Forensic Science International: Genetics 6 (2012) 381–386

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Germline mutations of STR-alleles include multi-step mutations as defined by sequencing of repeat and flanking regions

Eva-Maria Dauber [a,*], Adelgunde Kratzer [b], Franz Neuhuber [c], Walther Parson [d], Michael Klintschar [e,f], Walter Bär [b], Wolfgang R. Mayr [a]

[a] Department of Blood Group Serology and Transfusion Medicine, Medical University of Vienna, Austria
[b] Institute of Legal Medicine, University of Zurich, Switzerland
[c] Institute of Legal Medicine, University of Salzburg, Austria
[d] Institute of Legal Medicine, Innsbruck Medical University, Austria
[e] Institute of Legal Medicine, Medical University of Graz, Austria
[f] Institute for Forensic Medicine, Hannover Medical School, Germany

ARTICLE INFO

Article history:
Received 18 April 2011
Received in revised form 27 June 2011
Accepted 26 July 2011

Keywords:
Germline mutation
Mutation rate
Paternity testing
Mutational step
Exclusion of paternity
Genetic inconsistency

ABSTRACT

Well defined estimates of mutation rates are a prerequisite for the use of short tandem repeat (STR-) loci in relationship testing. We investigated 65 isolated genetic inconsistencies, which were observed within 50,796 allelic transfers at 23 STR-loci (ACTBP2 (SE33), CD4, CSF1PO, F13A1, F13B, FES, FGA, vWA, TH01, TPOX, D2S1338, D3S1358, D5S818, D7S820, D8S1132, D8S1179, D12S391, D13S317, D16S539, D17S976, D18S51, D19S433, D21S11) in Caucasoid families residing in Austria and Switzerland. Sequencing data of repeat and flanking regions and the median of all theoretically possible mutational steps showed valuable information to characterise the mutational events with regard to parental origin, change of repeat number (mutational step size) and direction of mutation (losses and gains of repeats). Apart from predominant *single-step* mutations including one case with a double genetic inconsistency, two *double-step* and two apparent *four-step* mutations could be identified. More losses than gains of repeats and more mutations originating from the paternal than the maternal lineage were observed (31 losses, 22 gains, 12 losses or gains and 47 paternal, 11 maternal mutations and 7 unclear of parental origin). The mutation in the paternal germline was 3.3 times higher than in the maternal germline. The results of our study show, that apart from the vast majority of *single-step* mutations rare multi-step mutations can be observed. Therefore, the interpretation of mutational events should not rigidly be restricted to the shortest possible mutational step, because rare but true multi-step mutations can easily be overlooked, if haplotype analysis is not possible.

1. Introduction

Short tandem repeat polymorphisms (STR) are widely used for relationship testing because selected microsatellite loci show a high degree of polymorphism and their analysis is feasible and reproducible even on low amounts of DNA or degraded human material. Furthermore, the application of these markers in relationship testing requires that their mode of inheritance is known and that good estimates of their mutation rate exist.

The polymorphism of STR-alleles is based on different numbers of tandemly repeated sequence motifs 2–6 bp in length [1], which are inherited in Mendelian manner. Isolated genetic inconsistencies, however, can occur, which are usually based on *de novo* mutations in one of the parental germlines [2]. The mutational process of STR-loci seems to be mainly due to polymerase slippage: during DNA replication the polymerase can dissociate from the nascent strand and the displaced loops reanneal incorrectly to a repetitive sequence, which results in abundant, unaligned repeat motifs either in the nascent or the copied strand. As a consequence, an insertion or deletion of tandem repeat motifs in the nascent strand occurs [3]. Additionally, recombinational events like unequal crossing over or gene conversion can also contribute to loss or gains of repeat motifs. These mutational mechanisms are mainly responsible for mutations of minisatellite markers [4], whereas polymerase slippage is the favoured mutational process for microsatellites [5,6]. Furthermore "indel slippage", the duplication of short sequences by insertions and substitutions, was inferred as additional mutation mechanism for short microsatellites in further models [7]. Small step sizes (loss or

* Corresponding author at: Department of Blood Group Serology and Transfusion Medicine, Medical University of Vienna, Wahringer Gurtel 18-20, A-1090 Wien, Austria. Tel.: +43 1 40400 5320; fax: +43 1 40400 5321.
E-mail address: eva-maria.dauber@meduniwien.ac.at (E.-M. Dauber).

1872-4973/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2011.07.015

gains of one or two repeats) seem to dominate over multi-step mutations [2,8], but also larger changes can occur [9–11].

The first estimates of *in vivo* germline mutations focused on di- or trinucleotide marker on a selected chromosome, but the natural variability of mutation rates between loci was not taken into account [1]. Mutation rates are frequently given without confidence intervals, which pretends a statistical accuracy that does not exist. Moreover, a statistical bias can occur, when meioses are only counted in case of mutational events.

The first experimental data on *in vivo* germline mutations of STR-markers, which are used for relationship testing, were case reports [12] or studies on a small number of markers and/or a small number of mutational events and meioses [2,13,14]. As a higher mutation rate was expected for the male germline, mutation studies aimed to estimate germline specific mutation rates. Haplotype analyses have the potential to disclose the parental origin of a germline mutation, but they require testing of flanking polymorphisms in families with at least two children [5] to obtain phase information. As this can hardly be achieved with classical trios (parents and one child), the shortest and most simply possible mutational steps were assumed as the most probable one [2,15]. The strong preponderance of *single-step* mutations over *multi-step* mutations was then concluded from these data and confirmed in a later study on 57 cases on four STR-loci, which were investigated by haplotype analysis [5,6].

As STR-alleles involved in germline mutations are rarely sequenced, it remains unclear, if the loss or gain of nucleotides occurred in the repeat or the flanking region. For example, a deletion of 4 bp in the flanking region cannot be distinguished from the loss of a tetrameric repeat without sequencing. Such a sequence variation was already described for alleles at the D13S317 locus [16]. Mutations in flanking regions can only be observed with primers, which encompass the variant sequence. The same applies to primer binding site mutations, which can result in a loss of amplification of a parentally inherited allele. These mutations are usually older and can be resolved after amplification with alternative primers [16–18]. For calculation of mutation rates these so-called "null alleles" have to be ruled out, because they can give incorrect results in calculation of mutation rates.

Further rare genetic phenomena can also cause problems in relationship testing. In rare cases of congenital chimerism one of the two genetically different cell lines can predominate in the sample and lead to genetic incompatibilities at loci located on different chromosomes. In uniparental disomy, however, genetic incompatibilities are restricted to the pair of homologues which derives from one parent in these cases [18,19]. Genetic inconsistencies involving null alleles, chimerism or uniparental disomy are not addressed in this study.

It is widely accepted, that non-parenthood is stated, when at least three genetic mismatches on markers located on different chromosomes exist, which generally results in a PI $< 1/1000$ [20]. A germline mutation has to be inferred in case of an isolated genetic inconsistency, when the probability of parenthood of all consistent loci amounts to >99.9999%. The aim of this study was to calculate locus- and germline specific mutation rates based on sequence data and to characterise the changes in sequence structure caused by these mutational events. Possible mutation mechanisms are discussed.

2. Materials and methods

2.1. Samples and DNA extraction

The samples were collected from children and their parents of Caucasoid origin, mostly trios or from single-parent families residing in Switzerland (area of Zurich) or Austria (area of Vienna, Salzburg, Innsbruck and Graz). STR-typing was carried out on these samples for relationship testing or other purposes. The DNA was extracted from peripheral blood or buccal swabs by the participating laboratories with different methods [21,22] or commercially available kits (Qiamp® DNA Blood Mini kit, Qiagen, Valencia, USA or NucleoSpin® Blood, Macherey-Nagel, Düren, Germany).

2.2. Initial and additional STR-analysis

The mismatches were primarily detected by the participating laboratories. The oldest samples were originally analysed with singleplex-PCR, subsequent polyacrylamide gel electrophoresis and silver staining combined with minsatellite markers and/or classical marker systems (e.g. red cell antigens, protein polymorphisms). Further samples were already initially tested with commercially available (AmpFlSTR® SGMplus® or Identifiler® PCR Amplification kit, Applied Biosystems, Foster City, USA) or *in-house* [23] multiplex-PCR kits. To bring all these heterogeneous sets of marker systems tested up to a common marker set, they all were subsequently analysed with a commercial 16-locus multiplex PCR-System (Powerplex16™ System, Promega, Madison, USA) except those, which have already been tested with the AmpFlSTR® Identifiler® kit before. These two 16-locus multiplex kits share 13 STR-loci, whereas the Penta D and Penta E loci are restricted to the Powerplex16™ System and the D2S1338 and the D19S433 loci to the AmpFlSTR® Identifiler® kit. The Penta D and Penta E loci were tested in a complementary duplex-PCR reaction [24] in those cases, which have been previously analysed with the AmpFlSTR® Identifiler® kit. Furthermore, the loci SE33 and D12S391 were additionally tested [23], if not included beforehand. As a result, a set of at least 20 STR-loci was analysed in all cases, in part complemented by D6S389, D8S1132 or minisatellite markers and conventional marker systems (red cell antigens, protein polymorphisms). In selected cases (ACTBP2_1, ACTBP2_9, D8S1179_1, D21S11_1, and CSF1PO_4) the Humantype Chimera and/or the Mentype Pentaplex ESS PCR Amplification Kit (Biotype, Dresden, Germany) were included, thus testing for nine (D2S1360, D3S1744, D4S2366, D5S2500, D6S474, D7S1517, D8S1132, D10S2325, D21S2055) or four additional loci (D1S1656, D2S441, D10S1248, D22S1045) respectively.

2.3. Null alleles

The following alternative, outlying primers were used for PCR to rule out possible null alleles in two cases of opposite homozygosity on genotype and sequence level (D8S1179_1 and FGA_1). The primers were designed with the Primer Express Software (Applied Biosystems) or taken from literature [25]:

D8S1179-alternative forward: 5′-TGTGCATTGTTGTTGGGAATG-3′
D8S1179-alternative reverse: 5′-CGGCCGGGCAACTTATATG-3′

FGA-alternative forward: 5′-ACTGGCATTCATGGAAGGCTGC-3′
FGA-alternative reverse: 5′-TCGGTTGTAGGTATTATCACGGTCTG-3′

2.4. Flanking microsatellites

The microsatellite-loci D8S1720 and D8S1799 flanking the D8S1179 locus (0.3 Mb centromeric and 3 Mb telomeric) have been analysed with the respective primer pairs (UniSTS:58041 and UniSTS:64440) in a singleplex PCR reaction (forward primer FAM-labelled) using the PCR protocols and cycling conditions of vWA and D13S317, respectively.