Int J Legal Med
DOI 10.1007/s00414-009-0408-x

ORIGINAL ARTICLE

Increasing the discrimination power of forensic STR testing by employing high-performance mass spectrometry, as illustrated in indigenous South African and Central Asian populations

Florian Pitterl · Konrad Schmidt · Gabriela Huber · Bettina Zimmermann · Rhena Delport · Sylvain Amory · Bertrand Ludes · Herbert Oberacher · Walther Parson

Received: 21 August 2009 / Accepted: 17 December 2009

Abstract Short tandem repeat (STR) typing has become the standard technique in forensic methodology for the identification of unknown samples. National DNA databases have been established that contain STR genotypes for intelligence purposes. Due to their success, national DNA databases have been growing so fast that the number of advantageous matches may become a logistic problem for the analysts. This is especially true for partial STR profiles as they display reduced discrimination power. To overcome this drawback, modified versions (so-called mini-STRs) of existing loci were introduced as well as new loci to improve the information content of (partial) STR profiles. We pursue an alternative approach that makes use of nucleotide variation within the amplified STR fragments, which can be discerned by mass spectrometry. We have developed an assay that determines molecular masses from crude STR amplicons which were purified and separated by a liquid chromatographic system directly hyphenated to an electrospray ionization mass spectrometer. We present here new population data of forensically relevant STRs in Khoisan and Yakut populations. These autochthonous groups were selected as they may harbor additional STR alleles that are rare or unobserved in modern humans from cosmopolitan areas, especially for the Khoisan, which are known to represent a very ancient human population. The analysis of the molecular mass of STRs offered a widened spectrum of allele variability escorted by enhanced forensic use. Thus, established STR data derived from fragment size analysis can still be used in casework or in the context of intelligence databasing.

Keywords Short tandem repeats · Khoisan · Yakut · Mass spectrometry · Liquid chromatography · Forensic science

Electronic supplementary material The online version of this article (doi:10.1007/s00414-009-0408-x) contains supplementary material, which is available to authorized users.

F. Pitterl · G. Huber · B. Zimmermann · S. Amory · H. Oberacher (✉) · W. Parson (✉)
Institute of Legal Medicine, Innsbruck Medical University,
Müllerstrasse 44,
6020 Innsbruck, Austria
e-mail: herbert.oberacher@i-med.ac.at
e-mail: walther.parson@i-med.ac.at

K. Schmidt
Department of Medical Genetics, Innsbruck Medical University,
6020 Innsbruck, Austria

R. Delport
Department of Family Medicine, School of Medicine,
University of Pretoria,
Pretoria 0002, South Africa

S. Amory · B. Ludes
Laboratory of Molecular Anthropology,
Institute of Legal Medicine, Strasbourg University,
67000 Strasbourg, France

S. Amory
The International Commission on Missing Persons,
71000 Sarajevo, Bosnia–Herzegovina

Introduction

In the early 1990s, DNA-based identification of biological samples using multiplex polymerase chain reaction (PCR) of short tandem repeats (STR) has revolutionized

Published online: 16 January 2010

the field of forensic genetics by thitherto unobserved assay sensitivity and specificity [1]. The application of STRs has soon spread worldwide, and today, a set of less than 30 loci is used daily in forensic laboratories in crime investigations. Scientific groups have been working toward standardization of STR loci and allele calling [2] to harmonize STR typing across borders. In the European context, the EU Council resolution 9192/01 meanwhile requires laboratories to use at least the European Standard Set (ESS) of loci for upload of STR profiles onto national DNA databases. The ESS panel includes the seven STR core loci D3S1358, D8S1179, D18S51, D21S11, FGA, TH01, and VWA, which have also been adopted by the Interpol DNA Monitoring Expert group (www.interpol.int/public/Forensic/dna/dnameg.asp) with the addition of the gender-specific locus amelogenin (International Standard Set of Loci). Recently, the European Network of Forensic Science Institutes (www.enfsi.org) has extended this set to include five additional STR loci [3, 4]. Laboratories have, however, been using commercial products that include additional, yet not requested STR loci such as D2S1338 and D19S433 (SGMplus and Identifiler, Applied Biosystems (AB), Foster City, CA, USA) or Penta D and Penta E (Powerplex 16, Promega, Madison, WI, USA). Also, SE33 has been added to commercial kits (e.g., SE-Filer, AB and Powerplex ES, Promega), as this marker is defined as core locus for the German National DNA Database. Thus, these additional loci are also stored in the DNA databases to contribute to the genetic information that is used for intelligence purposes in the forensic field.

The aforementioned developments apparently impacted research strategies in forensic genetics. Intelligence databases do have a retarding effect on the implementation of new genetic loci that would need to be brought in concert with existing data and databases. Among other reasons, this also prevented autosomal SNPs from being implemented in forensic practice so far [5]. An apparent retyping limitation is the lack of retained DNA extracts of database samples and the legal requirement to destroy DNA after typing in some countries. On the other hand, the forensic community has observed the need for increased discrimination power in the use of database genotypes, especially for partial DNA profiles and DNA mixtures in cases where multiple DNA databases are queried and compared, as is current practice in the so-called Prüm countries.

In this study, we propose to increase the discrimination power of forensic STR typing by changing the detection platform rather than implementing additional markers. As previously demonstrated [6], mass spectrometry offers the advantage of an increased discrimination power for STR typing compared to conventional capillary electrophoresis (CE). While the latter discriminates allele categories only on the basis of discernible allele sizes, identical fragment lengths can be differentiated with mass spectrometry when the underlying sequences differ. We have also shown that forensically relevant STR loci can be detected in multiplex format by ion-pair reversed-phase liquid chromatography–electrospray ionization time-of-flight mass spectrometry (ICEMS) to comply with the requirements of sensitive and informative forensic stain analysis [7]. The detection of nucleotide variability within STRs has the potential to aid in challenging forensic questions including missing persons and kinship analysis.

This study presents population data on the current ESS loci and the aforementioned additional STRs D2S1338, D16S539, D19S433, and SE33 in indigenous African and Asian populations, which we selected for their autochthonous status that would possibly show yet unobserved alleles due to their ancient genetic background compared to modern urban populations that are usually tested in forensic genetics. Increased forensic performance as retrieved through mass spectrometry and sequencing of alleles is demonstrated by descriptive analysis of the alleles.

Materials and methods

Samples and DNA extraction

Samples were received from unrelated individuals with informed consent. DNA was extracted from buccal swabs of 94 individuals from Central Yakutia using the phenol–chloroform method followed by purification with Cleanmix kit (Talent, Trieste, Italy). Details are described in [8]. Further, DNA was extracted from blood samples of 108 individuals from five Khoisan tribes (Ikongau, Kung, Kwe, Mahenge, Wiku; Republic of South Africa) as described in [9].

Chemicals and oligonucleotides

Acetonitrile (high-performance liquid chromatography (HPLC) gradient grade) and cyclohexyldimethylamine were obtained from Sigma-Aldrich (St. Louis, MO, USA). A stock solution (1.0 M) of cyclohexyldimethylammonium acetate (CycHDMAA) was prepared by titration of a 1.0 M solution of cyclohexyldimethylamine (Fluka, Buchs, Switzerland) with acetic acid (Fluka) at 5°C until pH 8.3 was reached. For preparation of all solutions, HPLC-grade water (Merck) was used. The sequences of the PCR primers were taken from [6]. All oligonucleotides were obtained from Sigma-Aldrich and quality-checked by ICEMS [10] before use in PCR.

Amplification for ICEMS STR analysis

The loci D3S1358, D21S11, D8S1179, vWA, and D16S539 were amplified within a recently published multiplexed PCR assay [7]. Singleplex amplification of TH01, FGA, D18S51, D2S1338, D19S433, and SE33 was performed as described in [6].

Experimental setup for ICEMS STR analysis

An Ultimate fully integrated capillary HPLC system (LC-Packings, Amsterdam, The Netherlands) in combination with a Famos micro-autosampler (LC-Packings) equipped with a 1 μl loop was used for all chromatographic experiments. The 50×0.2 mm id monolithic capillary column was prepared according to the published protocol [11]. The flow rate was set to 2.0 μl/min and the crude PCR mixtures were directly injected onto the column, which was flushed with 25 mM CycHDMAA (pH 8.3) containing 5.0% acetonitrile. A column temperature of 68°C was used to denature the amplicons into the corresponding single strands, which were separated using a gradient of 20.0% to 40.0% acetonitrile in 25 mM CycHDMAA (pH 8.3) within 5 min (for analysis of singleplex PCR products) or 10 min (for analysis of multiplexed PCR products), respectively. The gradient was started 5 min after the sample injection event. Eluting nucleic acids were detected online by negative electrospray ionization-mass spectrometry which was performed on a QSTAR XL mass spectrometer (AB) equipped with a modified TurboIonSpray source [10, 12]. Mass calibration and optimization of instrumental parameters was performed as described elsewhere [10, 12]. The spray voltage was set to 4.0 kV. Gas flows of 15 arbitrary units (nebulizer gas) and 45 arbitrary units (turbo gas) were employed. The temperature of the turbo gas was adjusted to 300°C. The accumulation time was set to 1 s and ten times bins were summed up. Mass spectra were recorded in the range between 800 and 1,200 on a personal computer operating with the Analyst QS software (service pack 8, AB). Deconvolution of raw mass spectra was performed with Bayesian Protein Reconstruct (BioAnalyst 1.1.1, AB).

Conventional STR analysis by CE and confirmative sequencing

Electrophoretic STR typing was carried out using the PCR amplification kits Identifiler, SEfiler, and Profiler Plus (all AB). PCR products were analyzed on an AB Prism 3100 Genetic Analyzer, and the generated data were analyzed as described in [13]. Sanger sequencing of a representative number of alleles was performed as described in [6].

Results and discussion

Mass spectrometric STR analysis

The current study reports STR data from indigenous African and Asian populations that are put in context with previously published data from a European sample set [6]. We applied ICEMS as genotyping platform. As reported earlier, ICEMS enables for highly efficient characterization of nucleic acid molecules [14–16]. Applying ICEMS for STR analysis, identical fragment sizes become discernible when the underlying nucleotide sequences of the alleles differ (SNPSTRs). This offers an increased number of discernible allele categories within an STR locus and thus provides more discrimination power in comparison to conventional fragment size determination by electrophoresis. The designation of newly observed STR alleles can be adapted to the existing nomenclature guidelines of STRs [2], and therefore, the data can be compared to previously generated data or stored in existing DNA intelligence databases.

All STR loci described herein were amplified in “mini-amplicon” format displaying reduced fragment sizes. This is known to result in an increased success rate when degraded DNA is investigated [17] and therefore highly appreciated in forensic analysis.

We selected 11 STR loci for our comparative population study; ten are included in the SGMplus kit (AB; D3S1358, D8S1179, D18S51, vWA, D21S11, FGA, D2S1338, D16S539, TH01, and D19S433) and therefore widely used in the forensic community and SE33 [18] which constitutes the most discriminating single STR locus and is used in Germany and neighbor countries for forensic purposes.

Observed alleles and nucleotide variability

Within the analyzed population samples, ICEMS STR typing detected nucleotide variations within all 11 investigated STR loci (D3S1358, D8S1179, D18S51, vWA, D21S11, FGA, D2S1338, D16S539, TH01, D19S433, and SE33). As illustrated in Online Resource 1, the frequency of nucleotide variants and the number of different kinds of variations observed varied between the different loci and the different population samples. Hence, the gain in allelic resolution in STRs accessible via mass spectrometric analysis is depending on the locus itself and also on the phylogenetic background of the population sample investigated. Within the herein reported population samples, D3S1358, D8S1179, vWA, D21S11, D2S1338, and SE33 were the most polymorphic loci regarding the number of resolvable STR allele categories due to the detection of nucleotide variability. This result is in concordance (for the overlapping loci D21S11, D3S1358, D8S1179, and vWA)

with recently published data obtained via mass spectrometric STR analysis of a Caucasian and an African American population sample [19].

In the following, the different types of nucleotide variants observed via ICEMS typing are discussed. The results are summarized in Table 1 and in Online Resource 1 and 2. Nucleotide variations were divided into three groups regarding the frequency of occurrence. "Common" variants occurred repeatedly within one locus and within different allele length categories. In D3S1358, for example, variants containing one or two (C > T) base exchanges were observed within alleles comprising of 14 to 18 repeat units (Online Resource 1). All "common" variations observed within the 11 loci are listed for each population sample individually in Table 1. If a kind of nucleotide variation occurred several times within one locus but exclusively within one certain allelic length category within the investigated population sample, the variation was denoted as "reduced distributed" variation. "Reduced distributed" variations were observed for the loci D18S51, vWA, FGA, D2S1338, D16S539, and D19S433. For example, in the locus D18S51, the variation (A > C) was only observed within Asian alleles comprising 14 repeat units (see also Online Resource 1(c)). All observed "reduced distributed variations" are listed in Table 1 in italics. All nucleotide variations which were observed infrequently (only once or twice in a population sample) were called "rare" variations. With exception of D19S433, D16S539, and D18S51, "rare" variations were detected in all loci. All "rare" variants observed within this study are listed in Online Resource 2. In Online Resource 1, all observed types of variations are displayed including the respective allele frequencies.

In order to confirm the results obtained by ICEMS and to elucidate the origin of the variability, a representative number of alleles (including all three abovementioned groups of variations) were analyzed by Sanger sequencing. All obtained sequences (N=1,042) are displayed in Online Resource 3. Sequence analysis revealed that alterations in the base composition of repeat building blocks and/or single base exchanges (in some cases SNPs) present within the region flanking the repeat units were the sources of nucleotide variability. For all "common" and "reduced distributed" variations observed in the present study, the sequencing results are presented in simplified form in Table 1.

Results of the concordance study

To verify the ICEMS results, all samples were genotyped with conventional CE. Although the two approaches used different primer designs in the majority of cases, allele calls based on the fragment size were confirmed. For the locus D19S433, the seeming difference between the ICEMS genotypes and the data generated by CE (CE alleles were generally two repeat units larger than the ICEMS genotypes) can be explained by the number of "AAGG" blocks that were included as repeat units in the Identifiler kit (AB) and in the ICEMS assay (Online Resource 3).

Three of the 11 investigated loci (D16S539, D19S433, and SE33) brought typing discrepancies (based on fragment size) represented by a number of 20 samples.

Comparison of the D16S539 genotypes obtained by ICEMS with the respective electrophoretically generated data brought disagreement in three Asian samples. Sequence analysis showed that in each case, a mismatch within the ICEMS forward primer binding region was the assumable cause for the discrepant typing results due to an amplification failure within the ICEMS assay (asterisked in Online Resource 1(h); sequence data are shown in Online Resource 4). Interestingly, this (C > G) base exchange causing the mismatch occurred exclusively in alleles comprising eight repeat units in the investigated Asian sample set. For the locus D19S433, the Identifiler kit (AB) failed to type allele D19S433*6 (13 alleles observed within the African sample set (108 samples), asterisked in Online Resource 1(j)). Sequence analysis revealed a (G > T) base exchange located inside the fragment amplified in the ICEMS assay (Online Resource 4). This (G > T) base exchange is the putative cause for the dropout of D19S433*6 within the Identifiler kit (AB, primer sequences for D19S433 not published). At STRBase (http://www.cstl.nist.gov/div831/strbase/), D19S433*6 (corresponds to D19S433*8 at STRBase) is reported as "variant allele" with a frequency of 1 in 5,367. Null alleles observed for D19S433 applying the SGMPlus kit (AB) were reported by Clayton et al. [20], but no information about the position of primer binding site mutations was given. In the locus SE33, four samples with African origin gave discordant typing results between the ICEMS assay and electrophoretic sizing. Sequence analysis of three of the four concerned alleles (for one sample, electrophoretic separation of the two alleles prior to sequence analysis was impossible due to a fragment size difference of only 1 bp) showed a four bases deletion located outside the ICEMS fragment (Online Resource 3). This deletion was putatively located inside the amplification product of the electrophoretic assay (SEfiler, AB; primer sequences not published) and therefore the origin of the reduced fragment length and the resulting discrepancy (Online Resource 3 and 4).

Performance of conventional STR analysis, MS analysis, and direct sequencing

In an additional set of experiments, a representative number of alleles were characterized by Sanger sequencing for each locus (sequencing results shown in Online Resource 3). In all cases, the nucleotide composition of the amplicon

Table 1 Summary of the most frequent nucleotide variations observed within 11 STR systems in a European, an Asian, and an African population sample

Locus	Population sample	Nucleotide variability within STR alleles identified by ICEMS	Source of nucleotide variability elucidated by sequence analysis
D3S1358 (Online Resource 1a)	Europe	n×(C > T) (n=1;2)	Variable ratio of **TAGA** to **CAGA**
	Asia	n×(C > T) (n=1;2)	See European sample set
	Africa	n×(C > T) (n=1;2)	See European sample set
D8S1179 (Online Resource 1b)	Europe	n×(A > G) (n=1;2)	Substitution(s) of **TCTA** by **TCTG**
	Asia	(A > G)	substitution of **TCTA** by **TCTG**
	Africa	n×(A > G) (n=1;2)	See European sample set
D18S51 (Online Resource 1c)	Europe	–	–
	Asia	*(A > C) (l= 14)*	Substitution of **AGAA** by **AGCA**
	Africa	(A > G)	Substitution of **AGAA** by **GGAA** or (A > G) base exchange in flanking region
vWA (Online Resource 1d)	Europe	(G > A)	Variable ratio of **TCTA** to **TCTG**
		(A > G, T > C, T > C) (l= 14)	Variable ratio of **TCTA** to **TCTG**, substitution of **TCTA** by **TCCA**, and (T > C) base exchange in flanking region
	Asia	See European sample set	See European sample set
	Africa	(G > A); n×(A > G) (n=1;2) *(A > G, T > C, T > C) (l= 14)*	See European sample set
D21S11 (Online Resource 1e)	Europe	(G > A)	Variable ratio of **TCTA** to **TCTG**
	Asia	n×(G > A) (n=1;2)	See European sample set
	Africa	(G > A); n×(A > G) (n=1;2)	See European sample set
FGA (Online Resource 1f)	Europe	–	–
	Asia	–	–
	Africa	(T > C)	Substitution of **TTCC** by **CTCC** or substitution of **CTTT** by **CCTT**
		(5×(T > C), 4×(T > G) (l= 42.2)	Substitutions of **CTTT** by **CTGT** and **CTTC** and excess of **TTCC**
D2S1338 (Online Resource 1g)	Europe	n×(T > G) (n=1–3)	Variable ratio of **TTCC** to **TGCC** and/or presence of one **GTCC**
	Asia	n×(T > G) (n=1–3)	See European Sample set
		(T > G,C > T) (l= 20)	Variable ratio of **TTCC** to **TGCC** and presence of one **TTTC**
	Africa	n×(T > G) (n=1–3); n×(G > T) (n=1–3)	See European sample set
D16S539 (Online Resource 1h)	Europe	(A > C)	SNP rs13265375
	Asia	(A > C)	SNP rs13265375
		(C > G) (l= 8)	(C > G) base exchange in flanking region (ICEMS primer binding region)
	Africa	(A > C)	SNP rs13265375
D19S433 (Online Resource 1j)	Europe	–	–
	Asia	–	–
	Africa	*(G > T) (l= 6)*	(G > T) base exchange in the flanking region
		(T > C) (l= 13)	(T > C) base exchange in the flanking region
SE33 (Online Resource 1k)	Europe	(G > A)	SNP rs9362477
		(A > G)	Substitution of **AAAG** by **GAAG**
	Asia	(G > A)	SNP rs9362477
	Africa	(G > A) (A > G)	See European sample set

All "common" variations identified via ICEMS analysis are listed (description in the text). "Reduced distributed" variations (observed only in alleles comprising of "l" repeat units) are in italics. A simplified description of the origin of variability determined via sequence analysis is added for each variant (repeat building blocks are in bold). Sequencing results can be found in Online Resource 3

determined with ICEMS was confirmed by sequence analysis. For a few samples, sequence analysis offered a better allelic discrimination than ICEMS. This observation can be explained by the inability of ICEMS to differentiate nucleic acid molecules that exhibit identical nucleotide compositions but varying nucleotide sequences. In the section below, this topic is discussed in detail.

To get a general estimation of the amount of genetic information accessible from STR loci via the different genotyping methods for the 11 herein discussed STRs, the allelic resolutions obtained by capillary gel electrophoresis, mass spectrometry, and Sanger sequencing were compared. In this comparison, 1,042 sequenced alleles with European, Asian, or African origin were included and combined to "one" sample. The number of discernible allele categories was used as a measure for the efficiency of the applied typing technology. It is important to note here that only the number of differentiable allele categories was considered and that the corresponding allele frequencies were disregarded in this qualitative comparison of the efficiencies of different STR typing strategies. The results of Sanger sequencing, which offers by definition the ultimate allelic resolution formed the basis for the comparison (sequence analysis corresponds to 100% of differentiable allele categories). In relation, the number of CE- and ICEMS-derived allele categories was expressed as percent values (Fig. 1). Fragment length determination by electrophoresis could only decipher portions of the genetic information present within the 11 STR loci. Through the simultaneous characterization of nucleotide and length polymorphisms via ICEMS analysis, the amount of accessible genetic information present within STRs was significantly increased for all 11 loci. In five STR systems (D3S1358, D18S1179, FGA, D16S539, and TH01), even the maximum number of discernible allele categories was reached. The remaining six loci harbored genetic information which was only accessible by sequence analysis. Nevertheless, the fraction of discernible allele categories by ICEMS analysis amounted to 80% or more compared to the sequence-derived categories (with exception of SE33).

The situation changes if the results obtained via sequence analysis are esteemed for each of the three population samples separately. In the European population sample five loci (SE33, D2S1338, D21S11, D19S433, and D8S1179), in the African sample set two loci (D21S11 and D8S1179), and in the Asian sample set only D21S11 exhibited increased allelic resolution via Sanger sequencing (Online Resource 3).

There is no doubt that the currently available sequencing technology would not be applicable to routine STR typing for multiple reasons, most importantly hands on time and cost. Taking the cost–value ratio into consideration, the slightly increased allelic resolution gained by sequencing would never justify its application for routine STR analysis. Although the ICEMS instrumentation would represent an investment and sequencing could be performed on existing platforms, the efforts required for allele separation and purification prior cycle sequencing and the incapacity of multiplexed analysis disqualify Sanger sequencing from routine STR analysis. The strength of ICEMS is the significantly increased allelic resolution compared to electrophoretic fragment size determination combined with low cost and short analysis time.

Comparison of the three population samples regarding population and forensic genetic aspects

The STR data obtained for the European, the Asian, and the African population sample show differences regarding the

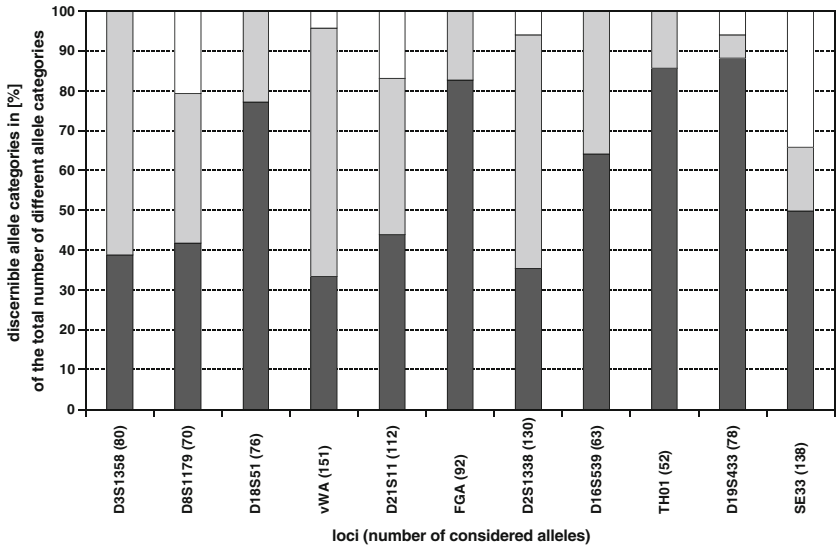

Fig. 1 Comparison of the number of differentiable allele categories by three different genotyping technologies. For each locus, a representative number of alleles (*numbers in parentheses*) were analyzed by electrophoresis, ICEMS, and Sanger sequencing. The number of allele categories discernible via electrophoresis (*black*) and ICEMS (*gray*) is compared to the number of differentiable allele classes by sequence analysis and expressed in percent. Sequence analysis (*white*) corresponds to 100% of discernible allele categories

observed allele size distributions and the corresponding frequencies considering the length variability (Online Resource 1). In the loci D3S1358, D8S1179, vWA, D2S1338, D16S539, and TH01, the three population samples exhibit only slight differences in the pattern of allele size distribution. For D18S51, D21S11, and FGA, a significant shift toward longer fragment sizes was observed for the African and/or Asian population sample in comparison to the data obtained for the European population sample. Interestingly, within the African sample set singular short allele sizes were observed in the loci D19S433 and SE33 (Online Resource 1). A similar situation—namely significant differences between the ethnicities—was observed regarding the presence of nucleotide variability within the STR loci. It is not unexpected to observe unusual sequence variants in our African population [21], as the tested individuals derive from autochthonous Khoisan tribes which had been living under rather isolated conditions, thus limiting the gene flow into other African populations and hence keeping the specific mutations which have accumulated over their long genetic history (more than 120,000 years) quite restricted to this population only. The Khoisan are reported to be one of five populations with the highest levels of genetic diversity based on variance estimates while more private alleles exist in Africa than in other regions [20]. The findings of this study are not necessarily representative for African populations in general. We were interested in investigating old populations with an expected higher STR allele diversity to increase the chance of finding unusual allele variants that can be differentiated by ICEMS. The loci D19S433 and FGA for example contained nucleotide variability only within the African sample set. However, for D16S539, ICEMS analysis brought alleles containing the same kind of base exchange (A > C) within all three investigated population samples. For D3S1358, allele categories containing one or two (C > T) base exchanges were observed within all three population samples. However, an additional allele variant of D3S1358 ("rare" variant (C > T,C > T,A > G,A > G)), which occurred exclusively in allele size 16, was observed only in the African sample set. A comprehensive summary of allele categories observed in the three population samples within the 11 STRs can be found in Table 1 and Online Resource 2.

For assessment of the forensic efficiency of each individual locus in the different population samples (~100 individuals each), we calculated the match probability, frequency of heterozygous samples, and the average power of exclusion (PE; Online Resource 5). The calculations were performed with the genotyping results taking into account only length variability (CE data) and additionally with the results including length and nucleotide variability observed via ICEMS analysis. In Fig. 2, the results of the comparison are exemplified for the PE values, for both CE and ICEMS data. For CE- and ICEMS-derived data, SE33 expectedly gave highest PE values. For CE-derived data D3S1358 brought lowest PE values for Asian and African samples (PE<0.45) and TH01 the lowest PE value for European samples (0.55). For D3S1358, D8S1179, D21S11, and FGA, strong inter-population fluctuation of PE values was observed. Taking nucleotide variability into consideration (white bars in Fig. 2), a significant increase of the PE values for almost all loci that contained nucleotide variations was observed. For the locus D3S1358, for example, the PE value showed a significant increase in all population samples (0.61 to 0.73 in the European, 0.44 to 0.68 in the African, and 0.38 to 0.53 in the Asian sample set), in the Asian sample set the PE of D21S11 jumped up from below 0.50 to more than 0.70, and for the African samples D2S1338 reached the same forensic efficiency as SE33 due to the detection of nucleotide variability (Fig. 2). The comparison clearly demonstrates the merit regarding the forensic efficiency of STR loci

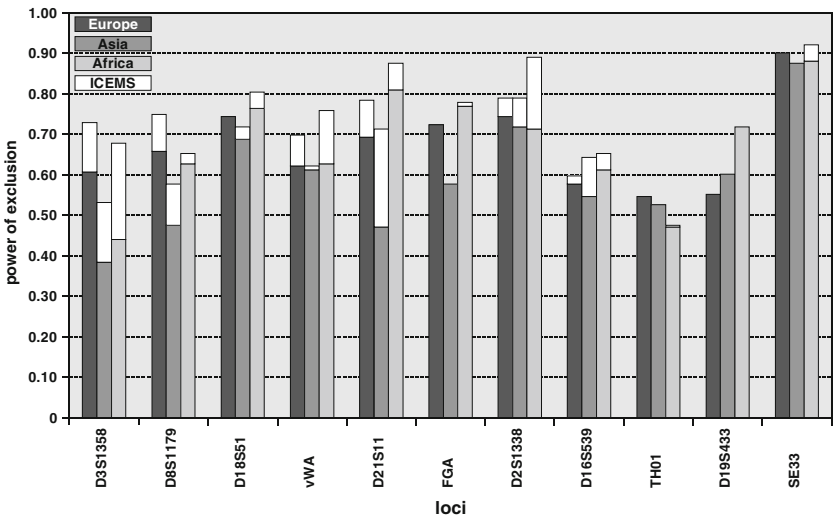

Fig. 2 Comparison of the forensic efficiency (expressed as power of exclusion) obtained via CE versus ICEMS analysis for 11 STR systems within a European, an Asian, and an African population sample. The *white bars* indicate the increase of the PE values when the additional allele categories discernible via mass spectrometric STR analysis are included in the calculation of PE

due to the detection of nucleotide variability by mass spectrometry.

Conclusions

In comparison with electrophoretic sizing, the ICEMS approach is able to characterize allele-specific length and nucleotide variability. We applied ICEMS to screen 11 STR loci that are commonly used in Europe (and the locus SE33) within an African and an Asian population sample and compared the results with data previously obtained for a European population sample. Within all 11 loci, nucleotide variations were detected and new allele variants were identified. A representative number of alleles (N=1,042) was sequenced to compare discriminatory allele categories. ICEMS clearly increased the amount of genetic information present in STR loci in contrast to CE, whereas sequence analysis brought only a slight increase in discrimination power in relation to ICEMS, which would not justify the additional cost and time effort to maximize the information content of STR typing. The new (population specific) allele variants observed in the course of the study (including base exchanges potentially present in primer binding regions of the loci D16S539 and D19S433) represent valuable data for forensic and population genetic inquiries.

Acknowledgments Financial support of the study by the Austrian Research Promotion Agency (FFG, "Österreichisches Sicherheitsforschungs-Förderprogramm" KIRAS—eine Initiative des Bundesministeriums für Verkehr, Innovation, Technologie (BMVIT), Projekt 813786) is acknowledged.

References

1. Jobling MA, Gill P (2004) Encoded evidence: DNA in forensic analysis. Nat Rev Genet 5:739–751, Erratum in: Nat Rev Genet 2005 6:246
2. Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B (1997) DNA recommendations. Further report of the DNA commission of the ISFH regarding the use of short tandem repeat systems. Int J Leg Med 110:175–176
3. Gill P, Fereday L, Morling N, Schneider PM (2006) The evolution of DNA databases—recommendations for new European STR loci. Forensic Sci Int 156:242–244
4. Gill P, Fereday L, Morling N, Schneider PM (2006) New multiplexes for Europe—amendments and clarification of strategic development. Forensic Sci Int 163:155–157
5. Gill P, Werrett DJ, Budowle B, Guerrieri R (2004) An assessment of whether SNPs will replace STRs in national DNA databases—joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). Sci Justice 44:51–53
6. Oberacher H, Pitterl F, Huber G, Niederstätter H, Parson W (2008) Increased forensic efficiency of DNA fingerprints through simultaneous resolution of length and nucleotide variability by high-performance mass spectrometry. Hum Mutat 29:427–432
7. Pitterl F, Niederstätter H, Huber G, Zimmermann B, Oberacher H, Parson W (2008) The next generation of DNA profiling—STR typing by multiplexed PCR—ion-pair RP LC-ESI time-of-flight MS. Electrophoresis 29:4739–4750
8. Keyser-Tracqui C, Crubézy E, Pamzsav H, Varga T, Ludes B (2006) Population origins in Mongolia: genetic structure analysis of ancient and modern DNA. Am J Phys Anthropol 131:272–281
9. Kraft HG, Lingenhel A, Pang RW, Delport R, Trommsdorff M, Vermaak H, Janus ED, Utermann G (1996) Frequency distributions of apolipoprotein(a) kringle IV repeat alleles and their effects on lipoprotein(a) levels in Caucasian, Asian, and African populations: the distribution of null alleles is non-random. Eur J Hum Genet 4:74–87
10. Oberacher H, Niederstätter H, Parson W (2005) Characterization of synthetic nucleic acids by electrospray ionization quadrupole time-of-flight mass spectrometry. J Mass Spectrom 40:932–945
11. Premstaller A, Oberacher H, Huber CG (2000) High-performance liquid chromatography-electrospray ionization mass spectrometry of single- and double-stranded nucleic acids using monolithic capillary columns. Anal Chem 72:4386–4393
12. Oberacher H, Niederstätter H, Casetta B, Parson W (2005) Detection of DNA sequence variations in homo- and heterozygous samples via molecular mass measurements by electrospray ionization time-of-flight mass spectrometry. Anal Chem 77:4999–5008
13. Grubwieser P, Zimmermann B, Niederstätter H, Pavlic M, Steinlechner M, Parson W (2007) Evaluation of an extended set of 15 candidate STR loci for paternity and kinship analysis in an Austrian population sample. Int J Legal Med 121:85–89
14. Huber CG, Oberacher H (2001) Analysis of nucleic acids by on-line liquid chromatography-mass spectrometry. Mass Spectrom Rev 20:310–343
15. Oberacher H, Parson W (2007) Forensic DNA fingerprinting by liquid chromatography–electrospray ionization mass spectrometry. Biotechniques 43:vii–xiii
16. Oberacher H (2008) On the use of different mass spectrometric techniques for characterization of sequence variability in genomic DNA. Anal Bioanal Chem 391:135–149
17. Grubwieser P, Mühlmann R, Berger B, Niederstätter H, Pavlic M, Parson W (2006) A new "miniSTR-multiplex" displaying reduced amplicon lengths for the analysis of degraded DNA. Int J Legal Med 120:115–120
18. Möller A, Brinkmann B (1994) Locus ACTBP2 (SE33). Sequencing data reveal considerable polymorphism. Int J Legal Med 106:262–267, Erratum in: Int J Legal Med 1997 110:46
19. Planz JV, Budowle B, Hall T, Eisenberg AJ, Sannes-Lowery KA, Hofstadler SA (2009) Enhancing resolution and statistical power by utilizing mass spectrometry for detection of SNPs within the short tandem repeats. Forensic Sci Int Gene Suppl 2:529–531. doi:10.1016/j.fsigss.2009.08.163
20. Clayton TM, Hill SM, Denton LA, Watson SK, Urquhart AJ (2004) Primer binding site mutations affecting the typing of STR loci contained within the AMPFlSTR SGM Plus kit. Forensic Sci Int 139:255–259
21. Tishkoff SA, Reed FA, Friedlaender FR et al (2009) The genetic structure and history of Africans and African Americans. Science 324:1035–1044