahti, E., Terwilliger, J.D., Peltonen, L. & Jarvela, I. (2002). *Identification of a variant associated with adult-type hypolactasia*. Nat.Genet. 30, 233-237.
- Evans, A. (1921). *The Palace of Minos; a comparative account of the successive stages of the early Cretan civilization as illustrated by the discoveries at Knossos*. Macmillan, London,.
- Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992). *Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data*. Genetics 131, 479-91.
- Excoffier, L. & Lischer, H.E. (2010). *Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows*. Mol Ecol Resour 10, 564-7.
- Fehren-Schmitz, L., Warnberg, O., Reindel, M., Seidenberg, V., Tomasto-Cagigao, E., Isla-Cuadrado, J. *et al.* (2011). *Diachronic investigations of mitochondrial and y-chromosomal genetic markers in pre-Columbian andean highlanders from South peru*. Annals of Human Genetics 75, 266-83.
- Finnilä, S., Lehtonen, M.S. & Majamaa, K. (2001). *Phylogenetic network for European mtDNA*. American journal of human genetics 68, 1475-84.
- Friedlaender, J., Schurr, T., Gentz, F., Koki, G., Friedlaender, F., Horvat, G. *et al.* (2005). *Expanding Southwest Pacific mitochondrial haplogroups P and Q*. Mol Biol Evol 22, 1506-17.
- Friedlaender, J.S., Friedlaender, F.R., Hodgson, J.A., Stoltz, M., Koki, G., Horvat, G. *et al.* (2007). *Melanesian mtDNA complexity*. PLoS One 2, e248.
- Fu, Y.X. (1997). *Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection*. Genetics 147, 915-925.
- Fujimoto, A., Ohashi, J., Nishida, N., Miyagawa, T., Morishita, Y., Tsunoda, T. *et al.* (2008). *A replication study confirmed the EDAR gene to be a major contributor to population differentiation regarding head hair thickness in Asia*. Hum Genet 124, 179-85.
- Gagneux, P., Boesch, C. & Woodruff, D.S. (1997). *Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair*. Mol.Ecol. 6, 861-868.
- Gao, S.Z., Yang, Y.D., Xu, Y., Zhang, Q.C., Zhu, H. & Zhou, H. (2007). *Tracing the genetic history of the Chinese people: mitochondrial DNA analysis of aneolithic population from the Lajia site*. Am J Phys Anthropol 133, 1128-36.

García-Garcerà, M., Gigli, E., Sanchez-Quinto, F., Ramirez, O., Calafell, F., Civit, S. & Lalueza-Fox, C. (2011). *Fragmentation of contaminant and endogenous DNA in ancient samples determined by shotgun sequencing; prospects for human palaeogenomics*. PLoS One 6, e24161.

Ghirotto, S., Tassi, F., Fumagalli, E., Colonna, V., Sandionigi, A., Lari, M. *et al.* (2013). *Origins and evolution of the Etruscans' mtDNA*. PLoS One 8, e55519.

Gilbert, M.T., Hansen, A.J., Willerslev, E., Rudbeck, L., Barnes, I., Lynnerup, N. & Cooper, A. (2003). *Characterization of Genetic Miscoding Lesions Caused by Postmortem Damage*. The American Journal of Human Genetics 72, 48-61.

Gilbert, M.T., Tomsho, L.P., Rendulic, S., Packard, M., Drautz, D.I., Sher, A. *et al.* (2007). *Whole-genome shotgun sequencing of mitochondria from ancient hair shafts*. Science 317, 1927-30.

Gilbert, M.T., Kivisild, T., Gronnow, B., Andersen, P.K., Metspalu, E., Reidla, M. *et al.* (2008). *Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland*. Science 320, 1787-1789.

GinkoMaps *GinkoMaps - Freie Digitale Landkarten*. <http://www.ginkgomaps.com>  
[Zugriff: Mai 2012]

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., Leproust, E.M., Brockman, W. *et al.* (2009). *Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing*. Nat Biotechnol 27, 182-9.

González-Ruiz, M., Santos, C., Jordana, X., Simon, M., Lalueza-Fox, C., Gigli, E. *et al.* (2012). *Tracing the origin of the east-west population admixture in the Altai region (Central Asia)*. PloS one 7, e48904.

Graf, J., Hodgson, R. & Van Daal, A. (2005). *Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation*. Hum Mutat 25, 278-84.

Graffelman, J. (2013). *HardyWeinberg: Graphical tests for Hardy-Weinberg equilibrium*. <http://CRAN.R-project.org/package=HardyWeinberg>

Grakov, B.N. (1978). *Die Skythen*.Dt. Verl. der Wiss., Berlin.

Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., Simons, J.F. *et al.* (2006). *Analysis of one million base pairs of Neanderthal DNA*. Nature 444, 330-6.

Green, R.E., Briggs, A.W., Krause, J., Prufer, K., Burbano, H.A., Siebauer, M. *et al.* (2009). *The Neandertal genome and ancient DNA authenticity*. Embo J 28, 2494-502.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M. *et al.* (2010). *A draft sequence of the Neandertal genome*. Science 328, 710-22.

- Grignani, P., Peloso, G., Achilli, A., Turchi, C., Tagliabracci, A., Alu, M. *et al.* (2006). *Subtyping mtDNA haplogroup H by SNaPshot minisequencing and its application in forensic individual identification.* Int J Legal Med 120, 151-6.
- Grjaznov, M.P. (1984). *Der Großkurgan von Arzan in Tuva, Südsibirien.* C. H. Beck, München.
- Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tanzer, M. *et al.* (2005). *Ancient DNA from the first European farmers in 7500-year-old Neolithic sites.* Science 310, 1016-8.
- Hamilton, M.B. (2009). *Population genetics.* Wiley-Blackwell, Chichester, UK ; Hoboken, NJ.
- Hapmapconsortium. (2003). *The International HapMap Project.* Nature 426, 789-96.
- Herodotus. *The history.* University of Chicago Press, Chicago.
- Herrnstadt, C., Elson, J.L., Fahy, E., Preston, G., Turnbull, D.M., Anderson, C. *et al.* (2002). *Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups.* Am J Hum Genet 70, 1152-71.
- Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A. & Wilson, A.C. (1984). *DNA sequences from the quagga, an extinct member of the horse family.* Nature 312, 282-4.
- Hippocrates. *De aere aquis et loci.* Harvard University Press; Heinemann, Cambridge; London, <http://data.perseus.org/catalog/urn:cts:greekLit:tlg0627.tlg002.perseus-eng2>. [Zugriff: 18.02.2014]
- Hofreiter, M., Capelli, C., Krings, M., Waits, L., Conard, N., Munzel, S. *et al.* (2002). *Ancient DNA analyses reveal high mitochondrial DNA sequence diversity and parallel morphological evolution of late pleistocene cave bears.* Molecular Biology and Evolution 19, 1244-50.
- Holden, C. & Mace, R. (1997). *Phylogenetic analysis of the evolution of lactose digestion in adults.* Hum.Biol. 69, 605-628.
- Horai, S., Kondo, R., Nakagawa-Hattori, Y., Hayashi, S., Sonoda, S. & Tajima, K. (1993). *Peopling of the Americas, founded by four major lineages of mitochondrial DNA.* Mol Biol Evol 10, 23-47.
- Hughes, D.A., Tang, K., Strotmann, R., Schoneberg, T., Prenen, J., Nilius, B. & Stoneking, M. (2008). *Parallel selection on TRPV6 in human populations.* PLoS One 3, e1686.
- Hughey, J.R., Paschou, P., Drineas, P., Mastropaolo, D., Lotakis, D.M., Navas, P.A. *et al.* (2013). *A European population in Minoan Bronze Age Crete.* Nat Commun 4, 1861.
- Illumina *Illumina / HiSeq 2500.*  
[http://www.illumina.com/systems/hiseq\\_2500\\_1500/performance\\_specifications.ilmn](http://www.illumina.com/systems/hiseq_2500_1500/performance_specifications.ilmn)  
[Zugriff: Feb. 2014]

Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. (2000). *Mitochondrial genome variation and the origin of modern humans*. Nature 408, 708-13.

Iontorrent PGM / LifeTechnologies / IonTorrent.

<http://www.lifetechnologies.com/order/catalog/product/4462921?CID=search-4462921>  
[Zugriff: Feb. 2014]

Itan, Y., Powell, A., Beaumont, M.A., Burger, J. & Thomas, M.G. (2009). *The origins of lactase persistence in Europe*. PLoS Comput Biol 5, e1000491.

Itan, Y., Jones, B.L., Ingram, C.J., Swallow, D.M. & Thomas, M.G. (2010). *A worldwide correlation of lactase persistence phenotype and genotypes*. BMC.Evol.Biol. 10, 36.

Johnson, P.H., Olson, C.B. & Goodman, M. (1985). *Isolation and characterization of deoxyribonucleic acid from tissue of the woolly mammoth, Mammuthus primigenius*. Comp Biochem Physiol B 81, 1045-51.

Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F. *et al.* (2012). *New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing*. Nat Commun 3, 698.

Keyser, C., Bouakaze, C., Crubezy, E., Nikolaev, V.G., Montagnon, D., Reis, T. & Ludes, B. (2009). *Ancient DNA provides new insights into the history of south Siberian Kurgan people*. Hum Genet 126, 395-410.

Kidd, K.K. (2012). *ALFRED - The Allele Frequency Database* <http://alfred.med.yale.edu>  
[Zugriff: Sept. 2013]

Kimura, R., Yamaguchi, T., Takeda, M., Kondo, O., Toma, T., Haneji, K. *et al.* (2009). *A common variation in EDAR is a genetic determinant of shovel-shaped incisors*. Am J Hum Genet 85, 528-35.

Kircher, M. & Kelso, J. (2010). *High-throughput DNA sequencing--concepts and limitations*. Bioessays 32, 524-36.

Kircher, M., Heyn, P. & Kelso, J. (2011). *Addressing challenges in the production and analysis of illumina sequencing data*. BMC Genomics 12, 382.

Kircher, M., Sawyer, S. & Meyer, M. (2012). *Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform*. Nucleic Acids Research 40, e3.

Kirsanow, K. & Burger, J. (2012). *Ancient human DNA*. Ann Anat 194, 121-32.

Kivisild, T., Tolk, H.V., Parik, J., Wang, Y., Papiha, S.S., Bandelt, H.J. & Villems, R. (2002). *The emerging limbs and twigs of the East Asian mtDNA tree*. Molecular Biology and Evolution 19, 1737-51.

Kloss-Brandstätter, A., Pacher, D., Schonherr, S., Weissensteiner, H., Binna, R., Specht, G. & Kronenberg, F. (2011). *HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups*. Human mutation 32, 25-32.

- Knapp, M. & Hofreiter, M. (2010). *Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives*. *Genes Immun* 1, 227-243.
- Kohl, P.L. (2007). *The making of bronze age Eurasia*. Cambridge University Press, Cambridge, UK ; New York.
- Kong, Q.P., Yao, Y.G., Sun, C., Bandelt, H.J., Zhu, C.L. & Zhang, Y.P. (2003). *Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences*. *American Journal of Human Genetics* 73, 671-6.
- Kong, Q.P., Bandelt, H.J., Sun, C., Yao, Y.G., Salas, A., Achilli, A. *et al.* (2006). *Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations*. *Hum Mol Genet* 15, 2076-86.
- Krause, J., Dear, P.H., Pollack, J.L., Slatkin, M., Spriggs, H., Barnes, I. *et al.* (2006). *Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae*. *Nature* 439, 724-7.
- Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R.E., Burbano, H.A. *et al.* (2007). *The Derived FOXP2 Variant of Modern Humans Was Shared with Neandertals*. *Current Biology* 17, 1908-1912.
- Krause, J., Briggs, A.W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A. & Paabo, S. (2010a). *A complete mtDNA genome of an early modern human from Kostenki, Russia*. *Curr Biol* 20, 231-6.
- Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P. & Paabo, S. (2010b). *The complete mitochondrial DNA genome of an unknown hominin from southern Siberia*. *Nature* 464, 894-7.
- Kreutzer, S. (2011). *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.
- Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M. & Paabo, S. (1997). *Neandertal DNA sequences and the origin of modern humans*. *Cell* 90, 19-30.
- Kuchay, R.A., Anwar, M., Thapa, B.R., Mahmood, A. & Mahmood, S. (2013). *Correlation of G/A -22018 single-nucleotide polymorphism with lactase activity and its usefulness in improving the diagnosis of adult-type hypolactasia among North Indian children*. *Genes Nutr* 8, 145-51.
- Kuehl, P., Zhang, J., Lin, Y., Lamba, J., Assem, M., Schuetz, J. *et al.* (2001). *Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression*. *Nat Genet* 27, 383-91.

- Lacan, M., Keyser, C., Ricaut, F.X., Brucato, N., Duranthon, F., Guilaine, J. *et al.* (2011). *Ancient DNA reveals male diffusion through the Neolithic Mediterranean route*. Proc Natl Acad Sci U S A 108, 9788-91.
- Lalueza-Fox, C., Sampietro, M.L., Gilbert, M.T., Castri, L., Facchini, F., Pettener, D. & Bertranpetit, J. (2004). *Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians*. Proc.R.Soc.Lond B Biol.Sci. 271, 941-947.
- Lamason, R.L., Mohideen, M.A., Mest, J.R., Wong, A.C., Norton, H.L., Aros, M.C. *et al.* (2005). *SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans*. Science 310, 1782-6.
- Leont'ev, N., Parzinger, H. & Nagler, A. (1996). *Die russisch-deutschen Ausgrabungen beim Berg Suchanicha am mittleren Enisej. Vorbericht der Kampagne 1995*. Eurasia Antiqua 2, 175ff.
- Li, C., Li, H., Cui, Y., Xie, C., Cai, D., Li, W. *et al.* (2010). *Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age*. BMC Biol 8, 15.
- Li, H., Gu, S., Cai, X., C Speed, W., J Pakstis, A., Golub, E.I. *et al.* (2008). *Ethnic Related Selection for an ADH Class I Variant within East Asia*. PLoS ONE 3, e1881.
- Maca-Meyer, N., Gonzalez, A.M., Larruga, J.M., Flores, C. & Cabrera, V.M. (2001). *Major genomic mitochondrial lineages delineate early human expansions*. BMC Genetics 2.
- Malashev, V.Y. & Yablonsky, L.T. (2004). *Early Nomads in the Southern Foothills of the Urals Based on Materials from the Pokrovka Burial-Ground*. Ancient Civilizations from Scythia to Siberia 10, 259-292.
- Malyarchuk, B.A., Grzybowski, T., Derenko, M.V., Czarny, J., Wozniak, M. & Miscicka-Sliwka, D. (2002). *Mitochondrial DNA variability in Poles and Russians*. Annals of Human Genetics 66, 261-83.
- Mannucci, A., Sullivan, K.M., Ivanov, P.L. & Gill, P. (1994). *Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin*. Int J Legal Med 106, 190-3.
- Mardis, E.R. (2008). *The impact of next-generation sequencing technology on genetics*. Trends Genet 24, 133-41.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bembien, L.A. *et al.* (2005). *Genome sequencing in microfabricated high-density picolitre reactors*. Nature 437, 376-80.
- Merriwether, D.A., Hodgson, J.A., Friedlaender, F.R., Allaby, R., Cerchio, S., Koki, G. & Friedlaender, J.S. (2005). *Ancient mitochondrial M haplogroups identified in the Southwest Pacific*. Proc Natl Acad Sci U S A 102, 13034-9.



- Metzker, M.L. (2010). *Sequencing technologies - the next generation*. Nat Rev Genet 11, 31-46.
- Meyer, M., Stenzel, U., Myles, S., Prufer, K. & Hofreiter, M. (2007). *Targeted high-throughput sequencing of tagged nucleic acid samples*. Nucleic Acids Research 35, e97.
- Meyer, M., Stenzel, U. & Hofreiter, M. (2008). *Parallel tagged sequencing on the 454 platform*. Nat Protoc 3, 267-78.
- Meyer, M. & Kircher, M. (2010). *Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing*. Cold Spring Harb Protoc 2010, pdb.prot5448-.
- Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S. *et al.* (2012). *A high-coverage genome sequence from an archaic Denisovan individual*. Science 338, 222-6.
- Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J.L. *et al.* (2014). *A mitochondrial genome sequence of a hominin from Sima de los Huesos*. Nature 505, 403-6.
- Miller, W., Drautz, D.I., Ratan, A., Pusey, B., Qi, J., Lesk, A.M. *et al.* (2008). *Sequencing the nuclear genome of the extinct woolly mammoth*. Nature 456, 387-390.
- Mitomap A human mitochondrial genome database. <http://www.mitomap.org>  
[Zugriff: 2012]
- Molodin, V.I., Pilipenko, A.S., Romaschenko, A.G., Zhuravlev, A.A., Trapezov, R.O., Chikisheva, T.A. & Pozdnyakov, D.V. (2012a). *Human migrations in the southern region of the West Siberian Plain during the Bronze Age: Archaeological, palaeogenetic and anthropological data*. In: *Population Dynamics in Prehistory and Early History - New Approaches Using Stable Isotopes and Genetics*, Kaiser, E.B., J.; Schier, W., ed. DE GRUYTER, 93-112.
- Molodin, V.I., Pilipenko, A.S., Zhuravlev, A.A., Trapezov, R.O. & Romashchenko, A.G. (2012b). *An Analysis of Mitochondrial DNA from the Pakhomovskaya Population of the Late Bronze Age, Western Siberia*. Archaeology, Ethnology & Anthropology of Eurasia 40, 62-69.
- Mooder, K.P., Schurr, T.G., Bamforth, F.J., Bazaliiski, V.I. & Savel'ev, N.A. (2006). *Population affinities of Neolithic Siberians: a snapshot from prehistoric Lake Baikal*. Am.J.Phys.Anthropol. 129, 349-361.
- Mordvintseva, V. (2013). *The Sarmatians: The Creation of Archaeological Evidence*. Oxford J Archaeol 32, 203-219.
- Mulcare, C.A., Weale, M.E., Jones, A.L., Connell, B., Zeitlyn, D., Tarekegn, A. *et al.* (2004). *The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans*. Am J Hum Genet 74, 1102 - 1110.
- Mullis, K.B. & Faloona, F.A. (1987). *Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction*. Methods Enzymol 155, 335-50.

- Nagy, D., Tömöry, G., Csányi, B., Bogácsi-Szabó, E., Czibula, Á., Priskin, K. *et al.* (2011). *Comparison of lactase persistence polymorphism in ancient and present-day Hungarian populations*. *American Journal of Physical Anthropology*, n/a-n/a.
- Nakano, M., Komatsu, J., Matsuura, S., Takashima, K., Katsura, S. & Mizuno, A. (2003). *Single-molecule PCR using water-in-oil emulsion*. *J Biotechnol* 102, 117-24.
- Nasab, H.V. (2011). *Reconstruction of early Neolithic/Bronze Age population diversity in the Shamanka II cemetery at Lake Baikal using mtDNA polymorphism* *Progress in Biological Sciences* 1, 29-35.
- Niehs\_Snps *NIEHS Environmental Genome Project* <http://egp.gs.washington.edu>  
[Zugriff: Sept. 2013]
- Nikitin, A.G., Sokhatsky, M.P., Kovalyukh, M.M. & Videiko, M.Y. (2010). *Comprehensive Site Chronology and Ancient Mitochondrial DNA Analysis from Verteba Cave – a Trypillian Culture Site of Eneolithic Ukraine*. *Interdisciplinaria Archaeologica Natural Sciences in Archaeology* 1, 9-18.
- Nikitin, A.G. (2011). *Bioarchaeological Analysis of Bronze Age Human Remains from the Podillya Region of Ukraine*. *Interdisciplinaria Archaeologica Natural Sciences in Archaeology* 2, 9-14.
- Nikitin, A.G., Newton, J.R. & Potekhina, I.D. (2012). *Mitochondrial haplogroup C in ancient mitochondrial DNA from Ukraine extends the presence of East Eurasian genetic lineages in Neolithic Central and Eastern Europe*. *J Hum Genet* 57, 610-2.
- Niu, B., Fu, L., Sun, S. & Li, W. (2010). *Artificial and natural duplicates in pyrosequencing reads of metagenomic data*. *BMC Bioinformatics* 11, 187.
- Noonan, J.P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J. *et al.* (2006). *Sequencing and analysis of Neanderthal genomic DNA*. *Science* 314, 1113-8.
- Norton, H.L., Kittles, R.A., Parra, E., Mckeigue, P., Mao, X., Cheng, K. *et al.* (2007). *Genetic Evidence for the Convergent Evolution of Light Skin in Europeans and East Asians*. *Molecular Biology and Evolution* 24, 710-722.
- Ohashi, J., Naka, I. & Tsuchiya, N. (2011). *The impact of natural selection on an ABCC11 SNP determining earwax type*. *Mol Biol Evol* 28, 849-57.
- Olalde, I., Allentoft, M.E., Sanchez-Quinto, F., Santpere, G., Chiang, C.W., Degiorgio, M. *et al.* (2014). *Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European*. *Nature*.
- Ou, C.Y., Moore, J.L. & Schochetman, G. (1991). *Use of Uv Irradiation to Reduce False Positivity in Polymerase Chain-Reaction*. *Biotechniques* 10, 442-&.
- Pääbo, S. (1989). *Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification*. *Proc Natl Acad Sci U S A* 86, 1939-43.

Pacificbiosciences *PacBio RS II / PacificBiosciences*.  
[http://files.pacb.com/pdf/PacBio\\_RS\\_II\\_Brochure.pdf](http://files.pacb.com/pdf/PacBio_RS_II_Brochure.pdf)  
[Zugriff: Feb. 2014]

Palanichamy, M.G., Sun, C., Agrawal, S., Bandelt, H.J., Kong, Q.P., Khan, F. *et al.* (2004). *Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia*. *American Journal of Human Genetics* 75, 966-78.

Parsons, T.J., Muniec, D.S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M.R. *et al.* (1997). *A high observed substitution rate in the human mitochondrial DNA control region*. *Nat Genet* 15, 363-8.

Parzinger, H. (2004). *Die Skythen*. C. H. Beck, München.

Parzinger, H. (2006). *Die frühen Völker Eurasiens*. C. H. Beck, München.

Peng, Y., Shi, H., Qi, X.-B., Xiao, C.-J., Zhong, H., Ma, R.-L. & Su, B. (2010). *The ADH1B Arg47His polymorphism in East Asian populations and expansion of rice domestication in history*. *BMC Evol Biol* 10, 15.

Pilipenko, A., Romaschenko, A., Molodin, V., Parzinger, H. & Kobzev, V. (2010). *Mitochondrial DNA studies of the Pazyryk people (4th to 3rd centuries BC) from northwestern Mongolia*. *Archaeological and Anthropological Sciences* 2, 231-236.

Pilipenko, A.S., Romashchenko, A.G., Molodin, V.I., Kulikov, I.V., Kobzev, V.F., Pozdnyakov, D.V. & Novikova, O.I. (2008). *Infant Burials in Dwellings at Chicha-1, in the Baraba Forest-Steppe: Results of DNA Analysis*. *Archaeology, Ethnology & Anthropology of Eurasia* 34, 57-67.

Pilipenko, A.S.M., V.I.; Romashchenko, A.G. (2012). *Mitochondrial DNA of a Late Neolithic Woman from Kaminnaia Cave (Gorny Altai)*. *Russian Journal of Genetics: Applied Research* 2, 214-221.

Poinar, H.N., Schwarz, C., Qi, J., Shapiro, B., Macphee, R.D., Buigues, B. *et al.* (2006). *Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA*. *Science* 311, 392-4.

Polosmak, N.V., Kundo, L.P., Balakina, G.G., Mamatyuk, V.I., Vasilyev, V.G., Karpova, E.V. *et al.* (2006). *Textiles from the "frozen" tombs in Gorny Altai 400–300 BC (An Integrated Study)*. In. Novosibirsk, Siberian Branch of the Russian Academy of Sciences.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S. *et al.* (2014). *The complete genome sequence of a Neanderthal from the Altai Mountains*. *Nature* 505, 43-9.

Quinlan, A.R., Stewart, D.A., Stromberg, M.P. & Marth, G.T. (2008). *Pyrobayes: an improved base caller for SNP discovery in pyrosequences*. *Nat Methods* 5, 179-81.

Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R. *et al.* (2004). *Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor*. *American Journal of Human Genetics* 74, 827-45.

R\_Core\_Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Wien. <http://www.R-project.org/>

Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I. *et al.* (2014). *Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans*. *Nature* 505, 87-91.

Rajeevan, H., Soundararajan, U., Kidd, J.R., Pakstis, A.J. & Kidd, K.K. (2012). *ALFRED: an allele frequency resource for research and teaching*. *Nucleic Acids Res* 40, D1010-5.

Ramirez, O., Gigli, E., Bover, P., Alcover, J.A., Bertranpetit, J., Castresana, J. & Lalueza-Fox, C. (2009). *Paleogenomics in a temperate environment: shotgun sequencing from an extinct Mediterranean caprine*. *PLoS One* 4, e5670.

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I. *et al.* (2010). *Ancient human genome sequence of an extinct Palaeo-Eskimo*. *Nature* 463, 757-62.

Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A. *et al.* (2011). *An Aboriginal Australian genome reveals separate human dispersals into Asia*. *Science* 334, 94-8.

Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., Degiorgio, M., Stafford, T.W., Jr. *et al.* (2014). *The genome of a Late Pleistocene human from a Clovis burial site in western Montana*. *Nature* 506, 225-9.

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y. *et al.* (2010). *Genetic history of an archaic hominin group from Denisova Cave in Siberia*. *Nature* 468, 1053-60.

Reynolds, J., Weir, B.S. & Cockerham, C.C. (1983). *Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance*. *Genetics* 105, 767-779.

Ricaut, F.X., Keyser-Tracqui, C., Bourgeois, J., Crubezy, E. & Ludes, B. (2004a). *Genetic analysis of a Scytho-Siberian skeleton and its implications for ancient Central Asian migrations*. *Hum.Biol.* 76, 109-125.

Ricaut, F.X., Keyser-Tracqui, C., Cammaert, L., Crubezy, E. & Ludes, B. (2004b). *Genetic analysis and ethnic affinities from two Scytho-Siberian skeletons*. *Am.J.Phys.Anthropol.* 123, 351-360.

Rodríguez-Robles, J.A., Jezkova, T. & Leal, M. (2010). *Climatic stability and genetic divergence in the tropical insular lizard *Anolis krugi*, the Puerto Rican 'Lagartijo Jardinero de la Montana'*. *Mol Ecol* 19, 1860-76.

Rogers, S.O. & Bendich, A.J. (1985). *Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues*. *Plant Mol Biol* 5, 69-76.

- Rolle, R. (1992). *Die skythenzeitlichen Mumienfunde von Pazyryk. Frostkonservierte Gräber aus dem Altaigebirge*. In: *Der Mann im Eis: Bericht über das Internationale Symposium 1992 in Innsbruck*, Höpfel, F.P., Werner; Spindler, Konrad, ed. Universität Innsbruck, Innsbruck, 334–358.
- Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. & Nyren, P. (1996). *Real-time DNA sequencing using detection of pyrophosphate release*. *Anal Biochem* 242, 84-9.
- Rootsforreal.Com *Genetic ancestry*. [www.rootsforreal.com](http://www.rootsforreal.com)  
[Zugriff: 2011]
- Rosenberg, N.A. & Nordborg, M. (2002). *Genealogical trees, coalescent theory and the analysis of genetic polymorphisms*. *Nat Rev Genet* 3, 380-90.
- Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M. *et al.* (2011). *An integrated semiconductor device enabling non-optical genome sequencing*. *Nature* 475, 348-52.
- Rousset, F. & Raymond, M. (1995). *Testing heterozygote excess and deficiency*. *Genetics* 140, 1413-9.
- Rudolph, P. (2007). *Palaeogenetische Untersuchungen der trichterbecherzeitlichen Skelettpopulation von Ostorf, Stadt Schwerin*. Diplomarbeit, Johannes Gutenberg-Universität, Mainz.
- Sabbagh, A., Darlu, P., Crouau-Roy, B. & Poloni, E.S. (2011). *Arylamine N-acetyltransferase 2 (NAT2) genetic diversity and traditional subsistence: a worldwide population survey*. *PLoS One* 6, e18507.
- Saleh, M., Vaillancourt, J.P., Graham, R.K., Huyck, M., Srinivasula, S.M., Alnemri, E.S. *et al.* (2004). *Differential modulation of endotoxin responsiveness by human caspase-12 polymorphisms*. *Nature* 429, 75-9.
- Samashev, Z. (2002). *Die "goldhütenden Greife" des Herodot und die archäologische Kultur der frühen Nomaden im kazachischen Altai – Skythenzeitliche Kurgane von Berel' und Tar Asu*. *Eurasia Antiqua* 8, 237–277.
- Sánchez-Quinto, F., Schroeder, H., Ramirez, O., Ávila-Arcos, María c., Pybus, M., Olalde, I. *et al.* (2012). *Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers*. *Current Biology*.
- Sanchez, J.J., Borsting, C., Hallenberg, C., Buchard, A., Hernandez, A. & Morling, N. (2003). *Multiplex PCR and minisequencing of SNPs--a model with 35 Y chromosome SNPs*. *Forensic Sci Int* 137, 74-84.
- Sanger, F., Nicklen, S. & Coulson, A.R. (1977). *DNA sequencing with chain-terminating inhibitors*. *Proc Natl Acad Sci U S A* 74, 5463-7.

- Schirmer, M., Toliat, M.R., Haberl, M., Suk, A., Kamdem, L.K., Klein, K. *et al.* (2006). *Genetic signature consistent with selection against the CYP3A4\*1B allele in non-African populations*. *Pharmacogenet Genomics* 16, 59-71.
- Shen, Y.C., Fan, J.H., Edenberg, H.J., Li, T.K., Cui, Y.H., Wang, Y.F. *et al.* (1997). *Polymorphism of ADH and ALDH genes among four ethnic groups in China and effects upon the risk for alcoholism*. *Alcohol Clin Exp Res* 21, 1272-7.
- Shendure, J. & Ji, H. (2008). *Next-generation DNA sequencing*. *Nat Biotechnol* 26, 1135-45.
- Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. & Sirotkin, K. (2001). *dbSNP: the NCBI database of genetic variation*. *Nucleic Acids Res* 29, 308-11.
- Shriver, M.D., Parra, E.J., Dios, S., Bonilla, C., Norton, H., Jovel, C. *et al.* (2003). *Skin pigmentation, biogeographical ancestry and admixture mapping*. *Hum Genet* 112, 387-99.
- Simoons, F.J. (1970). *Primary adult lactose intolerance and the milking habit: a problem of biologic and cultural interrelations*. *Am. J. Dig. Dis.* 15, 695.
- Skoglund, P., Malmstrom, H., Raghavan, M., Stora, J., Hall, P., Willerslev, E. *et al.* (2012). *Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe*. *Science* 336, 466-9.
- Smirnov, K.F. (1980). *Sauromates et Sarmates*. *Dialogues d'histoire ancienne* 6, 139-154.
- Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R. *et al.* (1986). *Fluorescence detection in automated DNA sequence analysis*. *Nature* 321, 674-9.
- Stiller, M., Knapp, M., Stenzel, U., Hofreiter, M. & Meyer, M. (2009). *Direct multiplex sequencing (DMPS)--a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA*. *Genome Research* 19, 1843-8.
- Stokowski, R.P., Pant, P.V., Dadd, T., Fereday, A., Hinds, D.A., Jarman, C. *et al.* (2007). *A genomewide association study of skin pigmentation in a South Asian population*. *Am J Hum Genet* 81, 1119-32.
- Sturm, R.A., Duffy, D.L., Zhao, Z.Z., Leite, F.P., Stark, M.S., Hayward, N.K. *et al.* (2008). *A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color*. *Am J Hum Genet* 82, 424-31.
- Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D.G., Mulligan, C.J. *et al.* (2007). *Beringian standstill and spread of Native American founders*. *PLoS One* 2, e829.
- Tanaka, M., Cabrera, V.M., Gonzalez, A.M., Larruga, J.M., Takeyasu, T., Fuku, N. *et al.* (2004). *Mitochondrial genome variation in eastern Asia and the peopling of Japan*. *Genome Res* 14, 1832-50.
- Tewhey, R., Warner, J.B., Nakano, M., Libby, B., Medkova, M., David, P.H. *et al.* (2009). *Microdroplet-based PCR enrichment for large-scale targeted sequencing*. *Nat Biotechnol* 27, 1025-31.

- The1000genomesconsortium, Abecasis, G.R., Auton, A., Brooks, L.D., Depristo, M.A., Durbin, R.M. *et al.* (2012). *An integrated map of genetic variation from 1,092 human genomes*. Nature 491, 56-65.
- Thompson, E.E., Kuttab-Boulos, H., Witonsky, D., Yang, L., Roe, B.A. & Di Rienzo, A. (2004). *CYP3A variation and the evolution of salt-sensitivity variants*. Am J Hum Genet 75, 1059-69.
- Torrioni, A., Schurr, T.G., Yang, C.C., Szathmary, E.J., Williams, R.C., Schanfield, M.S. *et al.* (1992). *Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations*. Genetics 130, 153-62.
- Torrioni, A., Sukernik, R.I., Schurr, T.G., Starikorskaya, Y.B., Cabell, M.F., Crawford, M.H. *et al.* (1993). *mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans*. American journal of human genetics 53, 591-608.
- Torrioni, A., Lott, M.T., Cabell, M.F., Chen, Y.S., Lavergne, L. & Wallace, D.C. (1994). *mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region*. Am.J.Hum.Genet. 55, 760-776.
- Torrioni, A., Bandelt, H.J., D'urbano, L., Lahermo, P., Moral, P., Sellitto, D. *et al.* (1998). *mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe*. Am.J Hum.Genet. 62, 1137-1152.
- Torrioni, A., Richards, M., Macaulay, V., Forster, P., Villems, R., Norby, S. *et al.* (2000). *mtDNA haplogroups and frequency patterns in Europe*. Am.J.Hum.Genet. 66, 1173-1177.
- Torrioni, A., Achilli, A., Macaulay, V., Richards, M. & Bandelt, H.J. (2006). *Harvesting the fruit of the human mtDNA tree*. Trends in genetics : TIG 22, 339-45.
- Toyoda, Y., Sakurai, A., Mitani, Y., Nakashima, M., Yoshiura, K., Nakagawa, H. *et al.* (2009). *Earwax, osmidrosis, and breast cancer: why does one SNP (538G>A) in the human ABC transporter ABCC11 gene determine earwax type?* Faseb J 23, 2001-13.
- Troy, C.S., Machugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P. *et al.* (2001). *Genetic evidence for Near-Eastern origins of European cattle*. Nature 410, 1088-1091.
- Van Oven, M. & Kayser, M. (2009). *Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation*. Human mutation 30, E386-94.
- Veeramah, K.R. & Hammer, M.F. (2014). *The impact of whole-genome sequencing on the reconstruction of human population history*. Nat Rev Genet 15, 149-62.
- Venables, W.N. & Ripley, B.D. (2002). *Modern Applied Statistics with S*. Springer, Voevoda, M.I., Romaschenko, A.G., Sitnikova, V.V., Shulgina, E.O. & Kobsev, V.F. (2000). *A Comparison of Mitochondrial DNA Polymorphism in Pazyryk and Modern Eurasian Populations*. Archaeology, Ethnology & Anthropology of Eurasia 4, 88-94.



Vormfelde, S.V., Sehrt, D., Toliat, M.R., Schirmer, M., Meineke, I., Tzvetkov, M. *et al.* (2007). *Genetic variation in the renal sodium transporters NKCC2, NCC, and ENaC in relation to the effects of loop diuretic drugs.* Clin Pharmacol Ther 82, 300-9.

Wall, J.D. & Kim, S.K. (2007). *Inconsistencies in Neanderthal Genomic DNA Sequences.* PLoS Genetics 3, 5.

Wall, T.L., Shea, S.H., Luczak, S.E., Cook, T.A. & Carr, L.G. (2005). *Genetic associations of alcohol dehydrogenase with alcohol use disorders and endophenotypes in white college students.* J Abnorm Psychol 114, 456-65.

Wang, H., Ding, K., Zhang, Y., Jin, L., Kullo, I.J. & He, F. (2007). *Comparative and evolutionary pharmacogenetics of ABCB1: complex signatures of positive selection on coding and regulatory regions.* Pharmacogenet Genomics 17, 667-78.

Wilde, S., Timpson, A., Kirsanow, K., Kaiser, C., Keyser, M., Unterländer, M. *et al.* (2014 *in press*). *Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y.* PNAS.

Xue, Y., Daly, A., Yngvadottir, B., Liu, M., Coop, G., Kim, Y. *et al.* (2006). *Spread of an inactive form of caspase-12 in humans is due to recent positive selection.* Am J Hum Genet 78, 659-70.

Yablonsky, L.T. (2000). *"Scythian Triad" and "Scythian World"*. In: *Kurgans, Ritual Sites, and Settlements - Eurasian Bronze and Iron Age*, Davis-Kimball, J., Murphy, E.M., Koryakova, L. & Yablonsky, T., eds. Archaeopress, Oxford, 3-8.

Yoshiura, K., Kinoshita, A., Ishida, T., Ninokata, A., Ishikawa, T., Kaname, T. *et al.* (2006). *A SNP in the ABCC11 gene is the determinant of human earwax type.* Nat Genet 38, 324-30.



# Anhang

## Abbildungsverzeichnis

### Abbildungen aus Teil A

Abb. A 1 Geographische Verteilung der Haplogruppen in modernen Bevölkerungen. ....	20
Abb. A 2 Verteilung der Fundplätze über die Eurasische Steppe .....	33
Abb. A 3 MDS-Plot der skythischen Populationen.....	38
Abb. A 4 MDS-Plot eisenzeitlicher und moderner Populationen der Eurasischen Steppe .....	40
Abb. A 5 MDS-Plot der Pazyryk-Gruppen .....	44
Abb. A 6 Frequenzen der abgeleiteten Allele der Pigmentierungsgene .....	49
Abb. A 7 Frequenzen der abgeleiteten Allele zweier <i>loci</i> , die mit Laktasepersistenz assoziiert sind .....	50
Abb. A 8 Frequenzen der abgeleiteten Allele von Genen, die mit einer Anpassung an Temperatur & Wasservorkommen assoziiert werden .....	51
Abb. A 9 Frequenzen abgeleiteter Allele von Genen, die mit der Umsetzung von Xenobiotika assoziiert sind ..	52
Abb. A 10 Haplogruppenverteilung der skythischen Populationen .....	59
Abb. A 11 Heutige Frequenzen des ADH1B*47His Allels .....	75

### Abbildungen aus Teil B

Abb. B 1 Aufbau eines Meyer-TAGs.....	108
Abb. B 2 dNTP (links) und ddNTP (rechts).....	114
Abb. B 3 Schematische Darstellung einer Pyrosequenzierreaktion. ....	119
Abb. B 4 Illumina <i>Bridge-PCR</i> .....	126
Abb. B 5 Illumina <i>sequencing by synthesis</i> .....	127
Abb. B 6 Ergebnis zur Illumina Amplikon Sequenzierung .....	142
Abb. B 7 Kontaminationsrate.....	145
Abb. B 8 Index Kontrolle durch Sanger-Sequenzierung.....	149

## Tabellenverzeichnis

### Tabellen aus Teil A

Tab. A 1 Probenliste nach Fundplatz und Region.....	16
Tab. A 2 phänotypische Eigenschaften der untersuchten nukleären loci .....	23
Tab. A 3 Haplogruppen.....	28
Tab. A 4 Literaturdaten .....	31
Tab. A 5 Probenanzahl pro Fundplatz nach Kultur inklusive Literaturdaten.....	32
Tab. A 6 Anteile west- und osteurasischer Linien .....	35
Tab. A 7 Haplotypen- und Nukleotiddiversität .....	36
Tab. A 8 $F_{ST}$ -Werte zwischen den skythischen Kulturgruppen.....	38
Tab. A 9 $F_{ST}$ -Werte zwischen skythischen Gruppen und modernen Populationen gleicher Regionen .....	39
Tab. A 10 Ergebnisse der SAMOVA.....	41
Tab. A 11 $F_u$ 's $F_S$ -Werte.....	42
Tab. A 12 Pazyryk-Population nach Region .....	43
Tab. A 13 $F_{ST}$ -Werte zwischen den Pazyryk-Gruppen nach Land .....	44
Tab. A 14 $F_{ST}$ -Werte zwischen den Pazyryk-Gruppen nach geographischer Subregion .....	44
Tab. A 15 $F_{ST}$ -Werte zwischen den skythischen Gruppen, bronzezeitlicher und neolithischer Populationen .....	46
Tab. A 16 Populationen der $F_{ST}$ -Tabelle.....	46
Tab. A 17 Ergebnisse zu phänotypischen Markern.....	47
Tab. A 18 Ergebnisse des $\chi^2$ -Test und des HWExact-Tests .....	53

### Tabellen aus Teil B

Tab. B 1 Sequenzierplattformen im Vergleich.....	91
Tab. B 2 Die Tabelle gibt die Anzahl an getesteten Proben aus den jeweiligen Fundregionen wieder .....	93
Tab. B 3 SNPs der kodierenden Region des mitochondrialen Genoms .....	95
Tab. B 4 Die nukleären Marker.....	96
Tab. B 5 Standard-PCR-Ansatz .....	98
Tab. B 6 Laufkonditionen eines Standard-PCR-Ansatzes .....	98
Tab. B 7 Primer für die Multiplex-PCR.....	100
Tab. B 8 Verteilung der Primer auf die drei Multiplex-PCR-Ansätze .....	102
Tab. B 9 mitochondrialer Multiplex-PCR-Ansatz .....	102
Tab. B 10 Laufkonditionen der mitochondrialen Multiplex-PCR.....	103
Tab. B 11 Primerpaare zur Amplifikation der nukleären Marker .....	103
Tab. B 12 nukleärer Multiplex-PCR-Ansatz.....	104
Tab. B 13 Laufkonditionen der nukleären Multiplex-PCR .....	104
Tab. B 14 <i>Blocking</i> -Oligos des Meyer Protokolls.....	106
Tab. B 15 Ansatz der <i>Blocking</i> -Oligos & der vereinigten <i>libraries</i> .....	106

---

Tab. B 16 PCR-Ansatz zur Amplifikation der DNA-Fragmente aus der Hybridisierungsreaktion .....	107
Tab. B 17 Laufkonditionen zur Amplifikation der <i>Capture</i> -Produkte .....	107
Tab. B 18 Reaktionsansatz für das <i>blunt-end repair</i> .....	111
Tab. B 19 Reaktionsansatz für die Ligation .....	111
Tab. B 20 Reaktionsansatz für das Adapter <i>fill-in</i> .....	112
Tab. B 21 Reaktionsansatz für die Dephosphorylierung .....	113
Tab. B 22 Ansatz für den Restriktionsverdau .....	113
Tab. B 23 <i>Cyclesequencing</i> -Ansatz .....	115
Tab. B 24 Laufkonditionen des <i>Cyclesequencing</i> .....	115
Tab. B 25 Adapter und Primer für die Präparation der Illumina <i>Library</i> .....	120
Tab. B 26 Positivkontrolle zur Illumina-Sequenzierung .....	121
Tab. B 27 Hybridisierungsansätze der Adapter und der Positivkontrolle .....	121
Tab. B 28 Programm zur Hybridisierung der Adapter und der Positivkontrolle .....	121
Tab. B 29 Reaktionsansatz des <i>Blunt End Repairs</i> .....	122
Tab. B 30 Reaktionsansatz zur Adapter Ligation .....	123
Tab. B 31 Reaktionsansatz zum Adapter <i>Fill-In</i> .....	123
Tab. B 32 Liste der verwendeten Indexe .....	124
Tab. B 33 Reaktionsansatz der Index-PCR .....	124
Tab. B 34 Laufkonditionen der Index-PCR .....	125
Tab. B 35 Analyse-Pipeline der Illumina Sequenzdaten .....	130
Tab. B 36 Probenanzahl nach Region .....	137
Tab. B 37 Amplifikationserfolg nach Region .....	138
Tab. B 38 Amplifikationserfolg nach PCR-Methode und Region .....	138
Tab. B 39 Amplifikationserfolg pro Fragment .....	139
Tab. B 40 Sequenziererfolg der Multiplex-PCRs .....	140
Tab. B 41 Erfolg der 454-Sequenzierungen .....	141
Tab. B 42 Anteil erfolgreich alignierter Sequenzen .....	142
Tab. B 43 Ergebnis der Illumina-Sequenzierung nach mitochondrialem Capture .....	143
Tab. B 44 Abdeckung des mitochondrialen Genoms .....	143
Tab. B 45 Kreuzkontamination 454 .....	146
Tab. B 46 Kreuzkontamination bei Illumina-Sequenzierung & <i>Capture</i> .....	147
Tab. B 47 Kreuzkontamination bei der Illumina-Sequenzierung .....	148
Tab. B 48 Haplogruppen nach SNPs der <i>coding region</i> und der HVR1 .....	150
Tab. B 49 Genotypen .....	152

## Tabellen aus dem Anhang

Tab. 1 Liste der verwendeten Chemikalien.....	195
Tab. 2 Liste der verwendeten Geräte .....	196
Tab. 3 SNPs der kodierenden Region für die Individuen der Fundplätze Barsučij Log (BL) und Aržan 2 (A)..	202
Tab. 4 SNPs der kodierenden Region für die Individuen der pazyrykzeitlichen Fundplätze des Ukok-Plateaus	203
Tab. 5 SNPs der kodierenden Region für die Individuen der Fundplätze des Chuya-Tals und aus Balik Sook .	204
Tab. 6 SNPs der kodierenden Region für die Individuen der Fundplätze aus dem kasachischen Altai .....	205
Tab. 7 SNPs der kodierenden Region für die Individuen der Zevakino-Cilikta-Kultur Ostkasachstans.....	205
Tab. 8 SNPs der kodierenden Region für die Individuen aus dem Westen der Eurasischen Steppe.....	206
Tab. 9 Beispiel für das Ranking nach HaploGrep.....	207
Tab. 10 Polymorphe Positionen und Haplogruppenzuordnung für die Individuen aus Barsučij Log .....	207
Tab. 11 Polymorphe Positionen und Haplogruppenzuordnung für die Individuen des Fundplatzes Aržan 2.....	208
Tab. 12 Polymorphe Positionen und Haplogruppenzuordnung für die Individuen aus dem russischen Altai ....	209
Tab. 13 Polymorphe Positionen und Haplogruppenzuordnung für die Individuen aus dem kasachischen Altai	210
Tab. 14 Polymorphe Positionen und Haplogruppenzuordnung für die Individuen Ostkasachstans.....	210
Tab. 15 Polymorphe Positionen und Haplogruppenzuordnung für die West-Skythen und Sarmaten.....	211
Tab. 16 Paarweise $F_{ST}$ -Werte zwischen den skythischen Populationen und die zugehörigen <i>p-values</i> .....	212
Tab. 17 Paarweise $F_{ST}$ -Werte zwischen skythischen und modernen Populationen mit <i>p-values</i> .....	213
Tab. 18 Paarweise $F_{ST}$ -Werte zw. skythischen, bronzezeitlichen und neolithischen Gruppen mit <i>p-values</i> .....	214
Tab. 19 <i>Reynold's distances</i> der skythischen Populationen ohne die Initialen Skythen .....	216
Tab. 20 <i>Reynold's distances</i> der eisenzeitlichen und modernen Populationen Eurasiens .....	216
Tab. 21 <i>Reynold's distances</i> der Pazyryk-Gruppen .....	216
Tab. 22 Frequenzen der abgeleiteten Allele der Skythen und der Vergleichspopulationen .....	217

Tab. 1 Liste der verwendeten Chemikalien

Chemikalien	Hersteller
Agarose	Invitrogen™, Life technologies, Darmstadt, Deutschland
Alconox	Sigma Aldrich, St. Louis, Missouri, USA
AmpliAq Gold™ DNA Polymerase (5 U/μl)	Applied Biosystems, Life technologies, Darmstadt, Deutschland
ATP (10mM)	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
Big Dye™ Terminator v1.1	Applied Biosystems, Life technologies, Darmstadt, Deutschland
BSA (20mg/ml)	Roche Diagnostics, Mannheim, Deutschland
Bst polymerase, großes Fragment (8 U/μl)	New England Biolabs GmbH, Frankfurt am Main, Deutschland
Calf-intestine phosphatase (10 U/μl)	New England Biolabs GmbH, Frankfurt am Main, Deutschland
Chloroform	Carl Roth GmbH, Karlsruhe, Deutschland
DNA-ExitusPlus™ IF	Applichem GmbH, Darmstadt, Deutschland
dNTPs (je 10 mM)	Qiagen, Hilden, Deutschland
EDTA (0,5 M, pH8)	Ambion®, Life technologies, Darmstadt, Deutschland
Ethanol	Carl Roth GmbH, Karlsruhe, Deutschland
Ethidiumbromid	Carl Roth GmbH, Karlsruhe, Deutschland
Exonuklease I	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
Exonuklease Puffer	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
GelRed™	BIOTREND, Köln, Deutschland
Gen-Ruler™ 50bp DNA-Ladder	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
HIDI-Formamid	Applied Biosystems, Life technologies, Darmstadt, Deutschland
HPLC-Wasser	ACROS Organics, Thermo Fisher Scientific, Schwerte, Deutschland
Hydroxylapatit	Sigma Aldrich, St. Louis, Missouri, USA
MgCl <sub>2</sub> (25 mM)	Applied Biosystems, Life technologies, Darmstadt, Deutschland
MinElute PCR purification kit	Qiagen, Hilden, Deutschland
MSB® Spin PCRapace Kit	STRATEC Molecular GmbH, Berlin, Deutschland
Multiplex PCR Kit	Qiagen, Hilden, Deutschland
Natriumhypochlorid DanKlorix	Palmolive-Colgate GmbH, Hamburg, Deutschland
NEBuffer (10x)	New England Biolabs GmbH, Frankfurt am Main, Deutschland
N-Laurylsarkosin	Merck, Darmstadt, Deutschland
Oligonukleotide	Biospring, Frankfurt am Main, Deutschland
PCR Gold Buffer (10x)	Applied Biosystems, Life technologies, Darmstadt, Deutschland
PEG-4000 (50%)	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
POP-6™, Performance Optimized Polymer 6	Applied Biosystems, Life technologies, Darmstadt, Deutschland
Proteinase K	Roche Diagnostics, Mannheim, Deutschland
Roti®-Phenol/Chloroform/Isoamylalkohol	Carl Roth GmbH, Karlsruhe, Deutschland
Seesand (SiO <sub>2</sub> )	Carl Roth GmbH, Karlsruhe, Deutschland
Sephadex G-50 Fine	GE Healthcare, Fairfield, Connecticut, USA
Shrimp Alkaline Phosphatase	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
Shrimp Alkaline Phosphatase Puffer	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
Stratagene Srf I Restriktionsenzym (10 U/μl)	Agilent Technologies, Santa Clara, California, USA
T4 DNA polymerase (5 U/μl)	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
T4 Ligase Puffer (10x)	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
T4 Ligase (5 U/μl)	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
T4 polynucleotide kinase (10 U/μl)	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
Tango Puffer (10x)	Fermentas, Thermo Fisher Scientific, Schwerte, Deutschland
ThermoPol Puffer (10x)	New England Biolabs GmbH, Frankfurt am Main, Deutschland
Tris-Puffer	Carl Roth GmbH, Karlsruhe, Deutschland
Universal Puffer (10x)	Agilent Technologies, Santa Clara, California, USA

**Tab. 2 Liste der verwendeten Geräte**

Gerät	Hersteller
3130 Genetic Analyzer	Applied Biosystems, Life technologies, Darmstadt,
3D-Wippschüttler GyroMini™	Carl Roth GmbH, Karlsruhe, Deutschland
Agilent 2100 Bioanalyzer	Agilent Technologies, Santa Clara, California, USA
Brutschrank	Memmert, Schwabach, Deutschland
Dremel Multi	Dremel, Utrecht, Niederlande
Elektronische Handsäge	MAFRA Maile + Hass GmbH, Schutterwald, Deutschland
Elektrophorese Gelkammer	Carl Roth GmbH, Karlsruhe, Deutschland
Kugel-Schwingmühle	Retsch, Haan, Deutschland
Mastercycler gradient	Eppendorf, Hamburg, Deutschland
Mastercycler	Eppendorf, Hamburg, Deutschland
Punktstrahlgerät	Harnisch + Rieth GmbH & Co.KG, Winterbach, Deutschland
Qubit® Fluorometer	Qubit®, Life technologies, Darmstadt, Deutschland
Thermomixer comfort	Eppendorf, Hamburg, Deutschland
UV-Tisch	Benda Laborgeräte, Wiesloch, Deutschland
Vortexer, MINI-Vortex	Kisker Biotech GmbH & Co. KG, Steinfurt, Deutschland
Zentrifuge Hettich Universal 320	Hettich, Tuttlingen, Deutschland
Zentrifuge Hettich Mikro 200	Hettich, Tuttlingen, Deutschland

## Alignment der HVR1-Konsensussequenzen

```

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
          16020   16030   16040   16050   16060   16070   16080   16090   16100   16110
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
NC_012920.seq ACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGGTATGTATTTCGTACATTACTGCCAGCC
BL1          .....C.....
BL2          .....
BL3          .....C.....
BL4          .....C.....
BL5          .....G.....C.....
BL6          .....
A2          .....
A3          .....
A4          .....
A5          .....
A6          .....
A7          .....
A8          .....
A9          .....
A10         .....
A11         .....
A14         .....
A17         .....
A19         .....A.....
A20         .....
A21         .....
Ak1_1       .....C.....
Ak4_1       .....
Ak5_1       .....
Ak5_4       .....
Ak5_5       .....G.....
Ak5_6       .....G.....
Ak5_7       .....C.....
Ak5_8       .....
Ala1        .....C.....
Ala2        .....C.....C.....
Ala4        .....
B1_1        .....
B1_2        .....A.....
B3_1        .....
BS1         .....
BS2         .....
Bt1         .....
Bt2         .....G.....C.....
D1          .....
J12_1       .....
J12_3       .....
J12_6       .....C.....
J12_7       .....
J12_8       .....
J12_9       .....
J22_1       .....
K1          .....C.....
Mo1         .....C.....
Mo2         .....
U1_1        .....
U1_2        .....
U2_1        .....C.....
U2_2        .....C.....
U4_1        .....G.....C.....
U4_4        .....
VK1_1       .....
Be2         .....
Be3         .....
Be4         .....
Be6         .....
Be8         .....
Be9         .....
Be11        .....C.....
Be12        .....
Be14        .....
Ta1         .....
Is1         .....
Is2         .....
Is4         .....
Ze2         .....
Ze3         .....C.....
Ze4         .....
Ze5         .....
Ze6         .....C.....
Ze7         .....
Ze8         .....
Ze9         .....

```

```

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      16020   16030   16040   16050   16060   16070   16080   16090   16100   16110
-----+-----+-----+-----+-----+-----+-----+-----+-----+
NC_012920.seq ACTATTCTCTGTTCTTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCC
PR1 .....
PR3 .....
PR4 .....
PR5 .....
PR6 .....
PR7 .....
PR8 .....
PR9 .....G.....C.....
PR10 .....
PR11 .....
PR13 .....
Ko11 .....
Ko12 .....
Ko13 .....
Ko15 .....T.....
Nov5 .....
Nov7 .....
Nov10 .....

```

```

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      16120   16130   16140   16150   16160   16170   16180   16190   16200   16210
-----+-----+-----+-----+-----+-----+-----+-----+-----+
NC_012920.seq ACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCTCCCATGCTTACAAGCAAGTACA
BL1 .....
BL2 .....
BL3 .....
BL4 .....
BL5 .....C.....C.....C.....
BL6 .....C.....
A2 .....
A3 .....
A4 .....
A5 .....T.....
A6 .....T.....
A7 .....C.....
A8 .....
A9 .....T.....
A10 .....
A11 .....
A14 .....C.....G.....T.....C.....
A17 .....C.....
A19 .....
A20 .....
A21 .....
Ak1_1 .....A.....
Ak4_1 .....
Ak5_1 .....A.....
Ak5_4 .....
Ak5_5 .....C.....CC.....C.....
Ak5_6 .....C.....CC.....C.....
Ak5_7 .....A.....
Ak5_8 .....
Ala1 .....A.....
Ala2 .....
Ala4 .....C.....C.....
B1_1 .....
B1_2 .....
B3_1 .....C.....G.....T.....C.....
BS1 .....C.....
BS2 .....
Bt1 .....C.....
Bt2 .....C.....C.....C.....
D1 .....C.....G.....T.....C.....
J12_1 .....C.....
J12_3 .....
J12_6 .....T.....
J12_7 .....C.....
J12_8 .....
J12_9 .....
J22_1 .....C.....
K1 .....
Mo1 .....CT.....
Mo2 .....
U1_1 .....TC.....
U1_2 .....A.....T.....
U2_1 .....G.....
U2_2 .....G.....
U4_1 .....C.....CC.....C.....
U4_4 .....
VK1_1 .....
Be2 .....
Be3 .....
Be4 .....
Be6 .....
Be8 .....
Be9 .....
Be11 .....A.....
Be12 .....
Be14 .....
Ta1 .....

```



```

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      16120   16130   16140   16150   16160   16170   16180   16190   16200   16210
-----+-----+-----+-----+-----+-----+-----+-----+-----+
NC_012920.seq ACCATGAATATTGTACGGTACCATAAACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAAACCCCTCCCATGCTTACAAGCAAGTACA
Is1          .....T.....
Is2          .....
Is4          .....
Ze2          .....C.....
Ze3          .....
Ze4          .....G.....
Ze5          .....A.....
Ze6          .....
Ze7          .....T.....
Ze8          .....
Ze9          .....T.....
PR1          .....T.....
PR3          .....A.....
PR4          .....CC...C.....
PR5          .....C.....
PR6          .....C...C.....
PR7          .....A.....C.....
PR8          .....C.....
PR9          .....C...C.....
PR10         .....T.....
PR11         .....C.....G.....T...C.....
PR13         .....T.....T.....
Ko11         .....C...C.....
Ko12         .....
Ko13         .....
Ko15         .....C.....T.....
Nov5         .....
Nov7         .....C.....
Nov10        .....C.....
-----+-----+-----+-----+-----+-----+-----+-----+-----+
      16220   16230   16240   16250   16260   16270   16280   16290   16300   16310
-----+-----+-----+-----+-----+-----+-----+-----+-----+
NC_012920.seq GCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAACTACCCACCCTTAACAGTACATAGTA
BL1          .....T.....C.....
BL2          .....T.....T.....G.....
BL3          .....T.....C...C.....
BL4          .....T.....C...C.....
BL5          .....
BL6          .....T.....T.....
A2          .....T.....T.....
A3          .....T.....
A4          .....T.....
A5          .....T.....C...C.....
A6          .....T.....T.....C.....
A7          .....C.....T.....
A8          .....
A9          .....T.....T.....C.....
A10         .....C.....T.....T...C.....
A11         .....T.....
A14         .....T.....
A17         .....
A19         .....T.....C.....
A20         .....T...G.....T.....T.....
A21         .....T.....T.....T...T.....
Ak1_1       .....T.....T.....T...C.....
Ak4_1       .....T.....T.....T...C.....
Ak5_1       .....T.....T.....C.....
Ak5_4       .....T.....T.....
Ak5_5       .....
Ak5_6       .....
Ak5_7       .....T.....C.....
Ak5_8       .....T.....T...C.....
Ala1        .....T.....T.....C.....
Ala2        .....T.....T.....C.....
Ala4        .....T.....T...T.....
B1_1        .....T.....
B1_2        .....T.....C.....
B3_1        .....T.....T.....
BS1         .....T.....T.....
BS2         .....
Bt1         .....T.....C.....
Bt2         .....
D1          .....T.....
J12_1       .....T.....C.....
J12_3       .....
J12_6       .....T.....T.....C.....
J12_7       .....T.....T.....T...C.....
J12_8       .....T.....T.....T...T.....
J12_9       .....T.....C.....T.....
J22_1       .....T.....T.....C.....
K1          .....
Mo1         .....T.....
Mo2         .....T.....C.....
U1_1        .....T.....T.....
U1_2        .....TC.....T.....T...C.....
U2_1        .....C.....T.....C.....
U2_2        .....C.....T.....C.....
U4_1        .....
U4_4        .....C.....C.....

```

# Anhang

```

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
          16220   16230   16240   16250   16260   16270   16280   16290   16300   16310
-----+-----+-----+-----+-----+-----+-----+-----+-----+
NC_012920.seq GCAATCAACCCCTCAACTATCACACATCAACTGCAACTCCAAGGCCACCCCTCACCCTAGGATACCAACAACCTACCCACCCCTTAACAGTACATAGTA
VK1_1         .....C.....
Be2          .....T.....
Be3          .....
Be4          .....T.....
Be6          .....T.....
Be8          .....T.....
Be9          .....T.....
Be11         .....T.....
Be12         .....C.....
Be14         .....T.....
Ta1          .....C.....
Is1          .....T.....
Is2          .....
Is4          .....
Ze2          .....T.....
Ze3          .....C.....
Ze4          .....T.....
Ze5          .....T.....
Ze6          .....T.....
Ze7          .....
Ze8          .....T.....
Ze9          .....T.....
PR1          .....
PR3          .....T.....
PR4          .....YC.....
PR5          .....T.....
PR6          .....A.....
PR7          .....T.....
PR8          .....T.....
PR9          .....
PR10         .....T.....
PR11         .....T.....
PR13         .....T.....
Kol1         .....T.....
Kol2         .....C.....
Kol3         .....
Kol5         .....T.....
Nov5         .....
Nov7         .....
Nov10        .....T.....

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
          16320   16330   16340   16350   16360   16370   16380   16390   16400
-----+-----+-----+-----+-----+-----+-----+-----+-----+
NC_012920.seq CATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCATGGATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCC
BL1          .....T.....
BL2          .....
BL3          .....T.....
BL4          .....T.....
BL5          .....C.....
BL6          .....A.....
A2           .....A.....
A3           .....C.....
A4           .....A.....
A5           .....T.....
A6           .....G.....
A7           .....
A8           .....A.....
A9           .....G.....
A10          .....A.....
A11          .....C.....
A14          .....
A17          .....
A19          .....T.....
A20          .....C.....
A21          .....A.....
Ak1_1        .....T.....
Ak4_1        .....A.....
Ak5_1        .....T.....
Ak5_4        .....A.....
Ak5_5        .....C.....
Ak5_6        .....C.....
Ak5_7        .....T.....
Ak5_8        .....A.....
Ala1         .....T.....
Ala2         .....T.....
Ala4         .....
B1_1         .....C.....
B1_2         .....C.....
B3_1         .....
BS1          .....
BS2          .....C.....
Bt1          .....
Bt2          .....C.....
D1           .....

```

```

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
          16320   16330   16340   16350   16360   16370   16380   16390   16400
-----+-----+-----+-----+-----+-----+-----+-----+-----+
NC_012920.seq CATAAAGCCATTACCGTACATAGCACATTACAGTCAAATCCCCTTCTCGTCCCATGGATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCC
J12_1 .....T.....
J12_3 .....T.....
J12_6 .....T.....
J12_7 .....T.....
J12_8 .....A.....
J12_9 .....C.....C.....
J22_1 .....T.....
K1 .....T.....
Mo1 .....A.....C.....
Mo2 .....G.....C.....
U1_1 .....T.....
U1_2 .....T.....
U2_1 .....T.....
U2_2 .....T.....
U4_1 .....T.....A.....
U4_4 .....T.....
VK1_1 .....T.....C.....
Be2 .....T.....C.....
Be3 .....T.....C.....
Be4 .....A.....C.....
Be6 .....A.....C.....
Be8 .....T.....C.....
Be9 .....A.....C.....
Be11 .....T.....C.....
Be12 .....T.....C.....
Be14 .....A.....C.....
Ta1 .....T.....C.....
Is1 .....T.....C.....
Is2 .....T.....C.....
Is4 .....T.....C.....
Ze2 .....T.....C.....
Ze3 .....T.....C.....
Ze4 .....T.....C.....A.....
Ze5 .....T.....C.....
Ze6 .....T.....C.....
Ze7 .....T.....C.....
Ze8 .....T.....C.....
Ze9 .....T.....G.....C.....
PR1 .....T.....G.....C.....
PR3 .....T.....C.....
PR4 .....T.....C.....
PR5 .....T.....C.....
PR6 .....T.....C.....G.....
PR7 .....T.....T.....C.....
PR8 .....T.....C.....
PR9 .....T.....C.....
PR10 .....T.....C.....
PR11 .....T.....C.....
PR13 .....T.....G.....C.....
Kol1 .....T.....A.....
Kol2 .....T.....C.....
Kol3 .....T.....C.....
Kol5 .....T.....C.....
Nov5 .....T.....C.....
Nov7 .....T.....C.....
Nov10 .....T.....C.....

```













## Haplogruppenzuordnung anhand von SNPs der kodierenden Region und der HVR1 des mitochondrialen Genoms

Die Haplogruppen wurden mit Hilfe des Programms *HaploGrep* bestimmt (KLOSS-BRANDSTÄTTER 2011), basierend auf dem *PhyloTree mtDNA tree Build 15* (VAN OVEN & KAYSER 2009). In der ersten Spalte sind die Proben angegeben, in der zweiten die Haplogruppen und in der dritten das von HaploGrep erstellte Ranking für die Haplogruppe. Bei gleichem Ranking wurde die basalere Haplogruppe angegeben. Haplogruppen, die keine 100%ige Übereinstimmung haben, tragen entweder eine Mutation, die in dem definierenden Haplotyp bisher nicht detektiert wurde, oder es fehlt eine Mutation. Steht eine zweite Haplogruppe in Klammer, wird für diese jeweils eine andere Mutation als privat angesehen (s. Tab. 7). Weiterhin sind die polymorphen Positionen und die Base relativ zur Referenzsequenz rCRS (*Accession number*: NC\_012920) angegeben.

**Tab. 9** Beispiel für das Ranking nach HaploGrep

Probe	Haplogruppe	Ranking	Grund
VK1_1	K1b2	96,6	16320T <i>local privat mutation</i>
	K1c2	91,3	12738G <i>local privat mutation</i>

**Tab. 10** Polymorphe Positionen und Haplogruppenzuordnung für die Individuen des Fundplatzes Barsučij Log

Probe	Haplogruppe	HaploGrep Ranking	Polymorphismen																				
BL1	C5+16093	100.0	4173G	7028T	10398G	10400T	10873C	11719A	12705T	13263G	14766T	14783C	16093C	16223T	16288C	16298C	16327T						
BL2	U5a1a2 (U5a2c3)	91.3 (91.0)	7028T	11719A	12308G	12346T	14766T	14793G	16256T	16270T	16309G												
BL3	C5+16093	100.0	4173G	7028T	10398G	10400T	10873C	11719A	12705T	13263G	14766T	14783C	16093C	16223T	16288C	16298C	16327T						
BL4	C5+16093	100.0	4173G	7028T	10398G	10400T	10873C	11719A	12705T	13263G	14766T	14783C	16093C	16223T	16288C	16298C	16327T						
BL5	U2e2	100.0	7028T	10876G	11719A	12308G	13734C	14766T	16051G	16092C	16129C	16183C	16189C	16362C									
BL6	A4	98.6	663G	4824G	7028T	11719A	12705T	14766T	16189C	16223T	16290T	16319A	16362C										

**Tab. 11 Polymorphe Positionen und Haplogruppenzuordnung für die Individuen des Fundplatzes Aržan 2**

Probe	Haplogruppe	HaploGrep Ranking	Polymorphismen																	
A2	A8	100.0	663G	4824G	7028T	11719A	12705T	14766T	16223T	16242T	16290T	16319A								
A3	G	100.0	4833G	7028T	10398G	10400T	10873C	11719A	12705T	14783C	16223T	16362C								
A4	D4b1a2a	100.0	3010A	4883T	5178A	7028T	8414T	10398G	10400T	10873C	11719A	12705T	13720T	14766T	14783C	14815T	16223T	16319A	16362C	
A5	C5b1	100.0	7028T	10398G	10400T	10873C	11719A	12705T	13263G	14766T	14783C	16148T	16223T	16288C	16298C	16327T				
A6	U5a1f1	100.0	7028T	11719A	2308G	14766T	14793G	16192T	16256T	16270T	16311C	16399G								
A7	Y1	93.7	7028T	8392A	10398G	11719A	12705T	12732C	14766T	16126C	16231C	16266T								
A8	U4a3	92.0	7028T	11719A	2308G	14766T	16327A	16356C	16362C											
A9	U5a1d2b	100.0	3027C	4924C	7028T	11719A	12308G	14766T	14793G	16192T	16256T	16270T	16304C	16399G						
A10	A11 (A8)	88.6 (88.0)	663G	4824G	6779G	7028T	11719A	12705T	14766T	16224C	16242T	16290T	16293C	16319A						
A11	G	100.0	4833G	7028T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16223T	16362C							
A14	T1a	100.0	4216C	4917G	7028T	11719A	14766T	16126C	16163G	16186T	16189C	16294T								
A17	H (H1+16189)	(64.0)	16189C																	
A19	C4	96.2	5171G	7028T	10398G	10400T	10873C	11719A	11969A	12705T	13263G	14766T	14783C	16111A	16223T	16298C	16327T			
A20	G2a	96.1	4833G	6340T	7028T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16223T	16227G	16262T	16278T	16362C			
A21	A4	95.7	663G	4824G	7028T	11719A	12705T	14766T	16223T	16286T	16290T	16319A	16362C							

**Tab. 12 Polymorphe Positionen und Haplogruppenzuordnung für die Individuen der pazyrkzeitlichen Fundplätze aus dem russischen Altai**

Ak = Ak Alacha; Ala = Alagail; B = Barburgazy; BS = Balik Sook; Bt = Baratal; D = Dcholin; J = Justyd; K = Kuturguntas; Mo = Moinak; U = Ulandryk; VK = Verch Kal'džin

Probe	Haplogruppe	HaploGrep Ranking	Polymorphismen																																					
Ak1_1	C4a1+16129	100.0	7028T	10398G	10400T	10873C	11719A	11969A	12705T	13263G	14766T	14783C	16093C	16129A	16223T	16298C	16327T																							
Ak4_1	A11 (A8)	91.7 (91.0)	663G	4824G	6779G	7028T	11719A	12705T	14766T	16242T	16290T	16293C	16319A																											
Ak5_1	C4	98.8	7028T	10398G	10400T	10873C	11719A	11969A	12705T	13263G	14766T	14783C	16129A	16223T	16298C	16327T																								
Ak5_4	D4b1	93.5	3010A	4883T	5178A	7028T	8414T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16223T	16239T	16319A	16362C																						
Ak5_5	U2e1a	94.7	7028T	8962G	10876G	11719A	11732C	12308G	13734C	14766T	16051G	16129C	16182C	16183C	16189C	16362C																								
Ak5_6	U2e1a	94.7	7028T	8962G	10876G	11719A	11732C	12308G	13734C	14766T	16051G	16129C	16182C	16183C	16189C	16362C																								
Ak5_7	C4a1+16129	100.0	7028T	10398G	10400T	10873C	11719A	11969A	12705T	13263G	14766T	14783C	16093C	16129A	16223T	16298C	16327T																							
Ak5_8	A11 (A8)	91.7 (91.0)	663G	4824G	6779G	7028T	11719A	12705T	14766T	16242T	16290T	16293C	16319A																											
Ala1	C4a1+16129	100.0	7028T	10398G	10400T	10873C	11719A	11969A	12705T	13263G	14766T	14783C	16093C	16129A	16223T	16298C	16327T																							
Ala2	C4	95.6	7028T	10398G	10400T	10873C	11719A	11969A	12705T	13263G	14766T	14783C	16086C	16223T	16287T	16298C	16327T																							
Ala4	T	93.6	4216C	4917G	7028T	11719A	14766T	16126C	16189C	16292T	16294T																													
B1_1	G	100.0	4833G	7028T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16223T	16362C																											
B1_2	D4m2	97.9	3010A	4883T	5178A	7028T	8414T	8989A	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16042A	16223T	16243C	16362C																					
B3_1	T1a1b	100.0	4216C	4917G	7028T	10143A	11719A	14766T	16126C	16163G	16186T	16189C	16294T																											
BS1	X2b	100.0	6371T	7028T	8393T	11719A	12705T	13708A	14766T	16189C	16223T	16278T																												
BS2	U4b1a4	100.0	7028T	11719A	11965T	12308G	14766T	16356C																																
Bt1	T2b+@16296	100.0	4216C	4917G	7028T	11719A	14766T	16126C	16294T	16304C																														
Bt2	U2e2a	100.0	4553C	7028T	10876G	11719A	12308G	13734C	14766T	16051G	16092C	16129C	16183C	16189C	16362C																									
D1	T1a1b	100.0	4216C	4917G	7028T	10143A	11719A	14766T	16126C	16163G	16186T	16189C	16294T																											
J12_1	T2b+@16296	100.0	4216C	4917G	7028T	11719A	14766T	16126C	16294T	16304C																														
J12_3	U7	91.8	7028T	10142T	11719A	12308G	14766T	16093C	16318T																															
J12_6	Z	100.0	7028T	9090C	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16185T	16223T	16260T	16298C																									
J12_7	T2b+@16296	100.0	4216C	4917G	7028T	11719A	14766T	16126C	16294T	16304C																														
J12_8	A8	97.9	663G	4824G	7028T	11719A	12705T	14766T	16223T	16242T	16278T	16290T	16319A																											
J12_9	G1a1	97.0	4833G	7028T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16223T	16263C	16325C	16362C																									
J22_1	T2b+@16296	100.0	4216C	4917G	7028T	11719A	14766T	16126C	16294T	16304C																														
K1	CRS																																							
Mo1	D4b1a2a1	97.9	3010A	4883T	5178A	7028T	8414T	10398G	10400T	10873C	11719A	12705T	13720T	14766T	14783C	14815T	16093C	16172C	16173T	16223T	16319A	16362C																		
Mo2	D4h4a	97.4	3010A	4883T	5178A	7028T	8414T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16223T	16311C	16316G	16362C																						
U1_1	W4a	91.1	7028T	8251A	8994A	11719A	11947G	12705T	14766T	16188T	16189C	16223T	16286T																											
U1_2	Z1a	98.1	7028T	9090C	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16129A	16185T	16223T	16224C	16260T	16294T	16298C																						
U2_1	F2a	92.5	6392C	7028T	11719A	12338C	13708A	14766T	16093C	16203G	16231C	16291T	16304C																											
U2_2	F2a	92.5	6392C	7028T	11719A	12338C	13708A	14766T	16093C	16203G	16231C	16291T	16304C																											
U4_1	U2e2	98.5	7028T	10876G	11719A	12308G	13734C	14766T	16051G	16092C	16129C	16182C	16183C	16189C	16362A																									
U4_4	K	100.0	11719A	12308G	14798C	16224C	16311C																																	
VK1_1	K1 (K1b2/K1c2)	89.7(96.5/93.3)	7028T	10398G	11719A	12308G	12738G	14766T	14798C	16224C	16311C	16320T																												

**Tab. 13 Polymorphe Positionen und Haplogruppenzuordnung für die Individuen der pazrykzeitlichen Fundplätze aus dem kasachischen Altai**

Be = Berel'; Ta = Tar Asu

Probe	Haplogruppe	HaploGrep Ranking	Polymorphismen																		
			3010A	4883T	5178A	6975C	7028T	8414T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16223T	16362C				
Be2	D4	96.0																			
Be3	CRS																				
Be4	A11 (A8)	91.7 (91.0)	663G	4824G	6779G	7028T	11719A	12705T	14766T	16242T	16290T	16293C	16319A								
Be6	A6	95.7	654C	663G	4553C	4824G	7028T	11719A	12705T	14766T	16223T	16290T	16319A	16362C							
Be8	D4	100.0	3010A	4883T	5178A	7028T	8273T	8414T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16223T	16362C				
Be9	A4f	100.0	663G	4824G	7028T	11719A	12705T	14766T	16223T	16290T	16292A	16319A	16362C								
Be11	C4a1+16129	100.0	7028T	10398G	10400T	10873C	11719A	11969A	12705T	13263G	14766T	14783C	16093C	16129A	16223T	16298C	16327T				
Be12	HV2	100.0	7028T	11935C	16217C																
Be14	A6	95.7	654C	663G	4553C	4824G	7028T	11719A	12705T	14766T	16223T	16290T	16319A	16362C							
Ta1	K2b1a	100.0	7028T	11719A	12308G	14766T	14798C	16224C	16270T	16311C											

**Tab. 14 Polymorphe Positionen und Haplogruppenzuordnung für die Individuen der Zevakino-Čilikta-Kultur Ostkasachstans**

Probe	Haplogruppe	HaploGrep Ranking	Polymorphismen																		
			3010A	4883T	5178A	7028T	8414T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16174T	16223T	16362C				
Is1	D4h1/ D4j8	je 100.0	3010A	4883T	5178A	7028T	8414T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16174T	16223T	16362C				
Is2	HV-CRS	100.0	7028T																		
Is4	H2a1	100.0	16354T																		
Ze2	T2b	100.0	4216C	4917G	7028T	11719A	14766T	16126C	16294T	16296T	16304C										
Ze3	K1a1 (K1b1b)	je 100.0	7028T	10398G	11719A	12308G	14766T	14798C	16093C	16224C	16311C										
Ze4	D4h4a (D4j+16311)	je 96.1	3010A	4883T	5178A	7028T	8414T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16171G	16223T	16311C	16362C			
Ze5	I	100.0	4529T	7028T	8251A	10034C	10398G	11719A	12705T	14766T	16129A	16223T	16391A								
Ze6	C4 (C4a1/C4b8/C4+152+16093)	98,6 (je 100.0)	7028T	10398G	10400T	10873C	11719A	11969A	12705T	13263G	14766T	14783C	16093C	16223T	16298C	16327T					
Ze7	U4a1	100.0	7028T	11719A	12308G	14766T	16134T	16356C													
Ze8	D4	100.0	3010A	4883T	5178A	7028T	8414T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16223T	16362C					
Ze9	D4j3	97.3	689C	3010A	4883T	5178A	7028T	8414T	10398G	10400T	10873C	11719A	12705T	14766T	14783C	16184T	16223T	16265C	16311C	16362C	

**Tab. 15 Polymorphe Positionen und Haplogruppenzuordnung für die skythenzeitlichen Individuen aus dem Westen der Eurasischen Steppe**

Pr = Pokrovka; Kol = Kolbino; Nov = Novozavedennoe

Probe	Haplogruppe	HaploGrep Ranking	Polymorphismen																	
Pr1	U3	83.4	7028T	12308G	16188T	16327T	16343G													
Pr3	M (M5/M7b'd/M13b)	(je 100.0)	10398G	10400T	16129A	16223T														
Pr4	U1a'c	100.0	16182C	16183C	16189C	16248Y	16249C													
Pr5	T	100.0	16126C	16294T																
Pr6	F1b	90.9	16183C	16189C	16232A	16249C	16304C	16311C	16399G											
Pr7	N1a1a1a1a	72.3	10398G	16147A	16172C	16189C	16223T	16248T	16320T	16355T										
Pr8	T2	100.0	16126C	16294T	16296T															
Pr9	U2e2	100.0	16051G	16092C	16129C	16183C	16189C	16362C												
Pr10	H2a1f	81.8	16193T	16264T	16354T															
Pr11	T1a	100.0	16126C	16163G	16186T	16189C	16294T													
Pr13	U5a1d2b	96.6	16192T	16256T	16270T	16304C	16311C	16399G												
Kol1	X4	100.0	6371T	7028T	11719A	12705T	14766T	16183C	16189C	16223T	16266T	16274A	16278T	16390A						
Kol2	H8c	100.0	13711A	16288C	16362C															
Kol3	U4	100.0	7028T	11719A	12308G	14766T	16356C													
Kol5	J2b1a	100.0	7028T	10398G	11719A	13708A	14766T	16069T	16126C	16193T	16278T									
Nov5	CRS																			
Nov7	H1c	84.7	477C	3010A	16136C															
Nov9	T2g1	100.0	4216C	4917G	7028T	11719A	14766T	14798C	14839G											
Nov10	X	92.9	6371T	7028T	11719A	12705T	14766T	16129C	16189C	16223T	16278T									

Paarweise  $F_{ST}$ -Werte mit zugehörigen  $p$ -values

*Distance method: pairwise difference; gamma-value  $\alpha = 0$ ; No. of permutations = 1000; Significance level = 0,05*

**Tab. 16** Paarweise  $F_{ST}$ -Werte zwischen den skythischen Populationen (links unten) und die zugehörigen  $p$ -values (rechts oben)

	TagarTes	Tagar	Aldy Bel	Pazyryk	Zevakino-Čilikta	Initiale Skythen	Skythen	Sarmaten
TagarTes			0.24902+-0.0123	0.29688+-0.0135	0.22754+-0.0124	0.83594+-0.0086	0.43848+-0.0164	0.36426+-0.0164
Tagar			0.16602+-0.0117	0.10059+-0.0088	0.07031+-0.0088	0.42285+-0.0156	0.03613+-0.0064	0.03711+-0.0050
Aldy Bel	0.01119	0.04039		0.72070+-0.0168	0.46777+-0.0151	0.80176+-0.0101	0.72852+-0.0161	0.06738+-0.0070
Pazyryk	0.00504	0.04204	-0.00959		0.67871+-0.0157	0.87500+-0.0104	0.46387+-0.0152	0.03418+-0.0054
Zevakino-Čilikta	0.01507	0.08830	-0.00223	-0.01111		0.59473+-0.0136	0.55371+-0.0127	0.03613+-0.0053
Initiale Skythen	-0.08808	0.03351	-0.05889	-0.07597	-0.02800		0.98926+-0.0026	0.95898+-0.0055
Skythen	-0.00181	0.07235	-0.01405	-0.00055	-0.00660	-0.12205		0.68066+-0.0121
Sarmaten	0.00428	0.09589	0.04650	0.04020	0.05125	-0.10115	-0.01468	

Tab. 17 Paarweise  $F_{ST}$ -Werte zwischen skythischen und modernen Populationen (links unten) und die dazugehörigen  $p$ -values (rechts oben)

	1	2	3	4	5	6	7	8	9
1		0.73145+-0.0153	0.27344+-0.0142	0.50879+-0.0139	0.17383+-0.0099	0.11523+-0.0106	0.03125+-0.0055	0.01270+-0.0034	0.00000+-0.0000
2	-0.00959		0.29590+-0.0158	0.71191+-0.0143	0.01758+-0.0046	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000
3	0.01119	0.00504		0.23730+-0.0140	0.26660+-0.0143	0.14551+-0.0108	0.18262+-0.0103	0.02246+-0.0051	0.00000+-0.0000
4	-0.00223	-0.01111	0.01507		0.19531+-0.0117	0.48145+-0.0140	0.21484+-0.0134	0.04492+-0.0059	0.01270+-0.0031
5	0.01508	0.01871	0.00767	0.01582		0.00098+-0.0010	0.00000+-0.0000	0.00000+-0.0000	0.02441+-0.0050
6	0.01401	0.01777	0.01281	-0.00286	0.03437		0.04688+-0.0079	0.00000+-0.0000	0.00000+-0.0000
7	0.03740	0.03619	0.01393	0.01587	0.05090	0.00953		0.01562+-0.0037	0.00000+-0.0000
8	0.04729	0.05384	0.03679	0.03767	0.09892	0.02612	0.02065		0.00000+-0.0000
9	0.07058	0.05786	0.05994	0.04839	0.01279	0.08136	0.10217	0.15671	
	Legende								
1	Aldy Bel (Tuva); N=15								
2	Pazyryk (Altai); N=71								
3	Tagar/Tes (Chakassien); N=16								
4	Zevakino-Čilikta; N=11								
5	West-Skythen+Sarmaten; N=34								
6	Altai modern; N=110								
7	Chakassien modern; N=53								
8	Tuva modern; N=90								
9	SW-Russland modern; N=198								

Tab. 18 Paarweise  $F_{ST}$ -Werte zwischen skythischen, bronzezeitlichen und neolithischen Populationen (links unten) und zugehörige  $p$ -values (rechts oben)

	1	2	3	4	5	6	7	8
1		0.61523+0.0154	0.18945+0.0126	0.30566+0.0149	0.35840+0.0141	0.02637+-0.0044	0.07324+-0.0064	0.17480+-0.0114
2	-0.00928		0.01660+0.0031	0.04492+0.0063	0.62891+-0.0151	0.00000+-0.0000	0.00098+-0.0010	0.16895+-0.0091
3	0.01515	0.01909		0.19238+-0.0146	0.49902+-0.0153	0.00000+-0.0000	0.00000+-0.0000	0.65332+-0.0153
4	0.00988	0.02778	0.01224		0.04785+-0.0059	0.00000+-0.0000	0.00000+-0.0000	0.84766+-0.0098
5	0.00865	-0.01083	-0.00516	0.05752		0.49902+-0.0131	0.14941+-0.0120	0.45703+-0.0141
6	0.06612	0.04747	0.07338	0.12438	-0.00745		0.12012+-0.0098	0.00391+-0.0019
7	0.04194	0.03978	0.08144	0.12233	0.03368	0.01778		0.00098+-0.0010
8	0.02349	0.01748	-0.00981	-0.02497	-0.00115	0.09254	0.10215	
9	0.09385	0.05963	-0.00360	0.07582	0.00680	0.11584	0.17067	-0.02828
10	0.19477	0.12696	0.15606	0.22578	0.13737	0.08013	0.13668	0.20484
11	0.01449	0.01544	0.03548	0.14475	0.00053	0.02073	-0.00008	0.12393
12	0.05112	0.02800	0.02550	0.04350	0.00541	0.06834	0.08113	-0.00015
13	0.05134	0.04759	0.08048	0.11516	0.01697	0.01553	0.01928	0.06471
14	0.10411	0.06710	0.02533	0.08167	0.05590	0.10669	0.16321	0.01202
15	0.05323	0.04952	0.00076	0.01905	0.02877	0.10633	0.13419	-0.01136
16	0.09379	0.07514	0.05015	0.02021	0.12911	0.15376	0.17149	0.02139
	<b>Legende</b>				<b>Quelle</b>			
1	Frühe Eisenzeit Ost (9.–7. Jh. v. Chr.); N=11							
2	Eisenzeit Ost (7.–3. Jh. v. Chr.); N=86							
3	Eisenzeit West (8.–3. Jh. v. Chr.); N=33							
4	Übergang Bronze- zu Eisenzeit Zentralasien (9.–8. Jh. v. Chr.); N=14				MOLODIN ET AL. 2012A			
5	Späte Bronzezeit Zentralasien (Ende des 2. Jt. v. Chr.); N=8				MOLODIN ET AL. 2012A, 2012B			
6	Mittlere Bronzezeit Zentralasien (3.–2. Jt. v. Chr.); N=46				MOLODIN ET AL. 2012A			
7	Frühe Bronzezeit Zentralasien (4.–3. Jt. v. Chr.); N=27				MOLODIN ET AL. 2012A			
8	Bronzezeit Chakassien (1800–800 v. Chr.); N=11				KEYSER ET AL. 2009			
9	Bronzezeit Kasachstan (1400–1000 v. Chr.); N=8				LALUEZA-FOX ET AL. 2004			
10	Bronzezeit Tarim Becken China (1980 v. Chr.); N=20				LI ET AL. 2010			
11	Bronzezeit Lajia China (1900 v. Chr.); N=13				GAO ET AL. 2007			
12	Neolithikum Nordwest Russland (5500 v. Chr.); N=11				DER SARKISSIAN ET AL. 2013			
13	Bronzezeit Nordwest Russland (1500 v. Chr.); N=23				DER SARKISSIAN ET AL. 2013			
14	Eneolithikum nördlicher Schwarzmeerraum (4500–3000 v. Chr.); N=10				WILDE ET AL. 2014			
15	Yamnaya; nördlicher Schwarzmeerraum (3000–2500 v. Chr.); N=28				WILDE ET AL. 2014			
16	Katakombengrabkultur; nördl. Schwarzmeerraum (2500–2000 v. Chr.); N=24				WILDE ET AL. 2014			



(Fortsetzung) Tab. 17 Paarweise  $F_{ST}$ -Werte zwischen skythischen, bronzzeitlichen und neolithischen Populationen (links unten) und zugehörige  $p$ -values (rechts oben)

	9	10	11	12	13	14	15	16
1	0.02344+-0.0046	0.00293+-0.0016	0.25098+-0.0119	0.11328+-0.0083	0.07910+-0.0095	0.00586+-0.0022	0.02637+-0.0050	0.00391+-0.0019
2	0.02344+-0.0041	0.00000+-0.0000	0.14648+-0.0115	0.05664+-0.0068	0.00195+-0.0014	0.00684+-0.0027	0.00000+-0.0000	0.00000+-0.0000
3	0.46289+-0.0190	0.00000+-0.0000	0.03613+-0.0060	0.11426+-0.0084	0.00000+-0.0000	0.12207+-0.0091	0.38770+-0.0121	0.00391+-0.0019
4	0.02441+-0.0053	0.00000+-0.0000	0.00000+-0.0000	0.11426+-0.0096	0.00293+-0.0016	0.00684+-0.0023	0.13574+-0.0099	0.16504+-0.0116
5	0.42285+-0.0156	0.01172+-0.0027	0.41504+-0.0203	0.36914+-0.0162	0.25586+-0.0147	0.12695+-0.0098	0.16699+-0.0120	0.00488+-0.0020
6	0.00781+-0.0031	0.00293+-0.0016	0.16797+-0.0097	0.01660+-0.0042	0.12695+-0.0098	0.00781+-0.0024	0.00000+-0.0000	0.00000+-0.0000
7	0.00000+-0.0000	0.00098+-0.0010	0.42383+-0.0199	0.01660+-0.0044	0.18555+-0.0149	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000
8	0.69922+-0.0162	0.00000+-0.0000	0.00098+-0.0010	0.44629+-0.0180	0.03516+-0.0065	0.31152+-0.0136	0.63672+-0.0172	0.18262+-0.0123
9		0.00000+-0.0000	0.00000+-0.0000	0.02344+-0.0036	0.01562+-0.0042	0.41211+-0.0157	0.73633+-0.0155	0.00195+-0.0014
10	0.27006		0.01074+-0.0036	0.00391+-0.0019	0.00781+-0.0031	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000
11	0.20750	0.13804		0.02051+-0.0045	0.34668+-0.0136	0.00000+-0.0000	0.00391+-0.0031	0.00000+-0.0000
12	0.10674	0.14423	0.10778		0.04199+-0.0062	0.00684+-0.0023	0.00977+-0.0029	0.00195+-0.0014
13	0.11759	0.09301	0.00378	0.07425		0.02148+-0.0047	0.00000+-0.0000	0.00000+-0.0000
14	0.00141	0.26720	0.17836	0.13823	0.10620		0.59570+-0.0141	0.06934+-0.0072
15	-0.02284	0.20365	0.09651	0.06922	0.09476	-0.01162		0.03516+-0.0068
16	0.12750	0.27328	0.17902	0.10913	0.14919	0.05812	0.03340	
	<b>Legende</b>				<b>Quelle</b>			
1	Frühe Eisenzeit Ost (9.–7. Jh. v. Chr.); N=11							
2	Eisenzeit Ost (7.–3. Jh. v. Chr.); N=86							
3	Eisenzeit West (8.–3. Jh. v. Chr.); N=33							
4	Übergang Bronze- zu Eisenzeit Zentralasien (9.–8. Jh. v. Chr.); N=14				MOLODIN ET AL. 2012A			
5	Späte Bronzezeit Zentralasien (Ende des 2. Jt. v. Chr.); N=8				MOLODIN ET AL. 2012A, 2012B			
6	Mittlere Bronzezeit Zentralasien (3.–2. Jt. v. Chr.); N=46				MOLODIN ET AL. 2012A			
7	Frühe Bronzezeit Zentralasien (4.–3. Jt. v. Chr.); N=27				MOLODIN ET AL. 2012A			
8	Bronzezeit Chakassien (1800–800 v. Chr.); N=11				KEYSER ET AL. 2009			
9	Bronzezeit Kasachstan (1400–1000 v. Chr.); N=8				LALUEZA-FOX ET AL. 2004			
10	Bronzezeit Tarim Becken China (1980 v. Chr.); N=20				LI ET AL. 2010			
11	Bronzezeit Lajia China (1900 v. Chr.); N=13				GAO ET AL. 2007			
12	Neolithikum Nordwest Russland (5500 v. Chr.); N=11				DER SARKISSIAN ET AL. 2013			
13	Bronzezeit Nordwest Russland (1500 v. Chr.); N=23				DER SARKISSIAN ET AL. 2013			
14	Eneolithikum nördlicher Schwarzmeerraum (4500–3000 v. Chr.); N=10				WILDE ET AL. 2014			
15	Yamnaya; nördlicher Schwarzmeerraum (3000–2500 v. Chr.); N=28				WILDE ET AL. 2014			
16	Katakombengrabkultur; nördl. Schwarzmeerraum (2500–2000 v. Chr.); N=24				WILDE ET AL. 2014			

## Reynold's distances

Matrix of coancestry coefficients as  $t/M = -\ln(1-F_{ST})$

**Tab. 19 Reynold's distances der skythischen Populationen ohne die Initialen Skythen**

	1	2	3	4	5	6
1 Aldy Bel	-0.00000					
2 Pazyryk	0	-0.00000				
3 Zevakino-Čilikta	0	0	-0.00000			
4 TagarTes	0.01126	0.00505	0.01519	-0.00000		
5 Skythen	0	0	0	0	-0.00000	
6 Sarmaten	0.04762	0.04103	0.05261	0.00429	0	-0.00000

**Tab. 20 Reynold's distances der eisenzeitlichen und modernen Populationen Eurasiens**

	1	2	3	4	5	6	7	8	9
1 ancient Tuva	-0.00000								
2 ancient Altai	0	-0.00000							
3 ancient Chakassien	0.01126	0.00505	-0.00000						
4 ancient Ostkasachstan	0	0	0.01519	-0.00000					
5 ancient West-Russland	0.01520	0.01889	0.00770	0.01595	-0.00000				
6 modern Tuva	0.04845	0.05535	0.03748	0.03840	0.10417	-0.00000			
7 modern Altai	0.01411	0.01793	0.01289	0	0.03497	0.02646	-0.00000		
8 modern Chakassien	0.03811	0.03686	0.01403	0.01600	0.05224	0.02086	0.00958	-0.00000	
9 modern SW-Russland	0.07320	0.05960	0.06181	0.04960	0.01287	0.17045	0.08486	0.10777	-0.00000

**Tab. 21 Reynold's distances der Pazyryk-Gruppen**

	1	2	3	4
1 Ukok_Plateau	0.00000			
2 Chuya_Tal	0.03058	0.00000		
3 kasachisches Altai	0	0.06149	0.00000	
4 mongolisches Altai	0	0.03167	0.00440	0.00000

## Frequenzen der abgeleiteten Allele für alle untersuchten Marker

Tab. 22 Frequenzen der abgeleiteten Allele der Skythen und der Vergleichspopulationen für die phänotypischen Marker

Marker	Pop	Freq(Der)	Marker	Pop	Freq(Der)	Marker	Pop	Freq(Der)	Marker	Pop	Freq(Der)		
HERC2	EU (120)	0,79	LCTa	EU (226)	0,73	EDAR	EU (120)	0,00	ADH1Ba	EU (42)	0,00		
	WSK (10)	0,60		WSK (8)	0,00		ZA (454)	0,76		WSK (8)	0,00		
	ZA (596)	0,06		ZA (226)	0,04		OSK (22)	0,55		ZA (570)	0,25		
	OSK (82)	0,30		OSK (82)	0,09		AS (80)	0,93		OSK (70)	0,01		
	AS (90)	0,00		AS (90)	0,00		FO (100)	0,85		AS (46)	0,67		
	FO (100)	0,09		FO (172)	0,07		ABCC11	EU (224)		0,13	FO (94)	0,17	
SLC24A5	EU (116)	1,00	LCTb	EU (170)	0,71	ABCC11	WSK (10)	0,00	ADH1Bb	EU (120)	0,00		
	WSK (4)	1,00		WSK (6)	0,00		ZA (714)	0,78		ZA (918)	0,40		
	ZA (534)	0,31		ZA (40)	0,10		OSK (80)	0,41		OSK (22)	0,00		
	OSK (24)	0,75		OSK (82)	0,11		AS (86)	0,98		AS (90)	0,77		
	AS (86)	0,01		AS (194)	0,00		FO (102)	0,93		FO (362)	0,04		
	FO (100)	0,21		FO (50)	0,06		CYP3A5	EU (222)		0,96	NAT2	EU (120)	0,44
SLC45A2	EU (116)	0,98	CYP3A4	EU (44)	0,98	CYP3A5	WSK (8)	1,00	NAT2	WSK (8)	0,25		
	WSK (12)	1,00		WSK (4)	1,00		ZA (40)	0,80		OSK (86)	0,26		
	ZA (578)	0,13		OSK (32)	1,00		OSK (62)	0,95		AS (90)	0,02		
	OSK (60)	0,57		AS (48)	1,00		AS (86)	0,66		AS (90)	0,02		
	AS (88)	0,01	SLC12A3	EU (120)	0,08	FO (50)	0,90	ABCB1a	EU (226)	0,45			
	FO (100)	0,16		OSK (42)	0,00	TRPV6b	EU (120)		0,94	WSK (4)	0,50		
TYR	EU (226)	0,43	AS (90)	0,00	WSK (10)		1,00		OSK (72)	0,38			
	WSK (8)	0,13	TRPV6c	EU (226)	0,92		ZA (40)	1,00	AS (86)	0,71			
	ZA (40)	0,13		WSK (6)	0,83		OSK (78)	1,00	ABCB1b	EU (226)	0,47		
	OSK (80)	0,16		OSK (54)	0,98		AS (88)	1,00		WSK (10)	0,50		
	AS (90)	0,00		FO (50)	1,00		FO (50)	1,00		OSK (66)	0,35		
	FO (100)	0,06		AS (86)	0,99	TRPV6c	EU (226)	0,92	AS (86)	0,62			
TYR	EU (226)	0,43		TRPV6c	WSK (6)		0,83	EU (226)	0,92	ABCB1c	EU (226)	0,57	
	WSK (8)	0,13	OSK (54)		0,98		WSK (8)	0,63	WSK (8)		0,63		
	ZA (40)	0,13	AS (86)		0,99		OSK (78)	0,49	OSK (78)		0,49		
	OSK (80)	0,16	TRPV6c		AS (86)		0,99	AS (84)	0,42	AS (84)	0,42		
	AS (90)	0,00			TRPV6c		AS (86)	0,99	CASP12	EU (120)	1,00	CASP12	EU (120)
	FO (100)	0,06				TRPV6c	AS (86)	0,99		OSK (22)	1,00		OSK (22)
		TRPV6c	AS (86)	0,99			AS (86)	0,99		AS (86)	0,99		

