The STR approach

B. Brinkmann, Institute of Legal Medicine, Münster, Germany

Introduction

Since the beginning of forensic hemogenetics, forensic scientists have been dreaming of a technology which should ideally fulfill a few basic requirements.
A high statistical efficiency is needed as well as a high suitability to the quality of the stain with its properties: degradation, contamination, inhibition etc. Also, an extreme sensitivity is desired especially if only microstains are present. Furthermore, the whole process involved needs a good success rate.
The invention/introduction of VNTR polymorphisms by Jeffreys has opened a wide door to these dream castles but we were still miles away from the ideal. There existed problems with the reproducibility especially of MLPs and there existed and still exist problems relative to the accuracy and precision of the SLPs. Although these technologies are statistcally unbelievably efficient, there are further problems concerning the sensitivity and degradation.
Microsatellite polymorphisms have been introduced by Edwards et al. (1991), Polymeropoulos et al. (1992), Tautz (1991) and others in the early nineties. At least theoretically, this DNA generation seems to easily resolve many of the disadvantages of earlier technologies. This will be investigated further.
By definition microsatellites or STRs (short tandem repeats) vary by repeat size between 2 and 7 bp (Edwards at al. 1991) with fragment sizes < 350 bp. Out of these dinucleotide polymorphisms suffer from the disadvantage of quite intensive slippage bands leading to patterns which are difficult to interpret in stain work. Tetranucleotide polymorphisms - thousands have been predicted to be present in the human genome (Weber and May 1992, Tautz and Rentz 1991) - have the advantages of no or negligable slippages and consecutive alleles can be more easily resolved. Therefore among those which have been elaborated for forensic application, tetra- and pentanucleotide polymorphisms

appear to be most promising.

Categories of microvariation

We subdivide into three categories of STR systems : (1) STRs with low microvariation mainly consisting of repeat related length variants. (2) STRs with intermediate microvariation where in addition to category (1) structure and length variations exist. (3) STRS with high microvariation showing extensive structure and length variations.
(1) Low microvariation: CD4 is a system of the first category with pentameric repeat units as basic motif (Fig. 1a). From allele 9 onwards the 4th repeat unit shows a T to C transition which is the only minor variation in this system. 11 consecutive alleles have been observed. - Black Africans show this multiplicity of 11 alleles, only 2 of them reaching the 20%-level (Fig. 2). Three rather different Caucasian populations (Germans, Turks, Moroccans) show quite similar profiles but are nearly reduced to a 3-allele model. Especially the German population nearly lacks all other alleles while the 2 other populations show some admixture of African alleles. Again the 3 Asian populations are very similar but are nearly reduced to a 2-allele model. From these differences one could conclude that these uneven distributions reflect different allele pools in the respective founder groups (Brinkmann et al. 1995).

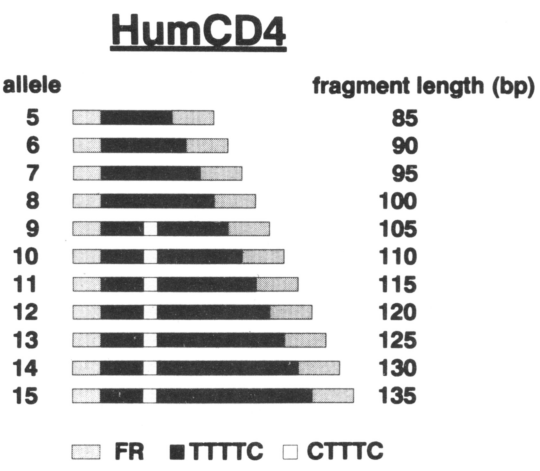
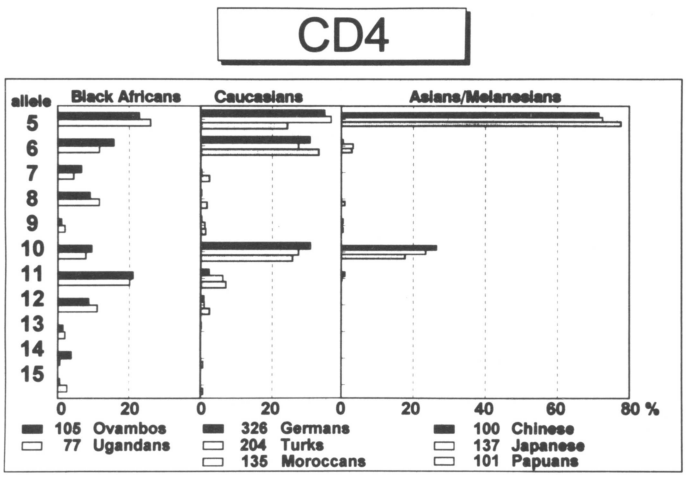


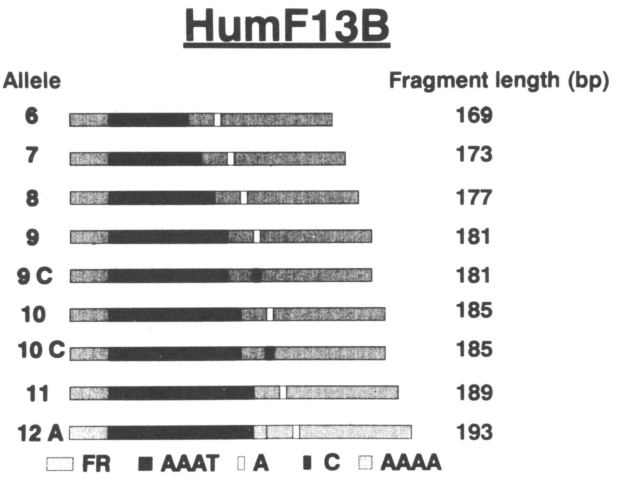
Fig. 1a) Schematic presentation of HumCD4 alleles 5 - 15. Nomenclature according to the number of repeats.
FR = flanking region.

Fig. 1b) Allele distribution of CD4 alleles in 9 populations differentiated into 3 major ethnic groups. The figure preceding each population = number of individuals.



HumF13B is an example of a tetranucleotide STR with low microvariation (Fig. 1b). There exist 6 regular alleles differing by one repeat unit and 2 rare A to C transversions in the 3' flanking region of 2 alleles and a further T to A transversion in allele 12 (Alper et al. 1995).
These variant alleles - all point mutations - can be easily resolved in native gels, not of course under denaturing conditions.

Fig. 2a) Schematic presentation of HumF13B alleles 6 - 12A detectable usinga non-denaturing gel system. FR = flanking region; C = cathodal; A = anodal.



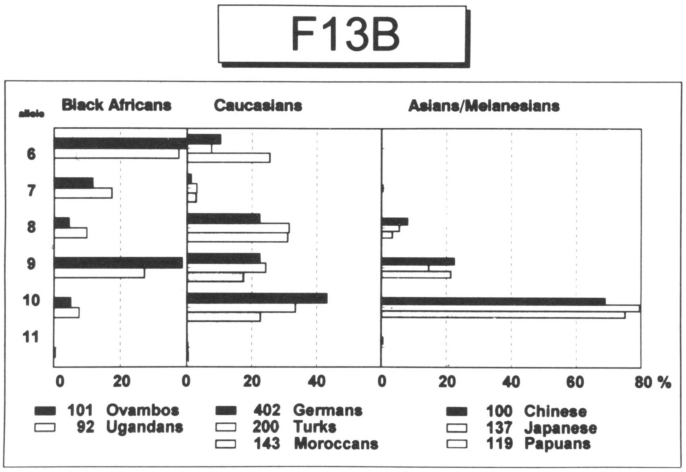


Fig. 2b) Allele distribution of HumF13B alleles. Mode of representation as in Fig. 1b.

Africans have 5 alleles with appreciable frequencies. Caucasian are reduced to 4 frequent alleles and a rare allele 7 with 2 % only in Germans. The 3 Asian populations are reduced to a 3-allele model. At least this looks like another founder effect. HumFES is an example of low microvariation with similarities to HumF13B. In addition to 6 regular repeat related variants there exist variants 10A and 11A with A to C transversion in the 5' flanking region (Möller et al. 1994). (A stands for anode and C for cathode because these alleles migrate either slower or faster relative to the consensus alleles of same length).

(2) Intermediate microvariation: VWA belongs to the intermediate microvariation group. There exist 2 alleles that differ basically. The consensus allele has a composite repetitive structure starting with 1 TCTA, continuing with 3 to 5 TCTGs and ending with the proper variable TCTA part. Allele 14 is unique and differs from the consensus at these 3 sites (Fig. 3a).

It seems to show until now only very few stepwise mutated daugh ter alleles in either population (Barber et al. 1995). There exists also no allele 14 with a consensus structure in either population. On the other hand, alleles 15 upwards show some length variation of the TCTG proportions (Fig. 3b). Therefore VWA alleles can be divided into 2 groups: 11 to 14 with considerable conservativeness of their structure and length with no or at most extremely low new mutations. Alleles 15 to 22 where we

have already observed 2 new mutations with either insertions or deletions of single repeats (Brinkmann et al. 1995) and a third one more recently.

Fig. 3a) Comparison between HumVWA consensus alleles and variant allele 14

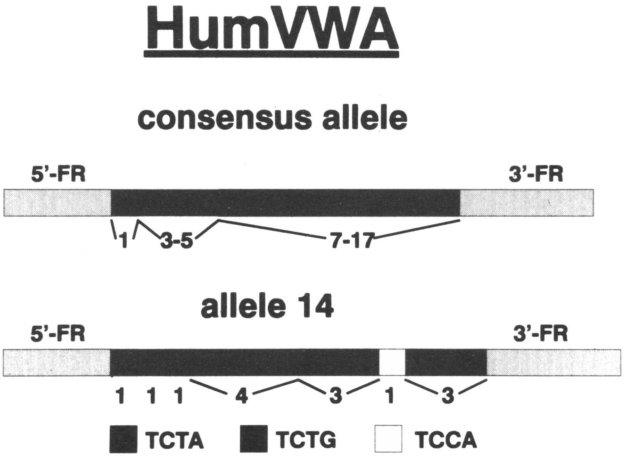
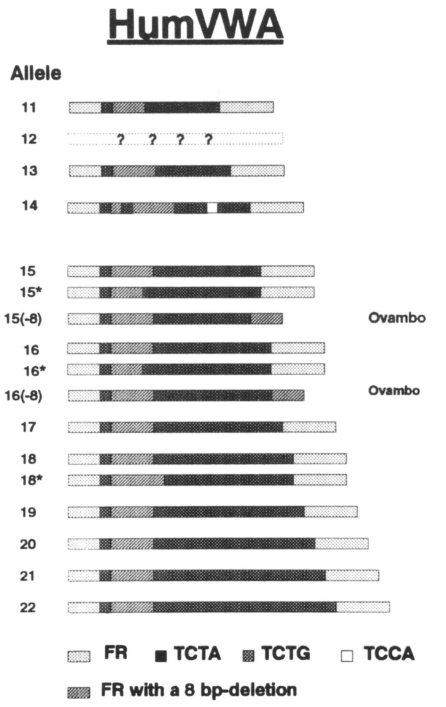


Fig. 3b) schematic presentation of known VWA alleles. Further explanation see text.



(3) High microvariation: ACTBP2 or SE33 belongs to group 3 and is one of the most complex systems. There are at least 3 series of alleles: Series 1 ranging between 12 and 21 with a more or less regular repeat structure (Fig. 4). Series 2 with the appearance of a hexamer within the repeat structure, but at different sites. These alleles range between 22 and 32. Series 3 with the insertion of 2 hexamer units, again at different sites within the repeat structure. Range 32 to 33. Because of this high complexity there have been early warnings as to the reproducability, especially under different gel conditions.

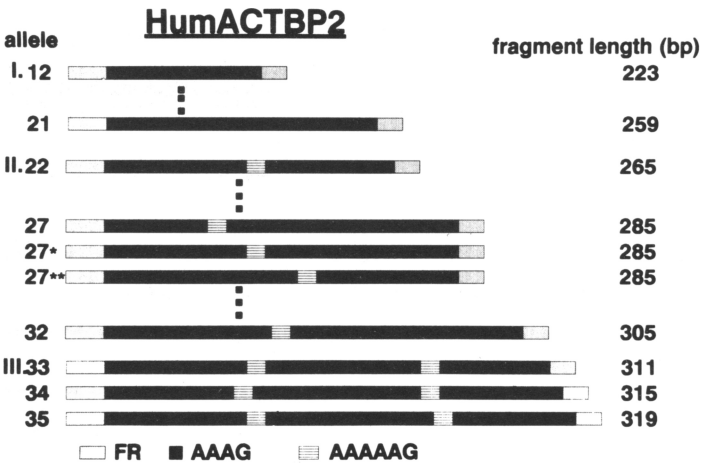


Fig. 4) Examples of 3 different classes of HumACTBP2 alleles.

Sequencing of 250 randomly selected alleles lead to approximately 100 structure and length variants while typing of 1200 was associated with about 30 electrophorecticly detectable alleles only. - The DI of 99.8% is equivalent to the combined DI of VWA, TH01, CD4 (Table 2).

The mean phenotype frequency between the old and the new nomenclature was even decreased by the factor 10. Apart from this extreme efficiency, sequencing would also lead to a high level of accuracy and specificity. The aforementioned systems are only selected examples. Further examples are HumTH01 for group (1) and D21S11 for group (3).

Table 1) ACTBP2: Increase of forensic efficiency values including sequencing data. MEC = mean exclusion chance.

HumACTBP2

n sequenced alleles: 250
n defined alleles: 98

Discrimination power: 99.8 %
(≙VWA, TH01, CD4)

MEC: 94.0 %

$\bar{x}$ phenotype (f) old: 1.0 %
$\bar{x}$ phenotype (f) new: 0.1 %

STR microvariation is caused by point mutations, repeat insertions/deletions and gross changes (Möller et al. 1994, Meyer et al. 1995, Brinkmann et al. 1995). The highest new mutation rate observed so far is associated with ACTBP but still lies under 0.8% which is approximately equivalent to some RFLPs. This figure seems to be extremely small in the low microvariation systems (Table 3). We have predominantly observed either insertions or deletions of single repeats (Brinkmann et al. 1995).

Table 2) Overview of forensically relevant parameters of STRs with low, intermediate and high microvariation.
MEC = mean exclusion chance; HWE = Hardy-Weinberg equilibrium.

STR Genetics

system		meioses (n=)	mutations n	mutations %	MEC (%) Caucasian data	HWE	homozyg excess
low	CD4	278	0	< 0.1%	41	yes	no
	FES	525	0		44		
	F13B	509	0		44		
	TH01	725	0		59		
inter-med	VWA	741	2	< 0.3%	63	yes	no
high	D21	529	1	< 1.0%	65	yes	no
	ACTBP2	508	4		89		

comb. 99.9 %

Differences of phenotype frequencies

Based on the observed differences of allele frequencies in major ethnic groups we calculated the combined phenotype frequencies

comparing different populations using 50 individuals randomly taken from each population sample (basic population). Their individual frequencies were calculated using 6 STRs. In the population to be compared, the frequency of each phenotype was calculated as well and also the ratio between both populations. Ratios greater than 1 indicate that the profile occurs more often in the basic population; e. g. 100 would indicate a frequency difference of 100-fold. Ratios smaller than 1 indicate that the profile is more common in the population compared. - Example (Table 4): The first Japanese individual - J1 - has a frequency of 10^{-8} and the German equivalent 10^{-11} leading to a ratio of 10^{3}. The mean value in this group is also ca. 10^{3}. In other words, the mean difference is ca. 1000. The other way around, i. e. starting with the Germans (basic population), the average ratio is 10^{6}.

Table 3) Example of combined phenotype frequency differences between Japanese and Germans. Ind. = Individual; J = Japanese; G = German; MV = mean value. Further explanation see text.

Phenotype (f)

(n = 6 systems)

Ind.	Jap.	Ger.	ratio	Ind.	Ger.	Jap.	ratio
J 1	10^{-8}	10^{-11}	10^{3}	G 1	10^{-9}	10^{-14}	10^{5}
J 2	10^{-9}	10^{-9}	10^{0}	G 2	10^{-6}	10^{-11}	10^{5}
J 3	10^{-7}	10^{-12}	10^{5}	G 3	10^{-8}	10^{-16}	10^{8}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
J 50	10^{-8}	10^{-10}	10^{2}	G 50	10^{-7}	10^{-13}	10^{6}
		MV =	10^{3}			MV =	10^{6}

Between German Caucasians and two other Caucasian populations this difference is less than one order of magnitude (Fig. 5). Compared to other major ethnic groups, the mean differences range between 10^{3} and 10^{8}.

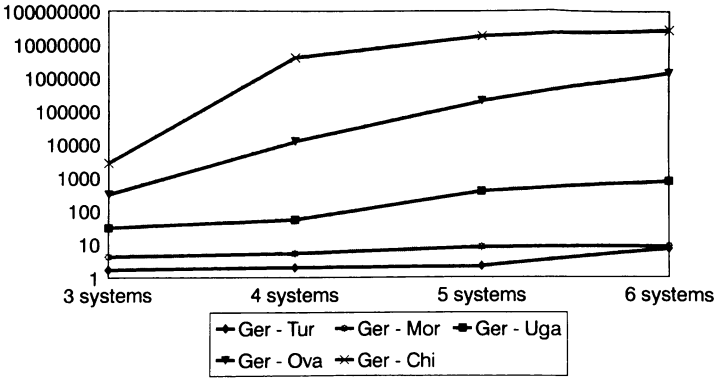


Fig. 5) Graphical presentation of combined phenotype frequency differences between Germans and other populations using 3-6 STRs. 3 systems = THO1, VWA, FES; 4 systems = THO1, VWA, FES, F13B; 5 systems = THO1, VWA, FES, F13B, CD4; 6 systems = THO1, VWA, FES, F13B, CD4, D21S11.

The reconstruction of phylogenetic trees involving 9 populations using the UPGMA method (average linkage analysis; Nei 1987) (DISPAN software kindly provided by Dr. Nei, Pennsylvania) leads to a phylogenetic tree grouping closely related populations (Fig. 6). They form a first category of pairwise relationships: Turks-Germans, Papuans-Aborigines, Japanese-Chinese, Ovambos-Ugandas. In the second category the Moroccans are grouped together with Turks and Germans and also, the Chinese-Japanese on the one side and the Papuans-Aborigines on the other side.

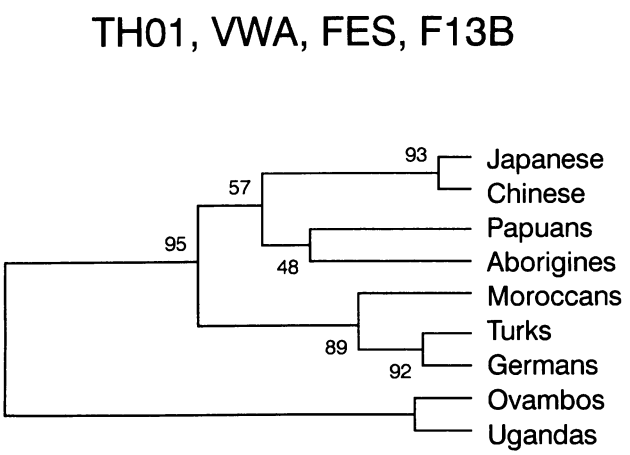


Fig. 6) Phylogenetic tree reconstructed using average linkage analysis (UPGMA). The numbers represent the bootstrap values.

To summarize: STRs seem to have some reciprocal properties:
1. There exist robust systems with a low microvariation which are usually easily typeable and associated with low mutation

rates and a high discrimination power between populations of different major ethnic groups. On the other hand there are highly variable systems. The definition of the respective alleles necessitates skill. Their mutation rates are still acceptable. Their possibility to distinguish between populations is low, but their forensic efficiency values are high. And there exist a series of intermediate microvariation systems with properties averaging between both extremes.

From these characteristics it can be derived that the selection of an ideal set of STRs will strongly depend on the question to be solved, i. e. in a mixed population consisting of different major ethnic groups, STRs with high microvariation could be useful in order to keep the resulting phenotype frequency differences low. Other parameters will be the through-put of labour, the machines available and the composition of the populations involved and of course the sample quality.

Literature

Alper B, Meyer E, Schürenkamp M, Brinkmann B (1995) HumFES/FPS and HumF13B: Turkish and German population data. Int J Legal Med (in the press)

Barber MD, Piercy RC, Anderson JF, Parkin BH (1995) Structural variation of novel alleles at the Hum vWA and HumFES/FPS short tandem repeat loci. Int J Legal Med 108:31-35

Brinkmann B, Möller A, Wiegand P (1995) Structure of new mutations in 2 STR systems. Int J Legal Med 107:201-203

Brinkmann B, Sajantila A, Goedde HW, Matsumoto H, Nishi K, Wiegand P (1995) Population genetic comparisons among eight populations using allele frequency and sequence data from three microsatellite loci. Eur J Hum Genet (in the press)

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49:746-756

Meyer E, Wiegand P, Rand SP, Kuhlmann D, Brack M, Brinkmann B (1995) Microsatellite polymorphisms reveal phylogenetic relationships in primates. J Mol Evol 41:10-14

Möller A, Meyer E, Brinkmann B (1994) Different types of structural variation in STRs: HumFES/FPS, HumVWA and HumD21S11. Int J Legal Med 106:319-323

Polymeropoulos MH, Rath DS, Xiao H, Merril CR (1992) Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2). Nucleic Acids Res 20:1432

Tautz D (1993) Notes on the definition and nomenclature of tandemly repetitive DNA sequences. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds) DNA fingerprinting: state of the science. Birkhäuser, Basel, pp 21-28

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12:4127-4138

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388-396