High-resolution forensic DNA typing

Construction, optimization, and validation of a human identity panel used with massively parallel sequencing and the development of a web-based genotyping software

DISSERTATION

presented to the Faculty of Biology of the Bielefeld University
to obtain the degree of

Doctor of Natural Sciences (Dr. rer. nat.)

submitted by
Sebastian Ganschow, M.Sc.

January 2020

The practical work of this thesis has been performed at LABCON-OWL GmbH, Bad Salzuflen from October 2015 to October 2018 and under the supervision of Prof. Dr. Carsten Tiemann.

Primary Reviewers	Prof. Dr. rer. nat. Jörn Kalinowski Microbial Genomics and Biotechnology, Center for Biotechnology (CeBiTec), Bielefeld University	Prof. Dr. rer. nat. Carsten Tiemann LABCON-OWL GmbH, Bad Salzuflen Bielefeld University of Applied Sciences
Secondary Reviewer	Prof. Dr. rer. nat. Alexander Goesmann Bioinformatics & Systems Biology, Gießen University	

Printed on acid-free permanent paper according to ISO 9706

Abstract

The field of forensic DNA typing has advanced rapidly over the past decades. Nowadays, short tandem repeats (STRs) are the markers of choice to identify the donor of biological evidence. This class of genetic variation consists of tandem-like repeated elements that are highly variable in both the number of repeat units and the repeat sequence. Analysis of length polymorphisms at STR loci currently almost exclusively relies on PCR amplification and subsequent fragment sizing using capillary electrophoresis – a "gold standard" with certain limitations as to the resolution of STR alleles and the separation of artificial products. With the ongoing advancement of DNA sequencing, the forensic community is exploring the opportunities of massively parallel sequencing (MPS) for high-resolution forensic DNA typing. MPS enables characterizing biological evidence in previously unimagined detail. Some attractive features of STR sequencing are the increased discrimination power compared to that of electrophoretic sizing and the ability to investigate a wide range of forensic markers in a single assay. Despite these benefits, the application of MPS to routine casework poses new challenges, which are addressed as part of this dissertation.

The monSTR identity panel was designed in response to demand for a medium-sized STR assay on the Illumina MiSeq platform, targeting 21 forensically important markers including the highly discriminative SE33 locus. This thematic complex describes the construction of a custom forensic MPS-STR assay from primer engineering through the optimization of thermocycling conditions. The Design of Experiments methodology pioneered in this context enables an experimentally practical and economically justifiable assay optimization. Statistical modeling revealed valuable insights that helped to understand the characteristics of the monSTR assay. Joint optimization of multiple process parameters resulted in a high-fidelity identity panel, characterized by a well-balanced amplification of STR loci, a high on-target ratio of sequence reads, and reduced formation of stutter products compared to standard PCR conditions. Developmental validation studies according to established forensic guidelines have explored the capabilities and limitations of this novel identity panel. One of the key findings was that monSTR generates complete and reproducible genotypes even with minute amounts of input DNA. Results have also demonstrated that STR alleles of multiple contributors in imbalanced mixed samples can be accurately resolved.

The bioinformatics analysis of STR sequencing data represents one of the main bottlenecks for the integration of MPS into standard casework laboratories. The present thesis introduces a novel open-access web application, toaSTR, that translates raw sequencing data into genetic profiles. The software engineering chapter provides insights into bioinformatics algorithms and the composition of application components. A novel stutter model proposed herein predicts and identifies artificial products originating from the analytical scheme. Sequence observations are automatically classified in order to assist in the interpretation of complex samples. Evidence from multiple studies has shown that toaSTR can precisely identify alleles from data obtained with various MPS platforms and identity panels. By emphasizing usability and versatility, toaSTR simplifies access to MPS data analysis for DNA laboratories without in-depth bioinformatics knowledge.

Wo kämen wir hin,
wenn alle sagten, wo kämen wir hin,
und niemand ginge, um einmal zu schauen,
wohin man käme, wenn man ginge?

– Kurt Marti (1921-2017)

Contents

List of abbreviations

AFLP	amplified fragment length polymorphism
ATemp	annealing temperature
ATime	annealing time
bp	base pair(s)
CE	capillary electrophoresis
CODIS	Combined DNA Index System
CV	coefficient of variation of the locus coverage
CR	calling ratio
Cyc	cycles
DAD	DNA-Analyse-Datei
dNTP	deoxyribonucleoside triphosphate
DOE	Design of Experiments
ESS	European Standard Set
ETime	extension time
GUI	graphical user interface
GEDNAP	German DNA Profiling
IEEE	Institute of Electrical and Electronic Engineers
InDel	insertion/deletion
ISFG	International Society for Forensic Genetics
kb	kilobases
LOD	limit of detection
LUS	longest uninterrupted stretch
MAF	minor component allele frequency
MLR	multiple linear regression
MPS	massively parallel sequencing
mtDNA	mitochondrial DNA
mPCR	multiplex polymerase chain reaction
NGS	next-generation sequencing
nt	nucleotide(s)
ORM	object relational mapper
PCR	polymerase chain reaction
QA	quality assurance
QC	quality control
RE	recognition element
RFLP	restriction fragment length polymorphism
RSM	response surface methodology
SDD	software design description
SLUS	second-longest uninterrupted stretch

List of abbreviations (continued)

SN	signal-to-noise ratio
SNP	single nucleotide polymorphism
SRS	software requirements specification
ST	stutter threshold
STR	short tandem repeat
SWGDAM	Scientific Working Group on DNA Analysis Methods
VNTR	variable number of tandem repeats

List of figures

List of tables

Preface

forensics /fəˈren.zɪks/ n. [L. *forum*, public center, *forensis*, in open court]
Scientific methods used in connection with the investigation of crime.
— *Oxford English Dictionary*

In ancient Rome, a forum was a public square used for judicial and other businesses. Legal proceedings, criminal investigations, verdicts of guilt or innocence, and the execution of sentences were carried out publicly on the marketplace. Nowadays, an important area of the criminal justice system – the scientific analysis of biological evidence – is located at the forensic DNA laboratory.

It was in the 1980s that dispersed tandem-repetitive sequences were discovered in certain regions of human deoxyribonucleic acid (DNA), which could differ from individual to individual – similar to a fingerprint. Molecular techniques for the examination of these polymorphic regions laid the foundation to perform human identity testing. Since that time, DNA typing methods have become a forensic key instrument and found widespread application: In paternity testing, comparing genetic profiles helps to identify a child's biological father; Mass disaster victims may be identified through the genetic profiles of their relatives; DNA traces found at a crime scene are compared to reference samples from the victim or suspects in order to reconstruct the course of events. Moreover, DNA testing not only plays a role in implicating the guilty but also in protecting the innocent. The Innocence Project's mission is to exonerate wrongly convicted people by re-testing old DNA evidence. Innocent prisoners, previously incarcerated for crimes they did not commit, continue to be released through modern DNA testing. More recently, cutting-edge DNA sequencing technology has found its way into forensic sciences, improving the resolution of genetic tests for the successful analysis of challenging samples. Given the complex data that is generated with these massively parallel sequencing (MPS) methods, tailored bioinformatics tools play an essential role in transforming sequence data into a reliable genetic profile.

This dissertation covers the construction, optimization, and validation of a novel molecular assay for human identification using MPS. In addition, the present work describes the development of an innovative web-based software tool for genetic profiling that promotes the integration of MPS into forensic laboratories. The overall structure of this thesis takes the form of four chapters:

Chapter 1 provides a brief review of the biology and technology of genetic markers used in forensic DNA analysis with a focus on short tandem repeats. Critical aspects of assay development are illuminated, and the Design of Experiments methodology is introduced as an economic framework for assay optimization. This chapter also presents MPS as an advanced technique for high-resolution DNA profiling and creates the thematic context of MPS with the field of bioinformatics for sequence data processing and result interpretation.

Chapter 2 describes the aims of the present work and derives research approaches.

Chapter 3 summarizes the results of the publications contributing to this cumulative dissertation and discusses their thematic relationship. Specimen copies of the papers may be found in the Publications section.

Chapter 4 offers a higher-level overview and gives an outlook on future work.

Parts of this thesis have been previously published (in chronological order):

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 [Status: published]

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 [Status: published]

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 [Status: published]

1 Introduction

1.1 DNA typing methods for human identification

The term **DNA typing**, or **DNA profiling**, refers to methods of identifying an individual through the analysis of patterns arising from differences in genetic markers. Generally, DNA typing is "a matter of comparing a question sample against a known sample" (Butler, 2010). In crime scene investigations, a known sample from the victim or suspect is compared to an evidential stain. If both samples share the same genetic type, an association might be made between the crime scene and the donor of the known sample. Since two human beings share approximately 99 % of their DNA (The 1000 Genomes Project Consortium, 2015), it is only one percent of the genome that is of interest to the forensic scientist. Genetic diversity partly manifests itself in phenotypic traits such as body height, eye color, and blood type (Rudin and Inman, 2002). More often, however, genetic differences must be investigated using analytic laboratory techniques.

The field of forensic biology has advanced quickly over the past decades. Prior to DNA typing becoming an established standard, forensic laboratories utilized several other serological and molecular techniques in order to characterize biological evidence and derive an individual profile. Important properties of a forensic test method are sensitivity, labor time, and the power of discrimination (i.e., the ability to differentiate between individuals). The following review will briefly discuss historic milestones and current techniques; however, it does not claim to be exhaustive.

1.1.1 Blood group typing

For many years, blood group typing was the only available forensic test. "Blood groups are related to antigen polymorphisms present on the surface of red blood cells. These antigens may be protein, carbohydrate, glycoprotein, or glycolipid differences that exist between people" (Butler, 2010). The first blood group system to be discovered was the ABO system (Landsteiner, 1900), which was also the first genetic tool used for distinguishing between individuals. Following this, several other blood group systems such as Rhesus, Kell, etc. were identified (Daniels, 2013), however, most are not of forensic importance. The ABO system discriminates four blood types: A, B, AB, and O. Each type denotes the presence or

absence of specific carbohydrate antigens present on erythrocytes as well as the associated immunoglobulin M antibodies present in the serum (Storry and Olsson, 2009). The oligosaccharide epitopes of the antigens are not coded by genes directly but synthesized by a set of glycosyltransferases, which are encoded by the ABO gene (Ferguson-Smith et al., 1976). The A and B alleles encode glycosyltransferases that produce the A and B epitopes, respectively. The O allele is considered to be non-functional because its protein product determines no detectable blood group antigen (Goubran, 2009). If an evidentiary bloodstain is found at a crime scene, the blood type of its donor can be determined using a rapid and simple serological test with anti-A anti-B antibodies that agglutinate antigen-positive red blood cells (Goubran, 2009). However, the result is of limited informative value. Although the frequency of the ABO phenotypes varies among different populations, either type A or type O is observed about 80 % of the time (Butler, 2010). Given this poor power of discrimination, ABO blood typing is useful for excluding an individual from being the donor of the sample, but it might not help to answer the question "Whose blood is it?" completely.

1.1.2 Analysis of restriction fragment length polymorphisms

With the discovery of hypervariable regions in the human genome, the use of DNA gained greater attention for human identification purposes. Jeffreys et al. (1985) described dispersed tandem-repetitive sequences across the genome, "showing multiallelic variation and correspondingly high heterozygosities". By using labeled probes targeting these so-called **minisatellites** in restriction fragment length polymorphism (RFLP) analysis, the produced pattern appears as an individual-specific "DNA fingerprint". Soon it was realized that this technique holds the potential to revolutionize forensic biology due to its high power of discrimination (Gill et al., 1985).

The term **satellite DNA** describes the secondary (or satellite) fraction that forms when separating sheared genomic DNA in density gradient centrifugation. Due to a different frequency of the bases adenine, cytosine, guanine, and thymine, large centromeric tandem repeats show a distinct buoyant density, detectable as a "satellite peak" (Beridze, 1986). Referring to this effect, shorter tandem repeats came to be known as minisatellites, also termed variable number of tandem repeats (VNTRs). They consist of a repeated sequence motif of approximately six to 100 bp in length. The observed polymorphisms mainly derive from allelic differences in the number of repeat units, resulting in alleles of about 400 to 1,000 bp (Vergnaud, 2000). In the pre-PCR era, VNTRs were routinely analyzed

using RFLP-based testing. The RFLP process (Kirby, 1993) involves the digestion of extracted DNA using a restriction enzyme such as *Hae*III. The DNA fragments are subsequently size-separated by means of agarose gel electrophoresis and a Southern blot (Southern, 1975) is performed to transfer the fragments from the gel to a nylon membrane. Radioactively labeled oligonucleotide probes, which contain the sequence motifs of VNTR loci, hybridize to the DNA attached on the nylon membrane. Finally, the radioactive probed membrane is exposed to X-ray film, resulting in a complex band pattern that suggests a fingerprint, similar to a barcode. Widely used VNTR markers such as D1S7, D10S28, and D17S79 show *Hae*III fragment sizes between 0.5 and 12 kb and a repeat unit length ranging from nine to 38 bp (Smith et al., 2000). Due to a large number of up to 30 alleles per locus, RFLP features an excellent power of discrimination of one in several million with four loci. However, this labor-intensive method requires relatively large amounts (> 50 ng) of high-molecular weight DNA (fragment size of 20 to 25 kb), which is a severe limitation since forensic evidence is often of limited quantity and degraded due to bacterial, biochemical, or oxidative processes (Butler et al., 2003). If degradation has taken place, high-molecular bands might be missed, leading to the absence (**dropout**) of an allele.

1.1.3 PCR-amplified minisatellites

With the advent of the polymerase chain reaction (PCR) in the late 1980s (Mullis et al., 1986), the forensic community began to embrace DNA testing. PCR has a significantly improved sensitivity compared to RFLP (Butler, 2010) and enables typing of polymorphic loci with much less amount of input DNA – even single-cell analysis has been demonstrated early (Jeffreys et al., 1988). One of the early PCR-based methods for human identity testing was the amplified fragment length polymorphism (AFLP). Instead of enzymatic digestion as with RFLP, this technique relies on PCR amplification of a VNTR region and subsequent size separation of alleles on stained agarose or polyacrylamide gels (Jeffreys et al., 1988). Alleles of an unknown sample are typed by comparison with an allelic ladder, that is, an artificial mixture of common alleles for a particular genetic marker. The most popular forensic AFLP marker, the minisatellite locus D1S80, contains a 16-bp repeat motif (Kloosterman et al., 1993). Its large allele length range of 18 to 42 repeats (430 – 814 bp) facilitates mixed-sample analysis, however, preferential amplification of shorter alleles may occur if the sample exhibits a large allele spread (Tully et al., 1993). Hence, large alleles might be missed. In addition, successful typing of minisatellites requires intact DNA of high-molecular weight in order to amplify full-length alleles.

1.1.4 Short tandem repeat genotyping

To overcome limitations associated with minisatellites, the forensic DNA field has transitioned on the PCR-amplification of **microsatellites**, also known as **short tandem repeats (STRs)**. In contrast to minisatellites, microsatellites consist of shorter repetitive elements such as $[CA]_n$ or $[TAGA]_n$ (Figure 1). While most microsatellites exhibit mono-, di-, tri and tetranucleotide motifs, repeats of five or six nucleotides are usually classified as microsatellites as well (Ellegren, 2004). STR markers have become the mainstay for identity testing for the following reasons:

- high sensitivity due to PCR amplification (Müller et al., 2018; Wiegand et al., 1993)
- reduced preferential amplification since the repeat size is smaller than VNTRs (Schmitt et al., 1994)
- capabilities for analyzing degraded samples with short amplicons (Butler et al., 2003; Kim et al., 2016)
- excellent discriminatory power of about one in billions or greater with eight to nine or more STR loci (Butler, 2010); and
- the ability to deconvolve human DNA mixtures (Clayton et al., 1998; Jäger et al., 2017)

As with other PCR-based techniques, disadvantages include the susceptibility to PCR inhibitors, failed amplification due to sequence mutations in the primer-binding region, and possible contamination with human DNA that does not originate from the sample.

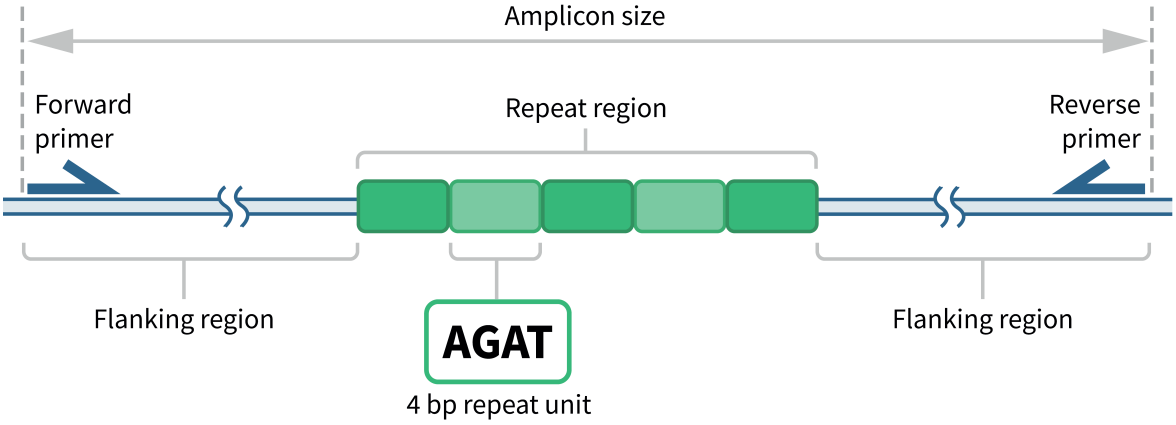


Figure 1. Schematic of a short tandem repeat (STR) locus. The exemplified repeat region consists of five tetrameric repeat units with the repeat motif AGAT. Binding sites for PCR primers are placed in conserved sequences surrounding the repeat stretch. The location of primers determines the length of the amplified flanking regions. Since the number of tandem repeat units varies among individuals, the overall amplicon size depends on the actual size of the repeat region as well.

The standard method for human DNA profiling involves the amplification of STR loci with fluorescently-tagged PCR primers followed by capillary electrophoresis (CE) size separation and length-based allele typing relative to a physical allelic ladder (Lazaruk et al., 1998). More recently, the application of MPS brings modern advantages to forensic DNA analysis. STR sequencing using MPS provides a higher genetic discrimination power compared to CE, since alleles of the same length are identified not only by the number of repeats but by the actual STR sequence on a nucleotide level (Bornman et al., 2012; Dalsgaard et al., 2013; Gettings et al., 2016). This high-resolution DNA typing promises benefits for the deconvolution of samples with multiple contributors and the analysis of partial STR profiles. The next sections are dedicated to the biology behind microsatellites as well as established and advanced technologies for STR typing.

Other classes of genetic markers that currently play a supplemental role in DNA typing include single nucleotide polymorphisms (SNPs) providing identity, ancestry, and phenotypic information (Churchill et al., 2016). Mitochondrial DNA (Holland and Parsons, 1999) enables the recovery of DNA information from challenged sources such as ancient material and allows the investigation of maternal inheritance. However, a full discussion of these marker classes lies beyond the scope of this thesis.

1.2 Biology and technology of short tandem repeat markers

1.2.1 Evolutionary dynamics of microsatellites in the human genome

Microsatellites are extremely common in eukaryotic genomes (Buschiazzo and Gemmell, 2006). Depending on how they are counted, there may be more than one million STR loci present in the human genome (Ellegren, 2004), constituting about three percent of the genetic information (Lander et al., 2001). Frequent mutations make them highly polymorphic in the number of repeat units and the repeat sequence (Gymrek and Erlich, 2013). The presence of multiple alleles and the ubiquitous occurrence explain their usefulness as molecular markers for genome mapping (Collins et al., 2003; Weissenbach et al., 1992) and population genetics (Bowcock et al., 1994; Sun et al., 2012). Due to their abundance and diversity in human populations, STRs have become the marker of choice for forensic DNA analysis (Jobling and Gill, 2004) and the development of national and international DNA databases (Gill et al., 2015). STR profiling continues to find its application in human and non-human cell line authentication for biomedical research and production (Reid et al., 2004) as well as the detection of cellular cross-contamination in biotechnological applications (Almeida et al., 2016). Moreover, expansions of triplet

repeats in various regions of their resident genes are implicated in the etiology of several human genetic diseases, including myotonic dystrophy and Huntington's disease (Mirkin, 2007). Another phenomenon termed microsatellite instability, that is, deletion mutations in microsatellite sequences arising from defects in the DNA mismatch repair system (Peltomäki et al., 1993), has been associated with hereditary nonpolyposis colorectal cancer (Lynch syndrome; Boland and Goel, 2010).

The vast majority of STRs are embedded in non-coding DNA (Ellegren, 2004), while some regional variation indicates a nearly twofold higher density near the ends of the chromosome arms in human and mouse genomes. The formation of an STR out of non-repetitive DNA was proposed to be a rare event with complex biology (Ananda et al., 2013). It requires "nontrivial mutations, such as the arrival of a transposable element or precise point mutations that destroy non-repetitive gaps between two short repeat stretches" (Willems et al., 2014). As with minisatellites, microsatellite polymorphisms mainly derive from variability in length with a strong tendency to mutate in whole repeat unit increments (Fondon et al., 2012). DNA polymerase-mediated replication slippage (Figure 2) is generally accepted as the main mechanism for length changes in repetitive DNA, resulting in the observed length polymorphisms between individuals of a population (Levinson and Gutman, 1987; Schlötterer and Tautz, 1992). According to this mechanism, during replication of a repeat tract, DNA polymerase stalling occurs at a pausing site such as a hairpin-like or quadruplex-like structure, which is formed frequently in STR sequences (Mirkin, 2007). Upon pausing, the polymerase dissociates from the primer-template-complex (Viguera et al., 2001). This allows a transient breathing of the replicating DNA strands. If out-of-register re-alignment occurs, this may either introduce a loop in the nascent or the template strand. Most slipped-strand loops should be recognized and correctly repaired by the DNA mismatch repair system (Dexheimer, 2013), resulting in no length change. If, however, errors are not repaired, re-association of the polymerase and renewed replication will lead to repeat expansion or contraction by one or more repeat units, respectively.

The susceptibility of a microsatellite to length mutations, hence the average number of alleles for a repeat locus, is correlated with the repeat unit size, the number of repeated units, and interruptions in the repeat sequence (Fondon et al., 2012; Legendre et al., 2007). Kelkar and colleagues (2010) experimentally determined the *in vitro* mutability of human mononucleotide and dinucleotide microsatellites as a function of repeats. While slippage rates were low at small repeat numbers, a significant shift above background slippage rates was observed when the repeat tract had a length of at least ten nt. Hence, the authors

proposed this critical threshold as an experimental definition to qualify a sequence consisting of short tandem-like repeats as a mutational hotspot and thus constituting a microsatellite. Similarly, Lai and Sun (2003) estimated the threshold size to be nine repeats for mononucleotide and four repeats for di- to hexanucleotide units. The rate of length mutation was found to be positively correlated with the number of repeat units. For common forensic STR markers, mutation rates of 10^{-2} to 10^{-4} per locus per generation have been reported (Butler, 2010).

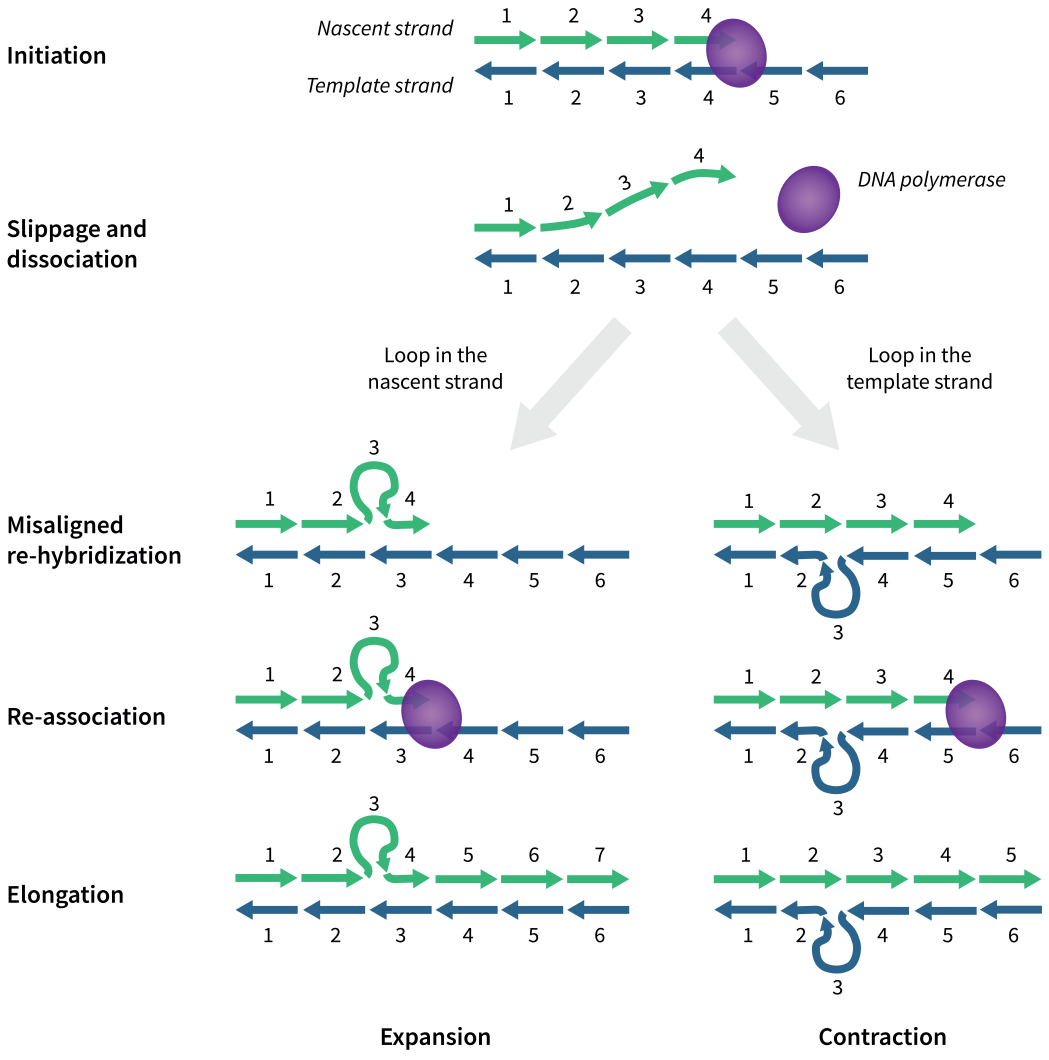


Figure 2. Model of microsatellite mutation based on DNA polymerase-mediated replication slippage. Repeat units are indicated by arrows and numbered within each strand. In the initiation step, the DNA polymerase has extended through four repeat units. If polymerase slippage occurs, this allows transient breathing of the replicating strands. Out-of-register re-alignment may subsequently introduce a loop in the nascent strand or the template strand. Re-association of the polymerase and continued elongation will then lead to repeat expansion or contraction, respectively.

Although being hypervariable spots in the genome, STRs do not expand into enormous arrays of repeated elements but might shrink and expand in a dynamic process (Ellegren, 2004). How is this circumstance solved? A well-received model of microsatellite evolution (Kruglyak et al., 1998) considers a balance between length and point mutations: While the former favors expansions of repeat arrays, the latter break long repeat arrays into smaller units. To add to the complexity, long alleles are suggested to be biased towards contraction (Xu et al., 2000), which may offer an explanation for the stationary allele distribution of STRs and may explain the scarcity of long microsatellites (Lai and Sun, 2003). In addition, point mutations occurring in a long repeat array can interrupt the repetitive sequence, leaving two shorter repeat stretches exhibiting lower mutation rates. Point mutations in the flanking regions have been described to hinder STR expansion as well (Santibáñez-Koref et al., 2001).

1.2.2 Forensically relevant autosomal STR loci

Markers used for human identity testing are found in non-coding regions of the DNA, either between genes or in introns (Butler, 2010). Thus, they do not code for phenotypic variation nor are they known to have any association with a genetic disease. Tri-, tetra-, and pentanucleotide repeats have become the most popular STR markers because length increments are large enough to allow resolving closely spaced alleles in electrophoretic separation. Moreover, longer repeats exhibit a lower tendency to form stutter artefacts during PCR amplification, a phenomenon that will be discussed in Section 1.3.1.

Over the years, a set of well-characterized STR markers has evolved that are in common use for routine DNA profiling (Table 1). STR allele frequencies have been established in the course of population genetic studies (Gettings et al., 2018; Gill and Evett, 1995). Allele frequencies enable biostatistical evaluation of genotypes to calculate probabilities of exclusion or inclusion when comparing DNA profiles. Upon building national forensic databases in the USA and Europe, the forensic community developed a selection of autosomal core loci enabling consistent databasing and exchange of DNA profiles. In the United States, the Combined DNA Index System (CODIS) was set up in 1997 and subsequently expanded with more markers (Hares, 2015; Hill, 2013). In order to harmonize DNA profiling results between the member states of the European Union, the European DNA Profiling (EDNAP) [1] group and the DNA Working Group of the European Network of

[1] https://www.isfg.org/EDNAP

Forensic Science Institutes (ENFSI)[2] recommended a European Standard Set (ESS) of markers. The ESS was later expanded to minimize the chance of adventitious matches and to enable the international exchange of DNA profiles (Schneider, 2005). Nowadays, the core loci of the expanded marker sets used in the USA and Europe are largely identical. The national German DNA Database (DNA-Analyse-Datei, DAD) was initially composed of five marker systems and has been expanded to a set of eight loci in 2001. As a unique feature, the DAD set includes the complex repeat system SE33 (Polymeropoulos et al., 1992), which is one of the most informative tetranucleotide loci studied to date. Its highly polymorphic and heterozygous nature exhibits complex length and sequence variation (Lászik et al., 2001; Rolf et al., 1997; Wiegand et al., 1993), which is useful for the analysis of mixed samples containing DNA of multiple donors. Variation in the flanking regions of SE33 adds to the extensive diversity (Gettings et al., 2015b). Conversely, the sequence complexity of SE33 may lead to an ambiguous demarcation of the repeat region for some alleles, complicating a consistent length-based and sequence-based nomenclature (Parson et al., 2016). In an MPS-based population study, Borsuk et al. (2018) observed allele-size dependent variance in coverage and related this phenomenon to the large size range of alleles from three to 49 repeats. The resulting heterozygous genotype imbalance makes the interpretation of sequence data challenging, since an imbalance might erroneously suggest a mixed sample or may even lead to an allele dropout. Moreover, unstable and AT-rich flanking regions render this locus a challenge to design multiplex-compatible PCR primers (Phillips et al., 2018). For genetic gender identification, STR database profiles can be supplemented with the X-Y homologous locus amelogenin. This gene codes for proteins found in tooth enamel and occurs on both heterochromosomes (Sullivan et al., 1993). PCR amplification commonly targets a deletion within intron one of the amelogenin gene carried by the X homolog. The Y-specific amplicon is six bp longer than the X-specific product. The amelogenin marker allows designating whether a sample originated from a male or a female source, which is useful for example in sexual assault cases.

STR markers are commonly subdivided into three categories, depending on their typical sequence structure (Gettings et al., 2015b): Simple repeats contain repeat units of identical length and sequence; Compound repeats consist of two or more adjacent simple repeats; Complex repeats may additionally contain interspersed non-repetitive elements and repeat stretches of variable length. Occasionally, so-called microvariants are observed, which contain incomplete repeats arising from mutations in the repeat region.

[2] http://enfsi.eu/

Table 1. Information on the autosomal STR core loci used in the United States, Europe, and Germany.

Locus	eCODIS [a]	eESS [b]	DAD [c]	Chromosome location	General sequence structure [d]	Category	Allele length range [e]	Repeat length range
D1S1656	●	●		1q42	CCTA $[TCTA]_n$ TCA $[TCTA]_n$	complex	9 – 21	36 – 84 nt
D2S441	●	●		2p14	$[TCTA]_n$ TCA $[TCTA]_n$	complex	8 – 17	32 – 68 nt
TPOX	●			2p25.3	$[AATG]_n$	simple	4 – 16	16 – 64 nt
D2S1338	●	(●)		2q35	$[GGAA]_n$ GGAC $[GGAA]_n$ $[GGCA]_n$	complex	11 – 28	44 – 112 nt
D3S1358	●	●	●	3p21.31	$[TCTA]_n$ $[TCTG]_n$ $[TCTA]_n$	compound	8 – 20	32 – 80 nt
FGA	●	●	●	4q31.3	$[GGAA]_n$ GGAG $[AAAG]_n$ AGAA AAAA $[GAAA]_n$	complex	12.2 – 51.2	62 – 206 nt
D5S818	●			5q23.2	$[ATCT]_n$	simple	6 – 18	24 – 72 nt
CSF1PO	●			5q33.1	$[ATCT]_n$	simple	5 – 16	20 – 64 nt
SE33		(●)	●	6q14	$[CTTT]_n$ TTCT $[CTTT]_n$	complex	3 – 39.2	12 – 158 nt
D7S820	●			7q21.11	$[TATC]_n$	simple	5 – 16	20 – 64 nt
D8S1179	●	●	●	8q24.13	$[TCTA]_n$ $[TCTG]_n$ $[TCTA]_n$	compound	7 – 20	28 – 80 nt
D10S1248	●	●		10q26.3	$[GGAA]_n$	simple	8 – 19	32 – 74 nt
TH01	●	●	●	11p15.5	$[AATG]_n$ ATG $[AATG]_n$	complex	3 – 14	12 – 56 nt
D12S391	●	●		12p13.2	$[AGAT]_n$ GAT $[AGAT]_n$ $[AGAC]_n$ AGAT	complex	15 – 26	60 – 104 nt
VWA	●	●	●	12p13.31	$[TAGA]_n$ $[CAGA]_n$ TAGA	compound	10 – 25	40 – 100 nt
D13S317	●			13q31.1	$[TATC]_n$	simple	5 – 17	20 – 68 nt
D16S539	●	(●)		16q24.1	$[GATA]_n$	simple	4 – 16	16 – 64 nt
D18S51	●	●	●	18q21.33	$[AGAA]_n$ AG	simple	7 – 39.2	28 – 158 nt
D19S433	●	(●)		19q12	$[CCTT]_n$ ccta $[CCTT]_n$ cttt $[CCTT]_n$	complex	5.2 – 20	30 – 88 nt
D21S11	●	●	●	21q21.1	$[TCTA]_n$ $[TCTG]_n$ $[TCTA]_n$ ta $[TCTA]_n$ tca $[TCTA]_n$ tccata $[TCTA]_n$ TA $[TCTA]_n$	complex	12 – 41.2	59 – 177 nt
D22S1045	●	●		22q12.3	$[ATT]_n$ ACT $[ATT]_n$	compound	8 – 20	24 – 60 nt
Amelogenin	●	(●)	●	Xp22.2, Yp11.2	-	InDel	X, Y	-

[a] Expanded Combined DNA Index System. [b] Expanded European Standard Set; parentheses indicate additional loci that are part of European STR kits. [c] DNA-Analyse-Datei (German core loci).
[d] Forward strand notation; nucleotide tracts within the repeat region that are not counted towards the allele designation are given in lower case letters. Due to the large variability of sequence alleles, the general sequence structure does not necessarily reflect the entire complexity of all observable alleles. [e] Obtained from STRBase (https://strbase.nist.gov/), accessed May 29, 2019.

For enhanced legibility, the repeat region sequence is usually expressed in a compressed notation. Core repeats are written in brackets with a number indicating the respective repeat count. The naming convention for STRs distinguishes markers that fall inside and outside of genes, respectively (Strachan and Read, 2018). For instance, the marker TH01 is located within intron 01 of the tyrosine hydroxylase gene. Other gene-associated loci reflect the gene name, such as CSF1PO (c-fms proto-oncogene for the CSF-1 receptor) or FGA (fibrinogen alpha chain). DNA markers that are not part of a gene have identifiers according to the chromosome on which the marker is located and the order of discovery. For instance, D13S317 stands for DNA marker, chromosome 13, single-copy sequence, 317th discovery.

1.2.3 STR genotyping by fragment length analysis

Analysis of length polymorphism at STR loci nowadays almost exclusively relies on PCR amplification and subsequent fragment sizing using capillary electrophoresis (CE). In order to reach a sufficient power of discrimination, multiple STR loci are simultaneously amplified through multiplex PCR with fluorescently-labeled primers, followed by amplicon length detection through size separation (Hill et al., 2011; Wang et al., 2012a). The separation is based on the electrophoretic mobility of DNA fragments. Longer fragments move slower through the polymer-filled capillary than smaller fragments. Allelic profiles at each locus are obtained by laser-induced multicolor fluorescence detection. The fluorescence intensity in each color channel is plotted against the amplicon size in an electropherogram (Figure 3), visualizing alleles as peaks. Assuming a single-person sample and a normal diploid set of chromosomes, one would expect to see one homozygote allele peak or two heterozygote allele peaks per STR marker, respectively. Numerical alleles are typically identified by comparing fragment lengths with a reference from an allelic ladder. In this way, allele lengths of multiple loci are compiled to an STR profile. A profile consists of a pair of numbers for each investigated locus representing the number of repeats of the called alleles. The DNA Commission of the International Society for Forensic Genetics (ISFG) has developed guidelines (Bär et al., 1997) for the numerical designations of length-based alleles. The naming convention for simple and compound repeats is straightforward: integer values indicate the number of complete repeat units. Microvariants are designated by the number of full repeat units separated by a decimal point and the number of additional nucleotides. Complex repeats may contain internal sequence variation that is not counted toward the allele designation. Full allele characterization, however, can only be achieved through sequence analysis. Discordances

between length-based designations and the actual repeat sequence may arise due to mobility shifts during CE separation (Fujii et al., 2016). InDels in amplified flanking regions may also cause discordant results, especially when comparing different primer sets that may or may not enclose an InDel site (Silvery et al., 2020; part of this dissertation).

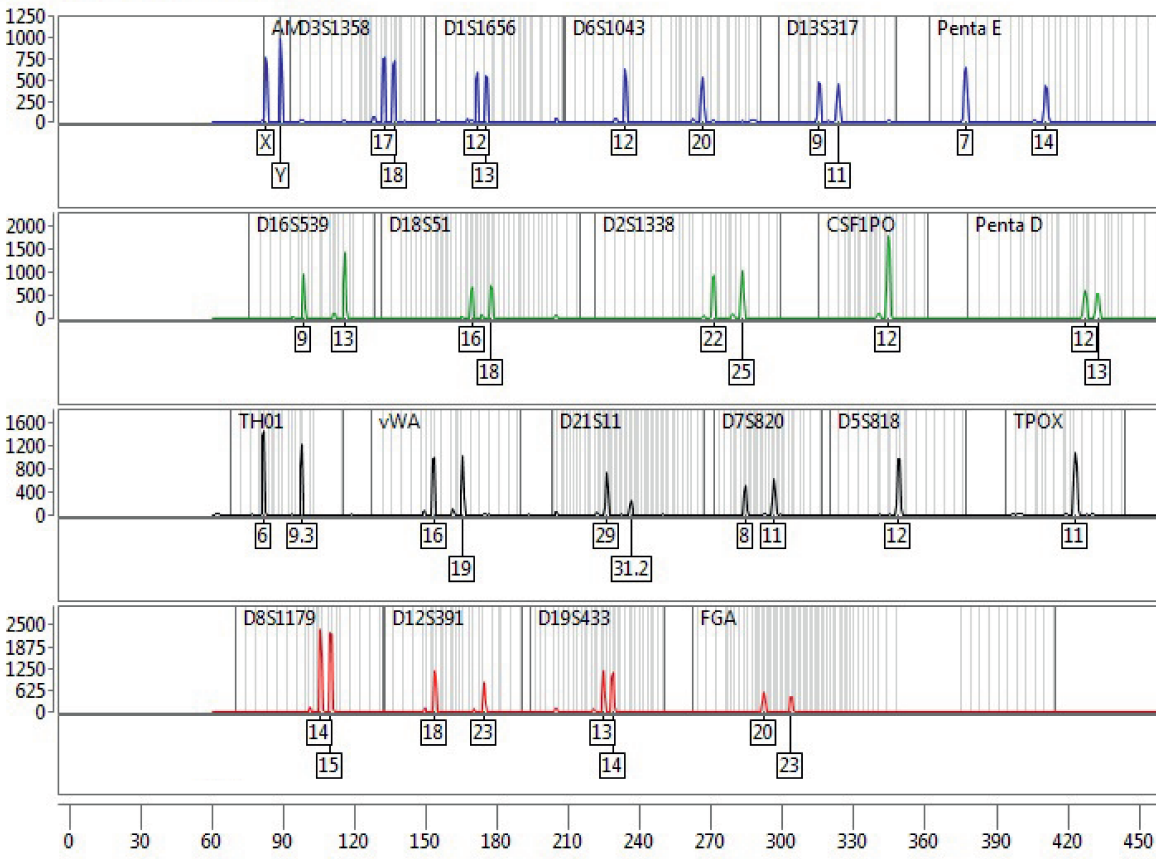


Figure 3. Sample electropherogram from a CE kit with 21 STR markers. Four dye channels are used (represented in blue, green, black and red). Grey vertical bars indicate the bins of the allelic ladder, a molecular ruler that enables to correlate fragment sizes to STR allele repeat numbers. Boxed values represent the called numerical alleles.

Length-based STR genotyping using CE requires STR amplicons to be spectrally and spatially distributed across multiple dye channels (Figure 3). Newer CE-based assays support a 6-dye chemistry format containing up to 30 mainly autosomal STR markers (Ludeman et al., 2018; Westen et al., 2014). Since the multiplexing capability is limited, analysis of additional markers such as X-STRs and Y-STRs involves iterative testing, which is complicated by the availability of only minute quantities of DNA material. This circumstance may force the forensic analyst to restrict the investigations. A complex spatial layout in each dye channel must ensure sufficient separation of markers to avoid

overlapping fragment lengths. Therefore, some amplicons are designed longer than would be necessary to reflect the information of the repeat region only. Amplicon lengths between 80 and 470 bp are common in commercial autosomal STR kits (Westen et al., 2014). Consequently, highly degraded samples or PCR inhibited samples, as encountered in criminal casework or mass disaster investigations, may produce partial STR profiles or inconclusive results (Hughes-Stamm et al., 2011). In this case, a "mini STR" approach (Butler et al., 2003) using reduced size STR amplicons may facilitate the analysis of degraded DNA evidence by improving PCR efficiency.

Several potential sources of error are directly connected with the CE-based procedure and may complicate STR genotyping. Occasionally, off-ladder alleles occur if a fragment size falls in between the bins of the allelic ladder or the allele peak is larger or smaller than the alleles spanning the allelic ladder range (Butler, 2014). These inconclusive results require manual intervention by a forensic analyst. Off-ladder alleles and discordances have also been described as a result of SNP-induced secondary structures of the amplified fragments affecting the electrophoretic mobility (Wang et al., 2012b). Further, technology- and biology-related issues are known to generate extraneous peaks in the electropherogram that need to be differentiated from true alleles to obtain a correct interpretation. Incomplete spectral separation may cause a pull-up or bleed-through of the signal between dye colors. Pull-up peaks, i.e. an elevated baseline in a neighbored dye channel, arise of overlapping emission spectra if an over-amplified sample is analyzed (Clayton et al., 1998). Another phenomenon known as dye blobs occurs when fluorescent dye molecules detach of the respective primers and migrate independently through the capillary. Biology-related artefact peaks arise from incomplete 3' adenylation (Clark, 1988) by the DNA polymerase used for amplifying STR loci, resulting in a split peak representing +A and -A products. The most common source of additional peaks, however, are **stutter** products (see Section 1.3.1). Caused by polymerase slippage during primer extension, these events can be recognized as minor peaks that differ from the amplified allele by multiples of the length of a repeat unit. A stutter percentage interpretational threshold may be set to distinguish a stutter product from a true allele (Butler, 2014). Nevertheless, the interpretation of imbalanced mixtures is highly compromised by stutter artefacts, since stutter peaks may be easily confused with allele peaks of a minor contributor.

1.2.4 High-resolution STR analysis by massively parallel sequencing

Massively parallel sequencing, also designated as next-generation sequencing (NGS), is a rapidly evolving technology. As a characteristic property, MPS enables sequencing of billions of short DNA fragments in parallel (Goodwin et al., 2016). MPS provides a much cheaper (in terms of cost-per-base) and higher-throughput alternative to sequencing DNA than traditional Sanger sequencing and has paved the way for sequencing complete genomes rapidly and at reasonable costs (Metzker, 2010). However, since DNA typing in its current legal, ethical, and economical setting focusses on selected genetic loci, forensic investigations typically involve targeted sequencing of PCR-amplified STRs rather than genome-wide analyses.

MPS surmounts many of the limitations encountered with PCR-CE typing (Table 2). High-density multiplexing of STR loci and pooling of multiple samples in a single sequencing run is possible. STR analyses can easily be linked with other forensic markers, such as SNPs, mtDNA or RNA in the same assay (Bruijns et al., 2018). This enables exciting possibilities of combining human identification with biogeographic ancestry information (Hussing et al., 2018), determination of phenotypical traits such as hair, skin, and eye color (Churchill et al., 2016), prediction of the chronological age (Parson, 2018), and characterization of tissues and body fluids (Zubakov et al., 2015). In addition to the allele length information gathered with electrophoretic sizing, MPS-derived genotypes ascertain nucleotide variations in the repeat region and nearby flanking regions (Børsting and Morling, 2015; Gettings et al., 2015a). As a result, equi-length alleles that differ by sequence, also termed isometric heterozygotes or **isoalleles**, can be resolved.

A wide range of sequence variation features in and around forensic STRs have been identified and annotated (Gettings et al., 2018). To date, the STR Sequence Guide (Phillips et al., 2018), a curated collection of forensic markers that have been characterized on a nucleotide level, lists 37 autosomal microsatellites and 41 X- and Y-chromosome microsatellites. It was demonstrated that the power of discrimination could be improved by differentiating the nucleotide sequences of alleles that share an identical length (Borsuk et al., 2018). For instance, a 250 % increase in alleles by sequence versus length has been observed at the locus D12S391 (Gettings et al., 2015b). Therefore, MPS is relevant especially for the deconvolution of complex mixtures, which often exhibit overlapping isoalleles and stutter products of multiple contributors. From a technical view, identifying STR amplicons by sequence, instead of spatial and spectral differences as with CE assays, simplifies the amplicon layout. This enables also typing degraded DNA with a set of small,

uniform amplicons (Kim et al., 2016). By design, MPS eliminates issues associated with electrophoretic separation, off-ladder phenomena, and the need for an allelic ladder at all.

While MPS offers a gain in information compared to CE, "the complexity of data processing and interpretation as well as the vulnerability to external factors like instrument performance, laboratory environment, staff training and quality levels of reagents and consumables may increase as well" (Köcher et al., 2018). Fundamental to massively parallel sequencing is the preparation of a so-called **sequencing library** that is compatible with the respective sequencing platform (Head et al., 2014). In the case of amplicon sequencing, **library preparation** involves the modification of PCR products with adapter oligonucleotides that mediate the immobilization of DNA fragments for clonal amplification and subsequent sequencing reactions. Adapters also contain sample-specific barcodes as well as binding sites for sequencing primers.

Table 2. Properties of conventional CE-based genotyping versus MPS.

CE	MPS
▪ Separation of amplification products by fragment size	▪ Loci can be identified by their DNA sequences
▪ Limited multiplex capability due to the spatial and spectral distribution of amplicons	▪ Greater flexibility in amplicon design. Simplifies typing degraded DNA with short amplicons
▪ Genotyping based on length only	▪ Genotyping by length and sequence
▪ Mixture analysis complicated by stutter products and overlapping isoalleles	▪ Differentiation of isoalleles and identification of stutter on a sequence level. Potential improvement to mixture analysis
▪ Relatively simple and established workflow	▪ More complex library preparation and sequencing workflow. Immense data requires bioinformatics solutions for analysis and interpretation
▪ Low cost per sample	▪ Pooling of samples needed to reduce cost-per-sample

Concerning data processing, generating an MPS-STR profile is fundamentally different from CE data analysis. As the extensive volume of sequence data is too immense to be handled manually, bioinformatics methods (see Section 1.4) are required to capture the relevant information and to aid in drawing conclusions. According to Liu and Harbison (2018), forensic MPS data may typically be processed in three stages. Primary data analysis includes base calling, computation of quality scores, adapter-trimming and

demultiplexing of barcoded sequences. This step produces an individual FASTQ file containing the sequence reads for each sample. Primary analysis is usually carried out on the respective sequencer. In the second step, raw sequence reads are subjected to quality control procedures such as trimming or filtering of low-quality reads in order to remove noise, which may otherwise bias the result interpretation. Stage three reflects the actual genotyping. Informative reads are assigned to STR loci through sequence alignment or sequence search. The subsequent allele calling step characterizes alleles by sequence and length and derives an STR profile. The number of reads for each allele (i.e., the allele coverage) corresponds with the peak height of CE-based methods. Alleles may be visualized as a stacked histogram (Figure 4), where each bar represents a sequence allele and the size of each bar indicates the allele coverage. This plot is analogous to an electropherogram; however, it allows to distinguish isoalleles and stutter by sequence.

The application of MPS to forensic casework also creates nomenclature challenges, as indicated by the ongoing discussion on the naming convention for STR profiles obtained with this new technological generation (Gettings et al., 2015b; Parson et al., 2016; Phillips et al., 2018; van der Gaag and de Knijff, 2015). MPS-derived alleles capture the detailed nucleotide sequence underlying the repeat region and flanking regions. This complexity cannot be accommodated by the numerical nomenclature conventions used with the CE method. Nevertheless, the nomenclature of MPS-derived STR alleles needs to be backward and parallel compatible with millions of CE-based profiles existing in DNA databases, as well as forward compatible with MPS data that will be generated with wider adoption of this technology. Therefore, the DNA Commission of the ISFG proposed a comprehensive STR nomenclature system (Parson et al., 2016), which unambiguously captures all of the information present in the STR sequence string. The format includes the locus name, CE-concordant allele name, chromosome number, human reference genome assembly version, repeat region coordinates of the reference allele, the STR sequence in compressed notation, and locations of flanking region variants between primer binding sites and the repeat region. An example of the proposed format would be

D13S317[CE12]-Chr13-GRCh38 82148025-82148068 [TATC]12 82148001-A; 82148069-T.

Shorthand identifiers for MPS alleles, analogous to numerical alleles used with fragment analysis, are yet to be defined (Gettings et al., 2019) and would preferably be curated by a centralized nomenclature commission to avoid ambiguous shorthand allele names (Phillips et al., 2018).

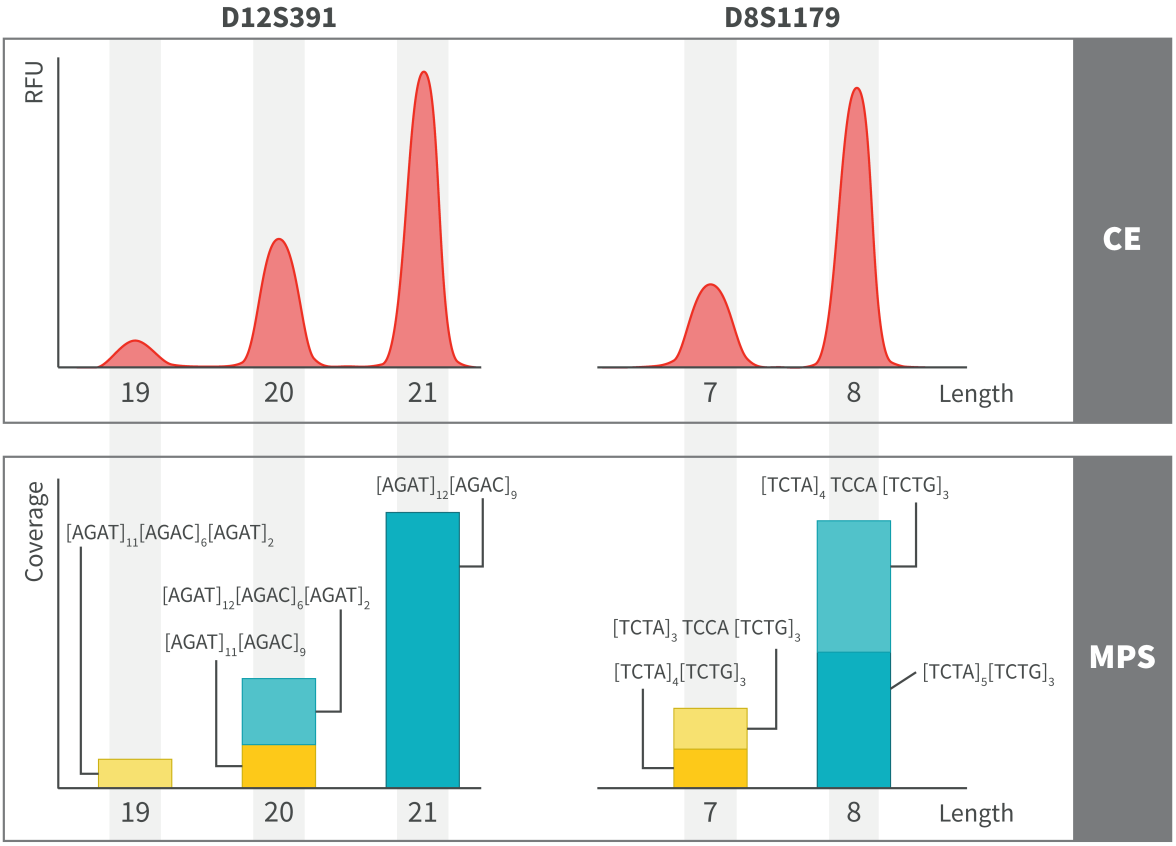


Figure 4. Comparison of length-based versus sequence-based genotypes for two exemplary STR loci. Top row: fragments appear as peaks in an electropherogram after electrophoretic separation. Bottom row: sequence-based alleles (blue bars) and stutter (yellow bars) obtained with MPS can be distinguished and visualized as a stacked histogram. Locus D12S391: With CE, the peak of length 20 might be (mis-) interpreted as a raised stutter peak of allele 21. However, MPS reveals that in position 20 a minor allele overlaps with N-1 stutter of the major allele 21. In position 19, N-1 stutter of the minor allele confirms this constellation. Locus D8S1179: CE would suggest a heterozygous allele of length 8 with N-1 stutter in position 7. However, two isoalleles of length 8 and their corresponding N-1 stutters of length 7 can be separated by sequence.

With regard to the actual implementation of MPS in forensic laboratories, the Illumina MiSeq platform and the Ion Torrent PGM/S5 sequencers are the preferred instruments (Alonso et al., 2017). At the time of writing, the Verogen (formerly Illumina) ForenSeq DNA Signature Prep Kit is the largest commercially available panel (Caratti et al., 2015; Xavier and Parson, 2017). It includes the expanded CODIS and expanded ESS core STR loci. This panel consists of two different primer mixes. The larger primer mix B of the ForenSeq kit contains 58 STRs and amelogenin as well as 94 identity informative SNPs, 56 ancestry informative SNPs, and 22 phenotypic informative SNPs for the inference of geographic ancestry and externally visible traits. Even though the highly polymorphic marker SE33 is

included in the primer mix (Borsuk et al., 2018), it is not officially supported and analyzed by the proprietary data analysis software Verogen UAS (Köcher et al., 2018). Previous publications described various applications of this assay including developmental validations (Jäger et al., 2017; Köcher et al., 2018) and population-scale studies (Gettings et al., 2018; Novroski et al., 2016). The ForenSeq system is not compatible with the RUO (research use only) version of the Illumina MiSeq sequencer but requires the dedicated Verogen MiSeq FGx Forensic Genomics System. Another drawback of this large MPS panel containing up to 231 markers is the lack of flexibility. It is not possible to modularize the analysis of markers according to the requirements of the present case, as all markers are targeted in the same amplification. In order to reach sufficient depth of coverage also for low-performing markers, the capacity to pool multiple samples in a single sequencing run is limited. The Promega PowerSeq Auto/Mito/Y system combines 22 autosomal STRs and amelogenin with 23 Y-chromosomal STR loci and the mtDNA control region (Faith and Scheible, 2016). This panel is largely consistent with CODIS and ESS core loci, however, it does not include the SE33 marker (Zeng et al., 2015a). PowerSeq has previously been used for MPS-based population studies (Gaag et al., 2016; Gettings et al., 2016) and paternity testing (Silva et al., 2018). Although developed by a commercial manufacturer, currently this panel is only available as a prototype version. Since Promega does not provide bioinformatics support, users must rely on open-access genotyping tools or commercial software for STR profiling. To date, only two in-house developed identity panels for the MiSeq platform have been published. The multiplex kit presented by Kim et al. (2016) targets 17 autosomal STRs and amelogenin. It uses small amplicons to improve STR profiles from degraded DNA samples. However, the selection of markers is not fully compatible with the European markers, since the polymorphic and highly discriminative STR systems D12S391 and SE33 are missing. The monSTR identity panel (Ganschow et al., 2019; part of this dissertation) covers 21 European core loci, including SE33, with medium-sized amplicons. The analytical performance of this assay was systematically optimized and thoroughly validated (Silvery et al., 2020). Data generated with the monSTR panel can be readily analyzed using the toaSTR genotyping software (Ganschow et al., 2018; part of this dissertation).

1.3 Developing an MPS-STR identity panel

The probability of identical alleles in two individuals decreases with the number of polymorphic loci examined (Butler, 2014). To generate an STR profile with a sufficiently low random match probability (i.e., the probability that a person sampled randomly from the population would have a particular STR profile), genotypes of multiple STR markers have to be combined. However, as biological evidence may be limited in quantity, it would not be appropriate to analyze each STR marker separately. Throughout this dissertation, the term **identity panel** will refer to a laboratory assay supporting the simultaneous analysis of multiple STR markers by means of MPS. The input of an identity panel is a purified DNA sample. The panel itself includes the co-amplification of STRs via multiplex PCR (mPCR), the preparation of amplification products for sequencing (library preparation), and the actual sequencing reactions. Subsequently, sequence data obtained with an identity panel will be subjected to bioinformatics software that transforms the sequence data into an STR profile (Figure 5).

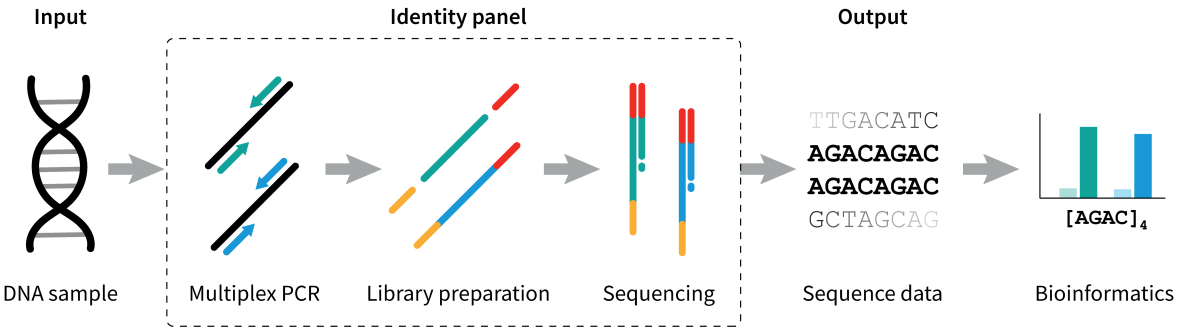


Figure 5. Components of a forensic identity panel. Genetic targets (blue, green) in a purified DNA sample are simultaneously amplified using multiplex PCR. A sequencing library is prepared by attaching adapters (red, yellow) to the amplification products. Sequencing of libraries yields extensive sequence data, which are transformed into an STR profile by bioinformatics software.

Each component in this process, as well as interactions between components, may impact the genotyping result. The following subsections will address important aspects of developing a robust identity panel. The topics discussed here include the construction of an mPCR assay and the optimization of critical assay parameters. In this context, the Design of Experiments methodology will be introduced as a systematic approach for optimizing multifactorial systems. Moreover, this section is dedicated to library preparation strategies and sequencing modes as well as the developmental validation of forensic DNA typing methods.

1.3.1 Multiplex PCR construction

mPCR (Chamberlain et al., 1988) is a variant of PCR in which two or more genetic loci are simultaneously amplified in a single reaction. Multiplex assays contain two to several hundreds of primer pairs targeting multiple regions of DNA. This technique is fundamental to conventional CE-based methods (Butler, 2005; Gibson-Daw et al., 2018; Shafique et al., 2016) as well as advanced STR typing using MPS (Fordyce et al., 2015; Kim et al., 2016; Müller et al., 2018; Zeng et al., 2015b). mPCR has greatly enabled forensic DNA analysis since very small amounts of DNA material may be used to obtain genotypes from multiple markers. The amount of labor required is reduced by typing loci in parallel rather than sequentially. Additionally, multiplex amplification can reveal false-negative reactions since each amplicon provides an internal control for the other amplified fragments (Edwards and Gibbs, 1994).

A multitude of parameters can affect the performance of an mPCR assay (Wallin et al., 2002). In order to achieve a robust amplification, an assay must be developed with overall consideration for the target region sequences, intended fragment sizes, dynamics of primers and other reaction components, and thermocycling conditions. Primary performance goals of a forensic multiplex assay would include:

(1) specific amplification of targeted loci, while allowing a certain tolerance for genotypic variation like mutations in primer binding sites;

(2) balanced amplification, hence equal depth of coverage across all loci (interlocus balance) and locus alleles (heterozygous balance); and

(3) low tendency to form stutter products and other technical artefacts (also referred to as noise) due to polymerase slippage and polymerase error, which can confuse result interpretation.

1.3.1.1 The stutter phenomenon

A phenomenon called **stutter** is a commonly encountered artefact in PCR-based analysis of STR loci and therefore will be briefly explained here. Amplification of repetitive DNA is highly prone to slippage errors during the polymerase-mediated replication (Bovo et al., 1999; Schlötterer and Tautz, 1992). The mechanism that is proposed to introduce this noise during *in vitro* amplification is also considered to be the principle of microsatellite instability *in vivo* (see Section 1.2.1). In the extension phase of PCR, the polymerase may fall off the primer-template complex, followed by dissociation of the template and nascent

DNA strands. Subsequent re-alignment of the double-strand may result in the insertion (forward stutter) or deletion (backward stutter) of one or more repeat units, respectively. Stutter has been well characterized for CE (Ludeman et al., 2018) and MPS (Aponte et al., 2015) workflows. Stutter products are most commonly observed in the N-1 position (loss of one repeat unit) of the parent allele being amplified and may be seen more rarely in the positions N-2 (loss of two repeat units) and N+1 (gain of one repeat unit) (Walsh et al., 1996). Therefore, the *in vitro* ratio of backward stutter was suggested to be much higher than that of forward stutter (Shinde et al., 2003). Due to the higher analytical sensitivity of MPS, sequence data occasionally display stutter products at the N-3 position (Aponte et al., 2015), while these fragments typically remain indistinguishable from background noise with CE.

The stutter ratio, i.e. the stutter percentage relative to the parent allele, has been attributed to the locus, the repeat motif, and the repeat structure (Aponte et al., 2015). It has been hypothesized that the stutter ratio increases with the number of repeat units and correlates with the length of long uninterrupted repeat stretches within an STR region (Brookes et al., 2012; Klintschar and Wiegand, 2003; Walsh et al., 1996). The longest uninterrupted stretch of repeats and adaptions on this concept have shown to be a better predictor of stutter ratio than the overall length of an allele (Bright et al., 2014; Gaag et al., 2016; Vilsen et al., 2018). Woerner et al. (2017a) investigated the effect of flanking region variation on the stutter ratio and found that the flanking haplotype may substantially impact the stutter ratio as well. Commonly, N-1 stutter ratios < 15 % are observed with CE, while values for MPS are similar (Gaag et al., 2016) or slightly higher (Aponte et al., 2015). Pentanucleotide loci exhibit lower levels of stutter than tetra- or trinucleotides (Bacher and Schumm, 1998).

Stutter products originating from the analytical scheme can be easily confused with true alleles and vice versa. In unbalanced mixtures, alleles of a minor contributor may overlap with stutter positions of the main component, which can complicate the interpretation of an STR profile. As stutter products are indistinguishable from biological alleles by length or by sequence alone, multiplex development should aim at finding reaction conditions where stutter formation is minimized (Olejniczak and Krzyzosiak, 2006; Seo et al., 2014, 2012; Walsh et al., 1996). Additionally, recent bioinformatics methods (Ganschow et al., 2018; part of this dissertation; Hoogenboom et al., 2017) attempt to predict the formation of stutter in order to filter or label corresponding reads in MPS data.

1.3.1.2 Primer engineering

Careful primer engineering is essential for the successful amplification of targets. Multiplexing does not only require a high target specificity, but also a set of primers with compatible thermodynamics properties including melting temperature, length, and GC content. Primers are designed to bind at unique, conserved sequences surrounding the repeat regions. Specific binding is important to avoid amplification of similar sequences. A mismatch in the primer binding site can result in a strongly imbalanced amplification or an allele dropout, depending on whether the mismatch occurs in the center of the primer or at the 3' end, respectively (Butler, 2014). Moreover, each additional primer in the reaction increases the potential of undesired primer interferences such as the formation of dimers, which deplete PCR reagent resources. Upfront bioinformatics can assist in the correct selection and validation of primers prior to expensive wet-lab experiments. Detailed sequence information for flanking regions of commonly used STR loci is available (Phillips et al., 2018). Automated primer designing tools (Kalendar et al., 2017; Shen et al., 2010) have been developed, which support the specific needs of multiplex primer design, such as handling multiple templates simultaneously.

Amplicon lengths need to be compatible with the read lengths supported by the respective MPS platform. Likewise, amplicons must not only include the pure repeat tract but also upstream and downstream flanking regions in order to uniquely assign reads to STR loci. Moreover, "including flanking regions in the analyzed sequence data will add to allelic diversity, could aid kinship interpretation, and may improve our understanding of mutational events and evolutionary history at these loci", as Gettings and colleagues (2015a) stated. Restrictions on primer placement due to unfavorable GC content or repetitive elements within the flanking regions may additionally require larger amplicon sizes. Recent increases in read length of MPS machines provide a greater degree of flexibility in amplicon design and allow to generate informative reads also for STR systems with larger alleles. The leading benchtop MPS platforms, Ion S5 (Thermo Fisher Scientific) and MiSeq (Illumina), currently reach read lengths of up to 600 bp (Illumina, 2018; Thermo Fisher Scientific, 2018).

When a set of candidate primers was designed, it is appropriate to perform a computer simulation of PCR. For a given set of primers and a DNA template, *in silico* PCR, also termed virtual PCR or ePCR, predicts which PCR products are synthesized (Yu and Zhang, 2011). Several tools are available, including Primer-BLAST (Ye et al., 2012), UCSC In-Silico PCR (Kuhn et al., 2013), and MFEprimer (Qu et al., 2012). MFEprimer is suitable for mPCR

analysis, accepting a large number of primers to check against a whole-genome sequence. The program determines which (cross-primer) amplicons are likely to be amplified and estimates the formation of primer-dimers and hairpin structures. Although greatly enabling economic primer design, *in silico* PCR does not replace *in vitro* testing of primers to verify an efficient and specific generation of PCR products. Empirical titration of primer concentrations is essential to obtain a balanced co-amplification of targets.

1.3.1.3 Reaction components

Along with primer design, also the PCR protocol plays an important role in a successful mPCR. The choice of DNA polymerase and buffer components greatly influence the yield, specificity, uniformity of amplification, and stutter formation (McPherson and Møller, 2006). During PCR, the DNA-dependent DNA polymerase catalyzes the synthesis of DNA in 5' → 3' direction. Several recombinant thermostable enzymes are available and have been extensively tested (McInerney et al., 2014; Olejniczak and Krzyzosiak, 2006). Polymerases can be compared in their properties such as speed (extension rate), fidelity (error rate), proofreading (3' → 5' exonuclease) activity, strand displacement (5' → 3' exonuclease) ability, and processivity. Here, processivity is a "measure of the affinity of the enzyme for the template strand. The stronger the interaction, the more processive the polymerase should be, and so the more DNA it will synthesize before it dissociates from the template" (McPherson and Møller, 2006). It has been predicted that higher processivity would allow less opportunity for a dissociation of the DNA strands during PCR and thus reduce the formation of stutter (Murray et al., 1993). However, the obtained results turned out to be inconclusive. Wu et al. (1998) used a mixture of *Taq* (*Thermus aquaticus*) DNA polymerase and *Pwo* (*Pyrococcus woesei*) DNA polymerase to amplify trinucleotide loci and explained the reduced occurrence of stutter products by the higher processivity of this polymerase blend. Conversely, the archaeal DNA polymerase KOD (*Pyrococcus kodakaraensis*), showing a processivity > 300 (Takagi et al., 1997), revealed an even higher generation of stutter (Giese et al., 2009; Seo et al., 2014). In a study investigating various polymerases with or without proofreading activity, Walsh et al. (1996) found that none of the enzymes resulted in an improved stutter ratio compared to *Taq*, which provides no proofreading capability. This appears plausible, since slipped strand mispairing, which causes stutter, still preserves a correct base pairing and should not trigger the excision of an incorrect nucleotide. Similarly, Olejniczak and Krzyzosiak (2006) described comparable stutter ratios across a wide spectrum of polymerases that differ in processivity, fidelity and nucleolytic activity. Viguera et al. (2001) reported that polymerases exhibiting a high

strand displacement activity, such as phage Φ29 DNA polymerase and *Bst* (*Bacillus stearothermophilus*) DNA polymerase, can proceed through a hairpin that is formed upon strand mispairing and does not slip. However, these enzymes are no viable solution to prevent stutter since their thermolability is not compatible with the thermal conditions of PCR. The Deep $Vent_R$ DNA polymerase with increased strand displacement activity is thermostable but was found to cause substantial allele dropout (Seo et al., 2014).

When many loci are simultaneously amplified, all products compete for the same limited pool of supplies such as enzyme and nucleotides. The dNTP requirement generally increases with the number of amplicons in the multiplex (Edwards and Gibbs, 1994). Hence, a proportional $MgCl_2$ concentration needs to be maintained since the DNA polymerase requires magnesium ions as a cofactor (besides the magnesium bound by the dNTPs and the DNA molecules). Varying the concentrations of polymerase, dNTPs, and Mg^{2+} ions have been found to significantly influence the total yield of PCR products but did not affect the stutter ratio (Olejniczak and Krzyzosiak, 2006).

The use of adjuvants in the reaction buffer may improve the amplification efficiency by inhibiting the formation of secondary structures, stabilizing the polymerase/template complex, or blocking matrix proteins, respectively (Henegariu et al., 1997). Popular enhancers are glycerol, formamide, betaine, bovine serum albumin (BSA), and dimethyl sulfoxide (DMSO). While formamide and betaine did not affect the stutter ratio in a study of Olejniczak and Krzyzosiak (2006), the presence of DMSO in the PCR mix has been linked to an increase in the formation of stutter products. Adding single-stranded DNA binding protein (SSB), which can inhibit polymerase slippage by stimulating the strand displacement activity (Viguera et al., 2001), and the macromolecular crowding agent polyethylene glycol (PEG) were found to decrease the stutter ratio (Seo et al., 2012).

1.3.1.4 Thermocycling conditions

In a comprehensive evaluation of thermocycling conditions, Henegariu et al. (1997) described the annealing temperature to be one of the most critical parameters influencing the specificity and efficiency of primer binding to the template DNA. Lowering the annealing temperature compared to singleplex conditions was required for co-amplification in multiplex mixtures. A potential risk for unspecific amplification at lower annealing temperatures may be overcome by the concurrent situation in the multiplex reaction. Generally, extension times should be increased with the number of loci amplified in the reaction to allow for the complete synthesis of amplicons by the DNA polymerase (Edwards and Gibbs, 1994). However, long extension and annealing times could again

provide an opportunity for unspecific amplification. Concerning the formation of stutter, Seo et al. (2014) hypothesized that "a lower temperature annealing/extension step reduces the likelihood of slippage during PCR by enhancing the stability of the DNA polymerase/template DNA complex or the stability of the generated duplex". A suppression of stutter products was also observed with a decreased denaturation temperature of 85 °C (Olejniczak and Krzyzosiak, 2006). A possible explanation for this might be that incompletely synthesized DNA strands will not dissociate upon polymerase slippage. Thus, the opportunity for slipped-strand misalignment will be reduced when the polymerase re-associates with the template strand and continues the synthesis of the new strand. While a certain number of PCR cycles are required to reach sufficient analytical sensitivity, extensive PCR cycling has been associated with an increased frequency of stutter products: The study of Bovo and colleagues (1999) supported the hypothesis that stutter is generated from improper reannealing of PCR products after PCR efficiency has reached the amplification plateau.

1.3.2 Library preparation and sequencing modes

The library preparation strategy may greatly influence genotyping results. Previous studies found that the downstream allele detection capability is dependent on "sequence read length and library preparation chemistry" (Warshauer et al., 2013), as unfavorable methods relying on enzymatic fragmentation may produce sequence reads that only partially span the repeat region and hence are not informational for allele calling. The effect of purification methods between library preparation steps has also been investigated (Riman et al., 2017). Column-based DNA purification was reported to increase the number of locus and allele dropout events as compared to the bead-based cleanup method. Commonly, adapter oligonucleotides are attached to the DNA fragments either by PCR or by enzymatic ligation. For example, Verogen ForenSeq libraries are generated in a two-step PCR procedure utilizing the Illumina Nextera XT DNA Library Preparation Kit (Illumina, 2015): During mPCR, forensic targets are co-amplified using primers with target-specific sequences and universal overhangs at the 3' ends; After bead-based clean-up, a second PCR is performed to attach barcoded adapter sequences using primers that are partially complementary to the overhangs introduced in the first PCR. Alternatively, library preparation with the Promega PowerSeq kit uses a combination of target-specific mPCR and enzymatic ligation of Illumina TruSeq adapters to the amplicons (Faith and Scheible, 2016). With the two-PCR procedure, it is possible to perform directional sequencing of only one DNA strand, while with the ligation strategy both strands will be

sequenced because the orientation in which adapters ligate is not directional (Børsting and Morling, 2015). In the latter case, the genotyping software must account for reverse-complementary reads to respect the strand notation when reporting sequence alleles.

Choosing the right sequencing mode is also important to obtain reliable genotyping results. Illumina sequencers support both single-end and paired-end sequencing. With the latter mode, DNA fragments are sequenced from both ends, generating a pair of reads representing both extremities of the same DNA fragment. Depending on the read length and fragment size, the 3' ends of paired reads overlap and can be assembled to an informative consensus sequence fully spanning the STR (Bushnell et al., 2017). However, merging of reads containing repetitive 3' ends, as commonly encountered when sequencing STRs, may lead to erroneous or ambiguous consensus sequences or loss of sequence data and hence jeopardize the allelic profile (Ganschow et al., 2018). Single-end sequencing eliminates the need for read merging but requires long reads completely spanning even long alleles. This comes at the cost of lower base qualities towards the end of a read due to the accumulation of base call errors (Edgar and Flyvbjerg, 2015). ForenSeq libraries are sequenced by *de facto* single-end sequencing of 351 x 31 nt (Illumina, 2015). Here, the forward read fully encompasses the amplicon, while the reverse read probably only serves as a control to correctly assign the corresponding fragment to the respective locus. Promega PowerSeq sequence data takes the form of 251 x 251 nt reads and requires merging to generate a consensus of overlapping read pairs (Gaag et al., 2016).

1.3.3 Conventional strategies for assay optimization

The complex interactions among the components of an mPCR assay make it unlikely that a standard set of reaction conditions would be optimal for all situations (Henegariu et al., 1997). Thus, every newly developed PCR application requires optimization of critical process variables to achieve good analytical performance. To add to the complexity, mPCR is only one element of an identity panel; The overall performance of an identity panel must be measured and optimized in its entirety to account for the interferences between mPCR, library preparation, and sequencing. For instance, it can be assumed that mPCR primer concentrations and size-dependent purification steps during library preparation will interact and exert a strong effect on the sequencing coverage of differently sized STR amplicons.

Surprisingly, except for the adjustment of primer concentrations, optimization of global assay conditions was not discussed in recent articles introducing in-house constructed MPS-STR identity panels (Kim et al., 2016; Zhao et al., 2016). However, several

reports on general mPCR development and optimization have been published. Henegariu et al. (1997) provided an extensive discussion of factors that can influence the amplification, including times and temperatures of the PCR protocol, the concentration of reaction components, and the use of adjuvants. mPCR optimization was understood as a step-by-step protocol of practical solutions to overcome commonly encountered problems. The report describes sequential alterations of cycling conditions and reaction components aiming at a higher yield of PCR products and an increase of reaction specificity. However, such a sequential optimization approach makes it difficult to understand potential interactions between factors. Butler (2005) suggested a development strategy focusing on careful primer design and empirical testing. The protocol included the evaluation of primer functionality and specificity as well as performing sensitivity and consistency studies. Optimization of thermocycling conditions was not subject of the discussion. Edwards and Gibbs (1994) emphasized the importance of similar reaction kinetics of all primers in the set. It was recommended to develop PCR conditions separately for each primer pair. The authors proposed to add primer pairs sequentially to the multiplex assay and to alter the conditions as necessary. However, choosing an appropriate experimental design would allow optimization of multiple PCR parameters in parallel rather than serially.

1.3.4 Introduction to Design of Experiments

While every scientific experiment has a design, some designs are more thoughtful than others. With well-designed experiments, it takes less time and effort to gain knowledge about the system under investigation. When designing a series of experiments to investigate the effect of multiple input variables (referred to as factors) on the process performance (response), there are essentially three possible approaches.

Under the common, intuitive approach, the investigator modifies **one factor at a time** (OFAT). It requires varying the first factor until an optimum is found while every other factor is held constant. Based on the optimal value of the first factor, the remaining factors are changed separately in successive experiments (Figure 6A). This method involves only a few measurements; as an example, an optimization series of three factors each at four discrete levels would include 4+3+3=10 experiments (replicates not included). However, this design does not thoroughly explore the space of possible solutions and will most likely lead to locally optimal target values, while the global optimum has not been reached (Pilipauskas, 1999). Moreover, it neglects interactions between factors: An isolated factor

may exert only little influence on the process performance, but rather the interaction of two factors may strongly affect the response.

The **matrix** method sequentially tests every factor against all levels of all other factors in order to include all possible combinations (Figure 6B). Such an exhaustive approach allows interactions between factors to be estimated but is inefficient because the number of experiments increases dramatically if many factors are considered. The optimization series exemplified above (three factors at four levels) would require 4^3=64 separate runs, excluding replicates.

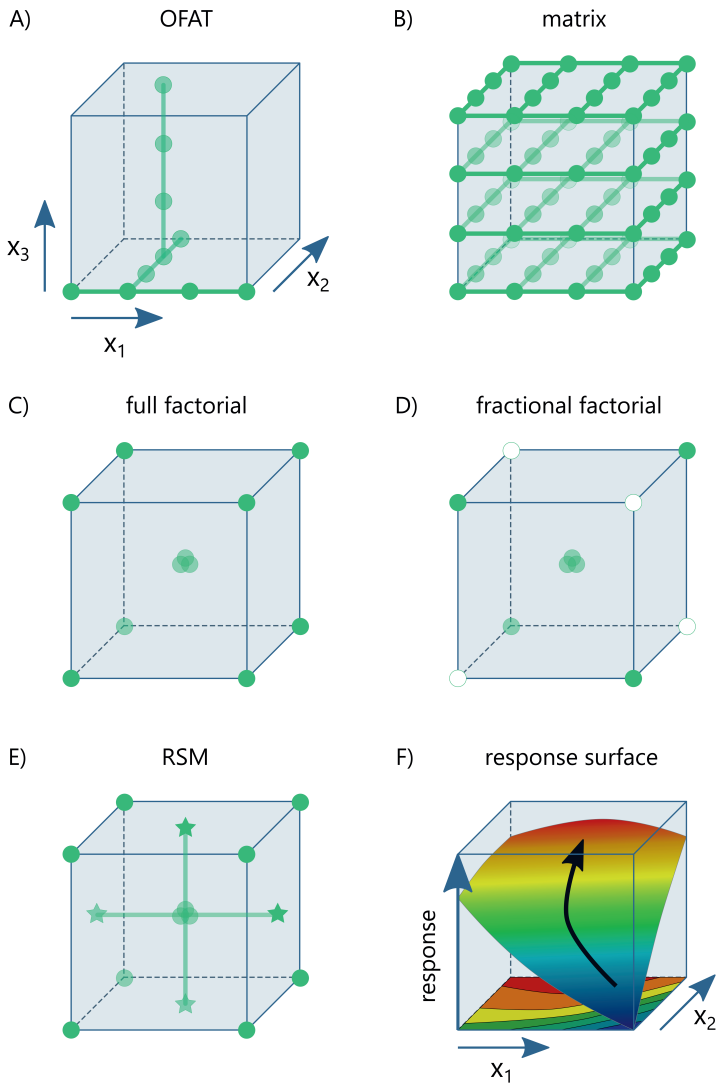


Figure 6. Illustration of the Design of Experiments methodology. A – E: graphic representations of common experimental designs with three factors x1, x2, and x3. F: exemplary response surface plot obtained from a response surface methodology (RSM) design. The response surface represents a map of the system that allows locating optimum conditions depending on variable settings of two factors (for visualization purposes, other factors are fixed at constant levels).

The **Design of Experiments** (DOE) methodology suggests a more sophisticated approach for the optimization of a multifactor process such as mPCR. Briefly, DOE is a framework that assists in planning, performing, and analyzing systematic trials. Statistically designed experiments enable the investigator to rapidly identify significant factors and complex interactions. DOE provides a steep learning curve in the sense that informative data can be obtained with a minimum number of experiments and minimal cost (Haaland, 1989). Experimental design techniques have been used successfully across numerous industries and research fields. First described by Fisher (1947) in an agricultural application, DOE has been widely used in the field of engineering (Ilzarbe et al., 2008). Other reports include applications in biotechnology (Rao et al., 2008; Toms et al., 2017) and molecular biology, where DOE was employed for the optimization of cDNA microarrays (Wrobel et al., 2003), PCR (Boleda et al., 1996; Caetano-Anollés, 1998; Cobb and Clarkson, 1994), mPCR (Dobay et al., 2009; Villarreal Camacho et al., 2013), and real-time PCR (Celani de Souza et al., 2011; Wadle et al., 2015). Application has also been found in the optimization of PCR conditions for microsatellite genotyping in a medical (Niens et al., 2005) and forensic context (Ballantyne et al., 2008). The work of Ganschow et al. (2019; part of this dissertation) was the first study to describe DOE methods in the context of MPS.

Depending on the objective of the study to be performed, different types of geometric experimental designs are available (Eriksson, 2008). Two-level full factorial designs are constructed by testing every factor at two levels. They are useful for early phase screening applications to determine the most significant factors that influence the system. The experimental space covered by a full factorial design in three factors at two levels would be a cube with the experiments on its eight corners (Figure 6C). It also includes a replicated center point in between the high and the low levels to assess the replicate error. The orthogonality of this design (i.e., each factor can be evaluated separately) allows for the estimation of linear effects as well as factor interactions (Altekar et al., 2006). However, such designs can lead to large numbers of experiments as the number of factors increases. Fractional factorial designs (Figure 6D) are balanced subsets, or fractions, of the full factorials. They are suitable for studies where four or more factors are of interest, such as in a robustness test or a screening study. Fractional factorial designs offer a reduction in measurements and still can analyze main effects and lower order interactions (Toms et al., 2017). Once the investigator is confident that the most influential factors and their relevant ranges were identified, an optimization study may be performed to locate the optimum conditions within the experimental region. Appropriate composite designs are part of the response surface methodology (RSM) design family (Figure 6E). They combine factorial

designs with additional levels to obtain precise information about the magnitude, influence, and interactions of the factors, including second-order (quadratic) effects (Eriksson, 2008). RSM designs provide a map of the system in the form of a response surface plot (Figure 6F), allowing to explore the behavior inside the limits of the region studied. Data from these experiments are used to build a predictive mathematical model. The model is a polynomial equation that describes a relationship between input factors and measured response values, weighted by regression coefficients. It allows for the approximation of responses based on a given combination of factor levels and thus is a useful tool to maximize the information available from limited or expensive experimental data (Box, 1954; Box and Draper, 1987).

1.3.5 Developmental validation of forensic DNA typing methods

Forensic DNA analysis methods supporting law enforcement have far-reaching consequences in convicting the guilty and protecting the innocent. In light of the complexity of DNA typing procedures, which have been outlined in the previous sections of this thesis, it is immediately evident that a rigorous quality assurance (QA) strategy is crucial to maintain the effectiveness of investigative tools and to provide confidence in the results obtained. QA measures must be in place at various levels, from the forensic DNA community level down to the individual laboratory and the interpretation of results (Schneider, 2007). At the laboratory level, one QA measure is the validation of the performance of methods and instruments. According to the DIN EN ISO/IEC 17025, which is the standard for forensic laboratory accreditation, validation is defined as "the confirmation by examination and the provision of objective evidence that the particular requirements for a specific intended use are fulfilled" (International Organization for Standardization, 2018). In other words, the primary purpose of validation is to demonstrate the robustness, reliability, and reproducibility of a newly implemented or modified method for forensic DNA testing.

There are two stages of validation – developmental and internal (Butler, 2011). **Developmental validation** is commonly performed by the manufacturer of a novel method such as an identity panel for detecting STR alleles. It involves the acquisition of data to determine the conditions and limitations of this method. **Internal validation** studies are conducted by an individual laboratory to verify that a previously developmentally validated procedure performs as expected.

There is no standardized validation strategy utilized across forensic DNA laboratories. However, numerous coordination efforts exist on a local, national, and international level (Butler, 2011). As one of these institutions, the Scientific Working Group on DNA Analysis Methods (SWGDAM) [3] issues the Validation Guidelines for DNA Analysis Methods. These guidelines are widely accepted in the forensic community and have been periodically revised to adapt to evolving technologies and laboratory practices. The most recent version (Scientific Working Group on DNA Analysis Methods, 2016) also addresses MPS technology and bioinformatics validation topics. SWGDAM guidelines suggest including the following major points for the developmental validation of an STR identity panel and encourage peer-reviewed publication of developmental validation studies.

Characterization of genetic markers

Determine and document the genomic location of the genetic markers, the mode of inheritance, polymorphisms in relevant population groups, and the technological basis for their detection. Collections describing the properties of well-established STR markers (Gettings et al., 2017; Phillips et al., 2018) greatly simplify this task.

Species specificity

Determine whether any genetic information from non-target species interferes with the DNA typing system. Typically, non-human DNA from domestic animals (such as dogs, cats or horses), bacteria, and yeast is tested, which may be prevalent in samples recovered from crime scenes (Iyengar and Hadi, 2014).

Sensitivity study

A dilution series of reference DNA with known genotype is conducted to determine the assay's dynamic range. DNA input quantities for PCR amplification between 1,000 pg and 7.8 pg might be evaluated. Here, the lower limit corresponds approximately with the genome content of a single human cell (6 pg). Low-level DNA, which is common in casework samples, increases the risk of allele imbalances and dropout due to stochastic effects during PCR (i.e., random sampling and amplification).

[3] https://www.swgdam.org/

Repeatability and reproducibility

Demonstration of the precision and accuracy of results when the analysis is performed by the same (repeatability) or among different (reproducibility) operator(s) and detection instrument(s). This study evaluates the comparability and consistency of results across both distance and time, which is essential for example to match a sample from a suspect with a previously sampled DNA database record.

Case-type samples

A DNA testing laboratory often encounters biologic material that contains PCR inhibitors or degraded DNA. The ability to obtain reliable DNA profiles from case-type situations is demonstrated using representative existing samples of closed forensic cases or mock specimens.

Mixture studies

STR identity panels should be able to detect mixed-source samples originating from more than one individual. Typically, two-person mixtures of known genotypes with minor contributions ranging from 50 % to 1 % are prepared. Mixture studies are performed to recover the major and minor contributor profiles and to estimate the limit of detection (LOD) based on read frequencies of the minor component alleles.

PCR-based studies

PCR-based studies involve varying thermal cycling conditions, the concentrations of primers, and other critical reagents to provide the required degree of specificity and robustness. Relevant aspects of mPCR assay development and optimization have been described earlier in this chapter.

MPS-specific studies

Assessment of effects related to the MPS instrumentation and library preparation, if not covered by general MPS operating procedures. Suggested studies address the effects of indexing of samples with subsequent demultiplexing, the extent of pooling libraries and the corresponding LOD, as well as the possibility of intra-run signal crosstalk and inter-run sample carryover.

Bioinformatics

Bioinformatics tools used for allele calling and result interpretation require validation prior to implementation. It is up to the forensic laboratory to determine the appropriate validation studies depending on the function and application of the software to identify its reliability and limitations.

In addition to the major studies outlined above and summarized in Table 3, developmental validation may also include stability studies assessing the effects of various substrates and environmental insults on DNA samples. A general examination of PCR products may evaluate the coverage balance of heterozygous alleles and characterize the observed stutter products (Butler, 2011).

Table 3. Short description of SWGDAM guidelines for the developmental validation of DNA typing methods.

Study	Purpose
Genetic markers	Characterization of genetic markers
Species specificity	Assess non-target species interferences
Sensitivity	Determine the dynamic range
Repeatability and reproducibility	Comparability and consistency of results
Case-type samples	Compatibility with case-type specimen
Mixture	Usability for mixture deconvolution
PCR	Amplification robustness and specificity
MPS	Library preparation and sequencing effects
Bioinformatics	Reliability and limitations of genotyping software

1.4 Bioinformatics analysis of STR sequencing data

1.4.1 Short-read sequence data poses a challenge to analyze STRs

Genotyping microsatellite repeats is fundamentally distinct from calling single nucleotide variations or InDels in non-repetitive regions of the genome. Due to the highly polymorphic nature of STRs including structural and sequence variations, the source DNA is virtually never identical to the reference. Nonetheless, allele calling must comply with a restrictive nomenclature to ensure the comparability of results obtained with different technologies. As Highnam et al. (2013) described, “microsatellite genotypes must be

assigned in terms of allele length or the number of sequenced bases within a read separating the non-repetitive flanking boundaries aligned to the reference, irrespective of intervening alignment gaps". Therefore, profiling STR variation poses remarkable challenges to mainstream aligners such as BWA (Burrows-Wheeler Aligner; Li and Durbin, 2009) or Bowtie (Langmead and Salzberg, 2012). First, STRs with similar or identical repeat sequences are found at many locations in the genome; hence, reads containing repetitive sequences can create ambiguities in the alignment and are susceptible to misplacement, which in turn can bias the interpretation of results (Treangen and Salzberg, 2011). Second, allelic expansions and contractions present a gapped alignment problem. Since mainstream aligners exhibit a trade-off between runtime and the tolerance to InDels (Li and Homer, 2010), high accuracy can lead to lengthy processing times. Third, non-informative reads that only partly encompass the repeat tract may confound downstream allele calling. Reads used for allelotyping must span the entire repeat tract plus some unique flanking region to confidently determine the allele length. Fourth, mainstream variant calling pipelines are not suitable to characterize complex repeat sequences and typically do not report repeat genotypes at all. Last, they fail to account for stutter noise. Stutter products are incorporated in PCR-based library preparation and require modeling of noise patterns in order to distinguish erroneous amplicons from true alleles and prevent over-interpretation of MPS data (Gymrek and Erlich, 2013).

1.4.2 Landscape of open-access tools for MPS-STR genotyping

Software developers have been working on various strategies to process MPS-STR data and provide added value over mainstream variant callers (Liu and Harbison, 2018). Generally, current tools fall into two categories: alignment approaches and sequence-search tools. Briefly, alignment-based tools perform a selective alignment or realignment of flanking regions to map STR loci, while sequence-search tools search for anchor sequences located in the conserved flanking regions. Open-access genotyping tools most relevant to forensic practice will be introduced in the following.

1.4.2.1 Alignment-based STR genotypers

lobSTR (Gymrek et al., 2012) is a command-line tool for STR profiling in whole genome sequencing datasets. Using an indexed reference genome and a BED file containing the set of markers to be analyzed, lobSTR aligns informative reads that fully encompass an STR locus. The alignment step internally uses extensive code parts of the short-read aligner BWA (Li and Durbin, 2009). Non-repetitive flanking regions are mapped to the genome to

characterize the STR position and length. Allele calling is based on the number of repeat units between the two flanking regions. To assess the genotype likelihood, lobSTR employs a locus-specific logistic regression model of stutter noise during the allele calling step. The model is built on haploid training data specific to the respective MPS platform and involves various sequence features including the size of the repeat motif, the length of the repeat tract, GC content of flanking regions, and SNPs within the STR (Gymrek et al., 2012). Alleles are given as the number of base pairs length difference from the reference sequence (e.g., -12/+4). Therefore, it is necessary to convert lobSTR results to the standard forensic notation given in the number of repeat units by using an external tool. Results do not provide information on the allelic sequence and structure. Visualization of STR sequence alignments is not part of the software. In a comparative analysis of STR profiling tools, Anvar et al. (2014) observed increasingly problematic alignments of lobSTR with more complex STR loci such as D13S317 and TH01. Although lobSTR performed well on genotyping diploid samples in the abovementioned study, the algorithm failed to accurately detect samples with multiple contributors. Considering the described limitations and the lacking graphical user interface (GUI), lobSTR may be regarded rather as part of an analysis pipeline than a full-featured tool for routine forensic genotyping.

RepeatSeq (Highnam et al., 2013) calls microsatellite repeat genotypes from an existing alignment of whole-genome resequencing data. Reads that map to a repeat region reference sequence are locally realigned around predefined repeat regions, while reads that do not completely span the repeat region are discarded. RepeatSeq estimates the probability of candidate genotypes using an empirically derived Bayesian error model based on findings of a previous population-scale analysis of microsatellite repeats in *D. melanogaster* (Fondon et al., 2012). RepeatSeq was designed to be integrated into existing genotyping pipelines and does not support a GUI or visualization of results. Therefore, it appears less suitable for direct routine application.

The command-line tool **STR Viper** (Cao et al., 2014) accepts an alignment in BAM/SAM format and analyses repeat-length variation from paired-end sequence read data. It uses the distance between linked reads that align in the flanking regions to infer an expansion or contraction of repeat units.

1.4.2.2 Sequence-search STR genotypers

Current legal requirements, as well as statistical, computational, ethical, economic and technical validation issues, prohibit genome-wide analyses for forensic purposes (Amorim and Pinto, 2018). As the aforementioned alignment-based tools are meant to work on

whole-genome data, they may be out of scope for forensic casework. Alignment tools may map a read to an incorrect location due to sequence similarities between repeat stretches of different STRs (Liu and Harbison, 2018). In response to these limitations, a class of sequence-search tools has been developed. Their allele calling algorithms are conceptually similar (Anvar et al., 2014; Bailey et al., 2017; Friis et al., 2016; Lee et al., 2017; Van Neste et al., 2014; Warshauer et al., 2013). Briefly, target reads are captured using an approximate pairwise search for anchoring sequences in the 5‘- and 3‘-flanking regions. Anchors are placed near or adjacent to the repeat region and have a typical length of 12 to 25 bp. Allowing inexact matches accounts for single-base substitutions within the anchor sequences. However, bioinformatics-related dropouts or null alleles may be experienced when an allele that is present in the raw sequence data is not detected due to InDels or multiple SNPs in the anchor sequence. If a read encompasses the full repeat region and contains these anchor sequences in the correct orientation, it is assigned to the respective locus. Next, the repeat sequence is isolated and clusters are generated based on sequence similarities. Read counts of a cluster can be used as a measure of allele abundance similar to fluorescence intensities in CE-based methods. Sequence alleles may be reduced to a length-based notation for comparison with results obtained by CE. However, sequence-search tools may not include the full amplicon length for genotyping, resulting in InDels not being counted as they would with CE if an InDel resides outside of the region that is spanned by the anchors.

STRait Razor (Warshauer et al., 2013) has been one of the first implementations of the sequence-search approach and is constantly further developed (Woerner et al., 2017b). This command-line application outputs alleles, sequences, and read counts to plain text files. Conversion of results into final genotypes requires copy-pasting the primary output into Excel workbooks provided with this software package. Only known STR sequence alleles are properly converted into the compressed bracketed notation using a lookup table that maps raw repeats and formatted sequences. Result interpretation such as automatic identification of stutter products is currently not supported by this software.

MyFLq (My-Forensic-Loci-queries) (Van Neste et al., 2014) consists of a MySQL database backend that holds known reference allele sequences, and a Python frontend to provide an interface for adding reference alleles to the database and analyzing MPS data. The program requires building a reference allele database from a training data set. For each reference allele, the algorithm determines the variable repeat region of a particular locus. Reads from an unknown sample are assigned to a locus based on the presence of both PCR primers as anchor sequences. Subsequently, identical reads are grouped, and alleles are

called by comparison of the repeat region of a group to the reference database. MyFLq has been implemented as a standalone web application and Illumina BaseSpace application, respectively. The GUI allows setting abundance thresholds and provides locus statistics such as frequencies and sequence similarities of allele candidates to help identify erroneous sequences.

The Python script **STRinNGS** (Friis et al., 2016) identifies sequence reads with STR loci in BAM or FASTQ files. It automatically reports STR allele names by length (i.e., the length of the repeat region divided by the length of the repeat unit) and parses the repeat structure. Additionally, STRinNGS can detect variation in the flanking regions. The current version of this software does not support locus-specific analytic thresholds or stutter modeling.

FDSTools (Hoogenboom et al., 2017) is a software package that is able to characterize and correct noise in the sequence data caused by stutter or other errors during amplification and sequencing. Allelic variants are detected and characterized by the TSSV tool (Anvar et al., 2014), which relies on adjacent anchor sequences flanking a target locus. With FDSTools, allele-specific noise profiles are built from a database of reference samples. The stutter model seeks to fit a polynomial function representing the relationship between the number of consecutive repeat units and the stutter ratio. When applied to case samples, read counts are corrected according to the previously generated error profiles. Results are visualized in interactive graphs.

Altius (Bailey et al., 2017) is a web application hosted in an AWS (Amazon Web Services) cloud environment. The implementation as a cloud-based service supports data storage, on-demand auto-scaling of computational power, and a multi-layered security concept including user access control, firewalls, and encryption of data. Altius allows for cross-platform usage since the user interacts with this software through browser-based GUI. Sequence allele nomenclature follows current ISFG considerations. However, Altius supports neither stutter modeling nor classification of results.

The web application **SeqMapper** (Lee et al., 2017) requires a user-supplied list of reference alleles. Mapping of reads to STR loci is supported at different levels of stringency such as matching the complete reference allele sequence, the repeat sequence or the primers. SeqMapper reports alleles matching the database and reads that could not be called for any of the analyzed loci. Result visualization, formatting of sequences, and stutter modeling are not supported by this software.

2 Aims of this work

MPS will play an emerging role in forensic science and casework. STR sequencing enables characterizing biological evidence in previously unimagined detail. However, some issues still need to be addressed to encourage greater acceptance of MPS within the forensic community. Currently, there exists no economic identity panel for routine STR sequencing on the Illumina MiSeq platform. Available commercial kits may appear over-sized for casework applications and require dedicated instrumentation. Moreover, the independent analysis of forensic MPS data remains an obstacle for DNA laboratories with limited bioinformatics expertise. Flexible open-access tools exist, however, drawbacks regarding their usability become apparent from a routine perspective.

This dissertation has two primary aims, each of which includes important milestones. First, the present work seeks to develop a bioinformatics tool that translates STR sequencing data into genetic profiles. It is intended to provide a user-friendly, platform-independent, and panel-agnostic software that generates valid genotypes and assists in data interpretation. This task includes the development and implementation of algorithms for locus-specific allele calling and sequence formatting according to the current forensic nomenclature. The herein proposed stutter model attempts to predict and identify the most frequent stutter variants and to enable automated allele calling.

The second objective of this dissertation is to develop a medium-sized human identity panel for the Illumina MiSeq platform, targeting 21 forensically important STR markers including the highly discriminative SE33 locus. This is the first work to introduce the Design of Experiments methodology for the optimization of critical process parameters of an MPS-STR assay. Systematic optimization studies aim at maximizing the assay performance with respect to amplification specificity, interlocus balance, and reduction of technical artefacts such as stutter. Developmental validation studies according to SWGDAM guidelines explore the capabilities and limitations of this novel identity panel.

3 Results and discussion

This chapter presents an overview of the three individual research articles contributing to this cumulative dissertation. Along with a discussion of their thematic relationship, this chapter provides supplementary information which has not been published yet. The first article, "toaSTR: A web application for forensic STR genotyping by massively parallel sequencing", introduces a novel software for STR genotyping based on MPS data. The toaSTR software stands out from other forensic tools: Its intuitive handling simplifies access to bioinformatics data analysis. The innovative sequence-based stutter model predicts and identifies common stutter variants and assists with the interpretation of STR profiles. Multiple studies evaluate the usability of this software by analyzing sequence data generated with both commercial and custom identity panels on common MPS platforms. Manuscripts two and three are titled "Development of a multiplex forensic identity panel for massively parallel sequencing and its systematic optimization using Design of Experiments" and "Developmental validation of the monSTR identity panel, a forensic STR multiplex assay for massively parallel sequencing". They cover the developmental process of a custom forensic MPS-STR assay – monSTR – from primer engineering through the optimization of thermocycling conditions and the validation of the assay performance according to established forensic guidelines. The highlight of this thematic complex is the transfer of the Design of Experiments methodology to MPS, enabling an experimentally practical and economically justifiable assay optimization. For bibliographic information and specimen copies of each paper, the reader is kindly referred to the Publications section of this thesis.

3.1 toaSTR — a web application for MPS-STR genotyping

As an accredited laboratory for forensics and routine diagnostics, LABCON-OWL attaches great importance to the validity and practicability of analytic procedures. Already at an early stage of implementing MPS for DNA testing in our laboratory, it was realized that mastering data analysis will be one of the major challenges to promote the routine application of STR sequencing. As was pointed out in the Introduction, existing open-access STR genotyping tools did not satisfy our functional and usability requirements since these programs often feature counterintuitive command-line interfaces and lack

assistance in data interpretation. Additionally, a European survey on forensic applications of MPS (Alonso et al., 2017) emphasized the need for an overarching software that is independent of the MPS platform. Therefore, the first aim of this thesis was to develop a novel software for analyzing STRs in MPS data, which addresses these limitations. The following sections describe the developmental process of the toaSTR genotyping software (Ganschow et al., 2018), including the specification of software requirements and the resulting software design.

3.1.1 Software requirements specification

The toaSTR software is intended to facilitate DNA typing based on targeted MPS data, integrating the analysis of STR loci and the reporting of STR profiles. This software is a novel approach to simplify the access to MPS-STR genotyping for investigators who do not want to delve deeper into bioinformatics. The developmental goals included to deliver a user-oriented, versatile tool with routine use in mind. Data analysis should support sequence data of popular MPS platforms generated with both commercial and in-house identity panels. Genotyping results are to be prepared for optimal human readability and fast interpretation without neglecting the high information density of MPS data. The reported STR profiles must be consistent with current nomenclature considerations to enable comparison with results obtained with parallel or orthogonal technologies.

In a typical use case of this software, a forensic geneticist has performed targeted sequencing of STRs on the Ion Torrent or Illumina platform and wants to transform the sequence data into STR profiles. The typical user has no extensive bioinformatics background and is therefore not able or willing to use neither CLI-based nor overly complex genotyping tools. The processed sample might originate from the DNA of a single donor or multiple contributors. Sample preparation may have been carried out using a commercial identity panel like the Verogen ForenSeq kit or an in-house developed custom panel. Demultiplexed sequence data are available either in FASTA or FASTQ format. The user provides the data file and selects the STR loci to be analyzed. After the software has analyzed the sequence data, the user reviews the called length-based and sequence-based alleles for each locus. Finally, the STR profile is stored in the sample management system or can be exported for external processing.

Based on this use scenario, a software requirements specification (SRS) was derived (Table 4), which documents the requirements that must be met by the software to be developed. The structure of the SRS is adapted on the IEEE Recommended Practice for Software Requirements Specifications, IEEE 830-1998 (IEEE, 1998).

Table 4. Software requirements specification for the toaSTR genotyping software.

Functional requirements	F01	Accept and validate short read sequence data in FASTA and FASTQ formats
	F02	Enable annotation of samples: sample name, notes, sample type
	F03	Sample management enabling to create, read, update, and delete analyses
	F04	Ability to configure custom sets of STR loci representing the composition of the identity panel to be analyzed
	F05	Use a sequence-search approach to map reads to STR loci
	F06	Detection and extraction of the repeat region of mapped reads
	F07	Reference-free, length-based analysis of alleles with the assignment of CE-concordant numerical allele designations
	F08	Reference-free formatting of the repeat region in a compressed bracketed notation
	F09	Reporting of genotypes according to the current nomenclature considerations of the ISFG
	F10	Automated classification of observations as allele, stutter, or artefact (i.e., non-stutter noise)
	F11	Ability to analyze single-source and multiple-source samples without *a priori* assumptions on the sample composition
	F12	Interactive histogram plots to visualize sequence observations and display annotations
	F13	Option to export results in CSV, XLSX, and PDF formats
Non-functional requirements	N01	Compatibility with sequence data of Ion Torrent and Illumina platforms
	N02	Compatibility with sequence data of commercial and in-house identity panels with minimum requirements on the amplicon design
	N03	Cross-platform compatibility with various operating systems such as Windows, Mac OS, Ubuntu, Android, and iOS
	N04	Cross-device accessibility of the software, supporting different device classes such as desktop computers, tablets, and smartphones
	N05	Intuitive usability of the software with a guided workflow and a graphical user interface (GUI)
	N06	Multi-user environment with private data compartments for every user
	N07	Security concept ensuring the integrity and privacy of data
	N08	Administration layer for managing the application and its users
	N09	Provide the software to the forensic community as an open-access tool
Database requirements	D01	Persistent storage and retrieval of input data files
	D02	Persistent storage and retrieval of marker panels and genotyping results
Performance requirements	P01	No extensive user-supplied computation power required
	P02	Typical analysis speed <1 min per 100k reads
	P03	Non-blocking application flow, i.e. the software is not blocked while an analysis is in progress

3.1.2 Software design description

The software requirements stated in the previous section were translated into a representation of software components, interfaces, and data. This software design description (SDD) is based on the IEEE Standard for Software Design Descriptions 1016-2009 (IEEE, 2009) and presents how the software is structured to completely satisfy the aforementioned requirements. Superscript identifiers ([F00], [N00], [D00], [P00]) refer to the specifications given in Table 4.

3.1.2.1 Core algorithms design

The analytic capabilities of the toaSTR software build upon four main bioinformatics algorithms. These include (1) detection of STRs in MPS data, (2) repeat region analysis, (3) stutter modeling, and (4) classification of observations.

STR detection

The STR detection algorithm implemented in toaSTR follows the sequence-search approach [F05]. STR detection is based on the approximate search of a pair of recognition elements (REs) in FASTA or FASTQ [F01] formatted reads (Figure 7). These locus-specific REs are oligonucleotide sequences located in the upstream and downstream flanking regions directly adjacent to the repeat region. Lengths of REs vary between 14 and 25 nt, depending on the sequence complexity of the flanking regions and the typically amplified flanking lengths in commercial identity panels [N01, N02]. The string-matching algorithm allows inexact matches (i.e., a Hamming distance of $d_H = 2$ per RE) in order to tolerate nucleotide substitutions while maintaining a high locus specificity. If a pair of REs are found on a read in a fixed order, the repeat region between them is extracted [F06] and validated based on the presence of the core repeat motif(s) of the respective STR locus. This validation step removes irrelevant sequence data that may have been captured due to flanking sequence homology. If necessary, reads are reverse complemented to meet forensic sequence orientation conventions. In the next step, reads with identical repeat regions are clustered to so-called observations. This neutral term, observation, will be used throughout the process as long as the cluster has not been classified. Low-abundant observations with a cluster size lower than the user-defined analytical threshold are considered background noise arising from stochastic amplification or sequencing errors. These observations are discarded to prevent the over-interpretation of data. The default analytical threshold of ten

reads is in consistence with the findings of Young et al. (2017), who considered this value generally sufficient for STR loci sequenced with a coverage of up to 5,000 reads.

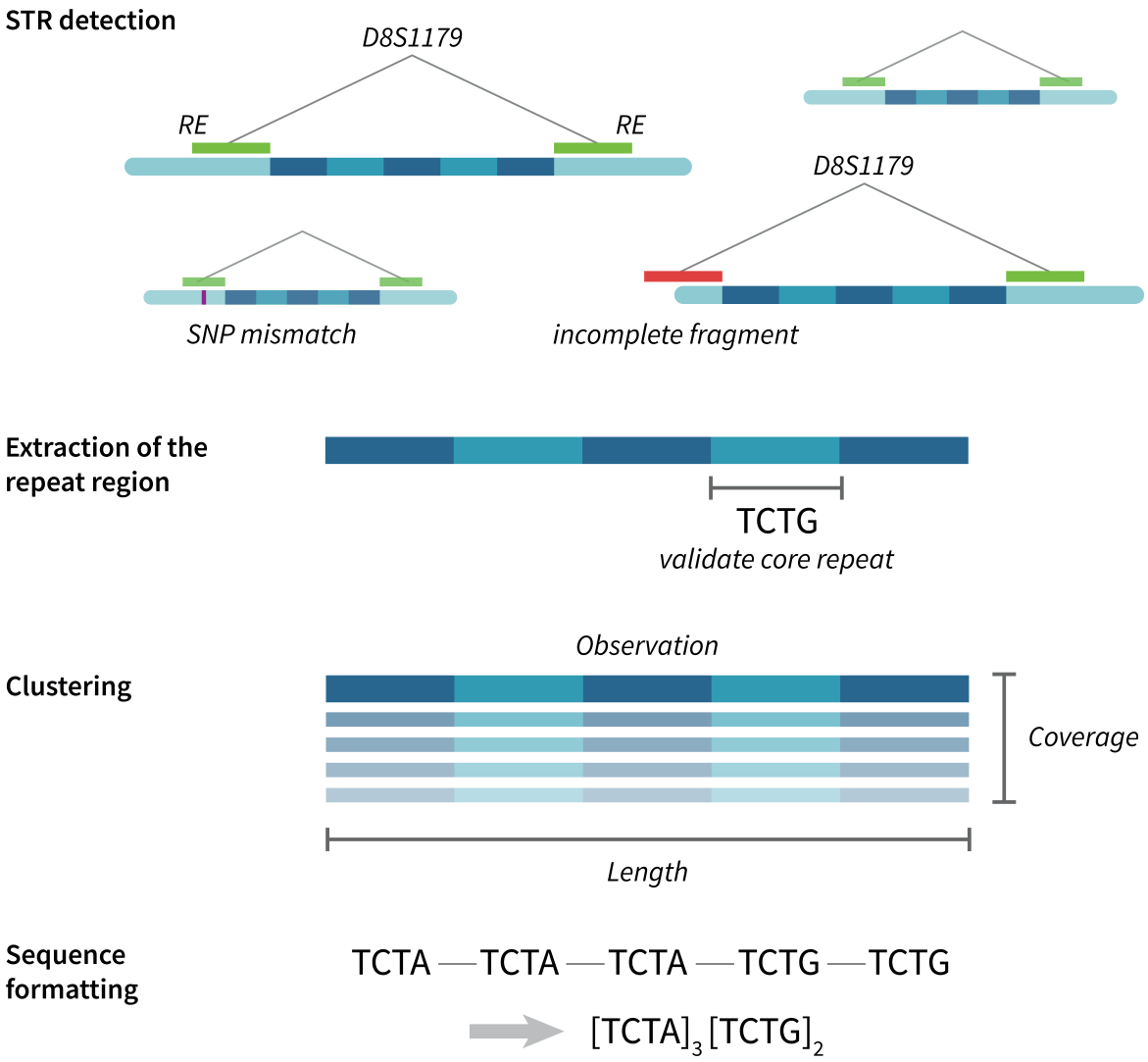


Figure 7. toaSTR detects STRs via locus-specific recognition elements (REs). REs have a certain mismatch tolerance while incomplete fragments are discarded. Repeat regions are extracted and validated for the presence of the core repeat motif. Identical repeat regions are clustered and form an observation. toaSTR characterizes the observation by length and coverage and formats the sequence.

Repeat region analysis

toaSTR's *de novo* analysis of the repeat region eliminates the need for a curated database of reference allele definitions [F07]. Therefore, both known and previously unknown sequence variants can be identified without a lookup table mapping raw sequences to formatted alleles. For length-based analysis of an observation, the CE-concordant length ℓ

is computed by the integer division $\ell = (L - N)/U$, with L: observation's length, N: number of non-STR associated bases within the repeat region, and U: repeat unit size. The remainder of this calculation is the microvariant nomenclature of incomplete repeats. A suffix is attached to the length designation if more than one observation with the same length is found at one locus (e.g., 8b). Sequence-based analysis involves bracketing [F08, F09] of the observation's sequence by writing core repeats in brackets with a number indicating the repeat counts (e.g., $[AGAT]_5$ $[AGAC]_3$). The formatting algorithm traverses along the observation's sequence and condenses repeat stretches to blocks by counting consecutive repetitions of the repeat motif(s) known for the respective locus.

Stutter modeling

The prediction of typical stutter variants and stutter frequencies aims at distinguishing artificial alleles from biological alleles. Stutter products appear during PCR-based amplification of STRs due to strand slippage-induced alterations of repeat numbers. The novel stutter model implemented here attempts to be an approximate computational representation of the biological stutter generation mechanism. It has previously been shown that the stutter ratio correlates with the length of long uninterrupted repeat stretches within an STR region (Brookes et al., 2012). Therefore, the stutter model assumes that sequence patterns of the most frequently observed stutter variants can be derived from the longest uninterrupted stretch (LUS) and the second-longest uninterrupted stretch of repeats (SLUS, only in compound or complex STRs) of the parent observation. In essence, the model computes virtual stutter variants for each observation found at a locus. Sequences of virtual stutter variants are built by contraction or elongation of the LUS and/or SLUS by one or more repeat units, respectively (Figure 8). The model considers forward, backward, and isometric stutter positions commonly described in MPS-STR analysis (Gaag et al., 2016; Young et al., 2017). The anticipated coverage for each virtual stutter variant is a function of the parent observation's coverage and the stutter threshold (ST) of the respective STR locus. STs are user-defined stutter ratios that are entered when a panel (that is, a user-defined collection of STR markers to be analyzed) is created. The commonly observed likelihood for stutter generation in different positions is reflected by a linear, quadratic, or cubic modulation of the ST term (Figure 8).

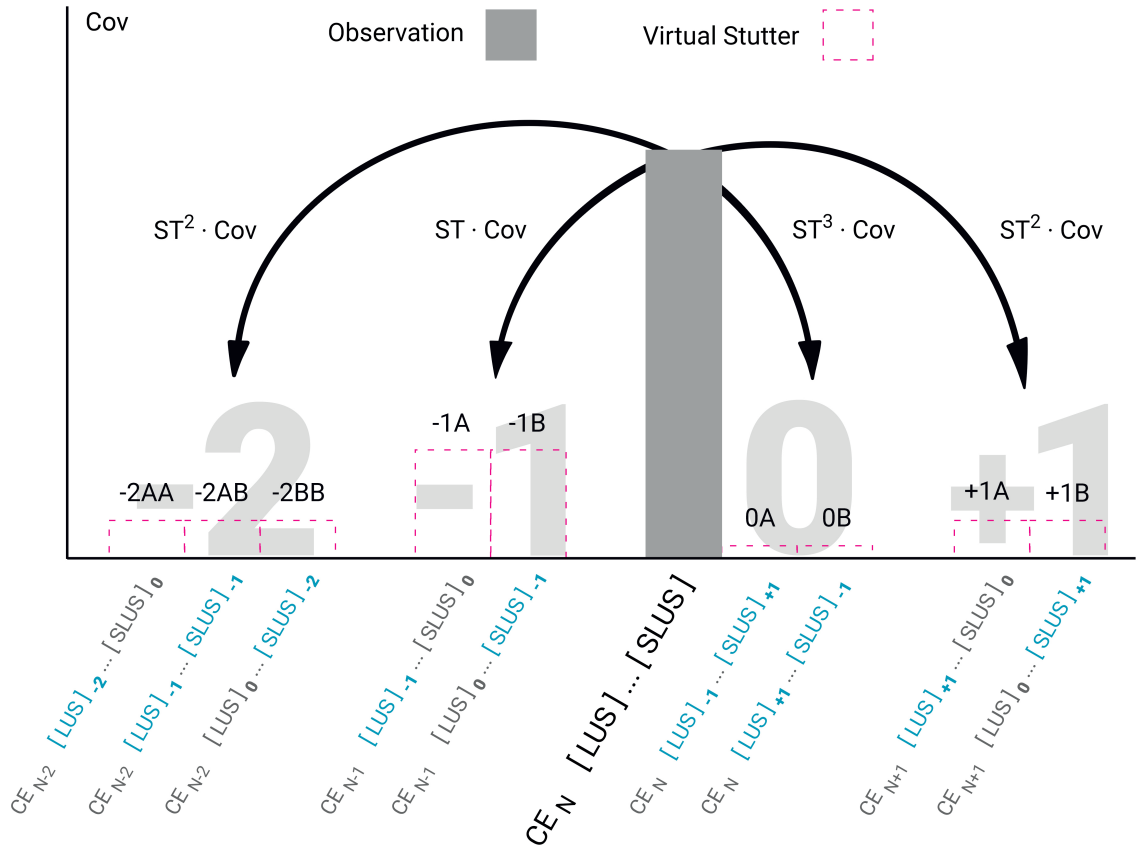


Figure 8. Schematic of the toaSTR stutter model. The algorithm considers nine virtual stutter variants in positions −2, −1, 0, and +1 relative to an observation's length N. Subscript numbers in the variant schemes (blue color) indicate the gain or loss of repeat elements in the longest uninterrupted stretch (LUS) or second-longest uninterrupted stretch (SLUS). Complex STRs may contain interspersed shorter repeat stretches and non-repetitive sequences (indicated by "…"). The virtual stutter coverage is a function of the stutter threshold (ST) for the respective STR system and the parent observation's coverage (Cov), as displayed at the arrows.

Classification of observations

The outcome of the stutter modeling process is again the basis for the classification of observations [F10, F11]. If a virtual stutter sequence matches with an observation actually being present in the sample, this virtual stutter amount is added to the so-called expected stutter value for the respective observation. The observation's class (allele, stutter, or artefact) is determined by comparing the expected stutter value and the observation's coverage, as described in Table 5. Additionally, the user-defined calling threshold protects against classifying low-abundance observations as true alleles, similar to a stochastic threshold known from CE-based methods (Scientific Working Group on DNA Analysis Methods, 2010).

Table 5. Classification of alleles, stutter, and artefacts in toaSTR. The class depends on the relation between the observation's coverage (Cov), the expected stutter (ES) value and the user-defined calling threshold (CT).

Class	Relation	Description
Stutter	Cov ≤ ES	Stutter is a product of PCR polymerase slippage in repetitive regions of biological alleles
Allele	Cov > ES, Cov > CT	Alleles are considered correctly amplified and sequenced DNA template of biological origin
Artefact	Cov > ES, Cov ≤ CT	Artefacts typically represent sequencing errors or low-level somatic mutations

3.1.2.2 Implementation

The toaSTR software has been implemented as a web application (web app) written in the Perl programming language and uses web technologies such as HTML, CSS, and JavaScript [N03, N04]. The production version of toaSTR is hosted by LABCON-OWL and has been made available to the forensic community as an open-access tool [N09]. The software is deployed as a FastCGI application running on an Apache web server, accessible through the internet: https://toastr.de/.

3.1.2.3 Architectural design

toaSTR was built using the Perl Catalyst [4] web application framework. Catalyst follows the Model-View-Controller (MVC) design pattern (Gamma, 2001), allowing to employ a modular program structure that partitions the system into separate components. Each component has own responsibilities assigned to it, such as the interaction with a data store, flow control through the application, and the presentation of content (Figure 9). Following this concept, modifying code in one component (e.g., how data is retrieved from a database) does not affect code in another component (e.g., how data is displayed to the user). Therefore, the application decomposition improves the readability, maintainability, and expandability of the program code.

The **model** is responsible for providing data, typically from a relational database. It performs CRUD (Create, Read, Update and Delete) operations on the data. toaSTR uses the DBIx::Class [5] engine as an object relational mapper (ORM) to access and modify content

[4] https://metacpan.org/pod/Catalyst
[5] https://metacpan.org/pod/DBIx::Class

from a database. An ORM maps a schema of database tables to classes in object-oriented programming. This abstraction layer allows to work with a SQL database by using Perl objects instead of SQL statements, thus nicely integrates writing application code and accessing databases. Moreover, using an ORM helps to prevent direct SQL injection attacks, i.e. executing unintended database commands or accessing data without proper authorization (Bundesamtes für Sicherheit in der Informationstechnik, 2006).

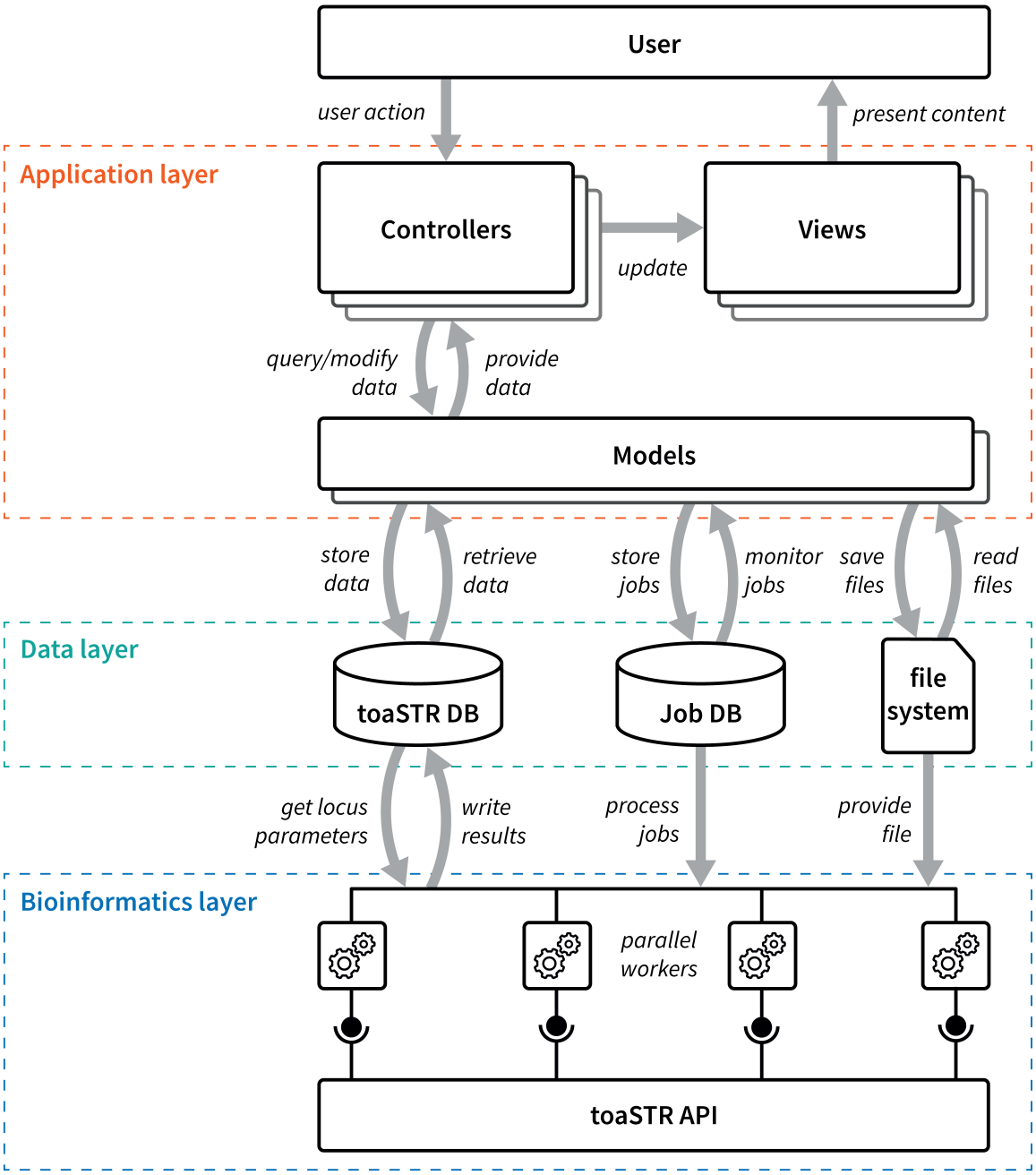


Figure 9. Structural decomposition diagram showing the high-level architectural design of the toaSTR software with interconnections between major subsystems and data repositories.

Views are responsible for presenting content to the user. Web content will typically be an HTML document, but a view may also generate a PDF or CSV file F13, a spreadsheet, or an e-mail. Views in toaSTR are built with the template processing engine Template::Toolkit [6]. Templates are static pages extended by templating directives such as variables and conditionals for displaying dynamic content. Genotyping reports as PDF files are produced using an additional templating plugin that allows the use of LaTeX markup. Templates help to maintain a clear separation of concerns between the structuring of content in the front-end and the application logic in the back-end. Formatting of content is again separated from the content's structure; toaSTR's design language (i.e., styling, colors, icons, and fonts) is based on the Materialize CSS framework [7] and Google's Material Design [8] principles. Materialize incorporates jQuery dialogs and animations, which provide feedback to the users and make toaSTR feel more like a native application. The underlying responsive system adapts the look and feel of toaSTR to virtually any screen size and allows for a unified user experience on various device classes. Diagrams for the visualization of results F12 are generated dynamically with the D3 (Data-Driven Documents) JavaScript library [9]. Instead of storing a large number of static images, all charts are drawn with D3 as interactive SVGs (Scalable Vector Graphics) at the time when a results page loads, driven by the data to be visualized. This reduces page loading times and storage requirements. The form construction and rendering framework HTML::FormHandler [10] is employed to set up forms and fields. toaSTR uses forms to receive user input F02 in many situations such as logging in to the application, starting an STR analysis or setting up a marker panel. Any user input is validated client-side and server-side before processing data in order to prevent the execution of potentially malicious code.

Controllers are responsible for the application flow control as well as the coordination of models and views. They handle requests and determine what actions the application takes when an event is raised by the user like clicking a button or accessing a URL (Uniform Resource Locator) with specific parameters. Figure 9 illustrates the interactions between the MVC components: On a user request, such as displaying the result of an analysis, the controller receives, interprets, and validates the user input. Next, the controller queries the model. The model then prepares the data and sends it back to the controller. Finally,

[6] https://metacpan.org/pod/Template::Toolkit
[7] https://materializecss.com/
[8] https://material.io/design/
[9] https://d3js.org/
[10] https://metacpan.org/pod/HTML::FormHandler

the controller creates or updates the view, which displays the requested content to the user. The logical domains of the toaSTR app are separated into five controllers (Admin, Panel, Root, Run, and User), each containing multiple actions (i.e., Perl subroutines). A brief description of controllers and responsibilities is given in Table 6.

Table 6. toaSTR controllers and responsibilities.

Controller	Responsibilities
Admin	App configuration, user administration, and system e-mails N08
Root	Display static content, error pages, and maintenance pages
Panel	Create, view, edit, and delete STR marker panels F04
Run	Start, edit, view, export, and delete analyses F03
User	User authentication and authorization, registration process, password reset, and user account self-management N06

On a web process timescale, the bioinformatics analysis of sequence data is a long-running task with an expected runtime between tens of seconds to a few minutes, depending on the amount of reads in the input file and the number of loci to be analyzed. Due to the request-response cycle of the web, a long execution time can block the application or even lead to a server connection timeout. Therefore, toaSTR utilizes the job queue system TheSchwartz [11] in the back-end that allows for a non-blocking execution of computationally expensive STR analysis, creating a rich and dynamic user experience P03. When a new genotyping task is started, the app generates a number of jobs equal to the number of loci to be analyzed and inserts the jobs into a separate job database. Worker processes in the background periodically monitor the job database, grab jobs, and execute them (Figure 9). Jobs that succeed are marked completed and removed from the queue. The number of workers depends on the available CPU threads. On a multi-core CPU, multiple jobs, hence multiple loci, can be processed in parallel with Parallel::Prefork [12], which greatly improves the time to result P02. A worker fetches all relevant information from the toaSTR database. It performs the actual analysis on the sequence data, including a search for the locus-specific recognition elements in the flanking regions, extraction and formatting of the repeat region, stutter modelling, and classification of observations, and stores the results in the database. The actual bioinformatics code is separated into a code

[11] https://metacpan.org/pod/TheSchwartz
[12] https://metacpan.org/pod/Parallel::Prefork

library (toaSTR API), since these functions are not necessarily coupled to the web app but might perspectively also be accessed from another interface like a command line.

3.1.2.4 Data design

toaSTR generates data that needs to be stored persistently and retrieved on request. These data include, for instance, information on the configuration of marker panels, previously performed analyses, and obtained STR profiles [D02]. Moreover, toaSTR depends on various administrative settings and requires access to user information to work properly and deliver a personalized user experience. toaSTR uses a MySQL [13] relational database management system with a core database containing twelve tables to represent all relevant entities such as analyses, results, alleles, users, etc. A separate database is used to support the job queueing system (Figure 9). A relational database system was chosen to organize relations and constraints between entities. For instance, one user may have multiple panels and one panel contains many markers. Referential functions maintain data consistency and integrity: Foreign key relationships cross-reference related data across tables; Cascading actions automatically delete or update matching rows in a child table upon updating or deleting a row in the parent table. The database layout including its properties and constraints is illustrated in the entity-relationship diagram available in Appendix A2. The uploaded sequence data files are converted to FASTA and stored on the server file system [D01] to enable re-starting an STR analysis with edited parameters without the need to re-upload the file.

3.1.2.5 User interface design

The toaSTR software is a browser-based application featuring a discovery-oriented graphical user interface [F02, F04, F12, N05]. Non-registered users can visit an information page and the registration form. Registered users get full access to the web app after logging in with their credentials. A tour through toaSTR with screenshots illustrating the common analytical workflow is available in Appendix A1.

3.1.2.6 Security design

The online available toaSTR web app employs a multi-layered security concept, relying on network security, host security, and application security. The application is hosted on a dedicated server located in a DIN ISO/IEC 27001 certified data center in Germany. The

[13] https://www.mysql.com/

server hardware is designed with multiple redundancies for high availability and reliability of this service. All application components (storage, database, and computing units) are located within a firewall-protected environment to control network traffic to and from toaSTR. The client-server communication is encrypted via Secure Socket Layer (SSL), thus preventing data in transit from being compromised or accessed by a third party. The app is multi-tenant [N06, N07], meaning that multiple users can use the app at the same time, while each user only has permission to its own data. Users are required to register for a personal account. Authorized users can log in with their username and password. Authentication is ensured via session cookies and access control when content is directly accessed (e.g., viewing an analysis by its ID). Any file uploads and user data are considered potentially malicious. Thus, form inputs are restricted by rules and validated on the client side as a feature of good user experience. Server-side validation prevents the processing of malicious data.

3.1.3 toaSTR proved to be a valid cross-platform genotyping software, simplifying the access to STR sequencing

The purpose of software development described in this dissertation was to deliver a versatile tool that is compatible with data of popular MPS platforms and can analyze both commercial and in-house identity panels. The toaSTR app presented here offers an intuitive graphical user-interface, sample management, comprehensive data analysis, and reporting of results following the nomenclature considerations of the ISFG. Taking full advantage of the high resolution enabled by MPS, sequence observations are automatically classified to assist with the interpretation of complex samples.

In an initial compatibility and concordance study (Ganschow et al., 2018), the robustness of this software has been assessed on a dataset containing reference DNA and GEDNAP (German DNA Profiling) ring trial samples. Results indicated the software's independent functionality; Precise allele calling was obtained with data from widely used MPS platforms (MiSeq RUO, MiSeq FGx, and Ion PGM), various custom and commercial STR amplification kits, five different library preparation chemistries, and multiple sequencing modes. In total, STR profiles showed 97 % concordance with CE-based results after automatic allele calling and 100 % concordance upon expert review. Simple, compound, and complex repeat structures were correctly detected and formatted in the common bracketed notation. The automatic classification of alleles, stutter, and artefacts was largely accurate and required only minor correction by the analyst. Sporadic

deviations resulted from an unexpectedly high stutter or a raised artefact coverage. The classification feature proved to be helpful for the interpretation of mixed-source samples, which at times are hard to deconvolve into individual STR profiles, since lengths and sequences of artificial and true alleles may be confusing (Marciano and Adelman, 2017). For instance, results for ring-trial sample 52S2 indicated a two-person constellation with up to four alleles per locus. At locus D21S11, two allele peaks were detected by CE; however, by making use of the sequence information gained by MPS, toaSTR identified three alleles including two isoalleles that share the same length but differ in sequence. Classification of observations greatly enabled the interpretation of imbalanced mixtures, in which alleles of a minor contributor were hidden in the stutter position of the major contributor. toaSTR also delivered data for the statistical metrics used during the development of the monSTR identity panel (Ganschow et al., 2019). Moreover, this software was involved in resolving allele calling discrepancies between CE-based and MPS-based data from 496 Spanish individuals with the Ion S5 system, acting as a "second opinion" to proprietary genotyping software (Barrio et al., 2019). In another dataset (Silvery et al., 2020), toaSTR analysis indicated one instance of discordance between CE and MPS methods at the SE33 locus. The CE-based genotype displayed a "< 3.2" allele, however, with MPS an allele "14" was seen. Subsequently, raw FASTQ files were investigated manually and a 63-bp-deletion was discovered within the flanking region. Here, toaSTR correctly characterized the repeat region. However, since toaSTR infers allele lengths from the repeat region only, InDels residing outside of toaSTR's recognition elements but within the fragment amplified by the CE kit may result in discrepant results between MPS and CE. This kind of discordance is, however, not unique to MPS, but can occur between different CE kits as well. This case illustrates that the sequence and positioning of recognition elements are of considerable importance for a correct sequence-based and length-based assignment of genotypes in accordance with forensic conventions. REs are likewise identifiers for an STR locus and boundaries for the repeat region to be reported and therefore may limit the detection abilities of sequence-search genotypers. Too permissive RE matching conditions may cause bioinformatics drop-in, i.e. a false positive calling of reads. If conditions are too strict, bioinformatics drop-out and imbalance may occur in case of an increased amplification error rate or sequencing error rate, incomplete reads, or multiple SNPs at the RE site. However, the generally observed high calling ratios in toaSTR (Silvery et al., 2020) indicated robust genotyping with the implemented REs. The author of this dissertation attended the 2019 STRAND Working Group meeting (Gettings et al., 2019) and agrees on the need for standardized start and stop

coordinates per locus. Standardized coordinates would unambiguously define the analyzable flanking ranges. Once this prerequisite is fulfilled, further extension of the toaSTR software may include the analysis of flanking region polymorphisms, which increase the allelic diversity (Gettings et al., 2018; Novroski et al., 2016; Phillips et al., 2018) and were also found to impact the stutter ratio (Woerner et al., 2017a).

toaSTR's stutter modeling algorithm recognizes nine stutter variants most frequently observed in MPS data. Although, in theory, stutter products may exhibit an even higher complexity, modeling stutter based on the LUS and SLUS patterns in the nominal positions N−2 to N+1 appeared to be a valid approximation of sequence-based stutter generation. Stutter in other repeat stretches or positions was found to be negligible (Ganschow et al., 2018; Silvery et al., 2020). Currently, the expected stutter values are calculated based on a locus-specific stutter threshold, which is determined empirically by the user based on reference samples. Adapting the stutter threshold on allele lengths or LUS lengths might further improve the prediction accuracy, as shown by Gaag et al. (2016). Recognizing stutter patterns with machine learning methods might be a promising strategy for future refinements of the stutter model. Hoogenboom et al. (2017) employed a linear regression model predicting the relationship between the repeat length and the stutter ratio. Gymrek et al. (2012) presented a logistic regression model of stutter noise built on haploid training data and involving features such as motif size, GC content, length and purity of the repeat region. Raz and colleagues (2018) have calibrated a Markov model for the prediction of stutter behavior using a synthetic library of STRs. However, training a model in the sense of supervised learning requires a relatively large, high-quality dataset of accurately labeled sequence observations (Hoogenboom et al., 2017) and the obtained model will likely be characteristic to stutter patterns of the MPS platform and the identity panel in use.

The web app concept picks up the idea of cloud computing and providing software-as-a-service (SaaS; Buxmann et al., 2008). Web apps for diagnostic and scientific purposes are in common use. As an example, the online tool HIV-GRADE [14] enables the prediction and interpretation of resistance against various drugs used in HIV therapy. The Y-Chromosome Haplotype Reference Database (YHRD) is a forensic web app that helps with the interpretation of genetic profiles from Y-chromosomal STRs and SNPs (Willuweit and Roewer, 2015). Web apps deliver many benefits compared to locally installed, native software. First, the output of a web app is rendered in a web browser. Since virtually any computer or mobile device features a web browser, a web app inherently supports most

[14] https://www.hiv-grade.de/

operating systems and device classes such as desktop computers, tablets, and smartphones. Second, instead of developing native software for each operating system (Windows, Ubuntu, macOS, Android, etc.), a web app covers many systems with a single code base, which greatly simplifies the maintainability of the software. Third, web apps are based on the client-server-model, meaning that data is processed and delivered via a web server and received by the client. Hence, users are not required to install the software or its dependencies. Since the computational work is mainly performed on the server, the application responds fast even on clients with low system specifications. On the negative side, legal and data privacy concerns might limit the usability of a web app for routine analysis of genetic data. Using the online version of toaSTR requires sending sensitive data over the internet and processing information on a remote server. Alternatively, the software and its numerous dependencies might be packaged as a virtual machine or a more lightweight Docker container [15], which would make the application portable and enable local deployment. A rewrite of the application code in JavaScript utilizing the ElectronJS [16] framework would allow the parallel deployment of toaSTR as a web app and cross-platform desktop app on a single code base.

In summary, toaSTR complements the landscape of existing open-access genotyping tools by combining cross-platform compatibility, a graphical user interface, high processing speed due to parallel processing, classification of results, and compliance with nomenclature considerations. With an emphasis on usability, toaSTR allows forensic experts to work with MPS data simply and efficiently. Table 7 highlights toaSTR's unique combination of features in comparison with previously published software packages.

[15] https://www.docker.com/
[16] https://electronjs.org/

Table 7. Comparison matrix of open-access software packages suitable for MPS-STR data analysis.

Approach	Software	Programming language(s)	Operating system	Input format	Graphical user interface	Parallel processing	Stutter modeling	Result visualization	ISFG allele nomenclature
alignment-based	lobSTR (Gymrek et al., 2012)	C/C++/R	Unix	FASTA, FASTQ, BAM	no	yes	yes	no	no
	RepeatSeq (Highnam et al., 2013)	C++/Python	Unix	BAM	no	yes	yes	no	no
	STR Viper (Cao et al., 2014)	Java	any	BAM, SAM	no	no	no	no	no
sequence-search	STRait Razor v3 (Woerner et al., 2017b)	C++	Windows, Unix	FASTQ	no	yes	no	(yes) [b]	(yes) [c]
	MyFLq (Van Neste et al., 2015)	Python	Web, Docker, Illumina BaseSpace [a]	FASTA, FASTQ	yes	yes	no	yes	no
	FDSTools (Hoogenboom et al., 2017)	Python	Unix	FASTA	no	no	yes	yes	no
	STRinNGS (Friis et al., 2016)	Python/R	Unix	FASTQ, BAM	no	no	no	yes	no
	SEQ Mapper (Lee et al., 2017)	.NET	Windows, Web	FASTA, FASTQ	yes	no	no	no	no
	Altius (Bailey et al., 2017)	Python	Web	FASTQ	yes	yes	no	yes	yes
	toaSTR (Ganschow et al., 2018)	**Perl**	**Web**	**FASTA, FASTQ**	**yes**	**yes**	**yes**	**yes**	**yes**

[a] Use of the BaseSpace application is restricted to Illumina MPS data

[b] Requires transferring results to a separate Excel workbook

[c] Formatting is supported only for known alleles present in the database

3.2 monSTR – a high-fidelity identity panel targeting twenty-one forensic markers

The second aim of the present thesis was to develop a mid-sized identity panel, monSTR, for the widely used MiSeq platform, offering an economical yet competitive alternative to large commercial panels. Panel development was divided into three subtasks: construction of the multiplex assay, systematic optimization of the assay performance, and developmental validation according to forensic guidelines.

3.2.1 Primer engineering and multiplex assay construction

A set of 21 forensically relevant markers commonly used in Europe and the USA is targeted by the monSTR panel, including D1S1656, D2S1338, D2S441, D3S1358, FGA, CSF1PO, SE33, D7S820, D8S1179, D10S1248, TH01, D12S391, vWA, D13S317, Penta E, D16S539, D18S51, D19S433, D21S11, Penta D, and amelogenin. Focusing on core CODIS and ESS loci facilitates pooling of more samples per sequencing run, which lowers costs per sample and makes this panel attractive for the application of MPS in a low-throughput scenario. The two pentamer systems, Penta D and Penta E, have been selected for their high power of discrimination and low levels of stutter (Bacher and Schumm, 1998). Inclusion of the highly informative marker SE33 is a unique feature of monSTR. While the Verogen ForenSeq system also contains primers targeting SE33, this marker is neither promoted by the manufacturer nor does the proprietary genotyping software Verogen UAS show results for SE33 (Köcher et al., 2018).

A multiplex-compatible set of primer pairs has been carefully engineered utilizing a combination of two bioinformatics tools, MPPrimer (Shen et al., 2010) and Primaclade (Gadberry et al., 2005) to meet specifically defined quality criteria. Specifications included (1) comparable thermodynamic properties of primers concerning length, GC content, and melting temperature, (2) no overlap of primer binding sites with frequent SNPs, (3) amplicon sizes compatible with MiSeq read lengths, (4) equally sized amplicons to prevent preferential amplification of shorter targets, and (5) maximization of the amplified flanking region variation to provide additional discriminatory features. In this workflow, MPPrimer was used to automatically suggest a set of candidate primers, while manual revision of primers, if necessary, was done with Primaclade. The final primer set was characterized by similar hybridization kinetics for all primers regarding balanced GC content and comparable melting temperatures within a range of 5 °C. The prediction of amplicon sequences and lengths, as well as the evaluation of potential non-specific (cross)

amplification between primer pairs, primer dimerization, and hairpin formation, was performed by *in silico* mPCR utilizing the MFEprimer software (Qu et al., 2012). Simulation of mPCR predicted a highly specific amplification of all 21 targets, while the formation of primer-dimers and secondary structures due to internal complementarity was largely avoided. Although primer design has arguably the most significant impact on multiplex PCR success, other components including concentrations of Mg^{2+}, dNTPs and the choice of the DNA polymerase may be contributing factors as well (Wallin et al., 2002). The Q5 High-Fidelity Master Mix (New England Biolabs, 2012) used for the monSTR identity panel was previously reported to achieve robust and highly specific mPCR results requiring little to no optimization of reaction components for mPCR applications (Menin and Nichols, 2013). The Q5 enzyme is among the most robust DNA polymerases on the market and has been successfully applied to the amplification of repetitive DNAs (Hommelsheim et al., 2014) and rapid PCR of STR markers (Romsos and Vallone, 2015). In preliminary studies of the present thesis, Q5 was compared to *Taq* polymerase to amplify a 10-plex STR assay (Ganschow, 2016). Results revealed a 10-fold reduction of artificial sequences containing random substitutions and decreased the formation of stutters up to 69 %. These results reflect those of Potapov and Ong (2017) who measured the lowest rate of base substitution errors with Q5 among various proofreading enzymes. Accordingly, Daunay et al. (2019) observed a reduction of stutter when PCR-amplifying microsatellites with the Q5 enzyme. Fragment analysis of monSTR amplicons from singleplex and multiplex reactions under standard PCR conditions suggested a solid basis for the successive optimization of the thermocycling protocol.

A dual-PCR library preparation strategy was chosen to have full control over the sequencing targets. In contrast to a random fragmentation followed by adapter ligation, the dual-PCR strategy yields only informative amplicons (Warshauer et al., 2013). In the first PCR, STR markers were amplified using target-specific primers with universal overhangs. Overhangs served as universal primer binding sites in the second PCR to introduce Illumina Nextera XT index adapter sequences. Amplicons were designed to have a mean size of 300 bp, while the longest known allele for Penta E was predicted to generate a 370 bp product. Therefore, monSTR amplicons are compatible with 251 x 251 nt paired-end sequencing, which is currently the maximum cycle number supported by the MiSeq Nano flow cell (Illumina, 2018). This flow cell type was chosen to meet the frequent requirement of forensic laboratories to process a small number of samples. The Nano flow cell is specified to yield 1,000,000 reads pairs per sequencing run and thus would be appropriate for pooling ten monSTR libraries, resulting in a typical locus coverage of about

5,000 reads on average (Barrio et al., 2019; Gettings et al., 2016). With the paired-end configuration, merging of overlapping read pairs is necessary to form an informative consensus read spanning the full STR. It was observed that merging may lead to ambiguous results, substantial read loss, and skewed genotypes when reads end within the repeat region (Ganschow et al., 2019, 2018). While not tested yet with monSTR, switching to 300 x 300 nt paired-end sequencing using a 600-cycle chemistry may improve merging results, since reads are more likely to end in non-repetitive flanking regions. The sequencing mode for monSTR has been changed in the course of the developmental validation to 400 nt single-end sequencing (Silvery et al., 2020). While this setup eliminates the need to merge reads, lower base qualities towards the end of reads were observed. Ignoring low-quality reads may add unreliable and random information to the dataset and lead to a false interpretation of results. Laboratory guidelines, for instance in clinical diagnostics (Aziz et al., 2015), therefore require quality control (QC) of raw MPS reads. However, to the best of the author's knowledge, there is no consensus in forensic genetics on quality indicators, thresholds, and best practices for the treatment of raw sequence data. Previous MPS-related publications did not report on QC of raw data at all (Gettings et al., 2018; Hoogenboom et al., 2017; Zeng et al., 2015b). Because of the lack of forensic QC guidelines, it was decided not to perform any quality treatment of the underlying raw data in the development and validation studies of the monSTR panel, since preliminary results indicated that different quality filtering algorithms and thresholds may have a substantial impact on the STR profile (unpublished data). However, the author wishes to enhance awareness for raw data QC in forensic genetics and encourages a debate on this topic. It should be explored whether current STR genotyping algorithms are capable of obtaining high-quality profiles even from lower quality sequence data. For instance, the STR detection algorithm of the toaSTR software (Ganschow et al., 2018) discards reads with too many errors and may be considered as an inherent quality filtering step. Experiences with successful standardization of raw data QC in other fields such as clinical laboratory practice (Gargis et al., 2012) may be transferred to the field of forensic genetics, probably guided by a central authority such as the QC platform STRidER (Bodner et al., 2016).

3.2.2 Design of Experiments enabled a systematic assay optimization

The development of an identity panel involves a rigorous experimental strategy to identify and optimize critical parameters. However, comprehensive PCR optimization has been prohibitively laborious and cost-intensive (Henegariu et al., 1997). A thorough exploration of experimental conditions depending on multiple influencing factors can quickly result

in dozens of experiments, which may discourage small laboratories in developing tailored in-house assays at all (Butler, 2005). Moreover, a one factor at a time approach ignores potential interactions between factors and will rarely identify true optimum conditions (Toms et al., 2017). Therefore, the optimization of a multi-factorial processes demands a systematic strategy.

Article two of the present thesis (Ganschow et al., 2019) describes the first attempt to apply the Design of Experiments methodology to the optimization of an MPS-based assay. In essence, DOE provides an organized approach to plan informative experiments, assess the joint influence of input factors on the system response, and create meaningful mathematical models of the system for decision making (Eriksson, 2008). Based on the experience of previous multiplex optimization studies (Boleda et al., 1996; Butler, 2005; Edwards and Gibbs, 1994; Henegariu et al., 1997; Menin and Nichols, 2013), four quantitative input factors were considered to have a strong effect on the fidelity of the monSTR panel: annealing temperature (ATemp), annealing time (ATime), extension time (ETime), and number of PCR cycles (Cyc). Since the performance of an MPS-STR assay cannot be measured intuitively, the optimization goals were translated into three representative responses that can be obtained from MPS-STR profiles: (1) the coefficient of variation of the locus coverage (CV) to determine the interlocus balance; (2) the calling ratio (CR) between on-target reads and total reads to reflect the amplification specificity; and (3) the signal-to-noise ratio (SN) of allelic observations to noise observations, which expresses the tendency to form stutter products and other technical artefacts. Based on samples of reference DNA, the four factors were varied according to an RSM experimental design with three levels per factor (low, center point, and high), incorporating 20 experiments with unique combinations of factor settings plus three replicates of center points (i.e., settings of the standard PCR protocol). STR profiles were generated with the toaSTR software (Ganschow et al., 2018). The creation of the experimental design and statistical models was performed using the MODDE Go software (Sartorius Stedim Data Analytics, Umeå, Sweden).

The DOE methodology provides several graphical diagnostic tools, as exemplified in Figure 10, which support the evaluation of raw data as well as the analysis and interpretation of models (Haaland, 1989). Replicate plots indicated that variation in the three center point replicates was much smaller than the variation in the entire investigation series, suggesting good control over the experimental procedure. One outlier experiment within the dataset, deviating by more than four standard deviations, was identified using normal probability plots and excluded prior to model building. Coefficient

plots visualized the effect of factors and interactions on the responses and helped to identify insignificant factors, which were subsequently removed by backward elimination during model refinement.

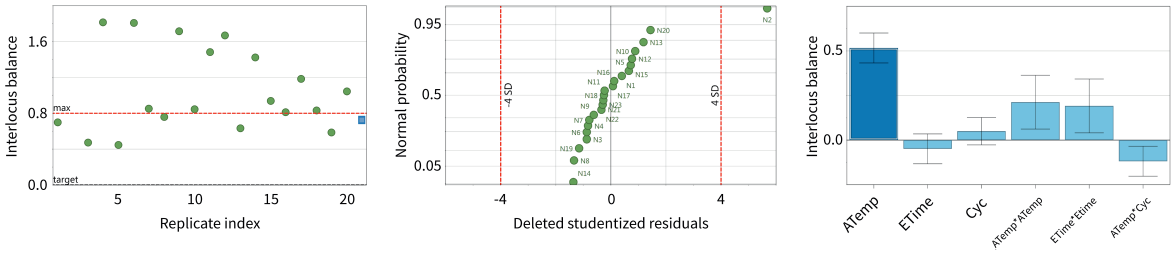


Figure 10. Graphical diagnostic tools of the DOE framework. Diagrams are shown exemplarily for the CV response. Left: replicate plot displaying response values of 23 experiments with varying factor settings. High overall variation (circles) versus small replicate variation (rectangles). Center: Normal probability plot indicating experiment N2 as an outlier. Right: Regression coefficient plot indicating a dominant ATemp factor, which increases CV by about 0.5 when ATemp changes from its standard condition to the high level.

Central to DOE is the generation of statistical models, which establish a relationship between the input factors and the measured response values (Eriksson, 2008). By applying multiple linear regression (MLR; Box and Draper, 1987) to the experimental data, a statistical model was generated for each of the three responses, incorporating first and second-order terms and factor interactions. MLR is based on fitting the regression model that minimizes the residual sum of squares of the respective response. A model can be written in the form $y = \beta \cdot X + \varepsilon$, where y is an n-dimensional vector of response values; β is a k-dimensional vector of regression coefficients; X is a $1 + n \times k$ matrix of model terms including the constant; ε is an n-dimensional vector of residuals. Model refinement based on the exclusion of insignificant model terms yielded the following model equations, which describe the behavior of the multiplex PCR system in the region studied:

$$CV = 0.79 + 0.52x_1 - 0.05x_3 + 0.05x_4 + 0.21x_1x_1 + 0.19x_3x_3 - 0.12x_1x_4 + \varepsilon$$

$$CR = 0.88 + 0.09x_1 - 0.02x_2 + 0.06x_3 - 0.11x_4 + 0.03x_1x_1 - 0.04x_3x_3 - 0.07x_4x_4 + 0.07x_1x_2 + 0.02x_1x_3 + 0.08x_1x_4 - 0.04x_2x_3 + \varepsilon$$

$$SN = 25.94 - 4.18x_1 - 0.72x_2 - 3.47x_3 - 2.61x_4 - 2.91x_2x_2 - 2.41x_3x_3 + 0.99x_2x_3 + 2.64x_3x_4 + \varepsilon$$

with x_1: ATemp, x_2: ATime, x_3: ETime, x_4: Cyc, ε: residual error. The regression coefficients were scaled and centered, therefore the constant terms relate to the estimated response value at the design center-point. The value of a regression coefficient indicates the response change as the factor changes from its standard condition to the high level.

Modeling revealed valuable insights that helped to understand the characteristics of the identity panel under investigation. Each of the four input factors had a significant impact on one or more responses, either as a main effect or as part of an interaction. The CV model was dominated by the factor ATemp (0.52 ± 0.08, $p = 1.2 \cdot 10^{-9}$), suggesting that high ATemp induced a stronger interlocus imbalance. A possible explanation for this might be that a more stringent ATemp in combination with competitive effects in a multiplex reaction may lead to preferential amplification of some targets. ATemp had also the strongest impact on CR (0.09 ± 0.03, $p = 6.37 \cdot 10^{-6}$), indicating that adjusting the temperature to higher levels improved the specificity of PCR products. Raising ATemp generally increases the stringency of primer-template hybridization since fewer mismatches at the priming site are tolerated (McPherson and Møller, 2006), which may explain decreased amplification of non-specific sequences. The negative effect for Cyc on CR (-0.11 ± 0.02, $p = 4.8 \cdot 10^{-7}$) was not unexpected since prior studies (Edwards and Gibbs, 1994; Henegariu et al., 1997) have noted that keeping the number of PCR cycles to the minimum required to generate sufficient product reduces the likelihood of PCR errors and the accumulation of non-specific products. Longer ETime was beneficial for a specific and complete amplification with regard to the CR model (0.06 ± 0.02, $p = 3.2 \cdot 10^{-5}$). This appears plausible, since multiplex PCR may require longer extension time for the polymerase to complete the synthesis of all products as the pool of enzyme and nucleotides becomes a limiting factor (Henegariu et al., 1997). The CR model was also characterized by several interactions between factors and second-order terms. For instance, the interaction of ATemp and ATime (0.06 ± 0.03, $p = 9.1 \cdot 10^{-5}$) appeared to have a positive effect on CR, which exemplifies the importance to investigate factor interactions in optimization strategies. Interestingly, ATemp (−4.18 ± 1.42, $p = 2.5 \cdot 10^{-5}$) was also the most important factor for SN. These findings support the hypothesis that a lower ATemp enhances the stability of the polymerase/template complex, which in turn reduces the likelihood of polymerase slippage, resulting in lower stutter ratios (Murray et al., 1993; Seo et al., 2014). In contrast, the SN model implied the generation of more technical artefacts upon increasing ETime (−3.46 ± 1.42, $p = 1.5 \cdot 10^{-5}$), which could be attributed to a longer timeframe for polymerase slippage events. ATime had a comparatively little influence on the models. In the literature, a longer annealing time has been associated with a higher risk of non-specific products (McPherson and Møller, 2006). Consistently, it was observed in the present study that extending ATime may reduce CR (−0.02 ± 0.02, $p = 0.05$). Moreover, a significant negative quadratic effect was included in the SN model (−2.91 ± 2.52, $p = 0.03$), which may be linked to a higher likelihood of polymerase/template dissociation.

The *p*-values for the complete models were determined to be p (CV) = $2.30 \cdot 10^{-8}$, p (CR) = $5.31 \cdot 10^{-7}$, p (SN) = $1.16 \cdot 10^{-6}$, indicating that the variance that can be explained by the models is significantly larger than the unmodelable (residual) variance. The models were subsequently used to predict the responses for any condition within the experimental space and to draw conclusions about optimal factor settings. Response contour plots (Figure 11) are graphical representations of the models, allowing an approximate localization of the optimal working zone that induces the best outcome for the respective response. As becomes apparent from the plots, low annealing temperature and reduced number of PCR cycles would be preferable for a low interlocus imbalance and high signal-to-noise ratio. Conversely, these conditions were predicted to have an unfavorable effect on the calling ratio. Therefore, the downhill simplex method (Nelder and Mead, 1965) was employed to find combinations of factor levels that jointly optimize all responses. The exact coordinates of the global optimum that minimizes CV and maximizes both CR and SN are given in Table 8. Optimal factor settings, i.e. a reduction of ATemp, ATime, and Cyc and an increased ETime, were predicted to improve all responses. Confirmation experiments employing the optimized PCR protocol for the monSTR identity panel indeed verified a substantial improvement of the assay performance compared to the standard PCR conditions before optimization. Results suggested that a high interlocus balance and low formation of stutter artefacts was achieved by choosing favorable thermocycling conditions, while the amplification specificity remained on a high level. After DOE optimization, the interlocus balance was further improved by titration of primer concentrations.

An efficient RSM experimental design, which required only a minimum number of experiments, supported the generation of realistic mathematical models. However, some deviations were observed between predicted and the confirmed values, which could be attributed to stochastic effects during the multiple library preparation steps as well as sequencing inter-run variation. Also, genotype-specific effects of the samples used throughout the study, such as the allele spread and heterozygosity, may add to the variation, since longer alleles tend to have a lower amplification specificity and a higher stutter ratio (Klintschar and Wiegand, 2003). Improvement of the modeling accuracy may be achieved by the inclusion of replicates of the individual experiments. Further improvements of the assay performance might be achieved by adjusting the composition and concentrations of reaction components. For instance, adjusting the Mg^{2+}/dNTP balance was found to increase the amplification specificity (Henegariu et al., 1997). Overall, this application of DOE to the optimization of the monSTR identity panel

strengthens the idea that computer-aided modeling can result in a viable approximation of complex molecular assays, as previously demonstrated for enzymatic assays (Altekar et al., 2006), stem cell bioprocessing (Toms et al., 2017), and cDNA microarrays (Wrobel et al., 2003). The DOE methodology enables forensic laboratories to develop and optimize custom identity panels with reasonable resources and communicate results transparently.

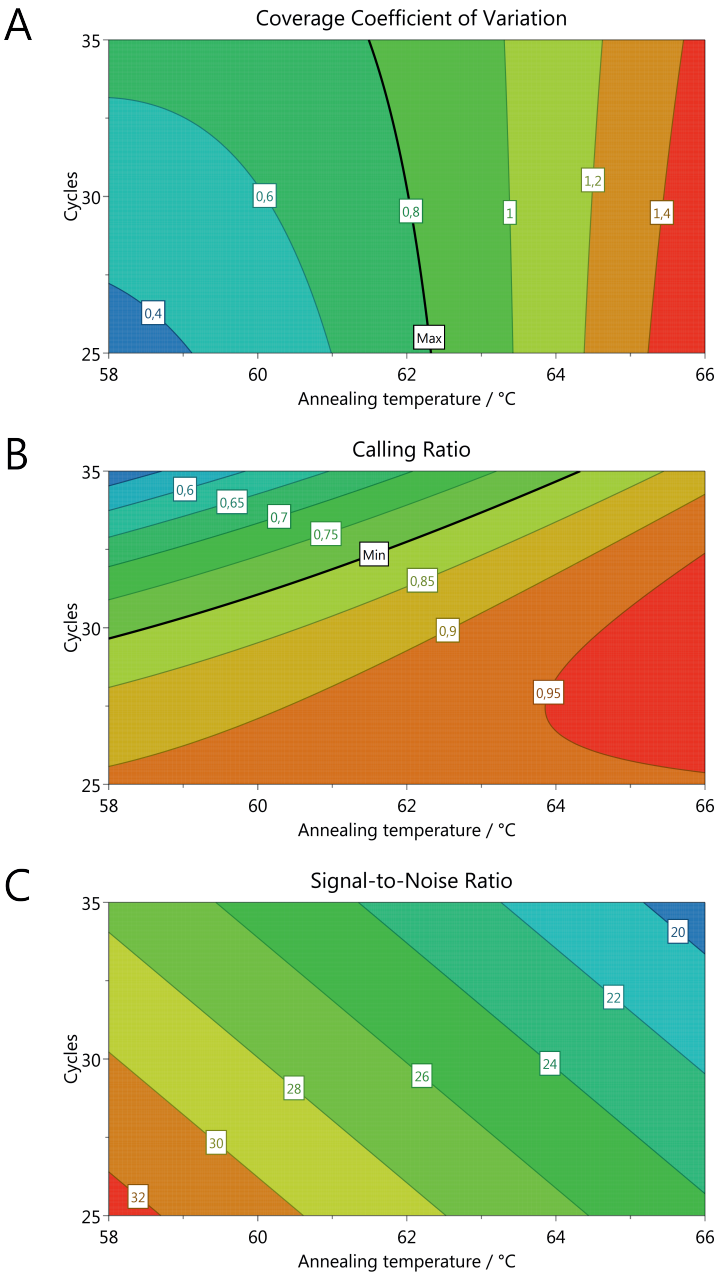


Figure 11. Response contour plots showing the predicted responses for (A) CV, (B) CR, and (C) SN as a function of ATemp and Cyc. As the factors ATemp and Cyc were identified as driving forces in all three models, ETime and ATime were fixed at their center-point levels for visualization purposes (ATime = 30 s, ETime = 60 s). Values in white boxes are given in the unit of the respective response.

Table 8. Comparison of input factors and assay performance before and after optimization. Response values for CV and SN substantially improved with the optimized protocol, while CR remained on a high level.

Protocol	ATemp	ATime	ETime	Cyc	CV	CR	SN
standard	62 °C	30 s	60 s	30	0.72 ± 0.01	0.92 ± 0.02	28 ± 1
optimized	58 °C	23 s	65 s	25	0.61 ± 0.01	0.87 ± 0.02	38 ± 4

3.2.3 monSTR's robustness was demonstrated by comprehensive developmental validation studies

It is common practice in forensic genetics that novel identity panels undergo a detailed evaluation and validation prior to routine application (Churchill et al., 2016; Guo et al., 2016; Jäger et al., 2017; Köcher et al., 2018; Müller et al., 2018; Xavier and Parson, 2017). The monSTR panel, in combination with the MiSeq platform and the open-access genotyping software toaSTR, was subjected to a rigorous "stress test" (Silvery et al., 2020) to validate its robustness and utility for routine application. Developmental validation studies described here were conducted in accordance with guidelines published by the Scientific Working Group on DNA Analysis Methods (SWGDAM; Scientific Working Group on DNA Analysis Methods, 2016). Robustness was assessed through concordance testing, sensitivity studies, evaluation of repeatability and reproducibility, mixture analysis, species testing, and investigation of forensic mock samples.

Maintaining backward compatibility of MPS allele calling with existing law enforcement STR databases is of great importance (Parson et al., 2016). Concordance testing of monSTR results revealed full concordant length- and sequenced-based genotypes in comparison with data from CE and parallel MPS technologies. When samples were handled by different operators, results obtained from reference DNA samples were fully reproducible with respect to genotyping accuracy, interlocus balance, amplification specificity, and signal-to-noise ratio. Only negligible variation was observed when samples were repeated within one sequencing run or across individual runs. Analysis of GEDNAP proficiency samples revealed largely concordant profiles between CE and MPS with 21 additional intra-allelic sequence variants (isoalleles) detected by MPS. Data suggested a good amplification efficiency and heterozygous balance of the SE33 marker, which was comparable to other loci even for SE33 genotypes exhibiting large alleles (> 30) and heterozygote allele spreads up to 5 repeat units. This is a remarkable outcome contrary to

that of Borsuk et al. (2018), who reported challenges of interpreting SE33 sequence data from a ForenSeq experiment, such as high sequence noise, low coverage of large alleles, and heterozygote imbalance. It seems possible that the high fidelity of SE33 in the monSTR panel benefitted from primer design and the thorough optimization of reaction conditions. However, there is a potential for bias due to limited sampling. Stutter ratios (N-1) observed from the validation data mostly ranged between 5 % and 20 % and were comparable to published values of other identity panels (Hussing et al., 2018). Locus D12S391 exhibited the highest stutter value (at maximum 26.9 %), while stutter as little as 0.4 % was seen at the Penta E locus.

At times, only minute quantities of DNA are available for STR analysis (Gill et al., 2015), which demands a highly sensitive test system. The monSTR panel was found to be competitive regarding the detection sensitivity and analyzability of low-input samples with commercially available MPS-STR panels (Jäger et al., 2017; Müller et al., 2018). While the monSTR panel has been optimized for amplification of 1,000 pg of total input DNA, full profiles with consistently balanced loci were obtained in a wide dynamic range of DNA inputs between 1,000 pg and 62.5 pg. Allele and locus dropouts were observed with 31.3 pg DNA input and lower. Still, 76.4 % of alleles were correctly called with as little as 7.8 pg DNA input amount, which approximately corresponds with the DNA content of a single nucleus. Good heterozygote balance (allele coverage ratio > 0.6) was generally achieved with a DNA input of 125 pg or higher. At low levels of DNA template (15.6 pg and 7.8 pg), allele and locus imbalances became more apparent and the proportion of on-target reads fell below 50 %. Imbalances are likely to be related to stochastic effects during PCR amplification (Butler, 2011) in combination with unequal sample loss upon bead-based size separation (Riman et al., 2017) and multiple pipetting steps during library preparation.

As mixed-source samples are frequently encountered in forensic casework (Balding, 2013), an identity panel should support the deconvolution of minor and major contributor genotypes. The mixture study examined the panel's ability to recover minor alleles at various ratios of 2-person mixtures (1:1, 1:4, 1:9, 1:19, and 1:49). Theoretical and observed mixing ratios were compared based on the minor component allele frequency (MAF). Unshared minor contributor alleles that did not overlap with major alleles or major stutter were accurately recovered in ratios 1:1 to 1:19. With a strongly imbalanced ratio of 1:49, still 94.4 % of minor contributor alleles were resolved, while allele dropouts were observed at the CSF1PO, D8S1179, and D19S1248 loci. Results suggested that MAF was a good estimator of the minor contribution; A precise prediction could be achieved even for the challenging 2 % (1:49) mixture ratio (MAF 2.0 % ± 0.7 %). In order to assess the species

specificity of monSTR, non-primate mammal samples of common domestic species (dog, cat, and horse) were analyzed. No evidence for species cross-reactivity was found. Spurious allele calls (CR < 0.05) were partly associated with the genotypes of the respective animal owners.

Taken together, data presented in this study lend support that the monSTR identity panel is suitable for STR sequencing and produces reliable and accurate data. A substantial set of experimental studies explored the capabilities and limitations of this assay. One of the initial objectives when developing monSTR was also to offer an economically attractive alternative to large panels for the MiSeq platform. For instance, the Verogen ForenSeq system was designed to be most cost-effective when analyzing 36 database or 12 casework samples per sequencing run in the low-throughput configuration (Verogen, 2018). This translates to reagent and sequencing costs of approximately EUR 69 to EUR 102 per sample for the ForenSeq kit. In comparison, the monSTR panel comes at much lower cost of EUR 47 per sample, which underlines its utility for forensic genetic laboratories who wish to focus on core STR markers.

4 Conclusion and outlook

Forensic DNA typing has come a long way from the first molecular techniques, through electrophoretic size-based profiling, to modern STR sequencing. MPS affords great promise for routine implementation in forensic sciences, including virtually unlimited multiplexing of markers, resolution of intra-allelic variation, and combination of different marker types. However, forensics is a conservative field due to its support of law enforcement, hence new technologies are introduced very carefully. The present study examined the question of how STR sequencing can be promoted to gain greater acceptance within the forensic community and proposed a holistic strategy for high-resolution DNA typing. A multi-faceted, interdisciplinary approach was followed, integrating forensic genetics, targeted massively parallel sequencing, experimental design methodology, bioinformatics, and software engineering. First, this dissertation presented the construction, optimization, and validation of a novel MPS-STR identity panel. The monSTR panel was designed in response to demand for a mid-sized STR assay on the Illumina MiSeq platform that captures European core STR loci plus the highly informative SE33 marker. Second, a software engineering project was undertaken to develop a new web application, toaSTR, that allows analyzing MPS results without in-depth bioinformatics knowledge, while taking full advantage of high-resolution sequence data.

Focusing on core STR loci commonly used for databasing in Europe, the monSTR assay complements the landscape of identity panels for the Illumina platform. It has been shown for the first time that the DOE methodology, previously applied mainly in the field of engineering, can be successfully transferred to the multifactorial optimization of an MPS-based assay. Statistical models of mPCR emerged as reliable predictors of universal reaction conditions. Jointly optimized parameters resulted in a high-fidelity identity panel, characterized by a well-balanced amplification of microsatellite loci, a high on-target ratio of sequence reads, and reduced formation of stutter products compared to standard PCR conditions. Comprehensive validation of the monSTR panel suggested robust genotyping capabilities, a large dynamic range, and deep mixture resolution. It is hoped that this research will contribute to a more efficient and transparent assay development in the field of forensic genetics and beyond, whose results can be better communicated than findings obtained with an intuitive approach. Further work might explore a modular combination

of the monSTR panel with other marker types. For instance, mRNA markers would allow determining the cellular source of a crime scene sample and may help to reconstruct the course of events. Preliminary studies indicated that monSTR amplicons and mRNA-derived products can be readily integrated into a parallel library preparation and sequencing workflow (unpublished data).

The analysis and interpretation of STR sequencing data represent one of the main challenges for the integration of MPS into standard casework laboratories. The toaSTR software strives to solve this issue with an intuitive, easy-to-use GUI and a streamlined workflow. Allele calling in both single source and mixed samples is supported by a novel stutter model that can identify artificial products originating from the analytical scheme. Evidence from multiple studies suggested that this independent tool can detect STR alleles in data generated with common MPS platforms and identity panels including commercial as well as in-house assays. By the end of 2019, more than 3,000 analyses have been performed by the forensic community using the open-access online version of toaSTR. LABCON-OWL has initiated the Euroformatics consortium and a collaboration with the German Federal Criminal Police Office (Bundeskriminalamt) to guide the further development of this software and to continue gathering support and feedback from its users. The roadmap of new features includes the extension of recognizable markers and the analysis of flanking region polymorphisms. Once a consensus has been reached within the forensic community regarding the nomenclature of sequence alleles and unified coordinates of STR loci, adjustments to the monSTR amplicon design and the regions targeted by toaSTR might become necessary. The web app concept of toaSTR picks up the current trend of using cloud computing services (Armbrust et al., 2010) to deliver scalable solutions for the analysis of big data. Following the example of the de.NBI Bioinformatics Cloud (Sczyrba et al., 2018), future studies might evaluate the idea to provide a suite of forensic software tools enabling quality control, genotyping, biostatistical analysis, etc. in a dedicated cloud infrastructure with strong emphasis on data security.

Possible applications of the developments described herein are not restricted to human identification. For example, panel and software may be applied to the authentication of human cell lines by STR analysis (Almeida et al., 2016) in order to prevent cell line misidentification in research and biotechnological applications. Moreover, the toaSTR software has shown to be suitable for the analysis of STR alleles in *Cannabis sativa* DNA, which may be useful to trace the origin of plants (manuscript in preparation). In a broader context, it appears attractive to adapt monSTR and toaSTR to alternative sequencing technology. Nanopore sequencing marks the onset of an era of portable, miniaturized

devices that may facilitate real-time genetic analysis at the crime scene (Plesivkova et al., 2019). Long read lengths associated with this technology can easily overspan repetitive structures and capture flanking region variation. Future research might explore the adaption of the monSTR library preparation to nanopore amplicon sequencing (Calus et al., 2018) and may study aspects of nanopore-specific error patterns (Harris et al., 2019) for allele calling with toaSTR. Current limitations for nanopore STR sequencing exist due to homopolymeric sequencing bias (Cornelis et al., 2018) and low base-calling accuracy, while the latter issue may be counteracted by consensus sequencing of intramolecular-ligated amplicons (Li et al., 2016). Due to its ability to directly detect methylation patterns in DNA from reads without extra laboratory techniques (Simpson et al., 2017), nanopore sequencing might augment forensic investigations through the analysis of age-informative epigenetic markers (Parson, 2018). Decreasing overall sequencing cost and high data output allow even for metagenomics analysis of forensically relevant bacterial communities, such as the epinecrome as a predictor for the postmortem interval (Pechal et al., 2014). Thus, the evolution of sequencing technology can be seen as a driver for the ongoing exciting transformation (Kayser and Parson, 2017) of forensic genetics into forensic genomics.

References

Almeida, J.L., Cole, K.D., Plant, A.L., 2016. Standards for Cell Line Authentication and Beyond. PLoS Biol. 14. https://doi.org/10.1371/journal.pbio.1002476

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

Altekar, M., Homon, C., Kashem, M., Mason, S., Nelson, R., Patnaude, L., Yingling, J., Taylor, P., 2006. Assay Optimization: A Statistical Design of Experiments Approach. J. Assoc. Lab. Autom. 11, 33–41. https://doi.org/10.1016/j.jala.2005.11.001

Amorim, A., Pinto, N., 2018. Big data in forensic genetics. Forensic Sci. Int. Genet. 37, 102–105. https://doi.org/10.1016/j.fsigen.2018.08.001

Ananda, G., Walsh, E., Jacob, K.D., Krasilnikova, M., Eckert, K.A., Chiaromonte, F., Makova, K.D., 2013. Distinct Mutational Behaviors Differentiate Short Tandem Repeats from Microsatellites in the Human Genome. Genome Biol. Evol. 5, 606–620. https://doi.org/10.1093/gbe/evs116

Anvar, S.Y., Gaag, K.J. van der, Heijden, J.W.F. van der, Veltrop, M.H.A.M., Vossen, R.H.A.M., Leeuw, R.H. de, Breukel, C., Buermans, H.P.J., Verbeek, J.S., Knijff, P. de, Dunnen, J.T. den, Laros, J.F.J., 2014. TSSV: a tool for characterization of complex allelic variants in pure and mixed genomes. Bioinformatics 30, 1651–1659. https://doi.org/10.1093/bioinformatics/btu068

Aponte, R.A., Gettings, K.B., Duewer, D.L., Coble, M.D., Vallone, P.M., 2015. Sequence-based analysis of stutter at STR loci: Characterization and utility. Forensic Sci. Int. Genet. Suppl. Ser. 5, e456–e458. https://doi.org/10.1016/j.fsigss.2015.09.181

Armbrust, M., Stoica, I., Zaharia, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., 2010. A view of cloud computing. Commun. ACM 53, 50. https://doi.org/10.1145/1721654.1721672

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. Arch. Pathol. Lab. Med. 139, 481–493. https://doi.org/10.5858/arpa.2014-0250-CP

Bacher, J., Schumm, J.W., 1998. Development of highly polymorphic pentanucleotide tandem repeat loci with low stutter. Profiles DNA 2, 3.

Bailey, S.F., Scheible, M.K., Williams, C., Silva, D.S.B.S., Hoggan, M., Eichman, C., Faith, S.A., 2017. Secure and robust cloud computing for high-throughput forensic microsatellite sequence analysis and databasing. Forensic Sci. Int. Genet. 31, 40–47. https://doi.org/10.1016/j.fsigen.2017.08.008

Balding, D.J., 2013. Evaluation of mixed-source, low-template DNA profiles in forensic science. Proc. Natl. Acad. Sci. 110, 12241–12246. https://doi.org/10.1073/pnas.1219739110

Ballantyne, K.N., Oorschot, R.A. van, Mitchell, R.J., 2008. Reduce optimisation time and effort: Taguchi experimental design methods. Forensic Sci. Int. Genet. Suppl. Ser. 1, 7–8. https://doi.org/10.1016/j.fsigss.2007.10.050

Bär, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Lincoln, P., Mayr, W., Olaisen, B., 1997. DNA recommendations. Int. J. Legal Med. 110, 175–176. https://doi.org/10.1007/s004140050061

Barrio, P.A., Martín, P., Alonso, A., Müller, P., Bodner, M., Berger, B., Parson, W., Budowle, B., 2019. Massively parallel sequence data of 31 autosomal STR loci from 496 Spanish individuals revealed concordance with CE-STR technology and enhanced discrimination power. Forensic Sci. Int. Genet. https://doi.org/10.1016/j.fsigen.2019.06.009

Beridze, T., 1986. Satellite DNA. Springer-Verlag, Berlin Heidelberg.

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 Sci. Int. Genet. 24, 97–102. https://doi.org/10.1016/j.fsigen.2016.06.008

Boland, C.R., Goel, A., 2010. Microsatellite Instability in Colorectal Cancer. Gastroenterology 138, 2073-2087.e3. https://doi.org/10.1053/j.gastro.2009.12.064

Boleda, M.D., Briones, P., Farrés, J., Tyfield, L., Pi, R., 1996. Experimental design: a useful tool for PCR Optimization. BioTechniques 21, 134–140.

Bornman, D., Hester, M., Schuetter, J., Kasoji, M., Minard-Smith, A., Barden, C., Nelson, S., Godbold, G., Baker, C., Yang, B., Walther, J., Tornes, I., Yan, P., Rodriguez, B., Bundschuh, R., Dickens, M., Young, B., Faith, S., 2012. Short-read, high-throughput sequencing technology for STR genotyping. BioTechniques. https://doi.org/10.2144/000113857

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.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R., Cavalli-Sforza, L.L., 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368, 455–457. https://doi.org/10.1038/368455a0

Box, G.E.P., 1954. The exploration and exploitation of Response surfaces. Biometrics 16–60.

Box, G.E.P., Draper, N.R., 1987. Empirical model-building and response surfaces. Wiley.

Bright, J.-A., Stevenson, K.E., Coble, M.D., Hill, C.R., Curran, J.M., Buckleton, J.S., 2014. Characterising the STR locus D6S1043 and examination of its effect on stutter rates. Forensic Sci. Int. Genet. 8, 20–23. https://doi.org/10.1016/j.fsigen.2013.06.012

Brookes, C., Bright, J.-A., Harbison, S., Buckleton, J., 2012. Characterising stutter in forensic STR multiplexes. Forensic Sci. Int. Genet. 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

Bundesamtes für Sicherheit in der Informationstechnik, 2006. Sicherheit von Webanwendungen - Maßnahmenkatalog und Best Practices.

Buschiazzo, E., Gemmell, N.J., 2006. The rise, fall and renaissance of microsatellites in eukaryotic genomes. BioEssays News Rev. Mol. Cell. Dev. Biol. 28, 1040–1050. https://doi.org/10.1002/bies.20470

Bushnell, B., Rood, J., Singer, E., 2017. BBMerge – Accurate paired shotgun read merging via overlap. PLOS ONE 12, e0185056. https://doi.org/10.1371/journal.pone.0185056

Butler, J.M., 2014. Advanced Topics in Forensic DNA Typing: Interpretation. Academic Press.

Butler, J.M., 2011. Advanced Topics in Forensic DNA Typing: Methodology. Academic Press.

Butler, J.M., 2010. Fundamentals of forensic DNA typing. Academic Press.

Butler, J.M., 2005. Constructing STR multiplex assays, in: Forensic DNA Typing Protocols. Springer, pp. 53–65.

Butler, J.M., Shen, Y., McCord, B.R., others, 2003. The development of reduced size STR amplicons as tools for analysis of degraded DNA. J. Forensic Sci. 48, 1054–1064.

Buxmann, P., Hess, T., Lehmann, S., 2008. Software as a Service. Wirtschaftsinformatik 50, 500–503. https://doi.org/10.1007/s11576-008-0095-0

Caetano-Anollés, G., 1998. DAF optimization using Taguchi methods and the effect of thermal cycling parameters on DNA amplification. BioTechniques 25, 472–6, 478–80.

Calus, S.T., Ijaz, U.Z., Pinto, A.J., 2018. NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform. GigaScience 7. https://doi.org/10.1093/gigascience/giy140

Cao, M.D., Tasker, E., Willadsen, K., Imelfort, M., Vishwanathan, S., Sureshkumar, S., Balasubramanian, S., Bodén, M., 2014. Inferring short tandem repeat variation from paired-end short reads. Nucleic Acids Res. 42, e16. https://doi.org/10.1093/nar/gkt1313

Caratti, S., Turrina, S., Ferrian, M., Cosentino, E., Leo, D.D., 2015. MiSeq FGx sequencing system: A new platform for forensic genetics. Forensic Sci. Int. Genet. Suppl. Ser. 5, e98–e100. https://doi.org/10.1016/j.fsigss.2015.09.040

Celani de Souza, H.J., Moyses, C.B., Pontes, F.J., Duarte, R.N., Sanches da Silva, C.E., Alberto, F.L., Ferreira, U.R., Silva, M.B., 2011. Molecular assay optimized by Taguchi experimental design method for venous thromboembolism investigation. Mol. Cell. Probes 25, 231–237. https://doi.org/10.1016/j.mcp.2011.08.001

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 Res. 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 Sci. Int. Genet. 20, 20–29. https://doi.org/10.1016/j.fsigen.2015.09.009

Clark, J.M., 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res. 16, 9677–9686. https://doi.org/10.1093/nar/16.20.9677

Clayton, T.M., Whitaker, J.P., Sparkes, R., Gill, P., 1998. Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Sci. Int. 91, 55–70. https://doi.org/10.1016/S0379-0738(97)00175-8

Cobb, B.D., Clarkson, J.M., 1994. A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods. Nucleic Acids Res. 22, 3801–3805.

Collins, J.R., Stephens, R.M., Gold, B., Long, B., Dean, M., Burt, S.K., 2003. An exhaustive DNA micro-satellite map of the human genome using high performance computing. Genomics 82, 10–19. https://doi.org/10.1016/S0888-7543(03)00076-4

Cornelis, S., Willems, S., Neste, C.V., Tytgat, O., Weymaere, J., Plaetsen, A.-S.V., Deforce, D., Nieuwerburgh, F.V., 2018. Forensic STR profiling using Oxford Nanopore Technologies' MinION sequencer. bioRxiv 433151. 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 Sci. Int. Genet. Suppl. Ser. 4, e218–e219. https://doi.org/10.1016/j.fsigss.2013.10.112

Daniels, G., 2013. Human Blood Groups. John Wiley & Sons.

Daunay, A., Duval, A., Baudrin, L.G., Buhard, O., Renault, V., Deleuze, J.-F., How-Kit, A., 2019. Low temperature isothermal amplification of microsatellites drastically reduces stutter artifact formation and improves microsatellite instability detection in cancer. Nucleic Acids Res. https://doi.org/10.1093/nar/gkz811

Dexheimer, T.S., 2013. DNA Repair Pathways and Mechanisms, in: Mathews, L.A., Cabarcas, S.M., Hurt, E.M. (Eds.), DNA Repair of Cancer Stem Cells. Springer Netherlands, Dordrecht, pp. 19–32. https://doi.org/10.1007/978-94-007-4590-2_2

Dobay, O., Juhász, E., Ungvári, A., Jeney, C., Amyes, S.G.B., Nagy, K., 2009. Taguchi optimisation of a multiplex pneumococcal serotyping PCR and description of 11 novel serotyping primers. Acta Microbiol. Immunol. Hung. 56, 327–338. https://doi.org/10.1556/AMicr.56.2009.4.3

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, M.C., Gibbs, R.A., 1994. Multiplex PCR: advantages, development, and applications. Genome Res. 3, S65–S75.

Ellegren, H., 2004. Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 5, 435–445. https://doi.org/10.1038/nrg1348

Eriksson, L., 2008. Design of Experiments: Principles and Applications. Umetrics Academy.

Faith, S.A., Scheible, M., 2016. Analyzing Data from Next Generation Sequencers Using the PowerSeq® Auto/Mito/Y System [WWW Document]. URL https://www.promega.de/resources/profiles-in-dna/2016/analyzing-data-from-next-generation-sequencers-using-the-powerseq-automitoy-system/ (accessed 8.7.17).

Ferguson-Smith, M.A., Aitken, D.A., Turleau, C., de Grouchy, J., 1976. Localisation of the human ABO: Np-1: AK-1 linkage group by regional assignment of AK-1 to 9q34. Hum. Genet. 34, 35–43.

Fisher, S.R.A., 1947. The design of experiments. Oliver and Boyd.

Fondon, J.W., Martin, A., Richards, S., Gibbs, R.A., Mittelman, D., 2012. Analysis of Microsatellite Variation in Drosophila melanogaster with Population-Scale Genome Sequencing. PLOS ONE 7, e33036. https://doi.org/10.1371/journal.pone.0033036

Fordyce, S.L., Mogensen, H.S., Børsting, C., Lagacé, R.E., Chang, C.-W., Rajagopalan, N., Morling, N., 2015. Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™. Forensic Sci. Int. Genet. 14, 132–140. https://doi.org/10.1016/j.fsigen.2014.09.020

Friis, S.L., Buchard, A., Rockenbauer, E., Børsting, C., Morling, N., 2016. Introduction of the Python script STRinNGS for analysis of STR regions in FASTQ or BAM files and expansion of the Danish STR sequence database to 11 STRs. Forensic Sci. Int. Genet. 21, 68–75. https://doi.org/10.1016/j.fsigen.2015.12.006

Fujii, K., Watahiki, H., Mita, Y., Iwashima, Y., Miyaguchi, H., Kitayama, T., Nakahara, H., Mizuno, N., Sekiguchi, K., 2016. Next-generation sequencing analysis of off-ladder alleles due to migration shift caused by sequence variation at D12S391 locus. Leg. Med. Tokyo Jpn. 22, 62–67. https://doi.org/10.1016/j.legalmed.2016.08.003

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

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

Gamma, E., 2001. Entwurfsmuster: Elemente wiederverwendbarer objektorientierter Software. Addison-Wesley.

Ganschow, S., 2016. Next Generation Sequencing für die forensische Genetik: Entwicklung und Evaluation eines High-Fidelity 10-plex STR-Assays. Presented at the 12th Annual Meeting of the DGAB, Bielefeld.

Ganschow, S., Silvery, J., Kalinowski, J., Tiemann, C., 2018. toaSTR: A web application for forensic STR genotyping by massively parallel sequencing. Forensic Sci. Int. Genet. 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 Sci. Int. Genet. 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., Ballard, D., Bodner, M., Borsuk, L., King, J., Parson, W., Phillips, C., 2019. Report from the STRAND Working Group on the 2019 STR sequence nomenclature meeting. Forensic Sci. Int. Genet. 43, 102165. https://doi.org/10.1016/j.fsigen.2019.102165

Gettings, K.B., Aponte, R.A., Kiesler, K.M., Vallone, P.M., 2015a. The next dimension in STR sequencing: Polymorphisms in flanking regions and their allelic associations. Forensic Sci. Int. Genet. Suppl. Ser. 5, e121–e123. https://doi.org/10.1016/j.fsigss.2015.09.049

Gettings, K.B., Aponte, R.A., Vallone, P.M., Butler, J.M., 2015b. STR allele sequence variation: Current knowledge and future issues. Forensic Sci. Int. Genet. 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 Sci. Int. Genet. 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 Sci. Int. Genet. 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 Sci. Int. Genet. 21, 15–21. https://doi.org/10.1016/j.fsigen.2015.11.005

Gibson-Daw, G., Crenshaw, K., McCord, B., 2018. Optimization of ultrahigh-speed multiplex PCR for forensic analysis. Anal. Bioanal. Chem. 410, 235–245. https://doi.org/10.1007/s00216-017-0715-x

Giese, H., Lam, R., Selden, R., Tan, E., 2009. Fast Multiplexed Polymerase Chain Reaction for Conventional and Microfluidic Short Tandem Repeat Analysis. J. Forensic Sci. 54, 1287–1296. https://doi.org/10.1111/j.1556-4029.2009.01200.x

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 Sci. Int. Genet. 18, 100–117. https://doi.org/10.1016/j.fsigen.2015.03.014

Gill, P., Jeffreys, A.J., Werrett, D.J., 1985. Forensic application of DNA `fingerprints'. Nature 318, 577–579. https://doi.org/10.1038/318577a0

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

Goubran, H., 2009. Blood group serology. ISBT Sci. Ser. 4, 1–5. https://doi.org/10.1111/j.1751-2824.2009.01206.x

Guo, F., Zhou, Y., Liu, F., Yu, J., Song, H., Shen, H., Zhao, B., Jia, F., Hou, G., Jiang, X., 2016. Evaluation of the Early Access STR Kit v1 on the Ion Torrent PGM™ platform. Forensic Sci. Int. Genet. 23, 111–120. https://doi.org/10.1016/j.fsigen.2016.04.004

Gymrek, M., Erlich, Y., 2013. Profiling Short Tandem Repeats from Short Reads, in: Shomron, N. (Ed.), Deep Sequencing Data Analysis, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 113–135. https://doi.org/10.1007/978-1-62703-514-9_7

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

Haaland, P.D., 1989. Experimental Design in Biotechnology. CRC Press.

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

Harris, R.S., Cechova, M., Makova, K.D., 2019. Noise-cancelling repeat finder: uncovering tandem repeats in error-prone long-read sequencing data. Bioinformatics 35, 4809–4811. https://doi.org/10.1093/bioinformatics/btz484

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.

Highnam, G., Franck, C., Martin, A., Stephens, C., Puthige, A., Mittelman, D., 2013. Accurate human microsatellite genotypes from high-throughput resequencing data using informed error profiles. Nucleic Acids Res. 41, e32–e32. https://doi.org/10.1093/nar/gks981

Hill, C.R., Duewer, D.L., Kline, M.C., Sprecher, C.J., McLaren, R.S., Rabbach, D.R., Krenke, B.E., Ensenberger, M.G., Fulmer, P.M., Storts, D.R., Butler, J.M., 2011. Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems. Forensic Sci. Int. Genet. 5, 269–275. https://doi.org/10.1016/j.fsigen.2010.03.014

Hill, J., 2013. Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis, in: Forensic DNA Analysis. CRC Press, pp. 181–198. https://doi.org/10.1201/b15361-13

Holland, M.M., Parsons, T.J., 1999. Mitochondrial DNA Sequence Analysis - Validation and Use for Forensic Casework. Forensic Sci. Rev. 11, 21–50.

Hommelsheim, C.M., Frantzeskakis, L., Huang, M., Ülker, B., 2014. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. Sci. Rep. 4, 1–13. https://doi.org/10.1038/srep05052

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 Sci. Int. Genet. 27, 27–40. https://doi.org/10.1016/j.fsigen.2016.11.007

Hughes-Stamm, S., Ashton, K., van Daal, A., 2011. Assessment of DNA degradation and the genotyping success of highly degraded samples. Int. J. Legal Med. 125, 341–348. https://doi.org/10.1007/s00414-010-0455-3

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 Sci. Res. 3, 111–123. https://doi.org/10.1080/20961790.2018.1446672

IEEE, 2009. Standard for Software Design Descriptions, IEEE 1016-2009.

IEEE, 1998. Recommended Practice for Software Requirements Specifications, IEEE 830-1998.

Illumina, 2018. MiSeq system specification sheet, Pub. No. 770-2011-001-B QB #6456.

Illumina, 2015. ForenSeq DNA Signature Prep Reference Guide, Catalog # TG-450-9001DOC.

Ilzarbe, L., Álvarez, M.J., Viles, E., Tanco, M., 2008. Practical applications of design of experiments in the field of engineering: a bibliographical review. Qual. Reliab. Eng. Int. 24, 417–428. https://doi.org/10.1002/qre.909

International Organization for Standardization, 2018. DIN EN ISO/IEC 17025:2018-03, Allgemeine Anforderungen an die Kompetenz von Prüf- und Kalibrierlaboratorien. Beuth Verlag GmbH. https://doi.org/10.31030/2731745

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 Sci. Int. Genet. 28, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Jeffreys, A.J., Wilson, V., Neumann, R., Keyte, J., 1988. Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. Nucleic Acids Res. 16, 10953–10971. https://doi.org/10.1093/nar/16.23.10953

Jeffreys, A.J., Wilson, V., Thein, S.L., 1985. Hypervariable 'minisatellite' regions in human DNA. Nature 314, 67–73. https://doi.org/10.1038/314067a0

Jobling, M.A., Gill, P., 2004. Encoded evidence: DNA in forensic analysis. Nat. Rev. Genet. 5, 739–751. https://doi.org/10.1038/nrg1455

Kalendar, R., Khassenov, B., Ramankulov, Y., Samuilova, O., Ivanov, K.I., 2017. FastPCR: An in silico tool for fast primer and probe design and advanced sequence analysis. Genomics 109, 312–319. https://doi.org/10.1016/j.ygeno.2017.05.005

Kayser, M., Parson, W., 2017. Transitioning from Forensic Genetics to Forensic Genomics. Genes 9, 3. https://doi.org/10.3390/genes9010003

Kelkar, Y.D., Strubczewski, N., Hile, S.E., Chiaromonte, F., Eckert, K.A., Makova, K.D., 2010. What Is a Microsatellite: A Computational and Experimental Definition Based upon Repeat Mutational Behavior at A/T and GT/AC Repeats. Genome Biol. Evol. 2, 620–635. https://doi.org/10.1093/gbe/evq046

Kim, E.H., Lee, H.Y., Yang, I.S., Jung, S.-E., Yang, W.I., Shin, K.-J., 2016. Massively parallel sequencing of 17 commonly used forensic autosomal STRs and amelogenin with small amplicons. Forensic Sci. Int. Genet. 22, 1–7. https://doi.org/10.1016/j.fsigen.2016.01.001

Kirby, L.T., 1993. DNA Fingerprinting: An Introduction. Oxford University Press.

Klintschar, M., Wiegand, P., 2003. Polymerase slippage in relation to the uniformity of tetrameric repeat stretches. Forensic Sci. Int. 135, 163–166. https://doi.org/10.1016/s0379-0738(03)00201-9

Kloosterman, A.D., Budowle, B., Daselaar, P., 1993. PCR-amplification and detection of the human D1S80 VNTR locus. Int. J. Legal Med. 105, 257–264. https://doi.org/10.1007/BF01370382

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 Sci. Int. Genet. 36, 77–85. https://doi.org/10.1016/j.fsigen.2018.05.007

Kruglyak, S., Durrett, R.T., Schug, M.D., Aquadro, C.F., 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc. Natl. Acad. Sci. 95, 10774–10778. https://doi.org/10.1073/pnas.95.18.10774

Kuhn, R.M., Haussler, D., Kent, W.J., 2013. The UCSC genome browser and associated tools. Brief. Bioinform. 14, 144–161. https://doi.org/10.1093/bib/bbs038

Lai, Y., Sun, F., 2003. The Relationship Between Microsatellite Slippage Mutation Rate and the Number of Repeat Units. Mol. Biol. Evol. 20, 2123–2131. https://doi.org/10.1093/molbev/msg228

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer,

K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Szustakowki, J., International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. https://doi.org/10.1038/35057062

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

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

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

Lazaruk, K., Walsh, P.S., Oaks, F., Gilbert, D., Rosenblum, B.B., Menchen, S., Scheibler, D., Wenz, H.M., Holt, C., Wallin, J., 1998. Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. Electrophoresis 19, 86–93. https://doi.org/10.1002/elps.1150190116

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 Sci. Int. Genet. 26, 66–69. https://doi.org/10.1016/j.fsigen.2016.10.006

Legendre, M., Pochet, N., Pak, T., Verstrepen, K.J., 2007. Sequence-based estimation of minisatellite and microsatellite repeat variability. Genome Res. 17, 1787–1796. https://doi.org/10.1101/gr.6554007

Levinson, G., Gutman, G.A., 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4, 203–221. https://doi.org/10.1093/oxfordjournals.molbev.a040442

Li, C., Chng, K.R., Boey, E.J.H., Ng, A.H.Q., Wilm, A., Nagarajan, N., 2016. INC-Seq: accurate single molecule reads using nanopore sequencing. GigaScience 5, 34. https://doi.org/10.1186/s13742-016-0140-7

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Homer, N., 2010. A survey of sequence alignment algorithms for next-generation sequencing. Brief. Bioinform. 11, 473–483. https://doi.org/10.1093/bib/bbq015

Liu, Y.-Y., Harbison, S., 2018. A review of bioinformatic methods for forensic DNA analyses. Forensic Sci. Int. Genet. 33, 117–128. https://doi.org/10.1016/j.fsigen.2017.12.005

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

Marciano, M.A., Adelman, J.D., 2017. PACE: Probabilistic Assessment for Contributor Estimation— A machine learning-based assessment of the number of contributors in DNA mixtures. Forensic Sci. Int. Genet. 27, 82–91. https://doi.org/10.1016/j.fsigen.2016.11.006

McInerney, P., Adams, P., Hadi, M.Z., 2014. Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. Mol. Biol. Int. 2014, 1–8. https://doi.org/10.1155/2014/287430

McPherson, M., Møller, S., 2006. PCR. Taylor & Francis Ltd., Hoboken.

Menin, J.F., Nichols, N., 2013. Multiplex PCR using Q5 High-Fidelity DNA Polymerase. N. Engl. Biolabs Appl. Note.

Metzker, M.L., 2010. Sequencing technologies — the next generation. Nat. Rev. Genet. 11, 31–46. https://doi.org/10.1038/nrg2626

Mirkin, S.M., 2007. Expandable DNA repeats and human disease. Nature 447, 932–940. https://doi.org/10.1038/nature05977

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 Sci. Int. Genet. 36, 95–103. https://doi.org/10.1016/j.fsigen.2018.06.016

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H., 1986. Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. Cold Spring Harb. Symp. Quant. Biol. 51, 263–273. https://doi.org/10.1101/SQB.1986.051.01.032

Murray, V., Monchawin, C., England, P.R., 1993. The determination of the sequences present in the shadow bands of a dinucleotide repeat PCR. Nucleic Acids Res. 21, 2395–2398.

Nelder, J.A., Mead, R., 1965. A Simplex Method for Function Minimization. Comput. J. 7, 308–313. https://doi.org/10.1093/comjnl/7.4.308

New England Biolabs, 2012. Protocol for Q5 Hot Start High-Fidelity 2X Master Mix [WWW Document]. URL https://www.neb.com/protocols/2012/08/30/protocol-for-q5-hot-start-high-fidelity-2x-master-mix-m0494 (accessed 7.20.18).

Niens, M., Spijker, G.T., Diepstra, A., Meerman, G.J.T., 2005. A factorial experiment for optimizing the PCR conditions in routine genotyping. Biotechnol. Appl. Biochem. 42, 157–162. https://doi.org/10.1042/BA20040192

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 Sci. Int. Genet. 25, 214–226. https://doi.org/10.1016/j.fsigen.2016.09.007

Olejniczak, M., Krzyzosiak, W.J., 2006. Genotyping of simple sequence repeats--factors implicated in shadow band generation revisited. Electrophoresis 27, 3724–3734. https://doi.org/10.1002/elps.200600136

Parson, W., 2018. Age Estimation with DNA: From Forensic DNA Fingerprinting to Forensic (Epi)Genomics: A Mini-Review. Gerontology 64, 326–332. https://doi.org/10.1159/000486239

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 Sci. Int. Genet. 22, 54–63. https://doi.org/10.1016/j.fsigen.2016.01.009

Pechal, J.L., Crippen, T.L., Benbow, M.E., Tarone, A.M., Dowd, S., Tomberlin, J.K., 2014. The potential use of bacterial community succession in forensics as described by high throughput metagenomic sequencing. Int. J. Legal Med. 128, 193–205. https://doi.org/10.1007/s00414-013-0872-1

Peltomäki, P., Lothe, R.A., Aaltonen, L.A., Pylkkänen, L., Nyström-Lahti, M., Seruca, R., David, L., Holm, R., Ryberg, D., Haugen, A., Brøgger, A., Børresen, A.-L., Chapelle, A. de la, 1993. Microsatellite Instability Is Associated with Tumors That Characterize the Hereditary Non-Polyposis Colorectal Carcinoma Syndrome. Cancer Res. 53, 5853–5855.

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 Sci. Int. Genet. 34, 162–169. https://doi.org/10.1016/j.fsigen.2018.02.017

Pilipauskas, D.R., 1999. Can the time from synthesis design to validated chemistry be shortened? Med. Res. Rev. 19, 463–474. https://doi.org/10.1002/(SICI)1098-1128(199909)19:5<463::AID-MED11>3.0.CO;2-3

Plesivkova, D., Richards, R., Harbison, S., 2019. A review of the potential of the MinION™ single-molecule sequencing system for forensic applications. Wiley Interdiscip. Rev. Forensic Sci. 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, 1432–1432. https://doi.org/10.1093/nar/20.6.1432

Potapov, V., Ong, J.L., 2017. Examining Sources of Error in PCR by Single-Molecule Sequencing. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0169774

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

Rao, R.S., Kumar, C.G., Prakasham, R.S., Hobbs, P.J., 2008. The Taguchi methodology as a statistical tool for biotechnological applications: A critical appraisal. Biotechnol. J. 3, 510–523. https://doi.org/10.1002/biot.200700201

Raz, O., Biezuner, T., Spiro, A., Amir, S., Milo, L., Titelman, A., Onn, A., Chapal-Ilani, N., Tao, L., Marx, T., Feige, U., Shapiro, E., 2018. Short Tandem Repeat stutter model inferred from direct measurement of in vitro stutter noise. bioRxiv 065110. https://doi.org/10.1101/065110

Reid, Y., Storts, D., Riss, T., Minor, L., 2004. Authentication of Human Cell Lines by STR DNA Profiling Analysis, in: Sittampalam, G.S., Coussens, N.P., Brimacombe, K., Grossman, A., Arkin, M., Auld, D., Austin, C., Baell, J., Bejcek, B., Caaveiro, J.M.M., Chung, T.D.Y., Dahlin, J.L., Devanaryan, V., Foley, T.L., Glicksman, M., Hall, M.D., Haas, J.V., Inglese, J., Iversen, P.W., Kahl, S.D., Kales, S.C., Lal-Nag, M., Li, Z., McGee, J., McManus, O., Riss, T., Trask, O.J., Weidner, J.R., Wildey, M.J., Xia, M., Xu, X. (Eds.), Assay Guidance Manual. Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda (MD).

Riman, S., Iyer, H., Borsuk, L., Gettings, K., Vallone, P., 2017. Investigating the effects of different library preparation protocols on STR sequencing. Forensic Sci. Int. Genet. Suppl. Ser. https://doi.org/10.1016/j.fsigss.2017.09.155

Rolf, B., Schürenkamp, M., Junge, A., Brinkmann, B., 1997. Sequence polymorphism at the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus. Int. J. Legal Med. 110, 69–72. https://doi.org/10.1007/s004140050033

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

Rudin, N., Inman, K., 2002. An introduction to forensic DNA analysis, 2nd ed. ed. CRC Press, Boca Raton, Fla.

Santibáñez-Koref, M.F., Gangeswaran, R., Hancock, J.M., 2001. A Relationship Between Lengths of Microsatellites and Nearby Substitution Rates in Mammalian Genomes. Mol. Biol. Evol. 18, 2119–2123. https://doi.org/10.1093/oxfordjournals.molbev.a003753

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

Schmitt, C., Schmutzler, A., Prinz, M., Staak, M., 1994. High sensitive DNA typing approaches for the analysis of forensic evidence: comparison of nested variable number of tandem repeats (VNTR) amplification and a short tandem repeats (STR) polymorphism. Forensic Sci. Int. 66, 129–141. https://doi.org/10.1016/0379-0738(94)90337-9

Schneider, P., 2005. Expansion of the European Standard Set of DNA Database Loci—the Current Situation. Profiles DNA. https://worldwide.promega.com/~/media/files/resources/profiles%20in%20dna/1201/expansion of the european standard set.pdf

Schneider, P.M., 2007. Scientific standards for studies in forensic genetics. Forensic Sci. Int., Research in Forensic Medicine and Forensic Sciences: Basic Principles and Application in Practice 165, 238–243. https://doi.org/10.1016/j.forsciint.2006.06.067

Scientific Working Group on DNA Analysis Methods, 2016. Validation Guidelines for DNA Analysis Methods.

Scientific Working Group on DNA Analysis Methods, 2010. SWGDAM Interpretation Guidelines for Autosomal STR Typing.

Sczyrba, A., Backofen, R., Belmann, P., Boutros, M., Eils, R., Eisenacher, M., Goesmann, A., Gru, B., Korbel, J., Kru, J., 2018. Federated de.NBI Cloud. https://www.denbi.de/images/About/Network/Poster/CAU_deNBI_Cloud.pdf

Seo, S.B., Ge, J., King, J.L., Budowle, B., 2014. Reduction of stutter ratios in short tandem repeat loci typing of low copy number DNA samples. Forensic Sci. Int. Genet. 8, 213–218. https://doi.org/10.1016/j.fsigen.2013.10.004

Seo, S.B., King, J., Budowle, B., 2012. A Systematic Approach for Reduction of STR Stutter Peak Heights from LCN Samples. Poster presentation at the 23rd International Symposium on Human Identification.

Shafique, M., Shahzad, M.S., Rahman, Z., Shan, M.A., Perveen, R., Shahzad, M., Hussain, M., Shahid, A.A., Husnain, T., 2016. Development of new PCR multiplex system by the simultaneous detection of 10 miniSTRs, SE33, Penta E, Penta D, and four Y-STRs. Int. J. Legal Med. https://doi.org/10.1007/s00414-016-1372-x

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

Shinde, D., Lai, Y., Sun, F., Arnheim, N., 2003. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. Nucleic Acids Res. 31, 974–980. https://doi.org/10.1093/nar/gkg178

Silva, D.S.B.S., Sawitzki, F.R., Scheible, M.K.R., Bailey, S.F., Alho, C.S., Faith, S.A., 2018. Paternity testing using massively parallel sequencing and the PowerSeq™ AUTO/Y system for short tandem repeat sequencing. Electrophoresis 39, 2669–2673. https://doi.org/10.1002/elps.201800072

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 Sci. Int. Genet. 46. https://doi.org/10.1016/j.fsigen.2020.102236

Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., Timp, W., 2017. Detecting DNA cytosine methylation using nanopore sequencing. Nat. Methods 14, 407–410. https://doi.org/10.1038/nmeth.4184

Smith, J., Moretti, T., DiZinno, J., 2000. DNA Typing Protocols: Molecular Biology and Forensic Analysis, 1st edition. ed. Eaton Publishing Company, Natick, MA.

Southern, E.M., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503–517. https://doi.org/10.1016/S0022-2836(75)80083-0

Storry, J.R., Olsson, M.L., 2009. The ABO blood group system revisited: a review and update. Immunohematology 25, 48–59.

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, 636–8, 640–1.

Sun, J.X., Helgason, A., Masson, G., Ebenesersdóttir, S.S., Li, H., Mallick, S., Gnerre, S., Patterson, N., Kong, A., Reich, D., Stefansson, K., 2012. A direct characterization of human mutation based on microsatellites. Nat. Genet. 44, 1161–1165. https://doi.org/10.1038/ng.2398

Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M., Imanaka, T., 1997. Characterization of DNA Polymerase from Pyrococcus sp. Strain KOD1 and Its Application to PCR. Appl Env. Microbiol 63, 7.

The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. Nature 526, 68–74. https://doi.org/10.1038/nature15393

Thermo Fisher Scientific, 2018. Ion GeneStudio S5 specification sheet, COL32218 0118.

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

Treangen, T.J., Salzberg, S.L., 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. 13, 36–46. https://doi.org/10.1038/nrg3117

Tully, G., Sullivan, K.M., Gill, P., 1993. Analysis of 6 VNTR loci by 'multiplex' PCR and automated fluorescent detection. Hum. Genet. 92, 554–562. https://doi.org/10.1007/BF00420938

van der Gaag, K.J., de Knijff, P., 2015. Forensic nomenclature for short tandem repeats updated for sequencing. Forensic Sci. Int. Genet. Suppl. Ser. 5, e542–e544. https://doi.org/10.1016/j.fsigss.2015.09.214

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

Vergnaud, G., 2000. Minisatellites: Mutability and Genome Architecture. Genome Res. 10, 899–907. https://doi.org/10.1101/gr.10.7.899

Verogen, 2018. Technical Note: MiSeq FGx Reagent Micro Kit [WWW Document]. URL https://verogen.com/wp-content/uploads/2018/07/MiSeq-FGx-Reagent-Micro-Kit-tech-note-VD2018016.pdf (accessed 11.29.19).

Viguera, E., Canceill, D., Ehrlich, S.D., 2001. Replication slippage involves DNA polymerase pausing and dissociation. EMBO J. 20, 2587–2595. https://doi.org/10.1093/emboj/20.10.2587

Villarreal Camacho, J.L., Mendoza Torres, E., Cadena, C., Prieto, J., Varela Prieto, L.L., Villanueva Torregroza, D.A., 2013. The use of factorial design, image analysis, and an efficiency calculation for multiplex PCR optimization. J. Clin. Lab. Anal. 27, 249–252. https://doi.org/10.1002/jcla.21595

Vilsen, S.B., Tvedebrink, T., Eriksen, P.S., Bøsting, C., Hussing, C., Mogensen, H.S., Morling, N., 2018. Stutter analysis of complex STR MPS data. Forensic Sci. Int. Genet. 35, 107–112. https://doi.org/10.1016/j.fsigen.2018.04.003

Wadle, S., Lehnert, M., Rubenwolf, S., Zengerle, R., von Stetten, F., 2015. Real-time PCR probe optimization using design of experiments approach. Biomol. Detect. Quantif. 7, 1–8. https://doi.org/10.1016/j.bdq.2015.12.002

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. https://doi.org/10.1520/JFS15205J

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., Chang, C.-W., Lagacé, R.E., Calandro, L.M., Hennessy, L.K., 2012a. Developmental Validation of the AmpFℓSTR® Identifiler® Plus PCR Amplification Kit: An Established Multiplex Assay with Improved Performance. J. Forensic Sci. 57, 453–465. https://doi.org/10.1111/j.1556-4029.2011.01963.x

Wang, D.Y., Green, R.L., Lagacé, R.E., Oldroyd, N.J., Hennessy, L.K., Mulero, J.J., 2012b. 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., 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 Sci. Int. Genet. 7, 409–417. https://doi.org/10.1016/j.fsigen.2013.04.005

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., Lathrop, M., 1992. A second-generation linkage map of the human genome. Nature 359, 794–801. https://doi.org/10.1038/359794a0

Westen, A.A., Kraaijenbrink, T., Medina, E.A.R. de, Harteveld, J., Willemse, P., Zuniga, S.B., Gaag, K.J. van der, Weiler, N.E.C., Warnaar, J., Kayser, M., Sijen, T., Knijff, P. de, 2014. Comparing six commercial autosomal STR kits in a large Dutch population sample. Forensic Sci. Int. Genet. 10, 55–63. https://doi.org/10.1016/j.fsigen.2014.01.008

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.

Willems, T., Gymrek, M., Highnam, G., Mittelman, D., Erlich, Y., 2014. The landscape of human STR variation. Genome Res. 24, 1894–1904. https://doi.org/10.1101/gr.177774.114

Willuweit, S., Roewer, L., 2015. The new Y Chromosome Haplotype Reference Database. Forensic Sci. Int. Genet. 15, 43–48. https://doi.org/10.1016/j.fsigen.2014.11.024

Woerner, A.E., King, J.L., Budowle, B., 2017a. Flanking Variation Influences Rates of Stutter in Simple Repeats. Genes 8, 329. https://doi.org/10.3390/genes8110329

Woerner, A.E., King, J.L., Budowle, B., 2017b. Fast STR allele identification with STRait Razor 3.0. Forensic Sci. Int. Genet. 30, 18–23. https://doi.org/10.1016/j.fsigen.2017.05.008

Wrobel, G., Schlingemann, J., Hummerich, L., Kramer, H., Lichter, P., Hahn, M., 2003. Optimization of high-density cDNA-microarray protocols by 'design of experiments.' Nucleic Acids Res. 31, e67–e67. https://doi.org/10.1093/nar/gng067

Wu, H., Zhang, S., Xiao, C., 1998. Shadow bands in PCR amplification of trinucleotide repeats and their elimination. Zhonghua Yi Xue Yi Chuan Xue Za Zhi Zhonghua Yixue Yichuanxue Zazhi Chin. J. Med. Genet. 15, 42–45.

Xavier, C., Parson, W., 2017. Evaluation of the Illumina ForenSeq™ DNA Signature Prep Kit – MPS forensic application for the MiSeq FGx™ benchtop sequencer. Forensic Sci. Int. Genet. 28, 188–194. https://doi.org/10.1016/j.fsigen.2017.02.018

Xu, Xin, Peng, M., Fang, Z., Xu, Xiping, 2000. The direction of microsatellite mutations is dependent upon allele length. Nat. Genet. 24, 396–399. https://doi.org/10.1038/74238

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13, 134. https://doi.org/10.1186/1471-2105-13-134

Young, B., King, J.L., Budowle, B., Armogida, L., 2017. A technique for setting analytical thresholds in massively parallel sequencing-based forensic DNA analysis. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0178005

Yu, B., Zhang, C., 2011. In silico PCR analysis. Methods Mol. Biol. Clifton NJ 760, 91–107. https://doi.org/10.1007/978-1-61779-176-5_6

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 Sci. Int. Genet. 16, 38–47. https://doi.org/10.1016/j.fsigen.2014.11.022

Zhao, X., Li, H., Wang, Z., Ma, K., Cao, Y., Liu, W., 2016. Massively parallel sequencing of 10 autosomal STRs in Chinese using the ion torrent personal genome machine (PGM). Forensic Sci. Int. Genet. 25, 34–38. https://doi.org/10.1016/j.fsigen.2016.07.014

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 Sci. Int. Genet. 17, 122–128. https://doi.org/10.1016/j.fsigen.2015.04.002

Appendix

A1 toaSTR software demonstration

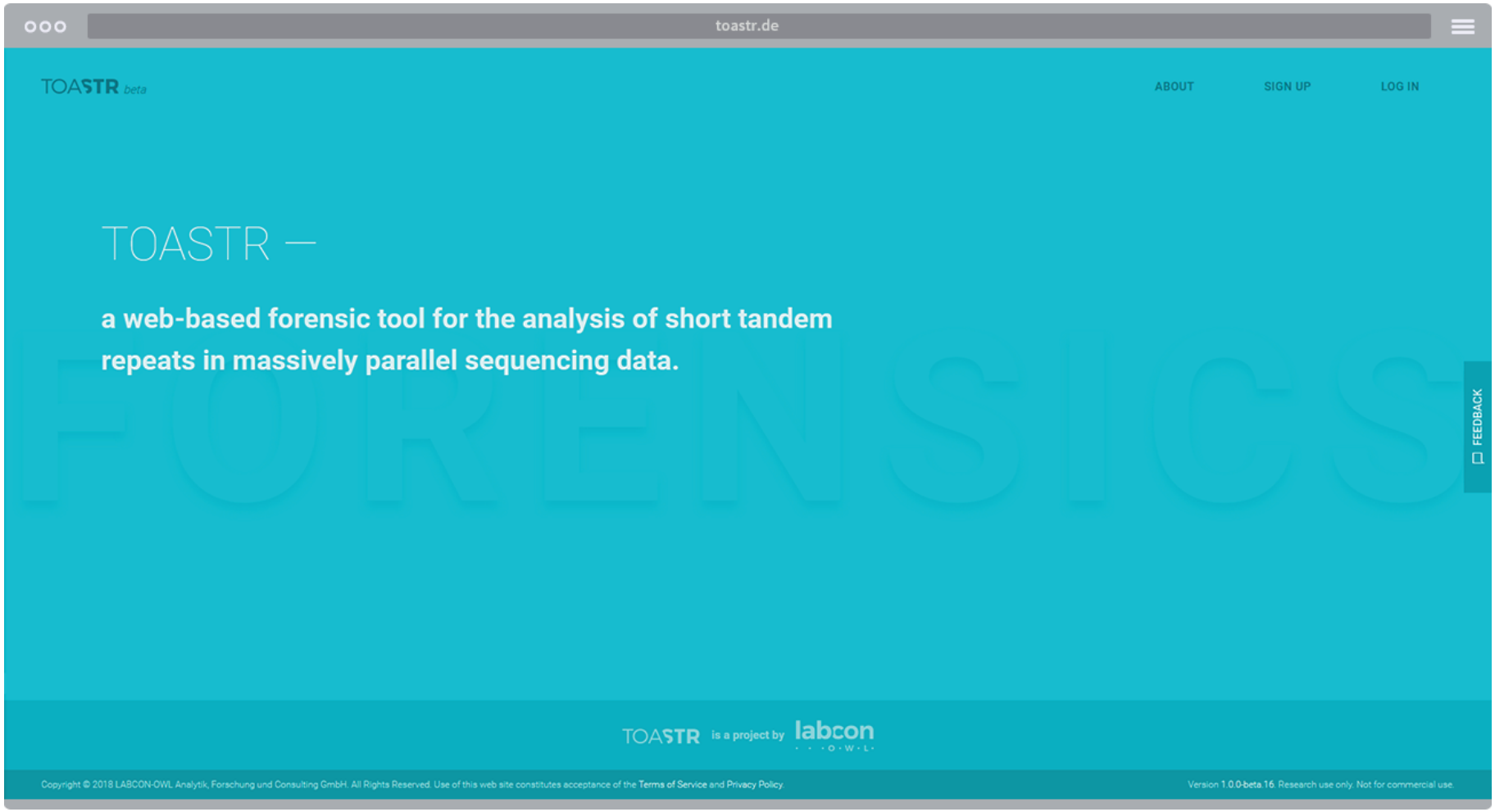


Figure 12. Homepage of the toaSTR web application

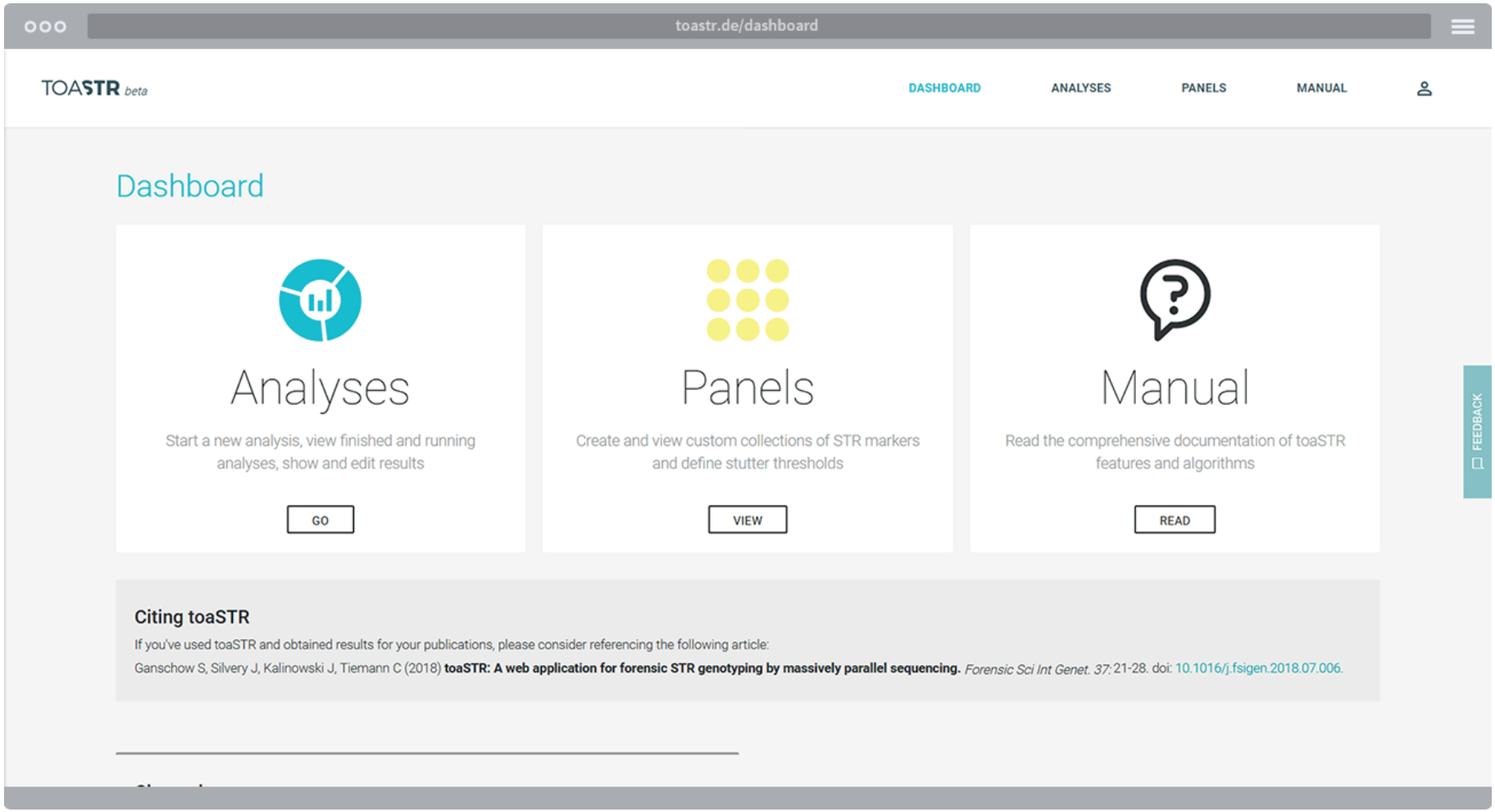


Figure 13. Dashboard page. After logging in as a user, all toaSTR functions are available via the central dashboard.

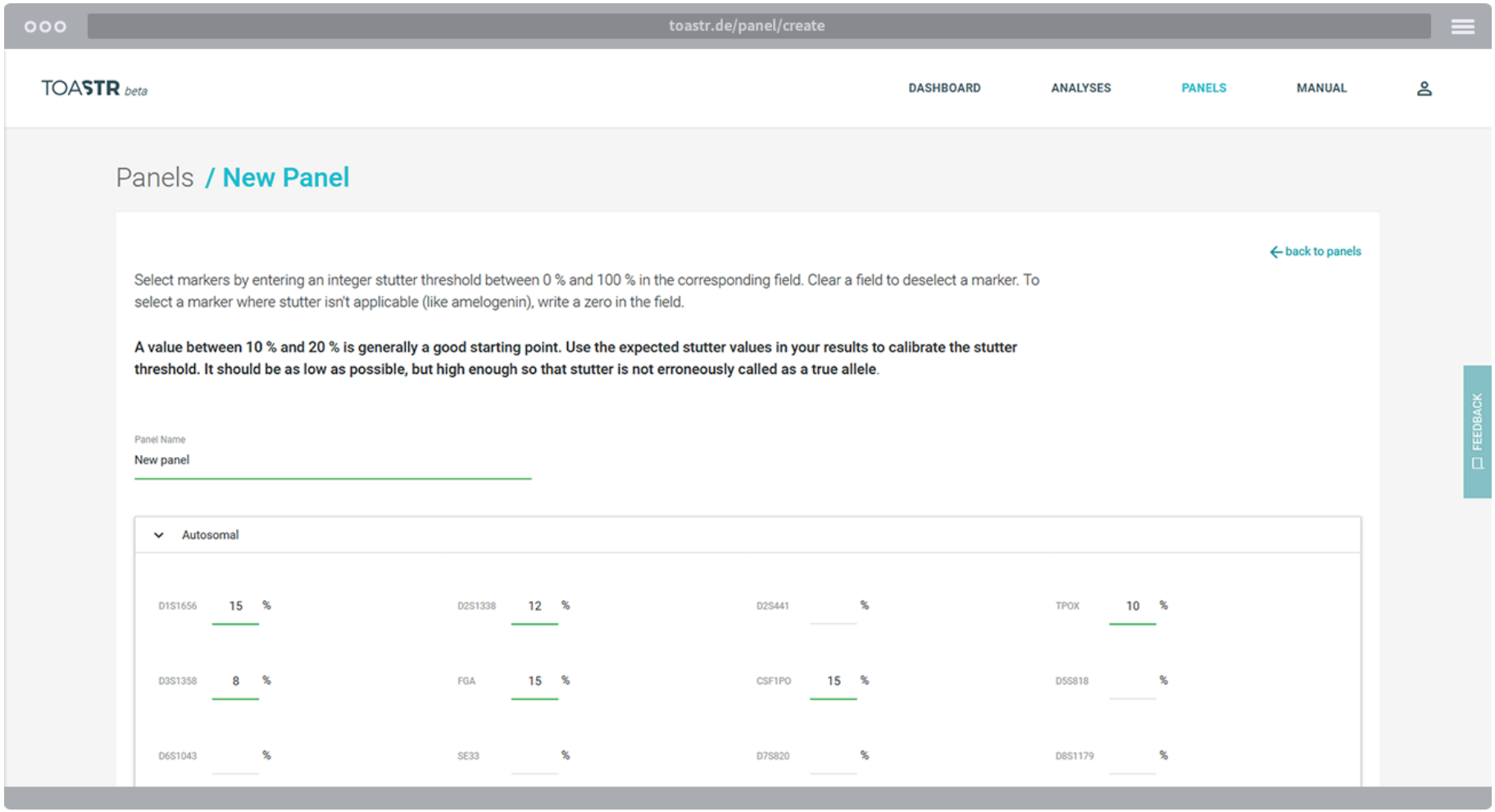


Figure 14. Designing panels in toaSTR. By creating a panel, the user can select STR markers to be analyzed. An individual stutter threshold can be defined for each marker.

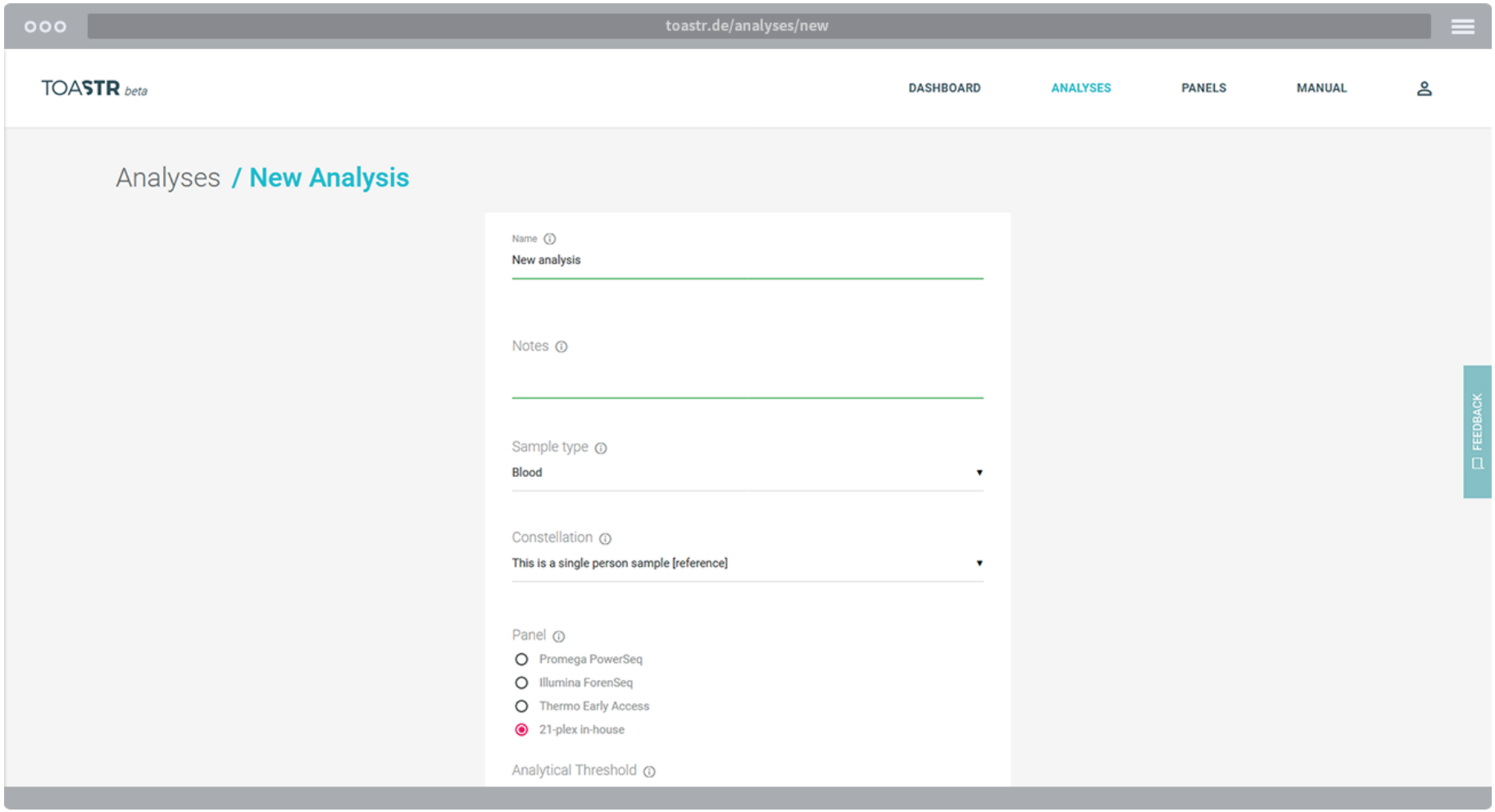
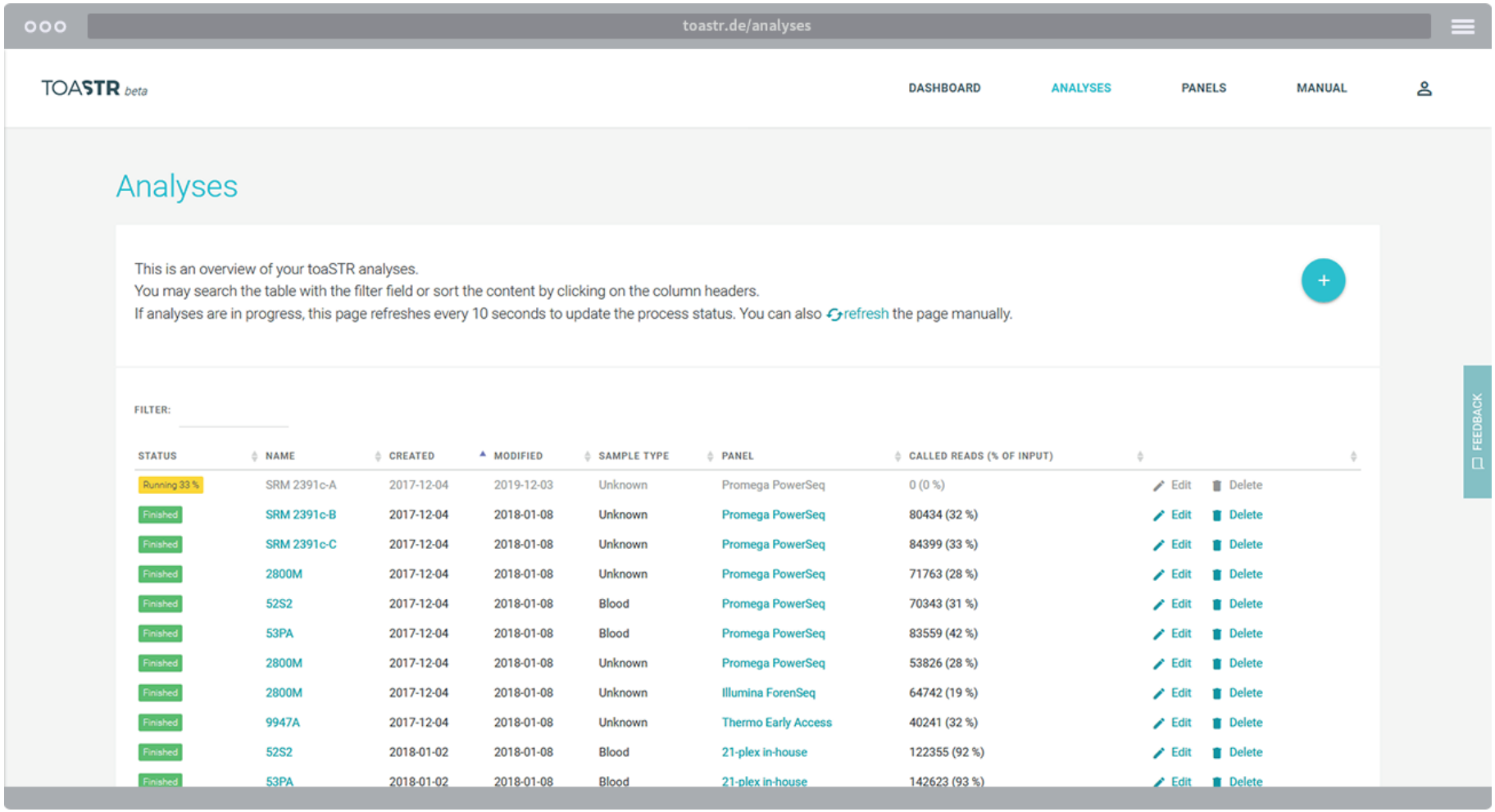


Figure 15. Starting a new analysis. Data upload is handled via a convenient web form that also enables to enter sample meta-data (e.g., sample name, notes, etc.). The user can select the pre-configured panel and adjust analysis parameters.



STATUS	NAME	CREATED	MODIFIED	SAMPLE TYPE	PANEL	CALLED READS (% OF INPUT)	
Running 33 %	SRM 2391c-A	2017-12-04	2019-12-03	Unknown	Promega PowerSeq	0 (0 %)	Edit Delete
Finished	SRM 2391c-B	2017-12-04	2018-01-08	Unknown	Promega PowerSeq	80434 (32 %)	Edit Delete
Finished	SRM 2391c-C	2017-12-04	2018-01-08	Unknown	Promega PowerSeq	84399 (33 %)	Edit Delete
Finished	2800M	2017-12-04	2018-01-08	Unknown	Promega PowerSeq	71763 (28 %)	Edit Delete
Finished	52S2	2017-12-04	2018-01-08	Blood	Promega PowerSeq	70343 (31 %)	Edit Delete
Finished	53PA	2017-12-04	2018-01-08	Blood	Promega PowerSeq	83559 (42 %)	Edit Delete
Finished	2800M	2017-12-04	2018-01-08	Unknown	Promega PowerSeq	53826 (28 %)	Edit Delete
Finished	2800M	2017-12-04	2018-01-08	Unknown	Illumina ForenSeq	64742 (19 %)	Edit Delete
Finished	9947A	2017-12-04	2018-01-08	Unknown	Thermo Early Access	40241 (32 %)	Edit Delete
Finished	52S2	2018-01-02	2018-01-08	Blood	21-plex in-house	122355 (92 %)	Edit Delete
Finished	53PA	2018-01-02	2018-01-08	Blood	21-plex in-house	142623 (93 %)	Edit Delete

Figure 16. The analyses overview page summarizes running and finished analyses.

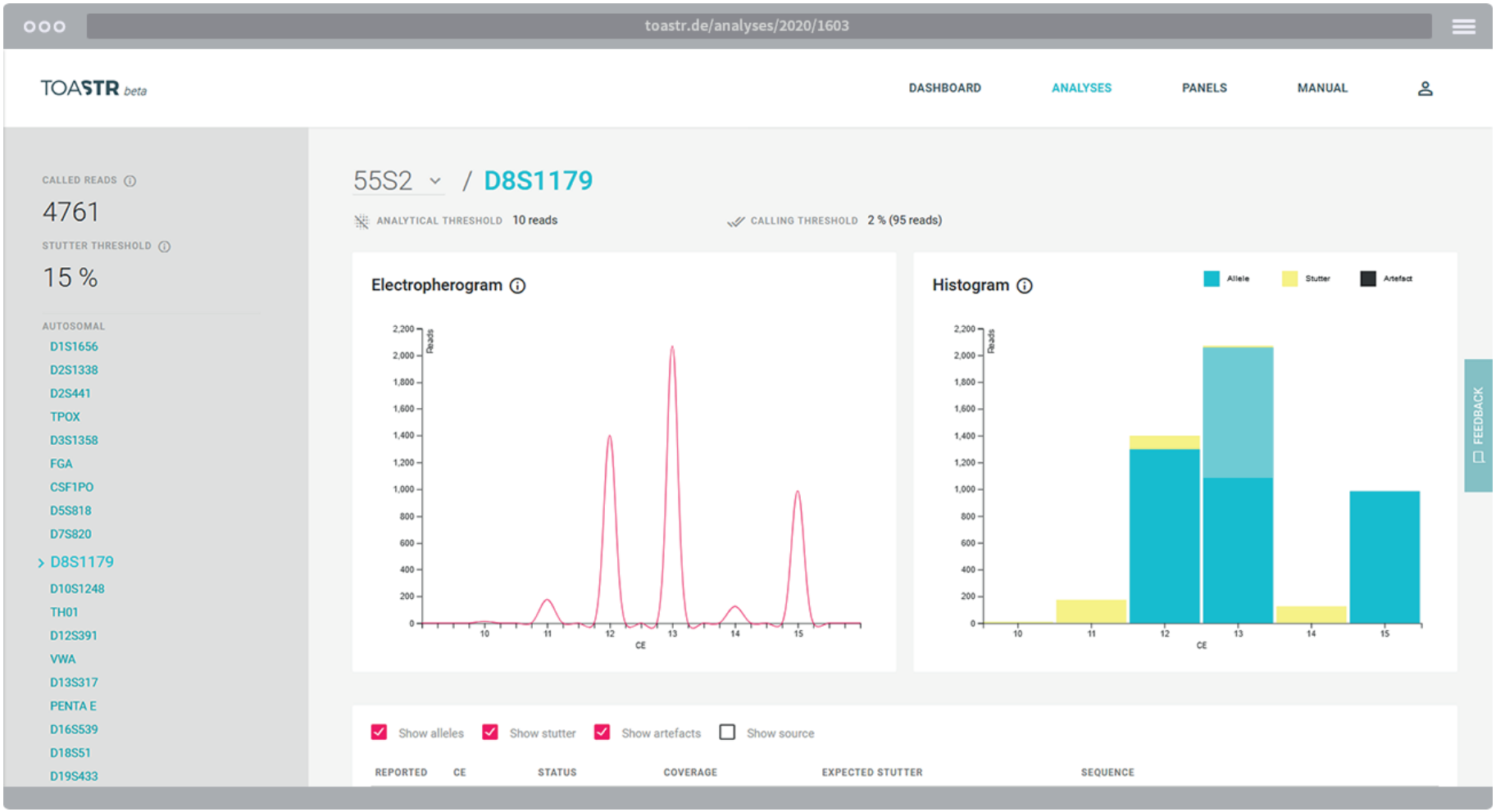


Figure 17. Exemplary results page (top). Sequence observations are visualized as an electropherogram simulation and an interactive histogram plot. The histogram color-codes alleles, stutter, and artefacts. Alleles sharing the same length but differ in sequence can be visually separated, as can be seen for the two isoalleles of length 13.

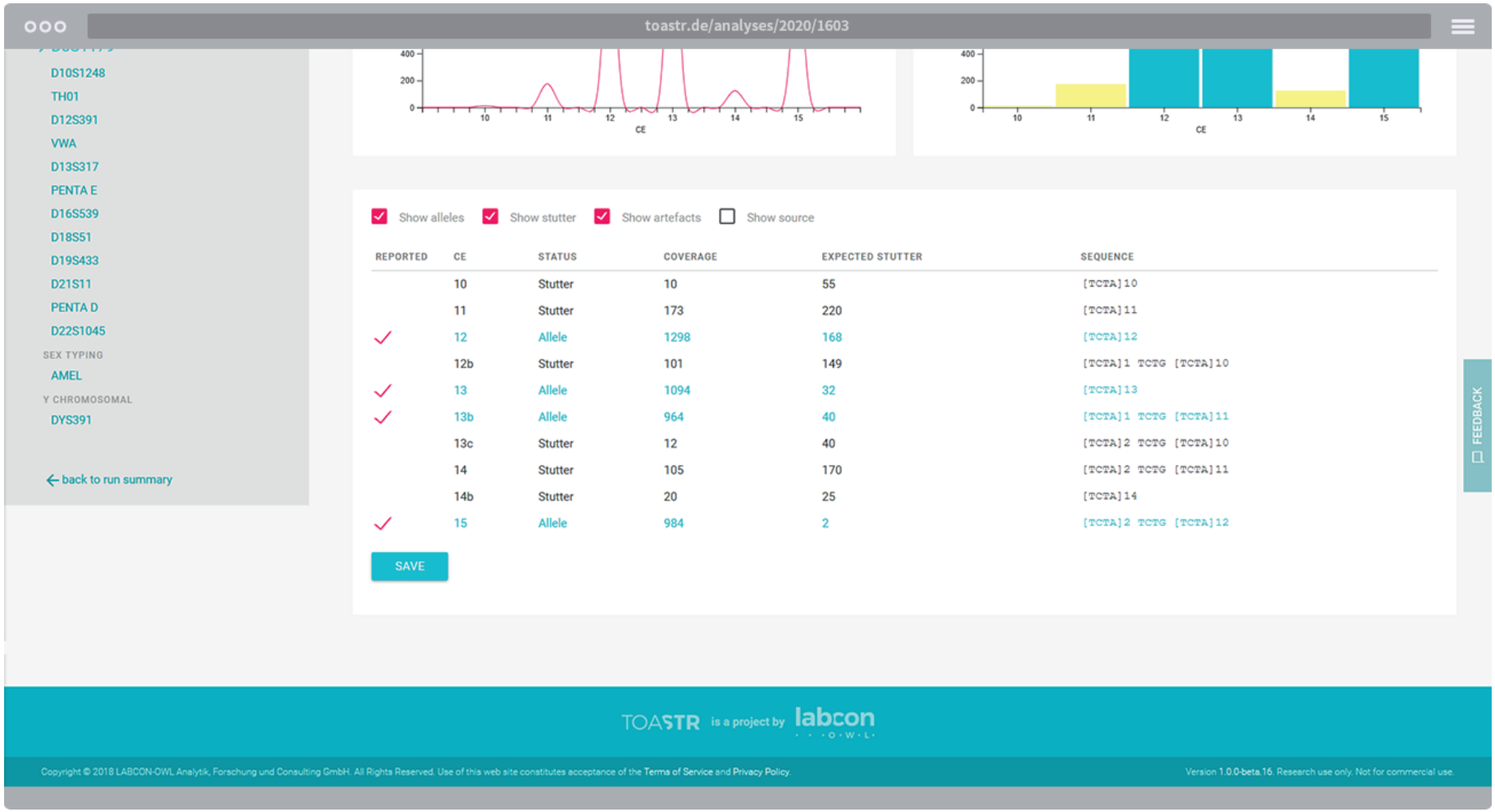


Figure 18. Exemplary results page (bottom). Tabular overview of automatically classified sequence observation for this locus.

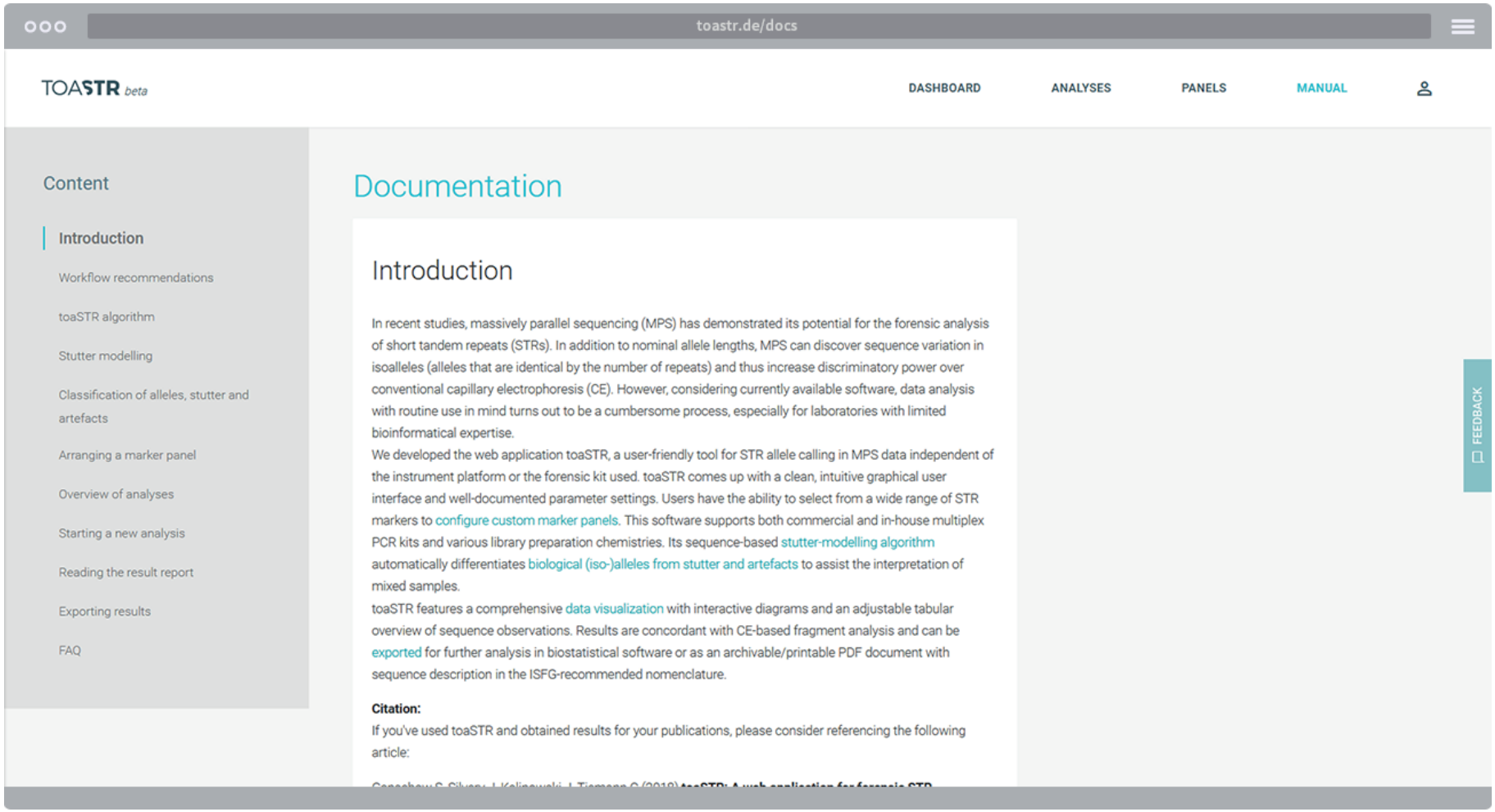


Figure 19. Comprehensive user manual.

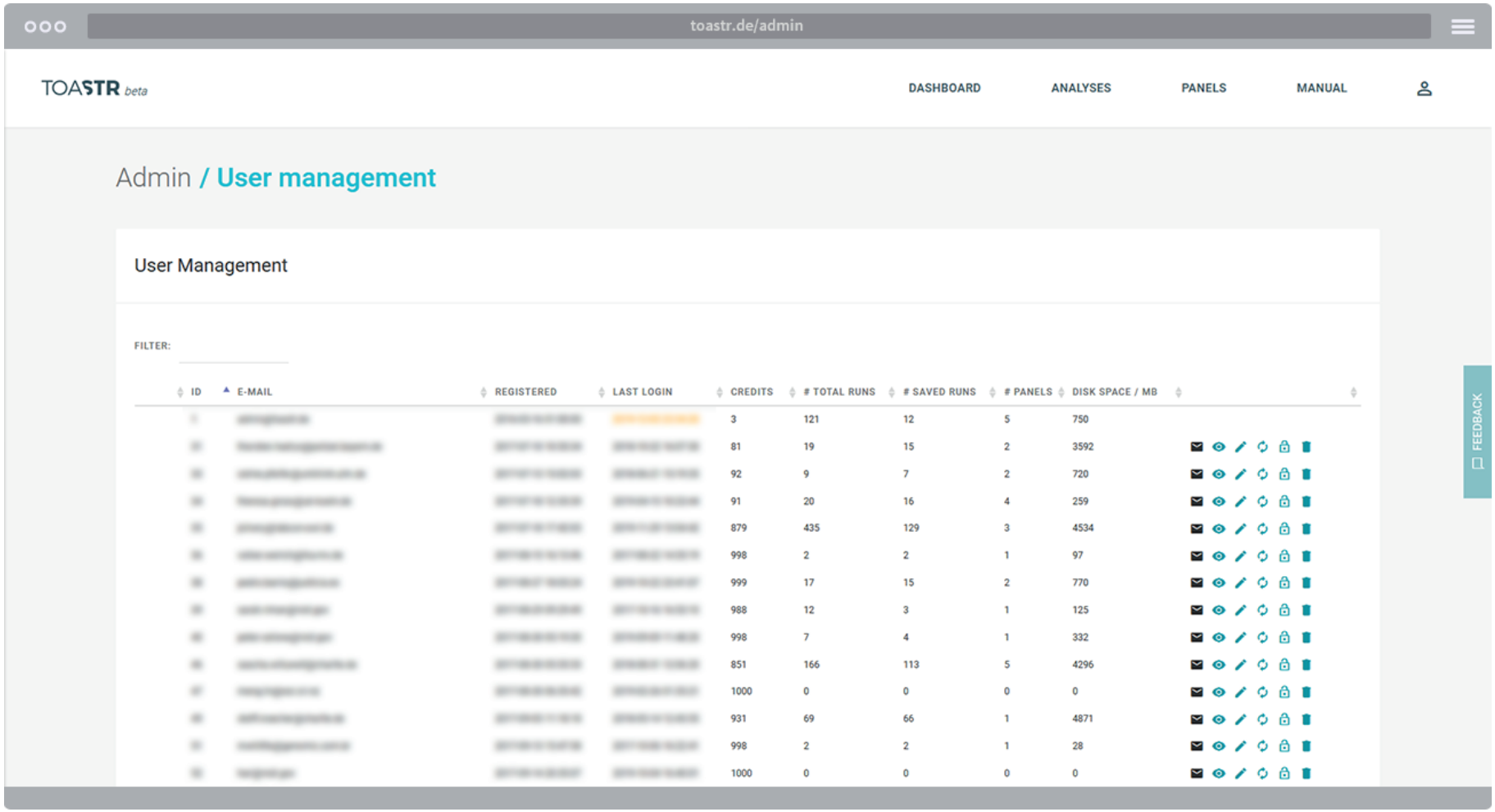


Figure 20. Administration interface for user management.

A2 Entity-relationship diagram of the toaSTR database

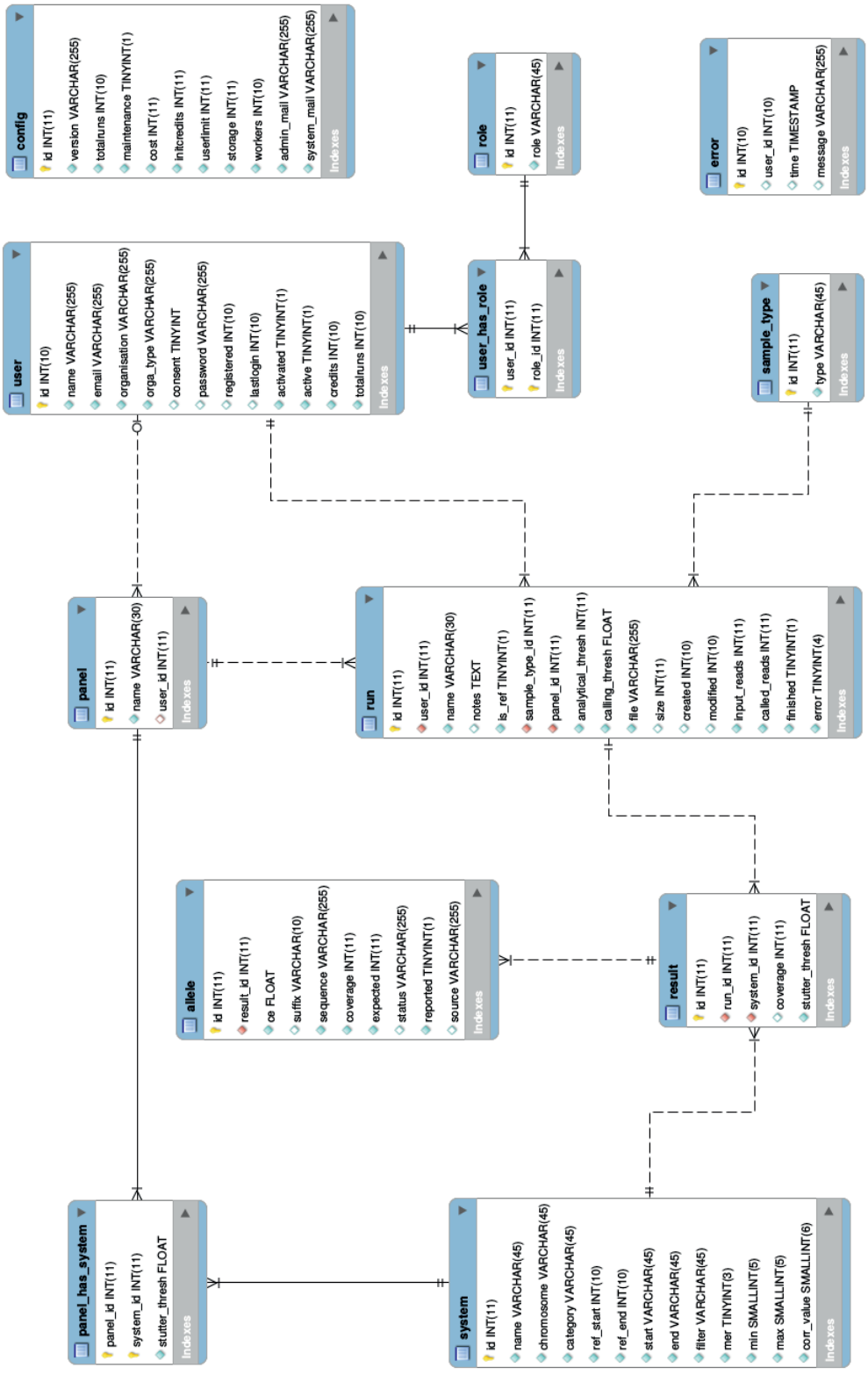


Figure 21. Entity-relationship diagram of the toaSTR database. Data is structured into 12 database tables reflecting the data entities (i.e., users, panels). Connecting lines indicate relationships between entities.

Danksagung

Diese Doktorarbeit war eine erkenntnisreiche wissenschaftliche Reise, auf der ich einen Eindruck vom unfassbaren Schatz menschlichen Wissens erlangen durfte und gleichzeitig eine Ahnung davon erhielt, wie viel mehr es noch zu entdecken gilt. Solch eine wunderbare Erfahrung wäre ohne zahlreiche Wegbegleiter und Förderer nicht denkbar gewesen. So möchte ich an dieser Stelle die Bemerkung anbringen, dass die Verdienste meiner Arbeit, wenn sie deren hat, zu einem bedeutenden Teil meinen Lehrmeistern und ihrer wohlwollenden Unterstützung zuzusprechen sind, indes für die Ungenauigkeiten und Mängel ich allein verantwortlich bin.

Mein herzlichster Dank gebührt meinem Gutachter und Betreuer bei LABCON-OWL, Herrn **Prof. Dr. Carsten Tiemann**. Gern erinnere ich mich an unsere zahlreichen wissenschaftlichen Diskussionen und Ihre wertvollen Ratschläge, durch die Sie maßgeblich zur inhaltlichen Ausrichtung und dem innovativen Charakter dieser Arbeit beigetragen haben. Vielen Dank für Ihr entgegengebrachtes Vertrauen und die Zurverfügungstellung eines inspirierenden und hochmodernen Arbeitsumfeldes, in dem sich meine Kreativität frei entfalten durfte. Dank Ihrer großzügigen Unterstützung ermöglichten Sie es mir, weltweit an Kongressen teilzunehmen und den Wert wissenschaftlicher Gemeinschaft zu erfahren. Für Ihr besonderes Interesse an meinem beruflichen und persönlichen Weiterkommen bin ich Ihnen zu großem Dank verpflichtet.

Meinem Doktorvater und Gutachter, Herrn **apl. Prof. Dr. Jörn Kalinowski**, danke ich herzlich für die Übernahme der universitären Betreuung. Vielen Dank für Deine Offenheit für das Thema meiner Arbeit, anregende fachliche Gespräche und Dein Engagement für unkomplizierte Rahmenbedingungen. Du hast mich sehr freundlich in Deine CeBiTec-Arbeitsgruppe eingeladen und großzügig an Eurer Expertise im Bereich der DNA-Sequenzierung teilhaben lassen.

Ein weiteres Dankeschön gilt Herrn **Prof. Dr. Alexander Goesmann** für die bereitwillige Übernahme des Zweitgutachtens. Ich danke Dir und Deinem Team für die konstruktiven, inspirierenden Diskurse rund um die Bioinformatik, durch die Ihr zum Gelingen dieser Arbeit beigetragen habt.

Auf zahlreichen Kongressen und Tagungen der forensischen Gemeinschaft haben sich neue Bekanntschaften entwickelt und Möglichkeiten zum regen wissenschaftlichen Austausch ergeben, für die ich sehr dankbar bin. Besonders erwähnen möchte ich an dieser Stelle Herrn **Prof. Dr. Peter Wiegand** und Herrn **ao. Univ.-Prof. Dr. Walther Parson**. Lieber Peter, Deine Offenheit und Erfahrung waren unentbehrliche Voraussetzungen für unser erfolgreiches Kooperationsprojekt und somit auch für den Rahmen meiner wissenschaftlichen Arbeit. Die bereichernden und motivierenden Begegnungen mit Dir werden mir immer in guter Erinnerung bleiben. Lieber Walther, ich bewundere Dein enormes Engagement für unsere Disziplin und danke Dir, dass Du jederzeit für den wissenschaftlichen Nachwuchs ansprechbar bist und mir wertvolle Gelegenheiten eröffnet hast, um mein akademisches Profil zu schärfen.

Meine **Schwesterinnen und Schwestern** bei LABCON-OWL haben mich seinerzeit als neuen Mitarbeiter und „Hahn im Korb" sehr wohlwollend aufgenommen. Ein spezieller Dank gilt Frau Janine Silvery und Frau Helena Siemens für die kollegiale Zusammenarbeit und unsere ehrgeizig vorangetriebenen gemeinsamen Forschungsprojekte.

Frau **Dr. Marta Diepenbroek** und Herrn **Nathaniel J. Faber** möchte ich für die hilfreichen Anmerkungen zum Manuskript danken.

Den Leitern, Mitarbeitern und Doktoranden aller Arbeitsgruppen am **CeBiTec** bin ich sehr verbunden für die angenehme Atmosphäre während unserer gemeinsamen Retreats und für den freigiebigen Erfahrungsaustausch unter Mitstreitern.

Liebe **Familie**, liebe **Jana** – ich danke Euch von Herzen für Eure bedingungslose Liebe und emotionale Unterstützung. Ihr seid unschätzbar wertvolle Ratgeber für mich und habt mir bei den kleinen und großen Entscheidungen im Leben stets Mut gemacht, meine Ziele zu verfolgen. Ein besonderer Dank gebührt meinen Eltern: Ihr habt mir unzählige, großartige Möglichkeiten gegeben mich weiterzuentwickeln und mit Beständigkeit und Geduld meine Begeisterung am Lernen gefördert.

Danke!

Erklärung

Hiermit erkläre ich, dass ich die hier vorliegende Dissertation eigenständig gemäß der Promotionsordnung der Fakultät für Biologie der Universität Bielefeld vom 1. Juni 2012 angefertigt habe.

Ich versichere, dass ich alle von mir benutzten Hilfsmittel und Quellen angegeben und keine Textabschnitte von Dritten oder eigenen Prüfungsarbeiten ohne Kennzeichnung übernommen habe.

Es wurden weder unmittelbar noch mittelbar geldwerte Leistungen an Dritte für Vermittlungstätigkeiten oder Arbeiten erbracht, die im Zusammenhang mit dem Inhalt der vorliegenden Dissertation stehen.

Weiterhin erkläre ich, dass die vorliegende Dissertation weder vollständig noch in Auszügen einer anderen Fakultät mit dem Ziel vorgelegt worden ist, einen akademischen Titel zu erwerben.

Ich bewerbe mich hiermit erstmalig um den Doktorgrad der Naturwissenschaften.

Berlin, den 15. Januar 2020

Sebastian Ganschow

Curriculum vitae

Sebastian Ganschow

Date and place of birth 03/16/1986, Salzkotten, Germany

Education

2015 – 2020 Ph.D., Faculty of Biology
Bielefeld University / LABCON-OWL GmbH, Bad Salzuflen
Concentrations: Forensic genetics, Assay development, Massively parallel sequencing, Bioinformatics

2012 – 2015 M.Sc. Biosystems Technology / Bioinformatics
Wildau Technical University of Applied Sciences (final grade: 1.4)
Concentrations: Biosensors, Massively parallel sequencing, Bioinformatics
Thesis: *Optimierung der Herstellung einer Sequenzierbank für die Ion-Torrent-Sequenzierung*

2008 – 2012 B.Sc. Biotechnology and Instrumentation Engineering
Bielefeld University of Applied Sciences (final grade: 1.3)
Concentrations: Cell culture, Bioreactors, Engineering
Thesis: *Hochzelldichte-Kultivierung und Charakterisierung einer CHO-Zelllinie im 2 Liter Perfusions-Bioreaktor mit Plattensedimenter*

Professional Experience

2015 – present Research associate
LABCON-OWL GmbH, Bad Salzuflen

Teaching Experience

2015 – 2016 Supervisor of the practical course “Molecular Diagnostics”
Bielefeld University of Applied Sciences

Service to the profession

Peer-review for the journal *Forensic Science International: Genetics*

Award

2017 Young Scientist Travel Award, ISFG Congress, Seoul

Publications

Conference proceedings

2019

toaSTR – a web app for STR sequencing
STRAND (STR Align, Name, Define) Working Group Meeting, London

Design of Experiments: Wie aus monSTR ein High-Performance Identity Panel wurde
39. Spurenworkshop: German DNA Profiling (GEDNAP), Jena

2018

toaSTR: a user-friendly tool for STR allele calling in MPS data
European Network of Forensic Science Institutes (ENFSI) DNA Working Group Meeting, Rome

toaSTR: die Web-App für eine einfache und effiziente STR-Genotypisierung mit MPS
38. Spurenworkshop: German DNA Profiling (GEDNAP), Basel

2017

toastr.de – the Swiss knife for next-generation STR genotyping
18th European Forensic DNA Working Group Meeting, Innsbruck

toaSTR: a web-based forensic tool for the analysis of short tandem repeats in massively parallel sequencing data
27th Congress of the International Society for Forensic Genetics (ISFG), Seoul

Next-Generation Sequencing für die forensische Genetik: Entwicklung und Evaluation eines 17-plex PCR-Kits
37. Spurenworkshop: German DNA Profiling (GEDNAP), Gießen

2016

Next Generation Sequencing für die forensische Genetik: Entwicklung und Evaluation eines High-Fidelity 10-plex STR-Assays
12. Jahrestagung der Deutschen Gesellschaft für Abstammungsbegutachtung (DGAB), Bielefeld

Original papers (specimen copies)

toaSTR: A web application for forensic STR genotyping by massively parallel sequencing

Title	toaSTR: A web application for forensic STR genotyping by massively parallel sequencing
Authors	Sebastian Ganschow [a], Janine Silvery [a], Jörn Kalinowski [b], Carsten Tiemann [a] [a] LABCON-OWL Analytik, Forschung und Consulting GmbH, Siemensstr. 40, 32105 Bad Salzuflen, Germany [b] Center for Biotechnology (CeBiTec), Bielefeld University, Sequenz 1, 33615 Bielefeld, Germany
Journal	Forensic Science International: Genetics
Journal ranking	4.884 (2018 Impact Factor)
Publication type	Research article
Publication year	2018
Publication status	Published journal article
DOI	https://doi.org/10.1016/j.fsigen.2018.07.006
Author contributions (CRediT taxonomy)	**Sebastian Ganschow:** Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Janine Silvery:** Formal Analysis, Investigation, Writing – review & editing. **Jörn Kalinowski:** Project administration, Supervision, Writing – review & editing. **Carsten Tiemann:** Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Forensic Science International: Genetics 37 (2018) 21–28

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

toaSTR: A web application for forensic STR genotyping by massively parallel sequencing

Sebastian Ganschow[a,*], Janine Silvery[a], Jörn Kalinowski[b], Carsten Tiemann[a]

[a] LABCON-OWL Analytik, Forschung und Consulting GmbH, Siemensstr. 40, 32105 Bad Salzuflen, Germany
[b] Center for Biotechnology (CeBiTec), Bielefeld University, Sequenz 1, 33615 Bielefeld, Germany

ARTICLE INFO

Keywords:
DNA typing
Short tandem repeat (STR)
Massively parallel sequencing (MPS)
Next-generation sequencing (NGS)
Web application

ABSTRACT

Massively parallel sequencing (MPS) is emerging within the forensic community as a promising technique for high-resolution short tandem repeat (STR) genotyping, discovering both length and sequence polymorphisms. Conversely, the application of MPS to routine casework poses new challenges to the DNA analyst in view of the complex sequence data that is generated with this technology.

We developed the web application *toaSTR* to help forensic experts work with MPS data simply and efficiently. An intuitive graphical user interface guides through the STR genotyping workflow. This versatile software handles data from various popular MPS platforms and supports both commercial and in-house multiplex PCR kits. Users can define locus-specific stutter thresholds and create custom sets of STR markers to be analyzed. toaSTR's innovative sequence-based stutter model predicts and identifies common stutter variants. The algorithm automatically differentiates biological (iso-)alleles from stutter and other artefacts to assist the interpretation of mixed samples. toaSTR features a comprehensive data visualization with interactive diagrams and a dynamic tabular overview of sequence observations. The software provides an interface to biostatistics tools and enables PDF result export in compliance with the sequence notation recommended by the International Society for Forensic Genetics (ISFG).

An initial compatibility and concordance study confirmed the software's independent functionality and precise allele calling with data of different MPS platforms, STR amplification kits, and library preparation chemistries. Discussion of genotyping results for single source and mixed samples demonstrates toaSTR's advantages and includes suggestions for future MPS software development.

The beta version of toaSTR is freely accessible at www.toastr.online.

1. Introduction

For the past decades, exploiting length-based variation in short tandem repeats (STRs) through capillary electrophoresis (CE) has been the 'gold standard' relied upon by forensic DNA laboratories. Massively parallel sequencing (MPS) has demonstrated its potential for high-resolution genotyping by providing benefits over CE analysis in terms of discovering sequence diversity of repeat motifs [1,2], increased multiplex capacity [3], and improved profiles from degraded samples by using short amplicons [4–6].

The additionally gained sequence information may be used to distinguish true (i.e., biological) alleles from non-allelic observations which result from the analytical scheme and do not originate from the sample [7]. The latter include stutter products and other artefacts that are generated during PCR-based library preparation and sequencing reactions. Stutter products arise during PCR amplification because of polymerase strand slippage. These amplicons are one or more repeat units longer or shorter in size than the true allele. Other artefacts may result from platform-specific sequencing errors like substitutions or InDels [8,9].

Furthermore, the identification of intra-allelic sequence variants may ease DNA mixture deconvolution in cases where contributors share isoalleles (alleles that are identical by the number of repeats but differ in sequence) or if the minor contributor's allele, hidden in a stutter position, has a different sequence than the stutter product of the major contributor. Using bioinformatics solutions, observations may be classified based on characteristic sequence structures [10], which likewise guards against over-interpretation of results in view of the high sensitivity of MPS.

Considering the large amount of data generated in a sequencing run,

* Corresponding author.
E-mail addresses: info@labcon-owl.de, sganschow@labcon-owl.de (S. Ganschow).

https://doi.org/10.1016/j.fsigen.2018.07.006
Received 31 January 2018; Received in revised form 4 July 2018; Accepted 5 July 2018
Available online 06 July 2018

analysis and reporting of forensic MPS data requires a dedicated software for reliable STR genotyping. As described by Børsting and Morling [11], the software should support user-definable options to alter locus-specific settings and adapt to in-house validation studies and different standards of laboratory accreditation. Further requirements comprise the ability to detect intra-repeat variations, compliance with nomenclature recommendations, and support for the interpretation of mixtures. A recent survey by Alonso et al. [12] on the current state of MPS in forensics indicated the need for an overarching software that is independent of the MPS platform. The authors of this survey also emphasized the importance of an interchangeable result format that provides compatibility between platforms and with established DNA databases.

The manufacturers of the popular benchtop sequencers MiSeq (Illumina, San Diego, CA) and Ion PGM/S5 (Thermo Fisher Scientific, San Francisco, CA) each provide their own genotyping software [13,14]. Usage of those tools is restricted to the associated commercial multiplex kits with defined STR marker selections and is limited to the respective sequencing device (technical support email communication). Thus, they do not meet the requirement of an independent solution.

Alternative programs, such as FDSTools [10], lobSTR [15], or STRait Razor [16], to name just a few, have also been proposed. A recent review article [17] covers freely available bioinformatic tools for the analyses of forensic STR markers. While being successfully applied in a research context, drawbacks of existing software become clear from a routine perspective. The usability of reviewed tools is limited by one or more of the following aspects: (1) a UNIX operating system is required, (2) the tool has to be handled via the command line interface, (3) results necessitate manual reworking (e.g., copy-pasting primary data output into a spreadsheet program), and (4) lack of assistance in data interpretation (no differentiation of biological alleles from non-allelic products).

Therefore, we developed a new web application to simplify the access to a user-friendly forensics tool for the integrated analysis and reporting of STR profiles with MPS. Our goal is to provide the forensic community with an overarching software with routine use in mind that works independently of the MPS platform and with both commercial and in-house multiplex PCR kits. Our solution offers an intuitive graphical user-interface, sample management, and comprehensive data visualization and takes full advantage of the high resolution enabled by MPS.

2. Materials and methods

toaSTR is a browser application that allows for cross-platform usage with all common operating systems and web browsers. The software is written in Perl and is based on the Catalyst MVC web application framework (catalystframework.org). A job queue system in the backend enables non-blocking parallel processing of analyses for a short time-to-result while delivering a rich user experience. A relational database system manages efficient storage and retrieval of results. The application employs current web technologies HTML5, CSS3, and JavaScript. Diagrams for data visualization are generated dynamically with the D3.js JavaScript library (d3js.org). toaSTR's responsive layout fits to desktops, tablets, and smartphone screens.

2.1. Security concept

The goal of information security is to maintain the confidentiality, integrity, and availability of data. We employed an industry standard security concept for toaSTR which involves different layers of security. The application is hosted on a dedicated server located in a DIN ISO/IEC 27001 certified data center in Germany. Server hardware is designed with multiple redundancies for high availability and reliability of this service. All application components (storage, database, and computing units) are located within a firewall-protected environment to control network traffic to and from toaSTR. The client-server communication is encrypted via Secure Socket Layer (SSL), thus preventing data in transit from being compromised or accessed by a third party. Usage of this application is only possible for registered users. After free registration at www.toastr.online, users get access to their password-protected personal account which allows managing own analyses, panels, and settings.

2.2. Data processing

Data upload is handled via a convenient web form that also enables to enter sample meta-data (e.g., sample name, notes, etc.) and to adjust analysis parameters. Accepted input for toaSTR is a raw or compressed (gzipped) FASTQ or FASTA file containing single-end reads or merged paired-end reads. Reads must span the complete repeat region plus a minimum of 30 nucleotides upstream and downstream the repeat region to be analyzed with toaSTR. Merging is only necessary after paired-end sequencing. It refers to the process of assembling a read pair by forming a consensus sequence of the overlapping 3′ ends and has been described elsewhere [18,19].

toaSTR employs the BBTools suite (jgi.doe.gov/data-and-tools/bbtools) for data pre-processing and calling of observations. Briefly, the algorithm validates sample input using the *testformat* program and performs a fast *k*-mer-based fuzzy search on all input reads in sense and antisense orientation with *bbduk*. A read is called for an STR system if it approximately matches a pair of internally defined recognition elements (REs) upstream and downstream of the repeat region in a fixed order. REs are oligonucleotide sequences with a length of 14–25 nt, being likewise identifiers for an STR locus and boundaries for the repeat region to be reported. Hence, their sequence and positioning is of considerable importance for a correct sequence-based and length-based genotype in accordance with forensic conventions. The valuable preliminary work of Gettings et al. [20] enabled an efficient RE design based on GRCh38 human reference genome sequences with annotated repeat regions and flanking region variation. Per default two mismatches (i.e., a Hamming distance of 2) are allowed in each RE to account for SNPs (single nucleotide polymorphisms). A read is rejected if it lacks the core repeat motif of the respective STR system (e.g., AGAT). If necessary, reads are reverse-complemented to meet sequence orientation conventions as described in Gettings et al. [20].

Reads with an identical repeat region are clustered to *observations*, counting their respective coverage (i.e., the cluster size). Low-abundant observations with a coverage below the user-defined *analytical threshold* (AT, default 10 reads) are considered background noise arising from stochastic sequencing errors. These observations are discarded to prevent over-interpretation of data. The AT can be set in a range from 0 to 100 reads.

Subsequent steps involve alignment-free analysis of the repeat structure and determination of the observation's CE-concordant length. toaSTR assigns a short name (i.e., the CE length) and a suffix if it finds more than one observation with the same length (e.g., 8b). The observation's sequence is formatted according to the common repeat notation (i.e., core repeats in brackets with a number indicating the repeat count). Formatting does not require error-prone manual interaction or a database of reference allele definitions. Thus, toaSTR is able to identify both known and previously unknown sequence variants.

Finally, each observation is classified into one of the three categories *allele*, *stutter*, or *artefact*, respectively. The classification bases on the output of toaSTR's sequence-based stutter prediction model, which will be described in the following sections.

2.3. Stutter modelling

In a bioinformatics context, the term *stutter modelling* refers to the approximate computational representation of a sequence-based stutter generation mechanism. This concept aims at the prediction of stutter

variants to identify non-allelic observations and distinguish them from biological alleles.

It has previously been shown that the stutter ratio correlates with the length of long uninterrupted repeat stretches within an STR region [21–24]. Given this relation, we postulate that the sequence pattern of the most frequently observed stutter variants derives from the longest uninterrupted stretch (LUS) and the second-longest uninterrupted stretch of repeats (SLUS, only in compound or complex STRs) of the parent observation. This assumption adapts on the concept of a stutter prediction algorithm described by Hoogenboom et al. [10].

toaSTR's stutter model considers the nine most frequent stutter products (Fig. 1). For every observation with a nominal length N, the algorithm generates *virtual stutter* sequences in position N − 2, N − 1 (backward), N ± 0 (isometric), and N + 1 (forward), as these are commonly described stutter positions in MPS-STR analysis [7,25]. Sequences of virtual stutter variants are built by contraction or elongation of the LUS and/or SLUS by one or more repeat units, respectively. Variants are abbreviated with an index that denotes the nominal position (−2, −1, 0, or +1) and an "A" for stutter in the LUS and "B" for stutter in the SLUS, respectively (e.g., +1B, −2AB).[1]

Virtual stutter coverage calculation includes the parent observation's coverage and the user-defined *stutter threshold* (ST) for the respective STR system. The ST stands for the anticipated N−1 stutter ratio. It should be determined empirically under defined conditions with replicates of well-characterized single-person reference samples. A value of 20% would be a conservative starting point. The expected stutter value (described in the next section) may be used to calibrate the ST. The ST should be as low as possible to sharply distinguish stutter from a possible allele with an identical sequence. Conversely, the ST has to be high enough so that stutter is not erroneously called as a true allele.

The commonly observed likelihood for stutter generation in different positions is reflected by a linear, quadratic, or cubic ST term: Generation of forward stutter (N + 1) generally has a lower likelihood than backward stutter (N − 1). Hence, the forward stutter calculation includes a quadratic stutter threshold term (ST^2) compared to a linear ST term in the backward stutter calculation. Isometric stutter (0A, 0B) may be interpreted as a series of +1 and −1 stutter, and therefore includes a cubic term (ST^3) in the virtual stutter calculation. Similarly, the quadratic ST is applied for N − 2 stutter, which can be considered as a series of −1 and −1 stutter.

2.4. Classification of observations as alleles, stutter, and artefacts

If a virtual stutter sequence matches an existing observation in the sample, the respective virtual stutter coverage is saved for that observation. The observation gets a source label (i.e., the stutter variant index) to trace back the stutter generation path. Virtual stutter products from multiple sources may be attributed to a single observation. The sum of all virtual stutter coverages for an observation is called *expected stutter* value. toaSTR does not filter reads marked as stutter nor manipulates input read counts; In fact, the observation's *status* solely depends on the relation between its coverage, the expected stutter value, and the user-defined calling threshold (CT), as described in Table 1. The CT (default 2%) is a relative value. It refers to the system coverage (i.e., the number of reads that were allocated to an STR system) and defines the minimum coverage needed for an observation to be qualified as a true allele. In other words, the CT differentiates true alleles from artefacts. Fig. 2 illustrates an exemplary classification process.

[1] Since the discrimination between allele, stutter, and artefact is not known at this point of analysis, also observations that are retrospectively identified as stutter can generate virtual stutter themselves. However, this assumption is supported by the fact that in a real-world PCR a stutter product may become a PCR template again, which is consequently subjected to polymerase slippage (thus leads to stutter).

2.5. Marker panels

In toaSTR, *panels* are customizable collections of STR markers. A panel defines which markers are to be analyzed and stores a user-defined stutter threshold (ST) for each marker. A marker is defined by preconfigured upstream and downstream recognition elements (REs) flanking the repeat region. They were designed to support both commercial and in-house developed STR PCR multiplex kits with minimum requirements on amplicon design. The RE's length, a maximum Hamming distance of 2, and the fixed matching order facilitates high stringency read calling while allowing a certain tolerance for possible SNPs in the recognition sites.

At present, users can select from 24 autosomal and one Y-chromosomal STR markers as well as amelogenin, including all loci of the expanded European Standard Set (ESS) and the expanded CODIS core set.

2.6. Compatibility and concordance study

An initial study was designed to test toaSTR's compatibility with sequence data from twelve reference and ring trial DNA samples. Data was generated with three different sequencing platforms and various combinations of forensic multiplex kits and library preparation chemistries (Table 2). Inter-assay and absolute concordance testing was performed by comparing toaSTR analysis results among each other or with CE fragment size and Sanger sequence data, respectively. Single source genomic DNA from Standard Reference Materials SRM 2391c (National Institute of Standards and Technology, Gaithersburg, MD), 2800M Control DNA (Promega, Madison, WI), and 9947A Control DNA (Applied Biosystems, Foster City, CA) were used as absolute references with known allele lengths and allele sequences, where appropriate. See Supplementary Table S1 for literature references.

2.7. Sample preparation

DNA from GEDNAP (German DNA Profiling) ring trial samples 52S2 (two person mixed sample) and 53PA (single person sample) was extracted using the High Pure PCR Template Preparation Kit (Roche, Switzerland). Amplification, library construction and sequencing were conducted as described in Table 2 according to the manufacturer's recommendations (manuscript for the in-house multiplex kit in preparation). Sequence data for samples 9–12 were obtained from Faith and Scheible [26]. Paired-end reads of samples 3–8 were merged with *bbmerge* [19] and parameters: qtrim = r trimq = 20 mismatches = 0 margin = 0 entropy = f.

Prior to data analysis, panels were created in toaSTR to select the markers to be analyzed and to define STs. FASTQ files were uploaded and annotated using toaSTR's upload form. Default values were used for AT (10 reads) and CT (2%), except for sample 1 (CT = 5%).

3. Results

3.1. Reporting of results

toaSTR's efficient algorithm takes less than one minute on average to process an input file with 100,000 reads. The analysis result summary displays user-provided sample information and parameter settings as well as the number of input reads and called reads. The genotype table gives the STR markers as selected in the panel with their respective coverages and the CE-like lengths of called alleles. Selecting a marker in the genotype table opens the detailed results view (Fig. 3). It shows all observations called for this system in a dynamic tabular and graphical view, respectively. The sequences table displays each observation's CE-like short description, the status (allele, stutter, or artefact), the detected coverage, the expected stutter value calculated by the stutter model, and the formatted STR sequence. Observations can be filtered based on their status. The optional 'source' column describes

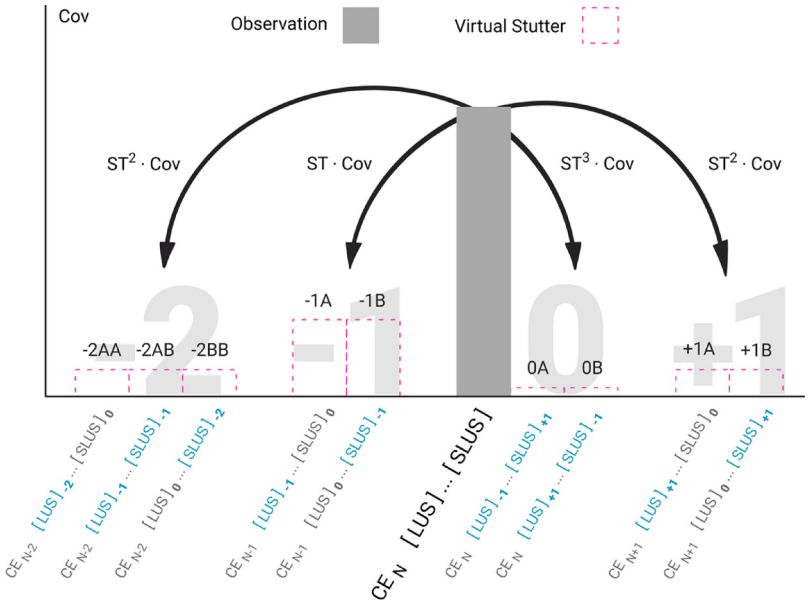


Fig. 1. The stutter modelling algorithm considers nine virtual stutter variants in positions −2, −1, 0, and +1 relative to an observation's length N. Subscript numbers in the variant schemes (blue color) indicate the gain or loss of repeat elements in the longest uninterrupted stretch (LUS) or second-longest uninterrupted stretch (SLUS, only in compound or complex STRs). Complex STRs may contain interspersed shorter repeat stretches and non-repetitive sequences (indicated by "…"). Calculation of the virtual stutter coverage includes the stutter threshold (ST) for the respective STR system and the parent observation's coverage (Cov), as indicated at the arrows.

Table 1
Classification of alleles, stutter, and artefacts depending on the relation between the observation's coverage (Cov), the expected stutter (ES) value and the user-defined calling threshold (CT).

Relation	Status	Description
Cov > ES, Cov > CT	Allele	Alleles are considered correctly amplified and sequenced DNA template of biological origin.
Cov > ES, Cov ≤ CT	Artefact	Artefacts mostly represent sequencing errors or low-level somatic mutations.
Cov ≤ ES	Stutter	Stutter is a product of PCR polymerase slippage in repetitive regions of biological alleles.

potential stutter generation paths identified by the stutter model. Unticking the 'reported' checkbox for an allele excludes it from the biostatistics export and parenthesizes it in the summary and the PDF report.

Data for each STR system is visualized in two diagrams (Fig. 3). The CE simulation resembles a capillary electropherogram and shows how marker peaks would appear in familiar fragment analysis. In this view, peak height represents the length-dependent coverage without any discrimination of (iso-)alleles, stutter, or artefacts. The MPS histogram groups observations by CE length in a stacked bar chart and colors them according to their status. Hovering over a bar displays the sequence and status in a tooltip.

toaSTR enables result export as an archivable and printable PDF file as well as comma-separated values (CSV format). The PDF report summarizes the analysis and contains detailed results for each marker. Sequence annotation herein follows the ISFG recommendations on a comprehensive nomenclature system [27], including the locus name, CE-like allele name, chromosome, human reference genome assembly version, repeat region coordinates of the reference allele, and the STR sequence in compressed notation. CSV export is an experimental interface to support downstream analysis with biostatistics software. The file contains observations that were classified as allele and selected to be reported. Since to date biostatistics software has not been adapted to sequence-based data, exported results are reduced to CE-like allele names and their corresponding allele heights (i.e., allele coverages).

A comprehensive user manual and a FAQ section are available on the toaSTR website.

3.2. Concordance testing

Robustness of toaSTR was assessed with a dataset of twelve reference and GEDNAP ring trial DNA samples. Four marker panels representing the PCR kits were created within the software and applied to the respective samples. Supplementary Table S1 consists of a set of Microsoft Excel spreadsheets (one for each reference DNA) which compare length and sequence alleles reported by toaSTR. Another spreadsheet contains panel configurations with stutter thresholds for the selected loci. Additionally, Supplementary Material S2 includes comprehensive PDF reports generated via the software's data export function.

Full STR profiles for up to 26 loci, depending on the amplification kit, were produced for all samples evaluated. A total of 2,010 observations were called and automatically classified by toaSTR, including 529 alleles, of which 512 alleles (96.8%) were immediately concordant to references. In total, 17 borderline allele calls, which marginally exceeded the expected stutter threshold or calling threshold, were manually excluded in the software during expert review. After review, length and sequence concordance reached 100%.

SRM 2391c Component A–C were prepared using the Promega PowerSeq Auto/Mito/Y primer set (not commercially available) and the Kapa Biosystems KAPA Hyper Prep library preparation kit. Unidirectional sequencing was performed on an Illumina MiSeq. Alleles were fully concordant to lengths and sequences as reported in the SRM Certificate of Analysis. Observation 17.1 of D1S1656 in sample 9 was manually classified as artefact (deletion of 'TG') since its coverage only marginally exceeded the calling threshold of 2%. Three other observations of D2S441 and D22S1045 in samples 10 and 11 were manually identified as forward stutter lightly exceeding the expected stutter value.

Male control DNA 2800M was paired-end sequenced on the MiSeq and MiSeq FGx, respectively. Multiplex PCR was performed with the Illumina ForenSeq, the Promega PowerSeq prototype, and our in-house 21-plex kit, respectively. Libraries were prepared as indicated in Table 2 using Illumina TruSeq LT, Nextera XT, or KAPA Hyper Prep. Fragment sizes reported by toaSTR were concordant to the 2800M product literature. Allele sequences were fully consistent across all 2800M samples. Two observations of D22S1045 in samples 5 and 12

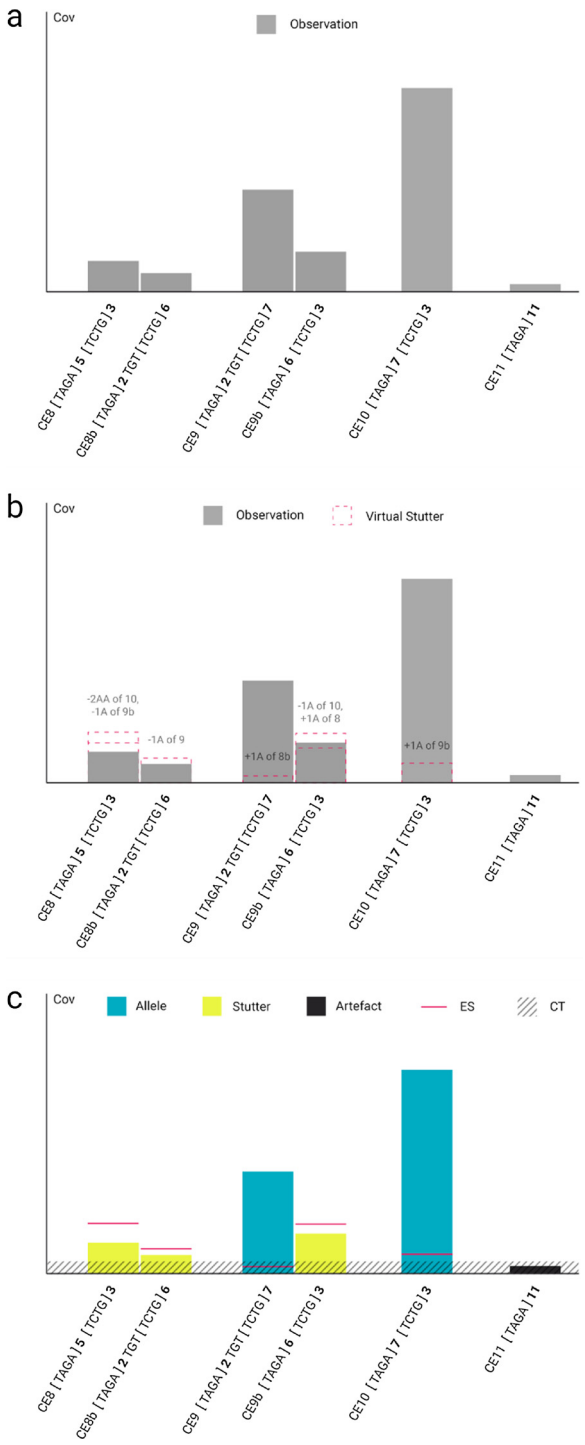


Fig. 2. Exemplary observation classification for one STR system. a) Observations called for this STR system. b) The stutter model identifies potential stutter products and assigns virtual stutter coverages (dashed boxes). Source labels (e.g., −2AA of 10) help to trace back the stutter generation path. c) Observation classification: CE8, CE8b, CE9b have a coverage below the expected stutter (ES) value and thus are classified as stutter. CE9, CE10 have a coverage above the expected stutter value and above the calling threshold (CT) and thus are classified as allele. CE11 has a coverage above the expected stutter value and below the calling threshold and thus is classified as artefact.

were manually classified as raised forward stutter. Although not officially part of the ForenSeq kit, alleles could also be called for the SE33 locus in sample 2.

STR markers of female DNA 9947A were amplified with the Thermo Fisher Early Access STR Kit v1. Ion AmpliSeq libraries were sequenced on the Ion PGM. The corresponding sample 1 showed a fully concordant profile. The D8S1179 marker was observed to be homozygous by fragment size (allele 13) but heterozygous by sequence with respective alleles $[TCTA]_1$ TCTG $[TCTA]_{11}$ and $[TCTA]_{13}$. Allele 19 for locus D1S1656 was manually excluded because both its low coverage and its sequence indicated an insertion artefact slightly above the calling threshold. Sequencing data from the Ion PGM (sample 1) showed the highest frequency of artefacts (Table 3). Loci D2S1338 and TPOX contained a large amount of observations arising from transitions (T ↔ C or A ↔ G). InDels were seen, for example, in artefact 12.2 of D10S1248, presumably coming from a 'GA' deletion of allele 13, or artefact 17.1 of VWA which can be explained by a 'C' insertion of allele 17 (cf. Supplement S1 and S2).

DNA of GEDNAP 52S2 (samples 3 and 6) and 53PA (samples 4 and 7) was prepared either with Promega PowerSeq/TruSeq LT or the in-house 21plex/Nextera XT. toaSTR performed equally well on both datasets and reported concordant allele sequences and lengths. Results for 52S2 indicated a two-person constellation with up to four alleles per locus. Marker D21S11 contained two isoalleles of length 30. Also in mixed samples, automatic differentiation of alleles against stutter and artefacts was accurate. Sporadic deviations resulting from unexpectedly high stutter or a raised artefact coverage were manually corrected.

4. Discussion

This software is a novel approach to simplify the access to MPS-STR genotyping for investigators who do not want to delve deeper into bioinformatics. Our developmental goals were to deliver a user-oriented, versatile tool that works independent of how data was generated in the wet-lab. Genotyping results are prepared for optimal human readability and fast interpretation without neglecting the high information density and sensitivity of MPS data.

Within the examined cohort, up to 94% of input reads were uniquely called for an STR locus (Table 3). Calling ratio for some samples was considerably lower, depending on whether the multiplex kit contains other markers (STRs, SNPs, mtDNA) that are currently not analyzed by toaSTR. Simple, compound, and complex repeat structures were automatically detected and formatted in the common bracketed notation. toaSTR's sequence compression works de novo, hence dispenses with the need for a database of known alleles. Reported sequences complied with forensic conventions as described by Gettings et al. [20]. One exception is the InDel-STR-locus Penta D, which shows a 13 bp insertion/deletion upstream of the repeat region. It required shifting the recognition element and including the InDel in the reported genotype in order to avoid a bioinformatic dropout in case of a deletion. Generally, the genotyping algorithm employed by toaSTR can be easily adapted to evolving forensic conventions by changing REs. Considerations of the DNA commission of the ISFG [27] include a consistent forward strand notation of STR sequences relative to the GRCh38 reference genome assembly, which would affect twenty-three STR loci historically aligned to the reverse strand. Since a change to the forward strand may result in a potential shift of the reading frame, reported repeat regions have to be strictly defined.

The stutter model appears to be a veritable approach to predict stutter. Instead of applying a fixed interpretation threshold, predicting stutter and respective coverages is carried out individually on a sequence basis. Based on the assumptions underlying the stutter generation model, results indicated that observations arising from stutter were reliably identified as such. Focusing on the LUS and SLUS and the nominal positions N − 2 to N + 1, respectively, appeared to be a valid approximation of sequence-based stutter generation; stutter in other repeat stretches or positions was negligible in this dataset. However, validation of the stutter model should be performed on a wider data basis. The goodness of fit of the stutter model could be measured by analyzing the deviation of expected stutter value and actual coverage of

Table 2
Dataset for compatibility and concordance testing. 2800M, 9947A: standard human genomic DNA; SRM2391c A/B/C: NIST Standard Reference Material 2391c Component A/B/C; 52S2, 53PA: GEDNAP 52/53 ring trial samples. PE: paired end, SE: single-end. R1: read 1, R2: read 2.

ID	Sample	Amplification kit	Library preparation kit	MPS platform	Sequencing reaction	toaSTR input
1	9947A	Thermo Fisher Early Access STR Kit v1	Ion AmpliSeq Library Kit 2.0	Ion PGM	Ion PGM Hi-Q Sequencing Kit SE 400 bases	R1
2	2800M	Illumina ForenSeq	DNA Signature Prep Kit	Illumina MiSeq FGx	MiSeq Reagent Kit v2 PE 351 × 31 bases	R1
3	52S2	Promega PowerSeq Auto/Mito	Illumina TruSeq LT	Illumina MiSeq	MiSeq Reagent Kit v2 PE 251 × 251 bases	Merged R1/R2
4	53PA					
5	2800M					
6	52S2	In-house 21-plex	Illumina Nextera XT	Illumina MiSeq	MiSeq Reagent Kit v2 PE 251 × 251 bases	Merged R1/R2
7	53PA					
8	2800M					
9	SRM 2391c A	Promega PowerSeq Auto/Mito/Y	Kapa Biosystems KAPA Hyper Prep Kit	Illumina MiSeq	MiSeq Reagent Kit v2 SE 300 bases	R1
10	SRM 2391c B					
11	SRM 2391c C					
12	2800M					

Fig. 3. Screenshot of the result page for marker D1S1656 displaying an electropherogram simulation and an interactive histogram for data visualization. The lower section gives automatically classified sequence observations and the option to override toaSTR's suggestions. Expected stutter values and potential sources of stutter enable traceability of theoretical stutter generation.

a stutter observation. Furthermore, the calculation of expected stutter could be refined by adapting the stutter threshold to the length of the repeat stretch. It was previously described by van der Gaag et al. [25] that the stutter ratio positively correlates with the number of repeat units and increases for alleles with longer uninterrupted stretches.

The subsequent automatic classification of alleles, stutter, and artefacts was largely accurate and required only minor correction by the analyst. Hence, this feature may be a helpful support for the interpretation of single-source and multiple-source samples. For example, at first sight, the two-person sample 3 (GEDNAP 52S2) showed two main 'peaks' at locus D21S11. However, making use of the sequence information gained by MPS, the software readily identified two isoalleles at position 30 and another allele at position 31. Additionally, stutter products and artefacts having the same nominal lengths as alleles were correctly distinguished based on sequence variation. The classification process may be further improved by flanking region

Table 3
toaSTR data analysis for twelve reference and ring trial samples. Please refer to Table 2 for information on sample preparation.

ID	Sample	Input reads	Called reads	Analyzed loci	Number of observations		
					Alleles (Concordant)	Stutter	Artefacts
1	9947A	122,634	40,241 (33%)	18	30/30	35	119
2	2800M	325,250	64,742 (20%)	26	47/47	72	32
3	52S2	222,787	70,343 (32%)	24	55/55	66	47
4	53PA	198,079	83,559 (42%)	24	41/41	71	50
5	2800M	191,378	53,826 (28%)	24	43/43	72	20
6	52S2	131,711	122,355 (93%)	21	52/52	89	43
7	53PA	152,266	142,623 (94%)	21	36/36	67	53
8	2800M	113,794	107,429 (94%)	21	41/41	73	21
9	SRM 2391c-A	250,000	134,821 (54%)	24	39/39	89	83
10	SRM 2391c-B	250,000	80,434 (32%)	24	44/44	85	42
11	SRM 2391c-C	250,000	84,399 (34%)	24	41/41	88	43
12	2800M	250,000	71,763 (29%)	24	43/43	74	47

analysis, which is currently not supported by toaSTR. Flanking region SNPs may be used to differentiate observations that even share an identical repeat region [28].

Default parameters for the analytical threshold and calling threshold proved to be suitable for Illumina data. In consistence with our settings, Young et al. [7] considered an AT of 10 reads generally sufficient for STR systems with coverages below 5000 reads. The default CT was set aspiringly low to 2% since this threshold also defines the lowest detectable minor contributor proportion. Ion PGM data (sample 1) contained a remarkable amount of InDel and substitution errors, which were visible as high frequency artefacts with a significant coverage. This circumstance required a CT setting of 5% for sample 1 in order to classify these observations correctly as artefacts. It was shown elsewhere that InDel errors accounted for the majority of errors introduced by the PGM [8]. Results have to be interpreted with care, however, as only a single sample from this platform was analyzed in the present study.

Like with other sequence-search approach tools, calling of observations is sensitive to the recognition element (RE) sequences and the mismatch tolerance. Reads have to span at least the complete repeat region plus both REs to be analyzed with toaSTR. There is a potential risk from two sides: Weak RE matching conditions may lead to bioinformatic drop-in, i.e. a false positive calling of reads. Bioinformatic drop-out and imbalance, i.e. missing an observation, may occur in case of an increased amplification error or sequencing error rate, incomplete reads, or multiple SNPs in the RE site. However, our study indicated a robust genotyping with toaSTR considering the high calling rates of 93%–94% with the 21-plex data, in which the toaSTR panel covers all markers contained in the kit. Surveillance mechanisms could be implemented to warn users if there is an indication of bioinformatic drop-out. Assessing sequence quality scores could add another automated filtering step, but would restrict input to FASTQ files. Subsequent multicentre evaluations are currently in the planning stage to extend testing on a wider data basis and define optimal bioinformatics settings for reliable genotyping.

When performing paired-end sequencing, toaSTR requires a manual pre-processing of sequence data. Read merging (i.e., generating an assembly of overlapping paired reads) can be done with cross-platform Java tools like BBMerge [19] or FLASH [18]. We do not recommend using the *StitchReads* parameter in the MiSeq Reporter software, since reads that cannot be merged result in two single reads. This may artificially increase the number of called reads and skew results. Own experiments suggested that, even with proper tools, merging repetitive sequences is prone to error (data not shown). Consequently, analysts may consider unidirectional (single-end) sequencing to avoid the need for merging, as demonstrated with samples 9–12. Results of sample 2 indicated that read 1 of ForenSeq data is sufficient for toaSTR input, as the quasi-unidirectional R1 already contains all necessary sequence information.

The concept of this software includes seamless integration into typical forensic LIMS (laboratory information management systems). Therefore, the software provides a data export interface. We suggested a clearly laid out PDF export format that can be used to download and save reported genotypes locally as a basis for STR profiling. Since biostatistics tools are already available, we decided not to implement probabilistic calculations in toaSTR. In future, the export function should faciliate data exchange with biostatistics tools. Currently, this is an experimental feature until the exchange format is clear and sequence allele frequency databases are available. Furthermore, discussions with biostatistics software developers have to clarify if filtering of observations should be applied before data export, how MPS information could be reduced for CE-oriented software that cannot handle sequence data, and how amplicon length information is correctly incorporated to support continuous statistical models.

5. Conclusion

toaSTR is a proposal how to reconcile high-resolution STR genotyping and user-friendliness. This software was designed for forensic investigators and does not require dedicated computing infrastructure nor bioinformatics expertise. It meets the demand for a platform-independent solution, complies with nomenclature recommendations and assists in the interpretation of results. The web app concept makes this software easily available while addressing data privacy and security concerns. toaSTR's intuitive access via the web browser may simplify handling of MPS in general.

Discussions have touched on possible future directions including analysis of flanking region variation, refining the stutter model, and improving input/output interfaces. This software is under continuous development. We encourage forensic laboratories to test toaSTR and greatly appreciate suggestions, bug reports and feature requests.

Conflict of interest

CT, JS, and SG are current employees of LABCON-OWL GmbH. The authors declare no other competing interests.

Role of funding

This work was supported by a grant from the ZIM funding program (Zentrales Innovationsprogramm Mittelstand) of the German Federal Ministry for Economic Affairs and Energy.

Acknowledgements

The authors would like to thank Theresa Groß (Institute of Legal Medicine, University Hospital of Cologne), Céline Pfeifer (Institute of Legal Medicine, University Hospital of Ulm), and Thorsten Hadrys

(Bavarian State Office of Criminal Investigation) for providing MPS data and giving their valuable feedback during the development of toaSTR.

Appendix A. Supplementary data

S1. Set of Microsoft Excel spreadsheets including description of sample preparation, toaSTR panel configurations, and reported alleles and sequences for each sample in comparison with references.

S2. toaSTR PDF export files for 12 samples.

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fsigen.2018.07.006.

References

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

[2] N.M.M. Novroski, J.L. King, J.D. Churchill, L.H. Seah, B. Budowle, Characterization of genetic sequence variation of 58 STR loci in four major population groups, Forensic Sci. Int. Genet. 25 (2016) 214–226, https://doi.org/10.1016/j.fsigen.2016.09.007.

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

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

[5] F. Guo, Y. Zhou, F. Liu, J. Yu, H. Song, H. Shen, B. Zhao, F. Jia, G. Hou, X. Jiang, Evaluation of the Early Access STR Kit v1 on the Ion Torrent PGM™ platform, Forensic Sci. Int. Genet. 23 (2016) 111–120, https://doi.org/10.1016/j.fsigen.2016.04.004.

[6] Z. Wang, D. Zhou, H. Wang, Z. Jia, J. Liu, X. Qian, C. Li, Y. Hou, Massively parallel sequencing of 32 forensic markers using the Precision ID GlobalFiler™ NGS STR Panel and the Ion PGM™ System, Forensic Sci. Int. Genet. 31 (2017) 126–134, https://doi.org/10.1016/j.fsigen.2017.09.004.

[7] B. Young, J.L. King, B. Budowle, L. Armogida, A technique for setting analytical thresholds in massively parallel sequencing-based forensic DNA analysis, PLoS One 12 (2017), https://doi.org/10.1371/journal.pone.0178005.

[8] L.M. Bragg, G. Stone, M.K. Butler, P. Hugenholtz, G.W. Tyson, Shining a light on dark sequencing: characterising errors in ion torrent PGM data, PLoS Comput. Biol. 9 (2013) e1003031, , https://doi.org/10.1371/journal.pcbi.1003031.

[9] M. Schirmer, U.Z. Ijaz, R. D'Amore, N. Hall, W.T. Sloan, C. Quince, Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform, Nucleic Acids Res. 43 (2015), https://doi.org/10.1093/nar/gku1341 e37–e37.

[10] J. Hoogenboom, K.J. van der Gaag, R.H. de Leeuw, T. Sijen, P. de Knijff, J.F.J. Laros, 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 Sci. Int. Genet. 27 (2017) 27–40, https://doi.org/10.1016/j.fsigen.2016.11.007.

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

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

[13] Illumina, ForenSeq™ Universal Analysis Software Guide, (2016) (Accessed 27 November 2017), https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/forenseq-universal-analysis-software/forenseq-universal-analysis-software-guide-15053876-01.pdf.

[14] Applied Biosystems, HID STR Genotyper Plugin User Guide, (2017) (Accessed 27 November 2017), https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0015879_HID_STR_Genotyper_Plugin_UG.pdf.

[15] M. Gymrek, D. Golan, S. Rosset, Y. Erlich, lobSTR: a short tandem repeat profiler for personal genomes, Genome Res. (2012), https://doi.org/10.1101/gr.135780.111.

[16] A.E. Woerner, J.L. King, B. Budowle, Fast STR allele identification with STRait Razor 3.0, Forensic Sci. Int. Genet. 30 (2017) 18–23, https://doi.org/10.1016/j.fsigen.2017.05.008.

[17] Y.-Y. Liu, S. Harbison, A review of bioinformatic methods for forensic DNA analyses, Forensic Sci. Int. Genet. 33 (2018) 117–128, https://doi.org/10.1016/j.fsigen.2017.12.005.

[18] T. Magoč, S.L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies, Bioinformatics 27 (2011) 2957–2963, https://doi.org/10.1093/bioinformatics/btr507.

[19] B. Bushnell, J. Rood, E. Singer, BBMerge – accurate paired shotgun read merging via overlap, PLoS One 12 (2017) e0185056, , https://doi.org/10.1371/journal.pone.0185056.

[20] K.B. Gettings, R.A. Aponte, P.M. Vallone, J.M. Butler, STR allele sequence variation: current knowledge and future issues, Forensic Sci. Int. Genet. 18 (2015) 118–130, https://doi.org/10.1016/j.fsigen.2015.06.005.

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

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

[23] C. Brookes, J.-A. Bright, S. Harbison, J. Buckleton, Characterising stutter in forensic STR multiplexes, Forensic Sci. Int. Genet. 6 (2012) 58–63, https://doi.org/10.1016/j.fsigen.2011.02.001.

[24] J.-A. Bright, K.E. Stevenson, M.D. Coble, C.R. Hill, J.M. Curran, J.S. Buckleton, Characterising the STR locus D6S1043 and examination of its effect on stutter rates, Forensic Sci. Int. Genet. 8 (2014) 20–23, https://doi.org/10.1016/j.fsigen.2013.06.012.

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

[26] S.A. Faith, M. Scheible, Analyzing Data from Next Generation Sequencers Using the PowerSeq® Auto/Mito/Y System, (2016) (Accessed 7 August 2017), https://www.promega.de/resources/profiles-in-dna/2016/analyzing-data-from-next-generation-sequencers-using-the-powerseq-automitoy-system/.

[27] W. Parson, D. Ballard, B. Budowle, J.M. Butler, K.B. Gettings, P. Gill, L. Gusmão, D.R. Hares, J.A. Irwin, J.L. King, P. de Knijff, N. Morling, M. Prinz, P.M. Schneider, C.V. Neste, S. Willuweit, C. Phillips, Massively parallel sequencing of Forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements, Forensic Sci. Int. Genet. 22 (2016) 54–63, https://doi.org/10.1016/j.fsigen.2016.01.009.

[28] X. Zeng, J.L. King, B. Budowle, Investigation of the STR loci noise distributions of PowerSeq™ Auto System, Croat. Med. J. 58 (2017) 214–221, https://doi.org/10.3325/cmj.2017.58.214.

Development of a multiplex forensic identity panel for massively parallel sequencing and its systematic optimization using design of experiments

Title	Development of a multiplex forensic identity panel for massively parallel sequencing and its systematic optimization using design of experiments
Authors	Sebastian Ganschow, Janine Silvery, Carsten Tiemann LABCON-OWL Analytik, Forschung und Consulting GmbH, Siemensstr. 40, 32105 Bad Salzuflen, Germany
Journal	Forensic Science International: Genetics
Journal ranking	4.884 (2018 Impact Factor)
Publication type	Research article
Publication year	2019
Publication status	Published journal article
DOI	https://doi.org/10.1016/j.fsigen.2018.11.023
Author contributions (CRediT taxonomy)	**Sebastian Ganschow:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing - original draft, Writing - review & editing. **Janine Silvery:** Data curation, Investigation, Validation, Writing - review & editing. **Carsten Tiemann:** Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing.

Forensic Science International: Genetics 39 (2019) 32–43

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Development of a multiplex forensic identity panel for massively parallel sequencing and its systematic optimization using design of experiments

Sebastian Ganschow*, Janine Silvery, Carsten Tiemann

LABCON-OWL Analytik, Forschung und Consulting GmbH, Siemensstr. 40, 32105 Bad Salzuflen, Germany

ARTICLE INFO

Keywords:
Massively parallel sequencing (MPS)
Short tandem repeat (STR)
Forensic genetics
Design of Experiments (DOE)
monSTR multiplex PCR assay
toaSTR software

ABSTRACT

The application of massively parallel sequencing (MPS) in forensic sciences enables high-resolution short tandem repeat (STR) genotyping for the characterization of biological evidence. While MPS supports multiplexing of a large number of forensic markers, the performance of an MPS-STR panel depends on good primer design and optimal PCR conditions. However, conventional strategies for multifactorial assay optimization are labor-intensive and do not necessarily allow the experimenter to identify optimum factor settings.

Here we describe our new multiplex PCR assay, *monSTR*, which supports the simultaneous amplification of 21 forensic markers followed by targeted sequencing on the Illumina MiSeq. The selection of STR markers adapts on the expanded European Standard Set (ESS), including the highly polymorphic locus SE33, for compatibility with existing forensic DNA databases. Primer engineering involved bioinformatics tools to create a multiplex-compatible primer set. Primer quality was evaluated in silico and in vitro. We demonstrate the systematic optimization of multiplex PCR thermocycling conditions using Design of Experiments (DOE) methodology. The objective was to yield a specific, balanced, low-noise amplification of forensic targets. A central composite face design of experiments enabled an efficient simultaneous investigation of multiple critical process parameters and their interactions. Optimal multiplex PCR conditions were predicted using software-aided modelling based on DOE data. Verification experiments suggested a balanced, reproducible amplification of all markers with reduced formation of artefacts. Fully concordant STR profiles were obtained for the investigated reference samples even with challenging input DNA concentrations. We found that application of DOE principles enabled an experimentally practical and economically justifiable assay development and optimization, even beyond the field of forensic genetics.

1. Introduction

Massively parallel sequencing (MPS) facilitates high-resolution forensic identity typing by discovering both length and sequence polymorphisms at short tandem repeat (STR) loci. Advantages of this technology over conventional capillary electrophoresis (CE) methods have been highlighted in recent studies: Exploiting intra-allelic sequence variation [1] provides additional information in complex forensic cases and mixed samples; Restrictions regarding the spatial and spectral amplicon layout are eliminated, allowing for an increased number of loci that can be analyzed simultaneously [2]; Genetic information from highly degraded DNA samples can be recovered using small amplicons (miniSTRs) [3]; High sensitivity is relevant especially for challenging casework samples with low quantity of biological evidence [4].

Various commercial [5,6], prototype [4,7], and custom [3,8,9] STR multiplex PCR kits for different MPS platforms have been developed and evaluated. Kits are characterized by the number and choice of loci and the amplicon design may support the analysis of SNPs and InDels in the flanking regions. Flanking variation may provide additional discriminatory features to facilitate mixture deconvolution [10]. For maximum compatibility with existing DNA databases, the multiplex composition should be largely consistent with national and international standard sets, respectively, such as the expanded European Standard Set (ESS) and CODIS core locus set [11]. Since MPS provides greater flexibility in terms of multiplex design, kits may be complemented with interesting forensic markers like SE33, which is one of the most informative tetranucleotide loci studied to date. SE33 is a core locus for the German national DNA database DAD and is also adopted by other laboratories in Europe [12]. Its highly polymorphic nature exhibits complex length and sequence polymorphisms [13,14]. Variation in the flanking regions add to the extensive diversity [15].

* Corresponding author.
E-mail addresses: info@labcon-owl.de, sganschow@labcon-owl.de (S. Ganschow).

https://doi.org/10.1016/j.fsigen.2018.11.023
Received 13 September 2018; Received in revised form 28 November 2018; Accepted 29 November 2018
Available online 30 November 2018
1872-4973/

Although this marker is included in various CE-based kits (GlobalFiler, NGM Select, ESSPlex, PowerPlex ESI/ESX 17), SE33 is not (officially) available in published MPS-STR panels at the time of writing.

1.1. Strategies for multiplex PCR development

A multitude of parameters can affect the performance of a PCR assay, including the design of primers, the composition of reagents, and the thermocycling protocol. Careful primer engineering is essential for successful amplification of targets. This is particularly true for multiplex assays, which require not only a high target specificity, but also a primer set with similar properties including melting temperature, GC content, and amplicon size. Furthermore, each additional primer increases the potential of undesired primer-template and primer-primer interactions in highly multiplexed reactions. Upfront bioinformatics can assist in the correct selection and validation of primers prior to expensive wet lab experiments. Automated primer designing tools have been developed which support the specific needs of multiplex primer design, such as handling multiple templates simultaneously. These programs include FastPCR [16], MPprimer [17], and others. MPprimer uses the popular Primer3 program [18] as primer design engine, which allows for a great flexibility to control sequence and primer parameters. To avoid primer dimerization, MPprimer examines the stability of complementarity of two primers based on thermodynamics.

When a set of candidate primers was designed, it is appropriate to perform a computer simulation of PCR. For a given set of primers and a DNA template, in silico PCR, also termed virtual PCR or ePCR, predicts which PCR products are synthesized [19]. A number of tools are available, including Primer-BLAST [20], UCSC In-Silico PCR [21], and MFEprimer [22]. MFEprimer is suitable for multiplex PCR analysis, accepting a large number of primers to check against a whole genome sequence. The program determines which (cross-primer) amplicons are likely to be amplified and estimates the formation of primer dimers and hairpin structures. However, in silico PCR does not replace in vitro testing of primers to ensure an efficient and specific amplification of targets.

The complex interactions among the components of a multiplex PCR assay make it unlikely that a standard set of reaction conditions would be optimal for all situations [23]. Thus, every newly developed PCR application requires optimization of critical process variables to achieve good analytical performance. Requirements for a forensic identity panel would include, among others: specific amplification of targeted loci, while allowing a certain tolerance for genotypic variation like mutations in primer binding sites; balanced amplification, hence equal depth of coverage across all loci (inter-locus balance) and locus alleles (heterozygous balance); low tendency to form stutter products and other technical artefacts (herein referred to as noise) due to polymerase slippage and polymerase error, which can be a problem impacting interpretation. Surprisingly, except for the adjustment of primer concentrations, optimization of global assay conditions was not discussed in recent articles introducing in-house constructed MPS-STR marker panels [3,9].

However, several reports on general multiplex PCR development and optimization have been published. Important aspects of those strategies are briefly summarized below. We can ignore issues regarding the spectral and spatial distribution of amplicons, since MPS eliminates these restrictions by design. Henegariu et al. [23] provided an extensive discussion of factors that can influence the amplification, including times and temperatures of the PCR protocol, concentration of reaction components, and the use of adjuvants. Multiplex PCR optimization was understood as a step-by-step protocol of practical solutions to overcome commonly encountered problems. The report describes sequential alterations of cycling conditions and reaction components aiming at a higher yield of PCR products and an increase of reaction specificity. However, we believe that such a sequential optimization approach makes it difficult to understand potential interactions between factors. Butler [24] suggested a development strategy focusing on careful primer design and empirical testing. The protocol included the evaluation of primer functionality and specificity as well as performing sensitivity and consistency studies. Optimization of thermocycling conditions was not subject of the discussion. Edwards and Gibbs [25] emphasized the importance of similar reaction kinetics of all primers in the set. It was recommended to develop PCR conditions separately for each primer pair. The authors proposed to add primer pairs sequentially to the multiplex assay and to alter the conditions as necessary. In our opinion, however, choosing an appropriate experimental design would allow optimization of PCR parameters in parallel rather than serially.

1.2. Introduction to experimental design

While every scientific experiment has a design, some designs are more thoughtful than others. With well-designed experiments, it takes less time and effort to gain knowledge about the system. When designing a series of experiments to investigate the effect of multiple input variables (referred to as *factors*) on the process performance (*response*), there are essentially three possible approaches.

Under the common, intuitive approach the investigator modifies one factor at a time (OFAT). It requires varying the first factor until an optimum is found while every other factor is held constant. Based on the optimal value of the first factor, the remaining factors are changed separately in successive experiments (Fig. 1A). This method involves only few measurements; as an example, an optimization series of three factors each at four discrete levels would include 4 + 3 + 3 = 10 experiments (replicates not included). However, this design does not thoroughly explore the space of possible solutions and will most likely lead to locally optimal target values, while the global optimum has not been reached [26]. Moreover, it neglects interactions between factors: An isolated factor may exert only little influence on the process performance, but rather the interaction of two factors may strongly affect the response.

The matrix method sequentially tests every factor against all levels of all other factors in order to include all possible combinations (Fig. 1B). Such an approach allows interactions between factors to be estimated but is inefficient because the number of experiments increases dramatically if many factors are considered. The optimization series exemplified above would require $4^3 = 64$ separate runs, excluding replicates.

The Design of Experiments (DOE) methodology suggests a more sophisticated approach for the optimization of a multifactor process, such as multiplex PCR. Recent reviews [27,28] provide an excellent introduction to this concept. Briefly, DOE is a framework that assists in planning and performing systematic trials. Statistically designed experiments enable the investigator to rapidly identify significant factors and complex interactions. DOE provides a steep learning curve in the sense that informative data can be obtained with a minimum number of assays and minimal cost [29]. Experimental design techniques have been used successfully across numerous industries and research fields. First described by Fisher [30] in the 1920's in an agricultural application, DOE has been widely used in the field of engineering [31]. Other reports include applications in biotechnology [28,32] and molecular biology, where DOE was employed for the optimization of cDNA microarrays [33], PCR [34–36], multiplex PCR [37,38], and real-time PCR [39,40]. Application has also been found in the optimization of PCR conditions for microsatellite genotyping in a medical [41] and forensic [42] context. However, DOE methods have not been described yet in the context of MPS.

Depending on the objective of the study to be performed, different types of geometric experimental designs are available [43]. A selection of classical designs is described here. Two-level full factorial designs are constructed by testing every factor at two levels. They are useful for early phase screening applications to determine the most significant

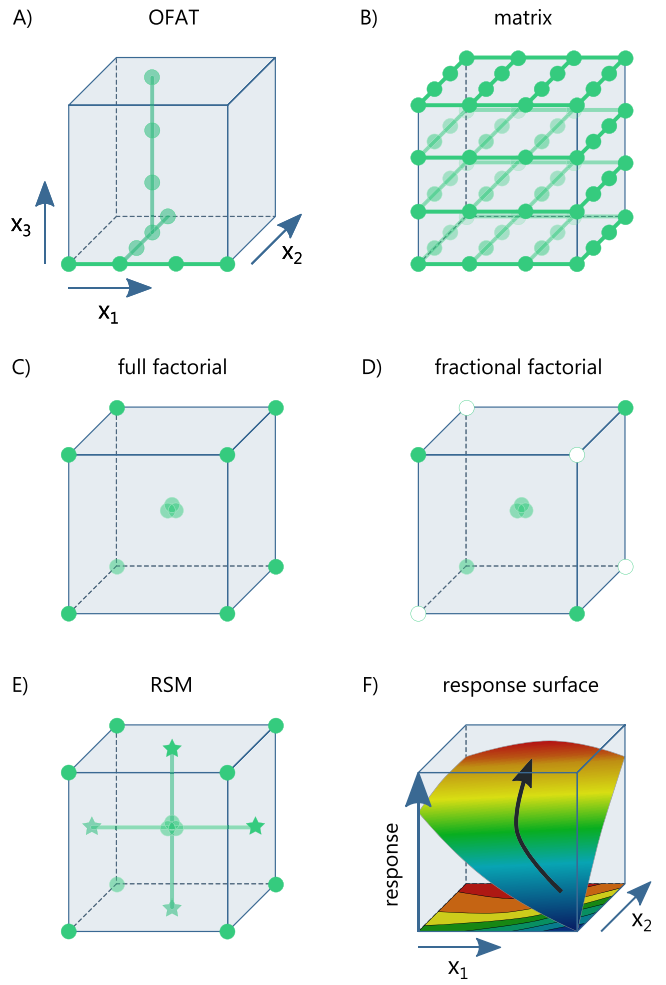


Fig. 1. A – E: graphic representations of experimental designs with three factors x_1, x_2, x_3. F: exemplary response surface plot obtained from an RSM design; The response surface represents a map of the system which allows to locate optimum conditions depending on variable settings of two factors (for visualization purposes, other factors are fixed at constant levels).

factors that influence the system. The experimental space covered by a full factorial design in three factors at two levels would be a cube with the experiments on its eight corners (Fig. 1C). It also includes a replicated center point in between the high and the low levels to assess the replicate error. The orthogonality of this design (i.e., each factor can be evaluated separately) allows for the estimation of linear effects as well as factor interactions [27]. However, such designs can lead to large numbers of experiments as the number of factors increases.

Fractional factorial designs (Fig. 1D) are balanced subsets, or fractions, of the full factorials. They are suitable for studies where four or more factors are of interest, such as in a robustness test or a screening study. Fractional factorial designs offer a reduction in measurements and still can analyze main effects and lower order interactions [28].

Once the investigator is confident that the most influential factors and their relevant ranges were identified, an optimization study may be performed to locate the optimum conditions within the experimental region. Appropriate composite designs are part of the response surface methodology (RSM) design family (Fig. 1E). They combine factorial designs with additional levels to obtain precise information about magnitude, influence, and interactions of the factors, including second-order (quadratic) effects [43]. RSM designs provide a map of the system in the form of a response surface plot (Fig. 1F), allowing to explore the behavior inside the limits of the region studied. Data from these experiments are used to build a predictive mathematical model. The model is a polynomial equation that describes a relationship between input factors and measured response values, weighted by regression coefficients. It allows for the approximation of responses based on a given combination of factor levels and thus is a useful tool to maximize the information available from limited experimental data [44].

In the present report, we propose an efficient and transparent strategy for MPS-STR multiplex assay development and optimization. Primers for our forensic identity panel *monSTR*, targeting 21 markers, were constructed and evaluated in parallel using a bioinformatics workflow. We demonstrate the use of DOE methods to identify optimum settings of critical thermocycling parameters. The aim was to maximize multiplex assay performance with respect to high amplification specificity and high interlocus balance as well as low formation of technical artefacts.

2. Materials and methods

2.1. Source of template DNA

Single-source human DNA LC3 was used for multiplex optimization experiments. For this purpose, whole blood was obtained from an anonymous donor with informed consent. DNA was extracted from 200 ml of blood using the High Pure PCR Template Preparation Kit (Roche, Switzerland) as recommended by the manufacturer. The quantity of recovered DNA was estimated using the Qubit 2.0 fluorometer with the Qubit dsDNA HS Assay Kit (Thermo Fisher, Foster City, CA). For concordance testing purposes, CE STR analysis of control DNA LC3 was accomplished using the PowerPlex 21 System (Promega, Madison, WI), PowerPlex ES SE33 Monoplex System (Promega), and VeriFiler Direct PCR Amplification Kit (Applied Biosystems, Foster City, CA) according to the manufacturers' protocols. Amplified product was separated in the 310 Genetic Analyzer (Thermo Fisher) and data was analyzed using GeneMapper ID software v3.2 (Thermo Fisher). 2800 M control DNA (Promega) and 9948 control DNA (Applied Biosystems) were used as complementary reference materials.

2.2. Selection of STR loci

The composition of monSTR was inspired by the expanded European Standard Set. This multiplex assay targets 21 forensic markers including D1S1656, D2S1338, D2S441, D3S1358, FGA, CSF1PO, SE33, D7S820, D8S1179, D10S1248, TH01, D12S391, vWA, D13S317, Penta E, D16S539, D18S51, D19S433, D21S11, Penta D, and amelogenin.

2.3. Primer engineering

We utilized a combination of bioinformatics tools to design and evaluate a primer set that meets our quality criteria for multiplex amplification of 21 targets.

First, MPprimer software v1.5 [45] was used to create an initial set of 42 candidate primers in parallel that are compatible for multiplexing. An MPprimer configuration file was set up to define primer conditions for each locus, including: 1) optimal primer size: 22 bp; 2) optimal melting temperature (Tm): 64 °C (estimated by the NEB Tm Calculator https://tmcalculator.neb.com/); 3) optimal GC content of primers: 50%; 4) predicted optimal amplicon size range: 250–350 bp, except for amelogenin: 150–200 bp. Other MPprimer parameters were left to their default values. Annotated reference sequences of 20 STR loci were obtained from Gettings et al. [15]. Sequence information of amelogenin was collected from NCBI GenBank (accession number NG_008011.1). SNP locations and allele frequencies were obtained from the 1000 Genomes Project Phase 3. SNP sites that may not overlap primers were specified in the MPprimer configuration file. Since the reference sequence allele used for primer design only represents one of the possible alleles, the desired amplicon size for each locus was adjusted according

to the average allele length reported on STRBase (https://strbase.nist.gov/).

Second, the candidate primer set was manually reviewed for an additional criterion that could not be implemented in the MPprimer process. Expected amplicons should contain as much flanking region variation as possible while taking into account the maximum amplicon size. If necessary, primers were re-constructed using the visual primer design tool Primaclade [46], while respecting the above-mentioned requirements.

Third, MFEprimer v3.0 [47] was used to predict amplicons and estimate potential non-specific (cross-) amplification between primer pairs, primer dimerization and hairpin formation. In silico PCR was performed with the revised set of 42 primers from step two. Human reference genome assembly GRCh38.p10 was selected as DNA reference. Unreasonable results were filtered out by setting the maximum amplicon size to 1000 bp and the minimum Tm to 50 °C. Other MFEprimer parameters were left to their default values.

Illumina Nextera XT overhang adapter sequences [48] were added to the 5'-ends of forward and reverse primers, respectively, to make amplicons compatible for sequencing on the Illumina platform. Primers were synthesized at TIB Molbiol (Berlin, Germany). The functionality of each primer pair was tested in singleplex PCR. Size and specificity of PCR products was examined using the Agilent Bioanalyzer 2100 with the DNA 7500 Kit (Agilent Technologies, Palo Alto, CA).

2.4. Experimental design

Multiplex PCR thermocycling conditions were to be optimized in order to yield a balanced, specific, low-noise amplification of targets. The number of potential critical process variables was narrowed down based on the experience of previous multiplex optimization studies [23–25]. We considered four quantitative input factors to have a significant effect on multiplex PCR performance: annealing temperature (ATemp), annealing time (ATime), extension time (ETime), and the number of PCR cycles (Cyc) (Table 1). The ranges in which factors were varied were established according to the general guidelines [49] for the NEB Q5 master mix, which was used for PCR setup.

We defined three statistical measures as response variables to express the optimization goals (Table 2): 1) The coefficient of variation of the coverage (CV) of all loci reflects the inter-locus amplification imbalance and should be minimal; 2) The calling ratio (CR) expresses the PCR specificity (i.e., the ratio between on-target reads and total reads) and should be close to one; 3) The average signal-to-noise ratio (SN) describes the mean coverage ratio of biological alleles and non-allelic observations (the latter including stutter and other technical artefacts originating from the analytical scheme). SN should be maximal.

The complexity of this biological process suggests interactions between factors and necessitates a quadratic mathematical regression model to obtain a reliable fit of the experimental data. Therefore, we selected a central composite face design (CCF) from the RSM design family, which enables the investigation of non-linear relationships and factor interactions while requiring minimum experimental effort. The four-factor CCF design geometrically corresponds to a regular four-dimensional hypercube. CCF is a fractional factorial design encoding three levels of each explored factor (low, medium, and high setting). It contains replicated center point experiments enabling to estimate the control over the experimental procedure, and star point experiments on the faces of the hypercube to assess second-order (quadratic) effects. The design resulted in twenty experiments with unique factor combinations plus three center point replicates, i.e. a total of 23 experiments. Located in the center of the experimental space, the center-point settings reflect standard thermocycling conditions as suggested by the manufacturer of Q5 PCR reagents [49]. To eliminate the effect of any nuisance influences, all experiments were done in a randomized order.

Table 1
Definition of variable process parameters (input factors) and investigated ranges.

Factor	Abbreviation	Range
Annealing temperature	ATemp	58–66 °C
Annealing time	ATime	20–40 s
Extension time	ETime	30–90 s
Number of cycles	Cyc	25–35

2.5. Multiplex PCR setup

Amplifications were performed as single-tube reactions in a volume of 25 µl, containing 12.5 µl Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA), 5 µl primer mix, 1 µl DNA template, and 6.5 µl water. The initial primer mix contained equimolar concentrations of each primer pair. After multiplex PCR optimization, concentrations for each primer pair were adjusted empirically.

One nanogram of LC3 template was used for all DOE experiments. For primer adjustment, serial dilutions of LC3, 2800 M, and 9948 were prepared, resulting in DNA input of 2 ng, 1 ng, and 500 pg DNA, respectively. Sensitivity study involved a dilution of 2800 M in duplicates to 1 ng, 500 pg, 250 pg, and 125 pg, respectively.

Thermal cycling was conducted in a Mastercycler pro S (Eppendorf, Germany) under the conditions described in Table 3. Settings for variable factors (ATime, ATemp ETime, Cyc) were defined by the experimental scheme (Table 4) and were subject of optimization.

2.6. Library preparation and sequencing

After multiplex PCR, amplicons were purified and left-side size-selected in a single step using Agencourt AMPure XP paramagnetic beads (Beckman Coulter, Jersey City, NJ) with a bead:sample ratio of 0.8 according to the manufacturer's protocol [50]. Indices and sequencing adapters were attached to purified amplicons in a limited cycle index PCR using the Nextera XT index kit (Illumina) under the following conditions: 98 °C for 30 s, 8 x (98 °C for 10 s, 62 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min A bead:sample ratio of 0.7 was used for the second PCR cleanup. The quality of purified libraries was evaluated by examining the size distribution and the absence of primer species on the Agilent Bioanalyzer 2100 with the DNA 7500 Kit (Agilent Technologies). Indexed DNA libraries were quantified using the NEBNext Library Quant Kit for Illumina (New England Biolabs), normalized to 4 nM, and pooled at equal volumes. After dilution to a final concentration of 10 pM, libraries were sequenced on the MiSeq platform using Reagent Kit v2 chemistry (Illumina) and performing 2 × 250 cycles of paired-end sequencing.

2.7. Data processing and analysis

Primary data analysis (base calling, quality scoring, and demultiplexing) was conducted using the MiSeq on-board software Real Time Analysis and MiSeq Reporter. Raw FASTQ files were subjected to quality filtering and merging using the bbmerge program [51] with parameters: qtrim = r trimq = 20 mismatches = 0 mininsert = 180.

STR genotyping and result interpretation were performed with *toaSTR* v1.0.0-beta14 [52] with default analysis parameters. Classification of allelic observations (signal) and non-allelic observations (noise) was done automatically based on toaSTR's stutter modelling algorithm and manually reviewed by an expert. Stutter thresholds used for the monSTR panel can be obtained from Supplementary Material S1. If more than three locus dropouts were observed in a profile, the respective experiment was excluded from analysis. Concordance of genotypes was checked against reference DNA certificates or CE results, respectively.

Creation of the CCF experimental design and statistical analysis

Table 2
Definition of responses with target optima, acceptable minimum and maximum values, and mathematical relations. SD: standard deviation, Cov: coverage, N: number of loci.

Response	Abbreviation	Minimum	Target	Maximum	Calculation
Coverage coefficient of variation	CV	–	0	0.8	$CV = \frac{SD_{Cov}}{Cov}$
Calling ratio	CR	0.8	1	–	$CR = \frac{Reads_{called}}{Reads_{input}}$
Average signal-to-noise ratio	SN	15	∞	–	$\frac{1}{N}\sum_{i=1}^{N}\frac{Cov_{Alleles,i}}{Cov_{Noise,i}}$

Table 3
Basic multiplex PCR thermocycling conditions and variable factors that were subject of optimization.

PCR step	Time / s	Temperature / °C	Cycles
Initial denaturation	30	98	1
Denaturation	10	98	variable (Cyc)
Annealing	variable (ATime)	variable (ATemp)	
Extension	variable (ETime)	72	
Final extension	120	72	1
Hold	infinite	4	–

Table 4
Matrix of experiments and response values included in the CCF design with four factors.

Experiment	ATemp / °C	ATime / s	ETime / s	Cyc	CV	CR	SN
N1	58	20	30	25	0.70	0.89	35.1
N2	66	20	30	25	2.80	0.97	18.5
N3	58	40	30	25	0.47	0.75	30.2
N4	66	20	90	25	1.81	0.97	12.4
N5	58	40	90	25	0.45	0.85	22.4
N6	66	40	90	25	1.80	0.97	13.0
N7	58	20	30	35	0.85	0.46	23.1
N8	58	40	30	35	0.76	0.40	22.7
N9	66	40	30	35	1.71	0.91	13.2
N10	58	20	90	35	0.84	0.72	21.9
N11	66	20	90	35	1.48	0.91	12.5
N12	66	40	90	35	1.66	0.90	14.8
N13	58	30	60	30	0.63	0.73	28.6
N14	66	30	60	30	1.43	0.96	18.8
N15	62	20	60	30	0.94	0.92	25.2
N16	62	40	60	30	0.81	0.91	20.1
N17	62	30	30	30	1.18	0.76	28.7
N18	62	30	90	30	0.83	0.91	17.6
N19	62	30	60	25	0.58	0.92	30.0
N20	62	30	60	35	1.04	0.69	21.2
N21	62	30	60	30	0.73	0.89	28.7
N22	62	30	60	30	0.72	0.93	27.5
N23	62	30	60	30	0.72	0.93	27.5

were performed using MODDE Go software v12.0.1. (Sartorius Stedim Data Analytics, Umeå, Sweden). Based on toaSTR output and observation classification, response values for CV, CR, and SN were calculated as described in Table 2 using Microsoft Excel. Response values were fed back into the MODDE software for model generation by means of multiple linear regression (MLR). MLR is a method for fitting a linear regression model to the observed data. A model can be written in the form $y = \beta \cdot X + \varepsilon$, where y is an n-dimensional vector of response values; β is a k-dimensional vector of regression coefficients; X is a 1 + n*k matrix of model terms including the constant; and ε is an n-dimensional vector of residuals. An equation system is said to be linear when it can be expressed as a linear combination of the β-coefficients, while it does not matter that the regression models may also include higher order polynomial (quadratic, qubic) x-terms. MODDE uses the singular value decomposition approach [43] to solve the equation system by minimizing the sum of squares of the distances between modelled and measured data points. Subsequent statistical validation of regression models was based on ANOVA. Evaluated parameters are presented in results Section 3.4 and in Supplementary Material S2.

3. Results

3.1. MPS-STR multiplex construction

We designed a multiplex PCR kit for the single-tube co-amplification of 21 forensic markers using a semi-automated primer engineering workflow. All primers fulfilled the criteria as set out in the primer engineering section. The target specific part of primers (without Illumina overhang) had a length between 19–27 bp (mean 22 bp) and an average GC content of 44% (30–55%) with melting temperatures in a narrow range around 64 °C (61–66 °C). Most amplicons were predicted to have a uniform size between 250 and 350 bp (Fig. 2). Loci with a wider distribution of length-based alleles (FGA, SE33, Penta E, and D21S11) may generate amplicons slightly larger or smaller in size depending on the number of repeats. The amelogenin fragment was intentionally designed shorter in order to serve also as a sample degradation control.

The amplicon layout was asymmetric by design. We intended to include as much flanking region variation as possible while complying with the upper amplicon size limit and other primer criteria. This partly resulted in unequal lengths of the left and right flanking regions. By adopting this approach, up to eleven SNP sites could be included per amplicon (Fig. 3). Supplementary Material S1 details SNP locations including rs numbers and allele frequencies as shown from 1000

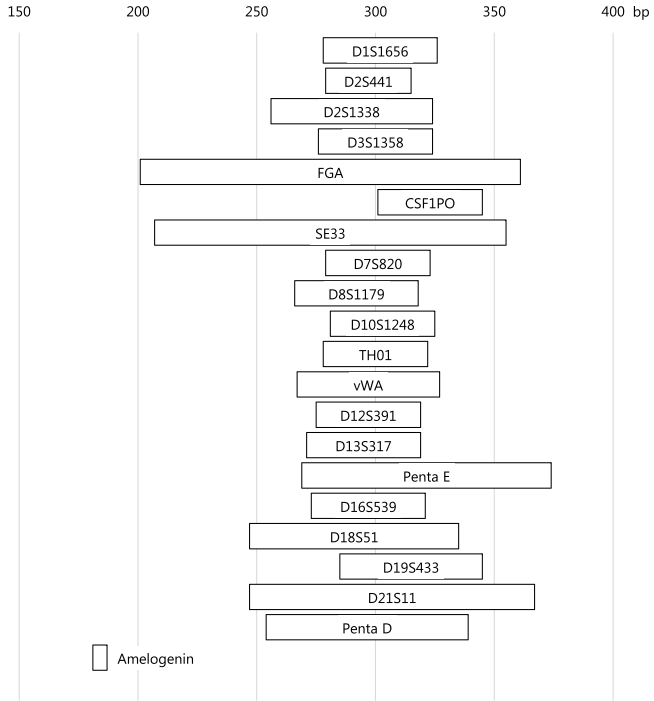


Fig. 2. PCR product sizes (without Illumina overhang) of the monSTR multiplex.

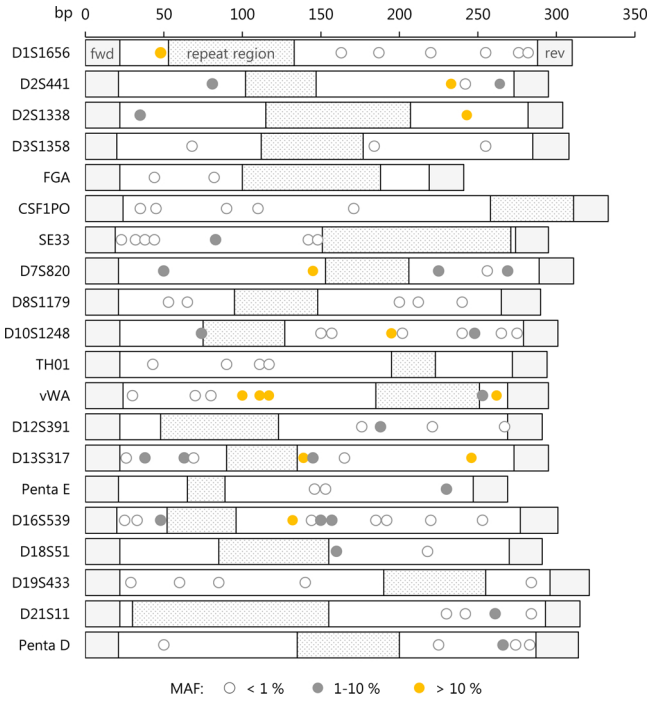


Fig. 3. Asymmetric amplicon design illustrating the target-specific primer binding sites (fwd, rev), repeat region, and amplified flanking regions. Each dot indicates a SNP location color-coded according to its 1000 Genomes global minor allele frequency (MAF). Size of the repeat region corresponds to the respective GRCh38 reference genome allele.

Genomes Project Phase 3. Most amplicons cover at least one SNP with a global minor allele frequency (MAF) ≥ 1%. The multiplex assay also includes twelve SNPs representing highly informative levels of flanking region variation, defined here as having a global MAF ≥ 10%. These SNPs are: rs4847015 in the D1S1656 locus; rs79534691 in the D2S441 locus; rs6736691 in the D2S1338 locus; rs16887642 in the D7S820 locus; rs2246512 in the D10S1248 locus; rs11063971, rs11063970, rs11063969, rs75219269 in the VWA locus; rs9546005, rs9531308 in the D13S317 locus, and rs11642858 in the D16539 locus.

In silico PCR with this primer set suggested a highly specific amplification and low tendency to form primer dimers and hairpin structures. MFEprimer predicted exactly 22 amplicons (20 STRs, amelogenin X and Y). Only two of 42 primers, D8S1179 forward and D12S391 forward, showed a potential to form a hairpin structure. Additionally, one potential heterodimer formed of the D18S51 reverse and amelogenin reverse primer was indicated. However, the lower binding stability ($Tm = 23\,°C$, $\Delta G = -3.14$ kcal/mol compared to $\Delta G < -20$ kcal/mol for all other amplicons) indicated that this structure would likely not exist under real PCR conditions. All targets were amplifiable in singleplex PCRs under standard conditions and the produced fragments were within the expected size range.

3.2. Sequencing run information

MiSeq sequencing runs for this study had a mean cluster density of 806 ± 50 K/mm^2. An average of 92.9 ± 1.2% of clusters passed filters. In total, 3.1 GB of data was obtained with 6.1 M reads. Among the reads, 87.3 ± 2.2% of bases satisfied a quality score of 30 (Q30). After merging of paired-end reads and quality filtering, 4.2 M reads were subjected to toaSTR analysis. In the DOE experimental series, the mean coverage per sample was 53,914 ± 17,541 reads and on average 2,165 ± 2,670 reads per locus were called in toaSTR. Confirmation and sensitivity studies showed a mean coverage per sample of 65,049 ± 2,068 reads and on average 2,562 ± 1,174 reads were called per locus.

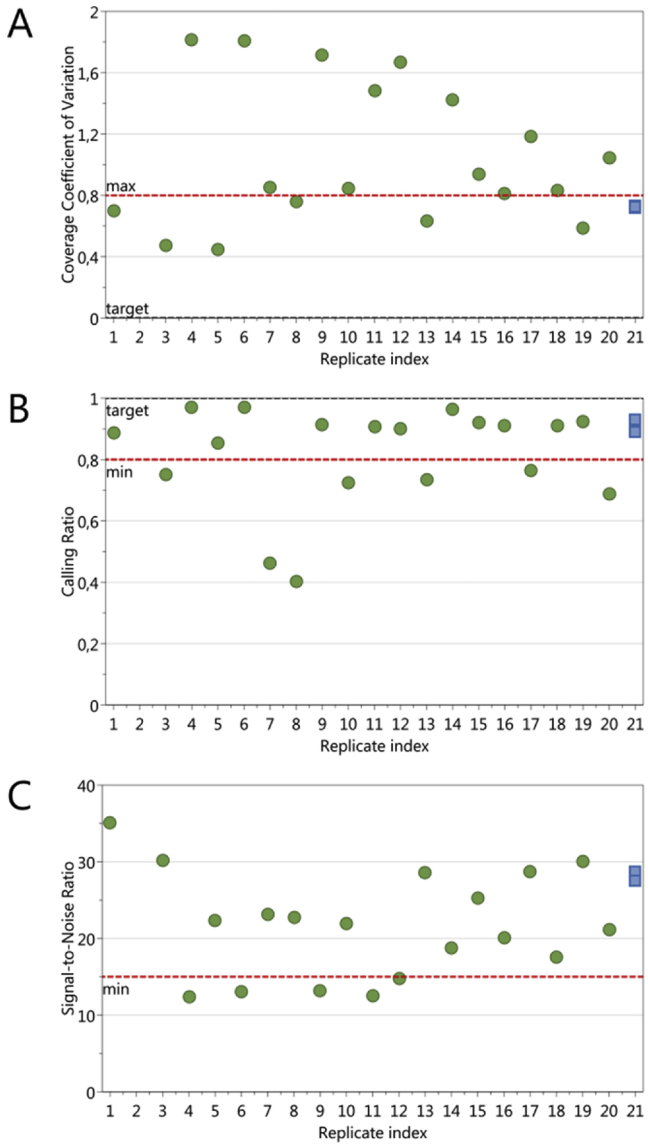


Fig. 4. Replicate plots displaying response values against the replicate index. Experiments with unique factor settings (circles) have a unique index and replicates (rectangles) share index number 21.

3.3. Evaluation of raw data

Twenty experiments with unique thermocycling conditions plus three replicates (N1–N23) were performed according to the experimental design (Table 4). Sample N2 was excluded from analysis due to dropout of six loci (CSF1PO, D10S1248, TH01, D12S391, vWA, amelogenin), which was probably a result of highly unfavorable PCR conditions in this experiment. Not more than two locus dropouts were observed in the other samples. The analysis of experimental data involved three basic steps, including (i) evaluation of raw data, (ii) regression model generation and analysis, and (iii) model interpretation.

Each experiment represented a unique combination of factor settings (Table 4). The obtained responses were visualized in replicate plots in order to estimate replicate variation versus overall variation. Replicate plots for the three responses coverage coefficient of variation (CV), calling ratio (CR), and average signal-to-noise ratio (SN) are shown in Fig. 4. In all cases the variation in the three center point replicates (N21–N23, blue rectangles) was much smaller than the variation in the entire investigation series, suggesting good control over the experimental procedure. Data also ensured that the expected range of

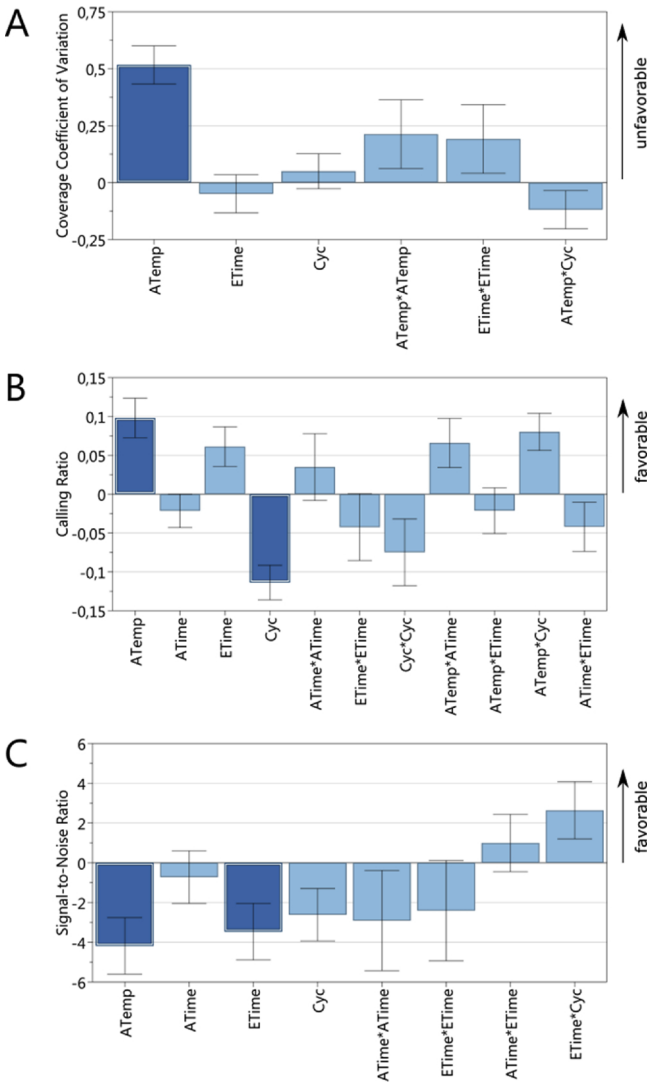


Fig. 5. Scaled and centered regression coefficients after model refinement. The height of a bar indicates the response change as the factor changes from its standard condition (i.e., the center point setting) to the high level. Depending on the response model, adjusting a factor to its high level corresponds to increasing or decreasing the response value, respectively. A high CV value (A) was unfavorable, while high CR and SN values (B, C) were favorable. Dominant factors are highlighted. An asterisk denotes an interaction between two factors or a quadratic effect. Superimposed 95% confidence intervals.

response values was covered.

3.4. Generation of multiplex PCR models and regression analysis

Statistical models were generated by applying multiple linear regression (MLR) to the experimental data. MLR is based on finding the regression model which minimizes the residual sum of squares of the respective response. Generally, a model of second order includes linear, quadratic, and interaction terms of factors which are weighted by their regression coefficients. The value of a coefficient indicates the response change as the factor moves from its standard condition (i.e., the center point setting). Coefficient plots (Fig. 5) were used to identify driving factors as well as insignificant terms in order to refine the models by excluding the latter. Confidence intervals give the uncertainty of coefficients and enable to estimate the significance of model terms. In case factors or interactions had no determinable effect, i.e. the confidence interval included zero, they were removed from the model, except otherwise stated.

Apparently, the CV model (Fig. 5A) was dominated by the factor ATemp (0.51 ± 0.08, $p = 1.2{\cdot}10^{-9}$). Rather large squared terms ATemp*ATemp and ETime*ETime were detected (an asterisk denotes an interaction between two factors or a quadratic effect). Although linear terms for ETime and Cyc were not significant as their confidence intervals included zero, they remained in the model since they were part of higher order terms or interaction terms. ATime was completely removed from the model as it had no detectable effect on CV.

Fig. 5B illustrates ATemp and Cyc had a strong impact on CR (0.10 ± 0.03, $p = 6.37{\cdot}10^{-6}$ and -0.11 ± 0.02, $p = 4.8{\cdot}10^{-7}$). These factors also participated in several quadratic and interaction terms in this model with partly contrary effects. ATime had a comparatively little influence. The CR model was characterized by a number of interactions between factors and second order terms.

Interestingly, ATemp (-4.18 ± 1.42, $p = 2.5{\cdot}10^{-5}$) was also the most important factor on SN (Fig. 5C). The second-most influential factor, ETime, amounted to -3.46 ± 1.42 ($p = 1.5{\cdot}10^{-4}$). A rather large interaction, ETime*Cyc, was also detected (2.64 ± 1.44, $p = 1.6{\cdot}10^{-3}$). The quadratic term ETime*ETime was a borderline case according to the confidence interval assessment ($p = 0.06$) but remained in the model since its exclusion would have decreased model quality. Comparable to the other models, the linear ATime term did not impact the target value of SN significantly ($p = 0.26$). However, a strong negative quadratic effect ATime*ATime was observed (-2.91 ± 2.52, $p = 0.03$).

Upon deletion of several insignificant regression terms, we analyzed commonly proposed statistical parameters ('summary of fit') to assess the model quality. The parameter R^2 (explained variation) is a measure how well a model fits current response data. An R^2 of 1 indicates a perfect model. The obtained R^2 values were 0.96, 0.98, and 0.95 for CV, CR, and SN, respectively. The Q^2 value (predicted variation) represents the goodness of prediction for future data and should be > 0.5 for a good model [43]. It amounted to 0.87, 0.91, and 0.82, respectively, for CV, CR, and SN. The model validity, which is based on the ANOVA lack of fit test, was above the common reference value of 0.25 for CR and SN, indicating valid models. A negative value for the CV model was considered an artifact arising from an extremely small replicate error due to almost identical replicates N21 – N23. The fourth parameter, reproducibility, is an indicator of the replicate error in relation to the overall variation. It amounted to 0.99, 0.98, and 0.99, respectively, and suggested excellent reproducibility and good control of the experimental procedure.

Normal probability plots (N-plots) displaying the distribution of residuals were used to search for deviating experiments (outliers) that might skew the models. Except for experiment N2, which has already been excluded previously due to a large number of locus dropouts, no outliers were detected. N-plots and complementary statistics are available in Supplementary Material S2.

Inserting regression coefficients yielded the following model equations, which describe the behavior of the multiplex PCR system in the region studied:

$$CV = 0.79 + 0.52x_1 - 0.05x_3 + 0.05x_4 + 0.21x_1x_1 + 0.19x_3x_3 - 0.12x_1x_4 + e$$

$$CR = 0.88 + 0.09x_1 - 0.02x_2 + 0.06x_3 - 0.11x_4 + 0.03x_1x_1 - 0.04x_3x_3 - 0.07x_4x_4 + 0.07x_1x_2 - 0.02x_1x_3 + 0.08x_1x_4 - 0.04x_2x_3 + e$$

$$SN = 25.94 - 4.18x_1 - 0.72x_2 - 3.47x_3 - 2.61x_4 - 2.91x_2x_2 - 2.41x_3x_3 + 0.99x_2x_3 + 2.64x_3x_4 + e$$

with x1: ATemp, x2: ATime, x3: ETime, x4: Cyc, e: residual error. The regression coefficients were scaled and centered, therefore the constant terms relate to the estimated response value at the design center-point.

The p-values for the models were determined to be p (CV) = $2.30{\cdot}10^{-8}$, p(CR) = $5.31{\cdot}10^{-7}$, p(SN) = $1.16{\cdot}10^{-6}$, indicating

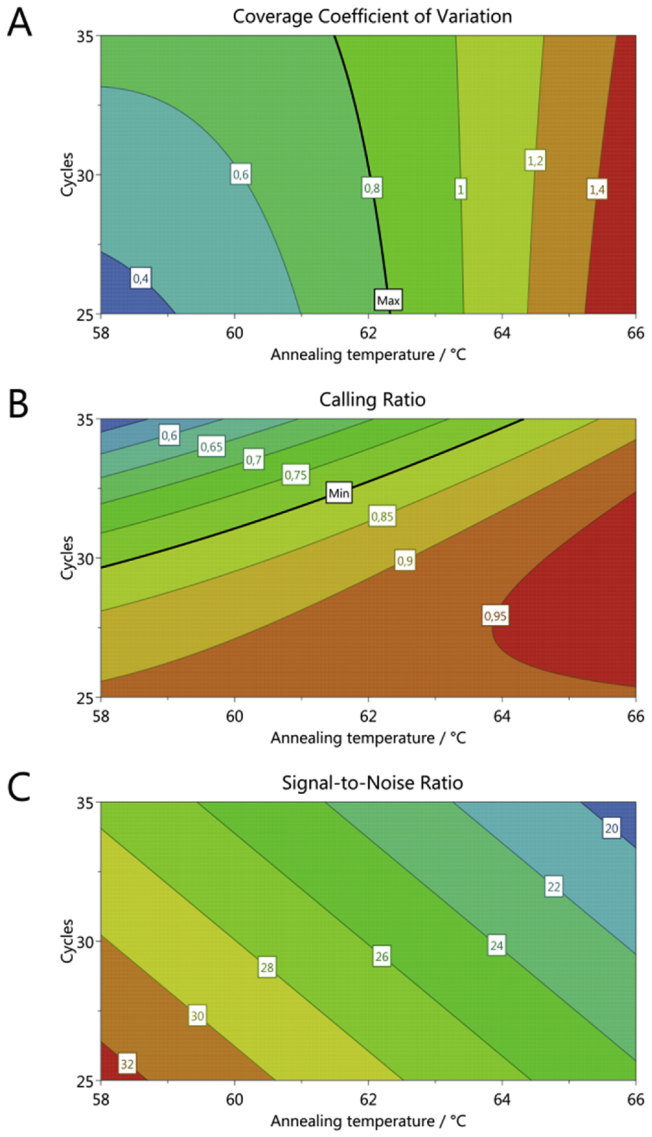


Fig. 6. Response contour plots showing the predicted responses as a function of annealing temperature and number of PCR cycles. Other factors were fixed at constant levels (annealing time = 30 s, extension time = 60 s). Values in white boxes are given in the unit of the respective response.

that the variance that can be explained by the models is significantly larger than the unmodellable (residual) variance. The models were subsequently used to predict the responses for any condition within the experimental space and to draw conclusions about optimal factor settings.

3.5. Model interpretation and optimization of thermocycling conditions

The main objective of an optimization approach guided by DOE is to convert experimental data into an informative map of the system under investigation. Response contour plots (Fig. 6) allow localization of the optimal working zone which will enable the best outcome of the respective response. After the factors ATemp and Cyc were identified as driving forces in all three models, ETime and ATime were fixed at their center-point levels for visualization purposes. Response contour plots of CV and CR showed strong curvature as a result of the quadratic model terms ATemp*ATemp and Cyc*Cyc, respectively. As indicated in the CV plot (Fig. 6A), low annealing temperature and reduced number of PCR cycles would be preferable for a low interlocus imbalance. A similar pattern was observed in the SN plot (Fig. 6C); ATemp and Cyc must be low in order to yield a high signal-to-noise ratio. Conversely, results for CR (Fig. 6B) showed that high ATemp and low Cyc favored a higher calling ratio, i.e. a higher ratio of on-target reads. Therefore, a simple visual inspection of the models' response contour plots did not immediately reveal experimental conditions which satisfy these partly opposing optimization goals simultaneously.

The MODDE optimizer tool was used to obtain more exact coordinates of the global optimum. The optimizer uses a Nelder-Mead 'simplex algorithm' together with the fitted models in order to find combinations of factor levels that jointly optimize all responses. The criteria were to minimize CV and maximize both CR and SN in compliance with the limits and targets described in Table 2. Optimization converged in the following factor settings: ATemp = 58 °C, ATime = 23 s, ETime = 65 s, Cyc = 25. The predicted response values for these settings are given in Table 5B.

Since the optimum for ATemp and Cyc was located at the edge of the investigated region, it was necessary to test if the true optimum may lie outside the experimental space. We re-centered the design around the optimized factor settings and performed a second set of RSM experiments in vitro with the following factor ranges: ATemp: 54–62 °C, ATime: 10–30 s, ETime: 30–90 s, Cyc: 20-30. No significant improvement was found with lower ATemp or Cyc values, hence the settings of the above described investigation series were retained (data not shown).

3.6. Experimental confirmation of modelled conditions

Confirmation experiments were conducted in duplicate to verify the predicted responses. The optimized thermocycling protocol was 98 °C for 30 s, 25 cycles of (98 °C/10 s, 58 °C/23 s, 72 °C/65 s), and 72 °C for 2 min. Observed response values revealed that SN could be clearly improved compared to the center-point conditions (i.e., the standard setup before optimization). The confirmed value (Table 5C) was close to the value predicted by the MODDE optimizer (Table 5B). CR was predicted to increase with the optimized settings, however, was slightly lower than the center-point value. CV has markedly improved during optimization, although the predicted value of 0.27 was not reached.

Up to this point, DOE experiments and confirmation of optimized settings were carried out with an equimolar mix of primers. Further improvement of the interlocus balance, hence a decrease in CV, was

Table 5
Response values of non-optimized settings (mean center-point results N21-N23) in comparison with predicted and experimentally confirmed response values after optimization. Precision given as 95% confidence interval (B) or standard deviation (A, C, D), respectively.

	Coverage Coefficient of Variation (CV)	Calling Ratio (CR)	Signal-to-Noise Ratio (SN)
Standard factor settings			
A) center-point	0.72 ± 0.01	0.92 ± 0.02	28 ± 1
Optimized factor settings			
B) predicted	0.27 ± 0.02	0.95 ± 0.05	32 ± 3
C) confirmed	0.61 ± 0.01	0.87 ± 0.02	38 ± 4
D) primer-adjusted	0.43 ± 0.03	0.93 ± 0.01	34 ± 5

Table 6
Adjusted primer pair concentrations per multiplex reaction.

Locus	Molar concentration / µM
Amelogenin	0.11
CSF1PO	0.08
D10S1248	0.08
D12S391	0.14
D13S317	0.08
D16S539	0.12
D18S51	0.08
D19S433	0.08
D1S1656	0.08
D21S11	0.08
D2S1338	0.10
D2S441	0.08
D3S1358	0.08
D7S820	0.08
D8S1179	0.12
FGA	0.10
Penta D	0.10
Penta E	0.10
SE33	0.10
TH01	0.08
vWA	0.10

expected by titrating primer concentrations in the multiplex reaction. Balancing of primer concentrations was performed empirically based on the average locus coverage distribution of LC3, 2800 M, and 9948 samples with varying DNA input amounts (2 ng, 1 ng, 500 pg). Table 6 presents the composition of the final primer mix. The total primer concentration per multiplex reaction was 2 µM.

The robustness of this multiplex assay was tested in a sensitivity study with varying dilutions of 2800 M DNA between 1 ng and 125 pg in duplicates. We observed a homogeneous amplification of all 21 targets for every dilution step, as can be seen from Fig. 7. Expressed as absolute numbers, the weakest represented locus, amelogenin, had a mean coverage of 1,109x ± 379x, while the strongest locus, D18S51, had an average coverage of 4,628x ± 646x, implying a coverage ratio of about factor three to five between the strongest and weakest locus. Primer-adjusted responses for CV, CR, and SN were calculated by averaging the four dilution steps (Table 5D). A CV of 0.43 ± 0.03 underlined that the interlocus balance was substantially improved upon primer adjustment, while CR and SN were still sufficiently high.

Genotypes were in full concordance with CE results for samples LC3, 2800 M, and 9948. No locus or allele dropouts were observed in any of the dilution steps down to 125 pg.

4. Discussion

The objective of this study was to propose an efficient, transparent workflow for multiplex PCR development and optimization, whose results can be better communicated than findings resulting from an intuitive approach. We demonstrated how primer engineering tools and the experimental design strategy can be combined to establish a forensic identity panel for MPS-STR genotyping.

4.1. Multiplex PCR assay development

The selection of forensic markers covered by monSTR is compatible with the German national DNA database (DAD) and contains markers commonly used in Europe and the USA. We decided not to include the trimeric locus D22S1045, although being part of the expanded ESS and CODIS set, since previous forensic MPS kit evaluations and CE experiments suggested weak performance regarding heterozygote allele balance [2], locus coverage [5], and stutter ratio [15]. Instead, we supported two pentanucleotide systems, Penta D and Penta E, which combine properties of high power of discrimination and low levels of stutter [53]. Additionally, this multiplex targets the highly polymorphic marker SE33, which is currently a unique feature of monSTR compared to other MPS-STR kits. Although SE33 is included in Illumina ForenSeq Signature Prep Kit primer mixes, it is not analyzed by proprietary UAS software [54]. While in-depth validation of locus performance was beyond the scope of the present work, preliminary results indicated a reasonable amplification of SE33 and a good heterozygous balance

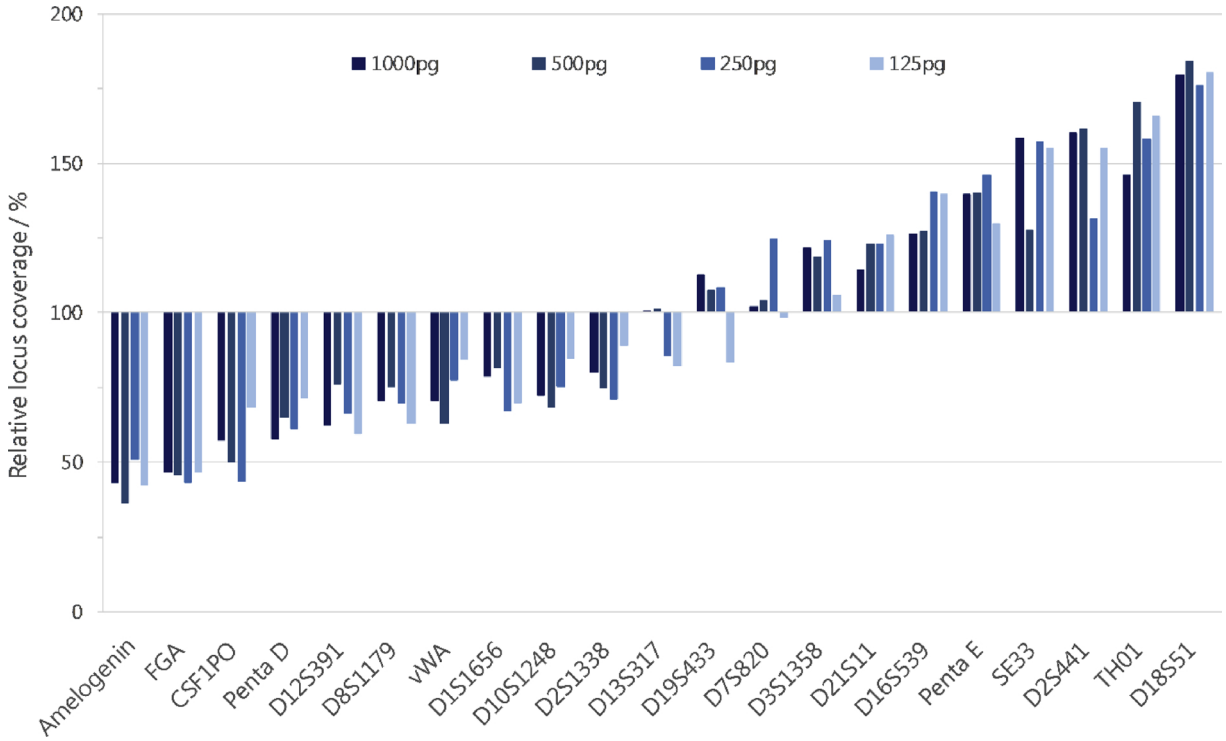


Fig. 7. Locus balance depending on the amount of 2800 M DNA input. Relative coverage of a given locus was calculated by dividing the locus coverage by the mean coverage of all included loci.

comparable to other loci even for genotypes exhibiting a larger allele spread.

With its medium size of 21 markers, monSTR may be an economic alternative to larger kits, since it faciliates pooling of more samples per sequencing run even on a smaller flow cell in a low-throughput scenario. monSTR sequence data can be analyzed with open-access STR genotyping software like toaSTR [52].

Careful design of primers is essential for the successful amplification of numerous loci in a single reaction. Non-specific products may arise if primers bind to template sites with similar sequences, and reduced amplification may occur due to primer-template mismatches [24]. A multiplex-compatible primer set was automatically designed based on GRCh38 reference sequences using the MPprimer program. In future, it might be possible to further increase the degree of automation for multiplex primer engineering. SNP database information could be linked to target sequences, so that candidate primers with too many SNPs in their binding site would be rejected automatically, and primer pairs that cover a maximum amount of flanking region variation would be favored. RExPrimer software [55] addresses some of these issues by integrating genomic variation databases, however, does not support batch construction of primer pairs for multiple templates.

The anticipated amplicon sizes were a tradeoff between compatibility with degraded DNA and the ability to include flanking variation. The monSTR multiplex covers twelve highly informative SNPs with a global MAF ≥ 10% that may provide additional discriminatory features. However, the analysis of flanking region variation using the monSTR multiplex assay was not part of the present study. The final primer set was characterized by similar hybridization kinetics for all primers regarding balanced GC content and comparable melting temperatures within a range of 5 °C. In silico PCR predicted a highly specific amplification of all 21 targets, while the formation of primer-dimers and secondary structures due to internal complementarity could be largely avoided. Results of in vitro testing in singleplex and multiplex reactions under standard PCR conditions suggested a solid basis for successive optimization of the thermocycling protocol.

4.2. DOE enabled efficient and transparent assay optimization

Multiplex PCR includes a multitude of parameters which need to be optimized [23]. Commonly, optimization relies on the sequential investigation of each variable, which involves a large number of experiments and substantial amounts of time in order to include all possible combinations [34]. Moreover, a sequential approach ignores potential interactions between factors and will rarely identify true optimum conditions. DOE methodology is virtually predestined for the systematic optimization of a multifactorial system [56]. In essence, DOE provides an organized approach to plan informative experiments, assess the joint influence of all factors, and create a meaningful map of the system for decision making. As illustrated in the present study, only 23 experiments were necessary to provide good coverage of the experimental space spanned by four input factors, saving time and costs. However, some work is required for familiarization with experimental design principles and software tools.

While DOE principles are commonly applied in the field of engineering, adaption to multiplex PCR required some additional effort. Since the outcome of an MPS-STR assay cannot be measured intuitively, we had to translate our optimization goals into three representative, measurable response values that allowed to perform statistical analysis. First, the inter-locus balance was reflected by the coefficient of variation of the coverage (CV). Large variation in coverage can impact sample throughput of MPS systems [2] or may lead to dropouts of alleles or complete loci. Second, PCR specificity was expressed by the calling ratio (CR) in toaSTR. CR expresses the proportion of on-target reads, which decreases if sequence data contains many reads of primer-dimers or non-specific PCR products. Third, signal-to-noise ratio (SN) is the coverage ratio of noise observations to allelic observations in toaSTR. We included this parameter to assess if undesired technical artefacts like stutter products (i.e., amplicons that are one or more repeat units longer or shorter in size than the parent allele) can be reduced by choosing favorable thermocycling conditions.

Calculated response values of the experiments were used for software-based generation of regression models for the three responses to identify optimal multiplex PCR parameters. Modeling first- and second-order terms and factor interactions is commonly expected to be sufficient to capture the most critical components of a biological process [28]. The curvature in response contour plots (Fig. 6) confirmed that it was reasonable to assume quadratic relationships in the model in order to obtain a reliable approximation of the multiplex PCR process. Four factors were varied according to a CCF experimental design incorporating twenty-three runs. Results indicated that each factor had significant impact on one or more responses, either as a main effect or as part of an interaction. Statistical analysis allowed not only to *identify* driving effects, but to *quantify* the influence of factors on the responses and describe these relations in mathematical models. During model refinement, models were reduced to the most significant terms. Removing an insignificant term was justified if it resulted in an increase in the model fitness, Q^2.

Conclusions that could be drawn from the coefficient plots were partly consistent with the literature but partly contradictory. Previously, the annealing temperature (ATemp) has been described as one of the most important factors on multiplex PCR performance [23]. Raising ATemp generally increases the stringency of primer-template hybridization, tolerating less mismatches at the priming site and leading to a decreased amplification of non-specific sequences [57]. This effect was also found in the CR model of our study, indicating that adjusting ATemp to higher levels increased the calling ratio, hence improved the specificity of PCR products. Conversely, results suggested that high ATemp was detrimental to CV, inducing a stronger inter-locus imbalance. A possible explanation for this might be that more stringent temperatures in combination with competitive effects in a multiplex reaction may lead to a preferential amplification of some targets. Locus dropouts were observed whenever the 'high' ATemp level (66 °C) was tested (experiments N2, N4, N6, N9, N11, N12, and N14). Primers for loci where dropouts were seen more frequently, particularly D13S317 and vWA, may undergo further optimization. Practically, however, developing primers for a multiplex assay targeting more than two loci always requires a reasonable compromise of primer properties. Using optimized settings, data suggested a well-balanced coverage without any dropouts even with low DNA input amounts.

A largely negative term for ATemp was also found in the SN model, implying more locus-specific PCR artefacts or less 'true' allele amplicons when increasing ATemp. Whether SN decreased due to an increase of noise observations or a decrease of signal observations was uncertain since at the same time the locus coverage was highly variable and hardly comparable. Further research should be undertaken to investigate this relation. Results support the hypothesis that a lower ATemp enhances the stability of the polymerase/template complex [58]. This stability would reduce the likelihood of polymerase slippage, resulting in lower stutter ratios. Previous work of Seo et al. [59] indicated that a lower temperature annealing/extension step was the primary factor in the stutter reduction observed when typing low copy number samples. In our work, the overall process performance was predicted to benefit the most from lowering ATemp to 58 °C, accepting that CR slightly decreased but still complied with the lower limit of 0.8.

Since the polymerase has some activity at ATemp, a longer annealing time (ATime) has been associated with a higher risk of non-specific products [57]. Consistent with the literature, we found that extending ATime reduced CR. Also a significant negative quadratic effect was observed in the SN model, which may be linked to a higher likelihood of polymerase/template dissociation. Interestingly, the interaction of ATemp and ATime appeared to have a positive effect on CR, which exemplifies the importance to investigate factor interactions in

optimization strategies. Changing ATime did not impact inter-locus balance. Compared to the non-optimized setting of 30 s, this factor was slightly reduced to 23 s.

As mentioned in the literature, multiplex PCR may require longer extension time (ETime) for the polymerase to complete the synthesis of all products, as the pool of enzyme and nucleotides becomes a limiting factor [23]. On the other hand, it was described that long ETime could provide opportunity for non-specific amplification [25]. The present study found that longer ETime was beneficial for a specific and complete amplification (see CR model). In contrast, the SN model implied the generation of more technical artefacts, which could be associated with an increased chance for polymerase slippage events due to longer ETime. An ETime of 65 s was predicted to jointly optimize all responses.

Prior studies have noted that keeping the number of PCR cycles (Cyc) to the minimum required to generate sufficient product reduces the likelihood of PCR errors and the accumulation of non-specific products [57]. Hence, the negative effect for Cyc, which dominated the CR and SN models, was not unexpected. It is somewhat surprising that the interaction ETime*Cyc was the only significant positive effect on SN. This result was counterintuitive since both linear factors were detrimental on this response. Further research is therefore suggested. For an optimal outcome, a reduction of Cyc to 25 was predicted. It should be noted that monSTR library preparation includes a second PCR step with 8 cycles, totaling 33 cycles of amplification before sequencing.

Optimal factor settings were predicted to improve all responses compared to the non-optimized standard thermocycling conditions. However, particularly the confirmed CV response exhibited a somewhat larger deviation from the predicted value (Table 5, B vs C). The observed deviation may be explained by stochastic effects during library preparation and by sequencing inter-run variation. Since the experiments for optimization, confirmation, and primer adjustment build upon each other, studies required serially performed sequencing runs. Future applications of DOE could benefit from a higher number of replicates for confirmation experiments. Also genotype-specific effects may play a role in the observed variability of responses. Genotype-specific characteristics (e.g., allele spread, heterozygosity) may affect CV and SN, since longer alleles tend to have a lower amplification efficiency and show a higher stutter ratio. SN as calculated in this study is sensitive to small noise changes when the noise coverage is close to zero and therefore SN will approach large values. However, given the complex nature of the biological process modelled here and the potential sources of variation within the analytical scheme, we found that confirmed responses were within acceptable differences to the predicted values. Results suggested that useful models of a multiplex PCR assay can be created based on statistically designed experiments. The thermocycling protocol was optimized with respect to CV and SN, while CR remained on a high level. It can therefore be assumed that the interlocus coverage balance can be improved not only by adjusting primer concentrations, but also by choosing favorable thermocycling conditions.

A series of dilution experiments with challenging input amounts of DNA revealed full STR profiles. No allele or locus dropouts were observed down to 125 pg, indicating a good sensitivity that has not yet been exhausted. Comparable sensitivity of MiSeq-based kits has been reported earlier [3,10,54]. However, with a small sample size of the present study, caution must be taken, as the findings do not replace a comprehensive kit validation. Further studies will have to validate monSTR's performance according to common validation guidelines including sensitivity, mixture, and casework studies (manuscript in preparation).

5. Conclusion

Targeting 20 STR markers including SE33 plus amelogenin, monSTR is a medium sized forensic identity panel that may complement the landscape of multiplex kits for the MiSeq platform. Using a chain of bioinformatics tools for economic primer engineering and testing, we constructed a highly specific and multiplex-compatible primer set.

This study has been one of the first attempts to apply Design of Experiments methodology to multiplex PCR optimization. An efficient RSM experimental design, which required only a minimum number of experiments, supported the generation of realistic multiplex PCR models. These models were successfully used to predict optimized settings for critical thermocycling parameters. In vitro experiments revealed that the assay performance was substantially improved.

Overall, this study strengthens the idea that computer-assisted modelling and simulation of biological mechanisms may reduce time and cost in the development and optimization stage of multifactorial processes such as multiplex PCR. We believe that DOE holds potential as an optimization tool for molecular analytic and diagnostic assays. Experimental design methods, either performed in-house or via an external service, enables laboratories to develop and optimize custom assays with reasonable resources.

Conflict of interest

CT, JS, and SG are current employees of LABCON-OWL. The authors declare no other competing interests.

Role of funding

This work was supported by a grant from the ZIM funding program (Zentrales Innovationsprogramm Mittelstand) of the German Federal Ministry for Economic Affairs and Energy (grant number ZF4453702).

CRediT authorship contribution statement

Sebastian Ganschow: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Writing - original draft, Writing - review & editing. **Janine Silvery:** Data curation, Funding acquisition, Investigation, Validation, Writing - review & editing. **Carsten Tiemann:** Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing.

Acknowledgements

We thank Prof. Dr. Dirk Lütkemeyer (Bielefeld University of Applied Sciences), Prof. Dr. Peter Wiegand (Institute of Legal Medicine, University Hospital of Ulm), and Dr. Robert Kulis-Horn of our lab for their excellent advice during the project.

Appendix B. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fsigen.2018.11.023.

References

[1] N.M.M. Novroski, J.L. King, J.D. Churchill, L.H. Seah, B. Budowle, Characterization of genetic sequence variation of 58 STR loci in four major population groups, Forensic Sci. Int. Genet. 25 (2016) 214–226, https://doi.org/10.1016/j.fsigen.2016.09.007.

[2] J.D. Churchill, S.E. Schmedes, J.L. King, B. Budowle, Evaluation of the illumina® Beta version ForenSeq™ DNA signature prep kit for use in genetic profiling, Forensic Sci. Int. Genet. 20 (2016) 20–29, https://doi.org/10.1016/j.fsigen.2015.09.009.

[3] E.H. Kim, H.Y. Lee, I.S. Yang, S.-E. Jung, W.I. Yang, K.-J. Shin, Massively parallel sequencing of 17 commonly used forensic autosomal STRs and amelogenin with small amplicons, Forensic Sci. Int. Genet. 22 (2016), https://doi.org/10.1016/j.fsigen.2016.01.001.

[4] F. Guo, Y. Zhou, F. Liu, J. Yu, H. Song, H. Shen, B. Zhao, F. Jia, G. Hou, X. Jiang, Evaluation of the early access STR kit v1 on the ion torrent PGM™ platform,

Forensic Sci. Int. Genet. 23 (2016) 111–120, https://doi.org/10.1016/j.fsigen.2016.04.004.
[5] P. Müller, A. Alonso, P.A. Barrio, B. Berger, M. Bodner, P. Martin, W. Parson, Systematic evaluation of the early access applied biosystems precision ID Globalfiler mixture ID and Globalfiler NGS STR panels for the ion S5 system, Forensic Sci. Int. Genet. 36 (2018) 95–103, https://doi.org/10.1016/j.fsigen.2018.06.016.
[6] S. Caratti, S. Turrina, M. Ferrian, E. Cosentino, D.D. Leo, MiSeq FGx sequencing system: a new platform for forensic genetics, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e98–e100, https://doi.org/10.1016/j.fsigss.2015.09.040.
[7] K.J. van der Gaag, R.H. de Leeuw, J. Hoogenboom, J. Patel, D.R. Storts, J.F.J. Laros, P. de Knijff, Massively parallel sequencing of short tandem repeats—population data and mixture analysis results for the PowerSeq™ system, Forensic Sci. Int. Genet. 24 (2016) 86–96, https://doi.org/10.1016/j.fsigen.2016.05.016.
[8] S.L. Fordyce, M.C. Ávila-Arcos, E. Rockenbauer, C. Børsting, R. Frank-Hansen, F.T. Petersen, E. Willerslev, A.J. Hansen, N. Morling, M.T.P. Gilbert, High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche genome sequencer FLX platform, BioTechniques 51 (2011) 127–133, https://doi.org/10.2144/000113721.
[9] X. Zhao, H. Li, Z. Wang, K. Ma, Y. Cao, W. Liu, Massively parallel sequencing of 10 autosomal STRs in Chinese using the ion torrent personal genome machine (PGM), Forensic Sci. Int. Genet. 25 (2016) 34–38, https://doi.org/10.1016/j.fsigen.2016.07.014.
[10] X. Zeng, J.L. King, M. Stoljarova, D.H. Warshauer, B.L. LaRue, A. Sajantila, J. Patel, D.R. Storts, B. Budowle, High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing, Forensic Sci. Int. Genet. 16 (2015) 38–47, https://doi.org/10.1016/j.fsigen.2014.11.022.
[11] ENFSI DNA Working Group, DNA Database Management. Review and Recommendations, (2016) (Accessed August 9, 2018), http://enfsi.eu/wp-content/uploads/2016/09/final_version_enfsi_2016_document_on_dna-database_management_0.pdf.
[12] J.M. Butler, Advanced Topics in Forensic DNA Typing: Methodology, Academic Press, 2011.
[13] P. Wiegand, B. Budowle, S. Rand, B. Brinkmann, Forensic validation of the STR systems SE 33 and TC 11, Int. J. Legal Med. 105 (1993) 315–320.
[14] L.A. Borsuk, K.B. Gettings, C.R. Steffen, K.M. Kiesler, P.M. Vallone, Sequencing of the highly polymorphic STR locus SE33, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e322–e323, https://doi.org/10.1016/j.fsigss.2017.09.125.
[15] K.B. Gettings, R.A. Aponte, P.M. Vallone, J.M. Butler, STR allele sequence variation: current knowledge and future issues, Forensic Sci. Int. Genet. 18 (2015) 118–130, https://doi.org/10.1016/j.fsigen.2015.06.005.
[16] R. Kalendar, B. Khassenov, Y. Ramankulov, O. Samuilova, K.I. Ivanov, FastPCR: an in silico tool for fast primer and probe design and advanced sequence analysis, Genomics 109 (2017) 312–319, https://doi.org/10.1016/j.ygeno.2017.05.005.
[17] Z. Shen, W. Qu, W. Wang, Y. Lu, Y. Wu, Z. Li, X. Hang, X. Wang, D. Zhao, C. Zhang, MPprimer: a program for reliable multiplex PCR primer design, BMC Bioinf. 11 (2010) 143, https://doi.org/10.1186/1471-2105-11-143.
[18] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3—new capabilities and interfaces, Nucleic Acids Res. 40 (2012) e115, https://doi.org/10.1093/nar/gks596.
[19] B. Yu, C. Zhang, In silico PCR analysis, Methods Mol. Biol. Clifton NJ 760 (2011) 91–107, https://doi.org/10.1007/978-1-61779-176-5_6.
[20] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T.L. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinf. 13 (2012) 134, https://doi.org/10.1186/1471-2105-13-134.
[21] R.M. Kuhn, D. Haussler, W.J. Kent, The UCSC genome browser and associated tools, Brief. Bioinform. 14 (2013) 144–161, https://doi.org/10.1093/bib/bbs038.
[22] W. Qu, Y. Zhou, Y. Zhang, Y. Lu, X. Wang, D. Zhao, Y. Yang, C. Zhang, MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity, Nucleic Acids Res. 40 (2012) W205–W208, https://doi.org/10.1093/nar/gks552.
[23] O. Henegariu, N.A. Heerema, S.R. Dlouhy, G.H. Vance, P.H. Vogt, Multiplex PCR: critical parameters and step-by-step protocol, Biotechniques 23 (1997) 504–511.
[24] J.M. Butler, Constructing STR Multiplex Assays, in: Forensic DNA Typing Protoc, Springer, 2005, pp. 53–65 (Accessed January 19, 2016), http://link.springer.com/protocol/10.1385/1-59259-867-6:053.
[25] M.C. Edwards, R.A. Gibbs, Multiplex PCR: advantages, development, and applications, Genome Res. 3 (1994) S65–S75.
[26] D.R. Pilipauskas, Can the time from synthesis design to validated chemistry be shortened? Med. Res. Rev. 19 (1999) 463–474, https://doi.org/10.1002/(SICI)1098-1128(199909)19:5<463::AID-MED11>3.0.CO;2-3.
[27] M. Altekar, C. Homon, M. Kashem, S. Mason, R. Nelson, L. Patnaude, J. Yingling, P. Taylor, Assay optimization: a statistical design of experiments approach, J. Assoc. Lab. Autom. 11 (2006) 33–41, https://doi.org/10.1016/j.jala.2005.11.001.
[28] D. Toms, R. Deardon, M. Ungrin, Climbing the mountain: experimental design for the efficient optimization of stem cell bioprocessing, J. Biol. Eng. 11 (2017), https://doi.org/10.1186/s13036-017-0078-z.
[29] P.D. Haaland, Experimental Design in Biotechnology, CRC Press, 1989.
[30] S.R.A. Fisher, The Design of Experiments, Oliver and Boyd, 1947.
[31] L. Ilzarbe, M.J. Álvarez, E. Viles, M. Tanco, Practical applications of design of experiments in the field of engineering: a bibliographical review, Qual. Reliab. Eng. Int. 24 (2008) 417–428, https://doi.org/10.1002/qre.909.
[32] R.S. Rao, C.G. Kumar, R.S. Prakasham, P.J. Hobbs, The Taguchi methodology as a statistical tool for biotechnological applications: a critical appraisal, Biotechnol. J. 3 (2008) 510–523, https://doi.org/10.1002/biot.200700201.
[33] G. Wrobel, J. Schlingemann, L. Hummerich, H. Kramer, P. Lichter, M. Hahn, Optimization of high-density cDNA-microarray protocols by 'design of experiments, Nucleic Acids Res. 31 (2003) e67, https://doi.org/10.1093/nar/gng067.
[34] B.D. Cobb, J.M. Clarkson, A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods, Nucleic Acids Res. 22 (1994) 3801–3805.
[35] M.D. Boleda, P. Briones, J. Farrés, L. Tyfield, R. Pi, Experimental design: a useful tool for PCR optimization, BioTechniques 21 (1996) 134–140.
[36] G. Caetano-Anollés, DAF optimization using Taguchi methods and the effect of thermal cycling parameters on DNA amplification, BioTechniques 25 (1998) 472–6, 478–80.
[37] J.L. Villarreal Camacho, E. Mendoza Torres, C. Cadena, J. Prieto, L.L. Varela Prieto, D.A. Villanueva Torregroza, The use of factorial design, image analysis, and an efficiency calculation for multiplex PCR optimization, J. Clin. Lab. Anal. 27 (2013) 249–252, https://doi.org/10.1002/jcla.21595.
[38] O. Dobay, E. Juhász, A. Ungvári, C. Jeney, S.G.B. Amyes, K. Nagy, Taguchi optimisation of a multiplex pneumococcal serotyping PCR and description of 11 novel serotyping primers, Acta Microbiol. Immunol. Hung. 56 (2009) 327–338, https://doi.org/10.1556/AMicr.56.2009.4.3.
[39] H.J. Celani de Souza, C.B. Moyses, F.J. Pontes, R.N. Duarte, C.E. Sanches da Silva, F.L. Alberto, U.R. Ferreira, M.B. Silva, Molecular assay optimized by Taguchi experimental design method for venous thromboembolism investigation, Mol. Cell. Probes 25 (2011) 231–237, https://doi.org/10.1016/j.mcp.2011.08.001.
[40] S. Wadle, M. Lehnert, S. Rubenwolf, R. Zengerle, F. von Stetten, Real-time PCR probe optimization using design of experiments approach, Biomol. Detect. Quantif. 7 (2015) 1–8, https://doi.org/10.1016/j.bdq.2015.12.002.
[41] M. Niens, G.T. Spijker, A. Diepstra, G.J.T. Meerman, A factorial experiment for optimizing the PCR conditions in routine genotyping, Biotechnol. Appl. Biochem. 42 (2005) 157–162, https://doi.org/10.1042/BA20040192.
[42] K.N. Ballantyne, R.A. van Oorschot, R.J. Mitchell, Reduce optimisation time and effort: Taguchi experimental design methods, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 7–8, https://doi.org/10.1016/j.fsigss.2007.10.050.
[43] L. Eriksson, Design of Experiments: Principles and Applications, Umetrics Academy, 2008.
[44] G.E.P. Box, The exploration and exploitation of response surfaces, Biometrics (1954) 16–60.
[45] Z. Shen, W. Qu, W. Wang, Y. Lu, Y. Wu, Z. Li, X. Hang, X. Wang, D. Zhao, C. Zhang, MPprimer: a program for reliable multiplex PCR primer design, BMC Bioinf. 11 (2010) 143.
[46] M.D. Gadberry, S.T. Malcomber, A.N. Doust, E.A. Kellogg, Primaclade—a flexible tool to find conserved PCR primers across multiple species, Bioinformatics 21 (2005) 1263–1264, https://doi.org/10.1093/bioinformatics/bti134.
[47] W. Qu, Y. Zhou, Y. Zhang, Y. Lu, X. Wang, D. Zhao, Y. Yang, C. Zhang, MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity, Nucleic Acids Res. 40 (2012) W205–W208, https://doi.org/10.1093/nar/gks552.
[48] Illumina, 16S Metagenomic Sequencing Library Preparation, Part # 15044223 Rev. B, (2013) (Accessed July 20, 2018), http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf.
[49] New England Biolabs, Protocol for Q5 Hot Start High-Fidelity 2X Master Mix, (2012) (Accessed July 20, 2018), https://www.neb.com/protocols/2012/08/30/protocol-for-q5-hot-start-high-fidelity-2x-master-mix-m0494.
[50] Beckmann Coulter, SPRIselect User Guide, PN B24965AA (2012).
[51] B. Bushnell, J. Rood, E. Singer, BBMerge—accurate paired shotgun read merging via overlap, PLoS One 12 (2017), https://doi.org/10.1371/journal.pone.0185056 e0185056.
[52] S. Ganschow, J. Silvery, J. Kalinowski, C. Tiemann, toaSTR: a web application for forensic STR genotyping by massively parallel sequencing, Forensic Sci. Int. Genet. 37 (2018) 21–28, https://doi.org/10.1016/j.fsigen.2018.07.006.
[53] J. Bacher, J.W. Schumm, Development of highly polymorphic pentanucleotide tandem repeat loci with low stutter, Profiles DNA 2 (1998) 3.
[54] S. Köcher, P. Müller, B. Berger, M. Bodner, W. Parson, L. Roewer, S. Willuweit, Inter-laboratory validation study of the ForenSeq™ DNA signature prep kit, Forensic Sci. Int. Genet. 36 (2018) 77–85, https://doi.org/10.1016/j.fsigen.2018.05.007.
[55] J. Piriyapongsa, C. Ngamphiw, A. Assawamakin, P. Wangkumhang, P. Suwannasri, U. Ruangrit, G. Agavatpanitch, S. Tongsima, RExPrimer: an integrated primer designing tool increases PCR effectiveness by avoiding 3' SNP-in-primer and mis-priming from structural variation, BMC Genomics 10 (2009) S4, https://doi.org/10.1186/1471-2164-10-S3-S4.
[56] J. Antony, Taguchi or classical design of experiments: a perspective from a practitioner, Sens. Rev. 26 (2006) 227–230, https://doi.org/10.1108/02602280610675519.
[57] M. McPherson, S. Møller, PCR, Taylor & Francis Ltd., Hoboken, 2006.
[58] V. Murray, C. Monchawin, P.R. England, The determination of the sequences present in the shadow bands of a dinucleotide repeat PCR, Nucleic Acids Res. 21 (1993) 2395–2398.
[59] S.B. Seo, J. Ge, J.L. King, B. Budowle, Reduction of stutter ratios in short tandem repeat loci typing of low copy number DNA samples, Forensic Sci. Int. Genet. 8 (2014) 213–218, https://doi.org/10.1016/j.fsigen.2013.10.004.

Developmental validation of the monSTR identity panel, a forensic STR multiplex assay for massively parallel sequencing

Title	Developmental validation of the monSTR identity panel, a forensic STR multiplex assay for massively parallel sequencing
Authors	Janine Silvery [a,b], Sebastian Ganschow [a], Peter Wiegand [b], Carsten Tiemann [a,c] [a] LABCON-OWL Analytik, Forschung und Consulting GmbH, Siemensstraße 40, 32105 Bad Salzuflen, Germany [b] Institute of Legal Medicine, University Hospital of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany [c] Faculty of Engineering and Mathematics, Bielefeld University of Applied Science, Interaktion 1, 33619 Bielefeld, Germany
Journal	Forensic Science International: Genetics
Journal ranking	4.884 (2018 Impact Factor)
Publication type	Research article
Publication year	2020
Publication status	Published journal article
DOI	https://doi.org/10.1016/j.fsigen.2020.102236
Author contributions (CRediT taxonomy)	**Janine Silvery:** Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Writing - original draft; Writing – review & editing. **Sebastian Ganschow:** Conceptualization; Data curation; Formal analysis; Methodology; Visualization; Writing – review & editing. **Peter Wiegand:** Funding acquisition; Supervision; Writing – review & editing. **Carsten Tiemann:** Funding acquisition; Project administration; Resources; Supervision; Writing – review & editing.

Forensic Science International: Genetics 46 (2020) 102236

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Developmental validation of the monSTR identity panel, a forensic STR multiplex assay for massively parallel sequencing

Janine Silvery[a,b,*], Sebastian Ganschow[a], Peter Wiegand[b], Carsten Tiemann[a,c]

[a] LABCON-OWL Analytik, Forschung und Consulting GmbH, Siemensstraße 40, 32105 Bad Salzuflen, Germany
[b] Institute of Legal Medicine, University Hospital of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany
[c] Faculty of Engineering and Mathematics, Bielefeld University of Applied Science, Interaktion 1, 33619 Bielefeld, Germany

ARTICLE INFO

Keywords:
Massively parallel sequencing (MPS)
Next-generation sequencing (NGS)
Short tandem repeat (STR)
Forensic genotyping
monSTR identity panel
SWGDAM validation

ABSTRACT

The 21-plex STR panel *monSTR* was designed for high-fidelity forensic genotyping on the Illumina MiSeq platform. In this study, the panel's performance was validated according to the recommended validation guidelines of the Scientific Working Group for DNA Analysis Methods (SWGDAM). Concordance, repeatability and reproducibility, sensitivity of detection, mixture analysis, species-specificity, and the ability to analyze mock samples were assessed. Sequence data was analyzed using the genotyping software *toaSTR*. The assay performance was evaluated by measuring the read on-target ratio, the genotype accuracy, the inter-locus balance, the heterozygosity balance, and the signal-to-noise ratio. Results showed that profiles of NIST reference DNA samples as well as GEDNAP proficiency samples were fully concordant with CE-based methods. In addition, inter-run and intra-run variation experiments indicated high precision. Furthermore, full profiles could be obtained using 62.5 pg of DNA input amount with proper inter-locus balance and read on-target ratio; 76.4% of alleles were correctly called with 7.8 pg DNA input amount. It was demonstrated that 94.4% of minor contributor alleles were resolved accurately in a 1:49 mixture. Results suggested that the minor contribution could be precisely calculated based on the minor component allele frequency. Validation results described here demonstrate that the monSTR forensic identity panel is a valid tool for forensic STR genotyping using massively parallel sequencing.

1. Introduction

For more than 20 years, forensic DNA analysis has routinely used short tandem repeat (STR) markers to unambiguously identify individuals because of their high discriminative power and multiplexing compatibility [1]. STRs, also known as microsatellites or simple sequence repeats (SSRs), are repetitive regions of DNA that differ in length and structure of their sequence motifs (usually 2–7 nucleotides) as well as in the number of their core repeats [2–4]. The gold standard method for forensic STR typing relies on the detection of allele lengths and includes multiplex PCR amplification of STR loci [5] followed by size separation of fragments using capillary electrophoresis (CE) [6]. CE typing proved to be highly efficient and cost-effective, while STR analysis cannot easily be combined with other types of forensic loci, e.g. single nucleotide polymorphisms (SNPs), mitochondrial DNA (mtDNA) or RNA markers, within the same assay.

In recent years, massively parallel sequencing (MPS) has become more and more attractive. This technology allows for the simultaneous evaluation of many genes and the generation of millions of short nucleic acid sequences of an individual in parallel. Nowadays, MPS is a mainstay for various applications in diagnostics and research [7–9]. In the field of forensic genetics, MPS may overcome CE-based limitations of STR typing [10]. MPS allows sequencing of many forensic genetic loci and different types of markers simultaneously with a high sequencing throughput in the same multiplex assay. In case of the Illumina MiSeq platform, this technique is based on unique adapter sequences that are added to both sides of the DNA fragment during the indexing process. This results in numerous different adapter combinations, which can be computationally separated based on the respective index sequence. Moreover, the technology provides a higher resolution compared to CE by detecting both allele length and sequence information of forensic STR markers. Thus, intra-allelic sequence variations, i.e. isoalleles, may be resolved [11] and the discrimination power for mixture deconvolution may be improved. MPS also provides greater flexibility with regard to the multiplexing process since there are fewer limitations concerning amplicon design and color dye channels, as known from CE

* Corresponding author at: LABCON-OWL Analytik, Forschung und Consulting GmbH, Siemensstraße 40, 32105 Bad Salzuflen, Germany.
E-mail addresses: jsilvery@labcon-owl.de, info@labcon-owl.de (J. Silvery).

https://doi.org/10.1016/j.fsigen.2020.102236
Received 16 September 2019; Received in revised form 7 January 2020; Accepted 12 January 2020
Available online 14 January 2020
1872-4973/

approaches. Due to the increasing use of MPS technology in forensic laboratories, there is a growing demand for multiplex kits supporting a straightforward workflow and implementation. Various multiplex PCR kits have been developed, evaluated, and validated. Studies on sensitivity, repeatability, mixture interpretation, and casework samples have highlighted potential applications for MPSSTR genotyping [12–15]. Furthermore, sequencing of STR loci provides new insights into sequence variation of established STR markers [16,17].

In a previous study, the development and systematic optimization of the monSTR forensic identity panel was described [18]. monSTR is designed for the simultaneous amplification of 21 common forensic loci followed by targeted sequencing on the Illumina MiSeq desktop sequencer. The selection of forensic markers covered by the monSTR panel are based on the expanded European Standard Set (ESS) [19], the expanded Combined DNA Index System (CODIS) core locus set by the Federal Bureau of Investigation (FBI) Laboratory [20,21], and also includes the highly polymorphic locus SE33 [22,23]. The following loci are included in the STR multiplex: D1S1656, D2S441, D2S1338, D3S1358, FGA, CSF1PO, SE33, D7S820, D8S1179, D10S1248, TH01, vWA, D12S391, D13S317, Penta E, D16S539, D18S51, D19S433, D21S11, Penta D, and the sex typing marker amelogenin.

Before a new assay can be implemented into a routine laboratory workflow, the performance must be validated [24,25]. Therefore, the aim of the present work was to validate the monSTR panel according to the recommended validation guidelines of the Scientific Working Group for DNA Analysis Methods (SWGDAM) [26]. In order to obtain a better understanding of the panel's characteristics and limitations, six experimental studies were performed including concordance testing, repeatability and reproducibility studies, sensitivity of detection, mixture analysis, species-specificity testing, and analysis of mock samples.

2. Materials and methods

2.1. Source of template DNA

The following DNA samples were used throughout the validation studies: NIST (National Institute of Standards and Technology) Standard Reference Material (SRM) 2391c A, B, and C with known allele lengths and sequences [27,28], human male control DNA 2800 M (Promega, Mannheim, Germany) with known allele lengths, and GEDNAP proficiency samples (German DNA profiling, http://www.gednap.org) 54 & 55 (single- and mixed source samples, respectively). For the species specificity study genomic DNA from three non-primate mammal organisms (dog, cat, and horse), which were collected on buccal swabs, and corresponding DNA samples of their owners were tested. This project was approved by the local ethical committee (University of Ulm, Germany, 267/18).

2.2. Sample preparation and data analysis

2.2.1. DNA extraction and quantification

GEDNAP proficiency samples 54 & 55 were acquired in 2017 and extracted using the High Pure PCR Template Preparation Kit (Roche, Mannheim, Germany) as recommended by the manufacturer. After successful extraction, DNA extracts were stored at −20 °C. Species specificity swab samples were extracted using the KingFisher Flex Purification System (Thermo Fisher, Waltham, MA) with the NucleoMag tissue kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. Buccal swab samples from the animal owners were extracted using the EZ1 Advanced System (Qiagen, Hilden, Germany) according to the manufacturer's recommendation. Recovered genomic DNA was quantified prior to amplification using the Qubit2.0 fluorometer with the Qubit dsDNA HS Assay Kit (Thermo Fisher).

2.2.2. Multiplex PCR setup

Multiplex amplification using the monSTR panel was carried out as a single-tube reaction, containing 12.5 µl Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Frankfurt, Germany), 5 µl monSTR primer mix, a variable amount of specific template DNA (7.8 pg – 10 ng, details in the specific method description), and molecular grade water in a volume of 25 µL. Thermal cycling was performed with a Mastercycler Pro S (Eppendorf, Hamburg, Germany) under the following conditions: 98 °C for 30 s; 25 cycles of 98 °C for 10 s, 58 °C for 23 s, and 72 °C for 65 s; final extension at 72 °C for 120 s with a 4 °C soak.

2.2.3. Library construction and sequencing

After multiplex PCR, amplicons were purified and size-selected using Agencourt AMPure XP paramagnetic beads (Beckman Coulter, Jersey City, NY) with a bead-to-sample ratio of 0.7 according to the manufacturer's protocol [29]. Next, a unique combination of indexed sequencing adapters of the Nextera XT index kit (Illumina, San Diego, CA) was added to the purified amplicons in a limited cycle index PCR (98 °C for 30 s; 8 cycles of 98 °C for 10 s, 62 °C for 30 s; 72 °C for 30 s; 72 °C for 120 s with a 4 °C soak). The libraries were then purified using Agencourt AMPure XP paramagnetic beads (Beckman Coulter) with a bead-to-sample ratio of 0.8. The quality of purified libraries was evaluated by examining the size distribution and the absence of primer species on an Agilent Bioanalyzer 2100 with the High Sensitivity DNA Kit (Agilent Technologies, Palo Alto, CA). To ensure uniform pooling of samples, indexed DNA libraries were subsequently quantified using the NEBNext Library Quant Kit for Illumina (New England Biolabs) according to the manufacturer's protocol, normalized to 2 nM, and pooled at equal volumes. After dilution to a final loading concentration of 10 pM, the libraries were sequenced on the MiSeq platform performing 400 cycles of single-end sequencing. Ninety-two libraries (including 76 libraries for validation purposes and 16 other libraries) were sequenced in the first run using MiSeq Reagent Kit v2 (Illumina). For inter-assay repeatability and reproducibility studies (five MiSeq runs in total), libraries were sequenced using the MiSeq Reagent Nano Kit v2 (Illumina) with 10 libraries per run.

2.2.4. Data processing and analysis

Primary data analysis (base calling, quality scoring, and demultiplexing) was conducted using the MiSeq on-board software Real Time Analysis and MiSeq Reporter. STR genotyping based on raw FASTQ files and result interpretation was performed using *toaSTR* v1.0.0-beta.16 (available at www.toastr.online) [30] applying an analytical threshold of ten reads and a calling threshold of 2%. A calling threshold of 2% indicates that a potential allele must have a minimum coverage of 2% of the total coverage of the respective locus. Otherwise, it would be classified as stutter or artefact, depending on its expected stutter value calculated by the stutter model. Sequence observations with a coverage lower than the analytical threshold (default 10 reads) were not considered for analysis. Prior to data analysis, a representation of the monSTR panel was created in toaSTR to select the markers to be analyzed and to define the respective stutter thresholds. A minimum stutter threshold of 5% was defined to account for fluctuations of low-level stutter. Classification of allelic observations (signal) and non-allelic observations (noise) was performed automatically based on toaSTR's stutter modelling algorithm and was manually reviewed. Statistical calculations and diagrams were generated using Microsoft Excel v16 and R v3.5.3.

2.2.5. Capillary electrophoresis

CE-based STR analysis of sample 2800 M and samples GEDNAP 54 & 55 was accomplished to obtain the size-based STR genotypes using the commercially available STR kits PowerPlex 21 System (Promega), PowerPlex ES SE33 Monoplex System (Promega), and VeriFiler Direct PCR Amplification Kit (Applied Biosystems, Foster City, CA) according

to the manufacturer's protocols. Amplified products were separated on the ABI PRISM 310 Genetic Analyzer (Thermo Fisher) and data was analyzed using GeneMapper ID software v3.2 (Thermo Fisher).

2.3. Concordance

Concordance was investigated using samples NIST SRM 2391c A, B, and C in triplicate with known allele and sequence information to assess the genotype accuracy of the monSTR panel in comparison to the certificate of analysis. The DNA input amount for this study was 1000 pg. For stutter analysis, the stutter ratio was calculated by dividing the N-1 stutter coverage by the coverage of the corresponding genuine allele. Based on the observed stutter ratios in the concordance study, stutter thresholds were defined in toaSTR to enable stutter modelling throughout this work.

2.4. Repeatability and reproducibility

The repeatability study was performed using 1000 pg of sample 2800 M as follows: (a) three independent libraries were prepared and sequenced in one MiSeq run (intra-run variation) and (b) three independent libraries were prepared and sequenced in three separate MiSeq runs (inter-run variation). For the reproducibility study, a second operator conducted separate intra- and inter-run variation experiments with 1000 pg of sample 2800 M. Since the performance of an MPS-STR panel cannot be measured intuitively, we translated our validation results into measurable values (Table 1) that allowed statistical analysis: first, the allele coverage ratio (ACR) was determined for heterozygous STR loci to evaluate the allelic balance under different conditions. Second, the locus balance was described by the coefficient of variation of the coverage (CV). The higher the CV value, the larger the variation in the locus coverages. Third, PCR specificity was assessed by the calling ratio (CR) in the toaSTR software, which expresses the proportion of on-target reads. Fourth, signal-to-noise ratio (SN) was calculated, which is the coverage ratio of allelic observations and noise observations such as stutter products and erroneous sequences. SN should be maximal. Furthermore, the evaluation of genotype accuracy was included. To ensure the comparability among the replicates with varying coverage per sample, replicates were computationally subsampled to the lowest input read coverage value, which was 70,000 reads (x) per sample.

2.5. Sensitivity

To evaluate the sensitivity of detection, eight serial dilutions (1000 pg, 500 pg, 250 pg, 125 pg, 62.5 pg, 31.3 pg, 15.6 pg, and 7.8 pg) of sample 2800 M were prepared and analyzed in triplicate. The statistical analysis included the evaluation of CV, CR, SN, and ACR (Table 1). All replicates were computationally sub-sampled to 42,000x per sample to ensure the comparability of the statistical analysis.

Table 1
Definitions and calculations of statistical parameters used throughout the validation studies. Statistical parameters were calculated by averaging replicates. SD = standard deviation; Cov = coverage; N = number of loci.

Parameter	Aim	Calculation
Allele coverage ratio (ACR)	Intra-locus balance	$ACR = \frac{Cov_{lowest\ coverage\ allele}}{Cov_{highest\ coverage\ allele}}$
Coverage coefficient of variation (CV)	Inter-locus imbalance	$CV = \frac{SD_{Cov}}{Cov}$
Calling ratio (CR)	PCR specificity	$CR = \frac{Called\ reads}{Input\ reads}$
Signal-to-noise ratio (SN)	Tendency to form stutter and artefacts	$SN = \frac{1}{N}\sum_{i=1}^{N} \frac{Cov_{alleles,i}}{Cov_{noise,i}}$
Minor component allele frequency over all included loci (MAF)	Theoretical minor contribution	$MAF = \frac{1}{N}\sum_{i=1}^{N} \frac{Cov_{minor\ alleles,i}}{Cov_{total\ alleles,i}}$

2.6. Mixtures

Two-person mixtures of SRM 2391c A (major component) and SRM 2391c C (minor component) were prepared at various ratios (1:1, 1:4, 1:9, 1:19, and 1:49) and analyzed in triplicate. Each multiplex reaction contained at least 50 pg DNA of the minor component. To achieve this, the total DNA input amount varied from 1000 pg to 2450 pg. For evaluation purposes, a total of 54 unique minor alleles (18 minor alleles in three replicates = 54 alleles per mixture ratio) were considered, which neither overlapped with alleles of the major component nor with major stutter products at N-2, N-1, and N + 1 positions (Supplementary Table 1). Sequence information was used to uniquely identify the minor alleles of each mixture. To compute the theoretical minor contribution, the minor allele frequency (MAF) was calculated across all loci by dividing the minor allele coverage by the total allele coverage (Table 1).

2.7. Mock samples

Artificially created forensic mock samples of GEDNAP 54 & 55 proficiency testing were used, including potential degraded samples and micro traces. The sample set consisted of ten single-source and four mixed-source samples (2–3 person mixtures). A detailed sample description can be found in Supplementary Table 3. For multiplex amplification, 1000 pg of template DNA was used. MPS results were compared to CE genotypes and the official GEDNAP results table. In order to assess the potential result variation between MPS and CE, intra-allelic sequencing variants (i.e. isoalleles) and discordances were counted.

2.8. Species specificity

Non-primate mammal samples (dog, cat, and horse) were amplified with a template DNA input amount of 10 ng and sequenced in triplicate. Furthermore, 1000 pg of genomic DNA of the respective animal owners, 1000 pg 2800 M control DNA (positive control), and a no template control (NTC, molecular grade water, Roche) were analyzed and compared to the obtained animal STR profiles. The statistical analysis included the evaluation of CR for each individual sample (Table 1).

3. Results

The aim of the present work was to validate the assay performance of the monSTR forensic identity panel according to the recommended validation guidelines of the Scientific Working Group for DNA Analysis Methods (SWGDAM). A total of six experiments including concordance testing, repeatability and reproducibility study, sensitivity of detection, mixture deconvolution, analysis of mock samples, and species specificity testing were conducted.

3.1. Sequencing run information

Sequencing runs had a mean cluster density of 753 K/mm2 (± 90 K/mm2) and 93.3% (± 1.5%) of generated clusters passed filters. In total, 6700 MB of data was obtained with a total of 17.5 million reads. 83% (± 6%) of sequenced bases were assigned a Phred quality score of 30 (Q30) or higher. Phasing and pre-phasing values were 0.07% (± 0.02%) and 0.05% (± 0.02%), respectively.

3.2. Stutter analysis

To quantify the N1 stutter ratio of monSTR loci, single-source samples of NIST SRM 2391c A, B, and C with 1000 pg DNA input amount were analyzed in triplicate. Calculated stutter was assessed

Table 2
Ranges of locus-specific N-1 stutter ratios observed with triplicate of NIST SRM 2391c A, B, and C as well as the resulting stutter thresholds set in the toaSTR genotyping software. N-1 stutter ratio was calculated by dividing the coverage of the stutter product in N-1 position by the coverage of the corresponding genuine allele.

STR marker	Range of N-1 stutter [%]	Stutter threshold in toaSTR [%]
D1S1656	3.1–19.1	20
D2S1338	10.9–17.7	18
D2S441	1.5 – 4.3	5
D3S1358	4.3–12.1	13
FGA	11.6–23.3	24
CSF1PO	1.6 – 4.6	5
SE33	6.4–17.3	12
D7S820	1.6–7.2	8
D8S1179	12.0–23.2	24
D10S1248	5.6 – 14.3	15
TH01	2.5–7.3	8
D12S391	4.6–26.9	27
vWA	9.9–14.9	15
D13S317	1.0–3.9	5
Penta E	0.4–6.8	7
D16S539	7.9–16.7	17
D18S51	5.8–16.3	17
D19S433	4.4–8.9	9
D21S11	2.0–10.1	11
Penta D	1.6 – 2.5	5

only if sequence observations were clearly distinguishable as stutter products and not masked by other alleles. Homozygous- and heterozygous genotypes were included in the calculation. The majority of STR markers covered by the monSTR panel produced stutter ratios between 5% and 20% (Table 2). The highest stutter ratio was observed for D12S391, with a maximum value of 26.9%. In contrast, the two pentamer repeat systems Penta E and Penta D had very low stutter ratios. In the toaSTR software, stutter thresholds were set above the highest observed stutter ratio of the respective STR system (minimum 5 %) to enable stutter modelling in the following studies. As an exception, the stutter threshold for SE33 was set to 12% since two out of three observed stutter ratios were lower than 10%.

3.3. Concordance

Samples NIST SRM 2391c A, B, and C were analyzed and subsequently compared to the certificate of analysis. In total, 189 loci (each 21 loci in three samples with each three replicates) with 342 alleles were examined. Results of all length- and sequenced-based genotypes were 100% concordant. Moreover, neither allele dropouts nor drop-ins were detected. Average coverage of sample SRM 2391c A, B, and C was 61,354x (± 36,948x), 55,780x (± 14,611x), and 60,543x (± 9,851x), respectively. The average coverage per locus was 2,922x (± 1,795x), 2,656x (± 740x), and 2,883x (± 514x), respectively.

3.4. Repeatability and reproducibility

The aim of this study was to quantify the genotyping variation within the monSTR panel when two different operators conducted intra- and interrun variation experiments of sample 2800 M. To ensure the comparability among the statistical analysis, all eight replicates were computationally sub-sampled to the lowest input read coverage value.

Within the same run, the average coverage per sample was 60,442x (± 1,431x) with a mean coverage per locus of 2,878x (± 290x) by operator one and 54,984x (± 598x) with a mean coverage per locus of 2,618x (± 259x) by operator two, respectively. Across three separate MiSeq runs by operator one and two, the average coverage per sample was 58,799x (± 3,024x) and 52,229x (± 4,677x) and on average 2,800x (± 589x) and 2,487x (± 574x) per locus were called, respectively. Concordant allele lengths and sequences across all replicates were detected. Low standard deviations were observed across all statistical parameters and operators (Table 3). Generally, inter-run variation was observed to be slightly higher than the variation between samples within one run.

Table 3
Averaged statistical parameters and respective standard deviations for repeatability and reproducibility testing. CV = coverage coefficient of variation; CR = calling ratio; SN = signal-to-noise ratio; ACR = allele coverage ratio; min = minimum observed value; max = maximum observed value.

	Operator A	Operator B	Operator A	Operator B
Parameter	Intra-run variation		Inter-run variation	
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

3.5. Sensitivity

With low DNA quantities, stochastic variations can occur during the amplification process. In order to determine the dynamic range of the monSTR panel where reliable genotypes can be obtained, varying 2800 M DNA input amounts were analyzed in triplicate (1000 pg, 500 pg, 250 pg, 125 pg, 62.5 pg, 31.3 pg, 15.6 pg, and 7.8 pg). To ensure the comparability among the statistical analysis, all replicates were prior to evaluation computationally sub-sampled to the lowest input read coverage value.

Full and concordant profiles were obtained with an input DNA amount of 1000 pg down to 62.5 pg (Table 4). Allelic and locus dropouts were observed with 31.3 pg DNA input and lower. In addition, balanced locus coverages with low corresponding standard deviations of all monSTR targets demonstrated homogeneous amplifications down to 62.5 pg (Fig. 1). Expressed in absolute numbers, the weakest locus over all eight dilutions was Penta D with a mean coverage of 427x (± 75x). The strongest locus D18S51 had an average coverage of 2,416x (± 446x), which results in a 5.6-fold difference in coverage between the least stable and most stable STR locus. The average coverage per locus over all dilutions was 1,054x (± 592x) and the average coverage per sample was 66,126x (± 36,250x). The obtained results for the three statistical parameters coverage coefficient of variation (CV), calling ratio (CR), and signal-to-noise ratio (SN) are shown in Fig. 2. Results suggested a good stability of the interlocus balance down to 62.5 pg. Conversely, DNA input amounts below 62.5 pg had a strong impact on some STR loci concerning interlocus balance. Relative locus coverages varied markedly and corresponding standard deviations increased (Fig. 1). An increased CV of 0.83 ± 0.20 at 15.6 pg underlined that the interlocus balance was substantially deteriorated at lower DNA

Table 4
Number and percentage of called alleles for the amplification of 2800 M control DNA at eight different dilution levels. 2800 M comprises a total of 41 alleles targeted by the monSTR panel. In total, 123 alleles and 63 loci (21 loci with three replicates) were assessed per dilution. N = number of alleles.

DNA input amount [pg]	Called alleles (N = 123)	Called [%]
1000	123	100
500	123	100
250	123	100
125	123	100
62.5	123	100
31.3	122	99.2
15.6	107	85.4
7.8	95	76.4

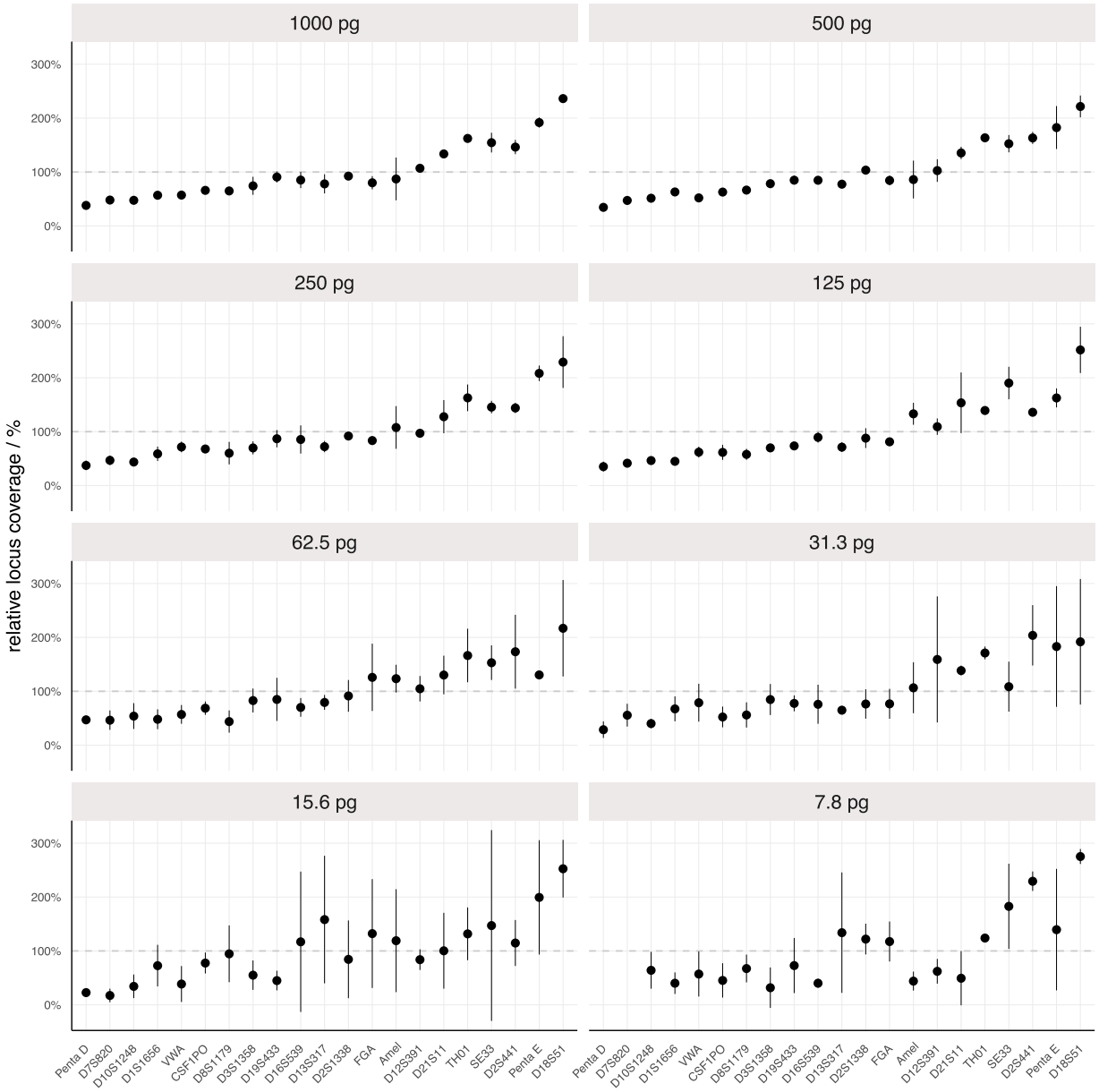


Fig. 1. Relative locus coverage depending on different DNA input amounts of 2800 M. The relative locus coverage of a respective STR locus was calculated by dividing the average locus coverage by the mean coverage of all included loci per dilution. Error bars represent the standard deviation. Dashed lines denote the mean locus coverage over all loci of the corresponding dilution. STR loci are listed in ascending order according to the average locus coverage over all dilutions. A balanced amplification was achieved with a DNA input amount down to 62.5 pg. Locus dropouts in two replicates were observed at 7.8 pg for Penta D and D7S820.

quantities, while CR also continuously decreased (Fig. 2). SN did not fluctuate remarkably across different dilutions, except for 7.8 pg DNA input.

The allele coverage ratio (ACR) was calculated to measure the heterozygous balance. Fig. 3 indicates that with DNA input amounts from 1000 pg down to 125 pg heterozygous loci were well balanced with an ACR greater than 0.6, except for locus D8S1179. Allelic imbalances (ACR < 0.6) without any specific pattern for a certain locus were observed for 62.5 pg of DNA input and lower, while the average ACR for this dilution level was 0.66 ($\pm$ 0.16). With less than 31.3 pg of DNA input, allelic imbalance became more apparent. The average ACR for the 15.6 pg dilution dropped to 0.46 ($\pm$ 0.23) and allelic dropouts were observed more frequently (Table 4). The STR locus with the highest allelic imbalance was D8S1179 with an average ACR of 0.58 ($\pm$ 0.13). The best ACR was observed for the locus SE33 (0.79 $\pm$ 0.19).

3.6. Mixtures

Biological samples containing DNA from more than one individual may be challenging to deconvolve and interpret. In order to test the ability to analyze mixtures, samples NIST SRM 2391c A (major component) and NIST SRM 2391c C (minor component) were combined in five different mixing ratios (1:1, 1:4, 1:9, 1:19, 1:49). In total, 54 unique minor alleles were included in the evaluation, which did not overlap with major contributor alleles or their stutter products (Supplementary Table 1). The objective of this study was to quantify potential deviations between theoretical and observed mixing ratios based on the minor component allele frequency (MAF). Moreover, the recovery of minor alleles especially for strongly imbalanced mixtures was evaluated.

In Fig. 4, the observed MAF was plotted against the theoretical mixing ratios. Generally, the MAF corresponded precisely to the theoretical minor contribution. At the 1:1 ratio (50%), the MAF was

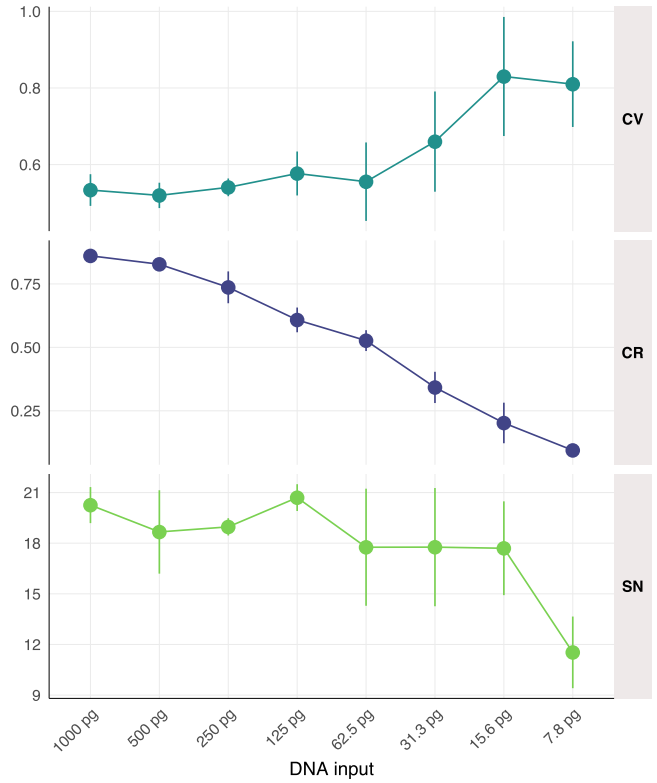


Fig. 2. Mean coverage coefficient of variation (CV), mean calling ratio (CR), and mean signal-tonoise ratio (SN) depending on different 2800 M DNA input amounts. Statistical results suggested a good amplification stability with monSTR down to 62.5 pg. Lower DNA quantities (< 62.5 pg) resulted in higher locus coverage variations, which is demonstrated by an increasing CV. In addition, the proportion of on-target reads (CR) fell below 50%. SN remained stable except for a drop at 7.8 pg.

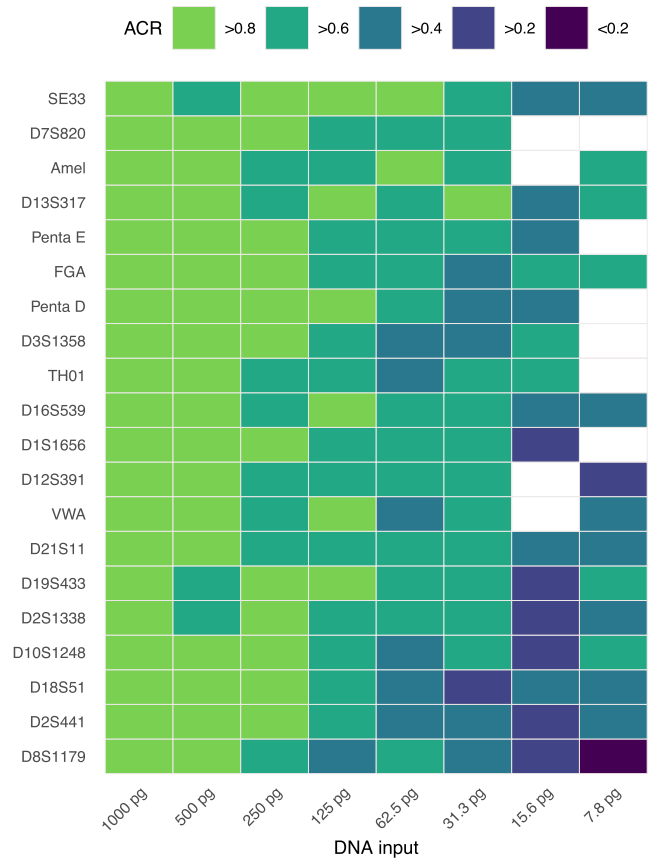


Fig. 3. Heat map of mean allele coverage ratios (ACR) over all heterozygous STR loci and different input amounts of 2800 M control DNA. Each tile represents the mean ACR over three replicates. For illustration purposes, ACR values were mapped to five distinct color groups (> 0.8, > 0.6, > 0.4, > 0.2, and < 0.2). White tiles indicate an allele dropout in two replicates. STR loci are sorted by the mean ACR over all dilution steps from top (balanced) to bottom (imbalanced). A balanced ACR (> 0.6) was achieved for all loci with a DNA input of 125 pg or higher, except for locus D8S1179.

calculated to be 53.7% (± 6.6%). A larger deviation was observed for a theoretical minor contribution of 20% (MAF 15.1% ± 4.1%). Although small variations regarding the observed minor contribution were seen, a precise prediction could be achieved for the challenging 2% mixture ratio (MAF 2.0% ± 0.7%). At the 1:1 ratio, the minor allele coverage reached 2,037x (± 797x), while minor allele coverages decreased with more imbalanced mixture ratios (Supplementary Table 2). All minor alleles were typed precisely with 100% concordance for mixture ratios 1:1 to 1:19. Results for the 1:49 ratio (2%) reflect the stochastic nature of amplifying lower DNA quantities, as the identification of the minor contributor became less accurate. Allelic dropouts were observed at the CSF1PO, D8S1179, and D19S1248 loci.

3.7. Mock samples

For the investigation of artificially created forensic mock samples, genomic DNA of GEDNAP 54 & 55 proficiency samples (Supplementary Table 3) were analyzed with MPS and CE. Genotyping results for all markers included in the monSTR panel were compared to CE results and the official GEDNAP result table, respectively.

Analysis of mock samples yielded largely concordant profiles between CE and MPS (Supplementary Table 3). In total, 723 alleles were differentiated by sequence, while 701 alleles were called by length (Table 5). The MPS technology has the potential to simultaneously identify STR loci by sequence. Thus, 21 additional intra-allelic sequence variants could be detected over CE. One allelic dropout in locus SE33 of sample 54-PA was seen with CE, while MPS yielded a fully concordant genotype compared to the reference. In addition, one allelic dropout in locus D21S11 of sample 55-S3 was seen in both CE and MPS results. Within the examined sample set, one instance of discordance between both technologies was found for sample 54S1 at the SE33 locus. The CE-based genotype displayed a " < 3.2″ allele, however with MPS an allele "14″ was seen. Raw FASTQ files were investigated manually and it was found that a 63 base pair deletion was present within the flanking region that caused the discordance between the two technologies. The most informative tetranucleotide locus SE33 is not (officially) available in published MPS-STR panels at the time of writing and can be considered as a special feature of monSTR. In order to get a better understanding of the performance of this locus, the ACR and locus coverage was evaluated (Table 6). Generally, ACR was observed to be high (> 0.8), except for sample 55-PC with an ACR of 0.69. Data suggested high locus coverages and well-balanced alleles even for sample 54-PC exhibiting a larger allele spread of 5 repeat units.

3.8. Species specificity

Human specificity of the monSTR panel was tested with nonhuman genomic DNA samples of common domestic species (dog, cat, and horse). In comparison to sample 2800 M with an averaged calling ratio (CR) of 86.82% (± 2.19%), the non-human samples yielded very low CR values (Fig. 5). The averaged CR for the cat, dog, and horse samples was 0.37% (± 0.40%), 0.09% (± 0.11%), and 0.08% (± 0.14%), respectively. No alleles above the analytical threshold (10 reads) were detected for each replicate of the NTC. For the dog samples, for instance, alleles were detected in CSF1PO, D7S820, TH01, D19S433, and D21S11, while for the horse samples only D10S1248 alleles were called

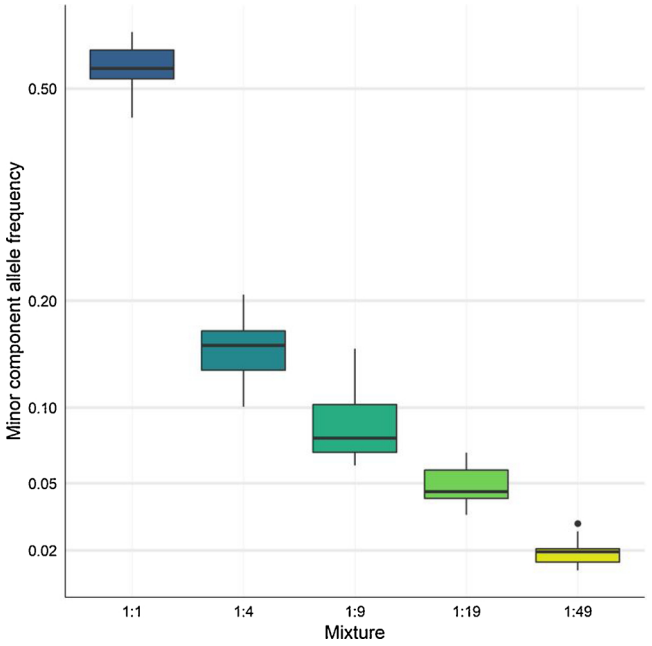


Fig. 4. Minor component allele frequency (MAF) for a total of 54 non-overlapping minor alleles (18 minor alleles in three replicates = 54 alleles per mixture ratio) depending on five different two-person mixtures of SRM 2391c A (major component) and SRM 2391c C (minor component). The MAF was calculated for each locus by dividing the minor allele coverage by the total allele coverage. The y-axis is square root transformed for better legibility of small values.

Table 5
Number of alleles observed for 21 STR loci included in the monSTR panel and corresponding CE-based STR kits for 14 GEDNAP proficiency samples in comparison to the reference GEDNAP results table. The percentage of total observed alleles relative to the number of reference alleles is given in parentheses.

Sample ID	Observed alleles			Isoalleles observed
	Reference	CE	MPS	with MPS
54-PA	42	41[a]	42	
54-PB	36	36	37	1
54-PC	42	42	42	
54-S1	86	86	91[b]	5
54-S2	39	39	39	
54-S3	41	41	41	
54-S4	64	64	68	4
55-PA	40	40	40	
55-PB	39	39	40	1
55-PC	40	40	40	
55-S1	40	40	41	1
55-S2	68	68	71	3
55-S3	88	87[a]	92[a]	5
55-S4	38	38	39	1
Total	703 (100%)	701 (99.7%)	723 (102.8%)	21

[a] In these samples, allelic dropouts were observed.
[b] In this MPS-sample, one allelic discordance at locus SE33 was observed.

occasionally. The cat samples yielded occasionally genotyping results for D1S1656, SE33, D7S820, TH01, vWA, D19S433, D19S433, D21S11, and amelogenin. Some but not all alleles called in the animal samples were identical in length and sequences with the genotypes of the corresponding owner (Supplementary Table 4). For instance, in one of the dog replicates, allele 10 was detected at locus CSF1PO, which was also detected for the respective animal owner. In contrast, allele 9.3 at locus TH01 (replicates one and three) was not observed at the corresponding owner's genotype.

Table 6
Heterozygous genotypes, ACR values, and coverages observed at the SE33 locus for GEDNAP 54 & 55 single source samples. The relative locus coverage of SE33 is given in parentheses and was calculated by dividing the locus coverage by the coverage of all 21 monSTR loci. ACR = allele coverage ratio.

Single-source sample	Heterozygous SE33 genotype	Allele spread	ACR	Locus coverage
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%)

4. Discussion

The aim of the present work was to perform a comprehensive developmental validation of the monSTR forensic identity panel [18], used in combination with the Illumina MiSeq platform and the open-access STR genotyping software toaSTR. Validation studies according to the recommended SWGDAM guidelines assessed (1) concordance, (2) sensitivity, (3) repeatability and reproducibility, (4) the ability to interpret mixtures, (5) mock samples, and (6) species-specificity. We utilized a set of statistical parameters (Table 1) to evaluate the validation results: The allele coverage ratio (ACR) was determined for heterozygous STR loci to estimate the allelic balance under different conditions. The coefficient of variation of the coverage (CV) describes the locus coverage balance; the higher the CV value, the larger the variation in the locus coverages, which can impact sample throughput of MPS systems [31] or may lead to allele or locus dropouts. PCR specificity was assessed by the calling ratio (CR) in the toaSTR software; CR expresses the proportion of on-target reads, which decreases if sequence data contains many reads of primer-dimers or non-specific PCR products. The signal-to-noise ratio (SN) reflects the coverage ratio of allelic observations and noise observations such as stutter products and erroneous sequences.

4.1. Performance of the monSTR forensic identity panel

The selection of forensic markers covered by monSTR is compatible with the expanded European Standard Set (ESS) and also contains markers commonly used for US databasing. Additionally, the highly polymorphic locus SE33 is implemented, which is a special feature of this multiplex panel compared to other MPS-STR kits [13,15]. While the Verogen ForenSeq system also contains primers targeting SE33, this marker is neither promoted by the manufacturer nor does the proprietary genotyping software Verogen UAS show results for SE33 [14,32,33]. Detailed analysis of the SE33 performance using artificially created mock samples indicated a satisfactory amplification with locus coverages comparable to other STR loci. Borsuk et al. [33] demonstrated in a comprehensive population study that heterozygous imbalance of SE33 increased in proportion to increased allele distances. Conversely, our results indicated good heterozygous balance even for genotypes exhibiting larger allele spreads except for one sample. However, these findings should be interpreted with caution, given the fact that mock samples were sequenced as single replicates.

During the amplification process of forensic STR loci, a certain quantity of minor products are commonly observed, which are caused by strand slippage of the DNA polymerase [34–36]. Stutter peaks may impact STR genotyping interpretation as it can be challenging to determine whether a minor product is a genuine allele from a minor contributor or a stutter product of an adjacent allele [37]. While in-depth stutter analysis was not the focus of the present study, N1 stutter

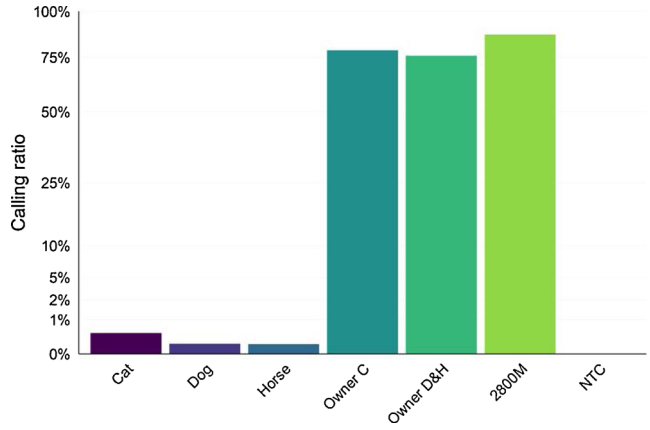


Fig. 5. Mean calling ratio (CR) depending on three different non-human samples (cat/C, dog/D, and horse/H), their corresponding owners, a 2800 M positive control, and a no template control (NTC). The y-axis is square root transformed for better legibility of small values.

ratios of monSTR loci were found to be comparable to those of Hussing et al. [38], who reported N-1 stutter ranges for the ForenSeq DNA Signature Prep Kit (Verogen). Pentamer STR systems such as Penta E showed the lowest stutter range (monSTR: 0.4–6.8%, ForenSeq: 2.0–12.3%), whereas high stutter ratios were seen for D12S317 (monSTR: 4.6–26.9%, ForenSeq: 3.1–24.8%). However, as stutter ratios depend on the genotype, in particular on the length of uninterrupted repeat stretches of alleles [39], the spectrum of stutter ratios observed in the present study might be incomplete due to limited sampling.

Robustness was assessed by the sensitivity study, the interpretation of two-person mixtures, and genotyping of various domestic animals. The multiplex panel was found to have a robust dynamic range using DNA inputs between 1000 pg and 62.5 pg. Within this range, complete and reproducible genotypes with balanced locus coverages were generated in full concordance with CE-STR typing. These findings correspond to Jäger et al. [40] who reported comparable sensitivity results using the ForenSeq DNA Signature Prep Kit in combination with the MiSeq FGx instrument and to the study of Müller et al. [15] who validated MPS kits for the Ion Torrent platform. Analysis of allele coverage ratios indicated well-balanced heterozygous alleles (ACR $>$ 0.6) using DNA inputs of at least 125 pg. DNA quantities below 62.5 pg had a negative impact on the inter-locus balance, as an increased variation in the coverage of certain loci was observed. DNA inputs of 15.6 pg and 7.8 pg caused allele and locus dropouts. These issues may arise from stochastic effects during the PCR process with low DNA input. Additionally, the library preparation including bead-based size separation and multiple pipetting steps could cause coverage variation. MPS approaches have a certain tolerance regarding both high and low performing markers without the drawback of saturated signal output that may be observed with CE methods. However, strong locus imbalances can impact sample throughput and genotyping results and should therefore be avoided. Stronger excess of primers resulting in the formation of primer species may explain the lower amplification specificity (CR) at low DNA input amounts.

We found that the observed minor allele frequency (MAF) was in good agreement with the true mixing ratio. All nonoverlapping minor alleles were detected down to a 5% minor contribution, which is comparable to results of previous studies using the ForenSeq DNA Signature Prep Kit [14,31,40,41] as well as studies on the Ion Torrent platform [42]. It is noteworthy that the minor contribution in a strongly imbalanced 1:49 mixture could be precisely recovered with very low variation across the replicates. Furthermore, results of the species study demonstrates monSTR's specificity for human MPS-STR identification. STR alleles that were occasionally detected in the animal samples (CR $<$ 0.5%) may stem from DNA of the animals' owner or other persons which have been in contact with the animals.

Evaluation of accuracy and precision in allele typing was conducted by analyzing repeatability, reproducibility and concordance between MPS and CE results of reference and forensic mock samples. Results demonstrate that the monSTR forensic identity panel produces reliable genotyping profiles comparable to commercially available MPS-STR kits [12,15,31]. We observed full length- and sequenced-based concordance in terms of the official NIST certificate of analysis [27] as well as a high precision across replicates of 2800 M, which were sequenced in different MiSeq runs and were prepared by the same or by different operators, respectively. The processing of 14 GEDNAP proficiency samples demonstrated the panel's compatibility with artificially created stains as all MPS-STR alleles were in concordance with CE results, except for an allele dropout in sample 55S3 and a discordance at locus SE33 in sample 54-S1. A " $<$ 3.2" allele was observed using the CE technology whereby allele "14" was detected with MPS. A 63 base pair flanking region deletion, which was covered by the CE kit and the monSTR panel, was found as the cause for the discordance. This discrepancy was resolved by the fact that CE-based allele calling considers the full amplicon length including the flanking region, whereas the allele calling using toaSTR is based on the repeat region. One of the major advantages of MPS in analyzing STRs is the ability to reveal intra-allele sequence variants (i.e. isoalleles) that are undetectable by traditional CE. This feature enabled the identification of 21 alleles exhibiting sequence variants. These variants may provide a greater discrimination power in distinguishing individuals and may help to deconvolute mixtures.

While the results of this developmental validation suggest that monSTR is a viable tool, some improvements might be necessary to optimize the workflow for daily laboratory routine. The protocol contains complex pipetting steps that could cause handling errors. Currently, the library preparation in a low throughput scenario requires a hands-on-time of approximately three hours, which is comparable to the commercially available ForenSeq DNA Signature Prep kit [40]. In order to increase sample throughput and reduce hands-on-time, further research will be undertaken to facilitate a fully automated solution. For example, a fully automated library preparation protocol for the ForenSeq DNA Signature Prep Kit was developed and evaluated by Laurent et al. [43]. In addition, complex DNA mixtures with three or more contributors as well as challenging samples from real crime scene scenarios need to be analyzed to get a better understanding of the panel's potential for forensic casework.

4.2. Impact of MPS raw read quality control

Standardization and quality management of MPS have been previously identified as crucial aspects for reliable and reproducible sequencing results [24]. For instance, laboratory guidelines for MPS-based applications in clinical diagnostics [44] and microbial ecology [45] demand quality control (QC) of raw reads as an important preprocessing step. Ignoring low quality reads may add unreliable and random information to the dataset and lead to false interpretation of results. To the best of our knowledge, there is no consensus in forensic genetics on quality indicators, thresholds, and best practices for the treatment of sequence data. Previous MPS-related publications did not report on QC of raw sequences [46–48].

A minimum quality score of Q20 has been proposed in a clinical context [49]. The Q score expresses the probability of an incorrect base call. Alternatively, Edgar and Flyvbjerg [50] suggested the expected number of errors, i.e. the sum of error probabilities in a read, as a better indicator of read quality than the average Q score. Sequencing errors can be reduced by quality trimming (truncating reads with low-quality bases), by quality filtering (discarding reads with low average base quality) [51], or by merging of overlapping paired-end reads and recalculation of quality scores [50]. These treatments are always a trade-off between loss of sequence data (sensitivity) and quality of the

remaining data (specificity) [52]. Tests on the dataset of the developmental validation studies presented here indicated that different quality filtering algorithms and thresholds may have a substantial impact on CV, CR, SN, and ACR (data not shown). Unfavorable filtering settings may skew the obtained STR profile. In view of the lack of forensic QC guidelines, we decided not to perform any quality treatment on the underlying raw data.

We would like to enhance awareness for raw data QC in forensic genetics and encourage a discussion on this topic. It may be discussed whether current genotyping algorithms are capable of obtaining high quality profiles even from low quality sequence data. For example, the toaSTR software maps a read to an STR locus if it approximately matches a pair of unique anchor sequences in the flanking regions. This mechanism may be considered as an inherent quality filtering step. Quality control offered on the autosomal STR online database and QC platform *STRidER* [53] (available at https://strider.online/) includes several *a posteriori* checks such as plausibility tests and statistical analyses on compiled autosomal STR genotype datasets. However, MPS raw data is currently considered only for specific investigations, e.g. to discern novel observations from errors. Experiences with successful standardization of raw data QC in other fields such as clinical laboratory practice [54] may be transferred to the field of forensic genetics, probably guided by a central authority such as STRidER.

Role of funding

This work was supported by a grant from the ZIM funding program (Zentrales Innovationsprogramm Mittelstand) of the German Federal Ministry for Economic Affairs and Energy (grant number ZF4453702).

CRediT authorship contribution statement

Janine Silvery: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Writing - original draft, Writing - review & editing. **Sebastian Ganschow:** Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing - review & editing. **Peter Wiegand:** Funding acquisition, Supervision, Writing - review & editing. **Carsten Tiemann:** Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing.

Declaration of Competing Interest

JS, SG and CT are current employees of LABCON-OWL GmbH. The authors declare no other competing interests.

Acknowledgements

The authors would like to thank Dr. Melanie Grabmüller (Institute of Legal Medicine, Bonn), Dr. Martin Bodner (Institute of Legal Medicine, Innsbruck), Dr. Robert Kulis-Horn (MVZ Labor Krone GbR, Bad Salzuflen) and Natalie Shoopinsky for their excellent advice during the project, as well as Helena Siemens (LABCON-OWL GmbH, Bad Salzuflen) for technical support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fsigen.2020.102236.

References

[1] M.A. Jobling, P. Gill, Encoded evidence: DNA in forensic analysis, Nat. Rev. Genet. 5 (2004) 739–751, https://doi.org/10.1038/nrg1455.
[2] J.M. Butler, Genetics and genomics of core short tandem repeat loci used in human identity testing, J. Forensic Sci. 51 (2006) 253–265, https://doi.org/10.1111/j.1556-4029.2006.00046.x.
[3] A. Urquhart, C.P. Kimpton, T.J. Downes, P. Gill, Variation in short tandem repeat sequences–a survey of twelve microsatellite loci for use as forensic identification markers, Int. J. Legal Med. 107 (1994) 13–20.
[4] Fundamentals of Forensic DNA Typing - 1st Edition, (n.d.). https://www.elsevier.com/books/fundamentals-of-forensic-dna-typing/butler/978-0-12-374999-4 (accessed February 26, 2019).
[5] K.B. Mullis, F.A. Faloona, Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction, Methods Enzymol. 155 (1987) 335–350.
[6] P. Gill, H. Haned, O. Bleka, O. Hansson, G. Dørum, T. Egeland, Genotyping and interpretation of STR-DNA: low-template, mixtures and database matches—twenty years of research and development, Forensic Sci. Int. Genet. 18 (2015) 100–117, https://doi.org/10.1016/j.fsigen.2015.03.014.
[7] L.G. Biesecker, W. Burke, I. Kohane, S.E. Plon, R. Zimmern, Next Generation Sequencing in the Clinic: Are we Ready? Nat. Rev. Genet. 13 (2012) 818–824, https://doi.org/10.1038/nrg3357.
[8] S.C. Schuster, Next-generation sequencing transforms today's biology, Nat. Methods 5 (2008) 16–18, https://doi.org/10.1038/nmeth1156.
[9] A.-M. Divne, H. Edlund, M. Allen, Forensic analysis of autosomal STR markers using Pyrosequencing, Forensic Sci. Int. Genet. 4 (2010) 122–129, https://doi.org/10.1016/j.fsigen.2009.07.004.
[10] C. Børsting, N. Morling, Next generation sequencing and its applications in forensic genetics, Forensic Sci. Int. Genet. 18 (2015) 78–89, https://doi.org/10.1016/j.fsigen.2015.02.002.
[11] K.B. Gettings, K.M. Kiesler, S.A. Faith, E. Montano, C.H. Baker, B.A. Young, R.A. Guerrieri, P.M. Vallone, Sequence variation of 22 autosomal STR loci detected by next generation sequencing, Forensic Sci. Int. Genet. 21 (2016) 15–21, https://doi.org/10.1016/j.fsigen.2015.11.005.
[12] X. Zeng, J. King, S. Hermanson, J. Patel, D.R. Storts, B. Budowle, An evaluation of the PowerSeq™ Auto System: A multiplex short tandem repeat marker kit compatible with massively parallel sequencing, Forensic Sci. Int. Genet. 19 (2015) 172–179, https://doi.org/10.1016/j.fsigen.2015.07.015.
[13] K.J. van der Gaag, R.H. de Leeuw, J. Hoogenboom, J. Patel, D.R. Storts, J.F.J. Laros, P. de Knijff, Massively parallel sequencing of short tandem repeats—Population data and mixture analysis results for the PowerSeq™ system, Forensic Sci. Int. Genet. 24 (2016) 86–96, https://doi.org/10.1016/j.fsigen.2016.05.016.
[14] S. Köcher, P. Müller, B. Berger, M. Bodner, W. Parson, L. Roewer, S. Willuweit, Inter-laboratory validation study of the ForenSeq™ DNA Signature Prep Kit, Forensic Sci. Int. Genet. 36 (2018) 77–85, https://doi.org/10.1016/j.fsigen.2018.05.007.
[15] P. Müller, A. Alonso, P.A. Barrio, B. Berger, M. Bodner, P. Martin, W. Parson, Systematic evaluation of the early access applied biosystems precision ID Globalfiler mixture ID and Globalfiler NGS STR panels for the ion S5 system, Forensic Sci. Int. Genet. 36 (2018) 95–103, https://doi.org/10.1016/j.fsigen.2018.06.016.
[16] K.B. Gettings, L.A. Borsuk, D. Ballard, M. Bodner, B. Budowle, L. Devesse, J. King, W. Parson, C. Phillips, P.M. Vallone, STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci, Forensic Sci. Int. Genet. 31 (2017) 111–117, https://doi.org/10.1016/j.fsigen.2017.08.017.
[17] C. Phillips, K.B. Gettings, J.L. King, D. Ballard, M. Bodner, L. Borsuk, W. Parson, "The devil's in the detail": release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide, Forensic Sci. Int. Genet. 34 (2018) 162–169, https://doi.org/10.1016/j.fsigen.2018.02.017.
[18] S. Ganschow, J. Silvery, C. Tiemann, Development of a multiplex forensic identity panel for massively parallel sequencing and its systematic optimization using design of experiments, Forensic Sci. Int. Genet. 39 (2019) 32–43, https://doi.org/10.1016/j.fsigen.2018.11.023.
[19] P.M. Schneider, Expansion of the European Standard Set of DNA Database Loci–the Current Situation, (2009) 2.
[20] D.R. Hares, Expanding the CODIS core loci in the United States, Forensic Sci. Int. Genet. 6 (2012) e52–e54, https://doi.org/10.1016/j.fsigen.2011.04.012.
[21] D.R. Hares, Selection and implementation of expanded CODIS core loci in the United States, Forensic Sci. Int. Genet. 17 (2015) 33–34, https://doi.org/10.1016/j.fsigen.2015.03.006.
[22] P. Gill, Role of short tandem repeat DNA in forensic casework in the UK—past, present, and future perspectives, BioTechniques. 32 (2002) 366–385, https://doi.org/10.2144/02322rv01.
[23] P. Wiegand, B. Budowle, S. Rand, B. Brinkmann, Forensic validation of the STR systems SE 33 and TC 11, Int. J. Legal Med. 105 (1993) 315–320.
[24] C. Endrullat, J. Glökler, P. Franke, M. Frohme, Standardization and quality management in next-generation sequencing, Appl. Transl. Genom. 10 (2016) 2–9, https://doi.org/10.1016/j.atg.2016.06.001.
[25] F.T. Peters, O.H. Drummer, F. Musshoff, Validation of new methods, Forensic Sci. Int. 165 (2007) 216–224, https://doi.org/10.1016/j.forsciint.2006.05.021.
[26] Scientific Working Group on DNA Analysis Methods, Validation Guidelines for DNA Analysis Methods, (2016).
[27] National Institute of Standards & Technology, Certificate of Analysis Standard Reference Material 2391c, (2018).
[28] C.R. Steffen, M.C. Kline, D.L. Duewer, P.M. Vallone, Establishing traceability to NIST SRM 2391c: PCR-Based DNA profiling standard, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e112–e113, https://doi.org/10.1016/j.fsigss.2015.09.045.
[29] Beckman Coulter, SPRIselect User Guide, PN B24965AA, (2012).
[30] S. Ganschow, J. Silvery, J. Kalinowski, C. Tiemann, toaSTR: a web application for forensic STR genotyping by massively parallel sequencing, Forensic Sci. Int. Genet. 37 (2018) 21–28, https://doi.org/10.1016/j.fsigen.2018.07.006.
[31] J.D. Churchill, S.E. Schmedes, J.L. King, B. Budowle, Evaluation of the Illumina ®

Beta Version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling, Forensic Sci. Int. Genet. 20 (2016) 20–29, https://doi.org/10.1016/j.fsigen.2015.09.009.
[32] N.M.M. Novroski, J.L. King, J.D. Churchill, L.H. Seah, B. Budowle, Characterization of genetic sequence variation of 58 STR loci in four major population groups, Forensic Sci. Int. Genet. 25 (2016) 214–226, https://doi.org/10.1016/j.fsigen.2016.09.007.
[33] L.A. Borsuk, K.B. Gettings, C.R. Steffen, K.M. Kiesler, P.M. Vallone, Sequence-based US population data for the SE33 locus, Electrophoresis. 39 (2018) 2694–2701, https://doi.org/10.1002/elps.201800091.
[34] E. Viguera, D. Canceill, S.D. Ehrlich, In vitro replication slippage by DNA polymerases from thermophilic organisms, J. Mol. Biol. 312 (2001) 323–333, https://doi.org/10.1006/jmbi.2001.4943.
[35] M. Klintschar, P. Wiegand, Polymerase slippage in relation to the uniformity of tetrameric repeat stretches, Forensic Sci. Int. 135 (2003) 163–166.
[36] C. Schlötterer, D. Tautz, Slippage synthesis of simple sequence DNA, Nucl Acids Res. 20 (1992) 211–215, https://doi.org/10.1093/nar/20.2.211.
[37] B. Budowle, A.J. Onorato, T.F. Callaghan, A. Della Manna, A.M. Gross, R.A. Guerrieri, J.C. Luttman, D.L. McClure, Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework, J. Forensic Sci. 54 (2009) 810–821, https://doi.org/10.1111/j.1556-4029.2009.01046.x.
[38] C. Hussing, C. Huber, R. Bytyci, H.S. Mogensen, N. Morling, C. Børsting, Sequencing of 231 forensic genetic markers using the MiSeq FGx™ forensic genomics system – an evaluation of the assay and software, Forensic Sci. Res. 3 (2018) 111–123, https://doi.org/10.1080/20961790.2018.1446672.
[39] C. Brookes, J.-A. Bright, S. Harbison, J. Buckleton, Characterising stutter in forensic STR multiplexes, Forensic Sci. Int. Genet. 6 (2012) 58–63, https://doi.org/10.1016/j.fsigen.2011.02.001.
[40] A.C. Jäger, M.L. Alvarez, C.P. Davis, E. Guzmán, Y. Han, L. Way, P. Walichiewicz, D. Silva, N. Pham, G. Caves, J. Bruand, F. Schlesinger, S.J.K. Pond, J. Varlaro, K.M. Stephens, C.L. Holt, Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories, Forensic Sci. Int. Genet. 28 (2017) 52–70, https://doi.org/10.1016/j.fsigen.2017.01.011.
[41] F. Guo, J. Yu, L. Zhang, J. Li, Massively parallel sequencing of forensic STRs and SNPs using the Illumina ® ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™ Forensic Genomics System, Forensic Sci. Int. Genet. 31 (2017) 135–148, https://doi.org/10.1016/j.fsigen.2017.09.003.
[42] F. Guo, Y. Zhou, F. Liu, J. Yu, H. Song, H. Shen, B. Zhao, F. Jia, G. Hou, X. Jiang, Evaluation of the Early Access STR Kit v1 on the Ion Torrent PGM™ platform, Forensic Sci. Int. Genet. 23 (2016) 111–120, https://doi.org/10.1016/j.fsigen.2016.04.004.
[43] F.X. Laurent, L. Ausset, M. Clot, S. Jullien, Y. Chantrel, C. Hollard, L. Pene, Automation of library preparation using Illumina ForenSeq kit for routine sequencing of casework samples, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e415–e417, https://doi.org/10.1016/j.fsigss.2017.09.156.
[44] N. Aziz, Q. Zhao, L. Bry, D.K. Driscoll, B. Funke, J.S. Gibson, W.W. Grody, M.R. Hegde, G.A. Hoeltge, D.G.B. Leonard, J.D. Merker, R. Nagarajan, L.A. Palicki, R.S. Robetorye, I. Schrijver, K.E. Weck, K.V. Voelkerding, College of american pathologists' laboratory standards for next-generation sequencing clinical tests, Arch. Pathol. Lab. Med. 139 (2015) 481–493, https://doi.org/10.5858/arpa.2014-0250-CP.
[45] N.A. Bokulich, S. Subramanian, J.J. Faith, D. Gevers, J.I. Gordon, R. Knight, D.A. Mills, J.G. Caporaso, Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing, Nat. Methods 10 (2013) 57–59, https://doi.org/10.1038/nmeth.2276.
[46] K.B. Gettings, L.A. Borsuk, C.R. Steffen, K.M. Kiesler, P.M. Vallone, Sequence-based U.S. Population data for 27 autosomal STR loci, Forensic Sci. Int. Genet. 37 (2018) 106–115, https://doi.org/10.1016/j.fsigen.2018.07.013.
[47] J. Hoogenboom, K.J. van der Gaag, R.H. de Leeuw, T. Sijen, P. de Knijff, J.F.J. Laros, 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 Sci. Int. Genet. 27 (2017) 27–40, https://doi.org/10.1016/j.fsigen.2016.11.007.
[48] X. Zeng, J.L. King, M. Stoljarova, D.H. Warshauer, B.L. LaRue, A. Sajantila, J. Patel, D.R. Storts, B. Budowle, High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing, Forensic Sci. Int. Genet. 16 (2015) 38–47, https://doi.org/10.1016/j.fsigen.2014.11.022.
[49] New York Department of Health, "Next Generation" Sequencing (NGS) Guidelines for Somatic Genetic Variant Detection, (2015).
[50] R.C. Edgar, H. Flyvbjerg, Error filtering, pair assembly and error correction for next-generation sequencing reads, Bioinformatics. 31 (2015) 3476–3482, https://doi.org/10.1093/bioinformatics/btv401.
[51] Sunyoung Kwon, Seunghyun Park, Byunghan Lee, Sungroh Yoon, In-depth analysis of interrelation between quality scores and real errors in illumina reads, 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, Osaka, 2013, pp. 635–638, , https://doi.org/10.1109/EMBC.2013.6609580.
[52] C. Del Fabbro, S. Scalabrin, M. Morgante, F.M. Giorgi, An extensive evaluation of read trimming effects on illumina NGS data analysis, PLoS One 8 (2013) e85024, , https://doi.org/10.1371/journal.pone.0085024.
[53] M. Bodner, I. Bastisch, J.M. Butler, R. Fimmers, P. Gill, L. Gusmão, N. Morling, C. Phillips, M. Prinz, P.M. Schneider, W. Parson, Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER), Forensic Sci. Int. Genet. 24 (2016) 97–102, https://doi.org/10.1016/j.fsigen.2016.06.008.
[54] A.S. Gargis, L. Kalman, M.W. Berry, D.P. Bick, D.P. Dimmock, T. Hambuch, F. Lu, E. Lyon, K.V. Voelkerding, B.A. Zehnbauer, R. Agarwala, S.F. Bennett, B. Chen, E.L.H. Chin, J.G. Compton, S. Das, D.H. Farkas, M.J. Ferber, B.H. Funke, M.R. Furtado, L.M. Ganova-Raeva, U. Geigenmüller, S.J. Gunselman, M.R. Hegde, P.L.F. Johnson, A. Kasarskis, S. Kulkarni, T. Lenk, C.S.J. Liu, M. Manion, T.A. Manolio, E.R. Mardis, J.D. Merker, M.S. Rajeevan, M.G. Reese, H.L. Rehm, B.B. Simen, J.M. Yeakley, J.M. Zook, I.M. Lubin, Assuring the quality of next-generation sequencing in clinical laboratory practice, Nat. Biotechnol. 30 (2012) 1033–1036, https://doi.org/10.1038/nbt.2403.