

Selektive Anreicherung nukleärer humaner DNA für die populationsgenetische Analyse bronze - und eisenzeitlicher Reiternomaden des eurasischen Steppengürtels

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften

Am Fachbereich Biologie
der Johannes Gutenberg-Universität Mainz

vorgelegt von

Melanie Groß
aus Heilbronn
geb. am 15.07.1983

Mainz 2019

Inhaltsverzeichnis

Zusammenfassung	iv
Summary	v
1 Einleitung.....	- 1 -
1.1 Ziele und Relevanz der Arbeit	- 2 -
1.2 Stand der Forschung	- 3 -
1.3 Forschungsfragen zur Populationshistorie skythischer Reiterverbände Eurasiens in der Eisenzeit	- 14 -
1.4 Methodische Ziele	- 17 -
2 Skelettmaterial und Methoden	- 19 -
2.1 Probenbearbeitung	- 19 -
2.1.1 Probenauswahl und Extraktion.....	- 20 -
2.1.2 <i>library</i> -Erstellung	- 22 -
2.1.3 PCR	- 23 -
2.1.4 Quantifizierung und Qualitätsbestimmung.....	- 25 -
2.2 Selektive Anreicherung nukleärer DNA	- 27 -
2.2.1 Vergleich verschiedener Anreicherungsverfahren.....	- 30 -
2.2.2 Neutrale Regionen	- 31 -
2.2.3 Funktionale Marker.....	- 31 -
2.2.4 <i>bait</i> -Design	- 32 -
2.2.5 Protokolle	- 35 -
2.3 NGS-Sequenzierung.....	- 37 -
2.4 Bioinformatische Datenauswertung	- 38 -
2.4.1 Rohdatenbearbeitung	- 38 -
2.4.2 Degradierungsmuster und Geschlechtsbestimmung.....	- 40 -
2.4.3 Kontaminationsbestimmung.....	- 40 -
2.4.4 Genotypisierung – ATLAS.....	- 41 -
3 Populationsgenetik.....	- 43 -
3.1 Hauptkomponentenanalyse – PCA.....	- 43 -
3.2 FST und F2	- 44 -
3.3 F3 und D-Statistik	- 45 -
3.4 Funktionale Marker.....	- 48 -
3.5 Nukleotiddiversität.....	- 54 -
4 Ergebnisse.....	- 55 -
4.1 Finale Probenauswahl für die Anreicherung.....	- 55 -
4.2 Wahl der Anreicherungsverfahren.....	- 60 -
4.3 Anreicherungsverfahren.....	- 63 -
4.4 Authentizität der Proben.....	- 66 -
4.5 Phänotypenbestimmung.....	- 68 -

4.6	Genetische Strukturen skythischer Populationen.....	- 79 -
5.	Diskussion.....	- 106 -
5.1	Effizienz der selektiven Anreicherung nukleärer DNA und Diskussion der finalen Probenauswahl	- 106 -
5.2	Pigmentierungsmuster und funktionale Marker	- 111 -
5.3	Populationsgenetik eisen- und bronzezeitlicher Steppenpopulationen Eurasiens.....	- 117 -
5.3.1	Genetische Kontinuität zwischen bronze- oder eisenzeitlichen Kulturgruppen und lokalen Populationen des heutigen Eurasiens.....	- 118 -
5.3.2	Homogenität der eisenzeitlichen Individuen Zentral – und Nord-Asiens zu den Individuen des Schwarzmeer-Raums und SW-Russlands	- 124 -
5.3.3	Das Verhältnis zu neolithischen und bronzezeitlichen Populationen der eurasischen Steppe.....	- 126 -
5.3.4	Ostasiatischer Einfluss auf nomadische Populationen im Übergang von der Bronze- zur Eisenzeit	- 135 -
5.3.5	Multiregionaler Ursprung eisenzeitlicher Steppenomaden-	138 -
	Literaturverzeichnis	- 143 -
	Anhang A	- 152 -
	Anhang B	- 154 -
	Tabellenverzeichnis	- 183 -
	Abbildungsverzeichnis	- 185 -
	Inhalt Daten-DVD	190

Zusammenfassung

In der vorliegenden Arbeit werden die Verhältnisse zwischen eisenzeitlichen reiternomadischen Kulturgruppen Eurasiens rekonstruiert. Für die Analyse wurden 34 Individuen (9. -2. Jh. v. Chr.) des eurasischen Steppenraums Zentralasiens und Südwest-Russlands ausgewählt. Zusätzlich konnten fünf bronzezeitliche Individuen aus dem westlichen Süd-Ural und aus der Wolga-Region, sowie zwei neolithische Individuen aus Novosibirsk analysiert werden. Mit Hilfe eines Anreicherungsverfahrens (*capture*) auf DNA-Bibliotheken wurden genomische Daten, bestehend aus neutralen Regionen und definierten SNP-Positionen, gezielt isoliert und sequenziert. Für die optimale selektive Anreicherung wurde zunächst ein Anreicherungsverfahren unter der Berücksichtigung der spezifischen Anforderungen von prähistorischem Skelettmaterial getestet und optimiert. Nach einer Hochdurchsatzsequenzierung und anschließender bioinformatischer Datenauswertung konnte eine durchschnittliche Abdeckung aller Zielregionen von 94,64 % und einer 5-80fachen Abdeckungstiefe erreicht werden. Mit dem erstellten Datensatz kann nachgewiesen werden, dass sich die Populationshistorie skythischer Reiternomaden Zentralasiens für die jeweiligen Kulturgruppen signifikant unterscheidet. Basierend auf den durchgeführten populationsgenetischen Analysen konnte eine Heterogenität der innerasiatischen Skythen identifiziert werden. Individuen des kasachischen Altai (Berel), des Ukok-Plateaus (Ak-Alacha) und des Chuya-Tals (Justyd) stellen keine homogene Gruppe dar. Für einzelne Individuen der Aldy-Bel-Kultur (Region Tuva) und Pazyryk-Kultur konnten deutliche ost-asiatische Bezüge nachgewiesen werden. Eine Sonderstellung innerhalb der kasachischen und sibirischen Skythen weisen Vertreter der Tagar-Kultur aus dem Minusinsker Becken auf. Die Distanz zu modernen zentral-asiatischen Populationen lässt auf eine isolierte Population mit geringem Genfluss zu lokalen Populationen schließen. Eine direkte Verwandtschaft mit Vertretern der zeitlich folgenden Pazyryk-Kultur kann ebenfalls nicht angenommen werden. Ein Vertreter der Saken aus Südost-Kasachstan belegt die Verbindung südlicher Steppennomaden nach Südwest-Asien. Für die frühen Sarmaten der Orenburg-Region ist von einem kontinuierlichen Genfluss mit östlichen Skythen Zentralasiens auszugehen, jedoch mit starken Einflüssen lokaler Populationen, was auf eine genetische Kontinuität im Bereich des südlichen Urals schließen lässt. Die kaukasische und osteuropäische Jäger-Sammler-Komponente in eisenzeitlichen Skythen Zentralasiens lässt auf eine Verwandtschaft mit Jamnaja-verwandten Populationen schließen. Aufgrund der hier durchgeführten Analysen ist von einem europäischen Ursprung der bronzezeitlichen Sintashta-Kultur auszugehen.

Summary

This thesis aims to investigate the relationships between the various cultural groups of Iron Age Eurasian nomads. 34 individuals originating from the Eurasian steppe and Southwestern Russia were selected for the analysis. In addition, five Bronze Age individuals from the western part of the Southern Ural means of an enrichment of DNA libraries (capture), target genomic data consisting of neutral regions and defined SNP positions were isolated and sequenced. To ensure optimum selective enrichment, an enrichment technique was initially tested and optimised, taking into consideration the specific requirements of prehistoric skeletal material. After high-throughput sequencing and subsequent bioinformatic data analysis, a mean target region coverage of 94.64% and a 5-80-fold sequencing depth was achieved. With the data obtained, it could be demonstrated that the population history of the Scythian horse nomads of Central Asia differs significantly between the various cultural groups. Based on the population genetic analyses performed for this thesis, heterogeneity was ascertained for the Central Asian Scythians. Individuals originating from Altai (Berel), the Ukok Plateau (Ak-Alacha) and the Chuya Valley (Justyd) do not represent a homogeneous group. Some individuals from the Aldy-Bel and Pazyryk cultures showed clear affinity to East Asia. Representatives of the Tagar culture from the Minusinsk Basin occupy a distinct position within the Kazakh and Siberian Scythians groups. The distance to modern Central Asian populations indicates an isolated group with a low level of gene flow to local populations. Similarly, direct affinity with representatives of the chronologically subsequent Pazyryk culture cannot be established. One representative of the Saka from South-eastern Kazakhstan documents the affinity of southern Steppe nomads with Southwestern Asia. With regard to the early Sarmatians from the Orenburg region, a continuous gene flow with the eastern Scythians of Central Asia must be assumed; however, with a strong influence through local populations, indicating a genetic continuity in the Southern Ural area. The Caucasian and Eastern European hunter-gatherer component in all Iron Age Scythians of Central Asia suggests genetic affinity to Jamnaja-related populations.

The population genetic analysis applied here indicate a European origin of the Bronze Age Sintasha culture.

1 Einleitung

Dank der Überlieferung antiker Schriftquellen Herodots und den spektakulären Funden in sibirischen Grabhügeln (Kurgane) aus Tuva wissen wir von den Reiternomaden der eurasischen Steppe, die sich seit der späten Bronzezeit zwischen dem Nordschwarzmeerraum bis ins Pamirgebirge verbreitet hatten. Ihre kriegerische Lebensweise, die beeindruckenden Goldfunde und eine für die Skythen typische Kunstform üben eine große Faszination auf den Betrachter aus. Wer waren diese Steppenreiter, die solch ein reiches Erbe hinterließen und im Laufe des 1. Jahrtausends die Geschichte Eurasiens wesentlich geprägt haben?

Trotz aller archäologischen Erkenntnisse und Studien der Populationsgenetik sind Fragen zur Verbreitung skythischer Kulturen und deren Verwandtschaft miteinander nach wie vor nicht eindeutig geklärt. Zahlreiche genetische Studien befassen sich mit der eurasischen Steppe als elementare Verbindung zwischen Westurasien und Nordost-Asien und den möglichen Migrationsbewegungen während der Bronze - und Eisenzeit. (Ricaud et al. 2004; Lalueza-Fox et al. 2004; Keyser et al. 2009; González-Ruiz et al. 2012; Haak et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015; Juras et al. 2017; Unterländer et al. 2017; Wang et al. 2018; Jeong et al. 2018).

Neueste Methoden der statistischen Genetik und paläogenetische Methoden erlauben es sich diesen Fragen in ihrer ganzen Komplexität zu widmen.

Im Zuge der neu publizierten Studien mit hoher Individuenzahl und genomischen Datensätzen ist es möglich, die Fragen zur Populationshistorie der Steppenreiter weitaus umfangreicher zu untersuchen als bisher. Mit statistischen Methoden wie beispielsweise der LASSER projizierte Hauptkomponenten-Analyse (Wang et al. 2014) im Vergleich mit weltweiten modernen Datensätzen (Hellenthal et al. 2014; Busby et al. 2015) und F-Statistiken (Patterson et al. 2012) mit einer Vielzahl an zeitgleichen Referenzindividuen können Fragen zu Verwandtschaftsverhältnissen von Kulturgruppen und möglichen Ursprungspopulationen präziser betrachtet werden.

In der vorliegenden Arbeit werden die Verhältnisse zwischen reiternomadischen Kulturgruppen Eurasiens rekonstruiert. Dazu werden mit Hilfe eines Anreicherungsverfahrens (*capture* oder *target enrichment*) auf DNA-Bibliotheken genomische Daten gezielt isoliert und sequenziert, um vergleichbare Positionen zwischen verschiedenen Individuen mit hoher Abdeckung (*coverage*) zu erzeugen.

Für die Analyse wurden Individuen des eurasischen Steppenraums Zentralasiens und Südwest-Russlands ausgewählt. Der überwiegende Teil analysierter Individuen konzentriert

sich auf die Region um das Altai-Gebirge. Dieses bildet eine natürliche Grenze zwischen Ost- und West-Eurasien und stellt daher einen wichtigen Raum dar, um Entwicklungen in Zentralasien nachvollziehen zu können. Frühe Sarmaten aus Pokrovka der Orenburg-Region, Skythen der Woronesch-Region (Kolbino) und Novozavedennoe im Nord-Kaukasus stehen stellvertretend für die westlichen Reiternomaden des eurasischen Steppengürtels.

Die Fundplätze der bronzezeitlichen Proben liegen mit Kamennyj Ambar im westlichen Süd-Ural und Kalinovka in der Wolga-Region. Sie sind der Sintaschta-Kultur bzw. Poltavka-Kultur zugeordnet. Für die Anreicherung konnten ebenfalls zwei neolithische Proben aus Novosibirsk in West-Sibirien gewonnen werden. Anhand der bronzezeitlichen Individuen kann eine mögliche Verbindung der skythischen Verbände aus dem zentralasiatischen Raum zu westlich lokalisierten vorangegangenen Kulturgruppen untersucht werden. Im Gegenzug erlauben weiter im Osten liegende, neolithische Funde eine Nähe der untersuchten Proben des Altai-Gebirges zu weitaus älteren Funden nachzuweisen.

Aufgrund der geringen Individuenzahl bronzezeitlicher Proben liegt der Schwerpunkt der Fragestellungen auf Entwicklungen während der Eisenzeit. Um Rückschlüsse auf Bevölkerungsbewegungen und mögliche Ursprungsgebiete ziehen zu können, ist der Vergleich mit Populationen vorangegangener Kulturgruppen jedoch entscheidend.

1.1 Ziele und Relevanz der Arbeit

Für die optimale selektive Anreicherung wird zunächst ein Anreicherungsverfahren der drei führenden Hersteller (MYbait s® von MYcroarray, NimbleGen von Roche und SureSelect^{XT} von Agilent) getestet, das in einem flüssigen Hybridisierungsansatz (*in-solution*) durchgeführt wird. Ziel ist es, am Ende der Testreihe ein Anreicherungs-kit auszuwählen, das in seiner Flexibilität und Effizienz optimal auf prähistorisches Probenmaterial angewendet werden kann. Durch das *capture*-Verfahren können die ausgewählten, neutralen Regionen sowie gezielte Markerpositionen mit einer hohen Abdeckung und Sequenzierentiefe in der anschließenden populationsgenetischen Analyse eingesetzt werden.

Die dann für diese weiterführende Untersuchung ausgewählten Proben verteilen sich über die zentralasiatische Steppe bis nach Südwest-Russland und umfassen verschiedene Gruppierungen, die mit der skythischen Kultur assoziiert werden. Im Zuge neuester Studien wird ein multiregionaler Ursprung der skythischen Reiternomaden diskutiert (Unterländer

et al. 2017; Damgaard et al. 2018; Narasimhan et al. 2018; Krzewinska et al. 2018). Eine wichtige Frage dabei ist das Ausmaß eines möglichen Genflusses zwischen west- und ost-eurasischen Gruppen der skythischen Kultur. Anhand von Vertretern der Tagar-, Pazyryk- und Aldy-Bel-Kultur, Saken, Sarmaten und frühen Skythen, die einen Zeitraum vom 9. Jhd. v. Chr – 2. Jhd. v. Chr. umfassen, kann die Verbindung der geografisch naheliegenden Kulturgruppen Südsibiriens untersucht sowie ihre Nähe zu Individuen Südwest-Russlands analysiert werden. Die vorliegende Arbeit liefert einen wichtigen Beitrag zur aktuellen Diskussion um einen polyzentristischen Ursprung skythischer Reitergruppen und die Verbindung der östlichen mit den westlich gelegenen Populationen über die Zeit. Des Weiteren konnte in aktuellen Publikationen nachgewiesen werden, dass sich die Populationshistorie signifikant unterscheidet, je nachdem welche kulturellen Gruppierungen im Einzelnen betrachtet werden. Die Bedeutung der Saken und der Tagar-Kultur im Kontext skythischer Populationsgeschichte wird in der vorliegenden Arbeit herausgearbeitet und ausführlich diskutiert.

1.2 Stand der Forschung

Migrationsbewegungen und Besiedlungsgeschichte der bronze- und eisenzeitlichen Steppe Eurasiens

Von der frühen Bronzezeit bis zur späten Eisenzeit verbinden nomadisch und semi-nomadisch lebende, reiternomadische Völker aufgrund ihrer hohen (Wang et al. 2018) Mobilität Europa und Asien. Das Verbreitungsgebiet dieser eng verbundenen Kulturgruppen erstreckt sich vom nordpontischen Raum bis ins Altai-Gebirge Sibiriens über ein Gebiet von 2000 – 3500 km. Welche Dimensionen die Migrationsbewegungen zwischen Europa und Asien über den zentral-asiatischen Raum angenommen haben, zeigen Mumien-Funde aus dem Tarim-Becken aus Xinjiang in West-China die deutlich kaukasische Merkmale aufweisen (J. P. Mallory 2000, S.237).

Die genetische Beziehung dieser kulturell verbundenen Gruppen und im Speziellen der Ursprung der als Skythen benannten Gruppen wird bis heute seitens der Archäologie und prähistorischen Populationsgenetik diskutiert (Allentoft et al. 2015; Juras et al. 2017; Unterländer et al. 2017; Damgaard et al. 2018; Narasimhan et al. 2018; Krzewinska et al. 2018). Um Migrationsbewegungen eisenzeitlicher Reiternomaden der eurasischen Steppe

nachzuvollziehen, ist ein Verständnis der kulturellen Entwicklungen in der Bronzezeit obligatorisch.

Die Bronzezeit wird aufgrund jeweils prägender Kulturentwicklungen in drei zeitliche Perioden unterteilt: die Früh-, Mittel-, und Späte Periode (Parzinger 2006, S.245ff). Ungefähr 3000 v. Chr. tritt die früh-bronzezeitliche Jamnaja-Kultur von der pontischen Steppe ausgehend größtenteils an die Stelle der europäischen Jäger-Sammler-Kulturen (JS) in Osteuropa. Im Altai-Gebirge und dem Minusinsker-Becken ist zur selben Zeit bereits die Afanasievo-Kultur verbreitet, deren archäologische Funde eine starke Verbindung zu Jamnaja-Gesellschaften aufweisen. Durch diese sehr hohe genetische Ähnlichkeit zur früh-bronzezeitlichen Afanasievo-Kultur im Minusinsker-Becken wird daher ebenso von einer Ostwärts-Bewegung der Jamnaja-Kultur ausgegangen (Allentoft et al. 2015). In West- und Zentralasien sind Jäger-Sammler-Kulturen noch für die frühe Bronzezeit nachweisbar. Die Sintascha-Kultur der Ural-Tobol-Steppe entsteht aus der Poltavka-Kultur und ist gekennzeichnet durch den Bau von Streitwagen und intensiver Pferdezucht und breitet sich von Europa Richtung Asien aus. Während in der frühen Bronzezeit noch kleinere Kulturverbände vorherrschen (Parzinger 2006), entwickeln sich in der Mittleren Bronzezeit größere Verbände, vor allem der Andronovo-Fedorovka-Kreis, erstreckt sich westlich des Urals bis zum Jenissei im Osten über ein großes Verbreitungsgebiet. Vertreter dieser Kultur expandieren von West- und Südsibirien in benachbarte Gebiete. Ihr Ursprung wird möglicherweise in Zentral-Kasachstan vermutet (Molodin et al. 2012). Migrationsbewegungen über Zentralasien in der Bronzezeit werden häufig basierend auf archäologischen Funden mit der Verbreitung der Andronovo-Kultur in Verbindung gebracht (Parzinger 2006). Die der Andronovo-Kultur nahestehende Srubnaja-Kultur entsteht koinzident und breitet sich vom nördlichen Schwarzmeerraum bis zur Nordgrenze des Kaspischen Meeres aus. Die Wolga-Ural-Region und die Wüstengebiete südlich des Aralsees scheinen Bereiche zu sein, wo Individuen der Srubnaja-Kultur und Vertreter der Andronovo-Kultur erstmals aufeinandertrafen (Bashilov, V.A. & Yablonsky 2000).

Im Laufe der mittleren bis späten Bronzezeit werden Vertreter der Andronovo-Kultur von lokalen Gruppierungen, wie der Karasuk-Kultur, abgelöst (Abb. 1). Die Karasuk-Kultur wird der späten Bronzezeit zugeordnet und folgt der Andronovo-Kultur in Südsibirien (Abb. 1). Als ihr Verbreitungsgebiet gilt die Region Tuva, das mittlere Jenissei-Gebiet bis zur Mongolei. Die eisenzeitliche Tagar-Kultur löst diese Verbände in der folgenden Zeit ab.

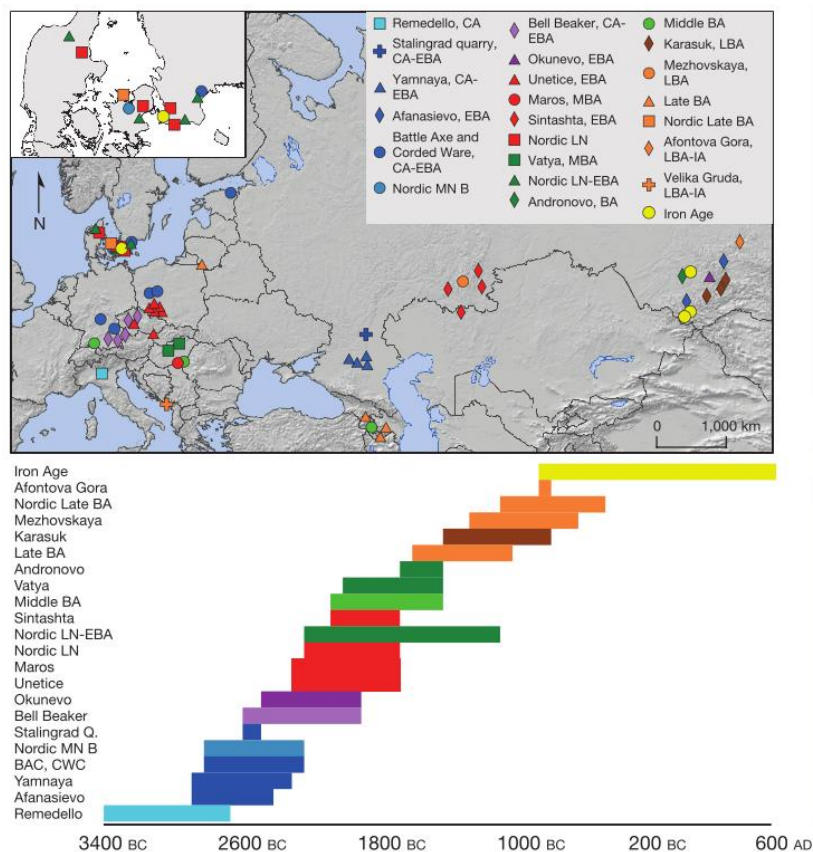


Abbildung 1: Geografische Verteilung und Zeitstrahl der eisen – und bronzezeitlichen Kulturguppen aus (Allentoft et al. 2015)

Eine der bekanntesten Gruppen der **Eisenzeit**, die Skythen, wird in Schriften Herodots erstmals erwähnt (Parzinger 2007). Die Namensgebung basiert auf Überlieferungen der Griechen, die diese Bezeichnung einer Vielzahl von nomadischen lebenden Reitervölkern, die von der nordpontischen Steppe bis Zentralasien verbreitet waren, gegeben hatten. Verbände des Nordschwarzmeerraums bezeichnet Herodot als Skythen, die weiter östlich beheimateten Gruppierungen als Sauromaten. Der Begriff der Saken wurde von den Persern geprägt, die damit Reiterstämme Mittelasiens mit einer den Skythen nahestehenden Sachkultur bezeichneten (Parzinger 2006, S. 542) (Abb. 1). Vom Übergang der Spätbronze bis zur Früheisenzeit begannen sich Merkmale einer skythischen Kultur herauszubilden. Mobile Reiterverbände verbreiteten sich in großen Teilen des eurasischen Steppenraums (Abb. 2) und es können die Anfänge des Reiterkriegernomadentums beobachtet werden (Parzinger 2006, S. 541ff). Dieser kulturelle Wandel und die Entstehung skythischer Gruppierungen ist gekennzeichnet durch besondere Lebens- und Wirtschaftsformen, dem charakteristischen skytho-sibirischen Tierstil, kunstvollen Goldobjekten als Grabbeigaben und monumentalen Grabhügeln (Kurgane) als verbindende Kulturelemente der Massageten,

Sauromaten, Saken und Skythen. Diese eisenzeitlichen Reiternomaden sind Teil der „*scytho-sibirischen Welt*“ mit ihrer skythisch geprägte Sachkultur als verbindendes Element, zeigen jedoch eine Vielzahl lokaler Variationen entsprechend ihrer geografischen Verbereitung. Lebens - und Wirtschaftsweise verändern sich mit dem Beginn der Skythenzeit erheblich und Vorläufer dieses Strukturwandels führen bis nach Tuva in Südsibirien an den östlichen Rand der eurasischen Steppe (Parzinger 2007).

Die ältere Eisenzeit vom 9./8. bis 3./2. Jh. v. Chr. ist durch die skythische Kultur zwischen der Pontischen Steppe und Kaspischem Meer geprägt. Im frühen 7. Jh. v. Chr. können Skythen erstmals im Nordschwarzmeerraum nachgewiesen werden und wenige Jahrzehnte später belegen Funde einer „skythisch geprägten Sachkultur“ in Ungarn und der Slowakei den westlichsten Rand ihres Verbreitungsgebietes. Während die Skythen die Kimmerer aus dem nordpontischen Raum vertreiben, führen ihre Kriegszüge durch den Kaukasus bis an die nordwestliche Grenze Vorderasiens im Iran. Im frühen 6. Jh. v. Chr. bewegen sich skythische Reitergruppen wieder Richtung Westen in den Nordschwarzmeerraum zurück.

Antike Schriftquellen legen nahe, dass der osteurasische Raum bis nach Sibirien für die Bevölkerungsverschiebungen der skythischen Periode eine wesentliche Rolle spielte. Die Verbindung des Nordschwarzmeerraums mit dem Gebiet östlich des Urals bis nach Tuva blieb nach Aussagen Herodots kontinuierlich bestehen. In Südsibirien können die ältesten Funde skythisch geprägter Kultur nachgewiesen werden (Parzinger 2007).

Der Skythenbegriff bezeichnet früheisenzeitliche Reiterkriegernomaden und ist nach Parzinger nicht im ethnischen Sinne zu verstehen, sondern dient der chronologischen Einordnung verschiedenster Steppenvölker und deren kultureller Zuordnung (Parzinger 2007).

Die Jüngere Eisenzeit beginnt mit der sogenannten hunno-sarmatischen Periode (Parzinger 2006, S. 542). Mit der Verbreitung der Wusun in Kasachstan und den Sarmaten im Nordschwarzmeerraum finden prägende Bevölkerungsverschiebungen in Inner-Asien statt. Es wird angenommen, dass gegen Ende des 3.Jhs. v. Chr. frühe Reiternomaden wie Skythen, Sauromaten und Saken durch Vertreter dieser hunno-sarmatischen Periode abgelöst wurden (Parzinger 2004).

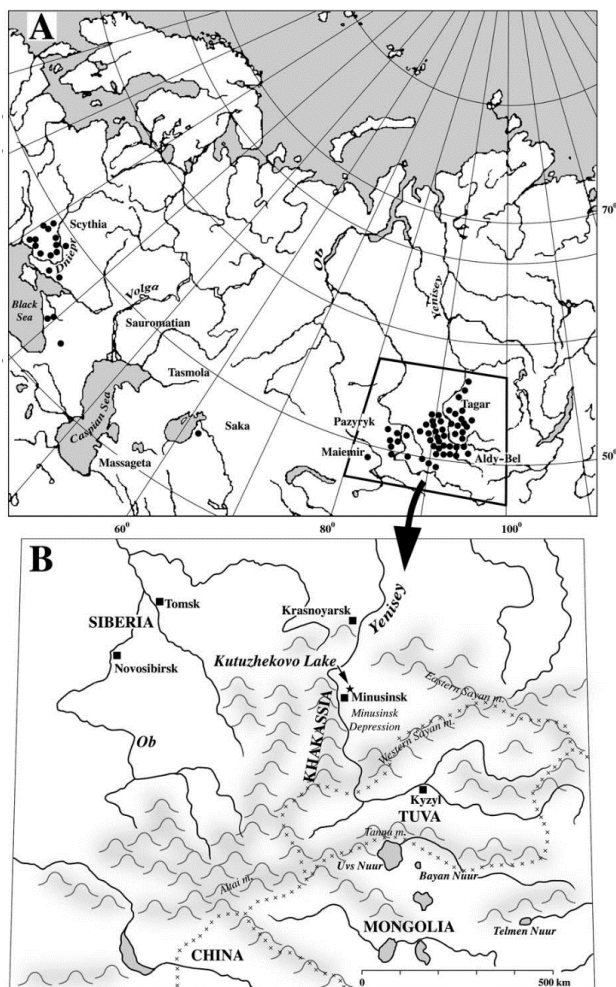


Abbildung 2: A) Skythische Kulturgruppen und ihre Verbreitung in Eurasien B) Detaillierte Ansicht der östlichen Fundplätze, Karte aus (van Geel et al. 2004)

Die Frage nach dem Ursprungsgebiet skythischer Kultur und die genetische Verbindung verschiedener Gruppen, die mit dem Skythentum assoziiert werden, wird bis heute diskutiert (Juras et al. 2017; Unterländer et al. 2017; Damgaard et al. 2018; Krzewinska et al. 2018). Seitens der Archäologie zeigen sich charakteristische Kulturelemente in Funden des Nordschwarzmeerraums ebenso wie im Norden Chinas des späten 9. Jahrhundert v. Chr. Die ost-asiatischen Funde gelten als die ältesten bislang bekannten skythischen Funde Eurasiens (Parzinger 2002).

Zunächst wird ein Ursprung im Nordschwarzmeerraum aufgrund archäologischer Funde und eine genetische Verbindung zu vorangegangenen, lokalen, bronzezeitlichen Individuen angenommen (Bashilov, V.A. & Yablonsky 2000; Parzinger 2006). Vieles deutet jedoch auf einen möglichen Ursprung östlich des Ural-Gebirges hin. Die Hypothese eines asiatischen bzw. sibirischen Ursprungs basiert auf der Ausgrabung des Kurgan Aržan 1 im russisch-

mongolischen Grenzgebiet in Tuva, der als der älteste Kurgan skythischer Kultur gilt und Teil der größten skythischen Grabnekropolen Eurasiens ist. Dieser Fund bekräftigt die Annahme, dass skythische Kulturgruppen weitaus früher östlich des Schwarzen Meeres verbreitet waren. Charakteristische Elemente des skythischen Tierstils konnten zudem nicht in bronzezeitlichen Gruppen im Schwarzmeerraum nachgewiesen werden, jedoch für Nomaden entlang des Jenissei der späten Bronzezeit (Bashilov, V.A. & Yablonsky 2000).

Im Zuge neuester Erkenntnisse eines multiregionalen Ursprungs skythischer Reiternomaden (Unterländer et al. 2017) ist die auf archäologischen Funden basierende dritte Hypothese des skythischen Ursprungs besonders interessant: Archäologen postulieren demnach eine polyzentristische Herkunft skythisch geprägter Kulturen. Durch den kontinuierlichen Austausch benachbarter, nomadischer Gruppen über ihr Verbreitungsgebiet hinaus konnten sich Formen des Tierstils unabhängig entwickelt haben und ausgetauscht worden sein (Bashilov, V.A. & Yablonsky 2000)

Studien basierend auf mitochondrialen Daten

Lange basierten alle genetischen Studien ausschließlich auf autosomalen STR-Analysen und der Bestimmung mitochondrialer oder Y-chromosomaler Haplogruppen.

Ein Großteil der Studien, die im Folgenden vorgestellt werden, konzentriert sich auf Funde des Altai-Gebirges und beruht zumeist auf einer geringen Individuenzahl.

Zusammenfassend lässt sich jedoch feststellen, dass für eisenzeitliche Funde ein nahezu ausgeglichenes Verhältnis an west- und osteurasischen Linien bestimmt werden konnte. So zeigt die Analyse von 36 Individuen aus Kasachstan, von der Bronze- bis zur frühen Eisenzeit bis zum 7. Jh. v. Chr., ausschließlich europäische Linien in Kasachstan (Lalueza-Fox et al. 2004). Eine Aufteilung der Haplotypen in westeurasische und osteurasische Linien tritt dann in der Eisenzeit auf.

Für eisenzeitliche Proben aus dem kasachischen Altai konnten eine der beiden typisierten Proben eher einem asiatischen Haplotyp zugeordnet werden (Clisson et al. 2002). Die Haplotypen U5a1 und HV2 dreier Individuen der Pazyryk-Kultur aus dem Altai-Gebirge der nordwestlichen Mongolei zeigen westeurasische Linien und unterstützen die These eines Austausches zwischen dem Altai und angrenzenden Gebieten in der frühen Eisenzeit (Pilipenko et al. 2010). Weitere Studien zu Individuen aus dem russischen, kasachischen und mongolischen Altai weisen ebenfalls west- sowie osteurasische Linien in der Eisenzeit nach (Clisson et al. 2002; Ricaut et al. 2004; Chikisheva et al. 2007; González-Ruiz et al. 2012).

Das Verhältnis von west- zu osteurasischen Linien scheint deutlich mit dem Übergang der Bronze- zur Eisenzeit in Verbindung zu stehen. Aus dem Krasnojarsk-Gebiet Süd - und Zentral-Sibiriens wurden 32 Individuen der Andronovo, Karasuk, Tagar und Tachtyk-Kultur analysiert. Hier liegt der Anteil an nachgewiesenen, westeurasischen Linien in der Bronzezeit bei 90 % und sinkt auf unter 70 % in der folgenden Eisenzeit, was für einen deutlichen Anstieg ost-asiatischer Linien nach dem 7. Jhd. v. Chr spricht. (Keyser et al. 2009).

Ein Fehlen osteurasischer Linien in der Bronzezeit Kasachstans verbunden mit einem sprunghaften Anstieg in der Eisenzeit (Lalueza-Fox et al. 2004) wurde ebenfalls in einer Studie von Comas und Kollegen für das Gorny-Altai festgestellt (Comas et al. 2004).

In Bezug auf die Ergebnisse der Bronzezeit konnte dies jedoch in einer späteren Studie von Moldin und Kollegen widerlegt werden, die bereits osteurasische Linien in der westsibirischen Waldsteppe der frühen bis mittleren Bronzezeit nachweisen konnten (Molodin et al. 2012). Für 19 Individuen der eisenzeitlichen Pazyryk-Kultur und drei bronzezeitliche Individuen des mongolischen Altai wurden hingegen westeurasische Linien erst mit dem Übergang zur Eisenzeit nachgewiesen (González-Ruiz et al. 2012) im Gegensatz zur westlichen Seite des Altais (Lalueza-Fox et al. 2004; Chikisheva et al. 2007).

Aus dem Rostov-on-Don Gebiets in Russland wurden 16 Skythen auf mitochondrialer Ebene analysiert und zeigen zu 31% ost-asiatische Linien und eine höhere Vielfalt ost-asiatischer Haplotypen im Vergleich zu den vorhergehenden Epochen in diesem Bereich. Die Autorin schließt auf einen genetischen Einfluss auf skythische Verbände aus Sibirien und SW-Asien, der durch eine Westwärts-Bewegung aus Osteurasien, während oder vor der Eisenzeit, entstanden sein könnte (Sarkissian 2011). Wohingegen ein signifikanter Zustrom an westeurasischen Linien nach Sibirien vom Beginn der Bronzezeit für eine Ostwärts-Bewegung Kurgan assoziierter Kulturgruppen aus Europa spricht (Keyser et al. 2009; González-Ruiz et al. 2012).

Der größte Datensatz Haplotypen-basierter, mitochondrialer Daten von 96 Repräsentanten skythischer Kultur der zentralasiatischen Steppe der Eisenzeit wurde von Unterländer und Kollegen publiziert (Unterländer et al. 2017). Basierend auf den mitochondrialen Daten konnte nur eine geringe genetische Distanz der östlichen Skythen untereinander festgestellt werden, ebenso zwischen der westlichen und östlichen Gruppe der Skythen. Die östlichen Skythen zeigen eine nahezu ausgeglichene Verteilung heutiger mitochondrialer Linien West- und Ost-Eurasiens und prähistorische Jäger-Sammler-Linien bereits in der frühen Eisenzeit. Der Anteil osteurasischer Linien ist bei den West-Skythen deutlich geringer bzw. konnte für Individuen der früh-skythischen Phase nicht detektiert werden. Der Nachweis

osteurasischer Linien in der Eisenzeit der Westlichen Skythen indiziert einen Genfluss bzw. Migrationsbewegungen aus Ost-Asien bis in die Don-Volga-Region Südrusslands.

Das erste vollständige mitochondriale Genom einer skythischen Probe des südlichen Urals wurde von Mathieson und Kollegen publiziert (Mathieson et al. 2015). Der mitochondriale Haplotyp wird ebenfalls einer osteurasischen Linie zugeordnet und über die Bestimmung der Y-chromosomale Linie kann eine Nähe zu Vertretern der Srubnaja-Kultur festgestellt werden.

Weitere 19 mitochondriale Genome aus der nord-pontischen Steppe (Moldawien und Ukraine) konnten durch ein *capture*-Verfahren gewonnen und erfolgreich sequenziert werden (Juras et al. 2017). Die Ergebnisse der Zusammensetzung der mitochondrialen Haplogruppen zeigen im Vergleich mit parallel existierenden Saken und Individuen der Pazyryk-Kultur eine große Ähnlichkeit. Der Anteil westeurasischer Linien nimmt in östlicher Richtung ab. Der Anteil osteurasischer Linien liegt insgesamt bei 26,3 %. Zudem zeigt sich eine große genetische Übereinstimmung mit Vertretern der Srubnaja-Kultur aus dem Nordschwarzmeer-Raum. Nach Ansicht der Autoren bestätigt dieses Ergebnis die These, dass die Srubnaja-Kultur die Vorgänger-Kultur der nord-pontischen Skythen ist. Die Haplogruppenverteilung teilt sich in drei Cluster möglicher Vorgängerpopulationen auf: Zum einen ist eine Nähe zu westlichen Jäger-Sammler-Kulturen basierend auf dem detektierten Haplotyp H5 erkennbar. Das zweite und dritte Cluster bilden frühe, neolithische Farmer und Linien osteurasischen Ursprungs. Hervorzuheben ist außerdem, dass für die Skythen Moldawiens eine Haplogruppe (M10) bestimmt wurde, die in prähistorischen Populationen Ost-Asiens und modernen Populationen des Altais zu finden ist.

Studien basierend auf genomischen Datensätzen

Die ersten genomweiten Daten einer eisenzeitlichen, skythischen Probe der Wolga-Steppe wurde im Jahr 2015 von Mathieson publiziert (Mathieson et al. 2015) und konnte durch ein Anreicherungsverfahren (1240k *capture*, (Haak et al. 2015)) gewonnen werden. Basierend auf ihrer Lokalisation in einer Hauptkomponenten-Analyse mit 777 modernen eurasischen Populationen und 221 prähistorischen Vergleichsindividuen und einer *ADMIXTURE*-Analyse kann eine Nähe zu bronzezeitlichen Steppenpopulationen (z. B. Russia_EBA, Poltavka, Jamnaja und Afanasievo) angenommen werden. Der mitochondriale Haplotyp der Probe wird einer osteurasischen Linie zugeordnet. Im Zuge dieser Studie wurden 37 weitere Vertreter bronzezeitlicher Kulturgruppen der frühen bis späten Bronzezeit Russ-

lands (Jamnaja, Poltavka, Potapovka und Srubnaja-Kultur) und östlichen Jäger-Sammlern analysiert (EHG), die in der vorliegenden Arbeit ebenfalls in einer PCA dargestellt und in den F-Statistiken mit weiteren Proben der Studie als Teil des Referenzdatensatzes (Lazaridis et al. 2016) mit einbezogen werden. Die Jamnaja der Samara - und Kalmykia-Region, Afanasievo des Altai-Gebirges und Vertreter der Poltavka-Kultur umspannen einen Zeitrahmen von 3300 – 2200 v. Chr. und können aufgrund ihrer genetischen Ähnlichkeit als bronzezeitliche Steppe-Individuen zusammengefasst werden. Diese Gruppe zeigt eine eindeutige, genetische Verwandtschaft in Richtung des Nahen Ostens. Für die Individuen der spätbronzezeitlichen Srubnaja-Kultur konnte zudem eine Verwandtschaft zu frühen, europäischen Farmern und neolithischen Proben Anatoliens nachgewiesen werden. Östlich des Urals konnte die Verbindung bereits für die Sintashta-Kultur festgestellt werden und steht möglicherweise für eine ostwärts gerichtete Migration von Vertretern der schnurkeramischen Kultur.

Von acht eisenzeitlichen Individuen aus Russland und Kasachstan konnten mittels *shotgun*-Sequenzierung die ersten genomischen Daten von Vertretern skythischer und sarmatischer Kulturgruppen analysiert werden (Unterländer et al. 2017). Die Individuen der Pazyryk-Kultur, Aldy Bel- und Zevkino-Čilikta-Stufe weisen geringe genetische Distanzen zueinander auf. Es wird außerdem die Hypothese eines multiregionalen Ursprungs der eisenzeitlichen Steppenpopulationen mit anschließendem, kontinuierlichem Genfluss von West nach Ost während der Eisenzeit angenommen. Des Weiteren konnte für alle eisenzeitlichen Populationen genetische Komponenten europäischer - und kaukasischer Jäger-Sammler nachgewiesen werden, ebenso wie für einige bronzezeitliche Kulturen (Srubnaja, Sintashta und Andronovo). Für Proben der Fundplätze Aržan, Berel, Ismailovo und Zevakino und alle weiteren eisenzeitlichen Gruppierungen ist eine östliche Komponente, die hauptsächlich in der sibirischen Population der Nganasanen vertreten ist, erkennbar. Die Sarmaten zeigen eine hohe genetische Ähnlichkeit mit Vertretern der östlichen Jamnaja- und Poltavka-Kultur der Bronzezeit. Für die Skythen kann eine genetische Verwandtschaft mit Vertretern der Jamnaja und Afanasievo-Kultur angenommen werden.

Allentoft und Kollegen konzentrieren sich in ihrer Studie auf Populationsbewegungen während der Bronzezeit in Eurasien. Aufgrund der Relevanz dieses Datensatzes für die vorliegende Arbeit wird die Studie ebenfalls in diesem Abschnitt aufgeführt. Basierend auf 101 Genomen des Neolithikums, der Bronze- und Eisenzeit mit niedriger Abdeckungstiefe belegen sie einen Genfluss von West- nach Ost-Eurasien während der Bronzezeit und anschließenden Vermischungsereignissen mit ost-asiatischen Populationen (Allentoft et al.

2015). Diese ostwärts gerichtete Expansion von Vertretern der Jamnaja-Kultur wird durch hohe genetische Ähnlichkeit dieser Kulturgruppe mit der früh-bronzezeitlichen Afanasievo-Kultur des Altai-Gebirges begründet. Eisenzeitliche Individuen Zentralasiens zeigen noch einen höheren Anteil westeurasischer Verwandtschaft im Vergleich zu lokalen, modernen Populationen.

Eine in der Bronzezeit Europas auftretende „kaukasische Komponente“ steht möglicherweise mit einer Expansion der Jamnaja-Kultur aus der Region des Kaspischen Meeres nach Europa in Verbindung.

Krzewinska und Kollegen konzentrieren sich in ihrer Studie auf 35 Individuen der pontisch-kaspischen Steppe der späten Bronze- (Srubnaja) und Eisenzeit, vertreten durch Skythen, Kimmerer und Sarmaten (Krzewinska et al. 2018). Vertreter der Srubnaja-Alakulkaya-Kultur sowie Kimmerer, Skythen und Sarmaten weisen eine hohe Diversität innerhalb ihrer jeweiligen Gruppierung auf. In einer *admixture*-Analyse konnte für die Kimmerer eine Nähe zu heutigen süd-asiatischen und nordost-asiatischen Populationen belegt werden, die mit dem Alter der Individuen zunimmt. Die PCA sowie f4-Statistik unterstützt die Affinität der Kimmerer zu west-sibirischen Populationen der Bronzezeit und zeigen eine starke Ähnlichkeit zu östlichen Sarmaten. Die Autoren sehen darin erstmals einen Nachweis, dass die Kimmerer nicht ursprünglich aus der Region der pontisch-kaspischen Steppe stammen, sondern vermuten deren Ursprung in Ostasien.

Die Skythen teilen sich in vier Untergruppen auf: in eine Nord- und Süd-Europäische Gruppe, sowie zwei Gruppen die hier Steppe- und Zentrales-Cluster genannt werden. In dem Steppe-Cluster, um das Gebiet des nördlichen Kaukasus, liegen in der Hauptkomponentenanalyse ebenfalls ein Vertreter der bronzezeitlichen Srubnaja-Kultur, ein Kimmerer und alle Sarmaten. Der chronologisch jüngste Vertreter der Kimmerer aus der Studie zeigt eine genetische Nähe zu eisenzeitlichen Individuen der Altai-Region, die in der Studie von Unterländer und Kollegen publiziert wurden (Unterländer et al. 2017). Trotz der Diversität der westlichen Skythen kann eine genetische Drift mit Vertretern der Andronovo-, Sintascha- und Afanasievo- Kultur im Vergleich zu Vertretern der Srubnaja-Alakulskaya-Kultur nachgewiesen werden. Die Autoren postulieren daher einen zentral-asiatischen Ursprung der Skythen, geprägt von genetischen Einflüssen lokaler Populationen und aus dem Nordschwarzmeerraum. Obwohl eine genetische Ähnlichkeit der eisenzeitlichen Individuen der kasachischen Steppe nachgewiesen werden konnte, konnten die Kimmerer nicht als Vorfahren der Skythen der pontisch-kaspischen Steppe festgestellt werden. Ebenso schließen die Autoren die Nomaden der westeurasischen Steppe als direkte Nachfahren der bronzezeitlichen Srubnaja-Alakulskaya-Population aus.

Für einzelne skythische Individuen kann keine südasiatische oder osteurasische Komponente nachgewiesen werden, was auf eine fehlende Mobilität aus Ost-Asien schließen lässt. Auf Basis der untersuchten Individuen wird in dieser Studie ein Zentrum der Migrationsbewegungen (*western nomadic horizon*) in der pontisch-kaspischen Steppe und südlich des Urals angenommen.

Mit 137 prähistorischen Genomen konnten Damgaard und Kollegen eine Zeitspanne von 4000 Jahren (2500 v. Chr. – 1500 n. Chr.) innerhalb der eurasischen Steppe von Europa bis in das Altai-Gebirge und den Tianshan untersuchen (Damgaard et al. 2018) und Fragen zu bronzezeitlichen Migrationsbewegungen diskutieren.

Die Autoren konnten eine starke Strukturierung innerhalb der skythischen Gruppe nachweisen, deren Ursprünge jeweils innerhalb spätbronzezeitlicher Viehhirten, europäischer Bauern und Jäger-Sammlern Südsibiriens zu finden sind. Die eisenzeitlichen Skythen teilen sich in eine Gruppe der Skythen aus Ungarn und eine weitere Gruppe auf, welche die Saken Zentralasiens beinhaltet. Die eisenzeitlichen Saken können aufgrund genetischer Ähnlichkeiten in drei weitere Untergruppen eingeteilt werden: i) Saken des Tianshan-Gebirges Turkestans, ii) die mit der Tagar-Kultur assoziierte Populationen Südsibiriens und iii) Saken der zentralen Steppe.

Während für die Skythen Ungarns eine genetische Nähe zu europäischen Bauern nachgewiesen werden konnte, jedoch ein Genfluss mit Zentralasien genetisch nicht sichtbar ist, zeigen die hier untersuchten Saken eine deutliche genetische Verwandtschaft mit südsibirischen Jäger-Sammler-Populationen. Für die Saken des Tianshan konnte zusätzlich eine genetische Komponente nachgewiesen werden, die mit neolithischen Populationen aus dem Iran in Verbindung gebracht wird.

Im Gegensatz zu Unterländer et al. (Unterländer et al. 2017) konnten Damgaard und Kollegen keinen Genfluss zwischen den untersuchten skythischen Populationen feststellen, jedoch Vermischungsereignisse mit spätbronzezeitlichen Viehhirten und jeweils entsprechenden lokalen Populationen, was die These eines multiplen Ursprungs unterstützt.

In einer weiteren Studie desselben Jahres analysieren die Autoren 74 vollständige Genome Osteuropas, Zentralasiens und Anatoliens vom Mesolithikum bis zum Mittelalter. Dabei konzentrieren sie sich auf die Migrationsbewegungen in den asiatischen Raum und den Beginn der Pferdezüchtung in der frühen Bronzezeit (Damgaard et al. 2018b). Die eisenzeitliche Probe aus Turkmenistan (~800 v. Chr.) zeigt eine genetische Nähe zu Jäger-Sammler-Populationen des Kaukasus (CHG) ebenso wie zu europäischen Farmern, was typisch für die späte Bronzezeit ist. Mögliche Vermischungsereignisse werden daher in die spätere Bronzezeit mit Vertretern der Jamnaja- bzw. Afanasievo-Kultur datiert. Eine weite-

re eisenzeitliche Probe aus Anatolien (~600 v.Chr.) liegt in der PCA mit ausgewählten modernen Populationen Eurasiens und prähistorischen Steppenpopulationen im Bereich der Individuen aus der frühen bis mittleren Bronzezeit (Anatolia EBA/MLBA) sowie JS des Kaukasus. Zusammenfassend sprechen sich die Autoren für einen geringen genetischen Einfluss früher Jamnaja-Viehzüchter auf Anatolien sowie Zentral- und Südasien in der Bronzezeit aus.

1.3 Forschungsfragen zur Populationshistorie skythischer Reiterverbände Eurasiens in der Eisenzeit

Basierend auf der bisherigen Datengrundlage und archäologischen Forschungen werden in der vorliegenden Arbeit die Verhältnisse zwischen reiternomadischen Kulturen rekonstruiert. Im Folgenden werden fünf Fragestellungen formuliert, in deren Hinblick die erzielten Ergebnisse eingeordnet und anschließend diskutiert werden sollen. Eine Genotypisierung ausgewählter funktionaler Marker liefert Aussagen über die Phänotypen einzelner Individuen und bringt zusätzliche Informationen zur genetischen Verwandtschaft mit heutigen, modernen Populationen.

Zur sprachlichen Vereinfachung werden in der vorliegenden Arbeit zentral-asiatische sowie eisenzeitliche Proben aus SW-Russland als Skythen oder skythische Proben/Individuen bezeichnet. Die Verwendung eines einheitlichen Begriffs des „*scytho-siberian triad*“ oder einer „skythischen Welt“ sind in der Archäologie aufgrund eigenständiger Merkmale der damit umschriebenen kulturellen Verbände nicht unumstritten (Bashilov, V.A. & Yablonsky 2000) Die Bezeichnung umfasst somit auch die frühen Sarmaten. Für relevante Fragestellungen und Diskussion der Ergebnisse werden die Sarmaten als eigenständige Kulturgruppe beschrieben.

A) Inwieweit ist eine genetische Kontinuität zwischen bronze- oder eisenzeitlichen Kulturgruppen zu lokalen Populationen des heutigen Eurasiens sichtbar?

Die Bestimmung genetischer Distanzen prähistorischer Kulturgruppen zu heutigen Populationen kann Aufschluss über mögliche Ursprungsgebiete oder genetische Einflüsse der untersuchten nomadischen Gruppen liefern. Eine Hauptkomponentenanalyse (PCA) mit modernen Populationen bietet die Möglichkeit, eine Korrelation der geographischen Loka-

lisation des jeweiligen Fundplatzes mit der Lage in der PCA zu überprüfen. Die *outgroup-f₃*-Statistik (Patterson et al. 2012) gibt Auskunft über die geteilte genetische Drift der zu vergleichenden Populationen.

B) Ist ein signifikanter Genfluss zwischen den eisenzeitlichen Individuen Zentral- und Nord-Asiens und den Individuen des Schwarzmeer-Raums und SW-Russlands nachweisbar?

Für die Individuen des Altai-Gebirges und die westlich gelegenen Proben aus SW-Russlands und des Schwarzmeerraums soll eine genetische Ähnlichkeit überprüft werden. Aufgrund der verschiedenen Kulturgruppen innerhalb der eisenzeitlichen Individuen kann dies Hinweise auf die Mobilität der Gruppen über den eurasischen Steppengürtel und Migrationsbewegungen liefern. Genetische Distanzen können außerdem auf verschiedene Ursprungspopulationen hindeuten.

C) In welchem Verhältnis stehen eisenzeitliche, skythische Gruppierungen zu anderen nomadischen Gruppen der Bronze- und Eisenzeit?

Auch die Frage nach genetischer Distanz bzw. Ähnlichkeit zu weiteren prähistorischen Kulturgruppen ist eng mit der Frage nach einem möglichen Ursprungsgebiet der Skythen verknüpft. Verwandtschaftsverhältnisse zu vorangegangenen Gruppierungen erlauben zudem Rückschlüsse auf Migrationsbewegungen oder Bevölkerungswechsel der jeweiligen Regionen. Die D-Statistik mit prähistorischen Kulturgruppen aus (Lazaridis et al. 2016) erlaubt die geteilte, genetische Drift zwischen verschiedenen Vergleichspopulationen und eisenzeitlichen oder bronzezeitlichen skythischen Gruppierungen zu ermitteln.

D) Ist die Verschiebung eines genetischen Einflusses aus Ost-Asien auf nomadische Populationen im Übergang von der Bronze- zur Eisenzeit erkennbar?

Der Übergang von der Bronze- zur Eisenzeit ist von kulturellen Umbrüchen geprägt. Populationsgenetische Studien konnten im Verlauf der Forschungshistorie einen Anstieg östlicher Linien von der Bronze- zur Eisenzeit nachweisen (Lalueza-Fox et al. 2004; Sarkissian 2011; Unterländer et al. 2017). Eine Expansion östlicher bzw. süd-östlicher Populationen zu Beginn der Eisenzeit kann als mögliches Szenario angenommen werden und soll in der vorliegenden Arbeit durch Vergleich mit westlichen, bronzezeitlichen Kulturgruppen

(Lazaridis et al. 2016) und der Nähe zu lokalen, modernen Populationen Zentral - und Ost-Asiens diskutiert werden (Hellenthal et al. 2014; Busby et al. 2015).

E) Welche möglichen Ursprungsgebiete können für westliche und östliche skythische Populationen angenommen werden? Bestätigt sich die Annahme eines multiregionalen Ursprungs skythischer Kulturgruppen mit anschließender Expansion?

Die Frage nach einem möglichen Ursprungsgebiet bzw. das Ausmaß einer genetischen Verwandtschaft skythischer Kulturgruppen konnte bislang noch nicht vollständig geklärt werden. Durch die Publikation von Martina Unterländer basierend auf mitochondrialen Haplogruppen und *shotgun*-Daten von 8 Individuen der hier vorgestellten eisenzeitlichen Fundplätze konnte ein möglicher multiregionaler Ursprung skythischer Kulturgruppen nachgewiesen werden (Unterländer et al. 2017). Mit dem erweiterten Datensatz dieser Arbeit von 41 Proben soll das Szenario überprüft werden. Hierfür werden die eisenzeitlichen, bronzezeitlichen und neolithischen Proben und 281 (bzw. 126) Vergleichsindividuen aus der Literatur (Lazaridis et al. 2016) herangezogen und mit den Ergebnissen vorangegangener Thesen zusammengeführt werden.

F) Welche Informationen liefern die Phänotypen der ausgewählten Marker?

Es werden 54 ausgewählte Positionen der funktionalen Marker des Anreicherungsverfahrens gezielt für jede Probe analysiert werden. Die Bestimmung der Genotypen der Pigmentierungs-assoziierten Marker erlauben ein optisches Erscheinungsbild der jeweiligen Individuen zu rekonstruieren. Zusätzlich werden Allelfrequenzen für 13 funktionale Marker bestimmt, deren abgeleitete Allele in ihrer Verbreitung zwischen Europa, Zentral- und Ost-Asien signifikant variieren. Die Ergebnisse dieser Phänotypenbestimmung liefern unterstützende Informationen zur genetischen Verwandtschaft der verschiedenen Kulturgruppen. Eine Affinität zu modernen west- bzw. osteurasischen Populationen kann für die Rekonstruktion der Herkunft der skythischen Gruppen herangezogen werden.

1.4 Methodische Ziele

Ein Anreicherungsverfahren (*capture* oder *target enrichment*) auf DNA-Bibliotheken, so genannten *libraries*, ermöglicht es, gesamt-genomische Daten zu sequenzieren und gezielt vergleichbare Positionen zwischen verschiedenen Individuen mit hoher Abdeckung (*coverage*) zu erzeugen. Die Fokussierung auf definierte Teilbereiche des gesamten Genoms ermöglicht aufgrund geringerer Kosten und Daten-Kapazitäten groß angelegte Studien mit einer hohen Individuenzahl umzusetzen.

Im Bereich der prähistorischen Forschung mit altem Skelettmaterial ist es aufgrund geringer endogener Molekül-Zahlen in DNA-Extrakten und teilweise degradiertem und kontaminiertem Material sehr aufwendig und kostspielig vollständige Genome mit hoher Abdeckung zu gewinnen. Das *capture*-Verfahren bietet eine Möglichkeit, gezielt spezifische DNA-Abschnitte für mehrere Individuen zu sequenzieren und so die Wahrscheinlichkeit eines großen Überschneidungsbereichs mit genomischen prähistorischen sowie modernen Referenzdaten zu schaffen.

In der vorliegenden Arbeit sollen neben ausgewählten funktionalen Markern auch neutrale Regionen des humanen Genoms angereichert werden. Damit sind nicht-kodierende Bereiche gemeint, die keine Muster genetischer Variation zeigen, nicht von natürlicher Selektion beeinflusst oder mit diesen Regionen verbunden sind und der freien Rekombination unterliegen. Die nicht-kodierenden DNA-Sequenzen sind am geeignetsten, um populationsgenetische Analysen durchzuführen, die auf einer Berechnung des *allele-frequency-spectrum*s (AFS) beruhen (Gutenkunst et al. 2009). Wenn mehrere Individuen einer Population untersucht werden, haben AFS/SFS-basierte Methoden den Vorteil, dass keine voll-genomischen Datensätze erforderlich sind, da die korrekte Abschätzung des Spektrums auch anhand weniger Mb möglich ist. Dafür ist allerdings eine hohe Abdeckung der sequenzierten Regionen erforderlich (Excoffier et al. 2013). Dies kann durch ein selektives *target enrichment* auch für prähistorisches Material erzielt werden.

Für die optimale selektive Anreicherung wird zunächst ein Anreicherungsverfahren der drei führenden Hersteller (MYbaits® von MYcroarray, NimbleGen von Roche und SureSelect^{XT} von Agilent) getestet, das in einem flüssigen Hybridisierungsansatz (*in-solution*) durchgeführt wird.

Ein methodisches Ziel seitens der Datenproduktion im Labor ist zunächst die Auswahl eines Anreicherungs-Kits, basierend auf einer hohen Effizienz des durchgeführten *captures*, einer maximale Abdeckungstiefe (*coverage*) der Sequenzabschnitte, einer Flexibilität im *bait-*

Design und die möglichst unabhängige Verwendung der benötigten Chemikalien. Das Protokoll des ausgewählten Herstellers kann dann durch Modifikationen an die Bedingungen alter DNA angepasst werden. Des Weiteren werden überwiegend eisenzeitliche Proben der eurasischen Steppe Zentralasiens und westeurasische, bronzezeitliche Proben auf ihren Erhaltungszustand und Authentizität hin getestet (*screening*). Für ein vollständiges *screening* einer Probe wird eine MiSeq-Sequenzierung einer erstellten DNA-*library* durchgeführt, um Rückschlüsse auf den endogenen DNA-Gehalt eines DNA-Extrakts zu erhalten. Die sequenzierte *library* wird ebenfalls mittels quantitativer *real-time* PCR (qPCR), dem Agilent Bioanalyzer- und einer Qubit-Messung quantifiziert.

Für die finale Probenauswahl werden dann neutrale Regionen und informative Polymorphismen in einem Ansatz angereichert und anschließend im *next-generation-sequencing* Verfahren sequenziert.

2 Skelettmaterial und Methoden

2.1 Probenbearbeitung

Insgesamt wurde DNA aus Knochenmaterial von 77 Proben aus Zentralasien extrahiert und in DNA-Bibliotheken (*libraries*) überführt, um weiterführende Analysen zur Qualitätsbestimmung der DNA vornehmen zu können. 41 der untersuchten Proben erfüllten die festgelegten Kriterien für eine Anreicherung nukleärer DNA mittels *capture enrichment* wie in Kapitel 3.1.3 beschrieben, oder wurden aufgrund ihrer thematischen Relevanz ausgewählt. 31 der ausgewählten Proben wurden erstmals in der Dissertation von Martina Unterländer bearbeitet (Unterländer 2014) eine weitere Probe aus der Frühen Bronzezeit wurde aus dem Dissertationsprojekt von Sandra Wilde (Wilde 2017) übernommen. Acht Proben der Fundplätze Solonzy 5, Ulžan, Kamennyj Ambar, Pokrovka und Berel werden erstmalig in der vorliegenden Arbeit untersucht.

Die Bearbeitung des bereits vorliegenden Knochenpulvers aller untersuchten Proben erfolgt zunächst in einem Labor an der Johannes Gutenberg-Universität Mainz, welches speziell für die Anforderungen der Kontaminationsvermeidung bei der Arbeit mit altem Knochenmaterial entworfen und erbaut worden ist. Räume der Probenvorbereitung bis hin zur Amplifikation extrahierter DNA und Bereiche, in denen amplifizierte DNA-Produkte bearbeitet werden, sind physikalisch voneinander getrennt. Ziel ist es, archäologisches Skelettmaterial vor kontaminierenden Umwelteinflüssen und bereits amplifizierten DNA-Molekülen zu schützen. Hierfür müssen bis zu den eigentlichen Laborräumen drei voneinander getrennte Räume (Schleusen) durchlaufen werden. Die frisch gewaschene Kleidung wird in der zweiten Schleuse abgelegt und gewechselt. Zudem wird eine Haube über die Haare gelegt, ein Mundschutz angelegt und zwei Paar Handschuhe übergezogen. In der letzten Schleuse werden Strümpfe und Handschuhe erneut gewechselt, Anzug, Überschuhe und ein Visier angezogen. Alle Gegenstände, Verbrauchsmaterialien, Geräte, Chemikalien und Knochen werden bereits in der ersten Schleuse oberflächlich gereinigt (Alconox®-anionisches Detergenz, Sigma Aldrich), wenn möglich in Lösung inkubiert (Natriumhypochlorid) und durch eine Durchreiche mit integriertem UV-Licht nach 8-stündiger Bestrahlung in Schleuse 2 überführt. Alle weiteren Methoden der Kontaminationsvermeidung und Bearbeitung von altem Knochenmaterial bis zur Erstellung der DNA *libraries* erfolgten nach den Richtlinien aus (Bollongino et al. 2013; Scheu et al. 2015).

Die Einführung des Skelettmaterials in das Labor sowie die Bearbeitung der Knochen bis zur Extraktion von DNA-Molekülen aus pulverisierten Knochenstücken wurde von Martina Unterländer durchgeführt und beschrieben (Unterländer 2014). Ein Individuum der Poltavka Kultur (Kal1) wurde erstmals von Sandra Wilde (Wilde 2014) bearbeitet. Eine Probe aus Berel (Be10) und Ulžan (Ul2) wurden in der Masterarbeit von Angela Gräfen (Gräfen 2009) beschrieben.

Alle Proben der Fundplätze Kamennyj Ambar 5 (KAMA) und Solonzy 5 (Sol5) wurden für die vorliegende Arbeit erstmals extrahiert bzw. gesägte Knochenfragmente pulverisiert und weiterbearbeitet.

2.1.1 Probenauswahl und Extraktion

Alles Zahn- und Knochenmaterial welches in vorangegangenen Amplifikationen auf nukleäre Marker mehrfach positive Ergebnisse erzielen konnte (Unterländer 2014, Wilde 2014) wurde zunächst für eine erneute Extraktion und Qualitätsbestimmung ausgewählt (Tab. 1).

Tabelle 1 Übersicht aller Proben, die für eine erste Analyse ausgewählt bzw. für die vorliegende Arbeit neu extrahiert wurden

Fundplatz	Gebiet	Alter	Kultur	Labor ID	Skelettelement	Extraktion
Aržan 2	Südsibirien, Region Tuva	7.–6.Jh.v.Chr.	Aldy-Bel	A1	Femur	16
				A4b	Zahn	8
				A6	Femur	6
				A7	Femur	16
				A8b	Zahn	8
				A9	Femur	12
				A10b	Zahn	8
				A19b	Femur	12
				A17	Zahn	8
				A20	Femur	9
A21	Humerus	9				
Ak-Alacha 1	Südsibirien, Altai Ukok Plateau	4.–3.Jh.v.Chr.	Pazyryk	Ak1_1	Femur	7
Ak1-2b				Tibia	14	
Ak-Alacha 4	Ak4-1			Femur	15	
Ak-Alacha 5	Südsibirien, Altai Ukok Plateau			AK5_1c	Zahn	10
				Ak5_4b	Zahn	10
		Ak5-6a	Femur	15		
			Ak5-7a	Tibia	15	

Fundplatz	Gebiet	Alter	Kultur	Labor ID	Skelettelement	Extraktion
				Ak5-8a	Femur	15
Berel	Ost-Kasachstan, Kasachischer Altai	4.-3.Jh.v.Chr.	Pazyryk	Be1	Zahn	8
				Be4	Zahn	8
				Be5	Schädel- knochen	14
				Be7	"Patella"	14
				Be9	"pars petrosa"	7
				Be10	pars petrosa	1_JK
				Be11	Femur	7
				Be12	Femur	7
Barsučij Log	Minusinsker Becker	5.Jh.v.Chr.	Tagar	Bl1	Femur	10
				Bl2	Femur	10
				Bl3	Femur	10
				Bl4	Femur	12
				Bl5	Femur	12
Ismailovo	Ostkasachstan	9.-7.Jh.v.Chr.	Zevakino- Čilikta	Is1	Tibia	8
				Is3	Clavicula	16
				Is4	Tibia	8
Justyd 12	Südsibirien, Čuja-Region	4.-3.Jh.v.Chr.	Pazyryk	J12-1a	Phalange	14
				J12-2a	Femur	14
				J12-3	Femur	14
				J12-6a	Femur	14
				J12-7a	Femur	15
				J12_8	Femur	8
				J12-9a	"Femur"	14
Kalinovka	SW-Russland	2700-2100 v. Chr.	Poltavka Kultur	Kal1	"Langknochen"	18
Kamennyi Am- bar 5	Kamennyj Ambar, Olgi- no Ural	2000 v. Chr.	Sintašta- Kultur	KAMA1	Langknochen	17
				KAMA2	Langknochen	17
				KAMA6	Langknochen	17
				KAMA7	Langknochen	17
				KAMA8	Langknochen	17
				KAMA10	Langknochen	17
Kolbino 1	SW-Russland	3.Jh.v.Chr.	Skythen	Kol4	Zahn	13
				Kol5	Zahn	13
Novozavedennoe 2	SW-Russland	8.-6.Jh.v.Chr.	Initiale Skythen	Nov7	Zahn	16
				Nov8	Zahn	16
Pokrovka 02	SW-Russland	5.-2.Jh.v.Chr.	Frühe Sarmat- ten	Pr1	Langknochen	11
				Pr2	Langknochen	11
				Pr3	Langknochen	11
				Pr4	Langknochen	11
				Pr5	Langknochen	11
				Pr6	Langknochen	11
				Pr7	Langknochen	11
				Pr8	Langknochen	11
				Pr9	Langknochen	11
				PR10	Langknochen	12

Fundplatz	Gebiet	Alter	Kultur	Labor ID	Skelettelement	Extraktion
				PR11	Langknochen	12
				PR12	Langknochen	12
				PR13	Langknochen	12
Solonzy 5	Westsibirien, Novosibirsk	Neolithisch		Sol5-1a	Zahn	13
				Sol5-9a	Zahn	13
Ulžan	SO- Kasachstan	5-4 Jhd. v. Chr.	Saken	Ul2	pars petrosa	1_JK
Zevakino	Ostkasachstan	9.-7.Jh.v.Chr.	Zevakino- Čilikta	Ze2	Tibia	8
				Ze3	Femur	8
				Ze4	"Tibia"	12
				Ze7	Femur	12
				Ze8	Femur	8 / 12
				Ze9a	Tibia	16
				Ze10a	Tibia	16

Die Lyse der Knochenstruktur und Isolierung von DNA-Molekülen aus bereits pulverisiertem Knochen- bzw. Zahnmaterial erfolgte mittels Phenol-Chloroform-Extraktion. Es wird ab dem ersten Bearbeitungsschritt eine Leerkontrolle (LK) mitgeführt, um eine mögliche Verunreinigung der verwendeten Reagenzien oder Arbeitsmaterialien überprüfen zu können. Als Einsatzmenge werden zwischen 0,2- 0,66g Knochenpulver mit 6,7 ml EDTA (0,5 M, pH 8), 30 µl Proteinase K und 250 µl Detergent (N-Laurylsarcosyl) angesetzt und für 48 Std. bei konstanten 36 °C auf einem Wippschüttler inkubiert. Um die freigesetzten DNA-Moleküle aus der Lösung zu isolieren wurde eine Phenol-Chloroform-Extraktion durchgeführt. Die sich dadurch abgesetzte wässrige Phase wird über Amicon® Ultra-15 (50 kDa) aufgereinigt und konzentriert. Das teilweise durch bodengelagerte Mineralien und Huminsäuren verfärbte Extrakt wird durch sukzessive Zugabe von max. 12 ml UV-HPLC-H₂O gewaschen und abzentrifugiert. Es werden 150-300 µl Extrakt je Probe in LoBind-Tubes (0,5 ml) je 50 µl überführt.

2.1.2 *library*-Erstellung

Bis zum Schritt der Amplifikation der *fill-in*-Produkte kann dasselbe Protokoll zur *library*-Erstellung für die Qualitäts- und Quantitätsbestimmung sowie für die *screening-libraries* der DNA-Extrakte verwenden werden. Die Einsatzmenge des jeweiligen DNA-Extrakts und die verwendeten PCR-Primer sind abweichend (siehe 3.1.3). Für *libraries*, die zunächst für eine *shotgun*-Sequenzierung auf dem Illumina MiSeq erstellt wurden, wurden jeweils 44-50

µl DNA-Extrakt eingesetzt. Für die folgenden *capture-libraries* variierte die Einsatzmenge zwischen 25 und 50 µl, in Abhängigkeit von der Anzahl an benötigten *libraries* und Verfügbarkeit des Extrakts.

Das Protokoll zur *library*-Erstellung basiert auf dem Protokoll aus (Kircher et al. 2012) mit folgenden Abweichungen:

- keine Behandlung mit USER™ Enzym-Mix
- die finale Konzentration des Adapter-Mix während der Ligation beträgt 1.25 µM pro Probe
- nach dem *fill-in* Schritt wird mit 1 µl je Probe eine 1:40 Verdünnung für die anschließende qPCR erstellt und kühl gelagert
- Mitführen einer synthetisch hergestellten Positivkontrolle mit bekannter Länge und Sequenzabfolge (Nonsense-Hybrid / NH) während der *library*-Erstellung zur Effizienzkontrolle der verwendeten Chemikalien und technischen Durchführung; *Libraries* aus Versuchsreihen mit $< 10^9$ Molekülen in der qPCR nach der *fill-in* Reaktion werden nicht verwendet. Herstellung und Nukleotidsequenz beschrieben in Kreuzer (2017) und im Anhang B1 und B2
- Mitführen einer Leerkontrolle (LK) aus 50 µl UV-bestrahltem Nuklease-freiem Wasser während der *library*-Erstellung und anschließender PCR bis zur Quantifizierung und Sequenzierung der *libraries* (siehe 3.1.4 und 3.4.4)
- Entfernen möglicher Adapter- und/oder Primerdimere über eine zusätzliche Aufreinigung der einzelnen *libraries* oder Probenpools über SPRI-*beads* mit einem *cut-off* bei > 100 bp (Agencourt® AMPure® XP, Beckmann Coulter)

2.1.3 PCR

Extraktionskontrolle

Nach jeder Extraktion erfolgte eine Amplifikation eines mitochondrialen Fragments (HVR IIIa; L16117 / H 16223 aus Wilde 2014) auf alle Extraktionsprodukte, um eine erfolgreiche Extraktion nachzuweisen und mitgeführte Leerkontrollen auf Kontaminationen zu überprüfen. Das Amplikon hat eine Länge von 162 bp inklusive der Primersequenzen. Der Reaktionsansatz der PCR besteht aus 2 µl DNA-Extrakt, 5 µl 1xGold Buffer, 1 µl dNTPs, 5 µl MgCl₂, 1 µl BSA, 2 µl je Primer, 31,5 µl H₂O unter folgenden Amplifikationsbedingungen: 6 min initiale Denaturierung bei 94 ° C, je 40 Zyklen eine Denaturierung bei 94 ° C für 40 Sek., eine Hybridisierung für 40 Sek. bei 58 ° C und eine Elongation für 40

Sekunden bei 72 ° C. Eine Positivkontrolle wurde außerhalb des Reinraumlabor in den Laborräumen zur Aufreinigung von PCR-Produkten kurz vor dem Start der PCR hinzugefügt.

Die Amplifikationsprodukte wurden direkt mittels Agarose-Gelelektrophorese gegen einen Längenstandard aufgetragen und die Länge und Stärke der Bande kontrolliert.

Screening-PCR

Die *fill-in* Produkte wurden in jeweils drei Parallelen mit einem Reaktionsvolumen von 50 µl in zwei unabhängigen PCR-Runden amplifiziert. Bereits bei der ersten Amplifikation, die im Reinraumlabor angesetzt und pipettiert wird, wird ein probenspezifischer Index (P7_*) vergeben und separat zur jeweiligen Probe hinzugefügt.

Es werden alle Primersequenzen mit 7 bp langen Indexen des dualen Indexsystems aus Kircher et al. (2012) verwendet.

Der Reaktionsansatz setzt sich wie folgt zusammen: 10-12 µl *library fill-in* Produkt, UV-HPLC-H₂O, AmpliTaq Gold® DNA-Polymerase (0.05U/µl), 1xGold Buffer, 2.5mM MgCl₂, 200µM je dNTP und 200nM je Primer. Die thermischen Konditionen betragen: Initial Denaturierung von 94 °C für 6 min gefolgt von 10-14 Zyklen Denaturierung bei 94 °C für 40 Sek., Annealing bei 60 °C für 40 Sek. einer Elongation bei 72 °C für 40 Sek. mit einem finalen Elongationsschritt bei 72 °C für 5 min.

Nach einer Aufreinigung mit dem MSB Spin PCR Repace Kit (Strattec) wird der zweite Index (P5_*) in einer weiteren PCR mit der Herculase-Polymerase (Agilent) und geringere Zyklenzahl hinzugefügt. Es werden 66,4 µl HPLC-H₂O, 20 µl 5x Herculase Puffer, 1 µl dNTPs (je 25mM), 2,3 µl je Primer, je 7 µl *library*-Produkt und 1 µl Herculase II Fusion Polymerase zu einem 100 µl Reaktionsansatz pipettiert. Die thermischen Konditionen hierfür sind: eine initiale Denaturierung bei 98 °C für 2 min, eine Denaturierung bei 98 °C für 30 Sek., die Hybridisierung erfolgte bei 60 °C für 30 Sek., eine Elongation für 30 Sek. bei 72 °C und eine finale Elongation bei 72 °C für 5 Sek.

Dadurch, dass der Index P5 erst in einer separaten, zweiten PCR hinzugefügt wird, bleibt die Kombinationsmöglichkeit beider Indexe für eine mögliche, erneute Sequenzierung flexibel. Zusätzlich können die *screening libraries* so ebenfalls für eine weitere nukleäre Anreicherung verwendet werden, da hierfür die Primersequenz P7 bis zur zweiten Amplifikation der *capture*-Produkte verkürzt bleibt. Dies ermöglicht eine effizientere Blockierung der Adapter-Sequenzen während der Hybridisierung.

Die Zyklen-Zahl und verwendeten Primer-Systeme sind für die erste und zweite Amplifikation abweichend (Tab. 2):

Tabelle 2 Verwendete Primer und Zyklenanzahl für die Amplifikation der *screening-* und *capture-libraries*; * = eine 8bp lange, spezifische Indexsequenz aus Kircher et al. (2012)

	PCR 1	PCR 2
Primer P5	Is7 (10 μ M)	P5_* / Is7 (10 μ M)
Primer P7	P7_* (10 μ M)	Is6 (10 μ M)
Zyklen	10	9-11

Capture-libraries

Der Reaktionsansatz beider Amplifikationsrunden ist wie bereits für die *screening-libraries* beschrieben. Abweichend wurde jede Probe für die Herculase-PCR in drei Parallelen amplifiziert und je 8-10 μ l Target eingesetzt. Anstelle des zweiten Index-Primers wird erneut der Primer Is7 verwendet und die Adapterstruktur somit nicht vervollständigt. Die Aufreinigung erfolgte dann ebenfalls mit MSB Spin PCRpace (Stratec).

Für alle bearbeiteten Proben aus Kamennyj Ambar (KAMA) wurden bereits in der ersten Amplifikationsrunde beide Indexsequenzen angehängt.

2.1.4 Quantifizierung und Qualitätsbestimmung

Die DNA-Konzentrationen der aufgereinigten PCR-Produkte (*libraries* oder *capture-Produkte*) werden mittels einer **Qubit®** Messung (dsDNA HS Assay, Invitrogen) und des Agilent **Bioanalyzer** (High Sensitivity Kit) gemessen. Durch ein Auftrennen der Moleküle nach ihrer Länge während einer Messung auf dem Bioanalyzer können zusätzlich die Fragmentlängen der Proben ermittelt werden. Mögliche Adapter- bzw. Primerdimere werden so erkannt und können anschließend durch eine zusätzliche Aufreinigung mit *AMPure XP Beads* (Agencourt) entfernt werden.

Mittels einer Sequenzierung auf dem **Illumina MiSeq** (50bp, *single-end*) wird eine Bestimmung des endogenen DNA-Gehalts der erstellten *libraries* ermöglicht. Die Sequenzierstrategien sind im Kapitel 3.3, die Analyse der Rohdaten im Kapitel 3.4.1 erläutert.

Um die genaue Anzahl endogener Moleküle einer Probe berechnen zu können, werden alle ermittelten Sequenzen gegen das menschliche Genom (GRCh37/hg19) aligniert und diese durch die Anzahl aller Sequenzen, die den Qualitätsfilter der Rohdatenauswertung passiert

haben, geteilt und in Prozent umgerechnet. Die Duplikationsrate wird bei der Berechnung berücksichtigt, da eine hohe Kopienzahl an *library*-Molekülen das Ergebnis verfälscht und zusätzlich auf eine niedrige Komplexität der Probe schließen lässt.

Mit einer quantitativen ***Real-Time PCR*** ist es möglich die genaue Anzahl an Molekülen zu berechnen, die während der *library*-Erstellung aus der erfolgreichen Adapter-Ligation hervorgegangen sind, ohne vorab eine Amplifikation durchführen zu müssen. So kann die tatsächliche Anzahl an einzigartigen DNA-Molekülen eines Extrakts berechnet werden. Deshalb wird für diese PCR das Primerpaar IS7/IS8 eingesetzt, da dieses an die unvollständige Adapterstruktur vor einer *library*-Amplifikation ligieren kann. Für den Mastermix wird der KAPA Sybr Fast Universal Mastermix (PeqLab, VWR International) verwendet. Die Amplifikation erfolgt auf dem Step One Plus™ Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific). Synthetisierte *library*-Moleküle (89bp) bekannter Konzentrationen werden zur Berechnung einer Standard-Reihe mitgeführt. Der genaue Reaktionsansatz, Versuchsaufbau und die Geräteeinstellungen sind der Dissertation von Susanne Kreutzer (Kreutzer 2017) zu entnehmen.

Da unterschiedliche Mengen an Knochenpulver je Probe in die Extraktionen eingesetzt wurden, wird die genaue Kopienzahl (*copy number* = CN) der qPCR unter Berücksichtigung der Einsatzmenge in die *library* für jede Probe korrigiert. Zunächst wurde berechnet wieviel mg Knochenpulver in 1 µl Extraktionsvolumen enthalten sind, dann wieviel mg in die *library* eingesetzt wurden. Die Kopienzahl pro µl wird auf 40 µl finale Kopienzahl hochgerechnet, um so ermitteln zu können, wie viele Moleküle je 1 mg Knochen enthalten sind.

So kann ein direkter Vergleich der Moleküle je *fill-in* Produkt der untersuchten Proben ermöglicht werden.

Setzt man die Molekülzahl zum endogenen DNA-Gehalt einer Probe ins Verhältnis, kann die Komplexität einer *library* bestimmt werden. Je höher der endogene DNA-Gehalt in Verbindung mit einer hohen Anzahl an Molekülen vor einer Amplifikation, desto größer ist die Wahrscheinlichkeit bei einer Sequenzierung nicht rudimentäre und einzigartige Sequenzinformation zu erhalten. Die quantitative Real-Time PCR wird ebenfalls genutzt, um *library*-Leerkontrollen auf mögliche Kontaminationen während der Bearbeitung zu testen.

Die Proben Be10 und Ul2 wurden zu einem späteren Zeitpunkt für die Bearbeitung ausgewählt. Da es für beide Proben möglich war Felsenbeine (*pars petrosa*) für die Analyse auszuwählen, wurde aufgrund der Annahme eines sehr guten Erhaltungsstand (Gamba et al.

2014; Pinhasi et al. 2015) am nur die Komplexität der DNA-Moleküle der erstellten *libraries* für eine nukleäre Anreicherung mittels qPCR kontrolliert (siehe 3.1.4)

Die Ergebnisse der hier beschriebenen Analysen und die finale Probenauswahl sind im Kapitel 4.1 und im Anhang B detailliert beschrieben und aufgelistet.

Anhand der *shotgun*-Sequenzierungen auf dem Illumina MiSeq kann zusätzlich eine molekulare Geschlechtsbestimmung der Proben durchgeführt werden (siehe Kapitel 3.4.2 und 5.1)

2.2 Selektive Anreicherung nukleärer DNA

Alle Verfahren einer selektiven Anreicherung genomischer DNA beruhen auf dem Prinzip, das gezielt definierte Bereiche der DNA für eine spätere Sequenzierung isoliert und nicht gewünschte Sequenzabschnitte schon im Vorfeld eliminiert werden können. Zum einen, um Sequenzierungen kosteneffizienter zu gestalten und des Weiteren können so untereinander vergleichbare Sequenzabschnitte in höherer Qualität erzielt werden.

Im Gegensatz zu einer *shotgun* Sequenzierung werden nur definierte DNA-Abschnitte, die aufgrund einer bestimmten Fragestellung ausgewählt werden, weiterbearbeitet und analysiert.

Vor allem bei der Arbeit mit prähistorischem Skelettmaterial ist es aufgrund geringen endogenen Molekül-Zahlen in DNA-Extrakten und teilweise degradiertem und kontaminiertem Material sehr aufwendig und kostspielig vollständige Genome mit hoher Abdeckung zu gewinnen. Gerade in diesem Bereich ist also die Möglichkeit interessant, gezielt DNA-Abschnitte zu sequenzieren die mit anderen Proben vergleichbar sind und einen Überschneidungsbereich mit Referenzdaten zeigen. Mittlerweile konnte in zahlreichen Publikationen erfolgreich ein *capture* oder *Target enrichment* auf alte DNA durchgeführt werden. Zunächst auf das mitochondriale Genom als *target* (Briggs et al. 2009a; Maricic et al. 2010; Krause et al. 2010), aber auch auf genomische DNA und einzelnen Chromosomen (Hodges et al. 2009; Briggs et al. 2009b; Burbano et al. 2010; Fu et al. 2013).

In der vorliegenden Arbeit sollen gezielt neutrale Regionen des humanen Genoms angereichert werden. Damit sind nicht-kodierende Bereiche gemeint, die keine Muster genetischer Variation zeigen, nicht von natürlicher Selektion beeinflusst oder mit diesen Regionen verbunden sind und der freien Rekombination unterliegen.

Die nicht-kodierenden DNA-Sequenzen sind am geeignetsten um populationsgenetische Analysen durchzuführen, die auf einer Berechnung des *allele-frequency-spectrum*s (AFS) beruhen (Gutenkunst et al. 2009; Veeramah et al. 2018). Wenn mehrere Individuen einer Population untersucht werden haben AFS/SFS-basierte Methoden den Vorteil, dass keine voll-genomischen Datensätze erforderlich sind, da die korrekte Abschätzung des Spektrums auch anhand weniger Mb möglich ist. Dafür ist allerdings eine hohe Abdeckung der sequenzierten Regionen erforderlich (Excoffier et al. 2013).

Der Ansatz basiert auf dem Konzept der Koaleszenztheorie, also der Frage welche Ereignisse rückblickend zu den beobachteten Daten führend konnten (Kingman 1982). In Abwesenheit von Selektion wird davon ausgegangen, dass Genorte bzw. Linien einer Population zufällig weitergegeben wurden und durch verschiedene Koaleszenzereignisse in einem letzten gemeinsamen Vorfahren münden. Jede Abweichung vom Modell der genetischen Drift ist nur durch Ereignisse erklärbar, die die gesamte Population beeinflusst haben, wie Migration/Immigration, anschließende Vermischung des Genpools, plötzlichem Wachstum der Population oder Reduzierung der Populationsgröße. Verwendet man also vermeintlich neutrale Regionen als Datengrundlage können genetische Prozesse der Populationshistorie untersucht und dadurch Rückschlüsse auf die Populationsdemografie gemacht werden.

Ziel ist es einen Datensatz zu schaffen der für diese nachfolgenden populationsgenetischen Methoden geeignet ist. Die Auswahl der neutralen Regionen erfolgte nach bestimmten Kriterien, die in Kapitel 2.2.2 beschrieben werden.

Zusätzlich wurden einzelnen SNP Positionen der Anreicherung hinzugefügt, die phänotypischen Merkmalen beeinflussen, mit populationsspezifischen Krankheitsbildern oder dem metabolischen Syndrom in Verbindung stehen (Kapitel 2.2.3). Eine Auflistung aller Positionen der funktionalen Marker findet sich im Anhang B3.

Um definierte Bereiche genomischer DNA gezielt zu selektieren und für eine Sequenzierung zur Verfügung zu stellen stehen laborseitig verschiedene Methoden zur Auswahl.

Bei *array*-basierten Anreicherungsverfahren hybridisieren die *DNA-library*-Moleküle an komplementäre *probes*, welche an Oberflächen sogenannter *microarrays* gebunden sind. (Hodges et al. 2007; Mamanova et al. 2010) Freie oder unspezifische DNA-Sequenzen werden anschließend ausgewaschen. Zunächst von NimbleGen (*HD1 array*) entwickelt, folgten bald vergleichbare *arrays* von Agilent, welche auch mit einer Sequenzierung auf dem Illumina Genome Analyzer kompatibel sind.

Anreicherungsverfahren, die in einem flüssigen Hybridisierungsansatz (*in-solution*) durchgeführt werden, verwenden auf einem microarray synthetisierte DNA-Oligonukleotide (*probes*), welche dann vom Array gelöst und mittels PCR amplifiziert werden (Abb. 3). Dabei wird ein T7-Promotor Bindestelle eingebaut. Dadurch können die Amplikons in biotinylierte RNA-Sequenzen (*baits*) transkribiert werden. Diese bilden mit zufällig geshearten, einzelsträngigen DNA-*library*-Molekülen während der Hybridisierung DNA-/RNA-Duplex-Strukturen aus. (Tewhey et al. 2009; Gnirke et al. 2009) Dabei werden die universellen Adaptersequenzen an den Enden der *library*-Moleküle durch korrespondierende Blocking-Oligos blockiert, um Bindungen an den Moleküleenden zu verhindern. Folgende Waschschrte bei definierten Temperaturen und Salzkonzentrationen dienen dazu, unspezifische DNA-RNA Bindungen und überschüssige Blocking-Oligos auszuwaschen. Mit Hilfe Streptavidin-überzogener, magnetischer *beads* können die biotin-markierten RNA-*baits* und die daran gebundene DNA gezielt isoliert werden. Vor einer anschließenden Sequenzierung werden die so gecapturten DNA-Sequenzen mittels PCR vervielfältigt und quantifiziert.

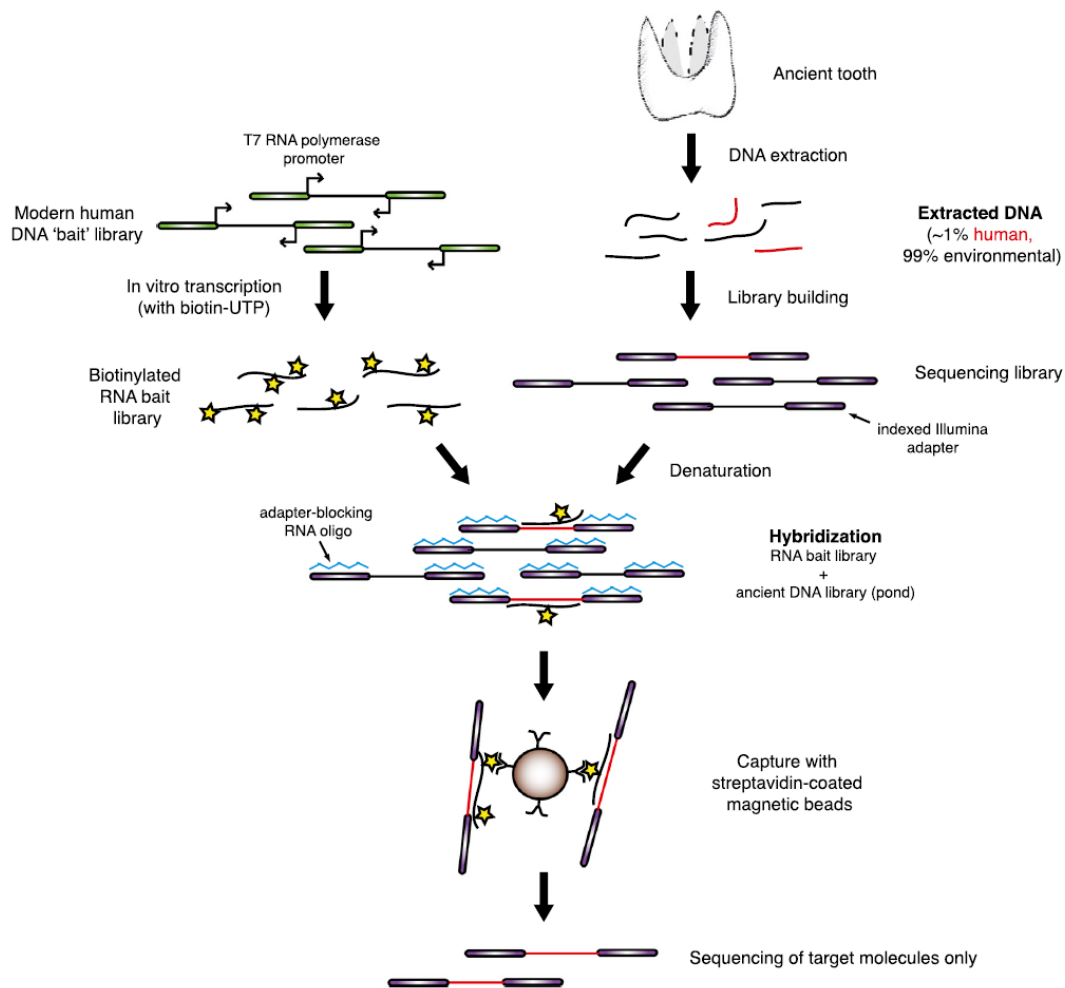


Abbildung 3 Grafischer Darstellung einer *bait-bead* basierten Anreicherung alter DNA in-solution.
Nach (Carpenter et al. 2013)

Zunächst soll ein für prähistorisches Skelettmaterial geeignetes Protokoll etabliert werden. Hierfür wurden drei verschiedene *in-solution* Anreicherungskits verschiedener Hersteller exemplarisch getestet. Detaillierte Beschreibungen des *bait*-Designs und Erläuterungen der Laborprotokolle finden sich in den Kapiteln 2.2.4 und 2.2.5, die Ergebnisse und die finale Auswahl werden im Kapitel 4.2 dargestellt.

2.2.1 Vergleich verschiedener Anreicherungsverfahren

Eine Anreicherung *in-solution* mit anschließender Isolierung der Fragmente über magnetische *beads* bietet speziell für die Arbeit mit prähistorischem DNA-Material viele Vorteile. Neben den klassischen Verbrauchsmaterialien wie Tubes, Falkons, Spitzen und PCR-Cycler werden keine zusätzlichen Geräte benötigt. Somit können alle Materialien gemäß unseren Standards UV-bestrahlt und speziell gereinigt werden. Das Kontaminationsrisiko kann so gemindert werden. Die transparente und vergleichsweise technisch einfache Durchführung

der Anreicherung erlaubt es die Versuche in gentechnischen Laboren der Universität selbst durchzuführen. Es ist kaum zusätzliche Laborgeräte notwendig. So können ein Transport und eine weitere Bearbeitung in fremden Laboren vermieden werden.

Es sollen Anreicherung-Kits von drei führenden Herstellern (MYbaits® von MYcroarray, NimbleGen von Roche und SureSelect^{XT} von Agilent) exemplarische im Hinblick auf Effizienz, Kosten, Durchführbarkeit und Transparenz im *bait*-Design und bei den verwendeten Chemikalien getestet werden. Die Auswahl der neutralen Regionen und funktionalen Marker unterscheidet sich zwischen den Anreicherungskits. Für die Versuche von SureSelect und NimbleGen wurden 155 Marker und 289 neutrale Regionen ausgewählt. Dem Anreicherungs-kit von MYbait wurden zusätzlich 155 Marker sowie 140 neutrale Regionen hinzugefügt. In den Kapiteln 2.2.2 und 2.2.3 ist die Auswahlstrategie für alle 429 Regionen und 269 Marker (Tabelle B3) aufgeführt. Je nach Design-Strategie der *baits*, die zwischen den Herstellern abweichend ist, mussten jeweils einzelne Loki ausgeschlossen werden. Die Anzahl und Größe der neutralen Regionen blieben davon unberührt. Eine Übersicht hierzu findet sich im Kapitel 4.2.

2.2.2 Neutrale Regionen

Die neutralen Regionen wurden von Christian Sell ausgewählt und sind in der Dissertation (Sell 2017) ausführlich beschrieben.

Mögliche neutrale Regionen werden aufgrund ihrer Nähe bzw. Distanz zu bekannten Genen (min. 0,5 cM) und dem Selektions-Koeffizienten von 0,95 (*background selection coefficient – BSC*) definiert. Die Kriterien wurden so festgelegt, dass theoretisch alle ausgewählten Regionen in einem Kopplungs-Ungleichgewicht liegen und als unabhängige Marker für koaleszenz-basierte Analysen verwendet werden können (Gronau et al. 2011).

Mit Hilfe des online Tools NRE (Arbiza et al. 2012) können neutrale Regionen des menschlichen Genoms identifiziert werden. Weitere Bearbeitungsschritte wurden mit internen Skripten durchgeführt und über verschiedene Optionen des *UCSC table* bzw. *genome browsers* (Karolchik 2004) kontrolliert und gefiltert. Alle ausgewählten Regionen sind 500 bp lang.

2.2.3 Funktionale Marker

Für die Anreicherung wurden insgesamt 155 bzw. 269 Marker ausgewählt, die phänotypisch für weitere Analysen von Bedeutung sind oder in ihrer Ausprägung eine funktionale

Relevanz zeigen. Eine Loki zeigen recht junge Selektionsmerkmale und weisen eine häufigere Verbreitung im europäischen oder asiatischen Raum auf (z.B. Lamason et al. 2005; Bryk et al. 2008; Hunt et al. 2008; Mathieson et al. 2015). Im weiteren Verlauf werden alle hier aufgeführten Marker als „funktionale Marker“ zusammengefasst. Alle Positionen wurden hauptsächlich in Zusammenarbeit mit Karola Kirsanow ausgewählt.

Zusätzlich wurden alle phänotypischen Marker aus der Dissertation von Martina Unterländer und Sandra Wilde hinzugefügt. Eine Liste aller Marker, welche im finalen Anreicherungs-Design verwendet wurden und ihrer chromosomalen Lokalisation ist im Anhang B3 eingefügt.

2.2.4 *bait*-Design

Da es sich bei allen hier getesteten Anreicherungskits um *in-solution* Verfahren handelt, besteht der größte Unterschied zwischen den Herstellern in der Beschaffenheit der eingesetzten *baits* und ihrer jeweilige Design-Strategie. Wichtige Aspekte hierzu sind in der nachfolgenden Tabelle (Tab. 3) aufgelistet:

Tabelle 3 Vergleich getesteter *bait*-Sequenzen der Hersteller Agilent, Roche und MYcroarray

Kit	SuSe	NG	MY
<i>bait</i> -Sequenz	RNA	DNA	RNA
<i>bait</i> -Länge [bp]	120	50-120	80
Anzahl einzigartiger <i>baits</i>	5700	k.A.	20 000
Anzahl gesamt	k.A.	bis zu 2E+06	5.79E+12
Kosten pro <i>capture</i> [€] ^a	260,75	507,08	162,17

^a(doppeltes *capture*, inkl. Steuer und Versand; Stand 2014)

Vorab werden alle SNP-Positionen über ihre zugehörige rs-Nummer auf den aktuellen Stand (hg19) gebracht.

SureSelect

Das *bait*-Design erfolgt durch den Anwender mit Hilfe des Web-Portals eArray (<https://earray.chem.agilent.com/earray/>) der Firma Agilent und greift auf alle aktuellen Informationen des USCS Human Genome Browsers zurück.

Um die SNP-Position effektiv abzudecken werden diese zunächst um 90 bp in 5' und 3' Richtung verlängert, die gewünschte Loci liegen mittig auf der 180 bp Region und sollen von je 3 *baits* abgedeckt werden (Abb. 4) (nach Wesolowska et al. 2011). Sequenzabschnitte die nach der Verlängerung nebeneinander liegen werden zu einem Abschnitt zusammengeführt.

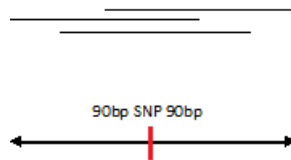


Abbildung 4: Bait -Design Strategie für SureSelect und MYbait

Mit der sogenannten Tiling-Frequenz kann definiert werden, nach wie vielen Basenpaar ein neues, sich überlagerndes *bait* beginnt und somit wie hoch die Abdeckung der einzelnen Basen einer Region ist. Über den Repeat-Masker von UCSC werden alle Regionen auf Repeat-reiche Sequenzen überprüft und angegeben. Es gibt dann verschiedenen Möglichkeiten das Design anzupassen: mit der Option „justified“ wird die Start- bzw. Stopp-Position der Regionen in *repeat*-reichen Regionen genau eingehalten, so dass der Abstand der überlagernden *baits* zueinander variieren kann. Für alle übrigen wird die Option „centered“ ausgewählt. So wird für den definierten Bereich versucht maximal ein Tiling von 3x zu erreichen. Die Regionen sind auf 199 bp verlängert.

Des Weiteren wird die *mapability* der generierten *bait* -Sequenzen über UCSC überprüft, d.h mit welcher Wahrscheinlichkeit eine *bait*-Sequenz exakt einmal mit einem entsprechenden Abschnitt auf hg19 zu 100 % übereinstimmt. Ein internes Skript gibt dann die *mapability* mit vier Werten für jedes *bait* an (S=1, einmal; S=0,25, viermal).

Regionen für die weniger als 3 *baits* generiert werden konnten, werden als „Orphans“ klassifiziert und das weitere *bait*-Design erfolgt ohne *masking*-Option.

Für das Design der neutralen Regionen wird ein 5xtiling festgelegt.

Da GC-arme Regionen weniger effizient hybridisieren wird am Ende des Prozesses ein Ausgleich GC-armer Regionen durch Zugabe zusätzlicher *baits* vorgenommen.

NimbleGen

Das Design wird nach Einsenden einer txt-Datei aller definierten SNP-Positionen und Sequenzbereiche neutraler Regionen vom Hersteller übernommen. Dafür werden die SNP-

Positionen um je 50 bp in beide Richtungen entlang des Strangs verlängert. Die *bait*-Länge ist variable und liegt durchschnittlich bei 75 bp. Für das hier beschriebene Design beträgt die durchschnittliche *bait*-Länge 74,28 bp, 60-94 bp gesamt. Einzigartige *bait*s werden mit Hilfe des SSAHA Algorithmus erstellt und dabei in zwei Kategorien unterteilt:

Match 1: Eine *bait*-Sequenz kann nur einmalig, mit weniger als 5 Baseninsertionen, Deletionen oder Substitutionen zwischen sich und dem zugrundeliegenden Genom, gegen das Referenzgenom aligniert werden. Match 5: *bait*-Sequenzen können bis zu 5x an verschiedenen Stellen im Genom alignieren.

Alle ausgewählten *bait*-Sequenzen könne als gff oder bed-file im UCSC Browser eingeladen und kontrolliert werden.

MYbait

Das Design wurde vom Hersteller MYcroarray© durchgeführt. Hierfür werden die SNP-Positionen um 60 bp in 5' und 3' Richtung verlängert und überlappende Regionen zusammengeführt. Insgesamt 19 Positionen funktionaler SNPs werden zu längeren Sequenzen gemerget (135-259 bp Länge).

Alle Sequenzabschnitte werden in 80 bp lange Intervalle, entsprechend der *bait*-Länge, mit einem Überlappungsbereich von 60bp (5x tiling) unterteilt und gegen da Referenzgenom geblastet. Die Original-Sequenz wird maskiert, um falsch-positive Treffer auszuschließen. Für jeden blast-Treffer wird eine Schmelztemperatur bestimmt und in 6 Temperaturbereiche unterteilt: 40-60, 60-62,5, 62,5-65, 65-67,5, 67,5-70 und > 70 °C. Auf diese vorläufig erstellten *bait*-Kandidaten werden nun drei Filter entsprechend unterschiedlich hoher Stringenzkriterien (Stringent) oder weniger konservativen Kriterien (relaxed 2 und 4) angelegt. Für alle *bait*-Sequenzen wurde der stringenteste Filter mit folgenden Bedingungen ausgewählt:

- kein Blast-Hit mit einer Schmelztemperatur unter 60 °C
- nicht mehr als 2 Hits bei 62,5 - 65 °C, oder 10 Hits im selben Intervall und mindestens einem verworfenen, angrenzenden Kandidaten
- nicht mehr als zwei Hits bei 65 – 67,5 °C und 10 Hits bei 62,5 - 65 °C und zwei angrenzenden, verworfenen Kandidaten auf jeder Seite
- nicht mehr als sein Hit über 70 °C und nicht mehr als ein Hit über 65 – 67,5 °C und 2 Hits bei 62,5 - 65 °C und zwei benachbarten, verworfenen Kandidaten in einem 80 bp Intervall auf jeder Seite

Für alle Regionen, welche funktionale Marker-Positionen enthalten, wurde ein $3\times$ tiling angewendet und ebenfalls der höchste Stringenzfilter ausgewählt.

2.2.5 Protokolle

Alle hier beschriebenen Anreicherungsverfahren basieren auf der grundlegenden Publikation von (Gnirke et al. 2009). Eine detaillierte Auflistung aller hierfür benötigten Pufferlösungen und Chemikalien ist in (Blumenstiel et al. 2010) nachzulesen.

Durch Vorversuche konnte festgestellt werden, dass der Einsatz mehrere *libraries* pro Probe die Komplexität der *capture*-Produkte deutlich erhöht. Durch unabhängig produzierte *libraries* einer Probe kann eine vollständige und tiefere Abdeckung der Zielregionen erreicht werden. Zudem konnte ebenfalls nachgewiesen werden, dass ein doppeltes *capture*, also ein erneutes Anreichern desselben *capture*-Produkts, die Spezifität und Anzahl der *reads* in den Zielregionen deutlich erhöht. In den Protokollen von NG und MB wird dieses Verfahren bereits für aDNA bzw. Proben niedriger Konzentration empfohlen.

Für einen Testlauf mit allen drei Kits wurde die Probe Ak1-1 (Ak-Alacha1) ausgewählt, da eine hohe Anzahl einzigartiger Moleküle in den *libraries* gemessen werden konnte ($6.21E+08$, Kapitel 5.1) und diese in vorherigen Anreicherungsversuchen konstant gute Ergebnisse erzielte. So ist eine Beeinflussung der Ergebnisse aufgrund eines schlechten Erhaltungszustands der Probe ausgeschlossen. Für die hier durchgeführten Versuche wurden 4-5 *libraries* zu insgesamt 3 Probenpools gemischt. Die Einsatzmenge lag bei jeweils 1350 ng (NG), 1150 ng (Suse) und 1500 ng (MY). Das Volumen der gepoolten *libraries* wird mit Hilfe einer *speed-vac* (Eppendorf Concentrator plus, Eppendorf) reduziert. Anschließend können die aufkonzentrierten Proben-Pools durch Zugabe von UV-HPLC- H_2O in der gewünschten μ l-Menge resuspendiert werden.

Eine detaillierte Auflistung der Probenpools für alle in der vorliegenden Arbeit beschriebenen *capture*-Produkte ist im Kapitel 4.3 zu finden.

Für die Versuchsreihe und für alle durchgeführten Anreicherungsversuche der hier beschrieben Proben wird eine Leerkontrolle aus UV-HPLC- H_2O oder einer mittels *speed-vac* aufkonzentrierten *library*-Leerkontrolle mitgeführt. Hierfür wird $\frac{1}{4}$ der für eine *library* mit $\sim 1 \mu$ g Einsatzmenge empfohlenen *bait*-Konzentration eingesetzt.

Blocking-Oligo-Mix

Die Zusammensetzung und Konzentration der Oligonukleotide, welche während der Hybridisierung an die einzelsträngigen Adaptersequenzen der *library*-Moleküle binden, ist für das MB und SuSe-Protokoll gleich und abweichend der Herstellerempfehlung. Mit dem Blocking-Oligo-Mix (BO-Mix) wird nur der reverse Strang eines Moleküls geblockt, die Index-Sequenz wird hierbei ausgespart. Der P5-Adapter wurde vor der Anreicherung nicht verlängert, sondern für die Hybridisierung noch unvollständig belassen. Der P7 Adapter, welcher bereits einen Index- und eine flowcell Bindestelle enthält, wird von zwei Blockern jeweils vor und nach der Indexsequenz geblockt. So kann der angesetzte Block-Mix unabhängig der verwendeten Indexe eingesetzt werden. Der Block-Mix setzt sich somit aus den drei folgenden Komponenten zusammen: Meyer_Bo4.P7.part1.R, Meyer_Bo6.P7.part2.R und BOIs1_revcomp. Die ersten beiden Sequenzen nach (Kircher et al. 2012), letztere wurde anhand der bekannten Adaptersequenz neu geordnet und angesetzt. Die finale Konzentration je Blocker-Sequenz im Ansatz beträgt 6,67 μM . (siehe Anhang B4)

Für die Anreicherung mit NG wird der Blockmix des Herstellers eingesetzt (TS-HE- *capture* Universal Oligo 1 / TS-INV_HE Index * Oligo). Durch die bekannten Illumina-Adaptersequenzen kann der Block-Mix auf unsere *libraries* angepasst werden.

Für die jeweiligen Versuche wurden die derzeit aktuellen Protokolle der Hersteller verwendet. Abweichend werden alle -Produkte mit dem MinElute Purification Kit (Qiagen) oder MSB ® Spin PCRapace (Invitex, Stratec Molecular) aufgereinigt. Relevante Unterschiede zwischen den Handbüchern im Ablauf, den Hybridisierungs-Konditionen und verwendeten Chemikalien sind in der nachfolgenden Tabelle (Tab. 4) dargestellt:

Tabelle 4 Unterschiede der Laborprotokolle der getesteten Anreicherungs-Kits

Kit	SuSe	NG	MY
Version	Target Enrichment Protocol XT	SeqCapEZ Library SR User's Guide v4.1	MYbait s user manual V.2.3.1
Hybridisierungstemp. [C°]	65	47	65
Hybridisierungszeit [Std.]	2 x 24	2 x 12-16	2 x 20-24
Waschtemp. [C°]	65	47	65
Inkubation <i>bead-bait</i> - Komplex [min]	30	45	45
Elution der DNA von den beads	0,1 N NaOH	On-bead-PCR	2 x 0,1 N NaOH ^a
PCR Polymerase	Herculase	Roche High-Fidelity Master-Mix	Herculase
PCR-Zyklen	2 x 14	1 x 5; 1 x 18	2x14

^a abweichend vom MYbait Handbuch

2.3 NGS-Sequenzierung

Screening

Für die oberflächliche *shotgun*-Sequenzierung aller erstellten DNA-*libraries* wurden die quantifizierten PCR-Produkte äquimolar zusammen pipettiert. Die erneut mit dem Agilent Bioanalyzer und Qubit Fluorometer quantifizierten Probenpools wurde auf dem Illumina MiSeq (100bp, SE) der Firma StarSeq an der Johannes Gutenberg-Universität Mainz auf jeweils einer *lane* sequenziert und anschließend nach ihrer Index-Kombination separiert.

Capture-Produkte

Entsprechend der Index-Kombinationen wurde ein gemeinsamer Probenpool aus 8-13 *capture*-Produkten erstellt und ebenfalls erneut mittels Bioanalyzer kontrolliert und quantifiziert. Die Sequenzierung erfolgte an der Johannes-Gutenberg-Universität auf dem Illumina HiSeq (2000/2500) des Instituts für Molekulare Genetik oder dem Illumina NextSeq 500™ der Firma StarSeq. Die kürzere Leselänge des NextSeq-Sequenzierers von 75 nt bzw. bp wird bei der bioinformatischen Datenauswertung bei den entsprechenden Proben berücksichtigt. (siehe Kapitel 3.4.1) In der folgenden Tabelle (Tab. 5) sind alle Konditionen der Sequenzierung für jede angereicherte Probe aufgelistet:

Tabelle 5 Plattform und Konditionen der sequenzierten *capture*-Produkte; PE= paired-end; bp= Basenpaar

Plattform	Konditionen	Proben
HiSeq	100bp, PE	Ak1-2, Ak5-6, Ze4, Sol5-1a, Kama7, Pr8, J12-1, Kama1
HiSeq Rapid	100bp, PE	A9, A19, A20, A21, AK5-8, Be1, Be10, Be11, Bl2, Bl3, Bl4, Bl5, J12-1, J12-6, J12-7, J12-8, Pr3, Pr4, Pr8, Pr10, Pr11, Pr12, Ze3, Nov7, Sol5-9a, Kal1, Ul2, Kol4, Kama6, Kama10
NextSeq	75bp, PE	Ak1-1, Be4, Be9, Be14, Bl1

2.4 Bioinformatische Datenauswertung

2.4.1 Rohdatenbearbeitung

Die Bearbeitung der Rohdaten erfolgt anhand einer standardisierten Pipeline und umfasst hier alle Prozesse bis zum Schritt der Genotypisierung. Eine detaillierte Beschreibung ist in der Dissertation von Christian Sell nachzulesen (Sell 2017).

Zunächst werden die Sequenzierdaten anhand ihrer jeweiligen Index-Kombination getrennt und im *fastq*-Format übermittelt. Daten einer *Paired-end* Sequenzierung sind in *forward* (R1) und *reverse* (R2) *reads* unterteilt und werden vor der Alignierung gegen das humane Referenzgenom GRCh37/hg19 zusammengeführt und entsprechende Paare gebildet. Das Alignment kann dann mit *bwa samse* (Li and Durbin 2009) für *single-end reads* durchgeführt werden.

Aufgrund der kürzeren Leselänge einer NextSeq-Sequenzierung wurde neben einem *samse*-Alignment zusätzlich *bwa sampe* (Li and Durbin 2009) für einige Proben eines NextSeq-Laufs, unter Berücksichtigung ihrer Fragmentlänge nach einer Bioanalyzer-Messung der *capture*-Produkte, durchgeführt. Im Unterschied zu *bwa samse* werden hierbei die identifizierten Paare beibehalten, die keinen überlappenden Sequenzbereich aufweisen und nicht zu einer einzelnen Sequenz zusammengeführt werden können. Beide *read*-Paare werden aufgrund ihrer Positionen auf dem Referenzgenom aligniert. Diese Paare müssen anschließend gefiltert werden, um nur sinnvolle Kombinationen beizubehalten (*proper-pairs*, *pp*). Das bedeutet, beide Partner befinden sich auf demselben Strang und in einem sinnvollen Abstand zueinander.

Nur bei Proben mit Längen $> \emptyset$ 83 bp wurde zusätzlich ein *sampe*-Alignment hinzugenommen. Aufgrund der geringen Rohdaten-Ausbeute der Proben Ze3 und Bl2, welche

vermutlich durch ein nicht-äquimolares Poolen der Proben entstanden ist, wurden ebenfalls bei beiden Proben zusätzlich pp-gefilterte *sampe-reads* analysiert, um die Ausbeute zu erhöhen.

Die folgende Tabelle (Tab. 6) stellt die einzelnen Schritte und die dafür verwendeten Parameter für die Rohdatenanalyse der MiSeq und HiSeq/NextSeq-Läufe zusammen:

Tabelle 6 Programme und Parameter der Rohdatenbearbeitung

	Programm	MiSeq	HiSeq/NextSeq
Adaptersequenzen schneiden	KeyAdapterTrimFastQ_cc.py ¹	-a 'AGATCGGAAGAGCACAGTCTGAACTCC' -k ", -c ",	
Qualitätsfilter	QualityFilterFastQ.py ¹	-c 15, -n 3	-c 15, -n 5
Paare zuordnen und bilden	remove_reads_no_mate.py ² fastq-join (ea-utils) ³	6bp Überschneidung bei 92% Übereinstimmung	
Alignment	BWA (aln, samse, sampe)		Standardparameter
Duplikate entfernen	MarkDuplicates (Picard-tools)		Standardparameter
Sortieren / Indexen	samtools (sort, index) ⁴		
Längenfilter	NGSutils ⁵		-minlen 30
Readgroup setzen	AddOrReplaceReadGroups.py (Picard-tools)	Benutzerdefinierte Namensgebung	
Indel-Realignment	The Genome Analysis Toolkit (GATK) ⁶		Standardparameter

¹ (Kircher 2012) ² (Sell 2017) ³ (Aronesty 2013) ⁴ (Li et al. 2009) ⁵ (Breese and Liu 2013) ⁶ (Depristo et al. 2011)

Für die Etablierung des hier verwendeten *capture*-Protokolls wurden einzelnen Proben mehrmals angereichert um spezielle Schritte im Laborprotokoll zu testen. Diese wurden zunächst getrennt in Hinblick auf Erfolg der Methoden hin ausgewertet (nicht dargestellt). Für die vorliegende Arbeit werden die BAM-files aller *runs* einer Probe vor dem Duplikate entfernen gemergt und gemeinsam analysiert.

Tabelle 11 im Kapitel 4.3 listet die Zusammensetzung der *capture*-Produkte je Probe auf. Die Abdeckung der Zielregionen und der Standardabweichung wird über ein internes *python*-Skript anhand eines bed-files der Regionen erstellt.

2.4.2 Degradierungsmuster und Geschlechtsbestimmung

Die für alte DNA typischen Deaminierungsmuster an den Molekülen können mit dem software-Paket **mapDamage** (Jónsson et al. 2013) bestimmt werden. In der Literatur hat sich die Angabe der C/T-Transitionssrate an der ersten Position am 5'-Ende eines Moleküls etabliert. Bei längeren Leseweiten einer Sequenzierung über das gesamte Molekül kann auch der Wert am 3' berechnet werden. Die Deaminierungsrate kann einen Hinweis auf die Authentizität und Qualität der bearbeiteten DNA liefern und wird in der vorliegenden Arbeit auf die jeweils finalen BAM-files einer Probe durchgeführt.

Die **Geschlechtsbestimmung** anhand molekularer Daten wurde auf die *screening-libraries* der MiSeq-Sequenzierungen nach (Skoglund et al. 2013) durchgeführt. Die Berechnung basiert auf der Anzahl an *reads*, die gegen das Y-Chromosom aligniert werden können, relativ zur Anzahl an *reads*, die gegen beide Geschlechtschromosomen alignieren. ($R_y = n_y / (n_x + n_y)$). Zusätzlich wird ein Konfidenzintervall (CI) für die Zuordnung zu den jeweiligen Gonosomen bestimmt.

Die Ergebnisse der genetischen Bestimmung stimmen in allen Fällen mit den vorliegenden morphologischen Klassifizierungen einer Probe überein und sind im Kapitel 4.1 dargestellt.

2.4.3 Kontaminationsbestimmung

Neben dem Mitführen von Leerkontrollen während der Probenbearbeitung im Labor stehen auch bioinformatische Methoden zu Verfügung, um mögliche Kontaminationen in den erstellten *libraries* bzw. *capture*-Produkten zu identifizieren. Für die hier vorliegende Arbeit wurden zwei verschiedene Methoden auf bereits zusammengeführte BAM-files der *capture*-Produkte angewendet, die für nachfolgende populationsgenetische Analysen verwendet und mit dem Programm ATLAS (Kapitel 2.4.4) recalibriert wurden.

contaMIX

Das Programm contaMIX v. 1.0-10 (Fu et al. 2013) schätzt unter Berücksichtigung von Sequenzierfehlern eine mögliche Kontamination innerhalb der mtDNA Sequenzen ab. Die Methode erstellt eine individuelle Konsensussequenz gegen die die mitochondrialen Sequenzen aligniert werden. Dies geschieht relativ zu einem Datensatz aus 311 weltweiten Mitogenomen, die eine mögliche Kontamination repräsentieren. Neben einer Fehlerrate erhält man als Ergebnis eine Kontaminationsrate und einen Authentizitätswert. Dieser gibt

einen relativen Wert an, mit dem die *reads* als authentisch eingeordnet werden können und wird hier in Prozent angegeben.

ANGSD

Das software-Paket ANGSND erlaubt die Abschätzung einer Kontaminationsrate auf der Basis von X-chromosomalen *reads* in männlichen Individuen (Korneliusson et al. 2014). ANGSND berechnet die Heterozygotenrate für jeden SNP und erwartet aufgrund der Autosomen-Konstellation XY nur ein Allel für jede Position. Jedes detektierte, zusätzliche Allel repräsentiert eine mögliche Kontamination. Die hierfür benötigten Informationen über polymorphe Positionen und Allelfrequenzen werden für das humane X-Chromosom zur Verfügung gestellt. Nach dem Erstellen eines binären *files* wird eine Abschätzung der Kontaminationsrate und eines p-Wert mittels exakter Fisher-Test und Jackknife-Verfahren ermittelt und berechnet.

2.4.4 Genotypisierung – ATLAS

Eine Genotypisierung erfolgt anhand des Software-Pakets ATLAS (*Analysis Tools for Low-coverage and Ancient Samples*), beschrieben in (Kousathanas et al. 2016) und erstmals angewendet in (Hofmanová et al. 2016). Es wurde speziell für Proben mit niedriger Basenabdeckung und Sequenzierdaten aus alter DNA entwickelt.

Zunächst werden mit ATLAS die für alte DNA-charakteristischen post-mortem Schädigungen (C – T und G –A Transitionen) ermittelt. Diese Schädigungsmuster (PMD) und mögliche Sequenzierfehler werden dann bei der Rekalibrierung der Basenqualität und der Genotypisierung mit einbezogen. Zudem ist eine Referenz-freie Rekalibrierung anhand haploider Regionen möglich, was Verschiebungen hin zum gewählten Referenzgenom verhindert. Für humane Proben eignen sich das mitochondriale Genom oder das X-Chromosom für männliche Individuen. In allen hier untersuchten Proben konnten aufgrund der geringen Anzahl allerdings keine mitochondrialen Sequenzen für den Rekalibrierungsschritt verwendet werden.

Zunächst werden für die Bestimmung der post-mortem Schäden und der Rekalibrierung alle DNA-Duplikate mit der Option `MarkDuplicates` von *picardtools* nicht vollständig entfernt, sondern nur als solche markiert. Beide Methoden, die in ATLAS für die Rekalibrierung zur Verfügung stehen, *recal* und *BQSR*, benötigen eine gewisse Anzahl an *reads* bzw. Abdeckung. Die Sequenz-Duplikate für diesen Schritt beizubehalten ermöglicht oftmals den geforderten Schwellenwert zu erreichen.

Eine Rekalibrierung mit *recal* war nur für zwei männliche Individuen möglich. Alle anderen Proben wurden dann mit BQSR kalibriert. Um ausreichend reads für Rekalibrierung mit BQSR zu erreichen, werden verschiedene Proben, die auf derselben *flowcell* sequenziert wurden für diesen Schritt gruppiert, die individuellen PMD Informationen aber beibehalten. Dies ist möglich, da die Abschätzung eines Sequenzierfehlers *lane* bzw. *flowcell* spezifisch ist.

Bei Proben mit verschiedenen *readgroups* werden diese zusammen bearbeitet, falls sie auf derselben flowcell, aber auf unterschiedlichen *lanes* sequenziert wurden.

Wurde eine Probe auf mehreren *flowcells* sequenziert, werden die entsprechenden BAM-files bis zum call getrennt bearbeitet. Bei allen vier Proben, für welche ein *sampe*-Alignment durchgeführt wurde, werden vor dem zusammenfügen der beiden BAM-files die nicht-überlappenden *sampe-reads* mit der Option *merge-reads* von ATLAS zu einer Sequenz zusammengeführt und dann gemeinsam bearbeitet. So werden alle Proben entsprechend der empfohlenen Vorgehensweise von ATLAS für *single-end* Daten analysiert.

Die Genotypisierung kann nach zwei verschiedenen Methoden durchgeführt werden:

- (1) MLE: *maximum likelihood estimate of diploid genotyp*, generiert diploide Genotypen
- (2) AP: *allelePresence call*, generiert haploide Genotypen mit dem wahrscheinlichsten Allel an der entsprechenden Position.

Mit ATALS werden rekalierte BAM-files und zwei vcf-Dateien der beiden oben genannten Methoden erstellt. Die weitere Bearbeitung der vcf-Dateien ist im Kapitel 3.2 beschrieben.

3 Populationsgenetik

3.1 Hauptkomponentenanalyse – PCA

Die Hauptkomponentenanalyse erstellt eine visuelle Verteilung der untersuchten Daten ähnlich einer geografischen Karte. Im Vergleich mit modernen oder prähistorischen Referenzgruppen können basierend auf den Positionen der Individuen genetische Nähe bzw. Distanz dargestellt werden, jedoch keine Migrationsereignisse. Die PCA soll einer ersten Orientierung dienen, wie sich die untersuchten Proben zueinander und zu den Referenzdatensätzen verhalten.

Die Hauptkomponentenanalyse wurde auf ATLAS rekaliibrierte BAM-files durchgeführt und mit LASER (Wang et al. 2014) in zwei Schritten generiert.

Zunächst wird der Referenzraum berechnet. Als Grundlagen hierfür dienen die modernen Datensätze von Hellenthal (Hellenthal et al. 2014) und Busby (Busby et al. 2015). Um die grafische Darstellung detaillierter und die Analyse schneller zu gestalten wurden alle Individuen aus Nord/Südamerika und afrikanische Populationen ausgeschlossen. Insgesamt wurde die PCA basierend auf 1731 Individuen aus den kombinierten Datensätzen von Hellenthal und Busby und 512367 Loki durchgeführt.

Nachdem Erstellen von *pileup*-files mit *samtools* können Informationen der BAM-files aller vorliegenden Daten durch eine PROCURSTUS Transformation (Wang and Zhan 2015) direkt in den Referenzraum projiziert werden, ohne eine Genotypisierung durchzuführen. Für die *pileup*-files wurden die Optionen *mapping quality* 30 und *base quality* 20 gewählt (Hofmanová et al. 2016). Neben den in der vorliegenden Arbeit produzierten Daten wurden zusätzlich geografisch und zeitlich relevante aDNA-Datensätze aus der Literatur, die als BAM-files vorliegen, mit *samtools pileup* bearbeitet und hinzugefügt. (Lazaridis et al. 2014; Fu et al. 2014; Jones et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015; Jones et al. 2017; Sikora et al. 2017). Eine Auflistung dieser Proben, insgesamt 112 Individuen, ist in der Tabelle B5 im Anhang B zu finden. Die Nomenklatur und Gruppierung der Referenzindividuen orientiert sich größtenteils an (Lazaridis et al. 2016). Gängige Zeitangaben und Bezeichnungen im englischen werden beibehalten, um die hier durchgeführten Analysen mit anderen Publikationen leichter vergleichbar zu machen.

3.2 FST und F2

Um genetische Distanzen zwischen möglichen Gruppen innerhalb der Steppenpopulationen und Distanzen in Relation zu modernen Populationen zu bestimmen werden zunächst paarweise Wright's **Fst-Werte** in verschiedenen Kombinationen berechnet (WRIGHT 1951). Es werden Fst-Werte nach Hudson berechnet, basierend auf der Publikation von (Bhatia et al. 2013). Die so errechneten paarweisen genetischen Differenzen zeigten sich am robustesten gegenüber verschiedenen Populationsgrößen. Der Fst-Wert ist ein Mittelwert aus allen errechneten Fst-Werten je gemeinsamen Locus zwischen zwei Populationen. Hierfür müssen die vcf-files der skythischen Populationen mit dem Referenzdatensatz zusammgeführt werden. Die Berechnung erfolgt auf die mit ATLAS erstellten, diploiden *calls*. Diese wurden zunächst auf eine Abdeckung der Positionen auf mindestens 2x mit einer Genotypqualität von 30 gefiltert und formatiert. Es werden ausschließlich die Regionen der Anreicherung verwendet, unter Ausschluss aller Positionen auf dem X- und Y-Chromosom.

Als Referenz dient der weltweite *1000 Genome Phase 3* Datensatz (1000 Genomes Project Consortium et al. 2015). Dieser wurde ebenfalls auf die *capture*-Regionen reduziert und zusätzlich nur die getesteten eurasischen Populationen isoliert. Die Bearbeitung der für die Arbeit erstellten vcf-files und des Referenzdatensatzes erfolgt mit *bcftools* (<https://github.com/samtools/bcftools>) unter Verwendung der Option `-m all`, um doppelte Positionen zu vermeiden und mehrerer Allele an einer Position zuzulassen

Skythischen Populationen wurden aufgrund ihrer geografischen Lage, ihrer zeitlichen Einordnung und Kulturzugehörigkeit gruppiert und untersucht. Basierend auf den Ergebnissen der PCA wurden auch Fst-Werte der östlichen Gruppe unter Ausschluss aller Proben aus Barsučij Log und Ulžan berechnet.

Je höher ein paarweiser Fst-Wert, desto größer wird eine genetische Distanz angenommen. Bei einem negativen Wert kann keine genetische Unterteilung der untersuchten Populations-Paare ausgeschlossen werden. Negative Fst-Werte sind in den Ergebnissen mit dem Wert 0 gleichgesetzt.

Zur Visualisierung der Daten werden neben tabellarischen Abbildungen die paarweisen Fst-Werte als *heatmap* dargestellt. Hierfür werden alle berechneten Fst-Werte als csv-Datei formatiert und ein Diagramm mit *heatmap.2* in R erstellt.

Die **F2-Statistik** ist Teil der F-Statistiken ((Reich et al. 2009) und die Berechnung erfolgt nach Patterson (Patterson et al. 2012). Der F2-Wert drückt aus wieviel genetische Drift zwischen zwei Populationen entstanden ist, in dem zum Beispiel Unterschiede in den Allelfrequenzen der Populationen bestimmt werden.

3.3 F3 und D-Statistik

F3 und D-Statistiken sind Methoden der statistischen Genomik, mit deren Hilfe genetische Drift basierend auf der Varianz der Allelfrequenzen zwischen definierten Populationen bestimmt werden kann. Unter der Annahme, dass die Phylogenie der untersuchten Populationen auf einer Baum-ähnlichen Struktur beruht und gemeinsame Drift eine geteilte Evolutionsgeschichte widerspiegelt. F3 wird mit der Formel $F3(C: A;B)$ beschrieben. Hierbei wird Population C als mögliches Ergebnis eines *admixture*-Ereignisses von Population A und B angenommen. Ein negativer Wert der Kalkulation unterstützt die Annahme der Verwandtschaft von C mit A und B.

Outgroup-F3-Statistik ist ein spezieller Fall der F-Statistiken. Hierbei wird eine Population als Außengruppe festgelegt, die nach der Abspaltung keinerlei Genfluss mit den beiden anderen Populationen aufweist. Positive Werte für F3 sind dann proportional zur Länge der gemeinsamen genetischen Drift zwischen Population A und B. Wird also eine weitere Population fixiert, z.B. A und Population B in verschiedenen Triplet-Kombinationen ausgetauscht, kann eine genetische Ähnlichkeit basierend auf den Allelfrequenzen zwischen Population A zu verschiedenen Testpopulationen für B, bestimmt werden. Als Außengruppe wird hier die Mbuti Population des Referenzdatensatzes verwendet.

Die D-Statistik der Form $D(A;B : C;D)$ ermöglicht, mit Population D als Außengruppe, verschiedene Topologien und damit die Richtung des Genflusses dreier Populationen zu testen. Eine signifikante Abweichung des F4-Wertes von 0, also der Nullhypothese, suggeriert einen Genfluss zwischen den Populationen A und B oder A und C. Ein positiver Wert entsteht bei einer gemeinsamen genetischen Abstammung von Population A und C (A,C : B,D), ein negativer Wert hingegen bei größerer Ähnlichkeit von B und C (B,C : A,D).

Somit kann mit der D-Statistik getestet werden, ob eine größere, genetische Ähnlichkeit der Population C mit Population A oder B besteht.

F3 - und D-Statistik sind Teil des Programms *Admixtools 4.1* (Patterson et al. 2012). Für F3 wird das Programm *qp3Pop*, für die D-Statistik *qpDstat* verwendet. In *Admixtools* ist das software-Paket *convertf* (Price et al. 2006) implementiert, womit binäre PLINK-Dateien (PLINK 1.9) zu EIGENSTRAT-Formaten konvertiert werden können (Purcell et al. 2018). Da pseudo-diploide Daten analysiert werden, muss die Option *inbreed-YES* für die Berechnung der F-Werte gesetzt werden.

Es werden die *allele-presence calls* der *alignments* gegen das gesamte Genom (hg19) der skythischen Proben verwendet. Diese werden zunächst nach Qualität (15) und Abdeckung (1x) gefiltert und pseudo-diploide Datensätze erstellt. Alle Positionen auf dem X- und Y-Chromosom sowie triallelische Positionen werden entfernt. Die Bearbeitung und das Zusammenfügen der Datensätze erfolgt mittels *bcftools* bzw. PLINK.

Als moderner Referenzdatensatz dient, wie bereits für die Hauptkomponentenanalyse, der kombinierte Datensatz von Hellenthal (Hellenthal et al. 2014) und Busby (Busby et al. 2015)

Der moderne Referenzdatensatz wird auf 23 Populationen reduziert, um eine Zuordnung zu West- bzw. Osteurasischen Gruppen zu ermöglichen (Tab. 7). Hier werden 7 europäische Populationen, 6 asiatisch-europäische Populationen und 11 asiatische Populationen ausgewählt.

Als Referenzdatensatz früherer oder zeitgleicher prähistorischer Kulturgruppen mit den hier bearbeiteten Individuen wird der Datensatz von Lazaridis et al. 2016 verwendet (Lazaridis et al. 2016). Die Gruppenzuordnung der 281 Individuen, die in der Publikation aus anderen Studien zusammengefasst bzw. erstmals beschrieben werden, wird weitestgehend beibehalten. Es werden jedoch relevante Gruppierungen der frühen und späten Bronzezeit (z.B. Steppe_MLBA) in ihre entsprechenden Kulturgruppen unterteilt. Die Populationsgröße wird auf maximal 5 Individuen beschränkt (insgesamt 126 Individuen), da bei früheren Analysen eine signifikante Verschiebung der höchsten f3-Werte bei starken Abweichungen der Populationsgrößen festgestellt werden konnte (Ergebnisse nicht dargestellt).

In beiden Referenzdatensätzen, sowie den skythischen Daten, werden die SNP-Positionen auf die gefilterten Positionen aus (Lazaridis et al. 2016) reduziert und das A1 Allele der Referenz forciert, um die hier dargestellten Ergebnisse mit anderen publizierten, populationsgenetischen Studien vergleichbar zu machen.

Da die Abdeckung des gesamten Genoms zwischen den einzelnen Proben stark variiert, basieren alle Statistiken auf einer unterschiedlichen Anzahl an SNP-Positionen nach dem

Qualitätsfiltern. Mit dem berechneten Z-Wert (Z-score) wird die Signifikanz jedes f3-Werts mittels *jackknife*-Verfahren unter Verwendung der empfohlenen Parameter bestimmt.

Basierend auf den PCA Ergebnissen werden einzelne Individuen nicht zu Fundplätzen gruppiert, sondern gesondert untersucht. Zuvor wurde in gezielten f3-Analysen getestet, ob sich die Proben mit einer Sonderstellung in der PCA in den f3-Werten widerspielen. Nur dann wird die Unterteilung beibehalten.

In Kapitel 4.6 werden die für die Fragestellung relevanten Ergebnisse der *outgroup*-f3 -und D-Statistiken beschrieben. Die höchsten 20 f3-Werte sind der Größe nach geordnet und werden als Diagramm mit entsprechenden Standardabweichungen dargestellt, signifikante f4-Werte in tabellarischer Form.

Tabelle 7 Auflistung der modernen Populationen aus dem Hellenthal/Busby Referenzdatensatz, die für f3-Analysen ausgewählt wurden

Population	Kontinent	Land	Gebiet
altai	Asien	Sibirien, Kasachstan, Mongolei	Zentral/Nord-Asien
brahui	Asien	Pakistan/Iran	Süd-Asien
buryat	Asien	Mongolei/Sibirien	Nord-Asien
daur	Asien	China/Mongolei	Ost-Asien
evenk	Asien	Sibirien, Mongolei, China	Nord/Ost-Asien
hannchina	Asien	China	Ost-Asien
iranian	Asien	Iran	Naher Osten
nganassan	Asien	Russland, Sibirien	Nord-Asien
syrian	Asien	Syrien	Naher Osten
tuva	Asien	Sibirien	Nord-Asien
uzbekistani	Asien	Uzbekistan	Zentralasien
balkar	Asien/Europa	Russland, Kasachstan	Kaukasus
georgian	Asien/Europa	Georgien	Kaukasus
northossetian	Asien/Europa	Nordossetien	Nordkaukasus
russian	Asien/Europa	Russland	Kaukasus/ Osteuropa
turkmen	Asien/Europa	Türkei	Südwest-Asien
belorussian	Europa	Belarus, Ukraine, Polen	Mittel/Osteuropa
bulgarian	Europa	Bulgarien	Südost-Europa
ceu	Europa	-	Europa
chuvash	Europa	Tschuwasien	Osteuropa
polish	Europa	Polen	Mittel/Osteuropa
romanian	Europa	Rumänien	Mittel/Osteuropa
ukrainian	Europa	Ukraine	Osteuropa

3.4 Funktionale Marker

Insgesamt 269 funktionale Marker wurden in das *capture*-Design integriert. Für weitere Analysen werden abzüglich der X - und Y-chromosomalen Positionen 193 Positionen für alle Proben mittels *vcftools* isoliert. Hierfür werden die diploiden *vcf*-files aus der ATLAS *pipeline* mit einer Abdeckung der Positionen von mindestens 2x und einer Genotyp-Qualität von 30 eingesetzt. Für die Proben Be11; be1, J12-1, Pr11, Pr12, Pr3, Bl5, Kama1, Kama6, Kama7, Kama10, Nov7, Bl4, Bl1, Bl2, Be14, Be4 und Pr8 wurden aufgrund der nicht ausreichend Abdeckung die Filterkriterien 1x und GQ 15 angewandt.

Insgesamt 54 Varianten wurden für eine detaillierte Bestimmung der Genotypen ausgewählt (Tab. 8). Davon sind 17 Positionen Teil des HIrisPlex-System zur Bestimmung der Augen – und Haarfarbe. Alle weiteren funktionalen Marker werden entsprechend ihres Einflusses auf einen bestimmten Phänotyp in 7 weitere Kategorien unterteilt: Hautpigmentierung, Laktasepersistenz, Xenobiotischer Metabolismus, Alkoholverträglichkeit, Risiko für Diabetes Typ2, Bluthochdruck und Klimaanpassung.

Für alle bronzezeitlichen Proben aus Kamennyj-Ambar, die frühen Sarmaten aus Pokrovka, die neolithischen Proben aus Solonzy5 und den Proben A19, A20, Be1, Be11, J12-1, J12-6 und J12-7 werden alle Phänotypen erstmals in der vorliegenden Arbeit bestimmt. Für die eisenzeitlichen *pars petrosa* aus Kasachstan Ul2 und Be10 wurden die Marker HERC2 (rs12913832), SLC24A und LCTb mittels Multiplex-PCR in (Gräfen 2009) bestimmt.

Alle Marker die in Verbindung mit einem erhöhten Risiko für Typ2 Diabetes und Bluthochdruck stehen, sowie die Marker rs1545397, rs885479, rs12203592, rs1800414, rs12896399 und rs2740574 werden für alle Proben erstmals beschrieben.

Aus allen hier beschriebenen Markern wurden gezielt 13 Marker ausgewählt, welche starke Varianzen der Allelfrequenzen zwischen europäischen und asiatischen Populationen zeigen. Errechnete Allelfrequenzen der West- und Ostskythen werden dann mit Frequenzen moderner Populationen verglichen. Die 95% Konfidenzintervalle werden durch Approximation durch Normalverteilung berechnet.

Die Allelfrequenzen moderner Population (EUR, EAS, SAS) der beschriebenen Marker sind im Methoden Teil, den Ergebnissen sowie der Diskussion aus *Ensembl GRCh37 release 88* (<http://grch37.ensembl.org/index.html>) entnommen. Die Allelfrequenzen zentral- und westasiatischer Populationen basieren auf den Daten der *Allele Frequency Database* ALFRED (Osier et al. 2002; Rajeevan et al. 2003; <http://alfred.med.yale.edu>). Es wurden entsprechend der Fundplatzverteilung der vorliegenden Proben folgende Populationen als zentral-asiatische Referenzgruppe (ZAS) ausgewählt: Kasachen, Kirgisen und Mongolen. Stellvertretend für

Nordwest-Russland (NWR) werden Individuen der Republik Komi (Komi-Zyrian), westliche des Urals ausgewählt.

Augen – und Haarpigmentierung

Zunächst wird eine Phänotypenbestimmung über die zugehörigen 24 Marker des Hirisplex-Systems durchgeführt (Walsh et al. 2014). Insgesamt 7 SNPs der HIRISPLEX sind nicht im *capture* enthalten. Alle Phänotypen definierenden Positionen wurden jedoch im Design berücksichtigt und sind enthalten. Eine präzise Bestimmung der Haut- und Augenfarbe ist daher möglich. Hierfür wird eine csv-Datei als Eingabe-Datei erstellt, welche alle Genotyp-Informationen der HIRISPLEX-Marker in der Form 0 für abweichendes Allel, 1 für heterozygot, 2 für homozygot oder NA für keine Angabe, enthält. Diese Datei kann unter dem online-Portal hochgeladen werden (<https://hirisplex.erasmusmc.nl/>). Die Ausgabe-Datei enthält alle Informationen zur Bestimmung der Haut – und Haarfarbe. Die Haarfarbe wird durch 3 Kategorien mit entsprechender Angabe der Prognosewahrscheinlichkeit(-p-value) angegeben, die Hautfarbe in 4 möglichen Kategorien mit entsprechender Wahrscheinlichkeit. Da alle Phänotypen durch mehrere Marker definiert werden, wird für alle Kategorien ein AUC-Wert bestimmt. Dieser gibt den Verlust der Genauigkeit einer phänotypischen Angabe an, wenn für entsprechende Marker einer Kategorie keine Angaben gemacht werden konnten. In der folgenden Tabelle sind alle hier beschriebenen funktionalen Marker entsprechend ihrer übergeordneten Kategorie aufgelistet.

Pigmentierung der Haut

Es wurden 7 Marker analysiert, welche mit der Pigmentierung der Haut assoziiert werden und deren Genotypen nach (Spichenok et al. 2011). in „nicht-dunkel“ und „nicht-weiß“ klassifiziert werden können.

Das abgeleitete Allel des Markers SLC45A2 (rs16891982) deutet auf eine helle Hautpigmentierung hin und ist in heutigen und alten europäischen Populationen unter Selektion. Die homozygote Form (G/G) ist ausschließlich in Europa detektiert.

Die Marker SLC24A5 (rs1426654) und TYR (rs1042602) stehen im Zusammenhang mit einer Depigmentierung der Haut in europäischen jedoch nicht in asiatischen Populationen (Lamason et al. 2005; Stokowski et al. 2007; Hancock et al. 2011). Die homozygote Form des ancestralen Status kommt nur in ost-asiatischen Population vor. Der Marker TYR ist nicht Teil des Klassifizierungssystems nach (Spichenok et al. 2011) und wird getrennt aufgelistet.

Zusätzlich wurde ein weiterer Marker bestimmt, der im Zusammenhang mit einer helleren Hautpigmentierung steht. Das abgeleitete Allel des Markers OCA2.3 (rs1800414) der OCA2-HERC2 wird ausschließlich in ost-asiatischen Populationen positiv selektiert. (Donnelly et al. 2012).

Laktasepersistenz

Die Fähigkeit unvergorene Milchprodukte auch im Erwachsenenalter verdauen zu können ist mit einer Mutation im MCM6-Gen assoziiert, welche jedoch Einfluss auf das Laktase-Gen LCT zeigt. Die beiden Marker 13910T (LCTa) und 22018A (LCTb) sind primär für den Haplotyp verantwortlich (Enattah et al. 2002) und standen innerhalb der letzten 10000-5000 Jahre unter starker Selektion (Bersaglieri et al. 2004; Burger et al. 2007). In ost-asiatischen Populationen kann ausschließlich die homozygot anzestrale Form nachgewiesen werden, wohingegen in europäischen Populationen die Allelfrequenzen beider Allele mit 49 % und 51 % nahezu ausgeglichen sind.

Xenobiotischer Metabolismus

In die Kategorie der Xenobiotika werden alle Marker eingeordnet, die mit der Verstoffwechslung körperfremder Stoffe assoziiert werden können.

Die Gene CYP3A4 und CYP3A5 codieren für Isoenzyme, die wesentlich am Abbau von Arzneimitteln beteiligt sind. CYP3A4 trägt zudem zur Verstoffwechslung von Vitamin D bei (Wang et al. 2013). Das rs2740574 Allel C (G) definiert den Haplotyp CYP3A4*1B. Das anzestrale Allel steht außerhalb Afrikas unter negativer Selektion und die Allelfrequenzen scheinen abhängig vom Breitengrad (Thompson et al. 2004). Rs776746 oder 6986A>G des CYP3A5-Gens kodiert für den Haplotyp CYP3A5*3. Das anzestrale Allel C ist mit 71 in ost-asiatischen Populationen und 94 % in europäischen Populationen am häufigsten.

Rs1801280 im NAT2 Gen kodiert für das Protein N-acetyltransferase, welches an der Verstoffwechslung und Deaktivierung vieler körperfremder, karzinogener Substanzen beteiligt ist. Je nach Genotyp können schnelle und langsame Acetylierungstypen unterschieden werden, wobei das abgeleitete Allel für rs1801280 einen langsamen Typus definiert. Eine positive Selektion wird mit der Transition zu ackerbaubetreibenden Gesellschaften und der damit verbundenen Ernährungsumstellung in Verbindung gebracht (Magalon et al. 2008). Es werden zwei SNPs für das NAT2 Gen bestimmt (rs1801280 und rs1041983), die in Kombination eine sensitive und akkurate Bestimmung des Phänotyps ermöglichen (Selinski et al. 2011). Für rs1041983 (C282T) ist das abgeleitete Allel mit einem langsameren Metabolismus assoziiert. Im Angiotensin-Gen werden die SNPs rs699 (M235T) und rs5051 (A-

6G) lokalisiert. Sie werden mit Erkrankungen, die durch Bluthochdruck verursacht werden, in Verbindung gebracht (Balam-Ortiz et al. 2012). In europäischen Populationen ist der heterozygote Genotyp mit 47 % für beide SNPs am verbreitetsten, während der homozygote Status für das ancestrale Allel in ost-europäischen Populationen vorherrscht.

Das ABCB1, oder Multi-Drug-Resistance-Gen, ist unter anderem für die Eliminierung körperfremder Substanzen verantwortlich. Durch veränderte Expression können Angriffe auf die Blut-Hirn-Schranke vermindert werden. Es werden 3 SNPs im AGT-Gen beschrieben, rs1128503, rs2032582 und rs1045642, welche unter Selektionsverdacht stehen (Wang et al. 2007). Für rs1128503 ist die Allelfrequenz für das abgeleitete Allel in europäischen Populationen mit 58 % gegen über 37 % in ost-asiatischen Populationen leicht erhöht. Bei ABCB1b ist das abgeleitete Allel T mit einer Frequenz von 2 % in Europa eher selten, jedoch in Ost-Asien mit 13 % etwas häufiger verbreitet.

Alkoholverträglichkeit

Der Marker ALDH2 (rs671) bzw. das gleichnamige Enzym, sorgt für den Abbau von toxischen Aldehyden. Eine Mutation des Acetaldehyddehydrogenase 2 bewirkt einen erhöhten Acetaldehyd-Spiegel im Körper und führt zu einer plötzlichen Ausdehnung der Blutgefäße in Gesicht und am Hals, Schwindel und Übelkeit, auch bekannt als „*alcohol flush*“ oder „*asian blush*“ (Yoshida et al. 1984; Oota et al. 2004). Träger dieses Polymorphismus erkranken weniger an Alkoholismus und an Alkohol bedingten Lebererkrankungen. Die heterozygote oder homozygote Form des abgeleiteten Allels ist nur in ost-asiatischen Populationen nachweisbar. Zusätzlich werden zwei weitere Mutationen ADH1Ba und ADH1Bb untersucht, welche mit dem Abbau von Alkohol assoziiert werden und in modernen asiatischen Populationen unter Selektion stehen (Peng et al. 2010).

Diabetes Typ2

Diabetes Typ 2 ist eine in westlichen Ländern häufig verbreitete Stoffwechselerkrankung und wird neben einer genetischen Prädisposition von äußeren Faktoren beeinflusst. Das Vorliegen von protektiven Allelen bzw. gewissen Risiko-Allelen kann die Ausprägung einer möglichen Erkrankung jedoch beeinflussen.

Basierend auf einer Risikobewertung für Diabetes Typ 2 nach (Cornelis et al. 2009) wird ein GRS-Wert je typisierte Probe berechnet. Acht der 10 beschriebenen Marker sind Teil des nukleären Anreicherungsverfahrens und können bei ausreichender Abdeckung beschrieben werden. Die genetische Risikobewertung (GRS-Wert) wird durch die Summe aller Risiko-Allele je Individuum definiert. Ein standardisierter Wert bezieht fehlende Posi-

tionen mit ein und wird wie folgt berechnet: $GRS\text{-Wert} = \frac{\text{Anzahl Risiko-Allele}}{\text{Anzahl typisierter Risiko-Allele}} \times \text{Anzahl max. Allele je zugrundeliegendem Model}$. Es wurden ausschließlich Individuen mit mindestens 3 vorhandenen Genotypen bestimmt.

Die beiden SNPs rs7903146 und rs11196205 des Gens TCF7L2 werden ebenfalls mit einem höheren Risiko für Typ 2 Diabetes assoziiert und sind zusätzlich aufgelistet. Sie stehen in starker Korrelation zueinander. (Chen et al. 2012). Für beide Marker ist in ost-asiatischen Populationen mit einer Frequenz von ca. 97 % für C bzw. G hauptsächlich der homozygot anzestrals bzw. abgeleitete Genotyp nachgewiesen. Die Allelfrequenzen in europäischen Populationen beider Allele sind nahezu ausgeglichen, wobei für rs11196205 die heterozygote Form am häufigsten ist.

Zöliakie und Bluthochdruck

SH2B3/ATXN2 (rs653178) und SH2B3 (R262W, rs3184504) werden mit einer Erkrankung an Zöliakie und einem Risiko für Bluthochdruck assoziiert und standen in prähistorischen eurasischen Populationen unter Selektion (Hunt et al. 2008; Mathieson et al. 2015). Zöliakie ist eine immunologische Erkrankung des Dünndarms und wird durch eine Unverträglichkeit von Getreideeiweiß ausgelöst. In ost-asiatische Populationen ist für beide Marker nur das anzestrals Allel nachweisbar.

Klimaanpassung

Der SNP rs653178 im Gen ABCC11 (c.538G>A / G180R) korreliert mit eher trockenem oder feuchten Zerumen sowie weniger Transpiration im Achselbereich und wird mit einer Anpassung an klimatische Verhältnisse in Verbindung gebracht (Yoshiura et al. 2006; Ohashi et al. 2011). Einige Populationen im asiatischen Raum weisen eine Mutation im Gen auf und besitzen somit kein funktionsfähiges Protein, was zu hellerem Ohrenschmalz sowie reduziertem Körpergeruch führt. In europäischen Populationen liegt die Allelfrequenz des anzestrals Allels bei 86 %.

EDAR beeinflusst die Entwicklung von ektodermalen Strukturen und kann mit drei Ausprägungsformen in Verbindung gebracht werden: Dicke und Struktur der Kopfbehaarung und Form der Schneidezähne (Bryk et al. 2008). Eine dichtere und vermehrte Behaarung kann möglicherweise einem Wärmeverlust über die Kopfhaut vorbeugen. Das abgeleitete Allel kommt fast ausschließlich in asiatischen Populationen und Amerika vor (EAS 87 % und EUR 1%).

Tabelle 8 Auflistung der nukleären, funktionalen Marker; Angabe der Kategorie, entsprechend ihrem assoziierten Phänotyps, Bezeichnung der Loki, rs-Nummer und chromosomale Position

Kategorie	Bezeichnung	rs-Nr.	CHR	Start	Stopp
Pigmentierung	HERC2	rs12913832	chr15	28365617	28365618
	OCA2	rs1545397	chr15	28187771	28187772
	SLC45A2.1	rs16891982	chr5	33951692	33951693
	SLC24A5	rs1426654	chr15	48426483	48426484
	SLC24A4	rs12896399	chr14	92773662	92773663
	MC1R.8	rs885479	chr16	89986153	89986154
	IRF4.1	rs12203592	chr6	396320	396321
	TYR.1	rs1042602	chr11	88911695	88911696
	OCA2.3	rs1800414	chr15	28197036	28197037
Laktasepersistenz	13910T (LCTa)	rs4988235	chr2	136608645	136608646
	22018A (LCTb)	rs182549	chr2	136616753	136616754
Xenobiotischer Metabolismus	NAT2	rs1801280	chr8	18257853	18257854
	NAT2	rs1041983	chr8	18257794	18257795
	ABCB1c	rs1045642	chr7	87138644	87138645
	ABCB1a	rs1128503	chr7	87179600	87179601
	ABCB1b	rs2032582	chr7	87160617	87160618
	CYP3A5	rs776746	chr7	99270538	99270539
	CYP3A4	rs2740574	chr7	99382095	99382096
	AGT	rs5051	chr1	230849871	230849872
	AGT	rs699	chr1	230845793	230845794
Alkoholverträglichkeit	ADH1Bb	rs1229984	chr4	100239318	100239319
	ADH1Ba	rs3811801	chr4	100244318	100244319
	ALDH2	rs671	chr12	112241765	112241766
Diabetes Typ2	TCF7L2	rs11196205	chr10	114807046	114807047
	TCF7L2	rs7903146	chr10	114758348	114758349
	CDKN2B	rs564398	chr9	22029546	22029547
	WFS1	rs10010131	chr4	6292914	6292915
	CDKAL1_a	rs7754840	chr6	20661249	20661250
	IGF2BP2	rs4402960	chr3	185511686	185511687
	PPARG	rs1801282	chr3	12393124	12393125
	KCNJ11	rs5215	chr11	17408629	17408630
	HHEX/ IDE	rs1111875	chr10	94462881	94462882
SLC30A8	rs13266634	chr8	118184782	118184783	
Zöliakie / Bluthochdruck	SH2B3/ATXN2	rs653178	chr12	112007755	112007756
	SH2B3	rs3184504	chr12	111884607	111884608
Klimaanpassung	ABCC11	rs17822931	chr16	48258197	48258198
	EDAR	rs3827760	chr2	109513600	109513601

3.5 Nukleotiddiversität

Die Nukleotiddiversität gibt eine Wahrscheinlichkeit an, dass sich zwei gewählte Nukleotide an einer Position zwischen zwei Sequenzen in allen Individuen einer Population unterscheiden. Der Mittelwert wird als Maß für die Variabilität innerhalb einer Population angenommen. Für die hier erstellten Gruppierungen wird mittels *vcftools* (Danecek et al. 2011) ein *site-pi*-Wert je Position zwischen den Individuen bestimmt. Der daraus errechnete Mittelwert inklusive Standardabweichung ist in Tabelle 17 und Abbildung 26 dargestellt.

Die Proben werden nach ihrer (A) zeitlicher Zuordnung und (B) Kulturstufe unterteilt:

(A) zeitlicher Zuordnung

eisenzeitliche Proben aus SW-Russland (westIA), eisenzeitliche Proben aus Sibirien und Kasachstan (Ost-IA), Proben aus der Bronzezeit (Bronze) und die beiden neolithische Proben aus West-Sibirien (Neo)

(B) Kulturstufe

Aldy-Bel, Pazyryk, Tagar, Sintašta, Sarmaten und Čilikta

In der Gruppierung nach Kulturangehörigkeit sind die Proben U12, Nov7, Kal1 und Kol4 nicht enthalten, da sie als einzelne Individuen die Saken, Initiale Skythen, Skythen und die Poltavka-Kultur repräsentieren.

4 Ergebnisse

4.1 Finale Probenauswahl für die Anreicherung

Die Auswahl der Proben für eine anschließende Anreicherung nukleärer Regionen erfolgte basierend auf den Ergebnissen der Qualitätsbestimmung der erstellten *libraries*.

Für die Proben Be10 und U12 standen Felsenbeinen als Probenmaterial zur Verfügung. DNA-Material, welches aus *pars petrosa* extrahiert werden kann, zeigt nachgewiesen einen überdurchschnittlich guten Erhaltungszustand (Rasmussen et al. 2014; Gamba et al. 2014; Pinhasi et al. 2015). Da die Bearbeitung der Proben zeitlich später erfolgte und man aufgrund des Skelettmaterials von einem sehr guten Erhaltungszustand der Proben ausgehen kann, liegt nur eine Bestimmung der Probenqualität mittels quantitativer PCR vor.

Quantitative Messungen und Qualitätsbestimmung der DNA-libraries

Basierend auf den Ergebnissen der *shotgun*-Sequenzierung auf dem *MiSeq* wurde, unter Berücksichtigung der Duplikationsrate, eine Auswahl der Proben getroffen. Die Duplikationsrate liefert bereits erste Hinweise auf die Komplexität einer *library*, welche die Effizienz einer nukleären Anreicherung maßgeblich beeinflusst. Lag die einzigartige Kopienzahl der *shotgun-libraries* bei der quantitativen Bestimmung der Molekülzahl vor der PCR zusätzlich über $1.00E+07$ Molekülen, wurde die Probe für die weitere Bearbeitung ausgewählt. Die Anzahl der erstellten *libraries* für die Anreicherung wurde aufgrund des endogenen DNA-Gehalts einer Probe in Kombination mit den qPCR-Ergebnissen festgelegt (siehe Tabelle 9 im folgenden Kapitel).

Die Ergebnisse der *shotgun*-Sequenzierung aller extrahierten Proben auf dem *MiSeq* mit Angabe des endogenen DNA-Gehalts und Duplikationsrate ist im Anhang Tabelle B6 zu finden.

Die durchschnittliche Molekülzahl der *libraries* aller ausgewählten Proben beträgt 10^8 einzigartige Moleküle in der quantitativen Messung mittels qPCR nach der *fill-in* Reaktion, was umgerechnet durchschnittlich 10^7 Adapter-ligierten Molekülen je mg Knochenpulver entspricht. Der endogene DNA-Gehalt der *screening-libraries* beträgt durchschnittlich 23,7 %

(3,36-71,11%). 31 Proben zeigen einen endogenen DNA-Gehalt von $> 10\%$. Die Proben Kal1, Kama10, Kol4, Nov7, Sol5-1, Pr3, Pr8, Pr11 und Pr12 liegen unter 10%. In den *fill-in*-Produkten konnten jedoch $> 10^7$ Molekülen bestimmt werden. Da die jeweilige Datierung und geografischen Lage für die Arbeit von großem Interesse ist wurden die Proben für eine Anreicherung ausgewählt. Die Probe J12-9a wurde trotz einem endogenen DNA-Gehalt von 15 % aufgrund der hohen Duplikationsrate von 85 % nicht weiterbearbeitet.

In der folgenden Abbildung (Abb. 5) ist die Kopienzahl der Moleküle per mg Knochen der ausgewählten Proben aus der qPCR-Quantifizierung gegen ihren endogenen DNA-Gehalt aufgetragen und grafisch dargestellt.

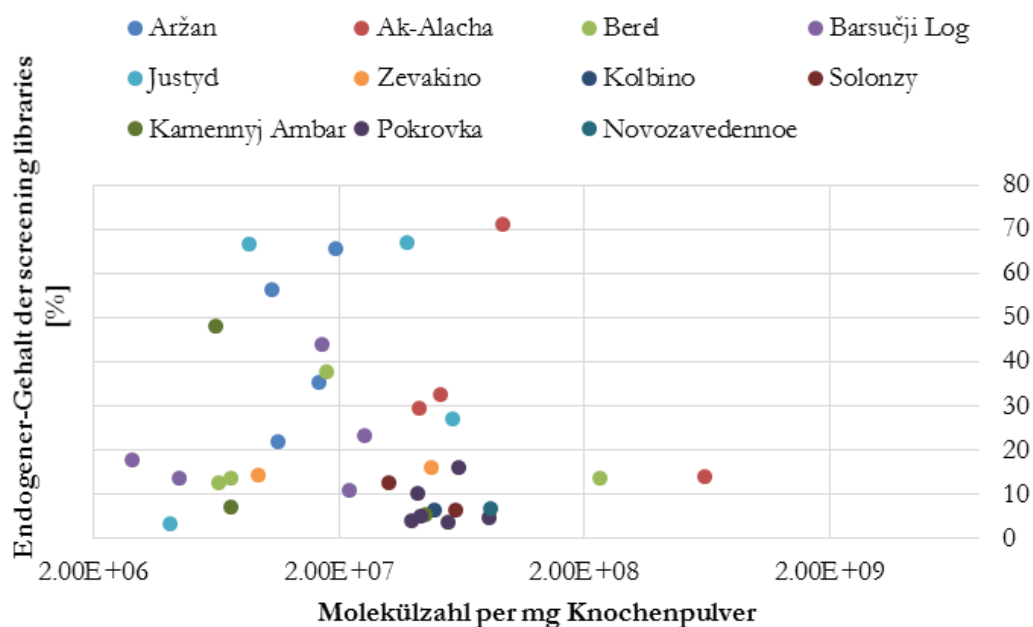


Abbildung 5: Darstellung des endogenen DNA-Gehalts und der umgerechneten Molekülzahl per mg extrahiertem Knochenpulver der *screening-libraries*, nach Fundplatz gruppiert, in logarithmischer Skalierung

Die molekulare Geschlechtsbestimmung wurde auf alle Proben, mit Ausnahme von Be10 und Ul2, basierend auf den Sequenzdaten der *screening-libraries* nach (Skoglund et al. 2013) durchgeführt. Als eindeutig weiblich konnten 8 der 41 Proben eingestuft werden, 31 als

männlich. Die Proben Kama10 und Bl2 konnten nicht eindeutig typisiert werden, entsprechen aber eher einem Ry-Wert der als männlich einzuordnen ist. Die Zuteilung der beiden Proben deckt sich mit der morphologischen Klassifizierung als männliche Individuen (Abb. 6).

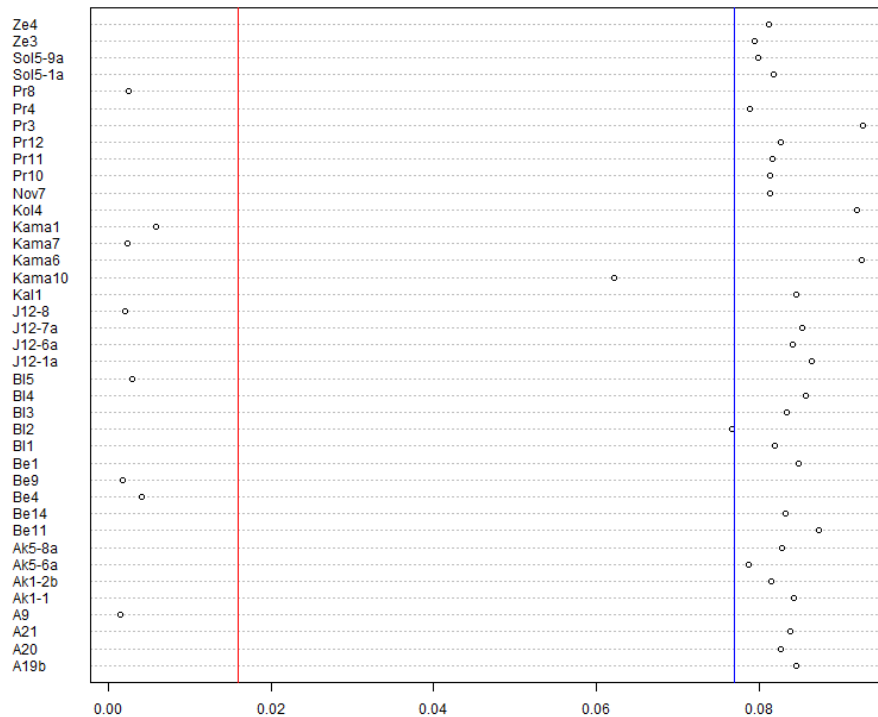


Abbildung 6: Ergebnisse der Molekularen Geschlechtsbestimmung der screening-libraries nach (Skoglund et al. 2013) y-Achse: sequenzierte Proben; x-Achse= Ry; Ry < 0,016 weiblich, Ry > 0,077 männlich

Tabelle 9 fasst die finale Probenauswahl je Fundplatz und Kulturzugehörigkeit mit allen Ergebnissen der Quantitätsbestimmungen und Geschlechtszuordnung zusammen.

Tabelle 9 Gruppierung der Proben je Fundplatz; Auflistung der MiSeq-*screening* Ergebnisse und qPCR-Messung: endogener DNA-Gehalt in % und Molekülzahl (MZ) pro mg Knochenpulver der *fill-in* Produkte

Fundplatz	Gebiet	Kultur	Probe (Geschlecht / end. DNA % / MZ qPCR)
Aržan 2	Südsibirien, Region Tuva	Aldy-Bel	A9 (XX / 21.85 / 1.13E+07); A19 (XX / 65.79 / 1.94E+07); A20 (XY / 35.41 / 1.67E+07); A21 (XY / 56.36 / 1.07E+07)
Ak-Alacha 1	Südsibirien, Altai Ukok Plateau	Pazyryk	Ak1_1 (XY / 13.94 / 6.21E+08); Ak1_2 (XY / 29.61 / 4.25E+07)
Ak-Alacha 5	Südsibirien, Altai Ukok Plateau	Pazyryk	Ak5_6 (XX / 32.72 / 5.19E+07); Ak5_8 (XX / 71.11 / 9.26E+07)
Berel	Ost- Kasachstan, Kasachischer Altai	Pazyryk	Be1 (XY / 45.37 / 7.29E+06); Be4 (XX / 13.7 / 6.46E+06); Be9 (XX / 12.79 / 2.37E+08); Be10 (XY / k.A. / 2.32E+08); Be11 (XY / 37.91 / 3.71E+07); Be14 (XY / 13.71 / 1.79E+07)
Barsučij Log	Minusinsker Becker	Tagar	Bl1 (XY / 11 / 2.19E+07); Bl2 (XY / 23.26 / 2.54E+07); Bl3 (XY / 44.16 / 1.71E+07); Bl4 (XY / 13.7 / 4.48E+06); Bl5 (XX / 17.84 / 2.89E+06)
Justyd XII	Südsibirien, Čuja-Region	Pazyryk	J12_1 (XX / 33,87 / 3.81E+07); J12_6 (XX / 67.04 / 5.85E+07); J12_7 (XY / 26.97 / 8.63E+06); J12_8 (XX / 66.82 / 4.09E+06)
Kalinovka	SW-Russland	Poltavka Kultur	Kal1 (XY / 8.52 / 1.08E+08)
Kamennyj Ambar 5	Kamennyi Ambar, Olgino Ural	Sintašta- Kultur	Kama1 (XY / 25.8 / k.A.); Kama6 (XY / 7.14 / 7.27E+06); Kama7 (XX / 48.04 / 6.32E+06); Kama10 (XY / 5.31 / 4.48E+07)
Kolbino	SW-Russland	Skythen	Kol4 (XY / 6.4 / 4.93E+07)
Novozavedennoe 2	SW-Russland	Initiale/West- Skythen	Nov7 (XY / 6.89 / 8.23E+07)
Pokrovka	SW-Russland	Früh- sarmatisch	Pr3 (XY / 3.56 / 5.57E+07); Pr4 (XY / 10.14 / 4.21E+07); Pr8 (XX / 4.63 / 8.14E+07); Pr10 (XY / 16.1 / 6.16E+07); Pr11 (XY / 4.03 / 3.95E+07); Pr12 (XY / 5.22 / 4.29E+07)
Solonzy 5	Westsibirien, Novosibirsk		Sol5_1a (XY / 6.4 / 6.01E+07); Sol5_9a (XY / 12.57 / 3.22E+07)
Ulžan	SO- Kasachstan	Saken	Ul2 (XY / k.A. / 7.15E+08)
Zevakino	Ostkasachstan	Zevakino- Čilikta	Ze3 (XY / 15.98 / 4.77E+07); Ze4 (XY / 14.43 / 9.39E+06)

Die folgende Abbildung zeigt die geografische Verteilung der ausgewählten Proben über die eurasische Steppe je Fundplatz und zeitlicher Zuordnung.

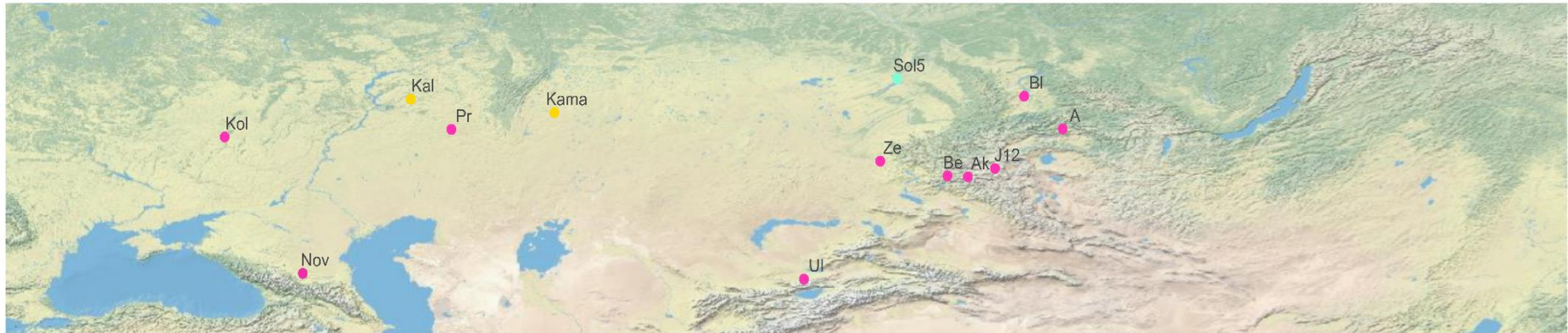


Abbildung 7: Geografische Verteilung der untersuchten Fundplätze über die eurasische Steppe

Kol= Kolbino (N=1), Nov=Novozavedennoe 2 (N=1), Kal= Kalinovka (N=1), Pr= Pokrovka (N=6), Kama= Kamennyj Ambar (N=4), Ul= Ulžan (N=1), Ze= Zevakino (N=2), Sol5= Solonzy 5 (N=2), Be= Berel (N=6), Ak= Ak-Alacha 1/5 (N=4), Bl= Barsučij Log (N=5), A= Aržan 2 (N=4); Farbgebung entspricht der Datierung: gelb=Bronzezeit, pink= Eisenzeit, blau= Neolithikum; erstellt in R (OpenStreetMap, ggplot2, mapplots2). Koordinaten in Tabelle B7 im Anhang

4.2 Wahl der Anreicherungsmethode

Design

Beim Vergleich der Design-Strategien der verschiedenen Hersteller soll die Handhabung und Nachvollziehbarkeit der gewählten *bait*-Sequenzen, die Stringenzkriterien und Abdeckung der neutralen Regionen und SNP-Positionen verglichen werden.

Im **NimbleGen** Design kann zwischen zwei *bait-libraries* gewählt werden. Für die durchgeführten Versuche wurde Match 1 ausgewählt, mit einer durchschnittlichen Abdeckung der Zielregionen von 82,8x. 99 % der gewünschten Zielregionen konnten mit *baits* abgedeckt werden, was insgesamt 174917 Basen entspricht.

Für die *bait-library* von **MYbait** wurde die Option A mit den stringentesten Filterkriterien ausgewählt. 14 SNP-Positionen konnten nur mit 1-2 anstatt 3 *baits* abgedeckt werden. Insgesamt wurden 245193 Basenpaare mit 7809660 *baits* abgedeckt. In 500 ng ssRNA sind demnach $1,1577E+13$ Kopien enthalten, was $1,19E+09$ einzigartigen *bait*-Sequenzen entspricht.

Im **SureSelect-Design** mit eArray konnten alle 289 neutralen Regionen in gewünschter *tiling*-Frequenz abgedeckt werden. Von den verlängerten Regionen der SNP-Positionen lagen 39 in repetitiven Regionen, 12 wurden als *orphan-baits* klassifiziert. 26 Regionen zeigen einen GC-Gehalt unter 30 % und die *bait*-Anzahl wurde hierfür verdoppelt. Es sind bis zu 57000 einzigartige *bait* -Sequenzen möglich. Die finale *bait* -Konzentration ist seitens des Herstellers nicht bekannt.

Zwei SNP-Positionen wurden in allen drei Designs verworfen: L142.1 und P43. Von den 9 SNPs, die nur mit jeweils einem *bait* abgedeckt werden, wurden 4 im NimbleGen-Design ausgeschlossen und 3 im Agilent-Design. Von den 5 SNPs, die mit nur jeweils 2 *baits* bei MB abgedeckt werden, wurden 1 SNP im NG-Design ausgeschlossen und 1 SNP im SuSe Design.

Auswertung der Sequenzierung

Neben der Transparenz und Flexibilität der Designstrategien wurde die Effizienz der verschiedenen Anreicherungsmethoden verglichen. Da exakt dieselbe Probe und vergleichbare

Konzentrationen in den Versuchen eingesetzt wurden, liegt der Schwerpunkt der Auswertung auf dem Vergleich der Abdeckungstiefe (coverage) und der Effizienz des jeweiligen Anreicherungskits. Je mehr Sequenzen (mit Duplikaten) im Verhältnis zur Gesamtheit der gegen hg19 alignierten Sequenzen in die *capture*-Regionen fallen, desto spezifischer die Anreicherungsverfahren.

Tabelle 10 Auswertung der Sequenzdaten der Probe Ak1-1 für die getesteten Anreicherungsverfahren verschiedener Hersteller; Aln. CAP=Anteil der Sequenzen, die gegen die *capture*-Regionen alignieren an alignierten Sequenzen gegen hg19; cov=Abdeckung in den *capture*-Regionen

	Aln. CAP [%] relativ zu hg19	Aln. CAP Mkdup	cov (min/max/mean)	Cov. CAP [%]
SuSe	76	145871	43.7/120.9/87.9	100
MB	72.8	260494	46.8/142.6/108.5	99,76
NG	91	111521	27.5/100.9/66.5	98,89

Signifikanter Unterschied des NG Protokolls zu den beiden anderen Verfahren ist die sehr niedrige Duplikationsrate. Trotz vergleichsweise kleiner Rohdatenmenge und wenigen Sequenzen zu Beginn der Analyse bleibt nach dem Duplikate entfernen eine hohe Sequenzanzahl übrig, die gegen die *capture*-Regionen aligniert werden kann (Tab. 10). Dies spart deutlich Kosten und Speicherkapazitäten und ist auf die niedrige Zyklenzahl der beiden PCR- Schritte zurückzuführen.

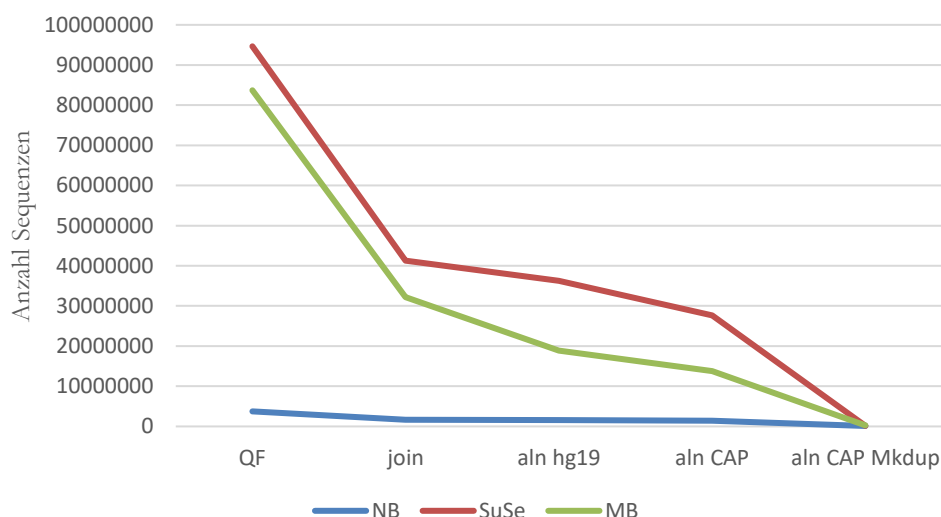


Abbildung 8: Vergleich der Sequenzzahlen der Probe Ak1-1 nach einzelnen Schritten der Rohdatenbearbeitung; OF=Qualitätsfilter; join=Sequenzpaare aln=aligniert gegen hg19 bzw. die entsprechenden *capture*-Regionen; Mkdup=ohne Duplikate

Die durchschnittliche Abdeckung der Zielregionen ist mit dem MB-Kit am höchsten, ebenso die minimalste Abdeckungstiefe bei 46,8 x. Die absolute Sequenzanzahl ohne Duplikate in den *capture*-Regionen ist mit dem MYbait -Kit ebenfalls am höchsten. Die Anzahl der Sequenzen, die im Verhältnis zur Gesamtanzahl der *reads* in den Zielregionen aligniert werden können ist im NimbleGen bzw. SureSelect-Kit höher. Bei allen drei Anreicherungsverfahren sind die *capture*-Regionen vollständig oder zu über 98 % abgedeckt.

Basierend auf den durchgeführten Analysen wurde das Anreicherungs-kit MYbait von MicroArray ausgewählt. Die einfache und flexible Handhabung im *bait*-Design sind ein klarer Vorteil, da hier *bait*-Länge und Art der Nukleotidsequenz (RNA/DNA) frei gewählt werden können. Das Design wird vom Hersteller übernommen und kann dann selbst flexibel modifiziert werden. Die *tiling*-Frequenz ist benutzerdefiniert. Zudem sind alle eingesetzten Chemikalien in ihrer jeweiligen Konzentration bekannt und separat lieferbar. Die *bait-libraries* können auch ohne Reagenzien geordert werden.

Die Spezifität des *captures* ist bei den Kits der anderen beiden Hersteller zwar höher, es besteht jedoch die Möglichkeiten dies laborseitig durch Hybridisierung- / Waschttemperatur und Salzkonzentrationen der Waschpuffer zu verbessern, weshalb letztendlich das MYbait-Kit ausgewählt wurde.

Das für die vorliegenden Proben verwendete Protokoll wurde im Laufe der Bearbeitung an die neue Version des MYbait Handbuches adaptiert oder Verbesserungen aus Publikationen entnommen. Folgende Änderungen wurden vorgenommen:

Die Zyklen-Zahl der beiden Amplifikationen nach den *capture*-Runden wurde auf 1 x 7-9 und 1 x 12-14 Zyklen angepasst. Für die Trennung der isolierten DNA von den *bait-bead*-Komplexen wurde am Ende einer jeden *capture*-Runde eine doppelte Elution durchgeführt. Als magnetische Beads werden die Dynabeads ® MyOne™ Streptavidin T1 (Thermo Fisher Scientific) eingesetzt.

4.3 Anreicherungserfolg

Für die Anreicherung wurden zunächst entsprechend der Probenqualität und Verfügbarkeit des Materials mehrere *libraries* erstellt und gepoolt, um die Komplexität und Einsatzmenge zu erhöhen. Die folgende Tabelle (Tab. 11) listet die Anzahl der gepoolten *libraries* je Proben und Einsatzmenge in ng auf.

Abdeckung der Zielregionen und des gesamten Genoms

Für die Beurteilung des *capture*-Erfolgs werden folgende Kriterien berücksichtigt:

- Anteil der Sequenzen die ausschließlich gegen die *capture*-Regionen aligniert werden können, im Verhältnis zu allen alignierten Sequenzen gegen die humane Referenz (Aln CAP [%] / hg19)
- Abdeckung der gesamten *capture*-Regionen (Cov CAP [%])
- Durchschnittliche Tiefe der Abdeckung mit Standardabweichung (Cov MW / SD)

Mit 31 der angereicherten Proben konnte eine Abdeckung der Zielregionen von über 5 x erzielt werden. In einem mittleren Abdeckungsbereich von 5 x – 20 x liegen 19 Proben. Acht Proben liegen mit einer Abdeckung zwischen 21,13 x und 54,7 x im oberen Bereich. Die Proben Ak5-8, Ak1-1, Be10 und U12 erreichen eine *coverage* von > 80 x. Wobei für die beiden Proben Be10 und U12 das DNA-Material aus Felsenbeinen gewonnen werden konnte. Im Vergleich aller *capture*-Produkte variiert die Höhe der Abdeckung stark untereinander und liegt zwischen 1,92 x und 124,95 x, durchschnittlich bei 22,82 x.

Die durchschnittliche Abdeckung der Zielregionen liegt bei 94,64 % mit der niedrigsten Abdeckung bei 65,10%.

Wenn Proben bereits nach der ersten *capture*-Runde sequenziert und mit Ergebnissen aus der zweiten Runden zusammengeführt wurden (Ak1-1, Be4, Be9, Be14), ist die Abdeckung des gesamten Genoms in % deutlich höher und liegt zwischen 7,67 und 18,06%, da das *capture* nach nur einem Durchlauf noch sehr unspezifisch ausfällt. Für *capture*-Produkte, die ausschließlich nach der zweiten Runde sequenziert wurden, ist das gesamte Genom zu 2,53 % abgedeckt.

Die folgende Tabelle (Tab. 11) fasst alle Ergebnisse zusammen.

Tabelle 11 Ergebnisse der Sequenzierung von 41 *capture*-Produkten; CAP= *capture*, Aln=aligniert gegen das Referenzgenom GRCh37/hg19; Mkdup= ohne Duplikate; cov=Abdeckung; MW/SD=Mittelwert / Standardabweichung; Frag.länge= Fragmentlänge

Probe	Laborcode	Anz. Libs	Einsatz CAP [ng]	Aln CAP [%] / hg 19	Aln CAP Mkdup	Cov MW	Cov SD	Cov CAP [%]	Cov hg19 [%]	Frag.länge MW
A19	NC9-3	3	1674	52.24	29628	12.84	5.12	99.59	3.46	119.84
A20	NC8-1 / NC8-2	4 / 4	2025	70.62	56267	22.19	8.34	99.59	2.67	109.18
A21	NC10-4	4	2199	61.62	46035	16.6	6.56	99.63	5.61	97.14
A9	NC8-3 / NC8-4	4 / 4	2000 / 2000	84.36	20889	8.96	3.7	95.58	1.15	113.83
Ak1_1	NC7a-1 / NC7b-1	4	1439	72.85	263198	106.4	35.6	99.76	18.06	107.73
Ak1_2	NC11-2	4	768	84.51	54907	19.34	7.75	99.66	2.39	94.37
Ak5_6	NC11-3	6	1683	81.94	68360	27.84	8.62	99.69	3.54	112.18
Ak5_8	NC9-2	4	1337	53.01	201318	88.05	26.26	99.76	8.4	121.47
Be1	NC9-1	4	1013	55.46	13485	5.37	2.74	97.61	2.82	107.33
Be10	NC10-10	4	530	73.92	245555	114.5	31.92	99.76	3.35	128.14
Be11	NC8-5	3	1960	41.74	9279	3.9	2.12	95.10	1.25	109.85
Be14	NC6b-2 / NC7a-3 / NC7b-3	3 / 3 / 3	348 / 1556	72.28	19630	7.53	4.43	97.69	8.53	80.67
Be4	NC6b-1 / NC7a-2 / NC7b-2	2 / 4 / 4	214 / 1601	58.54	12994	4.9	2.76	94.54	15.15	77.51
Be9	NC7a-7 / NC7b-7	5 / 5	1512	79.56	140673	54.17	15.29	96.12	7.67	98.84
Bl1	NC6b-4 / NC7b-4	4 / 4	2635	76.85	10942	4.23	2.48	94.63	3.02	82.38
Bl2	NC6-5 / NC7-6	2 / 2	1556	72.13	21751	5.24	2.57	97.92	3.17	106.91
Bl3	NC10-5	7	1930	65.32	33111	12.77	5.25	99.60	6.3	106.23
Bl4	NC10-8	3	645	67.46	4795	2.31	1.39	78.43	5.35	106.5
Bl5	NC10-9	4	766	69.88	4089	2.13	1.3	73.56	2.09	106.24
J12_1	NC9-8	5	1070	65.63	8505	3.53	2.05	92.27	2.15	108.16
J12_6	NC9-10	4	777	56.03	96416	39.43	12.26	99.74	3.65	111.8

Probe	Laborcode	Anz. Libs	Einsatz CAP [ng]	Aln CAP [%] / hg 19	Aln CAP Mkdup	Cov MW	Cov SD	Cov CAP [%]	Cov hg19 [%]	Frag.länge MW
J12_7	NC10-12	6	730	77.05	76725	33.97	15.61	99.64	2.21	122.98
J12_8	NC9-9	4	477	52.96	56608	23.59	6.76	95.95	6.72	111.18
Kal1	NC10-13	6	2250	83.56	36167	15.65	5.31	99.63	1.51	121.19
Kama1	NC10-7	5	750	70.73	3418	2.26	1.7	66.31	1.63	116.16
Kama10	CBI	5	1365	75.55	3398	1.92	1.18	65.10	2.43	103.25
Kama6	NC11-6	8	1351	89.22	3210	1.96	1.2	68.01	0.64	111.58
Kama7	NC11-5	9	1314	89.63	15042	6.53	3.08	94.72	0.25	111.61
Kol4	NC10-1	8	2800	81.66	39358	14.57	6.52	99.39	0.94	98.61
Nov7	NC10-2	8	1655	83.95	9404	3.6	2.04	93.48	0.83	98.19
Pr10	NC9-7	5	1059	69.64	86987	39.25	12.96	99.56	1.37	124.4
Pr11	NC9-5	6	2075	77.31	17785	6.73	3.23	98.56	1.14	101.7
Pr12	NC9-6	7	2027	74.28	25454	8.98	4.09	99.02	0.97	93.09
Pr3	NC9-4	8	1733	78.36	19056	7.8	3.6	98.96	0.58	111.76
Pr4	NC10-3	6	1900	79.85	41364	16.98	5.99	99.70	2.19	112.93
Pr8	NC9-11 / NC11-1	5 / 8	1054 / 1918	93.59	22384	5.79	3.07	93.69	0.47	82.9
Sol5_1a	NC11-4	7	1871	93.14	27783	10.59	4.51	99.48	0.81	102.33
Sol5_9a	NC10-14	8	2250	83.74	40956	16.16	5.93	99.72	1.94	108.41
UI2	NC10-11	4	1302	66.72	64765	125	38.09	99.78	2.93	120.39
Ze3	NC6-3 / NC7-5	5 / 5	522 / 1667	82.41	39588	21.13	10.36	99.67	1.54	109.01
Ze4	NC11-8	7	2208	88.11	28195	10.97	4.34	99.54	2.02	107.72

Eine vollständige Tabelle der Sequenziererergebnisse aller Teilschritte bei der Rohdatenauswertung ist im Anhang B8 zu finden.

4.4 Authentizität der Proben

Leerkontrollen

Alle Extraktionsleerkontrollen wurden mittels Gel-Elektrophorese auf amplifizierte DNA getestet und zweimal aufgetragen. Teilweise wurden Extraktionsleerkontrollen zusätzlich als Leerkontrolle für die *library*-Erstellung mitgeführt. Für alle Extraktionsleerkontrollen konnte keine Bande nachgewiesen werden. Alle *library*-Leerkontrollen wurden über eine Qubit - sowie Bioanalyzermessung kontrolliert und lieferten keine quantifizierbaren Ergebnisse. Die für jeden *capture*-Ansatz mitgeführten Leerkontrollen zeigten ebenfalls keine messbaren Resultate via Qubit – und Bioanalyzermessung.

Als zusätzliche Kontrolle wurden mitgeführte Leerkontrollen der *library*-Erstellung oder eines *capture*-Ansatzes zufällig ausgewählt und mittels qPCR oder MiSeq-Sequenzierung überprüft. Die totale Molekülzahl der *fill-in*-Produkte der mitgeführten Leerkontrollen während einer *library* liegt zwischen $1.75E+05$ - $8.69E+05$ Molekülen nach einer quantitativen Bestimmung mittels qPCR. Über eine MiSeq-Sequenzierung analysierten *library*-Leerkontrollen zeigen zwischen 22-281 alignierten Sequenzen.

Alle mittels qPCR quantifizierten, auffälligen Leerkontrollen der Anreicherungsverfahren wurden vollständig sequenziert, lieferten jedoch keine auswertbaren Ergebnisse.

contaMIX, ANGSD und mapDamage

Für alle hier bearbeiteten Proben wurde ein Authentizitätswert von > 91 % als Kriterium festgelegt. Eine Kontaminationsbestimmung anhand der mitochondrialen DNA konnte für alle Proben auf alle *capture*-Produkte durchgeführt werden. Drei der 41 getesteten Proben zeigen einen davon abweichenden Wert. Abzüglich der drei auffälligen Proben liegt der durchschnittliche Authentizitätswert der verbleibenden 38 Proben bei 97,84 %. Die Proben J12_1, A21 und B13 können mit einem Authentizitätswert von 85,6%, 70,06 % bzw. 75,42 % nicht als authentisch eingestuft werden, zeigen jedoch bei der Kontaminationsbestimmung basierend auf dem X-Chromosom keine Auffälligkeiten. Für die Probe A21 wurde zudem eine sehr niedrige Deaminierungsrate am 5'-Ende festgestellt, was auch an einem guten Erhaltungszustand des Probenmaterials liegen kann. Alle drei Proben werden in den

populationsgenetischen Analysen mitgeführt, jedoch in Hinblick auf eine mögliche Kontamination diskutiert. Für die Proben KAMA1 und Sol5-9a wurde ein Authentizitätswert von 91,7 % bzw. 91,4 % berechnet. Die Werte liegen knapp über dem festgelegten Schwellenwert und konnten nicht durch eine zweite Methode mit ANGSD verifiziert bzw. widerlegt werden. Vier Proben, Bl2, Sol5_1, Kol4 und Kal1 zeigen leicht erhöhte Kontaminationswerte in der Berechnung anhand des X-Chromosoms, werden jedoch mit contaMIX mit über 95 bzw. 99% als nicht kontaminiert eingestuft. Insgesamt konnten die Kontaminationsbestimmung mit ANGSD bei allen männlichen Individuen aufgrund der geringen Anzahl an alignierten Sequenzen auf dem X-Chromosom nur auf 13 von 33 männlichen Proben angewendet werden.

Die durchschnittliche Deaminierungsrate die für alle *capture*-Produkte bestimmt werden konnte liegt bei 15,13 %. Die niedrigste Rate mit 2,75 % wurde für die Probe Ak5-8 gemessen. Diese ist aufgrund allen weiteren Kontaminationsbestimmungen als authentisch einzuschätzen. Bei 10 Proben liegt die Deaminierungsrate unter 10 %. Für den überwiegenden Teil der Proben liegt die für aDNA charakteristische Deaminierung am 5'-Ende zwischen 10 % -33,38 %.

Die Ergebnisse der Proben mit auffälligen Werten in der Kontaminationsbestimmung sind in nachfolgender Tabelle (Tab. 12) dargestellt, eine Übersicht für alle Proben findet sich im Anhang B9.

Tabelle 12 Ergebnisse der Kontaminationsbestimmung anhand des mitochondrialen Genoms und des X-Chromosoms und Angabe der Deaminierungsrate; FR=geschätzte Fehlerrate; Auth.= Authentizität in %; Kon.=Kontamination; mD5=Deaminierungsrate am 5'-Ende eines Moleküls

Probe	contaMIX		ANGSD - Methode 1			ANGSD - Methode 2			mapDamage
	FR	Auth. [%]	Kon. [%]	SE	p-Wert	Kon. [%]	SE	p-Wert	mD5 [%]
J12_1	0.0031	85.61	2.07	0.02	0.3	0.01	0	1	10.42
A21	0.0081	70.06	1.33	0.01	0.232	4.22	0.03	0.197	5.98
Bl3	0.0056	75.42	0.01	0	0.35	1.27	0.02	1	17.46
Kama1	0.006	91.74	k.A.						12.47
Kol4	0.0023	99.51	12.10	0.05	0.06	0.01	0	1	34.02
Kal1	0.003	99.71	7.56	0.04	0	11.70	0.05	0.01	19.98
Sol5_9	0.0038	91.41	0.69	0.02	0.66	0.01	0	1	20.29
Sol5_1	0.0053	94.03	6.90	0.13	0.22	0.01	0.08	1	22.25
Bl2	0.0065	95.16	6.13	0.06	0.26	9.15	0.09	0.21	15.4

4.5 Phänotypenbestimmung

Insgesamt 193 funktionale Marker, abzüglich X und Y- chromosomalen Positionen, wurden im Rahmen des nukleären *captures* angereichert. Durchschnittlich 94,97 % (183,29; min. 100 – max. 193) der Marker konnten mit einer durchschnittlichen Abdeckung von 23,33x (1,84x-126,66x) erfolgreich sequenziert werden. Es wurden neben den Positionen, die für das HIrisplex-System relevant sind, 37 zusätzliche Marker ausgewählt und assoziierte Phänotypen definiert und beschrieben. Von diesen Loci konnten 97 % mit einer durchschnittlichen Abdeckung von 26,4x (1,77x-199,92x) angereichert werden (\bar{O} 35,9 Loci; min. 27 – max. 37)

Nach dem Filtern der *vcf*-files mit einer Abdeckung von 2x und einer Genotyp-Qualität von 30 sind von 37 der ausgewählten, funktionalen Marker durchschnittlich 57 % (\bar{O} 21,29 Positionen; 0 – max. 37) je Probe abgedeckt. Für eine Probe der Tagar-Kultur (Bl5) aus Barsučij Log und zwei bronzezeitlichen Proben aus Kamennyj-Ambar (Kama6 und Kama10) kann kein Phänotyp definiert werden.

Eine detaillierte Auflistung der Abdeckung aller Positionen inklusive Standardabweichung ist in der Tabelle B11 im Anhang zu finden.

Augen – und Haarpigmentierung

Für 24 Proben wurden die nach 2-facher Abdeckung und Genotyp-Qualität 30 gefilterten MLE-vcf Dateien verwendet. Da die Abdeckung der benötigten Marker bei den Proben Be1, Bl3, A9, Ze4, J12-1, Nov7, Kama7, Be4, Bl1, Bl2 und Pr8 für eine Phänotyp-Bestimmung nicht ausreicht, wurden hier die MLE-vcfs mit 1-facher Abdeckung und Genotyp-Qualität 15 verwendet. Diese Filterkriterien wurden für die Definition der Haarfarbe ebenfalls für die Proben Pr3, Pr11, Pr12, Kal1, Kol4, Sol5-1 und Ze3 angewandt. Wenn keine präzise Einordnung in eine der vier möglichen Kategorien für die Haarfarbe möglich ist, kann hier die Einteilung in „eher hell“ bzw. „eher dunkel“ herangezogen werden. Für die Proben Be11, Bl5, Kama1, Bl4, Kama6 und Kama10 konnte aufgrund der fehlenden Abdeckung keine Bestimmung durchgeführt werden.

Blaue Augen konnten nur in Kombination mit blondem Haar bzw. eher hellem Haar bei 5 der folgenden Proben festgestellt werden: Ak5-8, Bl3, Pr10, Kama7, und Ze3 (Tab. 13). Eher helles Haar mit braunen Augen zeigen nur die Proben Bl2 und Kol4. Die Zuordnung braune Augen und braunes Haar konnte für zwei Proben aus Pokrovka festgestellt werden.

Zwölf Proben zeigen die Kombination braune Augen und schwarzes Haar und weitere 14 Proben braune Augen und eher dunkles Haar. Eine detaillierte Auflistung aller Proben mit entsprechenden P-Werten der Augen- und Haarfarben ist im Anhang B10 zu finden.

Tabelle 13 Phänotyp-Kombinationen der Augen – und Haarfarbe mit dem HIrisPlex-System

<i>Augenfarbe</i>	<i>Haarfarbe</i>	<i>Probe</i>
<i>Blau</i>	<i>blond</i>	Ak5-8, Bl3, Pr10, Kama7, Ze3
<i>Braun</i>	<i>eher hell</i>	Bl2, Kol4
<i>Braun</i>	<i>braun</i>	Pr3, Pr12
<i>Braun</i>	<i>schwarz</i>	A21, Ak1-1, Ak1-2, Ak5-6, Be9, Be10, J12-6, J12-7, J12-8, Pr11, Sol5-1, Sol5-9
<i>Braun</i>	<i>eher dunkel</i>	A9, A19, A20, Bl1, Be1, Be4, Be14, J12-1, Nov7, Kal1, Pr3, Pr8, Ul2, Ze4

Pigmentierung der Haut

Die Auswertung von 6 angereicherten Markern nach (Spichenok et al. 2011), die mit einer Pigmentierung der Hautfarbe assoziiert werden können, ergibt eine helle bis intermediäre Hautfarbe für alle 22 Proben, die typisiert werden konnten (Tabelle B14).

Nach der Bestimmung der Hautpigmentierung basierend auf dem Marker TYR (rs1042602) zeigen die Proben Be9, Bl2, Bl3, J12-1, J12-7 und Pr3 im Gegensatz zu den als dunkelpigmentiert klassifizierten skythischen Proben einen intermediären Hauttyp (Tab. 14). Für die bronzezeitliche Probe aus Kalinovka und die neolithische Probe Sol5-1 aus Solonzy liegt das abgeleitete Allel homozygot vor, was auf eine deutlich hellere Pigmentierung hinweist. Das ausschließlich in ost-asiatischen Populationen positiv selektierte Allel des OCA2.3 Markers konnte bei drei der vier analysierten Proben aus Justyd detektiert werden (Tab. 14). Für SLC24A5 sind A9, A21, Be1, Be10, Ak5-6, Pr10 und die neolithische Probe Sol5-1 heterozygot. Nur J12-6 aus Justyd ist homozygot mit dem abgeleiteten Allel. Die homozygote Form kann ausschließlich in osteuropäischen Populationen nachgewiesen werden. Für den Marker SLC45A2 zeigen 12 der 27 typisierten Proben den homozygot aneztralen Status (C/C). Die in europäischen Populationen nahezu fixierte homozygote Form des selektierten Allels weisen Be10, Bl2, Bl3, Kol4, Pr10 und Ze3 auf. Heterozygot in rs16891982 sind A19, Ak5-6, Ak5-8, Be1, Be11, Be14, J12-8, Pr8 und Ze4. In der nachfolgenden Tabelle sind die Ergebnisse der Typisierung für TYR, OCA2.3, SLC45A2 und SLC24A5 aufgelistet.

Für die Probe Kama1, Be4, Bl1 und Nov7 konnte keiner der Phänotyp bestimmt werden.

Tabelle 14 Genotypen der nukleären Marker die mit der Pigmentierung der Haut assoziierte werden können

Kategorie	Hauptpigmentierung			
	TYR.1	OCA2.3	SLC24A5	SLC45A2.1
rs-Nr.	rs1042602	rs1800414	rs1426654	rs16891982
CHR	11	15	15	5
Position	88911696	28197037	48426484	33951693
anz. / abgel.	C/A	T/C	G/A	C/G
A9	C/C	T/T	A/G	
A19	C/C	T/T	A/A	C/G
A20	C/C	T/T	A/A	C/C
A21	C/C	T/T	A/G	C/C
Ak1-1	C/C	T/T	A/A	C/C
Ak1-2	C/C	T/T	A/A	C/C
Ak5-6	C/C	T/T	A/G	C/G
Ak5-8	C/C	T/T	A/A	C/G
Be1			A/G	C/G
Be9	C/A	T/T	A/A	C/C
Be10	C/C	T/T	A/G	G/G
Be11				C/G
Be14				C/G
Bl2	C/A		A/A	G/G
Bl3	C/A	T/T		G/G
J12-1	C/A	T/C		
J12-6	C/C	T/C	G/G	C/C
J12-7	C/A	T/C	A/A	C/C
J12-8	C/C	T/T	A/A	C/G
Kama7	C/C			
Kal1	A/A	T/T	A/A	C/C
Kol4	C/C	T/T	A/A	G/G
Sol5-1	A/A	T/T	A/G	
Sol5-9	C/C	T/T		C/C
U12	C/C	T/T	A/A	C/C
Pr3	C/A	T/T	A/A	
Pr4	C/C	T/T		
Pr8				C/G
Pr10	C/C	T/T	A/G	G/G
Pr11				C/C
Pr12	C/C	T/T	A/A	C/C
Zc3	C/C	T/T	A/A	G/G
Zc4	C/C	T/T	A/A	C/G

Laktasepersistenz

Für die Bestimmung der Milchverträglichkeit wurden alle Individuen in beiden mit Laktasepersistenz assoziierten Loci LCTa (rs4988235 -13910C>T) und LCTb (rs182549 -22018A>G) typisiert (Tabelle B12). Milchverträglichkeit zeigen A9, Bl1, Bl2, Sol5-9, Pr4, Pr11 und sind heterozygot an einer der beiden Positionen (Tabelle B12).

Für folgende Proben ist keine Bestimmung des Phänotyps möglich: Be1, Be4, Be11, Be14, J12-6, Nov7

Xenobiotischer Metabolismus

Für rs2740574 konnten nur 17 Individuen bestimmt werden (Tabelle B17). Mit Ausnahme von Be10, mit einem heterozygoten Genotyp, zeigen alle die homozygot abgeleitete Form. A19, Be1, Be4, Be10, Be14, AK1-1 und Ak1-2 sind heterozygot für CYP3A5 rs776746, alle weiteren 18 typisierten Individuen sind homozygot für das ancestrale Allel und entsprechen so dem Haplotyp CYP3A5*3.

Für rs1801280 sind ausschließlich Be9 und J12-7 homozygot für das abgeleitete Allel, während der homozygote Genotyp für das ancestrale Allel und der heterozygote Status bei jeweils 11 Individuen bestimmt werden konnte. Von 21 typisierten Individuen für C282T ist A9 homozygot im abgeleiteten Allel und 10 Individuen weisen einen heterozygoten Status auf.

Für die Mutationen im AGT-Gen konnte der häufiger in ost-asiatischen Populationen detektierte homozygote Genotyp für das jeweils ancestrale Allel für beide SNPs für die Individuen Ak1-1, Ak1-2, J12-6, U12, Pr4 und Pr10 bestimmt werden.

Für Bl1, Bl4, Kama1, Nov7, Pr8 ist keine Bestimmung eines xenobiotischen Markers für einen der aufgeführten SNPs möglich.

Das in Europa eher seltene T-Allel für rs2032582 (ABCB1b) zeigt nur ein Individuum aus Südsibirien (J12-8). Für 13 Individuen (A20, Ak1-1, Ak5-6, Ak5-8, Be9, be10, Bl3, J12-6, J12-7, J12-8, Kal1, U12 und Pr10) konnten Allele für alle 3 SNPs im ABCB1-Gen bestimmt werden (Tabelle B17).

Alkoholverträglichkeit

Alle typisierten Individuen zeigen die homozygote Form mit dem ancestralen Allel für jeden der drei analysierten Marker ADH1Ba, ADH1Bb und ALDH2.

Für die Proben Be14 und J12-1 konnte kein Phänotyp bestimmt werden. Die Genotypen der analysierten Proben sind in der Tabelle B18 aufgelistet.

Diabetes Typ2

Basierend auf dem Model nach (Cornelis et al. 2009) konnten acht der zehn Marker-Positionen abhängig von der individuellen Abdeckung für eine Berechnungen des standardisierten Risikowerts für Diabetes Typ 2 berücksichtigt werden. Der GRS-Wert aller analysierten Proben liegt durchschnittlich bei 8,26 (min. 4 – max. 12) und damit unterhalb des

Wertes von 11.0 der Kontrollgruppe (Personen weißer Hautfarbe mit europäische Wurzeln), die in der Studie nach (Cornelis et al. 2009) typisiert wurden (Tabelle B16).

Für: Be1, Be4, Be11, Bl1, Bl2, Bl4, J12-1, Kama1, Nov7 und Pr8 sind keine oder weniger als 3 Positionen abgedeckt und somit keine Berechnung eines Diabetes Typ 2 Risikos möglich.

Für rs7903146 zeigen die Individuen aus Ak-Alacha, Justyd und Pokrovka, A19, A21; Kol4, Sol5-1, Ze3 die homozygot abgeleitete Form und A9, Be14 und Bl3 sind heterozygot. Nur Sol5-9 weist einen ancestrale homozygoten Genotyp auf. Für den zweiten Marker im TCF7L2-Gen sind Ul2, Sol5-9 und Ak1-2 homozygot für das ancestrale Allel. Konnten mehrere Individuen je Fundplatz typisiert werden, zeigen sich homozygot abgeleitete und heterozygote Genotypen.

Für Be4, Be11, Bl1, Nov7, Sol5-1 und Kama10 konnte kein Phänotyp bestimmt werden.

Bluthochdruck

rs3184504 kann nur für insgesamt 19 Individuen bestimmt werden. Davon zeigen 5 einen heterozygoten Genotyp (Tabelle B18).

SH2B3/ATXN2 rs653178 dient als *proxy* für rs3184504. Alle Individuen sind entweder heterozygot oder homozygot für beide Positionen, sofern sie abgedeckt werden konnten. Die Individuen A9, Bl3, Bl4, Kama1, Kal1 und Pr8 sind heterozygot in rs653178. Alle weiteren Individuen, für die der Genotyp bestimmt werden konnte, sind homozygot für das ancestrale Allel.

Eine Ausnahme ist eine sarmatische Probe aus Pokrovka. Pr3 ist an beiden Positionen homozygot für das abgeleitete Allel, welches mit einer Frequenz von nur 0,3 % in ostasiatischen Populationen vorkommt. Für Be1, Be4, Be11, Bl2, J12-6, Kama7, Sol5-1, Pr4, Pr10 und Ze3 ist keine Typisierung möglich.

Klimaanpassung

Homozygot für das abgeleitete Allel von rs17822931 sind A21 und zwei Individuen aus Berel (Be9 und Be10). Für EDAR zeigen Bl3 und J12-6 die homozygote Form (G/G). Be9 und Be10 sind für diese Positionen heterozygot, ebenso wie Ak5-8, Be1, Be9, Be10, Be11, Bl1, Pr4 und Ze3 (Tabelle B13). Für Individuen Bl2, Bl4, Pr3, Pr8, Pr12, Proben aus So-

lonzy und Kamennyj, Nov7, J12-1 und Be14 konnte kein Genotyp aufgrund der fehlenden Abdeckung der Positionen bestimmt werden.

Zusätzlich wurden für 13 der bereits diskutierten Marker die Allelfrequenzen des abgeleiteten Allels bestimmt und mit modernen Referenzpopulationen verglichen (Tab. 15). Die ausgewählten SNPs zeigen eine deutliche Varianz der Allel-Verteilung zwischen Populationen europäischen Ursprungs und ost-asiatischen Populationen. Aufgrund der starken Schwankung der Abdeckung verschiedener Marker-Positionen ist die Typisierung der Phänotypen nicht für alle und nur für jeweils unterschiedliche Proben möglich. Die untersuchten skythischen Individuen werden für den Vergleich der Allelfrequenzen daher in zwei große Gruppen eingeteilt. Entsprechend der geografischen Lage ihres Fundplatzes werden alle eisenzeitlichen Proben, Initiale Skythen und frühen Sarmaten einer östlichen und westlichen Gruppe zugeordnet. Die Zusammensetzung der Gruppen variiert je nach Marker und der dafür typisierten Individuen. Aufgrund der zu geringen Individuenzahl werden die Genotypen der beiden neolithischen sowie aller bronzezeitlichen Proben gesondert diskutiert. Die Frequenzen moderner Populationen sind aus Ensembl GRCh37 release 88 (<http://grch37.ensembl.org/index.html>) entnommen. Die Allelfrequenzen zentral-asiatischer Populationen und NW-Russlands basieren auf den Daten der *Allele Frequency Database ALFRED* (Osier et al. 2002; Rajeevan et al. 2003; <http://alfred.med.yale.edu>).

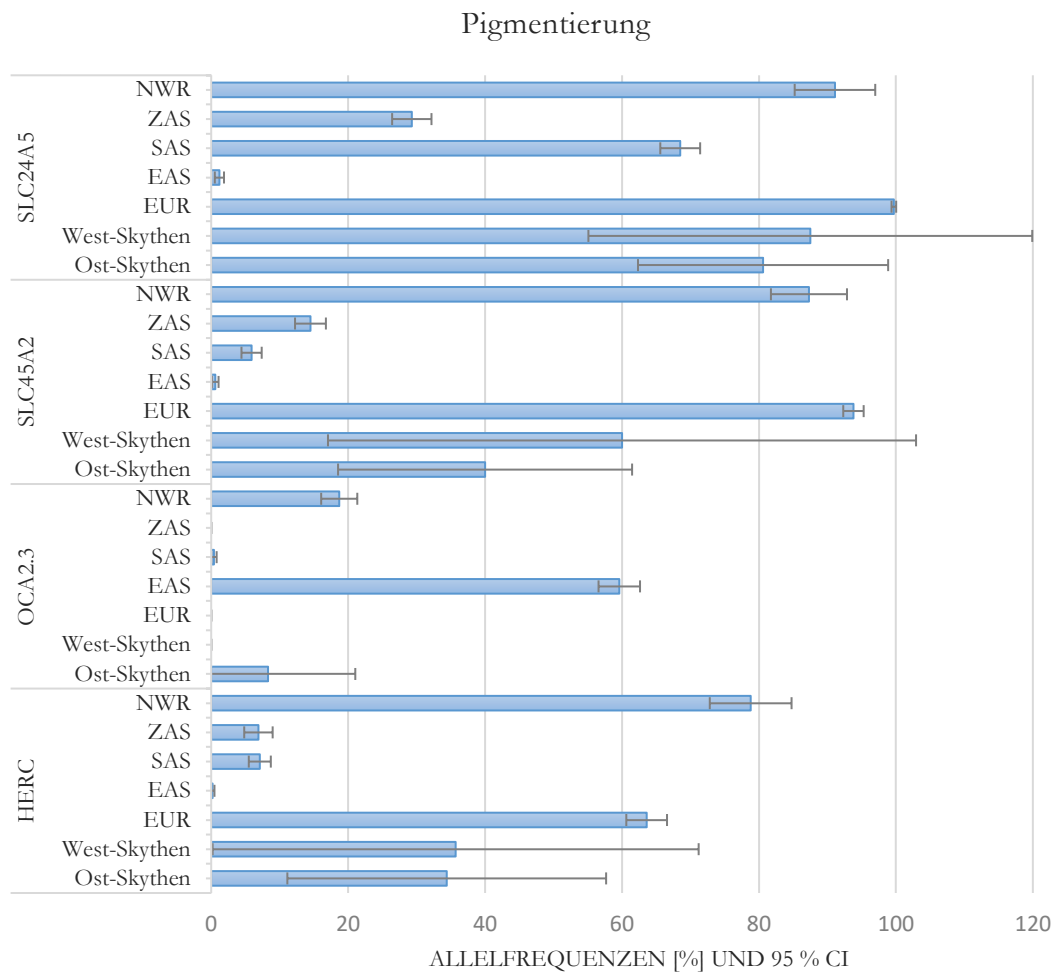


Abbildung 9: Allelfrequenzen für die abgeleiteten Allele der Pigmentierungsmarker; EUR = europäische Populationen; EAS = ost-asiatische Populationen; SAS = süd-asiatische Populationen; ZAS = zentral-asiatische Populationen; NWR = Population der Republik Komi aus NW-Russland

Die westliche und östliche Skythengruppe liegt mit den errechneten Allelfrequenzen der mit Pigmentierung assoziierten Marker zwischen europäischen und ostasiatischen Populationen, jedoch näher am europäischen Phänotyp (Abb. 9). Ausnahme ist hier die Allelfrequenz des abgeleiteten Allels für OCA2.3 der Ost-Skythen mit 8,3 %, welches in Europa und den westlichen Skythen nicht vorhanden ist.

Xenobiotischer Metabolismus

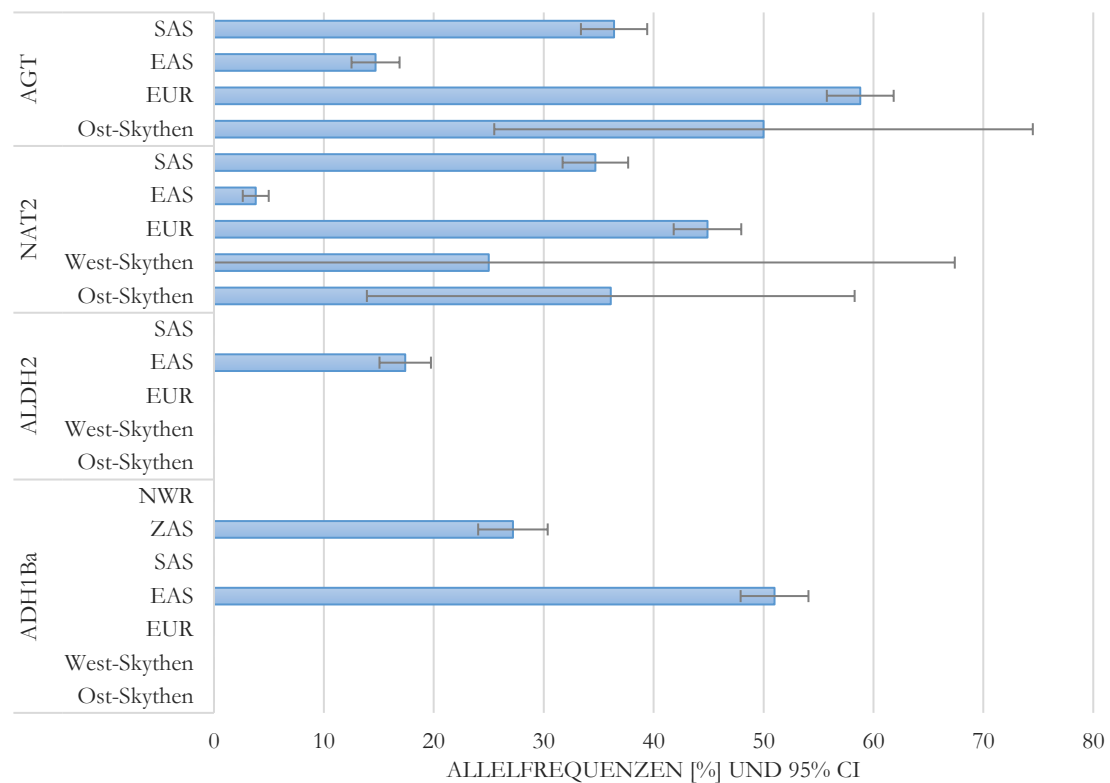


Abbildung 10: Allelfrequenzen für die abgeleiteten Allele der Marker, die mit der Verstoffwechslung von Xenobiotika assoziiert werden; EUR = europäische Populationen; EAS = ost-asiatische Populationen; SAS = süd-asiatische Populationen; ZAS = zentral-asiatische Populationen; NWR = Population der Republik Komi aus NW-Russland

Das abgeleitete Allel der Marker ADH1Ba und ALDH2 kann weder in europäischen Populationen noch in einer der beiden skythischen Gruppierungen nachgewiesen werden (Abb. 10). Für den AGT *locus* konnte die Allelfrequenz des abgeleiteten Allels nur für die Ost-Skythen bestimmt werden und liegt bei 50 %. Allelfrequenzen des NAT2-Markers liegen bei den Ostskythen (N=18) bei 36,1 % und bei 25 % bei den Westskythen (N=4).

Klimaanpassung & Laktasepersistenz

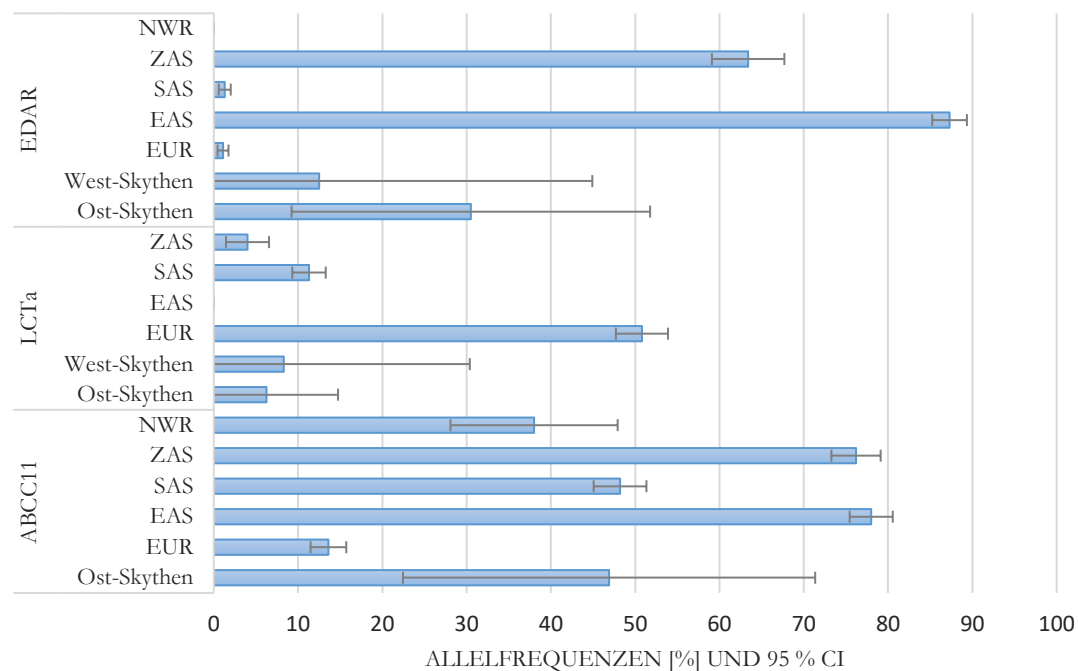


Abbildung 11: Allelfrequenzen der abgeleiteten Allele des Laktase-Markers und der Marker, die mit einer Anpassung an klimatische Verhältnisse assoziiert werden können; EUR = europäische Populationen; EAS = ost-asiatische Populationen; SAS = süd-asiatische Populationen; ZAS = zentral-asiatische Populationen; NWR = Population der Republik Komi aus NW-Russland

Für den Marker LCTa, dessen abgeleitetes Allel mit einer Milchverträglichkeit korreliert und in ost-asiatischen Populationen gar nicht und in zentral-asiatischen Population mit nur 4% auftritt, liegen die westlichen Skythen mit einer Allelfrequenz von 8,3 % und die östlichen Skythen mit einer Frequenz von 6,25 % nahe an der Frequenzverteilung zentral – und süd-asiatischer Populationen mit 4 % und 11.3 % (Abb. 11).

Die Allelfrequenzen der Marker EDAR und ABCC11 liegen zwischen europäischen und ost-asiatischen Populationen. Für das abgeleitete Allel des EDAR-Markers konnte für die Ost-Skythen eine deutlich höhere Frequenz als für die West-Skythen nachgewiesen werden.

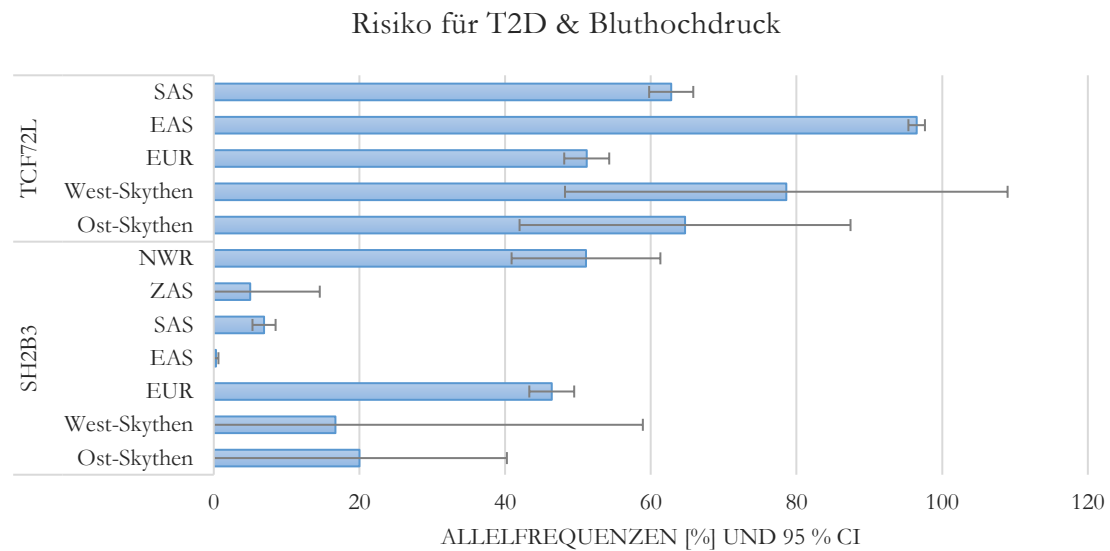


Abbildung 12: Allelfrequenzen der abgeleiteten Allele, die mit einem erhöhten Risiko für Diabetes Typ 2 und Bluthochdruck korrelieren; EUR = europäische Populationen; EAS = ost-asiatische Populationen; SAS = Süd-asiatische Populationen; ZAS = Zentral-asiatische Populationen; NWR = Population der Republik Komi aus NW-Russland

Für die Marker, die mit einem Risiko für Typ 2 Diabetes und Bluthochdruck assoziiert werden können, liegen die Frequenzen der selektierten Allele zwischen jenen von europäischen und ost-asiatischen bzw. zentral-asiatischen Populationen (Abb. 12).

Für die beiden neolithischen Proben aus West-Sibirien und die bronzzeitlichen Individuen aus SW-Russland konnten aufgrund fehlender Abdeckung der SNPs und einer geringen Individuenzahl je Zeitalter keine Allelfrequenzen berechnet werden. Für die Probe aus Kalinovka (Kal1) konnten jedoch 12 von 13 Positionen abgedeckt werden und für eine Probe der Woronesch-Region (Kama7) konnte der Genotyp an zwei Positionen bestimmt werden.

Eine Bestimmung des Genotyps für die west-sibirischen Proben war für 9 (Sol5-9) bzw. 5 (Sol5-1) Marker möglich

Kal1 sowie Sol5-1 und Sol5-9 entsprechend in Positionen mit stark selektierten Allelen zwischen europäischen und ost-asiatischen Populationen eher dem europäischen Phänotyp. Für rs12913832 zeigen die neolithischen Proben mit dem homozygot ancestralen Genotyp eine Nähe zu asiatischen Populationen, ebenso für den Marker LCTa mit der homozygot ancestralen Form (G/G) (Tabelle B12).

Tabelle 15 Allelfrequenzen der abgeleiteten Allele ausgewählter Marker; anz. = ancestrale Allele; abgl. = abgeleitetes Allel; N = Anzahl der Individuen; Allelfreq. = Allelfrequenzen; EUR = europäische Populationen; EAS = ost-asiatische Populationen; SAS = Süd-asiatische Populationen; ZAS = Zentral-asiatische Populationen; NWR = Population der Republik Komi aus NW-Russland

Marker	ABCC11		ADH1Ba		AGT		ALDH2		EDAR		HERC2		LCTa	
rs-Nr.	rs17822931		rs3811801		rs699		rs671		rs3827760		rs12913832		rs4988235	
Phänotyp	trockenes Zerumen/reduzierte Transpiration		Alkoholverträglichkeit		Xenobiotischer Metabolismus		Alkoholverträglichkeit		Struktur Kopfbehaarung		Augenfarbe		Laktasepersistenz	
anz./ abgl.	C/T		G/A		G/A		G/A		A / G		A / G		G/A	
	N	Allelfreq.	N	Allelfreq.	N	Allelfreq.	N	Allelfreq.	N	Allelfreq.	N	Allelfreq.	N	Allelfreq.
EUR	1006	0,136	1006	0	1006	0,588	1006	0	1006	0,011	1006	0,636	1006	0,508
EAS	1008	0,780	1008	0,510	1008	0,147	1008	0,174	1008	0,873	1008	0,002	1008	0
SAS	978	0,482	978	0	978	0,364	978	0	978	0,013	978	0,071	978	0,113
ZAS	814	0,762	760	0,272	-	-	-	-	482	0,634	572	0,069	226*	0,04
NWR	92	0,38	92	0	-	-	-	-	92	0	180	0,7879	-	-
Ost-Skythen	16	0,469	18	0	16	0,5	18	0	18	0,305	16	0,344	16	0,0625
West-Skythen	1	-	3	0	2	-	5	0	4	0,125	7	0,357	6	0,083

Marker	OCA2.3		TCF7L2		SH2B3		SLC45A2		SLC24A5		NAT2	
rs-Nr.	rs1800414		rs11196205		rs3184504		rs16891982		rs1426654		rs1801280	
Phänotyp	Hautpigmentierung		Risiko für Diabetes		Risiko für Bluthochdruck		Pigmentierung		Pigmentierung		Xenobiotischer Metabolismus	
anz./ abgl.	T/C		C/G		C/T		C/G		G/A		T/C	
	N	Allelfreq.	N	Allelfreq.	N	Allelfreq.	N	Allelfreq.	N	Allelfreq.	N	Allelfreq.
EUR	1006	0	1006	0,512	1006	0,464	1006	0,938	1000	0,997	1006	0,449
EAS	1008	0,596	1008	0,965	1008	0,003	1000	0,006	1008	0,012	1008	0,038
SAS	978	0,004	978	0,628	978	0,069	978	0,059	978	0,685	978	0,347
ZAS	92	0	-	-	20	0,050	938	0,145	964	0,293	-	-
NWR	840	0,187	-	-	92	0,511	138	0,873	90	0,911	-	-
Ost-Skythen	18	0,083	17	0,647	15	0,2	20	0,4	18	0,806	18	0,361
West-Skythen	5	0	7	0,786	3	0,167	5	0,6	4	0,875	4	0,250

4.6 Genetische Strukturen skythischer Populationen

Das Verhältnis skythischer Populationen zu modernen Populationen

Die in der Abbildung 13 dargestellte **PCA** zeigt alle für die vorliegende Arbeit angereicherten Proben, welche mittels LASER in einen reduzierten Referenzraum modernen Populationen aus (Hellenthal et al. 2014; Busby et al. 2015) projiziert wurden. Die Hauptkomponentenanalyse integriert zusätzlich Referenz-Individuen aus relevanten Studien alter DNA (aRef). Die ausführliche Darstellung dieser Kulturgruppen aus der Literatur ist in der nachfolgenden PCA in Abbildung 27 zu sehen. Die modernen Populationen sind in grau abgebildet. Die geografischen Bezeichnungen in der Abbildung dienen einer groben Orientierung. Eine detaillierte Darstellung der modernen Populationen in der PCA ist im Anhang B19 zu finden.

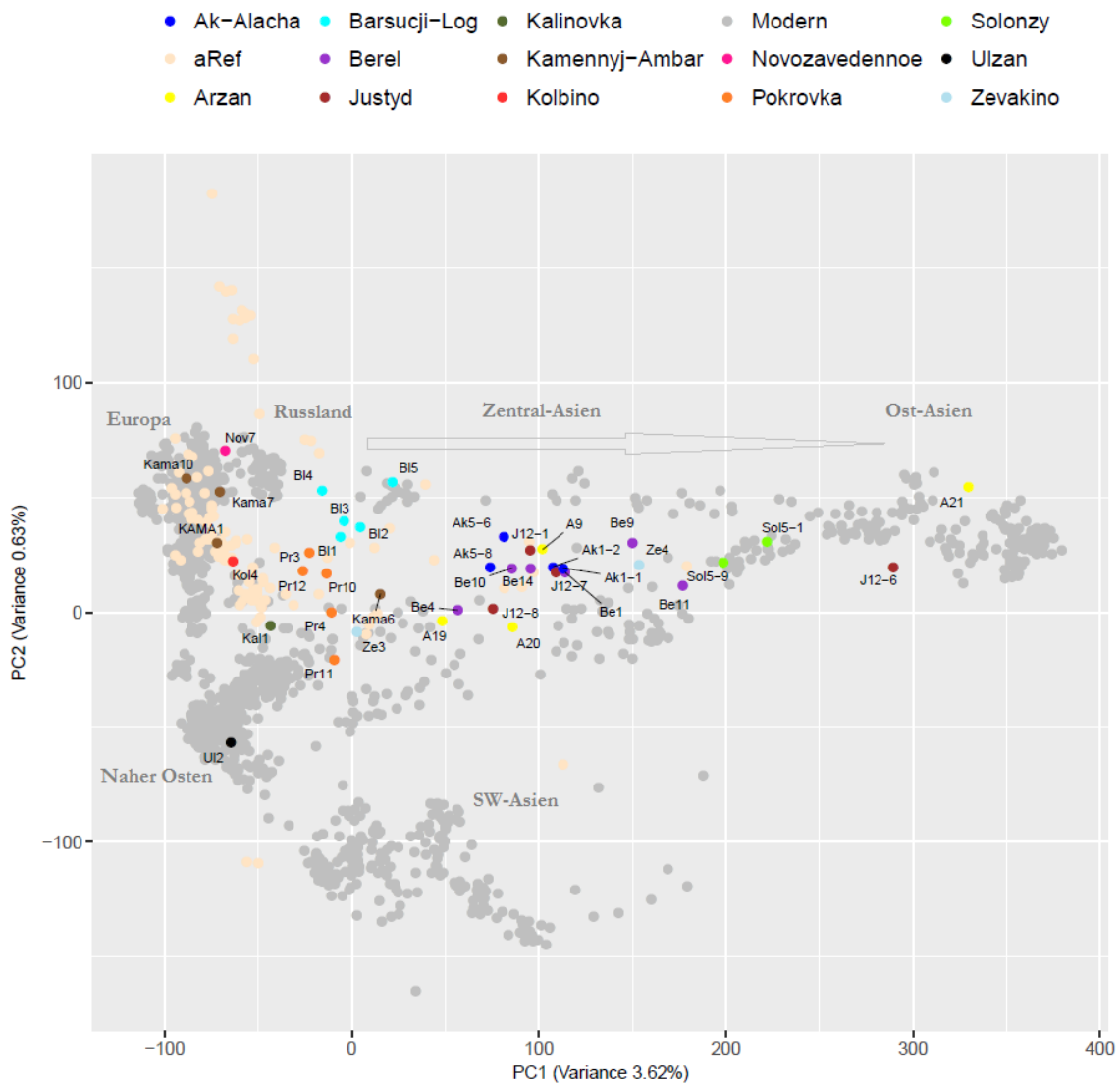


Abbildung 13: PCA neolithischer, bronze - und eisenzeitlicher Individuen der vorliegenden Arbeit und ausgewählter aDNA Referenzen =aRef (Lazaridis et al. 2014; Fu et al. 2014; Jones et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015; Jones et al. 2017; Sikora et al. 2017). Mittels LASER in einen Referenzraum moderner Populationen projiziert (Hellenthal et al. 2014; Busby et al. 2015); Farbgebung entsprechend der Fundplätze

Alle skythischen Individuen fallen in die Verteilung moderner Populationen (Abb. 13). Die Mehrheit der Proben erstreckt sich entlang der PCA1 von Osteuropa über den zentralasiatischen Raum nach Ost-Asien.

Die östlichen Proben der Aldy-Bel Kultur Südsibiriens der Tuva-Region, Vertreter der Pazyryk-Kultur Südsibiriens, des kasachischen Altai und einer ost-kasachischen Probe der Zevakino- Čilikta -Stufe bilden ein mittleres Cluster in Zentralasien beginnend von Tadschikistan bis Kirgisistan und dem Altai. Eine deutlich abweichende Position zeigt die ost-kasachische Probe Ze3 mit einer deutlichen Orientierung Richtung Osteuropa und der

Balkan Region und ihrer Entfernung zur zweiten Probe Ze4 desselben Fundplatzes. Die beiden süd-sibirischen Proben A21 und J12-6 liegen zwischen modernen Individuen der Mongolei und ost-asiatischen Populationen.

Die frühen Sarmaten aus Pokrovka in Südwest-Russland liegen im Grenzgebiet zwischen Osteuropa und Zentralasien und erstrecken sich bis in den persischen Raum und die Balkan-Region. Die bronzezeitliche Probe Kama6 aus Kamennyj-Ambar fällt in die Nähe der Probe Ze3 aus Ostkasachstan, jedoch liegen die drei weiteren Proben desselben Fundplatzes in Mittel- und Osteuropa. Die Proben der Tagar-Kultur aus dem Minusinsker Becken fallen in die Nähe der Tschuwaschen und liegen somit weit westlicher als ihr Fundplatz. Die den Saken zugeordnete Probe aus Ulžan in Süd-Ost-Kasachstan weicht in ihrer Position in der PCA ebenfalls stark von der geografischen Lage ihres Fundplatzes ab und liegt in der Verbreitung von modernen syrischen und iranischen Individuen.

Die bronzezeitliche Probe aus Kalinovka sowie die skythische Probe aus SW-Russland (Nov7) fallen in die Verteilung moderner osteuropäischer bzw. russischer Populationen.

Die beiden neolithischen Proben verteilen sich zwischen Burjaten und Nganasanen Sibiriens.

Basierend auf den Ergebnissen werden in weiteren Analysen die Individuen A21 und J12-6 und Proben der Tagar-Kultur ebenso wie die Probe aus Ulžan teilweise nicht mit anderen Proben desselben Fundplatzes bzw. derselben Zeitstufe gruppiert.

Die skythischen Individuen werden entsprechend ihrer zeitlichen und kulturellen Zuordnung für die Berechnung der **Fst-Werte** gruppiert. Eine ausführliche Tabelle aller berechneten Fst-Werte ist im Anhang *Ergänzende-Informationen_I1* zu finden. Eine Auflistung der Fst-Werte der skythischen Gruppierungen zueinander wird in Tabelle 16 dargestellt. Die folgende Abbildung zeigt die paarweisen Fst-Werte einzelner Gruppierungen der untersuchten Proben zueinander und zu allen asiatischen, CEU und TSI-Populationen des *1000 Genome Phase 3* Referenzdatensatzes (1000 Genomes Project Consortium et al. 2015). Die Abkürzungen für die verwendeten Populationen richten sich nach der Nomenklatur 1000 Genome-Projekts. Die Werte sind entsprechend ihrer Höhe farblich kodiert und werden in Form einer *heatmap* dargestellt. Weiß bis hellgelb entspricht einer hohen genetischen Distanz der Populationen, dunkelrot einer niedrigen Distanz. Ausgehend von den hier errechneten Fst-Werten zwischen modernen Populationen wird folgende Einteilung der Werte vorgenommen: bei einem Fst-Wert von 0-0,01 kann keine genetische Distanz angenom-

men werden, 0,01-0,04 zeigt eine geringe und 0,04-0,06 eine mittlere Distanz an, bei einem Fst-Wert von $> 0,06$ kann von zwei getrennten Populationen ausgegangen werden.

Aufgrund der starken Korrelation zwischen den berechneten Fst-Werte nach Hudson und Pattersons F2 (Patterson et al. 2012) miteinander werden ausschließlich Fst-Werte in der vorliegenden Arbeit dargestellt und diskutiert (Abb. 14).

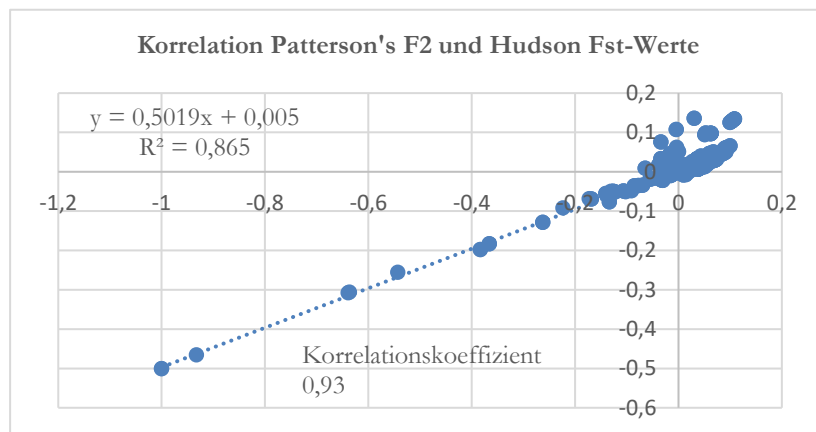


Abbildung 14: Korrelation der errechneten F2-Werte nach Patterson (Patterson et al. 2012) und Hudson Fst

Die modernen Populationen des Referenzdatensatzes zeigen zu erwartende und bekannte genetische Distanzen zueinander (Abb. 15)

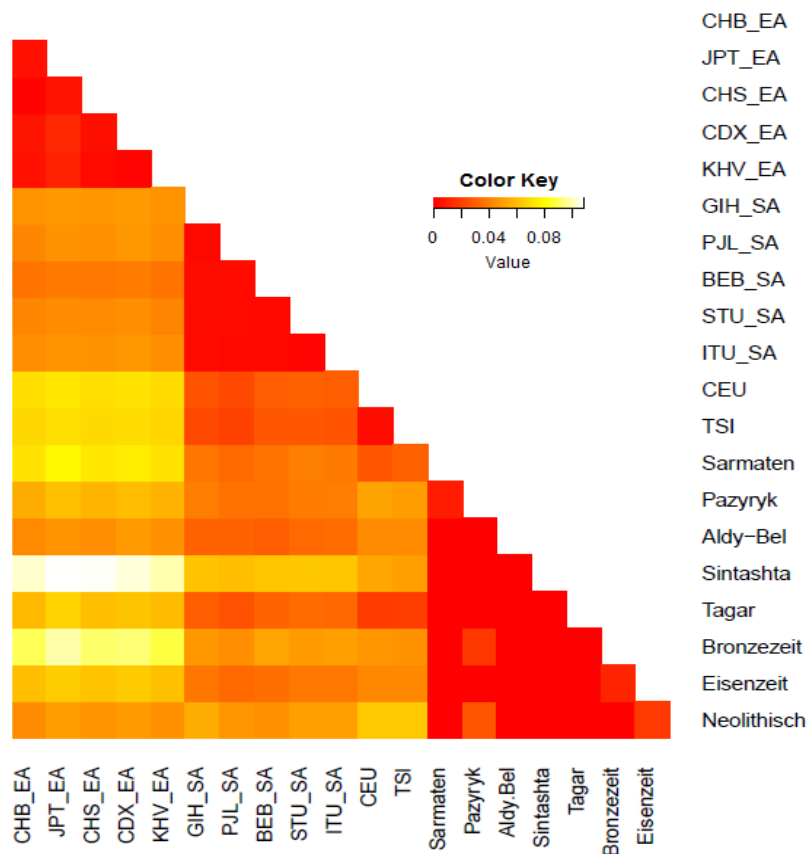


Abbildung 15: heatmap der berechneten Fst-Werte nach Hudson zwischen *1000 genomes* Populationen und skythischen Proben, gruppiert nach Zeit – und Kulturstufen; EA = ost-asiatische Populationen; SA = süd-asiatische Populationen

Die Distanzen zu den ost-asiatischen Populationen des modernen Datensatzes sind für die bronzezeitlichen Proben aus Kalinovka und Kamennyj Ambar am höchsten. Die beiden neolithischen Proben aus Solonzy zeigen die höchste Distanz (0,0637 und 0,0638) zu europäischen Populationen. Alle Proben, die in die Eisenzeit datiert werden, zu denen die Fundplätze Aržan, Ak-Alacha, Berel, Barsucij Log, Justyd, Kolbino, Novozavedennoe, Pokrovka, Ulžan und Zevakino gehören, weisen höhere genetische Distanzen (0,0603-0,0651) zu ost-asiatischen Populationen auf, als zu Populationen Süd-Asiens (0,0337-0,0383) oder Europas (0,0436; 0,0432). Betrachtet man nur die Individuen des Fundplatzes Kamennyj-Ambar der Sintashta-Kultur liegen die Fst-Werte $> 0,09$ zu allen ost-asiatischen Populationen. Wobei die genetische Distanz zu CEU/TSI geringer ist als zu allen Populationen Süd-Asiens. Proben des Fundplatzes Barsučij Log (Tagar-Kultur) in Sibirien und Individuen der Aldy-Bel Stufe aus Tuva zeigen höhere Fst-Werte im Vergleich

zu modernen Populationen, die der Lokalisation ihres Fundplatzes entsprechen. Proben der Tagar-Kultur weisen eine geringe genetische Distanz (0,0188 und 0,0197) zu europäischen Populationen auf und mittlere bis hohe Werte (0,0592- 0,0672) für alle ost-asiatischen Populationen. Wohingegen die Proben aus Tuva nur mittlere Fst-Werte (0,0439-0,0488) zu Populationen Ost-Asiens und Europa zeigen und die geringste Distanz zu süd-asiatischen Populationen (0,0299-0,0343).

Bei den Sarmaten aus Pokrovka werden die niedrigsten Fst-Werte für die europäischen Populationen berechnet (0,0279) und hohe Werte für alle Populationen Ost-Asiens (0,0717-0,0791). Die Gruppierung in Proben der Pazyryk-Kultur umfasst die meisten Proben. Hier zeigen sich mittlere bis hohe genetische Distanzen zu ost-asiatischen Populationen (0,0550-0,0611), jedoch niedrigere Fst-Werte für alle süd-asiatischen Populationen (0,0361-0,0400) im Vergleich zu Populationen Europas (0,0497-0,0518).

Basierend auf der geografischen Verteilung der Fundplätze wurden 23 moderne Populationen für die **f3-Statistik** ausgewählt. Eine Auflistung der höchsten 20 Werte der f3-Statistik inklusive Standardfehler, Z-Wert und SNP-Anzahl sind in den *Ergänzenden Informationen_I2* und *I3* tabellarisch dargestellt.

In den folgenden grafischen Darstellungen der f3-Werte sind jeweils die höchsten 20 Werte ausgewählt. Für alle f3-Statistiken mit dem modernen Referenzdatensatz von Hellenthal (Hellenthal et al. 2014) und Busby (Busby et al. 2015) sowie prähistorischen Kulturgruppen aus (Lazaridis et al. 2016) liegt kein Z-Wert < 3 vor. Die Einteilung der skythischen Proben nach Fundplätzen mit Ausnahme der PCA-Außenseiter wird beibehalten. Aufgrund der deutlichen Abweichung von ihrer geografischen Lokalisation und ihrem Verhältnis zu anderen Proben desselben Fundplatzes werden die Proben A21, J12-6, Ze3 und Ze4 gesondert betrachtet. Der durchschnittliche f3-Wert für Analysen mit modernen Populationen als Test Population liegt bei 0,277 (0,250 – 0,336) mit einem Standardfehler von \emptyset 0,007 (0,004 – 0,014). Werden prähistorische Populationen als Test in der *outgroup-f3*-Statistik eingesetzt, liegt der f3-Wert bei \emptyset 0,266 (0,235 – 0,303) mit einem Standardfehler von 0,003 – 0,030 bei den nach Fundplatz gruppierten, skythischen Individuen. Werden diese ebenfalls in die f3-Analyse mit einbezogen reichen die Werte von 0,243 bis 0,355 (*std. err:* 0,003 -0,033).

Genetische Drift zwischen eisenzeitlichen Proben SW-Russlands und modernen Populationen

Die Probe aus dem Oblast Voronezh, die frühen Sarmaten aus Orenburg, sowie Nov7 aus dem Nord-Kaukasus zeigen die fünf höchsten f3-Werte für Russland, Weißrussland, Polen, der Ukraine und Finnland in unterschiedlicher Reihenfolge (Abb. 16). Die niedrigsten f3-Werte zeigen osteurasische Populationen, der Nahen Osten und Populationen SW-Asiens. Mit Ausnahme der höchsten Werte fallen die f3-Werte in den Bereich des Standardfehlers voneinander.

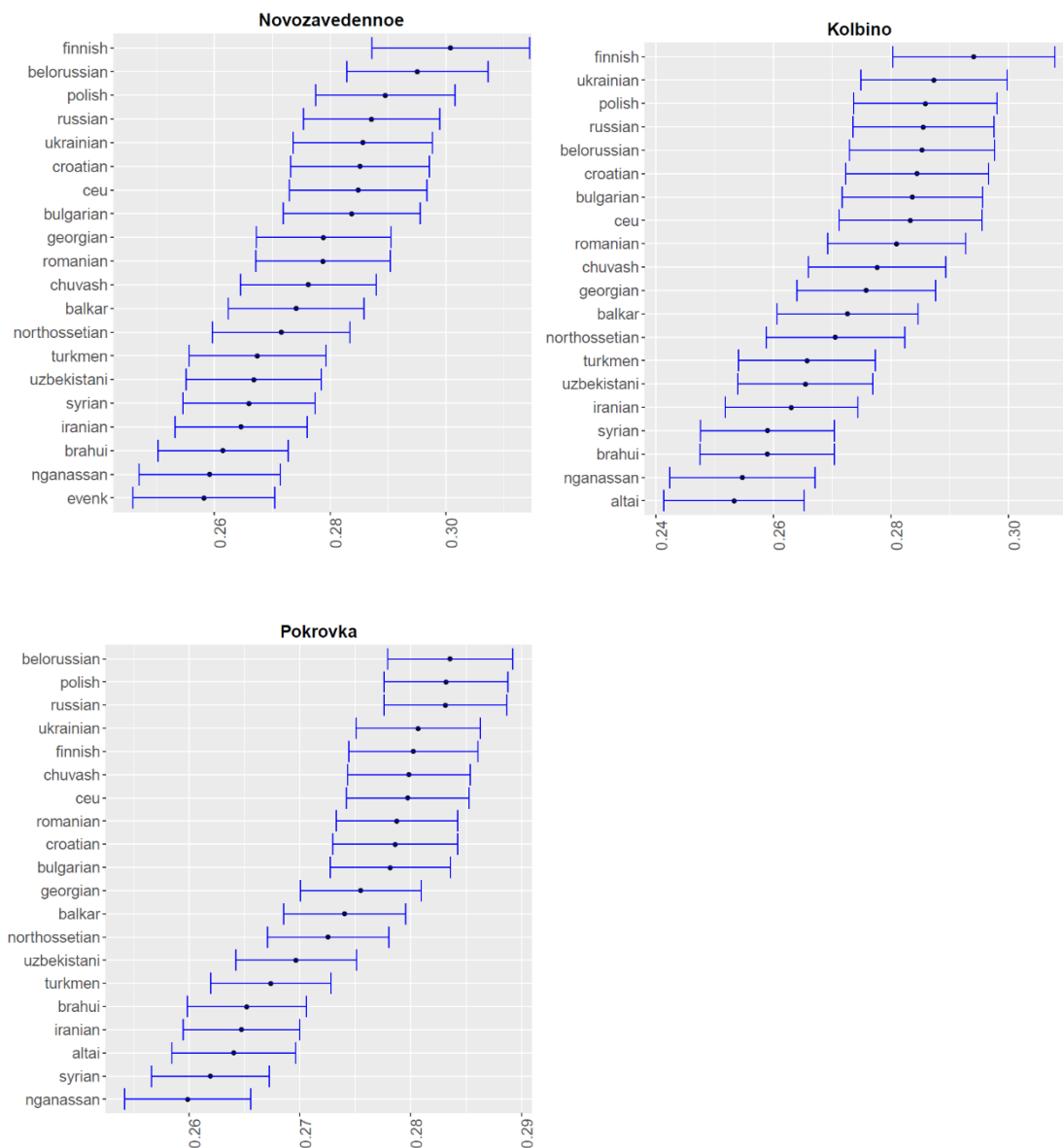


Abbildung 16: outgroup-f3-Statistik eisenzeitlicher Proben SW-Russlands im Vergleich zu modernen Referenzpopulationen der Form $f_3(6=S_Mbuti; \text{Modern, Skythen})$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Genetische Drift zwischen eisenzeitlichen Proben Zentralasiens und modernen Populationen

Die eisenzeitlichen Proben Zentralasiens können in die Fundplätze Aržan, Ak-Alacha, Barsučij Log und Justyd aufgeteilt werden.

Alle Proben aus dem Minusinsker Becken zeigen, wie in der PCA, die meiste genetische Drift mit Finnland, Polen, Russland/Weißrussland und CEU gefolgt von Ländern Osteuropas und des Balkans (Abb. 17). Die niedrigsten f_3 -Werte liegen bei den modernen osteuropäischen Populationen.

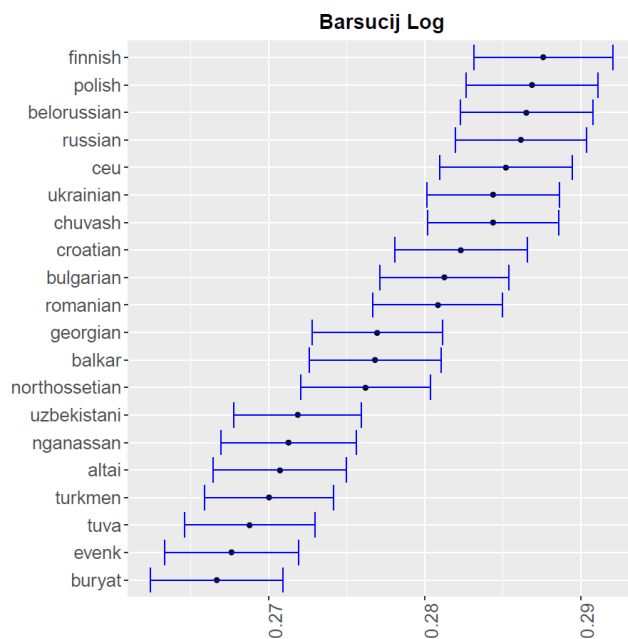


Abbildung 17: *outgroup-f3*-Statistik der Proben aus dem Minusinsker Becken im Vergleich zu modernen Referenzpopulationen der Form $f_3(6=S_Mbuti; Modern, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Individuen der Fundplätze Aržan, Ak-Alacha, und Justyd, die Außenseiter A21 und J12-6 ausgenommen, teilen die meiste genetische Drift mit Ewenken und Nganasanen und Populationen aus Tuva und des Altais bzw. Burjaten. (Abb. 18, 19, 20) Für die Proben der Region Tuva (Aržan) und des Chuya-Tals (Justyd) sind zusätzlich Tschuwaschen und Finnen unter den höchsten fünf f_3 -Werten.

Im Vergleich zu den restlichen Proben Aržans teilt die süd-sibirische Probe A21 die meiste genetische Drift mit Populationen Sibiriens und Ost-Asiens und mit einer deutlichen Differenz der f_3 -Werte für Zentralasien und Osteuropa (Abb. 18, links) Dieselbe Differenzierung der geteilten, genetischen Drift kann für J12-6 im Vergleich zu den andern Proben des Chuya-Tals festgestellt werden (Abb.19, links).

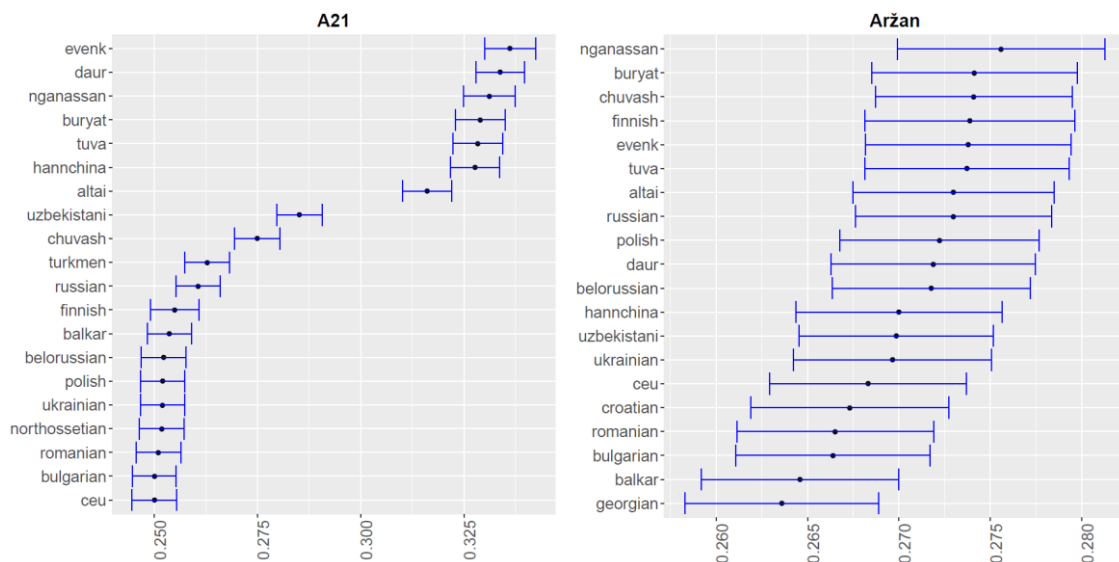


Abbildung 18: *outgroup-f3*-Statistik der Proben aus Tuva im Vergleich zu modernen Referenzpopulationen der Form $f_3(6=S_Mbuti; \text{Modern, Skythen})$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

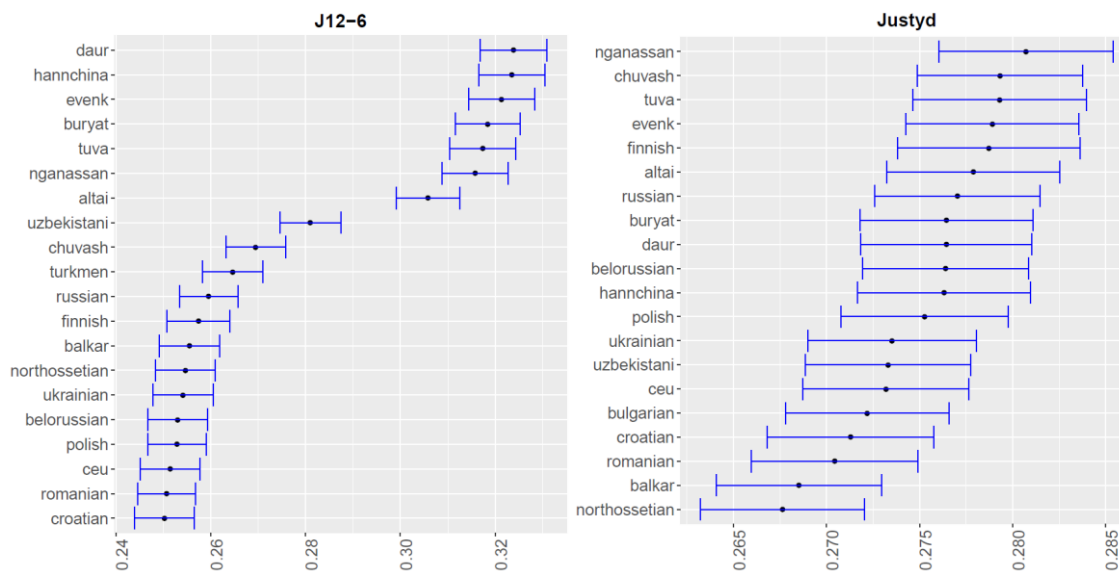


Abbildung 19: *outgroup-f3*-Statistik der Proben aus Justyd im Vergleich zu modernen Referenzpopulationen der Form $f_3(6=S_Mbuti; \text{Modern, Skythen})$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Die süd-sibirischen Proben aus Ak-Alacha teilen die meiste genetische Drift mit modernen Populationen Sibiriens und Ost-Asiens, gefolgt von Russland, Finnland, Weißrussland und Polen (Abb. 20). Die niedrigsten Werte liegen bei Populationen des Balkans.

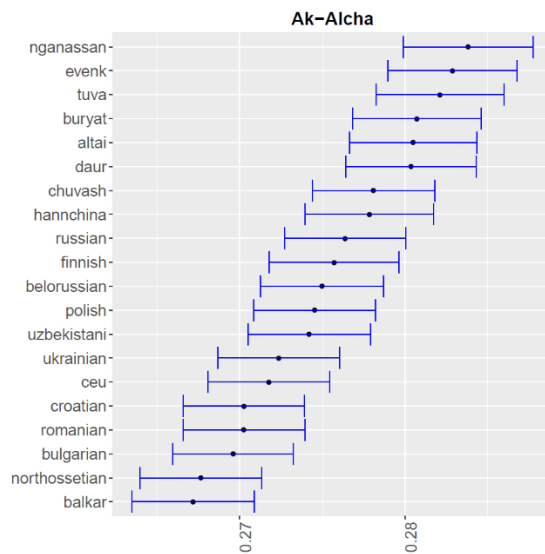


Abbildung 20: *outgroup-f3*-Statistik der Proben aus Ak-Alacha im Vergleich zu modernen Referenzpopulationen der Form $f_3(6=S_Mbuti; \text{Modern, Skythen})$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Drei Fundplätze liegen in Kasachstan: Ulžan, Zevakino und Berel. Während die Proben aus dem kasachischen Altai-Gebirge (Berel) die höchsten f_3 -Werte für osteurasische und ostasiatische Populationen zeigen, liegen für Ulžan 13 Populationen mit der höchsten genetischen Drift jeweils im Bereich des Standardfehlers voneinander (Abb. 21). Es ist jedoch eine deutliche Abgrenzung zu Populationen des Nahen Ostens und Zentralasiens erkennbar.

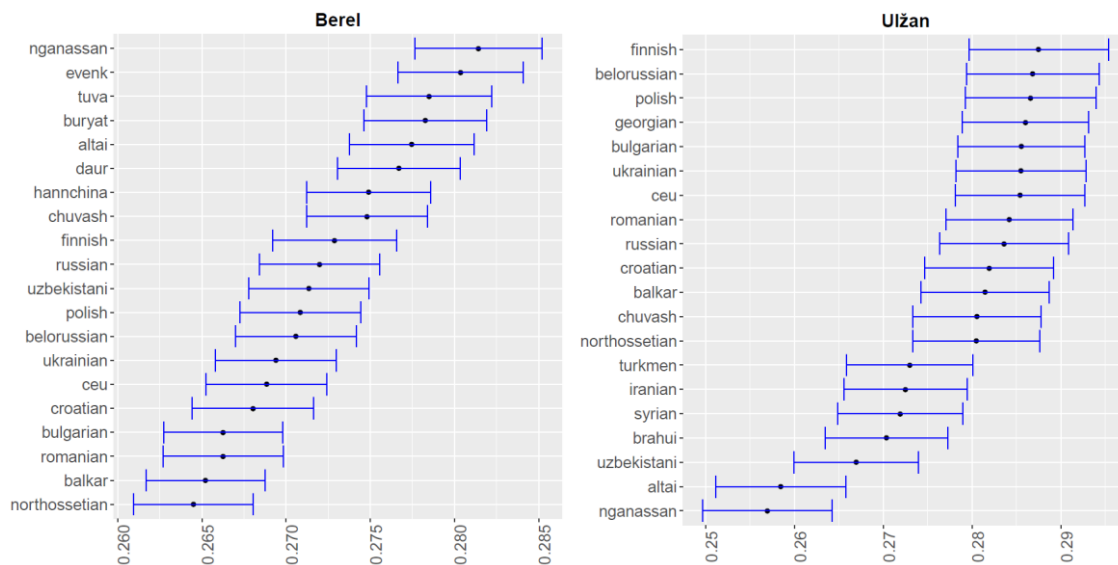


Abbildung 21: *outgroup-f3*-Statistik der kasachischen Proben im Vergleich zu modernen Referenzpopulationen der Form $f_3(6=S_Mbuti; \text{Modern, Skythen})$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Die ost-kasachische Probe Ze4 weist im Vergleich zu Ze3 desselben Fundplatzes die meiste genetische Drift mit osteurasischen und ost-asiatischen Populationen auf (Abb. 22). Ze3 zeigt durch hohe f_3 -Werte für Finnland, Weißrussland/Russland, Polen CEU, Kroatien und der Ukraine eine Nähe zu europäischen und osteuropäischen Populationen.

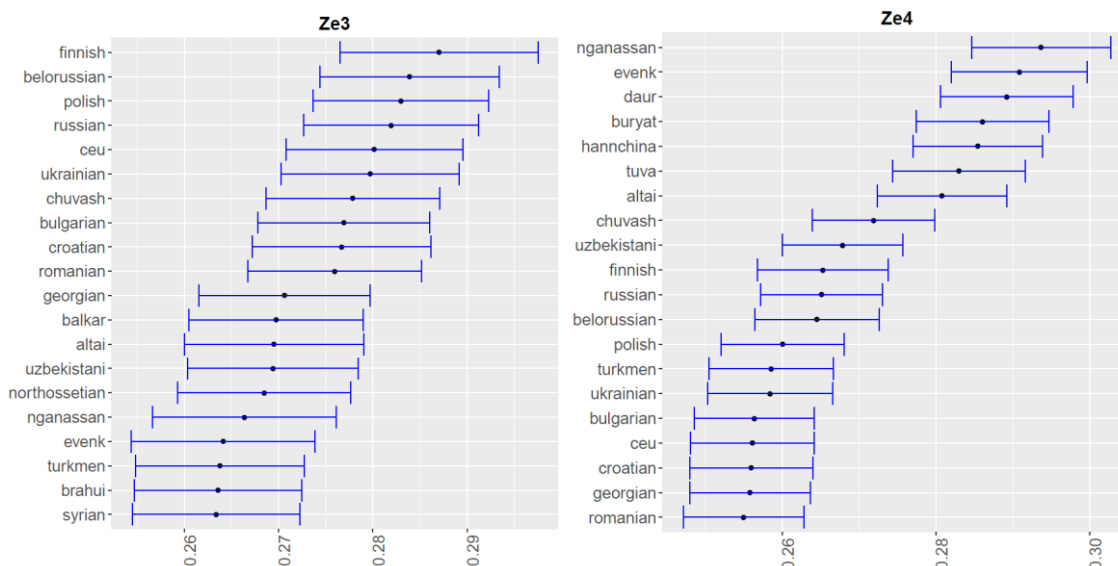


Abbildung 22: *outgroup-f3*-Statistik der ost-kasachischen Proben aus Zevakino im Vergleich zu modernen Referenzpopulationen der Form $f_3(6=S_Mbuti; \text{Modern, Skythen})$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Genetische Drift zwischen bronzezeitlichen Proben SW-Russlands und modernen Populationen

Die beiden Fundplätze der frühen bis mittleren Bronzezeit Kalinovka und Kamennyj Ambar weisen die fünf höchsten Werte der *outgroup-f3*-Statistik für CEU, Polen, Russland, Weißrussland und Finnland und die niedrigsten Werte für moderne, sibirische Populationen, Usbekistan und Populationen des Nahen Ostens auf (Abb. 23).

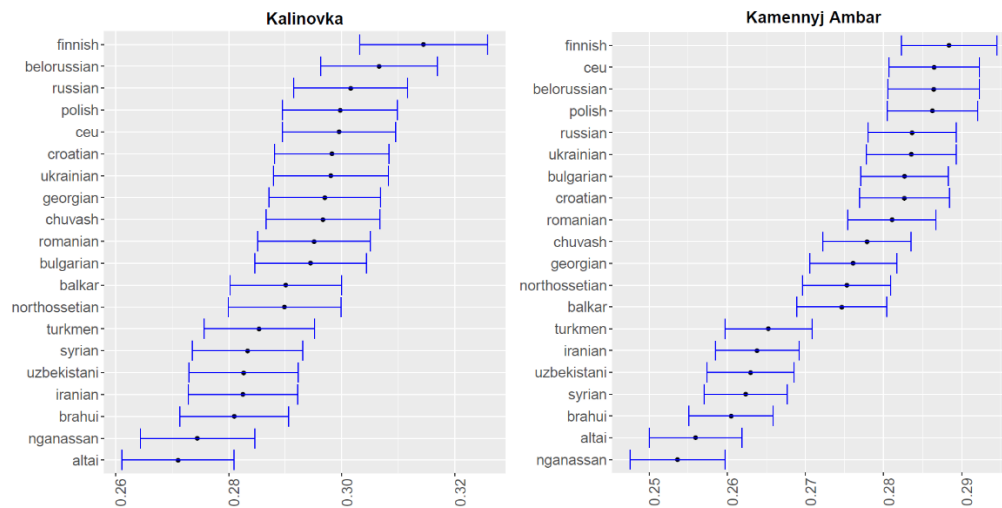


Abbildung 23: : *outgroup-f3*-Statistik bronzezeitlicher Proben SW-Russlands im Vergleich zu modernen Referenzpopulationen der Form $f_3(6=S_Mbuti; \text{Modern, Skythen})$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Werden die vier Individuen der Sintaschta Kultur im Trans-Ural einzeln getestet, sind sich unter den fünf höchsten *f3*-Werten für alle Individuen Polen, Russland und Weißrussland vertreten (*Ergänzende Informationen_I2*).

Genetische Drift zwischen neolithischen Proben Sibiriens und modernen Populationen

Für die neolithischen Proben aus Solonzy wird der höchste Anteil gemeinsamer genetischer Drift zu modernen Populationen Sibiriens und der Mongolei (Ewenken, Nganasanen und Burjaten) sowie ost-asiatischen Populationen gemessen (Abb. 24). Weiter sind mittlere *f3*-Werte zu zentralasiatischen Populationen festzustellen, gefolgt von osteuropäischen Populationen und dem Balkan.

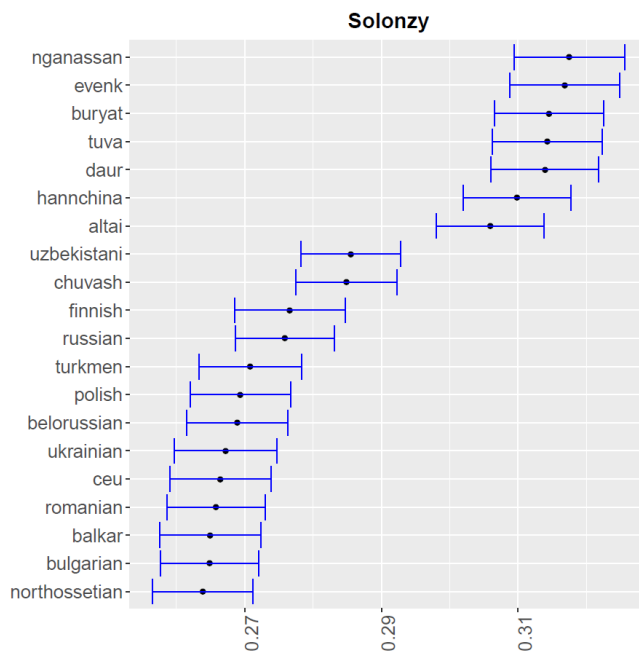


Abbildung 24: *outgroup-f3*-Statistik neolithischer Proben Sibiriens im Vergleich zu modernen Referenzpopulationen der Form $f3(6=S_Mbuti; \text{Modern, Skythen})$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Das Verhältnis der eisenzeitlichen Individuen Zentral – und Nord-Asiens zu den Individuen des Schwarzmeer-Raums und SW-Russlands

Für die skythischen Gruppierungen untereinander konnten ausschließlich **Fst-Werte** zwischen 0 – 0,0267 berechnet werden (Tab. 16). Wobei die Individuen der Pazyryk-Kultur im Vergleich zu den neolithischen Proben aus West-Sibirien und allen bronzzeitlichen Individuen aus SW-Russland positive Werte (0,0267 und 0,0180) erzielen. Hier kann also eine sehr geringe bzw. mittlere genetische Distanz angenommen werden kann. Alle eisenzeitlichen Proben, unabhängig ihrer räumlichen Verteilung, zeigen leicht erhöhte Fst-Werte zu neolithischen (0,0186) wie bronzzeitlichen (0,0116) Proben.

Tabelle 16 Paarweise Fst-Werte nach Hudson zwischen skythischen Gruppen verschiedener Kulturstufen und Datierungen; Anzahl der Individuen in Klammern (N)

	Sarmaten (6)	Pazyryk (14)	Aldy-Bel (4)	Sintaschta (4)	Tagar (5)	Bronzezeit (5)	Eisenzeit (34)
Pazyryk	0.0093						
Aldy-Bel	0	0					
Sintaschta	0	0	0				
Tagar	0	0	0	0			
Bronzezeit	0	0.0180	0	0	0		
Eisenzeit	0	0	0	0	0	0.0116	
Neolithisch	0	0.0267	0	0	0	0	0.0186

Die untersuchten Individuen werden zusätzlich entsprechend ihrer geografischen Lage in eine östliche bzw. westliche Gruppe eingeteilt. Zum einen nur aufgrund der Lokalisation ihres Fundplatzes, zum anderen aufgrund ihrer Datierung in die Eisenzeit.

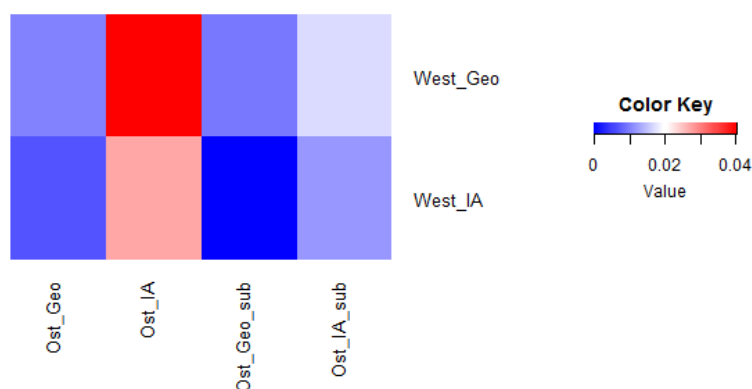


Abbildung 25: Parweise Fst-Werte skythischer Proben unterteilt in westliche und östliche Untergruppen entsprechend ihrer geografischen Lage oder Datierung in die Eisenzeit; **West_Geo**= Kolbino, Novozavedennoe, Pokrovka, Kamennyj; **Ost_Geo**= Zevakino, Berel, Ak-Alcha, Justyd, Aržan, Barsučij Log, Ul Ulžan žan, Solonzy; **West_IA**= Kolbino, Novozavedennoe, Pokrovka; **Ost_IA**= Zevakino, Berel, Ak-Alcha, Justyd, Aržan, Barsučij Log, Ulžan; **sub** = ohne Proben der Fundplätze Barsučij Log und Ulžan

In dieser Einteilung zeigen die Proben des osteurasischen Raums eine geringe genetische Distanz zu westeurasischen Gruppen jedoch mittlere Distanzen vergleicht man ausschließlich eisenzeitliche Individuen (Abb. 25). Werden in der rein geografischen Einteilungen die Fundplätze Barsučij Log und Ulžan ausgeschlossen, sinken die paarweisen Fst-Werte jeweils minimal. Im Vergleich der westlichen zu den eher östlichen eisenzeitlichen Gruppen, sinkt bei einem Ausschluss der beiden Fundplätze der Fst-Wert von 0,0271 auf 0,0119.

Der höchste Fst-Wert von 0,0403 zwischen östlichen, eisenzeitlichen Individuen und geografisch westlich liegenden Individuen ergibt sich dadurch, dass die zuletzt genannte Gruppierung auch die bronzezeitlichen Individuen umfasst.

Den niedrigsten π -Werte und somit die geringste **Nukleotiddiversität** zeigen bei einer zeitlichen Gruppierung der Proben die beiden neolithischen Proben mit 0,001521, die höchste Diversität liegt bei den Proben aus Kalinovka und Kamennyj-Ambar (Tab. 17). Wird eine Einteilung in die entsprechenden Kulturstufen vorgenommen, reichen die Diversitäts-Werte von 0,001643 für die Pazyryk-Kultur bis 0,007663 für Proben der Sintashta-Kultur. Für alle errechneten π -Werte liegt eine hohe Standardabweichung vor.

Tabelle 17 Nukleotiddiversität; sub = ohne Individuen aus Barsučij Log und Ulžan

	Probe	N	Ø π -Werte	SD
Zeit	Neolithisch	2	0.0015	0.0339
	Östl. Eisenzeit_sub	0	0.0017	0.0247
	Östl. Eisenzeit	2	0.0018	0.0253
		6		
	West. Eisenzeit	8	0.0018	0.0295
	Bronzezeit	5	0.0019	0.0380
Kultur	Pazyryk	1	0.0016	0.0252
		4		
	Aldy-Bel	4	0.0017	0.0293
	Sarmaten	6	0.0018	0.0306
	Čilikta	2	0.0024	0.0420
	Tagar	5	0.0031	0.0492
	Sintashta	4	0.0077	0.0853

Die folgende grafische Darstellung (Abb. 26) verdeutlicht den Anstieg der Nukleotiddiversität zwischen den Kulturgruppen der Eisen – und Bronzezeit.

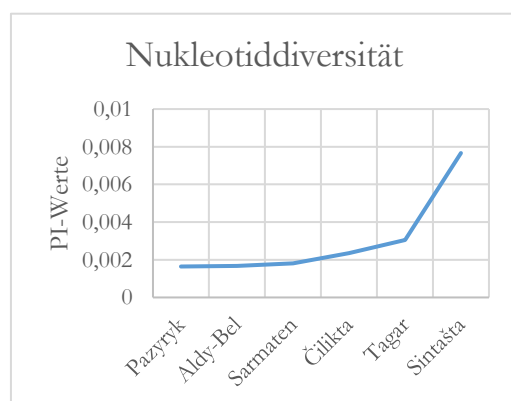


Abbildung 26: Grafische Darstellung der Mittelwerte der *site-pi* je Gruppierung in entsprechende Kulturstufen

Das Verhältnis eisenzeitlicher, skythischer Gruppierungen zu anderen prähistorischen Kulturgruppen

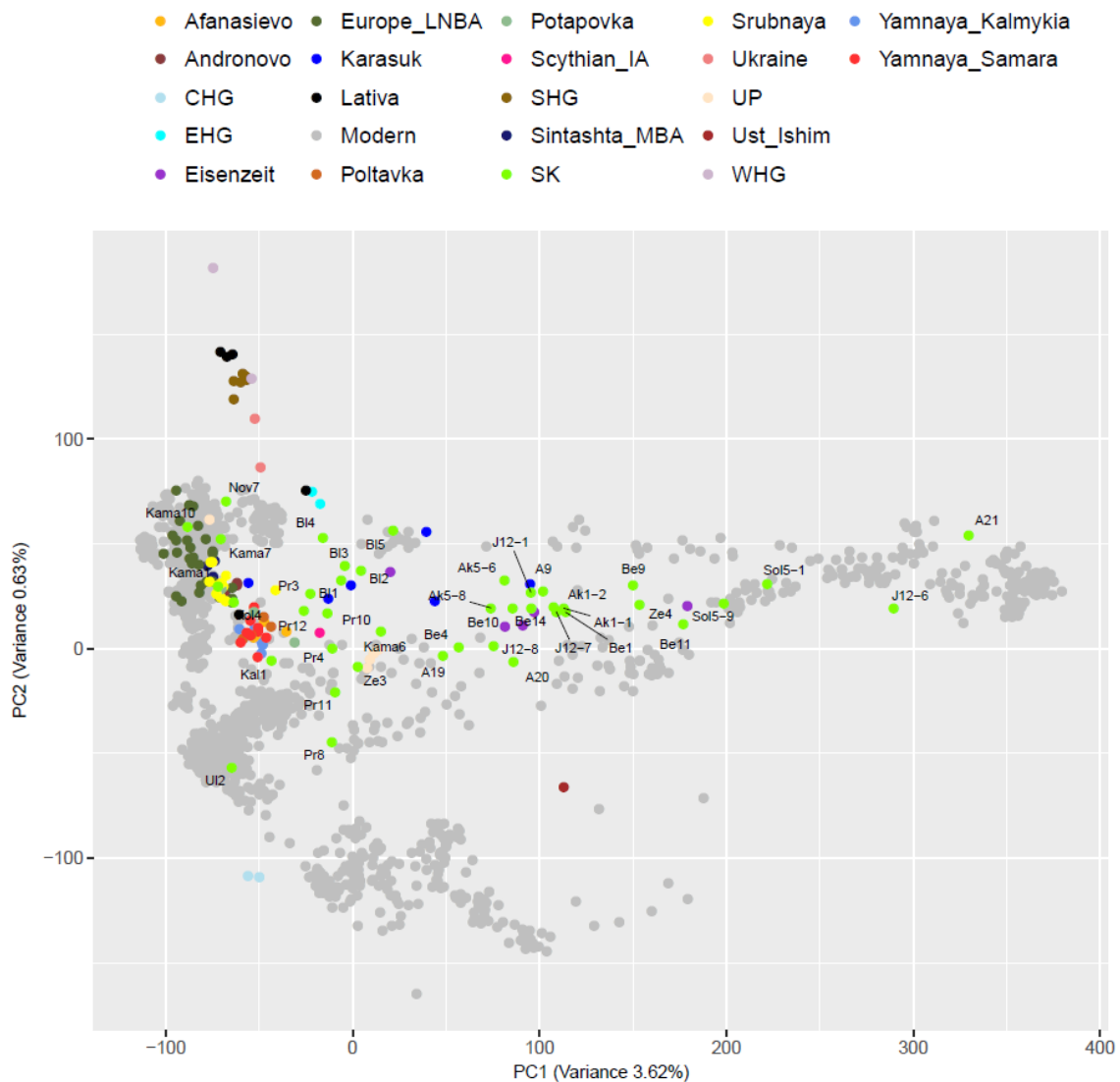


Abbildung 27: PCA der untersuchten Proben und zusätzlichen aDNA Referenzen (Lazaridis et al. 2014; Fu et al. 2014; Jones et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015; Jones et al. 2017; Sikora et al. 2017). SK = Individuen der vorliegenden Arbeit, inklusive bronzezeitlicher und neolithischer Individuen und Sarmaten; Aufteilung der bronzezeitlichen Steppenpopulationen der Referenzen in ihre Untergruppen

Die Bezeichnung und Klassifizierung der prähistorischen Referenzindividuen in der vorliegenden **PCA** (Abb. 27) wurden der Studie von (Lazaridis et al. 2016) entnommen und sind in der Tabelle B5 aufgelistet. Die Einteilung wird in der gesamten Arbeit beibehalten, um Vergleiche mit anderen Publikationen zu erleichtern. Abkürzungen für Zeitstufen richten sich nach den englischen Bezeichnungen Upper Paleolithic (UP), Iron Age (IA) und Bronze Age. (BA). Bei erstmaliger Nennung werden Abkürzungen für Jäger-Sammlerpopulationen der Referenzindividuen (CHG, EHG, SHG, WHG) ausgeschrieben, im wei-

teren Verlauf dann die gängigen Abkürzungen beibehalten. Für die PCA wird die Einteilung der Gruppen in ihre Unterkategorien wenn möglich entsprechend (Lazaridis et al. 2016) vorgenommen. Die für die vorliegende Arbeit analysierten Individuen sind einheitlich in hell-grün dargestellt. Die eisenzeitlichen Proben der Altai-Region aus Südsibirien (Verh-Uimon, Sary-Bel, Zhana-Aul) (Allentoft et al. 2015) verteilen sich über den zentralasiatischen Raum, zwischen Tadschikistan und Usbekistan, und liegen innerhalb der Verteilung der Proben aus Justyd, Berel und Aržan und Ak-Alacha. Individuen der Karasuk-Kultur (Allentoft et al. 2015) breiten sich in der vorliegenden PCA beginnend in Osteuropa bis in den zentral-asiatischen Raum aus und liegen zwischen frühen Sarmaten, Individuen der Tagar-Kultur sowie kasachischen und süd-sibirischen Fundplätzen. Die eisenzeitliche, skythische Probe aus dem südlichen Ural in Russland (Mathieson et al. 2015) fällt in die Nähe der Steppe-Individuen der frühen bis mittleren Bronzezeit und liegt im Verbreitungsgebiet der Sarmaten und Ze3. Die Individuen der Andronovo -, Sintashta - und Srubnaja-Kultur der späten bis mittleren Bronzezeit (Steppe_MLBA) liegen nord-westlicher und näher an modernen, europäischen Populationen. Die Probe aus Kalinovka (Kol4) und Kama1 liegen genau in diesem Cluster und überschneiden sich mit der Verteilung der europäischen Individuen des späten Neolithikums und Bronzezeit in der hier erstellten PCA (Allentoft et al. 2015; Mathieson et al. 2015). Mit Ausnahme von Kama6 liegen die Proben aus Kamennyj Ambar und Novozavedennoe ebenfalls in diesem Bereich zwischen modernen, osteuropäischen und europäischen Populationen.

Wurden die skythischen Proben für die *outgroup-f3*-Statistik jeweils einzeln im Test mit allen prähistorischen Populationen (Lazaridis et al. 2016) analysiert, ist die meiste genetische Drift zwischen den Skythen untereinander festzustellen. Eine Ausnahme bilden dabei die Probe Bl4 der Tagar-Stufe, zwei Sarmaten, der Sake und eine Probe aus Kalinovka. Hier finden sich nur sechs skythische Proben (Kol4, Kama6, Be1, A19, Be10 und Bl2) Für diese Proben konnten f3-Werte mit aDNA-Vergleichspopulationen innerhalb der höchsten 20 Werte berechnet werden. In der folgenden Tabelle (Tab. 18) sind diese Proben mit den entsprechenden prähistorischen Populationen dargestellt. Der Rang gibt die Position des f3-Werts innerhalb der höchsten 20 an.

In den *Ergänzenden Informationen_I3* sind alle Ergebnisse der *outgroup-f3*-Statistik mit dem Lazaridis-Datensatz als Test-Population für die höchsten 20 der f3-Werte aufgelistet. Die Standardfehler der f3-Werte sind auch hier sehr hoch und überwiegend nicht klar zu unterscheiden

Tabelle 18 Auflistung der *outgroup-f3*-Ergebnisse der skythischen Proben mit aDNA Referenzpopulationen (Lazaridis et al. 2016) mit prähistorischen Populationen unter den höchsten 20 Werten; AG = Außengruppe; f_3 = berechneter f_3 -Wert; std.err = Standardfehler, SNPs = Anzahl der für die Berechnung zugrundeliegenden SNPs, Rang = Position des dargestellten f_3 -Werts

Probe	Test	AG	f_3	std. err	SNPs	Rang
BI4	Andronovo	S_Mbuti	0.303	0.007	7920	20
BI4	WHG	S_Mbuti	0.303	0.008	593	19
BI4	Poltavka	S_Mbuti	0.303	0.007	7826	18
BI4	Steppe_IA	S_Mbuti	0.307	0.008	7436	17
BI4	Yamnaya-Samara	S_Mbuti	0.308	0.008	8021	16
BI4	Europe_LNBA	S_Mbuti	0.311	0.008	7558	15
BI4	SHG	S_Mbuti	0.313	0.007	8923	14
BI4	Loschbour	S_Mbuti	0.313	0.008	4776	13
BI4	Potapovka	S_Mbuti	0.313	0.010	7689	11
BI4	Srubnaja	S_Mbuti	0.317	0.008	5280	10
BI4	MA1	S_Mbuti	0.318	0.009	6765	9
BI4	Steppe_Eneolithic	S_Mbuti	0.322	0.009	6224	7
BI4	EHG	S_Mbuti	0.323	0.008	8335	6
BI4	Russia_EBA	S_Mbuti	0.324	0.015	2095	5
Kal1	Iberia_BA	S_Mbuti	0.303	0.015	2604	17
Kama10	Loschbour	S_Mbuti	0.285	0.007	41582	19
Kama10	Russia_EBA	S_Mbuti	0.285	0.014	19791	18
Kama1	Steppe_IA	S_Mbuti	0.285	0.009	31738	19
Kama1	Russia_EBA	S_Mbuti	0.288	0.017	17269	17
Kama1	Potapovka	S_Mbuti	0.292	0.011	27203	13
Kama6	EHG	S_Mbuti	0.313	0.013	23782	20
Kama6	Russia_EBA	S_Mbuti	0.321	0.026	12208	16
Kama7	Russia_EBA	S_Mbuti	0.330	0.043	7639	14
Pr11	Russia_EBA	S_Mbuti	0.319	0.022	1296	11
Pr3	Bichon	S_Mbuti	0.314	0.015	21322	15
UI2	Europe_LNBA	S_Mbuti	0.276	0.007	10994	18

Bei der *outgroup-f3*-Analyse der gemeinsamen genetischen Drift mit ausschließlich aDNA-Referenzpopulationen als Test sind mit wenigen Ausnahmen (J12-6 Barsučij Log, Berel, Ak-Alacha, Kamennyj Ambar und Solonzy), die f_3 -Werte aufgrund der hohen Standardfehler nicht eindeutig voneinander zu trennen. Die Auswertung der Ergebnisse erfolgt anhand der Einteilung der Fundplätze nach ihrer geografischen Lage und Zeitstufe. Die Individuen mit einer Sonderstellung in der PCA werden getrennt analysiert. Es sollen die 10 höchsten f_3 -Werte je Fundplatz näher betrachtet werden.

Die höchsten 20 *outgroup-f3*-Ergebnisse sind in den *Ergänzenden Informationen_I3* aufgelistet.

Genetische Drift zwischen östlichen Skythen der Eisenzeit und prähistorischen Kulturgruppen

Die Individuen der beiden Fundplätze Barsučij Log und Ulžan weichen in der PCA stark von ihrer geografischen Lokalisation ab. Die sibirischen Proben aus dem Minusinsker Becken teilen die höchsten f_3 -Werte mit östlichen Jäger-Sammlern, nordeuropäischen Jä-

ger-Sammlern (SHG) und Steppe-Bewohnern der frühen bis späten Bronzezeit (Abb. 28). Die ost-kasachische Probe weist sehr hohe Standardfehler der f_3 -Werte auf. Die meiste genetische Drift zeigt U12 mit Europe_LNBA, gefolgt von Russia_EBA, Vertretern der Jamnaja-Samara Kultur und nordeuropäischen Jäger-Sammlern (SHG). Erstmal ist für U12 eine chalkolitische Probe aus Anatolien mit einem hohen f_3 -Wert vertreten (Lazaridis et al. 2016).

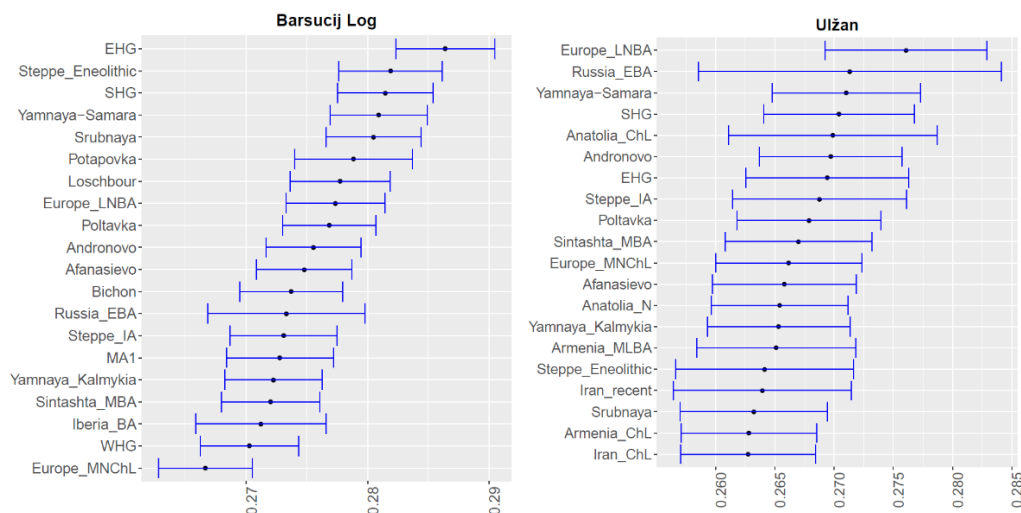


Abbildung 28: outgroup- f_3 -Statistik eisenzeitlicher, östlicher Proben im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Die süd-sibirische Probe J12-6 aus dem Chuya-Tal zeigt die meiste genetische Drift mit einem paläolithischen Individuum süd-zentral Sibiriens (MA1), der eisenzeitlichen, skythischen Probe, Individuen der Poltavka-Kultur, äneolithischen Steppebewohnern und östlichen Jäger-Sammlern (EHG) (Abb. 29). Für die verbleibenden Proben desselben Fundplatzes können die f_3 -Werte aufgrund hoher Standardabweichungen nicht eindeutig unterschieden werden. Es finden sich jedoch auch hier MA1, östliche Jäger-Sammler (EHG) und Steppe-Eneolithic unter den höchsten f_3 -Werten.

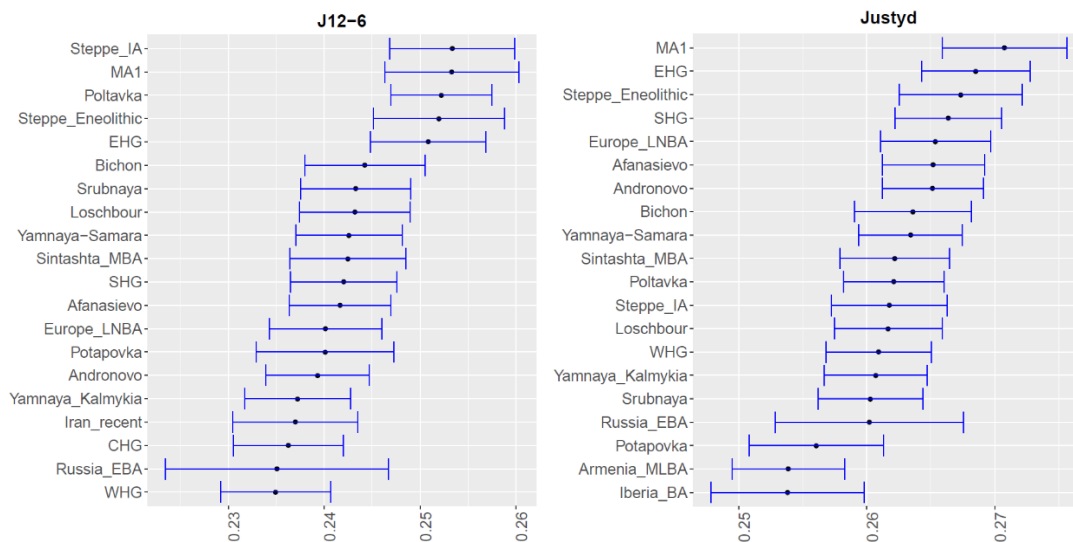


Abbildung 29: *outgroup-f3*-Statistik der Proben aus Justyd im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Die Probe A21 weist im Vergleich zu den gruppierten Individuen des Fundplatzes Aržan die meiste genetische Drift mit MA1 und östlichen Jäger-Sammlern auf. Die weiteren Proben des Fundplatzes Aržan hingegen mit äneolithischen Steppe-Individuen und MA1 (Abb. 30). Die weiteren f_3 -Werte liegen größtenteils im Bereich der Standardfehler der nächstfolgenden Werte.

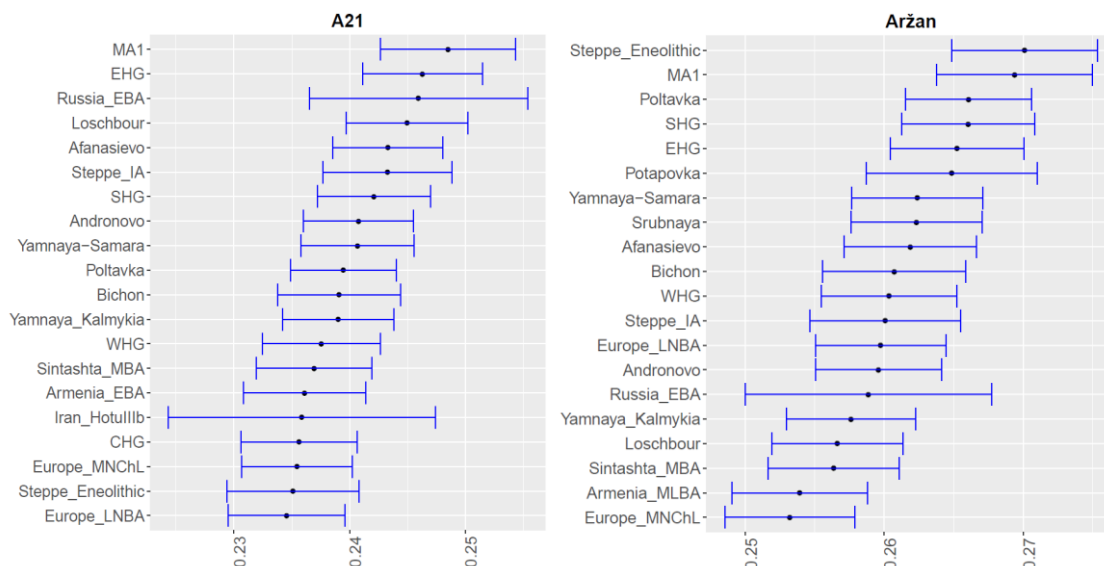


Abbildung 30: *outgroup-f3*-Statistik der Proben aus Aržan im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Die Proben des kasachischen Altai-Gebirges und dem süd-sibirischen Ukok-Plateau zeigen die meiste genetische Drift mit östlichen Jäger-Sammlern aus Karelia und Individuen der Jamnaja-Samara-Kultur des Äneolithikums (Abb. 31).

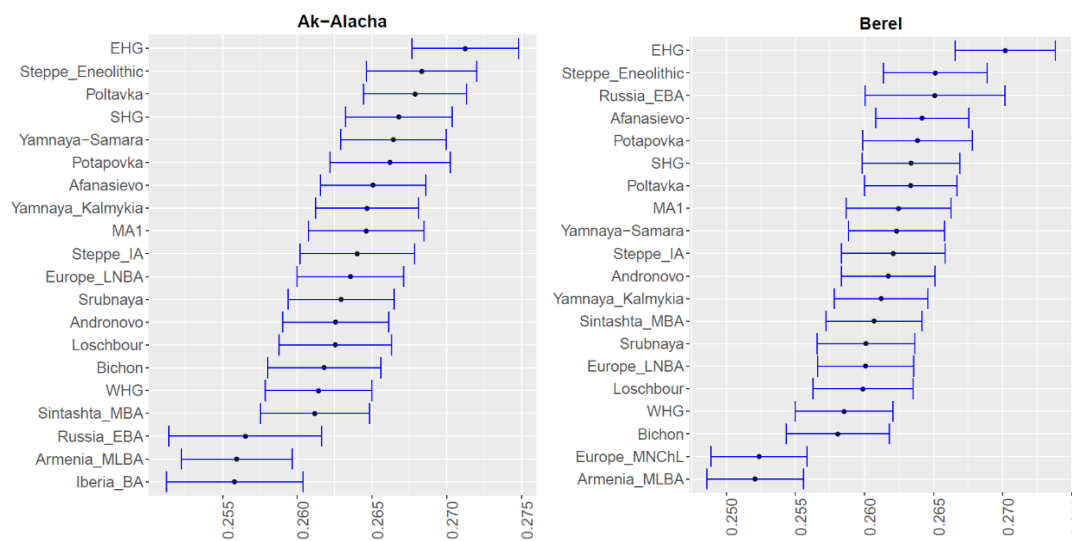


Abbildung 31: *outgroup-f3*-Statistik der Proben aus Aržan im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Die f_3 -Werte der beiden Proben Ze3 und Ze4 des ostkasachischen Fundplatzes weisen hohe Standardfehler in ihren f_3 -Werten auf (Tab.19). Beide Proben teilen die meiste genetische Drift mit Russia-EBA und östlichen Jägern-Sammlern. Ze3 weist hohe f_3 -Werte für SHG und europäischen Individuen des Neolithikums und der Bronzezeit auf, Ze4 hingegen für MA1 und Kulturgruppen der Steppe in der frühen Bronzezeit.

Tabelle 19 *outgroup-f3*-Statistik der Proben aus Zevakino im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind jeweils die höchsten 5 Werte; AG = Außen-gruppe; f_3 = berechneter f_3 -Wert; std.err = Standardfehler, Z = Z-score; SNPs = Anzahl der für die Berechnung zugrundeliegenden SNPs,

Probe	Test	AG	f_3	std. err	Z	SNPs
Ze4	Potapovka	S_Mbuti	0.253	0.009	26.851	5146
Ze4	MA1	S_Mbuti	0.255	0.009	29.916	6953
Ze4	Afanasiovo	S_Mbuti	0.256	0.007	39.154	8384
Ze4	Russia_EBA	S_Mbuti	0.257	0.015	17.688	2308
Ze4	EHG	S_Mbuti	0.260	0.007	37.518	8491
Ze3	Europe_LNBA	S_Mbuti	0.281	0.009	32.106	5909
Ze3	Steppe_Eneolithic	S_Mbuti	0.284	0.010	28.583	4706
Ze3	EHG	S_Mbuti	0.285	0.008	34.579	6091
Ze3	SHG	S_Mbuti	0.290	0.008	35.124	5557
Ze3	Russia_EBA	S_Mbuti	0.292	0.017	16.998	1700

Genetische Drift zwischen westlichen Skythen der Eisenzeit und prähistorischen Kulturgruppen

Die Gruppe der östlichen und westlichen Jäger-Sammler (EHG, WHG) sowie Individuen der Andronovo-Kultur sind bei den westlichen Proben der Eisenzeit, Kol4, Nov7 und Pokrovka unter den höchsten 10 f3-Werten zu finden (Abb. 32). Die frühen Sarmaten und die Skythische Probe aus SW-Russland besitzen die meiste genetische Drift mit SHG, Russia_EBA und Steppe-Eneolithic. Eine Nähe zu Individuen der Poltavka, Jamnaja und Sintashta-Kultur ist anhand der höchsten f3-Werte für Nov7 und Proben aus Pokrovka festzustellen.

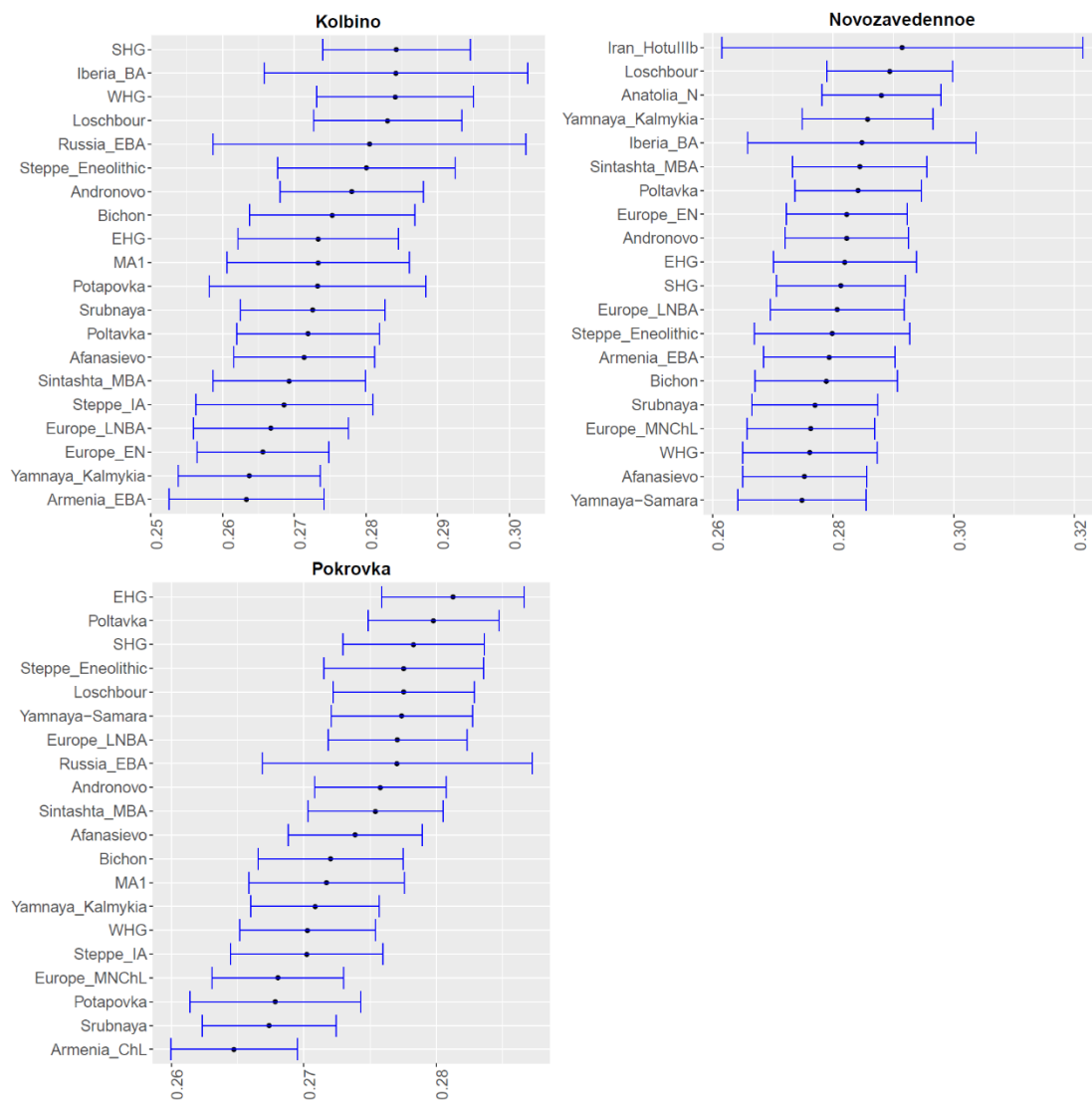


Abbildung 32: outgroup-f3-Statistik eisenzeitlicher Proben SW-Russlands im Vergleich zu prähistorischen Referenzpopulationen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Genetische Drift zwischen bronzezeitlichen Proben und prähistorischen Kulturgruppen

Die bronzezeitlichen Proben der Sintashta-Kultur zeigen die meiste genetische Drift mit einem früh-bronzezeitlichen Steppe-Bewohner aus dem heutigen Stalingrad in Russland (Allentoft et al. 2015). Die folgenden f3-Werte fallen in den jeweiligen Standardfehler voneinander (Abb. 33). Für die bronzezeitliche Probe aus Kalinovka ist aufgrund der hohen Standardfehler der f3-Werte keine aussagekräftige Beurteilung der f3-Ergebnisse möglich. Unter den höchsten 10 f3-Werten liegen Iberia-BA (WHG), Steppe-Individuen der frühen – bis späten Bronzezeit, östliche Jäger-Sammler (EHG), die skythische, eisenzeitliche Probe (Steppe_IA), MA1, SHG und äneolithische Steppenbewohner der Samara-Kultur (Mathieson et al. 2015).

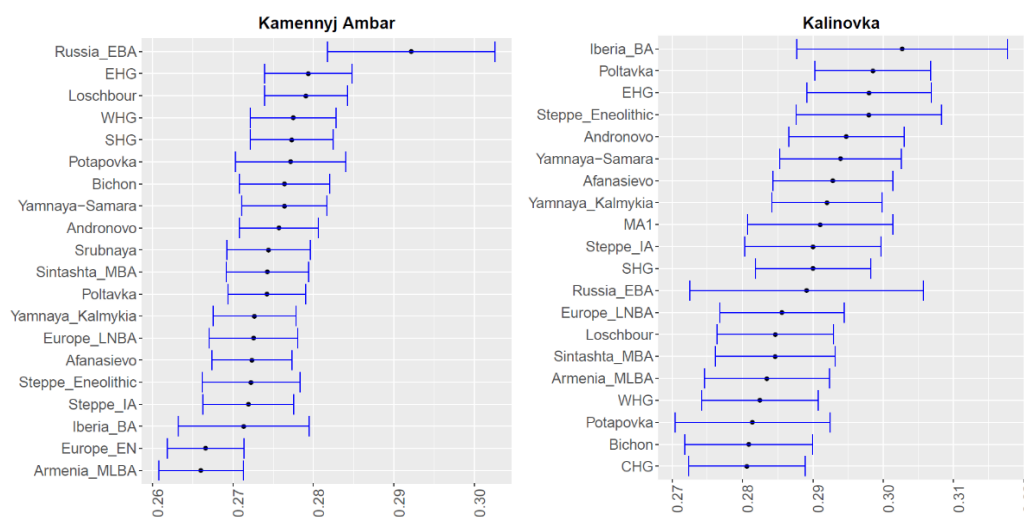


Abbildung 33: outgroup-f3-Statistik bronzezeitlicher Proben SW-Russlands im Vergleich zu prähistorischen Referenzpopulationen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Genetische Drift zwischen neolithischen Proben und prähistorischen Kulturgruppen

Die beiden neolithischen Proben Sol5-9 und Sol5-1 weisen die höchste genetische Drift mit MA-1 aus Mal'ta in Südsibirien aus dem Jungpaläolithikum auf (Raghavan et al. 2013) und mit östlichen Jäger-Sammlern auf, gefolgt von eisenzeitlichen und äneolithischen Steppeindividuen (Abb. 34).

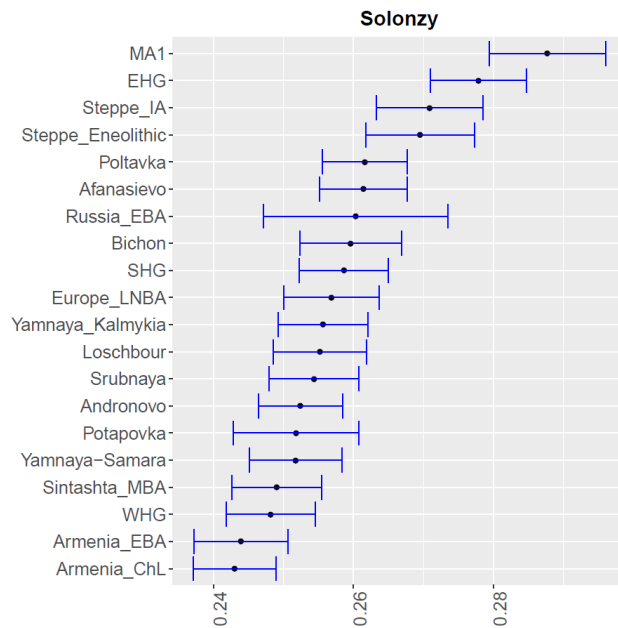


Abbildung 34: *outgroup-f3*-Statistik neolithischer Proben SW-Russlands im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler

Um die Position der Fundplätze Ulžan, Barsučij Log und Zevakino in der PCA, sowie die Sonderstellung der Proben J12-6 und A21 zu überprüfen, wurde eine *outgroup-f3*-Statistik wie oben durchgeführt. Es wurden jedoch die Proben J12-6 und A21 und zusätzlich die verbleibenden Proben des entsprechenden Fundplatzes als Referenz in die Analyse mit einbezogen. Die meiste genetische Drift teilen die untersuchten Individuen oder Fundplätze entsprechend ihrer Lage in der vorliegenden PCA (Abb. 13). Die Individuen der Tagar-Kultur zeigen zusätzlich eine Nähe zu den neolithischen Proben aus Solonzy und östlichen Jägern-Sammlern. In der folgenden Tabelle (Tab. 20) sind die höchsten 5 *f3*-Werte für die drei eisenzeitlichen Fundplätze, sowie die Proben Ze3, Ze4, A21 und J12-6 aufgelistet.

Tabelle 20 Auflistung der *outgroup-f3*-Ergebnisse der skythischen Proben mit aDNA Referenzpopulationen ((Lazaridis et al. 2016); die fünf höchsten f3-Werten sind aufgelistet; AG = Außen-gruppe; f_3 = berechneter f3-Wert; std.err = Standardfehler, Z = Z-score; SNPs = Anzahl der für die Berechnung zugrundeliegenden SNPs,

Probe	Test	AG	f_3	std. err	Z	SNPs
A21	Ak-Alacha	S_Mbuti	0.284	-0.007	43.356	12135
A21	Berel	S_Mbuti	0.287	0.007	44.028	12355
A21	Ze4	S_Mbuti	0.302	0.014	21.363	2613
A21	Solonzy	S_Mbuti	0.321	0.013	25.507	2838
A21	J12-6	S_Mbuti	0.324	0.012	26.195	3553
Barsučij	EHG	S_Mbuti	0.286	0.004	70.484	57872
Barsučij	Ze3	S_Mbuti	0.287	0.012	23.809	2515
Barsučij	Solonzy	S_Mbuti	0.288	0.010	29.568	3903
Barsučij	Nov7	S_Mbuti	0.288	0.017	17.432	1363
Barsučij	Kalinovka	S_Mbuti	0.296	0.012	24.618	2953
J12-6	Ak-Alacha	S_Mbuti	0.290	0.008	37.049	8166
J12-6	Ze3	S_Mbuti	0.290	0.018	16.483	1588
J12-6	Ze4	S_Mbuti	0.300	0.016	18.764	2118
J12-6	A21	S_Mbuti	0.324	0.012	26.195	3553
J12-6	Solonzy	S_Mbuti	0.338	0.015	23.009	2354
Ulžan	Europe_LNBA	S_Mbuti	0.276	0.007	40.466	10994
Ulžan	Pokrovka	S_Mbuti	0.278	0.011	25.708	3219
Ulžan	Ze3	S_Mbuti	0.285	0.017	16.519	1605
Ulžan	Nov	S_Mbuti	0.286	0.022	12.754	937
Ulžan	Kalinovka	S_Mbuti	0.301	0.016	19.036	1958
Ze3	Kamennyj	S_Mbuti	0.301	0.018	16.671	1416
Ze3	Kolbino	S_Mbuti	0.310	0.027	11.393	704
Ze3	Pokrovka	S_Mbuti	0.313	0.014	22.103	1871
Ze3	Nov	S_Mbuti	0.316	0.030	10.672	649
Ze3	Kalinovka	S_Mbuti	0.317	0.020	16.133	1266
Ze4	Justyd	S_Mbuti	0.296	0.012	25.409	3489
Ze4	Berel	S_Mbuti	0.296	0.010	30.375	5263
Ze4	J12-6	S_Mbuti	0.300	0.016	18.764	2118
Ze4	A21	S_Mbuti	0.302	0.014	21.363	2613
Ze4	Solonzy	S_Mbuti	0.310	0.017	18.547	1889

Populationsstrukturen innerhalb der Kulturgruppen der Steppenreiter und Vergleich zwischen prähistorischen Referenzdaten der Bronze – und Eisenzeit

Die prähistorischen Vergleichspopulationen (Lazaridis et al. 2016) welche in der f3-Statistik die meiste genetische Drift mit skythischen Proben teilen, wurden für die **D-Statistik** eingesetzt: CHG, EHG, SHG; WHG, Iberia_BA, MA1, Russia_EBA, Steppe_IA, Steppe_Eneolithic, Poltavka, Europe_LNBA, Jamnaja-Samara, Anatolia_ChL, Potapovka, Srubnaja und Afanasievo. Die Reiternomaden der vorliegenden Arbeit wurden hierfür in ihre entsprechende Kulturstufe gruppiert. Es wurde jede mögliche Kombination mit den 16 prähistorischen Kulturgruppen aus dem Lazardis-Datensatz getestet. Insgesamt konnten so 7510 Bäume generiert werden, davon erzielten 852 ein signifikantes Ergebnis. Alle Er-

gebnisse der D-Statistik mit einem signifikanten Z-Wert ($|Z| > 3$ oder $|Z| < 3$) sind in den *Ergänzenden Informationen_I4* aufgelistet.

Die D-Statistik wie in Kapitel 3.3 beschrieben kann mit der Form $D(\text{Skythen-A, Skythen-B, Skythen-C, AG})$ eingesetzt werden, um auf Populationsstrukturen innerhalb der in der vorliegenden Arbeit analysierten Proben zu testen. Hierfür wurde Mbuti wie bereits für die *outgroup-f₃*-Statistik erneut als Außengruppe (AG) verwendet und für Skythen A-C jeweils jede mögliche Kombination der hier vorgestellten Kulturgruppen eingesetzt. Für 11 Kombinationsmöglichkeiten konnten signifikante D-Statistiken erzielt werden (Tab. 21). Wobei ein negativer D-Wert eher eine gemeinsame genetische Abstammung von Population B zu C unterstützt als zwischen Population A und C und somit der Baum der Form $D(((A, C) B) AG)$ widerlegt werden kann.

Tabelle 21 Signifikante D-Statistiken ($|Z| > 3$ und $|Z| < 3$) der Form $D(A, B, C, AG)$

A	B	C	AG	D-Wert	SD	Z-Wert	SNPs
Sake	Pazyryk	Aldy-Bel	S_Mbuti	-0.00597	0.00192	-3.114	12702
Sake	Pazyryk	Solonzy	S_Mbuti	-0.00921	0.00258	-3.577	6576
Sintaschta	Pazyryk	Aldy-Bel	S_Mbuti	-0.00786	0.00180	-4.355	14302
Sintaschta	Aldy-Bel	Pazyryk	S_Mbuti	-0.00625	0.00169	-3.699	14302
Sintaschta	Čilikta	Solonzy	S_Mbuti	-0.01169	0.00372	-3.145	3567
Tagar	Solonzy	Aldy-Bel	S_Mbuti	-0.00868	0.00242	-3.585	7911
Pazyryk	Sarmaten	Sintaschta	S_Mbuti	-0.00711	0.00219	-3.241	9901
Solonzy	Tagar	Aldy-Bel	S_Mbuti	0.00868	0.00242	3.585	7911
Pazyryk	Sintaschta	Aldy-Bel	S_Mbuti	0.00786	0.00180	4.355	14302
Pazyryk	Sake	Aldy-Bel	S_Mbuti	0.00597	0.00192	3.114	12702
Aldy-Bel	Sintaschta	Pazyryk	S_Mbuti	0.00625	0.00169	3.699	14302

Basierend auf der D-Statistik teilen die Individuen der Pazyryk-Kultur eher eine gemeinsame Abstammung mit Individuen der Aldy-Bel-Kultur als mit der sakischen Proben (Ul2) oder den bronzezeitlichen Individuen der Sintaschta-Kultur. Die beiden neolithischen Proben aus Solonzy weisen eine höhere genetische Drift zur Pazyryk- und Aldy-Bel-Kultur und Individuen der Čilikta -Stufe auf als zu dem Sake aus Ulžan, den Proben der Sintaschta und Tagar-Kultur.

Bei einer D-Statistik der Form $D(\text{Skythen-A, Test, Skythen-C, AG})$ unterstützt ein positiver D-Wert eine gemeinsame Abstammung und eine höhere genetische Drift zwischen den beiden skythischen Populationen. Ein negativer D-Wert weist auf eine Nähe zwischen der

Test-Population zur skythischen Population C hin und bestätigt somit einen Baum der Form $((B, C) A) AG$). In der folgenden Tabelle (Tab. 22) sind alle Kombinationen mit einem signifikanten, negativen Wert aufgelistet.

Tabelle 22 Signifikante D-Statistiken ($|Z| < 3$) der Form D (A, Test, C, AG)

A	B	C	AG	D-Wert	SD	Z-Wert	SNPs
Pazyryk	EHG	Sintaschta	S_Mbuti	-0.00415	0.00117	-3.551	42697
Pazyryk	SHG	Sintaschta	S_Mbuti	-0.00395	0.00102	-3.883	42773
Pazyryk	WHG	Sintaschta	S_Mbuti	-0.00364	0.00116	-3.148	44080
Pazyryk	Poltavka	Sintaschta	S_Mbuti	-0.00325	0.00099	-3.302	44670
Pazyryk	Yamnaya-Samara	Sintaschta	S_Mbuti	-0.00373	0.00109	-3.434	41170
Aldy-Bel	SHG	Sintaschta	S_Mbuti	-0.00581	0.00175	-3.314	15509
Pazyryk	SHG	Skythen	S_Mbuti	-0.00682	0.00219	-3.120	8412
Pazyryk	EHG	Tagar	S_Mbuti	-0.00379	0.00074	-5.111	127214
Pazyryk	SHG	Tagar	S_Mbuti	-0.00228	0.00068	-3.357	127300
Pazyryk	Yamnaya-Samara	Tagar	S_Mbuti	-0.00254	0.00069	-3.680	122773

Wird ein Baum der Form D(Pazyryk, Test, Sintaschta, AG) getestet, teilen die bronzezeitlichen Individuen eine gemeinsame Abstammung mit allen Jäger-Sammler-Populationen, Individuen der Poltavka- und Jamnaja-Kultur. Die Proben der Tagar-Gruppe aus Barsučij Log weisen bei einer D-Statistik der Form (Pazyryk, Test, Tagar, AG) eine geteilte, genetische Drift mit EHG und SHG, sowie Vertretern der Jamnaja-Samara-Gruppe auf. Die Probe aus Kolbino zeigt eher eine Nähe zu SHG als zu Individuen der Pazyryk-Kultur.

Bäume der Form D(Test-A, Test-B, Skythen-C, AG) mit relevanten prähistorischen Kulturgruppen des Lazaridis-Datensatzes als Test-A bzw. Test-B ermöglichen eine gemeinsame Abstammung der jeweiligen skythischen Kulturgruppe zu einer der beiden Referenzpopulationen zu testen. Im Zuge der Hypothesendiskussion im Kapitel 5.3 werden relevante Ergebnisse dieser D-Statistik aufgeführt und im jeweiligen populationshistorischen Kontext diskutiert. Allerdings können nur Bäume mit einem signifikanten D-Wert ausgewertet werden, was bedeutet, dass nicht alle für die Diskussion sinnvollen Kombinationen mit einbezogen werden können.

5. Diskussion

5.1 Effizienz der selektiven Anreicherung nukleärer DNA und Diskussion der finalen Probenauswahl

Für ein *in-solution* Anreicherungsverfahren sind zum Zeitpunkt der durchgeführten Laborversuche (Stand 2014) Kits von verschiedenen Herstellern erhältlich. In den angebotenen Paketen ist jeweils eine Erstellung bzw. Produktion der *bait-library* und benötigten Chemikalien für die Anreicherung enthalten. In der vorliegenden Arbeit wurden zunächst MYbaits® von MYcroarray, NimbleGen von Roche und SureSelect^{XT} von Agilent exemplarisch anhand einer Probe getestet, deren nukleärer DNA-Gehalt bereits durch Vorversuche als hoch eingestuft werden konnte. Basierend auf den durchgeführten Analysen wurde das **Anreicherungskit MYbait** von MYcroarray ausgewählt. Die Spezifität des *captures*, also der prozentuale Anteil an Sequenzen, der gegen die *capture*-Regionen im Vergleich zum gesamten Genom (hg19) aligniert wird, liegt bei 72,8 % und ist bei den Kits der anderen beiden Hersteller mit 76 % und 91 % höher. Die minimale, maximale und durchschnittliche Abdeckungstiefe ist jedoch bei der Anwendung des MYbait-Kits am höchsten (46.8 x/142.6 x/108.5 x). Außerdem ist die flexible Handhabung im *bait*-Design ein klarer Vorteil, da hier *bait*-Länge und Art der Nukleotidsequenz (RNA/DNA) frei gewählt werden können. Auch die mögliche *tiling*-Frequenz ist benutzerdefiniert. Diese hat einen signifikanten Einfluss auf die Abdeckungstiefe und den *capture*-Erfolg (Tewhey et al. 2009)

Alle eingesetzten Chemikalien sind in ihrer jeweiligen Konzentration bekannt und separat lieferbar. So besteht die Möglichkeit Modifikationen am Protokoll vorzunehmen und gegebenenfalls an die Qualität der Proben anzupassen (Hofmanová et al. 2016; Kreutzer 2017)

Das für die vorliegenden Proben verwendete Protokoll wurde im Laufe der Bearbeitung an die neue Version des MYbait-Handbuches adaptiert oder Verbesserungen aus Publikationen entnommen. So wurde die Zyklen-Zahl der beiden Amplifikationen nach den *capture*-Runden reduziert und die quantifizierbare DNA-Konzentration am Ende einer jeden *capture*-Runde durch eine doppelte Elution der isolierten DNA von den *bait-bead*-Komplexen deutlich erhöht. Als magnetische Beads werden die Dynabeads® MyOne™ Streptavidin T1 (Thermo Fisher Scientific) eingesetzt. In aktuellen Handbüchern (ab v2.3) ist unterdessen eine *on-bead*-PCR standardisiert. Hierbei werden die *bait-bead*-Komplexe nicht aufgetrennt, sondern eine Amplifikation an den zunächst gebundenen Sequenzen durchgeführt, wodurch eine Denaturierung mittels NaOH entfällt.

Die Auswahl der Proben für eine anschließende Anreicherung nukleärer Regionen basiert auf den Ergebnissen der **Qualitätsbestimmung** der erstellten *libraries*. Insgesamt wurde Knochenmaterial von 77 Proben aus Zentralasien und SW-Russland extrahiert und in DNA-Bibliotheken überführt. Die anschließende Qualitätskontrolle, das sogenannte *screening*, umfasst eine MiSeq-Sequenzierung auf die erstellten *libraries*, um den endogenen DNA-Gehalt eines DNA-Extraktes zu ermitteln. Zusätzlich wird eine quantitative *real-time*-PCR auf die *fill-in*-Produkte der *library*-Erstellung vor der ersten Amplifikation durchgeführt, um eine absolute Molekülzahl des vorliegenden DNA-Extrakts zu bestimmen. Quantitative Messungen mittels Qubit und Agilent Bioanalyzer auf die amplifizierten *libraries* geben Hinweise auf die Konzentration und mögliche Dimer-Bildung durch Primer- und Adapterstrukturen. 41 Proben wurden aufgrund einer hohen Molekülzahl und hohem endogenen DNA-Gehalt ausgewählt. Die durchschnittliche Molekülzahl liegt bei 10^7 Adapterligierten Molekülen je mg Knochenpulver. Der endogene DNA-Gehalt der *screening-libraries* beträgt durchschnittlich 23,7 % (3,36-71,11%). Bei 31 Proben liegt der endogene DNA-Gehalt der *screening-libraries* bei $> 10\%$. Für die Proben Kal1, Kama10, Kol4, Nov7, Sol5-1, Pr3, Pr8, Pr11 und Pr12 liegt der DNA-Gehalt unter dem festgelegten Schwellenwert, in den *fill-in*-Produkten konnte jedoch eine Molekülzahl von $> 10^7$ bestimmt werden. Die Proben wurden aufgrund ihrer geografischen Lage und Datierung in die Bronzezeit trotzdem für ein weiteres Anreicherungsverfahren ausgewählt. Es wurden daher mindestens 5 und maximal 8 verschiedene *libraries* erstellt und für das anschließende *capture* gepoolt, um eine möglichst hohe molekulare Komplexität der Ausgangs-*library* für die Anreicherung zu erzielen.

Der molekulare Erhaltungszustand der Proben kann generell für prähistorisches Material als sehr gut bezeichnet werden.

Alle Proben, die aufgrund ihrer geografischen Lage der östlichen Gruppe zugeordnet werden konnten, wurden größtenteils aus Permafrostböden geborgen. In den durch Eis konservierten Gräbern des Pazyryk-Valley konnten Proben und Grabbeigaben in außergewöhnlich gutem Erhaltungszustand geborgen werden (Van Noten and Polosmak 1995). Die konstant niedrigen Temperaturen ermöglichen einen sehr guten DNA-Erhalt über einen längeren Zeitraum, da Schädigungsprozesse der DNA-Moleküle reduziert werden. Zusätzlich erfolgte eine Vorauswahl von 32 Proben aufgrund positiver Ergebnisse mehrerer Multiplex-PCRs auf nukleäre Marker (durchgeführt von Martina Unterländer).

Für 31 Proben konnte eine **Abdeckung der Zielregionen** von über 5 x erreicht werden. 19 Proben zeigen mit einer durchschnittlichen *coverage* von 5 - 20 x eine gute Abde-

ckungstiefe. Acht Proben liegen mit einer Abdeckung zwischen 21,13 x und 54,7 x in einem mittleren Bereich. Zwei Proben aus Ak-Alacha (Ak5-8, Ak1-1) und zwei kasachische Proben (Be10, Ul2) erreichen eine außergewöhnlich hohe *coverage* von > 80 x. Für die Proben Be10 und Ul2 konnte das DNA-Material aus Felsenbeinen gewonnen werden, die bereits in Studien als Skelettmaterial mit überdurchschnittlich gutem Erhaltungszustand und hoher Molekülzahl beschrieben wurden (Gamba et al. 2014; Pinhasi et al. 2015; Kreutzer 2017), was sich in der vorliegende Arbeit bestätigt und das Ergebnis der Sequenzierung erklärt. Für die Proben Ak5-8 und Ak1-1 wurden nur 4 libraries für die Anreicherung gepoolt und 1337 bzw. 1439 ng eingesetzt. Beide zeigen jedoch mit 201318 bzw. 263198 Sequenzen eine sehr hohe *read*-Anzahl ohne Duplikate, welche gegen die *capture*-Region aligniert werden kann. Vergleichbar mit 245555 Sequenzen der Felsenbein-Probe Be10. Während für Ak1-1 eine hohe Molekülzahl pro mg Knochenpulver von 6,21E+08 bestimmt werden konnte, weist die Probe Ak5-8 einen sehr hohen endogenen DNA-Gehalt von 71.11 % auf, was die vorliegenden Ergebnisse erklären kann.

Insgesamt variiert die Abdeckung aller *capture*-Produkte untereinander stark und liegt zwischen 1,92 x und 124,95 x, durchschnittlich bei 22,82 x. Die Spezifität des *captures* kann über die durchschnittliche Abdeckung der Zielregionen festgestellt werden und ist mit 65,10 % - 94,64 % sehr gut. Die Werte im unteren Bereich erklären sich dadurch, dass auch Proben bereits nach der ersten *capture*-Runde sequenziert und mit Ergebnissen aus der zweiten Runde zusammengeführt wurden (Ak1-1, Be4, Be9, Be14). Da das *capture* nach nur einem Durchlauf noch sehr unspezifisch ausfällt, kann noch keine maximale Spezifität erzielt werden. In diesen Fällen ist die Abdeckung des gesamten Genoms in % ebenfalls höher und liegt zwischen 7,67 % und 18,06%. Bei *capture*-Produkten, die ausschließlich nach der zweiten Runde sequenziert wurden, ist das gesamte Genom zu 2,53 % abgedeckt.

Die Abdeckung der gesamten Region in % korreliert nicht mit der durchschnittlichen *coverage* einer Probe. Es besteht auch kein direkter Zusammenhang zwischen der Effizienz des *captures* (alignierte Sequenzen in den *capture*-Regionen im Verhältnis zu allen alignierten *reads*) und der Abdeckung der Regionen. Schließt man Proben mit einer Abdeckung < 5 x für die Berechnung des Mittelwerts aus, erhöht sich der Wert nur minimal von 71,71 % auf 72,28 %

Während die vollständige Abdeckung der Zielregionen nicht mit dem endogenen DNA-Gehalt einer Probe korreliert, zeigen die Proben mit der höchsten *coverage* jedoch auch den höchsten Wert an Molekülen in den getesteten *fill-in*-Produkten der *screening-libraries*. Nachfolgende Abbildung (Abb. 35) zeigt deutlich, dass für die 5 Proben Ak5-8, Ak1-1, Be9,

Be10 und UI2 mit einer Abdeckung $> 50 \times$ ebenfalls die höchsten Molekülnzahlen per mg Knochenpulver der entsprechenden Extraktionen quantifiziert werden konnten. Eine vergleichbare Molekülnzahl wie bei Probe Ak5-8 konnte auch für drei weitere Proben gemessen werden. Die Tiefe der Abdeckung der Probe aus Ak-Alacha ist wahrscheinlich auf die Kombination der hohen Komplexität der DNA-*library* in Verbindung mit dem außergewöhnlich hohen endogenen DNA-Gehalt von 71,11 % zurückzuführen.

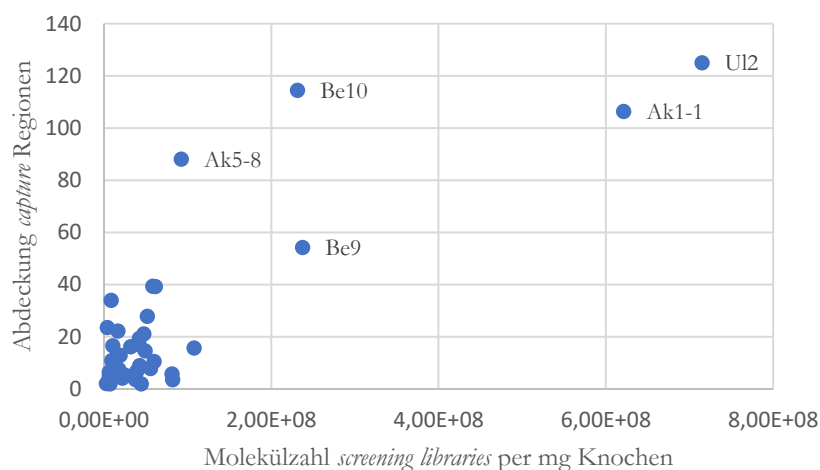


Abbildung 35 Abdeckung der Zielregionen im Verhältnis zur quantifizierten Molekülnzahl der *screening-libraries* einer Probe

Die laborseitig durchgeführten **Kontaminationskontrollen** zeigten keine quantifizierbaren Ergebnisse in der Qubit- oder Bioanalyser-Messung. Die detektierte Molekülnzahl der *library*-Leerkontrollen nach dem *fill-in* Schritt beträgt zwischen $1.75E+05$ - $8.69E+05$ Molekülen. Die über eine MiSeq-Sequenzierung analysierten *library*-Leerkontrollen zeigten zwischen 22-281 alignierten Sequenzen. Die detektierten Resultate der Leerkontrollen von qPCR und MiSeq-Sequenzierung wurde als unauffällig eingestuft. Die Bestimmung der Deaminierungsrate am 5'-Ende der aller Sequenzen der *capture*-Produkte zeigt eine durchschnittliche Rate von 15,13 %. Lediglich für die Probe Ak5-8 konnte eine niedrige Rate von $mD5=2,75 \%$ festgestellt werden. Diese Probe weist ebenfalls, wie bereits im oberen Abschnitt erwähnt, einen insgesamt sehr guten Erhaltungszustand auf. Basierend auf den Kontaminationsbestimmungen mittels *contaMIX* (Fu et al. 2013) und *ANGSD* (Korneliusen et al. 2014) ist die Probe als authentisch einzuschätzen.

Eine Kontaminationsbestimmung anhand der mitochondrialen DNA (contaMIX) zeigt für drei Proben Werte unter dem festgelegten Authentizitätswert von 91 %. Die Proben J12_1, A21 und B13 können mit einem Authentizitätswert von 85,6%, 70,06 % bzw. 75,42 % nicht als authentisch eingestuft werden, zeigen jedoch bei der Kontaminationsbestimmung basierend auf dem X-Chromosom keine Auffälligkeiten. Für die Probe A21 wurde zudem eine niedrige Deaminierungsrate von 5,98 % festgestellt. Eine X-Chromosom-basierte Kontaminationsbestimmung (ANGSD) zeigt erhöhte Werte für Kol4, Kal1, Sol5-1 und B12. Alle vier Proben sind mit einem Authentizitätswert von über 94 % durch die zweite Methode der Kontaminationsbestimmung und einer Deaminierungsrate mit über 10 bzw. 20 % als authentisch zu bewerten.

Alle Proben werden für die anschließende Bearbeitung beibehalten. Die Interpretation der Ergebnisse sollte jedoch unter Berücksichtigung einer möglichen Kontamination diskutiert werden.

Somit konnten 5 Individuen der Bronzezeit (Kama1, Kama6, Kama7, Kama10, Kal1), zwei neolithische Proben (Sol5-1, Sol5-9) und 34 eisenzeitliche Proben für populationsgenetische Analysen gewonnen werden. Die eisenzeitlichen Fundplätze in SW-Russland mit der Skythin aus Kolbino, einer Vertreterin der Initialen Skythen aus Novozavedennoe und die frühen Sarmaten aus Pokrovka können in eine westliche Gruppe eingeordnet werden. Die eisenzeitlichen Proben der süd-sibirischen Fundplätze Aržan, Ak-Alacha, Justyd, werden der Pazyryk-Kultur zugewiesen und werden den Ost-Skythen zugerechnet. Ebenso die kasachischen Proben aus dem Minusinsker Becken (Barsučij Log), dem Altai (Berel), Ulžan und Zevakino.

Trotz des allgemein sehr guten *capture*-Erfolgs für die 41 ausgewählten Proben, wurden für die nachfolgenden populationsgenetischen Analysen, mit Ausnahme der Bestimmung von genetischen Distanzen (F_{st}), die Alignments gegen das gesamte Referenzgenom (hg19) verwendet. Da die Vergleichspopulationen zumeist auf SNP-Daten (Mathieson et al. 2015; Lazaridis et al. 2016) oder genomischen Datensätzen mit sehr niedriger Abdeckung beruhen (Allentoft et al. 2015), kann durch Verwendung des Alignments gegen das gesamte Genom der Überschneidungsbereich mit Referenz-Daten bzw. die verfügbaren SNPs deutlich erhöht werden (Sell 2017).

5.2 Pigmentierungsmuster und funktionale Marker

Aus allen in Kapitel 4.5 beschriebenen Markern wurden gezielt 13 Marker ausgewählt, welche unter positiver Selektion in Europa oder Ost-Asien stehen und damit eine starke Varianz der Allelfrequenzen zwischen europäischen und asiatischen Populationen zeigen. Die errechneten Allelfrequenzen der West- und Ostskythen werden dann mit Frequenzen moderner Populationen verglichen.

Aufgrund einer nicht ausreichenden Abdeckung einiger Marker nach dem Filtern der MLE-*vif*-files mit einer Genotyp-Qualität von mindestens 30 und einer Abdeckungstiefe von 2x , bzw. GQ 15 und 1x für das HIrisplex-System, konnten eine Probe der Tagarkultur (Bl5) aus Barsučij Log und zwei bronzezeitlichen Proben aus Kamennyj-Ambar (Kama6 und Kama10) für keinen der im folgenden beschriebenen Marker typisiert werden.

Es wurden entsprechend der Fundplatzverteilung der vorliegenden Proben folgende Populationen als zentral-asiatische Referenzgruppe (ZAS) ausgewählt: Kasachen, Kirgisen und Mongolen. Stellvertretend für Nordwest-Russland (NWR) werden Individuen der Republik Komi (Komi-Zyrian), westlich des Urals ausgesucht.

Aufgrund der geringen Individuenzahl können für die bronzezeitlichen (N=2, ohne Kama1, Kama6 und Kama10) sowie die neolithischen Proben (N=2) keine Allelfrequenzen bestimmt werden. Die Genotypen der jeweiligen Proben werden in der Diskussion der einzelnen Marker gesondert aufgeführt. Auch die Individuenzahl für die Genotypen der West-Skythen variiert zwischen N=3 und N=7 und ist somit im Verhältnis zu modernen Referenzen sehr gering.

Laktasepersistenz

Das abgeleitete Allel des 13910T (LCTa) Markers steht in europäischen Populationen unter positiver Selektion und ist für die Umsetzung von Milchzucker verantwortlich (Bersaglieri et al. 2004). Mit einer Allelfrequenz von 8,3 % bei den West-Skythen (N=6) und 6,25 % der Ost-Skythen (N=16) entspricht die Verteilung am ehesten modernen zentral-asiatischen Populationen (4 %). Die beiden neolithischen Proben aus Solonzy sowie eine bronzezeitliche Probe SW-Russlands (Kal1) zeigen ebenfalls den homozygot anzestralen Status. Die Frequenzverteilung der skythischen Populationen unterscheidet sich von den detektierten Frequenzen aus der Dissertation von Martina Unterländer. Hier konnte für die westlichen Skythen nur die homozygot anzestrale Form festgestellt werden, für die östlichen Skythen jedoch eine Frequenz von 8,5 %. Sie ist somit höher und gegenläufig zu den

hier ermittelten Frequenzen. Allerdings liegt in beiden Arbeiten die Stichprobenzahl der West-Skythen bei $N=4$ bzw. $N=6$, weshalb die Werte bei einer größeren Stichprobe stark variieren könnten.

Eine nachweisbare Frequenz für das abgeleitete Allel tritt erstmals innerhalb der letzten 4000 Jahre auf (Burger et al. 2007; Gamba et al. 2014). Neueste Studien von Mathieson und Kollegen bestätigen dies anhand ihres Datensatzes von 230 prähistorischen Proben Westeurasiens. Das früheste, abgeleitete Allele des LCTa Markers konnte für einen europäischen Vertreter der Glockenbecher-Kultur nachgewiesen werden (Mathieson et al. 2015). In einer Studie von bronzezeitlichen Individuen Eurasiens desselben Jahres weisen die Autoren eine überraschend niedrige Frequenz von 10 % nach (Allentoft et al. 2015). Innerhalb der bronzezeitlichen Steppenkulturen konnten sie dabei die höchste Frequenz des abgeleiteten Allels für Vertreter der Jamnaja-Kultur feststellen und schließen daraus einen möglichen Ursprung der Laktosetoleranz in den Steppegebieten. In der Dissertation von Sandra Wilde zu äneolithischen und bronzezeitlichen Individuen der osteuropäischen Steppe des Nordpontikums konnte für 43 typisierte Proben ausschließlich der homozygot aneztrale Status nachgewiesen werden (Wilde 2014).

Pigmentierung

Die Augen-, Haar- und Hautfarbe ist vor allem in Westeurasiens hochvariabel und lässt neben dem Hinweis auf positive Selektion Rückschlüsse auf mögliche Migrationsbewegungen prähistorischer Populationen zu.

Basierend auf den Ergebnissen des HIRISplex-Systems zur Bestimmung der Augen- und Haarfarbe konnten blaue Augen in Kombination mit hellem Haar für eine bronzezeitliche Probe aus SW-Russland (Kama7), einen Vertreter der frühen Sarmaten (Pr10), eine Probe der Tagar-Kultur (Bl3), eine Probe der Pazyryk-Kultur des Ukok-Plateaus (Ak5-8) und eine kasachische Probe (Ze3) festgestellt werden. Dies ist insofern bemerkenswert, da alle Proben des Fundplatzes Barsučij Log sowie die Probe Ze3 in der Hauptkomponentenanalyse mit modernen Vergleichspopulationen eine deutlich westlichere Orientierung aufweisen als es der geografischen Lage ihres Fundplatzes entsprechen würde (siehe Kapitel 4.6 und 5.3). Den Phänotyp eher helles Haar und braune Augen zeigen eine weitere Probe der Tagar-Kultur (Bl2) sowie Kol4 aus SW-Russland. Die Ergebnisse aller weiteren Individuen sprechen für einen eher dunkleren Phänotyp von braun bis schwarzem Haar und einer braunen

Augenfarbe. Die bronzezeitliche Probe aus Kalinovka entspricht im Vergleich zu Kama7 ebenfalls einem dunkleren Erscheinungsbild, ebenso die beiden neolithischen Proben.

Das abgeleitete Allel in SLC45A2 wird mit einer hellen Pigmentierung der Haut assoziiert und ist in europäischen Populationen nahezu fixiert. Für die West-Skythen (N=5) konnte eine Allelfrequenz von 60 % und für die Ost-Skythen (N=20) 40 % bestimmt werden. Sie liegen damit noch unter der Frequenz von 93,8 % in europäischen Populationen und den berechneten Frequenzen für skythische Gruppierungen aus (Untertaler 2014). Eine neolithische Probe sowie eine bronzezeitliche Probe aus Kalinovka weisen für rs16891982 in SLC45A2 den homozygot ancestralen Status auf.

Vom Mesolithikum an steigt die Frequenz des abgeleiteten Allels innerhalb von ~ 3000 Jahren bis zu einer annähernden Fixierung in der Bronzezeit in Europa (Allentoft et al. 2015). In osteuropäischen Steppenpopulationen der Bronzezeit wurde allerdings eine noch niedrige Frequenz von 43,2 % des abgeleiteten Allels festgestellt (Wilde et al. 2014).

Für den Marker HERC2, welcher stark mit dem Phänotyp für eine blaue bzw. braune Augenfarbe korreliert, kann eine intermediäre Frequenz für bronzezeitliche Individuen Europas festgestellt werden, ist jedoch in Steppenpopulationen des nordpontischen Raums nicht nachweisbar (Allentoft et al. 2015). Eine niedrige Frequenz von 16 % konnten Wilde und Kollegen für bronzezeitliche Individuen des nord-pontischen Raums nachweisen (Wilde et al. 2014). Neun bronzezeitliche Individuen aus Chakassien in Südsibirien zeigen für das abgeleitete Allel jedoch eine Frequenz von 72 % (Bouakaze et al. 2009).

Für die Gruppe der westlichen (N=7) sowie östlichen Skythen (N=16) konnten vergleichbare Frequenzen für das abgeleitete Allel von 35,7 % bzw. 34,4 % ermittelt werden und liegen damit unter der Verteilung moderner europäischer (63,6 %), jedoch über zentral- und ost-asiatischen Populationen (6,9 % und 0,2 %). Die beiden neolithischen Proben aus West-Sibirien zeigen hier den homozygot ancestralen Status.

Für den Marker OCA2.3 konnte das abgeleitete Allel für die westlichen Individuen (N=5) gar nicht und für die Ost-Skythen mit 8,3 % nachgewiesen werden. Die Frequenzen liegen deutlich unter der ost-asiatischen Verbreitung des abgeleiteten Allels (59,6 %). Die West- und Ost-Skythen zeigen für diesen Marker eine deutliche Nähe zu modernen europäischen Populationen mit einer Allelfrequenz von 0 % bzw. modernen Individuen NW-Russlands (1,87 %). Für den Marker SLC24A5 stimmen die Frequenzen von 80,6 % für die Ost-Skythen (N=18) und 87,5 % für die West-Skythen (N=4) annähernd mit der Verbreitung in europäischen Populationen (99,7 %) überein.

Alkoholverträglichkeit

Das Enzym ALDH2 sorgt für den Abbau von toxischen Aldehyden, und eine Mutation kann zu einer typischen Reaktion wie Gesichtsrötung, Schwindel und Übelkeit führen (*alcohol flush^s*) (Yoshida et al. 1984; Oota et al. 2004). Das abgeleitete Allel steht unter positiver Selektion in ost-asiatischen Populationen. Zusätzlich werden zwei weitere Mutationen in dem ADH1B-Gen (ADH1Ba und ADH1Bb) untersucht, welche mit dem Abbau von Alkohol assoziiert werden und in modernen asiatischen Populationen unter Selektion stehen (Peng et al. 2010). Die Frequenz für das abgeleitete Allel liegt für ADH1Ba in zentralasiatischen Populationen bei 27,2 %, in ost-asiatischen Populationen bei 51 %, in europäischen und russischen Populationen bei 0%. Für zentral-asiatische Populationen ist keine Bestimmung der Allel-Frequenz für den Marker ALDH2 möglich, in Ost-Asien liegt die Frequenz jedoch bei 17,4 % und in Europa bei 0%. In modernen europäischen Populationen kann das abgeleitete Allel nicht nachgewiesen werden. Für alle drei Marker zeigen die typisierten Individuen die homozygot anzestrale Form und entsprechen damit eindeutig dem europäischen Phänotyp. Für ADH1Ba und ALDH2 konnten Allelfrequenzen der West-Skythen (N=3 bzw. N=5) und Ost-Skythen (N=18) bestimmt werden. Die bronzezeitliche Probe aus Kalinovka, Kama7 sowie die beiden bronzezeitlichen Proben aus Solonzy zeigen ebenfalls den homozygot anzestralen Status und entsprechen einem europäischen Phänotyp.

Xenobiotischer Metabolismus

Mit 50 % für AGT bei den Ostskythen (N=16) und 36,1 % für NAT2 der Ostskythen (N=18) und 25 % Westskythen (N=4) entsprechen die Allelfrequenzen für die abgeleiteten Allele für NAT2 und AGT eher dem europäischen Phänotyp.

Bluthochdruck und Diabetes Typ 2

Die Berechnung eines Risiko-Wertes für Diabetes Typ 2 beruht auf einem Modell nach (Cornelis et al. 2009) Acht von zehn benötigten Markern sind Teil des Anreicherungsverfahrens und konnten für Individuen mit ausreichender Abdeckung bestimmt werden. Der GRS-Wert liegt für die hier analysierten Individuen bei durchschnittlich 8,26 und damit unter dem Wert, der in der Studie für modernen Populationen als durchschnittlicher Risiko-Wert für Diabetes Typ 2 bestimmt wurde. Somit würden die skythischen Populationen ein deutlich geringeres Risiko aufweisen. Es gilt allerdings zu bedenken, dass neben geneti-

schen Faktoren auch Umwelteinflüsse für die individuelle Entwicklung von Diabetes Typ 2 verantwortlich sind (Cornelis et al. 2009). Außerdem konnten aufgrund fehlender Abdeckung für einige Proben nur 3 Genotypen bestimmt werden, was sicherlich zu einem ungenauen Ergebnis führen kann, obwohl der hier verwendete standardisierte GRS-Wert fehlende Positionen berücksichtigt.

Für rs7903146 des Gens TCF7L2, welches ebenfalls mit der Ausprägung von Diabetes Typ 2 assoziiert wird (Chen et al. 2012) konnten Allel-Frequenzen für West-Skythen (N=7) und Ost-Skythen (N=17) bestimmt werden. In modernen europäischen Populationen liegt die Frequenz des abgeleiteten Allels bei 51,2 % und in Ost- und Südasien bei 96,5 bzw. 62,8 %. Mit Frequenzen von 64,7 % der West-Skythen und 78,6 % der Ost-Skythen ist kein deutlicher Unterschied der Gruppen zueinander zu erkennen. Der Phänotyp nimmt einen intermediären Status zwischen europäischen und asiatischen Populationen ein. Eine Ausnahme bildet hier eine neolithische Probe aus Solonzy mit einem homozygoten Genotyp des ancestralen Allels. Dieser ist sehr selten und ist am ehesten in Europa nachzuweisen.

Die Allelfrequenzen für SHB3 (rs653178) wird mit einer Erkrankung an Zöliakie und einem Risiko für Bluthochdruck assoziiert und steht in prähistorischen eurasischen Populationen unter Selektion (Hunt et al. 2008; Mathieson et al. 2015). Die Frequenz für das abgeleitete Allel liegt in ost-asiatischen Populationen bei nur 0,3 %, in zentral-asiatischen Populationen bei 5 %, in Europa und der Population aus NW-Russland ist eine höhere Frequenz von 46,4 % und 51,1 % nachweisbar. Die Frequenzen des abgeleiteten Allels der West-Skythen (N=3) und Ost-Skythen (N=15) liegen mit 16,7 % und 20 % nah beieinander und entsprechen eher einem europäischen bis intermediären Phänotyp. Hier nimmt die sarmatische Probe Pr3 mit dem homozygot, abgeleiteten Status als einzige Probe eine Sonderstellung ein. Mathieson und Kollegen konnten für ATXN2/SH2B3 eine niedrigere Frequenz für westeurasische Populationen feststellen, als unter der Hypothese eines möglichen *selective sweeps* weiterer Varianten, die mit Zöliakie und einem damit verbundenen Schutz gegen Ergothionein-Mangel bei einer Ernährungsumstellung zu erwarten gewesen wären (Mathieson et al. 2015).

Klimaanpassung

Zwei Mutationen, die in Verbindung mit einer Anpassung an klimatische Verhältnisse stehen könnten, sind der SNP rs653178 im Gen ABCC11 (Yoshiura et al. 2006; Ohashi et al. 2011) und EDAR (Bryk et al. 2008). Das abgeleitete Allel im EDAR-Gen gilt zudem als Ziel eines *selective sweeps* in Ostasien (Kamberov et al. 2018). In ostasiatischen und zentralasiatischen Populationen liegt die Frequenz des abgeleiteten Allels des EDAR-Gens bei 87,3 % bzw. 63,4 % und in modernen europäischen Populationen bei 1,1 %. Für die Ostskythen (N=18) konnte eine Frequenz von 30,5 %, für die West-Skythen (N=4) von 12,5 % festgestellt werden. Der Phänotyp liegt damit zwischen zentral-asiatischen und europäischen Populationen, die West-Skythen jedoch näher bei Europa. Die in der vorliegenden Arbeit berechnete Allel-Frequenz liegt deutlich unter der detektierten Frequenz (55 %) von Martina Unterländer in eisenzeitlichen, östlichen Skythen (Unterländer 2014). Für nur zwei Proben der Bronze- und Eisenzeit Westeurasiens aus konnte das abgeleitete EDAR-Allel nachgewiesen werden: ein Vertreter der Afanasievo-Kultur (3300 -3000 v. Chr.) sowie der eisenzeitlichen, skythischen Probe (400 -200 v. Chr.) (Mathieson et al. 2015).

Für ABCC11 konnten nur Allelfrequenzen der Ost-Skythen (N=16) berechnet werden und nehmen mit einer Allelfrequenz des abgeleiteten Allels von 46,9 % eine intermediäre Position zwischen Europa (13,6%) und Zentral - / Ost-Asien (76,2 % / 78 %) ein.

Die Allelfrequenzen der westlichen Skythen-Gruppe der 13 diskutierten Marker entsprechen der Verbreitung in europäischen Populationen bzw. liegen zwischen der Allelverteilung in Zentralasien und Europa. Einzige Ausnahme stellt der Marker TCF7L2. Hier liegt die Allelfrequenz des abgeleiteten Allels zwischen modernen europäischen und ostasiatischen Populationen. Die Ost-Skythen nehmen in Bezug auf die festgestellten Phänotypen eine intermediäre Position zwischen modernen europäischen und zentral-asiatischen bzw. ost-asiatischen Populationen ein oder stehen europäischen Populationen näher, mit Ausnahme der Allelfrequenzen der Marker OCA2.3 und SLC45A2. Es sind somit nur geringe Unterschiede in der Frequenzverteilung zwischen den östlichen und westlichen Skythen der Eisenzeit erkennbar. Basierend auf der Auswertung der Phänotypen lässt sich für die zentral-asiatischen Fundplätze ein deutlicher Einfluss aus Europa bzw. ein geringer ost-asiatischer Zustrom für die Eisenzeit feststellen.

5.3 Populationsgenetik eisen- und bronzezeitlicher Steppenpopulationen Eurasiens

Um eine genetische Ähnlichkeit bronze- oder eisenzeitlicher Individuen der vorliegenden Arbeit zu modernen eurasischen Populationen zu untersuchen wurden zunächst alle Proben mittels LASER in einen entsprechenden Referenzraum moderner Populationen (Hellenthal et al. 2014; Busby et al. 2015) projiziert. Es galt festzustellen, inwieweit die einzelnen kulturellen Gruppierungen bzw. Fundplätze in ihrer geografischen Lage mit ihren Positionen in der PCA übereinstimmen und welche Nähe sie zu heutigen Populationen Eurasiens zeigen. Um die Frage nach einer Homogenität zwischen prähistorischen Reiternomaden der Bronze- und Eisenzeit und einem möglichen Ursprungsgebiet der Skythen diskutieren zu können, konnten zusätzlich geografisch und zeitlich relevante aDNA-Datensätze aus der Literatur hinzugefügt werden. (Lazaridis et al. 2014; Fu et al. 2014; Jones et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015; Jones et al. 2017; Sikora et al. 2017).

Anhand der *outgroup-f₃*-Statistik (Patterson et al. 2012) konnte die Höhe der geteilten genetischen Drift zwischen den jeweiligen Fundplätzen bzw. einzelnen Proben und den 23 ausgewählten modernen Populationen ermittelt werden. Als Referenzdatensatz früherer oder zeitgleicher prähistorischer Kulturgruppen zu den hier bearbeiteten Individuen dient der Datensatz von Lazaridis und Kollegen (Lazaridis et al. 2016).

Für die Bestimmung der genetischen Distanzen nach Hudson (Bhatia et al. 2013) wurden die Proben entsprechend ihrer Zugehörigkeit zu einer Kulturstufe bzw. ihrer zeitlichen Einordnung gruppiert und genetische Distanzen untereinander bzw. mit dem weltweiten *1000 Genome Phase 3 Datensatz* (1000 Genomes Project Consortium et al. 2015) verglichen.

Die Ergebnisse der drei populationsgenetischen Methoden werden jeweils für die eisenzeitlichen Ost- und West-Skythen, frühen Sarmaten, bronzezeitlichen und neolithischen Proben bzw. ihren entsprechenden Kulturgruppen diskutiert. Die visuelle Darstellung der genetischen Distanzen und Ähnlichkeiten anhand der Hauptkomponentenanalyse konnte für alle Individuen in der *outgroup-f₃*-Statistik bestätigt werden.

Phänotypische Merkmale einzelner Proben bzw. Allelfrequenzen der östlichen und westlichen Skythen fließen teilweise in die Diskussion der jeweiligen Fragestellungen mit ein.

Aufgrund der geringen Individuenzahl bronzezeitlicher Proben liegt der Schwerpunkt der Hypothesendiskussion auf der Populationshistorie Zentralasiens in der Eisenzeit. Das Verhältnis zu bronzezeitlichen Reiternomaden ist hierfür jedoch von großer Bedeutung.

5.3.1 Genetische Kontinuität zwischen bronze- oder eisenzeitlichen Kulturgruppen und lokalen Populationen des heutigen Eurasiens

In der Hauptkomponentenanalyse weisen alle eisenzeitlichen und bronzezeitlichen Individuen eine genetische Ähnlichkeit mit modernen Populationen auf und verteilen sich entsprechend der Fundplatz-Verteilung über den eurasischen Steppengürtel Zentralasiens bis nach Ost-Asien.

Heterogenität der innerasiatischen Skythen

Die Verwandtschaft der Reiterverbände des eurasischen Steppengürtels untereinander und ihre möglichen Ursprünge sind nach wie vor Grundlage neuester Publikationen (Juras et al. 2017; Unterländer et al. 2017; Damgaard et al. 2018; Krzewinska et al. 2018).

Aufgrund geringer genetischer Distanzen zwischen eisenzeitlichen skythischen Gruppierungen der west- und osteurasischen Steppe sowie zwischen zentral- und nord-asiatischen Kulturgruppen wird von einem kontinuierlichen Genfluss skythischer Bündnisse ausgegangen (Unterländer et al. 2017). Damgaard und Kollegen unterstützen die Annahme eines multiregionalen Ursprungs mit Vermischungsereignissen mit Nachfahren spätbronzezeitlicher Viehhirten und lokalen Populationen (Damgaard et al. 2018).

Für die eisenzeitlichen Skythen der pontischen Steppe (795 v. Chr. – 248 n. Chr.) konnte ebenfalls eine hohe Diversität nachgewiesen werden. Sie bilden in einer PCA vier Cluster entsprechend ihrer genetischen Ähnlichkeiten mit modernen Populationen aus: 1. eine nordeuropäische Gruppe (NE), 2. eine südeuropäische Gruppe (SE) 3. eine Gruppierung innerhalb der Steppe (SC) und 4. eine zentrale Gruppe mit Bezug zum heutigen Griechenland, Kroatien und der Türkei (CC) (Krzewinska et al. 2018).

In der vorliegenden Arbeit konnten zwar ebenfalls keine bzw. nur geringe genetische Distanzen zwischen den Kulturgruppen der zentral- und nordasiatischen Skythen errechnet werden, die Ergebnisse der PCA mit modernen Populationen sowie die Berechnung der genetischen Drift mittels *outgroup*- f_3 -Statistik verdeutlichen, dass die östlichen, eisenzeitlichen Skythen keine homogene Gruppe bilden.

Auch die hier untersuchten skythischen Individuen der zentral- und nord-asiatischen Steppe waren keine homogene Gruppierung hinsichtlich ihrer Ähnlichkeit zu modernen Populationen in der PCA, auch nicht innerhalb ihrer entsprechenden Kulturstufen. Individuen

der Pazyryk-Kultur Südsibiriens bilden ein zentral-asiatisches Cluster, wobei ein Individuum der Chuya-Region (J12-6) innerhalb der Verteilung moderner, ost-asiatischer Populationen liegt. Ein Vertreter der älteren Aldy-Bel-Kultur aus Aržan 2 weist ebenfalls diesen deutlichen ost-asiatischen Bezug auf. Die Resultate der F3-Statistik sind konsistent mit den Ergebnissen der PCA. Für A21 liegt Han-China unter den sechs höchsten f3-Werten, für die Probe J12-6 liegen die beiden höchsten f3-Werte bei ost-asiatischen Populationen und bestätigen im Vergleich mit den F3-Werten der gruppierten Fundplätze Aržan bzw. Justyd ihre Sonderstellung in der PCA. Des Weiteren ist die Probe J12-6 für den Marker SLC24A5 homozygot mit dem abgeleiteten Allel. Dieser Genotyp kann ausschließlich in osteuropäischen Populationen nachgewiesen werden. Die weiteren Fundplätze Aržan, Ak-Alacha, Justyd und Berel des „zentralasiatischen“ Clusters weisen die meiste genetische Drift mit Nganasanen, Ewenken und der Republik Tuwa (Ak-Alacha, Berel) bzw. Nganasanen, Burjaten, Tschuwaschen und der Republik Tuwa auf (Aržan, Justyd). Die Verbindung nach Sibirien zu den Nganasanen konnte auch anhand einer ADMIXTURE-Analyse von acht eisenzeitlichen Skythen, für Vertreter der Aldy-Bel, Zevakino-Čilikta und Pazyryk-Kultur nachgewiesen werden. Drei Individuen (Pr3, Be9 und Be11) wurden ebenfalls in der vorliegenden Arbeit analysiert (Unterländer et al. 2017). Zusätzlich ist für die Skythen der Eisenzeit eine Komponente nachweisbar, die in ost-asiatischen Populationen dominant ist.

Die beiden Vertreter der der ältesten hier untersuchten Kultur der Zevakino-Čilikta Studie aus Ost-Kasachstan bilden keine einheitliche Gruppe hinsichtlich ihrer Position in der PCA. Die breite geografische Streuung der ost-kasachischen Proben Ze3 und Ze4 konnte ebenfalls in der *outgroup-f3*-Statistik bestätigt werden. Während Ze3 eine deutliche Nähe zu russischen, europäischen und osteuropäischen Populationen zeigt, liegen die höchsten f3-Werte für Ze4 bei sibirischen und ost-asiatischen Populationen. Für SLC45A2 ist die Probe Ze3 homozygot im abgeleiteten Allel, welches in Europa nahezu fixiert ist. Auf Ebene der mitochondrialen Haplogruppen überwiegen für diesen Fundplatz westeurasische Linien, 54,5 % europäisch und 45,5 % ostasiatisch (Unterländer 2014). Dies lässt auf einen Einfluss der westlichen geprägten Andronovkultur der späten Bronzezeit schließen, der bereits für Individuen der Aldy-Bel- und der späteren Pazyryk-Kultur abnimmt. Das Zentrum früh-skythischer Kultur wird basierend auf archäologischen Funden in Tuva und Südsibirien angenommen. Einflüsse aus Zentralasien und Ostasien sind erst in darauf folgenden Gruppen sichtbar (Parzinger 2007). Für eisenzeitliche Steppe-Individuen konnte bereits ein westeurasischer Einfluss nachgewiesen werden, der zur damaligen Zeit deutlich höher war, als es für heutige Populationen derselben Regionen festzustellen ist (Allentoft et al. 2015).

Dies kann ebenfalls in den berechneten genetischen Distanzen der vorliegenden Arbeit bestätigt werden. Für die Berechnung der genetischen Distanzen zu modernen Referenzpopulationen wurden die Proben entsprechend ihrer kulturellen oder zeitlichen Zuordnung gruppiert. Das ermöglicht es Fundplätze mit einem oder wenigen Individuen einzubeziehen und dennoch Unterschiede möglichst kleiner Gruppierungen aufzuzeigen.

Eisenzeitliche Proben zeigen die geringsten genetischen Distanzen zu Populationen Süd-Asiens (0,0337-0,0383), im Vergleich jedoch höhere Distanzen zu Ostasien (0,0603-0,0651) als zu europäischen Populationen (0,0436 und 0,0432). Die Distanzen spiegeln eher eine Nähe zu Westurasien, als zu Ostasien wider. Die geringeren Distanzen zu süd-asiatischen Populationen können möglicherweise durch das Fehlen an zentral-asiatischen Referenzpopulationen erklärt werden.

Die Ost-Skythen (Pazyryk, Aldy-Bel) zeigen die geringsten genetischen Distanzen zu süd-asiatischen Populationen (\emptyset 0,0383 und 0,0322), die höchsten zu Ostasien (0,058 und 0,0460) und Europa (0,051 und 0,045). Die Differenz der berechneten Werte ist jedoch sehr gering und könnte auf Einflüsse aus Ost und West zurückzuführen sein. Der Genpool heutiger Populationen Sibiriens setzt sich ebenfalls aus ost- und westeurasischen Linien zusammen. Die Altai-Region Südsibiriens zeigt dabei den höchsten Anteil einer europäischen Komponente, die jedoch stark zwischen den untersuchten Populationen variiert (Cardona et al. 2014). Der Anteil westeurasischer Linien ist in westlichen und südlichen Teilen Sibiriens höher als in zentralen - und nordöstlichen Teilen (Derenko et al. 2014).

Eine Sonderstellung innerhalb der kasachischen und sibirischen Skythen weisen Vertreter der Tagar-Kultur aus dem Minusinsker Becken auf. Die Distanz zu modernen zentral-asiatischen Populationen lässt auf eine isolierte Population mit geringem Genfluss zu lokalen Populationen schließen. Auch kann eine Verwandtschaft mit Vertretern der zeitlich folgenden Pazyryk-Kultur nicht angenommen werden. Auch auf mitochondrialer Ebene weisen Vertreter der Tagar-Kultur innerhalb der Ost-Skythen hohe genetische Distanzen zu allen in der Studie bearbeiteten östlichen Skythen auf (Unterländer 2014) und zeigen eine Nähe zu den frühen Sarmaten, was in der vorliegenden Arbeit in der PCA bestätigt werden kann. Vertreter der Aldy-Bel-Kultur des Altai stehen bei einem Baum der Form D(Tagar, Solonzy, Aldy Bel, S_Mbuti) den neolithischen Proben aus Solonzy näher als den Proben des Minusinsker Beckens. Dies schließt eine genetische Verbindung der früheren Populationen aus Aržan mit der später folgenden Tagar-Kultur im Minusinsker Becken aus. Weitere Baum-Formationen in der D-Statistik erzielten für die Fundplätze leider keine weiteren signifikanten D-Werte $|Z| > 3$.

Die Nähe skythischer Proben des nord-pontischen Raums zu Vertretern der Tagar-Kultur aus Südsibirien im Vergleich zu Proben der Pazyryk-Kultur konnten auch Juras und Kollegen in ihrer Studie nachweisen (Juras et al. 2017).

Für Vertreter der innerasiatischen Saken, die teilweise mit der Tagar-Kultur assoziiert werden, konnte basierend auf genomischen Daten kein Genfluss mit ungarischen Skythen festgestellt werden (Damgaard et al. 2018), jedoch zwischen früh-skythischen Gruppen der nord-pontischen Region und Zentralasien (Unterländer et al. 2017).

Die Proben der Tagar-Kultur aus dem Minusinsker Becken zeigen entsprechend ihrer abweichenden Position in der PCA in Bezug auf ihren Fundplatz die meiste genetische Drift mit Finnland, Polen, Russland und Europa, gefolgt von Ländern Osteuropas und des Balkans. Proben des Fundplatzes Barsučij Log zeigen auch in den Fst-Werten eine deutliche Nähe zu europäischen Populationen mit den vergleichsweise geringen genetischen Distanzen zu CEU und TSI (0,0188 und 0,0197). Es ist somit ein deutlich westeurasischer Einfluss erkennbar, der mit einer Abstammung von bronzezeitlichen Viehhirten erklärt werden könnte (Damgaard et al. 2018). Ein westeurasischer Bezug wurde auch auf mitochondrialer Ebene für 12 Individuen der Tagar-Kultur aus Chakassien nachgewiesen (Keyser et al. 2009). Der ost-asiatische Einfluss für die Individuen der Tagar-Kultur (Unterländer 2014) kann anhand genomischer Daten nicht nachgewiesen werden und könnte auf eine Migration weiblicher Individuen aus Osteurasien hindeuten.

Kontinuität der westlichen Steppe

Trotz verbindender kultureller Elemente der frühen Skythen des nordpontischen Raums und eisenzeitlichen Nomadenreitern des eurasischen Steppengürtels ist von einer genetischen Distanz der westlichen und östlichen Gruppen auszugehen (Damgaard et al. 2018; Krzewinska et al. 2018), wobei ein kontinuierlicher Genfluss zwischen West und Ost teilweise angenommen wird (Unterländer et al. 2017; Krzewinska et al. 2018).

Anhand eines Vertreters der Initialen Skythen Südwest-Russlands (Nov7) aus der Orenburg-Region sowie eines Skythen (Kol4) der Woronesch-Region kann für SW-Russland ein Zeitraum von 800 -300 v. Chr. untersucht werden. Für zentralasiatische Skythen kann während der Eisenzeit eine Zunahme ostasiatischer Linien bestimmt werden. Für die beiden westlichen Skythen SW-Russlands kann die meiste genetische Drift mit Russland und Populationen Osteuropas nachgewiesen werden. Auch die Berechnung der genetischen Distanzen zu modernen Populationen zeigt keinen zentral- oder ost-asiatischen Einfluss in der späten Eisenzeit. Jedoch müssen die Ergebnisse aufgrund der geringen Probenzahl unter Vorbehalt interpretiert werden.

Die Migration der Sarmaten aus dem Osten bewirkte nach einer Phase der Koexistenz den Untergang der nord-pontischen Skythen (Parzinger 2007). Genomische Studien belegen eine Nähe zur östlichen Jamnaja-Kultur und der bronzezeitlichen Poltavka-Kultur. Für Sarmaten des südlichen Urals konnte eine genetische Kontinuität über einen Zeitraum von 300-500 Jahren belegt werden (Krzewinska et al. 2018). Die Sarmaten bilden mit eisenzeitlichen Nomaden der pontisch-kaspischen Steppe ein Cluster in der PCA, was für eine genetische Kontinuität im Bereich des südlichen Urals spricht. Die frühen Sarmaten aus Pokrovka weisen basierend auf ihrer Lage in der PCA einen zentral-asiatischen Einfluss auf und liegen süd-östlicher als ihr Fundort in SW-Russland. Die F3-Werte weisen zwar hohe Standardfehler auf, die meiste genetischen Drift teilen die Sarmaten jedoch mit Russland und Polen, gefolgt von osteuropäischen Populationen was wiederum mit der geografischen Lage des Fundplatzes korreliert. In der Dissertation von Martina Unterländer basierend auf mitochondrialen Daten können die geringen genetischen Distanzen zu modernen Referenzpopulationen SW-Russlands die vorliegenden Ergebnisse der *outgroup-f₃*-Analyse ebenfalls bestätigen (Unterländer 2014).

Die frühen Sarmaten als Vertreter der eisenzeitlichen West-Skythen weisen die geringsten genetischen Distanzen zu modernen europäischen Populationen auf (0,0392 und 0,0279) auf.

Europäischer Ursprung der bronzezeitlichen Sintaschta-Kultur

Die im südöstlichen Uralvorland erstmals aufgetretene und weit verbreitete Sintaschta-Kultur dehnte sich bis nach Europa aus und gilt als Vorläufer der Andronovo-Kultur in Asien. Es konnte eine genetische Nähe der Sintaschta-Kultur zu Vertretern der Schnurbandkeramik nachgewiesen werden (Allentoft et al. 2015). Ein möglicher gemeinsamer Ursprung beider kultureller Gruppen schließt eine Entstehung der Sintaschta-Kultur in Asien aus. Die Autoren postulieren vielmehr eine Ostwärts-Bewegung von Individuen diese Kulturgruppe.

Für die hier untersuchten bronzezeitlichen Individuen der Sintaschta-Kultur aus SW-Russland ist eine deutliche genetische Nähe zu europäischen Populationen erkennbar. Sie weisen jeweils die meiste genetische Drift mit Russland, Weißrussland, Polen, Finnland und CEU auf. Die Proben der Sintaschta-Kultur verteilen sich in der PCA über Europa und Osteuropa und mit Kama6 bis nach Zentralasien. Die Individuen aus Kamennyj Ambar bilden in der PCA keine homogene Gruppe. Diese Beobachtung konnte ebenfalls in einer weiteren Studie anhand 50 Vertretern der Sintaschta-Kultur derselben Fundgruppe bestä-

tigt werden (Narasimhan et al. 2018). Menschen verschiedenster Abstammungsverhältnisse könnten hier lange Zeit parallel zusammengelebt haben. Die Autoren weisen hier genetische Anteile frühbronzezeitlicher Steppenomadern, ein Muster west-sibirischer Jäger-Sammler-Kulturen sowie vereinzelt ost-asiatische Einflüsse nach. Die errechnete Nukleotiddiversität ist für Vertreter der Sintaschta-Kultur im Vergleich zu den anderen eisenzeitlichen Kulturgruppen dieser Arbeit am höchsten, was ebenfalls die Annahme einer stark heterogenen Gruppe stützt.

Der weitere bronzezeitliche Fundplatz Kalinovka des Oblast Kursk liegt an der ukrainischen Grenze, basierend auf der Position in der PCA zeigt Kal1 deutlich einen Einfluss aus Osteuropa. Alle bronzezeitlichen Individuen kumuliert zeigen ebenfalls hohe Distanzen zu ostasiatischen Populationen (0,0885-0,0994), jedoch vergleichbare Werte für Süd-Asien (0,0452-0,0510) und Europa (0,0477 und 0,0465). Werden die Fst-Werte nur für Vertreter der Sintaschta-Kultur berechnet, wird die genetische Distanz mit Werten über 0,09 für ost-asiatische Populationen noch deutlicher. Die Distanzen sind hier zu CEU und TSI geringer, als zu süd-asiatischen Populationen.

Verbindung südlicher Steppenomadern nach Südwest-Asien

Die Unterscheidung zwischen sakischen und skythischen Verbänden ist aufgrund ihrer verbindenden, kulturellen Elemente in schriftlichen Überlieferungen nicht immer eindeutig, sie schienen jedoch ein Bindeglied zwischen dem achaimenidischen-persischen Raum und der nord-östlich gelegenen eurasischen Steppe zu sein.

Die Saken bevölkerten Landstriche im südöstlichen Kasachstan und archäologische Funde belegen das erste Erscheinen skythisch-sakischer Kulturelemente bereits in der frühen Skythenzeit 7.-6. Jh. v. Chr. (Parzinger 2007). Sakische Stämme wanderten über den Tien Shan bis ins heutige Usbekistan. Auch im Pamir-Gebirge Tadschikistan und im Norden Pakistans bis zum Indus konnten skythisch-sakische Funde nachgewiesen werden. Aufgrund massiver Bevölkerungsveränderungen Innerasiens und einer Migration der Wusun wurden die Saken vermutlich am Ende des 3. Jh. v. Chr. aus ihrem Bereich vertrieben bzw. aufgelöst.

In der vorliegenden Arbeit kann für den Saken aus Ulžan aus Südkasachstan eine genetische Ähnlichkeit mit heutigen Populationen des Nahen Ostens identifiziert werden. Die Position des Saken in der PCA weicht massiv von der geografischen Lage seines Fundplatzes ab und liegt in der Verbreitung von modernen syrischen und iranischen Individuen. Der hier nachgewiesene, genetische Bezug zu Iran und Syrien kann mit dem Verbreitungsgebiet sakischer Kulturgruppen begründet werden und soll im Kapitel 5.3.3 näher erläutert

werden. Für Saken der Tian Shan-Region kann ein Genfluss von neolithischen Iranern nachgewiesen werden, was durch Vermischungsereignisse dieser südlichen Steppenomadern mit Vertretern des Bactria-Margina-Komplexes des heutigen Turkmenistans begründet werden kann (Damgaard et al. 2018).

Die Position der kasachischen Proben aus Ulžan in der PCA im Nahen Osten kann anhand der *outgroup-f3*-Statistik nicht eindeutig bestätigt werden. Hohe Standardfehler der F_3 -Werte machen eine genaue Zuordnung geteilter genetischer Drift nicht möglich. Die niedrigsten F_3 -Werte sind zwar zwischen Populationen Zentralasiens zu finden, die höchsten Werte treten jedoch für Weißrussland und Populationen Osteuropas auf. Auch phänotypisch zeigt die Probe an zwei Positionen erwähnenswerte Unterschiede zu allen anderen typisierten, eisenzeitlichen Proben. Für das Gen TCF7L2 konnte ein ancestral homozygoter Genotyp (T/T) bestimmt werden, welcher weltweit sehr selten und am ehesten in Europa (11,5 %) und süd-asiatischen Populationen nachzuweisen ist (8,8 %). Die den Saken zugeordnete Probe bildet auch bei beiden Markern, die in Verbindung mit Bluthochdruck und Zöliakie stehen, eine Ausnahme und zeigt an beiden Positionen den homozygoten Status (C/C) für das abgeleitete Allel. Dieser Genotyp kann für moderne ostasiatische Populationen nicht nachgewiesen werden, in Europa tritt er zu 21,3 % bzw. 21,7 % und in süd-asiatischen Populationen zu jeweils 1 % auf. Somit weist die Probe auf phänotypischer Seite einen deutlich europäischen und südasiatischen Bezug auf.

5.3.2 Homogenität der eisenzeitlichen Individuen Zentral – und Nord-Asiens zu den Individuen des Schwarzmeer-Raums und SW-Russlands

Hinsichtlich der Fragestellung einer Kontinuität zwischen westlichen und östlichen skythischen Populationen bestätigen die Ergebnisse der PCA mit modernen Populationen sowie der *outgroup-f3*-Statistik der vorliegenden Arbeit, dass auf genomischer Ebene zumindest kein exzessiver Genfluss zwischen Skythen der westlichen Steppe und der zentralasiatischen Steppe nachgewiesen werden kann (Damgaard et al. 2018). In der Studie von Damgaard und Kollegen wurden Skythen des Schwarzmeerraums, Vertreter der Tagar-Kultur Südsibiriens und Saken Zentralasiens und des Tian Shan analysiert. Wie im vorangegangenen Kapitel jedoch bereits erörtert, grenzen sich Saken und Vertreter der Tagar-Kultur deutlich von eisenzeitlichen Nomaden-Gruppen Zentralasiens ab. Mögliche geneti-

sche Strukturierung der Gruppen könnte in Bezug auf Skythen des NSR und Vertretern der eisenzeitlichen Pazyryk-Kultur abweichen. Die Abgrenzung der West-Skythen zu zentralasiatischen Populationen ist auch für die Skythen SW-Russlands erkennbar. Für einen Vertreter der initialen skythischen Phase durch Nov7 (8. – 6. Jh. v.Chr.) und ein Individuum der späten Phase aus Kolbino (3. Jh. v. Chr.) ist kein ost- bzw. zentralasiatischer Bezug erkennbar. Mit jeweils nur einzelnen Individuen kann auch von einem Stichproben-Artefakt ausgegangen werden. Die genetischen Differenzen werden jedoch auch durch genomische Daten von Skythen der west-pontischen Steppe bestätigt (Krzewinska et al. 2018). Einige Skythen der westlichen pontischen Steppe zeigen gar keine südasiatische oder osteurasische Komponente. Bemerkenswert ist hier, dass auch die Skythen der pontisch-kaspischen Steppe, also Vertreter der West-Skythen, keine homogene Gruppe bilden, sondern sich in vier Cluster unterteilen lassen mit unterschiedlichen Affinitäten nach Zentralasien. Ein Individuum der späten Skythenzeit gleicht jedoch eher den frühen östlichen Altai-Skythen und steht für eine mögliche Verbindung der West-Skythen mit Zentraleurasien (Krzewinska et al. 2018). Ein Bezug der nordpontischen Skythen zu Populationen des Altai-Gebirges konnte auch anhand mitochondrialer Haplotypen nachgewiesen werden (Juras et al. 2017). Die Mobilität aus dem Osten wird jedoch als sehr gering eingeschätzt. Als Hauptquelle der westlichen Nomaden werden die östlich-pontische Steppe und der südliche Ural angenommen. In der Eisenzeit wird vielmehr von einem Bevölkerungszug im Bereich des Minusinsker Becken aus südwestlicher Richtung ausgegangen und einer starken Bevölkerungskonzentration während der Skythenzeit. Der Reichtum des Altai-Gebirges zieht Skythen des Schwarzmeerraums sowie Perser an (Parzinger 2007).

Die genetische Distanz der westlichen eisenzeitlichen Skythen zu den eisenzeitlichen Populationen Zentral- und Nordasiens ist allerdings mit einem Wert von 0,0271 als sehr gering einzustufen. Mit Ausnahme der Allelfrequenzen des Markers SLC45A2 sind keine signifikanten Unterschiede in der Frequenzverteilung zwischen den östlichen und westlichen Skythen der Eisenzeit erkennbar. Aufgrund der geringen Größe der Stichprobe westlicher Skythen sollten diese Ergebnisse jedoch mit Vorsicht interpretiert werden.

Basierend auf der Diversität mitochondrialer Haplogruppen konnten ebenfalls keine signifikanten Unterschiede und nur geringe genetische Distanzen zwischen westlichen Skythen und skythischen Gruppierungen Zentralasiens festgestellt werden (Unterländer 2014). Hervorzuheben ist dabei, dass in der Publikation von Unterländer und Kollegen zwei Sarmaten stellvertretend für die westlichen Skythen herangezogen wurden.

Die geringen genetischen Distanzen, die sich auch auf nukleärer Ebene bestätigen, deuten, mit Ausnahme der Fundplätze Ulžan und Barsučij Log, somit auf einen kontinuierlichen Genfluss zwischen Sarmaten SW-Russlands mit östlichen Skythen Zentralasiens hin, jedoch mit starken Einflüssen lokaler Populationen.

Die genetischen Distanzen West- und Osteurasiens sind in modernen Populationen deutlich höher als in der Eisenzeit und sprechen dafür, dass die nomadische Steppebevölkerung eine hochmobile Gruppe darstellt.

Simulations-basierte Methoden unterstützen die Theorie eines konstanten Genflusses zwischen westlichen Sarmaten und östlichen Gruppierungen, der jedoch stärker von West nach Osteurasien angenommen werden kann (Unterländer et al. 2017).

5.3.3 Das Verhältnis zu neolithischen und bronzezeitlichen Populationen der eurasischen Steppe

Die genetische Identität eisenzeitlicher Reiternomaden des eurasischen Steppengürtels ist bis heute nicht geklärt. Die Nähe zu vorangegangenen prähistorischen Populationen der eurasischen Steppe kann hierbei einen Hinweis auf Migrationsbewegungen und Verwandtschaftsverhältnisse liefern. Die damit eng verbundene Frage nach dem möglichen Ursprungsgebiet skythischer Gruppen soll ausführlich im abschließenden Kapitel 5.3.5 diskutiert werden.

Über den zentralasiatischen Raum fanden beginnend mit dem Paläolithikum immer wieder Migrationsbewegungen statt (Palstra et al. 2015). Vermutlich zunächst aus östlicher Richtung und zu Beginn der Bronzezeit von West- nach Osteurasien (Li et al. 2010).

In der Bronzezeit fanden in der eurasischen Steppe zwischen 3000 - 1200 v. Chr. zwei entscheidende Migrationsbewegungen statt, die mit der Expansion der Jamnaja-Kultur in Verbindung stehen (Abb.36). Zum einen eine ost- und südostwärts gerichtete Verbreitung aus der westlichen Steppe, die eine Mischung aus EHG und kaukasischen Komponenten nach Zentralasien und Sibirien brachte und durch die genetische Ähnlichkeit mit Vertretern der Afanasievo-Kultur aus dem Altai belegt werden konnte (Allentoft et al. 2015). Im südsibirischen Altai erscheinen in darauffolgenden kulturellen Gruppierungen, wie der Karasuk-Kultur, zunehmend ost-asiatische Einflüsse. Diese ostwärts gerichtete Bewegung in die eurasische Steppe über Zentralasien nach Sibirien führte zu einer bis heute nachweisbaren Verbindung von Zentral-Europa zu Populationen der eurasischen Steppe (Haak et al.

2015), wo es anschließend nur zu einer geringen Vermischung mit ost-asiatischen Populationen gekommen ist (Allentoft et al. 2015).

Zu Beginn der Bronzezeit kann seitens der Archäologie sowie basierend auf genomischen Daten der Einfluss der Jamnaja-Kultur aus der pontisch-kaspischen Steppe auch in Europa belegt werden. Diese Jamnaja-Komponente ist dabei in Nordeuropa höher als in Südeuropa (Haak et al. 2015).

Das Auftreten westeurasischer Linien in Südasien wird vor und nach der Jamnaja-Kultur nachgewiesen und wird zwei unabhängigen Migrationsbewegungen zugeschrieben. Für das bronzezeitliche Anatolien konnte kein Nachweis von Steppe-Vorfahren gefunden werden. (Damgaard et al. 2018b).

Es stellt sich also die Frage, inwieweit die Steppe-Herkunft (Jones et al. 2015) für die einzelnen eisenzeitlichen Kulturgruppen der zentralasiatischen Populationen sichtbar ist. Kann ein exzessiver und anhaltender Genfluss zwischen westlichen und östlichen Gruppierungen bestätigt werden (Unterländer et al. 2017; Krzewinska et al. 2018)? Besteht eine nachweisbare Kontinuität zwischen bronzezeitlichen Populationen und lokalen eisenzeitlichen Individuen?

Der als Referenz verwendete prähistorische genomische Datensatz setzt sich aus überwiegend bronzezeitlichen Individuen der eurasischen Steppe, sowie europäischen und östlichen Jäger-Sammler-Populationen (JS) zusammen (Allentoft et al. 2015; Mathieson et al. 2015; Haak et al. 2015; Lazaridis et al. 2016). Die folgende Diskussion legt daher ebenfalls einen Schwerpunkt auf die Verbindung der eisenzeitlichen Skythen zu jenen Kulturgruppen.

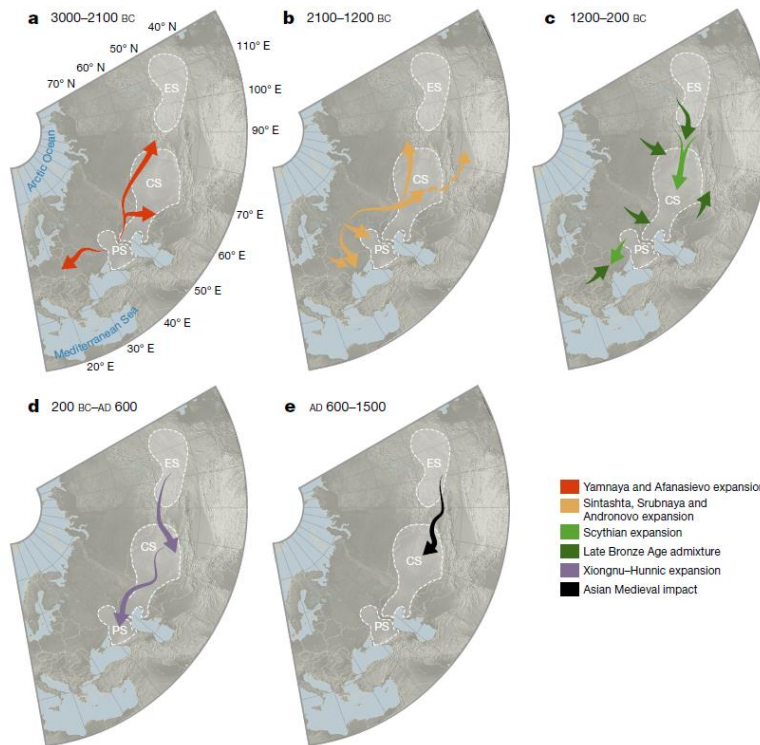


Abbildung 36 Schematische Darstellung der fünf großen Migrationsbewegungen der Pontischen Steppe (PS), zentralen Steppe (CS) sowie dem östlichen Steppenraum. a Expansion der Jamnaya- und Afanasievo-Kultur der frühen Bronzezeit b Spätbronzezeitliche Migration der Sintashta- und Andronovo-Kultur c Migrationsbewegungen der eisenzeitlichen Reiternomaden und Vermischungsereignisse in der späten Bronzezeit d und e Ausdehnung der Hunnen aus dem östlichen Steppenraum und asiatischer Einfluss im anschließenden Mittelalter; nach (Damgaard et al. 2018)

Betrachtet man die PCA (Abb. 27/36) der vorliegenden Proben mit prähistorischen Vergleichsgruppen (Lazaridis et al. 2014; Fu et al. 2014; Jones et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015; Jones et al. 2017; Sikora et al. 2017) (Lazaridis et al. 2016) fallen alle bronzezeitlichen Individuen, mit Ausnahme von Kama6, in die Verteilung europäischer und westeurasischer bronzezeitlicher Gruppierungen. Die drei eisenzeitlichen Proben aus Südsibirien (Allentoft et al. 2015) liegen im zentralasiatischen Cluster der Fundplätze Berel, Aržan, Justyd und Ak-Alacha. Die Vertreter der russischen Karasuk-Kultur (Allentoft et al. 2015) liegen im Verteilungsbereich der Proben des Minusinsker Beckens der Tagar-Kultur. Wie in den vorangegangenen Kapiteln bereits erwähnt, liegen die Proben deutlich westlicher als ihr Fundplatz in Sibirien.

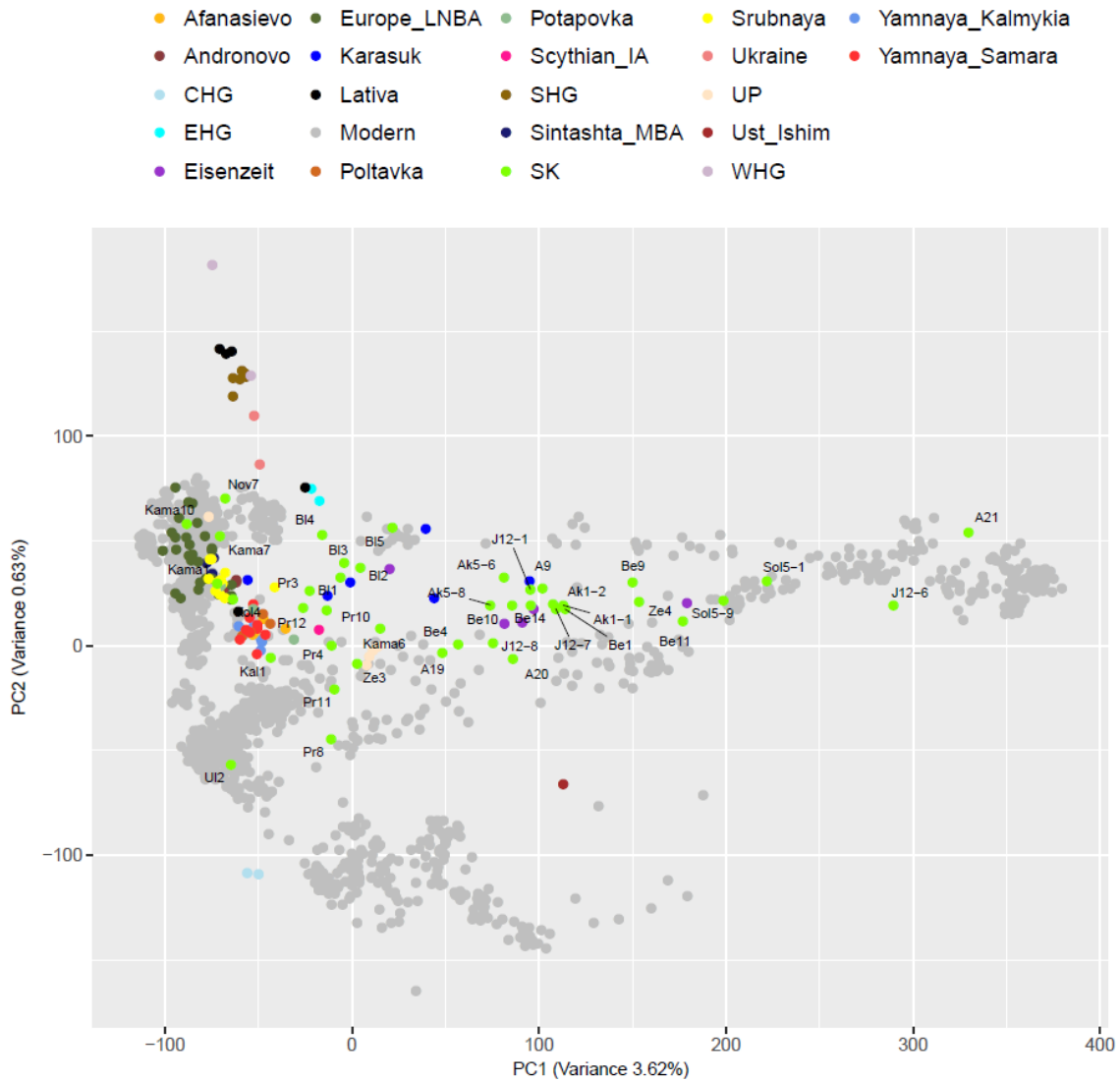


Abbildung 37 PCA der untersuchten Proben und zusätzlichen aDNA-Referenzen (Lazaridis et al. 2014; Fu et al. 2014; Jones et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015; Jones et al. 2017; Sikora et al. 2017). SK = Individuen der vorliegenden Arbeit, inklusive bronzezeitlicher und neolithischer Individuen und Sarmaten; Aufteilung der bronzezeitlichen Steppenpopulationen der Referenzen in ihre Untergruppen

Wie im Kapitel 5.3.1 bereits besprochen, stellen die skythischen Gruppen der Eisenzeit Zentralasiens keine homogene Gruppe dar. Die Historien der jeweiligen Kulturgruppen sollten unabhängig voneinander betrachtet werden. Im Folgenden werden deshalb der Sake, die Vertreter der Tagar-Kultur und eisenzeitliche, kasachische bzw. süd-sibirische Kulturgruppen gesondert diskutiert, ebenso die westlichen Skythen und frühen Sarmaten.

Bronzezeitliche Individuen

Die Sintaschta-Kultur verbreitete sich ausgehend vom Ural bis nach Asien und zeigt eine Affinität zu neolithischen Farmern Europas (Allentoft et al. 2015). Es konnte bereits ein gemeinsamer Ursprung der Sintaschta-Kultur mit den europäischen Schnurband-Keramikern festgestellt werden, was einen asiatischen Ursprung der Sintaschta-Kultur ausschließt. Die enge genetische Verwandtschaft zur Andronovo-Kultur, welche sich wiederum deutlich von Jamnaja und Afanasievo abgrenzt, weist auf eine Ostwärts-Bewegung von Vertretern der Sintaschta-Kultur hin.

Wird ein Baum der Form D(Pazyryk, Test, Sintaschta, AG) angewendet, teilen die bronzezeitlichen Individuen eine gemeinsame Abstammung mit allen Jäger-Sammler-Populationen, sowie den Individuen der Poltavka- und Jamnaja-Kultur. Dies zeigt, dass die bronzezeitliche Sintaschta-Kultur den lokalen, westeurasischen und europäischen, bronzezeitlichen Populationen sowie Jäger-Sammler-Gruppen (JS) näher ist, als skythischen Verbänden aus dem Osten in der Eisenzeit. Die Nähe bronzezeitlicher Populationen zu europäischen Jäger-Sammlern ist konsistent mit den Ergebnissen vorangegangener Publikationen (Allentoft et al. 2015; Mathieson et al. 2015).

Für östliche Skythen konnte bereits eine Nähe zur Srubnaja-Alakulskaya und Jamnaja-Kultur festgestellt werden (Unterländer et al. 2017). Das Individuum aus Kalinovka weist eher eine geteilte genetische Drift mit östlichen Jäger-Sammlern (EHG) anstatt mit Srubnaja auf.

Für die Individuen aus Kamennyj Ambar zeigt sich ebenfalls eine Nähe zu westlichen Jäger-Sammlern (WHG, SHG), aber auch zu Srubnaja, Afanasievo, Poltavka, Jamnaja-Samara anstatt mit Jäger-Sammlern des Kaukasus (CHG). Kama6 liegt außerhalb des Clusters der weiteren Proben des Fundplatzes und liegt im westlichen Teil des Verteilungsgebietes der eisenzeitlichen Skythen in der PCA. Dass Vertreter der Sintaschta-Kultur aus Kamennyj Ambar keine homogene Gruppe bilden und einzelne Individuen Komponenten Westsibirische-JS-Herkunft ebenso wie eine ostasiatische verwandte Abstammung zeigen, konnte ebenfalls in einer weiteren Studie von Narasimhan und Kollegen festgestellt werden (Narasimhan et al. 2018).

Während keine genetischen Distanzen anhand definierter mitochondrialer Haplogruppen zwischen allen skythischen Gruppen und der spätbronzezeitlichen Population Zentralasiens nachgewiesen werden konnten (Unterländer 2014), zeigen die östlichen Skythen der Pazyryk und Aldy Bel-Kultur als Ergebnis der D-Statistik eine größere Ähnlichkeit zueinander im Vergleich zu Vertretern der bronzezeitlichen Sintaschta-Kultur.

Westliche Skythen

In der Hauptkomponentenanalyse weisen die frühschythischen Individuen aus Novozavedennoe und der Skythe aus Kolbino eine deutliche Affinität zu bronzezeitlichen Steppenpopulationen auf. Ein ähnliches Bild zeigt sich für die West-Skythen in den Ergebnissen der D-Statistik mit der Form $D(\text{CHG}, \text{SHG}/\text{WHG}, \text{Kolbino}, \text{AG})$. Hier kann eher eine geteilte genetische Drift mit europäischen JS als mit JS des Kaukasus identifiziert werden.

Die Abgrenzung zum zentralasiatischen Raum wird des Weiteren in der geteilten gemeinsamen Herkunft mit SHG und nicht mit Vertretern der Pazyryk-Kultur deutlich. Dies kann zum einen auf einen unabhängigen Ursprung der westlichen Gruppen hindeuten oder unter der Annahme von Kontinuität zwischen der Bronze- und Eisenzeit für ein Ausbleiben eines starken ost-asiatischen Einflusses stehen.

Einen Genfluss zwischen europäischen Farmern und ungarischen Skythen konnte ebenfalls nachgewiesen werden. Dies deutete vielmehr auf eine Verwandtschaft mit spätbronzezeitlichen Viehzüchtern als mit der Jamnaja-Kultur der Bronzezeit hin (Damgaard et al. 2018). Skythische Populationen der pontisch-kaspischen Steppe teilen jedoch mehr genetische Drift mit Vertretern der Andronovo-, Afanasievo und Sintaschta-Kultur als mit Jamnaja, was einen zentral-asiatischen Ursprung nahe legen würde (Krzewinska et al. 2018).

Ostskythen

Aufgrund der ostwärts gerichteten, bronzezeitlichen Jamnaja-Bewegung tragen alle Steppenpopulationen Jäger-Sammler-Komponenten (Allentoft et al. 2015; Mathieson et al. 2015; Haak et al. 2015; Unterländer et al. 2017). Die eisenzeitlichen Skythen Zentralasiens zeigen eine kaukasische und osteuropäische JS-Sammler-Komponente, was für eine Nähe zu Jamnaja-verwandten Populationen spricht. Zusätzlich konnte eine Nähe zu ostasiatischen bzw. nordsibirischen Populationen nachgewiesen werden (Unterländer et al. 2017)

Eine eisenzeitliche Probe aus Turkmenistan (~800 v. Chr.) zeigt eine genetische Nähe zu Jäger-Sammler-Populationen des Kaukasus (CHG) ebenso wie zu europäischen Farmern, was typisch für die späte Bronzezeit ist (Narasimhan et al. 2018). Mögliche Vermischungsereignisse werden daher in die spätere Bronzezeit mit Vertretern der Jamnaja bzw. Afanasievo-Kultur datiert. Eine weitere eisenzeitliche Probe aus Anatolien (~600 v. Chr.) liegt in der PCA mit ausgewählten modernen Populationen Eurasiens und prähistorischen Steppenpopulationen im Bereich der frühen bis mittleren Individuen der Bronzezeit (Anatolia EBA/MLBA) sowie JS des Kaukasus. Zusammenfassend sprechen sich die Autoren für

einen geringen genetischen Einfluss früher Jamnaja-Viehzüchter auf Anatolien sowie Zentral- und Südasien in der Bronzezeit aus.

Ein signifikanter Anteil an europäischen und kaukasischen JS-Komponenten ist für alle Steppenpopulationen der hier analysierten prähistorischen Vergleichsgruppen (Lazaridis et al. 2016) von der frühen Bronzezeit bis zur Eisenzeit nachweisbar. (Unterländer et al. 2017). Das Auftreten dieser JS-Komponenten steht ebenfalls in Verbindung mit einer ostwärts gerichteten Migrationsbewegung der Jamnaja-Kultur über die eurasische Steppe bis nach Sibirien, die ein genetisches Substrat der Jäger-Sammler-Populationen mit sich trugen (Haak et al. 2015).

Es besteht ebenfalls eine deutliche Verbindung östlicher Skythen zu Vertretern der Andronovo-Kultur der westsibirischen Waldsteppe, was die These einer Ostwärts-Bewegung in der mittleren Bronzezeit stützt (Unterländer et al. 2017).

Die eisenzeitliche skythische Probe aus (Mathieson et al. 2015) liegt in einer Hauptkomponentenanalyse mit 777 modernen Individuen Westeurasiens und 221 prähistorischen Individuen ebenfalls bei den bronzezeitlichen Steppenpopulationen (Abb. 36).

Die Ergebnisse der *outgroup-f₃*-Analyse der eisenzeitlichen Individuen aus der vorliegenden Arbeit weisen alle eine geteilte genetische Drift zu östlichen JS-Populationen auf.

Die Nähe zu äneolithischen Steppe-Individuen ist auch für alle Ost-Skythen, hier inklusive der Individuen der Tagar-Kultur, festzustellen.

Vertreter der Tagar-Kultur

Für die Tagar-Kultur Südsibiriens ist eine erhöhte EHG-Verwandtschaft und der höchste Anteil an einer Abstammungskomponente bronzezeitlicher Viehzüchter im Vergleich zu weiteren innerasiatischen Skythen festgestellt worden (Damgaard et al. 2018). Auch in der vorliegenden Arbeit kann die Nähe zu östlichen JS (EHG) für Individuen aus Barsučij bestätigt werden. Zudem ist festzustellen, dass die Individuen der Tagar-Kultur in der PCA im Verbreitungsgebiet der Karasuk-Kultur liegen, aus welcher die Tagar-Kultur hervorgegangen ist (Parzinger 2006). Die Nähe zu Karasuk-Kultur sowie die deutliche Abgrenzung zu allen anderen eisenzeitlichen zentral-asiatischen Kulturgruppen lassen auf eine mögliche Ostwärts-Bewegung der Tagar-Kultur mit geringem Genfluss lokaler Populationen schließen. (Damgaard et al. 2018) Die Sonderstellung der Tagar-Kultur im Vergleich zu den östlichen Skythen kann auch auf mitochondrialer Ebene detektiert werden (Unterländer 2014). Die Karasuk-Kultur zeigt zudem eine Nähe zu skythischen Proben des nord-pontischen Raums (Juras et al. 2017).

Die Proben der Tagar-Gruppe aus Barsučij Log weisen bei einer D-Statistik der Form $D(\text{Pazyryk, Test, Tagar, AG})$ eine geteilte, genetische Drift mit EHG und SHG sowie Vertretern der Jamnaja-Samara-Gruppe auf und bestätigen die Ergebnisse der Hauptkomponenten-Analyse und somit eine genetische Prägung von Populationen der westeurasischen Steppe.

Der Sake

Für das mit der sakischen Kultur assoziierte Individuum aus Ulžan (U12) ist feststellen, dass es sich deutlich von allen anderen östlichen Gruppierungen abzugrenzen scheint und ein geringer Genfluss zu den östlichen Skythen bestanden haben muss.

In der D-Statistik der Form $D(\text{Skythen-A, Test, Skythen-C, AG})$ kann für die Pazyryk-Kultur eine größere Ähnlichkeit zu den beiden neolithischen Proben und zur Aldy-Bel-Kultur, als zur sakischen Probe nachgewiesen werden.

Das Verbreitungsgebiet sakischer Stämme reichte bis nach Nord-Pakistan (Parzinger 2007). Viele sakische Kurgane liegen im Pamir-Gebirge des heutigen Tadschikistan und verweisen basierend auf archäologischem Fundgut ins Siebenstromland und nach Südsibirien. Eine genetische Nähe zu sibirischen JS von Saken der Tian Shan-Region konnten Damgaard und Kollegen nachweisen (Damgaard et al. 2018), ebenso eine genetische Komponente die mit neolithischen Farmern Irans assoziiert werden kann. Die Autoren schließen daher auf einen Genfluss zwischen Saken des Tian Shan-Gebietes und Vertretern der Oasenkultur (Bactria-Margina-Komplexes) des heutigen östlichen Turkmenistans.

Die in dieser Studie analysierten Saken der zentralen Steppe und der Region um das Tian Shan-Gebirge bilden zwei genetisch differenzierte Gruppen aus mit unterschiedlichen Einflüssen bronzezeitlicher Steppenpopulationen. Ein nordwärts gerichteter Genfluss aus Usbekistan, Turkmenistan, Tadschikistan auf zentrale Steppengebiete konnte bereits für die Bronzezeit belegt werden (Narasimhan et al. 2018).

Die Sarmaten

Aus der sauromatischen Kultur entsteht die frühsarmatische Kultur, wobei mögliche Einflüsse aus dem Osten dabei nicht auszuschließen sind. Die Sarmaten fallen zwischen 3. und 2. Jh. v. Chr. in die nordpontische Steppe ein und mischen sich mit den dort ansässigen Skythen (Parzinger 2007). Im südlichen Ural am östlichen Rand der pontisch-kaspischen Steppe kann von einer genetischen Kontinuität der Sarmaten über eine Zeit von 300-500 Jahren ausgegangen werden (Krzewinska et al. 2018). Interessanterweise zeigen die in der zuletzt genannten Studie analysierten späten Sarmaten (4 -323 n. Chr.) eine geteilte geni-

sche Drift mit bronzezeitlichen Steppenpopulationen und nicht mit Kimmerern des Nordschwarzmeerraums. In einer Hauptkomponentenanalyse ist eine Nähe zu pontisch-kaspischen Skythen erkennbar, welche sich in einem Cluster mit Bezug zur Steppe gruppieren. Die frühen Sarmaten dieser Studie grenzen sich jedoch deutlich von zentralasiatischen, eisenzeitlichen Skythen ab (Unterländer et al. 2017; Krzewinska et al. 2018).

Dass eine geteilte Herkunft früher Sarmaten aus Südwest-Russland mit Vertretern der bronzezeitlichen Sintashta-Kultur östlich des Urals wahrscheinlicher ist, als mit der Pazyryk-Kultur des Altai, kann in der D-Statistik ebenfalls gezeigt werden. Die Nähe zu EHG zeigt sich für die frühen Sarmaten aus Pokrovka, ebenso wie eine genetische Ähnlichkeit zur Probe aus Loschbour, als Vertreter der westlichen JS (WHG).

Die frühen Sarmaten orientieren sich in der PCA deutlich nach Süd-Osten und liegen im Verteilungsbereich der bronzezeitlichen Jamnaja-Individuen des Referenzdatensatzes. Die genetische Ähnlichkeit zur bronzezeitlichen Jamnaja-Kultur und die bronzezeitliche Poltavka-Kultur konnte auch von Unterländer und Kollegen für die frühen Sarmaten herausgestellt werden (Unterländer et al. 2017).

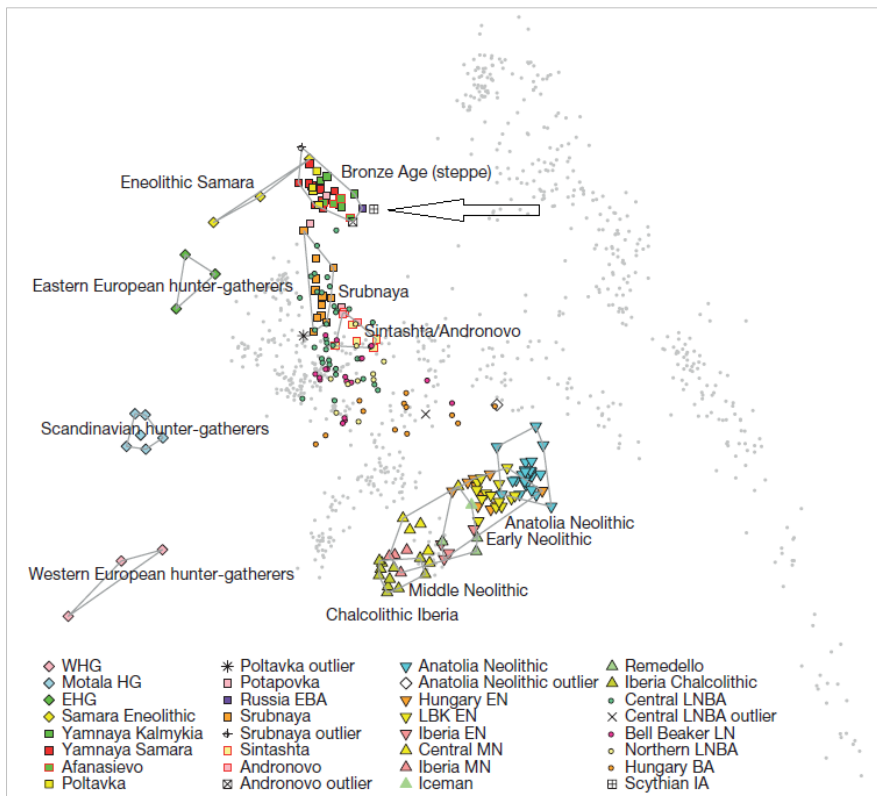


Abbildung 36: Hauptkomponentenanalyse mit 777 modernen Individuen Westeurasiens (grau), Projektion von 221 prähistorischen Individuen nach ihren entsprechenden Kulturstufen gruppiert; E/M/LN, Frühes/Mittleres/Spätes Neolithikum; LBK, Linearbandkeramik; E/WHG, Östlich - /Westliche Jäger-Sammler; EBA frühe Bronzezeit; IA, Eisenzeit; LNBA, Spätes Neolithikum und Bronzezeit (Mathieson et al. 2015).

5.3.4 Ostasiatischer Einfluss auf nomadische Populationen im Übergang von der Bronze- zur Eisenzeit

Drei wesentliche Fragen sind mit dem möglichen Anstieg ostasiatischer Linien in Westeurasiens verbunden: Besteht zwischen westlichen und östlichen nomadischen Populationen eine gemeinsame Herkunft mit anschließender *isolation by distance*? Oder ist von einem kontinuierlichen Genfluss mit ostasiatischen Populationen, also Migrationsbewegungen osteurasischer Populationen mit anschließender Vermischung mit lokalen Populationen auszugehen? Kann die These eines multiregionalen Ursprungs (Unterländer et al. 2017) gestützt werden? Der Anstieg ostasiatischer Komponenten im zentralasiatischen Raum lässt auf Vermischungsereignisse mit lokalen Populationen bzw. Migrationsbewegungen mit anschließendem Genfluss schließen.

Bisherige Studien belegen, dass mit Beginn der Eisenzeit ein Zustrom ostasiatischer Linien nachweisbar ist. Fasst man die Ergebnisse aus der Literatur betreffend der maternalen Historie zusammen, konnten ebenfalls ostasiatische Haplogruppen in der Bronzezeit für die Mongolei nachgewiesen werden (González-Ruiz et al. 2012); wohingegen kasachische Individuen und Proben aus Chakassien der späten Bronzezeit nur europäische Linien aufweisen (Lalueza-Fox et al. 2004; Keyser et al. 2009). In der Steppe West-Sibiriens zeigten untersuchte Proben von der frühen bis zur späten Bronzezeit bereits 60 % ostasiatische Linien (Molodin et al. 2012). Im nord-pontischen Raum des Äneolithikum und der Bronzezeit wurden ausschließlich westeurasische Linien nachgewiesen (Wilde et al. 2014).

Für dieselbe Region (Moldawien und Ukraine) weisen Juras und Kollegen in der Eisenzeit eine Mischung aus west- und osteurasischen Haplogruppen nach (Juras et al. 2017). Der Anteil ostasiatischer Linien liegt hier bei 26,3 %. Resultierend aus der Zusammensetzung der Haplogruppen (A und H8c) postulieren die Autoren eine Verbindung der West-Skythen des nord-pontischen Raums mit dem Altai und definieren dieses Gebiet als möglichen Ursprung des ost-asiatischen Einflusses.

Basierend auf mitochondrialen Daten der hier analysierten Skythen aus der Eisenzeit lassen sich 26,3 % osteurasische Linien in westlichen Skythen (inklusive Sarmaten) feststellen (Unterländer 2014). Für die östlichen Skythen konnten west- und osteurasische Linien in einem nahezu ausgeglichenen Verhältnis nachgewiesen werden. Allerdings zeigen die Proben der Fundplätze Novozavedennoe und Kolbino aus SW-Russlands ausschließlich westeurasische Linien. Da keine ostasiatische Komponente in westlichen, frühen und späten Skythen nachweisbar ist (Nov7, Kol4) (Unterländer 2014), jedoch für ost-kasachische Skythen eines vergleichbaren Zeitraums (Zevakino) kann, basierend auf den Ergebnissen dieser beiden Fundplätze, von einer *Isolation by distance* mit geringem Genfluss bis zur späten Eisenzeit in SW-Russland ausgegangen werden. Jedoch sollte aufgrund der geringen Probenzahl nur mit Vorsicht geschlussfolgert werden.

Auf phänotypischer Ebene entspricht die bronzezeitliche Probe aus Kalinovka, mit Ausnahme des Laktase-Gens (LCTa) und dem Pigmentierungsmarker SLC45A2, dem europäischen Phänotyp und damit den westlichen Skythen der Eisenzeit aus SW-Russland.

Anhand der *outgroup- β* -Statistik mit modernen Populationen konnte für die westlichen Proben der Bronze- (Kal1) und Eisenzeit (Nov7, Kol4) jeweils keine genetische Drift mit zentral-asiatischen bzw. Populationen aus Ost-Asien festgestellt werden.

Die frühen Sarmaten weisen mit 18,2 % einen leicht erhöhten Anteil asiatischer Linien auf (Unterländer 2014). Sie zeigen ostasiatische Haplogruppen, die vorab nicht in der westlichen eurasischen Steppe bestimmt werden konnten, was auf einen ständigen Genfluss aus

dem Osten hindeuten könnte, der in der Eisenzeit höher war als dies für heutige Populationen (8%) desselben Gebietes der Fall ist (Malyarchuk et al. 2002).

Neueste Studien gehen davon aus, dass östliche Linien in der Eisenzeit mit den Kimmerern in die pontisch-kaspische Steppe kamen (Krzewinska et al. 2018).

Die archäologischen Funde der Kimmerer und Skythen weisen eine große Übereinstimmung auf, Einflüsse aus dem ost-asiatischen Raum sind auf archäologischer Seite jedoch nicht ganz auszuschließen (Parzinger 2007). Für Kimmerer die zwischen ~ 1000 – 800 v. Chr. datieren, konnte eine größere Übereinstimmung mit west-sibirischen bronzezeitlichen Populationen belegt werden als zu Vertretern der Srubnaja-Kultur aus der pontisch-kaspischen Steppe. Schon im Übergang der Bronze- zur Eisenzeit kann für die Kimmerer des Schwarzmeerraums ebenfalls eine Nähe zu heutigen süd-asiatischen und nordost-asiatischen Populationen belegt werden, die mit dem Alter der Individuen zunimmt (Krzewinska et al. 2018).

Der nachweislich geringe Anteil ostasiatischer Linien auf mitochondrialer Ebene und leichte Ausdehnung der Sarmaten in den Nahen Osten bzw. den westlichen Teil der eurasischen Steppe unterstützen die Annahme, dass die Region südlich des Urals als Ort des kontinuierlichen Genflusses und Bindeglied der pontisch-kaspischen und kasachischen Steppe angesehen werden kann.

Studien basierend auf genomischen Daten stützen die Annahme, dass ein ost-asiatischer Einfluss bereits in der mittleren bis frühen Bronzezeit in der zentralen Steppe (Kasachstan) und dem Altai stattgefunden hat (Damgaard et al. 2018b).

Ein ostasiatischer Bezug kann ebenfalls für drei Individuen aus West-Sibirien (6200-4000 BCE) nachgewiesen werden. Diese repräsentieren eine neue Abstammungsmischung, die *West-Siberian-HG*-Komponente (Narasimhan et al. 2018), die auch in Usbekistan, Turkmenistan, Tadschikistan und der südlichen Steppe detektiert wurde. Die Komponente aus 30 % EHG, 50 % ANE (*ancient north eurasiens*), 20 % heutigem Ostasien soll eine genetische relative homogene Gruppe darstellen, die sich zwischen 2000-1400 BCE zwischen Osteuropa und trans-uralischer Steppe verbreitet hat. In der kasachischen Steppe und dem Minusinsker Becken entsteht während der mittleren bis späten Bronzezeit ein Cluster, genannt *Steppe-MLBA-East* aus einer Vermischung von Viehhirten der westlichen Steppe und Vorfahren dieser früheren *West-Siberian HG*. Für Individuen der an den Kaukasus angrenzenden Steppe Zone (Wang et al. 2018) kann zusätzlich eine mit anatolischen Ackerbauern verwandte Abstammung nachgewiesen werden.

Das Auftreten einer „östlichen Komponente“, welche hauptsächlich in Nganasanen aus Sibirien repräsentiert ist, konnte für eisenzeitliche Proben Zentralasiens der Aldy-Bel, Pazyryk, Zevakino-Čilikta-Kultur und in westlichen, frühen Sarmaten festgestellt werden (Unterländer et al. 2017). Zusätzlich zeigen die Skythen der Eisenzeit eine weitere Komponente, die in ost-asiatischen Populationen dominant ist. Während die genetische Verbindung nach Sibirien auch in bronzezeitlichen Populationen der Karasuk- und Mezhovskaya-Kultur präsent ist, konnte ein ost-asiatischer Einfluss ausschließlich für eisenzeitlichen Populationen detektiert werden. Beide Komponenten sind im Osten der eurasischen Steppe signifikant höher. Dieses für die skythische Periode typische Muster aus *admixture* mit ost-asiatischem Bezug zeigen ebenfalls Individuen der kasachischen Steppe der mittleren Bronzezeit in einer späteren Studie (Narasimhan et al. 2018).

Zusammenfassend lässt sich feststellen, dass ein östlicher Einfluss von der Bronze- zur Eisenzeit deutlich zunimmt, je weiter östlich sich die Proben in der eurasischen Steppe befinden. Für die zentralasiatische Steppe ist bereits in der mittleren und späten Bronzezeit ein Bezug zu Ostasien erkennbar.

Für die frühen Sarmaten lässt sich ein zentral-asiatischer Einfluss bzw. eine osteurasische Komponente feststellen. Dies spricht für einen kontinuierlichen Genfluss durch Expansion zentral-asiatischer Populationen oder für Migrationsbewegungen aus der östlichen Steppe bis nach SW-Russland.

5.3.5 Multiregionaler Ursprung eisenzeitlicher Steppenomadern

Das Verbreitungsgebiet der mit skythischer Kultur assoziierten Populationen reichte vom nord-pontischen Raum im Westen bis in den Osten der eurasischen Steppe. Seit Anbeginn der Forschung über die Skythenzeit stellt sich die Frage nach deren Ursprung und nach der Verbindung dieser einzelnen Reiternomadenvölker über eine beeindruckende Distanz hinweg. Über die Zeit wurde ein kaukasischer (Parzinger 2006), ostasiatischer und multiregionaler (Unterländer et al. 2017; Damgaard et al. 2018) Ursprung diskutiert. Eine kontinuierliche Verbindung der Schwarzmeer-Skythen zum Osten der eurasischen Steppe ist aus archäologischer Sicht eindeutig (Parzinger 2007). Die Möglichkeit eines polyzentristischen Ursprungs skythisch geprägter Kulturen mit einem ständigen kulturellen Austausch der Steppenpopulationen wurde aus archäologischer Sicht ebenfalls nie ganz ausgeschlossen (Bashilov, V.A. & Yablonsky 2000).

Neueste Studien anhand voll-genomischer Datensätze legen einen multiregionalen Ursprung skythischer Reiterverbände mit anhaltendem, aber geringem Genfluss nahe (Unterländer et al. 2017; Krzewinska et al. 2018). Damgaard und Kollegen belegen mit ihrem Datensatz die Hypothese eines multiregionalen Ursprungs, können jedoch einen kontinuierlichen Genfluss nicht bestätigen (Damgaard et al. 2018).

Es stellt sich also die Frage, ob ein multiregionaler Ursprung der skythischen Populationen mit dem vorliegenden Datensatz haltbar ist. Und hat dabei nur ein kultureller Austausch oder genetischer Genfluss zwischen skythischen Subkulturen stattgefunden?

Grundlegend für die Frage nach einem Ursprung der Skythen ist, inwieweit die einzelnen Kulturstufen selbst als homogene Gruppe betrachtet werden können. Basierend auf den Ergebnissen der vorliegenden Arbeit und wie bereits in vorangegangenen Kapiteln diskutiert, können die zeitgleich existierenden, skythischen Kulturgruppen des kasachischen Altai (Berel), des Ukok-Plateaus (Ak-Alacha) und des Chuya-Tals (Justyd) nicht als einheitliche Gruppierung betrachtet werden. Wie auch schon basierend auf mitochondrialen Daten festgestellt wurde, besteht eine größere genetische Distanz zwischen Individuen des Chuya-Tals und den geografisch naheliegenden Vertretern der Pazyryk-Kultur. Dies kann in der vorliegenden Arbeit durch nachweisbaren Bezug nach Ost-Asien der Probe J12-6 bestätigt werden. Die Nähe zu modernen ostasiatischen Populationen kann im vorliegenden Datensatz ebenfalls für einen deutlich älteren (700-600 v. Chr.) Vertreter der Aldy-Bel-Kultur aus Aržan nachgewiesen werden.

Die chronologisch folgende Tagar-Kultur des Minusinsker Beckens und der Sake aus Südostkasachstan grenzen sich genetisch voneinander wie auch von Vertretern der folgenden Pazyryk-Kultur ab.

Eine Feinstrukturierung innerasiatischer skythischer Populationen konnte auch für südsibirische Tagar, Saken der zentralen Steppe und des Tian Shan belegt werden (Damgaard et al. 2018). Weiterführende Analysen zeigen Vermischungsereignisse dieser drei Gruppen mit bronzezeitlichen Viehhirten und südsibirischen JS zu unterschiedlichen Anteilen. Die Autoren schlussfolgern demnach, dass kein intensiver Genfluss zwischen den skythischen Gruppen stattgefunden haben kann, jedoch Vermischungsereignisse mit entsprechenden lokalen Populationen.

In der vorliegenden Arbeit zeigen die West-Skythen eine Nähe zu vorangegangenen bronzezeitlichen Gruppen. Weder für den Vertreter der frühskythischen Kultur (800-600 v. Chr.) des Nord-Kaukasus noch für einen Vertreter der späten skythischen Periode SW-Russlands (300 v. Chr.) können osteurasische Linien (Unterländer 2014) oder genetische

Drift mit zentral- und/oder ost-asiatischen Populationen nachgewiesen werden. Dies spricht ebenfalls für eine genetische Kontinuität der skythischen Gruppierungen westlich des Urals mit möglichem Genfluss mit lokalen Populationen. Eine klare Abgrenzung ungarischer Skythen zu innerasiatischen, skythischen Gruppen zeigt sich durch einen nachgewiesenen Genfluss mit europäischen Ackerbauern. Die Saken teilen jedoch eine genetische Drift mit bronzezeitlichen Viehhirten und süd-sibirischen JS (Damgaard et al. 2018). Die etwas weiter östlich gelegenen Skythen des Nordschwarzmeerraums aus Moldawien und der Ukraine bilden ebenfalls keine homogene Gruppe. Die untersuchten Individuen teilen sich in drei Cluster auf, basierend auf ihren mitochondrialen Haplogruppen, die zum einen mit europäischen JS und früh-neolithische Farmern in Verbindung gebracht werden oder osteurasische Haplotypen repräsentieren. Es konnte ebenfalls eine genetische Affinität für alle Individuen mit der bronzezeitlichen Srubnaja-Kultur nachgewiesen werden. Krzewinska und Kollegen postulieren einen zentral-asiatischen Ursprung westlicher Skythen, da für die Individuen der pontisch-kaspischen Steppe eine gemeinsame genetische Drift mit der bronzezeitlichen Andronovo-, Afanasievo und Sintaschta-Kultur nachgewiesen werden konnte.

Für die frühen Sarmaten ist eine zentral-asiatische Verwandtschaft basierend auf den Positionierungen der Proben in der PCA anzunehmen. Dies ist durch die Evidenz einer sibirischen Komponente in den frühen Sarmaten aus Pokrovka belegt (Unterländer et al. 2017). Trotz geringen bis keinen nachweisbaren, genetischen Distanzen konnten Fst-Werte > 0 für Individuen der Pazyryk-Kultur zu den frühen Sarmaten (0,0093) sowie allen bronzezeitlichen Proben festgestellt werden (0,0180). Es wird von einer Westwärts-Bewegung der Sarmaten bis in den Nordschwarzmeerraum ausgegangen, mit einem anschließenden Zeitraum der Koexistenz von Skythen und Sarmaten, die dann die Vorherrschaft der Skythen mit der Zeit ablösten.

Die Nähe zu Jamnaja-verwandten Gruppierungen und der Nachweis einer nord-sibirischen Komponente lassen die Hypothese eines unabhängigen, multiregionalen Ursprungs nur unter der Annahme eines ständigen Genflusses über die eurasische Steppe zu.

Es kann außerdem eine große genetische Affinität zwischen der Schnurkeramik-Kultur Nordost-Europas und der Sintaschta-Kultur gezeigt werden (Allentoft et al. 2015), was gegen einen Ursprung der bronzezeitlichen Sintaschta in Zentralasien und dem Nahen Osten spricht. Dies könnte für eine mögliche, unabhängige Vorgängerpopulation der westlichen eisenzeitlichen Skythen sprechen.

Im Kapitel 5.3.3 wurde bereits die Verbindung zu östlichen Jäger-Sammler-Populationen für die vorliegenden Proben sowie eisenzeitliche Skythen aus der Literatur diskutiert (Mathieson et al. 2015; Unterländer et al. 2017) Dieses verbindenden Element zwischen den frühen Sarmaten und östlichen, skythischen Gruppierungen der Bronze – und Eisenzeit wird durch die Verbreitung der Jamnaja-Kultur aus der europäischen Steppe in den osteurasischen Raum transportiert (Haak et al. 2015).

Auch anhand des Datensatzes dieser Arbeit kann bestätigt werden, wie komplex die Migrationsbewegungen innerhalb des eurasischen Steppengürtels gewesen sein müssen. Skythen der westlichen Gebiete des Steppenraums weisen klare Vermischungsereignisse mit lokalen Vorgängerpopulationen auf. Ein Einfluss zentralasiatischer Gruppen ist für einzelne Individuen der westlichen Skythen nicht nachweisbar, es können jedoch für benachbarte Skythen osteurasische Komponenten detektiert werden. Die Verbindung zum östlichen Teil des Steppengürtels kann nicht ausgeschlossen werden. Für einzelne skythische Kulturgruppen Zentralasiens zeigen sich eine teilweise unabhängige Historie und kaum Vermischung mit angrenzenden Populationen.

Literaturverzeichnis

- 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA et al. (2015) A global reference for human genetic variation. *Nature* 526:68–74. doi: 10.1038/nature15393
- Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L et al. (2015) Population genomics of Bronze Age Eurasia. *Nature* 522:167–172. doi: 10.1038/nature14507
- Arbiza L, Zhong E and Keinan A (2012) NRE: a tool for exploring neutral loci in the human genome. *BMC Bioinformatics* 13:301. doi: 10.1186/1471-2105-13-301
- Aronesty E (2013) “Comparison of Sequencing Utility Programs.” *Open Bioinforma J.* doi: 10.2174/1875036201307010001
- Balam-Ortiz E, Esquivel-Villarreal A, Huerta-Hernandez D, Fernandez-Lopez JC, Alfaro-Ruiz L, Muñoz-Monroy O, Gutierrez R, Figueroa-Genis E, Carrillo K, Elizalde A et al. (2012) Hypercontrols in Genotype-Phenotype Analysis Reveal Ancestral Haplotypes Associated With Essential Hypertension. *Hypertension* 59:847–853. doi: 10.1161/HYPERTENSIONAHA.111.176453
- Bashilov, V.A. & Yablonsky LT (2000) Kurgans, Ritual Sites, and Settlements Eurasian Bronze and Iron Age. In: *In: Kurgans, Ritual Sites, and Settlements - Eurasian Bronze and Iron Age*, Davis-Kimball, J.M., E. M.; Koryakova, L.; Yablonsky, T. Archaeopress, Oxford
- Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE and Hirschhorn JN (2004) Genetic Signatures of Strong Recent Positive Selection at the Lactase Gene. *Am J Hum Genet* 74:1111–1120. doi: 10.1086/421051
- Bhatia G, Patterson N, Sankararaman S and Price a. L (2013) Estimating and interpreting F_{ST} : The impact of rare variants. *Genome Res* 23:1514–1521. doi: 10.1101/gr.154831.113
- Blumenstiel B, Cibulskis K, Fisher S, DeFelice M, Barry A, Fennell T, Abreu J, Minie B, Costello M, Young G et al. (2010) Targeted exon sequencing by in-solution hybrid selection. *Curr Protoc Hum Genet Chapter 18:Unit 18.4.* doi: 10.1002/0471142905.hg1804s66
- Bollongino R, Nehlich O, Richards MP, Orschiedt J, Thomas MG, Sell C, Fajkosova Z, Powell A and Burger J (2013) 2000 Years of Parallel Societies in Stone Age Central Europe. *Science* (80-) 342:479–481. doi: 10.1126/science.1245049
- Bouakaze C, Keyser C, Crubézy E, Montagnon D and 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. doi: 10.1007/s00414-009-0348-5
- Breese MR and Liu Y (2013) NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics* 29:494–496. doi: 10.1093/bioinformatics/bts731
- Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z et al. (2009a) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325:318–21. doi: 10.1126/science.1174462
- Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U and Pääbo S (2009b) Primer extension capture: targeted sequence retrieval from heavily degraded DNA sources.

J Vis Exp 1573. doi: 10.3791/1573

Bryk J, Hardouin E, Pugach I, Hughes D, Strotmann R, Stoneking M and Myles S (2008) Positive selection in East Asians for an EDAR allele that enhances NF- κ B activation. *PLoS One* 3:2–7. doi: 10.1371/journal.pone.0002209

Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, Good JM, Maricic T, Johnson PLF, Xuan Z et al. Targeted Investigation of the. 723–725.

Burger J, Kirchner M, Bramanti B, Haak W and Thomas MG (2007) Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc Natl Acad Sci U S A* 104:3736–41. doi: 10.1073/pnas.0607187104

Busby GB, Hellenthal G, Montinaro F, Tofanelli S, Bulayeva K, Rudan I, Zemunik T, Hayward C, Toncheva D, Karachanak-Yankova S et al. (2015) The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. *Curr Biol* 25:2518–2526. doi: 10.1016/j.cub.2015.08.007

Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, Rasmussen M, Gravel S, Guillén S, Nekhrizov G et al. (2013) Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am J Hum Genet* 93:852–64. doi: 10.1016/j.ajhg.2013.10.002

Chen R, Corona E, Sikora M, Dudley JT, Morgan AA, Moreno-Estrada A, Nilsen GB, Ruau D, Lincoln SE, Bustamante CD et al. (2012) Type 2 Diabetes Risk Alleles Demonstrate Extreme Directional Differentiation among Human Populations, Compared to Other Diseases. *PLOS Genet* 8:e1002621.

Chikisheva T, A. Gubina M, V. Kulikov I, Karafet T, Voevoda M and G. Romaschenko A (2007) A paleogenetic study of the prehistoric populations of the Altai. *Archaeol Ethnol Anthr Eurasia*. doi: 10.1134/S156301100704012

Clisson I, Keyser C, Francfort H-P, Crubezy E, Samashev Z and 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. doi: 10.1007/s00414-002-0295-x

Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F and Bertranpetit J (2004) Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur J Hum Genet* 12:495–504. doi: 10.1038/sj.ejhg.5201160

Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T, Hunter DJ and Hu FB (2009) Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. *Ann Intern Med* 150:541–550. doi: 10.7326/0003-4819-150-8-200904210-00008

Damgaard P de B, Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliussen T, Moreno-Mayar JV, Pedersen MW, Goldberg A, Usmanova E et al. (2018) 137 ancient human genomes from across the Eurasian steppes. *Nature* 557:369–374. doi: 10.1038/s41586-018-0094-2

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. doi: 10.1093/bioinformatics/btr330

de Barros Damgaard P, Martiniano R, Kamm J, Moreno-Mayar JV, Kroonen G, Peyrot M, Barjamovic G, Rasmussen S, Zacho C, Baimukhanov N et al. (2018) The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* (80-) 360:eaar7711. doi: 10.1126/science.aar7711

- Depristo MA, Banks E, Poplin R, Garimella K V., Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–501. doi: 10.1038/ng.806
- Donnelly MP, Paschou P, Grigorenko E, Gurwitz D, Barta C, Lu R-B, Zhukova O V, Kim J-J, Siniscalco M, New M et al. (2012) A global view of the OCA2-HERC2 region and pigmentation. *Hum Genet* 131:683—696. doi: 10.1007/s00439-011-1110-x
- Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L and Järvelä I (2002) Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 30:233—237. doi: 10.1038/ng826
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC and Foll M (2013) Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet*. doi: 10.1371/journal.pgen.1003905
- Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev A a., Johnson PLF, Aximu-Petri A, Prüfer K, de Filippo C et al. (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514:445–449. doi: 10.1038/nature13810
- Fu Q, Meyer M, Gao X, Stenzel U, Burbano H a, Kelso J and Pääbo S (2013) DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci U S A* 110:2223–7. doi: 10.1073/pnas.1221359110
- Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, Domboróczki L, Kóvári I, Pap I, Anders A et al. (2014) Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun* 5:5257. doi: 10.1038/ncomms6257
- Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C et al. (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27:182–9. doi: 10.1038/nbt.1523
- González-Ruiz M, Santos C, Jordana X, Simón M, Lalueza-Fox C, Gigli E, Aluja MP and Malgosa A (2012) Tracing the Origin of the East-West Population Admixture in the Altai Region (Central Asia). *PLoS One* 7:e48904.
- Gräfen A (2009) Development of a multiplex PCR for the assessment of ancient DNA preservation levels in prehistoric human remains. *Johannes Gutenberg Universität Mainz*
- Gronau I, Hubisz MJ, Gulko B, Danko CG and Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet* 43:1031–1035. doi: 10.1038/ng.937
- Gutenkunst RN, Hernandez RD, Williamson SH and Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*. doi: 10.1371/journal.pgen.1000695
- Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K et al. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. doi: 10.1038/nature14317
- Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R, Utermann G, Pritchard JK, Coop G and Di Rienzo A (2011) Adaptations to climate-mediated selective pressures in humans. *PLoS Genet* 7:e1001375. doi: 10.1371/journal.pgen.1001375

- Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D and Myers S (2014) A genetic atlas of human admixture history. *Science* 343:747–51. doi: 10.1126/science.1243518
- Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, Zhang MQ, Ye K, Bhattacharjee A, Brizuela L et al. (2009) High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. 1593–1605. doi: 10.1101/gr.095190.109.
- Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ et al. (2007) Genome-wide in situ exon capture for selective resequencing. *Nat Genet* 39:1522–7. doi: 10.1038/ng.2007.42
- Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Díez-Del-Molino D, Van Dorp L, López S, Kousathanas A, Link V et al. (2016) Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc Natl Acad Sci U S A*. doi: 10.1073/pnas.1523951113
- Hunt KA, Zhernakova A, Turner G, Heap GAR, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D et al. (2008) Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 40:395–402. doi: 10.1038/ng.102
- Jeong C, Wilkin S, Amgalantugs T, Bouwman AS, Taylor WTT, Hagan RW, Bromage S, Tsolmon S, Trachsel C, Grossmann J et al. (2018) Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proc Natl Acad Sci U S A* 115:E11248–E11255. doi: 10.1073/pnas.1813608115
- Jones ER, Gonzalez-ortez G, Connell S, Siska V, Eriksson A, Martiniano R, Mclaughlin RL, Llorente MG, Cassidy LM, Gamba C et al. (2015) Upper Palaeolithic genomes reveal deep roots of modern Eurasians. 1–8. doi: 10.1038/ncomms9912
- Jones ER, Zarina G, Moiseyev V, Lightfoot E, Nigst PR, Manica A, Pinhasi R and Bradley DG (2017) The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Curr Biol* 27:576–582. doi: 10.1016/j.cub.2016.12.060
- Jónsson H, Ginolhac A, Schubert M, Johnson PLF and Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29:1682–1684. doi: 10.1093/bioinformatics/btt193
- Juras A, Krzewińska M, Nikitin AG, Ehler E, Chyleński M, Łukasik S, Krenz-Niedbala M, Sinika V, Piontek J, Ivanova S et al. (2017) Diverse origin of mitochondrial lineages in Iron Age Black Sea Scythians. *Sci Rep* 7:1–10. doi: 10.1038/srep43950
- Kamberov YG, Wang S, Tan J, Gerbault P, Wark A, Tan L, Yang Y, Li S, Tang K, Chen H et al. (2018) Modeling Recent Human Evolution in Mice by Expression of a Selected EDAR Variant. *Cell* 152:691–702. doi: 10.1016/j.cell.2013.01.016
- Karolchik D (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32:493D–496. doi: 10.1093/nar/gkh103
- Keyser C, Bouakaze C, Crubézy E, Nikolaev VG, Montagnon D, Reis T and Ludes B (2009) Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum Genet* 126:395–410. doi: 10.1007/s00439-009-0683-0
- Kingman JFC (1982) The coalescent. *Stoch Process their Appl* 13:235–248. doi: https://doi.org/10.1016/0304-4149(82)90011-4
- Kircher M (2012) Analysis of High-Throughput Ancient DNA Sequencing Data. In: Shapiro B and Hofreiter M (eds) *Ancient DNA: Methods and Protocols*. Humana Press,

Totowa, NJ, pp 197–228

Kircher M, Sawyer S and Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 40:e3. doi: 10.1093/nar/gkr771

Korneliussen TS, Albrechtsen A and Nielsen R (2014) Open Access ANGSD : Analysis of Next Generation Sequencing Data. 1–13.

Kousathanas A, Leuenberger C, Link V, Sell C, Burger J and Wegmann D (2016) Inferring Heterozygosity from Ancient and Low Coverage Genomes. *Genetics*

Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A and Pääbo S (2010) A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol* 20:231–6. doi: 10.1016/j.cub.2009.11.068

Kreutzer S (2017) Populationsgenetische Analyse prähistorischer Individuen aus Griechenland. Johannes Gutenberg Universität Mainz

Krzewinska M, Kilinc GM, Juras A, Koptekin D, Chylenski M, Nikitin AG, Shcherbakov N, Shuteleva I, Leonova T, Kraeva L et al. (2018) Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. *Sci Adv* 4:eaat4457. doi: 10.1126/sciadv.aat4457

Lalueza-Fox C, Sampietro ML, Gilbert MTP, Castri L, Facchini F, Pettener D and Bertranpetit J (2004) Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proc Biol Sci* 271:941–7. doi: 10.1098/rspb.2004.2698

Lamason RL, Mohideen M-APK, Mest JR, Wong AC, Norton HL, Aros MC, Juryneć MJ, Mao X, Humphreville VR, Humbert JE et al. (2005) SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans. *Science* (80-) 310:1782 LP-1786.

Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M, Gamarra B, Sirak K et al. (2016) Genomic insights into the origin of farming in the ancient Near East. *Nature* 536:419–424. doi: 10.1038/nature19310

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M et al. (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513:409–413. doi: 10.1038/nature13673

Li C, Li H, Cui Y, Xie C, Cai D, Li W, Mair VH, Xu Z, Zhang Q, Abuduresule I 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. doi: 10.1186/1741-7007-8-15

Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. doi: 10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and Subgroup 1000 Genome Project Data Processing (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. doi: 10.1093/bioinformatics/btp352

Magalon H, Patin E, Austerlitz F, Hegay T, Aldashev A, Quintana-Murci L and Heyer E (2008) Population genetic diversity of the NAT2 gene supports a role of acetylation in human adaptation to farming in Central Asia. *Eur J Hum Genet* 16:243–51. doi: 10.1038/sj.ejhg.5201963

Malyarchuk BA, Grzybowski T, Derenko M V., Czarny J, Woźniak M and Miścicka-Śliwka

- D (2002) Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet* 66:261–283. doi: 10.1017/S0003480002001161
- Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J and Turner DJ (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7:111–8. doi: 10.1038/nmeth.1419
- Maricic T, Whitten M and Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5:e14004. doi: 10.1371/journal.pone.0014004
- Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M et al. (2015) Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528:499–503. doi: 10.1038/nature16152
- Meyer M and Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* doi: 10.1101/pdb.prot5448
- Molodin VI, Pilipenko AS, Romaschenko AG, Zhuravlev AA, Trapezov RO and Chikisheva TA Human migrations in the southern region of the West Siberian Plain during the Bronze Age : Archaeological , palaeogenetic and anthropological data. 93–112.
- Narasimhan VM, Patterson N, Moorjani P, Lazaridis I, Lipson M, Mallick S, Rohland N, Bernardos R, Nakatsuka N, Olalde I et al. (2018) The Genomic Formation of South and Central Asia. *BioRxiv* 292581. doi: 10.1101/292581
- Ohashi J, Naka I and Tsuchiya N (2011) The Impact of Natural Selection on an ABCC11 SNP Determining Earwax Type. *Mol Biol Evol* 28:849–857.
- Oota H, Pakstis AJ, Bonne-Tamir B, Goldman D, Grigorenko E, Kajuna SLB, Karoma NJ, Kungulilo S, Lu RB, Odunsi K et al. (2004) The evolution and population genetics of the ALDH2 locus: Random genetic drift, selection, and low levels of recombination. *Ann Hum Genet* 68:93–109. doi: 10.1046/j.1529-8817.2003.00060.x
- Osier M V, Cheung K-H, Kidd JR, Pakstis AJ, Miller PL and Kidd KK (2002) ALFRED: An allele frequency database for anthropology. *Am J Phys Anthropol* 119:77–83. doi: 10.1002/ajpa.10094
- Palstra FP, Heyer E and Austerlitz F (2015) Statistical Inference on Genetic Data Reveals the Complex Demographic History of Human Populations in Central Asia. *Mol Biol Evol.* doi: 10.1093/molbev/msv030
- Parzinger H (2006) *Die frühen Völker Eurasiens: vom Neolithikum bis zum Mittelalter.* Beck
- Parzinger H (2007) *Die Skythen.* C. H. Beck, München
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T and Reich D (2012) Ancient admixture in human history. *Genetics* 192:1065–1093. doi: 10.1534/genetics.112.145037
- Peng Y, Shi H, Qi X, Xiao C, Zhong H, Ma RZ and Su B (2010) The ADH1B Arg47His polymorphism in East Asian populations and expansion of rice domestication in history. *BMC Evol Biol* 10:15. doi: 10.1186/1471-2148-10-15
- Pinhasi R, Fernandes D, Sirak K, Novak M, Connell S, Alpaslan-Roodenberg S, Gerritsen F, Moiseyev V, Gromov A, Raczky P et al. (2015) Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLoS One* 10:e0129102.

- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA and Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ et al. (2018) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 81:559–575. doi: 10.1086/519795
- Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford Jr TW, Orlando L, Metspalu E et al. (2013) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*. doi: 10.1038/nature12736
- Rajeevan H, Osier M V., Cheung KH, Deng H, Druskin L, Heinzen R, Kidd JR, Stein S, Pakstis AJ, Tosches NP et al. (2003) ALFRED: The ALlele FREquency Database. Update. *Nucleic Acids Res* 31:270–271. doi: 10.1093/nar/gkg043
- Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW, Rasmussen S, Moltke I, Albrechtsen A, Doyle SM et al. (2014) The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506:225–9. doi: 10.1038/nature13025
- Reich D, Thangaraj K, Patterson N, Price AL and Singh L (2009) Reconstructing Indian Population History. *Nature* 461:489–494. doi: 10.1038/nature08365
- Ricaud F, Bourgeois J, Crubezy E and Ludes B (2004) Genetic Analysis of a Scytho-Siberian Skeleton and Its Implications for Ancie ...
- Sarkissian C Der (2011) Mitochondrial DNA in Ancient Human Populations of Europe Ecology and Evolutionary Biology.
- Scheu A, Powell A, Bollongino R, Vigne J-D, Tresset A, Çakırlar C, Benecke N and Burger J (2015) The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC Genet* 16:54. doi: 10.1186/s12863-015-0203-2
- Selinski S, Blaszkewicz M, Lehmann ML, Ovsianikov D, Moormann O, Guballa C, Kress A, Tru MC, Gerullis H, Otto T et al. (2011) Genotyping NAT2 with only two SNPs (rs1041983 and rs1801280) outperforms the tagging SNP rs1495741 and is equivalent to the conventional 7-SNP NAT2 genotype. *Pharmacogenet Genomics* 21:673–678. doi: 10.1097/FPC.0b013e3283493a23
- Sell C (2017) Addressing Challenges of Ancient DNA Sequence Data Obtained with Next Generation Methods. Johannes Gutenberg Universität Mainz
- Sikora M, Seguin-Orlando A, Sousa VC, Albrechtsen A, Korneliussen T, Ko A, Rasmussen S, Dupanloup I, Nigst PR, Bosch MD et al. (2017) Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* (80-.).
- Skoglund P, Storå J, Götherström A and Jakobsson M (2013) Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J Archaeol Sci* 40:4477–4482. doi: 10.1016/j.jas.2013.07.004
- Spichenok O, Budimlija ZM, Mitchell AA, Jenny A, Kovacevic L, Marjanovic D, Caragine T, Prinz M and Wurmbach E (2011) Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Sci Int Genet* 5:472–478. doi: 10.1016/j.fsigen.2010.10.005
- Stokowski RP, Pant PVK, Dadd T, Fereday A, Hinds DA, Jarman C, Filsell W, Ginger RS, Green MR, van der Ouderaa FJ et al. (2007) A genomewide association study of skin pigmentation in a South Asian population. *Am J Hum Genet* 81:1119—1132. doi:

10.1086/522235

Tewhey R, Nakano M, Wang X, Pabón-Peña C, Novak B, Giuffre A, Lin E, Happe S, Roberts DN, LeProust EM et al. (2009) Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biol* 10:R116. doi: 10.1186/gb-2009-10-10-r116

Thompson EE, Kuttub-Boulos H, Witonsky D, Yang L, Roe BA and Di Rienzo A (2004) CYP3A Variation and the Evolution of Salt-Sensitivity Variants. *Am J Hum Genet* 75:1059–1069. doi: <https://doi.org/10.1086/426406>

Unterländer M (2014) Populationsgenetik eisenzeitlicher Reiternomaden der Eurasischen Steppe: next generation sequencing alter DNA aus archäologischen humanen Skeletten des 1. Jahr-tausends vor Christus. Johannes Gutenberg Universität Mainz

Unterländer M, Palstra F, Lazaridis I, Pilipenko A, Hofmanová Z, Groß M, Sell C, Blöcher J, Kirsanow K, Rohland N et al. (2017) Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nat Commun*. doi: 10.1038/ncomms14615

van Geel B, Bokovenko NA, Burova ND, Chugunov K V, Dergachev VA, Dirksen VG, Kulkova M, Nagler A, Parzinger H, van der Plicht J et al. (2004) Climate change and the expansion of the Scythian culture after 850 BC: a hypothesis. *J Archaeol Sci* 31:1735–1742. doi: <https://doi.org/10.1016/j.jas.2004.05.004>

Veeramah KR, Rott A, Groß M, van Dorp L, López S, Kirsanow K, Sell C, Blöcher J, Wegmann D, Link V et al. (2018) Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. *Proc Natl Acad Sci* 201719880. doi: 10.1073/pnas.1719880115

Wagner DM, Klunk J, Harbeck M, Devault A, Waglechner N, Sahl JW, Enk J, Birdsell DN, Kuch M, Lumibao C et al. (2014) *Yersinia pestis* and the plague of Justinian 541-543 AD: a genomic analysis. *Lancet Infect Dis* 14:319–26. doi: 10.1016/S1473-3099(13)70323-2

Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L, Maeda H, Ishikawa T, Sijen T, de Knijff P et al. (2014) Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Sci Int Genet* 9:150–61. doi: 10.1016/j.fsigen.2013.12.006

Wang C-C, Reinhold SR, Kalmykov A, Wissgott A, Brandt G, Jeong C, Cheronet O, Ferry M, Harney E, Keating D et al. (2018) The genetic prehistory of the Greater Caucasus. *bioRxiv* 322347. doi: 10.1101/322347

Wang C and Zhan X (2015) LASER : Locating Ancestry from SEquence Reads version 2 . 02. 0–25.

Wang C, Zhan X, Bragg-Gresham J, Kang HM, Stambolian D, Chew EY, Branham KE, Heckenlively J, Study TF, Fulton R et al. (2014) Ancestry Estimation and Control of Population Stratification for Sequence-based Association Studies. *Nat Genet* 46:409–415. doi: 10.1038/ng.2924

Wang H, Ding K, Zhang Y, Jin L, Kullo IJ and He F (2007) Comparative and evolutionary pharmacogenetics of ABCB1: Complex signatures of positive selection on coding and regulatory regions. *Pharmacogenet Genomics* 17:667–678. doi: 10.1097/FPC.0b013e328165249f

Wang Z, Schuetz EG, Xu Y and Thummel KE (2013) Interplay between vitamin D and the drug metabolizing enzyme CYP3A4. *J Steroid Biochem Mol Biol* 136:54–58. doi: <https://doi.org/10.1016/j.jsbmb.2012.09.012>

Wesolowska a, Dalgaard MD, Borst L, Gautier L, Bak M, Weinhold N, Nielsen BF, Helt LR, Audouze K, Nersting J et al. (2011) Cost-effective multiplexing before capture allows screening of 25 000 clinically relevant SNPs in childhood acute lymphoblastic leukemia. *Leukemia* 25:1001–6. doi: 10.1038/leu.2011.32

Wilde S (2014) *Populationsgenetik kupfer- und bronzezeitlicher Bevölkerungen der osteuropäischen Steppe*. Johannes Gutenberg Universität Mainz

Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, Unterländer M, Hollfelder N, Potekhina ID, Schier W, Thomas MG et al. (2014) Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc Natl Acad Sci* 111:4832–4837. doi: 10.1073/pnas.1316513111

WRIGHT S (1951) The genetical structure of populations. *Ann Eugen* 15:323–354.

Yoshida A, Huang IY and Ikawa M (1984) Molecular abnormality of an inactive aldehyde dehydrogenase variant commonly found in Orientals. *Proc Natl Acad Sci* 81:258 LP-261.

Yoshiura KI, Kinoshita A, Ishida T, Ninokata A, Ishikawa T, Kaname T, Bannai M, Tokunaga K, Sonoda S, Komaki R et al. (2006) A SNP in the ABCC11 gene is the determinant of human earwax type. *Nat Genet* 38:324–330. doi: 10.1038/ng1733

Anhang A

Chemikalien und Geräte

Chemikalien	
Agencourt AMPure XP beads	Beckman Coulter
Agilent High Sensitivity DNA Reagent	Agilent Technologies
AmpliTaQ Gold ® Puffer II (10x)	Life Technologies TM
AmpliTaQ Gold ® DNA Polymerase	Life Technologies TM
ATP (100 mM)	Life Technologies TM
Bovine Serum Albumin (BSA) (20 mg/ml)	Roche Diagnostics
Bst Polymerase, Large Fragment (8 U/µl)	New England Biolabs GmbH
dNTPs (each 10 mM)	Qiagen
dNTPs (each 25 mM)	Agilent Technologies
Dynabeads ® MyOne TM Streptavidin T1 magnetische beads.	Thermo Fisher Scientific Inc.
EDTA (0.5 M), pH 8.0	Ambion ®
Ethanol 96% (vergällt)	Carl Roth GmbH + Co. KG
Herculase II fusion DNA Polymerase	Agilent Technologies
Herculase II Reaktionspuffer	Agilent Technologies
HPLC-H2O	Fisher Scientific GmbH
KAPA SYBR ® FAST QPCR MasterMix Universal Kapa	Biosystems/ Peqlab
MgCl ₂ (25 mM)	Life Technologies TM
1N NaOH Solution	Agilent Technologies
N-Laurylsarcosine	Merck Millipore Corporation
Nuclease-freies H ₂ O	Life Technologies TM
PEG-4000	Thermo Scientific TM
Proteinase-K Roche Diagnostics	Roche Diagnostics
Roti ® -Phenol/Chloroform/Isoamylalkohol	Roche Diagnostics
T4 DNA Ligase (5 U/µl)	Thermo Scientific TM
T4 DNA Ligase Buffer (10X)	Thermo Scientific TM
T4 DNA Polymerase (5 U/µl)	Thermo Scientific TM
T4 Polynukleotide Kinase	Invitrogen TM
Tango Buffer (10x)	Life Technologies TM
ThermoPol Buffer (10X)	New England Biolabs GmbH
Trichlormethan/Chloroform	Carl Roth GmbH + Co. KG
USER TM Enzym	New England Biolabs GmbH

Kits	
High Sensitivity DNA Kit	Agilent Technologies
MinElute ® PCR Purification Kit	Qiagen
MSB ® Spin PCRapace Kit	Stratec
Agilent ® Sure Select Target enrichment Kit	Agilent
SeqCap EZ Developer Library, 4 Reactions	Roche
MYbait s kit 1-48	MYcroarray®
QIAquick PCR Purification Kit	Qiagen
Qubit ® HS Assay Kit	Thermo Fisher Scientific Inc.
Geräte	
2100 Expert Bioanalyzer	Agilent Technologies
Eppendorf Concentrator plus Speed-vac	Eppendorf AG
Hettich Zentrifugen Mikro 200	Hettich Lab Technology
Hettich Zentrifugen Universal 320	Hettich Lab Technology
MasterCycler® Thermocycler	Eppendorf AG
Qubit® Fluorometer	Invitrogen™
StepOnePlus™ Real-Time PCR System	Applied Biosystems
Thermomixer comfort	Eppendorf AG
Verbrauchsmaterialien	
Amicon ® Ultra-15 (30, 50 kDa)	Merck Millipore Corporation
Astronaut caps	HANSA - TRADING HTH GmbH & Co. KG
DNA LoBind Tubes 0,5 ml, 1,5 ml	Eppendorf AG
DuPont™ Tyvek ® Protective Apparel	DuPont (via VWR International)
Eppendorf Safe-Lock Tubes 0,5 ml, 1,5 ml	Eppendorf AG
Falcon 15 ml, 50 ml	Falcon 15 ml, 50 ml SARSTEDT AG & Co
Einweg Handschuhe	HANSA - TRADING HTH GmbH & Co. KG
Gesichtsmaske	HANSA - TRADING HTH GmbH & Co. KG
Alconox ® detergent IF BC	Sigma-Aldrich Co.
DanKlorix Hygiene Reiniger	DanKlorix
DNA-ExitusPlus™ IF BC	AppliChem GmbH

Anhang B

B1 Adapter - und Primersequenzen nach (Kircher et al. 2012) und (Meyer and Kircher 2010)

<i>Bezeichnung</i>	<i>Sequenz</i>
IS1_adapter.P5	A*C*A*C*TCITTTCCCTACACGACGCTCTTCCG*A*I*C*T
IS2_adapter.P7	G*I*G*A*CTGGAGTTCAGACGTGTGCTCTTCCG*A*I*C*T
IS3_adapter.P5+P7	A*G*A*I*CGGAA*G*A*G*C
IS4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT
IS5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT
IS6	CAAGCAGAAGACGGCATAACGA
IS7	ACACTCTTTCCCTACACGAC
IS8	GTGACTGGAGTTCAGACGTGT
Read 1 (P7)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Index Read (P7)	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
Read 2 (P5)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P5 index sequencing	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

* bezeichnet eine PTO-Bindung

B2 Sequenz der Positivkontrolle (NH)

Bezeichnung	Sequenz (5' – 3')
Nonsense 1	CCATCTCATCCCTGCGTGTCAACTGACTAAACTAGGTGCCACGTTCGT-GAAAGTCTGACAACCTATCCCTGTGTGCCTTG
Nonsense 2	CAAGGCACACAGGGGATAGGTTGTCAGACTTTCACGACGTGGCACCTAGTT-TAGTCAGTTGACACGCAGGGATGAGATGG

B3 Auflistung ausgewählter funktionaler Marker im Anreicherungsverfahren

Bezeichnung	rs-Nummer	Chromosom	Start Position	Stop Position
SEC16B	rs10913469	chr1	177913518	177913519
NOTCH2	rs10923931	chr1	120517958	120517959
LEPR	rs1137100	chr1	66036440	66036441
SLC44A5	rs13373826	chr1	75970794	75970795
MTHFR	rs17367504	chr1	11862777	11862778
NEGR1	rs2568958	chr1	72765115	72765116
ADAM30	rs2641348	chr1	120437883	120437884
PROX1	rs340874	chr1	214159255	214159256
AGT	rs5051	chr1	230849871	230849872
PCSK9	rs562556	chr1	55524236	55524237
AGT	rs699	chr1	230845793	230845794
HHEX/ IDE	rs1111875	chr10	94462881	94462882
TCF7L2	rs11196205	chr10	114807046	114807047
CAMK1D/ CDC123	rs12779790	chr10	12328009	12328010
c19orf107	rs1530440	chr10	63524590	63524591
GBF1	rs3802678	chr10	104115394	104115395
TCF7L2	rs4506565	chr10	114756040	114756041
PLRP2	rs4751995	chr10	118397883	118397884
HHEX/IDE	rs5015480	chr10	94465558	94465559
TCF7L2	rs7901695	chr10	114754087	114754088
TCF7L2	rs7903146	chr10	114758348	114758349
HHEX/IDE	rs7923837	chr10	94481916	94481917
TYR.1	rs1042602	chr11	88911695	88911696
MTNR1B	rs10830963	chr11	92708709	92708710
MTCH2	rs10838738	chr11	47663048	47663049
TYR.2/3	rs1126809	chr11	89017960	89017961
MTNR1B	rs1387153	chr11	92673827	92673828
CENTD2	rs1552224	chr11	72433097	72433098
FADS2/FAD S3	rs174570	chr11	61597211	61597212
KCNQ1	rs2074196	chr11	2824799	2824800
KCNQ1	rs2237892	chr11	2839750	2839751
KCNQ1	rs231362	chr11	2691470	2691471
GRB2	rs2373115	chr11	78091149	78091150

Bezeichnung	rs-Nummer	Chromosom	Start Position	Stop Position
TPCN2.1	rs35264875	chr11	68846398	68846399
TPCN2.2	rs3829241	chr11	68855362	68855363
BDNF/LGR4 /LIN7C	rs4923461	chr11	27656909	27656910
CASP12	rs4971116	chr11	104763116	104763117
KCNJ11	rs5215	chr11	17408629	17408630
KCNJ11	rs5219	chr11	17409571	17409572
F2	rs5896	chr11	46745002	46745003
BDNF	rs6265	chr11	27679915	27679916
MADD/FOL H1	rs7395662	chr11	48518892	48518893
BDNF/LGR4 /LINC7	rs925946	chr11	27667201	27667202
DCD	rs1153188	chr12	55098995	55098996
KITLG.2	rs12821256	chr12	89328334	89328335
HMGA2	rs1531343	chr12	66174893	66174894
HMGA2	rs2612067	chr12	66170162	66170163
SH2B3	rs3184504	chr12	111884607	111884608
ENaC/SCNN 1a	rs3759324	chr12	6485660	6485661
KRT77	rs3782489	chr12	53086644	53086645
VDR	rs4516035	chr12	48299825	48299826
GNB3	rs5443	chr12	6954874	6954875
KITLG.1	rs642742	chr12	89299745	89299746
SH2B3/ATX N2	rs653178	chr12	112007755	112007756
ALDH2	rs671	chr12	112241765	112241766
OASL/HNF1 A	rs7957197	chr12	121460685	121460686
TSPAN8/LG R5	rs7961581	chr12	71663101	71663102
RPL21	rs10507380	chr13	27879525	27879526
RGS6	rs10149848	chr14	72855161	72855162
CYP17A1	rs1191548	chr14	30000681	30000682
SLC24A4	rs12896399	chr14	92773662	92773663
RAB27A.1	rs104894497	chr15	55520890	55520891
RAB27A.2	rs104894498	chr15	55516164	55516165
RAB27A.3	rs104894499	chr15	55516099	55516100
RAB27A.4	rs104894500	chr15	55516201	55516202
TRPM2	rs11070811	chr15	31394081	31394082
BCL2A1/ZF AND	rs11634397	chr15	80432221	80432222
HERC2	rs12913832	chr15	28365617	28365618
CYP1A2	rs1378942	chr15	75077366	75077367
SLC24A5	rs1426654	chr15	48426483	48426484
OCA2	rs1667394	chr15	28530181	28530182
MYO5A	rs16964944	chr15	52676391	52676392
MYO5A	rs1724577	chr15	52689630	52689631
TRPM1	rs17815774	chr15	31334361	31334362

Bezeichnung	rs-Nummer	Chromosom	Start Position	Stop Position
OCA2.1	rs1800401	chr15	28260052	28260053
OCA2.2	rs1800407	chr15	28230317	28230318
OCA2.3	rs1800414	chr15	28197036	28197037
RAB27A.5	rs28938176	chr15	55522620	55522621
MYO5A	rs4776053	chr15	52816194	52816195
CYP1A2	rs762551	chr15	75041916	75041917
PRC1	rs8042680	chr15	91521336	91521337
OCA2	rs1545397	chr15	28187771	28187772
MC1R.6	rs1110400	chr16	89986129	89986130
MC1R.4	rs11547464	chr16	89986090	89986091
SLC12A3	rs1529927	chr16	56904586	56904587
ABCC11	rs17822931	chr16	48258197	48258198
MC1R.1	rs1805005	chr16	89985843	89985844
MC1R.2	rs1805006	chr16	89985917	89985918
MC1R.5	rs1805007	chr16	89986116	89986117
MC1R.7	rs1805008	chr16	89986143	89986144
MC1R.9	rs1805009	chr16	89986545	89986546
MC1R.3	rs2228479	chr16	89985939	89985940
ENaC/SCNN1g	rs5718	chr16	23193864	23193865
SH2B1	rs7498665	chr16	28883240	28883241
FTO	rs8050136	chr16	53816274	53816275
MC1R.8	rs885479	chr16	89986153	89986154
FTO	rs9939609	chr16	53820526	53820527
PLCD3	rs12946454	chr17	43208120	43208121
ZNF652	rs16948048	chr17	47440465	47440466
KCNJ2	rs17779747	chr17	68494991	68494992
TCF2	rs4430796	chr17	36098039	36098040
TCF2/HNF1B	rs757210	chr17	36096514	36096515
MC4R	rs17782313	chr18	57851096	57851097
ANGPTL4	rs116843064	chr19	8429322	8429323
KCTD15	rs368794	chr19	34320451	34320452
APOE	rs429358	chr19	45411940	45411941
APOE	rs7412	chr19	45412078	45412079
BCL11A	rs10490072	chr2	60669930	60669931
LCTb	rs182549	chr2	136616753	136616754
BCL11A	rs243021	chr2	60584818	60584819
IRS1	rs2943641	chr2	227093744	227093745
EDAR	rs3827760	chr2	109513600	109513601
LCTa	rs4988235	chr2	136608645	136608646
TMEM18	rs6548238	chr2	634904	634905
IRS1	rs7578326	chr2	227020652	227020653
THADA	rs7578597	chr2	43732822	43732823
RBMS1/ITGB6	rs7593730	chr2	161171453	161171454
GCKR	rs780094	chr2	27741236	27741237
ALMS1	rs9309473	chr2	73743981	73743982

Bezeichnung	rs-Nummer	Chromosom	Start Position	Stop Position
ASIP.3	rs1015362	chr20	32738611	32738612
ASIP.4	rs4911414	chr20	32729443	32729444
ASIP.2	rs6058017	chr20	32856997	32856998
ADCY5	rs11708067	chr3	123065777	123065778
SYNC/PPARG	rs17036101	chr3	12277844	12277845
PPARG	rs1801282	chr3	12393124	12393125
MCF2L2	rs2314349	chr3	183091097	183091098
IGF2BP2	rs4402960	chr3	185511686	185511687
ADAMTS9	rs4607103	chr3	64711903	64711904
WFS1	rs10010131	chr4	6292914	6292915
TLR.10	rs11096955	chr4	38776106	38776107
TLR.10	rs11096957	chr4	38776490	38776491
NPY1R/NPY5R	rs11724320	chr4	164267922	164267923
ADH1Bb	rs1229984	chr4	100239318	100239319
FGF5	rs16998073	chr4	81184340	81184341
ADH1Ba	rs3811801	chr4	100244318	100244319
TLR.10	rs4129009	chr4	38774888	38774889
CORIN	rs4558836	chr4	47592649	47592650
TLR.1	rs4833095	chr4	38799709	38799710
TLR.1	rs5743618	chr4	38798647	38798648
TLR.6	rs5743810	chr4	38830349	38830350
UCP1	rs6536991	chr4	141481580	141481581
SCOC	rs6840033	chr4	141228860	141228861
CAMK2D	rs6850980	chr4	114710293	114710294
ADRB2	rs1042713	chr5	148206439	148206440
ADRB2	rs1042714	chr5	148206472	148206473
MTRR	rs162036	chr5	7885958	7885959
SLC45A2.1	rs16891982	chr5	33951692	33951693
SLC45A2.2	rs26722	chr5	33963869	33963870
SLC45A2	rs28777	chr5	33958958	33958959
SLC45A2	rs35391	chr5	33955672	33955673
ZBED3	rs4457053	chr5	76424948	76424949
PCSK1	rs6235	chr5	95728897	95728898
DTNBP1.1	rs1018381	chr6	15657069	15657070
DTNBP1.2	rs104893945	chr6	15627621	15627622
CDKAL1	rs10946398	chr6	20661033	20661034
IRF4.1	rs12203592	chr6	396320	396321
IRF4.2	rs1540771	chr6	466032	466033
TNXB	rs185819	chr6	32050066	32050067
TNXB/CREBL1	rs2269426	chr6	32076498	32076499
DTNBP1.3	rs2619522	chr6	15653648	15653649
DTNBP1.4	rs2619538	chr6	15665208	15665209
CDKAL1_a	rs7754840	chr6	20661249	20661250
CDKAL1_b	rs7756992	chr6	20679708	20679709
PPARD	rs9470015	chr6	35369083	35369084

Bezeichnung	rs-Nummer	Chromosom	Start Position	Stop Position
VEGFA	rs9472138	chr6	43811761	43811762
ABCB1c	rs1045642	chr7	87138644	87138645
ABCB1a	rs1128503	chr7	87179600	87179601
POU6F2	rs1525791	chr7	39156557	39156558
IL6	rs1800795	chr7	22766644	22766645
ABCB1b	rs2032582	chr7	87160617	87160618
TRIP6	rs2075756	chr7	100466440	100466441
DGKB	rs2191349	chr7	15064308	15064309
CYP3A4	rs2740574	chr7	99382095	99382096
TRPV6a	rs4987657	chr7	142574912	142574913
TRPV6b	rs4987667	chr7	142572907	142572908
TRPV6c	rs4987682	chr7	142569595	142569596
PON1	rs662	chr7	94937445	94937446
CYP3A5	rs776746	chr7	99270538	99270539
JAZF1	rs864745	chr7	28180555	28180556
GCK	rs917793	chr7	44245852	44245853
KLF14	rs972283	chr7	130466853	130466854
NAT2	rs1041983	chr8	18257794	18257795
SLC30A8	rs13266634	chr8	118184782	118184783
NAT2	rs1801280	chr8	18257853	18257854
ZMAT4	rs2722425	chr8	40484238	40484239
SMAD3	rs717834	chr8	19740867	19740868
PLEKHF2	rs896854	chr8	95960510	95960511
MYC/BC042052	rs9642880	chr8	128718067	128718068
CDKN2A/B	rs10811661	chr9	22134093	22134094
CHCHD9/TLE4	rs13292136	chr9	81952127	81952128
TYRP1.1	rs1408799	chr9	12672096	12672097
TYRP1.1	rs2733832	chr9	12704724	12704725
ASIP.1	rs491142	chr9	135327188	135327189
CDKN2B	rs564398	chr9	22029546	22029547
DUSP9	rs5945326	chrX	152899921	152899922
M304	rs13447352	chrY	22749852	22749853
M324	rs13447361	chrY	2821785	2821786
U106/S21/M405	rs16981293	chrY	8796077	8796078
P215	rs17307294	chrY	16638803	16638804
P224	rs17307398	chrY	17285992	17285993
M436	rs17315680	chrY	18747492	18747493
M184	rs20320	chrY	14898162	14898163
M168	rs2032595	chrY	14813990	14813991
M174	rs2032602	chrY	14954279	14954280
M172/P294	rs2032604	chrY	14969633	14969634
M73	rs2032634	chrY	21888873	21888875
M201	rs2032636	chrY	15027528	15027529
M86	rs2032643	chrY	21905916	21905917
M217	rs2032668	chrY	15437332	15437333

Bezeichnung	rs-Nummer	Chromosom	Start Position	Stop Position
M69	rs2032673	chrY	21894057	21894058
M175	rs2032678	chrY	15508705	15508710
M526	rs2033003	chrY	23550923	23550924
SRY10831.1	rs2534636	chrY	2657175	2657176
P58	rs34043621	chrY	14486666	14486667
L27	rs34126399	chrY	23021977	23021978
P312/S116	rs34276300	chrY	22157310	22157311
Tat	rs34442126	chrY	14922582	14922583
L497	rs35141399	chrY	17423319	17423320
M145	rs3848982	chrY	21717207	21717208
M12	rs3903	chrY	7583479	7583480
M15	rs3906	chrY	21724698	21724700
M17	rs3908	chrY	21733167	21733168
M20	rs3911	chrY	21733453	21733454
P143	rs4141886	chrY	14197866	14197867
L223	rs810801	chrY	6345147	6345148
L223	rs810801	chrY	9525231	9525232
M242	rs8179021	chrY	15018581	15018582
M96	rs9306841	chrY	21778997	21778998
M230	rs9341277	chrY	15481371	15481372
M231	rs9341278	chrY	15469723	15469724
M267	rs9341313	chrY	22741817	22741818
P297	rs9785702	chrY	18656507	18656508
P159	rs9785905	chrY	18097250	18097251
U1	rs9785956	chrY	15058877	15058878
L23	rs9785971	chrY	6753510	6753511
M412	rs9786140	chrY	8502235	8502236
M269	rs9786153	chrY	22739366	22739367
M343	rs9786184	chrY	2887823	2887824
L13	rs9786706	chrY	14698927	14698928
P99		chrY	14886385	14886386
P324/P260		chrY	17286005	17286006
M479		chrY	20834666	20834667
M35.1		chrY	21741702	21741703
L142.1		chrY	6753305	6753306
M407		chrY	2750407	2750408
M342		chrY	23243941	23243942
P15		chrY	23244025	23244026
P16_1		chrY	20668987	20668988
P16_2		chrY	20975189	20975190
L190		chrY	18759689	18759690
L140		chrY	7570858	7570859
L72		chrY	23243886	23243887
P38		chrY	14484378	14484379
P37.2		chrY	14491683	14491684
M410		chrY	2751677	2751678

Bezeichnung	rs-Nummer	Chromosom	Start Position	Stop Position
P256		chrY	8685229	8685230
M128		chrY	21767930	21767931
P43		chrY	21880972	21880973
M122		chrY	21764673	21764674
M458		chrY	24366463	24366464
L584		chrY	28731916	28731917
P40		chrY	14484393	14484394
M3	rs3894	chrY	17605756	17605757
M9	rs3900	chrY	20189644	20189645
M45	rs2032631	chrY	20327174	20327175
M89	rs2032652	chrY	20376700	20376701
M173	rs2032624	chrY	13535817	13535818
M216	rs2032666	chrY	13946957	13946958
RPS4Y (M130)	rs35284970	chrY	2794853	2794854

B4 Oligonukleotidsequenzen der Blocking-Oligos im Hybridisierungsansatz

Bezeichnung	Sequenz (5' – 3')
Meyer.Bo4.P7.part1.R ^a	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-Pho
Meyer.Bo6.P7.part2.R ^a	CAAGCAGAAGACGGCATACGAGAT-Pho
BOIs1.revcomp	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-Pho

^a nach (Kircher et al. 2012); Pho bezeichnet ein 3'-Phosphat

B5 Übersicht der Referenzdaten alter DNA für die PCA aus der Literatur

Referenz	Sample ID	Gruppen ID	Kultur	Alter	Land
Mathieson et al. 2015	I0231	Steppe_EMBA	Yamnaya-Samara	2921-2762 v. Chr.	Russland
Mathieson et al. 2015	I0357	Steppe_EMBA	Yamnaya-Samara	3090-2913 v. Chr.	Russland
Mathieson et al. 2015	I0429	Steppe_EMBA	Yamnaya-Samara	3339-2918 v. Chr.	Russland
Mathieson et al. 2015	I0439	Steppe_EMBA	Yamnaya-Samara	3321-2921 v. Chr.	Russland
Mathieson et al. 2015	I0370	Steppe_EMBA	Yamnaya-Samara	3300-2700 v. Chr.	Russland
Mathieson et al. 2015	I0438	Steppe_EMBA	Yamnaya-Samara	3021-2635 v. Chr.	Russland
Mathieson et al. 2015	I0444	Steppe_EMBA	Yamnaya-Samara	3335-2882 v. Chr.	Russland
Mathieson et al. 2015	I0443	Steppe_EMBA	Yamnaya-Samara	3300-2700 v. Chr.	Russland

Referenz	Sample ID	Gruppen ID	Kultur	Alter	Land
Mathieson et al. 2015	I0441	Steppe_EMBA	Yamnaya-Samara	3010-2622 v. Chr.	Russland
Mathieson et al. 2015	I0017	Motala_HG	Motala_HG	5721-5631 v. Chr.	Schweden
Mathieson et al. 2015	I0015	Motala_HG	Motala_HG	5964-5629 v. Chr.	Schweden
Mathieson et al. 2015	I0014	Motala_HG	Motala_HG	5964-5629 v. Chr.	Schweden
Mathieson et al. 2015	I0013	Motala_HG	Motala_HG	5964-5638 v. Chr.	Schweden
Mathieson et al. 2015	I0012	Motala_HG	Motala_HG	5714-5575 v. Chr.	Schweden
Mathieson et al. 2015	I0011	Motala_HG	Motala_HG	5714-5575 v. Chr.	Schweden
Mathieson et al. 2015	I1504	Europe_LNBA	Späte Bronzezeit	1270-1110 v. Chr.	Ungarn
Mathieson et al. 2015	I0440	Steppe_EMBA	Poltavka	2887-2666 v. Chr.	Russland
Mathieson et al. 2015	I0371	Steppe_EMBA	Poltavka	2872-2583 v. Chr.	Russland
Mathieson et al. 2015	I0374	Steppe_EMBA	Poltavka	2800-2200 v. Chr.	Russland
Mathieson et al. 2015	I0126	Steppe_EMBA	Poltavka	2867-2486 v. Chr.	Russland
Mathieson et al. 2015	I0430	Steppe_MLBA	Srubnaya	1850-1600 v. Chr.	Russland
Mathieson et al. 2015	I0431	Steppe_MLBA	Srubnaya	1850-1600 v. Chr.	Russland
Mathieson et al. 2015	I0424	Steppe_MLBA	Srubnaya	1850-1600 v. Chr.	Russland
Mathieson et al. 2015	I0422	Steppe_MLBA	Srubnaya	1850-1200 v. Chr.	Russland
Mathieson et al. 2015	I0358	Steppe_MLBA	Srubnaya	1906-1631 v. Chr.	Russland
Mathieson et al. 2015	I0359	Steppe_MLBA	Srubnaya	1850-1200 v. Chr.	Russland
Mathieson et al. 2015	I0360	Steppe_MLBA	Srubnaya	1850-1200 v. Chr.	Russland
Mathieson et al. 2015	I0361	Steppe_MLBA	Srubnaya	1850-1200 v. Chr.	Russland
Mathieson et al. 2015	I0232	Steppe_MLBA	Srubnaya	1850-1200 v. Chr.	Russland
Mathieson et al. 2015	I0234	Steppe_MLBA	Srubnaya	1850-1600 v. Chr.	Russland
Mathieson et al. 2015	I0235	Steppe_MLBA	Srubnaya	1850-1600 v. Chr.	Russland
Mathieson et al. 2015	I0247	Steppe_IA	Scythian_IA	375-203 v. Chr.	Russland
Mathieson et al. 2015	I0246	Steppe_EMBA	Potapovka	2469-1928 v. Chr.	Russland
Mathieson et al. 2015	I0418	Steppe_EMBA	Potapovka	2125-1769 v. Chr.	Russland
Mathieson et al. 2015	I0419	Steppe_EMBA	Potapovka	2200-1900 v. Chr.	Russland
Mathieson et al. 2015	I0061	EHG	Karelia_HG	~ 6850-6000 v. Chr.	Russland

Referenz	Sample ID	Gruppen ID	Kultur	Alter	Land
Mathieson et al. 2015	I0124	EHG	Samara_HG	~ 5657-5541 v. Chr.	Russland
Jones et al. 2015	KK1	CHG	CHG	7940-7600 v. Chr.	Georgien
Jones et al. 2015	SATP	CHG	CHG	11430-11180 v. Chr.	Georgien
Jones et al. 2017	ZVEJ25	Latvia_HG	HG	5841-5636 v. Chr.	Lettland
Jones et al. 2017	ZVEJ31	Latvia_MN2		4229-3800 v. Chr.	Lettland
Jones et al. 2017	ZVEJ27	Latvia_HG	HG	5302-4852 v. Chr.	Lettland
Jones et al. 2017	StPet2	Ukraine_HG	HG	9193-8641 v. Chr.	Ukraine
Jones et al. 2017	ZVEJ28	Latvia_LN1		3089-2676 v. Chr.	Lettland
Jones et al. 2017	StPet12	Ukraine_N1		4519-4343 v. Chr.	Ukraine
Jones et al. 2017	ZVEJ26	Latvia_MN1		4251-3976 v. Chr.	Lettland
Sikora et al. 2017	SI	OP	Oberes Paläolithikum	33875-31770 cal BP	Russland
Sikora et al. 2017	SII	OP	Oberes Paläolithikum	35283-33185 cal BP	Russland
Sikora et al. 2017	SIII	OP	Oberes Paläolithikum	35154-33031 cal BP	Russland
Sikora et al. 2017	SIV	OP	Oberes Paläolithikum	34485-33499 cal BP	Russland
Sikora et al. 2017	SVI	OP	Oberes Paläolithikum	730-850 cal BP	Russland
Allentoft et al. 2015	RISE109	Europe_LNBA	Unetice_EBA	1954-1772 v. Chr.	Polen
Allentoft et al. 2015	RISE150	Europe_LNBA	Unetice_EBA	1885-1693 v. Chr.	Polen
Allentoft et al. 2015	RISE154	Europe_LNBA	Unetice_EBA	1925-1765 v. Chr.	Polen
Allentoft et al. 2015	RISE175	Europe_LNBA	Nordic_BA	1395-1132 v. Chr.	Schweden
Allentoft et al. 2015	RISE179	Europe_LNBA	Nordic_LN	2010-1776 v. Chr.	Schweden
Allentoft et al. 2015	RISE210	Europe_LNBA	Nordic_BA	1432-1292 v. Chr.	Schweden
Allentoft et al. 2015	RISE240	Steppe_EMBA	Yamnaya_Kalmykia	2880-2632 v. Chr.	Russland
Allentoft et al. 2015	RISE247	Europe_LNBA	Vatya	1746-1611 v. Chr.	Ungarn
Allentoft et al. 2015	RISE254	Europe_LNBA	Vatya	2128-1909 v. Chr.	Ungarn
Allentoft et al. 2015	RISE276	Europe_LNBA	Nordic_LBA	794-547 v. Chr.	Dänemark
Allentoft et al. 2015	RISE391	Steppe_MLBA	Sintashta_MBA	2120-1887 v. Chr.	Kasachstan
Allentoft et al. 2015	RISE392	Steppe_MLBA	Sintashta_MBA	2126-1896 v. Chr.	Russland

Referenz	Sample ID	Gruppen ID	Kultur	Alter	Land
Allentoft et al. 2015	RISE394	Steppe_MLBA	Sintashta_MBA	1949-1754 v. Chr.	Russland
Allentoft et al. 2015	RISE395	Steppe_MLBA	Sintashta_MBA	1960-1756 v. Chr.	Russland
Allentoft et al. 2015	RISE434	Europe_LNBA	Corded_Ware_Germany	2880-2630 v. Chr.	Deutschland
Allentoft et al. 2015	RISE435	Europe_LNBA	Corded_Ware_Germany	2863-2498 v. Chr.	Deutschland
Allentoft et al. 2015	RISE436	Europe_LNBA	Corded_Ware_Germany	2868-2580 v. Chr.	Deutschland
Allentoft et al. 2015	RISE446	Europe_LNBA	Corded_Ware_Germany	2829-2465 v. Chr.	Deutschland
Allentoft et al. 2015	RISE47	Europe_LNBA	Nordic_BA	1499-1324 v. Chr.	Dänemark
Allentoft et al. 2015	RISE471	Europe_LNBA	Germany_Bronze_Age	k.A	Deutschland
Allentoft et al. 2015	RISE479	Europe_LNBA	Vatya	2000-1500 v. Chr.	Ungarn
Allentoft et al. 2015	RISE480	Europe_LNBA	Vatya	1700-1500 v. Chr.	Ungarn
Allentoft et al. 2015	RISE483	Europe_LNBA	Vatya	2000-1500 v. Chr.	Ungarn
Allentoft et al. 2015	RISE484	Europe_LNBA	Vatya	2000-1500 v. Chr.	Ungarn
Allentoft et al. 2015	RISE493	Karasuk	Karasuk	1531-1427 v. Chr.	Russland
Allentoft et al. 2015	RISE494	Karasuk	Karasuk	1416-1268 v. Chr.	Russland
Allentoft et al. 2015	RISE495	Karasuk	Karasuk	k.A	Russland
Allentoft et al. 2015	RISE496	Karasuk	Karasuk	1414-1261 v. Chr.	Russland
Allentoft et al. 2015	RISE499	Karasuk	Karasuk	k.A	Russland
Allentoft et al. 2015	RISE500	Steppe_MLBA	Andronovo	k.A	Russland
Allentoft et al. 2015	RISE502	Karasuk	Karasuk	1496-1306 v. Chr.	Russland
Allentoft et al. 2015	RISE503	Steppe_MLBA	Andronovo	1727-1511 v. Chr.	Russland
Allentoft et al. 2015	RISE505	Steppe_MLBA	Andronovo	1746-1626 v. Chr.	Russland
Allentoft et al. 2015	RISE508	Steppe_EMBA	Afanasievo	3331-2935 v. Chr.	Russland
Allentoft et al. 2015	RISE509	Steppe_EMBA	Afanasievo	2887-2677 v. Chr.	Russland
Allentoft et al. 2015	RISE510	Steppe_EMBA	Afanasievo	2851-2468 v. Chr.	Russland
Allentoft et al. 2015	RISE511	Steppe_EMBA	Afanasievo	2909-2679 v. Chr.	Russland
Allentoft et al. 2015	RISE512	Steppe_MLBA	Andronovo	1446-1298 v. Chr.	Russland
Allentoft et al. 2015	RISE546	Steppe_EMBA	Yamnaya_Kalmykia	k.A	Russland
Allentoft et al. 2015	RISE547	Steppe_EMBA	Yamnaya_Kalmykia	2887-2634 v. Chr.	Russland
Allentoft et al. 2015	RISE548	Steppe_EMBA	Yamnaya_Kalmykia	k.A	Russland

Referenz	Sample ID	Gruppen ID	Kultur	Alter	Land
Allentoft et al. 2015	RISE550	Steppe_EMBA	Yamnaya_Kalmykia	3334-2635 v. Chr.	Russland
Allentoft et al. 2015	RISE552	Steppe_EMBA	Yamnaya_Kalmykia	2849-2143 v. Chr.	Russland
Allentoft et al. 2015	RISE559	Europe_LNBA	Bell_Beaker_Germany	k.A	Deutschland
Allentoft et al. 2015	RISE560	Europe_LNBA	Bell_Beaker_Germany	k.A	Deutschland
Allentoft et al. 2015	RISE562	Europe_LNBA	Bell_Beaker_Germany	k.A	Deutschland
Allentoft et al. 2015	RISE563	Europe_LNBA	Bell_Beaker_Germany	k.A	Deutschland
Allentoft et al. 2015	RISE564	Europe_LNBA	Bell_Beaker_Germany	k.A	Deutschland
Allentoft et al. 2015	RISE577	Europe_LNBA	Unetice_EBA	k.A	Deutschland
Allentoft et al. 2015	RISE586	Europe_LNBA	Unetice_EBA	k.A	Deutschland
Allentoft et al. 2015	RISE61	Europe_LNBA	Nordic_MN_B	2851-2492 v. Chr.	Dänemark
Allentoft et al. 2015	RISE71	Europe_LNBA	Nordic_LN	2196-2023 v. Chr.	Dänemark
Allentoft et al. 2015	RISE97	Europe_LNBA	Nordic_LN	2025-1885 v. Chr.	Schweden
Allentoft et al. 2015	RISE555	Russia_EBA	Russia_EBA	2857-2497 v. Chr.	Russland
Allentoft et al. 2015	RISE602	Eisenzeit	Eisenzeit	k.A	Russland
Allentoft et al. 2015	RISE601	Eisenzeit	Eisenzeit	k.A	Russland
Allentoft et al. 2015	RISE600	Eisenzeit	Eisenzeit	k.A	Russland
Allentoft et al. 2015	RISE504	Eisenzeit	Eisenzeit	721-889 n. Chr.	Russland
Allentoft et al. 2015	RISE492	Eisenzeit	Eisenzeit	396-209	Russland
Lazaridis et al. 2014	Loschbour	WHG	Loschbour_HG	6210-5990 v. Chr.	Luxemburg
Fu et al. 2014	Ust_Ishim	Ust_Ishim	Ust_Ishim_HG	45530-40610 v. Chr.	Sibirien

B6 Screening-Ergebnisse der MiSeq-Sequenzierung aller extrahierten Proben

Auflistung der Proben mit Angabe der Laborkennzeichnung; Anz.Seq.QF= Anzahl der alignierten Sequenzen (hg19) nach dem Qualitätsfilter; endo.DNA= Sequenzen ohne Duplikate nach QF im Verhältnis zu allen ausgelesenen Sequenzen in %; Dup.rate= Duplikationsrate der *libraries* in %

Probenname	Lab.code	Anz. Seq. QF	endo. DNA [%]	Dup.rate [%]
A1	L14_1	179620	10.45	0.20
A10b	L_5 3	2607308	4.62	0.22
A17	L_5 5	2236417	9.03	0.82
A19b	L10_8	921511	65.79	0.29
A20	L_5 6	2004899	35.41	0.76
A21	L_5 7	2810095	56.36	1.37
A4b	L_5 1	2329514	0.60	0.31

Probennamen	Lab.code	Anz. Seq. QF	endo. DNA [%]	Dup.rate [%]
A6	L4_1	3124691	4.55	0.82
A8b	L_5_2	2999973	4.00	0.20
A9	L10_7	456903	21.85	0.17
Ak1_1	L6_1	1468613	13.94	0.07
Ak1-2b	L12_6	1068228	29.61	0.19
Ak4-1	L13_6	529973	19.06	0.16
Ak5_1c	L6_6	873257	12.59	0.33
Ak5_4b	L6_7	1018971	16.00	0.28
Ak5-6a	L13_7	474331	32.72	0.11
Ak5-7a	L13_8	656421	8.25	0.16
Ak5-8a	L13_9	596277	71.11	0.16
Be1	L4_3	4710027	48.15	2.28
Be1	L_5_8	2835898	45.37	3.27
Be11	L2	1065546	37.91	0.22
Be12	L2	528725	13.45	0.80
Be14	L4_5	2451348	13.71	1.77
Be5	L13_1	6136	0.10	0.00
Be7	L13_2	460788	0.46	0.00
Be9	L2	1468728	12.79	0.07
Bl1	L6_2	690214	11.00	0.37
Bl2	L6_3	487673	23.26	0.23
Bl3	L6_4	636143	44.16	0.53
Bl4	L10_5	612023	21.64	0.69
Bl5	L10_6	339118	17.84	0.74
Is1	L13_10	724276	5.35	0.19
Is3	L14_3	386863	1.14	0.09
Is4	L4_2	525120	2.24	0.88
Is4	L_5_10	2337586	2.21	0.84
J12_8	L6_5	817520	66.82	0.43
J12-1a	L12_3	202503	33.87	0.34
J12-2a	L12_1	591230	2.28	0.03
J12-3	L12_4	453035	0.88	0.03
J12-6a	L12_5	431023	67.04	0.14
J12-7a	L13_5	771792	26.97	0.15
J12-9a	L12_2	252535	15.03	89.52
Kal 1	LC8_1	57528	8.52	0.00
KAMA 1	LC6_1	342092	25.80	2.05
KAMA 10	LC6_7	513710	5.31	0.34
KAMA 2	LC6_2	414972	15.19	1.10
KAMA 3	LC6_3	636578	0.14	0.00
KAMA 6	LC6_4	464613	7.14	0.43
KAMA 7	LC6_5	645380	48.04	0.42
KAMA 8	LC6_6	459487	7.68	1.08
Kol4	L11_3	647105	6.83	0.11
Kol5	L11_4	636366	0.54	0.06
Nov7	L14_6	562547	6.89	0.16
Nov8	L14_7	546815	1.26	0.04
Pr1	L9_1	1104366	0.50	0.02
PR10	L10_1	706498	16.10	0.14
PR11	L10_2	749830	4.03	0.25
PR12	L10_3	679829	5.22	0.20
PR13	L10_4	726482	1.57	0.08
Pr2	L9_2	869589	0.36	0.00

Probennamen	Lab.code	Anz. Seq. QF	endo. DNA [%]	Dup.rate [%]
Pr3	L9_3	742546	3.56	0.20
Pr4	L9_4	672496	10.14	0.17
Pr5	L9_5	577434	3.36	0.09
Pr6	L9_6	404840	0.99	0.10
Pr7	L9_7	669500	0.73	0.02
Pr8	L9_8	854234	4.63	0.09
Pr9	L9_9	635867	1.41	0.12
Sol5-1a	L11_1	322280	6.40	0.13
Sol5-9a	L11_2	418590	12.57	0.20
Ze10a	L14_5	393744	9.70	0.10
Ze2	L4_6	3161082	7.34	0.35
Ze3	L4_7	2617489	15.98	0.83
Ze4	L10_9	507663	14.43	0.35
Ze7	L10_10	379427	4.19	0.36
Ze8	L4_8	4003234	3.09	0.24
Ze8b	L11_5	336594	1.61	0.15
Ze9a	L14_4	436331	1.13	0.06

B7 Koordinaten der Fundplätze für die Kartenerstellung in R

Population	Fundplatz	Längengrad	Breitengrad	Farbe	Gruppe
A	Aržan 2	93.711133	52.096067	maroon1	Eisenzeit
Ak	Ak-Alacha 1	87.546033	49.2954	maroon1	Eisenzeit
Be	Berel	86.20525	49.343533	maroon1	Eisenzeit
Bl	Barsučij Log	91.1971	54.008083	maroon1	Eisenzeit
J12	Justid XII	89.3	49.774533	maroon1	Eisenzeit
Ze	Zevakino	81.832167	50.20775	maroon1	Eisenzeit
Kol	Kolbino	39.183333	51.616667	maroon1	Eisenzeit
Sol5	Solonzy 5	82.927818	55.03923	aquamarine1	Neolithisch
Kama	Kamennyj Ambar 5	60.637378	53.055399	gold	Bronzezeit
Kal	Kalinovka	51.289	53.8338	gold	Bronzezeit
Ul	Ulžan	76.8681	43.299633	maroon1	Eisenzeit
Pr	Pokrovka	53.933333	52.066667	maroon1	Eisenzeit
Nov	Novozavedennoe 2	44.266667	43.633333	maroon2	Eisenzeit

Be10	NC10-10	18294012	17862115	97.64	86.54	6760605	4997559	1669355	245555	36.86
Be14	NC6b-2	17992876	16829129	93.53	54.13	2806613	2015905	308379	3651	
	NC7a-3	28630300	25862860	90.33	78.20	2055564	67086	1826659	14053	
	NC7b-3	41513416	38183206	91.98	61.59	6274822	4535298	1280424	9602	
	mergeSAMPE	88136592							19630	48.04
Bl1	NC6b-4	18274224	17110938	93.63	47.14	2598654	1997063	264933	5467	
	NC7b-4	32036778	30099199	93.95	53.30	4269364	3239138	823945	5890	
	mergeSAMPE	50311002							10942	49.1
Bl2	NC6-5	11364624	7948581	69.94	82.24	1980742	1428643	426265	11755	
	NC7-6	15199710	10940541	71.98	82.28	2711071	1955162	638266	12210	
	merge	26564334							21751	36.82
Bl3	NC10-5	30160146	29021344	96.22	86.11	9669325	6315676	2524260	33111	36.84
J12_6	NC9-10	15181058	14010306	92.29	95.59	4202858	2354940	1647481	96416	36.96
J12_7	NC10-12	20701730	17334489	83.73	92.94	5708758	4398350	956999	76725	36.95
J12_8	NC9-9	21812394	20368059	93.38	88.85	7745187	4101683	2809931	56608	36.98
Ze3	NC6-3	10505868	6706242	63.83	78.54	1354703	1116439	164533	7058	
	NC7-5	9718694	6419264	66.05	80.16	1385674	1095714	254604	21485	
	mergeSAMPE	20224562							39588	41.92
Ze4	NC11-8	38488430	35708391	92.78	89.24	12280951	10820789	792281	28195	36.87
Ko14	NC10-1	13069072	11927677	91.27	92.32	2988717	2440647	389665	39358	36.89
Sol5_1a	NC11-4	38984574	36041463	92.45	82.14	11228032	10457731	316394	27783	36.87
Sol5_9a	NC10-14	23742836	22830617	96.16	88.37	8100872	6783874	808866	40956	36.88
Kama7	NC11-5	2621188	2513181	95.88	91.62	1050658	941696	95429	15043	36.78
Kal1	NC10-13	22961228	21855113	95.18	80.07	6016706	5027454	580926	36167	36.96
U12	NC10-11	16276618	15609745	95.90	94.96	5812478	3878045	1783545	64765	36.92
Pr3	NC9-4	14459962	13026879	90.09	84.38	1499422	1174894	213143	19056	36.91
Pr4	NC10-3	26334786	24887320	94.50	80.60	5939810	4742872	801569	41364	36.94
Pr8	NC11-1	13583716	12177680	89.65	93.13	3213995	3008030	74017	15313	

	NC9-11	15827532	14104318	89.11	94.88	2434642	1948775	159547	7086	
	merge	29411248							22384	36.66
Pr10	NC9-7	14989177	13726976	91.58	87.63	2113122	1471642	612244	86987	36.99
Pr11	NC9-5	22310864	19108507	85.65	87.05	3652149	2823628	430583	17785	36.79
Pr12	NC9-6	21711261	18955793	87.31	93.93	2853383	2119515	434876	25454	36.75
Be11	NC8-5	7114120	5119508	71.96	82.53	678566	283202	370611	9279	36.78
J12_1	NC9-8	16488033	15097431	91.57	95.33	5303886	3480974	808497	8505	36.94
Bl4	NC10-8	52802340	48691187	92.21	80.99	11758066	7931732	1900921	4795	36.79
Kama1	NC10-7	13127002	12727075	96.95	87.60	4005316	2832966	531480	3418	36.87
Kama6	NC11-6	26869144	24753445	92.13	88.64	7589242	6771482	205495	3210	36.64
Kama10	CBI	94726458	79531892	83.96	90.84	16861823	12739565	859087	3398	36.73
Nov7	NC10-2	15155316	14582400	96.22	87.01	3389801	2845708	302191	9404	36.83
Bl5	NC10-9	20098630	19384409	96.45	87.81	4812843	3363061	756828	4089	36.74

B9 Ergebnisse der Kontaminationsbestimmung mit den Programmen contaMIX und ANGSD und Berechnung der Deaminierungsrate am 5'-Ende;

FR=geschätzte Fehlerrate; Auth.= Authentizität in %; Kon.=Kontamination

Probe	contaMIX		ANGSD - Methode 1			ANGSD - Methode 2			mapDamage
	FR	Auth. [%]	Kon. [%]	SE	p-Wert	Kon. [%]	SE	p-Wert	mD5 [%]
A19	0.01	99.44	0.01	0	1	0.01	0	1	8.5
A20	0.01	98.93	0.01	0	1	0.01	0	1	18.06
A21	0.01	70.06	1.33	0.01	0.232	4.22	0.03	0.197	6.12
A9	0.01	96.85							19.04
Ak1_1	0.01	99.92	3.64	0.01	0.00	4.32	0.02	0.001	6.33
Ak1_2	0	97.20	0.01	0	1	0.01	0	1	10.6
Ak5_6	0	98.73	0.01	0	1	0.01	0	1	7.85
Ak5_8	0.01	96.85	0.31	0	0.52	0.41	0.01	1	2.75
Be1	0.01	99.21	3.30	0.03	0.37	0.01	0	1	9.53
Be10	0	95.20	3.75	0.04	0.06	0.01	0	1	17.4
Be11	0	98.35	k.A.						18.90
Be14	0.01	99.92	1.82	0.02	0.3	0.01	0	1	11.39
Be4	0.01	99.79							9.68
Be9	0	98.83							16.62
Bl1	0.01	95.24	0.01	0	1	0.01	0	1	10.31
Bl2	0.01	95.16	6.13	0.06	0.26	9.15	0.09	0.21	15.77
Bl3	0.01	75.42	0.01	0	0.35	1.27	0.02	1	16.78
Bl4	0.01	98.39	k.A.						17.96
Bl5	0	99.62							22.64
J12_1	0	85.61	2.07	0.02	0.3	0.01	0	1	10.4
J12_6	0.01	99.90	0.01	0	0.62	0.01	0	1	8.34
J12_7	0.01	98.91	0.01	0	1	0.01	0	1	11
J12_8	0	97.80							4.3
Kal1	0	99.71	7.56	0.04	0	11.70	0.05	0.01	14.4
Kama1	0.01	91.74	k.A.						11.64
Kama10	0.01	99.82	k.A.						20.15
Kama6	0.01	99.65	k.A.						19.95
Kama7	0.01	99.18							17.65
Kol4	0	99.51	12.10	0.05	0.06	0.01	0	1	33.38
Nov7	0.01	99.85	k.A.						17.47
Pr10	0.01	97.27	0.01	0	1	0.01	0	1	7.5
Pr11	0.01	95.03	0.01	0.09	1	0.01	0.09	1	24
Pr12	0	98.89	0.01	0.09	1	0.01	0.09	1	21.8
Pr3	0	96.14	0.01	0.09	1	0.01	0.09	1	13.32
Pr4	0.01	99.28	0.94	0	1	0.01	0	1	11.51
Pr8	0.01	95.31							28.75
Sol5_1	0.01	94.03	6.90	0.13	0.22	0.01	0.08	1	21.23
Sol5_9	0	91.41	0.69	0.02	0.66	0.01	0	1	17.84
Ul2	0	98.71	0.30	0	1	0.01	0	1	31.69
Ze3	0	99.63	0.01	0	1	7.36	0.08	0.38	14.12
Ze4	0.01	98.48	0.01	0	1	0.01	0	1	13.44

B10 Phänotyp-Bestimmung der Augen- und Haarfarbe mit dem HIRISplex-System (Walsh et al. 2014) für alle skythischen Proben mit Angabe der P-Werte für jede Kategorie

NA= keine Angaben möglich

Probe	Augenfarbe			Haarfarbe					
	P-blau	P-Interm.	P-braun	P-blond	P-braun	P-rot	P-schwarz	P-hell	P-dunkel
Nov7	0.175	0.205	0.619	NA	NA	NA	NA	0.226	0.774
A19	0.034	0.066	0.901	NA	NA	NA	NA	0.301	0.699
A20	0.000	0.005	0.995	0.000	0.053	0.000	0.947	0.001	0.999
A21	0.000	0.004	0.996	0.002	0.064	0.000	0.934	0.003	0.997
A9	0.001	0.029	0.969	NA	NA	NA	NA	0.153	0.847
Ak1-1	0.016	0.054	0.930	0.002	0.109	0.000	0.889	0.004	0.996
Ak1-2	0.016	0.054	0.930	0.002	0.109	0.000	0.889	0.004	0.996
Ak5-6	0.000	0.008	0.992	0.036	0.272	0.000	0.693	0.074	0.926
Ak5-8	0.709	0.104	0.186	0.352	0.366	0.002	0.280	0.624	0.376
Be1	0.000	0.008	0.992	NA	NA	NA	NA	0.043	0.957
Be10	0.000	0.016	0.983	0.097	0.375	0.001	0.527	0.186	0.814
Be14	0.030	0.101	0.869	NA	NA	NA	NA	0.055	0.945
Be4	0.141	0.152	0.707	NA	NA	NA	NA	0.323	0.677
Be9	0.004	0.023	0.973	0.019	0.244	0.000	0.736	0.032	0.968
Bl1	0.060	0.122	0.818	0.029	0.445	0.001	0.525	0.064	0.936
Bl2	0.122	0.145	0.734	NA	NA	NA	NA	0.720	0.280
Bl3	0.958	0.030	0.013	0.718	0.234	0.011	0.037	0.944	0.056
J12-1a	0.002	0.032	0.966	NA	NA	NA	NA	0.218	0.782
J12-6	0.000	0.003	0.997	0.001	0.051	0.000	0.948	0.003	0.997
J12-7	0.000	0.003	0.997	0.006	0.226	0.000	0.769	0.007	0.993
J12-8	0.000	0.008	0.992	0.018	0.210	0.000	0.772	0.030	0.970
Kal1	0.007	0.043	0.949	0.006	0.429	0.000	0.564	0.012	0.988
Kama7	0.964	0.030	0.007	NA	NA	NA	NA	0.832	0.168
Kol4	0.154	0.158	0.688	0.403	0.448	0.003	0.146	0.800	0.200
Pr10	0.926	0.048	0.026	0.725	0.234	0.005	0.035	0.951	0.049
Pr11	0.000	0.012	0.988	0.036	0.400	0.000	0.564	0.069	0.931
Pr12	0.000	0.003	0.997	0.042	0.950	0.007	0.000	0.012	0.988
Pr3	0.175	0.205	0.619	0.114	0.659	0.003	0.224	0.317	0.683
Pr4	0.127	0.147	0.726	0.212	0.340	0.001	0.446	0.393	0.607
Pr8	0.000	0.014	0.985	NA	NA	NA	NA	0.045	0.955
Sol5-1a	0.001	0.021	0.978	0.003	0.098	0.000	0.899	0.109	0.891
Sol5-9	0.000	0.003	0.997	0.003	0.094	0.000	0.904	0.004	0.996
Ul2	0.008	0.029	0.963	0.040	0.416	0.000	0.544	0.124	0.876
Ze3	0.958	0.030	0.013	0.756	0.146	0.082	0.016	0.956	0.044
Ze4	0.000	0.008	0.992	0.058	0.406	0.006	0.530	0.116	0.884

Be11 Abdeckung aller angereicherten, funktionalen Marker und Auswahl der diskutierten Loci; cov = Abdeckung; SD = Standardabweichung

Probe	Funktionale Marker			Auswahl funktionaler Marker		
	cov	SD	Anzahl Loci	cov	SD	Anzahl Loci
A9	9.26	3.63	192	9.56	2.98	37
A19	12.73	4.53	192	12.83	4.37	37
A20	24.61	8.27	192	25.10	7.80	37
A21	16.33	5.62	193	16.28	5.09	37
Ak1-1	123.66	35.77	193	129.94	29.20	37
Ak1-2	18.89	6.68	193	19.61	5.83	37
Ak5-6	28.59	8.56	193	30.19	7.48	37
Ak5-8	96.67	28.03	193	104.00	21.17	37
Be4	4.99	2.57	188	4.83	2.25	36
Be9	52.28	14.72	193	54.25	12.36	37
Be10	109.63	33.12	193	116.50	29.47	37
Be11	4.17	2.16	186	4.37	2.32	36
Be1	5.48	2.48	192	5.42	2.34	37
Be14	7.78	3.67	192	8.31	3.86	37
Bl1	4.48	2.63	189	4.58	2.73	37
Bl2	5.39	2.35	100	6.14	2.06	37
Bl3	14.58	5.32	193	15.67	4.33	37
Bl4	2.28	1.38	166	2.03	1.26	32
Bl5	2.29	1.26	153	1.96	1.00	30
J12-1	4.01	2.11	183	4.85	2.64	34
J12-6	40.94	12.52	193	41.36	10.43	37
J12-7	40.62	16.34	193	44.44	16.97	37
J12-8	23.35	6.64	193	23.75	5.92	37
Kal1	16.23	5.03	192	19.89	4.93	37
Kama1	1.98	1.30	128	1.93	1.57	30
Kama6	1.91	1.18	132	1.77	0.85	27
Kama7	6.73	3.20	189	7.11	3.33	37
Kama10	1.84	1.03	142	1.81	1.00	27
Kol4	13.94	6.63	191	13.98	6.75	37
Nov7	4.01	2.30	186	4.36	2.68	37
Pr3	8.62	3.66	190	8.72	3.69	37
Pr4	16.69	5.75	193	16.92	5.45	37
Pr8	5.43	2.98	187	5.89	2.86	36
Pr10	38.32	13.37	191	40.61	12.74	37
Pr11	7.43	3.13	192	8.36	3.14	37
Pr12	8.98	3.84	192	9.50	3.49	37
Sol5-1	10.41	4.24	192	10.28	4.74	37
Sol5-9	16.11	6.02	193	15.61	5.56	37
Ul2	115.70	40.29	193	199.92	38.11	37
Ze3	16.66	7.94	191	17.06	7.39	37
Ze4	12.68	4.35	193	12.67	3.61	37

B12 Genotypen der Marker für Laktasepersistenz

Kategorie	Laktasepersistenz		
Bezeichnung	LCTa	LCTb	
rs-Nr.	rs4988235	rs182549	
CHR	2	2	
Position	136608646	136616754	
anz. / abgel.	G/A	C/T	persistent
A9		C/T	+
A19	G/G		-
A20	G/G	C/C	-
A21	G/G	C/C	-
Ak1-1	G/G	C/C	-
Ak1-2	G/G		-
Ak5-6	G/G	C/C	-
Ak5-8	G/G	C/C	-
Be9	G/G	C/C	-
Be10	G/G	C/C	-
Bl1		C/T	+
Bl2	G/A		+
Bl3	G/G		-
J12-6	G/G	C/C	-
J12-7	G/G	C/C	-
J12-8	G/G	C/C	-
Kal1	G/G	C/C	-
Kol4	G/G		-
Sol5-1	G/G		-
Sol5-9	G/G	C/T	+
Ul2	G/G	C/C	-
Pr3	G/G		-
Pr4	G/A	C/T	+
Pr8	G/G		-
Pr10	G/G	C/C	-
Pr11		C/T	+
Pr12	G/G		-
Zc3	G/G	C/C	-
Zc4		C/C	-

B13 Genotypen der Marker für Klimaanpassung

Kategorie	Klimaanpassung	
Bezeichnung	ABCC11	EDAR
rs-Nr.	rs17822931	rs3827760
CHR	16	2
Position	48258198	109513601
anz. / abgel.	C/T	A/G
A9	C/T	
A19		A/A
A20	C/C	A/A
A21	T/T	
Ak1-1	C/T	A/A
Ak1-2	C/T	A/A
Ak5-6	C/C	A/A
Ak5-8	C/T	A/G
Be1		A/G
Be4	C/T	
Be9	T/T	A/G
Be10	T/T	A/G

Kategorie	Klimaanpassung	
	ABCC11	EDAR
rs-Nr.	rs17822931	rs3827760
CHR	16	2
Position	48258198	109513601
anz. / abgel.	C/T	A/G
Be11		A/G
Bl1		A/G
Bl3	C/C	G/G
J12-6	C/T	G/G
J12-7	C/T	A/A
J12-8	C/T	A/A
Kal1	C/C	A/A
Kol4		A/A
Ul2	C/C	A/A
Pr4		A/G
Pr10	C/C	A/A
Pr11		A/A
Ze3		A/G
Ze4	C/T	A/A

B14 Genotypen der Marker für Hautpigmentierung nach (Spichenok et al. 2011)

Kategorie	Hautpigmentierung						
	SLC24A5	OCA2	MC1R.8	HERC2	IRF4.1	SLC45A2.1	
rs-Nr.	rs1426654	rs1545397	rs885479	rs12913832	rs12203592	rs16891982	
CHR	15	15	16	15	6	5	
Position	48426484	28187772	89986154	28365618	396321	33951693	Pigm.
anz. / abgl.	G/A	A/T	G/A	A/G	C/T	C/G	
A9	A/G	A/T			C/C		hell - interm.
A19	A/A	A/T	G/G	A/G	C/C	C/G	hell - interm.
A20	A/A	A/T	G/A	A/A	C/T	C/C	hell - interm.
A21	A/G	A/T	G/A	A/A	C/C	C/C	hell - interm.
Ak1-1	A/A		G/A	A/G	C/T	C/C	hell - interm.
Ak1-2	A/A	A/A	G/A	A/G	C/T	C/C	hell - interm.
Ak5-6	A/G	A/T	G/G	A/A	C/C	C/G	hell - interm.
Ak5-8	A/A	A/A	G/A	G/G	C/C	C/G	hell - interm.
Be1	A/G					C/G	hell - interm.
Be4				A/G			nicht möglich
Be9	A/A	A/T	G/A	A/G	C/C	C/C	hell - interm.
Be10	A/G	T/T	G/A	A/A	C/C	G/G	hell - interm.
Be11						C/G	hell - interm.
Be14			G/A	A/G	C/T	C/G	hell - interm.
Bl1		A/T					nicht

							möglich
Kategorie	Hauptpigmentierung						
Bezeichnung	SLC24A5	OCA2	MC1R.8	HERC2	IRF4.1	SLC45A2.1	
rs-Nr.	rs1426654	rs1545397	rs885479	rs12913832	rs12203592	rs16891982	
CHR	15	15	16	15	6	5	
Position	48426484	28187772	89986154	28365618	396321	33951693	Pigm.
anz. / abgl.	G/A	A/T	G/A	A/G	C/T	C/G	
Bl2	A/A	A/T				G/G	hell - interm.
Bl3		A/T	G/G		C/C	G/G	hell - interm.
J12-1							nicht möglich
J12-6	G/G	T/T	A/A	A/A	C/C	C/C	hell - interm.
J12-7	A/A	T/T	G/G	A/A	C/C	C/C	hell - interm.
J12-8	A/A	A/T	G/A	A/A	C/C	C/G	hell - interm.
Kama7					C/T		nicht möglich
Kal1	A/A	A/A	G/G	A/G	C/T	C/C	hell - interm.
Kol4	A/A	A/A		A/G	C/C	G/G	hell - interm.
Nov7					C/T		nicht möglich
Sol5-1	A/G	A/T	G/A	A/A	C/C		hell - interm.
Sol5-9		T/T	G/A	A/A	C/C	C/C	hell - interm.
Ul2	A/A	A/A	G/G	A/G	C/C	C/C	hell - interm.
Pr3	A/A	A/T		A/G	C/T		hell - interm.
Pr4			G/A	A/G			hell - interm.
Pr8			G/A	A/A		C/G	hell - interm.
Pr10	A/G	A/A	G/G	G/G	C/C	G/G	hell - interm.
Pr11			G/G	A/A		C/C	hell - interm.
Pr12	A/A	A/T	G/G	A/A	C/C	C/C	hell - interm.
Ze3	A/A	A/A	G/G	G/G	C/C	G/G	hell - interm.
Ze4	A/A	A/T	G/G		C/C	C/G	hell - interm.

B15 Genotypen der Marker für Hautpigmentierung

Kategorie	Hautpigmentierung			
	TYR.1	OCA2.3	SLC24A5	SLC45A2.1
rs-Nr.	rs1042602	rs1800414	rs1426654	rs16891982
CHR	11	15	15	5
Position	88911696	28197037	48426484	33951693
anz. / abgl.	C/A	T/C	G/A	C/G
A9	C/C	T/T	A/G	
A19	C/C	T/T	A/A	C/G
A20	C/C	T/T	A/A	C/C
A21	C/C	T/T	A/G	C/C
Ak1-1	C/C	T/T	A/A	C/C
Ak1-2	C/C	T/T	A/A	C/C
Ak5-6	C/C	T/T	A/G	C/G
Ak5-8	C/C	T/T	A/A	C/G
Be1			A/G	C/G
Be4				
Be9	C/A	T/T	A/A	C/C
Be10	C/C	T/T	A/G	G/G
Be11				C/G
Be14				C/G
Bl1				
Bl2	C/A		A/A	G/G
Bl3	C/A	T/T		G/G
J12-1	C/A	T/C		
J12-6	C/C	T/C	G/G	C/C
J12-7	C/A	T/C	A/A	C/C
J12-8	C/C	T/T	A/A	C/G
Kama7	C/C			
Kal1	A/A	T/T	A/A	C/C
Kol4	C/C	T/T	A/A	G/G
Nov7				
Sol5-1	A/A	T/T	A/G	
Sol5-9	C/C	T/T		C/C
Ul2	C/C	T/T	A/A	C/C
Pr3	C/A	T/T	A/A	
Pr4	C/C	T/T		
Pr8				C/G
Pr10	C/C	T/T	A/G	G/G
Pr11				C/C
Pr12	C/C	T/T	A/A	C/C
Ze3	C/C	T/T	A/A	G/G
Ze4	C/C	T/T	A/A	C/G

B16 Genotypen der Marker für ein erhöhtes Diabetes Risiko mit standardisierter Risikobewertung (GRS-Wert) nach (Cornelis et al. 2009)

Kategorie	Diabetes Typ2											
Bezeichnung	CDKN2B	SLC30A8	CDKAL1_a	IGF2BP2	WFS1	PPARG	KCNJ11	HHEX/ IDE			TCF7L2	TCF7L2
rs-Nr.	rs564398	rs13266634	rs7754840	rs4402960	rs10010131	rs1801282	rs5215	rs1111875			rs7903146	rs11196205
CHR	9	8	6	3	4	3	11	10			10	10
Position	22029547	118184783	20661250	185511687	6292915	12393125	17408630	94462882			114758349	114807047
R / KR	T/C	C/T	C/G	T/G	G/A	C/G	C/T	C/T	GRS	std GRS	T/C	C/G
A9	T/T	C/T		G/T	A/G				5	10.00	C/T	G/C
A19	T/T	C/T		G/G	A/G	C/G	C/T	T/T	6	6.86	C/C	G/G
A20	T/T	C/T	G/C	G/T	A/A	C/C	C/T	T/T	10	10.00	-	G/G
A21	T/T	C/T	G/C	G/T	A/G	C/C	C/T	T/T	9	9.00	C/C	G/C
Ak1-1	T/C	C/T	G/G	G/G	A/G	C/C	T/T	C/T	7	7.00	C/C	G/C
Ak1-2	T/C	C/T		G/G	A/G	C/C	T/T	C/T	6	6.86	C/C	C/C
Ak5-6	T/T	C/T	G/G	G/T	G/G	C/C	C/T	T/T	8	8.00	C/C	G/G
Ak5-8	T/T	C/T	G/C	G/T	A/G	C/G	C/C	C/T	10	10.00		G/C
Be9	T/C	C/T	G/C	G/G	G/G	C/C	T/T	C/T	8	8.00	C/C	G/G
Be10	T/T	C/T	G/C	G/G	A/G	C/C	C/C	C/T	10	10.00	C/C	G/C
Be14	T/C			G/G	A/G		C/T	C/T	4	6.40	C/T	
Bl3	T/T	C/T	G/G		A/G	C/C	C/T	C/T	8	9.14	C/T	G/C
J12-6	T/T	C/T	G/C	G/T	G/G	C/G	T/T	T/T	8	8.00	C/C	
J12-7	T/T	T/T	G/G	T/T	A/G	C/G	C/C	C/C	10	10.00	C/C	G/G
J12-8	T/C	C/T	G/G	G/T	A/G	C/C	C/C	T/T	8	8.00	C/C	G/G
Kama7		C/T				C/G	C/T	C/T	4	8.00	C/C	G/C
Kal1	T/C	C/C	G/G	G/T	A/G	C/C	C/T	T/T	6	6.00		G/C
Kol4	C/C	T/T	G/G	G/G		C/C	T/T	C/C	4	4.57	C/C	G/G
Sol5-1		C/C	G/C	G/T		C/G	C/T	T/T	6	8.00	C/C	G/G
Sol5-9	T/C		G/C	G/T	G/G	C/G	T/T	C/T	7	8.00	T/T	C/C
Ul2	C/C	C/C	G/C	G/T	G/G	C/C	C/T	C/C	11	11.00		C/C
Pr3		T/T	G/C			C/C			3	8.00	C/C	G/G
Pr4	T/C		G/C	G/G		C/C	T/T	C/T	5	6.67		G/G
Pr10	T/C	C/C	C/C	T/T	G/G	C/G	C/T	C/T	12	12.00	C/C	G/G
Pr11				G/G	A/A	C/G		C/T	2	4.00	C/C	G/C
Pr12	T/C	C/T	C/C	G/T		C/C	C/T		8	10.67		G/G
Ze3	T/C	C/C	G/G	G/T	A/G	C/C		C/T	8	9.14	C/C	G/G
Ze4	T/T		G/C		A/G	C/C	T/T	T/T	6	8.00		G/C

B17 Genotypen der Marker für Xenobiotischen Metabolismus

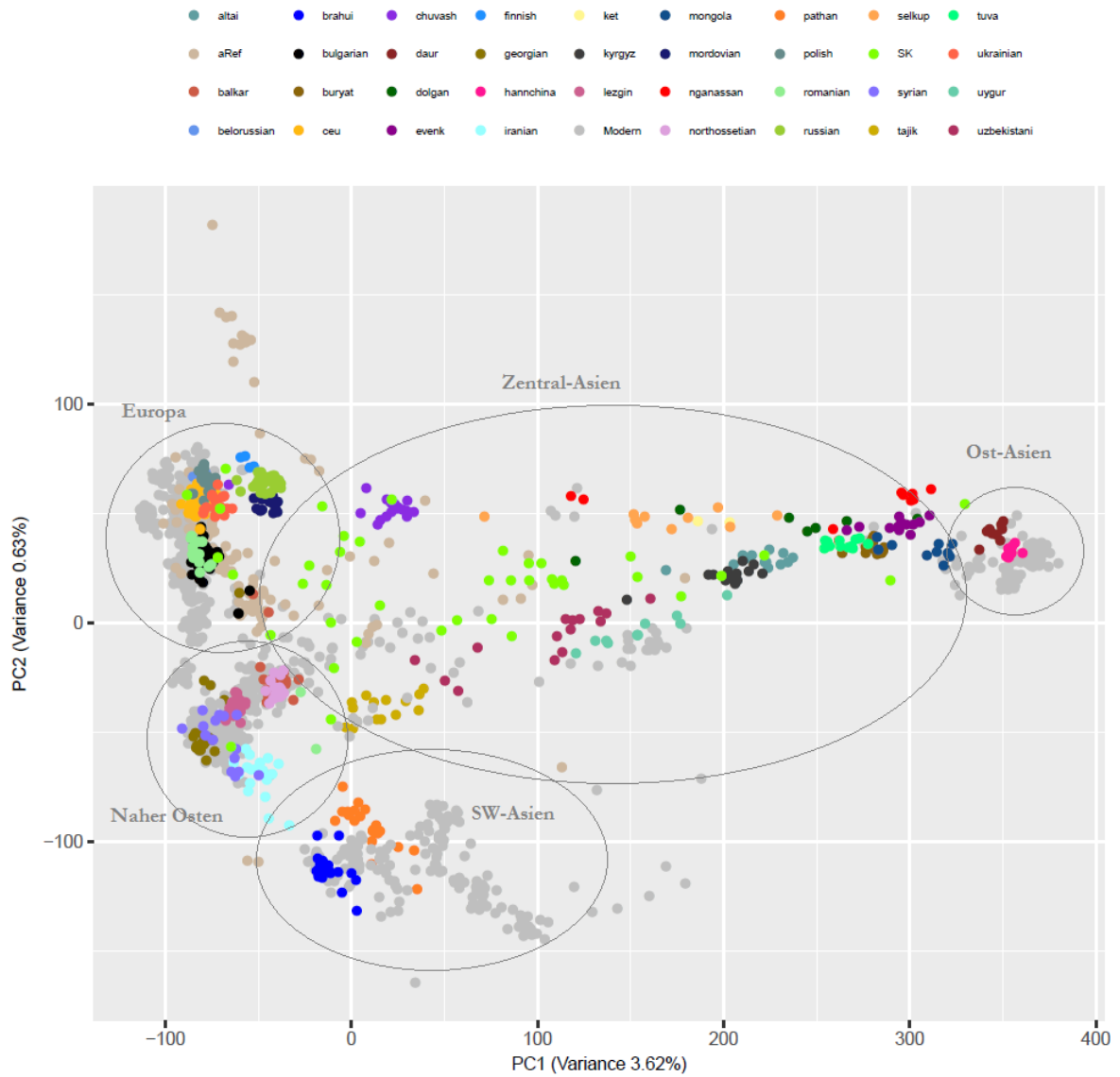
Kategorie	Xenobiotischer Metabolismus								
Bezeichnung	ABCB1a	ABCB1b	ABCB1c	NAT2	NAT2	AGT	AGT	CYP3A4	CYP3A5
rs-Nr.	rs1128503	rs2032582	rs1045642	rs1041983	rs1801280	rs699	rs5051	rs2740574	rs776746
CHR	7	7	7	8	8	1	1	7	7
Position	87179601	87160618	87138645	18257795	18257854	230845794	230849872	99382096	99270539
anz. / abgel.	G/A	C/A,T	G/A	C/T	T/C	G / A	T / C	C/T	T/C
A9	A/G	-	A/G	T/T	-	-	C/T	-	-
A19	A/G	-	A/G	C/T	T/T	A/G	C/T	-	C/T
A20	A/G	C/C	G/G	C/C	T/C	A/G	C/T	T/T	C/C
A21	-	C/C	G/G	C/C	T/T	A/G	C/T	T/T	C/C
Ak1-1	A/A	A/A	A/A	C/T	T/T	G/G	T/T	T/T	C/T
Ak1-2	-	A/A	A/A	C/T	T/T	G/G	T/T	T/T	C/T
Ak5-6	G/G	C/C	A/G	C/T	T/C	A/G	C/T	T/T	C/C
Ak5-8	A/A	A/A	A/A	C/C	T/T	A/A	C/C	T/T	C/C
Be1	-	-	-	-	-	-	-	-	C/T
Be4	-	-	-	-	-	-	-	-	C/T
Be9	A/G	A/C	A/A	C/C	C/C	A/G	C/T	T/T	C/C
Be10	A/G	A/C	A/A	C/C	T/C	A/A	C/C	C/T	C/T
Be11	-	-	-	-	T/C	-	-	-	-
Be14	-	-	-	-	-	-	-	-	C/T
Bl2	-	A/C	-	-	T/C	-	-	-	-
Bl3	G/G	A/C	A/G	C/C	T/C	A/G	C/T	-	C/C
J12-1	-	-	-	-	-	-	C/T	-	-
J12-6	G/G	A/C	G/G	C/T	T/T	G/G	T/T	T/T	C/C
J12-7	A/G	A/C	A/A	C/C	C/C	A/G	C/T	T/T	C/C
J12-8	G/G	C/T	G/G	C/C	T/C	A/A	C/C	T/T	C/C
Kama7	-	-	A/G	-	-	A/G	-	-	-
Kal1	A/G	A/C	A/A	C/T	T/C	A/G	C/T	T/T	C/C
Kol4	A/G	-	A/G	C/T	T/T	-	-	-	C/C
Sol5-1	-	-	-	-	-	-	-	-	C/C
Sol5-9	A/A	-	A/G	C/T	T/T	A/G	C/T	T/T	C/C
Ul2	A/G	A/C	A/G	C/C	T/T	G/G	T/T	T/T	C/C
Pr3	-	-	A/G	-	-	-	-	-	C/C
Pr4	A/A	A/A	-	C/T	T/C	-	T/T	T/T	C/C

Kategorie	Xenobiotischer Metabolismus								
Bezeichnung	ABCB1a	ABCB1b	ABCB1c	NAT2	NAT2	AGT	AGT	CYP3A4	CYP3A5
rs-Nr.	rs1128503	rs2032582	rs1045642	rs1041983	rs1801280	rs699	rs5051	rs2740574	rs776746
CHR	7	7	7	8	8	1	1	7	7
Position	87179601	87160618	87138645	18257795	18257854	230845794	230849872	99382096	99270539
anz. / abgel.	G/A	C/A,T	G/A	C/T	T/C	G / A	T / C	C/T	T/C
Pr10	A/A	A/A	A/G	C/C	T/C	G/G	T/T	-	C/C
Pr11	-	-	-	-	T/T	A/A	-	-	-
Pr12	-	-	-	-	-	-	-	T/T	-
Ze3	-	-	G/G	C/T	T/C	A/A	-	-	C/C
Ze4	-	A/C	-	C/C	T/C	A/G	C/T	T/T	-

B18 Genotypen der Marker für Alkoholverträglichkeit und Bluthochdruck

Kategorie	Alkoholverträglichkeit			Bluthochdruck	
	ADH1Bb	ADH1Ba	ALDH2	SH2B3	SH2B3/ATXN2
rs-Nr.	rs1229984	rs3811801	rs671	rs3184504	rs653178
CHR	4	4	12	12	12
Position	100239319	100244319	112241766	111884608	112007756
anz. / abgl.	C/T	G/A	G/A	C/T	T/C
A9	C/C	G/G		T/C	C/T
A19	C/C	G/G	G/G	-	T/T
A20	C/C	G/G	G/G	C/C	T/T
A21	C/C	G/G	G/G	-	T/T
Ak1-1	C/C	G/G	G/G	C/C	T/T
Ak1-2	C/C	G/G	G/G	C/C	T/T
Ak5-6	C/C	G/G	G/G	C/C	T/T
Ak5-8	C/C	G/G	G/G	C/C	T/T
Be1			G/G	-	-
Be4		G/G		-	-
Be9	C/C	G/G		C/C	T/T
Be10	C/C	G/G	G/G	C/C	T/T
Be11			G/G	-	
Bl1			G/G	-	T/T
Bl2			G/G		
Bl3	C/C	G/G	G/G	T/C	C/T
Bl4			G/G	T/C	C/T
J12-6	C/C	G/G		-	-
J12-7	C/C	G/G	G/G	C/C	T/T
J12-8	C/C	G/G	G/G	C/C	T/T
Kama1			G/G	T/C	C/T
Kama7		G/G		-	-
Kal1	C/C	G/G		-	C/T
Kol4	C/C		G/G	-	T/T
Nov7			G/G	-	T/T
Sol5-1	C/C	G/G		-	-
Sol5-9	C/C	G/G		-	T/T
Ul2	C/C	G/G		C/C	T/T
Pr3	C/C		G/G	T/T	C/C
Pr4	C/C		G/G	-	
Pr8			G/G	T/C	C/T
Pr10	C/C	G/G		-	-
Pr11		G/G	G/G	C/C	T/T
Pr12		G/G		C/C	-
Ze3	C/C	G/G	G/G	-	-
Ze4	C/C	G/G			T/T

B19 PCA der untersuchten Proben (SK) und zusätzlichen aDNA Referenzen (aRef); ausgewählte moderne Populationen sind hervorgehoben



Tabellenverzeichnis

Tabelle 1 Übersicht aller Proben, die für eine erste Analyse ausgewählt bzw. für die vorliegende Arbeit neu extrahiert wurden	- 20 -
Tabelle 2 Verwendete Primer und Zyklenanzahl für die Amplifikation der <i>screening</i>- und <i>capture-libraries</i>; * = eine 8bp lange, spezifische Indexsequenz aus Kircher et al. (2012)	- 25 -
Tabelle 3 Vergleich getesteter <i>bait</i>-Sequenzen der Hersteller Agilent, Roche und MYcroarray	- 32 -
Tabelle 4 Unterschiede der Laborprotokolle der getesteten Anreicherungs-Kits -	37 -
Tabelle 5 Plattform und Konditionen der sequenzierten <i>capture</i>-Produkte; PE= paired-end; bp= Basenpaar	- 38 -
Tabelle 6 Programme und Parameter der Rohdatenbearbeitung	- 39 -
Tabelle 7 Auflistung der modernen Populationen aus dem Hellenthal/Busby Referenzdatensatz, die für f3-Analysen ausgewählt wurden	- 47 -
Tabelle 8 Auflistung der nukleären, funktionalen Marker; Angabe der Kategorie, entsprechend ihrem assoziierten Phänotyps, Bezeichnung der Loki, rs-Nummer und chromosomale Position	- 53 -
Tabelle 9 Gruppierung der Proben je Fundplatz; Auflistung der MiSeq-<i>screening</i> Ergebnisse und qPCR-Messung: endogener DNA-Gehalt in % und Molekülzahl (MZ) pro mg Knochenpulver der <i>fill-in</i> Produkte	- 58 -
Tabelle 10 Auswertung der Sequenzdaten der Probe Ak1-1 für die getesteten Anreicherungsverfahren verschiedener Hersteller; Aln. CAP=Anteil der Sequenzen, die gegen die <i>capture</i>-Regionen alignieren an alignierten Sequenzen gegen hg19; cov=Abdeckung in den <i>capture</i>-Regionen	- 61 -
Tabelle 11 Ergebnisse der Sequenzierung von 41 <i>capture</i>-Produkten; CAP= <i>capture</i>, Aln=aligniert gegen das Referenzgenom GRCh37/hg19; Mkdup= ohne Duplikate; cov=Abdeckung; MW/SD=Mittelwert / Standardabweichung; Frag.länge= Fragmentlänge	- 64 -
Tabelle 12 Ergebnisse der Kontaminationsbestimmung anhand des mitochondrialen Genoms und des X-Chromosoms und Angabe der Deaminierungsrate; FR=geschätzte Fehlerrate; Auth.= Authentizität in %; Kon.=Kontamination; mD5=Deaminierungsrate am 5'-Ende eines Moleküls	- 67 -
Tabelle 13 Phänotyp-Kombinationen der Augen – und Haarfarbe mit dem HIrisPlex-System	- 69 -
Tabelle 14 Genotypen der nukleären Marker die mit der Pigmentierung der Haut assoziierte werden können	- 70 -
Tabelle 15 Allelfrequenzen der abgeleiteten Allele ausgewählter Marker; anz. = ancestrale Allel; abgl. = abgeleitetes Allel; N = Anzahl der Individuen; Allelfreq. = Allelfrequenzen; EUR = europäische Populationen; EAS = ost-asiatische	

Populationen; SAS = Süd-asiatische Populationen; ZAS = Zentral-asiatische Populationen; NWR = Population der Republik Komi aus NW-Russland.....- 78 -

Tabelle 16 Paarweise Fst-Werte nach Hudson zwischen skythischen Gruppen verschiedener Kulturstufen und Datierungen; Anzahl der Individuen in Klammern (N).....- 92 -

Tabelle 17 Nukleotiddiversität; sub = ohne Individuen aus Barsučij Log und Ulžan- 93 -

Tabelle 18 Auflistung der *outgroup-f3*-Ergebnisse der skythischen Proben mit aDNA Referenzpopulationen ((Lazaridis et al. 2016) mit prähistorischen Populationen unter den höchsten 20 Werten; AG = Außengruppe; f_3 = berechneter f_3 -Wert; std.err = Standardfehler, SNPs = Anzahl der für die Berechnung zugrundeliegenden SNPs, Rang = Position des dargestellten f_3 -Werts... - 96 -

Tabelle 19 *outgroup-f3*-Statistik der Proben aus Zevakino im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti)$; Test, Skythen); dargestellt sind jeweils die höchsten 5 Werte; AG = Außen-gruppe; f_3 = berechneter f_3 -Wert; std.err = Standardfehler, Z = Z-score; SNPs = Anzahl der für die Berechnung zugrundeliegenden SNPs,.....- 99 -

Tabelle 20 Auflistung der *outgroup-f3*-Ergebnisse der skythischen Proben mit aDNA Referenzpopulationen ((Lazaridis et al. 2016); die fünf höchsten f_3 -Werten sind aufgelistet; AG = Außen-gruppe; f_3 = berechneter f_3 -Wert; std.err = Standardfehler, Z = Z-score; SNPs = Anzahl der für die Berechnung zugrundeliegenden SNPs,.....- 103 -

Tabelle 21 Signifikante D-Statistiken ($|Z| > 3$ und $|Z| < 3$) der Form $D(A, B, C, AG)$- 104 -

Tabelle 22 Signifikante D-Statistiken ($|Z| < 3$) der Form $D(A, Test, C, AG)$...- 105 -

Abbildungsverzeichnis

Abbildung 1: Geografische Verteilung und Zeitstrahl der eisen – und bronzezeitlichen Kulturguppen aus (Allentoft et al. 2015)	5 -
Abbildung 2: A) Skythische Kulturgruppen und ihre Verbreitung in Eurasien B) Detaillierte Ansicht der östlichen Fundplätze, Karte aus (van Geel et al. 2004). -	7 -
Abbildung 3 Grafischer Darstellung einer <i>bait -bead</i> basierten Anreicherung alter DNA in-solution. Nach (Carpenter et al. 2013)	30 -
Abbildung 4: Bait -Design Strategie für SureSelect und MYbait	33 -
Abbildung 5: Darstellung des endogenen DNA-Gehalts und der umgerechneten Molekülzahl per mg extrahiertem Knochenpulver der <i>screening-libraries</i>, nach Fundplatz gruppiert, in logarithmischer Skalierung	56 -
Abbildung 6: Abbildung 6: Ergebnisse der Molekularen Geschlechtsbestimmung der <i>screening-libraries</i> nach (Skoglund et al. 2013) y-Achse: sequenzierte Proben; x-Achse= Ry; Ry < 0,016 weiblich, Ry > 0,077 männlich	57 -
Abbildung 7: Geografische Verteilung der untersuchten Fundplätze über die eurasische Steppe	59 -
Abbildung 8: Abbildung 8: Vergleich der Sequenzzahlen der Probe Ak1-1 nach einzelnen Schritten der Rohdatenbearbeitung; OF=Qualitätsfilter; join=Sequenzpaare aln=aligniert gegen hg19 bzw. die entsprechenden <i>capture</i>-Regionen; Mkdup=ohne Duplikate	61 -
Abbildung 9: Allelfrequenzen für die abgeleiteten Allele der Pigmentierungsmarker; EUR = europäische Populationen; EAS = ost-asiatische Populationen; SAS = süd-asiatische Populationen; ZAS = zentral-asiatische Populationen; NWR = Population der Republik Komi aus NW-Russland.....	74 -
Abbildung 10: Allelfrequenzen für die abgeleiteten Allele der Marker, die mit der Verstoffwechslung von Xenobiotika assoziiert werden; EUR = europäische Populationen; EAS = ost-asiatische Populationen; SAS = süd-asiatische Populationen; ZAS = zentral-asiatische Populationen; NWR = Population der Republik Komi aus NW-Russland	75 -
Abbildung 11: Allelfrequenzen der abgeleiteten Allele des Laktase-Markers und der Marker, die mit einer Anpassung an klimatische Verhältnisse assoziiert werden können; EUR = europäische Populationen; EAS = ost-asiatische Populationen; SAS = süd-asiatische Populationen; ZAS = zentral-asiatische Populationen; NWR = Population der Republik Komi aus NW-Russland.....	76 -
Abbildung 12: Allelfrequenzen der abgeleiteten Allele, die mit einem erhöhten Risiko für Diabetes Typ 2 und Bluthochdruck korrelieren; EUR = europäische Populationen; EAS = ost-asiatische Populationen; SAS = Süd-asiatische Populationen; ZAS = Zentral-asiatische Populationen; NWR = Population der Republik Komi aus NW-Russland	77 -

Abbildung 13: PCA neolithischer, bronze - und eisenzeitlicher Individuen der vorliegenden Arbeit und ausgewählter aDNA Referenzen =aRef (Lazaridis et al. 2014; Fu et al. 2014; Jones et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015; Jones et al. 2017; Sikora et al. 2017). Mittels LASER in einen Referenzraum moderner Populationen projiziert (Hellenthal et al. 2014; Busby et al. 2015); Farbgebung entsprechend der Fundplätze - 80 -

Abbildung 14: Korrelation der errechneten F2-Werte nach Patterson (Patterson et al. 2012) und Hudson Fst - 82 -

Abbildung 15: heatmap der berechneten Fst-Werte nach Hudson zwischen 1000 genomes Populationen und skythischen Proben, gruppiert nach Zeit – und Kulturstufen; EA = ost-asiatische Populationen; SA = süd-asiatische Populationen - 83 -

Abbildung 16: outgroup-f3-Statistik eisenzeitlicher Proben SW-Russlands im Vergleich zu modernen Referenzpopulationen der Form f3(6=S_Mbuti; Modern, Skythen); dargestellt sind die höchsten 20 Werte inklusive Standardfehler - 85 -

Abbildung 17: outgroup-f3-Statistik der Proben aus dem Minusinsker Becken im Vergleich zu modernen Referenzpopulationen der Form f3(6=S_Mbuti; Modern, Skythen); dargestellt sind die höchsten 20 Werte inklusive Standardfehler - 86 -

Abbildung 18: outgroup-f3-Statistik der Proben aus Tuva im Vergleich zu modernen Referenzpopulationen der Form f3(6=S_Mbuti; Modern, Skythen); dargestellt sind die höchsten 20 Werte inklusive Standardfehler..... - 87 -

Abbildung 19: outgroup-f3-Statistik der Proben aus Justyd im Vergleich zu modernen Referenzpopulationen der Form f3(6=S_Mbuti; Modern, Skythen); dargestellt sind die höchsten 20 Werte inklusive Standardfehler..... - 87 -

Abbildung 20: outgroup-f3-Statistik der Proben aus Ak-Alacha im Vergleich zu modernen Referenzpopulationen der Form f3(6=S_Mbuti; Modern, Skythen); dargestellt sind die höchsten 20 Werte inklusive Standardfehler..... - 88 -

Abbildung 21: outgroup-f3-Statistik der kasachischen Proben im Vergleich zu modernen Referenzpopulationen der Form f3(6=S_Mbuti; Modern, Skythen); dargestellt sind die höchsten 20 Werte inklusive Standardfehler..... - 89 -

Abbildung 22: outgroup-f3-Statistik der ost-kasachischen Proben aus Zevakino im Vergleich zu modernen Referenzpopulationen der Form f3(6=S_Mbuti; Modern, Skythen); dargestellt sind die höchsten 20 Werte inklusive Standardfehler - 89 -

Abbildung 23: : outgroup-f3-Statistik bronzezeitlicher Proben SW-Russlands im Vergleich zu modernen Referenzpopulationen der Form f3(6=S_Mbuti; Modern, Skythen); dargestellt sind die höchsten 20 Werte inklusive Standardfehler - 90 -

Abbildung 24: outgroup-f3-Statistik neolithischer Proben Sibiriens im Vergleich zu modernen Referenzpopulationen der Form f3(6=S_Mbuti; Modern, Skythen); dargestellt sind die höchsten 20 Werte inklusive Standardfehler..... - 91 -

Abbildung 25: Parweise Fst-Werte skythischer Proben unterteilt in westliche und östliche Untergruppen entsprechend ihrer geografischen Lage oder Datierung in die Eisenzeit; West_Geo= Kolbino, Novozavedennoe, Pokrovka, Kamennyj; Ost_Geo= Zevakino, Berel, Ak-Alcha, Justyd, Aržan, Barsučij Log, Ul Ulžan žan, Solonzy; West_IA= Kolbino, Novozavedennoe, Pokrovka; Ost_IA= Zevakino, Berel, Ak-Alcha, Justyd, Aržan, Barsučij Log, Ulžan; sub = ohne Proben der Fundplätze Barsučij Log und Ulžan.....- 92 -

Abbildung 26: Grafische Darstellung der Mittelwerte der *site-pi* je Gruppierung in entsprechende Kulturstufen.....- 93 -

Abbildung 27: PCA der untersuchten Proben und zusätzlichen aDNA Referenzen (Lazaridis et al. 2014; Fu et al. 2014; Jones et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015; Jones et al. 2017; Sikora et al. 2017). SK = Individuen der vorliegenden Arbeit, inklusive bronzezeitlicher und neolithischer Individuen und Sarmaten; Aufteilung der bronzezeitlichen Steppenpopulationen der Refrenzen in ihre Untergruppen.....- 94 -

Abbildung 28: *outgroup-f3*-Statistik eisenzeitlicher, östlicher Proben im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler.....- 97 -

Abbildung 29: *outgroup-f3*-Statistik der Proben aus Justyd im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler- 98 -

Abbildung 30: *outgroup-f3*-Statistik der Proben aus Aržan im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler- 98 -

Abbildung 31: *outgroup-f3*-Statistik der Proben aus Aržan im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler- 99 -

Abbildung 32: *outgroup-f3*-Statistik eisenzeitlicher Proben SW-Russlands im Vergleich zu prähistorischen Referenzpopulationen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler- 100 -

Abbildung 33: *outgroup-f3*-Statistik bronzezeitlicher Proben SW-Russlands im Vergleich zu prähistorischen Referenzpopulationen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler- 101 -

Abbildung 34: *outgroup-f3*-Statistik neolithischer Proben SW-Russlands im Vergleich zu prähistorischen Kulturgruppen der Form $f_3(6=S_Mbuti; Test, Skythen)$; dargestellt sind die höchsten 20 Werte inklusive Standardfehler - 102

-

Abbildung 35 Abdeckung der Zielregionen im Verhältnis zur quantifizierten Molekülzahl der *screening-libraries* einer Probe- 109 -

Abbildung 36 Schematische Darstellung der fünf großen Migrationsbewegungen der Pontischen Steppe (PS), zentralen Steppe (CS) sowie dem östlichen

Steppenraum. a Expansion der Jamnaja- und Afanasievo-Kultur der frühen Bronzezeit b Spätbronzezeitliche Migration der Sintashta- und Andronovo-Kultur c Migrationsbewegungen der eisenzeitlichen Reiternomaden und Vermischungsereignisse in der späten Bronzezeit d und e Ausdehnung der Hunnen aus dem östlichen Steppenraum und asiatischer Einfluss im anschließenden Mittelalter; nach (Damgaard et al. 2018) - 128 -

Abbildung 37 PCA der untersuchten Proben und zusätzlichen aDNA-Referenzen (Lazaridis et al. 2014; Fu et al. 2014; Jones et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015; Jones et al. 2017; Sikora et al. 2017). SK = Individuen der vorliegenden Arbeit, inklusive bronzezeitlicher und neolithischer Individuen und Sarmaten; Aufteilung der bronzezeitlichen Steppenpopulationen der Referenzen in ihre Untergruppen..... - 129 -

Inhalt Daten-DVD

1. Ergänzende Informationen I1-I4
2. vcf-Dateien aller Proben

Die Rohdaten, Alignments und gemergten *bam-files* können auf Nachfrage vom Institut für Anthropologie zur Verfügung gestellt werden.

Arbeitsgruppe Paläogenetik Mainz <anthro2@uni-mainz.de>