

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

DISSERTATION

Analyse von forensisch relevanten short tandem repeats (STRs) mittels massiver
paralleler Sequenzierung (MPS) und Vergleich zum Standardverfahren
Kapillarelektrophorese (CE)

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

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von
Steffi Bredemeyer
aus Jena

Datum der Promotion: 26.06.2022

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abkürzungen und Definitionen	II
Zusammenfassung	1
1 Abstract	1
2 Einführung	2
3 Zielstellung	5
4 Material und Methodik	5
4.1 Geräte und Technologien.....	5
4.2 Probenauswahl.....	6
4.3 STR-Typisierung mittels Kapillarelektrophorese (CE)	6
4.4 STR-Typisierung mittels massively parallel sequencing (MPS)	7
5 Ergebnisse	9
5.1 Konkordanz und Reproduzierbarkeit	9
5.2 Coverage, inter- und intra-locus balance	10
5.3 Sensitivität	12
5.4 Mischspuren.....	13
5.5 Alte DNA.....	14
5.6 Mutationen in Vater-Sohn-Paaren	15
5.7 Y-STR Allel-Diversität.....	15
6 Diskussion	16
7 Referenzen	20
Anteilerklärung/ Eidesstattliche Versicherung	24
Ausgewählte Publikationen	27
Publikation 1	27
Publikation 2	37
Publikation 3	52
Publikation 4	60
Lebenslauf	75
Komplette Publikationsliste	76
Danksagung	78

Abkürzungen und Definitionen

ACD	<i>allele coverage difference</i> (Differenz der <i>Coverage</i> zweier Allele)
ACR	<i>allele coverage ratio</i> (Verhältnis der <i>Coverage</i> zweier Allele)
aDNA	<i>ancient DNA</i> , Definition: sehr alte, teils prähistorische DNA
aSTR	<i>short tandem repeat</i> auf einem autosomalen Chromosom
BER	Abteilung Forensische Genetik, Institut für Rechtsmedizin, Charité – Universitätsmedizin
bp	Basenpaar
CE	<i>capillary electrophoresis</i> (Kapillarelektrophorese)
<i>Coverage</i>	Abdeckung, Definition: Anzahl der Sequenz-Reads, die einen bestimmten Locus oder ein bestimmtes Allel abdecken
DNA	<i>desoxyribonucleic acid</i> (Desoxyribonukleinsäure)
<i>drop-in</i>	zusätzlich detektiertes Allel (Artefakt), welches nicht von bekannten Spendern des DNA-Profiles stammt
<i>drop-out</i>	Allelausfall
ESS-Marker	<i>European Standard Set</i> , Definition: 12 europäische Standard-Marker, die bei der Erstellung von DNA-Profilen in Europa immer analysiert werden
<i>flanking region</i>	flankierende Region, Definition: Region außerhalb des <i>Repeat</i> -Musters eines short tandem repeats bis zur Erkennungssequenz der Primer
GEDNAP	<i>German DNA Profiling</i> (jährlich stattfindender Spuren-Ringversuch)
GMI	Institut für Gerichtliche Medizin der Medizinischen Universität Innsbruck
HB	<i>heterozygote balance/ intra-locus balance</i> (Ausgewogenheit der <i>Coverage</i> beider Allele eines Markers)
iSNP	<i>identity single nucleotide polymorphism</i> (Einzelnukleotidpolymorphismus, der zur Identitätsbestimmung genutzt wird)
IHZ	Isometrische Heterozygote
LR	<i>likelihood ratio</i> (Plausibilitätsquotient), Definition hier: Maß für die Beweiskraft bezüglich der Hypothese, dass zwei DNA-Profile vom selben Verdächtigen abgeleitet wurden
LUS	<i>longest uninterrupted repeat sequence</i> (längste ununterbrochene <i>Repeat</i> -Sequenz eines <i>short tandem repeats</i>)
µl	Mikroliter
MPS	<i>massively parallel sequencing</i> (massiv-parallele DNA-Sequenzierung)

mtDNA	mitochondriale Desoxyribonukleinsäure
ng	Nanogramm
NGS	<i>next generation sequencing</i> (Sequenzierung der nächsten Generation)
NIST	<i>National Institute of Standards and Technology</i>
PCR	<i>polymerase chain reaction</i> (Polymerasekettenreaktion)
pg	Pikogramm
<i>repeat region/ repeat</i>	Definition: repetitives Sequenzmuster innerhalb eines short tandem repeats
RM Y-STR	<i>rapidly mutating Y-chromosomal short tandem repeat</i> (schnell mutierendes short tandem repeat auf dem Y-Chromosom)
SF	<i>stutter filter threshold</i> (Messgrenze, die Stotterallele von „echten“ Allelen trennt)
SNP	<i>single nucleotide polymorphism</i> (Einzelnukleotidpolymorphismus)
STR	<i>short tandem repeat</i> (Mikrosatellit)
<i>threshold</i>	Messgrenze
UAS	<i>Universal Analysis Software</i>
X-STR	<i>short tandem repeat</i> auf einem X-Chromosom
Y-STR	<i>short tandem repeat</i> auf dem Y-Chromosom

Zusammenfassung

1 Abstract

Eine der grundlegenden Aufgaben der Forensischen Genetik besteht in der eindeutigen Zuordnung einer Tatortspur zu einer Person anhand des DNA-Profiles. Dazu werden kurze, sich wiederholende Bereiche der nichtcodierenden DNA untersucht, die als Mikrosatelliten oder *short tandem repeats* (STRs) bezeichnet werden. Sie unterscheiden sich von Person zu Person in der Anzahl ihrer Wiederholungen. Für die eindeutige Identifizierung einer Person ist die Kombination mehrerer STR-Marker in einem STR-Profil nötig, welches seit den 90er Jahren mit Hilfe der Kapillarelektrophorese (CE) erstellt wird. Diese Methode ist kostengünstig und zeitsparend, liefert allerdings nur die Information der unterschiedlichen Längen der Marker und erschwert die Kombination von mehr als 30 STR-Markern in einer Analyse. Um diese Nachteile zu eliminieren, wird seit einigen Jahren die Technologie *massively parallel sequencing* (MPS), auch *next generation sequencing* (NGS) genannt, in der Forensischen Genetik angewandt. Diese Technologie ermöglicht die gleichzeitige Sequenzierung von derzeit bis zu 58 STR-Markern sowie die Detektion zusätzlicher Sequenzvariationen dieser Marker. Meine Promotionsarbeit befasst sich mit der Analyse forensisch relevanter STR-Marker mittels MPS. Die technisch-forensische Validierung der Methode ist Gegenstand zweier Publikationen [1, 2]. Hier wurde die STR-Sequenzierung mit der Standard-Methode CE verglichen, da konkordante, reproduzierbare Ergebnisse für die Erstellung forensischer DNA-Profile von hoher Bedeutung sind. Dabei stellte sich heraus, dass MPS laborübergreifend konkordante und reproduzierbare Ergebnisse liefert. Die Qualität entspricht allerdings noch nicht vollständig derjenigen der CE-Analyse, was sich in vermehrten Allel-Ausfällen (*drop-outs*) und Stotterallelen sowie einer suboptimalen *inter-locus- und heterozygote balance* widerspiegelt. Des Weiteren wurden laborübergreifende Unterschiede der Geräteleistung der Sequenzierplattform aufgedeckt. Anschließend überprüfte ich, welche Vorteile die Sequenzinformation (MPS) gegenüber der Längeninformation (CE) für die forensische Anwendung bringt. Dazu publizierte ich eine weitere Studie [3] zur Sequenzierung Y-chromosomal lokalisierter STR-Profile von 64 Vater-Sohn-Paaren, um zu prüfen, ob MPS das Aufdecken von Sequenzvariationen und somit die weitere Unterscheidung männlicher Verwandter ermöglicht. Es wurden keine zusätzlichen Sequenzvariationen zwischen Vätern und ihren jeweiligen Söhnen detektiert. Jedoch wurden Sequenzvariationen zwischen nicht verwandten Männern festgestellt, wodurch die Allel-Diversität der Y-STR-Marker und damit das Individualisierungspotential erhöht wird. Die Sequenzinformation der STRs ermöglichte außerdem die genaue Lokalisation der Längenvariationen, was hilfreich ist, um Mutationsereignisse besser zu verstehen. Eine Erhöhung der Diversität wurde anhand der Konkordanzstudien auch für autosomale STRs festgestellt [1]. Dadurch kann die Auswertung von komplexen Mischspuren oder von komplexen Verwandtschaftsstrukturen erleichtert werden.

Meine Promotionsarbeit zeigt, dass der Vorteil der MPS-Technologie eindeutig im Informationsgewinn liegt und damit die forensische Spuren-Analyse auf ein neues Niveau hebt.

One of the basic tasks of forensic genetics is the unambiguous assignment of a crime scene trace to a person using DNA profiles. For this purpose, short, repetitive regions of non-coding DNA are examined. These are called microsatellites or short tandem repeats (STRs). STRs differ from person to person in the number of their repeats. The combination of several STR markers in one STR-profile is necessary to identify a person unambiguously. Since the 1990s STR-profiles have been generated using capillary electrophoresis (CE), which is a quick and cost-effective method. This method reveals the length-based information only and complicates the simultaneous analysis of more than 30 markers in one analysis. To eliminate these disadvantages, the technology of massively parallel sequencing (MPS), also called next generation sequencing (NGS), has been applied in forensic genetics. MPS enables the simultaneous sequencing of currently 58 STR markers and reveals additional sequence information. My thesis focuses on the analysis of forensically relevant STR markers using MPS. First, the method was validated and compared to CE as part of two publications [1, 2], since concordant, reproducible results are of high importance for forensic DNA profiling. The published results show concordant and reproducible results across participating laboratories. However, the quality does not yet match the quality of CE analysis, showing increased drop-outs and stutter alleles, a poorer inter-locus- and heterozygote balance. Furthermore, inter-laboratory differences in the performance of the sequencing devices were revealed. In a further publication [3] I reviewed the advantages of sequence information (MPS) over length information (CE). I sequenced 64 father-son pairs to test whether MPS allows the detection of sequence variation and thus a further distinction of male relatives. Here, no additional sequence variations were detected between fathers and their respective sons. However, sequence variation was found between unrelated males, resulting in an increase of allele diversity of the Y-STR-markers. In addition, sequence information allowed the exact localization of length variations, which is helpful to improve the understanding of mutational events. An increase of allele diversity was also detected for autosomal STRs during the concordance studies [1]. This can facilitate the evaluation of complex mixed traces or of complex kinship cases. My thesis shows that the advantage of the MPS technology for forensic analysis is the gain of detectable DNA variability. This is relevant for specific forensic cases, and generally raises forensic DNA analysis to a new level.

2 Einführung

Die Forensische Genetik beschäftigt sich mit der Untersuchung von biologischem Material zur Identifizierung von Personen anhand ihrer Desoxyribonukleinsäure (DNA), um Straftaten aufzuklären, Verwandtschaftsverhältnisse zu bestimmen oder unbekannt Verstorbene zu identifizieren. Die humane DNA wird dabei an kurzen nichtcodierenden Bereichen von repetitiven Sequenzmustern (*repeat region* oder *repeat*), umgeben von eindeutigen

Erkennungssequenzen, untersucht. Diese spezielle Art eines DNA-Polymorphismus wird Mikrosatellit oder engl. *short tandem repeat* (STR) genannt. Durch den wiederholenden Charakter dieser STRs kommt es während der Transkription der DNA zu einer erhöhten Fehlerrate und infolgedessen (meiotische Rekombination) zu einer hohen allelischen Variabilität innerhalb einer Population – also zu verschiedenen Ausprägungen dieses Merkmals (Allel). Die Länge einer Wiederholungseinheit der Sequenzmuster kann dabei 2 bis 7 Basenpaare (bp) betragen, wobei solche mit 4 bp (beispielsweise GATA) zu den häufigsten gehören [4]. Auf Grundlage ihrer Komposition werden STRs in die Kategorien einfach, zusammengesetzt oder komplex unterteilt: Bei identischer Länge und Sequenz aller Wiederholungseinheiten (z. B. [GATA]₁₀) handelt es sich um ein einfaches STR, zwei oder mehr benachbarte einfache Wiederholungen bilden einen zusammengesetzten STR (z. B. [TATA]₄[GATA]₆) und mehrere Wiederholungsblöcke mit variabler Länge sowie variablen Zwischensequenzen bilden einen komplexen STR [4]. In der Literatur finden wir neben STR auch die Begriffe STR-Marker und STR-Locus, hier wird auf die Definition bzw. Lokalisation der STRs durch die sie umgebenden Erkennungssequenzen hingewiesen.

Durch die Auswertung verschiedener, durch Fachkommissionen (wie z. B. ENFSI) festgelegter STRs auf den autosomalen Chromosomen (aSTRs) wird das so genannte genetische Profil einer Person eindeutig ermittelt (mit Ausnahme von eineiigen Zwillingen, die das gleiche genetische STR-Profil tragen). Für spezifischere Fragestellungen (beispielsweise bei der Abstammung) werden auch STRs auf dem Y- oder X-Chromosom (Y-/X-STRs) verwendet.

Seit den 1990er Jahren wird das genetische Profil einer Person oder einer Spur technisch mittels Kapillarelektrophorese (CE) festgestellt, die eine günstige und zeitsparende Analysemöglichkeit darstellt (Publikation 4 [5]). Dennoch weist diese Methode drei wesentliche Nachteile auf: (i) die Methode dient nur der Messung der Länge eines STRs, die tatsächliche Sequenzinformation bleibt somit unsichtbar, (ii) die Auftrennung verschiedener STRs in einer parallelen Mehrfachreaktion (Multiplex) anhand ihrer Länge bedingt, dass jeder getestete STR für die eindeutige Zuordnung in einem unterschiedlichen Längenbereich liegen bzw. farblich unterschiedlich markiert werden muss, (iii) die verlässliche Bestimmung ist auf eine Sequenzlänge von höchstens 600 bp beschränkt. Das heißt, bei einem maximal nutzbaren Bereich von ca. 80 bis 500 bp und der gleichzeitigen Nutzung von bis zu 6 Farbkanälen ist die Anzahl der möglichen kombinierten Marker zurzeit auf maximal 30 STR-Marker begrenzt [6-8].

So hat in den letzten Jahren die Technologie des *massively parallel sequencing* (MPS), auch *next generation sequencing* (NGS) genannt, Einzug in die Forensik gehalten mit dem Ziel, diese Nachteile zu eliminieren (Publikation 4 [5]). Mit Hilfe dieser Technologie können nicht nur die Längenvarianten des einzelnen STRs, sondern auch dessen Sequenzvarianten detektiert werden. Diese Sequenzvarianten können innerhalb der Repeat-Struktur eines STRs liegen oder außerhalb in der von den Primern flankierten Erkennungssequenz (*flanking region*). Verschiedene Studien konnten zeigen, dass zusätzliche Sequenzvarianten die Anzahl möglicher Allele eines STR-

Markers erheblich erhöhen können, womit die Wahrscheinlichkeit der eindeutigen Person-Spur-Zuordnung steigt [9-19]. Die Kenntnis der Erkennungssequenz in den Analyseergebnissen hebt die Beschränkung der Längenauftrennung auf und ermöglicht der MPS-Technologie die Kombination sehr vieler STR-Marker. Kommerziell werden bisher maximal 58 STR-Marker [20] kombiniert. Wie bei der CE gibt es auch bei MPS eine Beschränkung der maximalen Länge eines STRs für die sichere Bestimmbarkeit. Sie liegt zurzeit deutlich unter der der CE mit ca. 300 bp.

Um die MPS-Technologie in der Routinearbeit anwenden zu können, muss diese Methode zunächst umfänglich validiert werden und auf Kompatibilität mit der Standardmethode CE überprüft werden. In der Forensik sind für die Validierung die Reproduzierbarkeit und Konkordanz der Ergebnisse, die Sensitivität, die Auswertbarkeit von Mischspuren sowie von inhibierten Proben und Proben aus der Fallarbeit von Bedeutung. Darüber hinaus spielt bei der erfolgreichen Implementierung einer neuen Technologie in den Arbeitsablauf der forensischen DNA-Analyse die Konformität der Ergebnisse aus verschiedenen Labors und die Ausarbeitung einer einheitlichen Nomenklatur der neu detektierten, auf Sequenzvarianten basierenden Allele eine wichtige Rolle. Es wurden bereits Validierungsstudien zu diesen forensischen Kernthemen veröffentlicht [15, 21-31], die sich jedoch nicht mit der Problematik der Konformität zwischen den Laboren beschäftigen (Publikation 4 [5]).

Zusammen mit dem *National Institute of Toxicology and Forensic Sciences Spain* (Abteilung in Madrid) und der DNA-Abteilung des Instituts für Gerichtliche Medizin der Medizinischen Universität Innsbruck (GMI) arbeitete ich am Institut für Rechtsmedizin, Charité – Universitätsmedizin Berlin (BER) im EU-finanzierten Projekt *DNASeqEx*. Ziel dieses Projekts war es, die MPS-Technologie in der Forensik zu implementieren, die zwei derzeit verfügbaren MPS-Plattformen zu vergleichen und eine laborübergreifende Konformität der Ergebnisse zu prüfen (Publikation 1 [1] und 2 [2]). Im Review Paper (Publikation 4 [5]) haben alle Beteiligten des DNASeqEx Projekts den aktuellen Stand der Technik der STR-Sequenzierung in der Forensischen Genetik zusammengefasst.

Zusätzlich beschäftige ich mich mit der Durchführung von Y-chromosomalen Verwandtschaftsstudien mittels MPS, um die Sequenzvariabilität von Y-STRs zu ermitteln (Publikation 3 [3]). Hintergrund dieses Projektes ist, dass Tatortspuren oft Mischungen verschiedener Personen und damit verschiedener DNA aufweisen, die eine eindeutige Identifizierung einer einzelnen Person erschweren. Bei Straftaten gegen die sexuelle Selbstbestimmung werden vorwiegend Abstriche aus dem Körperinneren (Vagina, Anus, Mund) entnommen, die erwartungsgemäß fast ausschließlich aus Material der weiblichen Geschädigten bestehen. In der autosomalen STR-Analyse kann dann oft nur das weibliche Profil detektiert werden. Von solchen Spuren werden in unserer Abteilung gleichzeitig auch mann-spezifische Y-chromosomale STR-Analysen durchgeführt [32]. Aufgrund der Y-Chromosom-Spezifität kann oft ein „single source“ Y-Profil des männlichen Täters erstellt werden, auch wenn die männliche

Komponente im autosomalen Profil aufgrund des Mischungsverhältnisses nicht festgestellt werden konnte. Das Problem der Y-STRs zum Zweck der Identifizierung ist aber, dass männliche Verwandte einer paternalen Linie in mehreren Generationen auf Grund fehlender Rekombination oft das gleiche Y-Chromosom tragen. Deshalb kann ein Y-chromosomales STR-Profil oft nicht zwischen einem Tatverdächtigen und beispielsweise seinem Bruder oder Vater unterscheiden. Neben der Längensvariante (CE) lassen sich aber mit Hilfe der MPS-Technologie auch andere Sequenzveränderungen feststellen, die eine Unterscheidung zwischen Tatverdächtigem und seinen männlichen Verwandten wahrscheinlicher machen.

3 Zielstellung

Seit den 90er Jahren wird das STR-Profil einer Person oder einer Spur mittels der günstigen und zeitsparenden Analysetechnik CE erstellt (Review Paper: Publikation 4 [5]). Ziel meiner Promotionsarbeit ist es, forensisch bekannte STR-Marker mit der neueren Technologie MPS zu analysieren. Dabei steht die Validierung der Methode im Vordergrund. Für die Erstellung forensischer DNA-Profile im nationalen, europäischen und weltweiten Kontext sind konkordante, reproduzierbare Ergebnisse von hoher Bedeutung. Deshalb soll MPS zunächst mit der bestehenden Methode CE verglichen werden. Im Detail werden Konkordanz, Reproduzierbarkeit, Sensitivität sowie die Analyse von Mischspuren verglichen und überprüft. Wichtig ist dabei nicht nur der Vergleich der Daten im eigenen Labor, sondern auch, dass reproduzierbare Ergebnisse laborübergreifend generiert werden können (Publikation 1 [1] und 2 [2]). Des Weiteren soll überprüft werden, ob der Erhalt der Sequenzinformation für die Auswertung forensischer DNA-Profile ein nützlicher bzw. entscheidender Gewinn ist. Zu diesem Zweck werden Y-STR-Profile mehrerer Vater-Sohn-Paare sequenziert, um zu prüfen, ob die MPS-Analyse durch das Aufdecken von Sequenzvariationen die weitere Unterscheidung männlicher Verwandter ermöglicht (Publikation 3 [3]).

4 Material und Methodik

4.1 Geräte und Technologien

Zur MPS-Analyse gibt es derzeit zwei validierte Plattformen/Technologien:

- (i) *Ion Torrent* (Thermo Fisher Scientific, Waltham, USA): Halbleiterverfahren, um mittels Halbleiterchip-Detektion von Ionen eine unmittelbare nicht optische Sequenzierung durchzuführen
- (ii) *MiSeq/ MiSeq FGx* (Illumina, San Diego, USA): Optische *in situ* Detektion von vier verschiedenfarbig fluoreszierenden Kettenabbruchsubstraten entsprechend der vier Nukleinsäuren.

Für die Forensik stehen zurzeit drei kommerzielle MPS-STR-Kits zur Verfügung:

- (i) *ForenSeq DNA Signature Prep Kit* (Verogen, San Diego, USA) [20]

- 27 aSTRs, Amelogenin, 24 Y-S
- TRs, 7 X-STRs und 94 *identity single nucleotide polymorphisms (iSNPs)*
- kompatibel mit der Sequenzierplattform *MiSeq*
- Auswertung mit Universal Analysis Software (UAS)
- (ii) *Precision ID Globalfiler™ V2.0 NGS STR Panel v2* (Thermo Fisher Scientific) [33]
 - 31 aSTRs, 1 Y-STR und 3 Geschlechtsmarker (Amelogenin, SRY und rs2032678)
 - kompatibel mit der Sequenzierplattform Ion Torrent
 - Auswertung mit Torrent Suite Software und Converge Software
- (iii) *PowerSeq™ Auto/Mito/Y System* (Promega, Madison, USA) [34]
 - verschiedene Multiplex-Sets an: 22 aSTRs (+ Amelogenin), 23 Y-STRs, 10 Amplikons auf der Kontrollregion der mitochondrialen DNA (mtDNA)
 - kompatibel mit der Sequenzierplattform *MiSeq*
 - keine Analysesoftware: *open-source* oder eigenentwickelte Software.

4.2 Probenauswahl

Um zu ermöglichen, dass Ergebnisse laborübergreifend verglichen werden können, wurden Proben verwendet, die für alle Labore zugänglich sind (Publikation 1 [1]) bzw. Proben zwischen den Laboren geteilt (Publikation 2 [2]). Für Publikation 1 [1] wurden Proben von GEDNAP (German DNA Profiling - jährlich stattfindender Spuren-Ringversuch), standardisierte Referenzproben des NIST (*National Institute of Standards and Technology*, 2391c Komponenten A-F, 2372 Komponente A) sowie Kontroll-DNA Proben aus verschiedenen STR-Kits verwendet. In Publikation 2 [2] kamen wiederum NIST Referenzproben (2392c, Komponente A-C), GEDNAP-Proben sowie Kontroll-DNA Proben aus STR-Kits zum Einsatz. Zusätzlich wurden vier sehr alte DNA-Proben (*ancient DNA* - aDNA) analysiert, welche aus dem 5. bis 8. Jahrhundert stammen. Für diese Studie wurden am Institut für Gerichtliche Medizin der Medizinischen Universität Innsbruck (GMI) die Proben bis zu einem fertig sequenzierbaren Reagenz vorbereitet und anschließend an die acht teilnehmenden Labore zur Analyse versendet. Für Publikation 3 [3] wurden von 64 Vater-Sohn-Paaren Mundschleimhautabriebe genommen und daraus die DNA extrahiert.

4.3 STR-Typisierung mittels Kapillarelektrophorese (CE)

Das bis heute verwendete Standardverfahren zur Erstellung eines STR-Profiles ist die CE. Die Anreicherung der STR-Sequenzen erfolgt zunächst über eine Multiplex-PCR, bei der mehrere STR-Marker kombiniert werden.

In Publikation 1 [1] kamen verschiedene Kits für die Amplifikation der STRs zur Anwendung: *AmpFLSTR™ NGM Select™ PCR Amplification Kit* (Thermo Fisher Scientific), *AmpFLSTR™ NGM SElect™ Express Kit* (Thermo Fisher Scientific), *PowerPlex® ESX17* (Promega), *PowerPlex® 16 System* (Promega), *PowerPlex® 21 System* (Promega), *PowerPlex® Y23 System*

(Promega) sowie das *Investigator Argus X-12 Kit* (Qiagen, Hilden, Deutschland). In Publikation 3 [3] wurde hauptsächlich das *PowerPlex® Y23 System* (Promega) verwendet. Zur Analyse von schnell mutierenden Y-STRs (*rapidly mutating* oder RM Y-STRs) wurde ein modifiziertes *RM Yplex Assay* von Alghafri et al. [35] (siehe *Supplementary File S1* Publikation 3 [3]) angewandt, indem der hochpolymorphe Marker DYS464 statt des Markers DYS518 einbezogen wurde.

In der PCR werden die Amplifikate mit bis zu sechs verschiedenen Fluoreszenzfarbstoffen markiert [6, 7]. Im elektrophoretischen Schritt wird dann die Laufzeit jedes fluoreszenzmarkierten Fragments bis zur Laserdetektion gemessen und mittels Kennkurve zu einer Fragmentlänge korreliert. Die unterschiedliche Farbmarkierung ermöglicht, dass im Anschluss mehrere Marker derselben Größe im Elektropherogramm in verschiedenen Farbkanälen separiert werden können. Alle Amplifikate werden im Elektropherogramm als Peak dargestellt. Die Position des Peaks wird im Abgleich mit einer Allelleiter einem bestimmten Allel und somit der spezifischen Repeat-Anzahl zugeordnet. Die CE wurde in Publikation 1 [1] und Publikation 3 [3] am *Applied Biosystems 3130xl Genetic Analyzer* bzw. am *Applied Biosystems 3500 Genetic Analyzer* (Thermo Fisher Scientific) durchgeführt, in Publikation 2 [2] am *Applied Biosystems Prism 3500XL Genetic Analyzer* (Thermo Fisher Scientific).

4.4 STR-Typisierung mittels massively parallel sequencing (MPS)

Alle Sequenzierreaktionen für diese Arbeit wurden am *MiSeq FGx* (Illumina) durchgeführt. Beim *MiSeq FGx* System handelt es sich um eine spezialisierte Sequenzierplattform für die forensische Anwendung. Dem forensischen Anwender sind dabei alle Möglichkeiten der normalen *MiSeq* Plattform im Research Modus gegeben sowie die zusätzliche Anwendung des Forensik Modus, welcher nur mit der Nutzung des *ForenSeq DNA Signature Prep Kits* ausführbar ist. Die dazu gehörige UAS ermöglicht die Auswertung aller im Kit enthaltenen Marker.

Die eigentliche Sequenzierreaktion findet im *MiSeq* auf einem Glasobjektträger, der *Flow Cell*, in einer integrierten Kapillare statt. Um gewünschte DNA-Sequenzen auf dem *MiSeq* zu sequenzieren, müssen diese an die *Flow Cell* binden. Dazu ist zunächst die sogenannte *Library* Präparation nötig, welche in Publikation 1 [1], Publikation 2 [2] und Publikation 3 [3] mit Hilfe des *ForenSeq DNA Signature Prep Kits* durchgeführt wurde. Hierbei werden die Enden der Ziel-Sequenzen mit P5/P7-Adaptoren, *Indices* und Bindungsstellen für die Sequenzierprimer versehen, wonach dieser Komplex als *Library* bezeichnet wird. Die angehängten P5/P7-Adapter sind komplementär zu Oligomeren auf der *Flow Cell* und ermöglichen somit, dass die Sequenzen an die *Flow Cell* binden. *Indices* dienen der späteren bioinformatischen Zuordnung der Sequenzen zur jeweiligen Probe. Da für die Sequenzierung bis zu 96 Proben in einem *Pool* vereint werden, wird jede Probe mit einem spezifischen Indexpaar versehen. Das Anhängen der Sequenzierprimerbindestellen, *Indices* und Adapter erfolgt bei der *ForenSeq*-Chemie über das Verfahren der *nested PCR* mit zwei nacheinander laufenden PCRs. Anschließend ist eine

Aufreinigung nötig, bei der alle überschüssigen Primer, Primerdimere bzw. Adapterdimere und restlichen Reaktionskomponenten entfernt werden, sodass nur vollständige Libraries in der Lösung verbleiben. Die Aufreinigung erfolgt beim *ForenSeq DNA Signature Prep Kit* über magnetische *Beads*. Um eine ausgewogene Abdeckung aller Proben zu garantieren, werden alle Libraries normalisiert. Das heißt, die Konzentration aller *Libraries* wird auf den gleichen Level gebracht, was hier wiederum mit Hilfe der magnetischen *Beads* erreicht wird. Nun werden alle *Libraries* in einem *Pool* vereint, um sie anschließend in den *MiSeq FGx* zu laden. Die Probe wird dazu in die Sequenzier-Cartridge pipettiert, die alle weiteren benötigten Chemikalien enthält.

Auf der Flow Cell findet vor der eigentlichen Sequenzierung die *Cluster*-Generierung statt. Dabei wird jedes ursprüngliche PCR-Produkt bis zu 1000 Mal in einem *Cluster* mit einer Größe von 1 μm vervielfältigt. Zum einen wird durch diese Vervielfältigung ein ausreichend starkes Lichtsignal bei der Sequenzierung erzeugt, zum anderen werden dadurch Sequenzierfehler deutlich. Die anschließende Sequenzierung erfolgt zyklusweise nach dem *Sequencing-by-Synthesis* Prinzip [36]. Ein Zyklus der Sequenzierung besteht im Wesentlichen aus drei Schritten und benötigt fünf bis sechs Minuten: (i) Einbau eines Nukleotids, welches mit einem reversiblen Terminator und einem reversiblen Fluorophor versehen ist. (ii) Detektion der Basen-spezifischen Wellenlänge (530 & 660 nm) nach Anregung mit einem Diodenlaser durch zwei Digitalkameras. (iii) Waschen und Entfernen des reversiblen Terminators und des Fluorophors, sodass im nächsten Zyklus ein neues Nukleotid eingebaut werden kann. Pro Zyklus wird somit eine Base jedes *Clusters* sequenziert. Von einem *Cluster* werden alle Kopien zusammen als sogenannter *Read* detektiert. Es werden Millionen von *Reads* generiert, die alle ursprünglichen PCR-Fragmente repräsentieren. Die Entstehung dieser Millionen *Reads* verleiht der Methode den Namen *massively parallel sequencing*. Zusätzlich werden von jedem Cluster drei weitere *Reads* erzeugt, die für die bioinformatische Zuordnung der *Reads* zu einem Marker und zu einer Probe notwendig sind. So erhält man von jedem *Cluster* insgesamt vier *Reads*. *Read 1* stellt, wie bereits beschrieben, die gewünschte Zielsequenz dar. *Read 2* ist eine Sequenzierung derselben Sequenz in die entgegengesetzte Richtung, was als *paired-end* Sequenzierung bezeichnet wird. Diese Art der Sequenzierung ermöglicht durch eine Überlagerung von *Read 1* und *2*, Sequenzierfehler zu eliminieren oder längere Sequenzen (mit weniger Überlappung) zu sequenzieren. Zwei der vier *Reads* repräsentieren somit die Zielsequenz. Die beiden anderen *Reads* ergeben sich aus der Sequenzierung der beiden Indices. Über die Position auf der Flow Cell werden die vier *Reads* einander zugeordnet und anschließend in einer Datei zusammengeführt. Die Sequenzierchemie (*Reagent Cartridge*) bestimmt, wie viele Basen sequenziert werden können. Beim *ForenSeq DNA Signature Prep Kit* ist die Readlänge festgelegt: 351 bp werden von *Read 1*, 31 bp von *Read 2* und jeweils acht bp werden von beiden Indices sequenziert. Diese ungleiche Verteilung der Readlänge von *Read 1* und *2* stellt eine Ausnahme dar, ist aber wichtig, um die gesamte Länge eines STRs ohne Unterbrechung durch *Read 1* abzudecken und dadurch die korrekte

Anzahl der *Repeats* im STR wiederzugeben. Der kurze *Read 2* ist notwendig, um die Sequenzen einem STR-Marker eindeutig zuordnen zu können. Bei der Auswertung der Daten wird die Abdeckung (*coverage*) jeder Sequenz als Read-Anzahl angegeben. Die Auswertung ist für Publikation 1 [1] und Publikation 2 [2] mit Hilfe der UAS erfolgt, für Publikation 3 [3] mit Hilfe der im Rahmen des DNaseqEx Projekts entwickelten Software *wintermute* (open-source Software, verfügbar unter <https://github.com/545ch4/wintermute>, letzter Zugriff am 24.02.2021).

5 Ergebnisse

Publikation 1 [1] und Publikation 2 [2] behandeln die Validierung und Evaluierung des *ForenSeq DNA Signature Prep Kits* am *MiSeq FGx* System. Publikation 1 [1] konzentriert sich dabei auf die Validierung zwischen zwei Laboren mit den Schwerpunkten Konkordanz, Reproduzierbarkeit, Sensitivität und Mischspurenanalyse. Hierbei wurden alle Schritte von der Probenextraktion bis zur Sequenzierung am *MiSeq FGx* in jedem Labor selbstständig durchgeführt. Publikation 2 [2] evaluiert dieselben Schwerpunkte sowie die Analyse von aDNA zwischen acht teilnehmenden Laboren. Um die vielen Einflussfaktoren während der komplexen *Library* Präparation zu eliminieren und die Vergleichbarkeit des *MiSeq FGx* Systems deutlich zu machen, wurde hier nur die Sequenzierung an sich in jedem Labor selbstständig durchgeführt. Die *Library* wurde dazu im GMI für die Sequenzierung vorbereitet und dann an die sieben weiteren teilnehmenden Institute versendet. In Publikation 3 [3] wurden Y-STRs von 64 Vater-Sohn-Paaren analysiert und auf Sequenzvariationen sowohl der verwandten Personen als auch der nicht verwandten Personen untersucht.

5.1 Konkordanz und Reproduzierbarkeit

Die Konkordanz des *ForenSeq DNA Signature Prep Kits* wurde sowohl zum Standardverfahren CE, zu bekannten Referenzprofilen sowie zwischen den Resultaten der verschiedenen Labore geprüft (Publikation 1 [1] und 2 [2]).

In Publikation 1 [1] wurden MPS-Ergebnisse von 372 (GMI) / 288 (BER) aSTRs, 42 (GMI) / 126 (BER) X-STRs und 198 (GMI und BER) Y-STRs mit den CE-Ergebnissen verglichen. Alle Ergebnisse waren konkordant bis auf drei Unstimmigkeiten bei den Ergebnissen von BER: Hier wurden in drei Markern (vWA, DYS385ab, DXS10135) Stotterallele auf Grund stärkerer Amplifikation vom Filter der UAS nicht erkannt und somit fälschlicherweise als Allel detektiert (SF). Bei Stotterallelen handelt es sich um PCR-Produkte, die während der PCR durch Strangrutschen (*slippage*) der Polymerase entstehen und typischerweise eine Wiederholungseinheit kleiner sind als das echte Allel. Bei der Konkordanz-Prüfung der MPS-Ergebnisse zwischen den beiden teilnehmenden Instituten wurden 504 aSTRs, 126 Y-STRs und 240 X-STRs verglichen. Zusätzlich zu den bereits vorher beobachteten Stotterallelen wurde hier ein *drop-out* im Marker DXS10103 beobachtet (siehe Tabelle 3, *Supplementary data* Publikation 1 [1]).

In Publikation 2 [2] wurden MPS-Ergebnisse von 162 aSTRs, 120 Y-STRs und 42 X-STRs mit Referenzprofilen und MPS-Ergebnisse der acht teilnehmenden Institute miteinander verglichen. Alle aSTRs und Y-STRs waren vollständig konkordant. Bei den X-STRs wurden acht Diskrepanzen beobachtet, alle im Marker DXS10103: Dieser wies zwei Allel- und sechs Locus-Ausfälle (*drop-out* beider Allele eines Loci) auf. Stotterallele, welche über dem *threshold* der UAS liegen, wurden hier nicht beobachtet (siehe Tabelle S4 Publikation 2 [2]).

In Publikation 1 [1] wurde anhand der Analyse von duplizierten Proben zusätzlich die Reproduzierbarkeit in 1008 aSTRs, 252 X-STRs und 480 Y-STRs (von beiden Instituten) überprüft. Dabei waren die Ergebnisse in 30 Fällen nicht reproduzierbar: Stotterallele traten in fünf aSTRs, zwei X-STRs und neun Y-STRs auf. Auch die Problematik des Markers DXS10103 wurde hier erneut bestätigt, für welchen 13 *drop-outs* detektiert wurden. Ein weiterer *drop-out* wurde im Marker vWA beobachtet (siehe Tabelle 3, *Supplementary data* Publikation 1 [1]).

Im Vergleich mit den CE-Ergebnissen wurden unter den MPS-Ergebnissen 17 isometrische Heterozygoten (IHZ, Publikation 1 [1]) beobachtet. Eine IHZ ist ein heterozygoter STR-Marker, dessen Allele die gleiche Länge, aber eine unterschiedliche Sequenz aufweisen. Diese IHZs können mittels CE nicht nachgewiesen werden, weshalb MPS hier einen Informationsgewinn bringt. Dabei stellte sich heraus, dass einige dieser IHZs in der flankierenden Region liegen, was nicht deutlich in der UAS gekennzeichnet ist (siehe *Supplementary data* Publikation 1 [1]).

In Publikation 3 [3] wurden zur Untersuchung der Konkordanz die MPS-Ergebnisse von 64 Vater-Sohn-Paaren (87 Personen insgesamt) mit CE-Ergebnissen des *PowerPlex® Y23 Systems* verglichen. Zusätzlich wurden 13 schnell mutierende Y-STRs (*rapidly mutating STRs* – RM Y-STRs) mit einem modifizierten *RM Yplex-Assay* untersucht (Verwendung von DYS464 anstelle von DYS518, siehe *Supplementary File S1* Publikation 3 [3]), der von Alghafri et al. [35] entwickelt wurde. Dadurch konnten 22 Y-STRs überlappend mit beiden Analysemethoden (CE und MPS) analysiert werden (siehe *Supplementary File S2* Publikation 3 [3]). Insgesamt wurden 1.914 Y-STR-Maker verglichen, wovon 1.880 (98 %) konkordant waren. Diskrepanzen kamen auf Grund eines *drop-outs* im Marker DYS392, mehrerer überexprimierter Stotterallele (DYS 385ab, DYS392, DYS456, DYS576, DYF387S1 und DYS612) sowie weniger Sequenz-*drop-ins* (DYS437 und DY448) zustande. Bei Sequenz-*drop-ins* handelt es sich um zusätzliche Amplifikationen des gleichen numerischen Allels, welche sich nur anhand ihrer Sequenz unterscheiden. Diese wiesen alle eine Peak Height Ratio (PHR) zwischen 10 % und 13 % auf, weshalb sie von der Software *wintermute* als Allel bewertet wurden.

5.2 Coverage, inter- und intra-locus balance

Die abweichenden Clusterdichten bei Sequenzierung der gleichen *Library* an acht verschiedenen *MiSeq FGx* Systemen wiesen auf eine unterschiedliche Performance der Geräte hin: Im Durchschnitt lag die Clusterdichte bei 645 K/mm², das Minimum lag bei 432 K/mm², das Maximum bei 862 K/mm² (siehe Tabelle 1 Publikation 2 [2]). Weiterhin wurde geprüft, ob diese

Abweichungen der Clusterdichte sich auch auf die Locus-Coverage, die *heterozygote (intra-locus) balance* (HB) und somit die Dateninterpretation auswirken.

Da Spurenproben selten eindeutige STR-Profile einer einzelnen Person enthalten, ist es wichtig, dass sowohl die STR-Marker in einem Kit untereinander (*inter-locus balance*), als auch die beiden Allele eines heterozygoten Markers (HB) in ihrer *Coverage* gut ausbalanciert sind, um eine Interpretation der Ergebnisse zu ermöglichen. In Publikation 1 [1] wurde zur Beurteilung der *inter-locus balance* die Standardabweichung σ (gemäß [37, 38]) für jeden Locus jeder Probe in 68 MPS-Experimenten (34/58 STR-Marker) berechnet und mit bekannten CE-Werten verglichen [38]. Der hier beobachtete Medianwert betrug 49 % mit einer annähernd normalen Verteilung (siehe Abbildung 1 Publikation 1 [1]). Für CE-Experimente wurde ein Median von 28 % für 16 autosomale Loci ermittelt [38]. In Publikation 2 [2] wurde zur Abschätzung der *inter-locus balance* die *Coverage* der einzelnen STR-Loci betrachtet. In Publikation 2 [2] variierte die durchschnittliche *Coverage* pro Locus über alle Läufe von 49.387 *Reads* (TH01) zu 314 *Reads* (DXS10103, Tabelle S5 Publikation 2 [2]). Diese Zahlen zeigen ein deutliches Ungleichgewicht der *Inter-Locus balance*. Die abweichende Clusterdichte unter den Instituten spiegelte sich auch in Abweichungen der Locus-Coverage und der Sequenzier-Tiefe des gesamten Laufs (Gesamtanzahl der *Reads/Run*) wider: Die durchschnittliche Locus-Coverage reichte von 7.437 *Reads/Locus (Run 2)* bis zu 19.044 *Reads/Locus (Run 7)*, die durchschnittliche Sequenzier-Tiefe reichte von 438.786 *Reads/Run (Run 2)* bis zu 1.123.576 *Reads/Run (Run 7)*, siehe Tabellen 1 und S5 Publikation 2 [2]).

Zur Betrachtung der HB wurde in Publikation 1 [1] die *allele coverage difference* (ACD) berechnet. Eine ACD nahe 0 bedeutet, dass die beiden Allele eines heterozygoten Markers nahezu gleich stark abgedeckt sind. Eine ACD von 1 bedeutet, dass eines der beiden Allele ausgefallen ist. Die ACD wurde in dieser Studie für zwei Proben (NIST2372A und 2800M) für jeweils 6 unterschiedliche Verdünnungen, ausgehend vom Optimum 1000 pg bis hin zu 31 pg, berechnet (siehe Abbildung 3 Publikation 1 [1]). Wie zu erwarten, zeigte sich bei höherem DNA-Einsatz auch eine bessere HB. Bis zu einem DNA-Einsatz von 250 pg erzielten die meisten Marker eine ACD $< 0,3$. Dabei waren vWA, Amelogenin und D1S1656 die Marker mit der durchschnittlich höchsten ACD. D7S820, D8S1179 und TH01 zeigten mit einer durchschnittlichen geringen ACD eine ausgeglichene HB. In Publikation 2 [2] wurde die HB als *allele coverage ratio* (ACR) bei einem DNA-Einsatz von 1 ng berechnet, wobei hier umgekehrt zur ACD eine ACR nahe 0 auf ein *drop-out* hindeutet und eine ACR nahe 1 eine gute HB beschreibt. Hier wiesen alle Marker eine gute HB mit einer ACR $\geq 0,6$ (*threshold* der UAS) auf mit Ausnahme von D22S1045 (ACR = 0,43) und D5S818 (ACR = 0,57). Das bereits beschriebene Ungleichgewicht der *inter-locus balance* schien keinen großen Einfluss auf die HB zu haben: So wurde für TH01 die höchste Locus-*Coverage* und eine gute HB (ACR = 0,95) ermittelt, wobei D19S443 bei den meisten Läufen eine sehr niedrige *Coverage*, aber trotzdem eine gute HB (ACR = 0,82) zeigte (siehe Tabelle S6, Abbildung 1 Publikation 2 [2]). Die

beobachteten Varianzen der Locus-Coverage und der Clusterdichte zwischen den Instituten hatten nur eine geringe Auswirkung auf die HB (siehe Tabellen 1 und S6, Abbildung 1 Publikation 2 [2]).

5.3 Sensitivität

Der optimale DNA-Einsatz beim *ForenSeq DNA Signature Prep Kit* liegt bei 1 ng auf 5 ml. Spurenproben weisen oft einen niedrigeren DNA-Gehalt auf, weshalb die Sensitivität jedes forensischen STR-Kits systematisch validiert werden muss. Dazu wurden zwei unterschiedliche Proben in sechs Verdünnungsstufen (1000 pg, 500 pg, 250 pg, 125 pg, 63 pg, 31 pg) jeweils in Triplikaten und in zwei Sequenzierläufen pro Labor analysiert (siehe Tabelle 1 Publikation 1 [1]). Hierbei wurde sowohl die Ergebnisqualität (*drop-outs* und HB) in Abhängigkeit von der Verdünnung (1000 pg – 31 pg), als auch in Abhängigkeit von der Probenanzahl pro Sequenzierlauf (20 vs. 38 Proben) untersucht. Zusätzlich wurde überprüft, ob Unterschiede zwischen den beiden Proben auftreten und ob ein höherer finaler *Library*-Einsatz auf dem *MiSeq FGx* (10 µl BER vs. 7 µl GMI) die Sensitivität verbessern kann (Publikation 1 [1]). In Publikation 2 [2] wurde eine Mischprobe ebenfalls in 6 Verdünnungsstufen (1000 pg, 500 pg, 250 pg, 125 pg, 63 pg, 31 pg) jeweils in Duplikaten von allen acht teilnehmenden Instituten sequenziert (siehe Tabelle S1 Publikation 2 [2]) und auf *drop-outs* und Unterschiede zwischen den Instituten untersucht.

Abbildung 2 in Publikation 1 [1] zeigt die durchschnittliche Anzahl von Locus-Ausfällen bei 12 betrachteten ESS-Loci (12 europäische Standard-Marker, die bei der Erstellung von DNA-Profilen in Europa immer analysiert werden) bei den unterschiedlichen Verdünnungsstufen. Wie zu erwarten, stiegen die *drop-outs* mit abnehmendem DNA-Einsatz an. Dabei wurde der Einfluss der erhöhten Probenanzahl (38 vs. 20 Proben pro Sequenzierlauf) deutlich: Bei einem DNA-Einsatz von 63 pg und der gleichzeitigen Analyse von 38 Proben wurden hier nur 66,7 % der Allele (im Durchschnitt 8 von 12 Markern) detektiert, bei der gleichzeitigen Analyse von nur 20 Proben wurden bei gleichem DNA-Einsatz über 90 % der Allele (im Durchschnitt 11 von 12 Markern) detektiert. Außerdem wurden bei der gleichzeitigen Sequenzierung von 38 Proben erste *drop-outs* bereits bei einem DNA-Einsatz von 125 pg beobachtet (vs. 63 pg DNA-Einsatz bei 20 Proben). Zwischen der Probenart wurden keine Abweichungen deutlich. Im Gegensatz zu der erwarteten erhöhten Empfindlichkeit bei einem finalen *Library Input* von 10 µl (BER) statt 7 µl (GMI) zeigten sich mehr Locus-Ausfälle in BER als in GMI (Publikation 1 [1]). Hier wurde bereits eine geräteabhängige Varianz vermutet, die sich in Publikation 2 [2] bestätigte, wo *Run 2* deutlich schlechtere Ergebnisse im Vergleich mit den anderen Instituten (*Run 0-1, 3-8*) erzielte (siehe Abbildung 2, Tabelle S8 Publikation 2 [2]). Zusätzlich wurde in Publikation 1 [1] die HB über die verschiedenen Verdünnungsstufen betrachtet, was bereits im Abschnitt 5.2 beschrieben wurde.

In Publikation 2 [2] wurden 26 Proben pro Sequenzierlauf sequenziert und alle im *ForenSeq DNA Signature Prep Kit* enthaltenen Marker betrachtet. Für aSTRs wurde in Publikation 2 [2] ein erster *drop-out* bei 250 pg beobachtet. Bei Y-STRs erschienen erste *drop-outs* bereits bei 500 pg, bei X-STRs bereits bei 1000 pg DNA-Einsatz. Autosomale STRs wurden bis zu einem DNA-Einsatz von 62 pg in über 90 % korrekt typisiert, X- und Y-STRs bis zu einem DNA-Einsatz von 125 pg. Bei einem DNA-Einsatz von 31 pg wurden 71 % der aSTRs, 60,3 % der Y-STRs und 81,1 % der X-STRs fehlerfrei typisiert (siehe Tabelle S8, Abbildung S1 Publikation 2 [2]). Dabei zeigten sich zwischen den Instituten vergleichbare Sensitivitäts-Level mit Ausnahme von *Run 2*, wo deutlich höhere *drop-outs* sichtbar waren (siehe Tabelle S8, Abbildungen 2 und S1 Publikation 2 [2]).

5.4 Mischspuren

Wie bereits erwähnt, weisen Spurenproben selten eindeutige STR-Profile einer Einzelperson auf, sondern häufig Mischprofile mehrerer Personen. Zu diesem Zweck wurden in Publikation 1 [1] und Publikation 2 [2] verschiedene Mischungen getestet. Publikation 1 [1] untersuchte Mischungsverhältnisse von 1:1, 1:5, 1:10, 1:15 und 1:20, Publikation 2 [2] Mischungsverhältnisse von 83,3:16,7, 90,9:9,1, 93,7:6,3 und 95,2:4,8. Zum Zweck der besseren Vergleichbarkeit mit Publikation 1 [1] wird das Verhältnis 83,3:16,7 im Nachfolgenden als 1:5, 90,9:9,1 als 1:10, 93,7:6,3 als 1:15 und 95,2:4,8 als 1:20 beschrieben. In Publikation 1 [1] wurden dabei Mann-Frau und Mann-Mann Mischungen analysiert. Bei männlich/weiblichen DNA-Mischungen wurde jeweils die weibliche als auch die männliche Komponente im Überschuss untersucht. Publikation 2 [2] wählte Mann-Frau Mischungen, wobei die männliche Komponente immer im Überschuss eingesetzt wurde. Zusätzlich wurde in Publikation 1 [1] das Verhalten bei sogenannten Extremmischungen mit einem Verhältnis von 1:100, 1:500 und 1:1000 betrachtet, wobei die weibliche Komponente im Überschuss eingesetzt wurde.

Um zu beurteilen, wie gut die Mischungsverhältnisse in den MPS-Resultaten wiedergegeben werden, wurde das tatsächliche Verhältnis zwischen der *Read*-Anzahl von Minor- und Hauptkomponente der MPS-Ergebnisse berechnet und in Abbildung 4 (Publikation 1 [1]) für alle Marker in Bezug zum anfangs eingestellten Verhältnis abgebildet. Insgesamt wurden die eingestellten Verhältnisse gut reproduziert. Beim eingestellten Verhältnis 1:1 waren die Ergebnisse beider Institute gleichermaßen um das Zielverhältnis gestreut. Bei den eingestellten Verhältnissen 1:5, 1:10, 1:15 und 1:20 war die Abweichung vom Zielverhältnis am BER größer als am GMI. Abbildung 5 (Publikation 1 [1]) zeigt die beobachteten Mischungsverhältnisse über alle geeigneten Marker unter Verwendung von 1 ng und 2 ng DNA-Einsatz bei einem Zielverhältnis von 1:20. Die unterschiedlichen DNA-Mengen hatten dabei keinen Einfluss auf die beobachteten Mischungsverhältnisse.

In Publikation 2 [2] wurde zur Beurteilung der Mischungen berechnet, wieviel Prozent der Allele der Minor-Komponente korrekt typisiert wurde (siehe Abbildung 2, Tabelle 2 Publikation 2 [2]).

100 % der Minor-Komponente konnten beim Mischungsverhältnis von 1:5 korrekt typisiert werden, 94,4 % beim Verhältnis von 1:10, 76,7 % beim Verhältnis von 1:15 und 45 % beim Verhältnis von 1:20. In Tabelle S9 und Abbildung S2b sind ebenfalls die beobachteten Mischungsverhältnisse dargestellt (Publikation 2 [2]). Die beobachteten Mischungsverhältnisse liegen jeweils nah am eingestellten Mischungsverhältnis, mit einer durchgehend leichten Überexpression der Minor-Komponente. Abweichungen der beobachteten Mischungsverhältnisse zwischen den Instituten werden erst ab einem Mischungsverhältnis von 1:20 sichtbar.

Abbildung 6 (Publikation 1 [1]) zeigt wieviel Prozent der männlichen Allele (Minor-Komponente) bei den Extremmischungen (1:100, 1:500, 1:1000) mittels MPS und mittels CE detektiert werden konnten. Es wird deutlich, dass mittels MPS bei allen drei Mischungsverhältnissen prozentual weniger Allele typisiert werden konnten als mittels CE, wobei der Unterschied bei einem Verhältnis von 1:100 (33 % CE vs. ca. 14 % MPS) und 1:500 (19 % CE vs. ca. 3 % MPS) deutlich größer ist als bei 1:1000 (6% CE vs. ca. 3 % MPS). Zwischen 1 ng und 2 ng DNA-Input sowie im Vergleich zwischen den Instituten wurden keine wesentlichen Abweichungen beobachtet.

5.5 Alte DNA

Die Analyse von aDNA-Proben stellt den DNA-Analytiker vor verschiedene Herausforderungen: Auf Grund ihres Alters von mehreren tausend Jahren sind sie oft stark degradiert, haben einen niedrigen DNA-Gehalt und weisen möglicherweise Kontaminationen auf, die bereits vor der Arbeit im Labor entstanden sein können. Um die Robustheit und Leistung des *ForenSeq DNA Signature Prep Kits* an hochgradig degradierter DNA zu untersuchen, wurden in Publikation 2 [2] vier aDNA-Proben (FA10013B01A, FA10026B01A, FA10030T01A, FA10058T01B) aus dem Mittelalter analysiert. Da bei degradierten Proben durch die Analyse von *single nucleotide polymorphisms* (SNPs) häufig mehr Informationen gewonnen werden können als durch die Analyse von STRs, wurden für diese Proben zusätzlich die 94 im *ForenSeq DNA Signature Prep Kit* enthaltenen iSNPs ausgewertet. Wie zu erwarten, konnte im Vergleich zu Proben mit hoher Qualität, eine geringere Anzahl von Markern korrekt typisiert werden: Für Probe FA10013B01A wurden nur 2 iSNPs von 152 Markern (entspricht 1,3 %), für Probe FA10026B01A 136 von 151 Markern (90,1 %), für Probe FA10030T01A 116 von 128 Markern (90,6 %) und für Probe FA10058T01B 147 von 152 Markern (96,7 %) erfolgreich typisiert. Zusätzlich wurden die CE-basierten und MPS-basierten Ergebnisse auf ihre Beweiskraft verglichen. Dazu wurde die *likelihood ratio* (LR) berechnet. Diese ist ein Maß für die Beweiskraft bezüglich der Hypothese, dass zwei DNA-Profile vom selben Verdächtigen abgeleitet wurden. LRs wurden für CE-basierte Daten für die 15 aSTRs des *AmpFISTR NGM Select Kits* und für MPS-basierte Daten sowohl für die 27 aSTRs des *ForenSeq DNA Signature Prep Kits* als auch für eine Teilmenge von 49 iSNPs berechnet (siehe Tabellen S15-S17 Publikation 2 [2]). Auf Grund der höheren Anzahl gleichzeitig analysierter Marker und der erfolgreichen Typisierung der meisten von ihnen

zeigen MPS-basierte Ergebnisse deutlich höhere LR's als CE-basierte Ergebnisse. Jedoch fiel auf, dass bei der Analyse der Probe FA10013B01A mittels CE zumindest ein Teilprofil erstellt werden konnte, während die Analyse mittels MPS keine Ergebnisse brachte (siehe Tabellen S10-1S2 Publikation 2 [2]).

5.6 Mutationen in Vater-Sohn-Paaren

In Publikation 3 [3] wurden Y-STRs von 64 Vater-Sohn-Paaren sowohl mit CE als auch mit MPS auf das Vorhandensein von Mutationen zwischen Vater und Sohn untersucht. Insgesamt wurden unter Verwendung des *PowerPlex® Y23 Systems* (CE-Analyse) unter 22 Y-STR-Markern 13 Mutationen beobachtet (siehe *Supplementary File S2* Publikation 3 [3]). Alle Mutationen waren entweder auf den Verlust oder den Gewinn einer Wiederholungseinheit zurückzuführen mit Ausnahme des Mutationsereignisses bei Y-GATA-H4, das durch eine Duplikation verursacht wurde.

Durch die MPS-Analyse konnten 12 der 13 beobachteten Mutationen bestätigt werden (siehe *Supplementary File S2* Publikation 3 [3]). Die Mutation im Marker DYS458 konnte nicht bestätigt werden, da dieser Marker nicht im *ForenSeq DNA Signature Prep Kit* enthalten ist. Dafür wurde eine zusätzliche Mutation im Marker DYS612 beobachtet, welcher nicht im *PowerPlex® Y23 Systems* enthalten ist. Alle Mutationen waren wiederum längenbasiert (mit Ausnahme der bereits erwähnten Duplikation von Y-GATA-H4), zusätzliche sequenzbasierte Mutationen wurden nicht gefunden. Jedoch ermöglichte die Sequenzierung die genaue Lokalisation der Längenvariation. So wurde festgestellt, dass alle Mutationen jeweils in der längsten ununterbrochenen Repeat-Einheit des STRs auftraten, welche aus der Literatur als *LUS* (*longest uninterrupted repeat sequence*) bekannt ist [39].

Bei der Analyse derselben Proben mit dem modifizierten *RM Yplex-Assay* (13 Y-STR-Marker mit bis zu 24 Y-STR-Allelen) wurden acht Mutationsereignisse beobachtet (siehe *Supplementary File S2* Publikation 3 [3]). Hier konnte auch die Mutation im Marker DYS612 bestätigt werden, welche mit der MPS-Analyse gefunden wurde. Alle verbleibenden sieben Mutationen traten in Markern auf, die weder im *PowerPlex® Y23 System* noch im *ForenSeq DNA Signature Prep Kit* enthalten sind. Jeder dieser Marker zählt zu den schnell mutierenden Y-STRs (RM Y-STRs) mit einer durchschnittlichen Mutationsrate von 10^{-2} pro Locus pro Generation [40], zu den *multicopy* Markern (besitzen auf Grund der palindromischen Struktur des Y-Chromosoms mehrere Kopien) sowie zu den komplexen STRs (siehe *Supplementary File S2* Publikation 3 [3]). Wieder waren alle Mutationsereignisse auf den Verlust oder den Gewinn einer Wiederholungseinheit zurückzuführen.

5.7 Y-STR Allel-Diversität

In Publikation 3 [3] wurde zusätzlich untersucht, ob die Sequenzinformationen eine höhere Allel-Diversität zeigen als die Längeninformationen der CE. Hierfür wurden alle 29 Familien

zusammen betrachtet (siehe *Supplementary File S2* Publikation 3 [3]). Mit Hilfe der CE-Analyse wurden 112 Allele detektiert. Mittels MPS-Analyse wurden 151 Allele detektiert, wovon 128 Allele Längenvarianten und 23 Allele Sequenzvarianten waren, sogenannte isometrische Allele (siehe Abbildung 1 Publikation 3 [3]). Isometrische Allele wurden in den Markern DYF387S1, DYS389II, DYS448, DYS481, DYS570, DYS612 und DYS635 gefunden. Die meisten dieser Marker sind durch eine komplexe Repeat-Struktur gekennzeichnet (DYF387S1, DYS389II, DYS448, DYS612, DYS637), zwei von ihnen durch eine einfache Repeat-Struktur (DYS481, DYS570).

6 Diskussion

Die Ergebnisse der drei Publikationen zeigen, dass MPS konkordante und reproduzierbare Ergebnisse erstellen kann. Das *ForenSeq DNA Signature Prep Kit* lieferte größtenteils konkordante Ergebnisse, sowohl zu den CE-Ergebnissen als auch im Vergleich der Ergebnisse zwischen unterschiedlichen Instituten. Da es sich beim *ForenSeq DNA Signature Prep Kit* um ein kommerzielles Produkt handelt, kam es bereits in verschiedenen Studien zur Anwendung [15, 21-27], die ebenfalls die Konkordanz zur CE bestätigten. Ausnahmen sind unter anderem überexprimierte Stotterallele, wobei dieses Problem auch bei der CE zu beobachten ist [41]. Jedoch wurde in der UAS für manche STRs ein *Stotter-threshold* von bis zu 50 % gewählt, während bei CE-Analysen ein *threshold* von höchstens 15 % verwendet wird [4]. Die leicht erhöhten Stotterallele der MPS-Ergebnisse lassen sich durch die zusätzlichen PCR-Zyklen, die für diese Methode notwendig sind, erklären. Weitere Diskordanzen waren auf Allel- oder Locus-Ausfälle (*drop-out* beider Allele eines Loci) zurückzuführen, welche hauptsächlich den Marker DXS10103 betrafen. Die Problematik des Markers DXS10103 wurde bereits in vorherigen Studien [15, 21-23, 25, 26] beschrieben, was zeigt, dass dieser Marker für den Einsatz in der Forensik noch optimiert werden muss. Die beobachteten Sequenz-*drop-ins* in Publikation 3 [3] sind auf die dort verwendete Software *wintermute* zurückzuführen. Dieselben *drop-ins* wurden auch durch Auswertung mit der UAS beobachtet, von dieser aber nicht als Allel bewertet.

Die für die forensische Fallarbeit notwendige Konkordanz der Ergebnisse wurde bestätigt. Ist die Qualität der MPS-Ergebnisse aber auch mit den Ergebnissen der Standardmethode CE vergleichbar? Das scheint noch nicht der Fall zu sein, denn selbst bei standardisierten Präparationsbedingungen (1 ng DNA-Input, 32 Fallarbeitsproben in einem Pool, bis zu 96 Referenzproben in einem Pool) wurden in Publikation 1 [1] und Publikation 2 [2] *drop-outs*, überexprimierte Stotter und schwach balancierte Marker beobachtet. Außerdem wurden in Publikation 2 [2] abweichende Leistungen der unterschiedlichen *MiSeq FGx*-Geräte der verschiedenen Institute aufgedeckt. Abweichungen zeigten sich vor allem in der Clusterdichte, der Locus-Coverage, der Sequenzier-Tiefe sowie bei der Sensitivität. Da durch die identische Vorbereitung der Reagenzien durch das GMI Laborablaufunterschiede weitestgehend minimiert

wurden, kommen als Verursacher der Abweichungen vor allem unterschiedliche Geräteleistungen in Betracht.

Im Vergleich mit aus der Literatur bekannten CE-Ergebnissen zeigte sich, dass MPS-Ergebnisse eine deutlich schlechtere *inter-locus balance* aufweisen (Publikation 1 [1]). Dieses Ungleichgewicht wurde bereits in früheren Studien beobachtet [21, 22]. Vermutlich haben auch hier die beiden PCRs der *Library* Präparation (anstatt einer PCR bei der CE-Analyse) einen negativen Einfluss auf die Balance der Marker. Ähnliches folgern auch England et al. [18], die die PCR-Produkte des *AmpFI STR Identifiler* und des *PowerPlex Y Kits* von 12 Proben mittels MPS sequenzierten und mit den CE-Ergebnissen verglichen. Des Weiteren kann sowohl Länge als auch Zusammensetzung der Erkennungssequenz durch verschieden hohe Fehlerraten zu deutlichen Unterschieden in der Detektierbarkeit valider *Reads* führen – sowohl für die Cluster-Bildung als auch während der bioinformatischen Analyse der Rohdaten.

Die bereits beschriebenen Varianzen zeigten jedoch kaum eine Auswirkung auf die HB (Publikation 2 [2]). In Publikation 1 [1] wurde bis zu einem DNA-Einsatz von 250 pg für die meisten Marker eine ausgeglichene HB ermittelt. Publikation 2 [2] bestätigt diese Ergebnisse, wobei hier eine gute HB bis zu einer $ACR \geq 0,6$ beschrieben wurde, was dem gewählten *threshold* der Firma Illumina in der UAS entspricht. Auch wenn die meisten Marker deutlich bessere Balancen vorweisen, wurde dieser niedrige *threshold* vermutlich auf Grund einiger weniger Marker (vor allem D5S818 und D22S1045) gewählt, die häufiger eine schwache Balance aufweisen, was sich auch in anderen Studien zeigte [15, 19, 21, 22, 24, 26, 42, 43].

In Publikation 2 [2] schien das Sensitivitäts-Level zwischen den Instituten bis zu einem DNA-Einsatz von 250 pg ähnlich zu sein. Ab 125 pg DNA-Einsatz wurden jedoch Unterschiede zwischen den Instituten festgestellt, wobei vor allem für *Run 2* eine deutlich schlechtere Sensitivität aufgedeckt wurde. Auch die Ergebnisse von Publikation 1 [1] weisen auf geringe Unterschiede zwischen den beiden Instituten hin. In Publikation 1 [1] wurde außerdem gezeigt, wie anfällig MPS-Analysen auf die Sequenzierung einer höherer Probenanzahl reagieren. Das erklärt sich aus dem begrenzten Platz für die *Reads* auf der Flow Cell. Werden mehr Proben sequenziert, verbleibt im Durchschnitt weniger Raum und damit weniger *Reads* pro Probe, wodurch die Sensitivität sinkt. Dieser Punkt sollte in der Kostenkalkulation beachtet werden, da weniger Proben pro Lauf natürlich höhere Kosten pro Probe bedeuten. Die Ergebnisse aus Publikation 1 [1] zeigen, dass LDOs bei einem DNA-Einsatz von 63 pg und darunter zu erwarten sind (beim gleichzeitigen Sequenzieren von 20 Proben), was in guter Übereinstimmung mit der CE-Analyse steht [44]. In Publikation 2 [2] zeigten sich erste *drop-outs* jedoch schon bei 1000 pg DNA-Einsatz in den X-STRs. Dabei muss beachtet werden, dass in Publikation 2 [2] alle im *ForenSeq DNA Signature Prep Kit* enthaltenen Marker in die Sensitivitätsanalyse einbezogen wurden, während in Publikation 1 [1] nur ESS-Marker, nur der Durchschnitt von drei Replikaten (somit nicht die absolute Anzahl von *drop-outs*) und nur Locus-Ausfälle (*drop-out* beider Allele eines Loci, nicht *Allel-drop-outs*) betrachtet wurden. Deshalb muss davon ausgegangen werden,

dass die Sensitivität nicht der von CE-Kits entspricht, sobald man alle verfügbaren Marker einbezieht. Bei dem Verzicht auf festgelegte Marker ergibt sich nichtsdestotrotz eine höhere Aussagekraft der MPS-Ergebnisse aufgrund der sehr viel höheren Anzahl von STRs gegenüber der CE. So wurde gezeigt, dass bei einem DNA-Einsatz von 125 pg noch mehr als 90 % der Allele detektiert wurden (Publikation 2 [2]). Diese Ergebnisse sind vergleichbar mit anderen Studien zur Sensitivität des *ForenSeq Signature Prep Kits* [21, 22, 25, 26] sowie mit anderen Studien zur Sensitivität anderer STR-MPS-Kits [30, 45]. Kann man hingegen nicht auf konkrete STRs verzichten (beispielsweise beim Abgleich von Profilen aus der DNA-Analyse-Datei des Bundeskriminalamts oder von bereits analysierten Spurenproben) ist die höhere Ausfallwahrscheinlichkeit der MPS-Technologie möglicherweise kritisch. Deshalb wurde in Publikation 1 [1] die Sensitivität ausschließlich von ESS-Markern betrachtet. Hierfür wurden bei einem DNA-Einsatz von 125 pg ebenfalls mehr als 90 % der Allele detektiert. Erst bei einem DNA-Einsatz von 63 pg und der gleichzeitigen Analyse von 38 Proben wurde eine hohe Ausfallrate von 33,3 % (im Durchschnitt vier von 12 Markern) beobachtet. Bei der gleichzeitigen Analyse von nur 20 Proben verbesserte sich diese Ausfallrate jedoch auf weniger als 10 % (im Durchschnitt einer von 12 Markern).

Die eingestellten Mischungsverhältnisse waren bis zu einem Verhältnis von 1:20 gut reproduzierbar (Publikation 1 [1] und 2 [2]). In beiden Laboratorien konnte die Nebenkompente in allen Mann-Frau und Mann-Mann Mischungen bei Verhältnissen bis zu 1:20 nachgewiesen werden, was die Ergebnisse früherer Veröffentlichungen bestätigt [21, 23, 25, 26]. Publikation 1 [1] zeigt, dass bei Mischungen von 1:100 (33 % CE vs. ca. 14 % MPS) und 1:500 (19 % CE vs. ca. 3 % MPS) mittels CE prozentual mehr Allele der männlichen Minor-Kompente detektiert wurden. Dieser Informationsvorteil der CE liegt in der hier fehlenden Primer-Konkurrenz zwischen männlicher und weiblicher DNA begründet. So kombiniert das *ForenSeq DNA Signature Prep Kit* verschiedene Marker in einer großen Multiplex-PCR, während aSTRs und Y-STRs beim CE-Ansatz separat amplifiziert werden [46, 47]. Um Y-Profile auch in unbalancierten Mischungsverhältnissen mittels MPS zu generieren, benötigt man auch hier einen separaten Ansatz von autosomalen und Y-chromosomalen STRs. Diese Möglichkeit liefert z. B. das *PowerSeqTM System* [34].

Im Zwischenfazit muss festgestellt werden, dass die Qualität der MPS-Technologie für die Auswertung von forensisch relevanten STRs geringer ist als die der konventionellen CE-Methode. Es stellt sich nun die Frage, ob MPS trotzdem einen Vorteil gegenüber der Standardmethode CE ausweist und ob damit eventuell die Qualitätseinbußen kompensiert werden können.

Einer der großen Vorteile der MPS-Technologie ist die gleichzeitige Analyse von vielen (hunderte bis tausende) Markern gegenüber der limitierten Anzahl der CE-Technologie (Publikation 4 [5]). Dementsprechend zeigte sich auch bei der Auswertung der vier aDNA-Proben in Publikation 2 [2] sowie in anderen Studien [48-50], dass MPS einen deutlichen

Informationsgewinn gegenüber der CE-Analyse ermöglichen kann. Dieser Informationsgewinn beruht nicht nur auf der stark erhöhten Anzahl gleichzeitig analysierter Marker der MPS-Analyse, sondern auch auf der erfolgreichen Typisierung der meisten dieser Marker ($\geq 87\%$ der aSTRs, $\geq 78\%$ der Y-STRs, $\geq 68\%$ der X-STRs, und $\geq 92\%$ der iSNPs). Daher wurde die Beweiskraft eines DNA-Profiles für drei der analysierten aDNA-Proben mittels MPS gegenüber den CE-Analysen deutlich erhöht (siehe Tabellen S15-S17, Abbildung S5 Publikation 2 [2]). Eine Ausnahme bildete die Referenz-Probe FA10013B01A, bei welcher stochastische Effekte auf Grund des geringen DNA-Gehalts oder eine mögliche Inhibierung des MPS-Ansatzes die Ursache für das suboptimale MPS-Ergebnis sind. Inhibierungsprobleme bei der MPS-Analyse von aDNA-Proben wurde bereits in früheren Studien beobachtet [51, 52].

Ein weiterer Zugewinn der MPS-Technologie im Vergleich zur CE-Technologie besteht in der Verfügbarkeit der Sequenzinformation (Publikation 4 [5]). In Publikation 1 [1] wurden 17 isometrische Heterozygote in 37 Proben detektiert. Diese isometrischen Heterozygoten erhöhen zusätzlich die Diversität und somit die Beweiskraft eines DNA-Profiles. In Publikation 3 [3] wurde durch die Sequenzinformation eine Zunahme der Allel-Diversität für sieben Y-STR-Marker festgestellt. Dieser Zugewinn ist jedoch für die einfache Spur-Personen Bewertung zu vernachlässigen, da mit der CE-Analyse von nur 12 ESS-Markern bereits eine statistische Auftrittswahrscheinlichkeit in der europäischen Referenzpopulation von 30 Milliarden erreicht wird [53]. Für die Zuordnung der einzelnen Verursacher einer komplexen Mischspur [54] und bei Untersuchung komplexer Verwandtschaftsstrukturen bietet eine höhere Diversität jedoch einen erheblichen Vorteil. So könnte sie bei der Analyse von Y-chromosomalen Markern einer paternalen Linie dienlich sein, da durch die fehlende Rekombination des Y-Chromosoms die Allel-Diversität bei den Y-STR-Markern gegenüber den autosomalen STR-Markern deutlich geringer ausfällt. Die Sequenzierung von 64 Vater-Sohn-Paaren lieferte jedoch keine weiteren Unterscheidungsmöglichkeiten zwischen den männlichen Verwandten im Vergleich zu den CE-Profilen (Publikation 3 [3]). Die erhöhte Allel-Diversität dient somit nur der weiteren Unterscheidung nicht verwandter Männer. Dieses Ergebnis war jedoch zu erwarten, da ein gewöhnlicher Basenaustausch, der eine Unterscheidung von Vater und Sohn möglich macht, mit einer Häufigkeit von ca. 1 zu 1 Million auftritt [55]. Meine Untersuchungen bestätigen die von Huszar et al. [54] geäußerte Schlussfolgerung, dass die Identifizierung von männlichen Spurenlegern mit Y-STRs durch die Anwendung von forensischen MPS-Ansätzen nicht wesentlich verbessert werden kann. Durch die CE-Analyse von schnell mutierenden Y-STRs der gleichen Proben wurden jedoch acht zusätzliche Mutationen detektiert (Publikation 3 [3]). Für die Differenzierung von Vater-Sohn-Paaren scheint daher die zusätzliche Analyse von schnell mutierenden Y-STRs mittels CE ein vielversprechenderer Ansatz zu sein. Erst kürzlich haben Ralf et al. [56] 12 neue schnell mutierende Y-STRs mit einer mittleren Mutationsrate von $2,6 \times 10^{-2}$ vorgestellt, die eine hohe Unterscheidungskraft demonstrieren. Jedoch ermöglichte die Sequenzierung es, die genaue Lokalisation der Längenvariation zu ermitteln. So bestätigen

unsere Ergebnisse (Publikation 3 [3]) frühere Beobachtungen, dass Tetranukleotid-*Repeats* anfälliger für Mutationen sind als andere repetitive Motive [40, 56, 57], Mutationsereignisse meist auf den LUS eines zusammengesetzten/komplexen STR-Markers beschränkt sind [29, 39, 40, 57], die LUS-Länge mit der Mutationsfähigkeit eines STR-Markers korreliert [40, 56-59] und dass isometrische Sequenzvarianten meist an komplexen STR-Markern auftreten [9, 10, 54]. Diese Informationen können helfen, Mutationsereignisse besser zu verstehen und dadurch tiefgehende Stammbaum- und Vaterschaftsanalysen unter Verwendung von Y-STRs zu verbessern.

Zusammenfassend kann festgestellt werden, dass die MPS-Methode konkordante und reproduzierbare Ergebnisse liefert, wobei die Qualität nicht ganz derjenigen von CE-Ergebnissen entspricht. Davon betroffen sind vor allem die *inter-locus balance*, aber auch höher gewählte *thresholds* bei der HB und den Stotterallelen. Allel- und Locus-*drop-outs* können sogar bei optimalen Probenmaterial und Präparationsbedingungen auftreten. Außerdem konnten unterschiedliche Geräteleistungen festgestellt werden, weshalb eine laborinterne Validierung immer notwendig ist. Der Vorteil der MPS-Technologie liegt eindeutig im Informationsgewinn, der durch die Sequenzinformation und die deutlich erhöhte Anzahl gleichzeitig analysierter Marker aus einer beweisereheblichen DNA-Probe generiert wird. Diese Vorteile sind vor allem für spezifische Fragestellungen relevant, wie z. B. für die Untersuchung und Zuordnung komplexer Mischspuren oder die Analyse von aDNA. Vielversprechend scheint außerdem die MPS-Analyse von SNPs oder der gesamtgenomischen DNA-Sequenz von Mitochondrien zu sein, um weitere forensische Fragestellungen z. B. im Rahmen der sog. erweiterten DNA-Analyse zu klären (Publikation 4 [5]). Für die statistisch eindeutige Zuordnung einer Einzelspur zu einer Vergleichsperson ist die MPS-Analyse jedoch nicht notwendig.

7 Referenzen

1. **Köcher S**, Müller P, Berger P, Bodner M, Parson W, Roewer L, Willuweit S, Consortium TD, Inter-laboratory validation study of the ForenSeq DNA Signature Prep Kit. *Forensic Sci Int Genet*, 2018. 36: p. 77-85.
2. Müller P, Sell C, Hadrys T, Hedman J, **Bredemeyer S**, Laurent F-X, Roewer L, Achtruth S, Sidstedt M, Sijen T, Trimborn M, Weiler N, Willuweit S, Bastisch I, Parson W, SeqForSTR-Consortium t, Inter-laboratory study on standardized MPS libraries: evaluation of performance, concordance, and sensitivity using mixtures and degraded DNA. *International Journal of Legal Medicine*, 2019.
3. **Bredemeyer S**, Roewer L, Willuweit S, Next generation sequencing of Y-STRs in father-son pairs and comparison with traditional capillary electrophoresis. *Forensic Sciences Research*, 2021: p. 1-6.
4. Butler JM, *Fundamentals of Forensic DNA Typing*. 2009: Academic Press Inc. 520.
5. Alonso A, Barrio PA, Muller P, **Köcher S**, Berger B, Martin P, Bodner M, Willuweit S, Parson W, Roewer L, Budowle B, Current state-of-art of STR sequencing in forensic genetics. *Electrophoresis*, 2018. 0: p. 1-14.
6. Martin P, de Simon LF, Luque G, Farfan MJ, Alonso A, Improving DNA data exchange: validation studies on a single 6 dye STR kit with 24 loci. *Forensic Sci Int Genet*, 2014. 13: p. 68-78.
7. Wang DY, Gopinath S, Lagace RE, Norona W, Hennessy LK, Short ML, Mulero JJ, Developmental validation of the GlobalFiler® Express PCR Amplification Kit: A 6-dye multiplex assay for the direct amplification of reference samples. *Forensic Sci Int Genet*, 2015. 19: p. 148-155.
8. Butler JM and Hill CR, *Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis*. *Forensic Sci Rev*, 2012. 24(1): p. 15-26.

9. Gettings KB, Kiesler KM, Faith SA, Montano E, Baker CH, Young BA, Guerrieri RA, Vallone PM, Sequence variation of 22 autosomal STR loci detected by next generation sequencing. *Forensic Sci Int Genet*, 2016. 21: p. 15-21.
10. Novroski NM, King JL, Churchill JD, Seah LH, Budowle B, Characterization of genetic sequence variation of 58 STR loci in four major population groups. *Forensic Sci Int Genet*, 2016. 25: p. 214-226.
11. Churchill JD, Chang J, Ge J, Rajagopalan N, Wootton SC, Chang CW, Lagace R, Liao W, King JL, Budowle B, Blind study evaluation illustrates utility of the Ion PGM system for use in human identity DNA typing. *Croat Med J*, 2015. 56(3): p. 218-229.
12. Wendt FR, Churchill JD, Novroski NM, King JL, Ng J, Oldt RF, McCulloh KL, Weise JA, Smith DG, Kanthaswamy S, Budowle B, Genetic analysis of the Yavapai Native Americans from West-Central Arizona using the Illumina MiSeq FGx forensic genomics system. *Forensic Sci Int Genet*, 2016. 24: p. 18-23.
13. Fordyce SL, Avila-Arcos MC, Rockenbauer E, Borsting C, Frank-Hansen R, Petersen FT, Willerslev E, Hansen AJ, Morling N, Gilbert MT, High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. *Biotechniques*, 2011. 51(2): p. 127-133.
14. Scheible M, Loreille O, Just R, Irwin J, Short tandem repeat typing on the 454 platform: strategies and considerations for targeted sequencing of common forensic markers. *Forensic Sci Int Genet*, 2014. 12: p. 107-119.
15. Almalki N, Chow HY, Sharma V, Hart K, Siegel D, Wurmbach E, Systematic assessment of the performance of illumina's MiSeq FGx forensic genomics system. *Electrophoresis*, 2017. 38(6): p. 846-854.
16. Zhao X, Ma K, Li H, Cao Y, Liu W, Zhou H, Ping Y, Multiplex Y-STRs analysis using the ion torrent personal genome machine (PGM). *Forensic Sci Int Genet*, 2015. 19: p. 192-196.
17. Gettings KB, Aponte RA, Vallone PM, Butler JM, STR allele sequence variation: Current knowledge and future issues. *Forensic Sci Int Genet*, 2015. 18: p. 118-130.
18. England R, Curnow N, Liu A, Stacey J, Harbison S, Massively parallel sequencing of Identifiler and PowerPlex ® Y amplified forensic samples. *Forensic Sci Int Genet Supplement Series*, 2015. 5: p. e210-e212.
19. Devesse L, Ballard D, Davenport L, Riethorst I, Mason-Buck G, Syndercombe Court D, Concordance of the ForenSeq system and characterisation of sequence-specific autosomal STR alleles across two major population groups. *Forensic Sci Int Genet*, 2018. 34: p. 57-61.
20. Verogen, ForenSeq™ DNA Signature Prep Reference Guide, Available from: <https://verogen.com/wp-content/uploads/2020/08/forenseq-dna-signature-prep-reference-guide-VD2018005-c.pdf>, 2020, (Last Access: 27.02.2021)
21. Churchill JD, Schmedes SE, King JL, Budowle B, Evaluation of the Illumina® Beta Version ForenSeq DNA Signature Prep Kit for use in genetic profiling. *Forensic Sci Int Genet*, 2016. 20: p. 20-29.
22. Silvia AL, Shugarts N, Smith J, A preliminary assessment of the ForenSeq FGx System: next generation sequencing of an STR and SNP multiplex. *Int J Legal Med*, 2017. 131(1): p. 73-86.
23. Xavier C and Parson W, Evaluation of the Illumina ForenSeq DNA Signature Prep Kit - MPS forensic application for the MiSeq FGx benchtop sequencer. *Forensic Sci Int Genet*, 2017. 28: p. 188-194.
24. Just RS, Moreno LI, Smerick JB, Irwin JA, Performance and concordance of the ForenSeq system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens. *Forensic Sci Int Genet*, 2017. 28: p. 1-9.
25. Jager AC, Alvarez ML, Davis CP, Guzman E, Han Y, Way L, Walichiewicz P, Silva D, Pham N, Caves G, Bruand J, Schlesinger F, Pond SJ, Varlaro J, Stephens KM, Holt CL, Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. *Forensic Sci Int Genet*, 2017. 28: p. 52-70.
26. Guo F, Yu J, Zhang L, Li J, Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq DNA Signature Prep Kit on the MiSeq FGx Forensic Genomics System. *Forensic Sci Int Genet*, 2017. 31: p. 135-148.
27. Sharma V, Chow HY, Siegel D, Wurmbach E, Qualitative and quantitative assessment of Illumina's forensic STR and SNP kits on MiSeq FGx. *PLoS One*, 2017. 12(11): p. e0187932.
28. Zeng X, King J, Hermanson S, Patel J, Storts DR, Budowle B, An evaluation of the PowerSeq Auto System: A multiplex short tandem repeat marker kit compatible with massively parallel sequencing. *Forensic Sci Int Genet*, 2015. 19: p. 172-179.
29. van der Gaag KJ, de Leeuw RH, Hoogenboom J, Patel J, Storts DR, Laros JFJ, de Knijff P, Massively parallel sequencing of short tandem repeats-Population data and mixture analysis results for the PowerSeq system. *Forensic Sci Int Genet*, 2016. 24: p. 86-96.

30. Wang Z, Zhou D, Wang H, Jia Z, Liu J, Qian X, Li C, Hou Y, Massively parallel sequencing of 32 forensic markers using the Precision ID GlobalFiler NGS STR Panel and the Ion PGM System. *Forensic Sci Int Genet*, 2017. 31: p. 126-134.
31. Meiklejohn KA and Robertson JM, Evaluation of the Precision ID Identity Panel for the Ion Torrent™ PGM™ sequencer. *Forensic Sci Int Genet*, 2017. 31: p. 48-56.
32. Purps J, Geppert M, Nagy M, Roewer L, Validation of a combined autosomal/Y-chromosomal STR approach for analyzing typical biological stains in sexual-assault cases. *Forensic Sci Int Genet*, 2015. 19: p. 238-242.
33. Thermo Fisher Scientific, Precision ID GlobalFiler™ NGS STR Panel v2 with the HID Ion S5™/HID Ion GeneStudio™ S5 System Application Guide, Available from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0016129_PrecisionIDSTRIonS5_UG.pdf, 2018, (Last Access: 27.02.2021)
34. Faith S and Scheible M, Analyzing Data from Next Generation Sequencers Using the PowerSeq® Auto/Mito/Y System. , Available from: <https://www.promega.de/resources/profiles-in-dna/2016/analyzing-data-from-next-generation-sequencers-using-the-powerseq-automitoy-system/> 2016, (Last Access: 27.02.2021)
35. Alghafri R, Goodwin W, Ralf A, Kayser M, Hadi S, A novel multiplex assay for simultaneously analysing 13 rapidly mutating Y-STRs. *Forensic Sci Int Genet*, 2015. 17: p. 91-98.
36. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara ECM, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovskiy Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurler ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R and Smith AJ, Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 2008. 456(7218): p. 53-59.
37. Gill P, Sparkes R, Fereday L, Werrett DJ, Report of the European Network of Forensic Science Institutes (ENSFI): formulation and testing of principles to evaluate STR multiplexes. *Forensic Sci Int*, 2000. 108(1): p. 1-29.
38. Debernardi A, Suzanne E, Formant A, Pene L, Dufour AB, Lobry JR, One year variability of peak heights, heterozygous balance and inter-locus balance for the DNA positive control of AmpFISTR® Identifier® STR kit. *Forensic Sci Int Genet*, 2011. 5(1): p. 43-49.
39. Just RS and Irwin JA, Use of the LUS in sequence allele designations to facilitate probabilistic genotyping of NGS-based STR typing results. *Forensic Sci Int Genet*, 2018. 34: p. 197-205.
40. Ballantyne KN, Goedbloed M, Fang R, Schaap O, Lao O, Wollstein A, Choi Y, van Duijn K, Vermeulen M, Brauer S, Decorte R, Poetsch M, von Wurmb-Schwark N, de Knijff P, Labuda D, Vezina H, Knoblauch H, Lessig R, Roewer L, Ploski R, Dobosz T, Henke L, Henke J, Furtado MR, Kayser M, Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am J Hum Genet*, 2010. 87(3): p. 341-353.
41. Chen M, Zhang J, Zhao J, Chen T, Liu Z, Cheng F, Fan Q, Yan J, Comparison of CE- and MPS-based analyses of forensic markers in a single cell after whole genome amplification. *Forensic Sci Int Genet*, 2020. 45: p. 102211.
42. Churchill JD, Novroski NMM, King JL, Seah LH, Budowle B, Population and performance analyses of four major populations with Illumina's FGx Forensic Genomics System. *Forensic Sci Int Genet*, 2017. 30: p. 81-92.
43. Moreno LI, Galusha MB, Just R, A closer look at Verogen's Forenseq™ DNA Signature Prep kit autosomal and Y-STR data for streamlined analysis of routine reference samples. *Electrophoresis*, 2018. 39(21): p. 2685-2693.

44. Green RL, Lagacé RE, Oldroyd NJ, Hennessy LK, Mulero JJ, Developmental validation of the AmpFLSTR® NGM SElect™ PCR Amplification Kit: A next-generation STR multiplex with the SE33 locus. *Forensic Sci Int Genet*, 2012. 7(1): p. 41-51.
45. Guo F, Zhou Y, Liu F, Yu J, Song H, Shen H, Zhao B, Jia F, Hou G, Jiang X, Evaluation of the Early Access STR Kit v1 on the Ion Torrent PGM™ platform. *Forensic Sci Int Genet*, 2016. 23: p. 111-120.
46. Thompson JM, Ewing MM, Frank WE, Pogemiller JJ, Nolde CA, Koehler DJ, Shaffer AM, Rabbach DR, Fulmer PM, Sprecher CJ, Storts DR, Developmental validation of the PowerPlex® Y23 System: a single multiplex Y-STR analysis system for casework and database samples. *Forensic Sci Int Genet*, 2013. 7(2): p. 240-250.
47. Gopinath S, Zhong C, Nguyen V, Ge J, Lagace RE, Short ML, Mulero JJ, Developmental validation of the Yfiler® Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications. *Forensic Sci Int Genet*, 2016. 24: p. 164-175.
48. Calafell F, Anglada R, Bonet N, Gonzalez-Ruiz M, Prats-Munoz G, Rasal R, Lalueza-Fox C, Bertranpetit J, Malgosa A, Casals F, An assessment of a massively parallel sequencing approach for the identification of individuals from mass graves of the Spanish Civil War (1936-1939). *Electrophoresis*, 2016. 37(21): p. 2841-2847.
49. Fattorini P, Previdere C, Carboni I, Marrubini G, Sorcaburu-Cigliero S, Grignani P, Bertoglio B, Vatta P, Ricci U, Performance of the ForenSeq™ DNA Signature Prep kit on highly degraded samples. *Electrophoresis*, 2017. 38(8): p. 1163-1174.
50. Kulstein G, Hadrys T, Wiegand P, As solid as a rock-comparison of CE- and MPS-based analyses of the petrosal bone as a source of DNA for forensic identification of challenging cranial bones. *Int J Legal Med*, 2018. 132(1): p. 13-24.
51. Elwick K, Zeng X, King J, Budowle B, Hughes-Stamm S, Comparative tolerance of two massively parallel sequencing systems to common PCR inhibitors. *Int J Legal Med*, 2018. 132(4): p. 983-995.
52. Sidstedt M, Steffen CR, Kiesler KM, Vallone PM, Rådström P, Hedman J, The impact of common PCR inhibitors on forensic MPS analysis. *Forensic Sci Int Genet*, 2019. 40: p. 182-191.
53. Ulbrich W, Anslinger K, Bäßler G, Eckert M, Fimmers R, Hohoff C, Kraft M, Leuker C, Molsberger G, Pich U, Razbin S, Schneider H, Templin M, Wächter A, Weirich V, Zierdt H, Schneider PM, Gemeinsame Empfehlungen der Projektgruppe „Biostatistische DNA-Berechnungen“ und der Spurenkommission zur biostatistischen Bewertung von DNA-analytischen Befunden. *Rechtsmedizin*, 2016. 26(4): p. 291-298.
54. Huszar TI, Jobling MA, Wetton JH, A phylogenetic framework facilitates Y-STR variant discovery and classification via massively parallel sequencing. *Forensic Sci Int Genet*, 2018. 35: p. 97-106.
55. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR, A global reference for human genetic variation. *Nature*, 2015. 526(7571): p. 68-74.
56. Ralf A, Lubach D, Kousouri N, Winkler C, Schulz I, Roewer L, Purps J, Lessig R, Krajewski P, Ploski R, Dobosz T, Henke L, Henke J, Larmuseau MHD, Kayser M, Identification and characterization of novel rapidly mutating Y-chromosomal short tandem repeat markers. *Human Mutation*, 2020. 41(9): p. 1680-1696.
57. Eckert KA and Hile SE, Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Mol Carcinog*, 2009. 48(4): p. 379-388.
58. Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kruger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A, Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet*, 2000. 66(5): p. 1580-1588.
59. Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C, A comprehensive survey of human Y-chromosomal microsatellites. *Am J Hum Genet*, 2004. 74(6): p. 1183-1197.

Anteilerklärung/ Eidesstattliche Versicherung

„Ich, Steffi Bredemeyer, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Analyse von forensisch relevanten short tandem repeats (STRs) mittels massiver paralleler Sequenzierung (MPS) und Vergleich zum Standardverfahren Kapillarelektrophorese (CE)“/ „Analysis of forensically relevant short tandem repeats (STRs) using massively parallel sequencing (MPS) and comparison to the standard procedure capillary electrophoresis (CE)“ 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/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilerklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

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

Datum

Steffi Bredemeyer

Anteilerklärung an den erfolgten Publikationen

Steffi Bredemeyer hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1:

Köcher S*, Müller P*, Berger B, Bodner M, Parson W, Roewer L, Willuweit S, on behalf of the DNASEQEX Consortium, Inter-laboratory validation study of the ForenSeq™ DNA Signature Prep Kit. *Forensic Science International Genetics*, 2018. 36: p. 77-85.

Impact Factor: 4,884

*Diese Autoren haben zu gleichen Teilen zu dieser Arbeit beigetragen.

Beitrag im Einzelnen: selbstständige Durchführung der Experimente am BER, selbstständige Auswertung der Ergebnisse am BER, Gegenüberstellung der Ergebnisse von BER und GMI mit statistischer Auswertung (außer Berechnung der Standardabweichung σ der *inter-locus balance* für Fig. 1), Manuskripterstellung (Korrektur und Ergänzungen durch Koautoren), Grafikerstellung von Fig. 2, Fig. 3, Fig. 6, Erstellung aller Tabellen und der *Supplementary data*.

Publikation 2:

Müller P, Sell C, Hadrys T, Hedman J, **Bredemeyer S**, Laurent FX, Roewer L, Achtruth S, Sidstedt M, Sijen T, Trimborn M, Weiler N, Willuweit S, Bastisch I, Parson W, the SeqForSTR-Consortium, Inter-laboratory study on standardized MPS libraries: evaluation of performance, concordance, and sensitivity using mixtures and degraded DNA. *International Journal of Legal Medicine*, 2020. 134: p. 185–198.

Beitrag im Einzelnen: Durchführung der Sequenzierung am Referenzlabor Charité - Berlin, Korrekturen am Manuskript

Publikation 3:

Bredemeyer S, Roewer L, Willuweit S, Next Generation Sequencing of Y-STRs in father-son pairs and comparison with traditional capillary electrophoresis. *Forensic Sciences Research*, 2021. p.1-6.

Beitrag im Einzelnen: selbstständige Durchführung aller Experimente, Manuskripterstellung (Korrektur und Ergänzungen durch Koautoren), Datenauswertung

Publikation 4:

Alonso A, Barrio PA, Müller P, **Köcher S**, Berger B, Martin P, Bodner M, Willuweit S, Parson W, Roewer L, Budowle B, Current state-of-art of STR sequencing in forensic genetics. Electrophoresis, 2018. 39: p. 2655–2668.

Impact Factor: 2,754

Beitrag im Einzelnen: anteilige Manuskripterstellung und Korrekturen am Manuskript

Datum

Steffi Bredemeyer

Ausgewählte Publikationen

Publikation 1

Köcher S*, Müller P*, Berger B, Bodner M, Parson W, Roewer L, Willuweit S, on behalf of the DNASEQEX Consortium, Inter-laboratory validation study of the ForenSeqTM DNA Signature Prep Kit, Forensic Science International Genetics, 2018. 36: p. 77-85.

*Diese Autoren haben zu gleichen Teilen zu dieser Arbeit beigetragen.

<https://doi.org/10.1016/j.fsigen.2018.05.007>

Publikation 2

Müller P, Sell C, Hadrys T, Hedman J, **Bredemeyer S**, Laurent FX, Roewer L, Achtruth S, Sidstedt M, Sijen T, Trimborn M, Weiler N, Willuweit S, Bastisch I, Parson W, the SeqForSTR-Consortium, Inter-laboratory study on standardized MPS libraries: evaluation of performance, concordance, and sensitivity using mixtures and degraded DNA. *International Journal of Legal Medicine*, 2020. 134: p. 185–198.

<https://doi.org/10.1007/s00414-019-02201-2>



Inter-laboratory study on standardized MPS libraries: evaluation of performance, concordance, and sensitivity using mixtures and degraded DNA

Petra Müller¹ · Christian Sell² · Thorsten Hadrys³ · Johannes Hedman^{4,5} · Steffi Bredemeyer⁶ · Francois-Xavier Laurent⁷ · Lutz Roewer⁶ · Sabrina Achtrouth⁸ · Maja Sidstedt^{4,5} · Titia Sijen⁹ · Marc Trimborn⁸ · Natalie Weiler⁹ · Sascha Willuweit⁶ · Ingo Bastisch² · Walther Parson^{1,10}  · the SeqForSTR-Consortium

Received: 4 September 2019 / Accepted: 29 October 2019 / Published online: 19 November 2019
© The Author(s) 2019

Abstract

We present results from an inter-laboratory massively parallel sequencing (MPS) study in the framework of the SeqForSTRs project to evaluate forensically relevant parameters, such as performance, concordance, and sensitivity, using a standardized sequencing library including reference material, mixtures, and ancient DNA samples. The standardized library was prepared using the ForenSeq DNA Signature Prep Kit (primer mix A). The library was shared between eight European laboratories located in Austria, France, Germany, The Netherlands, and Sweden to perform MPS on their particular MiSeq FGx sequencers. Despite variation in performance between sequencing runs, all laboratories obtained quality metrics that fell within the manufacturer's recommended ranges. Furthermore, differences in locus coverage did not inevitably adversely affect heterozygous balance. Inter-laboratory concordance showed 100% concordant genotypes for the included autosomal and Y-STRs, and still, X-STR concordance exceeded 83%. The exclusive reasons for X-STR discordances were drop-outs at DXS10103. Sensitivity experiments demonstrated that correct allele calling varied between sequencing instruments in particular for lower DNA amounts (≤ 125 pg). The analysis of compromised DNA samples showed the drop-out of one sample (FA10013B01A) while for the remaining three degraded DNA samples MPS was able to successfully type $\geq 87\%$ of all aSTRs, $\geq 78\%$ of all Y-STRs, $\geq 68\%$ of all X-STRs, and $\geq 92\%$ of all iSNPs demonstrating that MPS is a promising tool for human identity testing, which in return, has to undergo rigorous in-house validation before it can be implemented into forensic routine casework.

Keywords Short tandem repeats · Massively parallel sequencing · ForenSeq DNA Signature Prep Kit · Inter-laboratory study · Collaborative study

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00414-019-02201-2>) contains supplementary material, which is available to authorized users.

✉ Walther Parson
walther.parson@i-med.ac.at

¹ Institute of Legal Medicine, Medical University of Innsbruck, Müllerstraße 44, 6020 Innsbruck, Austria

² Federal Criminal Police Office, Wiesbaden, Germany

³ Institute of Forensic Sciences, DNA Department, Bavarian State Criminal Police Office, Munich, Germany

⁴ Swedish National Forensic Centre (NFC), Linköping, Sweden

⁵ Applied Microbiology, Department of Chemistry, Lund University, Lund, Sweden

⁶ Institute of Legal Medicine and Forensic Sciences, Charité – Universitätsmedizin Berlin, Berlin, Germany

⁷ Institut National de Police Scientifique, Laboratoire de Police Scientifique de Lyon, Ecully Cedex, France

⁸ The Police President in Berlin, Forensic Science Institute, Berlin, Germany

⁹ Biological Traces, Netherlands Forensic Institute, Laan van Ypenburg 6, 2497 GB The Hague, The Netherlands

¹⁰ Forensic Science Program, The Pennsylvania State University, State College, PA, USA

Introduction

With the emergence of massively parallel sequencing (MPS) technologies, molecular genetic tools have been developed to characterise the nucleotide sequence of short tandem repeat (STR) markers that have so far been analysed using fragment sizing by capillary electrophoresis (CE) [1–7]. As is common practise in forensic genetics, such novel tools undergo detailed evaluation [3, 8–12] and validation [1, 13, 14] prior to their application in casework. Recent population studies have shown that sequencing of STRs lead to an increased power of discrimination compared to commonly used CE sizing by a) illustrating micro-variant structures and b) identifying sequence variations located in the repeat- (isometric variants; alleles of identical size but different in internal sequence) or flanking region [15–20]. Moreover, current MPS-kits provide the option to multiplex various loci simultaneously within one reaction, such as STRs (autosomal, Y- and X-chromosomal), mitochondrial DNA (control region) as well as single nucleotide polymorphisms (SNPs) that provide estimates of identity, phenotypic traits, biogeographical and ancestry information [21–24]. The inclusion of SNPs adds the benefit [25–27] that these markers allow for shorter amplicon design, which can support human identification of degraded or challenging DNA samples [28–30]. Although published studies have demonstrated that MPS is a promising tool for forensic applications, little is known about the method's variability and potential differences in the limit of detection between sequencing platforms from the same supplier. Inter-laboratory studies are however highly relevant to understand the performance of this new technology when applied to real casework settings. Since the workflow for MPS-based sequence analysis of forensic genetic markers is more complex than for electrophoresis-based systems, we decided to reduce the complexity of this study to libraries prepared at the organising laboratory (OL) from forensically relevant samples that were shipped to the participating laboratories in the following way: In the framework of the SeqForSTRs project (EU Project ISF-Nr. IZ25-5793-2015-30 2017-2020, [31]) the ForenSeq DNA Signature Prep Kit [22] was used for collaborative validation experiments and population studies. As a general benchmark for the comparability of the results between laboratories, and, to reduce the inherent complexity of MPS experiments derived from multiple steps during manual library preparation, the SeqForSTRs consortium performed a collaborative exercise analysing a centrally prepared, standardized sequencing library. The library was distributed between eight participating laboratories located in Austria, France, Germany, The Netherlands, and Sweden and were analysed on their respective MiSeq FGx sequencing platforms (Illumina, San Diego, USA;

[32]). In order to monitor the effect of shipment, one library aliquot was sent to one of the participants, immediately returned to the OL and analysed (run 8) to compare the results to those of the test run prior to shipment (run 0).

We note that only STR genotype calls were considered here in agreement with the aim of the SeqForSTRs project, except for ancient DNA samples, where also SNP information was analysed to evaluate their performance in highly degraded DNA. To rate the performance of iSNPs in compromised samples additional SNP information was examined for standard reference material 2391c (components A–C). The library contained selected DNA samples (control DNAs, reference material, mixtures and degraded DNA) to test for instrument variability, concordance, and sensitivity.

Materials and methods

This collaborative study combined data generated in nine sequencing runs that were conducted in eight laboratories located in Austria, France, Germany, The Netherlands, and Sweden (Table 1). To ensure uniform starting conditions a standardized sequencing library consisting of 25 selected DNA samples, a negative and a positive amplification control (see “Sample selection” section) was prepared using the ForenSeq DNA Signature Prep Kit (Verogen, San Diego, USA) at the OL. To assess the quality of sequencing products an initial test run (run 0) was performed at the OL (Table 1) prior to shipment. Pooled sequencing libraries were shipped with the United Parcel Service (unchilled and under ambient conditions) and sequenced by all participants using their respective MiSeq FGx sequencers (Illumina). To monitor transport conditions (under ambient temperatures) one aliquot of the library was shipped by the OL to one of the participants, returned to the OL upon receipt and analysed (run 8).

Sample selection

Concordance, mixture, and sensitivity

Concordance was assessed using three single donor standard reference materials (SRM): 2391c, component A–C [33] purchased from the National Institute of Standards and Technology (NIST; Gaithersburg, USA), control DNA 2800M (Promega, Madison, USA), and two samples from previous GEDNAP (German DNA Profiling; <https://www.gednap.org/>) proficiency tests (G49-S4, single source; G49-S1, two-person mixture).

Two-person mixtures were prepared using control DNAs 2800M (major contributor; Promega) and 9947A (minor contributor; Thermo Fisher Scientific (TFS), Waltham, USA) at a stock concentration of 9 ng/μL and five ratios (M500 = 1:1; M166 = 83.3:16.7; M91 = 90.9:9.1; M62.5 = 93.7:6.3 and

Table 1 Run and quality metrics information for all sequencing runs performed within the SeqforSTRs study. The resulting data were analysed and presented in anonymous form. An initial test run (run 0) was performed at the organising laboratory to assess the quality of

sequencing products. Prior to sequencing each laboratory loaded 7 μL of pooled library on the MiSeq FGx instrument according to the manufacturer's recommendation (Verogen)

Run	Cluster density (K/mm ²)	Cluster passing filter (%)	Phasing (%)	Pre-phasing (%)	Total no. of reads (performance testing) Single source samples	UAS version	NTC
UAS Guide (range)	400–1650 K/mm ²	$\geq 80\%$	$\leq 0.25\%$	$\leq 0.15\%$			
0	706	96.13	0.141	0.102	704,959	1.2.16337	Negative
1	684	97.71	0.180	0.099	847,323	1.2.16337	Negative
2	432	97.25	0.168	0.144	438,786	1.2.16337	Negative
3	818	94.79	0.171	0.103	948,395	1.2.16173	Negative
4	603	97.65	0.193	0.084	641,752	1.2.16337	Negative
5	587	96.19	0.166	0.113	579,248	1.2.16337	Negative
6	486	97.76	0.226	0.102	650,416	1.2.16173	Negative
7	862	93.86	0.181	0.100	1,123,576	1.2.16337	Negative
8	625	97.03	0.192	0.130	729,121	1.2.16337	Negative
Mean	645	96.49	0.185	0.109	740,397		
SD	141	1.38	0.020	0.018	205,448		
Median	614	97.03	0.181	0.102	689,769		
Min	432	93.86	0.141	0.084	438,786		
Max	862	97.76	0.226	0.144	1,123,576		

M47.6 = 95.2:4.8). The initial intention was to prepare the mixtures with a male minor contributor however, the female fraction (9947A) was used as minor contributor mistakenly, which affected the ability to test male minor contributions in the mixture analyses. All mixtures were diluted accordingly to 1 ng/ μL in molecular grade water prior to library preparation. Mixture study consisted of M166, M91, M62.5 and M47.6, which were analysed as singletons only. Sensitivity was assessed using mixture M500 prepared to a final DNA input of 1000 pg, 500 pg, 250 pg, 125 pg, 62 pg and 31 pg in molecular grade water and analysed in duplicate.

Ancient DNA samples

To determine the robustness and performance of large MPS panels on highly degraded DNA, four ancient DNA (aDNA) samples were included. Libraries were prepared from extracted DNA from two bone (femur: FA10013B01A, humerus: FA10026B01A) and two tooth samples (molar: FA10030T01A, incisor: FA10058T01B) according to [22], respectively. Ancient samples dated between the fifth to eighth century ([34]) and were taken from earlier studies [35].

DNA quantification

The amount of genomic DNA was determined with a real-time PCR assay targeting specific AluYb8 sequences [36]. A spiked in vitro mutagenized and cloned part of the *human retinoblastoma susceptibility protein 1 (RBI)* gene was co-

amplified as internal amplification positive control (pRBI_{IPC}) according to [37], updated in [35]. Calibration curve analysis covered a DNA input range from 10 ng to 169.5 fg per reaction and was analysed in duplicate. The reaction volume was 10 μL consisting of 5 μL TaqMan Fast Universal PCR mix (TFS), 3 μL primer probe premix (made in-house) and 2 μL extracted DNA sample. Thermal cycler protocol consisted of an initial denaturation at 95 °C for 20 s followed by 40 cycles of denaturation at 95 °C for 3 s and annealing/elongation at 60 °C for 30 s. Samples were run on an Applied Biosystems 7500 Fast Real-Time PCR Instrument (TFS) using the HID Real-Time PCR Software v 2.3. Kinetic information for the pRBI_{IPC} system yielded no indication of inhibition during DNA amplification.

Library preparation and sequencing

Each laboratory obtained one 2 mL screw-cap tube (Sarstedt, Nümbrecht, Germany; labelled: *SeqForSTRs, SeqPool 20 μL , ForenSeq Kit, 29.11.2017 PM*) containing 20 μL pooled sequencing library. The final library pool was made of 25 selected DNAs, a negative and a positive amplification control. To ensure sufficient library volume OL simultaneously prepared three library pools, including the same samples and index adapters (i5 and i7) using the ForenSeq DNA Signature Prep Kit, primer mix A (Verogen, [22]) in one 96-well plate (Table S1). Library preparation, purification and normalization were performed according to [22]. All samples were amplified with 1 ng DNA according to [22], except for

G49-S4 (2.6 ng DNA), the samples for the sensitivity study (serial dilution: 1000 pg–31 pg DNA) and aDNA samples.

Prior to pooling, 30 μ L of each library sample (present in pools 1, 2 and 3; Table S1) was joined into a 0.8 mL deep-well storage plate (one sample/well). Libraries were pooled, diluted and denatured following [22] before loading into a MiSeq FGx Reagent Cartridge and sequenced on the MiSeq FGx instrument ([32], Table S2). To achieve uniform designation for all sequencing runs OL provided a text-based sample sheet including relevant information on, e.g. sample name as well as adapter combination needed for demultiplexing and data analysis (Table S3).

Data analysis

Capillary electrophoresis

STR typing was performed at OL with PowerPlex ESX 17 (Promega [38]), PowerPlex 16 System (Promega [39]), PowerPlex 21 System (Promega [40]), Investigator Argus X-12 (Qiagen, Hilden, Germany [41]), AmpFISTR Yfiler (TFS [42]), Genderplex (an in-house developed sex-typing assay [35, 43]) or AmpFISTR NGM Select Kit (TFS [44]). Amplification was performed on an Applied Biosystems GeneAmp 9700 thermal cycler (TFS) following the manufacturer's or recommended protocols [35, 38–44]. PCR fragments were separated and detected using an Applied Biosystems Prism 3500XL Genetic Analyzer (TFS). The analysis of size-based STR fragments was conducted at OL with GeneMapper ID-X software, version 1.2 (TFS) by applying in-house validated dye thresholds: blue – 50 relative fluorescence units (RFU), green – 80 RFU, yellow – 100 RFU, red – 100 RFU.

Universal Analysis Software

Analysis of MPS data utilized the ForenSeq Universal Analysis Software, version 1.2 (Verogen, exact version see Table 1) for allele and genotype calling by applying the manufacturer's default analysis settings [45]. Universal Analysis Software (UAS) called alleles and genotypes based on the number of reads by applying a specified analytical (AT; 1.5%) and interpretation threshold (IT; 4.5%), except for DYS398II (AT: 5%; IT: 15%), DYS448 (AT: 3.3%; IT: 10%) and DYS635 (AT: 3.3%; IT: 10%) [45]. Single nucleotide polymorphism (SNP) analysis of aDNA samples were manually revised and no-call genotype corrected according to [11]. For example, no-call SNPs that fell below the manufacturer's default IT but in the range of 20 to 29 reads were manually corrected and genotype results put in brackets. No-call SNPs that fell below the manufacturer's default IT but showed reads \geq 30 were manually corrected and called. All laboratories applied UAS using the preinstalled default

settings. Following to sequencing, an Excel-based genotype report-file was generated at each laboratory and sent to the OL for further analysis. Provided data were analysed at OL and reported in anonymous form.

Statistical analyses

Figures, diagrams and statistical analyses were generated using Microsoft Excel, GraphPad Prism, version 7.04 for Windows and IBM SPSS, version 24 [46].

Allele frequencies and statistical calculations

Allele frequencies used for genotype frequency calculations were taken from SPSmart v5.1.1 [47–50]. Allele frequencies for SRM 2391c (components A–C) [33] were selected according to the manufacturer's provided information on geographic origin of the respective DNA sample (2391cA: European, 2391cB: American, 2391cC: Oceania). Random match probability (RMP) estimates were calculated according to the formulae 4.1. [51]. Alternatively, likelihood ratios (LR) were used, which is a measure of the power of proof regarding the hypothesis that two DNA profiles were derived from the same suspect and is known as the inverse of the RMP [51, 52].

Results and discussion

Sequencing run information

The results of the collaborative exercise were analysed using UAS, which provided the following general information on the sequencing runs (recommended values in brackets [45]; Table 1): cluster density (400–1650 K/mm²), cluster passing filter (\geq 80%), phasing (\leq 0.25%) and pre-phasing (\leq 0.15%). Mean cluster density plus standard deviation (SD) was calculated for all runs and amounted on average to 645 K/mm² (SD = 141 K/mm²) with a minimum of 432 K/mm² (run 2) and a maximum of 862 K/mm² (run 7; Table 1). The library was analysed at OL after preparation (run 0) and after shipment (run 8). The cluster density of the post-shipment run (run 8) yielded 88.5% of the initial cluster density value (run 0). Still, this cluster density fell within the range of the cluster densities of the other sequencing runs. Mean values (SD in brackets) for cluster passing filter, phasing and pre-phasing amounted on average to 96.49% (1.38%), 0.185% (0.020%) and 0.109% (0.018%) with a minimum of 93.86% (run 7), 0.141% (run 0) and 0.084% (run 4) and a maximum of 97.76% (run 6), 0.226% (run 6) and 0.144% (run 2), respectively (Table 1). Based on these findings, we conducted a more detailed analysis of all datasets to examine possible effects of cluster density variation on the overall data interpretation quality. Below, we present and discuss inter-laboratory concordance ([Inter-](#)

laboratory concordance section), locus coverage (Locus coverage section) and heterozygous balance (Heterozygous balance section). All negative amplification controls yielded no detectable sequences.

Inter-laboratory concordance

Inter-laboratory concordance was assessed by comparing a) genotype profiles to known reference profiles (SRM 2391c A–C, control DNA 2800M) [22, 33] or previously obtained MPS results (GEDNAP samples G49-S1 and G49-S4; data not shown) and b) genotype from identical library preparation and run in different laboratories. All runs obtained fully concordant autosomal STR (aSTR) and Y-STR results to known reference and between sequencing runs for SRM 2391c A–C, control DNA 2800M, G49-S1 and G49-S4 (Table S4). X-STR analysis revealed full concordance to known reference and between sequencing runs for all runs, except for runs 1 (G49-S1), 2 (2391cA, 2391cC), 4 (2800M) and 6 (2800M, G49-S1). X-STR concordance for control DNA 2800M, SRM 2391cA, 2391cC and G49-S1 was 85.7% (runs 4 and 6: 6 out of 7 alleles; both runs), 92.3% (run 2: 12 out of 13 alleles), 85.7% (run 2: 6 out of 7 alleles) and 91.7% (runs 1 and 6: 11 out of 12 alleles; both runs), respectively (Table S4). All six instances of discordances were due to allelic drop-out at DXS10103 (control DNA 2800M: allele 18 (hemizygous genotype); SRM 2391cA: allele 19 (genotype: 18, 19); SRM 2391cC: allele 19 (hemizygous genotype) and G49-S1: allele 18 (genotype: 18, 20) (Table S4)). Interestingly, for G49-S1 runs 0, 2–5 and 8 obtained a marker drop-out at DX10103 resulting in 83.3% concordance, while only run 7 correctly typed both alleles that were known from reference (Table S4). In addition, DXS10103 was reported to be sensitive to locus drop-out in earlier studies [1, 8, 13, 53, 54] most likely due to underperformance of this marker during amplification. Therefore, further optimization would be required as recommended by [16, 53].

Performance testing

Cluster density is known to be an important but also critical parameter that affects overall run quality, the total amount of sequencing reads and reads passing filter. For example, while lower cluster density does not necessarily adversely affect data quality, it predominantly leads to decreased data output. In contrast, overclustering potentially leads to poor run performance including the introduction of sequencing artefacts, elevated sequencing error rates together with lowered sequencing data output [55]. Based on the observed cluster density variation (Table 1, “Sequencing run information” section) we checked for possible differences in locus coverage and intra-locus balance (heterozygous genotype). The sample set analysed in

this respect consisted of five single source samples (control DNA 2800M, SRM 2391c A–C and GEDNAP G49-S4).

Locus coverage

Locus coverage was calculated for each STR marker and sequencing run (Table S5). Among aSTRs, TH01 (49,387; 13,501) showed the highest mean number of reads while D19S443 (3691; 1003) performed the least in terms of locus coverage (Table S5). In line with these findings for the ForenSeq DNA Signature Prep Kit, Müller et al., 2018 [4] observed comparable results for TH01 and D19S443 evaluating both Globalfiler MPS-kits (i.e. the Early Access Applied Biosystems Precision ID Globalfiler Mixture ID and the Globalfiler NGS STR Panels) on the Ion S5 instrument [4]. Although using a centrally prepared sequencing library we observed locus coverage variation for single source samples sequenced on the eight MiSeq FGx sequencers (Tables 1 and S5). The mean number of reads/run was 740,397 (205,448) ranging from 438,786 reads (run 2) to 1,123,576 reads (run 7) relating to the afore-mentioned single source samples (Table 1). Run 2 yielded lowest locus coverage over all STRs, whereas run 7 displayed highest locus coverage for nearly all STRs (55 out of 59 STRs; 93.2%) (Table S5).

Heterozygous balance

Heterozygous balance (HB), also known as intra-locus balance, was calculated (for each sample and locus) for heterozygous- and isometric genotypes by dividing the lower number of allele reads by the higher number of allele reads [56]. Average HB/run and mean HB/locus was calculated for 27 aSTRs, Amelogenin, two Y-STRs and six X-STRs. Results were comparable for each particular locus between all sequencing runs (Table S6). All markers showed $HB \geq 0.60$ (manufacturer’s default threshold settings) except for D22S1045 and D5S818 that yielded lower HB values, 0.43 (0.02) and 0.57 (0.02), respectively (Table S6). These findings were in line with earlier studies that reported imbalances at D22S1045 [3, 12–15, 17, 22, 54, 57] and D5S818 [53, 54, 57]. Six aSTRs (D4S2408, CSF1PO, D7S820, D8S1179, D9S1122 and TH01), one Y-STR (DYF387S1) and three X-STRs (DXS10135, DXS10074 and HPRTB) displayed $HB \geq 0.90$ (Table S6). These findings were consistent with previous studies [3, 13]. Results for afore-mentioned aSTRs and one X-STR were in line with Silvia et al., 2017 [54], except for DXS10135 and DXS10074.

Apparently, this analysis showed that low locus coverage did not necessarily increase heterozygous imbalance (Fig. 1). Among aSTRs, TH01 showed highest locus coverage along with highly balanced allele calls (0.95, 0.01). In contrast, D19S443 displayed lowest locus coverages for most sequencing runs with relatively balanced genotype calls (0.82, 0.03)

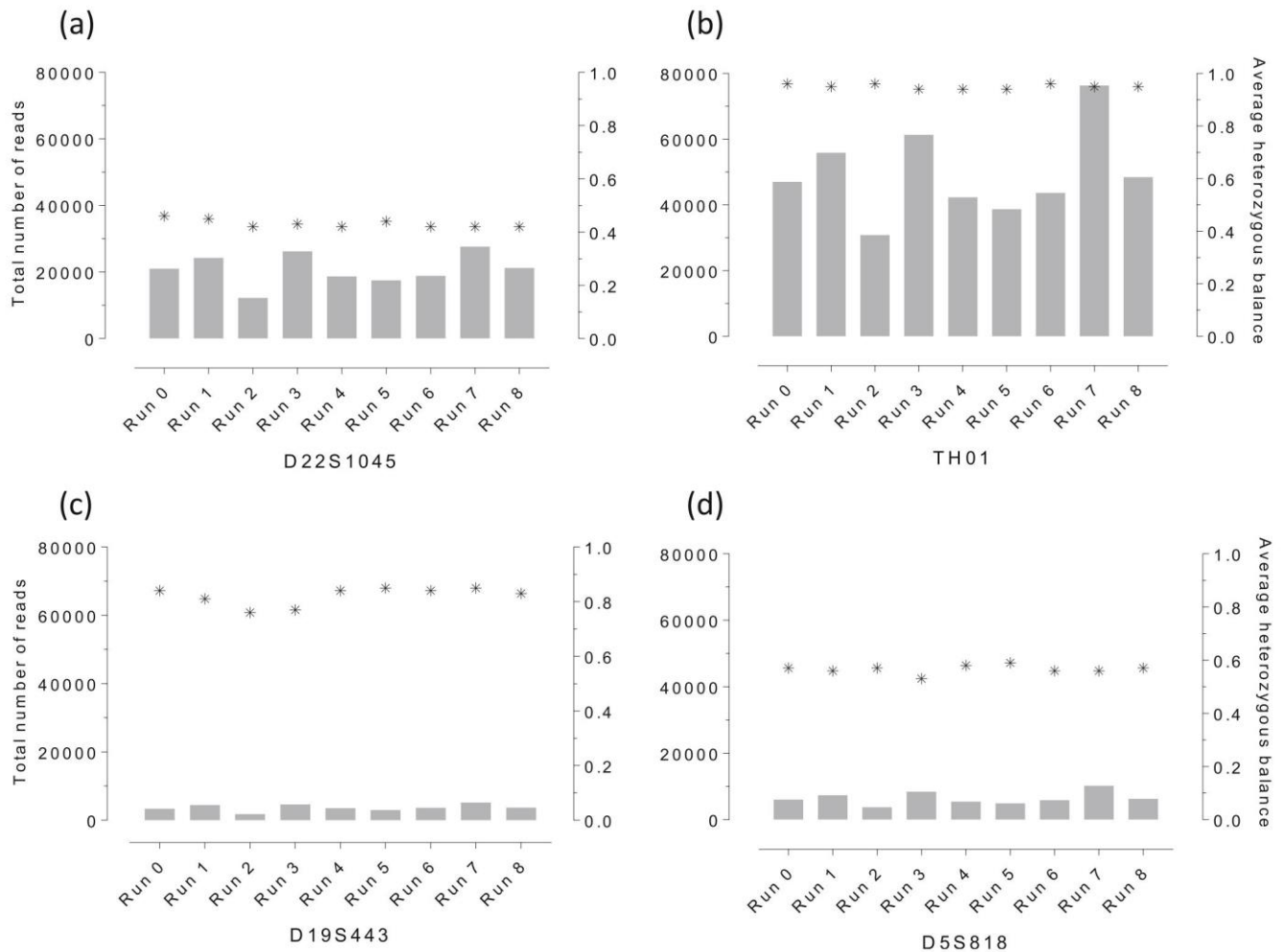


Fig. 1 Inter-laboratory investigation of possible correlations between locus coverage differences and heterozygous imbalances. Total locus coverage and average heterozygous balance (HB) were calculated using five single source samples (SRM 2391c A-C, control DNA 2800M and G49-S1; amplified with 1 ng DNA). Against expectations, we found that varying locus coverage did not affect heterozygous balance. For example,

(Table S6, Fig. 1). Observed differences in locus coverage as a result of cluster density variation had only minor effects on the fluctuation of HB (Tables 1 and S6).

Sensitivity

Sensitivity of the ForenSeq DNA Signature Prep Kit was assessed with decreasing DNA input of sample M500 ranging from 1000 pg to 31 pg (final DNA input). Alleles were called based on read counts when they exceeded the manufacturer's default IT [1, 45]. The full profile of M500 consisted of 80 aSTR alleles, 26 Y-STR alleles and 17 X-STR alleles.

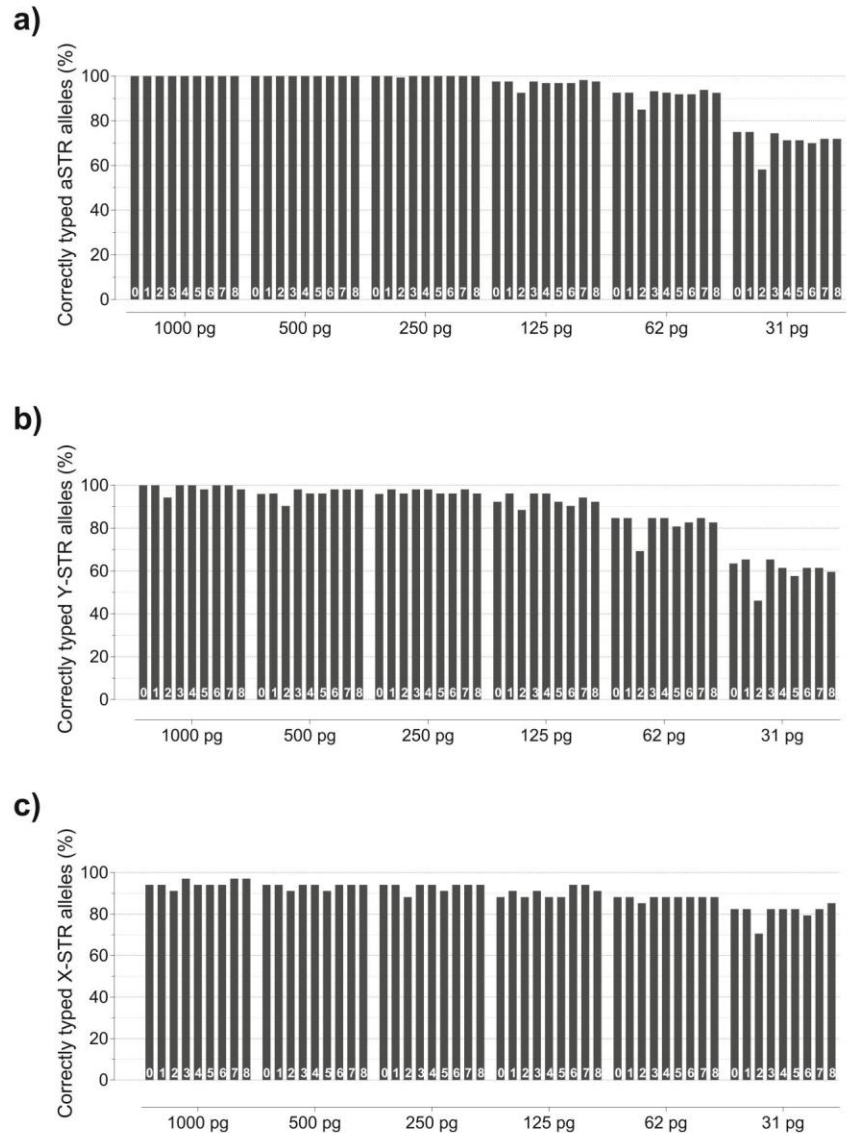
The mean aSTR sensitivity success rate over all runs was 100% down to 500 pg and still $\geq 91.8\%$ (2.6%) down to 62 pg DNA input (Table S7). On average, all sequencing runs successfully typed $\geq 93.2\%$ (2.8%) of all Y-STRs down to 125 pg DNA (Table S7). Furthermore, X-STR analysis showed

a D22S1045 and **d** D5S818 were found to show $HB \leq 0.60$ but mean locus coverages of 20,838 reads (D22S1045) and 6460 reads (D5S818). In contrast, **b** TH01 and **c** D19S443 were found to be highly balanced forensic markers though showing obvious differences in locus coverage (mean locus coverage of 49,387 reads (TH01) and 3691 reads (D19S443); Table S5)

sensitivity levels $\geq 90.5\%$ (2.5%) using 125 pg DNA (Table S7). For all runs, the average sensitivity level was 71.0% (5.1%), 60.3% (5.8%) and 81.1% (4.2%) for aSTRs, Y-STRs and X-STRs, respectively, using 31 pg DNA (Table S7). These findings were in agreement with Churchill et al., 2016 [3] who reported yields of more than 94% of all alleles using 100 pg DNA amplified with the ForenSeq DNA Signature Prep Kit.

To investigate potential instrument-related differences in sensitivity we carefully examined aSTR, Y-STRs and X-STR results for each sequencing run individually. Results for aSTRs revealed that all runs obtained full-profiles down to 250 pg, except for run 2 (drop-out at D19S443, allele 13, for 250 pg) (Table S8, Fig. 2a). The results showed that all runs obtained comparable sensitivity levels for all classes of markers included in the ForenSeq DNA Signature Prep Kit, except for run 2 (Table S8, Figs. 2 and S1). Particularly for

Fig. 2 Sensitivity summary showing sensitivity levels for each sequencing run (numbers on the bottom indicate the respective sequencing run) and marker class included in the ForenSeq DNA Signature Prep Kit: **a** autosomal STRs (aSTRs), **b** Y-STRs, and **c** X-STRs. Sample M500 was prepared at ratio 1:1 using control DNAs 9947A and 2800M. A serial dilution was prepared at 1000 pg, 500 pg, 250 pg, 125 pg, 62 pg, and 31 pg total DNA input and amplified in duplicate. Results indicate that all sequencing runs obtained comparable sensitivity levels for all classes of markers, except for run 2. Particularly for DNA input amounts of ≤ 62 pg sensitivity levels of run 2 differed distinctly from the other sequencing runs



lower DNA input amounts (≤ 62 pg) sensitivity levels of run 2 differed distinctly from the others (Fig. 2a–c). Still, run 2 success rates were 58.1%, 46.2% and 70.6% for aSTRs, Y-STRs and X-STRs using as little as 31 pg DNA, respectively (Table S7, Figs. 2 and S1). In general, results for the ForenSeq DNA Signature Prep Kit aSTR sensitivity were comparable to CE-based STR kits [58–61] and in line with earlier published MPS studies [1, 3, 4, 14, 54].

Mixtures

We defined mixtures here as DNA profiles showing more than two alleles in at least two forensic STR markers [62], based on biological material from two or more persons contributing to the sample being tested [63]. Mixture study was performed using samples M166, M91, M62.5 and M47.6 prepared from control DNAs 9947A (minor contributor) and 2800M (major

contributor) at four ratios (83.3:16.7; 90.9:9.1; 93.7:6.3 and 95.2:4.8). The total DNA input amount for each mixture was 1 ng. Within each mixture we identified aSTR alleles that were consistent with the minor contributor genotype (9947A). Note, only those 12 aSTRs with genotype calls that were distinguishable from the major contributor’s genotype were included for mixture analysis (D1S1656, TPOX, D2S1338, D3S1358, D5S818, D8S1179, vWA, PentaE, D16S539, D18S51, D21S11 and D22S1045) yielding a total of 41 distinguishable alleles (Table 2 a and b). Alleles were called based on their reads by applying the manufacturer’s default thresholds. On average, the ForenSeq DNA Signature Prep Kit was able to detect 100%, 94.4% (1.7%), 76.7% (5.0%) and 45% (5.0%) of the minor contributor’s alleles for M166, M91, M62.5 and M47.6, respectively (Table 2b). Furthermore, for M62.5 and M47.6 we observed a noticeable decrease in the proportion of correctly assigned alleles

Table 2 Summary of mixture analysis showing the total number of alleles observed for mixtures M166, M91, M62.5, and M47.6 prepared at four ratios (83.3:16.7; 90.9:9.1; 93.7:6.3, and 95.2:4.8). Mixtures were prepared using control DNAs 9947A (minor component) and 2800M (major component) amplified with 1 ng total DNA (analysed as singletons). Including only distinguishable autosomal STRs, we were

able to identify 41 potential alleles per mixture at forensic markers D1S1656, TPOX, D2S1338, D3S1358, D5S818, D8S1179, vWA, PentaE, D16S539, D18S51, D21S11, and D22S1045. **a** depicts results obtained for all sequencing runs individually while **b** summarizes the average percentage of correctly typed alleles per minor contributor (dark grey header) and mixture (light grey header)

a)		Mixture ratio 2800M:9947A (%)		Total DNA input/ component (pg)		Total number of observed alleles			Allele drop-out		Total no. of correctly typed alleles (%)	
Run	Reference	Mixture	major:minor	2800M	9947A	Mixture	2800M	9947A	2800M	9947A	Mixture	Minor contributor
0						41	21	20	0	0	41 (100)	20 (100)
1						41	21	20	0	0	41 (100)	20 (100)
2						41	21	20	0	0	41 (100)	20 (100)
3						41	21	20	0	0	41 (100)	20 (100)
4		M166	83.3:16.7	833	167	41	21	20	0	0	41 (100)	20 (100)
5						41	21	20	0	0	41 (100)	20 (100)
6						41	21	20	0	0	41 (100)	20 (100)
7						41	21	20	0	0	41 (100)	20 (100)
8						41	21	20	0	0	41 (100)	20 (100)
aSTRs												
0						40	21	19	0	1	40 (97.6)	19 (95)
1						40	21	19	0	1	40 (97.6)	19 (95)
2						40	21	19	0	1	40 (97.6)	19 (95)
3						40	21	19	0	1	40 (97.6)	19 (95)
4		M91	90.9:9.1	909	91	39	21	18	0	2	39 (95.1)	18 (90)
5						40	21	19	0	1	40 (97.6)	19 (95)
6						40	21	19	0	1	40 (97.6)	19 (95)
7						40	21	19	0	1	40 (97.6)	19 (95)
8						40	21	19	0	1	40 (97.6)	19 (95)
0						37	21	16	0	4	37 (90.2)	16 (80)
1						37	21	16	0	4	37 (90.2)	16 (80)
2						36	21	15	0	5	36 (87.8)	15 (75)
3						36	21	15	0	5	36 (87.8)	15 (75)
4		M62.5	93.7:6.3	937.5	62.5	36	21	15	0	5	36 (87.8)	15 (75)
5						34	21	13	0	7	34 (82.9)	13 (65)
6						37	21	16	0	4	37 (90.2)	16 (80)
7						37	21	16	0	4	37 (90.2)	16 (80)
8						37	21	16	0	4	37 (90.2)	16 (80)
0						29	21	8	0	12	29 (70.7)	8 (40)
1						30	21	9	0	11	30 (73.2)	9 (45)
2						29	21	8	0	12	29 (70.7)	8 (40)
3						30	21	9	0	11	30 (73.2)	9 (45)
4		M47.6	95.2:4.8	952.4	47.6	31	21	10	0	10	31 (75.6)	10 (50)
5						30	21	9	0	11	30 (73.2)	9 (45)
6						32	21	11	0	9	32 (78.1)	11 (55)
7						30	21	9	0	11	30 (73.2)	9 (45)
8						29	21	8	0	12	29 (70.7)	8 (40)

b)		Overall correctly typed minor contributor's alleles (%)				Overall correctly typed mixture alleles (%)			
Sample	/	Range		Mean (SD)	/	Range		Mean (SD)	
		Min	Max			Δ	Range		
M166	/	100	0	100 (0.0)	/	100	0	100 (0.0)	
M91	90	95	5	94.4 (1.7)	95.1	97.6	2.4	97.3 (0.8)	
M62.5	65	80	15	76.7 (5.0)	82.9	90.2	7.3	88.6 (2.4)	
M47.6	40	55	15	45.0 (5.0)	70.7	78.0	7.3	73.2 (2.4)	

(Figure S2a), which is most likely related to the decline in minor contributor's DNA input, while the major contributors DNA input remained almost unchanged (Table 2a, Figure S2b). Referring to M47.6 (ratio: 95.2:4.8) we were able to type 73.2% (2.4%) of all alleles. These findings for mixture analysis were comparable to Jäger et al., 2017 [1], who observed on average 73 allele calls out of 83 alleles (or 87.9%) at 5% minor contributor DNA input including 27 aSTRs. However, when detecting the minor contributors alleles, findings for M91 demonstrated that 5% variation (ranging from 90% to 95%) was obtained between all sequencing runs, which increased to 15% for M62.5 (ranging from 65% to 80%) and M47.6 (ranging from 40% to 55%), respectively (Table 2b, Figure S2a). Still, results for M47.6 showed that on average all runs were able to detect 45% (5%) of the minor contributor's alleles using as little as 47.6 pg DNA (Table 2b).

Again, despite the constant DNA input of 1 ng, clear differences in the total number of sequencing reads were obtained between the participating laboratories (Figure S2b). Figure S2b shows the read-proportion for each mixture component of M166, M91, M62.5, and M47.6 in relation to DNA input. As expected, reads of the minor component decreased with lowered DNA input while those of the major component

accumulated with increasing DNA amount, except for M47.6. Generally, in proportion to DNA input the fraction of the minor contributor 9947A received higher number of reads compared to those of 2800M (Table S9). The observed mixture ratios were equivalent for all mixtures and between sequencing runs, except for M47.6, run 4 (Table S9). Although all mixtures were prepared using 1 ng DNA M47.6 showed an additional drop in the total number of reads for control DNA 2800M. However, this drop approximated the theoretical mixture ratio that would have been expected (Table S9, Figure S2b). It should be noted that the decrease in sequencing reads of control DNA 2800M is more likely related to sample preparation or laboratory procedure than to differences between sequencing instruments.

Ancient DNA samples

The analysis of aDNA is associated with possible contamination at various stages of the process, some of which are difficult to control, e.g. contamination prior to receiving the sample in the laboratory. Also, aDNA is characterised by some degree of degradation that depends on various factors such as the general age of the specimen - although this is not following

a strictly linear correlation - and environmental (storage) conditions. Usually ancient samples contain lower DNA quantity compared to contemporary samples [64]. To gather information on the ForenSeq DNA Signature Prep Kit’s performance on highly degraded DNA samples we included four compromised specimens from the early medieval times. As expected, we observed lowered percentage of correctly typed loci compared to high quality samples with broad variation in success rates between the samples (success rates ranged from 1.3% to 96.7%, Table 3, Figure S3). Genotypes were called when they exceeded the manufacturer’s default threshold [45]. Manual revision of the non-called STR loci (automatically by the software) was performed only for aDNA samples (all runs) if the locus information in the provided sample-details-report Excel-table indicated “INC” (genotype) in combination with the comment “interpretation threshold” (QC indicator). INC (inconclusive) indicated ambiguous genotype results that were not reported in the Excel-table. Non-called SNPs that indicated “INC” (genotype) in combination with “interpretation threshold” (QC indicator) and fell in the range of 20 to 29 reads were manually corrected and genotype results put in brackets (Universal Analysis Software section).

Furthermore, the manual revision of homozygous STR genotype calls that indicated the comment “interpretation threshold” (QC indicator) was performed as described above, except for alleles in stutter position.

On average, the results for the ancient samples FA10013B01A, FA10026B01A, FA10030T01A, and FA10058T01B demonstrated that MPS was able to recover 2 out of 152 markers (1.3%), 136 out of 151 markers (90.1%), 116 out of 128 markers (90.6%) and 147 out of 152 markers (96.7%), respectively (Table 3, Figure S3).

The lowest percentage of correctly typed loci was obtained for FA10013B01A dating from the seventh–eighth century showing drop-out rates of 98.7% (150 out of 152 markers failed amplification) (Tables 5 and S10–S12, Figure S3). Only identity SNPs (iSNP) rs1109037 (genotype: GG; all runs) and rs6444724 (genotype: CC; runs 0, 1, 3, 4, and 7) were successfully typed. Manual revision of non-called iSNP genotype calls was performed as described in “Universal Analysis Software” section for rs6444724 (genotype: CC; runs 5, 6, and 8), rs740598 (genotype: A; for runs 1, 6; genotype: G; for run 7) and rs964681 for run 1 (genotype: TT) (Table S13).

Table 3 Summary of ancient DNA analysis using the ForenSeq DNA Signature Prep Kit amplified for 27 autosomal STRs (aSTRs), 24 Y-STRs, 7 X-STRs, and 94 identity SNPs (iSNPs). The sample set consisted of four compromised samples dating from the fifth to eighth century (for more details see Table S10). Concordance was assessed by comparing genotype results between sequencing runs and to reference

profile (if available, see Tables S10–S12). Overall markers indicate the mean percentage (standard deviation [SD] in brackets) of concordantly typed markers for each ancient DNA sample. The average number of typed markers per sample was calculated based on the mean percentage of concordantly typed markers. The maximum number of markers using the ForenSeq DNA Signature Prep Kit (primer mix A) is 152

Sample	Successfully typed markers per MPS run (%)										Successfully typed markers over all runs			
	Marker	Run										Mean (SD) [%]	Mean no. of typed markers/sample	No. of typed markers vs. total no. of markers
		0	1	2	3	4	5	6	7	8				
FA10013B01A	<i>aSTRs</i>	0	0	0	0	0	0	0	0	0	0	0	/	2/152
	<i>Y-STRs</i>	0	0	0	0	0	0	0	0	0	0	0	/	
	<i>X-STRs</i>	0	0	0	0	0	0	0	0	0	0	0	/	
	<i>iSNPs</i>	2.1	2.1	1.1	2.1	2.1	1.1	1.1	2.1	1.1	1.7 (0.5)	2		
FA10026B01A	<i>aSTRs</i>	92.6	96.3	85.2	96.3	92.6	88.9	92.6	92.6	92.2 (3.4)	25	136/151 ^a		
	<i>Y-STRs</i>	79.2	83.3	66.7	83.3	83.3	79.2	75.0	79.2	78.7 (5.3)	19			
	<i>X-STRs</i>	71.4	71.4	57.1	71.4	71.4	71.4	57.1	71.4	68.2 (6.3)	5			
	<i>iSNPs</i>	93.6	94.6	92.5	93.6	92.5	93.6	92.5	93.6	93.5 (0.8)	87			
FA10030T01A	<i>aSTRs</i>	85.2	88.9	81.5	92.6	88.9	88.9	81.5	88.9	87.3 (3.8)	24	116/128 ^b		
	<i>Y-STRs</i>	♀	♀	♀	♀	♀	♀	♀	♀	/	/			
	<i>X-STRs</i>	85.7	85.7	71.4	85.7	85.7	71.4	85.7	85.7	82.5 (6.3)	6			
	<i>iSNPs</i>	92.6	92.6	86.2	93.6	92.6	92.6	91.5	93.6	92.0 (2.3)	86			
FA10058T01B	<i>aSTRs</i>	96.3	100	96.3	100	100	96.3	96.3	100	97.9 (2.0)	26	147/152		
	<i>Y-STRs</i>	100	100	95.8	100	100	100	95.8	100	99.1 (1.9)	24			
	<i>X-STRs</i>	85.7	85.7	85.7	85.7	85.7	85.7	85.7	85.7	85.7 (0.0)	6			
	<i>iSNPs</i>	96.8	97.8	95.7	97.9	96.8	96.8	98.9	97.9	96.8	97.3 (0.9)		91	

^a Exclusion of rs722290 from concordance calculation due to tri-allelic genotype; therefore, the total number of markers is 151 instead of 152

^b Sample derived from a female individual; therefore, the total number of markers recovered with the ForenSeq DNA Signature Prep Kit was 128 instead of 152

The DNA profile of FA10026B01A was concordant between all runs and to known allele calls derived from CE (if available), except for D3S1358 (CE: 15; MPS: 14, 15), D12S391 (CE: 19; MPS: 19, 19.3), D21S11 (CE: 30; MPS: 29, 30) and DYS385a-b (CE: allele 14; MPS: 11, 14) (Tables S10–S11). The careful revision of previously generated CE data for sample FA10026B01A indicated peaks at D3S1358, D12S391, D21S11 and DYS385 that were either located in stutter position and/or fell below the in-house analytical threshold of 50 RFU. Based on these findings, and, because of the comparison of MPS and CE data, the most likely reason for the observed differences is allele drop-out. Results for all runs at locus D13S317 revealed the presence of three different isometric variants of allele 11, viz. (a) sequence identical to the reference sequence as taken from the updated “Forensic STR Sequence Guide v4” file [65] originally published in [66] and available online at <https://strider.online> [67], (b) allele 11 containing rs9546005-T, and (c) allele 11 containing a C>A transversion at nucleotide position GRCh38 CHR13:82148040 together with rs9546005-T, respectively (Tables S10 and S14). Furthermore, run 3 showed two isomeric variants of DYS385a-b, allele 11 (a) sequence identical to the reference sequence as taken from [65, 67], and (b) allele 11 containing a G>A transition located at GRCh38 CHRY:18639770 (18639770-G showing 107 reads and 18639770-A showing 38 reads), respectively (Table S11). A possible reason for sequence variants in degraded samples is based on the misincorporation of nucleotides during enzymatic amplification due to cytosine deamination during DNA degradation [68–71]. One instance of discordance between sequencing runs was found for run 2 showing allelic drop-out of the longer allele at D19S443 (MPS: 13, 14) (Table S10). Identity SNP results for FA10026B01A were fully concordant between sequencing runs, except for rs1015250 and rs722290 (Table S13). Manual revision of rs1015250 no-call iSNP revealed a potential G allele for runs 3–7 showing reads that accounted for 20%, 16%, 16%, 18%, and 14% of the called iSNP-allele reads, respectively. However, non-called iSNP, allele G, was found for rs1015250 for runs 0–2 and 8 too with reads \leq 19. Data analysis revealed one tri-allelic genotype (iSNP genotype: C, G, (A)) at rs722290 for runs 0, 1 and 7. Therefore, rs722290 was excluded from concordance evaluation (Table S13). Observed reads for rs722290, allele A, added up to 4.6%, 4.2%, and 5.6% of allele-reads with highest intensity for runs 0, 1, and 7, respectively (data not shown). Locus drop-out was found for all runs at rs13182883, rs354439, rs719366, and at rs13218440 (runs 0, 2, 3, 5, and 6), rs1736442 (runs 0, 2, 4 and 5), and rs7041158 (run 2) (Table S13). The average mean over all runs for FA10026B01A was 92.2% (3.4%),

78.7% (5.3%), 68.2% (6.3%), and 93.5% (0.8%) of typed aSTRs, Y-STRs, X-STRs and iSNPs, respectively (Table 3, Figure S3).

Results for FA10030T01A were fully concordant between runs and to known reference (if available), except for D7S820 (CE: 8; MPS: 8, 10), D19S433 (CE: 12; MPS: 12, 13). Results for D22S1045 revealed the drop-out of the longer allele within all runs, except for run 3 (CE: 16, 17) (Table S10). In addition, all runs showed drop-out of the shorter allele at DXS8378 (CE: 11, 13) (Table S12). According to previously obtained CE results and to [34], human remains of FA10030T01A were derived from a female individual. Data analysis revealed one Y-STR call at DYS391, allele 11, most likely due to contamination during handling of the sample or laboratory procedure (Table S11). However, the total number of allele-reads for FA10030T01A at DYS391, allele 11, was \leq 49 (Table S11). Identity SNP results for FA10030T01A were fully concordant between runs, except for rs1294331, rs354439, and rs1382387 showing the drop-out of allele A, allele T and allele G, respectively (Table S13). Locus drop-out was observed at rs2920816 (all runs), rs719366 (runs 0, 2–8), rs13182883 (runs 0–2, 4, 6–8) and rs13218440 (runs 1–6, 8). Overall, profile of FA10030T01A consisted of 87.3% (3.8%), 82.5% (6.3%), and 92.0% (2.3%) concordantly typed aSTRs, X-STRs, and iSNPs, respectively (Table 3, Figure S3).

The highest percentage of correctly typed loci was found for FA10058T01B, which was the eldest aDNA sample, dating from the fifth century showing 97.9% (2.0%), 99.1% (1.9%), 85.7% (0.0%), and 97.3% (0.9%) of concordantly typed aSTRs, Y-STRs, X-STRs, and iSNPs, respectively. Results were fully concordant between runs and to known reference (if available) for all 152 markers (Tables S10–S13, Figure S3). Drop-out was obtained for runs 2 and 5 at rs13218440 (Table S13). In addition, MPS was able to identify homozygous genotype as isometric heterozygous genotype for FA10058T01B at D8S1179, allele 13 (repeat structure variants [TCTA] [TCTG] [TCTA]₁₁ and [TCTA]₁₃ containing a G>A transition located at GRCh38 CHR8:124894872) and DYF387S1, allele 38 (repeat structure variants [AAAG]₃ [GTAG] [GAAG]₄ [AAAG]₂ [GAAG] [AAAG]₂ [GAAG]₉ [AAAG]₈ [AAAA] [AAAG]₇ and [AAAG]₃ [GTAG] [GAAG]₄ [AAAG]₂ [GAAG] [AAAG]₂ [GAAG]₁₀ [AAAG]₇ [AAAA] [AAAG]₇ both containing a G>A transition located at GRCh38 CHRY:23785484) (Table S14) [65, 67]. Earlier studies reported sequence variants at D8S1179 [4, 15, 72, 73], which was in line with our findings for this marker. However, at the time of writing, no references were found for DYF387S1. Discordances between sequencing runs were observed for three iSNPs at rs1355366, rs13182883, and rs13218440 due to the drop-out of allele G, allele A and allele A, respectively (Table S13). During manual revision of non-called iSNPs we observed reads that were below 1.8% of the called iSNP-allele reads at rs251934 (allele T) and rs1821380

(allele A) and were thus considered as noise. Runs 2 and 5 obtained locus drop-out at rs13218440 (Table S13).

In general, our results support the findings of earlier publications [74–76] that the application of MPS for STR analysis of degraded and challenging DNA samples can be beneficial. As expected and due to the nature of SNP-containing amplicons [25, 26], the ancient DNA data showed that, in contrast to STRs, the majority of sequencing runs obtained constantly high SNP recovery rates, which was in line with previous reports [8, 75] (Table 3). To evaluate the validity of the obtained results and to assess the impact of iSNPs vs. aSTRs we calculated the RMPs and LR_s for each aDNA sample (if applicable) as well as for reference material SRM 2391c, components A–C (Tables S15–S18). RMPs for iSNPs were calculated from a subset of 49 iSNP that are included in the validated SNPforID 52-plex panel [77]. Note that rs1355366 was excluded for RMP calculation due to the ambiguous genotype results. The calculated RMPs for 27 aSTRs and 49 iSNPs for aDNAs ranged from 1.65E-32 to 7.15E-27 and from 1.37E-29 to 8.55E-28, respectively (Tables S15–S17), whereas the corresponding RMP values for SRM 2391c A–C ranged from 3.54E-43 to 2.48E-34 (aSTRs) and from 9.89E-31 to 1.40E-28 (iSNPs) (Table S18). As expected, the data of the reference material SRM 2391c A–C showed clearly higher LR levels for aSTRs than for the iSNP subset (Figure S4) [52, 78–80]. In the aDNA samples, however, the STR-based LR_s were roughly in the same range as those obtained for iSNPs. Furthermore, there was no pronounced difference in iSNPs LR levels between reference and aDNA samples (Figure S4) [25, 52]. This relative decline of the STR-based LR_s compared to the constant levels of SNP-based LR_s can be an indication for a noticeably better performance of SNPs compared to STRs when analysing old and compromised DNA samples with MPS approaches. However, it is important to mention that the LR for SRM 2391cB was markedly higher than those of components A and C representing an above average LR value and that only a small number of compromised samples were included. Therefore, these data can be regarded as preliminary results indicating that SNPs are a preferable choice of markers, especially for heavily degraded DNA. Nevertheless, this has to be shown in more comprehensive MPS-studies.

Additionally, we contrasted CE-based with MPS-based aDNA genotyping. For this comparison, we calculated LR_s (if applicable) for the AmpFISTR NGM Select Kit, the AmpFISTR NGM Select loci as included in the ForenSeq DNA Signature Prep Kit and for the entire set of aSTR as well as for a subset of 49 iSNP loci included in the latter kit (Tables S15–S17). The results clearly show the benefit of the large number of markers included in the MPS-based assay that resulted in much higher LR_s than CE did (Tables S15–S17). In general, we observed comparable MPS-based LR_s for all aDNA samples and classes of markers suggesting stable performance of the ForenSeq DNA Signature Prep Kit on compromised samples (Figure S5). Overall, the predominant reason for STR discordance between

MPS and CE results within aDNA samples was due to additional allele calls obtained using MPS that lead to an increased power of discrimination, except for FA10013B01A, which showed total drop-out in MPS analysis but a partial profile using CE (Tables S10–S12). Possible reasons could be that the CE analysis was performed eight years earlier and the remaining DNA extract meanwhile frozen, stochastic effects due to low DNA input in general (Table S10) or inhibition of the MPS approach as described in [28, 81]. For instance, using MPS the average profile of FA10026B01A and FA10058T01B was composed of 25 aSTRs, 19 Y-STRs, 5-XSTRs and 87 iSNPs and 26 aSTRs, 24 Y-STRs, 6 X-STRs and 91 iSNPs, respectively (Table 3, Figure S5). However, CE analysis was able to recover 4 aSTRs, 5 Y-STRs and 15 aSTRs, 12 Y-STRs plus 1 X-STR for FA10026B01A and FA10058T01B, respectively (Tables S10–S11, Figure S5). Therefore, it seems important to note that such high numbers of different loci cannot be typed with CE-based STR kits within a single run. This indicates the usefulness of MPS-based technologies for the analysis of compromised samples especially if the original material is limited.

Conclusions

The presented results are the first MPS-data collected in a collaborative exercise performed among eight independent European laboratories using sequencing instruments of the same supplier. This inter-laboratory study conducted in the framework of the SeqForSTRs project [31] described validation experiments for MPS STR analysis using a standardized sequencing library prepared with the ForenSeq DNA Signature Prep Kit and sequenced on eight different MiSeq FGx instruments. The primary intention of this study was to reduce the inherent complexity derived from multiple steps during manual library preparation to a minimum to test for instrument variation about which knowledge is scarce (including the effect of transport conditions). This was enabled by centralized library preparation using the ForenSeq DNA Signature Prep Kit (primer mix A). Despite broad observed variation in instrument performance, all laboratories obtained run quality metrics that fell within the manufacturer's recommended range. Importantly, obtained locus coverage differences did not necessarily adversely affect heterozygous balances. Inter-laboratory concordance revealed 100% concordance for autosomal- and Y-STRs and yet 85.7% for X-STRs due to drop-out of one allele plus 83.3% due to locus drop-out at marker DXS10103. Mean success rates of all sensitivity runs with 125 pg DNA input were 96.9% (1.7%), 93.2% (2.8%), and 90.5% (2.5%) for aSTR, Y-STR, and X-STR alleles and were comparable to [1, 3, 53, 54] as well as to results using MPS-kits from other suppliers [9, 82].

Interestingly, sensitivity results for each particular run showed that especially for DNA input amounts of < 125 pg the ability to correctly type a STR profile might be dependent on the particular detection limit of the sequencing instrument. Sensitivity results showed that between laboratories the variation in successfully typed aSTRs (highest vs. lowest number of correctly typed alleles) increased with decreasing DNA input and added up to 9% (62 pg) and 17% (31 pg), respectively. These results indicate the importance of performing rigorous in-house validation before implementing MPS into forensic casework applications. Results showed that MPS is a promising tool for the identification of compromised DNA samples and were in line with [8, 75]. The analysis of degraded DNA samples using the ForenSeq DNA Signature Prep Kit showed the total drop-out of FA10013B01A; however, MPS was able to successfully type $\geq 87\%$ of all aSTRs, $\geq 78\%$ of all Y-STRs, $\geq 68\%$ of all X-STRs, and $\geq 92\%$ of all iSNPs for the remaining three compromised DNA samples. Again, this high number of successfully typed loci was not achieved using conventional CE-based technologies.

Acknowledgements The authors would like to thank Burkhard Berger, Harald Niederstätter, and Martin Bodner for helpful discussions, as well as Alexandra Kaindl-Lindinger (all Institute of Legal Medicine, Medical University of Innsbruck) and Angelika Fürst (Bavarian State Criminal Police Office, Munich) for technical support.

Funding information Open access funding provided by University of Innsbruck and Medical University of Innsbruck. This study received support from the European Union, grant agreement number IZ25-5793-2015-30 2017-2020 - Evaluierung des Next Generation Sequencing (NGS) für die Anwendbarkeit zur DNA-Analytik in der polizeilichen Praxis von Bund und Ländern – Sequencing of forensic STRs (SeqForSTRs) and grant agreement number HOME/2014/ISFP/AG/LAWX/4000007135 under the Internal Security Funding Police programme of the European Commission-Directorate General Justice and Home Affairs (DNASeqEx).

Compliance with ethical standards Research involving human participants and/or animals: does not apply; the study is using standard reference material and DNA extracted from old human remains (> 800 years);

Conflict of interest The authors declare that they have no conflict of interest.

Disclaimer This publication reflects the views only of the authors, and the European Commission cannot be held responsible for any use, which may be made of the information contained therein.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Jäger AC, Alvarez ML, Davis CP, Guzman E, Han Y, Way L et al (2017) Developmental validation of the MiSeq FGx Forensic Genomics System for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Sci Int Genet* 28:52–70
- Churchill JD, Chang J, Ge J, Rajagopalan N, Wootton SC, Chang CW, Lagacé R, Liao W, King JL, Budowle B (2015) Blind study evaluation illustrates utility of the Ion PGM system for use in human identity DNA typing. *Croat Med J* 56:218–229
- Churchill JD, Schmedes SE, King JL, Budowle B (2016) Evaluation of the Illumina Beta Version ForenSeq DNA Signature Prep Kit for use in genetic profiling. *Forensic Sci Int Genet* 20:20–29
- Müller P, Alonso A, Barrio PA, Berger B, Bodner M, Martin P, Parson W, DNASEQEX Consortium (2018) Systematic evaluation of the early access applied biosystems precision ID Globalfiler mixture ID and Globalfiler NGS STR panels for the ion S5 system. *Forensic Sci Int Genet* 36:95–103
- Van Neste C, Van Nieuwerburgh F, Van Hoofstat D, Deforce D (2012) Forensic STR analysis using massive parallel sequencing. *Forensic Sci Int Genet* 6:810–818
- Scheible M, Loreille O, Just R, Irwin J (2014) Short tandem repeat typing on the 454 platform: strategies and considerations for targeted sequencing of common forensic markers. *Forensic Sci Int Genet* 12:107–119
- Scheible M, Loreille O, Just R, Irwin J (2011) Short tandem repeat sequencing on the 454 platform. *Forensic Sci Int Genet Suppl Ser 3*: e357–e3e8
- Xavier C, Parson W (2017) Evaluation of the Illumina ForenSeq DNA Signature Prep Kit - MPS forensic application for the MiSeq FGx benchtop sequencer. *Forensic Sci Int Genet* 28:188–194
- Guo F, Zhou Y, Liu F, Yu J, Song H, Shen H, Zhao B, Jia F, Hou G, Jiang X (2016) Evaluation of the Early Access STR Kit v1 on the Ion Torrent PGM platform. *Forensic Sci Int Genet* 23:111–120
- Strobl C, Eduardoff M, Bus MM, Allen M, Parson W (2018) Evaluation of the precision ID whole MtDNA genome panel for forensic analyses. *Forensic Sci Int Genet* 35:21–25
- Eduardoff M, Gross TE, Santos C, de la Puente M, Ballard D, Strobl C, Børsting C, Morling N, Fusco L, Hussing C, Egyed B, Souto L, Uacyisrael J, Syndercombe Court D, Carracedo Á, Lareu MV, Schneider PM, Parson W, Phillips C, EUROFORGEN-NoE Consortium (2016) Inter-laboratory evaluation of the EUROFORGEN Global ancestry-informative SNP panel by massively parallel sequencing using the Ion PGM. *Forensic Sci Int Genet* 23:178–189
- Just RS, Moreno LI, Smerick JB, Irwin JA (2017) Performance and concordance of the ForenSeq system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens. *Forensic Sci Int Genet* 28:1–9
- Köcher S, Müller P, Berger B, Bodner M, Parson W, Roewer L, Willuweit S, DNASEqEx Consortium (2018) Inter-laboratory validation study of the ForenSeq DNA Signature Prep Kit. *Forensic Sci Int Genet* 36:77–85
- Moreno LI, Galusha MB, Just R (2018) A closer look at Verogen's ForenSeq DNA Signature Prep kit autosomal and Y-STR data for streamlined analysis of routine reference samples. *Electrophoresis* 39:2685–2693
- Devesse L, Ballard D, Davenport L, Riethorst I, Mason-Buck G, Syndercombe CD (2017) Concordance of the ForenSeq system and characterisation of sequence-specific autosomal STR alleles across two major population groups. *Forensic Sci Int Genet* 34:57–61
- Phillips C, Devesse L, Ballard D, Weert L, MI P, Melis S et al (2018) Global patterns of STR sequence variation: sequencing the

- CEPH human genome diversity panel for 58 forensic STRs using the Illumina ForenSeq DNA Signature Prep Kit. *Electrophoresis* 39:2708–2724
17. Churchill JD, Novroski NMM, King JL, Seah LH, Budowle B (2017) Population and performance analyses of four major populations with Illumina's FGx Forensic Genomics System. *Forensic Sci Int Genet* 30:81–92
 18. Gettings KB, Borsuk LA, Steffen CR, Kiesler KM, Vallone PM (2018) Sequence-based U.S. population data for 27 autosomal STR loci. *Forensic Sci Int Genet* 37:106–115
 19. Novroski NMM, King JL, Churchill JD, Seah LH, Budowle B (2016) Characterization of genetic sequence variation of 58 STR loci in four major population groups. *Forensic Sci Int Genet* 25: 214–226
 20. van der Gaag KJ, de Leeuw RH, Hoogenboom J, Patel J, Storts DR, Laros JFJ et al (2016) Massively parallel sequencing of short tandem repeats—Population data and mixture analysis results for the PowerSeq system. *Forensic Sci Int Genet* 24:86–96
 21. Faith SA, Scheible M. Analyzing data from next generation sequencers using the PowerSeq Auto/Mito/Y System. Promega Corporation Web site. <http://at.promega.com/resources/profiles-in-dna/2016/analyzing-data-from-next-generation-sequencers-using-the-powerseq-automitoy-system/> (Updated 2016). 2016.
 22. Illumina. ForenSeq DNA Signature Prep Kit, Reference Guide (Part # 15049528 v01). 2015.
 23. ThermoFisherScientific. Precision ID GlobalFiler NGS STR Panel v2 with the HID Ion S5/HID Ion GeneStudio S5 System: template preparation and sequencing, Quick Reference (Pub. No. MAN0016252, Rev. B.0). 2018.
 24. ThermoFisherScientific (2018) Expand your forensics workflow with the Precision ID NGS System, Brochure
 25. Butler JM, Coble MD, Vallone PM (2007) STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol* 3:200–205
 26. Brookes AJ (1999) The essence of SNPs. *Gene* 234:177–186
 27. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SL, Karoma NJ et al (2006) Developing a SNP panel for forensic identification of individuals. *Forensic Sci Int* 164:20–32
 28. Elwick K, Zeng X, King J, Budowle B, Hughes-Stamm S (2018) Comparative tolerance of two massively parallel sequencing systems to common PCR inhibitors. *Int J Legal Med* 132:983–995
 29. Zhang Q, Zhou Z, Liu Q, Liu L, Shao L, Zhang M, Ding X, Gao Y, Wang S (2018) Evaluation of the performance of Illumina's ForenSeq system on serially degraded samples. *Electrophoresis* 39:2674–2684
 30. Zeng X, Elwick K, Mayes C, Takahashi M, King JL, Gangitano D et al (2018) Assessment of impact of DNA extraction methods on analysis of human remain samples on massively parallel sequencing success. *Int J Legal Med*
 31. Bundeskriminalamt, Fachbereich-IZ-25, Berlin, Germany. EU Project ISF, Nr. IZ25-5793-2015-30 2017-2020 Evaluierung des Next Generation Sequencing (NGS) für die Anwendbarkeit zur DNA-Analytik in der polizeilichen Praxis von Bund und Ländern – Sequencing of forensic STRs (SeqforSTRs)
 32. Illumina (2015) MiSeq FGx Instrument, Reference Guide (Part# 15050524, Rev. C)
 33. NationalInstituteofStandards&Technology. Standard Reference Material 2391c PCR-based DNA profiling standard (Certificate of Analysis). 2011, Certificate Revision History: 03 April 2015
 34. McGlynn G (2007) Using 13C-, 15N-, and 18O stable isotope analysis of human bone tissue to identify transhumance, high altitude habitation and reconstruct palaeodiet for the early medieval Alpine population at Volders. Ludwig-Maximilians-University at München, Austria
 35. Bauer CM, Niederstätter H, McGlynn G, Stadler H, Parson W (2013) Comparison of morphological and molecular genetic sex-typing on mediaeval human skeletal remains. *Forensic Sci Int Genet* 7:581–586
 36. Walker JA, Hedges DJ, Perodeau BP, Landry KE, Stoilova N, Laborde ME, Shewale J, Sinha SK, Batzer MA (2005) Multiplex polymerase chain reaction for simultaneous quantitation of human nuclear, mitochondrial, and male Y-chromosome DNA: application in human identification. *Anal Biochem* 337:89–97
 37. Niederstätter H, Köchl S, Grubwieser P, Pavlic M, Steinlechner M, Parson W (2007) A modular real-time PCR concept for determining the quantity and quality of human nuclear and mitochondrial DNA. *Forensic Sci Int Genet* 1:29–34
 38. Promega. PowerPlex ESX 17 Fast System for use on the Applied Biosystems Genetic Analyzers, Technical Manual. September, 17
 39. Promega. PowerPlex 16 System, Technical Manual. Revised 5/16
 40. Promega. PowerPlex 21 System for use on the Applied Biosystems Genetic Analyzers, Technical Manual. Revised 4/17
 41. Qiagen (2013) Investigator Argus X-12 Handbook
 42. ThermoFisherScientific. AppliedBiosystems AmpFISTR Yfiler PCR Amplification Kit, User Guide (Pub.no. 4358101, Rev. J). 2014:234
 43. Esteve Codina A, Niederstatter H, Parson W (2009) “GenderPlex” a PCR multiplex for reliable gender determination of degraded human DNA samples and complex gender constellations. *Int J Legal Med* 123:459–464
 44. ThermoFisherScientific (2015) AppliedBiosystems AmpFISTR NGM Select PCR Amplification Kit, User Guide (Pub. no. 4458841, Rev. F)
 45. Illumina (2016) ForenSeq Universal Analysis Software Guide (Document # 15053876 v01)
 46. IBM-Corp (2016) IBM SPSS Statistics for Windows, Version 24.0. Armonk
 47. Amigo J, Salas A, Phillips C, Carracedo A (2008) SPSmart: adapting population based SNP genotype databases for fast and comprehensive web access. *BMC Bioinformatics* 9:428
 48. Amigo J, Phillips C, Salas T, Formoso L, Carracedo A, Lareu M (2009) pop.STR—An online population frequency browser for established and new forensic STRs. *Forensic Sci Int Genet Suppl Ser* 2:361–362
 49. Amigo J, Phillips C, Lareu M, Carracedo A (2008) The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project. *Int J Legal Med* 122:435–440
 50. Amigo J, Phillips C, Salas A, Carracedo A (2009) Viability of in-house datamarting approaches for population genetics analysis of SNP genotypes. *BMC Bioinformatics* 10(Suppl 3):S5
 51. National Research C (1996) The evaluation of forensic DNA evidence. The National Academies Press, Washington, DC
 52. Butler JM (2005) Forensic DNA typing: biology, technology, and genetics of STR markers, 2nd edn. Elsevier Academic Press, New York
 53. Guo F, Yu J, Zhang L, Li J (2017) Massively parallel sequencing of forensic STRs and SNPs using the Illumina ForenSeq DNA Signature Prep Kit on the MiSeq FGx Forensic Genomics System. *Forensic Sci Int Genet* 31:135–148
 54. Silvia AL, Shugarts N, Smith J (2017) A preliminary assessment of the ForenSeq FGx System: next generation sequencing of an STR and SNP multiplex. *Int J Legal Med* 131:73–86
 55. Illumina (2016) Optimizing cluster density on Illumina Sequencing Systems - understanding cluster density limitations and strategies for preventing under- and overclustering
 56. Gill P, Sparkes R, Fereday L, Werrett DJ (2000) Report of the European Network of Forensic Science Institutes (ENSFI): formulation and testing of principles to evaluate STR multiplexes. *Forensic Sci Int* 108:1–29
 57. Almalki N, Chow HY, Sharma V, Hart K, Siegel D, Wumbach E (2017) Systematic assessment of the performance of illumina's MiSeq FGx forensic genomics system. *Electrophoresis* 38:846–854

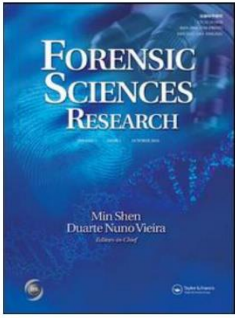
58. Martin P, de Simon LF, Luque G, Farfan MJ, Alonso A (2014) Improving DNA data exchange: validation studies on a single dye STR kit with 24 loci. *Forensic Sci Int Genet* 13:68–78
59. Oostdik K, Lenz K, Nye J, Schelling K, Yet D, Bruski S, Strong J, Buchanan C, Sutton J, Linner J, Frazier N, Young H, Matthies L, Sage A, Hahn J, Wells R, Williams N, Price M, Koehler J, Staples M, Swango KL, Hill C, Oyerly K, Duke W, Katzilierakis L, Ensenberger MG, Bourdeau JM, Sprecher CJ, Krenke B, Storts DR (2014) Developmental validation of the PowerPlex Fusion System for analysis of casework and reference samples: a 24-locus multiplex for new database standards. *Forensic Sci Int Genet* 12:69–76
60. Ensenberger MG, Lenz KA, Matthies LK, Hadinoto GM, Schienman JE, Przech AJ, Morganti MW, Renstrom DT, Baker VM, Gawrys KM, Hoogendoorn M, Steffen CR, Martin P, Alonso A, Olson HR, Sprecher CJ, Storts DR (2016) Developmental validation of the PowerPlex Fusion 6C System. *Forensic Sci Int Genet* 21:134–144
61. Wang DY, Gopinath S, Lagace RE, Norona W, Hennessy LK, Short ML et al (2015) Developmental validation of the GlobalFiler Express PCR Amplification Kit: a 6-dye multiplex assay for the direct amplification of reference samples. *Forensic Sci Int Genet* 19:148–155
62. Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, Rothämel T, Schmitter H, Schneider H, Brinkmann B (2009) The German Stain Commission: recommendations for the interpretation of mixed stains. *Int J Legal Med* 123:1–5
63. Butler JM, Kline MC, Coble MD (2018) NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned. *Forensic Sci Int Genet* 37:81–94
64. Marciniak S, Klunk J, Devault A, Enk J, Poinar HN (2015) Ancient human genomics: the methodology behind reconstructing evolutionary pathways. *J Hum Evol* 79:21–34
65. Phillips C, Gettings KB, King JL, Ballard D, Bodner M, Borsuk L, Parson W (2018) "The devil's in the detail": Release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide. *Forensic Sci Int Genet* 34:162–169
66. Parson W, Ballard D, Budowle B, Butler JM, Gettings KB, Gill P, Gusmão L, Hares DR, Irwin JA, King JL, Knijff P, Morling N, Prinz M, Schneider PM, Neste CV, Willuweit S, Phillips C (2016) Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Sci Int Genet* 22:54–63
67. Bodner M, Bastisch I, Butler JM, Fimmers R, Gill P, Gusmão L et al (2016) Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). *Forensic Sci Int Genet* 24:97–102
68. Dabney J, Meyer M, Pääbo S (2013) Ancient DNA damage. *Cold Spring Harb Perspect Biol* 5
69. Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res* 29:4793–4799
70. Stiller M, Green RE, Ronan M, Simons JF, Du L, He W et al (2006) Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proc Natl Acad Sci U S A* 103:13578–13584
71. Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prufer K et al (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* 104:14616–14621
72. Gettings KB, Aponte RA, Vallone PM, Butler JM (2015) STR allele sequence variation: current knowledge and future issues. *Forensic Sci Int Genet* 18:118–130
73. Gettings KB, Kiesler KM, Faith SA, Montano E, Baker CH, Young BA, Guerrieri RA, Vallone PM (2016) Sequence variation of 22 autosomal STR loci detected by next generation sequencing. *Forensic Sci Int Genet* 21:15–21
74. Calafell F, Anglada R, Bonet N, Gonzalez-Ruiz M, Prats-Munoz G, Rasal R et al (2016) An assessment of a massively parallel sequencing approach for the identification of individuals from mass graves of the Spanish Civil War (1936–1939). *Electrophoresis* 37:2841–2847
75. Fattorini P, Previderé C, Carboni I, Marrubini G, Sorçaburu-Cigliero S, Grignani P, Bertoglio B, Vatta P, Ricci U (2017) Performance of the ForenSeq™ DNA Signature Prep kit on highly degraded samples. *Electrophoresis* 38:1163–1174
76. Kulstein G, Hadrys T, Wiegand P (2018) As solid as a rock—comparison of CE- and MPS-based analyses of the petrosal bone as a source of DNA for forensic identification of challenging cranial bones. *Int J Legal Med* 132:13–24
77. Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M et al (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27:1713–1724
78. Gill P (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med* 114:204–210
79. Gill P, Werrett DJ, Budowle B, Guerrieri R (2004) An assessment of whether SNPs will replace STRs in national DNA databases—joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDM). *Sci Justice* 44:51–53
80. Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B (1999) The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis* 20:1682–1696
81. Sidstedt M, Steffen CR, Kiesler KM, Vallone PM, Radstrom P, Hedman J (2019) The impact of common PCR inhibitors on forensic MPS analysis. *Forensic Sci Int Genet* 40:182–191
82. Wang Z, Zhou D, Wang H, Jia Z, Liu J, Qian X et al (2017) Massively parallel sequencing of 32 forensic markers using the Precision ID GlobalFiler NGS STR Panel and the Ion PGM System. *Forensic Sci Int Genet* 31:126–134

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Publikation 3

Bredemeyer S, Roewer L, Willuweit S, Next Generation Sequencing of Y-STRs in father-son pairs and comparison with traditional capillary electrophoresis. *Forensic Sciences Research*, 2021. p. 1-6

<https://doi.org/10.1080/20961790.2021.1898078>



Next generation sequencing of Y-STRs in father-son pairs and comparison with traditional capillary electrophoresis

Steffi Bredemeyer, Lutz Roewer & Sascha Willuweit

To cite this article: Steffi Bredemeyer, Lutz Roewer & Sascha Willuweit (2021): Next generation sequencing of Y-STRs in father-son pairs and comparison with traditional capillary electrophoresis, Forensic Sciences Research, DOI: [10.1080/20961790.2021.1898078](https://doi.org/10.1080/20961790.2021.1898078)

To link to this article: <https://doi.org/10.1080/20961790.2021.1898078>



© 2021 The Author(s). Published by Taylor & Francis Group on behalf of the Academy of Forensic Science.



View supplementary material [↗](#)



Published online: 16 Apr 2021.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Next generation sequencing of Y-STRs in father-son pairs and comparison with traditional capillary electrophoresis

Steffi Bredemeyer, Lutz Roewer and Sascha Willuweit

Department of Forensic Genetics, Institute of Legal Medicine and Forensic Sciences, Charité – Universitätsmedizin Berlin, Berlin, Germany

ABSTRACT

To evaluate the promising advantages of massively parallel sequencing (MPS) in our casework, we analysed a total of 33 Y-chromosomal short tandem repeats (Y-STRs) with traditional capillary electrophoresis (CE) and 25 Y-STRs using the newer MPS technology. We studied the outcome of both technologies in 64 father-son pairs using stock and custom-designed kits. Current MPS technology confirmed the 13 mutational events observed with CE and improved our understanding of the complex nature of STR mutations. By detecting isometric sequence variants between unrelated males, we show that sequencing Y-STRs using MPS can boost discrimination power.

ARTICLE HISTORY

Received 4 December 2020
Accepted 28 February 2021

KEYWORDS

Forensic sciences; forensic genetics; Y-STRs; rapidly mutating Y-STRs; sequence polymorphism; isometric sequence variants; massively parallel sequencing

Introduction

Y-chromosomal short tandem repeats (Y-STRs) are especially helpful to analyse the male proportion of complex female-male mixtures, which is a typical scenario of samples in sexual assault cases. While autosomal STRs can easily be used to identify individuals, Y-STR profiles tend to be identical amongst patrilineal relatives because of their non-recombining inheritance mode along paternal lineages. Therefore, a male perpetrator most likely cannot be differentiated from his brother or father using Y-STRs [1]. While this feature of Y-STRs is detrimental when trying to differentiate paternally related individuals, it is beneficial for analysing deep-rooted pedigrees [2] and familial searching [3].

Currently, the standard technique used in forensic genetics is to type STRs with capillary electrophoresis (CE) analysis. This analysis method is fast (profiles can be generated within one day) and affordable for the average crime laboratory. Furthermore, the resulting data handling is straightforward, as STR length alleles are defined by the number of repetitive motifs by most standard software packages. This enables practitioners to report and compare results very easily. In contrast, massively parallel sequencing (MPS) is more time-consuming (at least two days hands-on in the laboratory and approximately 30 h

sequencing time) and expensive, depending on how many samples are pooled per run [4]. However, MPS has some advantages compared with CE analysis: (1) the lack of specific shared fluorescent markers makes it possible to multiplex more than 100 hundred markers in one reaction instead about 30 markers with CE analysis [5–7] and (2) MPS reveals the actual sequence of the variation rather than just its length, and is therefore able to detect sequence variants within alleles (isometric variants), which are not distinguishable with CE analysis. Sequence variants can appear as intra-repeat single nucleotide polymorphisms (SNPs), SNPs within the flanking region, indels, or repeat pattern variants. Using these sequence variants, two individuals could be distinguished with one marker even if they have the same nominal allele [8] and are thus indistinguishable using CE. Various studies have shown an increase in the number of alleles per locus when these isometric variants are analysed by MPS, which results in a higher genetic diversity of STRs [8–22]. Additionally, sequencing can cover flanking sequences and is thus able to detect SNPs that are not detectable with CE [7,9,23,24]. However, SNPs have a relatively low mutation rate of 10^{-8} per base per generation [25] and are therefore unlikely to contribute to the separation of male relatives.

CONTACT Steffi Bredemeyer steffi.bredemeyer@gmail.com

Supplemental data for this article are available online at <https://doi.org/10.1080/20961790.2021.1898078>

© 2021 The Author(s). Published by Taylor & Francis Group on behalf of the Academy of Forensic Science.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To determine if Y-STR sequence variants may help to improve the differentiation of paternally-related males in forensic casework, we compared the CE and MPS analysis methods based on mutations at 33 Y-STRs and 25 Y-STR markers of 64 father-son pairs.

Materials and methods

Sample preparation

We collected buccal swap samples from 64 father-son pairs from European populations. Our analyses are based on self-declared family relations. DNA extraction was performed on the BioRobot® EZ1 (Qiagen, Hilden, Germany) using the EZ1 DNA Blood 200µL Kit (Qiagen). Quantification of DNA samples was performed with the Rotor-Gene Q (Qiagen) and the Investigator Quantiplex Kit (Qiagen) or with the Quantus™ Fluorometer (Promega Corporation, Madison, WI, USA) and the QuantiFluor® ONE dsDNA Kit (Promega Corporation).

CE analysis

A total of 33 Y-STR markers were amplified by combining 22 Y-STR markers of the PowerPlex® Y23 System (Promega Corporation) and 11 Y-STR markers of an adapted RM-Yplex assay (13 Y-STRs in total) based on Alfghafri et al. [26]. The modified primer mix can be found in [Supplementary File S1](#). Instead of the DYS518 marker that was used by Alfghafri et al. [26], the marker DYS464 was added to the multiplex PCR. This multi-copy Y-STR marker with at least 4 copies on the Y chromosome was selected for its outstanding diversity [27,28]. The primer sequence for DYS464 was taken from Redd et al. [29] and labelled with ATTO 550. PCR amplification was performed in a final volume of 15µL per reaction: 7.5µL Platinum™ Multiplex PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA), 3.3µL primer mix according to [Supplementary File S1](#), 3.2µL PCR grade water, and 1µL DNA template (1 ng/µL). To confirm a duplication observed within the marker Y-GATA-H4 in one sample we performed an additional PCR using the Yfiler™ Plus PCR Amplification Kit (Thermo Fisher Scientific) with this sample and the sample of the associated son. CE analysis was carried out on the 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and data were analysed using the GeneMapper IDX version 1.4 software (Thermo Fisher Scientific).

MPS analysis

Library preparation for sequencing analysis was performed with the ForenSeq™ DNA Signature

Prep Kit (Verogen, San Diego, CA, USA) according to the manufacturer's instructions. For PCR amplification, primer mix A was used, which includes 25 Y-STR markers. All samples were sequenced with the MiSeq FGx (Illumina, San Diego, CA, USA) using the MiSeq FGx Reagent Kit (Verogen). For better coverage, the library input volume was increased from 7µL to 10µL. Data analysis was done using an in-house analysis pipeline (as explained below), allowing us to analyse the marker DYS456, which is included in the primer mix of the ForenSeq™ DNA Signature Prep Kit but excluded from the ForenSeq™ Universal Analysis Software (UAS, Verogen). Allele-stutter differentiation was done following the forward-, back-, and double-back-stutter models given by Bright et al. [30] and the vendor-provided expected stutter percentages (same for UAS and the in-house pipeline). All results with less than 10 reads were discarded.

In-house analysis pipeline

To be able to assess the analysis done by the UAS software, we re-analysed the raw FASTQ files using a bioinformatic pipeline called “wintermute” (open-source software available online at <https://github.com/545ch4/wintermute>, publication in preparation). This tool was prototyped as part of the EU funded DNaseqEx (DNA-STR Massive Sequencing & International Information Exchange) project, which aimed to enable the user to understand the variability and composition of the huge number of single reads/sequences within a MPS FASTQ file. In the first step, the software assigns each read to one or more targets (derived from the target primer that was used). All sequences of those sequence-buckets are then multi-aligned to generate one or more consensus sequence that represents the sequence-bucket. The software does not generate reliable results for the flanking regions of the Y-STRs used here. As the sequence quality declines over read length (approx. beginning at 120 bp), the software was unable to generate merged flanking sequences of longer Y-STR (e.g. DYS448 and DYS389II). Therefore, we decided to exclude all flanking sequences for reasons of consistency.

Results

We used CE analysis to examine 22 Y-STR markers of 64 father-son pairs using the PowerPlex® Y23 System and used MPS analysis for 25 Y-STRs of the same sample set using the ForenSeq™ DNA Signature Prep Kit. DYS393 and DYS458 are included in the

PowerPlex® Y23 System, but not in the ForenSeq™ DNA Signature Prep Kit. DYF387S1, DYS460, DYS505, DYS522, and DYS612 are included in the ForenSeq™ DNA Signature Prep Kit, but not in the PowerPlex® Y23 System. In total, 20 Y-STR markers were analysed using both kits (Supplementary File S2). Comparing MPS results with CE results, all allele designations were concordant with the exception of one drop-out (in DYS392) because of low coverage of this sample. There were some stutters above the stutter filter, namely in DYS385ab, DYS392, DYS456, and DYS576, and some additional sequence variants (drop-ins) slightly above a peak height ratio (PHR) of 10% in DYS437 and DYS448. All drop-ins were below a PHR of 13% and were therefore not calculated as “real” alleles. Notably, the two alleles of the DYS385ab marker were sometimes unbalanced in the MPS analysis (minimum 21.5% PHR for DYS385b). Additionally, a duplication at Y-GATA-H4 in one sample was observed with both CE and MPS (MPS: 495 reads, 76% PHR for the CE allele 10).

Furthermore, we used CE to examine 13 rapidly mutated Y-STRs using a modified RM-Yplex assay developed by Alghafri et al. [26] (including DYS464 in place of DYS518) of the same sample set. DYF399S1, DYF403S1, DYF404S1, DYS449, DYS464, DYS526, DYS547, DYS626, and DYS627 are included in the RM-Yplex assay, but not in the ForenSeq™ DNA Signature Prep Kit. DYF387S1, DYS570, DYS576, and DYS612 are included in both the RM-Yplex assay and ForenSeq™ DNA Signature Prep Kit (Supplementary File S2). Similar to our previous results, all allele designations were concordant, except some stutters above the stutter filter in DYF387S1 and DYS612.

In total, 13 mutations were detected when analysing 22 Y-STR markers of 64 father-son pairs with CE using the PowerPlex® Y23 System (Supplementary File S2): two events at marker DYS19, two events at DYS448, and one event each at DYS385ab, DYS389I, DYS389II, DYS456, DYS458, DYS481, DYS549, DYS643, and Y-GATA-H4. All mutations were either one-repeat losses or one-repeat gains, with the exception of the mutational event at Y-GATA-H4, which was caused by a duplication. Mutational events occurred at markers with simple repeat structures (DYS385ab, DYS456, DYS458, DYS481, DYS549, DYS643, Y-GATA-H4), at markers with complex repeat structures (DYS389I, DYS389II, DYS448), and at one marker with a compound repeat structure (DYS19). They were observed in sequences with a trimeric repeat structure (DYS481), a tetrameric repeat structure (DYS19, DYS385ab, DYS389I, DYS389II, DYS456, DYS458, Y-GATA-H4), a pentameric repeat structure (DYS643), and a hexameric repeat structure (DYS448).

Using MPS analysis (ForenSeq™ DNA Signature Prep Kit, 27 Y-STR loci) we confirmed 12 of 13

mutational events in the same cohort (DYS19, DYS448, DYS385ab, DYS389I, DYS389II, DYS456, DYS481, DYS549, DYS643, and Y-GATA-H4; Supplementary File S2). Two mutations at DYS458 and DYS612 could not be confirmed because of the absence of these markers either in the ForenSeq™ DNA Signature Prep Kit or the PowerPlex® Y23 System. All mutations observed with MPS analysis were length-based mutations with the exception of the duplication at Y-GATA-H4 mentioned earlier. No additional sequence-based mutations were found. However, by sequencing these Y-STRs, we were able to determine the specific part of the repetitive stretch that causes the length variation: DYS19—the first repeat (of two compound repeats), DYS389I—the first repeat (of two compound repeats), DYS389II—the third repeat (of four compound repeats), DYS448—the first repeat (of three compound repeats), and DYS612—the third repeat (of three compound repeats) (Supplementary File S2). All mutational events occurred at the longest uninterrupted repeat sequence (LUS) [31].

By analysing the same sample set with our modified RM-Yplex assay (13 Y-STR markers with up to 24 Y-STR alleles), we identified eight mutational events (Supplementary File S2) and were able to confirm the mutational event at DYS612 found with the ForenSeq™ DNA Signature Prep Kit. All remaining seven mutations occurred at markers not included in the PowerPlex® Y23 System nor in the ForenSeq™ DNA Signature Prep Kit: two at DYF399S1, three at DYF403S1, one at DYF404S1, and one at DYS526. Each of these markers, categorized as RM Y-STR with an average mutation rate of 10^{-2} per locus per generation [32], belongs to the Y-STR multicopy marker and contains a complex repeat structure (Supplementary File S2). All observed mutations showed either one-repeat losses or one-repeat gains.

We calculated the number of alleles per marker for CE analysis (PowerPlex® Y23 System) and for MPS analysis (ForenSeq™ DNA Signature Prep Kit) by looking at all 29 families combined (Supplementary File S2) and identified 112 alleles in our CE analysis results and 151 alleles in our MPS analysis results. Of these 151 alleles detected by MPS, 128 alleles were length-based alleles, which are detectable with CE and MPS in the same manner. The remaining 23 isometric alleles were found to be different based only on their sequence, as they were not detectable with CE: eight at DYF387S1, six at DYS389II, two each at DYS448 and DYS481, three at DYS570, and one each at DYS612 and DYS635 (Figure 1, Supplementary File S2). Most of these markers were defined by a complex repeat structure (DYF387S1, DYS389II, DYS448, DYS612, DYS637), whereas only two of them by a simple repeat structure (DYS481, DYS570).

When looking for the specific part of the repetitive stretch that caused the length variation, we found that more repeats than only the LUS tended to vary in complex markers: two repeats each in DYF387S1 and DYS448, three repeats in DYS389II, and one repeat (LUS) in DYS635. Additionally, alleles longer than 23 (CE length-based) at DYS635 carried an insertion of [TACA]2[TAGA]4 (Supplementary File S2).

Discussion

In summary, we identified the same 13 mutational events in 64 father-son pairs with CE analysis (PowerPlex® Y23 System) and with MPS analysis. Compared with traditional CE analysis, we were unable to differentiate more father-son pairs with MPS analysis than with CE using the PowerPlex® Y23 System. All mutational events detected by MPS were length-based variants, which is in concordance with CE analysis.

Mutations were observed at all types of markers, including simple and complex repeat structures and trimeric to hexameric repeat structures. Our results suggest that tetranucleotide repeats are more prone to mutations than other motifs [32–34], mutation events are mostly restricted to the LUS of a compound/complex STR marker [31,32,34,35], and the LUS length correlates with the mutability of an STR marker [32–34,36–38].

Unsurprisingly, when looking at the sequence variants between unrelated males, we found that the observed length variance is mainly explained by changes in the repetitive structure of a Y-STR. Hence, Just et al. showed that the combination of a length-based allele and LUS length as a designator can represent more than 80% of the variability detected by sequencing [31]. Our data concur with their observations, as nearly all isometric sequence

variants could be transformed using this method without losing any information. Only one exception was found at marker DYS389II, where four different sequence variants (CE length-based allele 30) would be denoted as two alleles (two times CE allele 30 + LUS allele 11 and two times CE allele 30 + LUS allele 12). This finding may help improve existing simplified mutation models (SMM) and phylogenetic separation used in population genetics.

Our results are in agreement with earlier findings [9,39,40] that isometric sequence variants mostly occur at complex STR markers. The largest increases for additional sequence-based alleles were found at DYF387S1 (consistent with [9,41]) and DYS389II (consistent with [39,41–43]). This could be very important for deep-rooted pedigree and paternity analysis using Y-STRs because specific knowledge of the mutation event could facilitate the comparison of more complex mutation models to the simplified single step mutation model [44].

As Huszar et al. predicted [39], we could not find any SNP or indel because of the associated lower mutation rates of these mutational events. We therefore agree with their conclusion that male identification using Y-STRs may not be greatly advanced by applying forensic MPS approaches [39]. In our MPS analysis, we found the same number of mutational events as with CE analysis. However, when analysing the dataset with RM Y-STRs (CE), we found eight additional mutations. Thus, for differentiation of father-son pairs, analysing more RM Y-STRs with CE seems to be a much more promising approach than just sequencing the common Y-STRs. Recently, Ralf et al. proposed 12 new RM Y-STRs with a mean mutation rate of 2.6×10^{-2} , demonstrating high discrimination power [34]. We also agree with Huszar et al. that applying MPS to RM Y-STRs is expected to increase discriminatory power as allele diversity increases. However, MPS analysis using the

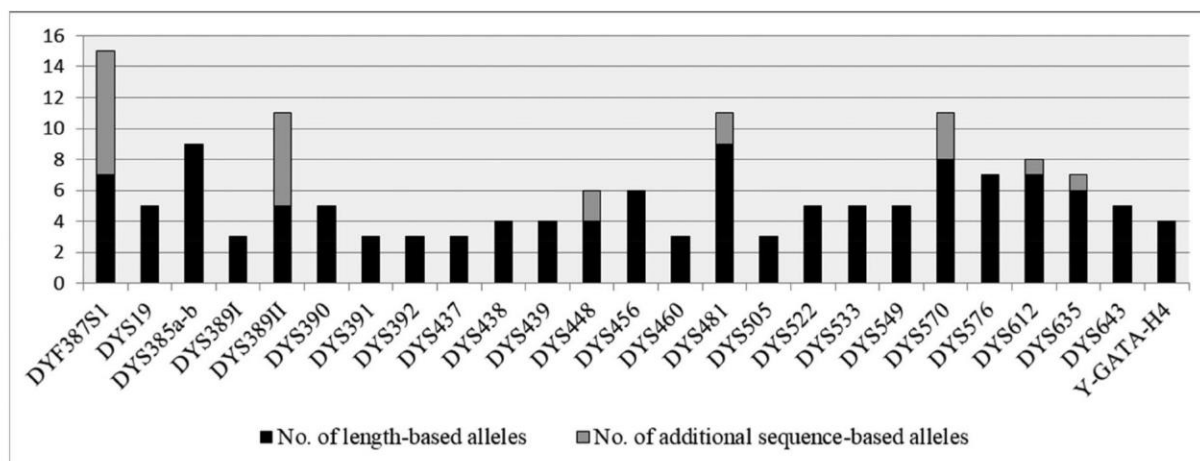


Figure 1. Observed allele increase with massively parallel sequencing (MPS) analysis using the ForenSeq™ DNA Signature Prep Kit (over 29 families).

ForenSeq™ DNA Signature Prep Kit includes the results of autosomal STRs, Y-STRs, X-STRs, and several SNPs at the same time. This would allow for easy individualization, except in female-male DNA mixtures where the male proportion is masked by the female component.

Though we did not find any additional sequence-based mutations between fathers and sons, we found many additional sequence-based alleles between unrelated males. These isometric variants are helpful to exclude a suspect or to deconvolute mixtures of multiple male contributors [39,43], and to detect private or rare mutations pointing to a remote patrilinear relation.

One major drawback of MPS is the amount of novel and challenging bioinformatics methodology necessary for analysing new data. When restricted to vendor-provided tools and protocols, researchers may miss important new features of this technology. Besides basic measurements like sequence reads, coverage, and quality assessment, the major obstacle is a missing internationally standardized nomenclature [45], which would enable laboratories to exchange MPS data without an unnecessary reduction to CE alleles. In summary, we believe that Y-STR MPS analysis is much more demanding for the forensic crime laboratory than traditional CE analysis. Together with the much higher cost of MPS, this could explain the rather hesitant adoption of this new technology in the forensic genetics field. CE will likely remain the standard method for genetic profiling for the next several years, even though especially at Y-STRs MPS could reveal much more information and help to understand the mechanisms of mutational events beyond simplified models like SMM [44].

Acknowledgments

The authors would like to thank all volunteers. The authors would like to thank DNASEqEx members Antonio Alonso, Pedro Barrio, and Pablo Martin (National Institute of Toxicology and Forensic Sciences, Madrid Department, Spain), Walther Parson, Petra Müller, Martin Bodner, and Burkhard Berger (Institute of Legal Medicine, Medical University of Innsbruck, Austria), and Bruce Budowle (Center for Human Identification, University of North Texas Health Science Center, USA; Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia).

Author contributions

Steffi Bredemeyer conducted the experiments and wrote the manuscript. Sascha Willuweit programmed the in-house analysis pipeline (including stutter calculation) for the analysis of raw data. Subsequent data analysis was performed by Steffi Bredemeyer. Sascha Willuweit helped to draft the manuscript. Lutz Roewer helped with the design of the study and corrected the manuscript.

Compliance with ethical standards

Ethical review for recruitment and analysis was provided by the Charité – Universitätsmedizin Berlin, Forensic Genetics Department. Informed consent was provided by all participants.

Funding

The DNASEqEx project was supported by the Internal Security Funding Police program of the European Commission-Directorate General Migration and Home Affairs under the European Commission [grant number HOME/2014/ISFP/AG/LAWX/4000007135].

ORCID

Lutz Roewer  <http://orcid.org/0000-0001-9383-4941>

Sascha Willuweit  <http://orcid.org/0000-0003-1844-8375>

References

- [1] Roewer L. Y chromosome STR typing in crime case-work. *Forensic Sci Med Pathol.* 2009;5:77–84.
- [2] Kayser M, Vermeulen M, Knoblach H, et al. Relating two deep-rooted pedigrees from Central Germany by high-resolution Y-STR haplotyping. *Forensic Sci Int Genet.* 2007;1:125–128.
- [3] Curran JM, Buckleton JS. Effectiveness of familial searches. *Sci Justice.* 2008;48:164–167.
- [4] Alonso A, Barrio PA, Muller P, et al. Current state-of-art of STR sequencing in forensic genetics. *Electrophoresis.* 2018;39:2614–2655.
- [5] Bai R, Liu Y, Li Z, et al. Developmental validation of a novel 5 dye Y-STR system comprising the 27 YfilerPlus loci. *Sci Rep.* 2016;6:29557.
- [6] Martin P, de Simon LF, Luque G, et al. Improving DNA data exchange: validation studies on a single 6 dye STR kit with 24 loci. *Forensic Sci Int Genet.* 2014;13:68–78.
- [7] Köcher S, Müller P, Berger P, et al. Inter-laboratory validation study of the ForenSeq™ DNA Signature Prep Kit. *Forensic Sci Int Genet.* 2018;36:77–85.
- [8] Warshauer DH, Churchill JD, Novroski N, et al. Novel Y-chromosome short tandem repeat variants detected through the use of massively parallel sequencing. *Genom Proteom Bioinf.* 2015;13:250–257.
- [9] Novroski NM, King JL, Churchill JD, et al. Characterization of genetic sequence variation of 58 STR loci in four major population groups. *Forensic Sci Int Genet.* 2016;25:214–226.
- [10] Planz JV, Sannes-Lowery KA, Duncan DD, et al. Automated analysis of sequence polymorphism in STR alleles by PCR and direct electrospray ionization mass spectrometry. *Forensic Sci Int Genet.* 2012;6:594–606.
- [11] Pitterl F, Schmidt K, Huber G, et al. Increasing the discrimination power of forensic STR testing by employing high-performance mass spectrometry, as illustrated in indigenous South African and Central Asian populations. *Int J Legal Med.* 2010;124:551–558.
- [12] Geppert M, Edelmann J, Lessig R. The Y-chromosomal STRs DYS481, DYS570, DYS576 and DYS643. *Legal Med.* 2009;11:S109–S110.

- [13] D'Amato ME, Ehrenreich L, Cloete K, et al. Characterization of the highly discriminatory loci DYS449, DYS481, DYS518, DYS612, DYS626, DYS644 and DYS710. *Forensic Sci Int Genet.* 2010;4:104–110.
- [14] Ballantyne KN, Ralf A, Aboukhalid R, et al. Toward male individualization with rapidly mutating y-chromosomal short tandem repeats. *Hum Mutat.* 2014;35:1021–1032.
- [15] Churchill JD, Chang J, Ge J, et al. Blind study evaluation illustrates utility of the Ion PGM™ system for use in human identity DNA typing. *Croat Med J.* 2015;56:218–229.
- [16] Wendt FR, Churchill JD, Novroski NM, et al. Genetic analysis of the Yavapai Native Americans from West-Central Arizona using the Illumina MiSeq FGx™ forensic genomics system. *Forensic Sci Int Genet.* 2016;24:18–23.
- [17] Ruitberg CM, Reeder DJ, Butler JM. STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res.* 2001;29:320–322.
- [18] Fordyce SL, Avila-Arcos MC, Rockenbauer E, et al. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. *Biotechniques.* 2011;51:127–133.
- [19] Fordyce SL, Mogensen HS, Borsting C, et al. Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™. *Forensic Sci Int Genet.* 2015;14:132–140.
- [20] Scheible M, Loreille O, Just R, et al. Short tandem repeat typing on the 454 platform: strategies and considerations for targeted sequencing of common forensic markers. *Forensic Sci Int Genet.* 2014;12:107–119.
- [21] Zeng X, King JL, Stoljarova M, et al. High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing. *Forensic Sci Int Genet.* 2015;16:38–47.
- [22] Kim EH, Lee HY, Kwon SY, et al. Sequence-based diversity of 23 autosomal STR loci in Koreans investigated using an in-house massively parallel sequencing panel. *Forensic Sci Int Genet.* 2017;30:134–140.
- [23] Devesse L, Ballard D, Davenport L, et al. Concordance of the ForenSeq™ system and characterisation of sequence-specific autosomal STR alleles across two major population groups. *Forensic Sci Int Genet.* 2018;34:57–61.
- [24] Wendt FR, King JL, Novroski NM, et al. Flanking region variation of ForenSeq™ DNA Signature Prep Kit STR and SNP loci in Yavapai Native Americans. *Forensic Sci Int Genet.* 2017;28:146–154.
- [25] Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. *Genetics.* 2000;156:297–304.
- [26] Alghafri R, Goodwin W, Ralf A, et al. A novel multiplex assay for simultaneously analysing 13 rapidly mutating Y-STRs. *Forensic Sci Int Genet.* 2015;17:91–98.
- [27] Butler JM, Schoske RUS. Population data for the multi-copy Y-STR locus DYS464. *J Forensic Sci.* 2005;50:1–2005103.
- [28] Rodig H, Grum M, Grimmecke HD. Population study and evaluation of 20 Y-chromosome STR loci in Germans. *Int J Legal Med.* 2007;121:24–27.
- [29] Redd AJ, Agellon AB, Kearney VA, et al. Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci Int.* 2002;130:97–111.
- [30] Bright JA, Curran JM, Buckleton JS. Modelling PowerPlex® Y stutter and artefacts. *Forensic Sci Int Genet.* 2014;11:126–136.
- [31] Just RS, Irwin JA. Use of the LUS in sequence allele designations to facilitate probabilistic genotyping of NGS-based STR typing results. *Forensic Sci Int Genet.* 2018;34:197–205.
- [32] Ballantyne KN, Goedbloed M, Fang R, et al. Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am J Hum Genet.* 2010;87:341–353.
- [33] Eckert KA, Hile SE. Every microsatellite is different: intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Mol Carcinog.* 2009;48:379–388.
- [34] Ralf A, Lubach D, Kousouri N, et al. Identification and characterization of novel rapidly mutating Y-chromosomal short tandem repeat markers. *Hum Mutat.* 2020;41:1680–1696.
- [35] van der Gaag KJ, de Leeuw RH, Hoogenboom J, et al. Massively parallel sequencing of short tandem repeats-Population data and mixture analysis results for the PowerSeq™ system. *Forensic Sci Int Genet.* 2016;24:86–96.
- [36] Brinkmann B, Klitschar M, Neuhuber F, et al. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet.* 1998;62:1408–1415.
- [37] Kayser M, Roewer L, Hedman M, et al. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet.* 2000;66:1580–1588.
- [38] Kayser M, Kittler R, Erler A, et al. A comprehensive survey of human Y-chromosomal microsatellites. *Am J Hum Genet.* 2004;74:1183–1197.
- [39] Huszar TI, Jobling MA, Wetton JH. A phylogenetic framework facilitates Y-STR variant discovery and classification via massively parallel sequencing. *Forensic Sci Int Genet.* 2018;35:97–106.
- [40] Gettings KB, Kiesler KM, Faith SA, et al. Sequence variation of 22 autosomal STR loci detected by next generation sequencing. *Forensic Sci Int Genet.* 2016;21:15–21.
- [41] Just RS, Moreno LI, Smerick JB, et al. Performance and concordance of the ForenSeq™ system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens. *Forensic Sci Int Genet.* 2017;28:1–9.
- [42] Simayijiang H, Morling N, Borsting C. Sequencing of human identification markers in an Uyghur population using the MiSeq FGx™ Forensic Genomics System. *Forensic Sci Res.* 2020:1–9.
- [43] Kwon SY, Lee HY, Kim EH, et al. Investigation into the sequence structure of 23 Y chromosome STR loci using massively parallel sequencing. *Forensic Sci Int Genet.* 2016;25:132–141.
- [44] Caliebe A, Jochens A, Krawczak M, et al. A Markov chain description of the stepwise mutation model: local and global behaviour of the allele process. *J Theor Biol.* 2010;266:336–342.
- [45] Parson W, Ballard D, Budowle B, et al. Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Sci Int Genet.* 2016;22:54–63.

Publikation 4

Alonso A, Barrio PA, Müller P, Köcher S, Berger B, Martin P, Bodner M, Willuweit S, Parson W, Roewer L, Budowle B, Current state-of-art of STR sequencing in forensic genetics. *Electrophoresis*, 2018. 39: p. 2655–2668.

<https://doi.org/10.1002/elps.201800030>

Lebenslauf

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

Komplette Publikationsliste

Bredemeyer S, Roewer L, Willuweit S, Next Generation Sequencing of Y-STRs in father-son pairs and comparison with traditional capillary electrophoresis. *Forensic Sciences Research*, 2021. p. 1-6.

Müller P, Sell C, Hadrys T, Hedman J, **Bredemeyer S**, Laurent FX, Roewer L, Achtruth S, Sidstedt M, Sijen T, Trimborn M, Weiler N, Willuweit S, Bastisch I, Parson W, the SeqForSTR-Consortium, Inter-laboratory study on standardized MPS libraries: evaluation of performance, concordance, and sensitivity using mixtures and degraded DNA. *International Journal of Legal Medicine*, 2020. 134: p. 185–198.

Köcher S, Müller P, Berger B, Bodner M, Parson W, Roewer L, Willuweit S, on behalf of the DNASEQEX Consortium, Inter-laboratory validation study of the ForenSeq™ DNA Signature Prep Kit. *Forensic Science International Genetics*, 2018. 36: p. 77-85. Impact Factor: 4,884

Alonso A, Barrio PA, Müller P, **Köcher S**, Berger B, Martin P, Bodner M, Willuweit S, Parson W, Roewer L, Budowle B, Current state-of-art of STR sequencing in forensic genetics. *Electrophoresis*, 2018. 39: p. 2655–2668. Impact Factor: 2,754

Vorträge

Köcher S, Roewer L, Nagy M, Willuweit S, The DNASEqEx Consortium, Willuweit S, Neuer NGS-Prototyp – Reduktion auf das Wesentliche. In 38. Spurenworkshop, Basel, Februar 2018

Köcher S, Willuweit S, Nagy M, Roewer L, Erste Ergebnisse des DNASEqEx-Projekts aus Berlin – Informationsgewinn durch MPS? In 13. Jahrestagung der Deutschen Gesellschaft für Abstammungsbegutachtung (DGAB) und der Deutschsprachigen Arbeitsgruppe der International Society of Forensic Genetics (ISFG), Münster, Juni 2017

Roewer L, **Köcher S**, Comparison between CE and MPS technology for the analysis of forensic STRs. In 6th Investigator Forum, Prag, April 2017

Köcher S, Willuweit S, Nagy M, Roewer L, Erste Ergebnisse des DNASEqEx-Projekts aus Berlin – Informationsgewinn durch MPS? In 37. Spurenworkshop, Gießen, Februar 2017

Trimborn M, **Köcher S**, Schulz S, Next generation sequencing Die Zukunft der forensischen DNA-Analytik? In Symposium Forensische DNA-Analyse, Berlin, Dezember 2016

Poster

Qian X, Muller J, **Bredemeyer S**, Fabiani M, Kampmann ML, Leeuw R, Willuweit S, Didier M, Stephens K, Parson W, Morling N; Hou Y, Roewer L, Holt C, Børsting C, Knijff P, Passessment of the DNASEqEx 29 Y STR Panel Designed for the MiSeq FGx[®]. In 28th Congress of the International Society for Forensic Genetics, Prag, September 2019

Müller P, Berger B, Bodner M, Alonso A, Barrio P, Martin P, **Köcher S**, Roewer S, Willuweit S, Budowle B, The DNASEQEX Consortium, Parson W, Systematic Evaluation of Massively Parallel STR Sequencing in the DNASEqEx Project. In 27th Congress of the International Society for Forensic Genetics, Seoul, September 2017

Köcher S, Nagy M, Searching for New Methylation Sites for Tissue Identification, In 26th Congress of the International Society for Forensic Genetics, Krakau, September 2015

Danksagung

Abschließend möchte ich allen Personen danken, die mich bei der Erstellung dieser Arbeit unterstützt haben.

Allen voran danke ich meinem Doktorvater Prof. Dr. Lutz Roewer, der mir die einmalige Möglichkeit gegeben hat, in diesem spannenden Thema zu promovieren. Ich danke dir für die unkomplizierte Betreuung und das Vertrauen, dass du mir jederzeit entgegengebracht hast.

Ich möchte mich auch bei Frau PD Dr. Nagy bedanken, die mir damals als Praktikantin den Eintritt in die Abteilung Forensische Genetik ermöglichte. Auch ihr danke ich für das Vertrauen in mich sowie jegliche Unterstützung meiner wissenschaftlichen Arbeit und die Bemühungen um meine Finanzierung.

Mein besonderer Dank gilt Sascha Willuweit. Sascha, ich danke dir für die großartige Zusammenarbeit, für jegliche Unterstützung, Motivation und das Voranbringen meiner Arbeit. Deine bioinformatische Zuarbeit hat die Arbeit sehr viel interessanter und informativer gemacht.

Maria Seidel und Jessica Rothe, ich danke euch für konstruktive Kritik an meiner Arbeit und eure hilfreichen Ratschläge. Judith Weber und Josephine Purps, vielen Dank für die schöne Zeit in unserem gemeinsamen Raum und eure aufbauenden Worte in schwierigen Zeiten. Euch allen danke ich für die umfangreiche Einführung in die Laborarbeit.

Beim gesamten Team der Forensischen Genetik bedanke ich mich für die freundliche Aufnahme in ihren Reihen, die technische Unterstützung und die generelle Hilfsbereitschaft.

Außerdem bedanke ich mich bei allen Kooperationspartnern unseres Projekts DNaseqEx für die freundliche und gute Zusammenarbeit.

Josef Steilen danke ich für die Unterstützung in wissenschaftlichen Fragestellungen.

Zu guter Letzt danke ich meiner Familie. Danke, dass ihr mir so viele Freiräume verschafft habt. Ohne euch hätte ich die Arbeit nicht fertigstellen können. Vor allem danke ich meinem Mann. Danke Conrad, dass du mir so viel abgenommen hast. Danke, für die Unterstützung bei der Formatierung, Verbesserungen von Formulierungen, aber vor allem für deinen unerschöpflichen Rückhalt und all die Ermutigung in schwierigen Zeiten.