Development of New STRs for Forensic Casework: Criteria for Selection, Sequencing & Population Data and Forensic Validation

Angel Carracedo and M.V. Lareu
Institute of Legal Medicine. University of Santiago de Compostela, Spain

1. INTRODUCTION

STRs are currently the markers which are most commonly used for Forensic purposes and they are likely to continue being widely used in the forthcoming years. But, do we really need more STRs?

There are advantages and disadvantages in deciding to explore new STRs for forensic casework. On the one hand, the STRs which are currently used are probably not the best ones since many of them were proposed when knowledge on this subject was in its early stages. On the other hand criminal databases are being established in many countries with prefixed STRs and National QC and accreditation systems can not handle the never ending number of markers.

When we decided to explore new STRs this was mainly due to our belief that now STRs were really necessary for specific forensic purposes.

The general criteria for selecting STRs include characteristics such as high heterozygosity, low stutter, few and low extra-peaks, robustness, easy multiplexing and a low mutation rate, but the importance of these criteria depends on the specific forensic purposes. For instance low mutation rate STRs are important for paternity testing but not so important for stain analysis. There are at least three types of STRs especially needed in the forensic world:

1. Short STRs for typing degraded materials
2. STRs with low stuttering characteristics for analyzing mixtures
3. Y chromosome STRs for analyzing male-female mixtures especially in sexual crimes

2. SELECTION OF STRS

Short tandem repeats are widely distributed throughout the human genome, occurring with a frequency of 1 locus every 6-10 kb (l). Although STRs are composed of tandemly repeated sequences of 1 to 6 bases in length, tetranucleotide and pentanucleotide repeats are preferred by forensic scientists due to their low stutter characteristics.

Polymorphic tetranucleotide repeats are common in thc human genome and therefore as a first step in our selection process a survey in genome databases including Genbank, CHLC and EMBL was carried out using, a heterozygosity of over 75% and the possibility of designing common amplification conditions as criteria for selection. So, 40 STRs were selected.

50 samples were amplified for these 40 STRs and a further selection was carried out taking the real hetero-zygosity, and the robutsness for common amplification conditions into consideration. It is interesting to comment that the published heterozygosity for all 40 systems according to the genome databases was 20% lower than the observed heterozygosity, and in some STRs was really critical, going from a published rate of 90% to a real rate lower than 60%. This could be in part due to the low variation of the population used in our study, the Galician population which is located on the westernmost edge of continental Europe. This low variation is also reflected in other STRs and in mtDNA.

After this second step a total of 11 STRs were selected for a more detailed study. The selected systems are shown in Figure 1. All of them were teranucleotide repeats and most of them were detected using GATA probes.

The primers used in this study are shown in Figure 2. Most of them are the primers already described and only a few were modified in order to use common amplification and sequencing conditions. PCR and sequencing conditions are shown in Figure 3.

A population genetic study was performed in 70 to 160 unrelated individuals from Galicia for each of the eleven STRs. Samples were typed by comparison with ladders made up of a cocktail of samples containing most of the alleles observed for each system.

Fig. 4 summarizes the characteristics of the systems. Their size ranges from 100 to 350 bp, the number of observed alleles from 7 to 20, and the heterozygosity from 0.69 to 0.90 with a discrimination power of over 0.9. The discrimination power of each individual system is over 90% in most of these STRs.

Because of their high heterozygosity and ideal characteristics, 6 systems were selected for a further study. These systems are D1S1656, D6S477, D9S302, D12S391, D18S535 and D22S683. All of them have high heterozygosities, they are robust, they can be easily included in multiplexes and in addition some of them have low stutter characteristics.

Our aim was then to sequence the observed alleles in order to find out about the structure of each system, the production of sequenced allelic ladders, and the performance of population and validation studies.

3. SEQUENCING DATA OF SELECTED STRS

All the observed alleles, including many common alleles of the same length were sequenced and the corresponding nomenclature was established according to the recommendations of the DNA Commission of the ISFH (2).

STRs range from very simple ones to very complex ones. Among the selected systems we found almost all varieties of STRs, from simple STRs such as the D18S535 to complex ones such as the D9S302 system.

The sequence structure of the different systems is shown in Figure 5 to Figure 10

D1S1656 is a compound STR with the basic sequence structure shown in Figure 5. (TAGA)n repeats form the basis of the system. Only the shortest alleles do not contain a tetranucleotide (TAGG) at the end of the array. Allele 12 may or may not show the (TAGG) unit. From allele 15, intermediate .3 alleles are observed due to a (TGA) insertion in the middle of the (TAGA) units.

The sequence structure of D6S477 is shown in Figure 6. Regular D6S477 alleles consist of (TCTA)n repeats. Intermediate .2 alleles show only one (TA) repeat instead of the 2 (TA) repeats observed in regular alleles. This is a very interesting STR for understanding the origin of variation in STRs. It is also a very robust system although the heterozygosity is slightly lower than 80% in Galicians.

D9S302 is a complex STR. Its sequence structure is shown in Fig 7. The complexity of this system is intricate, therefore it is difficult to represent the whole variation of the system. The polymorphism of this system consists of variations of (ATCT) n units separated by different motifs. This originates 17 different alleles and a heterozygosity of over 90% in Caucasians.

D12S391 is a compound STR consisting of (AGAT) and (AGAC) repents. The basic structure of this STR is shown in Fig 8. Up to now 15 different allelic groups were found. The shortest alleles (from allele number 15 to number 18) have variations in the (AGAT) repeat unit exclusively, whereas the rest of the sequence is constant. The intermediate .3 alleles are originated due to the presence of one (GAT) unit between the (AGAT) repeats. The complexity of this system increases in larger alleles with variation in the number of both (AGAT) and (AGAC) motifs. Additional variation in the largest allele (from number 20 to number 26) occurs due to the presence or absence of an AGAT unit at the end of the tandem array.

D18S535 is the simplest allele of this group and it is composed of (GATA)n repeats. The basic structure of this STR is shown in Figure 9. In the Galician population we found eight common alleles differing in the number of (GATA) repeats. We haven't found any intermediate alleles up to now. This system is a short and robust STR with very low stutter characteristics.

D22S683 is a compound STR consisting of (TATC)n and (TA)n repeats (Figure 10). A variable number of (TA)n repeats inserted within the (TATAC) units originates regular and .2 alleles in most of the allelic classes. In addition some rare .3 alleles were observed due to the insertion of an (ATC) in the variable (TACT)n part. D22S683 has shown to have an extremely low percentage of stutters, lower than all the tetranucleotides that we are using in casework.

4. POPULATION STUDIES

Allelic ladders were constructed using regularly spaced sequenced alleles and the corresponding population studies performed in the population of Galicia (NW Spain). Heterozygosities are over 0.8 in most of the populations studied until now (Figure 11) but more population studies are needed to know the distribution of these STRs in human populations. D12S391 and D1S1656 have been further studied by our group and other groups and therefore more population data is available. These two systems have similar heterozygosities (around 0.9) in all the population groups studied with some variation in the distribution of frequencies in the main population groups, but with very similar profiles in Caucasians. All the populations studied up to now for these STRs were found to be in Hardy-Weinberg equilibrium. In addition these STRs were shown to be independent of each other in all the studies performed.

5. MUTATION RATE

The number of meiosis studied is too low to make conclusions about the mutation rate of the majority of these STRs. Up until now (Figure 12) we have observed 3 mutations for the D12S391 locus and 1 mutation in the D1S1696 locus, given that the former has a mutation rate of 0.21 and the latter 0.16. A similar mutation rate was observed for D12S391 in the study performed by Brinkmann *et al*. (3). As it happens, in the latter study the mutation event was repeat related owing to a single-step mutation. If we use the model developed in the aforementioned study correlating the mutation rate and the geometric mean of the repeat length we can expect a rate between 0.1 and 0.4% for our six systems.

6. STABILITY IN DEGRADED SAMPLES

It has been shown in a previous study performed by our group (4), that there is a direct relationship between the average size of each PCR system and positive results when typing series of degraded samples. We are currently performing studies on the stability of these markers in degraded samples. Preliminary studies are showing, as expected, that stability is directly related to size and very good results are being obtained for the shorter systems (D18S535 and D1S1656). For this reason, we are currently looking for alternative primers for some of these systems. D12S391 and D6S477 can be reduced 100 bp in size using new primers.

7. ARTIFACTUAL BANDS AND STUTTERS

Apart from stutters, only insignificant artifactual bands were observed in all of these STRs. A stutter is characteristically 1 repeat unit or 4bp less than its associated alleles and is mainly caused by slippage of the *Taq* enzyme during the amplification process. The proportion of the stutter product relative to the main allele (percent stutter) is measured by dividing the height of the stutter peak by the height of the main allele peak. The calculation of the percentage of stutter for each individual system is important in forensic genetics for the establishment of guidelines for automation and especially for analyzing mixtures. Many factors influence the percentage of the stutter. These factors include the characteristics of the system itself, but also the PCR conditions and reagents and factors determining stringency during the post-PCR and pre-electrophoresis handling of samples (5). Furthermore, the method for estimating the peak area also influences the final result.

We have measured hundreds of samples using common standard procedures (and using standard *Taq*) and we have observed results similar to other published results (5, 6). D18S535 and D22S683 show a percent stutter lower than the other STRs used in our lab (Figure 13), even better than other tetranucleotides such as TH01 and FES which are known to have good stuttering characteristics. As a result these two systems (Figure 14) were, the ones which had a lower stutter percentage and they were therefore classified as systems with low stuttering characteristics, D6S477 and D1S1656 were included in the group of systems with medium stuttering characteristics together with others such as FIBRA or D21S11. Finally D12S391 has a high percent stutter and it was included in the group of high stutter characteristics together with SE33 or VWA31/A but D12 behaved better than these two systems.

Each allele within a locus displays a percent stutter that is quite reproducible. The average standard deviation ranges from 0.3 to 0.9. This parameter is of crucial importance for the evaluation of mixed samples.

8. BAND SHIFT ANALYSIS

It has recently been shown by Gill *et al*. (7) that it is possible to detect single base-pair differences between samples by using measurements of band shift, relative to allelic ladders run on the same gel, something like a "floating window method". The size of each STR band detected in an amplified sample is compared with the closest equivalent ladder band size and, in addition, the absolute shift from the appropriate ladder band to each pair of STR bands is also calculated. Alleles are designated where the differences in band size are within 0.5 bp and the absolute shift is no more than 0.5 bp. Intermediate alleles can be determined accurately by following this rule but not all STRs behave well on this regard, as the mobility of the strands can be strongly influenced even in denaturing conditions by factors such as slight differences in the AC/TG proportions (8). Therefore, checking for band shift analysis is necessary. All the systems included in this study have been shown to behave well and therefore the (0.5 bp rule is applicable in all these systems (9).

9. VALIDATION

Validation studies following similar parameters to those recommended by TWGDAM were carried out. These include robustness, stability, mixtures, non-human studies, mutation rate and checking for independence with other loci. In our opinion the final validation of a system cannot be carried out by individual groups and companies and should always be performed by an internationally established validation group. In Europe a final assessment and intercomparison exercises are usually performed by the EDNAP group, a working group of the ISFH.

D12S391 and D1S1656 were tested in the 1997 exercises performed by the group (10). Both of these systems showed excellent reproducibility between labs. In the D12S391 system a rare 18.3 allele was erroneously scored as 19 by three labs. On further analysis using the 0.5 bp rule, the correct designation was obtained. This exercise emphasized the importance of using the criteria described and the importance of achieving standards in the labs for the detection of alleles varying by 1 bp.

10. DISSEMINATION OF RESULTS. PUBLICATION OF DATA

Sequence, validation and population data on these systems has been published or submitted for publication (11- 14). Sequenced allelic ladders as well as research data are freely available to researchers and commercial companies on request

11. FUTURE DEVELOPMENTS

In connection with these STRs we aim to complete the population data for each of these STRs and above all to design new primers for shortening some of these STRs and in this way design multiplexes for degraded samples. At this moment in time we don't intend to develop new STRs in autosomal STRs but to contribute to the analysis of the variation in the Y chromosome. In this sense we have also sequenced most of the STRs in current use in our laboratory, developed allelic ladders as well as multiplexes which are now being validated (15).

ACKNOWLEDGMENTS

The authors wish to thank all those who collaborated in the Forensic Genetic Service of the Institute of Legal Medicine of Santiago for their tireless efforts. Special thanks are due to Sandra Barral who performed most of the sequencing work on these new STRs.

REFERENCES

1. Beckman J.S., Weber J.L.(1992) Survey of Human and Rat Microsatellites. *Genomics*, **12**:627-631.
2. Bar W, Brinkmann B, Budowle B, Carracedo A, Gill P., Lincoln P., Mayr W., Olaisen B. DNA recommendations. (1997) Further Report of the DNA Commission of ISFH Regarding thc Use of Short Tandem Repeat Systems, *Int. J. Legal Med.* **110**:175-176
3. Brinkmann B., Klintschar M., Neuhuber F., Huhne J., Rolf B. (1998) Mutation Rate in Human Microsatellites: Influence of the Structure and Length of the Tandem Repeat. *Am. J. Hum. Genet.* **62**:1408-1415.
4. Alvarez-Garcia A., Munoz I., Pestoni C., Lareu M.V., Rodriguez-Calvo M.S., Carracedo A. (1996) Effect of Environmental Factors on PCR-DNA Analysis from Dental Pulp. *Int J Legal Med* 1996; 109:125-129.
5. Meldgaard M., Morling N. (1997) Detection and Quantitative Characterization of Artificial Extra Peaks Following Polymerase Chain Reaction Amplification of 14 Short Tandem Repeat Systems Used in Forensic Investigations. *Electrophoresis* **18**:1928-1935.
6. Schumm J., Bacher J., Hennes L., Gu T., Micka K., Sprecher C., Lins A., Amiott E., Rabbach D., Taylor J., Tereba A. Pentanucleotide repeats: highly polymorphic genetic markers displaying minimal stutter artifact. The Second European Symposium on Human Identification. Insbruck, June 9-12, 1998. (Proceedings in press).
7. Gill P., Urquhart A., Millican E., Oldroyd N., Watson S., Sparkes R., Kimpton CP. A New Method of STR Interpretation Using Inferential Logic, *Int. J. Legal Med.*, **109**:14-22.
8. Saitoh H., Ueda S., Kurosaki K., Kiuchi M. (1998) The Different Mobility of Complementary Starnds Depends on the Proportion AC/GT. *Forensic Sci. Int.*, **91**:81-90.
9. Phillips C.P., Syndercombe Court D., Lareu M.V., Haaskamp J., Carracedo A. (1998) Band Shift Analysis of Three Base-Pair Repent Alleles in the Short Tandem Repeat Locus. *Forensic Sci. Int.*, **93**:79-88.
10. Gill P., d'Aloja E., Dupuy B., Jangblad M., Johnsson V., Kloosterman A.D., Kratzer, Lareu M.V., Meldegaard M., Phillips C., Pfitzinger H., Rand S., Sabatier M., Scheithauer R., Schmitter H., Schneider P., Vide M.C. (1998, in press) Report of the European DNA Profiling Group (EDNAP)- An Investigation of the Hypervariable STR Loci ACTBP2., APOA11 and D11S554 and the Compound Loci D12S391 and D1S1656. *Forensic Sci. Int.*
11. Lareu M.V., Pestoni C., Schurenkamp M., Rand S., Brinkmann B., Carracedo A. (1996) A Highly Variable STR at the D12S391 Locus. *Int. J. Legal Med.*, **109**:134-138.
12. Lareu M.V., Barral S., Salas A., Pestoni C., Carracedo A. (1998) Sequence Variation of a Hypervariable Short Tandem Repeat at the D1S1656 Locus. *Int. J. Legal Med.* **111**:244-277
13. Lareu M.V., Barral S., Salas A., Carracedo A. (1998, in press) Sequence Variation of a Hypervariable Tandem Repeat at the D18S535 Locus, *Int. J. Legal Med*.
14. Lareu M.V., Barral S., Salas A., Carrcedo A. (1998, in press) Sequence Variation of Two Hypervariable Short Tandem Repeats at the D22S683 and D6S477 Loci. *Int. J. Legal Med.*
15. Pestoni C., Cal M.L., Lareu M.V., Rodriguez-Calvo M.S., Carrecedo A. (1998, in press) Y Chromosome STR Haplotypes: Genetic and Sequencing Data of the Galician Population (NW Spain). *Int. J. Legal Med*.

NAME	N	No.ALLELES	SIZE (bp)	HETEROZYGO EXP.	OBS.	CE	PD
D1S549	96	10	157-193	0.79	0.75	0.59	0.93
D1S1656	125	13	123-163	0.89	0.90	0.77	0.98
D2S436	70	10	179-203	0.70	0.71	0.46	0.87
D3S1754	97	9	162-194	0.76	0.71	0.56	0.91
D6S477	70	12	206-240	0.82	0.75	0.63	0.94
D9S302	105	17	255-353	0.90	0.90	0.77	0.98
D12S375	70	7	279-303	0.73	0.70	0.47	0.87
D12S391	166	18	209-253	0.90	0.89	0.75	0.97
D18S535	129	8	129-157	0.78	0.80	0.57	0.91
D19S433	79	7	199-221	0.75	0.69	0.52	0.89
D22S683	114	20	168-206	0.91	0.89	0.81	0.94

Figure 1: STR loci selected in the second step

PRIMERS

D1S549	CAAAGAGGACATGTGTTTGTG TACCAGCAATGGGTAGTATGG
D1S1656	GTGTTGCTCAAGGGTCAACT GAGAAATAGAATCACTAGGGAACC
D2S436	GATATGGGAGCAACATGAGC GGAATCAACTTTCAGTATAAACCC
D3S1754	ACGCTTTTAAGGGGTTTTTG CATTTCTGATCTGGAACGCT
D6S477	GATTTGCCATGATAGATGGC GGGGGATATCTCAAACAACC
D9S302	GGGGACAGACTCCAGATACC GCGACAGAGTGAAACCTTGT
D12S375	TTGTTGAGGGTCTTTCTCCA TCTTCTTATTTGGAAAAGTAACCC
D12S391	AACAGGATCAATGGATGCAT TGGCTTTTAGACCTGGACTG
D18S535	TCATGTGACAAAAGCCACAC GACAGAAATATAGATGAGAATGCA
D19S433	CCTGGGCAACAGAATAAGAT TAGGTTTTTAAGGAACAGGTGG
D22S683	AACAAAACAAAACAAAACAAACA GGTGGAAATGCCTCATGTAG

Figure 2: Primers for the STRs selected in the second step

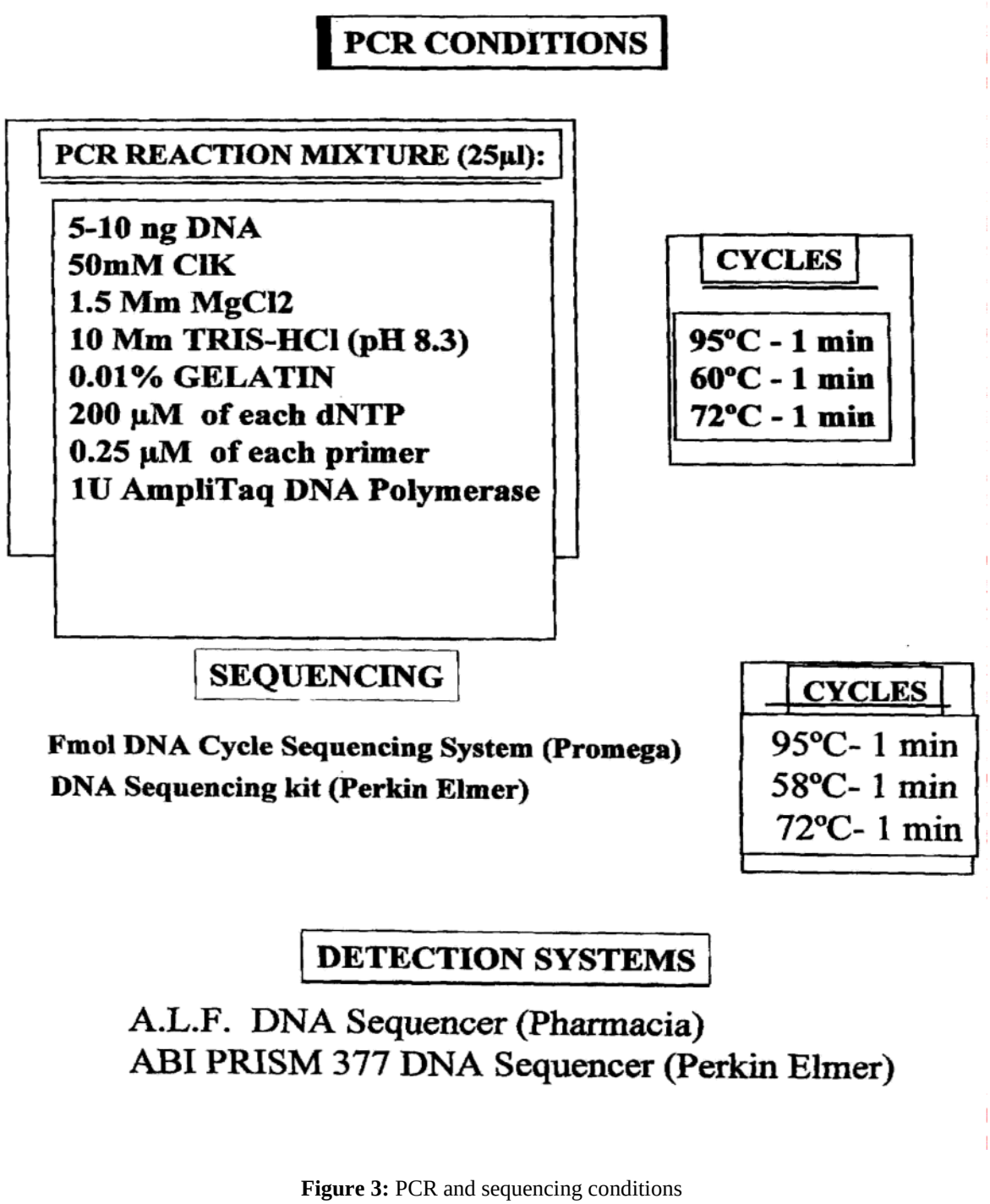

Figure 3: PCR and sequencing conditions

STR	SIZE(bp)	HETEROZ.	ROBUSTNESS	STUTTERING	MULTIPLEX
D18S535	129-157	0.80	+++	LOW	+++
D1S1656	123-163	0.90	+++	MEDIUM	+++
D12S391	209-253	0.89	+++	HIGH	+++
D22S683	168-206	0.89	+++	LOW	+++
D6S477	206-240	0.75	+++	MEDIUM	+++
D9S302	255-353	0.90	+++	MEDIUM	+++

Figure 4: Characteristics of the 6 systems selected

D1S1656

ALLELE	SIZE(bp)	COMPOSITION $(TAGA)_N$ $(TGA)_{0-1}$ $(TAGA)_N$ $(TAGG)_{0-1}$
10	129	(TAGA)10 (TG)5
11	133	(TAGA)11 (TG)5
12	137	(TAGA)12 (TG)5
12^2	137	(TAGA)11 (TAGG)1 (TG)5
13	141	(TAGA)12 (TAGG)1 (TG)5
14	145	(TAGA)13 (TAGG)1 (TG)5
15	149	(TAGA)14 (TAGG)1 (TG)5
15.3	152	(TAGA)4 (TGA)1 (TAGA)10 (TAGG)1 (TG)5
16	153	(TAGA)15 (TAGG)1 (TG)5
16.3	156	(TAGA)4 (TGA)1 (TAGA)10 (TAGG)1 (TG)5
17	157	(TAGA)16 (TAGG)1 (TG)5
17.3	160	(TAGA)4 (TGA)1 (TAGA)12 (TAGG)1 (TG)5
18.3	164	(TAGA)4 (TGA)1 (TAGA)13 (TAGG)1 (TG)5
19.3	168	(TAGA)4 (TGA)1 (TAGA)14 (TAGG)1 (TG)5

Figure 5

D6S477

(TCTA)n (TA)1 (TCTA)n (TA)0-1 (TCTA)n

ALLELES	SIZE (bp)	
13.2	206	(TCTA)3 (TA)1 (TCTA)10
14.2	210	(TCTA)3 (TA)1 (TCTA)11
15	212	(TCTA)2 (TA)1 (TCTA)2 (TA)1 (TATC)10
15.2	214	(TCTA)3 (TA)1 (TCTA)12
16	216	(TCTA)2 (TA)1 (TCTA)2 (TA)1 (TATC)11
17	220	(TCTA)2 (TA)1 (TCTA)2 (TA)1 (TATC)12
17.2	222	(TCTA)3 (TA)1 (TCTA)14
18	224	(TCTA)2 (TA)1 (TCTA)2 (TA)1 (TATC)13
19	228	(TCTA)2 (TA)1 (TCTA)2 (TA)1 (TATC)14
20	232	(TCTA)2 (TA)1 (TCTA)2 (TA)1 (TATC)15
21	236	(TCTA)2 (TA)1 (TCTA)2 (TA)1 (TATC)16
22	240	(TCTA)2 (TA)1 (TCTA)2 (TA)1 (TATC)17

Figure 6

D9S302

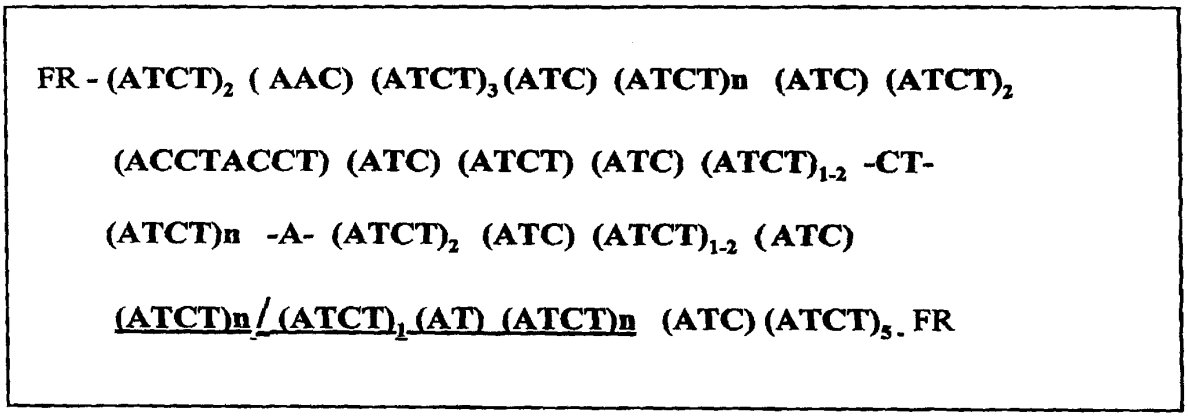

Figure 7

D12S391

ALLELE	SEQUENCE
15 (209bp)	(AGAT)8 (AGAC)6 (AGAT)
16 (213bp)	(AGAT)9 (AGAC)6 (AGAT)
17 (217bp)	(AGAT)10 (AGAC)6 (AGAT)
17.3 (220bp)	(AGAT)1 (GAT)1 (AGAT)8 (AGAC)7 (AGAT)
18 (221bp)	(AGAT)11 (AGAC)7 (AGAT)
18.3 (224bp)	(AGAT)1 (GAT)1 (AGAT)9 (AGAC)7 (AGAT)
19 (225bp)	(AGAT)12 (AGAC)6 (AGAT)
19^2 (225bp)	(AGAT)11 (AGAC)7 (AGAT)
19.3 (228bp)	(AGAT)1 (GAT)1(AGAT)10(AGAC)7 (AGAT)
19.3 (228bp)	(AGAT)5 (GAT)1 (AGAT)7 (AGAC)7 (AGAT)
20 (229bp)	(AGAT)11 (AGAC)8 (AGAT)
20^2 (229bp)	(AGAT)11 (AGAC)9
21 (233bp)	(AGAT)14 (AGAC)6 (AGAT)
21^2 (233bp)	(AGAT)12 (AGAC)9
22 (237bp)	(AGAT)15 (AGAC)6 (AGAT)
22^2 (237bp)	(AGAT)14 (AGAC)7 (AGAT)
22^3 (237bp)	(AGAT)13 (AGAC)9
22^4 (237bp)	(AGAT)12 (AGAC)10
23 (241bp)	(AGAT)14 (AGAC)8 (AGAT)
23^2 (241bp)	(AGAT)14 (AGAC)9
23^3 (241bp)	(AGAT)13 (AGAC)10
24 (245bp)	(AGAT)15 (AGAC)8 (AGAT)
24^2 (245bp)	(AGAT)14 (AGAC)9 (AGAT)
24^3 (245bp)	(AGAT)15 (AGAC)9
25 (249bp)	(AGAT)16 (AGAC)8 (AGAT)
25^2 (249bp)	(AGAT)16 (AGAC)9
26 (253bp)	(AGAT)17 (AGAC)8 (AGAT)
26^2 (253bp)	(AGAT)17 (AGAC)9

Figure 8

D18S535

(GATA)1 (GACA)1 (GATA)1 (GAT)1 (GATA)n

ALLELE - SIZE		COMPOSITION
9	(130bp)	(GATA)9
10	(134bp)	(GATA)10
11	(138bp)	(GATA)11
12	(142bp)	(GATA)12
13	(146bp)	(GATA)13
14	(150bp)	(GATA)14
15	(154bp)	(GATA)15
16	(158bp)	(GATA)16

Figure 9

D22S683

$(TA)_1\ (TATC)_1\ (TA)_1\ (TATC)_1\ (TA)_{1.....}(TATC)_N\ (ATC)_{0-1}\ (TATC)_N$

ALLELE	-SIZE-bp	COMPOSITION
12	168	(TA)4 (TATC)3 (TATC)7
13	172	(TA)4 (TATC)3 (TATC)8
13.2	174	(TA)3 (TATC)2 (TATC)10
14	176	(TA)4 (TATC)3 (TATC)9
14.2	178	(TA)3 (TATC)2 (TATC)11
15	180	(TA)4 (TATC)3 (TATC)10
15.2[1]	182	(TA)3 (TATC)2 (TATC)12
15.2[2]	182	(TA)5 (TATC)4 (TATC)9
16	184	(TA)4 (TATC)3 (TATC)11
16.2[1]	186	(TA)3 (TATC)2 (TATC)13
16.2[2]	186	(TA)7 (TATC)6 (TATC)7
17	188	(TA)4 (TATC)3 (TATC)12
17.1	189	(TA)3 (TATC)2 (TATC)11 (ATC)1 (TATC)2
17.2	190	(TA)7 (TATC)6 (TATC)8
18	192	(TA)4 (TATC)3 (TATC)13
18.2	194	(TA)7 (TATC)6 (TATC)9
18.3	195	(TA)4 (TATC)3 (TATC)11(ATC)1 (TATC)2
19	1 96	(TA)6 (TATC)5 (TATC)11
19.2	198	(TA)7 (TATC)6 (TATC)10
20	200	(TA)8 (TATC)7 (TATC)9
20.2	202	(TA)7 (TATC)6 (TATC)11
21	204	(TA)8 (TATC)7 (TATC)10

Figure 10

STR	SIZE(bp)	CAUCASIAN	CHINESE	AFROCARIBBEAN
D18S535	129-157	0.800	-	-
D1S1656	123-163	0.900	0.894	0.913
D12S391	209-253	0.899	0.890	0.902
D22S683	168-206	0. 890	-	-
D6S477	206-240	0.752	-	-
D9S302	255-353	0.890	-	-

Figure 11: Heterozygosity of the systems selected in the different populations

STR	SIZE (bp)	No. MEIOSIS	No. MUTATIONS	MUTATION RATE (%)
D18S535	129-157	65	0	-
D1S1656	123-163	620	1	0.16
D12S391	209-253	1462	3	0.21
D22S683	168-206	50	0	-
D6S477	206-240	50	0	-
D9S302	255-353	40	0	-

Figure 12: Number of mutations observed and mutation rate

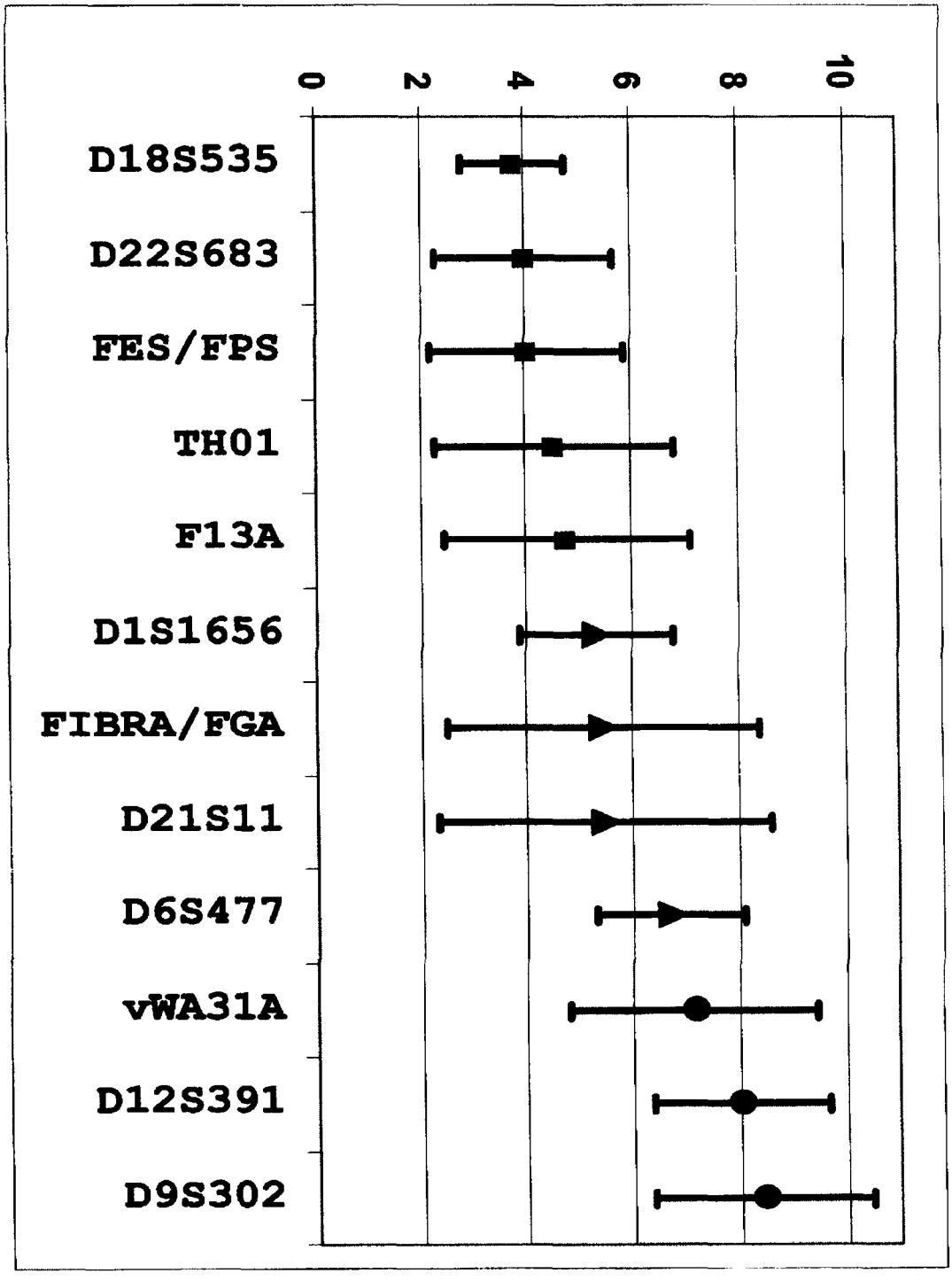

Figure 13: Percent stutter and standard deviation for each STR

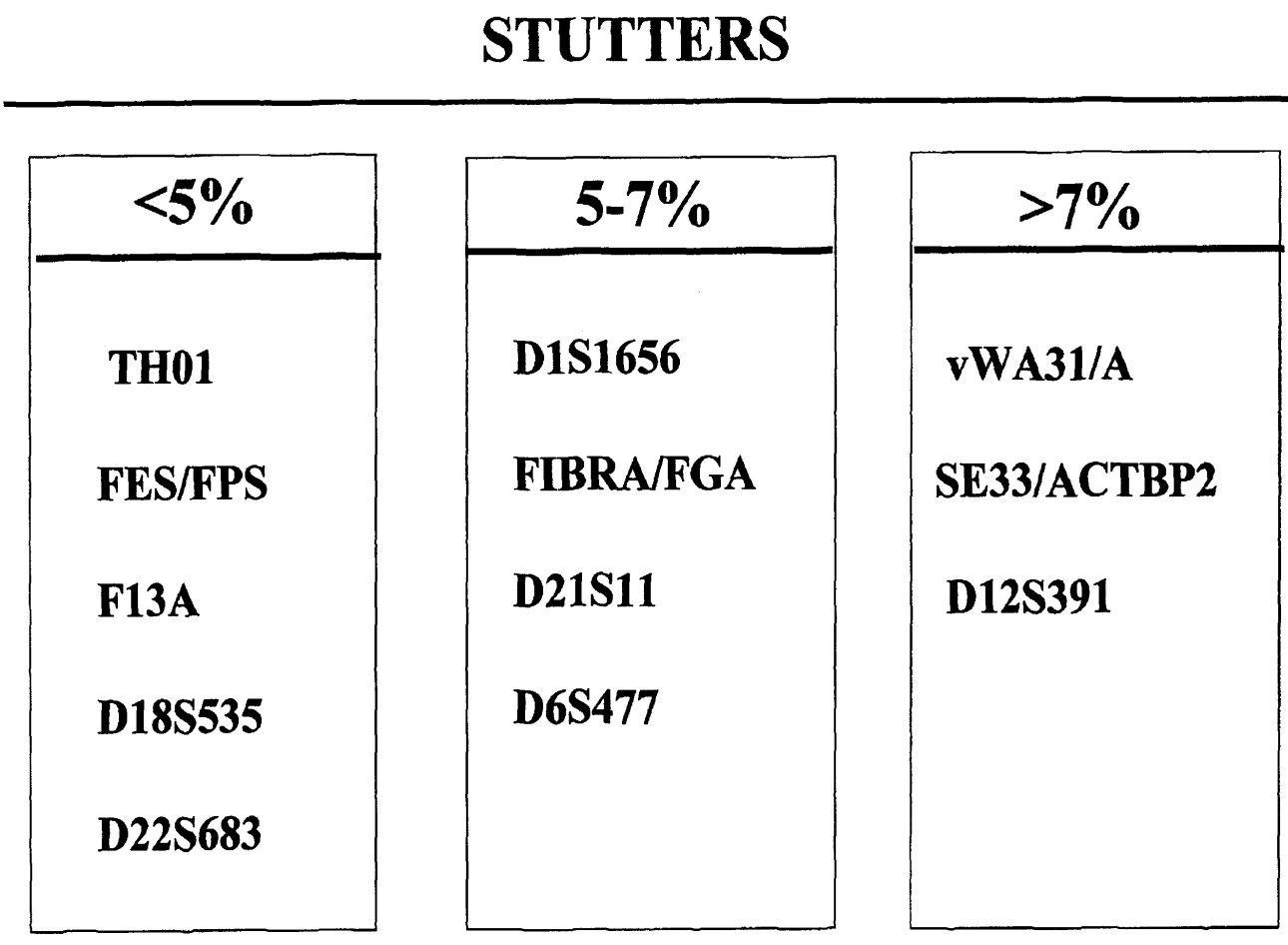

Figure 14: Stutter characteristics of some systems used in our laboratory

Table 3: Paternity Indices

Microsatellites

	CSF1PO	TH01	TPOX	FES	F13A01	vWA	D13S317	D7S820	D16S539	D6S366	FABP
PI	3.3	4.8	1.8	2	2.4	1.7	4.9	3.6	2	2.5	1.7
PI_T	3.3	15.8	28.5	57	137	232	1140	4104	8208	$2.0X10^4$	$3.4X10^4$

Minisatellites

	D1S80	YNH24	TBQ7	EFD52	LH1	MS1	PH30	CEB42
PI	16	5	9.2	5.8	31.2	5	6.2	6.5
PI_t	16	80	736	4269	$1.3x10^5$	$6.6x10^5$	$4.1x10^6$	$2.6x10^7$

PI: paternity Indices
PIt: Accumulative Paternity Indices

Table 4: Cumulative Paternity Indices

	Y-Specific STRs	Autosomal STRs	AmpFLP	SLPs	Total
PI	33	$3.4x10^4$	16	$1.7x10^6$	$3.2x10^{13}$