Universität Ulm
Medizinische Fakultät

Validierung eines Short Tandem Repeat-basierten Analyseverfahrens mittels Massively Parallel Sequencing zur Identifizierung forensischer Proben

DISSERTATION

zur Erlangung des Doktorgrades der

Humanbiologie der Medizinischen Fakultät der Universität Ulm

Janine Silvery

geb. in Bad Kreuznach

2022

Amtierender Dekan: Prof. Dr. Thomas Wirth

1. Gutachter: Prof. Dr. Peter Wiegand
2. Gutachterin: Prof. Dr. Sabine Lutz-Bonengel

Tag der Promotion: 11. November 2022

Für Frederik und Mama.
In Liebe und Dankbarkeit!

Inhaltsverzeichnis

Seite

Abkürzungsverzeichnis ..... **VI**
Glossar ..... **VIII**
Tabellenverzeichnis ..... **X**
Abbildungsverzeichnis ..... **XII**
Publizierte Ergebnisse dieser Dissertation ..... **XIII**
1 Einleitung **1**
1.1 Hochvariable repetitive Regionen 2
1.1.1 Forensisch genutzte Short Tandem Repeats 5
1.2 Molekulargenetische Methoden der STR-Typisierung 7
1.2.1 Multiplex-PCR und Kapillarelektrophorese 7
1.2.2 Sequenzierung der nächsten Generation 9
1.2.3 Massively Parallel Sequencing in der forensischen Molekulargenetik 11
1.2.4 Kommerziell erhältliche MPS-Panels 13
1.2.5 Hochdurchsatzsequenzierung mittels MiSeq-Plattform 14
1.3 Entwicklung und Optimierung von monSTR 17
1.4 Methodenvalidierung 20
1.5 Ziel der Arbeit 22
2 Material und Methoden **23**
2.1 Material 24
2.1.1 Reagenzien und Enzyme 24
2.1.2 Verbrauchsmaterialien 24
2.1.3 Kommerzielle Kits 25
2.1.4 Laborequipment und Analysegeräte 25
2.1.5 Software und Webseiten 26
2.1.6 Biologisches Referenzmaterial 27
2.1.7 Biologisches Probenmaterial 27
2.1.8 Ethikvotum 27
2.2 Methoden 27
2.2.1 Auswahl der forensischen STR-Systeme und Primerdesign 27
2.2.2 Anfügen der illuminaspezifischen Überhangadapter 29
2.2.3 DNA-Isolierung 32
2.2.4 Quantifizierung genomischer DNA 33
2.2.5 Massively Parallel Sequencing mittels monSTR 33
2.2.6 Kapillarelektrophorese 47
2.2.7 Experimenteller Studienaufbau zur Validierung von monSTR 47
3 Ergebnisse **56**
3.1 Funktionalität der monSTR-Primer 56
3.2 Library-Preparation mittels monSTR 57

3.3 Sequenzierqualität der durchgeführten MPS-Läufe 60
3.4 Ergebnisse der Validierung des monSTR-Panels 61
3.4.1 Ermittlung des Stutter-Thresholds 61
3.4.2 Konkordanz 62
3.4.3 Wiederholbarkeit und Reproduzierbarkeit 63
3.4.4 Sensitivität 64
3.4.5 DNA-Mischungen 71
3.4.6 Forensische Ringversuchsproben 73
3.4.7 Speziesspezifität 78
3.4.8 Zusammenfassung der Validierungsergebnisse 79
4 Diskussion 82
4.1 Analytische Leistungsfähigkeit des monSTR-Panels 84
4.2 Validierung des monSTR-Panels 86
4.3 Optimierungsmöglichkeiten und Ausblick 93
5 Zusammenfassung 97
6 Literaturverzeichnis 99
Danksagung 111
Lebenslauf 112

Abkürzungsverzeichnis

ATemp	Annealing Temperature
ATime	Annealing Time
ACR	Allele Coverage Ratio
bp	Basenpaar(e)
BKA	Bundeskriminalamt
CCD	Charged Coupled Device
CE	Capillay electrophoresis (dt. Kapillarelektrophorese)
CODIS	Combined DNA Index System
Cov	Coverage
CV	Coefficient of variation of the locus coverage
CR	Calling ratio
DAD	DNA-Analyse-Datei
DAkkS	Deutsche Akkreditierungsstelle GmbH
dNTP	Desoxyribonukleosid-Triphosphat
ddNTP	Didesoxyribonukleosid-Triphosphate
DoE	Design of Experiment (dt. Statistische Versuchsplanung)
ENFSI	European Network of Forensic Science Institutes
ESS	European Standard Set
ETime	Extension Time
FU	Fluoreszenzintensität
fwd	Forward
GB	Gigabasen
GEDNAP	German DNA Profiling
IEM	Illumina Experiment Manager
InDel	Insertion/ Deletion
ISFG	International Society of Forensic Genetics
kb	Kilobasenpaar(e)
kV	Kilovolt
MAF	Minor component allele frequency (dt. Allelfrequenz der Nebenkomponente)
min	Minute

Max.	Maximum
Min.	Minimum
Mio.	Millionen
MPS	Massively Parallel Sequencing
mtDNA	Mitochondriale DNA
mPCR	Multiplex Polymerase Chain Reaction
NGS	Next-Generation Sequencing
nt	Nukleotid(e)
NTC	No Template Control
PCR	Polymerase Chain Reaction (dt. Polymerase-Kettenreaktion)
pM	Pikomolar
RT	Raumtemperatur
rev	Reverse
RFLP	Restriction fragment length polymorphism
rfu	Relative Fluorescence Unit (dt. Relative Fluoreszenz Einheit)
rpm	Revolution per Minute (dt.: Umdrehungen pro Minute)
QC	Qualitätskontrolle
qPCR	Quantitative (Real-Time) Polymerase Chain Reaction
s	Sekunde
SBS	Sequencing-by-synthesis
SN	Signal-to-Noise Verhältnis
SNP	Single nucleotide polymorphism
SOP	Standard Operation Protocol (dt. Arbeitsanweisungsprotokoll)
SPRI	Solid Phase Reversible Immobilization
STR	Short tandem repeat
SWGDAM	Scientific Working Group on DNA Analysis Methods
Tm	Schmelztemperatur
µm	Mikrometer
VNTR	Variable number of tandem repeats

Glossar

Im folgenden Glossar werden einzelne Fachbegriffe aufgeführt, die im Zusammenhang der vorliegenden Dissertation mit *Massively Parallel Sequencing* (MPS) häufig verwendet werden:

Amplicon-Sequencing	Hochdurchsatz-Sequenzierung von DNA-Fragmenten, die zuvor durch herkömmliche PCR amplifiziert werden.
Cluster Density	Dieser Parameter wird in der Einheit k/mm^2 angegeben und stellt die Dichte (engl. *Density*) der klonalen *Cluster* auf der Fließzelle (*Flow Cell)* dar. Die *Cluster Density* korreliert mit Laufqualität und Gesamtdatenoutput. Es wird daher eine hohe Dichte angestrebt, um den Gesamtdatenoutput zu maximieren. Für das in dieser Dissertation beschriebene MiSeq Reagent Kit v2 (Illumina) liegt der Optimalwert bei 800 k/mm^2.
(Depth of) Coverage	Die durchschnittliche Anzahl an *Reads*, die mit jeder Base im Template alignieren. Eine *Coverage* von 30x (*Depth*) bedeutet beispielsweise, dass im Durchschnitt jede Base im Template von 30 *Reads* erfasst wurde.
Cluster	Ein einzelner Bereich von identischen DNA-Sequenzen auf der *Flow Cell*.
Cluster Passing Filter	Dieser Filter umfasst einen gewissen Prozentsatz an *Reads*, die den spezifischen Qualitätsfilter passieren und somit in die weitere bioinformatische Auswertung einbezogen werden konnten.
FASTQ	Eine Datei mit Sequenzen einzelner *Reads* und entsprechenden Qualitätsindikatoren für jede einzelne Base. Das *FASTQ*-File wird typischerweise bei der primären MPS-Datenanalyse erzeugt.
Fehlerrate	Die Fehlerrate gibt den jeweiligen Prozentsatz an Sequenzierfehlern im MPS-Lauf an und wird auf Basis der sequenzierten *PhiX*-Kontrolle ermittelt, die als interne Kontrolle mit einer bekannten Nukleotidsequenz mitgeführt wurde.
Flow Cell	Hier findet während der Sequenzierung die *Cluster*-Generierung statt. Eine *Flow Cell* ist ein planarer Glasträger, welche mit den zu den *Index*-Adaptern komplementären

	Oligos beschichtet ist, die eine komplementäre Bindung ermöglichen.
Indexing	Anfügen individueller Barcode-Markierungen (*Indices*) verschiedener *Libraries* für den MPS-Lauf, so dass diese zwar gemeinsam auf dem Gerät analysiert, aber im Anschluss getrennt ausgewertet werden können.
Library	Sequenzierbibliothek, welche alle zu sequenzierenden DNA-Abschnitte beinhaltet und am Ende eines komplexen Prozesses für die eigentliche Sequenzierung bereit ist (Plural: *Libraries*).
Massively Parallel Sequencing	Auch bekannt als *Next-Generation Sequencing* (NGS) oder Hochdursatzsequenzierung. Die „nächste Generation“ der DNA-Sequenzierung (nach der Methode nach Sanger). Technologie zur massiven parallelen Sequenzierung von Nukleinsäuren.
Paired End Sequencing	Hierbei handelt es sich um ein Sequenzierverfahren, bei dem die Analyse von DNA-Fragmenten einer *Library* von beiden Richtungen erfolgt (3´- 5‘ und 5´- 3‘-Richtung).
Panel	Zusammenfassung mehrerer ähnlicher Targets (z.B. *Short Tandem Repeats*, STR) in einem Sequenzieransatz.
PhiX	*PhiX* ist eine adapter-ligierte, kommerziell erhältliche *Library* der Firma Illumina, die als interne Kontrolle in geringer Konzentration für Sequenzierläufe Verwendung findet, um eine optimierte Bewertung technischer Artefakte innerhalb der Sequenzierung zu ermöglichen. Die durchschnittliche Fragmentgröße der *Library* beträgt 500 Basenpaare (bp).
Q-Score	Prognose über die Wahrscheinlichkeit eines fehlerhaften Baseneinbaus, die während der Sequenzierung für jeden Zyklus erfasst wird. Der Wert drückt die durchschnittliche Wahrscheinlichkeit eines falschen Baseneinbaus innerhalb der Sequenzierung aus. Ein *Q-Score* von 20 (Q20) steht dabei bspw. für einen falschen je 100 Baseneinbauten, was einer Sequenziergenauigkeit von 99 % entspricht. Bei einer höheren Genauigkeit von 99,9 % (Q30) liegt die Wahrscheinlichkeit eines falschen Baseneinbaus bei 1:1000.
Read	Eine mittels MPS-Technologie generierte Nukleotidsequenz.
Single end Sequencing	Sequenzierung eines DNA-Fragments in nur einer Richtung von einem Ende (*Read* 1) zum anderen.

Tabellenverzeichnis

Tabelle 1: Übersicht autosomaler STR-Kernloci der Vereinigten Staaten von Amerika, Europa und Deutschland.....6

Tabelle 2: Vergleich der heute gängigen STR-Analyseverfahren.....12

Tabelle 3: Verwendete Reagenzien und Enzyme.....24

Tabelle 4: Verwendete Verbrauchsmaterialien.....24

Tabelle 5: Verwendete kommerzielle Kits.....25

Tabelle 6: Verwendetes Laborequipment und Analysegeräte.....25

Tabelle 7: Verwendete Software und Webseiten.....26

Tabelle 8: Forward und reverse Primersequenzen des Multiplexpanels monSTR.....31

Tabelle 9: Temperaturprotokoll der mPCR.....36

Tabelle 10: Temperaturprotokoll der Index-PCR.....38

Tabelle 11: Detailliertes Pipettierschema für die Verdünnung der Library vor der Sequenzierung.....42

Tabelle 12: Übersicht der für die Validierung verwendeten Sequenzierreagenzien inkl. der vom Hersteller angegebenen Qualitätswerte.....43

Tabelle 13: Definition und Berechnung der statistischen Parameter zur Bewertung der Leistungsfähigkeit des monSTR-Panels.....46

Tabelle 14: Autosomales STR-Profil der SRM 2391c A und C gemäß Analysezertifikat des Herstellers.....51

Tabelle 15: Probenübersicht der GEDNAP-Einzelpersonenproben 54 bis 61, inkl. Probentyp und zugrundeliegende DNA-Konzentration.....53

Tabelle 16: Probenübersicht der GEDNAP-Spurenproben 54 bis 61, inkl. Probentyp und zugrundeliegende DNA-Konzentration.....54

Tabelle 17: Qualitätsparameter der Sequenzierläufe bei Verwendung von monSTR.....60

Tabelle 18: Analysebereich des N-1-Stutters aller STR-Systeme des monSTR-Panels.....62

Tabelle 19: Durchschnittliche Coverages inkl. Standardabweichungen der Konkordanzstudie.....63

Tabelle 20: Ergebnisse der statistischen Parameter inkl. Standardabweichungen für das Intra- und Inter-Run Experiment.....64

Tabelle 21: Erfolgreich typisierte Allele der 2800M innerhalb acht verschiedener Verdünnungsstufen. 65

Tabelle 22: Durchschnittliche Coverage und erfolgreich typisierte Allele der Referenzprobe SRM2391c C (Nebenkomponente) in Abhängigkeit des jeweiligen Mischungsverhältnisses. 73

Tabelle 23: Anzahl der mittels monSTR und CE typisierten (Iso-) Allele für insgesamt 23 GEDNAP-Personenproben im Vergleich zur Referenz. 75

Tabelle 24: Anzahl der mittels monSTR und CE typisierten (Iso-) Allele für insgesamt 32 GEDNAP-Spurenproben im Vergleich zur Referenz. 76

Tabelle 25: Heterozygoter Genotyp, ACR und Coverage des SE33-Lokus innerhalb der GEDNAP-Einzelpersonenproben. 77

Abbildungsverzeichnis

Abbildung 1: Schematische Darstellung eines Short Tandem Repeat (STR)-Systems. 3

Abbildung 2: Schematische Darstellung der Bridge amplification. 15

Abbildung 3: Schematische Darstellung des Sequencing-by-synthesis Ansatzes. 16

Abbildung 4: Diagramm des monSTR-Workflows. 35

Abbildung 5: Elektrophoretische Auftrennung der 2800M Kontroll-DNA in unterschiedlichen Stadien der Library-Preparation. 59

Abbildung 6: Relative Lokus-Coverage in Abhängigkeit verschiedener DNA-Konzentrationen der 2800M. 67

Abbildung 7: Mittelwerte der statistischen Parameter CV, CR und SN in Abhängigkeit verschiedener 2800M-Konzentrationen. 68

Abbildung 8: Heatmap der durchschnittlichen ACR über alle heterozygoten STR-Systeme der 2800M innerhalb verschiedener Konzentrationsstufen. 70

Abbildung 9: Durchschnittliche MAF unterschiedlicher Mischungsverhältnisse der SRM 2391c A (Hauptkomponente) und SRM 2391c C (Nebenkomponente). 72

Abbildung 10: Durchschnittliche CR in Abhängigkeit dreier Haustierarten, den jeweiligen Tierhaltern, der 2800M und der Negativkontrolle. 79

Abbildung 11: Library-Preparation mittels monSTR und MiSeq. 81

Publizierte Ergebnisse dieser Dissertation

Teile der vorliegenden Dissertation wurden in einer begutachteten Fachzeitschrift veröffentlicht:

Janine Silvery, Sebastian Ganschow, Peter Wiegand, Carsten Tiemann. *Developmental validation of the monSTR identity panel, a forensic STR multiplex assay for massively parallel sequencing*, Forensic Science International: Genetics 46 102236 (2020). DOI: https://doi.org/10.1016/j.fsigen.2020.102236

1 Einleitung

Seit ihrer Einführung hat sich die DNA-Analyse sehr dynamisch entwickelt und sich auch in der forensischen Genetik als zuverlässige und unersetzliche Methode etabliert. Gegenwärtig werden für forensische Fragestellungen, wie z.B. die eindeutige Identifizierung von Personen aber auch für Abstammungsuntersuchungen, bestimmte Abschnitte in der Mikrosatelliten-DNA mit kurzen („short"), hintereinander angeordneten („tandem") Wiederholungsmotiven („repeats") untersucht, deren Wiederholungsanzahlen sich von Individuum zu Individuum unterscheiden können. Diese sind als Längenpolymorphismen bekannt. Der Grundstein für die forensisch-molekularbiologische Untersuchung polymorpher DNA-Regionen wurde in den 1980er Jahren mit der Beschreibung von Abschnitten in der Minisatelliten-DNA, den Restriktionsfragmentlängenpolymorphismen (engl. *Restriction fragment length polymorphism*, RFLP), populär als „genetischer Fingerabdruck" bezeichnet, durch Sir Alec Jeffreys (Jeffreys et al., 1985) gelegt. Die Einführung des genetischen Fingerabdrucks hatte die Ablösung einer Reihe klassischer serologischer Methoden, u.a. die Bestimmung von Blutgruppensystemen (LANDSTEINER, 1900), zur Folge, die bis zu diesem Zeitpunkt die einzige Möglichkeit geboten hatten, (Blut-)Spuren aus einem Kriminalfall Individuen zuzuordnen. Einer der Vorteile moderner Verfahren zur forensischen Individualisierung gegenüber den vormals üblichen Blutgruppenanalysen ist die Analyse der nukleären DNA, die in allen kernhaltigen Zellen vorhanden ist. Außerdem ist die nukleäre DNA als Untersuchungsmatrix haltbarer, was es ermöglicht, auch noch nach Jahren genetisches Material aus einer Spur zu gewinnen. Die Entdeckung der Polymerase-Kettenreaktion (engl. *Polymerase Chain Reaction*, PCR) in den späten 1980er Jahren (Mullis and Faloona, 1987) setzte den Grundstein für die Entwicklung der modernen Molekularbiologie. Überdies bewirkte ihre Einführung in der forensischen Genetik erhebliche Fortschritte, da durch Einsatz der PCR die Analyse selbst minimaler Spurenmengen ermöglicht wurde.

Seit einigen Jahren wird die hochauflösende Sequenziertechnologie (*Massively Parallel Sequencing*, MPS, ebenfalls bekannt als *Next-Generation Sequencing*, NGS oder Hochdurchsatzsequenzierung) nicht zuletzt im Feld der forensischen Genetik eingesetzt, wodurch sich die Erfolgsaussichten bei der Untersuchung schwieriger, degradierter und/ oder

gering ausgeprägter Spuren abermals verbessert haben. Angesichts der Komplexität der Daten, die mittels MPS erzeugt werden, spielen für die erfolgreiche Bearbeitung dieser Daten sowohl die dafür eingesetzten Analyseverfahren als auch entsprechende bioinformatische Softwarelösungen eine entscheidende Rolle.

1.1 Hochvariable repetitive Regionen

Die forensische Genetik spielt in straf- und zivilrechtlichen Streitfällen eine entscheidende Rolle, da Gerichtsurteile in Strafverfahren regelmäßig forensische Befunde als wesentliche Beweismittel einbeziehen. Das sog. „Austauschprinzip“, formuliert von Edmond Locard in den 1930er Jahren, besagt zudem, dass beim Kontakt zweier Objekte und/ oder Personen wechselseitig Spuren des Kontakts aneinander hinterlassen werden (Locard, 1930), wodurch am Tatort vorfindliche Spuren zu einem Schlüsselelement in der forensischen Fallarbeit werden können, da sie Rückschlüsse auf Beteiligte und Modus stattgehabter Kontakte ermöglichen. Jedes nukleinsäurehaltige biologische Material, wie z.B. Blut, Sperma, Speichel, Knochen, Haare oder Zähne kann dabei genetische Information liefern.

Mit der Entdeckung repetitiver Regionen im menschlichen Genom gewann die Verwendung von DNA zu Zwecken der Identifizierung von Menschen zunehmend an Bedeutung. Genomische Regionen, die hochrepetitive Sequenzen enthalten, werden als „Satelliten-DNA“ bezeichnet. Diese wird konstituiert durch zahlreiche Wiederholungen kurzer Basenmotive (ca. 3-5 Basenpaare, bp). Je nach Gesamtlänge einer repetitiven Sequenz wird zwischen den in Zentromer- und Telomerbereichen liegenden Satelliten (100 Kilobasenpaare, kb, bis mehrere Megabasenpaare), Minisatelliten (*Variable number of tandem repeats*, VNTR, 0,1 bis 20 kb) und Mikrosatelliten (20 bis 150 bp, auf allen Chromosomen lokalisiert) unterschieden.

Seit Einführung der PCR-Technologie beruht die heute gängige forensische DNA-Analyse üblicherweise auf der Typisierung der zu den Mikrosatelliten zählenden *Short tandem repeats* (STRs, Singular: STR) (McCord et al., 2019). STRs wurden im Genom von *Drosophila virilis* bereits 1984 von Tautz und Renz entdeckt (Tautz and Renz, 1984) und 1989 ebenfalls von Tautz als geeignete Marker zur Identifizierung, Kartierung sowie in populationsgenetischen Studien verwendet (Tautz, 1989). STR-Systeme haben sich seit ihrer Entdeckung als genetische Marker der Wahl etabliert. Sie enthalten, flankiert von nicht-repetitiven Sequenzen, Wiederholungen ein- und desselben wiederkehrenden kurzen Sequenzmotivs (Abbildung 1)

und sind über das gesamte Genom verteilt (Edwards et al., 1991). STRs befinden sich in nicht kodierenden Bereichen der DNA, sind evolutiv neutral und haben somit bis auf wenige Ausnahmen keinen Einfluss auf den Phänotyp (Wyner et al., 2020). Charakteristisch sind tandemartige Wiederholungen kurzer Basenmotive, die entsprechend ihrer Länge als Di-, Tri-, Tetra-, Penta- bzw. Hexanukleotid-Marker bezeichnet werden. Da STR-Systeme eine hohe Variabilität innerhalb der Bevölkerung aufweisen, bezeichnet man diese Loci als polymorph und zählt sie zu den multiallelischen Merkmalsystemen. Durch ihre polymorphe Natur besteht eine große individuelle Variabilität in der Anzahl der Motivwiederholungen. Dies entspricht der Vielzahl beobachteter genetischer Ausprägungen (Allele) und ermöglicht – bei Kombination hinreichend vieler Systeme – ein praktisch für einen Menschen einzigartiges DNA-Profil (eineiige Zwillinge bilden hier eine wichtige Ausnahme, da ihre STR-Profile identisch sind).

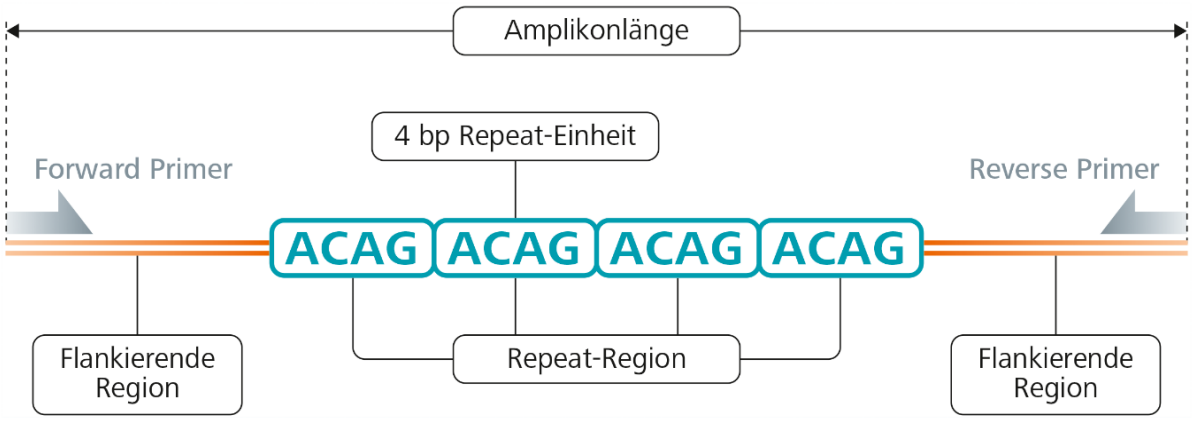

Abbildung 1: Schematische Darstellung eines Short Tandem Repeat (STR)-Systems. Die hier dargestellte Repeat-Region besteht aus vier tetrameren Repeat-Einheiten mit dem Basenmotiv ACAG. Die Bindestellen für die PCR-Primer liegen innerhalb der flankierenden Region und bestimmen die gesamte Amplikonlänge des STR-Systems. Da sich die Anzahl repetitiver Einheiten von Individuum zu Individuum unterscheiden kann, hängt die Gesamtgröße des Amplikons von der individuellen Repeat-Region ab. bp = Basenpaar; A = Adenin; C = Cytosin; G = Guanin.

Die Nomenklatur der STR-Systeme beruht auf der Anzahl der variablen Wiederholungseinheiten, wodurch eine Klassifizierung zwischen einfachen (*simple*), zusammengesetzten (*compound*) sowie komplexen (*complex*) Repeatstrukturen vorgenommen wurde (Gill et al., 1997). Das Interesse an Mikrosatelliten in der Spurenkunde begründete sich zunächst vor allem in ihrer geringen Größe (120-400 bp). Angesichts ihrer kurzen

Fragmentlängen sind sie weniger stark von Degradationsprozessen bzw. Fragmentierung betroffen und lassen sich mit höherer Erfolgsquote in einer sogenannten Multiplex-PCR (mPCR) amplifizieren (Kimpton et al., 1993). Aus diesem Grund hat sich die Typisierung forensischer STR-Marker zur zentralen Methode der forensischen Genetik entwickelt.

Neben den zuvor beschriebenen multiallelischen Merkmalsystemen sind die biallelischen Systeme zu nennen, die aktuell unterstützend bzw. ergänzend zur STR-Analyse eingesetzt werden. Hierzu zählen die sogenannten Einzelnukleotid-Polymorphismen (engl. *Single nucleotide polymorphisms,* SNPs) und die Insertions-Deletions Polymorphismen (InDels). Im Unterschied zu den Längenpolymorphismen der Mikrosatelliten-DNA sind SNPs Beispiele für Sequenzpolymorphismen. Ein SNP ist eine Variation eines einzelnen Basenpaares in einem komplementären DNA-Doppelstrang, der in der Population gewöhnlich in nur zwei (von vier möglichen) Ausprägungen auftritt (triallelische SNPs wurden jedoch beobachtet ((Hodgkinson and Eyre-Walker, 2010)). SNPs besitzen einen durchschnittlichen Abstand zueinander von 0,3 bis 1,2 kb. Dies ergibt eine geschätzte Gesamtzahl der im gesamten, humanen Genom vorkommenden SNPs von 1,5 x 10^7 (Bezogen am 03.04.2022 von Bionumbers, https://bionumbers.hms.harvard.edu/bionumber.aspx?id=103895&ver=7&trm=SNPs&org=).

Die Entstehung derartiger Punktmutationen wird durch unterschiedliche Mechanismen beeinflusst, die einen Austausch des Nukleotids an dieser Position zur Folge haben. Begründet durch ihre niedrige Mutationsrate mit durchschnittlich 10^{-8} (Amorim and Pereira, 2005) im Vergleich zu den bisher aufgeführten STRs, bei denen jene zwischen 10^{-2} und 10^{-4} liegt (Brinkmann et al., 1998), eignen sich SNPs hervorragend für populationsgenetische und populationsgeographische Analysen. Fragmentlängenpolymorphismen, die durch kombinierte Insertions-Deletions-Mutationen weniger Nukleotide entstehen, werden als InDels bezeichnet. Derzeit sind über 2000 InDels im humanen Genom bekannt (Weber et al., 2002).

Die genetische Information in der Zelle ist auf Autosomen und Geschlechtschromosomen (sog. Gonosomen) sowie dem mitochondrialen Genom verteilt. Auf letzteren beiden konnten zudem forensisch nutzbare Markersysteme und Bereiche detektiert und charakterisiert werden, die sich aufgrund ihres besonderen Vererbungsmusters - die mitochondriale DNA (mtDNA) wird an alle Nachkommen über die maternale Linie und das Y-Chromosom an männliche Nachkommen über die paternale Linie vererbt - sehr gut zur Klärung spezieller forensischer Fragestellungen eignen (z.B. komplexe Verwandtschaftsverhältnisse über mehrere Generationen). Die paternale Vererbung des Y-Chromosoms und das Fehlen eines Rekombinationspartners in der Meiose

bedingen, dass dieses Chromosom nahezu unverändert in der väterlichen Linie weitervererbt wird. Analog zu den Autosomen befinden sich auf dem Y-Chromosom ebenfalls STRs, die im Gegensatz zu autosomalen STRs ohne Rekombination als sog. „Haplotypen“ vererbt werden (Roewer et al., 1992). Der Vorteil der Analyse Y-chromosomaler Marker offenbart sich bei der forensisch-molekularbiologischen Untersuchung von Spuren bei Sexualdelikten. Bei dieser Deliktform werden an Vaginalabstrichen häufig Mischspuren festgestellt, die nur wenige Zellen des männlichen Täters in einem überwiegenden Hauptanteil weiblicher Epithelzellen des Opfers enthalten. Auf diese Weise kann der männliche Anteil des Y-Chromosoms selektiv analysiert werden. Überdies enthält das X-Chromosom STRs, die für spezielle genetische Fragestellungen von besonderer Bedeutung sind. Hier sind vor allem die Aufklärung komplizierter Verwandtschaftsverhältnisse, wie beispielsweise Inzest- oder Defizienzfällen, in denen nicht alle für die Untersuchung wichtigen Personen zur Verfügung stehen, zu nennen (Szibor, 2007).

1.1.1 Forensisch genutzte Short Tandem Repeats

Im Laufe der Jahre haben sich eine Reihe gut charakterisierter STR-Marker etabliert, die routinemäßig für die Erstellung forensischer DNA-Profile verwendet werden. Allelfrequenzen, die im Rahmen von populationsgenetischen Studien ermittelt wurden (Gettings et al., 2018; Gill and Evett, 1995), ermöglichen dabei die biostatistische Bewertung der Genotypenhäufigkeiten zur Berechnung von Verursacherschaftswahrscheinlichkeiten und Likelihood-Ratios beim Vergleich von DNA-Profilen. Gemeinsame Bemühungen verschiedener forensischer Arbeitsgruppen führten zu einer Einigung über die in Europa (*European Standard Set*, ESS) und den Vereinigten Staaten von Amerika (*Combined DNA Index System*, CODIS) zu analysierenden Locus-Kombinationen (Tabelle 1), die von mehreren kommerziell erhältlichen STR-Kits abgedeckt werden (Ludeman et al., 2018; Oostdik et al., 2014).

Diesem Standardisierungsprozess folgte die Einrichtung nationaler DNA-Datenbanken in mehreren Ländern weltweit sowie die Veröffentlichung umfangreicher Studien zu PCR-basierten technischen Artefakte der ausgewählten Kernloci. Die DNA-Analyse-Datei (DAD) ist die im April 1998 zur Speicherung von genetischen DNA-Profilen eingerichtete Datenbank in Deutschland, die vom Bundeskriminalamt (BKA) zentral betrieben wird. Innerhalb dieser

Datenbank werden sowohl genetische Profile von bekannten Personen (sog. Personendatensätze) als auch von Spurenträgern, die von unbekannten Personen stammen (sog. Spurendatensätze), registriert und stetig miteinander abgeglichen. DAD-Datensätze umfassten ursprünglich fünf STR-Systeme und wurden 2001 auf eine Kombination von acht STR-Loci erweitert. Ein Alleinstellungsmerkmal der DAD gegenüber anderen nationalen DNA-Datenbanken ist die Verwendung des SE33-Systems (Polymeropoulos et al., 1992), das zu den informativsten und komplexesten bekannten Tetranukleotid-Markern gehört. Seine hochgradig polymorphe und heterogene Natur weist aufgrund einer überdurchschnittlich hohen Mutationsrate eine komplexe Längen- und Sequenzvariation auf (Lászik et al., 2001; Wiegand et al., 1993), die für das Diskriminierungspotential forensischer DNA-Typisierungen sehr nützlich ist. Andererseits kann die komplexe Struktur des Lokus die CE- und MPS-Technologie stören (Wang et al., 2012). Die Variation in den flankierenden Regionen von SE33 trägt zusätzlich zum hohen Informationsgehalt dieses Systems bei (Gettings et al., 2015).

Tabelle 1: Übersicht autosomaler STR-Kernloci der Vereinigten Staaten von Amerika, Europa und Deutschland. In der Übersicht sind zusätzlich die jeweilige chromosomale Lokalisation sowie die Allelbereiche angegeben. CODIS = *Combined DNA Index System;* ESS = *European Standard Set;* DAD = DNA-Analyse-Datei. Bezogen am 24.03.2022 von STRBase (https://strbase.nist.gov/).

STR-Lokus	CODIS	ESS	DAD	chrom. Lokalisation	Allelbereich
D1S1656	✓	✓		1q42	9 - 21
TPOX	✓			2p25.3	4 - 16
D2S441	✓	✓		2q14	8 - 17
D2S1338	✓			2q35	11 - 28
D3S1358	✓	✓	✓	3p21.31	8 - 20
FGA	✓	✓	✓	4q31.3	12.2 – 51.2
D5S818	✓			5q23.2	6 - 18
CSF1PO	✓			5q33.1	5 - 16
SE33			✓	6p14	3 – 39.2
D7S820	✓			7q21.11	5 - 16
D8S1179	✓	✓	✓	8q24.13	7 - 20
D10S1248	✓	✓		10q26.3	8 - 19
TH01	✓	✓	✓	11p15.5	3 - 14
vWA	✓	✓	✓	12p13.31	10 - 25
D12S391	✓	✓		12p13.2	15 - 26
D13S317	✓			13q31.1	5 - 17
D16S539	✓			16q21.1	4 - 16
D18S51	✓	✓	✓	18q21.33	7 – 39.2
D19S433	✓			19q12	5.2 - 20
D21S11	✓	✓	✓	21q21.1	12 – 41.2
D22S1045	✓			22q12.3	8 - 20
Amelogenin	✓		✓	Xp22.2, Yp11.2	X, Y

Zur genetischen Geschlechtsbestimmung dient in kommerziellen Analysekits der X-Y-homologe Marker im AMELX bzw. AMELY-Gen, das für Proteine im Zahnschmelz codiert und auf beiden Gonosomen vorkommt (Sullivan et al., 1993). Die PCR-Amplifikation zielt dabei auf eine Deletion im Intron 1 des AMELX-Gens ab und ergibt ein Amplifikationsprodukt von 106 bp. Das Y-spezifische Amplikon aus dem AMELY-Gen ist sechs bp länger und erzeugt ein Produkt von 112 bp. Anhand dieses Markers im DNA-Profil lässt sich somit feststellen, ob eine männliche oder weibliche DNA-Probe analysiert wurde.

Wie bereits erwähnt, setzen sich STR-Systeme aus einfachen, komplexen oder zusammengesetzten Repeatstrukturen zusammen. In vielen Systemen werden sog. Mikrovarianten beobachtet, die unvollständige Wiederholungen enthalten, die auf Mutationen in der jeweiligen Wiederholungsregion zurückzuführen sind. Die Namenskonvention für STR-Systeme unterscheidet zwischen Markern, die innerhalb und außerhalb von Genen liegen (Strachan and Read, 2018). Außerhalb von Genen liegende STR-Marker haben Kennungen entsprechend der chromosomalen Lokalisation, auf der sich der Marker befindet (z.B. D21S11: DNA-Marker, Chromosom 21, Single-Copy-Sequenz, 11. untersuchter Marker auf Chromosom 21). Dagegen weisen innerhalb von Genen liegende STR-Marker auf das entsprechende Gen hin, z.B. FGA (Fibrinogen-Alpha-Kette).

1.2 Molekulargenetische Methoden der STR-Typisierung

1.2.1 Multiplex-PCR und Kapillarelektrophorese

Aus sichergestelltem biologischen Spurenmaterial wird DNA extrahiert und anschließend Menge und Qualität der isolierten autosomalen und Y-chromosomalen DNA bestimmt. Die vorherrschende Methode zur Erstellung von DNA-Profilen umfasst die Amplifikation forensisch relevanter, hochpolymorpher STR-Systeme mit fluoreszenz-markierten PCR-Primern. Hierbei werden mehrere (meist 16 oder mehr) STR-Regionen gleichzeitig untersucht, um die biostatistische Aussagekraft des DNA-Profils zu erhöhen (Chamberlain et al., 1988). Gegenwärtig können mit kommerziell erhältlichen Multiplex-Kits bis zu 26 STR-Marker gleichzeitig in einer einzigen PCR-Reaktion angereichert werden. Im Anschluss daran werden die PCR-Produkte entsprechend ihrer Amplikongröße mittels Kapillarelektrophorese (engl.

Capillary electrophoresis, CE) aufgetrennt. Diese Auftrennung findet in Kapillaren statt, die im Durchmesser zwischen 50 und 100 Mikrometer (µm) messen. Folgend werden die mit unterschiedlichen Fluoreszenzfarbstoffen markierten PCR-Produkte nach Anlegen einer Spannung in die mit Polymerlösung befüllten Kapillaren injiziert. Die Trennung basiert dabei auf der differentiellen elektrophoretischen Mobilität der DNA-Fragmente: Längere Fragmente bewegen sich langsamer durch die Kapillare als kleinere Fragmente. Eine sich daran anschließende Detektion der Moleküle findet vollautomatisiert im Endbereich der Kapillaren mit Hilfe eines fokussierten Argon-Laser statt. Dieser induziert die in die Fragmente eingebauten Fluorophore zur Emission von Fluoreszenzsignalen, die von einer hochauflösenden *Charged Coupled Device*- (CCD) Kamera detektiert werden. Im Zuge dessen werden die Fluoreszenzintensitäten jedes Farbkanals in elektronische Signale umgewandelt sowie von einer speziellen Software rechnerisch verarbeitet und als sog. Elektropherogramm graphisch dargestellt, in dem die in relativen Einheiten gemessene Fluoreszenzintensität (engl. *Relative Fluorescence Unit*, rfu) gegen die Amplikongröße aufgetragen ist. Ausgehend von einer Einzelprobe mit diploidem Chromosomensatz würde pro STR-Marker ein bzw. zwei Peaks (ein Peak repräsentiert dabei ein DNA-Fragment einer bestimmten Größe) bei homozygoter bzw. heterozygoter Lokuskonfiguration auftreten. Die Zuordnung der Fragmente zu einem bestimmten STR-System und die genaue Bestimmung ihrer Länge gelingt hierbei durch in der Software vorgegebene Einstellungen (Zuordnung aller Systeme zu jeweils einem bestimmten Farbkanal) sowie durch den Vergleich der Fragmentlängen mit einer parallel mitgeführten sog. „Allelleiter“ (eine kommerziell erhältliche Zusammenstellung aller für ein System bekannten Fragmentlängen). Auf diese Weise können alle Fragmente durch die kombinierten Eigenschaften „Farbkanal“ und „Länge in bp“ eindeutig voneinander unterschieden und zu einem STR-Profil zusammengestellt werden.

Elektropherogramme, die im Rahmen der forensischen STR-Typisierung gewonnen werden, enthalten angrenzend an echte Allel-Peaks regelmäßig deutlich geringer ausgeprägte Artefakt-Peaks (meist ≤ 15 % des rfu-Werts des allelischen Peaks), die um eine Repeateinheit kürzere DNA-Fragmente repräsentieren und als „N-1 Stutter“ oder „Stutterpeaks“ bezeichnet werden. Die Ursache jener Stutterpeaks ist die Entstehung sog. „*slippage*“-Artefakte, denen ein „Verrutschen“ des Matrizenstrangs in der DNA-Polymerase während des PCR-Prozesses zugrunde liegt (Bovo et al., 1999). Daher erfordert die Interpretation komplexer DNA-Mischungen zur Erkennung von Stutter-Peaks sowie zur Abgrenzung von gering ausgeprägten

echten allelischen Peaks ein gutes Verständnis der Stutter-Charakteristika der eingesetzten Analysemethode. Zu diesem Zweck wird eine detaillierte Analyse und Charakterisierung von Stutter-Phänomenen in Einzelproben empfohlen.

1.2.2 Sequenzierung der nächsten Generation

Die steigende Notwendigkeit an immer größer werdenden Mengen an Nukleinsäuren bis hin zu gesamten Genomen zur Sequenzierung, bedingte einen wachsenden Bedarf an einer im Vergleich zur herkömmlichen Methode schnelleren, vergleichbar genauen, aber kostengünstigeren Sequenzierungsmethode, wodurch die Entwicklung neuer Sequenzierungstechnologien vorangetrieben wurde. Die kostengünstige Generierung großer Mengen an Sequenzdaten ist dabei einer der Vorteile der neuen Methoden zum massiv parallelen Sequenzieren gegenüber der klassischen Sequenziermethode nach Frederick Sanger (Sanger et al., 1977). Letztere wurde 1977 entwickelt und das Prinzip basiert auf dem Einbau eines Didesoxyribonukleosid-Triphosphat (ddNTP) in eine wachsende DNA-Kette, wodurch eine Verlängerung des DNA-Stranges durch die DNA-Polymerase verhindert wird. Diese Methode, auch benannt als Kettenabbruch-Methode, gilt als Sequenziertechnologie der "ersten Generation" und dominierte lange den Markt.

Als Alternative zur Sanger-Sequenzierung wurde die Pyrosequenzierung im Jahre 1996 als „*Sequencing-by-synthesis*“ (SBS) Methode vorgestellt (Ronaghi, 2001). SBS-Techniken hängen dabei von einzelnen chemischen Signalen ab, die jeweils den Einbau eines oder mehrerer Desoxyribonukleosid-Triphosphate (dNTP) in den neu synthetisierten DNA-Strang darstellen. Daher werden der Sequenzierungsreaktion nacheinander dNTPs zur Verfügung gestellt. Der Ablauf einer Pyrosequenzierung gliedert sich in mehrere Phasen, in denen mittels einer Enzymkaskade unter Beteiligung der Enzyme DNA-Polymerase, ATP-Sulfurylase, Luziferase und Apyrase sichtbares Licht erzeugt wird. Dieses wird von einer CCD-Kamera in Echtzeit erfasst, sodass eine Elektrophorese der Sequenzierprodukte (wie bei der Sanger-Methode) nicht erforderlich ist und die chronologisch gemessene Lichtintensität, die den Einbau bestimmter Basen anzeigt, stattdessen in einem sogenannten Pyrogramm dargestellt wird. Die Pyrosequenzierung lief im Vergleich zur Sanger-Sequenzierung kostengünstiger und deutlich schneller ab, weshalb diese Methode für die mtDNA-Sequenzierung (Andréasson et al., 2002) und später für die STR-Sequenzierung (Divne et al., 2010) eingesetzt wurde. Obwohl

die Pyrosequenzierung im Hinblick auf die praktische Anwendung nicht weit verbreitet war, hatte die innovative Idee dieser Sequenzierung „in Echtzeit“ große Auswirkungen auf theoretischer Ebene und gilt als Meilenstein in der Fortentwicklung der DNA-Sequenzierungstechnologie mit der Einführung der MPS-Techniken.

Die im Jahre 2005 erste kommerziell verfügbar gewordene Sequenzierungsplattform der sog. „nächsten Generation“, der *Genome Sequencer 20* (454 Life Sciences), nutzte die Pyrosequenzierung als technologische Grundlage (Margulies et al., 2005). Seither konnten weitere Hochdurchsatz-Plattformen und Sequenziermethoden entwickelt und in den Markt eingeführt werden. Mit der Verbreitung von Benchtop-Sequenziergeräten wie den *Ion Torrent PGM* (Thermo Fisher Scientific), *MiSeq* (Illumina) oder *MiSeq FGx* (Verogen) wurde die Technologie zunehmend „demokratisiert“, d.h. auch für kleinere Labore finanziell tragbar. Somit konnte der Einsatz dieser Technologie beispielsweise im Bereich der klinischen Diagnostik realisiert werden. MPS ermöglicht die simultane Sequenzierung mehrerer Millionen DNA-Fragmente in relativ kurzer Zeit und ist somit nicht von den Einschränkungen der klassischen Sanger-Sequenzierung betroffen. Mit der Einführung derartiger Hochdurchsatz-Verfahren reduzierten sich Kosten und Aufwand für die Sequenzierung größerer Mengen DNA bis hin zu ganzen Genomen bei geringer werdendem Arbeitsaufwand und steigender Qualität der Daten (Goodwin et al., 2016; Margulies et al., 2005). Neue Anwendungsmöglichkeiten, wie z.B. im Bereich der personalisierten Medizin oder der nicht-invasiven Pränataldiagnostik, rückten die Sequenzierung in den Fokus molekularbiologischer und medizinischer Forschung (Mardis, 2011; Yuan et al., 2013). Die Möglichkeit, das gesamte Genom vieler verwandter Organismen zu sequenzieren, hat zudem die Durchführung groß angelegter evolutionsbiologischer Studien ermöglicht, die noch vor wenigen Jahren nicht realisierbar gewesen wären.

Mit der Einführung des *minION* im Jahr 2014 stellte die Firma Oxford Nanopore Technologies die nachfolgende Generation des Sequenzierens in Form hochmobiler und per USB-Schnittstelle steuer- und versorgbarer *Long-Read*-Sequenzierplattformen vor, die Nukleinsäuresequenzierungen in Echtzeit und ohne vorgeschaltete Amplifikation und bei vergleichsweise moderaten Anschaffungskosten der Geräte ermöglicht (Kchouk et al., 2017; Liu et al., 2012; Zaaijer et al., 2017). Als Nachfolger der Sequenzierung durch Synthese (z.B. Illumina-Sequenzierung) wird diese *Long-Read*-Sequenzierungstechnologie auch als *Third-Generation*-Sequenzierungsmethode eingestuft.

1.2.3 Massively Parallel Sequencing in der forensischen Molekulargenetik

Neben der CE hat sich die MPS-Technologie im Bereich der forensischen Molekularbiologie ebenfalls als hilfreiches Werkzeug erwiesen (Børsting and Morling, 2015). Die Analyse klassischer forensischer STR-Marker mittels MPS weist dabei mehrere Vorteile gegenüber der üblichen CE-Analyse auf (Tabelle 2). Beispielsweise ermöglicht MPS eine bessere genetische Trennschärfe („*power of discrimination*"), da Allele nicht nur anhand der Anzahl ihrer Wiederholungseinheiten, mithin der Fragementlänge, sondern zusätzlich anhand der tatsächlichen STR-Sequenz auf Nukleotidebene charakterisiert werden können (Bornman et al., 2012; Dalsgaard et al., 2013; Gettings et al., 2016). Analog zur CE-basierten Typisierung ist ein Multiplexing möglich, bei dem jedoch wesentlich mehr STR-Systeme und damit mehr Unterscheidungsmerkmale erfasst werden können, da keine Limitationen hinsichtlich des Amplikon-Layouts und der Farbkanäle zu berücksichtigen sind. Darüber hinaus sind Amplikon-Längen flexibler gestaltbar, da die Zuordnung nicht mehr anhand der Kombination von Farbstoffmarkierung und Fragmentlänge, sondern mittels der Nukleotidsequenz erfolgt. Für das Multiplex-Design bedeutet dies mehr Freiheitsgrade bei der Auswahl optimaler Primer-Bindestellen für ein einheitliches Amplikon-Design. So kann die präferenzielle Amplifikation kurzer Fragmente vermieden werden, um ein möglichst vollständiges Profil aus z.B. degradierter DNA zu erhalten.

Da die PCR-Produkte bei diesem Verfahren anhand ihrer Sequenz einem STR-System zugeordnet werden, entfällt die Größenseparation inklusive ihrer Fehlerquellen, wie dem Einfluss der elektrophoretischen Mobilität durch etwaige Sekundärstrukturausbildungen von DNA-Fragmenten. Eine Allelleiter als Längenstandard wird für die Auswertung der genetischen Daten nicht benötigt. Anhand der ermittelten Sequenzinformation können Allele identischer Länge (sog. Isoallele) z.B. mit Hilfe einer Punktmutation differenziert und von Artefakten abgegrenzt werden. Neben der eigentlichen Repeat-Region lassen sich zusätzlich Polymorphismen in den flankierenden Bereichen in die Auswertung einbeziehen, wodurch das Diskriminierungspotential abermals erhöht wird (Gettings et al., 2016; van der Gaag et al., 2016).

Mehrere Studien zeigen die Erhöhung der effektiven Anzahl unterscheidbarer Allele durch die Sequenzierung der STR-Motive und ihrer Variationen (Dalsgaard 2013, Gettings 2016). Insbesondere bei Spurenmischproben und Defizienzfällen in der Abstammungsanalyse

erweisen sich die enorme Sensitivität der Technologie und das verbesserte Auflösungsvermögen als vorteilhaft (Ma et al., 2016; Rockenbauer et al., 2014). Weiterhin ermöglicht die Detektion seltener Sequenzvarianten in der Abstammungsanalyse die verbesserte Detektion verwandtschaftlicher Verhältnisse besonders in nur teilweise zu untersuchenden oder komplexen Stammbäumen.

Tabelle 2: Vergleich der heute gängigen STR-Analyseverfahren.

Kapillarelektrophorese (CE)	**Massively Parallel Sequencing (MPS)**
▪ Unterscheidung von Stutter und echten Allelen in Mischspuren oft nicht eindeutig	▪ STR-Allele können zusätzlich anhand der Nukleotidsequenz identifiziert werden. Dadurch Verbesserung der Mischspurenanalyse
▪ Limitierte Multiplexfähigkeit der STR-Marker	▪ Größere Flexibilität bei der Gestaltung von Amplikons und der Multiplexkapazität
▪ Einfacher Laborworkflow, der schnell in die Routine etabliert werden kann	▪ Komplexere methodische Arbeitsabläufe und bioinformatische Auswertung der Sequenzdaten erforderlich
▪ Insgesamt niedrige Kosten pro Probe	▪ Pooling von Proben erforderlich, um die Kosten pro Probe zu senken

Der Einsatz von MPS als vergleichsweise junge Sequenziertechnologie bietet enorme Vorteile hinsichtlich Sensitivität und Durchsatz, birgt aber gleichermaßen nicht zu unterschätzende technische Risiken und Herausforderungen. Als neuartiges Analysewerkzeug in der forensischen Genetik findet diese Technologie in der forensischen Routinediagnostik noch keine verbreitete Anwendung und ist daher in seinen Eigenschaften nicht vollständig charakterisiert. Die erste Akkreditierung nach DIN EN ISO/ IEC 17025 in Europa für die routinemäßige STR-Typisierung mittels MPS hat das forensische DNA-Labor der Universität Leiden erhalten (Bruijns et al., 2018).

Während in verschiedenen Laboren die einzelnen Arbeitsschritte mittlerweile bereits gut etabliert sind, ist die Generierung, Speicherung, Auswertung und Interpretation der MPS-Sequenzdaten mit großen Herausforderungen verbunden, die insbesondere bioinformatische wie auch finanzielle Aspekte betreffen. Einige dieser Herausforderungen wurden in einer vom *DNA Seq Ex*-Konsortium in Zusammenarbeit mit der Arbeitsgruppe des *European Network of Forensic Science Institutes* (ENFSI) durchgeführten Umfrage ermittelt (Alonso et al., 2017): 1) das Fehlen einer einheitlichen Nomenklatur; 2) die mangelnde Kompatibilität mit bestehenden

nationalen DNA-Datenbanken; 3) das Fehlen genetischer Populationsdaten zur Unterstützung statistischer Berechnungen und 4) die deutlich höheren Kosten pro Probe im Vergleich zur CE-Technologie.

In Anbetracht der großen Datenmengen, die mittels MPS generiert und im Anschluss analysiert werden müssen sowie den damit verbundenen Schwierigkeiten und dem erhöhten organisatorischen Aufwand, ergriffen mehrere forensische Gesellschaften und Arbeitsgruppen (Parson et al., 2016; Scientific Working Group on DNA Analysis Methods, 2016) und verschiedene internationale Forschungsverbünde (Gettings et al., 2017; Köcher et al., 2018) die Initiative, zur Bewältigung der mit dieser Technologie verbundenen Herausforderungen beizutragen. So hat die *International Society of Forensic Genetics* (ISFG) Kriterien für MPS-STR Sequenzdatenanalysen herausgegeben und sich mit der Notwendigkeit eines konsistenten und plattformunabhängigen Nomenklatursystems befasst (Parson et al., 2016). Die *Scientific Working Group on DNA Analysis Methods* (SWGDAM) hat der forensischen Community vor einigen Jahren bereits überarbeitete Richtlinien für DNA-Analysemethoden und MPS-bezogener Validierungsstudien zur Verfügung gestellt (Scientific Working Group on DNA Analysis Methods, 2016).

1.2.4 Kommerziell erhältliche MPS-Panels

Bisher wurden drei kommerzielle STR-Multiplexpanels entwickelt, die dabei überwiegend Marker des amerikanischen *Combined DNA Index System* (CODIS) und die Kern-Loci des *European Standard Set* (ESS) abbilden. Teilweise wurden diese Panels vom Hersteller bereits wieder überarbeitet. Die Firma Verogen hat mit dem *ForenSeq DNA Signature Prep Kit* ein kommerzielles Analysepanel mit über 200 forensischen Markern vorgestellt, das für den Einsatz auf dem *MiSeq FGx* Gerät (Churchill et al., 2016) ausgelegt ist. Für die herkömmliche Version des *MiSeq*-Geräts (Illumina) hat die Firma Promega das *PowerSeq 46GY* Panel entwickelt, welches insgesamt 23 autosomale und 23 Y-STR-Loci kombiniert. Das *Precision ID Globalfiler* Panel der Firma Thermo Fisher umfasst insgesamt 35 autosomale STR-Loci und ist ausschließlich mit der Geräteplattform *S5-System* (Thermo Fisher) kompatibel. Viele unterschiedliche Arbeitsgruppen haben bereits zur ausführlichen Validierung der zuvor aufgeführten MPS-Panels beigetragen und die Ergebnisse der wissenschaftlichen Community

zur Verfügung gestellt (Churchill et al., 2016; Köcher et al., 2018; Müller et al., 2018; van der Gaag et al., 2016).

Wie bereits erwähnt, zählt das System SE33 aufgrund seiner hochgradig polymorphen und heterogenen Struktur zu den informativsten und komplexesten untersuchten Tetranukleotid-Markern (Polymeropoulos et al., 1992). Abgesehen von dem *ForenSeq Signature Prep Kit* sind derzeit in keinem anderen kommerziell erhältlichen Multiplexpanel Primersequenzen für SE33 enthalten. Allerdings wird dieser Marker im Verogen-Panel aufgrund inkonsistenter Genotypisierungsergebnisse nicht innerhalb der für mit diesem Kit generierten Ergebnisse optimierten Software *Universal Analysis Software* (AUS) ausgewertet. Nach Aussage der Firma Verogen soll SE33 allerdings Teil des bislang nicht kommerziell erhältlichen *ForenSeq MainstAY SE Kits* werden (Vortrag im Rahmen 42. Spurenworkshops der Spurenkommission und der Deutschen Gesellschaft für Rechtsmedizin).

1.2.5 Hochdurchsatzsequenzierung mittels MiSeq-Plattform

Das charakteristische Prinzip aller Sequenzierplattformen ist die Einteilung in die Prozessabschnitte *Library-Preparation*, *Template-Preparation*, Hochdurchsatz-Sequenzierung und bioinformatische Auswertung der Sequenzdaten (Head et al., 2014). Dabei unterscheidet die Kombination spezifischer Protokolle eine Technologie von der jeweils Anderen. Im Rahmen dieser Dissertation wurde für die Hochdurchsatz-Sequenzierung die *MiSeq*-Plattform (Illumina) verwendet, aufgrund dessen im nachfolgenden Abschnitt der für dieses Gerät zugrundeliegende Sequenzierprozess beschrieben wird.

Der *MiSeq*, dessen zugrundeliegende Sequenzierchemie auf einem zyklischen, reversiblen Kettenabbruch und der zeitgleichen Zugabe aller vier unterschiedlich-markierter und 3‘-blockierter Nukleotide innerhalb eines Sequenzierzyklus beruht (Ju et al., 2006), ist im Jahr 2011 auf den Markt gekommen. Diese Technologie hat sich bewiesen, da Illumina auf eine lange Erfolgsgeschichte einzelner Hochdurchsatz-Geräte, wie z.B. den *HiSeq* (Goodwin et al., 2016), zurückblicken kann. Bei dem *MiSeq* handelt es sich um eine „Low-Throughput“ Version des *HiSeqs* mit einer deutlich kürzeren Sequenzierlaufzeit, je nach verwendetem Protokoll. Zusätzlich bietet der *MiSeq* eine Leselänge von 300 bp sowie eine Sequenzierkapazität von bis zu 15 Gigabasen (GB). Charakteristisch ist das Zerlegen der genomischer DNA in kurze Fragmente, weshalb der Workflow mit der ersten mPCR beginnt, innerhalb derer die

forensischen Ziel-Targets amplifiziert werden. Die nachfolgende Index-PCR dient der Erstellung einer Sequenzierbibliothek (*Library*), indem plattformspezifische Oligonukleotide, sog. Indexadapter, an die Enden der DNA-Fragmente ligiert werden. Diese Indexadapter werden zur eindeutigen Markierung einer bestimmten *Library* verwendet, wodurch das Pooling von bis zu 96 Proben in Abhängigkeit vom jeweils verwendeten Indexing-Kit innerhalb eines Sequenzierlaufs ermöglicht wird. Die klonale Amplifikation der generierten *Templates*, in der Millionen identische Kopien erzeugt werden, erfolgt *in vitro* in einer polymeraseabhängigen Reaktion (*Template-Preparation*). Anschließend läuft die Hochdurchsatz-Sequenzierung als automatisierter, sich wiederholender Prozess ab, in dem komplementäre Nukleotide schrittweise entlang des *Template*-Stranges eingebaut werden.

Bei der Illumina-Technologie kommt für diesen Zweck die sog. *Bridge amplification* (Abbildung 2) zur Generierung räumlich getrennter *Cluster* auf der Oberfläche einer Fließzelle zum Einsatz (Glenn, 2011). Dieser planare Reaktionsträger ist mit Primern beschichtet, die komplementär zu den Indexadaptern der *Templates* sind. Hybridisiert nun ein einzelsträngiges *Template* an einen Primer, wird es kopiert und anschließend denaturiert, wobei der neu synthetisierte Strang kovalent an der Oberfläche gebunden bleibt. Nachfolgend bildet das freie Ende des Strangs eine „Brücke“ durch Hybridisierung mit einem weiteren noch freien Primer auf der Oberfläche der Fließzelle, womit eine erneute Strangsynthese eingeleitet wird. Auf diese Weise entstehen aus jedem einzelnen *Template* viele gebundene *Cluster*.

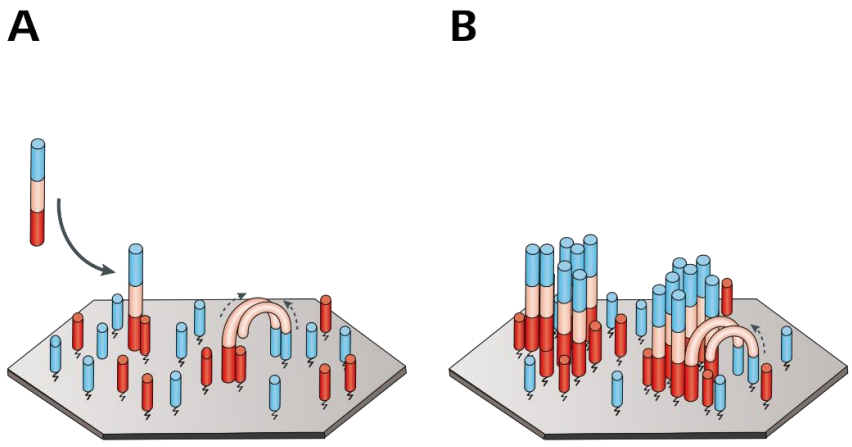

Abbildung 2: Schematische Darstellung der Bridge amplification. A) Hybridisierung des DNA-*Templates* über eine rote Indexadaptersequenz an komplementäre Primer, die auf dem planaren Reaktionsträger immobilisiert sind. **B)** das freie Ende des DNA-*Templates* bildet im Anschluss eine Brücke, um an die zweite (blaue) Primersequenz zu hybridisieren und die Amplifikation einzuleiten. Nach mehreren Amplifikationszyklen wird ein Produkt von mehreren Millionen klonalen *Clustern* generiert (aus Goodwin et al., 2016, Seite 336).

Im Rahmen der daran anknüpfenden parallelen Hochdurchsatzsequenzierung werden fluoreszenzmarkierte und reversibel mit einem Terminator versehene dNTPs eingesetzt (Ansorge, 2009; van Dijk et al., 2014). Anhand des Fluoreszenzsignals aus jedem *Cluster* wird die Identität jeder eingebauten Base in Echtzeit mittels des *Sequencing-by-synthesis*-(SBS) Ansatz und einer CCD-Kamera bestimmt (Abbildung 3). In jeder Sequenzierrunde wird die Fließzelle mit allen vier Nukleotiden durchströmt und ein komplementäres Nukleotid entsprechend der Sequenzen der *Cluster* eingebaut. Durch chemische Abspaltung des Terminators kann der Einbau der nächsten komplementären Base ermöglicht werden. Der Nachteil der SBS-Technik sind verhältnismäßig kurze Leselängen von ca. 400-500 bp, da eine 100 %ige Effizienz des Baseneinbaus und der Basenabspaltung in jedem Sequenzierzyklus nur schwer zu erreichen ist. Angesichts der Tatsache, dass das *Base Calling* durch direkte Messung der jeweiligen Signalintensität erfolgt, fallen die Fehlerraten der Illumina-Technologie im Vergleich zu anderen MPS-Geräten stark reduziert aus (Ross et al., 2013).

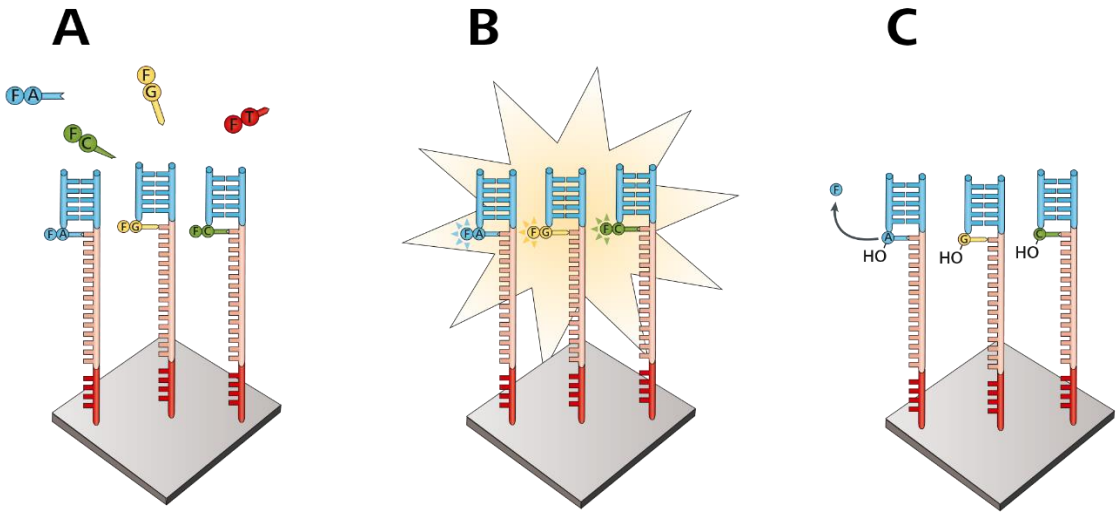

Abbildung 3: Schematische Darstellung des Sequencing-by-synthesis Ansatzes. A) Zugabe aller vier fluoreszenzmarkierter Nukleotide, die reversibel an der 3'-Hydroxylgruppe durch einen Terminator blockiert sind und komplementärer Einbau eines einzelnen Nukleotids in das DNA-*Template*, welches fest auf der *Flow Cell* immobilisiert ist. **B)** Fluoreszenzsignal nach jedem erfolgreichen Nukleotideinbau; **C)** Chemische Abspaltung des Terminators, wodurch der Zyklus erneut beginnen kann (aus Goodwin et al., 2016, Seite 340).

Die Mehrheit der forensischen Labore verwendet für die anschließende bioinformatische Analyse der Sequenzdaten die von dem jeweiligen Hersteller bereitgestellte Analysesoftware.

Die *Universal Analysis Software* der Firma Illumina stellt beispielsweise Informationen zu bestimmten Qualitätskriterien zur Verfügung, wodurch eine Interpretation des Sequenzierlaufs vor der eigentlichen Auswertung der STR-Sequenzdaten ermöglicht wird. Im Laufe der Zeit gelangen jedoch nicht nur Innovationen hinsichtlich neuer MPS-Methoden, sondern auch im Hinblick auf Software-Tools zur Analyse von STR-Sequenzdaten. Dazu gehören beispielsweise *lobSTR* (Gymrek et al., 2012), *STRait Razor* (Warshauer et al., 2013), *MyFLq Framework* (Van Neste et al., 2014), *SEQ Mapper* (Lee et al., 2017) und *FDStools* (Hoogenboom et al., 2017), die frei zugänglich sind. Allerdings setzen diese Tools bioinformatische Grundkenntnisse voraus, da die Anwender mittels einer Kommando-Zeile navigieren müssen (Huszar et al., 2021). Für die Auswertung der in dieser Dissertation generierten Sequenzdaten wurde die Web-basierte Software *toaSTR* (Ganschow et al., 2018) verwendet, die im Rahmen einer Vorstudie zu dieser Arbeit entwickelt wurde. Angesichts einer intuitiven und nutzerfreundlichen Oberfläche für die Analyse von Sequenzdaten mittels FASTQ- oder FASTA-Dateien, erfordert diese Software keine bioinformatischen Vorkenntnisse (s. auch Kapitel 2.2.5.7).

1.3 Entwicklung und Optimierung von monSTR

Mittels Einsatzes der mPCR können mehrere genetische Zielregionen parallel amplifiziert werden. Die Leistungsfähigkeit des Analysepanels hängt dabei von einer Reihe kritischer PCR-Faktoren ab, weshalb man auch von einem multifaktoriellen Prozess spricht. Zu diesen Faktoren zählen neben der Auswahl und Konzentration geeigneter Reaktionskomponenten auch die Zyklusbedingungen und -anzahl des zugrundeliegenden PCR-Protokolls. Gleichzeitig müssen die verwendeten Primersequenzen ähnliche *Annealing*-Temperaturen aufweisen, um innerhalb eines Multiplexansatzes parallel amplifiziert werden zu können.

Von einem forensischen STR-Multiplex-Kit erwarten die Anwender z.B. eine spezifische und ausbalancierte Amplifikation über alle STR-Systeme. Gleichzeitig sollen dabei möglichst wenige technische Artefakte (z.B. Stutter) und unspezifische Produkte entstehen. Diese Eigenschaften sind jedoch unter Verwendung der jeweiligen Herstellerprotokolle zumeist nicht erfüllt, weshalb zunächst eine Optimierung der einzelnen Reaktionskomponenten erforderlich ist. Derartige Laborversuche haben sich in der Vergangenheit als äußerst aufwendig und kostenintensiv erwiesen, da die gleichzeitige Optimierung unterschiedlicher Komponenten

zahlreiche Einzelexperimente voraussetzen (Henegariu et al., 1997). Es wird zwischen zwei Strategien zur Optimierung einer mPCR unterschieden:

- Der **One-factor-at-a-time-** (OFAT) Ansatz nimmt die Optimierung nur eines Faktors vor, während alle anderen Einflussfaktoren konstant bleiben. Daraufhin wird die beste Einstellung beibehalten und der nächste Faktor untersucht. Vorteilhaft an diesem Verfahren ist die einfache, schnelle und relativ kostengünstige Vorgehensweise. Demgegenüber stehen das wenig systematische Vorgehen sowie die Tatsache, dass das Optimum zumeist nur durch Zufall entdeckt wird. Ferner bleiben die Wechselwirkungen der Einflussfaktoren untereinander unbekannt, wodurch das tatsächliche PCR-Optimum zumeist nicht ermittelt wird (Toms et al., 2017).

- Die **statistische Versuchsplanung** (engl.: *Design of Experiments*, DoE) ist ein Modell für die systematische Planung und Optimierung multifaktorieller Prozesse. Statistisch angelegte Experimente ermöglichen es, signifikante Faktoren und komplexe Wechselwirkungen zu identifizieren. DoE erlaubt darüber hinaus eine steile Lernkurve, d.h. dass hochinformative Daten mit einer minimal möglichen Anzahl von Versuchen bei minimal möglichen Kosten erhalten werden können. In geometrisch designten Experimenten werden mehrere Einflussfaktoren gleichzeitig variiert und im Rahmen einer statistischen Auswertung ein funktionaler Zusammenhang zwischen den jeweiligen Einflussfaktoren beschrieben (Eriksson et al., 2008). Infolgedessen kann eine gesicherte Aussage getroffen werden, ob die untersuchten Faktoren einen signifikanten Einfluss auf das Ergebnis haben. Gleichzeitig werden Wechselwirkungen unter den Faktoren berechnet und analysiert. Die anschließende Visualisierung des statistischen Models ist vergleichbar mit einer Landkarte des experimentellen Raumes, der es erlaubt, die optimale Kombination der Einflussfaktoren zu lokalisieren. Jedoch ist ein deutlich höherer Versuchsaufwand zu verzeichnen, da alle möglichen Einzelkombinationen untersucht werden müssen.

Begründet durch den Bedarf an einer kostengünstigen und dennoch wettbewerbsfähigen Alternative zu derzeit kommerziell erhältlichen forensischen MPS-Multiplexpanels, wurde das STR-Multiplexpanel *monSTR* innerhalb einer vorbereitenden Studie zur vorliegenden Arbeit

entwickelt und mittels des DoE-Ansatzes optimiert (Ganschow et al., 2019). Im Rahmen dieser Studie wurde der erste Versuch beschrieben, die statistische Versuchsplanung auf die Optimierung eines forensischen MPS-Multiplexpanels anzuwenden. Unter Zuhilfenahme biostatistischer Tools wurde zunächst ein effizientes Set aus STR-Primern entwickelt und die Funktionalität der Primer individuell überprüft. *MonSTR* wurde für die Sequenzierung auf der Illumina *MiSeq*-Plattform konzipiert und ermöglicht die gleichzeitige Amplifikation von insgesamt 20 autosomalen STR-Loci und Amelogenin innerhalb einer einzigen Reaktion. Folgende STR-Marker wurden in diesem Multiplexpanel berücksichtigt: D1S1656, D2S1338, D2S441, D3S1358, FGA, CSF1PO, SE33, D7S820, D8S1179, D10S1248, TH01, D12S391, vWA, D13S317, Penta E, D16S539, D18S51, D19S433, D21S11, Penta D, und Amelogenin. Ein Alleinstellungsmerkmal stellt dabei die Implementierung des hoch polymorphen Markers SE33 dar, das innerhalb anderer kommerziell verfügbarer MPS-STR-Multiplexpanels bisher nicht berücksichtigt wurde (Müller et al., 2018; van der Gaag et al., 2016).

Basierend auf den Erfahrungen früherer Multiplex-Optimierungsstudien wurden zunächst vier quantitative Inputfaktoren definiert (Ganschow et al., 2019), die einen Einfluss auf die mPCR haben: 1) *Annealing-Temperatur* (ATemp); 2) *Annealing*-Zeit (ATime); 3) Extension-Zeit (ETime) und 4) Anzahl der PCR-Zyklen (Cyc). Bedingt durch die nicht direkt messbare Leistungsfähigkeit eines MPS-STR-Multiplexpanels, wurden die Optimierungsziele in drei repräsentative Ergebnisdimensionen übersetzt, die aus den resultierenden STR-Profilen gewonnen werden konnten: 1) der Variationskoeffizient der Lokus-*Coverage* (engl.: *Coefficient of variation of the locus coverage*, CV), um die Interlocus-Balance zu bestimmen; 2) die *Calling Ratio* (CR) zwischen *On-Target-Reads* und Gesamt-Reads, um die Amplifikationsspezifität widerzuspiegeln und 3) die *Signal-to-Noise Ratio* (SN) von Allelen zu Artefakten, die die Tendenz zur Bildung von Stutterprodukten beschreibt.

Auf Basis der experimentellen Daten wurde ein Regressionsmodell entwickelt und zur Ermittlung einer optimalen Kombination von Faktoreinstellungen verwendet. Eine Reduzierung von ATemp, ATime und Cyc und eine Erhöhung von ETime erzielte in diesem Zusammenhang eine wesentliche Verbesserung der Leistungsfähigkeit von *monSTR* im Vergleich zu den Standard-PCR-Bedingungen vor der Optimierung. Durch die Wahl günstiger Thermocycling-Bedingungen konnte somit eine hohe Interlocus-Balance und eine geringe Bildung von Stutter-Artefakten erreicht werden, während die Amplifikationsspezifität auf einem hohen Niveau blieb.

1.4 Methodenvalidierung

In Anbetracht der Komplexität der heutzutage gängigen forensischen DNA-Analysemethoden können eine strikte Prüfung und Überwachung der Qualität der verwendeten Instrumente, einzelner Arbeitsprozesse und der Kompetenz des eingesetzten Fachpersonals als eine zwingende Voraussetzung angesehen werden, um das Vertrauen in die Ergebnisse zu gewährleisten und sind überdies in Teilen sogar gesetzlich gefordert. Für diese Überprüfung ist in Deutschland seit Januar 2010 die Deutsche Akkreditierungsstelle (DAkkS) als nationale Akkreditierungsbehörde zuständig. Der Begriff Akkreditierung stammt dabei von dem lateinischen Verb *accredere* ab, das als „Glauben schenken“ zu übersetzen ist. Bei einer Akkreditierung handelt es sich somit um eine Bestätigung von Kompetenz und Normkonformität und fordert ein Verfahren, bei dem eine unabhängige Instanz bestätigt, dass ein System oder Prozesse nicht nur konform zu den anzuwendenden Vorschriften und Normen sind, sondern Funktionen auch durch spezifische Fachkenntnisse nachweislich erfüllt werden können.

In nach DIN EN ISO/ IEC 17025 akkreditierten forensisch-genetischen Prüflaboren ist es erforderlich, neu entwickelte Analysekits vor ihrem routinemäßigen Einsatz einer detaillierten Methodenvalidierung zu unterziehen. Das Hauptziel ist dabei der erfolgreiche Nachweis der Robustheit, Zuverlässigkeit und Reproduzierbarkeit bzw. Konkordanz (Endrullat et al., 2016a; Peters et al., 2007). Im Allgemeinen können zwei Stufen der Validierung unterschieden werden: die Entwicklungsvalidierung (engl.: *Developmental Validation*) und die interne Validierung (engl. *Internal Validation*) (Scientific Working Group on DNA Analysis Methods, 2016). Erstere umfasst die Validierung neuer STR-Systeme oder Analysekits und Technologien, wohingegen eine interne Validierung der Überprüfung bereits etablierter Verfahren im eigenen Labor dient.

Aufgrund derzeit fehlender allgemeingültiger Validierungsstrategien, die für alle forensisch-genetischen Prüflabor verbindlich sind, existieren zahlreiche Institutionen auf lokaler, nationaler und internationaler Ebene, die entsprechende Richtlinien veröffentlicht haben (s. Kapitel 1.2.3), darunter die US-amerikanische *Scientific Working Group on DNA Analysis Methods* (SWGDAM). Letztere Richtlinien sind in der forensischen Gemeinschaft offiziell anerkannt und werden regelmäßig überarbeitet, sodass sie stets auf dem aktuellen Stand der sich weiterentwickelnden Technologien sind. Daher erfasste die neueste Version der Richtlinie

auch die Validierung forensischer MPS-Analysepanels (Scientific Working Group on DNA Analysis Methods, 2016). Da die SWGDAM-Richtlinien die Grundlage der im Rahmen der in dieser Dissertation beschriebenen Arbeiten durchgeführten Methodenvalidierung bildet (s. Kapitel 2.2.7), wird im Folgenden auf die zugrundeliegenden Validierungsaspekte und -untersuchungen eingegangen:

Konkordanz	Untersuchung der Fähigkeit einer Messung, Ergebnisse zu liefern, die einem Soll- oder Vergleichswert entsprechen.
Sensitivität	Beurteilung der Nachweisgrenze eines Panels mittels einer Verdünnungsreihe einer bekannten Referenz-DNA.
Wiederholbarkeit/ Reproduzierbarkeit	Prüfung der Präzision einer Analysemethode, wenn diese von derselben (Wiederholbarkeit) oder von verschiedenen (Reproduzierbarkeit) Personen durchgeführt wird.
Mischungsverhältnisse	Untersuchung von Mischproben unterschiedlicher Mischverhältnisse zur Bestimmung der Nachweisgrenze basierend auf den Allelen der Nebenkomponenten.
Speziesspezifität	Prüfung der Humanspezifität durch Analyse genomischer DNA verbreiteter Haus- und Nutztierarten (z.B. Hund, Katze und Pferd) (Iyengar and Hadi, 2014).
Forensische (Ringversuchs-)Proben	Untersuchung repräsentativer forensischer Proben aus abgeschlossenen Realfällen oder Ringversuchen (externe Leistungsbewertung eines Labors bei der Durchführung routinemäßig eingesetzter Analyseverfahren, unter Vorgabe eines standardisierten Probensatzes bekannter DNA-Profile, der in einem vorgegebenen Zeitraum analysiert werden muss).

1.5 Ziel der Arbeit

Aller Voraussicht nach wird die Bedeutung der MPS-Technologie in der forensischen Genetik stetig zunehmen. Die Sequenzierung von STR-Systemen ermöglicht eine wesentlich detailliertere genetische Charakterisierung biologischer Spurenträger als es unter Verwendung der klassischen Kapillarelektrophorese möglich ist. Nichtsdestotrotz bedarf es der Bewältigung weiterer Hindernisse, um eine flächendeckendere Akzeptanz dieser Technologie in forensischen Laboren zu erreichen. Derzeit existiert z.B. keine wirtschaftlich attraktive Alternative zu kommerziell erhältlichen Panels für den routinemäßigen Einsatz auf der Illumina *MiSeq*-Plattform, welches einen einfachen Laborworkflow ermöglicht. Darüber hinaus wird die zeitaufwendige bioinformatische Auswertung der generierten Sequenzdaten oftmals zu einer Engstelle für den Arbeitsfluss in den Laboren.

Wie in Kapitel 1.3 erwähnt, wurde im Rahmen einer Studie (nicht Teil der in dieser Dissertation beschriebenen Arbeiten) das forensische MPS-Multiplex-STR-Panel *monSTR* zur Personenidentifizierung entwickelt und mittels statistischer Versuchsplanung optimiert. *MonSTR* amplifiziert dabei die wichtigsten forensisch relevanten STR-Kernloci des ESS inkl. SE33. In der Entwicklungsstudie wurde erstmalig die erfolgreiche Anwendung der statistischen Versuchsplanung auf die multifaktorielle Optimierung eines forensischen MPS-Panels dokumentiert. Es konnte so ein Multiplexpanel mit hoher Zuverlässigkeit, ausgewogener Amplifikation über alle STR-Loci einschließlich SE33 und eine geringere Bildung von Stutterprodukten entwickelt werden.

Nach DIN EN ISO/ IEC 17025 akkreditierte forensisch-genetische Labor müssen neu entwickelte Analysemethoden und -materialien vor ihrem routinemäßigen Einsatz einer detaillierten Validierung unterziehen. Das Ziel dieser Arbeit bestand darin, die Leistungsfähigkeit von *monSTR* gemäß den empfohlenen SWGDAM-Richtlinien umfassend zu bewerten und validieren. Für diesen Zweck wurden insgesamt sechs qualitätsrelevante Aspekte der Methode experimentell untersucht, um die Robustheit und Präzision des entwickelten Panels für die spätere Routineanwendung umfassend zu erproben. Mittels der zuvor entwickelten *toaSTR*-Software wurden die Sequenzdaten analysiert sowie die Leistungsfähigkeit des Multiplexpanels durch die Bewertung der Amplifikationsspezifität, der Interlokus-Balance, der Heterozygotie-Balance und der Neigung zur Bildung technischer Artefakte (z.B. Stutter) bewertet.

2 Material und Methoden

Der praktische Teil der vorliegenden Dissertation wurde als externe Arbeit im Zeitraum von November 2017 bis Juni 2021 an der LABCON-OWL Analytik, Forschung und Consulting GmbH, Bad Salzuflen, angefertigt und vom Geschäftsführer, Herrn Prof. Dr. rer. nat. Carsten Tiemann, vor Ort betreut.

Dieses Projekt wurde in einem Zeitraum von insgesamt drei Jahren durch das Zentrale Innovationsprogramm Mittelstand (ZIM) des deutschen Bundesministeriums für Wirtschaft und Energie finanziell gefördert (Förderkennzeichen: ZF4453702).

In diesem Kapitel werden die verwendeten Materialien aufgelistet und die angewandten molekularbiologischen Methoden beschrieben. Die molekulargenetischen Analysen wurden gemäß den Bestimmungen der DIN EN ISO/ IEC 17025 durchgeführt und die entsprechenden Protokolle teilweise der jeweiligen Standard-Arbeitsanweisungen (engl.: *Standard Operating Procedure*, SOP) entnommen. Die verwendeten Analysekits, spezielle Softwaretools sowie Fachbegriffe liegen in englischer Sprache und kursiver Schreibweise entsprechend den Herstellerangaben vor.

2.1 Material

2.1.1 Reagenzien und Enzyme

Tabelle 3: Verwendete Reagenzien und Enzyme.

Reagenzien und Enzyme	Hersteller
Absolut-Ethanol 99.8 +%	Thermo Fisher Scientific, Mannheim
Beckman Coulter™ Agencourt AMPure XP	Beckman Coulter GmbH, Krefeld
Buffer EB (dt. Elutionspuffer)	QIAGEN GmbH, Hilden
GeneScan™ 500 LIZ™ dye Size Standard	Applied Biosystems, Darmstadt
Hi-Di™ Formamid	Applied Biosystems, Darmstadt
Macherey-Nagel™ Proteinase K	MACHEREY-NAGEL GmbH & Co. KG, Düren
Nuclease-freies Wasser	VWR International GmbH, Darmstadt
PhiX Control v3	Illumina, San Diego
POP-7™ for 3130/3130*xl* Genetic Analyzers	Applied Biosystems, Darmstadt
ROTI® Nukleinsäurefrei eXtra, 500 mL	Carl Roth GmbH & Co. KG, Karlsruhe
Sodium hydroxide solution, 10 M (Natriumhydroxid-Lösung)	Sigma-Aldrich Produktions GmbH, Steinheim
Tween® 20, viscous liquid, Molecular Biology Grade	Promega GmbH, Walldorf
WEN Internal Lane Standard 500	Promega GmbH, Walldorf

2.1.2 Verbrauchsmaterialien

Tabelle 4: Verwendete Verbrauchsmaterialien.

Verbrauchsmaterialien	Hersteller
Applied Biosystems™ Plate Septa, 96 well	Applied Biosystems, Darmstadt
DNA LoBind® Tubes, 0,5mL, 1,5mL	Eppendorf SE, Hamburg
Forensik-Abstrichtupfer, rund, ISO 18385, 125 mm, Viskose	SARSTEDT AG & Co. KG, Nümbrecht
Infinium BeadChip Wash Rack	Illumina, San Diego
Infinium BeadChip Glass Tray	Illumina, San Diego
KIMTECH SCIENCE Präzisionstücher	Kimberly-Clark GmbH, Koblenz
neoLab® | Eiswanne Midi aus PU, 4 l	neoLab Migge GmbH, Heidelberg
NucleoSpin Forensic Filters, bulk	MACHEREY-NAGEL GmbH & Co. KG, Düren
One Touch Pipettenspitzen 10, 100, 1000	SorensonTM BioScience, Murray
PCR-Tubes, 8er-Streifen, 0,2mL	Kisker Biotech GmbH & Co. KG, Steinfurt
PCR-Platten für Roche® LightCycler® 480 I+II	Kisker Biotech GmbH & Co. KG, Steinfurt
POP-6™ Polymer for 3130/3130*xl* Genetic Analyzers	Applied Biosystems, Darmstadt
Qubit™ Assay Tubes	Thermo Fisher Scientific, Mannheim
Ratiomed Einmal-Skalpelle	Ratiomed-Krankenhaushygieneartikel-GmbH, Schloss Holte-Stukenbrock
Serologische Pipette ROTILABO®, 10 mL	Carl Roth GmbH & Co. KG, Karlsruhe
Zentrifugenröhrchen 15+50 mL (Falkons)	Greiner Bio-One International GmbH, Kremsmünster

2.1.3 Kommerzielle Kits

Tabelle 5: Verwendete kommerzielle Kits.

Kommerzielle Kits	Hersteller
Agilent DNA 7500 Reagent Kit	Agilent Technologies Deutschland GmbH, Waldbronn
Agilent High Sensitivity DNA Kit	Agilent Technologies Deutschland GmbH, Waldbronn
EZ1 DNA Investigator Kit (48)	QIAGEN GmbH, Hilden
Gasket Kit for Chip Priming Station	Agilent Technologies Deutschland GmbH, Waldbronn
Hi-Di™ Formamid	Thermo Fisher Scientific, Mannheim
High Pure PCR Template Preparation Kit	Roche Diagnostics GmbH, Mannheim
MiSeq Reagent Kit V2	Illumina, San Diego
MiSeq Reagent Nano Kit V2	Illumina, San Diego
NEBNext® Library Quant Kit for Illumina®	New England Biolabs, Frankfurt a.M.
Nextera XT Index Kit	Illumina, San Diego
NucleoMag Tissue Kit	MACHEREY-NAGEL GmbH & Co. KG, Düren
PowerPlex® 21 System	Promega GmbH, Walldorf
PowerPlex® ES Monoplex System, SE33	Promega GmbH, Walldorf
Q5® Hot Start High-Fidelity 2X Master Mix	New England Biolabs, Frankfurt a.M.
Qubit™ dsDNA HS-Assay-Kit	Thermo Fisher Scientific, Mannheim
VeriFiler™ Direct PCR Amplification Kit	Applied Biosystems, Darmstadt

2.1.4 Laborequipment und Analysegeräte

Tabelle 6: Verwendetes Laborequipment und Analysegeräte.

Laborequipment und Analysegeräte	Hersteller
96S Super Magnet, SPRIPlate	Alpaqua Engineering, Massachusetts
ABI PRISM® 310 Genetic Analyzer	Thermo Fisher Scientific, Mannheim
Agilent 2100 Bioanalyzer System	Agilent Technologies Deutschland GmbH, Waldbronn
Blockthermostat Grant QBT1	labfish rental solution GmbH, Hamburg
Dosierhilfe, 10 mL	Brand GmbH + Co. KG, Wertheim
ErgoOne® 8-Kanal Pipetten (0,5-10 µl; 10-100 µl; 100-1000 µl)	Starlab International GmbH, Hamburg
SICCO Exsikkator	Bohlender GmbH, Grünsfeld
EZ1 Advanced System	QIAGEN GmbH, Hilden
Gefrierschränke	Siemens, München
Glasflasche, 1 Liter	Schott AG, Mainz
KingFisher Flex Purification System	Thermo Fisher Scientific, Mannheim
Kühlschränke	Siemens, München
Labortimer	A. Hartenstein GmbH, Würzburg
LightCycler® 480 Instrument II	Roche Diagnostics GmbH, Mannheim
Mastercycler® Pro S	Eppendorf SE, Hamburg
MiSeq System	Illumina, San Diego

Laborequipment und Analysegeräte	Hersteller
Pipette (0,5-10 µl; 10 -100 µl; 100-1000 µl)	Eppendorf SE, Hamburg
Pipettierhilfe	Boeckel & Co. GmbH & Co. KG, Hamburg
Qubit 2.0 Fluorometer	Thermo Fisher Scientific, Mannheim
Scheren	Carl Roth GmbH & Co. KG, Karlsruhe
Tischzentrifugen	SIGMA Laborzentrifugen GmbH, Osterode
TruSeq Index Plate	Illumina, San Diego
Vacuum Gaspumpe	VWR International GmbH, Darmstadt
Vortex Mixer	IKA®-Werke GmbH & CO. KG, Staufen
Zentrifugen	Hettich | Andreas Hettich GmbH & Co. KG, Tuttlingen

2.1.5 Software und Webseiten

Tabelle 7: Verwendete Software und Webseiten.

Software und Webseiten	Hersteller
2100 Expert Software	Agilent Technologies, Kalifornien
ABI PRISM 310 Genetic Analyzer Data Collection Software	Applied Biosystems, Darmstadt
Adobe Photoshop 23.0.2	Adobe Systems Software Ireland Limited, Ireland
Adobe InDesign 17.0.1	Adobe Systems Software Ireland Limited, Ireland
BBtools	BBTools Google Group (https://jgi.doe.gov/data-and-tools/bbtools/)
FastQC	Babraham Bioinformatics, Cambridge
GenBank®	National Center for Biotechnology Information, Bethesda MD
GeneMapper® ID software v3.2	Thermo Fisher Scientific, Mannheim
Illumina Experiment Manager (IEM)	Illumina, San Diego
Illumina Sequencing Analysis Viewer (SAV)	Illumina, San Diego
Lightcycler® 480 Software	Roche, Mannheim
MFEprimer-3.0	https://mfeprimer3.igenetech.com/
Microsoft Excel und Word	Microsoft Cooperation, Washington
MPprimer v1.5	https://doi.org/10.1186/1471-2105-11-143
NEB Tm Calculator	New England BioLabs, Frankfurt a.M.
Primaclade	http://www.umsl.edu/~biology/Kellogg/primaclade.html
R v3.5.3	Foundation for Statistical Computing, Vienna
STRBase	National Institute of Standards and Technology, Gaithersburg
toaSTR v1.0.0-beta.16	LABCON-OWL GmbH, Bad Salzuflen

2.1.6 Biologisches Referenzmaterial

Für den praktischen Teil dieser Dissertation wurden die folgenden Referenzmaterialien und Ringversuchsproben verwendet:

1. *Standard Reference Material* (SRM) 2391c A - C des *National Institute of Standard and Technology* (NIST, USA) (National Institute of Standards & Technology, 2018; Steffen et al., 2015)
2. Männliche 2800M Kontroll-DNA (DD7101, Promega)
3. Forensische Ringversuche 54 & 55, 56 & 57, 58 & 59 und 60 & 61 aus den Jahren 2017 bis 2020 der *German DNA Profiling Group* (GEDNAP, https://www.gednap.org/de/)

2.1.7 Biologisches Probenmaterial

Für die Speziesspezifitätsstudie wurden insgesamt drei Haustierarten (Hund, Katze und Pferd) und deren Haltern DNA in Form von Mundschleimhautabstrichen entnommen. Dazu wurden Mundschleimhautzellen von jeder Wangeninnenseite mittels einer sterilen Bakteriette (Sarstedt) durch Abreiben für 30 s gesammelt. Die Abstriche wurden nach der Entnahme getrocknet und bis zur weiteren Verarbeitung in einem Exsikkator gelagert.

2.1.8 Ethikvotum

Ein positives Votum der Medizinischen Ethikkommission des Universitätsklinikums Ulm, Deutschland, liegt für diese Dissertation vor (Antrag 267/18).

2.2 Methoden

2.2.1 Auswahl der forensischen STR-Systeme und Primerdesign

Die Auswahl der STR-Marker für das forensische MPS Multiplex-Panel, welches nachfolgend *monSTR* bezeichnet wird, umfasst die folgenden 21 STR-Marker: D1S1656, D2S441, D2S1338, D3S1358, FGA, CSF1PO, SE33, D7S820, D8S1179, D10S1248, TH01, vWA,

D12S391, D13S317, Penta E, D16S539, D18S51, D19S433, D21S11, Penta D und Amelogenin. In der vorliegenden Arbeit wurde für die Amplifikation ein spezifisches Primerset auf der Grundlage der zuvor aufgeführten STR-Marker verwendet, welches im Rahmen einer Vorstudie entwickelt und optimiert wurde (Ganschow et al., 2019). Für das Primerdesign wurden die folgenden bioinformatischen Tools verwendet:

1. *MPprimer*

Mittels der Software *MPprimer* (Shen et al., 2010) wurden zunächst 42 Kandidaten-Primer parallel generiert, die für ein Multiplexing geeignet sind. Folgende Qualitätskriterien wurden dabei innerhalb der Konfigurationsdatei für jeden STR-Marker festgelegt:

1. Vergleichbare thermodynamische Eigenschaften der Primer in Bezug auf Länge (22 bp), GC-Gehalt (50 %) und Schmelztemperatur (64 °C, *NEB Tm Calculator*, s. 2.1.5).
2. Amplikongrößen, die für eine Sequenzierung auf der *MiSeq*-Plattform kompatibel sind (250-350 bp). Das System Amelogenin bildet hier eine Ausnahme, da es kürzer konfiguriert wurde, um als Degradationskontrolle für die Amplifikation forensischer Proben zu dienen.
3. Keine Überlappung der Primerbindestellen mit häufig vorkommenden SNPs. Die entsprechenden Positionen wurden innerhalb der Konfigurationsdatei angegeben.
4. Ähnliche Amplikongrößen, um eine bevorzugte Amplifikation kürzerer Targetsequenzen durch die DNA-Polymerase zu verhindern.

Alle anderen Parameter wurden in ihren Standardeinstellungen belassen. Die Referenzsequenzen von 20 STR-Systemen einschließlich annotierter flankierender Regionen wurden von Gettings et al. (2015) bezogen. Die Sequenzinformation des geschlechtsspezifischen Markers Amelogenin wurden aus der *NCBI GenBank* Datenbank entnommen (Zugangsnummer NG_008011.1). SNP-Positionen und Allelhäufigkeiten wurden aus dem Datensatz des *1000 Genomes Project Phase 3* erhalten. Das verwendete Referenzsequenz-Allel sollte nur eines der möglichen Allele darstellen, weshalb die gewünschte Amplikongröße für jeden STR-Lokus entsprechend der durchschnittlichen Allellänge gemäß *STRBase* (s. Kapitel 2.1.5) angepasst wurde.

2. *Primaclade*

Die Kandidaten-Primer wurden im Anschluss manuell auf ein zusätzliches Kriterium überprüft, welches im *MPprimer*-Prozess nicht implementiert werden konnte. Die erwarteten Amplikons sollten möglichst viel Variation innerhalb der flankierenden STR-Region aufzeigen, wobei die maximale Amplikongröße zu berücksichtigen galt. Daher wurden ggf. einzelne Primer mit Hilfe des visuellen Tools *Primaclade* (Gadberry et al., 2005) neu konstruiert, wobei die oben genannten Anforderungen eingehalten wurden.

3. *MFEprimer*

Zur theoretischen Abschätzung unspezifischer (Kreuz-) Amplifikation zwischen einzelnen Primerpaaren, Primer-Dimerisierungen und Haarnadelbildungen wurde die Software *MFEprimer* (Qu et al., 2012) verwendet. Eine *in-silico* PCR wurde mit dem überarbeiteten Primer-Satz durchgeführt. Als DNA-Referenz wurde das menschliche Referenzgenom *GRCh38.p10* ausgewählt. Unspezifische Ergebnisse wurden herausgefiltert, indem die maximale Amplikongröße auf 1000 bp und die minimale Schmelztemperatur auf 50 °C gesetzt wurde. Alle anderen Parameter wurden auf ihren Standardeinstellungen belassen.

2.2.2 Anfügen der illuminaspezifischen Überhangadapter

Nextera XT Überhang-Adaptersequenzen (15044223 Rev. B, Illumina) wurden im Anschluss an die 5'-Enden der finalen *Forward*- und *Reverse*primer angefügt, um die Kompatibilität der Amplikons auf der *MiSeq*-Plattform zu gewährleisten:

Forward-Überhangadapter:

5’ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG- [Targetspezifische Sequenz]

Reverse-Überhangadapter:

5’ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG- [Targetspezifische Sequenz]

Die Primersequenzen wurden bei TIB Molbiol (Berlin, Deutschland) synthetisiert und in lyophilisierter Form bezogen. Vor der Verwendung wurden die Oligonukleotide gemäß Herstellerprotokoll mittels Elutionspuffer zunächst auf eine 100 µM Stammlösung

rekonstituiert. Hiernach wurden jeweils für *Forward*- und *Reverse*primer desselben Targets 10 µM Gebrauchslösungen hergestellt und mittels Singleplex-PCR auf einem *Mastercycler Pro S* auf Eignung überprüft. Größe und Spezifität des daraus resultierenden PCR-Produkts wurde mit einem *Agilent Bioanalyzer 2100* unter Verwendung des *DNA 7500 Kits* (G2938-90024 Rev. C) untersucht. Die Gebrauchslösungen wurden bis zur weiteren Verwendung bei -20° C gelagert. Die finalen Primer-Sequenzen und die Molaritäten im finalen *monSTR*-Plex sind in Tabelle 8 angegeben. Die Primer-Molaritäten waren im Rahmen der Optimierungsstudie bestimmt worden (Ganschow et al., 2019).

Tabelle 8: Forward und reverse Primersequenzen des Multiplexpanels monSTR. Das *monSTR*-Panel umfasst insgesamt 21 Zielregionen und die finalen Primersequenzen wurden bei TIB Molbiol bezogen. nt = Nukleotid; µM = Mikromolar; fwd = foreward; rev = reverse.

STR-Marker	Orientierung	Primersequenz in 5' → 3' Richtung	Primerlänge	Molarität im Multiplex
D1S1656	fwd	CCCCATATAAGTTCAAGCCTGT	22 nt	0,08 µM
	rev	CCAGGAAAGAGAGAAACCATGT	22 nt	
D2S441	fwd	GAGATTCTGTTCCTGAGCCCT	21 nt	0,08 µM
	rev	ACCACACCCAGCCATAAATAAC	22 nt	
D2S1338	fwd	GACTTCATGGTCCTGACTACAG	22 nt	0,10 µM
	rev	TTGTCAGTGTTCATGCCTACAT	22 nt	
D3S1358	fwd	GCCATATTCACTTGCCCACT	20 nt	0,08 µM
	rev	TCATGCCTATAATCCCAGGTACT	23 nt	
FGA	fwd	TGCAGGGCATAACATTATCCAA	22 nt	0,10 µM
	rev	AAGTAGCTGCTGAGTGATTTGT	22 nt	
CSF1PO	fwd	GAGGTAAAGGTGTCTTAAAGTGAG	24 nt	0,08 µM
	rev	GACCCTGTTCTAAGTACTTCCT	22 nt	
SE33	fwd	CCATTGCACTCCAATCTGG	19 nt	0,10 µM
	rev	CTCCCCTACCGCTATAGTAAC	21 nt	
D7S820	fwd	GACTACAGGCATGTGCTACTG	21 nt	0,08 µM
	rev	AACGATTCCACATTTATCCTCA	22 nt	
D8S1179	fwd	CTTTGCCTGAGTTTTGCTCAG	21 nt	0,12 µM
	rev	TCCAGTTTCTTTTACCAAATTGTGT	25 nt	
D10S1248	fwd	AAAAGCAAACCTGAGCATTAGC	22 nt	0,08 µM
	rev	AACCATACTTTTTCCCTCAGGT	22 nt	
TH01	fwd	CACCACATTTCAATCAAGGTCC	22 nt	0,08 µM
	rev	GGGGCAAAATTCAAAGGGTATC	22 nt	
vWA	fwd	ATTACTAGTCTCCTACAATGTGCC	24 nt	0,10 µM
	rev	GGACAGATGATAAATACATAGGATGG	26 nt	
D12S391	fwd	GGCTTCTCCAGAGAGAAAGAAT	22 nt	0,14 µM
	rev	CCACCTGGGATGGATAATTACA	22 nt	
D13S317	fwd	GGATGTGGAGGAGAGTTCATTT	22 nt	0,08 µM
	rev	GTCTGAAAGTACAAGTGGGGAA	22 nt	
Penta E	fwd	GAGATTGAAGTGAGCCGAGAT	21 nt	0,10 µM
	rev	GTAGCCTCAAACATCAGTTTCC	22 nt	
D16S539	fwd	CAGATCCCAAGCTCTTCCTC	20 nt	0,12 µM
	rev	CGAAAGTGATGCCATAGACTTAAA	24 nt	
D18S51	fwd	TTCTTGAGCCCAGAAGGTTAAG	22 nt	0,08 µM
	rev	TAAGGTGGACATGTTGGCTTC	21 nt	
D19S433	fwd	GCAACATAGTGAGACCTCTTCT	22 nt	0,08 µM
	rev	GCAGTGATTTCTGATATTTTGGTGC	25 nt	
D21S11	fwd	GTTGTTCTAAGGGCTTCAGACT	22 nt	0,08 µM
	rev	TCAATGTTCTCCAGAGACAGAC	22 nt	
Penta D	fwd	TAAATAGCCAGGCATGGTGAG	21 nt	0,10 µM
	rev	AGGTCATGATTTTGTGATATCTAAGAA	27 nt	
Amelogenin	fwd	CCCTGGTTATATCAACTTCAGCTAT	25 nt	0,11 µM
	rev	ATCAGAGCTTAAACTGGGAAGC	22 nt	

2.2.3 DNA-Isolierung

Im Rahmen der vorliegenden Arbeit wurden drei unterschiedliche DNA-Extraktionsmethoden verwendet:

1. *HighPure PCR Template Preparation Kit*

Die manuelle DNA-Extraktion der GEDNAP-Ringversuchsproben erfolgte unter Verwendung des *HighPure PCR Template Preparation Kit* (Ref. 11796828001). Dazu wurden die Watteträger mit Hilfe eines sterilen Skalpells in sechs gleiche Stücke geschnitten und diese jeweils in Reaktionsgefäße überführt. Handelte es sich um authentisches Spurenmaterial (z.B. Kondom mit Sperma, usw.) wurden zunächst Abriebe mittels eines sterilen Watteträgers angefertigt. Diese wurde anschließend mit Hilfe eines Skalpells von ihren Kunststoffstielen abgetrennt und in sechs etwa gleich große Stücke geschnitten. Ein Stück wurde daraufhin unmittelbar in die Extraktion eingesetzt und die anderen Stücke für weitere Laborschritte (nicht Teil dieser Dissertation) zurückgestellt. Die Zellen wurden zunächst mittels *Proteinase K* behandelt und die DNA in mehreren Waschschritten auf einer Silica-Membran gereinigt. Der weitere Arbeitsablauf erfolgte gemäß dem Herstellerprotokoll (Version 26, November 2019). Die genomische DNA wurde in 50 µL eluiert und die gewonnenen Extrakte bis zur weiteren Verarbeitung bei -20 °C gelagert.

2. *NucleoMag Tissue Kit*

Die für die Prüfung auf Speziesspezifität verwendeten Mundschleimhautabriebe typischer Haus- und Nutztierarten (Hund, Pferd und Katze) wurden mittels *NucleoMag Tissue Kit* (REF 744300.4) unter Einsatz des *KingFisher Flex* Geräts nach Herstellerangaben extrahiert und in 70 µL eluiert. Das Extraktionsverfahren basiert dabei auf Affinitätschromatographie vermittels beschichteter, magnetischer *Beads*. Die DNA-Extrakte wurden bis zur weiteren Verarbeitung bei -20 °C gelagert.

3. *EZ1 DNA Investigator Kit*

Mundschleimhautabstriche der Tierhalter wurden mit Hilfe des Extraktionsrobotors *EZ1 Advanced System* unter Verwendung des *EZ1 DNA Investigator Kits* extrahiert. Dazu wurden die Köpfe der Watteträger mit einem Skalpell von ihren Kunststoffstielen abgetrennt, in

Reaktionsgefäße überführt und anschließend nach Herstellerangaben automatisch extrahiert. Die genomische DNA wurde in 50 µL eluiert und die DNA-Extrakte bis zur weiteren Verarbeitung bei -20 °C gelagert.

2.2.4 Quantifizierung genomischer DNA

Die extrahierte DNA wurde mittels des *Qubit 2.0 Fluorometers* quantifiziert indem die Konzentration der genomischen DNA in Anwesenheit eines DNA-spezifischen Fluoreszenzfarbstoffes bestimmt wurde. Der Farbstoff interkaliert dabei in die Doppelhelixstruktur der DNA, wodurch sich seine Fluoreszenzintensität signifikant erhöht und ein spezifisches Signal gemessen werden kann. Die Konzentration wird dann anhand einer zuvor aus Proben bekannter Konzentrationen generierten Kalibrierungskurve bestimmt. Zur Quantifizierung der finalen MPS-*Libraries* wurde ebenfalls das beschriebene Verfahren eingesetzt.

Durchführung:
Zur Quantifizierung der genomischen DNA wurde für die vorliegende Arbeit das *Qubit DNA HS Assay Kit* (Q32851) entsprechend den Herstellerangaben bei Einsatz von 2 µL DNA-Extrakt verwendet.

2.2.5 Massively Parallel Sequencing mittels monSTR

Das Prinzip des *Massively Parallel Sequencing* (MPS) wurde in Kapitel 1.2.5 ausführlich beschrieben, ein Glossar auf Seite VIII ff. führt zudem in die wichtigsten Begriffe und Konzepte ein.
Die *Library-Preparation* mittels *monSTR* umfasst gängige molekularbiologische Methoden (Abbildung 4). Es wurde eine zielgerichtete („targeted“) Sequenzierstrategie verfolgt, die sich auf die Sequenzierung spezifischer Zielregionen (Amplikons) konzentriert. Dabei werden zunächst kurze DNA-Fragmente (<500 bp) in einer Multiplex-PCR (mPCR) amplifiziert, wobei spezifische Überhänge an den Primersequenzen in der zweiten Index-PCR als universelle Primerbindestellen dienen, um *Nextera XT* Adaptersequenzen (Illumina) zu ligieren.

Einen Schritt innerhalb der *Library-Preparation*, der eine hohe Aufmerksamkeit verlangt, stellt die Aufreinigung des PCR-Produkts dar. Dieser Prozess ist zwingend notwendig, um aus der spezifischen *Library* ungebundene Primer und residuales Enzym zu entfernen und jene so für den nächsten Arbeitsschritt vorzubereiten. Im Rahmen des *monSTR*-Workflows fand unmittelbar nach jeder Amplifikation die Aufreinigung des PCR-Produkts unter Einsatz von *Agencourt AMPure XP Beads* statt. Die Qualitätskontrolle der *Library* wurde im Anschluss mittels der *DNA 7500* und *Qubit dsDNA High Sensitivity Kits* durchgeführt. Für die finale Quantifizierung der *Library* wurde die quantitative Real-Time PCR (qPCR) unter Verwendung eines Kits von New England Biolabs eingesetzt, die als die genaueste und effektivste Methode zur Quantifizierung gilt. Wenn die *Libraries* auf die gewünschte Konzentration eingestellt und einer finalen Qualitätskontrolle unterzogen wurden, begannen die Vorbereitungen für die Hochdurchsatzsequenzierung entsprechend eines zuvor angelegten *Sample Sheets*. Im Rahmen dieser Arbeit wurde ausschließlich das *MiSeq*-Sequenziergerät der Firma Illumina eingesetzt. Die Auswertung der Sequenzdaten wurde unter Zuhilfenahme der Analysesoftware *toaSTR* (Ganschow et al., 2018) durchgeführt, um einen maximalen Informationsgehalt zu gewährleisten.

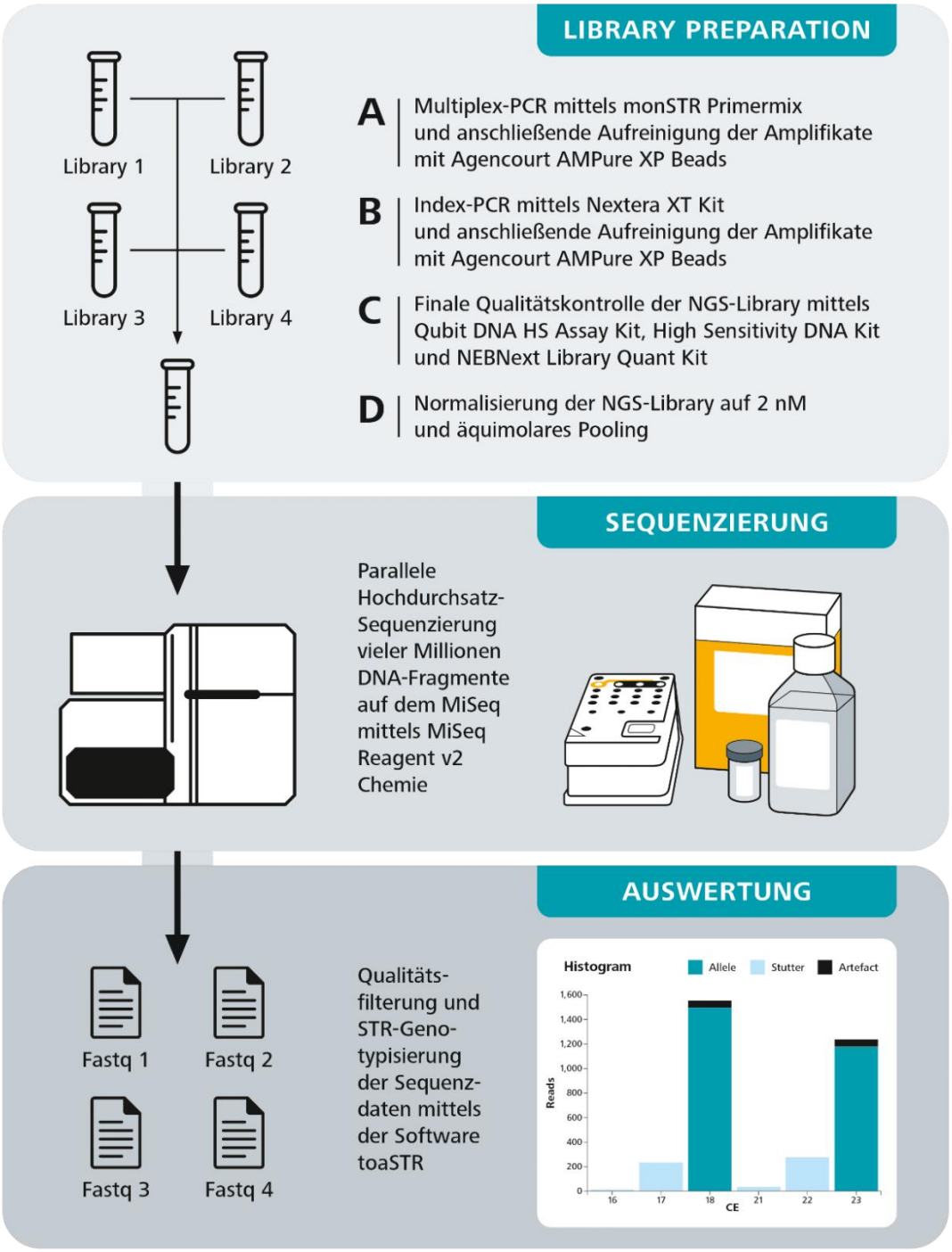

Abbildung 4: Diagramm des monSTR-Workflows. Die genomische DNA durchläuft die einzelnen Schritte der *Library-Preparation*, die sich aus der Multiplex- und Index-PCR zusammensetzen. Jede *Library* (in dieser Abbildung wurden zur besseren Übersicht insgesamt vier *Libraries* dargestellt) wird nach Amplifikation mittels *Agencourt AMPure XP Beads* aufgereinigt und anschließend mittels unterschiedlicher molekularbiologischer Methoden quantifiziert. Sobald die einzelnen *Libraries* normalisiert wurden, werden sie zu einem Pool zusammengefasst, in die entsprechende Position einer Illumina v2 Kartusche pipettiert und auf dem *MiSeq* sequenziert. Die anschließende bioinformatische Auswertung setzt sich aus der Qualitätsfilterung und der STR-Genotypisierung mittels der Software *toaSTR* (Ganschow et al., 2018) zusammen.

2.2.5.1 Multiplex-PCR

Die initiale Fragmentierung der genomischen DNA erfolgte durch eine mPCR. Hierbei handelt es sich um einen PCR-Ansatz, bei dem alle im *monSTR*-Panel enthaltenen Primerpaare (Tabelle 9) im selben Ansatz eingesetzt werden und eine parallele, simultane Amplifikation der entsprechenden Targetsequenzen ermöglichen. Wie bereits beschrieben, enthalten die *monSTR*-Primer sowohl die lokusspezifischen Sequenzen sowie einen Illumina spezifischen Überhangadapter, der als spätere Bindestelle der Indexadapter dient (s. Kapitel 2.2.2).

Durchführung:
Zunächst wurde ein Mastermix bestehend aus je 12,5 µL *Q5 Hot Start High-Fidelity 2X Mastermix* (M0515) und 5 µL *monSTR* Primermix pro Probe hergestellt. Die eingesetzte DNA-Menge für im Rahmen dieser Arbeit untersuchte Proben lag zwischen 7,8 pg und 10 ng. Der Ansatz wurde anschließend mit Nuklease-freiem Wasser auf ein Endvolumen von 25 µL aufgefüllt. Die Amplifikation erfolgte in einem *MasterCycler Pro S* gemäß nachstehendem Temperaturprofil. Die Lagerung der mPCR-Produkte erfolgte für eine kurze Zeit (< 24h) bei 4 °C oder für einen längeren Zeitraum (> 24h) bei -20 °C.

Tabelle 9: Temperaturprotokoll der mPCR.

Temperaturprofil	Dauer	PCR-Zyklen
98 °C	30 s	1 x
98 °C	10 s	25 x
58 °C	23 s	
72 °C	65 s	
72 °C	120 s	1 x
4 °C	∞	-

2.2.5.2 Aufreinigung der Amplifikate

Die Aufreinigung von Amplifikaten als Teil der *Library-Preparation* hat einen erheblichen Einfluss auf den Erfolg des Sequenzierlaufs, da z.B. ungebundene Primer oder Adapterdimere als Nebenprodukte der mPCR resultieren und als unspezifische Targets die Effizienz der folgenden Sequenzierung reduzieren können. Ziel dieses Schritts ist es daher, unspezifische Fragmente zu entfernen. Die Aufreinigung der Amplifikate erfolgte mit Hilfe von *Agencourt*

AMPure XP Magnetpartikeln. Diese Methode arbeitet nach dem Prinzip der *Solid Phase Reversible Immobilization* (SPRI), die auf ihrer Oberfläche freie Carboxygruppen tragen (DeAngelis et al., 1995). In Anwesenheit von Polyethylenglycol (PEG) und Natriumchlorid (NaCl) erlauben sie eine reversible Bindung der DNA. Durch Zugabe von Elutionspuffer (Niedrigsalzpuffer) kann diese wieder gelöst werden, wodurch eine größenbasierte Selektion innerhalb der Lösung gewährleistet wird. Die Größe der DNA-Fragmente, die von den magnetischen *Beads* gebunden werden, ist dabei abhängig von dem eingesetzten Verhältnis zwischen *Beads* und Amplifikat. Je niedriger das Volumenverhältnis, desto größer die im Eluat verbleibenden DNA-Fragmente. Kleinere Fragmente werden im Überstand zurückgehalten und i.d.R. verworfen. Im Rahmen der vorliegenden Arbeit wurden einige *in-house* Änderungen am Herstellerprotokoll (B24965AA) vorgenommen.

Durchführung:

Aufgrund der Länge der zu selektierenden DNA-Fragmente wurde für die Aufreinigung der mPCR-Produkte ein Verhältnis von 0,8 gewählt (entspricht 18 µL *Beads* und 22 µL mPCR-Produkt). Zunächst wurden in einem PCR-Streifen je Tube 18 µL resuspendierte und an Raumtemperatur (RT) adaptierte *AMPure XP Beads* vorgelegt und 22 µL mPCR-Produkt hinzugefügt. Der gesamte Ansatz wurde dann resuspendiert und für 5 min bei RT inkubiert. Nachfolgend wurde der Ansatz auf eine *SPRIPlate 96 Super Magnet Plate* gestellt und für 5 min inkubiert, wodurch sich die Magnetpartikel an der Gefäßwand ansammeln. Der klare Überstand wurde mit einer Multikanalpipette ohne das Partikelpellet zu stören abpipettiert, verworfen und zu jeder Probe 200 µL 80 %iges Ethanol (frisch angesetzt) hinzugefügt. Der Ethanol-Überstand wurde nach 30–60 s abpipettiert, verworfen und der Waschschritt wiederholt. Die PCR-Streifen wurden anschließend von der Magnetplatte genommen und für 3 min bei RT getrocknet, um die Ethanolreste vollständig zu entfernen. Das ringförmige Magnetpellet wurde in 24 µL Elutionspuffer resuspendiert und für weitere 2 min bei RT inkubiert. Der Ansatz wurde erneut für 2 min auf die Magnetplatte gestellt. 22 µL des klaren Überstands wurden in neue PCR 8er-Streifen überführt und die aufgereinigten Produkte entweder bei -20 °C gelagert oder sofort weiterverwendet.

2.2.5.3 Ligation der Index-Adapter

Unter Verwendung des *Nextera XT Index Kits* wurde in diesem Schritt jede aufgereinigte MPS-*Library* mit einer individuellen Kombination von Indexadaptern am 3‘- und 5‘-Ende markiert, um die spätere parallele Sequenzierung mehrerer Proben auf derselben *Flow Cell* zu ermöglichen. Die im Kit verwendeten Primerpaare enthalten 8 nt lange Illumina-Adapter (i7 und i5), welche der Anheftung der Fragmente auf der Oberfläche der Fließzelle dienen und die Bindestellen für die Sequenzierprimer tragen. Im Rahmen des *Demultiplexing* können mittels der individuellen Adapterkombinationen alle gelesenen DNA-Fragmente derjenigen Probe zugeordnet werden, aus der sie stammen.

Durchführung:
Zunächst wurde ein Mastermix bestehend aus 5 µL aufgereinigtem mPCR-Produkt, 12,5 µL *Q5 Hot Start High-Fidelity 2X Mastermix* und 2,5 µL Nuklease-freiem Wasser in PCR 8er-Streifen angesetzt. Auf einer *TruSeq Index Plate* wurden die *Nextera XT Index Primer* 1 (N7xx) und 2 (S5xx) entsprechend platziert und je 2,5 µL Primer 1 und 2 zum Mastermix hinzugefügt. Das Endvolumen betrug 25 µL. Die Index-PCR wurde gemäß nachstehendem Temperaturprotokoll auf einem *Mastercycler Pro S* durchgeführt.

Tabelle 10: Temperaturprotokoll der Index-PCR.

Temperaturprofil	Dauer	PCR-Zyklen
98 °C	30 s	1 x
98 °C	10 s	8 x
62 °C	30 s	
72 °C	30 s	
72 °C	120 s	1 x
4 °C	∞	-

Im Anschluss an die Adapter-Ligation fand eine weitere Aufreinigung mittels *Agencourt AMPure XP Beads* gemäß dem in Kapitel 2.2.5.2 beschriebenen Protokolls statt. Es wurde ein Verhältnis von 0,7 gewählt, was insgesamt 16 µL *Beads* und 22 µL Ligationsprodukt entspricht. Die *Libraries* wurden bei -20 °C gelagert oder sofort für die weitere Bearbeitung verwendet.

2.2.5.4 Qualitätskontrolle mittels Agilent Bioanalyzer

Bei dem *2100 Bioanalyzer Instrument* handelt es sich um ein Mikrofluidik-Gelelektrophoresesystem zur Bestimmung der Produktgröße, Qualität und Konzentration von Nukleinsäuren. Basierend auf Prinzipien der Mikrofluidik werden die Proben in einen speziellen Chip, der mit einer Gelmatrix befüllt wurde, pipettiert und elektrophoretisch aufgetrennt. Mittels in der Gelmatrix eingelagerter Fluoreszenzfarbstoffe werden die Nukleinsäurefragmente detektiert und gegen einen mitgeführten Größenstandard abgeglichen.

Durchführung:
Zur Qualitätsbestimmung der *Libraries* wurde das *High Sensitivity DNA Kit* (G2938-90321 Rev. B) entsprechend den Herstellerangaben verwendet. Das Kit ermöglicht es, 11 Proben parallel zu analysieren und Fragmentgrößen im Bereich von 50 bis 10380 bp zu messen. Die Messergebnisse wurden in Form eines Reports (.pdf-Datei) gespeichert und enthielten die Informationen zu den jeweiligen Fragmentgrößen der analysierten *Libraries*.

2.2.5.5 Quantifizierung der Libraries mittels qPCR

Eine möglichst exakte Konzentrationsbestimmung einer hergestellten *Library* ist für eine maximale Datenausbeute unerlässlich. Speziell für die Sequenzierung auf der Illumina-Plattform ist eine genaue Quantifizierung der *Libraries* entscheidend, um eine optimale *Cluster Density* zu erzielen. Die zuvor beschriebene elektrophoretische Auftrennung kann dabei nicht zwischen DNA-Fragmenten unterscheiden, an die erfolgreich i5 und i7 Adapter ligiert wurden und solchen, die keine Adapter tragen. Aus diesem Grund wurde die qPCR eingesetzt, die als die genaueste und effektivste Methode zur Quantifizierung von MPS-*Libraries* gilt. Die Bindestellen der Primer, die für diese qPCR verwendet wurden, befinden sich ausschließlich auf den i5 und i7 Adapter-Sequenzen. Dadurch wird sichergestellt, dass ausschließlich DNA-Fragmente mit ligierten Adaptern amplifiziert und somit quantifiziert werden.

Durchführung:
Die *Libraries* wurden mittels *NEBNext Library Quant Kit for Illumina* auf einem *LightCycler 480 Instrument* gemäß Herstellerangaben (Version 6.0_8/21) quantifiziert. Die Sequenzen der

im Kit mitgelieferten qPCR-Primer sind in der Sequenz der *Nextera XT Indexadapter* enthalten. Als Fluoreszenzfarbstoff wurde *SYBR Green* verwendet, welches während der Elongationsphase der qPCR in die doppelsträngige DNA interkaliert. Zur Erstellung einer Standardkurve wurden parallel vier im Kit enthaltene Standards unterschiedlicher Konzentrationen (1 pM, 0,1 pM, 0,01 pM und 0,001 pM) in gleicher Weise wie die Proben prozessiert und mitgeführt. Die einzelnen *Libraries* wurden mittels *Dilution Buffer* 1:100.000 verdünnt und in Triplikaten gemessen, um Messungenauigkeiten zu kompensieren. Die Daten der gemessenen Fluoreszenzintensitäten für die einzelnen *Libraries* wurden mittels der gerätespezifischen *Lightcycler 480 Software* erfasst und graphisch ausgewertet. Die Konzentration der NGS-*Libraries* wurde anhand der im Herstellerprotokoll angegebenen Formel berechnet. Hierfür wurde eine durchschnittliche *Library*-Größe der *monSTR*-Amplikons von 430 bp angenommen.

2.2.5.6 Hochdurchsatzsequenzierung

Die *Libraries* wurden gemäß des Illumina Herstellerprotokolls (Ref. 1000000061014 v00) für die Sequenzierung vorbereitet und auf dem *MiSeq* sequenziert.

Durchführung:

1. *Erstellung der Sample Plate*

Zur Durchführung eines Sequenzierlaufs auf dem *MiSeq* ist es zunächst notwendig, eine *Sample Plate* sowie ein *Sample Sheet* zu erstellen, in denen die einzelnen Sequenzierparameter und die Zuordnung der Indexadapter zu den jeweiligen Proben aufgelistet sind. Beide Dokumente wurden mit der Software *Illumina Experiment Manager* (IEM) erstellt. In der IEM Software wurde das Feld „*Create Sample Plate*“ ausgewählt und anschließend das Library-Prep Kit „*Nextera XT*“ bestätigt. Ein spezifischer Plattenname wurde im Feld „*Unique Plate Name*“ eingetragen und als Index-Read „*2*“ eingetragen. Im Tabulator „*Plate*“ wurden die Spalten/ Zeilen entsprechend den verwendeten Indexadaptern mittels Dropdown-Menü ausgefüllt. Die NGS-*Libraries* wurden anschließend entsprechend ihrer Indexkombination in der *Sample Plate* eingetragen und schließlich das Dokument gespeichert.

2. Erstellung des Sample Sheets

Für die Erstellung des *Sample Sheets* wurde innerhalb der IEM Software das Feld „*Create Sample Sheet*“ und im Folgeschritt der „*MiSeq*“ ausgewählt. Hernach wurden in der Kategorie „*Other*“ die Anwendung „*FASTQ Only*“ eingegeben. Unter „*Reagent Cartridge Barcode*“ wurde die individuelle Kartuschen-Nummer des verwendeten *MiSeq Reagent Kits* eingetragen und das Kit “*Nextera XT*“ ausgewählt. Für die Index-Reads wurde „*2*“ eingetragen. Zusätzlich wurden weitere Informationen wie der Experimentname, das Datum und der Bearbeitername hinterlegt. Als Read Type wurde „*Single-Read*“ ausgewählt sowie die benötigte Leselänge angeben (Single-Read: 400). Die zuvor erstellte *Sample Plate* wurde unter „*Select Plate*“ geöffnet und eingeladen. Sobald der Status „*Valid*“ erschien, wurde über „*Finish*“ das *Sample Sheet* gespeichert und auf einen USB Stick kopiert. Über „*Manage Files*“ wurde das zuvor generierte *Sample Sheet* auf den *MiSeq* geladen.

3. Pooling und Normalisierung der Libraries

Die *Libraries* wurden im Anschluss äquimolar gepoolt, um eine gleichmäßige Verteilung auf der Fließzelle zu gewährleisten. Hierzu wurde ausgehend von den in Kapitel 2.2.4.5 ermittelten Konzentrationen jede *Library* mittels Elutionspuffer auf 2 oder 4 nM eingestellt. Im Anschluss wurden je 5 µL der verdünnten *Library* zu einem äquimolaren Pool kombiniert. Der *Library*-Pool wurde im Anschluss unter Einsatz eines *Qubit 2.0 Fluorometers* in Dreifachbestimmung quantifiziert und der daraus resultierende Mittelwert für die weitere Bearbeitung herangezogen. Die weitere Bearbeitung des *Library*-Pools bestand aus einer Normalisierung auf 10 pM. Zusätzlich wurde der Pool mit 20 % *PhiX* (Begriffserklärung s. Glossar) versehen. Die weiteren Arbeitsschritte zur Verdünnung des MPS-Pools wurden wie im Pipettierschema angegeben durchgeführt (Tabelle 11).

4. Start des MiSeq-Systems und Post-Run Wash

Eine *MiSeq v2* Reagenzienkartusche für 500 Sequenzier-Zyklen wurde im Anschluss entweder 1 h in einem Wasserbad oder über Nacht bei 4 °C aufgetaut und zur Durchmischung der Sequenzierreagenzien zehnmal invertiert. Über den *MiSeq*-Desktop wurde das Gerät für die Sequenzierung vorbereitet. Zunächst erfolgte eine Reinigung der Fließzelle mit hochreinem Wasser entsprechend der Herstellerangaben zur Entfernung von Salzen der Pufferlösung, in der die Fließzelle gelagert wurde. Anschließend wurden 600 µL *Library/ PhiX*-Gemisch in das

Proben-Reservoir der Sequenzierkartusche geladen. Das Sequenziergerät wurde mit der *Flow Cell*, der PR2-Pufferflasche und der Sequenzierkartusche bestückt und ein *Pre-Run* Check initiiert. Anschließend wurde der Sequenziervorgang gestartet. Nach Beendigung der Sequenzierung wurden die Rohdaten (für jede Probe wurde ein FASTQ-File generiert) auf eine lokale Festplatte kopiert und deren Inhalte für die spätere bioinformatische Analyse auf einem Linux-basierten Server gespeichert. Im Anschluss fand der *Post-Run-Wash* mit einer Lösung von *Tween 20* entsprechend der Herstellerangaben statt.

Tabelle 11: Detailliertes Pipettierschema für die Verdünnung der Library vor der Sequenzierung. Um technische Artefakte innerhalb der Sequenzierung besser beurteilen zu können, wurde die *PhiX-Library* (Illumina) als interne Kontrolle mitgeführt.

I	**Verdünnung des Library-Pools auf 20 pM** Ansatz in einem 1,5 mL LoBind-Tube: 1. 5 µL Library-Pool (4 nM) + 5 µL NaOH (0,2 M) 2. Kurz vortexen 3. 1 min bei 2000 rpm zentrifugieren 4. 5 min bei RT inkubieren 5. 990 µL HT1 hinzufügen 6. Reaktionsgefäß in Kühlblock stellen	**Verdünnung der PhiX-Kontrolle auf 20 pM** Ansatz in einem 1,5 mL LoBind-Tube: 1. 2 µL PhiX + 3 µL Elutionspuffer + 5 µL NaOH (0,2 M) 2. Kurz vortexen 3. 1 min bei 2000 rpm zentrifugieren 4. 5 min bei RT inkubieren 5. 990 µL HT1 hinzufügen 6. Reaktionsgefäß in Kühlblock stellen
II	**Verdünnung des Library-Pools auf 10 pM** Ansatz in einem 1,5 mL LoBind-Tube: 1. 300 µL Library (20 pM) + 300 µL HT1 2. Vortexen und anzentrifugieren 3. Reaktionsgefäß in einen Kühlblock stellen	**Verdünnung der PhiX-Kontrolle auf 10 pM** Ansatz in einem 1,5 mL LoBind-Tube: 1. 300 µL PhiX (20 pM) + 300 µL HT1 2. Vortexen und anzentrifugieren 3. Reaktionsgefäß in einen Kühlblock stellen
III	**Kombination von Library und PhiX (20%) und Denaturierung** Ansatz in einem 1,5 mL LoBind-Tube: 1. 480 µL Library (10 pM) + 120 µL PhiX (10 pM) kombinieren 2. Reaktionsgefäß für 2 min bei 96° C inkubieren 3. Reaktionsgefäß 2x invertieren und anzentrifugieren 4. Reaktionsgefäß in einen Kühlblock stellen	
IV	**Beladung der MiSeq v2 Reagenzienkartusche** 1. 600 µL kombinierter Library/PhiX-Pool werden in das entsprechende Well („*Load Sample*“) der v2 Reagenzienkartusche pipettiert	

2.2.5.7 Bioinformatische Auswertung der Sequenzdaten

1. *Real Time Analysis Viewer*

Die Analyse der Rohdaten (*base calling, quality scoring* und *Demultiplexing*) wurde mit der auf dem *MiSeq* integrierten Software *Sequencing Analysis Viewer* und dem *MiSeq Reporter* durchgeführt. Für eine erste Beurteilung der Qualität wurden die Parameter *Cluster Density, Q-Score, Cluster Passing Filter* und die Fehlerrate (Begriffserklärungen s. Glossar) analysiert. Dabei kann die nachfolgende, direkte Abhängigkeit der Qualitätsparameter festgestellt werden: Je dichter/ höher die *Cluster Density*, desto mehr *Reads* wurden generiert und entsprechend höher fällt der *Cluster Passing Filter*-Wert aus. Kommt es zu keiner eindeutigen Zuordnung des Fluoreszenzsignals zu den Basen, lässt dies den Rückschluss zu, dass die *Cluster Densitiy* zu hoch ist, d.h. der MPS-Lauf überclustert wurde. Dies wird anhand einer hohen Fehlerrate sowie einem geringen Q30-Wert festgestellt. Die optimalen Kriterien der einzelnen Qualitätsparameter sind dabei abhängig von der jeweils verwendeten Sequenzierchemie sowie der Länge des sequenzierten *Reads* (Tabelle 12).

Tabelle 12: Übersicht der für die Validierung verwendeten Sequenzierreagenzien inkl. der vom Hersteller angegebenen Qualitätswerte. Bezogen am 15.01.2022 von Illumina, https://emea.illumina.com/systems/sequencing-platforms/miseq/specifications.html.

Sequenzierchemie	Zyklen	Cluster Density	Total Reads	Output	Q30-Wert
MiSeq Reagent Nano Kit v2	500	800 k/mm^2	1 Mio.	0,5 Gb	>75 %
MiSeq Reagent Kit v2	500	800 k/mm^2	15 Mio.	7,5 Gb	>75 %

2. *Quality Filtering* und *Subsampling*

Das bioinformatische *Quality Filtering* und *Supsampling* der resultierenden MPS-Sequenzdaten wurde durch Herrn Dr. rer. nat. Sebastian Ganschow (Bioinformatiker der LABCON-OWL GmbH) durchgeführt. Für die Qualitätsüberprüfung der MPS-Rohdaten wurde die Software *FastQC* (Kapitel 2.1.5) eingesetzt. Um die statistische Vergleichbarkeit der einzelnen Replikate untereinander zu gewährleisten, wurden im Anschluss die Rohdaten der Sensitivitäts- und Reprodizierbarkeitsstudie (Kapitel 2.2.7.2 und 2.2.7.3) einem bioinformatischem *Supsampling* unterzogen, welches durch das Tool *reformat* aus dem *BBtools* Softwarepaket (Kapitel 2.1.5) geleistet wurde. *reformat.sh* gibt dabei nach Eingabe des folgenden Befehls eine vorgegebene Anzahl zufälliger Reads innerhalb einer Probe aus:

"reformat.sh in=$IN.FQ out=$OUT.FQ samplereadstarget=n"

wobei "*$IN.FQ*" die Ausgangs-Fastq-Datei, "*$OUT.FQ*" die Ausgabedatei und "*n*" die vordefinierte Menge an *Reads* darstellt, die entsprechend der minimal ermittelten *Coverage* über alle eingeschlossenen Proben erzielt werden konnte.

3. *toaSTR*

Die Interpretation der STR-Genotypen auf der Grundlage der Fastq-Rohdaten erfolgte mittels der Software *toaSTR* (Ganschow et al., 2018). Der *Analytical-Threshold* wurde auf zehn *Reads* und der *Calling-Threshold* auf 2 % eingestellt. Ein *Calling-Threshold* von 2 % bedeutet, dass ein Allel eine *Coverage* von mindestens 2 % der Gesamt-*Coverage* des jeweiligen STR-Lokus aufweisen muss, um als echtes biologisches Allel klassifiziert zu werden. Andernfalls würde die Beobachtung als Stutter oder Artefakt eingestuft werden, je nach erwartetem Stutter-Wert, der durch das Stutter-Modell berechnet wird. Sequenzbeobachtungen mit einer *Coverage* unterhalb des *Analytical-Threshold* wurden innerhalb der Analyse nicht berücksichtigt. Vor der eigentlichen Auswertung der Sequenzdaten wurde in *toaSTR* ein entsprechendes *monSTR*-Panel hinterlegt. Dazu wurden die im Panel implementierten STR-Systeme ausgewählt und die gem. Kapitel 3.4.1 ermittelten N-1 *Stutter-Thresholds* definiert. Es wurde dabei ein N-1 Stutter von mindestens 5 % festgelegt. Die Klassifizierung von allelischen (*Signal*) und nichtallelischen Beobachtungen (*Noise*) erfolgte automatisch auf Grundlage des *toaSTR*-Algorithmus (Referenz) zur Stuttermodellierung und wurde für jede Probe zusätzlich manuell überprüft.

2.2.5.8 Statistische Analyse der Sequenzdaten

Die statistische Auswertung der Sequenzdaten (Tabelle 13) erfolgte mittels *Microsoft Excel*. Bei der Analyse mehrerer Replikate wurden im Anschluss ein Mittelwert gebildet und die jeweilige Standardabweichung berechnet.

1. *Allele Coverage Ratio* (ACR)

Kürzere Allele werden zumeist bevorzugt amplifiziert als längere Allele (= präferentielle Amplifikation). Das Verhältnis der *Coverages* heterozygoter Allele ist dabei ein Maß für die

erfolgreiche Detektion und abhängig vom untersuchten STR-System, dem Längenunterschied beider Allele und der eingesetzten DNA-Menge. Im Rahmen dieser Arbeit wurde für die Bewertung heterozygoter Einzelproben die *Allele Coverage Ratio* (ACR) für jedes heterozygote STR-System bestimmt. Hierfür wurde der Quotient aus den *Coverages* des Allels mit schwächerer und desjenigen mit stärkerer Intensität gebildet. Weisen beispielsweise beide heterozygoten Allele die gleiche *Coverage* auf, ergibt sich ein Verhältnis von eins. Ein zunehmendes *Coverage*-Ungleichgewicht über beide heterozygoten Allele führt dementsprechend zu einer kleineren ACR.

2. *Coverage coefficient of variation* (CV)

Für die statistische Bewertung des Amplifikationsgleichgewichts über alle im *monSTR*-Panel berücksichtigten STR-Systeme wurde der Variationskoeffizient der *Coverage* (*Coverage coefficient of variation*, CV) analysiert. Je größer dieser Wert ist, desto größer ist das Amplifikationsungleichgewicht und die Variation innerhalb der Lokus-*Coverages*. Der CV-Wert sollte somit minimal (nahe 0) sein.

3. *Calling-Ratio* (CR)

Des Weiteren sollte die Spezifität der Amplifikation mittels *monSTR* statistisch ermittelt werden. Für die Bewertung der Amplifikationsspezifität wurde die *Calling Ratio* (CR) berechnet. Für diesen Zweck wurde der Quotient aus den in *toaSTR* angegebenen *Called Reads* einer Probe und den *Input Reads* gebildet. CR drückt dabei den Anteil an *On-Target-Reads* (= Anzahl an Sequenzen, die eindeutig einem STR-System zugeordnet wurden) einer Probe aus.

4. *Signal-to-Noise Ratio* (SN)

Während der Amplifikation forensischer STRs können technische Artefakte entstehen, die die eindeutige Interpretation des Genotyps beeinträchtigen. Für die Bewertung dieser technischen Artefakte (z.B. die Tendenz zur Bildung von Stutterprodukten) wurde das *Signal-to-Noise Ratio* (SN) berechnet. Die Berechnung erfolgt dabei durch Division der durchschnittlichen *Coverages* echter biologischer Allele und nicht-allelischer Beobachtungen innerhalb einer Probe. Der SN-Wert sollte so hoch wie möglich sein.

5. *Minor Component Allele Frequency* (MAF)

Biologische Proben, die die genomische DNA von mehr als einem Individuum enthalten, werden in der forensisch-genetischen Analytik als Mischspuren bezeichnet und stellen ggf. eine Herausforderung für die Interpretation der STR-Profile dar. Aus diesem Grund sollte ein forensisches Analysekit die Detektion der Haupt- und Nebenkomponente in Mischspuren ermöglichen. Im Rahmen von Validierungen werden Mischungsstudien durchgeführt, um u.a. die Nachweisgrenze basierend auf der Allel-*Coverage* der Nebenkomponente zu bestimmen. Für diese Arbeit wurden unterschiedliche Mischungsverhältnisse der DNA zweier Personen erzeugt und mittels *monSTR* analysiert. Zur Bewertung des in der DNA-Mischung tatsächlich vorliegenden Verhältnisses wurde die *Minor Component Allele Frequency* (MAF) über alle eingeschlossenen STR-Loci ermittelt. Hierzu wurde der Quotient aus der Allel*coverage* der Nebenkomponente und der *Coverage* aller Allele innerhalb einer Verdünnung gebildet.

Tabelle 13: Definition und Berechnung der statistischen Parameter zur Bewertung der Leistungsfähigkeit des monSTR-Panels. Statistische Parameter wurden durch die Bildung des Mittelwertes über alle Replikate einer Probe berechnet. SD = Standardabweichung; Cov = *Coverage*; N = Anzahl der STR-Systeme.

Parameter	Abkürzung	Ziel	Berechnung
Allele coverage ratio	ACR	Intra-Lokus Balance	$\mathrm{ACR} = \frac{\mathrm{Cov}_{\text{Allel niedrigere Coverage}}}{\mathrm{Cov}_{\text{Allel höhere Coverage}}}$
Coverage coefficient of variation	CV	Inter-Lokus Imbalance	$\mathrm{CV} = \frac{\mathrm{SD}_{\mathrm{Cov}}}{\overline{\mathrm{Cov}}}$
Calling ratio	CR	PCR-Spezifität	$\mathrm{CR} = \frac{\text{Called reads}}{\text{Input reads}}$
Signal to noise ratio	SN	Tendenz Artefakte und Stutter zu produzieren	$\mathrm{SN} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\mathrm{Cov}_{\mathrm{alleles,i}}}{\mathrm{Cov}_{\mathrm{noise,i}}}$
Minor component allele frequency	MAF	Theoretisches Mischungsverhältnis der Nebenkomponente	$\mathrm{MAF} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\mathrm{Cov}_{\text{minor alleles,i}}}{\mathrm{Cov}_{\text{total alleles,i}}}$

2.2.5.9 Visualisierung der Daten

Für die Visualisierung der Daten wurde das Statistikprogramm *R* verwendet. Die Abbildungen 6 bis 10 (s. Kapitel 3) wurden durch Herrn Dr. rer. nat. Sebastian Ganschow erstellt und der Autorin für die Interpretation zur Verfügung gestellt.

2.2.6 Kapillarelektrophorese

Um die mittels *monSTR* generierten MPS-Daten zu verifizieren, wurden die 2800M sowie alle Ringversuchsproben (s. Kapitel 2.2.7.5) zusätzlich mittels Kapillarelektrophorese typisiert und die generierten STR-Profile verglichen.

Durchführung:
Um alle im *monSTR*-Panel berücksichtigten STR-Systeme zu typisieren, erfolgte die Amplifikation unter Verwendung der kommerziell erhältlichen STR-Multiplexkits *Powerplex 21 System, PowerPlex ES SE33 Monoplex System* und *VeriFiler Direct PCR Amplification Kit* nach jeweiligem Herstellerprotokoll auf einem *Mastercycler Pro S*. Die Produktauftrennung und Detektion der Amplifikate wurde auf einem *ABI Prism 310 Genetic Analyzer* durchgeführt. Für die Kapillarelektrophorese wurde 1 µL des PCR-Produkts zu 24 µL *Hi-Di Formamid* und 0,5 µL Größenstandard *WEN ILS 500* gegeben. Die PCR-Streifen wurden verschlossen und für 5 min bei 92 °C zur Denaturierung inkubiert. Nach der Denaturierung wurden die Proben für 5 min auf Eis gekühlt. Die Proben wurden anschließend nacheinander für 3 s bei 1,2 Kilovolt in die Kapillare injiziert. Die verschiedenen Multiplexprodukte wurden nach dem Prinzip der Elektrophorese mit *POP-6* als Trennmatrix entsprechend ihrer Größe durch Anlegen einer elektrischen Spannung aufgetrennt. Die resultierenden Daten wurden in der Software *GeneMapper ID* ausgewertet. Der stochastische Threshold für heterozygote Peaks wurde innerhalb der *GeneMapper* Software auf 100 und der für homozygote Peaks auf 200 relative Fluoreszenzeinheiten (rfu) gesetzt.

2.2.7 Experimenteller Studienaufbau zur Validierung von monSTR

Nach DIN EN ISO/ IEC 17025 akkreditierte forensisch-genetische Labor müssen neu entwickelte Analysemethoden und -materialien vor ihrem routinemäßigen Einsatz einer detaillierten Validierung unterziehen. Das Ziel dieser Validierungsstudie war es daher, die Leistungsfähigkeit von *monSTR* gemäß den forensisch anerkannten Richtlinien der *Scientific Working Group on DNA Analysis Methods* (SWGDAM) zu messen (Scientific Working Group on DNA Analysis Methods, 2016). Es wurden insgesamt sechs Studien durchgeführt, um die Robustheit des Panels umfassend zu bewerten. Mit Ausnahme der GEDNAP-

Ringversuchsproben 56 bis 61 aus den Jahren 2018-2020 wurden alle 76 *Libraries* zusammen mit 16 weiteren DNA-*Libraries* (nicht in dieser Dissertation beschrieben) innerhalb eines Sequenzierlaufs unter Verwendung des *MiSeq Reagent Kit v2* sequenziert. Der finale *Library*-Pool wurde auf eine Konzentration von 10 pM eingestellt und auf dem *MiSeq* mit 1 x 400 Zyklen *Single-End Sequencing* sequenziert. Für die Reproduzierbarkeit wurden insgesamt fünf weitere MPS-Läufe mit jeweils 10 *Libraries* pro Lauf mittels *MiSeq Reagent Nano Kit v2* sequenziert. Die Analyse der Rohdaten wurde wie in Kapitel 2.2.5.7 und 2.2.5.8 beschrieben durchgeführt und die Interpretation der Sequenzdaten erfolgte mittels *toaSTR*.

2.2.7.1 Konkordanz

Ziel dieser Studie war die vergleichende Analyse unterschiedlicher Proben mittels CE und MPS. Für diesen Zweck wurden insgesamt 1.000 pg der Referenzproben SRM 2391c A, B und C (NIST) in Dreifachbestimmung mittels *monSTR* amplifiziert und die generierten MPS-Daten im Anschluss mit den Allel- und Sequenzinformationen des NIST-Analysezertifikats verglichen (National Institute of Standards & Technology, 2018; Steffen et al., 2015). Gleichzeitig diente diese Studie zur Berechnung des prozentualen N-1 Stutters von *monSTR*. Der N-1 Stutter wurde für insgesamt 21 STR-Loci berechnet, indem der Quotient aus der *Coverage* des N-1 Stutters und der *Coverage* des entsprechenden Allels gebildet wurde. Er wurde nur dann berechnet, wenn die Sequenzinformation eindeutig als Stutter zu erkennen war und nicht mit anderen Allelen überlappte. Sowohl homozygote- und heterozygote STR-Systeme wurden in die Analyse einbezogen. Der durchschnittliche N-1 Stutter wurde in *toaSTR* hinterlegt und bildete die Grundlage für alle weiteren Validierungsstudien.

2.2.7.2 Wiederholbarkeit und Reproduzierbarkeit

Für eine günstige Beurteilung der Leistungsfähigkeit eines Analysekits ist es entscheidend, dass sowohl von unterschiedlichen Personen als auch in unabhängigen Analyseläufe vergleichbare Ergebnisse produziert werden. Ziel dieser Studie war es daher, die Reproduzierbarkeit des *monSTR*-Panels zu überprüfen. Die 2800M Kontroll-DNA diente als Referenzprobe, von der insgesamt 1.000 pg eingesetzt wurden. Folgender Studienaufbau wurde gewählt:

1. *Intra-Run*-Experiment durch Person 1

Drei einzelne Replikate der 2800M wurden innerhalb einer *Library-Preparation* gleichzeitig durch Person 1 prozessiert und anschließend innerhalb eines *MiSeq*-Laufs parallel sequenziert.

2. *Inter-Run*-Experiment durch Person 1

Drei einzelne Replikate der 2800M wurden innerhalb einer *Library-Preparation* gleichzeitig durch Person 1 prozessiert und innerhalb dreier unterschiedlicher *MiSeq*-Läufe sequenziert.

3. *Intra-* und *Inter-Run*-Experiment durch Person 2

Intra- und *Inter-Run*-Experimente wurden wie zuvor beschrieben durch Person 2 durchgeführt.

Zur besseren Vergleichbarkeit der statistischen Auswertung bezogen auf die einzelnen Replikate wurde über alle 2800M-Replikate ein bioinformatisches *Subsampling* der ermittelten *Coverages* gemäß Kapitel 2.2.5.7 durchgeführt. Alle *FASTQ*-Files dieser Studie wiesen vor dem Upload in *toaSTR* eine *Coverage* von 70.000 *Reads* auf. Die statistische Auswertung der Sequenzdaten bezog sich auf die Berechnung von ACR, CV, CR und SN (Tabelle 13).

2.2.7.3 Sensitivität

Durch die Sensitivitätsstudie sollte eine Aussage darüber getroffen werden, welche Mindest-Inputmenge an genomischer DNA für die Amplifikation mittels *monSTR* eingesetzt werden muss, um noch valide Ergebnisse zu erzielen. Zur Beurteilung der Nachweisgrenze wurde eine Verdünnungsreihe der 2800M angesetzt und jede Konzentration in Dreifachbestimmung analysiert (1.000 pg, 500 pg, 250 pg, 125 pg, 62,5 pg, 31,3 pg, 15,6 pg und 7,8 pg). Zunächst wurden dabei die 1.000 pg und 500 pg Verdünnungen mittels Elutionspuffer präzise eingestellt und die Konzentrationen mittels *Qubit 2.0 Fluorometer* überprüft. Alle weiteren Konzentrationsstufen (250 pg, 125 pg, 62,5 pg, 31,3 pg, 15,6 pg und 7,8 pg) wurden anschließend durch eine serielle 1:2 Verdünnung mittels Elutionspuffer hergestellt. Bei Amplifikation der 2800M Kontroll-DNA mittels *monSTR* können insgesamt maximal 41 Allele typisiert werden (DD7101). Da jede Konzentrationsstufe in Dreifachbestimmung analysiert wurde, konnten somit insgesamt 123 Allele je Verdünnung in die statistische Berechnung einbezogen werden. Zur besseren Vergleichbarkeit der statistischen Auswertung bezogen auf

die einzelnen Replikate, wurde über alle 2800M-Replikate ein bioinformatisches *Subsampling* der ermittelten *Coverages* durchgeführt (s. Kapitel 2.2.5.7). Alle *FASTQ*-Files wiesen vor dem Upload *toaSTR* eine *Coverage* von 42,000 *Reads* auf. Die Auswertung wurde für CV, CR, SN und ACR durchgeführt (Tabelle 13).

2.2.7.4 DNA-Mischungen

Forensische Proben, zu denen mehr als eine Person beigetragen hat (sog. „Mischspuren"), sind in der forensischen Routinearbeit ein häufig anzutreffendes Phänomen. Im Rahmen von Validierungen können Mischungsstudien bei der Interpretation der DNA-Profile von Mischspuren helfen und Aufschluss über das zugrundeliegende Mischungsverhältnis einschließlich der Genotypen von Haupt- und Nebenkomponente geben. Um die Leistungsfähigkeit von *monSTR* bei einer zugrundeliegenden 2-Personenmischung zu testen, wurden anhand zweier Referenzproben bekannter Genotypen die folgenden Mischungsverhältnisse in Dreifachbestimmung analysiert: 1:1, 1:4, 1:9, 1:19 und 1:49. Als Hauptkomponente diente dabei SRM 2391c A und als Nebenkomponente SRM 2391c C. Da in die mPCR mindestens 50 pg genomische DNA der Nebenkomponente eingesetzt werden sollten, variierte der DNA-Input innerhalb dieser Studie von 1.000 pg bis 2.450 pg. Um die Nachweisgrenze der Nebenkomponente bei Amplifikation mittels *monSTR* zu bestimmen, wurden die generierten Allel- und Sequenzinformationen mit dem NIST-Analysezertifikat (National Institute of Standards & Technology, 2018; Steffen et al., 2015) verglichen. In der statistischen Auswertung wurden dabei nur diejenigen Allele der Nebenkomponente berücksichtigt, die weder mit den Allelen der Hauptkomponente noch mit deren Stutterprodukten an den Positionen N-2, N-1 und N+1 überlappten (Tabelle 14). Insgesamt konnten je Replikat 18 eindeutige Allele der Nebenkomponente (über alle drei Replikate je Mischungsverhältnis somit insgesamt 54 Allele) in der statistischen Auswertung berücksichtigt werden. Um die Abweichung zwischen theoretischem und tatsächlich beobachtetem Mischungsverhältnis zu quantifizieren, wurde über alle eindeutigen Allele die *Minor Component Allele Frequency* (MAF, Tabelle 13) berechnet.

Tabelle 14: Autosomales STR-Profil der SRM 2391c A und C gemäß Analysezertifikat des Herstellers. Graue Kästen zeigen insgesamt 18 eindeutige Allele der Nebenkomponente, die weder mit Allelen der Hauptkomponente noch mit deren Stutterprodukten in N-2, N-1 und N+1 Positionen überlappen. Da die Analyse jedes Mischungsverhältnisses in Dreifachbestimmung erfolgte, konnten insgesamt 54 Allele in die Auswertung einbezogen werden (18 Allele * 3 Replikate = 54 Allele).

	SRM 2391c A - Hauptkomponente				**SRM 2391c C - Nebenkomponente**			
	Allel 1		Allel 2		Allel 1		Allel 2	
	Länge	*Sequenz*	*Länge*	*Sequenz*	*Länge*	*Sequenz*	*Länge*	*Sequenz*
D1S1656	17.3	[TAGA]4 TGA [TAGA]12 TAGG TGTG TGTG TG			11	[TAGA]11 TGTG TGTG TG	15	[TAGA]14 TAGG TGTG TGTG TG
D2S1338	18	[TGCC]6 [TTCC]12	23	[TGCC]7 [TTCC]13 GTCC [TTCC]2	19	[TGCC]7 [TTCC]12		
D2S441	10	[TCTA]10	10	[TCTA]8 TCTG [TCTA]1	10	[TCTA]8 TCTG [TCTA]1		
D3S1358	15	TCTA [TCTG]2 [TCTA]12	16	TCTA [TCTG]3 [TCTA]12	16	TCTA [TCTG]3 [TCTA]12	18	TCTA [TCTG]3 [TCTA]14
FGA	21	[TTTC]3 TTTT TTCT [CTTT]13 CTCC [TTCC]2	23	[TTTC]3 TTTT TTCT [CTTT]15 CTCC [TTCC]2	24	[TTTC]3 TTTT TTCT [CTTT]16 CTCC [TTCC]2	26	[TTTC]3 TTTT TTCT [CTTT]18 CTCC [TTCC]2
CSF1PO	10	[AGAT]10			10	[AGAT]10	12	[AGAT]12
SE33	16	[AAAG]2 AG [AAAG]3 AG [AAAG]16 G [AAAG]3 AG	18	[AAAG]2 AG [AAAG]3 AG [AAAG]18 G [AAAG]3 AG	28,2	[AAAG]2 AG [AAAG]3 AG [AAAG]10 AAAAAG [AAAG]17 GAAGG [AAAG]2 AG	31,2	[AAAG]2 AG [AAAG]3 AG [AAAG]9 AAAAAG [AAAG]21 GAAGG [AAAG]2 AG
D7S820	11	[GATA]11			10	[GATA]10	12	[GATA]12
D8S1179	13	[TCTA]13	14	[TCTA]2 TCTG [TCTA]11	10	[TCTA]10	17	[TCTA]2 TCTG [TCTA]14
D10S1248	15	[GGAA]15	16	[GGAA]16	12	[GGAA]12	16	[GGAA]16
TH01	8	[AATG]8	9.3	[AATG]6 ATG [AATG]3	6	[AATG]6	8	[AATG]8
D12S391	18.3	AGAT GAT [AGAT]9 [AGAC]7 AGAT	22	[AGAT]13 [AGAC]8 AGAT	19	[AGAT]13 [AGAC]5 AGAT	23	[AGAT]12 [AGAC]11
VWA	18	TCTA [TCTG]4 [TCTA]13	19	TCTA [TCTG]4 [TCTA]14	16	TCTA [TCTG]4 [TCTA]11	18	TCTA [TCTG]4 [TCTA]13
D13S317	8	[TATC]8			11	[TATC]11		
PENTA E	5	[AAAGA]5	10	[AAAGA]10	12	[AAAGA]12	13	[AAAGA]13
D16S539	10	[GATA]10	11	[GATA]11	10	[GATA]10		
D18S51	12	[AGAA]12	15	[AGAA]15	16	[AGAA]16	19	[AGAA]19
D19S433	13	[AAGG] AAAG [AAGG] TAGG [AAGG]11	14	[AAGG] AAAG [AAGG] TAGG [AAGG]12	13.2	[AAGG] AA [AAGG] TAGG [AAGG]12	15.2	[AAGG] AA [AAGG] TAGG [AAGG]14
D21S11	28	[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]10	32.2	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]12 TA TCTA	29	[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11	30	[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11
PENTA D	9	[AAAGA]9	13	[AAAGA]13	10	[AAAGA]10	11	[AAAGA]11
AMEL	X				X		Y	

2.2.7.5 Forensische Ringversuchsproben

Um das *monSTR*-Panel als effizientes Instrument für die Identifizierung von Individuen einsetzen zu können, muss es verlässliche und präzise Sequenzdaten auch aus realistischen forensischen Spurenproben generieren können. Im Rahmen dieser Studie sollte die Leistungsfähigkeit von *monSTR* bei Einsatz für forensisch Proben untersucht werden. Zu diesem Zweck wurden artifiziell hergestellte Spurenproben aus den internationalen forensischen Ringversuchen der *German DNA Profiling Group* (GEDNAP) im Zeitraum von 2017 bis 2020 zusätzlich mittels *monSTR* analysiert. Die Ringversuchsproben wurden im Rahmen der bestehenden Akkreditierung innerhalb der forensischen Routineanalytik mittels CE erfolgreich typisiert. Der für diese Dissertation verwendete Probensatz umfasste insgesamt 24 Einzelpersonenproben und 32 Spurenproben, die sich aus Einzelproben und DNA-Mischungen von bis zu 3 Personen zusammensetzten. Eine detaillierte Probenbeschreibung einschließlich Angaben über die gemessenen DNA-Konzentrationen sind Tabelle 15 und Tabelle 16 zu entnehmen. Aufgrund des jährlichen Rhythmus wurden die Ringversuche jeweils zusammengefasst bearbeitet und auf dem *MiSeq* sequenziert. Für die Amplifikation mittels *monSTR* wurden, soweit möglich, insgesamt 1.000 pg genomische DNA eingesetzt. Aufgrund der zu niedrigen DNA-Konzentration wurde für die Proben 57-PA, 58-S3, 58-S4, 59-S4, 60-S1, 60-S3 und 61-S1 der maximale Input von 7,5 µL in die mPCR gewählt. Die generierten MPS-Sequenzdaten wurden mit den ermittelten CE-Ergebnissen sowie den offiziellen GEDNAP-Ergebnissen hinsichtlich Konkordanz verglichen. Um potenzielle Abweichungen zwischen MPS und CE zu bewerten, wurden zusätzlich intra-allelische Sequenziervarianten (sog. Isoallele) und Diskordanzen gezählt.

Tabelle 15: Probenübersicht der GEDNAP-Einzelpersonenproben 54 bis 61, inkl. Probentyp und zugrundeliegende DNA-Konzentration. Da es sich hier um eine Auflistung aller analysierten Einzelpersonen handelt, befindet sich in der Probe das genetische STR-Profil einer Person. P = GEDNAP-Personenprobe; A-C = Nummerierung der Personenproben.

Jahr	Proben ID	Probentyp	DNA Konzentration [ng/µL]	Anzahl Personen in der Probe
2017	54-PA	Blut auf Copan Swab	0,313	1
	54-PB	Blut auf Zellstofftupfer	0,351	1
	54-PC	Blut auf Zellstofftupfer	0,329	1
	55-PA	Blut auf Zellstofftupfer	0,663	1
	55-PB	Blut auf Zellstofftupfer	0,615	1
	55-PC	Blut auf Zellstofftupfer	0,367	1
2018	56-PA	Blut auf Zellstofftupfer	0,587	1
	56-PB	Blut auf Zellstofftupfer	0,489	1
	56-PC	Blut auf Zellstofftupfer	0,265	1
	57-PA	Blut auf Zellstofftupfer	0,186	1
	57-PB	Blut auf Baumwollstoff (2005)	0,393	1
	57-PC	Blut auf Zellstofftupfer	0,254	1
2019	58-PA	Blut auf forensischen Abriebtupfer	0,407	1
	58-PB	Blut auf Zellstofftupfer	0,713	1
	58-PC	Blut auf Zellstofftupfer	1,17	1
	59-PA	Blut auf Zellstofftupfer	0,207	1
	59-PB	Blut auf Zellstofftupfer	0,191	1
	59-PC	Blut auf Zellstofftupfer	1,33	1
2020	60-PA	Blut auf Zellstofftupfer (2016)	0,76	1
	60-PB	Blut auf Zellstofftupfer (2016)	0,389	1
	60-PC	Blut auf Zellstofftupfer (2018)	0,137	1
	61-PA	Blut auf Zellstofftupfer (2016)	0,454	1
	61-PB	Blut auf Zellstofftupfer (2018)	0,386	1
	61-PC	Blut auf Zellstofftupfer (2009)	1,09	1

Tabelle 16: Probenübersicht der GEDNAP-Spurenproben 54 bis 61, inkl. Probentyp und zugrundeliegende DNA-Konzentration. Da es sich hier um eine Auflistung aller analysierten Spurenpersonen handelt, kann das genetische STR-Profil von bis zu drei Personen detektiert werden. Die rechte Spalte gibt an, aus wie vielen Personen die Spurenprobe besteht (1, 2 oder 3 Personen). In runden Klammern ist das Jahr angegeben, in dem die Probe angelegt wurde. S = GEDNAP-Spurenprobe; 1-4 = Nummerierung der Spurenproben.

Jahr	Proben ID	Probentyp	DNA Konzentration [ng/µL]	Anzahl Personen in der Probe
2017	54-S1	Blut auf Zellstofftupfer	0,535	3
	54-S2	Blut auf Baumwollstoff (2004)	0,185	1
	54-S3	Sperma auf Toilettenpapier	0,202	1
	54-S4	Blut auf Zellstofftupfer	0,319	2
	55-S1	Kondom mit Sperma	0,313	1
	55-S2	Blut auf Zellstofftupfer	0,298	2
	55-S3	Blut auf Zellstofftupfer	0,357	3
	55-S4	Blut auf Baumwollstoff (2005)	0,635	1
2018	56-S1	Blut auf Natrontablette	0,36	1
	56-S2	Blut auf Mikrofasertuch	0,286	1
	56-S3	Sperma auf Zellstofftupfer	0,07	1
	56-S4	Blut auf Zellstofftupfer	0,417	2
	57-S1	Blut auf Strohpellet	1,27	1
	57-S2	Blut-/Spermamischung auf Zellstofftupfer	0,223	1
	57-S3	Blut auf Zellstofftupfer	0,547	3
	57-S4	Blut auf Zigarettenpapier	0,333	2
2019	58-S1	Blut auf Tampon	0,189	1
	58-S2	Blut auf Zellstofftupfer	0,535	2
	58-S3	Blut auf Zellstofftupfer	0,481	3
	58-S4	Kondom mit Sperma	0,07	1
	59-S1	Blut auf Zellstofftupfer	0,497	2
	59-S2	Blut auf Zigarettenpapier	0,51	1
	59-S3	Blut auf Zellstofftupfer	0,378	3
	59-S4	Sperma auf Toilettenpapier	0,052	1
2020	60-S1	Blut auf Sand (2008)	0,073	1
	60-S2	Blutmischung auf Zellette (2016)	0,334	2
	60-S3	Sperma auf Wattestäbchen (2019)	<0,5	1
	60-S4	Blutmischung auf Zellette (2016)	0,404	3
	61-S1	Sperma auf Toilettenpapier (2020)	0,059	1
	61-S2	Blut auf Muschelnudeln (2018)	0,554	1
	61-S3	Blutmischung auf Zellette (2016)	0,577	3
	61-S4	Blutmischung auf Zellette	0,571	2

2.2.7.6 Speziesspezifität

In dieser Studie sollte das Multiplexpanel *monSTR* auch dahingehend überprüft werden, ob andere biologische DNA-Quellen seine Leistungsfähigkeit beeinträchtigen. Es wurden Maulschleimhautabriebe dreier Haus- und Nutztierarten (Hund, Katze und Pferd) mit sterilen Watteträgern (Sarstedt) angefertigt und mittels *monSTR* analysiert. Die Amplifikation erfolgte in Dreifachbestimmung mit einem DNA-Input von 10 ng je Probe. Zusätzlich wurden 1.000 pg genomische DNA der jeweiligen Halter, 1.000 pg 2800M (Positivkontrolle) und eine Negativkontrolle (Elutionspuffer) analysiert und mit den genetischen Sequenzdaten der Tierproben verglichen. In Bezug auf die statistische Analyse wurde zunächst über die einzelnen Replikate der Mittelwert gebildet und im Anschluss gemäß Tabelle 13 der Parameter CR berechnet.

3 Ergebnisse

Das Ziel der in der vorliegenden Dissertation beschriebenen Arbeit war die Evaluation des forensischen STR-Multiplexpanels *monSTR* auf Grundlage der geltenden forensischen Richtlinien der *Scientific Working Group on DNA Analysis Methods* (SWGDAM) (Scientific Working Group on DNA Analysis Methods, 2016; Silvery et al., 2020). Für diesen Zweck wurden insgesamt sechs Validierungsläufe (Konkordanz, Reproduzierbarkeit und Wiederholbarkeit, Sensitivität, die Analyse unterschiedlicher Mischungsverhältnisse, die Untersuchung forensischer Ringversuchsproben und Speziesspezifität) durchgeführt, um die Leistungsfähigkeit des Multiplexpanels umfassend zu bewerten.

3.1 Funktionalität der monSTR-Primer

Das *monSTR*-Panel amplifiziert insgesamt 21 forensische STR-Systeme, einschließlich des hochinformativen Systems SE33 sowie des geschlechtsspezifischen Markers Amelogenin (Ganschow et al., 2019) und ist kompatibel mit der *MiSeq*-Plattform der Firma Illumina. Folgende STR-Systeme sind im *monSTR*-Panel enthalten: D1S1656, D2S1338, D2S441, D3S1358, FGA, CSF1PO, SE33, D7S820, D8S1179, D10S1248, TH01, D12S391, vWA, D13S317, Penta E, D16S539, D18S51, D19S433, D21S11, Penta D und Amelogenin. Unter Zuhilfenahme zweier bioinformatischer Tools (Gadberry et al., 2005; Shen et al., 2010) wurde im Rahmen einer Vorstudie ein spezifisches Primerset entwickelt, welches speziell definierte Qualitätsanforderungen für eine Multiplex-PCR erfüllen sollte (Ganschow et al., 2019). Die targetspezifische Sequenz der Primer wiesen eine Länge zwischen 19-27 bp (Mittelwert 22 bp) und einen durchschnittlichen GC-Gehalt von 44 % (30-55 %) mit einer durchschnittlichen Schmelztemperatur von 64 °C (61-66 °C) auf. Für die meisten Amplikons wurde eine Größe zwischen 250 und 350 bp gewählt. STR-Systeme wie z.B. FGA, SE33, Penta E und D21S11 können eine größere Anzahl verschiedener Allele aufweisen, wodurch je nach Anzahl der davon umfassten Repeateinheiten größere oder kleinere Amplikons erzeugt werden. Das längste bekannte Allel für Penta E generiert z.B. ein 370 bp langes Produkt. Das Fragment für den geschlechtsspezifischen Marker Amelogenin wurde im Rahmen des *Primerengineerings*

absichtlich kürzer designt, um zusätzlich als Degradationskontrolle für die Amplifikation forensischer Proben zu dienen. Die *in-silico* PCR unter Verwendung der Software *MFEprimer* (Qu et al., 2012) wies auf eine hochspezifische Amplifikation aller 21 *monSTR* Zielregionen hin. Darüber hinaus zeigte die PCR-Simulation gleichzeitig, dass die Tendenz zur Bildung von Primerdimeren oder unspezifischen Produkten über alle Zieltargets gering war. Bevor die Primerpaare im Multiplex-Ansatz kombiniert wurden, wurden diese in Singleplex-Ansätzen amplifiziert und das entstandene PCR-Produkt überprüft. Die anschließende gelelektrophoretische Auftrennung der Amplifikationsprodukte zeigte, dass die jeweils zu erwartenden Produktgrößen erfolgreich nachgewiesen werden konnten.

3.2 Library-Preparation mittels monSTR

Anhand elektrophoretischer Auftrennungen der 2800M, die zu unterschiedlichen Zeitpunkten der *Library-Preparation* generiert wurden, soll im Folgenden die erfolgreiche Optimierung dieses komplexen Prozesses demonstriert werden (Abbildung 5). Eine erste Qualitätskontrolle der MPS-*Library* erfolgte nach der Multiplex-PCR (mPCR). Hier erscheint neben den beiden Kontroll-Markern bei 50 bp und 10.380 bp ein 110 bp langer Primerdimer-Peak mit einer Fluoreszenzintensität (engl.: *Fluorescence Units*, FU) von 80 (Abbildung 5A). Der Zustand des die MPS-*Library* enthaltenden Produkts unmittelbar nach der ersten größenselektiven Aufreinigung mittels *Agencourt AMPure XP* Magnetpartikeln ist in Abbildung 5B dargestellt. Das Ziel dieses Arbeitsschritts, die vollständige Entfernung unspezifischer und/ oder ungebundener Fragmengte, wurde erreicht, da der o.g. Primerdimer-Peak (110 bp) vollständig entfernt wurde. Die spezifische MPS-*Library* ist schwach vom Hintergrund zu unterscheiden. Innerhalb der Index-PCR wird, wie beschrieben, an beide Enden der *Library* ein 35 bp langer Indexadapter angefügt, weshalb sich die targetspezifischen Amplikons insgesamt um ca. 70 bp verlängern. Die elektrophoretische Auftrennung der *Library* nach der Index-PCR ist in Abbildung 5C gezeigt. Neben einem Primerdimer-Peak bei 185 bp (ca. 30 FU) ist die spezifische *Library* im Elektropherogram nun deutlich als Peak zu erkennen. Da die *monSTR-Library* erst unmittelbar nach der Index-PCR erkennbar wird, kann erst nach diesem Schritt regulierend in den Prozess eingegriffen werden, sollten z.B. noch zu viele Primerdimer-Artefakte auftreten oder unspezifische PCR-Produkte amplifiziert worden sein. Da für die meisten Amplikons im Rahmen des Primerdesigns eine einheitliche Größe gewählt wurde, sind

nicht alle 21 Zielregionen als separate Signale zu erkennen. Die *Library* bewegt sich im Größenbereich von ca. 350 bp bis 500 bp. Der höchste Amplikon-Peak weist 30 FU und eine Länge von ca. 400 bp auf. Der Primerdimer-Peak (185 bp) konnte innerhalb der zweiten größenspezifischen Aufreinigung mittels *Agencourt AMPure XP Beads* vollständig entfernt werden, sodass für die nachfolgende Sequenzierung die spezifische und aufgereinigte MPS-*Library* zur Verfügung stand (Abbildung 5D).

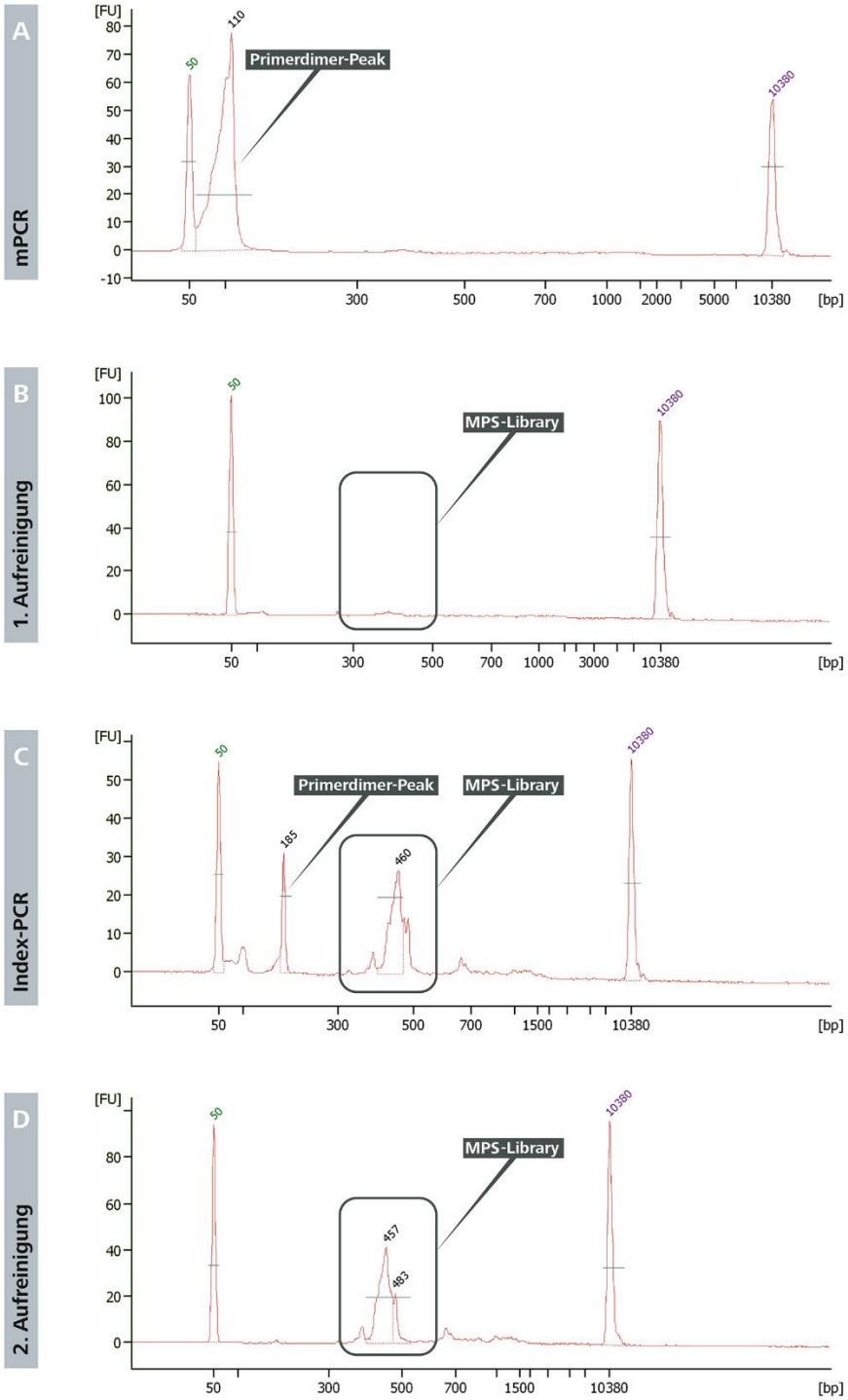

Abbildung 5: Elektrophoretische Auftrennung der 2800M Kontroll-DNA in unterschiedlichen Stadien der Library-Preparation. Auf der Y-Achse ist die Fluoreszenz-Intensität und auf der X-Achse die Fragmentlänge in Basenpaaren aufgetragen. Peaks bei 50 bp und 10380 bp repräsentieren jeweils Kontrollmarker. **A)** *Library* unmittelbar nach der mPCR. **B)** *Library* nach der ersten Aufreinigung mittels *Agencourt AMPure XP* Magnetpartikeln. **C)** *Library* nach der Index-PCR und **D)** *Library* nach der zweiten beadbasierten Aufreinigung. FU = *Fluorescence Units*; bp = Basenpaaren.

3.3 Sequenzierqualität der durchgeführten MPS-Läufe

Die Software *MiSeq Sequencing Analysis Viewer* erstellt nach Fertigstellung jedes Sequenzierlaufs einen Analysereport über einzelne Sequenzierparameter, der eine Bewertung des Laufes schon vor der eigentlichen bioinformatischen Auswertung der Sequenzdaten erlaubt (Tabelle 17). Es wurden insgesamt sechs MPS-Sequenzierungen unter Einsatz des *MiSeq Reagent Kits v2* (1 x 400 bp, Lauf 1) und des *MiSeq Reagent Nano Kits* (1 x 400 bp, Läufe 2-6) durchgeführt. Im ersten Sequenzierlauf wurden insgesamt 76 und in den Läufen zwei bis sechs jeweils 10 MPS-*Libraries* sequenziert. Über alle MPS-Läufe hinweg konnten durchschnittlich insgesamt 6.700 MB an Sequenzdaten und 17,5 Mio. *Reads* generiert werden. Die Sequenzierläufe wiesen eine mittlere *Cluster density* von 753 K/mm2 (± 90 K/mm2) und einen *Cluster Passing Filter* Wert von 93,3 % (± 1,5 %) auf. Für 83 % (± 6 %) aller sequenzierten Basen wurde ein *Q-Score* von 30 erzielt. In allen *MiSeq*-Läufen wurde die *PhiX*-Kontrolle mitgeführt und eine darauf basierende Fehlerrate von 6,44 % (± 3,66 %) ermittelt.

Tabelle 17: Qualitätsparameter der Sequenzierläufe bei Verwendung von monSTR. Dargestellt sind die einzelnen MPS-Läufe, die im Rahmen der Validierungsstudie durchgeführt wurden und deren zugehörige Qualitätsparameter.

	Illumina Chemie	Sequenzierzyklen	Cluster Density	Clusters Passing Filter	Yield	Reads Passing Filter	Basen ≥ Q30	Fehlerrate
			K/mm²	*%*	*MB*	*M*	*%*	*%*
Lauf 1	v2	1x400	716	91,9	5000	12,6	71,3	8,88
Lauf 2	v2 Nano	1x400	908	92,0	300	1,1	86,5	1,71
Lauf 3	v2 Nano	1x400	803	92,6	300	1,0	88,2	1,70
Lauf 4	v2 Nano	1x400	662	94,8	300	0,9	83,1	8,81
Lauf 5	v2 Nano	1x400	686	95,4	400	0,9	86,5	8,71
Lauf 6	v2 Nano	1x400	744	93,1	400	1,0	82,6	8,80

3.4 Ergebnisse der Validierung des monSTR-Panels

Die im Folgenden aufgeführten Ergebnisse erlauben die Bewertung der Robustheit und Validität von *monSTR* gemäß den Referenz-Richtlinien der SWGDAM. Für diesen Zweck sind insgesamt sechs definierte Evaluationen geplant und durchgeführt worden.

3.4.1 Ermittlung des Stutter-Thresholds

Für diesen Teil der Studie wurde der prozentuale N-1 Stutter für jedes STR-System des *monSTR*-Panels bestimmt, indem der Quotient der *Coverage* des Stutters und der *Coverage* des entsprechenden biologischen Allels gebildet wurde. Als Referenzprobe diente das *Standard Referenz Material* (SRM) 2391c A, B und C des *National Institute of Standard and Technology* (NIST, USA), was jeweils in Dreifachbestimmung analysiert wurde.

Die Mehrheit der STR-Systeme bildete Stutter-Produkte an N-1 Position mit einer Signalintensität zwischen 5 % und 20 % (Tabelle 18) des allelischen Peaks. Die höchste Stutteraktivität wurde im STR-System D12S391 mit einem Maximum von 26,9 % gemessen. Im Gegensatz dazu produzierten die beiden Pentamer-Repeat-Systeme Penta E und Penta D vergleichsweise niedrige Stutterwerte (durchschnittlicher Stutter-Wert für Penta E/ Penta D: 3 %/ 2 % der Intensität des allelischen Peaks). In *toaSTR* wurde der lokusspezifische Stutter-*Threshold* des jeweiligen STR-Systems dabei jeweils oberhalb des ermittelten Maximums gesetzt (mindestens jedoch bei 5 %). Der ermittelte *Stutter-Threshold* für das *monSTR*-Panel wurde anschließend in die *toaSTR* Software übernommen und bildete somit die Grundlage für die Interpretation aller folgenden Validierungsstudien. Eine Ausnahme stellte lediglich der in *toaSTR* eingestellte Stutter-*Threshold* für das System SE33 dar. Da zwei von drei Replikaten einen Stutterwert unterhalb von 10 % aufwiesen, wurde der Stutter-*Threshold* dieses STR-Systems auf 12 % gesetzt.

Tabelle 18: Analysebereich des N-1-Stutters aller STR-Systeme des monSTR-Panels. Als Referenzproben dienten die NIST SRM 2391c A, B und C in Dreifachbestimmung. Der daraus resultierende N-1 Stutter-*Threshold* wurde innerhalb der Software *toaSTR* (Ganschow et al., 2018) eingestellt und für alle folgenden Studien verwendet. Der *Threshold* wurde berechnet, indem die *Coverage* des Stutters in N-1-Position durch die *Coverage* des zugehörigen echten Allels dividiert wurde. Min = Minimum; Max = Maximum.

STR-Marker	% N-1 Stutter Mittelwert	% N-1 Stutter Min.	% N-1 Stutter Max.	Eingestellter N-1 Stutter in toaSTR
D1S1656	10	3,1	19,1	20 %
D2S1338	15	10,9	17,7	18 %
D2S441	3	1,5	4,3	5 %
D3S1358	8	4,3	12,1	13 %
FGA	18	11,6	23,3	24 %
CSF1PO	3	1,6	4,6	5 %
SE33	10	6,4	17,3	12 %
D7S820	4	1,6	7,2	8 %
D8S1179	18	12,0	23,2	24 %
D10S1248	10	5,6	14,3	15 %
TH01	4	2,5	7,3	8 %
D12S391	17	4,6	26,9	27 %
vWA	13	9,9	14,9	15 %
D13S317	2	1,0	3,9	5 %
Penta E	3	0,4	6,8	7 %
D16S539	10	7,9	16,7	17 %
D18S51	11	5,8	16,3	17 %
D19S433	7	4,4	8,9	9 %
D21S11	6	2,0	10,1	11 %
Penta D	2	1,6	2,5	5 %

3.4.2 Konkordanz

Die präzise und reproduzierbare Typisierung aller in einem Analysekit implementierten STR-Systeme ist für eine Methode mit befriedigender Leistungsfähigkeit unbedingt zu fordern. Das Ziel dieses Teils der Arbeit war es daher, vollständig konkordante Ergebnisse zwischen den mittels *monSTR* generierten MPS-Daten und den Sollwerten der Referenzproben zu erzielen. Als Kontrollmaterial wurde das SRM 2391c A, B und C (NIST) verwendet, dessen Eigenschaften (Fragmentlängen und -sequenzdaten) vom Hersteller vollständig publiziert wurden (National Institute of Standards & Technology, 2018; Steffen et al., 2015).

Da das *monSTR*-Panel insgesamt 21 STR-Systeme umfasst und die SRM-Referenzproben jeweils in Triplikaten untersucht wurden, konnten für diese Evaluation insgesamt 189 STR-Systeme und 342 Allele in die Auswertung einbezogen werden. Die Ergebnisse zeigen, dass

sowohl hinsichtlich der gemessenen Fragmentlängen als auch der Sequenzen über alle 21 *monSTR*-Targets eine 100 %ige Konkordanz zu den vorgegebenen Sollwerten erzielt wurde. In keinem STR-System traten Allel-Dropouts/ -Dropins auf. Die durchschnittlichen *Coverages* der untersuchten Referenzproben inklusive der erzielten durchschnittlichen *Coverages* pro STR-System sind in Tabelle 19 aufgeführt.

Tabelle 19: Durchschnittliche Coverages inkl. Standardabweichungen der Konkordanzstudie.

NIST-Referenzprobe	**Coverage (x) pro Probe** Mittelwert	**Coverage (x) pro STR-System** Mittelwert
SRM 2391c A	61.354x ± 36.948x	2.922x ± 1.795x
SRM 2391c B	55.780x ± 14.611x	2.656x ± 740x
SRM 2391c C	60.543x ± 9.851x	2.883x ± 514x

3.4.3 Wiederholbarkeit und Reproduzierbarkeit

Während der massiven parallelen Sequenzierung zahlreicher MPS-*Libraries* innerhalb eines Sequenzierlaufes können molekulare Wechselwirkungen und Sekundärstrukturausbildungen die Amplifikation einzelner DNA-Fragmente beeinträchtigen. Bedingt durch die Bearbeitung verschiedener *Libraries* durch unterschiedliche Labormitarbeiter können interindividuelle Unterschiede zwischen *Library-Preparations* auftreten, die sich auch in den Sequenzdaten widerspiegeln. Eine Kalibrierung der Analysemethode ist daher im Rahmen einer Validierung unerlässlich, um eine vollständige Vergleichbarkeit der Ergebnisse untereinander zu ermöglichen und die Robustheit der Methode zu bestätigen. Ziel dieses Teils der Arbeit war es daher, Unterschiede innerhalb von *Library-Preparations* mittels *monSTR* zu quantifizieren, nachdem zwei unterschiedliche Personen MPS-*Libraries* hergestellt hatten. Zu diesem Zweck wurden *Intra-* und *Inter-Run*-Experimente mit Hilfe der Referenz 2800M durchgeführt.
Innerhalb desselben *MiSeq*-Laufs durch Person 1 (*Intra-Run* Experiment) betrug die durchschnittliche *Coverage* pro Probe 60.442x (± 1.431x) mit einer durchschnittlichen *Coverage* pro STR-System von 2.878x (± 290x). Durch Person 2 konnte im *Intra-Run* Experiment eine durchschnittliche *Coverage* pro Probe von 54.984x (± 598x) und eine durchschnittliche *Coverage* pro STR-System von 2.618x (± 259x) erzielt werden. Wurden drei separate *MiSeq*-Läufe innerhalb des *Inter-Run* Experiments durch Person 1 prozessiert, betrug die durchschnittliche *Coverage* pro Probe 58.799x (± 3.024x) und die durchschnittliche

Coverage pro STR-System lag bei 2.800x (± 589x). Führte Person 2 das *Inter-Run* Experiment durch, konnte eine durchschnittliche *Coverage* pro Probe von 52.229x (± 4.677x) und pro STR-System von 2.487x (± 574x) erzielt werden. Der Vergleich zwischen den Sollwerten der Referenzprobe und den generierten STR-Sequenzdaten ergab eine vollständige Übereinstimmung der Allellängen. Gleichzeitig wurden die statistischen Parameter *Coverage Coefficient of Variation* (CV), *Calling Ratio* (CR), *Signal-to-Noise Ratio* (SN) und *Allele Coverage Ratio* (ACR) berechnet (s. Kapitel 2.2.5.8, Tabelle 13) und die Ergebnisse in Tabelle 20 dargestellt. Alle statistischen Parameter zeigen zwischen Person 1 und 2 vergleichbare Werte mit jeweils geringen Standardabweichungen. Zwischen unterschiedlichen Sequenzierläufen (*Inter-Run* Experiment) war die Standardabweichung über alle statistischen Parameter höher als innerhalb des *Intra-Run* Experiments. Wurde die 2800M durch zwei unterschiedliche Personen bearbeitet, waren die Ergebnisse der statistischen Parameter vollständig reproduzierbar.

Tabelle 20: Ergebnisse der statistischen Parameter inkl. Standardabweichungen für das Intra- und Inter-Run Experiment. CV = *Coverage Coefficient of Variation*; CR = *Calling Ratio*; SN = *Signal-to-Noise Ratio*; ACR = *Allele Coverage Ratio*; Min = minimaler beobachteter Wert; Max = maximaler beobachteter Wert.

Statistische Parameter	Person 1	Person 2	Person 1	Person 2
	Intra-Run Experiment		Inter-Run Experiment	
CV	0,54 ± 0,04	0,51 ± 0,03	0,51 ± 0,12	0,54 ± 0,10
CR	0,86 ± 0,02	0,85 ± 0,01	0,84 ± 0,04	0,82 ± 0,03
SN	18,1 ± 0,8	16,2 ± 1,2	17,2 ± 0,9	17,4 ± 0,7
ACR Min.	0,78 ± 0,09	0,73 ± 0,01	0,72 ± 0,10	0,71 ± 0,03
ACR Max.	0,99 ± 0,04	0,99 ± 0,03	0,99 ± 0,03	0,99 ± 0,04

3.4.4 Sensitivität

In diesem Teil der Validierungsstudie sollte die minimale Menge einzusetzender DNA für *monSTR* definiert werden, bei der noch zuverlässige STR-Profile generiert werden. Es wurde eine 8-stufige Verdünnungsreihe der Referenzprobe 2800M von 1.000 pg bis 7,8 pg angesetzt und jede Konzentrationsstufe als Triplikat analysiert.

Das *monSTR*-Panel erzielte innerhalb dieser Evaluation vollständige und konkordante STR-Profile bis zu einer DNA-Inputmenge von 62,5 pg (Tabelle 21). Bei einer DNA-Inputmenge von 31,3 pg trat ein Allel-Dropout im System D12S391 auf, so dass in dieser Verdünnungsstufe

insgesamt 99,2 % aller Allele erfolgreich typisiert wurden. Ab einer DNA-Inputmenge von 15,6 pg traten über alle Replikate hinweg insgesamt 16 Allel-Dropouts und ein vollständiger Lokus-Dropout des heterozygoten Systems D12S391 auf. Mehrere vollständige Lokus-Dropouts wurden in den STR-Systemen Penta D und D7S820 bei einer DNA-Inputmenge von 7,8 pg gemessen. Dennoch konnten auch innerhalb der niedrigsten Verdünnungsstufe insgesamt noch 76,4 % aller Allele korrekt typisiert werden (Teilprofil).

Tabelle 21: Erfolgreich typisierte Allele der 2800M innerhalb acht verschiedener Verdünnungsstufen. Die Referenzprobe umfasst in Bezug auf das monSTR-Panel insgesamt 41 Allele. Da jede Konzentration in Dreifachbestimmung analysiert wurde, konnten insgesamt 123 Allele und 63 STR-Loci pro Verdünnungsstufe in der Bewertung berücksichtigt werden. N = Anzahl der Allele; pg = Pikogramm.

DNA-Input [pg]	Detektierte Allele [N=123]	% Detektierte Allele
1.000	123	100
500	123	100
250	123	100
125	123	100
62,5	123	100
31,3	122	99,2
15,6	105	85,4
7,8	94	76,4

Die Analyse der relativen Lokus-*Coverage* über alle *monSTR*-Targets und Verdünnungsstufen sollte Aufschluss über eine homogene Amplifikation geben. Die relative Lokus-*Coverage* wurde berechnet, indem der Quotient aus der durchschnittlichen Lokus-*Coverage* und der durchschnittlichen *Coverage* aller STR-Loci einer Konzentrationsstufe gebildet wurde. Eine ausbalancierte Lokus-*Coverage* mit zugehörigen niedrigen Standardabweichungen wurde bis zu einem DNA-Input von 62,5 pg beobachtet, was eine homogene Amplifikation über alle *monSTR*-Targets bis zu dieser Verdünnungsstufe demonstriert (Abbildung 6). Die Abbildung zeigt deutlich, dass die relativen Lokus-*Coverages* ab einem DNA-Input unter 62,5 pg variierten und die entsprechenden Standardabweichungen ansteigen. In absoluten Zahlen dargestellt, war der STR-Marker mit der geringsten Qualität über alle Verdünnungsstufen das Pentamer-System Penta D, mit einer durchschnittlichen *Coverage* von 427x (± 75x). Über alle Verdünnungsstufen zeigte der STR-Marker D18S51 mit einer durchschnittliche *Coverage* von 2.416x (± 446x) die beste Qualität, was einer 5,6-fachen *Coverage*-Differenz zwischen diesen beiden STR-Systemen entspricht. Die durchschnittliche *Coverage* je STR-Marker, über alle

acht Konzentrationsstufen, lag bei 1.054x (± 592x) und die durchschnittliche *Coverage* pro Probe bei 66.126x (± 36.250x).

Die Ergebnisse der statistischen Parameter zeigen eine gute Stabilität der Interlokus-Balance (CV) bis zu einem DNA-Input von 62,5 pg (Abbildung 7). Je größer der CV-Wert, desto größer das Ungleichgewicht der Lokus-*Coverages*. Der höchste CV-Wert wurde bei einem DNA-Input von 15,6 pg beobachtet (0,83 ± 0,20). Erwartungsgemäß wurde die Interlokus-Balance bei geringeren DNA-Inputmengen deutlich instabiler. Der Anteil der *On-Target-Reads* (CR) nahm je Verdünnungsstufe kontinuierlich ab und fiel ab einem DNA-Input von 31,3 pg unter 50 %. Der Parameter SN schwankte leicht bis zu einem DNA-Input von 15,6 pg und fiel bei 7,8 pg DNA-Input auf 11,5 ± 2,12.

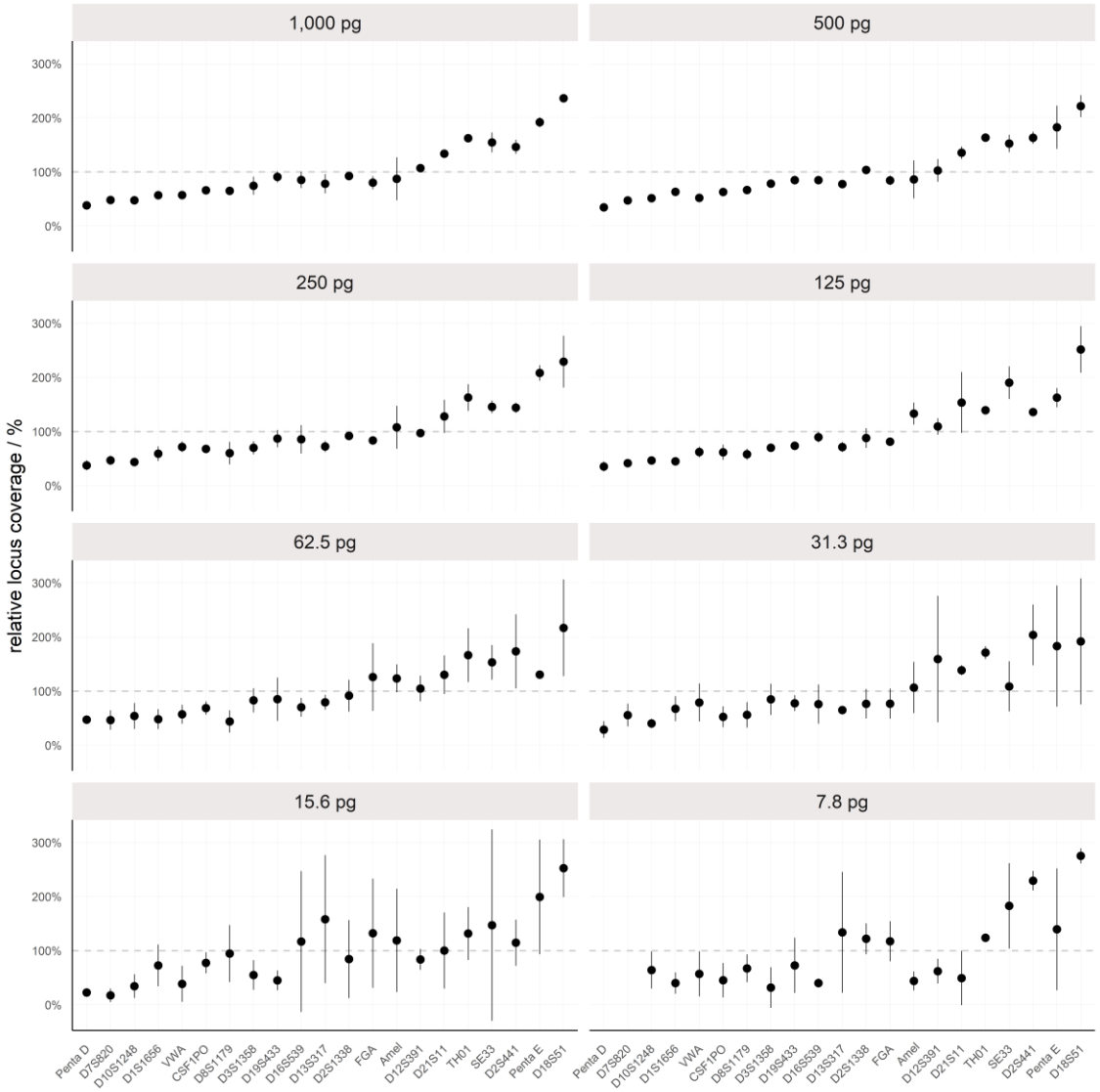

Abbildung 6: Relative Lokus-Coverage in Abhängigkeit verschiedener DNA-Konzentrationen der 2800M. Zur Beurteilung der Nachweisgrenze wurde eine Verdünnungsreihe (1.000 pg – 7,8 pg) der 2800M angesetzt und jede Konzentration in Dreifachbestimmung analysiert. Im Anschluss wurde ein Mittelwert gebildet und die jeweilige Standardabweichung berechnet. Die relative *Coverage* eines STR-Systems wurde berechnet, indem die durchschnittliche Lokus-*Coverage* durch die mittlere *Coverage* aller einbezogenen STR-Loci pro Verdünnung dividiert wurde. Die Fehlerbalken stellen die entsprechenden Standardabweichungen dar. Gestrichelte Linien zeigen die mittlere Lokus-*Coverage* über alle STR-Systeme in der entsprechenden Verdünnungsstufe an. Die STR-Loci sind in aufsteigender Reihenfolge nach der durchschnittlichen Lokus-*Coverage* über alle Verdünnungen aufgelistet. Amel = Amelogenin; pg = Pikogramm. (aus Silvery et al., 2020, Seite 5).

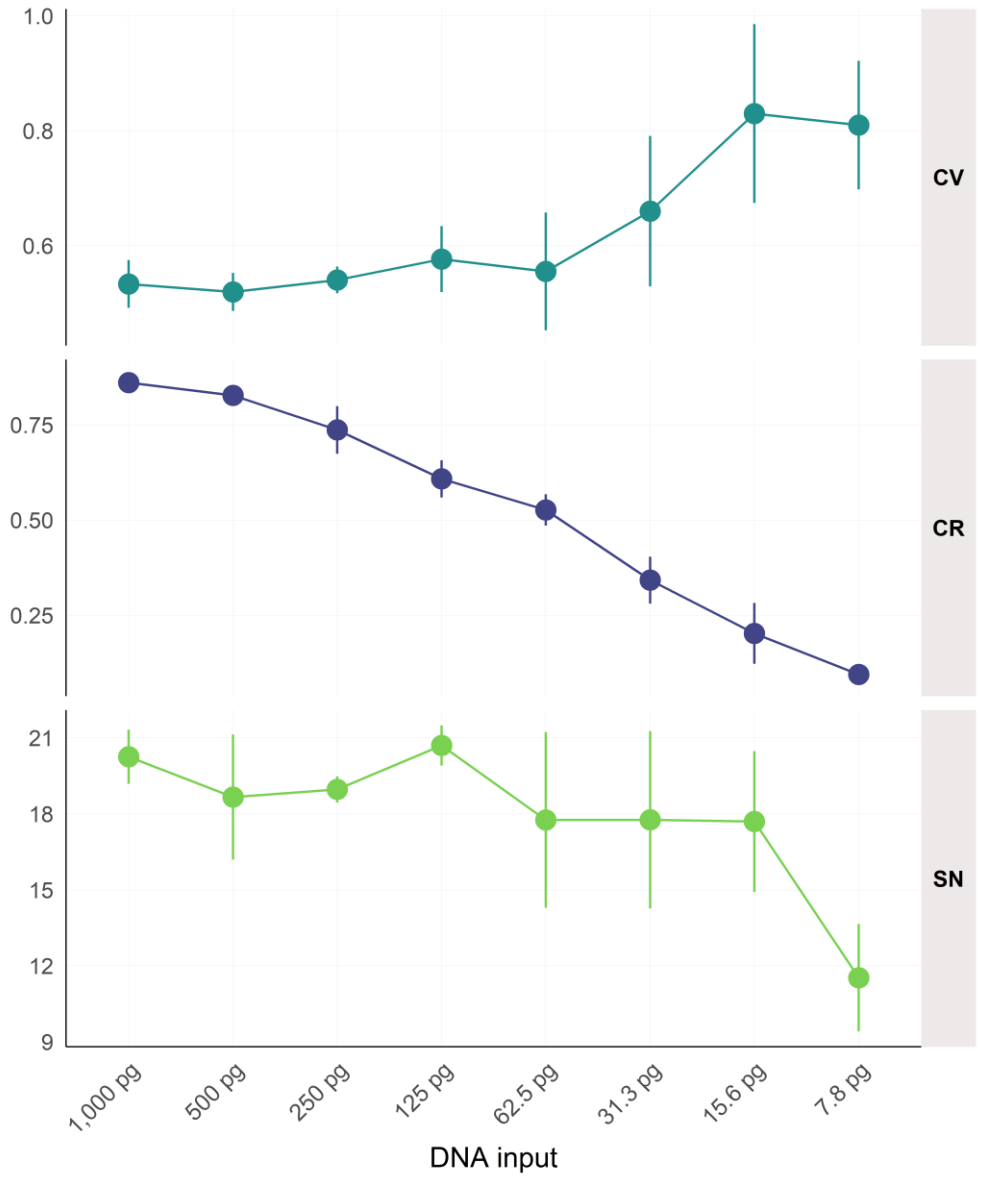

Abbildung 7: Mittelwerte der statistischen Parameter CV, CR und SN in Abhängigkeit verschiedener 2800M-Konzentrationen. Jede Konzentration wurde in Dreifachbestimmung analysiert, im Anschluss ein Mittelwert gebildet und die jeweilige Standardabweichung berechnet. Die Ergebnisse deuten auf eine gute Amplifikationsstabilität mittels *monSTR* bis zu einem DNA-Input von 62,5 pg hin. Geringere DNA-Mengen (< 62,5 pg) führten erwartungsgemäß zu höheren Lokus-*Coverage* Variationen, was sich u.a. in einem steigenden CV-Wert widerspiegelt. Außerdem fiel der Anteil der *On-Target-Reads* (CR) unter 50 %. SN blieb bis 15,6 pg stabil. CV = *Coverage Coefficient of Variation*; CR = *Calling Ratio*; SN = *Signal-to-Noise Ratio*; Amel = Amelogenin; pg = Pikogramm. (aus Silvery et al., 2020, Seite 6)

Als Maß für die Heterozygotiebalance wurde über alle heterozygoten STR-Systeme je Verdünnungsstufe die *Allel Coverage Ratio* (ACR) berechnet (s. Kapitel 2.2.5.8, Tabelle 13). Aus der Heat-Map (Abbildung 8) geht hervor, dass bei DNA-Inputmengen zwischen 1.000 pg und 125 pg heterozygote STR-Systeme ausbalanciert waren und eine durchschnittliche

ACR > 0,6 aufwiesen. Das System D8S1179 bildet hier eine Ausnahme, da bei einer DNA-Inputmenge von 125 pg die durchschnittliche ACR bei > 0,4 lag. Mit zunehmender Verdünnung der Referenzprobe nahmen die ACRs ab und Ungleichgewichte der heterozygoten Merkmale (ACR < 0,6) wurden ab einem DNA-Input von 62,5 pg beobachtet. Es konnte dabei jedoch kein spezifisches Muster für ein bestimmtes STR-System festgestellt werden. Bei einem DNA-Input von 62,5 pg betrug die durchschnittliche ACR 0,66 (± 0,16). Ab einem DNA-Input von 31,3 pg wurde das Ungleichgewicht der heterozygoten Merkmale ausgeprägter. Zusätzlich wurden ab einer Inputmenge von 15,6 pg auch allelische Dropouts häufiger beobachtet, weshalb die mittlere ACR für diese Verdünnungsstufe bei 0,46 (± 0,23) lag. Das STR-System D8S1179 wies über alle acht Verdünnungsstufen hinweg das größte Ungleichgewicht heterozygoter Marker mit einer durchschnittlichen ACR von 0,58 (± 0,13) auf. Das hochpolymorphe STR-System SE33 wies hingegen die stabilste durchschnittliche ACR (0,79 ± 0,19) auf.

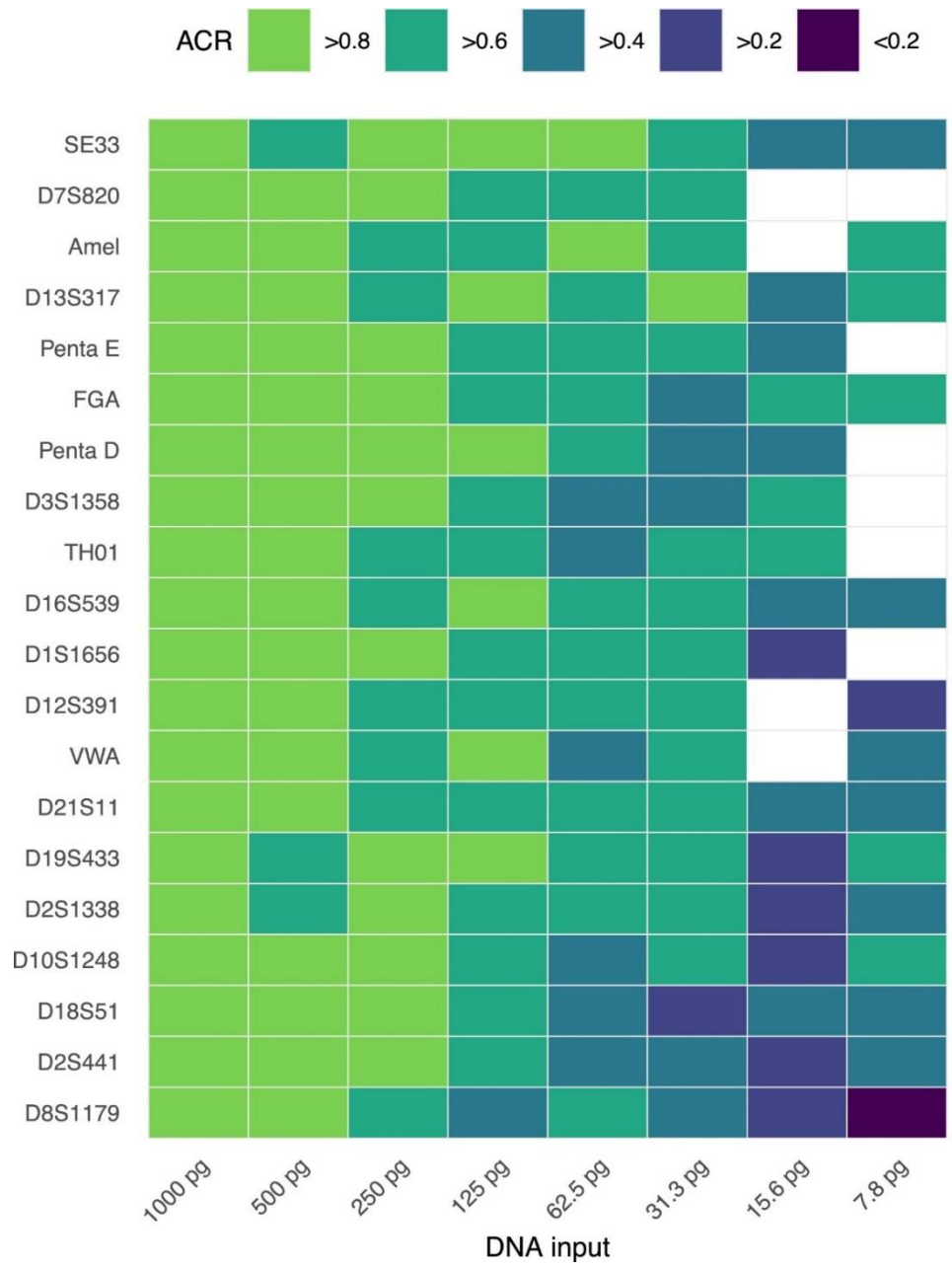

Abbildung 8: Heatmap der durchschnittlichen ACR über alle heterozygoten STR-Systeme der 2800M innerhalb verschiedener Konzentrationsstufen. Jede farbige Kachel repräsentiert dabei die durchschnittliche ACR über drei Replikate. Zur Veranschaulichung wurden die ACR-Werte fünf verschiedenen Farbgruppen zugeordnet (> 0,8, > 0,6, > 0,4, > 0,2 und < 0,2). Je heller dabei die Kachel, desto besser ist die ACR. Weiße Kacheln demonstrieren einen Allel-Ausfall in zwei von drei Replikaten. Heterozygote STR-Systeme sind nach ihrer durchschnittlichen ACR über alle acht Verdünnungsstufen von oben (ausbalanciert) nach unten (unausgeglichen) sortiert. Eine ausbalancierte ACR (> 0,6) wurde für alle STRs bis zu einem DNA-Input von 125 pg oder höher erreicht, außer für das STR-System D8S1179. ACR = *Allele Coverage Ratio*; Amel = Amelogenin; pg = Pikogramm. (aus Silvery et al., 2020, Seite 6).

3.4.5 DNA-Mischungen

Am Tatort sichergestellte Spuren können unterschiedliche Körperflüssigkeiten und/ oder Gewebe enthalten, die dabei ggf. von mehr als einer Person stammen können. Derartige Spurenmischungen werden in der forensischen Fallarbeit häufig angetroffen und ihre Dekonvolution kann sehr problematisch sein. Um die Sicherheit in Bezug auf die erfolgreiche Komponentenzerlegung heterogener DNA-Mischungen von *monSTR* zu testen, wurden unterschiedliche Mischungsverhältnisse (1:1, 1:4, 1:9, 1:19, 1:49) der Referenzproben SRM 2391c A (Hauptkomponente) und SRM 2391c C (Nebenkomponente) analysiert. In der statistischen Auswertung wurden nur diejenigen Allele der Nebenkomponente berücksichtigt, die weder mit den Allelen der Hauptkomponente noch mit deren Stutterprodukten überlappten (s. Kapitel 2.2.7.4, Tabelle 14) Um das tatsächlich beobachtete Mischungsverhältnis zu bewerten, wurde die *Minor Component Allele Frequency* (MAF) berechnet (s. Kapitel 2.2.5.8, Tabelle 13).

In Abbildung 9 wurde die MAF gegen das theoretisch vorliegende Mischungsverhältnis aufgetragen. Aus der Boxplot-Darstellung geht hervor, dass die MAF im Allgemeinen genau dem theoretisch vorliegenden Mischungsverhältnis entsprach. Bei einem vorgegebenen Mischungsverhältnis von 1:1 (50 % Anteil der Nebenkomponente) betrug die gemessene MAF 53,7 % (± 6,6 %). Eine größere Abweichung zwischen vorgegebenem und beobachtetem Mischungsverhältnis wurde bei einem Anteil von 20 % (1:4) beobachtet (MAF 15,1 % ± 4,1 %). Die Ergebnisse verdeutlichen, dass die ermittelte MAF ein guter Indikator für den prozentualen Anteil der Nebenkomponente darstellt. Obwohl bei zunehmender Verdünnung der Nebenkomponente Abweichungen in Bezug auf die MAF beobachtet wurden, konnte eine präzise Vorhersage für die anspruchsvolle 2 %-Mischung erreicht werden (MAF 2,0 % ± 0,7 %).

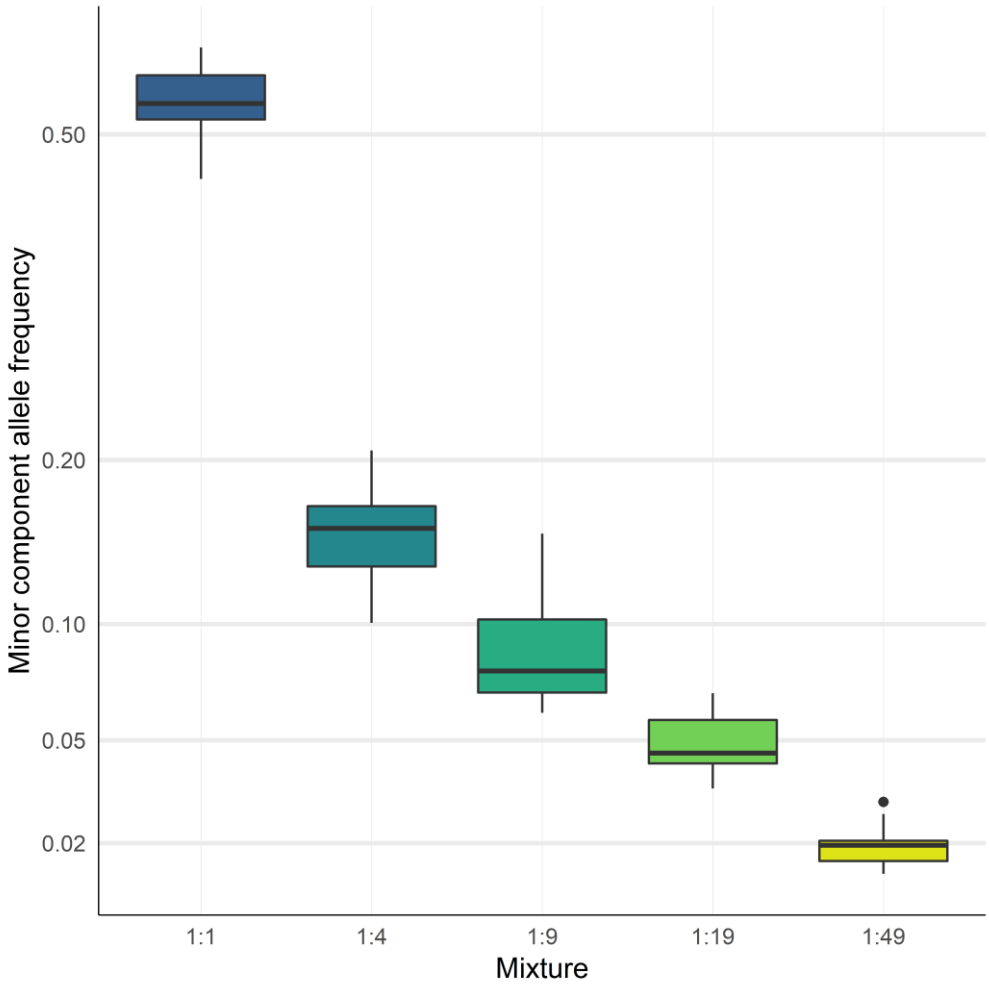

Abbildung 9: Durchschnittliche MAF unterschiedlicher Mischungsverhältnisse der SRM 2391c A (Hauptkomponente) und SRM 2391c C (Nebenkomponente). Für die Berechnung der MAF (Y-Achse) konnten je Mischungsverhältnis (X-Achse) insgesamt 54 eindeutige Allele der Nebenkomponente bewertet werden, da jedes Replikat in Dreifachbestimmung untersucht wurde. Die MAF wurde für jedes STR-System berechnet, indem die Allel-*Coverage* der Nebenkomponente durch die Allel-*Coverage* aller Systeme dividiert wurde. Die y-Achse ist zur besseren Lesbarkeit logarithmisch aufgetragen. MAF = *Minor Allele Frequency* (aus Silvery et al., 2020, Seite 7).

Bei einem Mischungsverhältnis von 1:1 (50 % Anteil der Nebenkomponente) konnte eine durchschnittliche Allel-*Coverage* von 2.037x (± 797x) der Nebenkomponente beobachtet werden (Tabelle 22). Mit kleiner werdendem Mischungsverhältnis nahm die Allel-*Coverage* kontinuierlich ab. So trat bei einem Mischungsverhältnis von 1:19 eine Allel-*Coverage* von 105x (± 17x) auf. Für die Mischungsverhältnisse 1:1 bis 1:19 wurden alle SRM 2391c C-Allele 100 % konkordant und präzise typisiert. Die Ergebnisse für das Mischungsverhältnis 1:49 (2 %) spiegeln die stochastische Natur der Amplifikation geringerer DNA-Mengen wider, da die Identifizierung der Nebenkomponente innerhalb dieser Mischung ungenau wurde. Es

wurden insgesamt 94,4 % der Allele der Nebenkomponente erfolgreich detektiert, obwohl Allel-Dropouts innerhalb der STR-Systeme CSF1PO, D8S1179 und D19S1248 beobachtet wurden.

Tabelle 22: Durchschnittliche Coverage und erfolgreich typisierte Allele der Referenzprobe SRM2391c C (Nebenkomponente) in Abhängigkeit des jeweiligen Mischungsverhältnisses.

Theoretisches Mischungsverhältnis	Mittlere Allel-Coverage (x) der Nebenkomponente	Detektierte Allele der Nebenkomponente (N=54)
1:1 (50 %)	2.037x ± 797x	100 %
1:4 (20 %)	470x ± 120x	100 %
1:9 (10 %)	254x ± 143x	100 %
1:19 (5 %)	105x ± 17x	100 %
1:49 (2 %)	42x ± 17x	94,4 %

3.4.6 Forensische Ringversuchsproben

Für die Untersuchung forensischer Ringversuchsproben wurde die genomische DNA der GEDNAP-Ringversuchsproben 54 bis 61 parallel mittels *monSTR* und der klassischen CE analysiert. Die MPS-Daten wurden im Anschluss sowohl mit den Ergebnissen der CE als auch mit den offiziellen GEDNAP-Sollwerten auf Konkordanz überprüft.

Im Rahmen dieser Evaluation wurden insgesamt 55 GEDNAP-Proben analysiert. Die Auswertung ergab überwiegend konkordante Profile zwischen der CE- und MPS-Technologie. Die Gesamtanzahl der mittels *monSTR* typisierten Allele betrug 2.730 und insgesamt 2.643 Allele konnten durch die CE erfolgreich typisiert werden (Tabelle 23 und Tabelle 24). Die MPS-Technologie bietet im Vergleich zur CE eine deutlich höhere Auflösung, da sowohl die Allellänge sowie die Sequenzinformation des untersuchten forensischen STR-Systems zur Verfügung stehen. Auf diese Weise können intra-allelische Sequenzvariationen (sog. Isoallele) identifiziert werden. Im untersuchten GEDNAP-Datensatz konnten insgesamt 85 zusätzliche Isoallele mittels MPS erfolgreich analysiert werden. Ein Allel-Dropout im System SE33 der Probe GEDNAP 54-PA war lediglich in der CE nachweisbar, während das *monSTR*-Panel im Vergleich zur offiziellen Ergebnistabelle den vollständig konkordanten triallelischen Genotypen 14, 15 und 26.2 lieferte. Ein weiterer Allel-Dropout im System D21S11 der Mischspur-Probe GEDNAP 55-S3 (offizieller Genotyp laut GEDNAP-Ergebnistabelle: 28, 29, 30, 31, 31.2) trat sowohl bei CE- als auch MPS-Analyse auf. Die GEDNAP-Probe 56-PC zeigte im STR-System vWA den Genotyp 16 und 17, welcher nur durch das *monSTR*-Panel

erfolgreich detektiert wurde. Innerhalb des untersuchten Probensatzes wurde eine Diskordanz zwischen den verwendeten Technologien für die Probe GEDNAP 54-S1 im System SE33 beobachtet. Der durch CE ermittelte Genotyp wies ein als "< 3.2" bezeichnetes Allel auf, wohingegen mittels *monSTR* das Allel "14" typisiert wurde. Es erfolgte daraufhin eine manuelle bioinformatische Analyse der Sequenzdateien. Eine 63 bp-Deletion im flankierenden Bereich des SE33-Lokus konnte hier als Ursache dieser Diskordanz nachgewiesen werden. Die GEDNAP-Probe 61-PB wurde aufgrund der Auswahl einer falschen Indexkombination aus der Auswertung entfernt.

Die Implementierung des hochpolymorphen STR-Markers SE33 in das *monSTR*-Panel konnte erfolgreich umgesetzt werden. Um ein besseres Verständnis für die Leistungsfähigkeit dieses STR-Systems zu erhalten, wurde die ACR und die relative Lokus-*Coverage* über alle GEDNAP-Einzelpersonenproben ausgewertet (Tabelle 25). Die Daten deuten insgesamt auf eine sehr gute Amplifikationseffizienz und Heterozygotiebalance des SE33-Markers hin. Selbst bei Genotypen mit Allellängen > 29 und einer heterozygoten Alleldistanz von bis zu 13 Repeateinheiten konnten ACRs zwischen 0,99 und 0,54 gemessen werden. Diese Ergebnisse sind vergleichbar mit den ACR-Werten anderer STR-Systeme. Wurde die Distanz der heterozygoten Allele innerhalb einer Probe deutlich größer (z.B. GEDNAP 56-S2; Alleldistanz: 14,2; ACR: 0,41), wurden Ungleichgewichte zwischen den Allelen häufiger, wodurch die ACR unter 0,5 fiel.

Tabelle 23: Anzahl der mittels monSTR und CE typisierten (Iso-) Allele für insgesamt 23 GEDNAP-Personenproben im Vergleich zur Referenz. Der Prozentsatz der insgesamt beobachteten *monSTR*- und CE-Allele im Verhältnis zur Referenz ist zusätzlich aufgeführt. P = GEDNAP-Personenprobe; A-C = Nummerierung der Personenproben; CE = *Capillary electrophoresis*.

Jahr	**Probenzeichnung**	**Detektierte Allele**			**detektierte Isoallele**
		Referenz	CE	monSTR	monSTR
2017	54-PA	42	41	42	
	54-PB	36	36	37	1
	54-PC	42	42	42	
	55-PA	40	40	40	
	55-PB	39	39	40	1
	55-PC	40	40	40	
2018	56-PA	39	39	39	
	56-PB	40	40	40	
	56-PC	35	34	36	1
	57-PA	35	35	38	3
	57-PB	40	40	40	
	57-PC	39	39	39	
2019	58-PA	39	39	40	1
	58-PB	36	36	36	
	58-PC	38	38	39	1
	59-PA	39	39	40	1
	59-PB	37	37	40	3
	59-PC	38	38	40	2
2020	60-PA	39	39	39	
	60-PB	41	41	42	1
	60-PC	36	36	37	1
	61-PA	38	38	40	2
	61-PC	35	35	37	2
	Gesamt	**883**	**881**	**903**	**20**
		100%	**100%**	**102%**	

Tabelle 24: Anzahl der mittels monSTR und CE typisierten (Iso-) Allele für insgesamt 32 GEDNAP-Spurenproben im Vergleich zur Referenz. Der Prozentsatz der insgesamt beobachteten *monSTR*- und CE-Allele im Verhältnis zur Referenz ist zusätzlich angegeben. S = GEDNAP-Spurenprobe; 1-4 = Nummerierung der Spurenproben; CE = *Capillary electrophoresis*.

Jahr	Probenzeichnung	Detektierte Allele			detektierte Isoallele
		Referenz	CE	monSTR	monSTR
2017	54-S1	86	86	91	5
	54-S2	39	39	39	
	54-S3	41	41	41	
	54-S4	64	64	68	4
	55-S1	40	40	41	1
	55-S2	68	68	71	3
	55-S3	88	87	92	5
	55-S4	38	38	39	1
2018	56-S1	38	38	38	
	56-S2	37	37	37	
	56-S3	38	38	38	
	56-S4	61	61	65	4
	57-S1	42	42	42	
	57-S2	39	39	40	1
	57-S3	83	83	90	7
	57-S4	67	67	68	1
2019	58-S1	36	36	37	1
	58-S2	63	63	65	2
	58-S3	88	88	93	5
	58-S4	39	39	41	2
	59-S1	67	67	69	2
	59-S2	38	38	39	1
	59-S3	80	80	83	3
	59-S4	37	37	39	2
2020	60-S1	38	38	38	
	60-S2	66	66	67	1
	60-S3	38	38	38	
	60-S4	80	80	82	2
	61-S1	40	40	40	
	61-S2	38	38	38	
	61-S3	84	84	91	7
	61-S4	62	62	67	5
	Gesamt	**1763**	**1762**	**1827**	**65**
		100%	**99,94%**	**103,63%**	

Tabelle 25: Heterozygoter Genotyp, ACR und Coverage des SE33-Lokus innerhalb der GEDNAP-Einzelpersonenproben. Die relative *Coverage* ist in Klammern angegeben und wurde berechnet, indem die SE33 Lokus-*Coverage* durch die *Coverage* aller 21 *monSTR*-Systeme dividiert wurde. ACR = *Allele Coverage Ratio*; P = Personenprobe; A-C = Nummerierung der Personenproben; S = Spur; 1-3 = Nummerierung der Spurenproben; b = intraallelische Sequenzvariante (sog. Isoallel).

Jahr	GEDNAP-Einzelprobe	Heterozygoter SE33 Genotyp	Alleldistanz	ACR	(%) Lokus-Coverage
2017	54-PB	19 / 19b	0	0,97	6.641x (8,4%)
	54-PC	25.2 / 30.2	5	0,97	5.689x (7,4%)
	54-S2	14.3 / 15	0,1	0,97	7.497x (9,6%)
	54-S3	28.2 / 29.2	1	0,85	9.749x (9,7%)
	55-PA	27.2 / 28.2	1	0,99	3.917x (6,0%)
	55-PB	19 / 22.2	3,2	0,91	6.483x (8,4%)
	55-PC	27.2 / 32.2	5	0,69	4.672x (5,3%)
	55-S1	16 / 16b	0	0,93	2.400x (5,6%)
	55-S4	16 / 19	3	0,96	5.958x (9,1%)
2018	56-PA	15 / 26.2	11,2	0,54	19.357x (12,9%)
	56-PB	16 / 31.2	15,2	0,26	13.068x (11,4%)
	56-PC	16 / 18	2	0,89	19.605x (15,2%)
	56-S2	15 / 29.2	14,2	0,41	16.085x (11,2%)
	56-S3	15 / 15.2	0,2	0,58	22.204x (13,7%)
	57-PA	19 / 26.2	7,2	0,76	16.774x (12,9%)
	57-PB	15 / 26.2	11,2	0,56	16.903x (14,5%)
	57-PC	20 / 28.2	8,2	0,6	12.512x (9,6%)
	57-S2	19 / 23.2	4,2	0,81	20.397x (14,6%)
2019	58-PA	18 / 20	2	0,9	5.832x (10,5%)
	58-PB	27.2 / 29.2	2	0,89	4.618x (7,87%)
	58-PC	15 / 26.2	11,2	0,54	19.357x (12,9%)
	58-S1	27.2 / 31.2	4	0,91	3.785x (7,7%)
	58-S4	16 / 29.2	13,2	0,83	4.252x (7,7%)
	59-PA	19 / 29.2	10,2	0,78	5.655x (8,3%)
	59-PB	27.2 / 27.2b	0	0,96	4.790x (7,5%)
	59-PC	18 / 29.2	11,2	0,95	5.660x (8,5%)
	59-S2	20.2 / 24.2	4,2	0,97	6.771x (10,2%)
	59-S4	26.2 / 26.2b	0	0,87	5.535x (9,8%)
2020	60-PA	20 / 20.2	0,2	0,9	5.148x (7,1%)
	60-PB	18 / 24.2	6,2	0,93	6.610x (8,1%)
	60-PC	27.2 / 27.2b	0	0,99	6.508x (7,6%)
	60-S1	16 / 29.2	13,2	0,85	6.717x (7,4%)
	60-S3	27.2 / 28.2	1	0,73	438x (6,2%)
	61-PA	14 / 17	3	0,61	5.544x (6,7%)
	61-PC	17 / 26.2	9,2	0,85	5.694x (8,7%)
	61-S1	19 / 24.2	5,2	0,89	5.492x (10,1%)
	61-S2	18 / 29.2	11,2	0,88	8.745x (6,1%)

3.4.7 Speziesspezifität

Um die Humanspezifität von *monSTR* zu bewerten, wurden genomische DNA-Proben verschiedener Haus- und Nutztierarten (Hund, Katze und Pferd) analysiert. Um eventuelle Kontaminationen identifizieren zu können, wurde zusätzlich das DNA-Profil der Tierhalter erstellt. Als Maß für die Spezifität innerhalb der PCR wurde von jeder Probe die CR berechnet (s. Kapitel 2.2.5.8, Tabelle 13).

Im Vergleich zur Referenzprobe 2800M mit einer durchschnittlichen CR von 86,82 % (± 2,19 %) ergaben die DNA-Proben der Haustiere niedrige CR-Werte (Abbildung 10). Die durchschnittliche CR für die Katzen-DNA betrug 0,37 % (± 0,40 %), die für die Hunde-DNA 0,09 % (± 0,11 %) und für die Pferde-DNA 0,08 % (± 0,14 %). Es konnten somit keine Hinweise auf eine starke spezifische Amplifikation mit DNA dieser Tierarten gefunden werden. Für kein Replikat der mitgeführten Negativkontrolle (NTC) wurden Allele oberhalb des in *toaSTR* eingestellten analytischen Schwellenwerts von zehn *Reads* detektiert. Sequenzdaten der Hundeproben wiesen z. B. Allele in CSF1PO, D7S820, TH01, D19S433 und D21S11 auf, während bei den Pferdeproben lediglich Allele im STR-System D10S1248 typisiert wurden. Die Sequenzdaten der Katzenproben lieferten Typisierungsergebnisse für D1S1656, SE33, D7S820, TH01, vWA, D19S433, D19S433, D21S11, und Amelogenin. Einige Allele, die innerhalb der Tierproben detektiert wurden, stimmten zudem sowohl in der Länge sowie in der Sequenz mit den Genotypen der entsprechenden Halter überein. In einem Replikat einer Hundeprobe wurde z.B. im STR-System CSF1PO das Allel 10 detektiert, welches auch seine Halterin besitzt. Im Gegensatz dazu konnte für die Replikate eins bis drei im Lokus TH01 das Allel 9.3 der Halterin nicht zugeordnet werden.

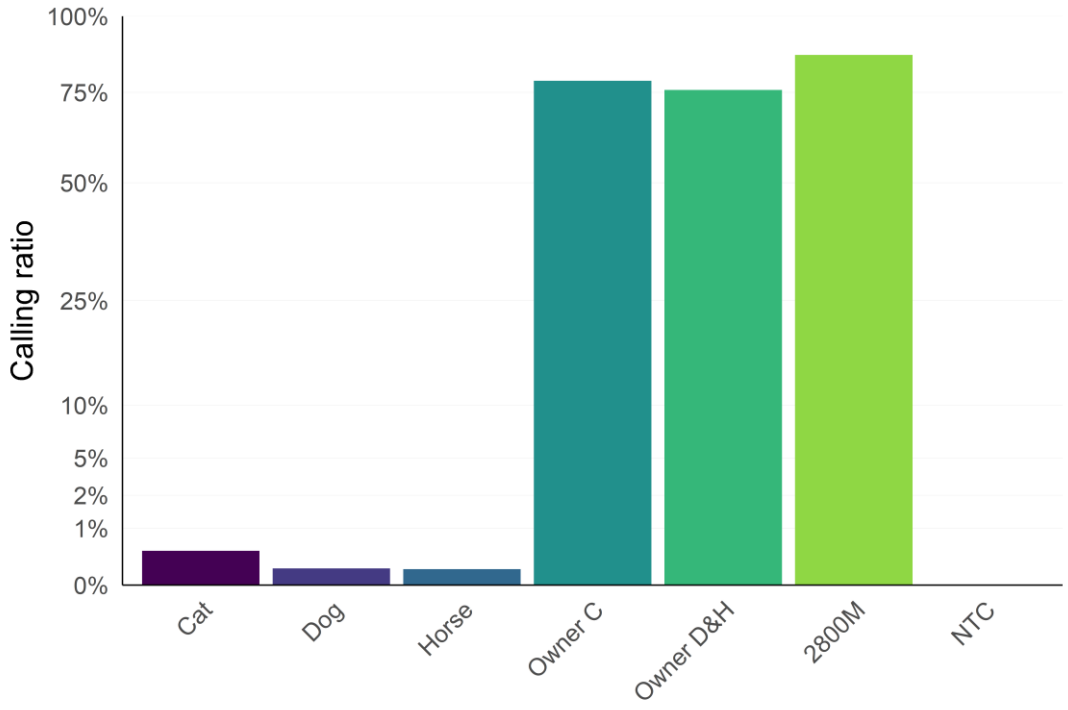

Abbildung 10: Durchschnittliche CR in Abhängigkeit dreier Haustierarten, den jeweiligen Tierhaltern, der 2800M und der Negativkontrolle. Die y-Achse wurde zur besseren Übersicht logarithmiert. CR = *Calling Ratio*; C = *Cat*/ Katze; D = *Dog*/ Hund; H = *Horse*/ Pferd; NTC = *No Template Control*/ Negativkontrolle, (aus Silvery et al., 2020, Seite 8).

3.4.8 Zusammenfassung der Validierungsergebnisse

Innerhalb dieser Validierung wurde die Leistungsfähigkeit des *monSTR*-Panels gemäß den empfohlenen SWGDAM-Richtlinien bewertet. Für diesen Zweck wurden insgesamt sechs Teilevaluationen durchgeführt (Konkordanz, Wiederholbarkeit und Reproduzierbarkeit, Sensitivität, Mischungsanalyse, die Eignung zur Analyse forensischer Ringversuchsproben und Speziesspezifität) und die statistischen Parameter CV, CR, ACR, SN und MAF analysiert. Die Ergebnisse zeigen, dass die genetischen Profile der SRM-Referenzproben (NIST) sowie die der GEDNAP-Ringversuchsproben eine nahezu vollständige Konkordanz zu den CE-basierten Genotypen aufweisen. Innerhalb der *Intra-* und *Inter-Run* Experimente wurde eine hohe Präzision sowohl zwischen unterschiedlichen MPS-Läufen als auch innerhalb derselben Sequenzierläufe ermittelt. Vollständige STR-Profile, eine ausgeglichene Interlokus-Balance und ein gutes CR-Verhältnis konnten mit einem DNA-Input von 62,5 pg DNA im Rahmen der Sensitivitätsstudie ermittelt werden. Gleichzeitig wurden insgesamt 76,4 % der Allele bei

einem DNA-Input von 7,8 pg erfolgreich typisiert. Die Analyse unterschiedlicher DNA-Mischungsverhältnisse zeigte, dass 94,4 % der Allele der Nebenkomponente innerhalb der 1:49-Mischung korrekt aufgelöst wurden. Die Ergebnisse deuten zudem darauf hin, dass der tatsächliche Anteil der Nebenkomponente auf der Grundlage der MAF genau berechnet werden konnte. Hinsichtlich der Speziesspezifität konnte keine spezifische Anreicherung bei DNA der getesteten Tierarten (Hund, Katze und Pferd) ausgemacht werden. Die zuvor beschriebenen Validierungsergebnisse veranschaulichen, dass das forensische Multiplexpanel *monSTR* ein valides Analyseverfahren für die forensische STR-Genotypisierung mittels *Massive Parallel Sequencing* darstellt.

Für die manuelle *Library-Preparation* von ca. 10-16 forensischen Proben werden etwa drei Stunden *Hands-on-Time* benötigt (Abbildung 11). Das Protokoll ist derzeit auf eine manuelle Probenbearbeitung ausgerichtet und enthält einige komplexe Arbeitsschritte.



Abbildung 11: Library-Preparation mittels monSTR und MiSeq. A) Es werden zwei Amplifikationsrunden durchgeführt, um zunächst die forensischen STR-Loci mittels spezifischer Primersequenzen zu amplifizieren und dabei gleichzeitig Überhangadapter (Illumina) anzubringen (mPCR) und um probenspezifische Indexadapter (sog. Barcodes) anzufügen (Index-PCR). Jede amplifizierte und eindeutig markierte *Library* wird anschließend mittels *AMPure Beads* aufgereinigt und mittels molekularbiologischer Standardmethoden quantifiziert. **B)** Im Anschluss an die Normalisierung werden die MPS-*Libraries* innerhalb eines Reaktionsgefäßes gepoolt, in die *MiSeq v2* Reagenzienkartusche pipettiert und in das *MiSeq* Sequenziergerät gebracht. Die generierten *Reads* werden anhand ihrer jeweiligen Indexsequenzen demultiplexed und in der Software *FastQC* einer Qualitätskontrolle unterzogen. Die forensischen Sequenzierdaten werden anschließend mit *toaSTR* (Ganschow et al., 2018) ausgewertet. mPCR = *Multiplex Polymerase Chain Reaction*.

4 Diskussion

Seit etwa 10 Jahren hat das *Massively Parallel Sequencing* (MPS), auch bekannt als *Next-Generation Sequencing* (NGS), immer mehr an Bedeutung gewonnen. Diese Technologie ermöglicht die simultane Sequenzierung mehrerer hundert Millionen Basenpaare innerhalb einer DNA-Sequenz. Gegenwärtig kommt die MPS-Technologie in vielen verschiedenen Bereichen der Diagnostik und Forschung zum Einsatz (Biesecker et al., 2012; Divne et al., 2010; Schuster, 2008). Im Feld der forensischen Genetik ermöglicht MPS u.a. die Sequenzierung von bis zu 35 STR-Loci und zusätzlicher forensischer Marker innerhalb eines einzigen Multiplex-Ansatzes und ist somit verschiedenen Limitationen der kapillarelektrophoretischen Auftrennung von STR-Markern nicht unterworfen (Børsting and Morling, 2015). Zum Zeitpunkt der Niederschrift sind, die aktuelle Implementierung dieser Technologie in forensischen Laboren betreffend, die Geräte der „Benchtop"-Kategorie *MiSeq* (Illumina) und *Ion Torrent S5* (Thermo Fisher) die vorherrschenden Instrumente (Alonso et al., 2017). Im Falle der *MiSeq*-Plattform basiert die Technik auf individuellen Adaptersequenzen (s. Kapitel 1.2.5), die während des *Indexing* Prozesses im Rahmen der *Library-Preparation* an beiden Enden des DNA-Fragments ligiert werden.

Aufgrund des zunehmenden Einsatzes der MPS-Technologie in forensischen Laboren ist der Bedarf an Multiplex-Kits gestiegen, die einen schnellen und effizienten Arbeitsablauf gewährleisten. Es wurden bis dato unterschiedliche Multiplex-Kits entwickelt und deren Leistungsfähigkeit ausführlich validiert (Köcher et al., 2018; Müller et al., 2018; van der Gaag et al., 2016; Zeng et al., 2015b). Das *Forenseq DNA Signature Prep Kit* der Firma Verogen stellt derzeit das größte kommerziell verfügbare MPS-Panel dar. Dieses Multiplexpanel besteht aus zwei verschiedenen Primer-Mixen, wobei der umfangreichere Mix B neben 58 autosomalen STR-Systemen und Amelogenin auch 172 informative SNPs abdeckt. Obgleich spezifische Primer für SE33 integriert wurden, wird dieses STR-System weder vom Hersteller beworben noch liefert die bioinformatische Auswertung mittels der *Universal Analysis Software* (UAS, Verogen) dem Nutzer Ergebnisse für SE33 (Borsuk et al., 2018; Köcher et al., 2018; Novroski et al., 2016). Zudem ist das *ForenSeq*-Panel nicht mit der RUO- (*Research Use Only*) Version des *MiSeq*-Systems (Illumina) kompatibel, sondern erfordert die Nutzung des *MiSeq FGx*

Forensic Genomics Systems. Einen weiteren Nachteil dieses MPS-Panels stellt die mangelnde Flexibilität bei der bioinformatischen Auswertung der Sequenzdaten dar. Da alle Marker innerhalb derselben Amplifikation erfasst werden, ist nicht möglich, die Analyse entsprechend den Anforderungen eines vorliegenden Falles zu modularisieren.

Weiterhin ist das *PowerSeq 46 GY System* (Promega) zu nennen, bei dem es sich um ein weiteres verfügbares kommerzielles MPS-Panel handelt, das Reagenzien zur Amplifikation autosomaler und Y-chromosomaler STRs enthält. Dieses Multiplexpanel umfasst die Kernloci des erweiterten *European Standard Set* (ESS) (Schneider, 2009) und des erweiterten *Combined DNA Index System* (CODIS) *Core Locus Set des Federal Bureau of Investigation* (FBI) *Laboratory* (Hares, 2015, 2012) mit Ausnahme des SE33-Markers (Zeng et al., 2015a).

Um eine kostengünstige und dennoch wettbewerbsfähige Alternative zu den bisher auf dem Markt verfügbaren kommerziell erhältlichen MPS-Panels anbieten zu können, wurde in einer eigenen Vorstudie zu den in dieser Dissertation beschriebenen Arbeiten die Entwicklung und systematische Optimierung des forensischen Multiplexpanels *monSTR* beschrieben (Ganschow et al., 2019). *MonSTR* ermöglicht die gleichzeitige Amplifikation von insgesamt 20 STR-Loci und Amelogenin innerhalb einer einzigen Reaktion und wurde für die Sequenzierung auf der Illumina *MiSeq*-Plattform optimiert. Die Auswahl der von *monSTR* abgedeckten forensischen Marker basiert dabei auf den ESS- und CODIS Kernloci. Eine Besonderheit stellt die Implementierung des hoch polymorphen Markers SE33 dar (Polymeropoulos et al., 1992), da dieses STR-System innerhalb anderer MPS-STR-Multiplexpanels bisher nicht berücksichtigt wurde (Müller et al., 2018; van der Gaag et al., 2016). Durch seine hochgradig polymorphe und heterozygote Natur weist dieser Tetranukleotidmarker eine komplexe Längen- und Sequenzvariation auf (Borsuk et al., 2018; Lászik et al., 2001; Wiegand et al., 1993), welche vor allem für die Dekonvolution komplexer Mischspuren von Bedeutung ist. Aufgrund dessen ist SE33 unerlässlich für die DNA-Profilmeldung an die DNA-Analyse-Datei (DAD) und in der Folge äußerst interessant im Bereich der Spurenroutine innerhalb Deutschlands. Die beiden Pentamer-Systeme Penta D und Penta E wurden aufgrund ihrer hohen Diskriminationskraft und der geringen Anfälligkeit für die Ausbildung von Stutterprodukten ausgewählt (Bacher und Schumm, 1998).

4.1 Analytische Leistungsfähigkeit des monSTR-Panels

Bei der Vorbereitung von MPS-*Libraries* aus genomischen DNA-Proben sind, neben einem erfolgreichen und spezifischen Primerdesign, weitere Faktoren entscheidend. Hierzu zählt beispielsweise die sorgfältige Auswahl einer zuverlässigen DNA-Polymerase (Wallin et al., 2002). Für die Amplifikation mittels *monSTR* wurde der *Q5 High-Fidelity Master Mix* der Firma New England Biolabs eingesetzt, der die *Proofreading*-Polymerase Q5 enthält. Menin und Nichols konnten 2013 bereits aufzeigen, dass diese Polymerase robuste und hochspezifische Multiplex-PCR Ergebnisse liefert, die kaum Optimierung weiterer Reaktionsbedingungen erfordern (Menin and Nichols, 2013). Ebenjenes Enzym gehört zu den robustesten DNA-Polymerasen auf dem Markt und wurde bereits erfolgreich für die schnelle PCR von STR-Markern eingesetzt (Romsos and Vallone, 2015). Aufgrund der Kompatibilität der mittels *monSTR* amplifizierten Produkte wurde für die Sequenzierung der *Libraries* die *MiSeq v2* Reagenzienkartusche verwendet, die die Sequenzierung von insgesamt bis zu 500 Basenpaaren gestattet. Die *v2*-Chemie liegt in zwei unterschiedlichen Kartuschentypen vor: als Nano-Fließzelle mit einem maximalen Output von 1 Mio. *Reads* und mit einem maximalen Output von 15 Mio. *Reads*. Beide Kartuschen wurden im Rahmen der vorliegenden Dissertation verwendet. Für die Entwicklungs- und Optimierungsstudie wurden die mit *monSTR* amplifizierten Produkte mittels 2 x 251 Zyklen *Paired-End*-Sequenzierung unter Verwendung der Nano-Fließzelle sequenziert (Ganschow et al., 2019). Im Anschluss sieht die bioinformatische Auswertung der Sequenzdaten zunächst die Bildung einer Konsensussequenz (= *Read-Merging*) vor, welche die gesamte STR-Region abdeckt. Das erforderliche *Read-Merging* führte jedoch zu mehrdeutigen Ergebnissen und zu einem erheblichen Verlust an Sequenzdaten (Ganschow et al., 2019, 2018). In der Konsequenz wurde für die Sequenzierung der Zielregionen im Rahmen der *monSTR*-Validierung die *Single-End*-Sequenzierstrategie gewählt, bei der gezeigt werden konnte, dass das Multiplexpanel auch mittels der *Single-End*-Sequenzierung von 1 x 400 Basenpaaren kompatibel ist (Silvery et al., 2020). Unter Verwendung dieser Konfiguration entfällt die Notwendigkeit des bioinformatischen *Read-Merging*. Allerdings muss darauf hingewiesen werden, dass bei diesem Setup eine zum Ende des *Reads* hin abnehmende Basenqualität beobachtet wurde, die zu unzuverlässigen Ergebnissen führen kann. Eine daraus resultierende mögliche Konsequenz ist die falsche Rekonstruktion der zugrundeliegenden Genotypen. Illumina hat bereits eine überarbeitete

Chemie zur Erhöhung der *Cluster Density* sowie zur Verbesserung des *Q-Scores* in den Markt eingeführt (*v3*-Chemie), wodurch eine tiefere Sequenzierung der Zielsequenzen und eine damit einhergehende Verbesserung der Sequenzdaten möglich ist. Die kostenintensivere *v3*-Kartusche wurde für die in dieser Dissertation beschriebenen Arbeiten jedoch nicht verwendet. In Anbetracht der verbesserten Reaktionsbedingungen soll die Kompatibilität des *monSTR*-Panels mit der Illumina *v3*-Chemie zeitnah getestet werden.

Als entscheidende Aspekte für zuverlässige und reproduzierbare Sequenzierungsergebnisse wurden bereits die Standardisierung und das Qualitätsmanagement bei der Generierung von MPS-Daten identifiziert (Endrullat et al., 2016b). Laborrichtlinien für MPS-basierte Anwendungen in der klinischen Diagnostik (Aziz et al., 2015) und der mikrobiellen Ökologie (Bokulich et al., 2013) fordern beispielsweise eine Qualitätskontrolle (QC) der Sequenzrohdaten zur Vorbereitung auf die anschließende Auswertung. Das Ignorieren von *Reads* geringer Qualität kann dabei dem Datensatz unzuverlässige und/ oder zufällige Informationen hinzufügen und zu einer falschen Interpretation der Ergebnisse führen. Derzeit herrscht im Feld der forensischen Genetik kein Konsens über die Verwendung bestimmter Qualitätsparameter, *Thresholds* oder bioinformatischer Softwarelösungen für die Prozessierung der Sequenzrohdaten. Von Seiten der *International Society of Forensic Genetics* (ISFG) wird zum gegenwärtigen Zeitpunkt lediglich eine Empfehlung hinsichtlich einer einheitlichen Nomenklatur ausgesprochen, um die Kompatibilität bestehender Datensätze zu gewährleisten (Parson et al., 2016). Überdies wurde innerhalb MPS-bezogener Studien in der Vergangenheit nicht über die Verwendung bestimmter Qualitätsparameter hinsichtlich der Sequenzrohdaten berichtet (Gettings et al., 2018; Hoogenboom et al., 2017; Zeng et al., 2015b). Im klinischen Kontext wurde z.B. ein *Q-Score* (s. Glossar) von mindestens Q20 vorgeschlagen (New York Department of Health, 2015). Bei einem Score von Q30 enthalten die generierten *Reads* abhängig von der Amplikonlänge, der verwendeten Plattform und Sequenzierchemie dann praktisch keine Fehler. Daher gilt Q30 als Maßstab für die Qualität des *Massively Parallel Sequencings*. Als Alternative zum durchschnittlichen *Q-Score* schlugen Edgar und Flyvbjerg (2015) die erwartete Anzahl von Fehlern innerhalb eines *Reads*, d.h. die Summe der Fehlerwahrscheinlichkeiten in einem *Read*, als besseren Indikator für die Lesequalität vor. Dabei auftretende Sequenzierfehler können durch *Quality Trimming* (Gezieltes Abschneiden von *Reads* mit Basen geringer Qualität), *Quality Filtering* (Gezieltes Verwerfen von Reads mit niedriger durchschnittlicher Basenqualität) (Sunyoung Kwon et al., 2013) oder durch *Read-*

Merging überlappender *Paired-End-Reads* und die darauf basierende Neuberechnung des *Q-Scores* (Edgar and Flyvbjerg, 2015) reduziert werden.

Die beschriebenen bioinformatischen Applikationen stellen immer einen Kompromiss zwischen dem Verlust von Sequenzdaten (Sensitivität) und der Qualität der verbleibenden Daten (Spezifität) (Del Fabbro et al., 2013) dar. Angesichts des Fehlens aktueller forensischer Richtlinien wurde im Rahmen dieser Validierungsstudie auf eine Qualitätskontrolle der zugrundeliegenden MPS-Rohdaten verzichtet. Interne Studien mit dem zugrundeliegenden Datensatz zeigten jedoch, dass unterschiedliche Qualitätsfilteralgorithmen und *Thresholds* erhebliche Auswirkungen auf die statistischen Parameter *Coverage coefficient of variation* (CV), *Calling Ratio* (CR), *Signal to Noise Ratio* (SN) und *Allele Coverage Ratio* (ACR) haben können. Ungünstige Filtereinstellungen können sogar zu fehlerhaften Analysen eines vorliegenden STR-Profils führen. Die für die Auswertung der Sequenzdaten verwendete Software *toaSTR* (Ganschow et al., 2018) ordnet beispielsweise einen *Read* einem bestimmten STR-Lokus erst dann zu, wenn er mit einem Paar eindeutiger Erkennungselemente übereinstimmt, die in der flankierenden Regionen positioniert sind. Dieser Mechanismus kann als inhärenter Schritt der Qualitätsfilterung betrachtet werden, da diese Erkennungselemente einerseits als Identifikatoren eines bestimmten STR-Lokus dienen und andererseits die Grenzen der entsprechenden Repeatregion aufzeigen. Die Qualitätskontrolle, die auf der autosomalen STR-Online-Datenbank und Plattform *STRidER* (Bodner et al., 2016) angeboten wird, umfasst mehrere Plausibilitätstests und die statistische Analyse der beobachteten autosomalen STR-Genotypen. Bisher beschriebene Erfahrungen der erfolgreichen Qualitätskontrolle von MPS-Rohdaten in anderen Bereichen wie z.B. der klinischen Diagnostik (Gargis et al., 2012) könnten mittels einer zentralen Software auf die forensischen Genetik übertragbar gemacht werden, sodass das Feld ebenfalls zu einem Konsens gelangen kann und ggf. bindende Richtlinien erarbeitet werden können.

4.2 Validierung des monSTR-Panels

Wie in vielen Anwendungsbereichen ist es auch in der forensischen Genetik gängige Praxis, neu entwickelte Analysekits vor ihrem routinemäßigen Einsatz einem detaillierten Validierungsprozess zu unterziehen. Auf dieser Grundlage wurde innerhalb der in dieser Dissertation beschriebenen Arbeiten die Validierung des forensischen Multiplexpanels

monSTR gemäß den empfohlenen Richtlinien der *Scientific Working Group for DNA Analysis Methods* (SWGDAM) durchgeführt, die in der forensischen Community internationale Anerkennung finden (Scientific Working Group on DNA Analysis Methods, 2016). Zu diesem Zweck wurden die nachstehenden Validierungsstudien geplant und bearbeitet: Konkordanz, Sensitivität, Wiederholbarkeit und Reproduzierbarkeit, die Analyse unterschiedlicher DNA-Mischungsverhältnisse, die Sequenzierung forensischer Ringversuchsproben und Speziesspezifität (Silvery et al., 2020).

Während der Amplifikation forensischer STR-Marker kann die Entstehung einer gewissen Menge Nebenprodukte kontinuierlich beobachtet werden. Dieses Phänomen wird durch das Dissoziieren der DNA-Polymerase während der Replikation verursacht (Bovo et al., 1999; Klintschar and Wiegand, 2003; Schlötterer and Tautz, 1992; Viguera et al., 2001). Typischerweise werden Stutterprudukte in CE-generierten Daten beobachtet, die eine Repeateinheit kürzer (N-1 Stutter) oder länger (N+1 Stutter) im Vergleich zu dem echten biologischen Allel sind (Walsh et al., 1996). Weitere Stutterprodukte (z. B. N-2, N-3, usw.) werden innerhalb der kapillarelektrophoretischen Auftrennung nicht beobachtet, da das Signal für diese seltenen Artefakte zumeist nicht von der Hintergrundfluoreszenz unterschieden werden kann. Unter Verwendung der MPS-Technologie können mehrere Stutterprodukte (N-1, N+1, N-2, N-3, usw.) gleichzeitig analysiert werden, insbesondere, wenn die erzielte *Coverage* einer Probe sehr hoch ist. Diese Stutterprodukte können sich auf die Interpretation der Ergebnisse auswirken. Innerhalb einer komplexen DNA-Mischung ist es daher schwierig festzustellen, ob es sich bei einem derartigen Nebenprodukt um das echte Allel einer Nebenkomponente oder um den Stutter eines benachbarten Allels handelt (Budowle et al., 2009). Wenngleich eine umfassende Analyse der durch Amplifikation mittels *monSTR* verursachten Stutterartefakte nicht Teil der vorliegenden Validierung war, wurde anhand des *Standard Reference Materials* (SRM) 2391c A, B und C der prozentuale N-1 Stutter ermittelt. In diesem Zusammenhang konnte gezeigt werden, dass der mittels *monSTR* beobachtete Stutter zwischen 5 % und 20 % lag und damit vergleichbar mit den Ergebnissen von Hussing et al. (2018) ist. Die Arbeitsgruppe beschrieb dabei N-1 Stutter-*Thresholds* für das *ForenSeq DNA Signature Prep Kit*. Das Pentamer-System Penta E produzierte den niedrigsten Stutter (*monSTR*: 0,4 - 6,8 %; *ForenSeq*: 2,0 - 12,3 %), während für das System D12S317 hohe Stutterprodukte beobachtet wurden (*monSTR*: 4,6 - 26,9 %; *ForenSeq*: 3,1 - 24,8 %). Aufgrund der Abhängigkeit der beobachteten Stutterprodukte vom jeweiligen Genotypen, insbesondere

von der Länge der ununterbrochenen Repeateinheiten der STR-Allele (Brookes et al., 2012), ist das beobachtete Probenspektrum bedingt durch die begrenzte Stichprobenanzahl statistisch nicht aussagekräftig. Für eine detailliertere Beurteilung der mittels *monSTR* produzierten N-1 Stutter ist die Analyse weiterer Proben notwendig.

Zur Beurteilung der Robustheit von *monSTR* wurden Validierungsstudien zur Sensitivitätsgrenze, unterschiedlichen Mischungsverhältnissen und der Sequenzierung einiger nichtmenschlicher Proben durchgeführt. Häufig stehen für die forensische STR-Analyse nur geringe Mengen an genomischer DNA zur Verfügung (Gill et al., 2015), so dass eine erfolgreiche Untersuchung eine hohe Sensitivität der eingesetzten Analysekits erfordert. Valide Angaben zur Sensitivität eines STR-Multiplexpanels sind von entscheidender Bedeutung für dessen Einsatz in der forensischen Routineanalytik, da DNA-Proben aus der realen Fallarbeit häufig durch Umweltexposition und andere Prozesse degradiert sind, weshalb nur wenige Fragmente in einer Probe von analysierbarer Länge vorhanden sein können. Hinsichtlich seiner Nachweisgrenze erwies sich *monSTR* als konkurrenzfähig zu den zuvor aufgeführten kommerziell erhältlichen MPS-Multiplexpanels (Jäger et al., 2017; Müller et al., 2018). Während *monSTR* innerhalb der Entwicklungsstudie zunächst für die Amplifikation von 1.000 pg DNA-Input optimiert wurde (Ganschow et al., 2019), wurden in dieser Validierungsstudie vollständige STR-Profile mit DNA-Inputmengen bis herab zu 62,5 pg generiert. Die Arbeitsgruppe von Jäger et al. (2017) berichtete vergleichbare Sensitivitätsdaten unter Verwendung des *ForenSeq DNA Signature Prep Kits* in Kombination mit dem *MiSeq FGx*. Müller et al. (2018) validierte MPS-Multiplexkits für die *Ion Torrent*-Plattform und erzielte Sensitivitätsergebnisse, die ebenfalls mit den in dieser Dissertation berichteten Ergebnissen vergleichbar sind. Innerhalb des dynamischen Analysebereichs von 1.000 pg bis 62,5 pg wurden vollständige und reproduzierbare STR-Genotypen der Referenzprobe 2800M mit einer jeweils ausbalancierten *Coverage* über alle STR-Loci generiert. Zugleich stimmen die Sequenzierergebnisse vollumfänglich mit den Ergebnissen der CE STR-Typisierung überein. DNA-Inputmengen von 15,6 pg und 7,8 pg führten zu Allel- und Lokus-Dropouts. Dennoch wurden insgesamt 76,4 % aller Allele innerhalb der niedrigsten Verdünnungsstufe (7,8 pg) korrekt typisiert. Diese Ergebnisse können durch stochastische Effekte während der Amplifikation geringer DNA-Inputmengen erklärt werden (Gill et al., 2015). Darüber hinaus könnte die aufwendige *Library-Preparation* die verschiedene manuelle, variationsanfällige

Pipettierschritte umfasst, einschließlich *bead*-basierter Aufreinigungen zusätzliche Schwankungen innerhalb der Lokus-*Coverages* verursachen (Riman et al., 2017).

Im Vergleich zur CE-basierten STR-Typisierung verfügt die MPS-Technologie über eine gewisse Toleranz in Bezug auf forensische Marker, die in der Sequenzierung eine geringere oder bessere Signalqualität aufweisen, da die Schwierigkeit mit der Interpretation, die bei überladenen, gesättigten STR-CE-Profilen auftritt, hier nicht beobachtet wird. Nichtsdestotrotz können starke Schwankungen einzelner Lokus-*Coverages* die MPS-Sequenzierung und anschließende Ableitung des Genotyps beeinträchtigen und sollten durch eine vorherige umfangreiche Optimierungsstudie des verwendeten Analysekits unbedingt vermieden werden. Bei der Amplifikation geringerer DNA-Inputmengen liegt immer ein Überschuss an targetspezifischen Primern vor, wodurch die Bildung sog. Primerdimere begünstigt wird. Dadurch kann auch die in dieser Studie beobachtete niedrigere Amplifikationsspezifität (*Calling Ratio,* CR) ab einer DNA-Inputmenge von 31,3 pg erklärt werden. Gleichzeitig wurden ab einem DNA-Input von 15,6 pg bzw. 7,8 pg erwartungsgemäß Allel- bzw. Lokus-Dropouts beobachtet, weshalb der Anteil der *On-Target-Reads* (CR) unter 50 % fiel.

Heterozygote STR-Systeme weisen zwei Allele auf, die sich hinsichtlich ihrer Repeateinheiten und *Coverages* unterscheiden. Unterschiedliche *Coverages* können dabei Aufschluss über die Menge an amplifizierter DNA einer Probe geben. Werden beispielsweise lediglich geringere Mengen an DNA amplifiziert, ist aufgrund von stochastischen Effekten während der PCR mit einer höheren Variation innerhalb der *Coverages* beider heterozygoter Allele zu rechnen. Im Rahmen der Sensitivitätsstudie wurde die *Allele Coverage Ratio* (ACR) für alle heterozygoten STR-Systeme bestimmt. Eine den forensischen Anforderungen entsprechende ACR > 0,6 wurde bis zu einem DNA-Input von 125 pg erreicht. DNA-Inputmengen unter 62,5 pg wirkten sich negativ auf die ACR einzelner STR-Systeme aus, was durch eine erhöhte Variation innerhalb der *Coverages* beobachtet werden konnte.

Das Potential von *monSTR*, den Nachweis der Allele der Nebenkomponente einer DNA-Mischung zu ermöglichen, wurde im Rahmen der Mischungsstudie untersucht. Die Grundlage hierfür bildeten DNA-Mischungen zweier Referenzproben, die als Haupt- und Nebenkomponente eingesetzt wurden, in fünf unterschiedlichen Verhältnissen. Die Validierungsstudie ergab, dass *monSTR* alle Allele der Nebenkomponente bis zu einem Mischungsanteil der Nebenkomponente von 5 % erfolgreich detektieren konnte. Diese Ergebnisse sind mit denjenigen anderer Arbeitsgruppen in Bezug auf das *ForenSeq DNA*

Signature Prep Kit (Churchill et al., 2016; Guo et al., 2017; Jäger et al., 2017; Köcher et al., 2018) sowie aus Studien auf der *Ion Torrent*-Plattform (Guo et al., 2016) vergleichbar. Innerhalb der anspruchsvollsten DNA-Mischung (2 %, 1:49) konnten noch 94,4 % aller Allele der Nebenkomponente mit einer insgesamt geringen Variation über die drei untersuchten Replikate präzise typisiert werden. Dropouts wurden in den STR-Systemen CSF1PO, D8S1179 und D19S1248 beobachtet. Zusätzlich wurde das tatsächlich in der DNA-Mischung vorliegende Mischungsverhältnis auf Grundlage der *Minor Component Allele Frequency* (MAF) geschätzt. Die Ergebnisse demonstrieren, dass dieser statistische Parameter ein insgesamt nützlicher Indikator für die Bewertung des tatsächlich vorliegenden Mischungsverhältnisses war und *monSTR* Zwei-Personen-Mischungen erfolgreich dekonvolvieren konnte.

Darüber hinaus ist das *monSTR*-Panel für die eindeutige Identifizierung humaner STRs geeignet. Dies konnte innerhalb der Speziesspezifität gezeigt werden, da keine eindeutigen Hinweise auf eine spezifische Amplifikation mit DNA der getesteten Spezies gefunden wurden. Die dabei in geringer Ausprägung darstellbaren STR-Allele in den Tierproben (CR < 0,5 %) sind größtenteils auf kontaminante DNA der Halter oder anderer Personen zurückzuführen, die mit den Tieren in engem Kontakt standen. Die hier generierten Ergebnisse sind zudem nicht unerwartet, da für mehrere forensische PCR-Multiplexkits bereits die Entstehung von STR-Signale für einige nicht-menschliche Spezies beschrieben wurde (Hennessy et al., 2014; McLaren et al., 2014). Grundlage für diese Validierungsstudie waren jedoch Proben gängiger Haus- und Nutztierarten. Zur Erzielung einer besseren Aussage hinsichtlich Speziesspezifität von *monSTR*, sollte das Probenkollektiv in weiterführenden Studien um zusätzliche Tierarten erweitert werden. Die Arbeitsgruppe um Jäger et al. (2017) untersuchte z.B. im Rahmen der Evaluation des *MiSeq FGx Forensic Genomics Systems* die genomische DNA von Hausschwein (*Sus scrofa domesticus*), Hamster (*Cricetinae*) und Rind (*Bos taurus*). Hierzu könnten in weiterführenden Studien zunächst mittels *in-silico* PCR die theoretischen Amplifikationsergebnisse unterschiedlicher Tiergenome in Bezug auf die spezifischen *monSTR*-Primer simuliert werden.

Ein weiteres Ziel der Validierung bestand darin, die Genauigkeit und Präzision des forensischen Multiplexpanels zu bewerten. Zu diesem Zweck wurde eine umfassende Analyse hinsichtlich *Intra*- und *Inter-Run* Präzision und Konkordanz zwischen MPS- und CE-Ergebnissen in Bezug auf unterschiedliche Referenz- und forensische Ringversuchsproben durchgeführt. Zusammenfassend kann herausgestellt werden, dass für unterschiedliche Referenzproben

vollständig konkordante und zuverlässige Sequenzdaten generiert wurden. Diese Ergebnisse sind mit kommerziell verfügbaren MPS STR-Kits vergleichbar (Churchill et al., 2016; Müller et al., 2018; Zeng et al., 2015b). Darüber hinaus wurde eine vollständige längen- und sequenzbasierte Konkordanz der 2800M hinsichtlich der bekannten Sollwerte (National Institute of Standards & Technology, 2018; Steffen et al., 2015) ermittelt. Zudem belegen die Sequenzierergebnisse, dass das Multiplexpanel *monSTR* und das *MiSeq*-System bei mehreren *Library-Preparations* und über verschiedene Sequenzierläufe konsistente und präzise Ergebnisse liefert. Wenn die *Library-Preparation* von verschiedenen Personen durchgeführt wurde, waren die Ergebnisse hinsichtlich der statistischen Parameter CV, ACR, CR und SN vollständig reproduzierbar. Die statistischen Parameter zeigten lediglich eine geringe Variation, sobald die MPS-Libraries innerhalb eines Sequenzierlaufs oder über einzelne Läufe hinweg wiederholt wurden. Variationen über verschiedene MPS-Läufe hinweg können ggf. durch weitere Optimierungsstudien behoben, oder durch die Umstellung auf eine automatisierte *Library-Preparation* möglicherweise vermieden werden (s. Kapitel 4.3).

Die Analyse von forensischen Ringversuchsproben der *German DNA Profiling Group* (GEDNAP) bewies die Kompatibilität des Multiplexpanels mit artifiziell hergestellten Spuren. Es konnte eine vollständige Konkordanz zwischen MPS und CE-Ergebnissen erzielt werden. Als Ausnahme sei an dieser Stelle eine Diskordanz im Lokus SE33 der Probe 54-S1 zu nennen. In der Probe 54-S1 wurde mittels der CE ein *Off-Ladder* Allel (<3.2) beobachtet, wohingegen mittels MPS-Technologie das Allel "14" nachgewiesen wurde. Ursächlich für diese Diskordanz wurde eine Deletion von 63 bp innerhalb der flankierenden Region des SE33-Lokus festgestellt, die sowohl durch das CE-Kit als auch das *monSTR*-Panel abgedeckt wurde. Die Diskrepanz wurde bioinformatisch ausgewertet und lässt sich dadurch erklären, dass die Allelbestimmung der CE-basierten kapillarelektrophoretischen Auftrennung die gesamte Amplikonlänge einschließlich der flankierenden Region berücksichtigt, wohingegen die Allelbestimmung unter Zuhilfenahme der Software *toaSTR* (Ganschow et al., 2018) ausschließlich auf der Bestimmung der Repeatregion eines STR-Lokus erfolgt. Dementsprechend können Insertionen und/ oder Deletionen (InDels), die sich außerhalb der Erkennungselemente von *toaSTR*, aber innerhalb des vom verwendeten CE-Kit amplifizierten Fragments befinden, zu abweichenden Ergebnissen hinsichtlich MPS und CE führen. Diese Art der Diskordanz ist jedoch nicht nur zwischen CE und MPS zu beobachten, sondern kann auch zwischen verschiedenen CE-Kits auftreten. *ToaSTR* hat die Repeatregion der Probe 54-S1 korrekt charakterisiert. Die

vorliegende Diskordanz verdeutlicht dabei die Bedeutung der Positionierung dieser spezifischen Erkennungselemente für eine korrekte sequenz- und längenbasierte Zuordnung der Genotypen (Ganschow et al., 2018). Der Fall verdeutlicht gleichzeitig den besonderen Informationsgehalt der flankierenden STR-Region (Novroski et al., 2016; Phillips et al., 2018). Allerdings unterstützt *toaSTR* derzeit noch nicht die Analyse der flankierenden Region, weshalb für die Auswertung weitere bioinformatische Tools einbezogen werden müssen.

Einer der vielversprechendsten Vorteile der MPS-Technologie, welcher im Rahmen dieser Validation bestätigt wurde, ist das Potenzial, intra-allelische Sequenzvarianten (sog. Isoallele) nachzuweisen und zu identifizieren (Gettings et al., 2016, p. 22, 2015; Parson et al., 2016; Warshauer et al., 2015). Isoallele ermöglichen dabei eine bessere Auflösung der genetischen Unterscheidung von Individuen und können bei der Interpretation komplexer DNA-Mischungen von entscheidender Bedeutung sein. Dieser Vorteil ermöglichte die Identifizierung von insgesamt 85 zusätzlichen Isoallelen im zugrundeliegenden GEDNAP-Datensatz.

Die Implementierung des SE33-Lokus in das *monSTR*-Panel stellt eine Besonderheit dar, weshalb bei der Auswertung der Ringversuchsproben ein spezieller Fokus auf die Leistungsfähigkeit dieses STR-Systems gelegt wurde. Insgesamt deuten die Validierungsergebnisse auf eine hohe Amplifikationseffizienz und Heterozygotie-Balance (ACR > 0,8) dieses Markers hin, welche selbst bei Genotypen mit längeren Allelen (> 30) und größeren Alleldistanzen von bis zu 13 Wiederholungseinheiten (z.B. 60-S1) vergleichbar mit anderen STR-Systemen war. Dies lässt sich als ein gutes Ergebnis im Gegensatz zu den Resultaten von Borsuk et al. (2018) bewerten, die von „Herausforderungen“ bei der Interpretation von SE33-Sequenzdaten in Bezug auf das *ForenSeq*-Panel berichteten. Innerhalb einer umfassenden Bevölkerungsstudie wiesen die Autoren der Publikation nach, dass heterozygote Imbalancen mit zunehmenden Alleldistanzen im SE33-System häufiger auftraten. Es wird angenommen, dass die gute Leistungsfähigkeit von SE33 im *monSTR*-Panel auf die Optimierung der Reaktionsbedingungen als Folge einer statistischen Versuchsplanung sowie dem effektiven Primerengineering zurückzuführen ist (Ganschow et al., 2019). Die Ergebnisse sollten jedoch weiter validiert werden, da die GEDNAP-Ringversuchsproben bisher nur als einzelnes Replikat sequenziert wurden.

4.3 Optimierungsmöglichkeiten und Ausblick

Die ursprüngliche Zielstellung für die Entwicklung von *monSTR* war es, eine wirtschaftlich attraktive Alternative zu kommerziell erhältlichen Panels für die *MiSeq*-Plattform anzubieten (Ganschow et al., 2019). In der Nachbetrachtung belegen die Ergebnisse der Validierung eine Eignung des forensischen Multiplexpanels für die STR-Sequenzierung mittels *Massively Parallel Sequencing* sowie eine Lieferung von zuverlässigen Sequenzierdaten. In einer Reihe experimenteller Studien wurde die Leistungsfähigkeit des Multiplexkits umfassend untersucht. Im Zuge dessen konnte demonstriert werden, dass *monSTR* eine adäquate Anwendung für die forensische Fallarbeit darstellt. Demgegenüber wurden die Grenzen des Multiplexkits aufgezeigt, die zukünftig einige Modifikationen erfordern, um den Arbeitsablauf für die tägliche forensische Laborroutine weiter zu optimieren.

Das Protokoll ist derzeit auf eine manuelle Probenbearbeitung ausgerichtet und enthält einige komplexe Arbeitsschritte, wodurch Fehler und/ oder Kontaminationen innerhalb des Probensatzes auftreten können. Diese Tatsache kann die Implementierung des Workflows für Routineanalysen in einem forensischen Labor mit hohem Probenaufkommen erschweren. Zur Lösung dieses Problems im Hinblick auf zukünftige Forschungsarbeiten am *monSTR*-Panel, wird derzeit an einem vollautomatisierten Arbeitsablauf gearbeitet, der alle Schritte der *Library-Preparation* abdeckt. Erste Erfahrungswerte konnten hierzu auf dem *Biomek i7* (Beckman Coulter) erzielt werden, der im Labor für die Bearbeitung der SARS-CoV-2 Vollgenomsequenzierung routinemäßig eingesetzt wird. Resultierend daraus wurde bereits ein Skript für die *bead*-basierte Aufreinigung der MPS-*Libraries* geschrieben und auf dem *Biomek i7* implementiert, dessen Funktion in Wasserläufen bereits erfolgreich verifiziert werden konnte. Ein vollautomatisiertes Protokoll der *Library-Preparation* wurde im forensischen Kontext für das *ForenSeq DNA Signature Prep Kit* von Laurent et al. (2017) entwickelt.

Darüber hinaus müssen weitere Studien hinsichtlich komplexer DNA-Mischungen durchgeführt werden, um ein detaillierteres Verständnis zur Fähigkeit des *monSTR*-Panels komplexere DNA-Mischungen zu dekonvolvieren, zu erhalten. Im Zuge dieser Validierung wurden zunächst nur Zwei-Personenmischungen von Referenzproben in unterschiedlichen Mischungsverhältnissen analysiert. Diese bilden die Realität in der forensischen Fallarbeit nur unzureichend ab. DNA-Mischungen mehrerer Personen/ Spezies und/ oder Geweben sind eher

repräsentativ für einen Tatort und sollten aus diesem Grund von dem verwendeten Analysekit analysiert werden können.

Zur Erzeugung eines STR-Profils mittels CE werden in der Regel DNA-Fragmente von ca. 100 bis 450 bp benötigt (Romsos and Vallone, 2015). Je schlechter dabei der Zustand einer Probe ist, z.B. in Folge einer Exposition mit Umwelteinflüssen sowie einer Verunreinigung mit Umweltsubstanzen, desto schwieriger gelingt die erfolgreiche Analyse eines DNA-Abschnitts bedingt durch fortgeschrittene Degradationsprozesse oder die Gegenwart von PCR-Inhibitoren (Alaeddini et al., 2010). Begründet durch diesen Umstand führt ein Signalverlust, vor allem längere Fragmente betreffend, zu einem genetischen Teilprofil mit partiellen oder kompletten Lokus-Dropouts. Daher ist die weitergehende Untersuchung realistischer forensischer Spuren eine zwingende Voraussetzung, um ein noch besseres Verständnis hinsichtlich der Leistungsfähigkeit von *monSTR* für die forensische Fallarbeit zu entwickeln. Die interne Sequenzierung einer kleinen Stichprobe forensischer Spurenträger (z.B. Fingernägel, Knochenmaterial, Haare, usw.) hat ergeben, dass das Multiplexpanel bei der Amplifikation an seine analytische Grenze stößt. STR-Teilprofile und eine nicht zufriedenstellende Proben- und Lokus-*Coverage* traten wiederholt auf (nicht Teil dieser Dissertation). Als mögliche Erklärung sei auf die in Kapitel 2.2.1 definierten Qualitätskriterien der Primersequenzen verwiesen. Die *monSTR* Amplikons wurden mittels zweier bioinformatischer Tools derartig entwickelt, dass sie eine mittlere Amplikongröße von 250-350 bp aufweisen (Ganschow et al., 2019). Aus diesem Grund stellt die Bearbeitung von Spurenmaterial ausreichender Menge die Amplifikation mittels *monSTR* vor keine besondere Herausforderung. DNA-Spuren mit einer reduzierten Qualität und/ oder Quantität (sog. Low-Template-DNA) unterliegen jedoch durch ihre geringe Anzahl an Ausgangsmolekülen von Beginn an stochastischen Effekten wie z.B. Amplifikationsausfällen oder Imbalancen von heterozygoten Allelen (Gill, 2001). Da die Fragmentgrößen von Low-Template-DNA Proben deutlich kürzer sein können, kann der Amplifikationserfolg unter Verwendung von *monSTR* aufgrund der darin erfassten Amplikongrößen deutlich eingeschränkt sein. Einige kommerziell erhältliche STR-Kits verfolgen aus diesem Grund beim Assay-Design bereits die Strategie, die Größe der Amplikons zu reduzieren. Ein wesentlicher Vorteil dieser sog. „Mini-STRs“ ist, dass sie weniger empfindlich gegenüber PCR-Inhibitoren und DNA-Degradation sind (Coble and Butler, 2005). Das *ForenSeq DNA Signature Prep Kit* ist hier als Beispiel zu nennen, welches kurze Fragmente amplifiziert. Durch eine entsprechende Anpassung und Optimierung der *monSTR*-

Primer und einer damit einhergehenden Reduzierung der Amplikongrößen, wäre eine erfolgreiche Amplifikation und Sequenzierung herausfordernder Spuren möglich.

Seit einigen Jahren steigt das spurenkundliche Interesse an molekulargenetischen Alternativen zur Bestimmung der Spurenart bzw. zur Dekonvolution von Spurenmischungen mittels der Analyse ausgewählter gewebe- bzw. körperflüssigkeitsspezifisch exprimierter *messenger* RNAs (mRNAs) kontinuierlich. Der Einsatz von mRNAs bietet verschiedene Vorteile gegenüber klassischen chemischen oder immunchromatographischen Vortests: die Methode ist sensitiver, spezifischer, anwendbar auf eine größere Anzahl verschiedener Matrices (v.a. Vaginalsekret) und erlaubt zugleich die simultane Analyse von RNA und DNA aus einer einzigen Probe, wodurch Arbeitsabläufe im Labor beschleunigt und wertvolles Probematerial eingespart werden können. Speziell in Spurenfällen, in denen konventionelle Vortestverfahren an ihre Grenzen stoßen, wie z.B. bei Fällen mit latenten und durch Umweltexposition degradierten Antragungen sowie insbesondere Mischspuren kann die Analyse von mRNAs noch zu aussagekräftigen Ergebnissen führen. Es sind bereits zahlreiche Ziel-mRNAs, die gewebespezifisch oder vornehmlich in einem Gewebetyp/ Zellart exprimiert werden, identifiziert worden (Fleming and Harbison, 2010; Juusola and Ballantyne, 2007; Lindenbergh et al., 2013, 2012), die bereits als Multiplex-Assays mit nachfolgender quantitativer PCR (qPCR) oder CE im Routineeinsatz Anwendung finden (van den Berge et al., 2016). Ferner wurden für diesen Zweck bereits optimierte DNA/ RNA-Co-Extraktionsverfahren etabliert (Kulstein et al., 2016). Im Jahre 2015 entwickelten Hanson et al. den Prototypen eines RNA-Kits mit insgesamt 34 Markern zur Gewebedifferenzierung für die *Illumina* MiSeq Plattform und demonstrierten somit die prinzipielle Durchführbarkeit der RNA-Sequenzierung ausgewählter Marker im forensischen Kontext (Hanson et al., 2015). Dieser Befund wurde inzwischen durch erste internationale Ringversuche bestätigt (Ingold et al., 2018). MPS besitzt gegenüber der CE und der Genexpressionsanalyse mittels qPCR zusätzlich den Vorteil, dass die sequenzbasierte Personenidentifizierung und die RNA-basierte Gewebecharakterisierung prinzipiell in einem kombinierten Verfahren erfolgen können. So können individualisierende SNPs aus RNA-Sequenzen verwendet werden, um Komponenten einer Mischspur verschiedenen in Frage kommenden Donoren zuzuordnen (Ingold et al., 2017). Auch für diese Anwendung erbrachte ein erster internationaler Ringversuch vielversprechende Ergebnisse (Ingold et al., 2020). Zudem wird die Etablierung eines MPS-Verfahrens zur simultanen DNA-basierten Personenidentifizierung und RNA-basierten Gewebedifferenzierung von

forensischen Proben angestrebt. Prinzipiell ist diese Methode zur parallelen Detektion von DNA und RNA-Markern geeignet, wie schon 2015 in einer *proof-of-principle*-Studie zur simultanen Sequenzierung von 10 STR-Systemen und 14 mRNA-Markern auf dem *PGM*-System gezeigt werden konnte (Zubakov et al., 2015). Eine weiterführende Forschungsarbeit beschäftigt sich aus diesem Grund zum gegenwärtigen Zeitpunkt mit einer modularen Kombination des *monSTR*-Panels und mRNA-Markern, um zusätzlich forensisch relevante Gewebe und Körperflüssigkeiten zu identifizieren. Als Grundlage dienten hierbei die mRNA-Marker aus den Publikationen von Hanson et al. (2015) und Zubakov et al. (2015) sowie die Marker des Kapillarelektrophoreseverfahrens von van den Berge et al. (2016). Interne Vorstudien haben bereits belegt, dass die PCR-Produkte aus beiden Ansätzen problemlos parallel innerhalb einer *Library-Preparation* bearbeitet und kombiniert auf dem *MiSeq* sequenziert werden konnten.

Mit der Einführung des *minION* stellte die Firma Oxford Nanopore Technologies eine neue Generation hochmobiler und USB-betriebener *Long-Read*-Sequenzierplattformen vor, die Nukleinsäuresequenzierungen in Echtzeit, ohne vorgeschaltete Amplifikation und bei vergleichsweise sehr erschwinglichen Anschaffungskosten ermöglicht (Jain et al., 2016; Plesivkova et al., 2019). Im Gegensatz zur Sequenzierung durch Synthese (z.B. Illumina-Sequenzierung) wird diese *Long-Read*-Sequenzierungstechnologie aufgrund ihrer Neuartigkeit auch als „Sequenzierung der dritten Generation“ eingestuft. Durch die langen Leselängen können repetitive Strukturen besser bewältigt und Variationen innerhalb der flankierenden Region erfasst werden. Gegenwärtig bestehen jedoch weiterhin Einschränkungen für die STR-Sequenzierung bedingt durch das Auftreten von Homopolymeren und die zu geringe Genauigkeit beim *Base-Calling* (Cornelis et al., 2018). Erste Versuche, die Sequenzierung mittels Nanopore für die STR-Profilerstellung zu verwenden, wurden vor kurzem publiziert (Asogawa et al., 2020; Cornelis et al., 2018). Zukünftige Forschungsarbeiten an *monSTR* könnten auf die Adaption der *Library-Preparation* für ein *Long-Read*-Sequenziergerät wie dem *minION* abzielen, da mittels der Nanopore-Technologie auch kurze Amplikons sequenziert werden können. Erste Erfahrungen konnten hierzu bereits gewonnen werden, da diese neuartige Technologie im hiesigen Labor für die SARS-CoV-2 Vollgenomsequenzierung eingesetzt wird.

5 Zusammenfassung

Für die forensische Molekulargenetik hat sich *Massively Parallel Sequencing* (MPS) bereits als hilfreiches Werkzeug erwiesen. Mittels dieser Technologie können nicht nur zahlreiche forensisch-relevante Marker innerhalb eines Analysenlaufs, sondern zugleich für mehrere Proben parallel untersucht werden. Dies führt insgesamt zu geringeren Kosten pro analysierter Probe im Vergleich zur Kombination herkömmlich verwendeter Routineverfahren. Die forensische DNA-Typisierung auf Basis der Kapillarelektrophorese (engl.: *Capillary Electrophoresis*, CE) ist gegenwärtig die meistgenutzt Methode in forensisch-genetischen Routinelaboren, obwohl sie hinsichtlich Auflösung, Skalierbarkeit und Probendurchsatz limitiert ist. Im Vergleich zur CE bietet MPS eine wesentlich höhere genetische Trennschärfe, da Allele nicht nur anhand der Anzahl ihrer Wiederholungseinheiten, sondern zusätzlich aufgrund der basengenauen Nukleotidsequenz der STR-Abschnitte identifiziert werden. Dadurch können nicht zuletzt komplexe Mischprofile erfolgversprechender dekonvolviert werden.

Das forensische MPS-Multiplexpanel *monSTR* wurde für die gezielte, hochauflösende STR-Typisierung auf der Illumina *MiSeq*-Plattform entwickelt und amplifiziert die wichtigsten STR-Kernloci des *European Standard Sets* (ESS) sowie das hochinformative System SE33. Darüber hinaus stellt es eine kostengünstige und dennoch wettbewerbsfähige Alternative zu kommerziell erhältlichen MPS-Panels dar. Für die *Library-Preparation* wurde eine amplikonbasierte Strategie gewählt, die sich auf die Sequenzierung spezifischer Zielregionen (Amplikons) konzentriert.

Ziel der in der vorliegenden Dissertation beschriebenen Arbeiten war die umfassende Validierung von *monSTR* auf Grundlage der forensisch wesentlichen Richtlinien der *Scientific Working Group on DNA Analysis Methods* (SWGDAM). Für diesen Zweck wurden experimentelle Studien zu insgesamt sechs Qualitätsaspekten (Konkordanz, Reproduzierbarkeit und Wiederholbarkeit, Sensitivität, die Analyse unterschiedlicher Mischungsverhältnisse, die Untersuchung forensischer Ringversuchsproben und Speziesspezifität) geplant und durchgeführt, um die Robustheit und analytische Grenzen des Multiplexpanels umfassend zu bewerten. Die Leistungsfähigkeit wurde anhand der statistischen Parameter *Allele Coverage*

Ratio (ACR), *Coverage coefficient of variation* (CV), *Calling Ratio* (CR), *Signal to Noise Ratio (SN*) und *Minor Component Allele Frequency* (MAF) bewertet.

Die Ergebnisse verdeutlichen, dass die genetischen Profile der untersuchten Referenz- und Ringversuchsproben eine nahezu vollständige Konkordanz zu den mittels CE bestimmten Genotypen aufweisen. Innerhalb der *Intra-* und *Inter-Run* Experimente wurde eine hohe Präzision sowohl zwischen unterschiedlichen MPS-Läufen als auch innerhalb derselben Sequenzierläufe festgestellt. Vollständige STR-Profile, eine ausgeglichene Interlokus-Balance und ein gutes CR-Verhältnis konnten mit DNA-Inputmengen ab 62,5 pg DNA im Rahmen der Sensitivitätsstudie erhalten werden. Gleichzeitig wurden insgesamt 76,4 % der Allele bei einer DNA-Inputmenge von 7,8 pg erfolgreich typisiert. Die Analyse unterschiedlicher DNA-Mischungsverhältnisse zeigte, dass 94,4 % der Allele der Nebenkomponente innerhalb der 1:49-Mischung korrekt zugeordnet wurden. Zudem deuten die Ergebnisse darauf hin, dass der tatsächliche Anteil der Nebenkomponente auf der Grundlage der *minor allele frequency* genau berechnet werden konnte. Bei der Prüfung auf Speziesspezifität der Methode resultierten keine signifikanten Signale bei Untersuchung der DNA verschiedener verbreiteter Tierarten (Hund, Katze und Pferd). Die zuvor beschriebenen Validierungsergebnisse veranschaulichen, dass das forensische Multiplexpanel *monSTR* ein valides Analyseverfahren für die forensische STR-Genotypisierung mittels *Massive Parallel Sequencing* darstellt.

Begründet durch die im Rahmen dieser Validierung aufgezeigten Grenzen der Methode, die zukünftig einige Modifikationen erforderlich machen, sollte der Arbeitsablauf für die tägliche forensische Laborroutine in nachfolgenden Forschungsarbeiten weiter optimiert werden. Gleichzeitig sollen neue Sequenzierplattformen, wie beispielsweise die *Long-Read-*Sequenziertechnologie mittels Nanopore, auf die *Library-Preparation* mittels *monSTR* adaptiert werden.

6 Literaturverzeichnis

Alaeddini, R., Walsh, S.J., Abbas, A., 2010. Molecular studies of time- and environment-dependent effects on bone DNA survival. Australian Journal of Forensic Sciences 42, 211–220. https://doi.org/10.1080/00450611003758264

Alonso, A., Müller, P., Roewer, L., Willuweit, S., Budowle, B., Parson, W., 2017. European survey on forensic applications of massively parallel sequencing. Forensic Sci Int Genet 29, e23–e25. https://doi.org/10.1016/j.fsigen.2017.04.017

Amorim, A., Pereira, L., 2005. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. Forensic Sci. Int. 150, 17–21. https://doi.org/10.1016/j.forsciint.2004.06.018

Andréasson, H., Asp, A., Alderborn, A., Gyllensten, U., Allen, M., 2002. Mitochondrial sequence analysis for forensic identification using pyrosequencing technology. BioTechniques 32, 124–126, 128, 130–133.

Ansorge, W.J., 2009. Next-generation DNA sequencing techniques. N Biotechnol 25, 195–203. https://doi.org/10.1016/j.nbt.2008.12.009

Asogawa, M., Ohno, A., Nakagawa, S., Ochiai, E., Katahira, Y., Sudo, M., Osawa, M., Sugisawa, M., Imanishi, T., 2020. Human short tandem repeat identification using a nanopore-based DNA sequencer: a pilot study. J Hum Genet 65, 21–24. https://doi.org/10.1038/s10038-019-0688-z

Aziz, N., Zhao, Q., Bry, L., Driscoll, D.K., Funke, B., Gibson, J.S., Grody, W.W., Hegde, M.R., Hoeltge, G.A., Leonard, D.G.B., Merker, J.D., Nagarajan, R., Palicki, L.A., Robetorye, R.S., Schrijver, I., Weck, K.E., Voelkerding, K.V., 2015. College of American Pathologists' Laboratory Standards for Next-Generation Sequencing Clinical Tests. Archives of Pathology & Laboratory Medicine 139, 481–493. https://doi.org/10.5858/arpa.2014-0250-CP

Biesecker, L.G., Burke, W., Kohane, I., Plon, S.E., Zimmern, R., 2012. Next Generation Sequencing in the Clinic: Are we Ready? Nat Rev Genet 13, 818–824. https://doi.org/10.1038/nrg3357

Bodner, M., Bastisch, I., Butler, J.M., Fimmers, R., Gill, P., Gusmão, L., Morling, N., Phillips, C., Prinz, M., Schneider, P.M., Parson, W., 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 Science International: Genetics 24, 97–102. https://doi.org/10.1016/j.fsigen.2016.06.008

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A., Caporaso, J.G., 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods 10, 57–59. https://doi.org/10.1038/nmeth.2276

Bornman, D.M., Hester, M.E., Schuetter, J.M., Kasoji, M.D., Minard-Smith, A., Barden, C.A., Nelson, S.C., Godbold, G.D., Baker, C.H., Yang, B., Walther, J.E., Tornes, I.E., Yan, P.S., Rodriguez, B., Bundschuh, R., Dickens, M.L., Young, B.A., Faith, S.A., 2012. Short-read, high-throughput sequencing technology for STR genotyping. Biotech Rapid Dispatches 2012, 1–6.

Børsting, C., Morling, N., 2015. Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet 18, 78–89. https://doi.org/10.1016/j.fsigen.2015.02.002

Borsuk, L.A., Gettings, K.B., Steffen, C.R., Kiesler, K.M., Vallone, P.M., 2018. Sequence-based US population data for the SE33 locus. ELECTROPHORESIS 39, 2694–2701. https://doi.org/10.1002/elps.201800091

Bovo, D., Rugge, M., Shiao, Y.H., 1999. Origin of spurious multiple bands in the amplification of microsatellite sequences. Mol Pathol 52, 50–51. https://doi.org/10.1136/mp.52.1.50

Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., Rolf, B., 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am. J. Hum. Genet. 62, 1408–1415. https://doi.org/10.1086/301869

Brookes, C., Bright, J.-A., Harbison, S., Buckleton, J., 2012. Characterising stutter in forensic STR multiplexes. Forensic Science International: Genetics 6, 58–63. https://doi.org/10.1016/j.fsigen.2011.02.001

Bruijns, B., Tiggelaar, R., Gardeniers, H., 2018. Massively parallel sequencing techniques for forensics: A review. ELECTROPHORESIS 39, 2642–2654. https://doi.org/10.1002/elps.201800082

Budowle, B., Onorato, A.J., Callaghan, T.F., Della Manna, A., Gross, A.M., Guerrieri, R.A., Luttman, J.C., McClure, D.L., 2009. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. J. Forensic Sci. 54, 810–821. https://doi.org/10.1111/j.1556-4029.2009.01046.x

Chamberlain, J.S., Gibbs, R.A., Rainer, J.E., Nguyen, P.N., Thomas, C., 1988. Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. Nucleic Acids Research 16, 11141–11156. https://doi.org/10.1093/nar/16.23.11141

Churchill, J.D., Schmedes, S.E., King, J.L., Budowle, B., 2016. Evaluation of the Illumina ® Beta Version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling. Forensic Science International: Genetics 20, 20–29. https://doi.org/10.1016/j.fsigen.2015.09.009

Coble, M.D., Butler, J.M., 2005. Characterization of new miniSTR loci to aid analysis of degraded DNA. J Forensic Sci 50, 43–53.

Cornelis, S., Willems, S., Van Neste, C., Tytgat, O., Weymaere, J., Vander Plaetsen, A.-S., Deforce, D., Van Nieuwerburgh, F., 2018. Forensic STR profiling using Oxford Nanopore Technologies' MinION sequencer (preprint). Genomics. https://doi.org/10.1101/433151

Dalsgaard, S., Rockenbauer, E., Gelardi, C., Børsting, C., Fordyce, S.L., Morling, N., 2013. Characterization of mutations and sequence variations in complex STR loci by second generation sequencing. Forensic Science International: Genetics Supplement Series 4, e218–e219. https://doi.org/10.1016/j.fsigss.2013.10.112

DeAngelis, M.M., Wang, D.G., Hawkins, T.L., 1995. Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Res 23, 4742–4743.

Del Fabbro, C., Scalabrin, S., Morgante, M., Giorgi, F.M., 2013. An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. PLoS ONE 8, e85024. https://doi.org/10.1371/journal.pone.0085024

Divne, A.-M., Edlund, H., Allen, M., 2010. Forensic analysis of autosomal STR markers using Pyrosequencing. Forensic Sci Int Genet 4, 122–129. https://doi.org/10.1016/j.fsigen.2009.07.004

Edgar, R.C., Flyvbjerg, H., 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31, 3476–3482. https://doi.org/10.1093/bioinformatics/btv401

Edwards, A., Civitello, A., Hammond, H.A., Caskey, C.T., 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am. J. Hum. Genet. 49, 746–756.

Endrullat, C., Glökler, J., Franke, P., Frohme, M., 2016a. Standardization and quality management in next-generation sequencing. Applied & Translational Genomics 10, 2–9. https://doi.org/10.1016/j.atg.2016.06.001

Endrullat, C., Glökler, J., Franke, P., Frohme, M., 2016b. Standardization and quality management in next-generation sequencing. Applied & Translational Genomics 10, 2–9. https://doi.org/10.1016/j.atg.2016.06.001

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wikström, C., Wold, S., 2008. Design of experiments: principles and applications. Umetrics, Umeå.

Fleming, R.I., Harbison, S., 2010. The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. Forensic Sci Int Genet 4, 244–256. https://doi.org/10.1016/j.fsigen.2009.10.006

Gadberry, M.D., Malcomber, S.T., Doust, A.N., Kellogg, E.A., 2005. Primaclade--a flexible tool to find conserved PCR primers across multiple species. Bioinformatics 21, 1263–1264. https://doi.org/10.1093/bioinformatics/bti134

Ganschow, S., Silvery, J., Kalinowski, J., Tiemann, C., 2018. toaSTR: A web application for forensic STR genotyping by massively parallel sequencing. Forensic Science International: Genetics 37, 21–28. https://doi.org/10.1016/j.fsigen.2018.07.006

Ganschow, S., Silvery, J., Tiemann, C., 2019. Development of a multiplex forensic identity panel for massively parallel sequencing and its systematic optimization using design of experiments. Forensic Science International: Genetics 39, 32–43. https://doi.org/10.1016/j.fsigen.2018.11.023

Gargis, A.S., Kalman, L., Berry, M.W., Bick, D.P., Dimmock, D.P., Hambuch, T., Lu, F., Lyon, E., Voelkerding, K.V., Zehnbauer, B.A., Agarwala, R., Bennett, S.F., Chen, B., Chin, E.L.H., Compton, J.G., Das, S., Farkas, D.H., Ferber, M.J., Funke, B.H., Furtado, M.R., Ganova-Raeva, L.M., Geigenmüller, U., Gunselman, S.J., Hegde, M.R., Johnson, P.L.F., Kasarskis, A., Kulkarni, S., Lenk, T., Liu, C.S.J., Manion, M., Manolio, T.A., Mardis, E.R., Merker, J.D., Rajeevan, M.S., Reese, M.G., Rehm, H.L., Simen, B.B., Yeakley, J.M., Zook, J.M., Lubin, I.M., 2012. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol 30, 1033–1036. https://doi.org/10.1038/nbt.2403

Gettings, K.B., Aponte, R.A., Vallone, P.M., Butler, J.M., 2015. STR allele sequence variation: Current knowledge and future issues. Forensic Science International: Genetics 18, 118–130. https://doi.org/10.1016/j.fsigen.2015.06.005

Gettings, K.B., Borsuk, L.A., Ballard, D., Bodner, M., Budowle, B., Devesse, L., King, J., Parson, W., Phillips, C., Vallone, P.M., 2017. STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci. Forensic Science International: Genetics 31, 111–117. https://doi.org/10.1016/j.fsigen.2017.08.017

Gettings, K.B., Borsuk, L.A., Steffen, C.R., Kiesler, K.M., Vallone, P.M., 2018. Sequence-based U.S. population data for 27 autosomal STR loci. Forensic Science International: Genetics 37, 106–115. https://doi.org/10.1016/j.fsigen.2018.07.013

Gettings, K.B., Kiesler, K.M., Faith, S.A., Montano, E., Baker, C.H., Young, B.A., Guerrieri, R.A., Vallone, P.M., 2016. Sequence variation of 22 autosomal STR loci detected by next generation sequencing. Forensic Science International: Genetics 21, 15–21. https://doi.org/10.1016/j.fsigen.2015.11.005

Gill, P., 2001. Application of low copy number DNA profiling. Croat Med J 42, 229–232.

Gill, P., Brinkmann, B., d'Aloja, E., Andersen, J., Bar, W., Carracedo, A., Dupuy, B., Eriksen, B., Jangblad, M., Johnsson, V., Kloosterman, A.D., Lincoln, P., Morling, N., Rand, S., Sabatier, M., Scheithauer, R., Schneider, P., Vide, M.C., 1997. Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. Forensic Science International 87, 185–192. https://doi.org/10.1016/S0379-0738(97)00111-4

Gill, P., Evett, I., 1995. Population genetics of short tandem repeat (STR) loci. Genetica 96, 69–87. https://doi.org/10.1007/BF01441153

Gill, P., Haned, H., Bleka, O., Hansson, O., Dørum, G., Egeland, T., 2015. Genotyping and interpretation of STR-DNA: Low-template, mixtures and database matches—Twenty years of research and development. Forensic Science International: Genetics 18, 100–117. https://doi.org/10.1016/j.fsigen.2015.03.014

Glenn, T.C., 2011. Field guide to next-generation DNA sequencers. Mol Ecol Resour 11, 759–769. https://doi.org/10.1111/j.1755-0998.2011.03024.x

Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 17, 333–351. https://doi.org/10.1038/nrg.2016.49

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 Science International: Genetics 31, 135–148. https://doi.org/10.1016/j.fsigen.2017.09.003

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 Science International: Genetics 23, 111–120. https://doi.org/10.1016/j.fsigen.2016.04.004

Gymrek, M., Golan, D., Rosset, S., Erlich, Y., 2012. lobSTR: A short tandem repeat profiler for personal genomes. Genome Res. 22, 1154–1162. https://doi.org/10.1101/gr.135780.111

Hanson, E., Ingold, S., Haas, C., Ballantyne, J., 2015. Targeted multiplexed next generation RNA sequencing assay for tissue source determination of forensic samples. Forensic Science International: Genetics Supplement Series 5, e441–e443. https://doi.org/10.1016/j.fsigss.2015.09.175

Hares, D.R., 2015. Selection and implementation of expanded CODIS core loci in the United States. Forensic Science International: Genetics 17, 33–34. https://doi.org/10.1016/j.fsigen.2015.03.006

Hares, D.R., 2012. Expanding the CODIS core loci in the United States. Forensic Science International: Genetics 6, e52–e54. https://doi.org/10.1016/j.fsigen.2011.04.012

Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R., Ordoukhanian, P., 2014. Library construction for next-generation sequencing: Overviews and challenges. Biotechniques 56, 61-passim. https://doi.org/10.2144/000114133

Henegariu, O., Heerema, N.A., Dlouhy, S.R., Vance, G.H., Vogt, P.H., 1997. Multiplex PCR: critical parameters and step-by-step protocol. Biotechniques 23, 504–511.

Hennessy, L.K., Mehendale, N., Chear, K., Jovanovich, S., Williams, S., Park, C., Gangano, S., 2014. Developmental validation of the GlobalFiler® express kit, a 24-marker STR assay, on the RapidHIT® System. Forensic Science International: Genetics 13, 247–258. https://doi.org/10.1016/j.fsigen.2014.08.011

Hodgkinson, A., Eyre-Walker, A., 2010. Human triallelic sites: evidence for a new mutational mechanism? Genetics 184, 233–241. https://doi.org/10.1534/genetics.109.110510

Hoogenboom, J., van der Gaag, K.J., de Leeuw, R.H., Sijen, T., de Knijff, P., Laros, J.F.J., 2017. FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise. Forensic Science International: Genetics 27, 27–40. https://doi.org/10.1016/j.fsigen.2016.11.007

Hussing, C., Huber, C., Bytyci, R., Mogensen, H.S., Morling, N., Børsting, C., 2018. Sequencing of 231 forensic genetic markers using the MiSeq FGx™ forensic genomics system – an evaluation of the assay and software. Forensic Sciences Research 3, 111–123. https://doi.org/10.1080/20961790.2018.1446672

Huszar, T.I., Gettings, K.B., Vallone, P.M., 2021. An Introductory Overview of Open-Source and Commercial Software Options for the Analysis of Forensic Sequencing Data. Genes 12, 1739. https://doi.org/10.3390/genes12111739

Ingold, S., Dørum, G., Hanson, E., Ballard, D., Berti, A., Gettings, K.B., Giangasparo, F., Kampmann, M.-L., Laurent, F.-X., Morling, N., Parson, W., Steffen, C.R., Ulus, A., van den Berge, M., van der Gaag, K.J., Verdoliva, V., Xavier, C., Ballantyne, J., Haas, C., 2020. Body fluid identification and assignment to donors using a targeted mRNA massively parallel sequencing approach - results of a second EUROFORGEN / EDNAP collaborative exercise. Forensic Sci Int Genet 45, 102208. https://doi.org/10.1016/j.fsigen.2019.102208

Ingold, S., Dørum, G., Hanson, E., Berti, A., Branicki, W., Brito, P., Elsmore, P., Gettings, K.B., Giangasparo, F., Gross, T.E., Hansen, S., Hanssen, E.N., Kampmann, M.-L., Kayser, M., Laurent, F.-X., Morling, N., Mosquera-Miguel, A., Parson, W., Phillips, C., Porto, M.J., Pośpiech, E., Roeder, A.D., Schneider, P.M., Schulze Johann, K., Steffen, C.R., Syndercombe-Court, D., Trautmann, M., van den Berge, M., van der Gaag, K.J., Vannier, J., Verdoliva, V., Vidaki, A., Xavier, C., Ballantyne, J., Haas, C., 2018. Body fluid identification using a targeted mRNA massively parallel sequencing approach – results of a EUROFORGEN/EDNAP collaborative exercise. Forensic Science International: Genetics 34, 105–115. https://doi.org/10.1016/j.fsigen.2018.01.002

Ingold, S., Haas, C., Dørum, G., Hanson, E., Ballantyne, J., 2017. Association of a body fluid with a DNA profile by targeted RNA/DNA deep sequencing. Forensic Science International: Genetics Supplement Series 6, e112–e113. https://doi.org/10.1016/j.fsigss.2017.09.037

Iyengar, A., Hadi, S., 2014. Use of non-human DNA analysis in forensic science: a mini review. Med Sci Law 54, 41–50. https://doi.org/10.1177/0025802413487522

Jäger, A.C., Alvarez, M.L., Davis, C.P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S.J.K., Varlaro, J., Stephens, K.M., Holt, C.L., 2017. Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. Forensic Science International: Genetics 28, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Jain, M., Olsen, H.E., Paten, B., Akeson, M., 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biology 17, 239. https://doi.org/10.1186/s13059-016-1103-0

Jeffreys, A.J., Wilson, V., Thein, S.L., 1985. Hypervariable “minisatellite” regions in human DNA. Nature 314, 67–73.

Ju, J., Kim, D.H., Bi, L., Meng, Q., Bai, X., Li, Z., Li, X., Marma, M.S., Shi, S., Wu, J., Edwards, J.R., Romu, A., Turro, N.J., 2006. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. Proc Natl Acad Sci U S A 103, 19635–19640. https://doi.org/10.1073/pnas.0609513103

Juusola, J., Ballantyne, J., 2007. mRNA profiling for body fluid identification by multiplex quantitative RT-PCR. J. Forensic Sci. 52, 1252–1262. https://doi.org/10.1111/j.1556-4029.2007.00550.x

Kchouk, M., Gibrat, J.F., Elloumi, M., 2017. Generations of Sequencing Technologies: From First to Next Generation. Biol Med (Aligarh) 09. https://doi.org/10.4172/0974-8369.1000395

Kimpton, C.P., Gill, P., Walton, A., Urquhart, A., Millican, E.S., Adams, M., 1993. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Methods Appl. 3, 13–22.

Klintschar, M., Wiegand, P., 2003. Polymerase slippage in relation to the uniformity of tetrameric repeat stretches. Forensic Sci. Int. 135, 163–166.

Köcher, S., Müller, P., Berger, B., Bodner, M., Parson, W., Roewer, L., Willuweit, S., 2018. Inter-laboratory validation study of the ForenSeq™ DNA Signature Prep Kit. Forensic Science International: Genetics 36, 77–85. https://doi.org/10.1016/j.fsigen.2018.05.007

Kulstein, G., Marienfeld, R., Miltner, E., Wiegand, P., 2016. Automation of DNA and miRNA co-extraction for miRNA-based identification of human body fluids and tissues. Electrophoresis 37, 2742–2750. https://doi.org/10.1002/elps.201600365

LANDSTEINER, K., 1900. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. Zentralbl. Bakteriol. 27, 357–362.

Lászik, A., Sótonyi, P., Rand, S., Hohoff, C., 2001. Frequency data for the STR locus ACTBP2 (SE33) in eight populations. Int J Leg Med 115, 94–96. https://doi.org/10.1007/s004140100209

Laurent, F.X., Ausset, L., Clot, M., Jullien, S., Chantrel, Y., Hollard, C., Pene, L., 2017. Automation of library preparation using Illumina ForenSeq kit for routine sequencing of casework samples. Forensic Science International: Genetics Supplement Series 6, e415–e417. https://doi.org/10.1016/j.fsigss.2017.09.156

Lee, J.C.-I., Tseng, B., Chang, L.-K., Linacre, A., 2017. SEQ Mapper: A DNA sequence searching tool for massively parallel sequencing data. Forensic Science International: Genetics 26, 66–69. https://doi.org/10.1016/j.fsigen.2016.10.006

Lindenbergh, A., de Pagter, M., Ramdayal, G., Visser, M., Zubakov, D., Kayser, M., Sijen, T., 2012. A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. Forensic Science International: Genetics 6, 565–577. https://doi.org/10.1016/j.fsigen.2012.01.009

Lindenbergh, A., Maaskant, P., Sijen, T., 2013. Implementation of RNA profiling in forensic casework. Forensic Science International: Genetics 7, 159–166. https://doi.org/10.1016/j.fsigen.2012.09.003

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012, 251364. https://doi.org/10.1155/2012/251364

Locard, E., 1930. The Analysis of Dust Traces. Part I. The American Journal of Police Science 1, 276–298. https://doi.org/10.2307/1147154

Ludeman, M.J., Zhong, C., Mulero, J.J., Lagacé, R.E., Hennessy, L.K., Short, M.L., Wang, D.Y., 2018. Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. Int J Legal Med 132, 1555–1573. https://doi.org/10.1007/s00414-018-1817-5

Ma, Y., Kuang, J.-Z., Nie, T.-G., Zhu, W., Yang, Z., 2016. Next generation sequencing: Improved resolution for paternal/maternal duos analysis. Forensic Sci Int Genet 24, 83–85. https://doi.org/10.1016/j.fsigen.2016.05.015

Mardis, E.R., 2011. A decade's perspective on DNA sequencing technology. Nature 470, 198–203. https://doi.org/10.1038/nature09796

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, Chun Heen, Ho, Chun He, Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380. https://doi.org/10.1038/nature03959

McCord, B.R., Gauthier, Q., Cho, S., Roig, M.N., Gibson-Daw, G.C., Young, B., Taglia, F., Zapico, S.C., Mariot, R.F., Lee, S.B., Duncan, G., 2019. Forensic DNA Analysis. Anal Chem 91, 673–688. https://doi.org/10.1021/acs.analchem.8b05318

McLaren, R.S., Bourdeau-Heller, J., Patel, J., Thompson, J.M., Pagram, J., Loake, T., Beesley, D., Pirttimaa, M., Hill, C.R., Duewer, D.L., Kline, M.C., Butler, J.M., Storts, D.R., 2014. Developmental validation of the PowerPlex® ESI 16/17 Fast and PowerPlex® ESX 16/17 Fast Systems. Forensic Science International: Genetics 13, 195–205. https://doi.org/10.1016/j.fsigen.2014.08.004

Menin, J.F., Nichols, N., 2013. Multiplex PCR using Q5® High-Fidelity DNA Polymerase [WWW Document]. URL https://www.semanticscholar.org/paper/Multiplex-PCR-using-Q5%C2%AE-High-Fidelity-DNA-Menin-Nichols/637c0262c1bff0de8a5185963fb4c93d820d1bc6 (accessed 1.20.22).

Müller, P., Alonso, A., Barrio, P.A., Berger, B., Bodner, M., Martin, P., Parson, W., 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 Science International: Genetics 36, 95–103. https://doi.org/10.1016/j.fsigen.2018.06.016

Mullis, K.B., Faloona, F.A., 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155, 335–350.

National Institute of Standards & Technology, 2018. Certificate of Analysis Standard Reference Material 2391c.

New York Department of Health, 2015. "Next Generation" Sequencing (NGS) guidelines for somatic genetic variant detection.

Novroski, N.M.M., King, J.L., Churchill, J.D., Seah, L.H., Budowle, B., 2016. Characterization of genetic sequence variation of 58 STR loci in four major population groups. Forensic Science International: Genetics 25, 214–226. https://doi.org/10.1016/j.fsigen.2016.09.007

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, K.L., Hill, C., Oyerly, K., Duke, W., Katzilierakis, L., Ensenberger, M.G., Bourdeau, J.M., Sprecher, C.J., Krenke, B., Storts, D.R., 2014. Developmental validation of the PowerPlex® Fusion System for analysis of casework and reference samples: A 24-locus multiplex for new database standards. Forensic Science International: Genetics 12, 69–76. https://doi.org/10.1016/j.fsigen.2014.04.013

Parson, W., Ballard, D., Budowle, B., Butler, J.M., Gettings, K.B., Gill, P., Gusmão, L., Hares, D.R., Irwin, J.A., King, J.L., Knijff, P. de, Morling, N., Prinz, M., Schneider, P.M., Neste, C.V., 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 Science International: Genetics 22, 54–63. https://doi.org/10.1016/j.fsigen.2016.01.009

Peters, F.T., Drummer, O.H., Musshoff, F., 2007. Validation of new methods. Forensic Science International, Research in Forensic Medicine and Forensic Sciences: Basic Principles and Application in Practice 165, 216–224. https://doi.org/10.1016/j.forsciint.2006.05.021

Phillips, C., Gettings, K.B., King, J.L., 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 Science International: Genetics. https://doi.org/10.1016/j.fsigen.2018.02.017

Plesivkova, D., Richards, R., Harbison, S., 2019. A review of the potential of the MinION™ single-molecule sequencing system for forensic applications. Wiley Interdisciplinary Reviews: Forensic Science 1, e1323. https://doi.org/10.1002/wfs2.1323

Polymeropoulos, M.H., Rath, D.S., Xiao, H., Merril, C.R., 1992. Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2). Nucleic Acids Res 20, 1432. https://doi.org/10.1093/nar/20.6.1432

Qu, W., Zhou, Y., Zhang, Y., Lu, Y., Wang, X., Zhao, D., Yang, Y., Zhang, C., 2012. MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity. Nucleic Acids Res 40, W205–W208. https://doi.org/10.1093/nar/gks552

Riman, S., Iyer, H., Borsuk, L., Gettings, K., Vallone, P., 2017. Investigating the effects of different library preparation protocols on STR sequencing. Forensic Science International: Genetics Supplement Series 6, e418–e420. https://doi.org/10.1016/j.fsigss.2017.09.155

Rockenbauer, E., Hansen, S., Mikkelsen, M., Børsting, C., Morling, N., 2014. Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing. Forensic Sci Int Genet 8, 68–72. https://doi.org/10.1016/j.fsigen.2013.06.011

Roewer, L., Arnemann, J., K Spurr, N., H Grzeschik, K., Epplen, J., 1992. Roewer L, Arnemann J, Spurr NK, Grzeschik KH, Epplen JT. Simple repeat sequences on the human Y-Chromosome are equally polymorphic as their autosomal counterparts. Hum Genet 89: 389-394. Human genetics 89, 389–94. https://doi.org/10.1007/BF00194309

Romsos, E.L., Vallone, P.M., 2015. Rapid PCR of STR markers: Applications to human identification. Forensic Sci Int Genet 18, 90–99. https://doi.org/10.1016/j.fsigen.2015.04.008

Ronaghi, M., 2001. Pyrosequencing sheds light on DNA sequencing. Genome Res. 11, 3–11.

Ross, M.G., Russ, C., Costello, M., Hollinger, A., Lennon, N.J., Hegarty, R., Nusbaum, C., Jaffe, D.B., 2013. Characterizing and measuring bias in sequence data. Genome Biology 14, R51. https://doi.org/10.1186/gb-2013-14-5-r51

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467.

Schlötterer, C., Tautz, D., 1992. Slippage synthesis of simple sequence DNA. Nucl Acids Res 20, 211–215. https://doi.org/10.1093/nar/20.2.211

Schneider, P.M., 2009. Expansion of the European Standard Set of DNA Database Loci--the Current Situation 2.

Schuster, S.C., 2008. Next-generation sequencing transforms today's biology. Nat. Methods 5, 16–18. https://doi.org/10.1038/nmeth1156

Scientific Working Group on DNA Analysis Methods, 2016. Validation Guidelines for DNA Analysis Methods [WWW Document].

Shen, Z., Qu, W., Wang, W., Lu, Y., Wu, Y., Li, Z., Hang, X., Wang, X., Zhao, D., Zhang, C., 2010. MPprimer: a program for reliable multiplex PCR primer design. BMC Bioinformatics 11, 143. https://doi.org/10.1186/1471-2105-11-143

Silvery, J., Ganschow, S., Wiegand, P., Tiemann, C., 2020. Developmental validation of the monSTR identity panel, a forensic STR multiplex assay for massively parallel sequencing. Forensic Science International: Genetics 46, 102236. https://doi.org/10.1016/j.fsigen.2020.102236

Steffen, C.R., Kline, M.C., Duewer, D.L., Vallone, P.M., 2015. Establishing traceability to NIST SRM 2391c: PCR-Based DNA Profiling Standard. Forensic Science International: Genetics Supplement Series 5, e112–e113. https://doi.org/10.1016/j.fsigss.2015.09.045

Strachan, T., Read, A., 2018. Human Molecular Genetics. Garland Science.

Sullivan, K.M., Mannucci, A., Kimpton, C.P., Gill, P., 1993. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. Biotechniques 15, 636–638, 640–641.

Sunyoung Kwon, Seunghyun Park, Byunghan Lee, Sungroh Yoon, 2013. In-depth analysis of interrelation between quality scores and real errors in illumina reads, in: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Presented at the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, Osaka, pp. 635–638. https://doi.org/10.1109/EMBC.2013.6609580

Szibor, R., 2007. X-chromosomal markers: past, present and future. Forensic Sci Int Genet 1, 93–99. https://doi.org/10.1016/j.fsigen.2007.03.003

Tautz, D., 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17, 6463–6471.

Tautz, D., Renz, M., 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12, 4127–4138. https://doi.org/10.1093/nar/12.10.4127

Toms, D., Deardon, R., Ungrin, M., 2017. Climbing the mountain: experimental design for the efficient optimization of stem cell bioprocessing. J Biol Eng 11, 35. https://doi.org/10.1186/s13036-017-0078-z

van den Berge, M., Bhoelai, B., Harteveld, J., Matai, A., Sijen, T., 2016. Advancing forensic RNA typing: On non-target secretions, a nasal mucosa marker, a differential co-extraction protocol and the sensitivity of DNA and RNA profiling. Forensic Science International: Genetics 20, 119–129. https://doi.org/10.1016/j.fsigen.2015.10.011

van der Gaag, K.J., de Leeuw, R.H., Hoogenboom, J., Patel, J., Storts, D.R., Laros, J.F.J., de Knijff, P., 2016. Massively parallel sequencing of short tandem repeats—Population data and mixture analysis results for the PowerSeq™ system. Forensic Science International: Genetics 24, 86–96. https://doi.org/10.1016/j.fsigen.2016.05.016

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C., 2014. Ten years of next-generation sequencing technology. Trends Genet 30, 418–426. https://doi.org/10.1016/j.tig.2014.07.001

Van Neste, C., Vandewoestyne, M., Van Criekinge, W., Deforce, D., Van Nieuwerburgh, F., 2014. My-Forensic-Loci-queries (MyFLq) framework for analysis of forensic STR data generated by massive parallel sequencing. Forensic Sci Int Genet 9, 1–8. https://doi.org/10.1016/j.fsigen.2013.10.012

Viguera, E., Canceill, D., Ehrlich, S.D., 2001. In vitro replication slippage by DNA polymerases from thermophilic organisms. Journal of Molecular Biology 312, 323–333. https://doi.org/10.1006/jmbi.2001.4943

Wallin, J.M., Holt, C.L., Lazaruk, K.D., Nguyen, T.H., Walsh, P.S., 2002. Constructing universal multiplex PCR systems for comparative genotyping. J Forensic Sci 47, 52–65.

Walsh, P.S., Fildes, N.J., Reynolds, R., 1996. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Res 24, 2807–2812. https://doi.org/10.1093/nar/24.14.2807

Wang, D.Y., Green, R.L., Lagacé, R.E., Oldroyd, N.J., Hennessy, L.K., Mulero, J.J., 2012. Identification and secondary structure analysis of a region affecting electrophoretic mobility of the STR locus SE33. Forensic Sci Int Genet 6, 310–316. https://doi.org/10.1016/j.fsigen.2011.06.008

Warshauer, D.H., Churchill, J.D., Novroski, N., King, J.L., Budowle, B., 2015. Novel Y-chromosome Short Tandem Repeat Variants Detected Through the Use of Massively Parallel Sequencing. Genomics, Proteomics & Bioinformatics 13, 250–257. https://doi.org/10.1016/j.gpb.2015.08.001

Warshauer, D.H., Lin, D., Hari, K., Jain, R., Davis, C., LaRue, B., King, J.L., Budowle, B., 2013. STRait Razor: A length-based forensic STR allele-calling tool for use with second generation sequencing data. Forensic Science International: Genetics 7, 409–417. https://doi.org/10.1016/j.fsigen.2013.04.005

Weber, J.L., David, D., Heil, J., Fan, Y., Zhao, C., Marth, G., 2002. Human Diallelic Insertion/Deletion Polymorphisms. Am J Hum Genet 71, 854–862.

Wiegand, P., Budowle, B., Rand, S., Brinkmann, B., 1993. Forensic validation of the STR systems SE 33 and TC 11. Int. J. Legal Med. 105, 315–320.

Wyner, N., Barash, M., McNevin, D., 2020. Forensic Autosomal Short Tandem Repeats and Their Potential Association With Phenotype. Front. Genet. 11, 884. https://doi.org/10.3389/fgene.2020.00884

Yuan, Y., Jiang, F., Hua, S., Du, B., Hao, Y., Ye, L., Liu, J., Feng, K., Huang, X., Yi, X., Wang, W., Yang, L., Mu, F., Liu, C., Liang, Y., 2013. Feasibility study of semiconductor sequencing for noninvasive prenatal detection of fetal aneuploidy. Clin. Chem. 59, 846–849. https://doi.org/10.1373/clinchem.2012.196725

Zaaijer, S., Gordon, A., Speyer, D., Piccone, R., Groen, S.C., Erlich, Y., 2017. Rapid re-identification of human samples using portable DNA sequencing. Elife 6, e27798. https://doi.org/10.7554/eLife.27798

Zeng, X., King, J., Hermanson, S., Patel, J., Storts, D.R., Budowle, B., 2015a. An evaluation of the PowerSeq™ Auto System: A multiplex short tandem repeat marker kit compatible with massively parallel sequencing. Forensic Sci Int Genet 19, 172–179. https://doi.org/10.1016/j.fsigen.2015.07.015

Zeng, X., King, J.L., Stoljarova, M., Warshauer, D.H., LaRue, B.L., Sajantila, A., Patel, J., Storts, D.R., Budowle, B., 2015b. High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing. Forensic Science International: Genetics 16, 38–47. https://doi.org/10.1016/j.fsigen.2014.11.022

Zubakov, D., Kokmeijer, I., Ralf, A., Rajagopalan, N., Calandro, L., Wootton, S., Langit, R., Chang, C., Lagace, R., Kayser, M., 2015. Towards simultaneous individual and tissue identification: A proof-of-principle study on parallel sequencing of STRs, amelogenin, and mRNAs with the Ion Torrent PGM. Forensic Science International: Genetics 17, 122–128. https://doi.org/10.1016/j.fsigen.2015.04.002

Danksagung

Aus Gründen des Datenschutzes in der Veröffentlichungsversion entfernt.

Lebenslauf

Aus Gründen des Datenschutzes in der Veröffentlichungsversion entfernt.