

Aus dem Institut für Rechtsmedizin
Abteilung Forensische Genetik
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Hierarchische Y-SNP Analyse: Phylogeographische Daten für die
forensische Praxis und populationsgenetische Forschung

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Maria Geppert

aus Berlin

Datum der Promotion: 22.06.2014

Inhaltsverzeichnis

1	Zusammenfassung/Abstract.....	1
2	Einführung	2
2.1	Der Standard der forensischen DNA-Analyse.....	2
2.2	Uniparental vererbte DNA-Marker – Y-Chromosom und mtDNA.....	3
2.3	Single Nucleotide Polymorphisms (SNPs).....	4
2.4	Y-chromosomale SNPs.....	5
3	Zielstellung	10
4	Methodik.....	10
4.1	Multiplex-Analysen für phylogenetisch informative Y-SNPs	10
4.2	Minisequenzierung mittels SNaPshot®-Analyse	11
5	Ergebnisse	12
5.1	Anwendungsbezogene Adaption der SNP-Analyse	12
5.2	Adaption der hierarchischen Y-SNP Analyse für degradierte DNA und die phylogeographische Analyse	12
5.3	Phylogeographische Analysen am Beispiel einer endogamen und einer gemischten Population aus Lateinamerika	15
6	Diskussion.....	16
6.1	Einschätzung zu den Möglichkeiten und Grenzen der biogeographischen Herkunftsvorhersage.....	16
6.2	Diskussion der phylogeographischen Ergebnisse der Populationsstudien in Lateinamerika	19
6.3	Einschätzung der praktischen forensischen Anwendung von Y-SNPs	21
7	Fazit.....	23
8	Literaturverzeichnis	24
9	Eidesstattliche Versicherung/Anteilserklärung.....	27
10	Ausgewählte Publikationen	31
10.1	Publikation 1.....	31

10.2	Publikation 2.....	33
10.3	Publikation 3.....	35
10.4	Publikation 4.....	37
10.5	Publikation 5.....	39
11	Lebenslauf.....	41
12	Komplette Publikationsliste	43
13	Danksagung.....	47

1 Zusammenfassung/Abstract

Die Anforderungen an molekulargenetische Untersuchungen sind in den letzten Jahren gewachsen, selbst aus schwierigem Proben- oder Spurenmaterial mit wenig und fragmentierter DNA sollen möglichst aufschlussreiche, genetische Informationen gewonnen werden. Y-chromosomal lokalisierte *Single Nucleotide Polymorphisms* (SNPs) können einen wichtigen Beitrag zur Erweiterung des forensischen Analysespektrums leisten, das bisher nahezu ausschließlich auf einem DNA-Mustervergleich von Spur und Referenzprobe basiert. In der vorliegenden Arbeit wird ein hierarchischer Analyseansatz vorgestellt, mit dem die durch Y-SNPs definierten Haplogruppen entlang der Phylogenie des Y-Chromosoms typisiert werden können und damit jede männliche DNA genau einem zeitlich und räumlich definierten Ast des genetischen Stammbaums des modernen Menschen zugeordnet werden kann. Dabei bietet die Kombination aus PCR-Multiplexanalysen und Minisequenzierung eine schnelle und effektive Untersuchungsmethode. Anhand von zwei Fallstudien, in denen Skelettfunde mit forensischem und archäogenetischem Hintergrund untersucht wurden, wird dargestellt, dass Y-SNPs sich besonders für Untersuchungen von degradiertem DNA-Material eignen. Es wird gezeigt, dass sich mit Y-SNPs aufgrund ihrer Populationspezifität eine Vorhersage der geographischen Herkunft unbekannter DNA aus Spurenmaterial vornehmen lässt, und zwar besser als mit jedem anderen einzelnen genetischen Marker. Diese charakteristische Information aus dem Bereich der *DNA intelligence* kann bei Fällen ohne Vergleichsproben („cold cases“), sowie bei der Aufklärung von Vermisstenfällen zur sinnvollen Eingrenzung des Personenkreises verwendet werden. Um die geographische Verteilung der Haplogruppen zu ermitteln, ist die Durchführung von regionalen Populationsstudien wie sie hier für indigene und moderne Populationen durchgeführt wurden, unerlässlich. Dass diese Studien darüber hinaus auch zum Verständnis von Migration, Populationsgenese und prähistorischen sowie historischen Ereignissen in der Menschheitsgeschichte beitragen können, wird hier am Beispiel der Besiedlung Südamerikas demonstriert.

A challenge of forensic casework is to meet the increasing demands on molecular analyses. Even low copy or strongly fragmented DNA should provide helpful genetic information for the purpose of solving criminal cases. The analysis of Y-linked Single Nucleotide Polymorphisms (SNPs) is a new approach to comply with most difficult cases. The method extends the current forensic DNA analysis spectrum by providing DNA intelligence information. In this work, a hierarchical assay to analyze SNP defined haplogroups along the phylogeny of the Y chromosome is presented, assigning each male DNA to a temporal and

spatial specified branch of the genetic genealogy of modern humans. The combination of PCR-multiplexes and minisequencing provides an effective and sensitive approach. With the application of these techniques to two case studies, where skeletal remains were investigated with forensic and archaeogenetic background, it is shown that Y-SNPs are valuable genetic markers to analyze especially degraded DNA material. Furthermore, it could be demonstrated that due to the population specificity of Y-SNPs, the geographical origin of unknown male DNA can be identified much better than with any other single genetic marker. This “intelligence” information can support police investigations in “cold cases” in which no suspect is known and in missing person cases to narrow down the number of concerned persons. Population studies are essential for revealing the phylogeographical distribution of Y-chromosomal haplogroups. Here, autochthonous and admixed populations were examined for this purpose. The results of these studies also contribute to the understanding of the past demography of populations shaped by founder effects, migration, admixture and bottlenecks. This is illustrated by the example of the peopling of South America.

2 Einführung

2.1 Der Standard der forensischen DNA-Analyse

Die Aufgaben der forensischen DNA-Analyse sind vielfältig, wobei die Grundlage aller Anwendungen die Untersuchung von biologischem Material mit dem Ziel der Zuordnung zu einer Person ist (individuelle Identifizierung). Die Standardmethoden beruhen dabei auf dem Vergleich von DNA-Mustern, sogenannte STR-Profilen. So werden beispielsweise im Rahmen der Aufklärung von Straftaten DNA-Profilen, die von Spurenmaterial erstellt wurden, mit denen von potentiellen Tatverdächtigen verglichen. Diese Referenzproben werden direkt oder über DNA-Datenbanken, die in der Mehrzahl der europäischen Länder errichtet wurden, recherchiert und abgeglichen. Nach demselben Prinzip werden mit konventionellen Methoden bei nicht identifizierbaren Toten oder Körperteilen (Vermisstenfälle, Katastrophenfälle) DNA-Profilen mit denen von potentiellen Verwandten (direkte Identifizierung) oder aber Gegenständen der zu identifizierenden Person (indirekte Identifizierung) auf Übereinstimmung geprüft [1]. In der Abstammungsbegutachtung können mit Hilfe der DNA-Analyse Verwandtschaftsverhältnisse, wie z.B. Vaterschaften geklärt werden.

Grundlegend für die Erstellung eines DNA-Profiles, des sogenannten genetischen Fingerabdrucks, ist die Analyse von nichtkodierenden Sequenzabschnitten des humanen Genoms. Es werden repetitive Sequenzen (STRs – *short tandem repeats*) des autosomalen DNA-Bereichs untersucht, in denen sich ein bestimmtes Motiv (z.B. GATA) unterschiedlich

oft wiederholt. Die Kombination der Anzahl von Wiederholungen in den einzelnen STRs ergibt ein individualspezifisches Profil mit dem Personen eindeutig voneinander unterschieden werden können. Die eigentliche STR-Analyse erfolgt über die Längenmessung der entsprechenden Sequenzabschnitte. Diese Methode ist vollständig etabliert in der forensischen Routine und es stehen für ihre Anwendung kommerziell vertriebene Kits zur Verfügung, die eine parallele Untersuchung von mehreren STR-Systemen erlauben.

2.2 Uniparental vererbte DNA-Marker – Y-Chromosom und mtDNA

Für besondere Fragestellungen kann die Analyse von uniparental vererbten, genetischen Markern erforderlich sein. Dafür steht zum einen das Y-Chromosom zu Verfügung. Der überwiegende Teil des Y-Chromosoms (die MSY, *male specific region of the Y chromosome*) unterliegt anders als autosomale Abschnitte des Genoms einem klonalen Erbgang in väterlicher Linie, der ohne klassische Rekombination zwischen Chromosomenpaaren väterlicher und mütterlicher Herkunft und damit ohne eine regelmäßige Neuordnung von Merkmalskombinationen abläuft [2]. Aufgrund der unveränderten Weitergabe des Y-Chromosoms von Generation zu Generation werden paternale Linien konstituiert, die einzig durch Mutationen verändert werden. Genau wie im autosomalen Sequenzbereich gibt es STR-Polymorphismen auf dem Y-Chromosom, die für forensische und populationsgenetische Untersuchungen eingesetzt werden und deren Analyse seit langem fester Bestandteil der forensischen Praxis ist. Ihre hohen Mutationsraten von durchschnittlich 0,2% per STR per Generation (wie auch bei autosomalen STRs) [3] führen zu einer hohen Variabilität der Y-chromosomalen Haplotypen. In der Forensik wird die Y-chromosomale Haplotypisierung vor allem für Sexualstraftaten angewendet, in denen eine Differenzierung von unbalancierten Mann-Frau-DNA-Mischspuren notwendig ist [4]. In den meisten Fällen lässt nur die zusätzliche Analyse der mann-spezifischen Y-STRs einen DNA-Mustervergleich von Spur und potentiell männlichem Spurenverursacher zu. Darüber hinaus werden Y-STRs in der Abstammungsdiagnostik zur Analyse komplexer Stammbäume in Defizienzfällen eingesetzt.

Der zweite uniparental vererbte Marker des humanen Genoms ist das extranukleäre, mitochondriale Genom (mtDNA). Dieses liegt in bis zu tausendfacher Kopie in einer Zelle vor und ist besonders robust gegenüber Degradation. Für Fälle in denen STR-Polymorphismen aufgrund der starken Fragmentierung der Kern-DNA nicht mehr amplifiziert werden können, bietet die Untersuchung der mtDNA eine realistische Chance wertvolle genetische Information zu einer Probe zu erhalten. Es muss, analog zum Y-Chromosom, allerdings berücksichtigt werden, dass mtDNA nur maternal vererbt wird und sich daraus

mütterliche Abstammungslinien ableiten. Diese lineare Vererbung kann jedoch gerade komplexe Abstammungsverhältnisse über mehrere Generationen aufklären [5]. Da das mitochondriale Genom (wie auch die MSY) frei von Rekombination ist, erfolgt die Differenzierung der mitochondrialen Linien ausschließlich durch auftretende Mutationen. Diese werden untersucht, in dem die hypervariablen Abschnitte der Kontrollregion, die Kontrollregion im Ganzen oder gleich die gesamte mtDNA sequenziert werden. Anschließend kann das mtDNA-Profil für den Mustervergleich erstellt werden, indem die beobachteten Mutationen mit einer festgelegten Referenzsequenz (*revised Cambridge Reference Sequence - rCRS*) abgeglichen und nach Nomenklaturregeln annotiert werden [5].

2.3 Single Nucleotide Polymorphisms (SNPs)

Neben den bereits beschriebenen STR-Polymorphismen stellen SNPs, eine weitere Form genetischer Variation dar, bei der das Mutationsereignis zur Variation eines einzelnen Nukleotids im Genom führt. Im Vergleich zu STR-Systemen, bei denen längere Sequenzbereiche (100-400 Basenpaaren) untersucht werden, können für die Untersuchung von SNPs prinzipiell kurze Sequenzbereiche amplifiziert werden (unter 100 Basenpaaren). Dadurch eignen sie sich im Besonderen für die Analyse von degradiertem Untersuchungsmaterial wie es z.B. bei Skelettfunden vorliegt. Die in der Populationsgenetik und Forensik eingesetzten SNPs sind meistens bi-allelisch, weisen demnach zwei Merkmalszustände, ein ursprüngliches (anzestrales) und ein mutiertes Allel, auf. SNPs sind zahlreich und weit verbreitet im Genom (in der dbSNP-Datenbank sind im Build 137 53,567,890 SNPs gelistet, *release* 26.06.2012). Die in der Forensik angewandten SNPs werden in vier Kategorien unterteilt [6]. Unter der Bezeichnung *Individual Identification SNPs* (IISNPs) werden SNPs zusammengefasst, die dazu dienen Individuen voneinander zu unterscheiden. Um eine ähnlich hohe Diskriminierung wie mit multiallelischen STR-Markern zu erreichen, werden erstaunlich wenige bi-allelische SNPs benötigt. Im Rahmen des SNPforID Projektes wurde bereits beispielhaft ein Analyseset von 52 autosomalen SNP-Markern zur humanen Identifizierung veröffentlicht [7]. *Lineage Informative SNPs* (LISNPs) umfassen gekoppelte SNPs, die einen sogenannten Haplotypen bilden. LISNPs werden für evolutionäre Untersuchungen und für die Abstammungsanalyse eingesetzt. Marker dieser Klasse sind beispielsweise auf dem Y-Chromosom und im mitochondrialen Genom zu finden. In einer Studie von Haas [8] zur Abstammung des Schweizer Volkshelden Jörg Jenatsch (1596-1639) wird beispielhaft gezeigt, dass die kombinierte SNP/STR Analyse des Y-Chromosoms die Rekonstruktion von Stammbäumen anhand des Vergleichs von DNA-Mustern der Vor- und Nachfahren ermöglicht. Als *Phenotype Informative SNPs* (PISNPs)

werden Polymorphismen bezeichnet, die Vorhersagen über den Phänotyp eines Individuums zulassen, beispielsweise über Hautfarbe, Haarfarbe oder Augenfarbe. Die Arbeitsgruppe von Kayser [9] veröffentlichte das erste forensisch relevante Set von autosomalen SNP-Markern, mit deren Analyse eine statistisch abgesicherte Aussage über die Wahrscheinlichkeit einer bestimmten Augenfarbe (für blaue und braune Augen kann eine Vorhersagegenauigkeit von über 90% erreicht werden) getroffen werden kann. Die vierte Kategorie umfasst SNPs, die Hinweise auf die geographische Herkunft eines Individuums zulassen. Diese werden als *Ancestry Informative SNPs* (AISNPs) bezeichnet. In diese Gruppe fallen zum einen ausgewählte autosomale SNPs, die z.B. selektionsbedingt populationspezifische Mutationsmuster aufweisen [10]. Auch selektionsneutrale Y-SNPs werden als AISNPs bezeichnet, da sie durch Drift, kulturelle und geographische Einflüsse erhebliche Frequenzunterschiede zwischen Populationen aufweisen.

Während die in den ersten beiden Kategorien zusammengefassten SNPs (IISNP und LISNPs) in ihrer Anwendung der klassischen forensischen DNA-Analytik – dem Abgleich von DNA-Mustern entsprechen, handelt es sich bei den PISNPs und den AISNPs um Marker, deren Untersuchung Informationen über den DNA-Träger selbst (z.B. einen möglichen Spurenverursacher) geben. Diese vom ursprünglichen Konzept des DNA Profiling abweichende Strategie wird einem erweiterten Methodenspektrum zugerechnet, das im angelsächsischen als *DNA intelligence* bezeichnet wird [11]. *DNA intelligence* bietet weiterführende Untersuchungsansätze für Fall-Konstellationen, in denen keine Vergleichsprobe zur Verfügung steht (z.B. sogenannte *cold cases* oder Identifizierung unbekannter Personen bzw. Leichenteile in der Kriminalistik). Unter diesem Aspekt soll in dem nächsten Abschnitt näher auf Y-chromosomale SNPs und ihre Funktion als AISNPs eingegangen werden.

2.4 Y-chromosomale SNPs

Die in der Forensik untersuchten Y-chromosomalen SNPs liegen wie auch die übrigen forensischen DNA-Systeme im nicht-codierenden Bereich. Deshalb handelt es sich um neutrale Mutationen, die keine Auswirkung auf die Fitness eines Individuums haben, daher nicht selektiert werden und nachweisbar im Genpool der Nachfahren erhalten bleiben, sofern sie nicht durch z.B. Drift verloren gehen. Zusammen mit der um Größenordnungen geringeren Mutationsrate von SNPs (10^{-8} im Vergleich zu STRs mit 10^{-3} [12]), die mit einer geringeren Wahrscheinlichkeit von Rück- oder Parallelmutationen einhergeht und der weitgehend rekombinationsfreien Vererbung des Y-Chromosoms, können Verwandtschaftsbeziehungen

der männlichen Abstammungslinien, bis hin zum MRCA (*most recent common ancestor*) rekonstruiert werden. Männer die identische SNP-Muster haben besitzen eindeutig eine gemeinsame Herkunft durch Abstammung (*identity by descent*- IBD), während identische Y-STR Profile sowohl IBD als auch IBS (*identity by state*) sein können [13]. Die evolutionäre Beziehung der Y-SNPs zueinander kann in einem robusten phylogenetischen Stammbaum abgebildet werden [14-16]. Die Verzweigungen des Baumes werden dabei durch mindestens einen phylogenetisch informativen SNP definiert und repräsentieren genealogische Abstammungslinien des Y-Chromosoms, die als Haplogruppen bezeichnet werden. Im Jahr 2002 veröffentlichte das YCC (*Y Chromosome Consortium*) einen umfassenden Stammbaum (der 2008 überarbeitet publiziert wurde [15]), der nach dem Prinzip maximaler Parsimonie konstruiert und anhand von Außengruppenvergleichen (mit z.B. Schimpanse, Gorilla oder Orang-Utan) gewurzelt wurde. Gleichzeitig wurde eine bindende Nomenklatur für die Haplogruppen der Y-Phylogenie vorgeschlagen [14]. Dieser Nomenklatur folgend, tragen die Hauptäste des Baumes Großbuchstaben (z.B. Q) und alle Untergruppen werden alternierend mit Zahlen und Kleinbuchstaben benannt (z.B. Q1a3a, für eine Haplogruppe vier Knoten weiter). Der phylogenetische Baum des Y-Chromosoms besteht gegenwärtig aus 20 Hauptästen, die das Gerüst des Baumes bilden und mit den Buchstaben A bis T benannt sind [15, 16]. Jeder dieser Hauptäste endet in mehrfach dichotomen Verzweigungen, die durch weitere SNPs in einer hierarchischen Abfolge definiert werden. In einer erst kürzlich erschienenen Studie von Van Geystelen et al. [16] wurde eine aktualisierte Version des phylogenetischen Baumes mit 359 Haplogruppen und 721 Y-SNP Marker vorgestellt. Die Y-SNPs wurden dabei anhand von *whole genome sequencing* Daten auf ihre Richtigkeit geprüft und neu detektierte Y-SNPs wurden dem Baum hinzugefügt. Um der zunehmenden Komplexität gerecht zu werden, geht die heute gebräuchlichere Bezeichnung der Haplogruppen aus der Haplogruppe und den sie definierenden SNP selbst hervor (z.B. Q-M3 statt Q1a3a). Haplogruppen, die nicht durch mindestens einen SNP definiert sind, bezeichnet man als Paragruppe. Darunter zusammengefasste Y-Chromosomen gehören einer Haplogruppe an, aber nicht einer ihrer Untergruppen. Zur Unterscheidung der Paragruppen sind diese im Baum mit einem Stern gekennzeichnet (z.B. C3*).

Der demografische Prozess von Migrationen, Kolonisationen und Sesshaftwerden, der die Besiedlung aller Erdteile durch den anatomisch modernen Menschen im Laufe der letzten 40,000 Jahre ermöglichte, begründet, dass die Y-SNPs definierten Haplogruppen nicht zufällig zwischen den Populationen verteilt sind, sondern eine spezifische Struktur innerhalb und zwischen den Kontinenten, Sprachgruppen und Ethnien aufweisen [17]. Nach

der *Out of Africa* Theorie [18], hat sich der anatomisch moderne Mensch in Afrika entwickelt und verbreitete sich von dort aus über die ganze Welt. Im Zuge der Migration nach Asien, Ozeanien, Europa und Amerika traten neue Y-chromosomale Mutationen (SNPs) auf, die bei kontinuierlicher Vererbung neue genealogische Linien (Haplogruppen) hervorbrachten. So bildeten sich im zeitlichen Abstand zu der ersten afrikanischen Auswanderung und damit von der Wurzel des Stammbaumes viele kontinent- oder regionenspezifische Y-Haplogruppen aus. Der phylogenetische Stammbaum des Y-Chromosoms reflektiert damit die Evolution menschlicher Diversität [17]. Die ersten beiden Äste A und B sind auf Populationen aus Afrika beschränkt. Ihre phylogenetische Position, ihr Verbreitungsmuster und ihre Diversität, aufgrund zahlreicher genetischer Variationen, deuten auf eine frühe Diversifikation und eine frühe, weite Verteilung innerhalb Afrikas hin [17]. Das weitere, zeitlich jüngere Gerüst des Stammbaums spaltet sich in die drei großen Cluster C, DE und F (siehe Abbildung 1), mit der letzten gemeinsamen Mutation am Y-SNP-Marker M168 (und alternative Y-SNPs). In diesem Teil des Baumes befinden sich weitere afrikanische und alle nicht-afrikanischen Phylae [17, 18]. Die Haplogruppe C findet man in Ostasien, Ozeanien und Amerika, Haplogruppe D ist weitestgehend auf Ostasien begrenzt, Haplogruppe E ist in Afrika, Westasien und Europa präsent, während Haplogruppe F mit ihren zahlreichen Untergruppen (G, H, I, J, K, L, M, N, O, P, Q, R, S, T; siehe Abbildung 1) stark differenziert ist und eine weltweite Verbreitung außerhalb Afrikas zeigt [18]. Die Haplogruppe G ist weit verbreitet in Europa und tritt vielerorts mit relativ geringen Frequenzen von 2% bis 10% auf (siehe www.yhrd.org/search/SNPs). Die Haplogruppe H evolvierte vermutlich im heutigen nördlichen Indien und verteilt sich bis heute über den gesamten indischen Subkontinent [15]. Sie umfasst aber auch die von dort abgewanderten Roma-Populationen. Haplogruppe I ist eine der beiden Haupthaplogruppen Europas. Die Untergruppen I1 und I2 sind weit verbreitet in Europa und andernorts nahezu abwesend, wobei I1 die höchsten Frequenzen in Nord-Europa aufweist (bis zu 50%) und I2 in Osteuropa und den Balkanstaaten häufig ist [15]. Die Schwesterhaplogruppe von I, Haplogruppe J spaltet sich ebenfalls in zwei geographisch recht deutlich voneinander abgegrenzte Untergruppen auf. Haplogruppe J1 ist am häufigsten in Europa (in Südeuropa bis 30%) und J2 ist auf der arabischen Halbinsel, im Mittleren Osten, in Nordafrika und Äthiopien vertreten [15]. Die Haplogruppe K umfasst die Untergruppen M, NO, P und S. Sie ist selbst selten und wurde bisher auf dem indischen Subkontinent, in Ozeanien, Indonesien und Australien beobachtet. Die Haplogruppe M ist ebenfalls selten und ihre geographische Verteilung beschränkt sich auf Ozeanien und Ostindonesien. Dagegen ist die Haplogruppe N weit verbreitet in Zentralasien, Sibirien (bis zu 90%) und Nord-Ost-

Europa, wobei sie andernorts abwesend oder nur marginal vertreten ist [15]. Die Haplogruppe O ist eine asiatische Haplogruppe, die mit hohen Frequenzen von über 80% in Ostasien beobachtet wird. Die Haplogruppe P umfasst die Subhaplogruppen Q und R. Während die Haplogruppe Q in Sibirien, Zentralasien sowie als Haupthaplogruppe der indigenen Völker Amerikas zu finden ist, gehört die Mehrheit der europäischen Y-Chromosomen der Haplogruppe R an. Die Haplogruppe S tritt mit nur geringen Frequenzen in Ozeanien und Indonesien auf [15]. Die Haplogruppen L und T gehen auf einen gemeinsamen Knoten zurück [16]. Die Haplogruppe L ist dabei eng mit dem indischen Subkontinent verknüpft und zeigt dort die höchsten Frequenzen. Sie ist aber auch vereinzelt in Europa, dem Mittleren Osten, Zentralasien und Nordafrika vertreten. Die Haplogruppe T ist dagegen nur mit geringen Frequenzen im Mittleren Osten, Afrika und Europa zu finden [15]. Generell ist der historische und demografische Kontext zu beachten. So finden sich viele westeuropäische Linien in Nord- und viele südeuropäische Linien in Lateinamerika, chinesische und indische Linien in Melanesien, europäische Linien in Australien und Südafrika. Mit den Recherche-Tools der YHRD (www.yhrd.org) ergibt sich ein guter Überblick über die Migration der Neuzeit und ihre Folgen für die Verteilungsmuster Y-chromosomaler Haplogruppen.

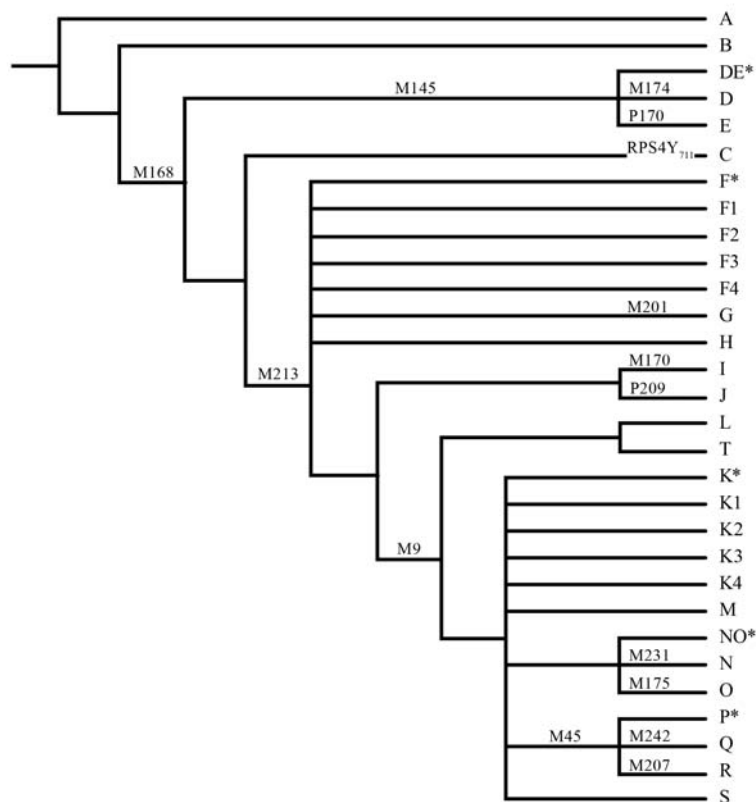


Abbildung 1: Phylogenetischer Stammbaum des Y-Chromosoms. Modifiziert nach [15, 16]. Es sind alle Haupthaplogruppen mit Großbuchstaben bezeichnet. Abstammungslinien, die mit Y-SNPs bezeichnet sind, können mit den hier vorgestellten Multiplex-Reaktionen typisiert werden. Die Länge der Zweige repräsentiert nicht das Alter der Abstammungslinien.

Aufgrund dieser spezifischen geographischen Verteilung der Haplogruppen innerhalb der Y-Phylogenie, lässt die Untersuchung von Y-SNPs Rückschlüsse auf die Herkunft einer Y-chromosomalen Abstammungslinie und der Personen, die ihr angehören zu [19]. Diese Art der Untersuchung, in der die geographische Verteilung von Abstammungslinien eines phylogenetischen Stammbaums beurteilt wird, bezeichnet man als Phylogeographie [18]. Wie schon durch Brion und Wetton [20, 21] diskutiert, handelt es sich um ein Verfahren der *DNA intelligence*, das in der Forensik für Kriminalfälle, Skelettfunde oder Vermisstenfälle eingesetzt werden kann, in denen keine Referenzprobe(n) von Vergleichspersonen zur Verfügung stehen. Für eine phylogeographische Untersuchung werden die Frequenzen der Haplogruppen und ihre Verbreitungsmuster untersucht. Haplogruppen weisen in der Regel eine hohe Frequenz am Ort des Mutationsereignisses (SNP) auf und mit zunehmender Entfernung von diesem Ort nimmt ihre Häufigkeit graduell ab (*clines*) [18]. Vom Zentrum der Haplogruppe mit hoher Frequenz kann dabei zum einen eine radiale Ausstrahlung (z.B. R-U106) oder zum anderen eine gerichtete Verbreitung ausgehen (wie z.B. bei der Roma-Population, Haplogruppe H) (www.yhrd.org). So ergeben sich klare geographische Strukturen, die eine Vorhersage des wahrscheinlichsten geographischen Ursprungs ermöglichen. Das Y-Chromosom weist im Vergleich zum autosomalen und mtDNA Genom die stärksten genetischen Unterschiede mit zunehmender, geographischer Distanz auf [22].

Eine wichtige Grundlage zur Auswertung und Anwendung phylogeographischer Y-chromosomaler Daten, neben der aufwendigen Literaturrecherche, stellt die forensische Referenzdatenbank YHRD – *Y Chromosome Haplotype Reference Database* (www.yhrd.org) dar, die an der Charité seit dem Jahr 2000 kuratiert wird [23]. Sie ist ein Repositorium für Daten aus Y-chromosomalen Populationsstudien von geprüfter forensischer Qualität und umfasst derzeit (*release 44*) 114.256 Haplotypen (verschiedene Y-Chromosomen, die mit Y-STRs analysiert wurden), davon sind 11.996 auch mit Y-SNP typisiert. Mit der stetig wachsenden Zahl an Populationsstudien wird das Bild der geographischen Verteilung der Y-Haplogruppen immer klarer. Als webbasierte, freizugängliche Datenbank bietet die YHRD die Möglichkeit einen Haplotypen (Y-STR Profil), der beispielsweise in einer forensischen Probe detektiert wurde, mit oder ohne Haplogruppe in den hinterlegten Populationsdaten zu suchen, um dessen Häufigkeit und geographische Verbreitung zu ermitteln. Man erhält daraufhin eine Aufstellung darüber, wie oft dieser Haplotyp (oder die Haplogruppe) in der Datenbank vorkommt und mit welchen Frequenzen er in Populationen oder Metapopulationen beobachtet wurde. Eine Kartierungsfunktion unterstützt die Visualisierung der Ergebnisse. Anhand der Treffer-Statistik kann dann eine statistisch gestützte Vorhersage über die

wahrscheinlichste Herkunftspopulation vorgenommen werden. Dabei lassen gut untersuchte Haplogruppen, wie beispielsweise E oder R in der Regel genauere Einschätzungen zu als weniger gut dokumentierte Haplogruppen, wie S und T.

3 Zielstellung

Im Fokus der eingereichten Arbeiten stehen die Analyse von Y-chromosomalen SNPs mittels Minisequenzierung und die Erstellung einer routinerelevanten Analysetechnik. Ziel ist es herauszuarbeiten, inwiefern diese Untersuchung von Y-SNPs, im Kontext eines spezifischen forensischen Sachverhalts, Schlussfolgerungen zur geographischen Herkunft der DNA von bisher unbekanntem männlichen Personen erlaubt. Zum einen soll aus methodischer Sicht getestet werden, welchen Beitrag die Analyse von Y-SNPs für die Bearbeitung von degradierten Proben von z.B. Skelettfunden leisten kann. Zum anderen soll gezeigt werden, welche statistisch validierte Aussage zur geographischen Herkunft einer extrahierten DNA möglich ist. Mit der Etablierung eines hierarchischen Y-SNP Ansatzes kommt eine sensitive und hochauflösende Technik zur Untersuchung von Y-SNPs zum Einsatz. Es werden anwendungsbezogene Strategien erarbeitet und diese im praktischen Gebrauch anhand von Fallbeispielen getestet. Es soll darauf eingegangen werden, welche Voraussetzungen für die forensische Anwendung der Y-SNPs erfüllt sein müssen. Darüber hinaus soll dargestellt werden, welchen Stellenwert Populationsdaten für die Interpretation der Y-chromosomalen Daten besitzen und welche neuen Erkenntnisse für populationsgenetische Fragestellungen gewonnen werden können.

4 Methodik

4.1 Multiplex-Analysen für phylogenetisch informative Y-SNPs

Um den Mutationsstatus einzelner SNPs untersuchen zu können, wird zunächst die den SNP einschließende Sequenz in einer *Polymerase Chain Reaction* (PCR) vervielfältigt. Da in der Regel eine Untersuchung mehrerer SNPs zur exakten Bestimmung der Y-chromosomalen Haplogruppe notwendig und auch vorteilhaft ist, können verschiedene Sequenzen eines Genoms in einer gleichzeitigen, sogenannten Multiplex-PCR amplifiziert werden. Auf diese Weise wird weniger DNA verbraucht als in mehreren Einzelreaktionen und der technische sowie der zeitliche Aufwand sind geringer.

Im Rahmen der Dissertation wurden Daten zu unterschiedlichen humanen Populationen erhoben. Darum war es notwendig aufgrund der bereits beschriebenen Populationsspezifität der Y-chromosomalen SNPs zwei verschiedene Multiplex-Ansätze zu entwerfen, um einen

Einsatz für die unterschiedlichen Anwendungsgebiete zu ermöglichen. Beiden gemeinsam ist das von Brion et al. [24] postulierte Prinzip der hierarchischen Analyse, bei der systematisch entlang des phylogenetischen Baums zunächst die generische Haplogruppe und anschließend gezielt die spezifische Untergruppe mit einer entsprechenden, hoch auflösenden Multiplex-PCR typisiert wird. Mit diesem Ansatz wird die Anzahl der zu typisierenden Marker gering gehalten. Desweiteren gewährleistet die Typisierung von mehreren, hierarchisch im Baum angeordneten, phylogenetischen SNPs die sichere Zuordnung einer Probe zu einer Haplogruppe. Widersprüchliche Analyseergebnisse, die auf mögliche Verunreinigungen oder sehr seltene Rückmutationen zurückzuführen sein können, werden so erkannt.

Nach diesem Konzept wurde ein Set von sechs Multiplex-Reaktionen entworfen, mit dem vorrangig Marker untersucht werden können, die für Eurasien typisch sind. Multiplex I und Multiplex II dienen als generische Multiplex-Reaktionen, um insgesamt 12 Haupthaplogruppen zu bestimmen (siehe Publikation 1). Die vier spezifischen Multiplex-Reaktionen können herangezogen werden, um die in Europa weit verbreiteten Haplogruppen E (5 Untergruppen), I (4 Untergruppen), J (4 Untergruppen) und R (6 Untergruppen; [8]) weiter aufzulösen (siehe Publikation 5). Desweiteren wurden drei Multiplex-Reaktionen entwickelt um indigene Populationen Südamerikas zu studieren, die nur die Haplogruppen Q und C zeigen. Auch hier wird zunächst eine generische Multiplex-PCR (SA Major) angewendet mit der insgesamt sieben Haplogruppen bestimmt werden können (siehe Publikation 4). Diese Multiplex-Reaktion dient auch dazu nicht-indigene Haplogruppen zu detektieren. Für eine spezifischere Analyse der indigenen Haplogruppen stehen in diesem Fall zwei weitere Ansätze (SA SpecQ und SA SpecC) zur Verfügung, in denen vier Untergruppen von Haplogruppe Q und sechs Untergruppen von Haplogruppe C bestimmt werden können (siehe Publikation 4).

4.2 Minisequenzierung mittels SNaPshot®-Analyse

In den vorliegenden Publikationen wurde zur Untersuchung der Y-SNPs das SNaPshot®-Multiplex System der Firma Applied Biosystems eingesetzt, das bereits in anderen Y-chromosomalen Studien zur Anwendung kam [24-26]. Dieses ermöglicht die zeitgleiche Typisierung verschiedener SNPs. Im Anschluss an die Multiplex-PCR erfolgt die sogenannte Minisequenzierung oder auch *single base extension* (SBE) Reaktion, in der ein Primer unmittelbar an die SNP-Position bindet und nur um das Nukleotid des SNPs verlängert wird. Wie in der Sanger-Sequenzierung beruht die Methode auf einer Kettenabbruch-Synthese und zielt darauf ab, mithilfe von fluoreszenzmarkierten ddNTPs das exakte Nukleotid an einer

bestimmten Sequenzposition abzulesen und nicht wie für die STR-Analyse üblich, allein die Länge eines Fragments zu ermitteln. Die in der SBE-Reaktion eingesetzten, artifiziellen ddNTPs führen zum Kettenabbruch, da am 3'Ende die Hydroxylgruppe fehlt, an die das nachfolgende Nukleotid gebunden wird. Sowohl nach der PCR, als auch nach der SBE-Reaktion erfolgt ein enzymatischer Aufreinigungs-Schritt, um überschüssige dNTPs, ddNTPs und Primer zu entfernen. Jedes der vier ddNTPs ist mit einem anderen Fluoreszenzfarbstoff markiert. Bei der gleichzeitigen Analyse mehrerer SNPs kann es zum Auftreten von identischen Mutationsmustern und somit zur selben farblichen Markierung kommen. Um dennoch eine Unterscheidung und parallele Typisierung der SNP-Positionen vornehmen zu können, wird die Länge der SBE-Primer mit unterschiedlich langen, nicht bindenden Sequenz-Anhängen variiert. In der Kapillarelektrophorese werden die fluoreszenz-markierten Primerfragmente detektiert und anhand ihrer spezifischen Längen und Farben unterschieden.

5 Ergebnisse

5.1 Anwendungsbezogene Adaption der SNP-Analyse

In Publikation 1 wurde ein Protokoll für die Analyse von phylogenetischen SNPs in einem hierarchischen Multiplex-Assay unter Verwendung der SNaPshot® Multiplex Systems (Applied Biosystems) erarbeitet. Schritt-für-Schritt erfolgt die Anleitung für die Adaption der SNaPshot®-Reagenzien basierten Methode für eigene, anwendungsbezogene SNP-Analysen. Dabei werden präzise Anleitungen zur Auswahl der Y-SNPs, dem Primerdesign, der Evaluation der Primer, der Datenanalyse und der Dateninterpretation mit Hinweisen auf mögliche Schwierigkeiten gegeben. Darüber hinaus wird beispielhaft eine phylogeographische Analyse zur Vorhersage der geographischen Herkunft einer unbekannt Person anhand der Ergebnisse der beschriebenen Methode erläutert.

5.2 Adaption der hierarchischen Y-SNP Analyse für degradierte DNA und die phylogeographische Analyse

In der ersten Fallstudie zur biogeographischen Herkunftsprädiktion (Publikation 2) wurden sieben mittelalterliche Skelette untersucht, die 1985 im Norden Spaniens, in der Provinz Aragonien entdeckt wurden. Historische oder archäologische Aufzeichnungen zu den Gräbern waren nicht bekannt. Unter Einbeziehung der Radiocarbon-Datierung auf 776-991 u.Z. wurde die Hypothese aufgestellt, dass es sich bei den Skeletten um historische Persönlichkeiten, eventuell sogar aus der königlichen Dynastie Aragons handelte. Anthropologische und genetische Untersuchungen sollten zeigen, ob Rückschlüsse auf Liegezeit, Geschlecht, eventuelle familiäre Beziehungen und mögliche Herkunft gezogen werden können. Die Y-

SNP Analyse stellte dabei einen Teil der Gesamtuntersuchungen dar. Die anthropologischen und genetischen Analysen ergaben, dass es sich bei allen Individuen um Männer handelte. Drei der untersuchten sieben Skelette wiesen nach der Extraktion DNA-Konzentrationen auf, die Analysen an nukleären DNA-Markern zuließen. Die autosomalen STR-Analysen ergaben für diese drei Skelette Teil- und auch Vollprofile. Direkte verwandtschaftliche Beziehungen wie Vaterschaft oder geschwisterliche Verwandtschaft konnten nicht nachgewiesen werden. Mitochondriale Analysen waren für sechs der sieben Skelette erfolgreich. Die mtDNA Resultate wiesen darauf hin, dass zwei der Skelette maternal verwandte Cousins waren, während die Übrigen unverwandt schienen. Die identifizierten mtDNA Haplogruppen sind mit jeweils etwa 30% (Haplogruppe H) und 8% (Haplogruppe U5a) in der heutigen spanischen Population vertreten (Publikation 2). Die Y-SNP Analysen wurden an zwei der bereits zuvor erfolgreich untersuchten Skelettproben (Oberschenkelknochen) durchgeführt. Obwohl es sich um in Europa aufgefundene Skelette handelte, wurde die generische Multiplex-Reaktion SA Major (Publikation 2) als Analyse-Ansatz gewählt, da dieser sich in vorangegangenen Populationsstudien als sehr sensitiv gegenüber geringen DNA-Konzentrationen zeigte. Aufgrund der geringen verbliebenen DNA-Menge (mit ohnehin geringer DNA-Konzentration) aus den vorangegangenen DNA-Untersuchungen konnte in der Multiplexanalyse (SA Major) nur ein Marker (nicht-informativ) in einem der beiden untersuchten Oberschenkelknochen typisiert werden. Aus diesem Grund wurden nachfolgend Einzelanalysen durchgeführt, in denen gezielt Haplogruppen untersucht wurden, die in der rezenten spanischen Population mit einer erhöhten Häufigkeit auftreten und zusätzlich in den bereits vorhandenen Standardprotokollen die kürzesten PCR-Fragmente aufwiesen. Mit diesem Ansatz gelang es die Knochenproben für Y-SNPs zu typisieren und beide Skelette der Haplogruppe R (M207) zuzuordnen. Bei einem Skelett war es darüber hinaus möglich die spezifische Untergruppe R-M269 (R1b1b2) zu bestimmen. Haplogruppe R zeigt im Allgemeinen eine hohe Frequenz in europäischen Populationen, die Untergruppe R-M269 hat die höchsten Frequenzen in Westeuropa, mit einem sich abschwächenden Gradienten in Richtung Mitteleuropa. Wie Abbildung 2 zeigt, weisen die nordspanischen YHRD-Stichproben die höchsten relativen Frequenzen für R-M269 auf.

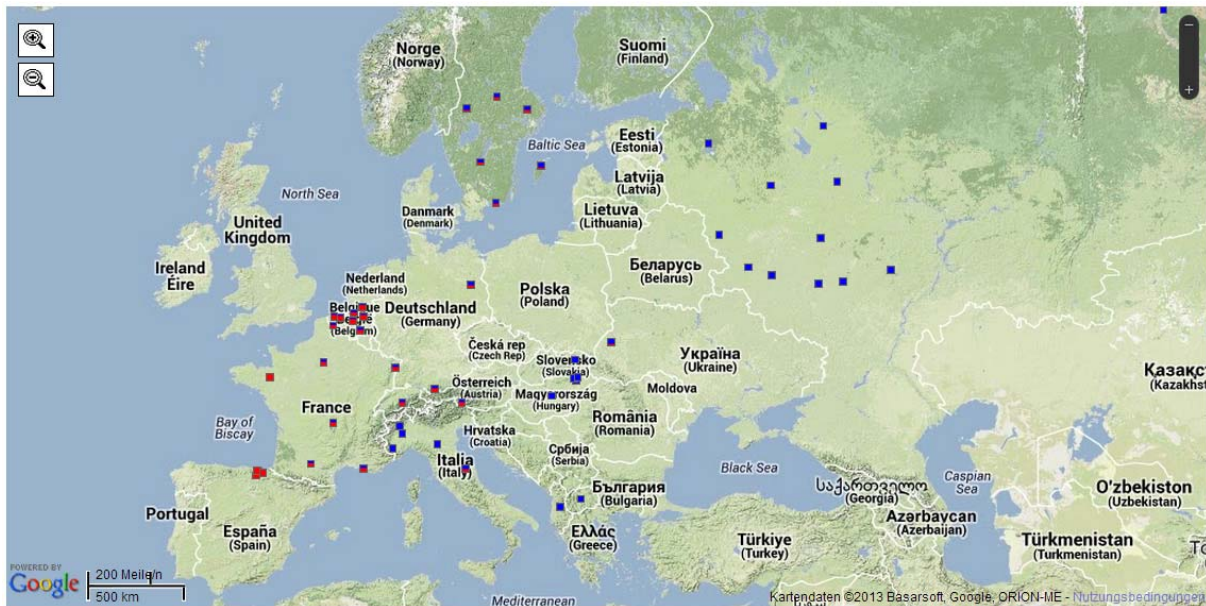


Abbildung 2: Geographische Verteilung der Haplogruppe R-M269 in der YHRD-Datenbank (www.yhrd.org). Ausschnitt der europäischen Treffer. Rote Quadrate stellen die Stichproben (\varnothing n=150) dar, in denen die Haplogruppe R-M269 beobachtet wurde. Das Verhältnis von rot zu blau korreliert mit der relativen Frequenz der Haplogruppe. Blaue Quadrate repräsentieren Geopositionen, an denen Haplotypen mit SNP Informationen vorhanden sind, aber die gesuchte Haplogruppe nicht beobachtet wurde.

In der zweiten Fallstudie (Publikation 3) war Gegenstand der Untersuchung ein Skelett, das in einer Berliner Kleingartenanlage gefunden wurde. Es sollte geklärt werden, ob es sich um eine kürzlich verstorbene Person oder um eine ältere Leiche, möglicherweise aus dem Zweiten Weltkrieg oder der Nachkriegszeit handelt. In der rechtsmedizinischen Begutachtung wurde ein Defekt des Schädels festgestellt, der mit einem Einschuss vereinbar war. Weiterhin gab es weder klare Anhaltspunkte zum Geschlecht noch zur Herkunft. Das Alter der Leiche wurde auf 40-46 Jahre eingegrenzt und mithilfe der Radiokarbon-Datierung wurde eine Liegezeit von mehr als 60 Jahren ermittelt. Die anschließende DNA-Untersuchung war daher das Mittel der Wahl, um die Herkunft und damit die mutmaßliche Liegezeit der menschlichen Überreste zu klären. Die genetische Untersuchung wies zwar nach, dass es sich um ein männliches Skelett handelte, eine Identifizierung im Abgleich mit der Vermisstendatei war aber nicht möglich. Daher wurden Y-chromosomale Untersuchungen angeschlossen, um die Herkunftsregion einschränken zu können. Bei dem Fund eines Skeletts in Berlin ist es zunächst einmal wahrscheinlicher, dass dieses eine europäische Herkunft hat. Aus diesem Grund wurde das eurasische Panel für die Bearbeitung des Falls verwendet. Dem Skelett konnte Haplogruppe N-Tat (N1c) zugeordnet werden, welche sich mit den höchsten Frequenzen in Finnland, im Baltikum und in Russland, insbesondere im nördlichen sowie mittleren Sibirien wiederfindet, wo die Frequenzen bis zu 90% erreichen. In Berlin, dem Fundort des Skelettes, kommt Haplogruppe N hingegen nur mit einer Frequenz von 1-3% vor (Publikation 3).

5.3 Phylogeographische Analysen am Beispiel einer endogamen und einer gemischten Population aus Lateinamerika

Im Rahmen der bisher größten Y-chromosomal basierten Studie zu südamerikanischen Populationen, unter Leitung unserer Arbeitsgruppe, wurden 1011 Individuen mit Y-STRs und Y-SNPs analysiert [27]. In dieser Zusammenarbeit wurde auch die indigene Volksgruppe der Waorani untersucht, die in Ecuador im Amazonastiefland beheimatet ist (Publikation 4). Die ehemals nomadischen Sammler und Jäger sind heute weitestgehend sesshaft. Sie haben eigene kulturelle Bräuche und distinkte physische Merkmale, wie z.B. eine durchschnittlich kleine, äußerst muskuläre Statur. Für die Y-SNP Analysen wurden die Multiplex-Assays SA Major, SA SpecC und SA SpecQ verwendet. Alle Individuen konnten mit diesem Typisierungsansatz einer Haplogruppe zugeordnet werden. Es wurden insgesamt 65 Männer aus zwei verschiedenen Waorani Dörfern untersucht. Während die Mehrzahl der untersuchten Individuen (n=42) laut Selbstauskunft zu der Gruppe der Waorani gehörte, waren auch andere ethnische Gruppen vertreten – 15 Kichwa, 2 Shuar, 4 Mestizo und 2 Afro-Ecuadorianer. Es wurden insgesamt vier verschiedene Haplogruppen detektiert. Ein Großteil der untersuchten Männer (85%, n=55) wiesen Haplogruppe Q-M3 (Q1a3a) auf, welche die dominante *founder* Haplogruppe Südamerikas ist und die höchste Frequenz in südamerikanischen Ureinwohnern aufweist. In 9% (n=6) der Stichprobe konnten Haplogruppen detektiert werden, die nicht autochthon sind, sondern auf europäische Einwanderer - Haplogruppe R (n=4) und afrikanische Sklaven – Haplogruppe E (n=2) zurückzuführen sind. Unerwartet war der Fund von Haplogruppe C-M217 (C3*) in vier der untersuchten Individuen. Diese ursprünglich aus Asien stammende Haplogruppe ist bisher nur ein weiteres Mal bei amerikanischen Ureinwohnern beobachtet worden, in zwei männlichen Individuen der kolumbianischen Population der Wayuu. Allerdings wurden bei diesen Individuen nicht alle Untergruppen von C-M217 typisiert, da diese damals zum Teil noch nicht beschrieben waren, so dass die Zugehörigkeit zur Paragruppe C-M217 nicht eindeutig geklärt ist. Damit ist der Nachweis von C-M217 in Südamerika bislang einzigartig. Er lässt sich mit den bisherigen Modellen zur Besiedlung des Kontinents nicht eindeutig erklären.

In der zweiten Populationsstudie wurde eine Mestizo-Population untersucht (Publikation 5) die eine rezente Misch-Population des zentralamerikanischen Staates Nicaragua repräsentiert. Um die hohe genetische Variabilität der untersuchten Stichprobe zu erfassen, mussten bis auf die Multiplex-Reaktion R (eine Zuordnung zu R1a oder R1b wurde schon mit Multiplex II des eurasischen Panels erreicht) und SpecC alle Multiplex-Assays für die Typisierung eingesetzt werden. Dennoch blieben 4% der untersuchten Individuen ohne spezifische

Haplogruppenzuordnung, da z.B. seltenere Haplogruppen wie S und T mit unspezifischen geographischen Verteilungen nicht in den verwendeten Multiplex-Reaktionen implementiert worden waren. Fast 47% der untersuchten 165 unverwandten Männer aus vielen Teilen des Landes konnten der eurasischen Haplogruppe R (anteilig 6,5% R1a und 93,5% R1b) zugeordnet werden. Eine weitere Haplogruppe eurasischen Ursprungs findet sich mit Haplogruppe J in 13% der Proben. Weitere 15% der untersuchten Mestizos gehören Haplogruppe E an, deren Untergruppen sich zum einen von Nordafrika (E-M35, 10% innerhalb der untersuchten Population) aus in Europa, in den Mittleren Osten und in Asien verbreitet haben und zum anderen insbesondere in Westafrika zu finden sind (E-M2, ca. 5%). Lediglich 15% der untersuchten Mestizo-Population weisen die für amerikanische Ureinwohner typischen Haplogruppen Q-M346 und Q-M3 auf. Die restlichen 6% der untersuchten Population verteilen sich auf die ebenfalls in Eurasien zum Teil weit verbreiteten Haplogruppen G und I. Zusätzlich zur Häufigkeitsverteilung der Y-SNPs konnte gezeigt werden, dass die Laboranalyse für die eindeutige Bestimmung von Haplogruppen alternativlos ist. Computer-Programme, die auf der Basis von Y-STRs Haplogruppen vorhersagen, weisen zum Teil hohe Fehlerraten auf (Publikation 5).

6 Diskussion

6.1 Einschätzung zu den Möglichkeiten und Grenzen der biogeographischen Herkunftsvorhersage

Die angeführten Studien zeigen, dass es prinzipiell möglich ist eine Vorhersage zur geographischen Herkunft einer unbekannt Person mittels Y-SNP Analyse zu treffen, um relevante Informationen für forensische Ermittlungen oder auch anthropologische Untersuchungen zu erhalten. Jede Mutation auf dem Y-Chromosom hat einen Ort und eine Zeit ihres ersten Auftretens. Damit kann die Herkunft und die Verbreitung einer Patriline, definiert durch diese Mutation, verfolgt werden. Wie stark die geographische Herkunftsregion eingeschränkt werden kann, hängt von der Häufigkeit und der Verbreitung der entsprechenden Haplogruppe ab. Die kontinentale Herkunft einer Probe ist aber in den meisten Fällen zweifelsfrei zu bestimmen.

In der Studie zu den spanischen Skelettfunden (Publikation 2) wurde zum einen die generische Haplogruppe R-M207 detektiert und zum anderen deren Untergruppe R-M269 (R1b1b2). Die Paragruppe R ist eurasischer Herkunft, lässt aber kaum eine weitere Eingrenzung der geographischen Herkunftspopulation zu. Es ist eine sehr gut untersuchte Haplogruppe, die aber sehr häufig und weltweit verbreitet ist (in der YHRD haben 3476 von

11996 Individuen Haplogruppe R aus 101 von 173 Populationen, *release 44*). Betrachtet man hingegen die Untergruppe R-M269 (R1b1b2), so grenzt sich diese deutlich von der zweiten großen Untergruppe R-M198 (R1a1a) ab. Beide Untergruppen bilden distinkte Cluster, die geographisch deutlich voneinander unterschieden werden können. Während R-M198 mit den höchsten Frequenzen in Osteuropa (in Russland mit ca. 50%) auftritt, ist R-M269 am stärksten in Süd- und Westeuropa vertreten. In Großbritannien beispielsweise gehören im Schnitt 80% der Haplogruppe R-M269 an, in Frankreich und Spanien liegen die Frequenzen bei 50 bis 60% [28]. Haplogruppe R-M269 ist demnach eine an der Atlantikküste verbreitete Haplogruppe. Eine noch genauere geographische Eingrenzung lässt sich mit Y-SNPs nicht vornehmen, hierzu wären Y-STRs nötig. Im Kontext der Studie, lässt sich sagen, dass der Fundort in etwa auch dem Herkunftsort entspricht. Um eine genauere Aussage treffen zu können, müssten weitere Untergruppen von R-M269 untersucht werden. Beispielsweise die Untergruppe R-S116, auf die immerhin noch 30-60% der R-M269 Individuen von der Iberischen Halbinsel entfallen [28]. In der hier vorgelegten Arbeit musste aber auf diese Analysen verzichtet werden, da nicht ausreichend DNA zur Verfügung stand. Es zeigt sich, dass die Y-SNP Analyse nur ein, wenn auch wichtiger, da faktischer Mosaikstein der interdisziplinären Analysen zur Herkunft und Abstammung der aufgefundenen Skelette sein kann.

Seltene und geographisch stark eingegrenzte Haplogruppen, die gut mit Populationsstudien belegt sind, lassen genauere Aussagen über die Herkunftspopulation zu, wie die zweite Fallstudie zeigen konnte. Von Haplogruppe N-Tat (N1c), wie sie in dem Skelett aus Berlin detektiert wurde, ist bekannt, dass sie in Ost-Asien entstanden ist und sich dann über Nordsibirien bis nach Ost-Europa ausgebreitet hat [29]. Heute findet man Haplogruppe N-Tat (N1c) vorrangig in Populationen aus Nord-Sibirien mit einer Frequenz von bis zu 90% und mit einer sehr klaren Verteilung in Europa, gekennzeichnet durch einen Ost-West-Gradienten in Skandinavien (Finnland ca. 60%, Schweden ca. 20%) ebenso wie zwischen den Baltischen Staaten (ca. 30%) und Polen (ca. 4%) [29] (Publikation 3). Im Hinblick auf diese Häufigkeitsverteilung ist eine osteuropäische Abstammung des Skelettes von größerer Wahrscheinlichkeit in Relation zur Abstammung aus einer anderen geographischen Region. Als Komponente eines erweiterten forensischen Gutachtens, wurde daher die Schlussfolgerung gezogen, dass es sich bei dem Toten um einen russischen Soldaten oder Kriegsgefangenen gehandelt haben könnte und nicht um ein deutsches Mord- oder Unfallopfer aus jüngerer Zeit. Die Polizei stellte daraufhin ihre Ermittlungen ein und übergab das Skelett der Kriegsgräberfürsorge.

Dieses Beispiel zeigt, wie auch schon bei Brion [20] beschrieben, dass durch die Bestimmung einer spezifischen Haplogruppe über die Haupthaplogruppe (z.B. R) hinaus eine präzisere Eingrenzung der Herkunftspopulation vorgenommen werden kann. Haplogruppen sind im Allgemeinen nie vollkommen auf eine Region beschränkt, sondern zeigen wie schon beschrieben oft einen Ausgangspunkt mit hohen Frequenzen sowie radial und graduell abnehmende Frequenzen. Zudem werden prähistorisch entstandene Cluster wie z.B. in Europa oder Südamerika durch die zunehmende Mobilität der Menschen komplex verändert. Bezugnehmend auf die forensische Begutachtung muss sich die Interpretation eines Untersuchungsergebnisses daher auf Wahrscheinlichkeitsaussagen beschränken. Diese können mit dem Bayes'schen Theorem und der Berechnung des Likelihood-Quotienten sowie der Verwendung einer adäquaten Datenbasis statistisch gestützt werden. Dabei wird die Häufigkeit des Auftretens der festgestellten Haplogruppe in den zu berücksichtigenden Regionen in Relation zueinander gesetzt, um eine Aussage zu treffen um wie viel wahrscheinlicher die Herkunft der Y-chromosomalen Linie aus der einen im Gegensatz zu der anderen Region ist. Für die zweite Fallstudie zu dem Skelettfund in Berlin würde eine solche Berechnung folgendermaßen aussehen: ausgehend von der Hypothese, dass es sich um ein deutsches Mordopfer oder aber einen russischen Kriegsbeteiligten handelt und diese beiden Annahmen gleich wahrscheinlich sind, kann die a priori Wahrscheinlichkeit auf 50% gesetzt werden. Durch eine Recherche in den Populationsstichproben der YHRD und dem Einbeziehen der aufgefundenen N-Tat Probe selbst ergibt sich in der westeuropäischen Metapopulation eine Frequenz von 0,01 $[(33+1)/(3423+1)]$ und in der osteuropäischen Metapopulation eine Frequenz von 0,12 $[(80+1)/(697+1)]$ (*release 44*, www.yhrd.org). Aus der Division jeweiligen Einzel-Frequenzen dieser Metapopulationen durch die Summe der Frequenzen beider Populationen zusammen lassen sich die Wahrscheinlichkeitswerte ableiten. Für die westeuropäische Metapopulation ergibt sich ein Wert von 0,08 (0,01/0,13) und für die osteuropäische Metapopulation ein Wert von 0,92 (0,12/0,13). Damit ist es elf Mal wahrscheinlicher, dass das Skelett osteuropäischer Abstammung ist an Stelle westeuropäischer.

Da in den vorliegenden Arbeiten ausschließlich Analysen des Y-Chromosoms vorgestellt wurden, ist die Vorhersage des geographischen Ursprungs auf die Beschreibung von Patrilinien begrenzt. Bei Wetton et al. [21] und Brion et al. [20] sowie in einer erst kürzlich publizierten Studie von van Oven et al. [30] wird in diesem Kontext diskutiert, ob die zusätzliche Untersuchung autosomaler AISNPs und mitochondrialer DNA eine umfassendere Rekonstruktion der geographischen Herkunft zulassen könnte. Allerdings liegen noch keine forensischen Studien oder Fallbeispiele zur Herkunftsprädiktion basierend auf der

kombinierten DNA-Analyse dieser drei unabhängigen molekulargenetischen Marker vor, so dass die Anwendbarkeit und der zusätzliche Informationsgewinn für die polizeiliche Fallarbeit schwierig einzuschätzen sind.

6.2 Diskussion der phylogeographischen Ergebnisse der Populationsstudien in Lateinamerika

Die heutige genetische Variabilität Amerikas ist durch die späte prähistorische Besiedlung durch eine relativ kleine ostasiatische Bevölkerungsgruppe und die massive europäische Kolonialisierung nach der Entdeckung des Kontinents durch Christoph Kolumbus (1492) geprägt. Nach der Beringia-Theorie [18], stammen alle indigenen Bevölkerungsgruppen Amerikas aus Asien und sind über eine heute nicht mehr existierende Landbrücke (Beringia) im nördlichen Beringmeer eingewandert. In den meisten archäologischen und genetischen Studien wird die erste Besiedlung auf den Zeitraum vor 20,000 bis 12,000 Jahren datiert. Erkenntnisse aus der Paläoklimatologie trugen stark zur Eingrenzung dieses Einwanderungs-Zeitraums bei. Es gilt als gesichert, dass Amerika damit der letzte Kontinent war, der von Menschen besiedelt worden ist [18]. Eine Ausbreitung der Besiedlung von Nord- über Zentral- nach Südamerika kann durch genetische Analysen belegt werden. Es gibt allerdings eine Reihe von Theorien zur Anzahl der Migrationswellen, den Wanderrouten innerhalb Amerikas und der Datierung [18, 27]. Die ursprünglichen Besiedlungsstrukturen wurden durch die massive europäische Einwanderung und die Dezimierung der indigenen Völker durch Kriege, aber vor allem durch Krankheiten signifikant verändert und können nur noch lückenhaft durch archäologische Funde und genetische Untersuchungen rekonstruiert werden. Letztere stützen sich häufig auf Y-chromosomale DNA-Marker, da diese wie bereits angeführt eine Filterung von Mutationsmustern nach Herkunftsort und Zeit ermöglichen. So können post-kolumbianisch eingewanderte Linien aus Europa (Siedler) und Afrika (Sklavenhandel) von autochthonen (indigenen) Linien asiatischer Herkunft unterschieden werden. Die Unterscheidung asiatischer Haplogruppen indigener Amerikaner von rezenten asiatischen Einwanderern ist hingegen nur im Kontext möglich und bedarf deshalb zusätzlicher Informationen, wie z.B. Angaben zu den untersuchten Probanden, ihrer Vorfahren und Hinweise auf Kontakt mit stammesfremden Personen. Um die Herkunft der in der Waorani-Studie detektierten C-M217 Linien zu klären, wurden anhand von genetischen Analysen und persönlich dokumentierten Auskünften der Probanden Stammbäume erstellt. Diese reflektieren die traditionell endogame Familienstruktur der untersuchten Dörfer. Es zeigten sich fast ausschließlich gemeinsame patrilineare Abstammungslinien (siehe Untersuchung von Y-STRs in [27]). Zudem belegt die Ethnogenese des Stammes die isolierte

Lebensweise der untersuchten Individuen (Erstkontakt mit Europäern erst 1958) [31]. In diesem Kontext ist es möglich, zu schlussfolgern, dass die gefundenen C-M217 (C3*) Linien nicht von rezenten Asiaten eingetragen wurden, sondern dass es sich hier um eine weitere *founder* Haplogruppe der indigenen Bevölkerung Südamerikas neben Q-M3 handelt. In der bisher größten Y-Chromosom basierten Studie Südamerikas [27] wurde das isolierte Auftreten von C-M217 in der indigenen Bevölkerung Ecuadors bestätigt. Zehn weitere indigene Individuen (Kichwa) aus derselben Amazonasprovinz konnten der Haplogruppe C-M217 zugeordnet werden. Alle übrigen ethnischen Gruppen aus Argentinien, Brasilien, Bolivien, Kolumbien, Peru und Venezuela wiesen ausnahmslos Haplogruppe Q-M3, sowie andere Q-Subtypen auf. Haplogruppe C-M217 bildet somit ein distinktes Cluster an der nördlichen Westküste Südamerikas. Ein mögliches Szenario für den Ursprung der Haplogruppe ist eine weitere auf ca. 6000 Jahre datierte Einwanderungswelle aus Ost-Asien über den Pazifik oder entlang der amerikanischen Pazifikküste [27].

Die Waorani/Kichwa-Studie in Ecuador verdeutlicht, dass die Interpretation der Y-Phylogeographie unbedingt in Abhängigkeit vom Kontext, d.h. im Zusammenhang mit zusätzlichen Informationen zur untersuchten Population durchgeführt werden sollte. Haplogruppe Q-M3 galt bis zu unserer Studie als die einzige *founder* Haplogruppe indigener Völker Amerikas, da sich frühere Studien zum einen nur auf die Untersuchung dieser einen Haplogruppe beschränkten und zum anderen alle nicht Q-M3 Individuen a priori als rezente Einwanderer aussortiert wurden. Nur eine umfangreiche Untersuchung einer großen Stichprobe, die den ganzen Kontinent umfasste, eine standardisierte, hochauflösende Analyse-Methode und das Einbeziehen von Stammbäumen sowie weiterer Informationen ermöglichte die Entdeckung von Haplogruppe C-M217 und die Kartierung bislang unbekannter präkolumbianischer genetischer Variabilität in der indigenen Bevölkerung Südamerikas.

In der zweiten Populationsstudie wurde eine Misch-Population (Mestizo) des mittelamerikanischen Staates Nicaragua (Publikation 5) untersucht. Im Gegensatz zur Populationsstudie in Ecuador und Südamerika wurden Individuen einbezogen, die in urbanen Zentren leben. Die Mestizo-Population ist somit viel stärkeren Ein- und Abwanderungen ausgesetzt, die zu einer regelmäßigen Veränderung des Genpools führen, als die traditionell, nahezu isoliert lebenden Gruppen Ecuadors. Dieser Unterschied zwischen der traditionell lebenden und der urbanen Population wird in den Ergebnissen der Y-SNP Analyse deutlich. In der Populationsstichprobe Ecuadors wiesen über 90% der Individuen einen indigenen Ursprung auf. In der Mischpopulation aus Nicaragua gehören gerade 15% der untersuchten

Individuen indigenen Patrilineen (nur 2 Individuen einer gemeinsamen) an, die Übrigen sind europäischer oder afrikanischer Abstammung aus post-kolumbianischer Besiedlung. Vergleicht man die Populationszusammensetzung von Nicaragua mit der der Iberischen Halbinsel, insbesondere Spanien, wird eine ähnliche Frequenzverteilung der nicht-indigenen Y-chromosomalen Haplogruppen deutlich (Publikation 5). Diese Struktur kann auf die Kolonisierung des Landes überwiegend durch spanische Einwanderer nach 1492 zurückgeführt werden.

Die beiden unterschiedlichen Populationsstudien zeigen wie wichtig es ist, auch traditionelle, autochthone Bevölkerungsgruppen zu untersuchen, um ursprüngliche Besiedlungen zu erkennen und historische demografische Verbreitungsmuster zu verstehen. Nur aufgrund der starken geographischen Differenzierung des Y-Chromosoms können nicht-indigene Haplogruppen auch in Mischpopulationen Zentral- oder Südamerikas sehr gut identifiziert werden und eine Trennung von prä- und postkolumbianischer Einwanderung nachvollzogen werden. Mit den Studien zu lateinamerikanischen Populationen konnte gezeigt werden, dass endogame und gemischte Populationen unterschiedliche genetische Mischungsgrade in Abhängigkeit von ihrer Ethnogenese aufweisen. Beide Formen existieren und müssen in forensisch genutzten Datenbanken wie der YHRD repräsentiert sein, um die Interpretation von Y-SNP Daten zu verbessern.

6.3 Einschätzung der praktischen forensischen Anwendung von Y-SNPs

Die Entwicklung von Untersuchungsmethoden am Y-Chromosom schreitet stetig voran. Auf dem Gebiet der Y-SNPs wird intensiv geforscht und publiziert (in der Literaturdatenbank PubMed findet man zum jetzigen Zeitpunkt 521 Studien zu Y-STRs und 59 zu Y-SNPs), allerdings wurde bislang kein forensischer Standard festgelegt, wie es für die Kernsysteme der (Y-) STRs seit Mitte der 90er Jahre der Fall ist. Kurzzeitig kam es zum kommerziellen Vertrieb eines hierarchischen Multiplex-Assays und dessen Anwendung [21]. Die Methode war aber an die Technologie der Durchflusszytometrie gekoppelt, die nicht zur Grundausstattung von forensischen Laboren zählt. Damit kann wohl auch begründet werden, warum sich diese Technik nicht durchsetzen konnte. Es bleibt daher jedem einzelnen Labor überlassen, eine für sich geeignete und an Forschungsschwerpunkte oder forensische Fragestellungen angepasste Strategie aus den publizierten Studien herauszufiltern und geeignete Typisierungsansätze für Y-SNPs zu entwerfen.

In den vorliegenden Studien konnte gezeigt werden, dass bei der Erstellung eines eigenen Analyse-Ansatzes eine vorangestellte Literaturrecherche (auch online auf www.yhrd.org)

notwendig ist, um zu klären welche Haplogruppen in der zu untersuchenden Population oder Populationsstichprobe mit welcher Häufigkeit vertreten sind. SNP-Marker der Haplogruppen mit der höchsten Frequenz und damit des wahrscheinlichsten Auftretens sollten im Anschluss an die Recherche in einer oder mehreren Multiplex-PCRs entsprechend ihrer Anordnung im phylogenetischen Baum zusammenstellt werden. Die SNP-Typisierung ist damit ein modulares System, dass anwendungsbezogen angepasst werden kann und muss. Der hierarchische Ansatz wurde erstmals von Brion [24] beschrieben und hat sich als Typisierungs-Methode auch in nachfolgenden Studien bewährt [25, 30]. Dem gegenübergestellt ist eine andere Herangehensweise, in der möglichst viele SNPs in einer einzigen Multiplex-Reaktion untersucht werden [20, 26]. Die Argumentation für diesen alternativen Ansatz ist der insgesamt geringere DNA-Einsatz, was bei forensischen Proben oft von Relevanz ist. Die hier beschriebenen eigenen Erfahrungen bei der Konzeption von Y-SNP Ansätzen haben allerdings gezeigt, dass die Ergebnisse von Multiplex-Reaktionen mit wenigen SNPs verlässlicher und leichter zu interpretieren sind. Der Farb-Hintergrund von kleinen Multiplexreaktionen ist erfahrungsgemäß geringer, da weniger unspezifische Produkte aufgrund von unspezifischen Primerbindungen oder Primerdimeren entstehen. Die Vermeidung von umfangreichen Multiplex-Reaktionen mit mehr als 12-15 simultan untersuchten SNPs ist bei der Verwendung des hierarchischen Ansatzes auch leicht umsetzbar. So ist der sinnvolle Einsatz der hierarchischen Methode für forensische Proben zu empfehlen, um die ohnehin schwierige Begutachtung der genetischen Analysen von Spuren so überschaubar wie möglich zu gestalten. Gerade Mischproben (Mann-Mann) zeigen bei sehr großen Multiplex-Reaktionen weniger eindeutige Auswertungen.

Im Hinblick auf die praktische Umsetzung können SNPs entsprechend der Ausstattung eines Labors mit verschiedenen Analysemethoden (z.B. *Pyrosequencing*, *Real-Time-PCR*, Massenspektrometrie, Microarray, KASPar SNP *Genotyping System*, *next generation sequencing*) untersucht werden. Die hier verwendete Methode der Multiplex-PCR in Kombination mit einer Minisequenzierung hat sich bereits in zahlreichen Studien zur Y-SNP Typisierungen bewährt [24-26, 30]. Sie ermöglicht die Kombination von mehreren SNPs, ist schnell, effizient und gut in die forensische Routine integrierbar, da die technischen Voraussetzungen denen anderer routinemäßig eingesetzter forensischer Methoden entsprechen (Thermocycler und Kapillarelektrophorese). Die Methode der Minisequenzierung erfüllt somit alle Voraussetzungen, um sich als forensischer Standard in der Analyse von Y-SNPs zu etablieren. Allgemein lässt sich sagen, dass durch die Variation nur eines einzelnen Nukleotids die Amplifikatlänge in der Analyse im Vergleich zu allen anderen

bereits in der forensischen Fallarbeit verwendeten Polymorphismen sehr kurz gehalten werden kann. Dies macht Y-SNPs zu wichtigen molekulargenetischen Markern, um schwieriges Probenmaterial mit degradiertem DNA zu untersuchen, wie in den Fallbeispielen demonstriert werden konnte. Allerdings zeigen die Ergebnisse zu den spanischen Skelettfunden auch, dass in Fällen mit besonders schwierigem oder begrenzt vorhandenem Untersuchungsmaterial die üblicherweise verwendete Multiplex-Analyse weniger geeignet sein kann und der Einsatz von Einzelanalysen der relevanten SNP-Marker vorzuziehen ist. In der Analyse der stark degradierten Proben konnte nur einer von 12 Markern in der Multiplexanalyse SA Major (Publikation 2) amplifiziert werden. Dies ist darauf zurückzuführen, dass in einer Multiplexreaktion die parallel eingesetzten Primerpaare in unmittelbarer Konkurrenz um die DNA stehen. So sind Multiplex-Ansätze zwar mit einem geringeren Zeitaufwand verbunden, sie benötigen aber auch eine höhere Ausgangskonzentration der DNA als Einzelanalysen und führen daher bei geringen Probenkonzentrationen bzw. stark fragmentierter DNA von sehr alten oder schlecht erhaltenen Skelett-Proben, wie im beschriebenen Beispiel, nur vereinzelt zu einem Ergebnis. Zudem lassen Einzelansätze in solchen Fällen eine deutlich bessere Anpassung der Analyse-Bedingungen für die SNP-Marker zu, um dennoch eine erfolgreiche Typisierung zu ermöglichen.

Neben der bisher gebräuchlichen Genotypisierung einzelner Marker, rückt zudem immer mehr die Sequenzierung gesamt-genomischer Abschnitte in den Fokus der Untersuchungen. Ganz aktuell zeigt eine Veröffentlichung aus der Arbeitsgruppe von Tyler-Smith [32], welchen zusätzlichen Informationsgewinn man mit der Re-Sequenzierung der mann-spezifischen Regionen des gesamten Y-Chromosoms erreicht. In dem untersuchten Sequenzbereich von nahezu 9 Megabasen konnten allein 5865 SNP-Marker identifiziert werden. Mit der fortschreitenden Entwicklung von neuen Technologien und den sinkenden Kosten für die Analysen wird die Bedeutung des *next generation sequencing* für die Untersuchung des Y-Chromosoms weiter zunehmen.

7 Fazit

Mit den Y-SNPs wird hier eine neue Kategorie forensischer Marker vorgestellt, die wie phänotypisch-assoziierte autosomale SNPs nicht vorrangig zur direkten Identifizierung durch Mustervergleich, sondern zur Beschreibung der Eigenschaften einer unbekanntem DNA und damit ihres Trägers eingesetzt werden können. Damit erweitern Y-SNPs die Möglichkeiten der Untersuchung von beweiserheblichen Spuren und generieren Informationen bei

schwierigen Fragestellungen der forensischen Praxis. Mit den Y-SNPs hat man lineare Marker zur Hand, die auch bei starker Degradation des Untersuchungsmaterials Analyseergebnisse liefern können und somit den Zugang zu problematischen Proben gewähren. Y-SNP Analysen können sowohl über Generationen hinweg genealogische Linien identifizieren, um Abstammungsfragen zu untersuchen als auch phylogeographische Informationen zur Eingrenzung der geographischen Herkunft (Kontinent oder Region) einer unbekanntes männlichen DNA liefern. Die Untersuchung der Phylogeographie des Y-Chromosoms bietet über die Standardmethoden der forensischen Praxis hinaus einen neuen strategischen Ansatz im Bereich der *DNA intelligence*, der es erlaubt in ungelösten Kriminalfällen ohne Vergleichsprobe den Personenkreis von möglichen Tatverdächtigen einzuschränken. In Vermisstenfällen oder Identifizierungsfällen von unbekanntes Toten oder Skeletten kann der Hinweis auf die geographische Herkunft neue Anstöße für polizeiliche Ermittlungen geben. Im Hinblick auf die populationsgenetische Forschung tragen Y-SNPs durch ihre feste Verankerung im phylogenetischen Stammbaum des Menschen zum Verständnis der menschlichen Evolution bei.

8 Literaturverzeichnis

- [1] M. Kayser, P. M. Schneider, DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations, *Forensic Sci Int Genet.* 3 (2009) 154-61.
- [2] M. A. Jobling, C. Tyler-Smith, The human Y chromosome: an evolutionary marker comes of age, *Nat Rev Genet.* 4 (2003) 598-612.
- [3] K. N. Ballantyne, M. Goedbloed, R. Fang, et al., Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications, *Am J Hum Genet.* 87 (2010) 341-53.
- [4] L. Roewer, Y chromosome STR typing in crime casework, *Forensic Sci Med Pathol.* 5 (2009) 77-84.
- [5] W. Parson, Bedeutung der mtDNA-Analyse für forensische Fragestellungen, *Rechtsmedizin.* 19 (2009) 183-194.
- [6] B. Budowle, A. van Daal, Forensically relevant SNP classes, *Biotechniques.* 44 (2008) 603-8, 610.
- [7] J. J. Sanchez, C. Phillips, C. Borsting, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, *Electrophoresis.* 27 (2006) 1713-24.
- [8] C. Haas, N. Shved, F. Rühli, et al., Y-chromosomal analysis of skeletal remains of Swiss national hero Jörg Jenatsch (1596-1639), *Forensic Sci Int Genet.* (in press).

- [9] S. Walsh, F. Liu, K. N. Ballantyne, et al., IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information, *Forensic Sci Int Genet.* 5 (2011) 170-180.
- [10] M. Fondevila, C. Phillips, C. Santos, et al., Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies, *Forensic Sci Int Genet.* 7 (2013) 63-74.
- [11] O. Ribaux, A. Girod, S. J. Walsh, et al., Forensic intelligence and crime analysis, *Law, Probability and Risk.* 2 (2003) 47-60.
- [12] M. F. Hammer, S. L. Zegura, The Human Y Chromosome Haplogroup Tree: Nomenclature and Phylogeography of Its Major Divisions, *Annu Rev Anthropol.* 31 (2002) 303-321.
- [13] L. Roewer, Populationsgenetik des Y-Chromosoms, *Medizinische Genetik.* 20 (2008) 288-292.
- [14] A nomenclature system for the tree of human Y-chromosomal binary haplogroups, *Genome Res.* 12 (2002) 339-48.
- [15] T. M. Karafet, F. L. Mendez, M. B. Meilerman, et al., New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, *Genome Res.* 18 (2008) 830-8.
- [16] A. Van Geystelen, R. Decorte, M. H. Larmuseau, Updating the Y-chromosomal phylogenetic tree for forensic applications based on whole genome SNPs, *Forensic Sci Int Genet.* (2013).
- [17] P. A. Underhill, G. Passarino, A. A. Lin, et al., The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations, *Ann Hum Genet.* 65 (2001) 43-62.
- [18] M. Jobling, M. E. Hurles, C. Tyler-Smith, *Human Evolutionary Genetics: origins, peoples and disease*, Garland Science Publishing, London/New York, 2004, p. 523.
- [19] M. A. Jobling, Y-chromosomal SNP haplotype diversity in forensic analysis, *Forensic Sci Int.* 118 (2001) 158-62.
- [20] M. Brion, J. J. Sanchez, K. Balogh, et al., Introduction of an single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages, *Electrophoresis.* 26 (2005) 4411-20.
- [21] J. H. Wetton, K. W. Tsang, H. Khan, Inferring the population of origin of DNA evidence within the UK by allele-specific hybridization of Y-SNPs, *Forensic Sci Int.* 152 (2005) 45-53.

- [22] M. A. Jobling, The impact of recent events on human genetic diversity, *Philos Trans R Soc Lond B Biol Sci.* 367 (2012) 793-9.
- [23] S. Willuweit, L. Roewer, Y chromosome haplotype reference database (YHRD): update, *Forensic Sci Int Genet.* 1 (2007) 83-7.
- [24] M. Brion, B. Sobrino, A. Blanco-Verea, et al., Hierarchical analysis of 30 Y-chromosome SNPs in European populations, *Int J Legal Med.* 119 (2005) 10-5.
- [25] V. Onofri, F. Alessandrini, C. Turchi, et al., Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs, *Forensic Sci Int.* 157 (2006) 23-35.
- [26] J. J. Sanchez, C. Borsting, C. Hallenberg, et al., Multiplex PCR and minisequencing of SNPs--a model with 35 Y chromosome SNPs, *Forensic Sci Int.* 137 (2003) 74-84.
- [27] L. Roewer, M. Nothnagel, L. Gusmao, et al., Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans, *PLoS Genet.* 9(4): e1003460. doi:10.1371/journal.pgen.1003460 (2013).
- [28] N. M. Myres, S. Rootsi, A. A. Lin, et al., A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe, *Eur J Hum Genet.* (2011).
- [29] S. Rootsi, L. A. Zhivotovsky, M. Baldovic, et al., A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe, *Eur J Hum Genet.* 15 (2007) 204-11.
- [30] M. van Oven, A. Ralf, M. Kayser, An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups, *Int J Legal Med.* 125 (2011) 879-85.
- [31] S. Beckerman, P. I. Erickson, J. Yost, et al., Life histories, blood revenge, and reproductive success among the Waorani of Ecuador, *Proc Natl Acad Sci U S A.* 106 (2009) 8134-9.
- [32] W. Wei, Q. Ayub, Y. Chen, et al., A calibrated human Y-chromosomal phylogeny based on resequencing, *Genome Res.* 23 (2013) 388-95.

9 Eidesstattliche Versicherung/Anteilserklärung

„Ich, Maria Geppert, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Hierarchische Y-SNP Analyse: Phylogeographische Daten für die forensische Praxis und populationsgenetische Forschung“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche in korrekter Zitierung (siehe „Uniform Requirements for Manuscripts (URM)“ des ICMJE -www.icmje.org) kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) entsprechen den URM (s.o) und werden von mir verantwortet.

Meine Anteile an den ausgewählten Publikationen entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem Betreuer, angegeben sind. Sämtliche Publikationen, die aus dieser Dissertation hervorgegangen sind und bei denen ich Autor bin, entsprechen den URM (s.o) und werden von mir verantwortet.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156,161 des Strafgesetzbuches) sind mir bekannt und bewusst.

Berlin, den

Maria Geppert

Dipl. Biol. Maria Geppert hatte folgenden Anteil an den vorgelegten Publikationen:

Publikation 1:

M. Geppert, L. Roewer (2012), SNaPshot[®] minisequencing analysis of multiple ancestry-informative Y-SNPs using capillary electrophoresis, DNA Electrophoresis protocols for forensic genetics, Humana Press, Methods Mol Biol. 830, 2012, 127-40

(90 Prozent)

Beitrag im Einzelnen: Konzeption und Umsetzung der Analysestrategie, selbständige Durchführung der Experimente, selbständige Auswertung der Ergebnisse, selbständige Manuskripterstellung.

Publikation 2:

C. Nunez, C. Sosa, M. Baeta, **M. Geppert**, M. Turnbough, N. Phillips, Y. Casalod, M. Bolea, R. Roby, B. Budowle, B. Martínez-Jarreta, Genetic analysis of 7 medieval skeletons from the Aragonese Pyrenees, Croat Med J. 52(3) (2011); 336-343. Impact Faktor: 1,796

(10 Prozent)

Beitrag im Einzelnen: Durchführung der Experimente zur SNP-Typisierung, Auswertung der SNP-Typisierung Ergebnisse, Beteiligung an der Manuskripterstellung.

Publikation 3:

M. Geppert, J. Rothe, S. Willuweit, M. Nagy, L. Roewer, Geographische Herkunftsbestimmung unbekannter DNA-Spuren, Rechtsmedizin. 20 (2010) 270-274. Impact Faktor: 0,814

(50 Prozent)

Beitrag im Einzelnen: selbständige Durchführung der Experimente zur SNP-Typisierung, Auswertung der Ergebnisse, Manuskripterstellung.

Publikation 4:

M. Geppert, M. Baeta, C. Nunez, B. Martinez-Jarreta, S. Zweynert, O.W. Cruz, F. Gonzalez-Andrade, J. Gonzalez-Solorzano, M. Nagy, L. Roewer, Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America, Forensic Sci Int Genet. 5 (2011) 100-104. Impact Faktor: 3,082

(80 Prozent)

Beitrag im Einzelnen: Konzeption und Umsetzung der Analysestrategie, selbständige Durchführung der Experimente, selbständige Auswertung der Ergebnisse, Manuskripterstellung.

Publikation 5:

C. Núñez, **M. Geppert**, M. Baeta, L. Roewer, B. Martínez-Jarreta (2012), Y chromosome haplogroup diversity in a Mestizo population of Nicaragua, Forensic Sci Int Genet. 6(6):e192-5. Impact Faktor: 3,082

(10 Prozent)

Beitrag im Einzelnen: Anleitung der Experimente zur SNP-Typisierung und Auswertung, Design neuer Multiplex-Assays für die erweiterte SNP-Typisierung als Anpassung an die untersuchten Proben, Korrektur des Manuskripts.

Berlin, den

Maria Geppert

10 Ausgewählte Publikationen

10.1 Publikation 1

M. Geppert, L. Roewer (2012), SNaPshot[®] minisequencing analysis of multiple ancestry-informative Y-SNPs using capillary electrophoresis, DNA Electrophoresis protocols for forensic genetics, Humana Press, Methods Mol Biol. 830, 2012, 127-40

http://dx.doi.org/10.1007/978-1-61779-461-2_9

10.2 Publikation 2

C. Núñez, C. Sosa, M. Baeta, **M. Geppert**, M. Turnbough, N. Phillips, Y. Casalod, M. Bolea, R. Roby, B. Budowle, B. Martínez-Jarreta, Genetic analysis of 7 medieval skeletons from the Aragonese Pyrenees, *Croat Med J.* 52(3) (2011); 336-343.

<http://dx.doi.org/10.3325/cmj.2011.52.336>

10.3 Publikation 3

M. Geppert, J. Rothe, S. Willuweit, M. Nagy, L. Roewer, Geographische
Herkunftsbestimmung unbekannter DNA-Spuren, Rechtsmedizin. 20 (2010) 270-274.

<http://dx.doi.org/10.1007/s00194-010-0692-2>

10.4 Publikation 4

M. Geppert, M. Baeta, C. Núñez, B. Martinez-Jarreta, S. Zweynert, O.W. Cruz, F. Gonzalez-Andrade, J. Gonzalez-Solorzano, M. Nagy, L. Roewer, Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America, *Forensic Sci Int: Genet.* 5 (2011) 100-104.

<http://dx.doi.org/10.1016/j.fsigen.2010.08.016>

10.5 Publikation 5

C. Núñez, **M. Geppert**, M. Baeta, L. Roewer, B. Martínez-Jarreta (2012), Y chromosome haplogroup diversity in a Mestizo population of Nicaragua, *Forensic Sci Int: Genet.* 6(6):e192-5

<http://dx.doi.org/10.1016/j.fsigen.2012.06.011>

11 Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

12 Komplette Publikationsliste

M. Geppert, J. Edelman, R. Lessig, The Y-chromosomal STRs DYS481, DYS570, DYS576 and DYS643, *Legal Medicine*. 11 (2009) S109-S110.

M. Baeta, C. Núñez, F. González-Andrade, C. Sosa, Y. Casalod, M. Bolea, S. Zweynert, O.W. Vacas Cruz, J. González-Solorzano, **M. Geppert**, L. Roewer and B. Martínez-Jarreta, Mitochondrial analysis revealed high homogeneity in the Waorani population—The last nomadic group of hunter-gatherers from Ecuador, *Forensic Sci Int Genet Suppl Series*. 2 (1) (2009), S313-314.

M. Geppert, J. Rothe, S. Willuweit, M. Nagy, L. Roewer, Geografische Herkunftsbestimmung unbekannter DNA-Spuren, *Rechtsmedizin*. 20 (2010) 270-274.

M. Geppert, M. Baeta, C. Nunez, B. Martinez-Jarreta, S. Zweynert, O.W. Cruz, F. Gonzalez-Andrade, J. Gonzalez-Solorzano, M. Nagy, L. Roewer, Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America, *Forensic Sci Int Genet*. 5 (2011) 100-104.

C. Nunez, C. Sosa, M. Baeta, **M. Geppert**, M. Turnbough, N. Phillips, Y. Casalod, M. Bolea, R. Roby, B. Budowle, B. Martínez-Jarreta, Genetic analysis of 7 medieval skeletons from the Aragonese Pyrenees, *Croat Med J*. 52(3) (2011); 336-343.

J.Purps, **M. Geppert**, M. Nagy, L. Roewer, Evaluation of the IrisPlex eye colour prediction tool in a German population sample, *Forensic Sci Int Genet Suppl Series*. 3 (1) (2011); e202-e203

M. Geppert, L. Roewer (2012), SNaPshot[®] minisequencing analysis of multiple ancestry-informative Y-SNPs using capillary electrophoresis, *DNA Electrophoresis protocols for forensic genetics*, Humana Press, *Methods Mol Biol*. 830, 2012, 127-40

L. Roewer, **M. Geppert** (2012), 'Interpretation Guidelines of a Standard Y-chromosome STR 17-plex PCR-CE Assay for Crime Casework.', *Methods Mol Biol*. 830, 2012, 43-56

C. Núñez, **M. Geppert**, M. Baeta, L. Roewer, B. Martínez-Jarreta (2012), Y chromosome haplogroup diversity in a Mestizo population of Nicaragua, *Forensic Sci Int Genet.* 6(6):e192-5.

M. Nothnagel, R. Szibor, O. Vollrath, C. Augustin, J. Edelmann, **M. Geppert**, C. Alves, L. Gusmao, M. Vennemann, Y. Hou, U. Immel, S. D.Inturri, H. Luo, S. Lutz-Bonengel, C. Robino, L. Roewer, B. Rolf, J. Sanft, K. Shin, J. Sim, P. E.Wiegand, C. Winkler, M. Krawczak, S. Hering, (2012) Collaborative genetic mapping of 12 forensic short tandem repeat (STR) loci on the human X chromosome. *Forensic Sci Int Genet.* 6(6):778-84.

C. Haas, N. Shved, F. Rühli, C. Papageorgopoulou, J. Purps, **M. Geppert**, S. Willuweit, L. Roewer, M. Krawczak (2013), Y-Chromosomal analysis of skeletal remains of Swiss national hero Jörg Jenatsch (1596-1639), *Forensic Sci Int Genet.* (in press).

H. Niederstätter, B. Berger, D. Erhart, S. Willuweit, **M. Geppert**, C. Gassner, H. Schennach, W. Parson, L. Roewer (2013) Multiple recurrent mutations at four human Y-chromosomal single nucleotide polymorphism sites in a 37bp sequence on the ARSDP1 sequence, *Forensic Sci Int Genet.* (in press).

L. Roewer, M. Nothnagel, L. Gusmao, V. Gomes, M. González, D. Corach, A. Sala, E. Alechine, T.J. Palha, N. Santos, A. Ribeiro-dos-Santos, **M. Geppert**, S. Willuweit, M. Nagy, S. Zweynert, M. Baeta, C. Núñez, B. Martínez-Jarreta, F. Gonzalez-Andrade, E. Fagundes de Carvalho, D. Aparecida da Silva, J.J. Builes, D. Turbón, A.M. Lopez Parra, E. Arroyo-Pardo, U. Toscanini, L. Borjas, C. Barletta, E. Ewart, S. Santos, M. Krawczak, Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans, *PLoS Genet.* 9(4): e1003460. doi:10.1371/journal.pgen.1003460 (2013).

Vorträge

M. Geppert, C.Hahne, S. Willuweit, M. Nagy, S.Schmidt, M. Tsokos, L. Roewer - Y-STR und Y-SNP Bestimmungen eines menschlichen Skeletts zur Eingrenzung der geographischen Herkunft, 29. Spurenworkshop in Verbindung mit der Deutschen Gesellschaft für Rechtsmedizin e.V. sowie der Spurenkommission der DGRM, 27./28. Februar 2009, Münster

M. Geppert, M. Baeta, C. Núñez, S. Zweynert, O. Vacas, J. González-Solorzano, S. Willuweit, M. Nagy, B. Martínez-Jarreta, F. Gonzáles-Andrade, L. Roewer - Inferring the demographic history of the last nomad hunter –gatherer population in Ecuador – the Waorani-

using lineal and recombining DNA markers, 23rd World Congress International Society of Forensic Genetics, 14.-18. September 2009, Buenos Aires, Argentinien

M. Geppert, L. Roewer - Herkunftsbestimmung von männlichen DNA Spuren bei Sexualdelikten durch Analyse von Y-chromosomalen SNPs, 31. Spurenworkshop in Verbindung mit der Deutschen Gesellschaft für Rechtsmedizin e.V. sowie der Spurenkommission der DGRM, 25.-26. Februar 2011, Hamburg

Poster

M. Geppert, M. Nagy, L. Roewer - Hierarchical Y-SNP assay to study indigenous populations of South America, Haploid DNA markers in forensic genetics - 7th International Y Chromosome User Workshop and 4th International EMPOP meeting, 22.-24. April 2010, Berlin

M. Geppert, L. Roewer, M. Nagy, Where sexual assailants leave their DNA profiles: Highlighting the importance of testing Y-chromosomal markers, 24rd World Congress International Society of Forensic Genetics, 29.-03. August/September 2011, Wien, Österreich

J. Purps, **M. Geppert**, M. Nagy M, L. Roewer, Evaluation of the irisplex eye color prediction tool in a German population sample, 24rd World Congress International Society of Forensic Genetics, 29.-03. August/September 2011, Wien, Österreich

Rothe, J.¹, Watkins, N.E., Jr.², **Geppert**, M.¹, Nagy, M.¹, Visual OMPtm software; an excellent tool for effective probe/primer design in complex PCR reactions and allele specific assays, 24rd World Congress International Society of Forensic Genetics, 29.-03. August/September 2011, Wien, Österreich

13 Danksagung

Ich möchte an dieser Stelle allen Personen danken, die zu der Entstehung dieser Arbeit beigetragen haben.

Mein besonderer Dank gilt dabei meinem Doktorvater Prof. Dr. Lutz Roewer, der mir es ermöglicht hat meine Promotion in der Abteilung Forensische Genetik durchzuführen. Lutz, du hast mich mit sehr viel Engagement und guten Ideen betreut. Ich danke dir für dein Vertrauen und die tolle Zusammenarbeit in so vielen interessanten Projekten. Deine Begeisterung für die Wissenschaft überträgt sich und motiviert.

Bei Frau PD Dr. Marion Nagy möchte ich mich ebenfalls herzlich bedanken für ihr Vertrauen in meine Arbeit, die wertvollen Hinweise zu dieser Arbeit und die Unterstützung in den unliebsamen organisatorischen Fragen der letzten Jahre.

Beiden möchte ich auch dafür danken, dass ich bereits an so zahlreichen, überaus wertvollen Kongressen teilnehmen durfte, um unsere gemeinsamen Forschungsprojekte zu präsentieren.

Bei dem gesamten Team der Forensischen Genetik möchte ich mich für die sehr angenehme Arbeitsatmosphäre, die gute Zusammenarbeit und für den gegenseitigen Austausch danken.

Dabei danke ich Petra Anders, Bärbel Henske, Carmen Krüger, Petra Otremba und Corinna Weyer vor allem für die technische Unterstützung und die ständige Hilfsbereitschaft bei den Problemen des Laboralltags.

Bei meinen Kollegen, Jessica Rothe, Sascha Willuweit, Patricia Entz, Jessica Teschner, Judith Zander und Josephine Purps möchte ich mich für die vielen anregenden Diskussion und die produktive wissenschaftliche Zusammenarbeit bedanken.

Zu guter Letzt gilt meiner Familie großer Dank für ihren Zuspruch und die ausdauernde Unterstützung, die entscheidend zum Vollenden dieser Arbeit beigetragen haben. Meinem Freund Ronald danke ich im Besonderen für sein Verständnis, das Schaffen von Freiräumen sowie das Korrekturlesen und die grafische Hilfestellung bei dieser Arbeit.