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)

Sir: It is unlikely that SNPs will replace STRs as the preferred method of testing of forensic samples and database samples in the near to medium future. This is because STRs have several advantages:

a) Extensive databases already exist that are STR based. There would be a significant cost involved in retyping of existing samples with SNPs. For most jurisdictions retyping would not be an option because samples are not retained. Use of different marker systems would compromise our ability to carry out cross-border comparisons in the short to medium term.

b) The cost advantages for SNPs are not obvious at present. Although platforms and biochemistry are being developed that enable extremely high throughput, this research is carried out primarily for the pharmaceutical industry for drug discovery and is not directly applicable to forensic use. This is because forensic scientists have to contend with small-degraded samples. There are substantial difficulties associated with building large multiplexes. It is easy to be swept along with the tide of new technology, simply because it is something different. However, we must take a step back and objectively evaluate the reasons for implementing a new system. New does not necessarily mean better or applicable. We should remember that SNPs were initially developed for forensic application in the late 1980s and early 1990s, but were superseded by STRs because of their marked superiority – the discussion is not new. Since then, the SNP platform and biochemistry have undergone development to merit fresh consideration within the context of high throughput and associated reduced costs. However, concurrent parallel research continues to miniaturise and to reduce the costs of STR platforms as well. Consequently, more time is needed to assess the new developments.

c) Whereas just 10-16 STR loci suffice for most existing forensic applications, between 50-100 SNP loci will be required to fulfil the same purpose. Building such large multiplexes is a significant hurdle at present. Although there is significant research in this area, a validated solution has not yet emerged.

d) One area where STRs have a significant advantage is in mixture interpretation, because each STR locus contains multiple alleles. Consequently, a two person mixture that consists of up to 4 different alleles may easily be interpreted. SNPs are limited in that they are biallelic. As a result, interpretation is more limited. A SNP profile mixture may be characterised because there are more heterozygotes across the panel than expected. But the resolution of specific contributors is more difficult because only two allelic states exist for many SNPs. In addition there will be a desire for a SNP based test to be quantitative, and some current SNP assays are not quantitative. On the other hand, the absence of stutter artefacts with SNPs is a benefit for some interpretation scenarios. In addition, in common with STRs, it is envisaged that SNPs will also be susceptible to phenomena associated with low copy number (LCN) analysis (currently used by some laboratories as a method to analyse highly degraded or limited DNA samples). In particular, drop-out and the increased risk of spurious contamination will also occur with SNPs. Any validation programme for LCN analysis will need to address these issues in detail.

e) This does not mean to say that SNPs will not fulfil an important purpose in the forensic armoury, but they are likely to be used for specific purposes only. The ability to analyse highly degraded samples will be enhanced with autosomal SNPs because of the potential to use primers that reside close to the target sites, thereby minimising the amplicon sizes.

f) SNPs will be used for some specialist applications – in particular, mtDNA (currently used for bone, teeth, hair shaft); Y-chromosome DNA (used to analyze the male component from mixed stains and to possibly elucidate ethnicity); and common-place physical characteristics (skin, hair, eye colour)

Selection of autosomal SNPs

We do not envisage replacement of STRs by SNPs in the short to medium term for reasons outlined above. However, autosomal SNPs have been very useful in some circumstances. In particular, body parts of mass disaster victims may be highly degraded, hence SNPs will offer the opportunity to carry out further analysis where STRs fail to give a result, as demonstrated by analysis of samples from the World Trade Centre. However, we envisage that in the case of mass disaster analysis STRs will be used first, and SNPs will be used to supplement STRs where needed. We note that in relation to parentage analysis and family reconstruction, STRs have proven to be highly successful in the past e.g. Waco disaster and various air disasters. However, even if the DNA is high quality there are occasions when there are insufficient family members available to achieve a high level of confidence with an association. To achieve this purpose, either new STRs could be developed, or alternatively, existing STRs could be supplemented with a SNP panel. Consequently, SNPs could begin to serve a dual purpose – to improve power of exclusion for family reconstruction and also to provide an additional level of support to analyse highly degraded material (i.e. where STRs fail). There also are efforts for modifying existing STR panels by decreasing the size of amplicons by designing new primers.

Neither mass-disaster nor paternity analysis is dependent upon national DNA databases. This means that for the initial introduction, standardisation is not a necessity. Provided that analyses are carried out using the same set of loci then standardisation is achieved by default on a per-case basis.

Regardless, it is difficult to predict the medium to long term future (>5–10 years). Therefore, it may be desirable at some stage in the future to supplement the existing STRs with some SNPs. Encouraging global participation in studies on a

method(s) based on a few loci would foster support and develop foundations, as well as produce peer reviewed publications.

The technology is still maturing and the subject of intensive research, so no central method(s) or specified SNPs have emerged in the forensic community. There are numerous methods currently being evaluated. Currently, it is not clear which platforms or biochemistry are likely to be the best. However, it is both likely and desirable that multiple options will arise. For example, solutions for casework stains at scenes of crime are likely to be different to solutions to analyse large numbers of reference samples in the laboratory. It is implicit that results from different methods must be directly comparable and interchangeable.

Consequently, the SNP selection process must take account of different methodologies that are available. Thus, sequence data on candidate loci that are selected should be made freely available to laboratories working in this area so that they may be evaluated using different methodologies. To facilitate the process, and to promote standardisation of SNPs it may be necessary to focus on readily available loci.

The process of standardisation can follow a specific route proposed here. It will require co-ordination by a committee of experts. In Europe, experts will be drawn from the ENFSI and in North America experts will be drawn from SWGDAM. These will comprise the SNP standardisation group. Information flow will follow by exchange of ENFSI members and SWGDAM members at each other's meetings. Eventually, a standard set of loci will be chosen by mutual agreement

It is proposed to place information onto the National Institute of Science and Technology (NIST) web-site (http://www.cstl.nist.gov/biotech/strbase/SNP.htm) in order to inform the forensic community of progress and to openly invite suggestions for consideration by the SNP standardisation group. Any laboratory may submit SNPs to the group using the downloadable proforma (Figure 1). The GENBANK designation of a SNP will be required along with an analysis of some major population groups to provide preliminary frequencies of the polymorphism. Information about the primer design, biochemistry and multiplexing capabilities will also be required. This information will be placed onto the web-site. Participant laboratories will be encouraged to develop assays for these markers. A good marker is one that will work in different formats and is highly polymorphic in a major ethnic group(s). The standardisation group will encourage feedback and at a minimum will publish reports on the Internet. Eventually a body of information will be achieved that will enable a selection process to be carried out on the basis of the criteria set above (along with a list of recommended methods).

Non-autosomal SNPs

Standardisation of non-autosomal SNPs is much more straightforward. For example, the mtDNA molecule is small and population structure is more evident so any useful SNPs discovered will quickly be adopted for use. There is probably a larger choice of Y-chromosome SNPs and their selection will be dependent upon the ethnic group(s) that predominate and the substructure within a population. Such Y SNP panels may well vary among geographic areas. Standardisation may not be a necessity because it is not envisaged that Y chromosome markers will be used in national DNA databases. Alternatively Y chromosome markers may be applied to missing persons databases where paternal lineage is relevant. For both mtDNA and Y chromosomal DNA it is likely that labs will select markers from a standard panel that are relevant for their specific populations.

Peter Gill, David J. Werrett
Forensic Science Service, Trident Court, 2960 Solihull Parkway, Birmingham UK
Bruce Budowle, Richard Guerrieri
FBI, Quantico, VA, USA

Figure 1 SNP proforma showing submission details.

SNP/Locus Name /TSC Designation : TSC 0421768

General Information

GenBank Entries	NCBI Submitted SNP ID NCBI Reference SNP ID	Genomic location
NT_008046 NT_008046.13 AC027258.2	ss 2384956	Chr 8
	rs 1542931	91,669,649

General Details

Polymorphism (ancestral/derived, if known)	Orientation with respect to GenBank accession (+/-)
C/G	plus strand

Figure 1 SNP proforma showing submission details (continued).

Reference Sequence (FASTA format; at least 200 bp on either side of the SNP for primer/assay design)

5' flank: ttagagattt gaggtggggt aaggggcatg gtgttcgggg acaacattgt cctctgggag
ctcagcatct tcttgataga tgagtcatgc acatggagaa aattgaatat ttttcttttc
ttttttttt ttctttttt tgagacagag tcttgctctg tctcccaggc tggagcgcag
tggtgcaatc tcggctcacc gcaacctcca cctcccaggt tcaagagatt ctcctgcctc
cgcctcccga gtagctggga ctacaggcgc ttgccaccat gcccggctaa ttttttgggg
ttttgttcag cagagacagg gtttcactct gttagccagg atggtctcga tctcctgacc
tcatgatcca cctgcctcgg cctcccaaag tactgggatt acaggcgtga gccactgcac
ctggcgaaaa ttgattaatt ttcaacatca acgtcaacag cacaactctg ctacagcgaa
gagctacagc aaatgaagca tcaataattt tcaatg
Observed: S(c/g)
3' flank: gttcacaatg caagaggcat cactaagcca aatgagtgag aaaacaccct aatgcattag
agttctgagg aggaagaggt ct

Population Allele Frequencies

Panel (No. of Individuals)	Allele 1 Frequency	Allele 2 Frequency	References
FSS N. European (86)	C = 73%	G = 27%	1
FSS Indo-Pakistani (33)	C = 74%	G = 26%	1
FSS Afro-Caribbean (29)	C = 62%	G = 38%	1

Detection Protocols

Detection Protocol	PCR Primer/Probe Sequences	Reference
FSS URP Autosomal SNP Detection Protocol	Forward Primer(s) : gatgcctcttgcattgtgaac(g/c)	1
	Reverse Primers : gctcaacagcacaactctgctacagc	
	Detection probe : n/a	
	Forward Primer(s) :	
	Reverse Primers :	
	Detection probe :	
	Forward Primer(s) :	
	Reverse Primers :	
	Detection probe :	
	Forward Primer(s) :	
	Reverse Primers :	
	Detection probe :	
	Forward Primer(s) :	
	Reverse Primers :	
	Detection probe :	

General Comments (e.g., utility/usage in multiplexes, multi-copy locus, equivalent to other markers-Y SNPs, present in commercial assay)

Present in FSS Autosomal Snp Multiplex Assay (1)
Amplicon size in PCR (1) : 127bp
PCR Cycling Conditions : 95c 11mins ; 94c 30s,
60c 15s, 72c 15s, 60c 15s, 72c , 60c 15s, 72c 30s x6 Cycles
94c 30s, 76c 105s x29 Cycles
94c 60s, 60c 30s, 72c 60s x3 Cycles
60c 10mins
URP Sequences : Forward Universal Sequence 1:
cgacgtggtggatgtgctatgatgcctcttgcattgtgaacg
Forward Universal Sequence 2:
tgacgtggctgacctgagacgatgcctcttgcattgtgaacc
Reverse Universal Sequence :
caagctggtggctgtgcaaggctcaacagcacaactctgctacagc