

## STRs vs. SNPs: thoughts on the future of forensic DNA testing

John M. Butler · Michael D. Coble ·  
Peter M. Vallone

Accepted: 4 January 2007 / Published online: 12 September 2007  
© Humana Press Inc. 2007

**Abstract** Largely due to technological progress coming from the Human Genome and International HapMap Projects, the issue has been raised in recent years within the forensic DNA typing community of the potential for single nucleotide polymorphism (SNP) markers as possible replacements of the currently used short tandem repeat (STR) loci. Our human identity testing project team at the U.S. National Institute of Standards and Technology (NIST) has explored numerous SNP and STR loci and assays as well as developing miniSTRs for degraded DNA samples. Based on their power of discrimination, use in deciphering mixture components, and ability to be combined in multiplex assays in order to recover information from low amounts of biological

material, we believe that STRs rather than SNPs will fulfill the dominant role in human identity testing for the foreseeable future. However, SNPs may play a useful role in specialized applications such as mitochondrial DNA (mtDNA) testing, Y-SNPs as lineage markers, ancestry informative markers (AIMs), the prediction of phenotypic traits, and other potential niche forensic casework applications.

**Keywords** DNA · DNA typing · DNA profiling · Short tandem repeat · Single nucleotide polymorphism · MiniSTR · STR · SNP · mtDNA

---

*Official disclaimer:* Contribution of the U.S. National Institute of Standards and Technology. Not subject to copyright. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose. This work was funded in part by the National Institute of Justice through interagency agreement 2003-IJ-R-029 with the NIST Office of Law Enforcement Standards.

---

J. M. Butler (✉) · M. D. Coble · P. M. Vallone  
National Institute of Standards and Technology, Biochemical  
Science Division, 100 Bureau Drive, Mail Stop 8311, Building  
227, Room A243, Gaithersburg, MD 20899, USA  
e-mail: john.butler@nist.gov

*Present Address:*

M. D. Coble  
Armed Forces DNA Identification Laboratory, Research Section,  
1413 Research Blvd., Bldg 101, Rockville, MD 20850, USA

### Introduction

Both length and sequence genetic variation exist in human populations. These forms of variation enable forensic DNA testing because many different alleles can exist in non-coding regions of the genome. When information from multiple unlinked genetic markers is combined, high powers of discrimination are possible. Length variation, in the form of short tandem repeat (STR) markers, has become the primary means for forensic DNA profiling over the past decade [1]. Large national DNA databases now exist containing millions of STR profiles based on a few core STR markers [2, 3].

While current STR systems and DNA databases are working well, the question often posed is “what does the future hold for forensic DNA testing?” Will another set of genetic markers, such as single nucleotide polymorphisms (SNPs), replace the core STR loci already so effectively in use? These questions were addressed in 2000 by the Research and Development Working Group of the National Commission on the Future of DNA Evidence [4] and will

be considered here briefly based on the current state of the science with STRs and SNPs.

For a number of years, largely due to technology progress coming from the Human Genome and International HapMap Projects [5], the primary potential replacement for STRs has usually been proposed to be SNPs, which are sequence variants that occur on average every several hundred bases throughout the human genome [6]. Abundant SNP loci have been characterized and studied in various human populations [7].

Most articles to date discussing the potential applications of SNPs in forensic DNA testing have focused on technology reviews [8, 9] or the number of markers needed to generate equivalent powers of discrimination in single source samples [10–12]. Recently work with improving the level of multiplex amplification has been addressed [13, 14], as has the selection of optimal SNP loci for use in forensic panels [15]. However, direct comparisons between SNPs and STRs suggest that SNPs are not ready to replace STRs as the workhorse of forensic DNA identification markers.

### Potential SNP advantages

The primary reason provided for considering SNPs with forensic applications centers around the fact that a higher recovery of information from degraded DNA samples is theoretically possible since a smaller target region is needed. Only a single nucleotide needs to be measured with SNP markers instead of an array of nucleotides—sometimes hundreds of nucleotides in length—as with STRs (Fig. 1). While this argument has been made for many years, a recent direct comparison for SNPs and STRs found that STR markers, when shortened to miniSTRs, perform quite well on degraded DNA material [13]. A wide variety of STR loci exist, and ones with moderate allele ranges and small amplicon sizes can work effectively with compromised DNA samples. Primer redesign to create smaller amplicons with core STR loci [16] and a number of new STR loci [17] has been described recently.

Another advantage of SNPs is that they possess mutation rates approximately 100 thousand times lower than STRs ( $10^{-8}$  vs.  $10^{-3}$ ). Thus, theoretically SNPs, being more stable in terms of inheritance, could aid parentage testing in some cases or kinship analysis such as is performed with identifying mass disaster victims [18]. However, Amorim and Pereira [19] note some unexpected drawbacks of using SNPs in forensic kinship investigations based on simulations. They predicted that a battery of 45 SNPs would produce a higher frequency of cases where statistical evidence would be inconclusive when applied to routine paternity investigations [19].

### Significant SNP disadvantages

Several significant disadvantages exist with SNP markers when considered as a possible replacement for currently used STR loci with the top two being the number of loci needed and the inability to easily decipher mixtures. First, because SNPs are not as polymorphic as STRs, more SNPs are required to reach equivalent powers of discrimination or random match probabilities. Numbers on the order of 40–60 SNPs have been suggested in order to approximate the power of 13–15 STR loci as are commonly in use today [10, 12, 19].

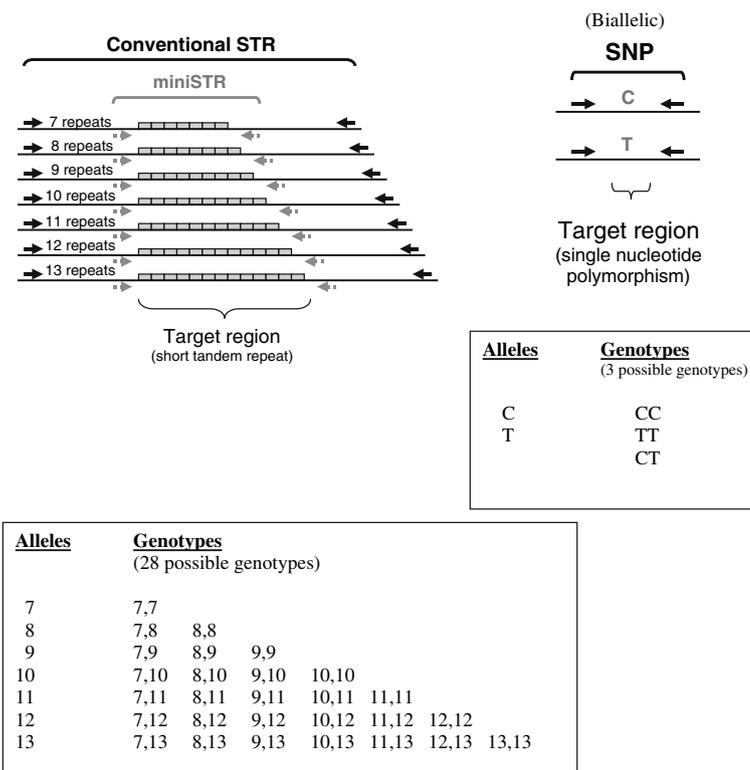
Remember that 15 STRs can be routinely amplified simultaneously in a single multiplex amplification reaction from minimal amounts (e.g., 500 pg) of DNA template using commercially available kits such as PowerPlex 16 and Identifiler. While multiplex PCR amplification of such a large number of SNPs (e.g., ~50) has only recently been demonstrated in a research setting [14], routine production and commercialization of robust assays containing upwards of 100 oligonucleotide PCR primers will not be trivial. Likewise, the expense of examining more loci will be higher.

Perhaps more importantly data interpretation becomes increasingly difficult with more loci and amplification products. Issues with locus drop out will become more significant when three to five times more loci are involved in comparison to traditional STR typing. In addition, assays with a larger number of loci are more sensitive to the quantity and integrity of the input DNA template particularly when trying to amplify limited DNA materials.

Current SNP typing for use in HapMap population studies (e.g., 7) or other clinical genetics projects involve attempting to type many thousands of SNP loci in a relatively small number of samples. If a few dozen or even hundreds of loci fail to produce a result on a sample, these data are excluded from the final analysis or further attempts are made using a replenishable supply of DNA. This type of data loss when attempting to perform a direct comparison between a suspect and evidence is undesirable or even unacceptable under the current paradigm of sample matching performed with STR typing on only a dozen or so loci. Will investigators or the courts care that a fraction of the SNP loci attempted failed to produce a result? With limited amounts of starting material in many situations, there may not be opportunities to repeat the testing in an effort to recover the lost loci. Furthermore, the loci missing on reference samples may be different from those that failed on the evidentiary material leaving even less of an overlap of successfully typed loci for comparison purposes.

When attempting to analyze a greater number of genetic loci, there will be an increased complexity of data to be examined. Depending on the SNP detection platform, the

**Fig. 1** Comparison of (A) STRs and (B) SNPs in terms of the number of possible allele combinations and relative size of the target region



data to be interpreted will vary. With so many loci being typed, software interpretation will become more reliant on expert computer systems as the primary form of data analysis. More loci mean more peak signals and the potential for more artifacts. Thus, detailed data analysis will become more tedious and practically impossible without validated expert systems. Although the argument has been made that data interpretation will be easier with SNPs because they do not possess stutter products or microvariants, in a certain sense these biological amplification artifacts provide greater confidence in results—i.e., that a measured peak with a stutter product is truly an STR allele rather than a spike or a fluorescent dye artifact.

However, from our point of view, the most significant disadvantage of SNP typing is that the limited number of alleles (typically two) for each SNP locus limits or prevents reliable mixture interpretation. A major advantage with STRs in a forensic setting is that many possible alleles exist providing the possibility that the multiple contributors to a mixture will have distinguishable (non-overlapping) alleles. Figure 2 shows SNP and STR typing data obtained on the same mixed DNA sample. While the STR results clearly suggest more than one contributor based on the number of alleles present at multiple loci, an analyst would find it difficult to determine that a mixture is present based on the six SNP loci shown.

Another challenge preventing routine application of SNPs today is that multiple possible detection platforms

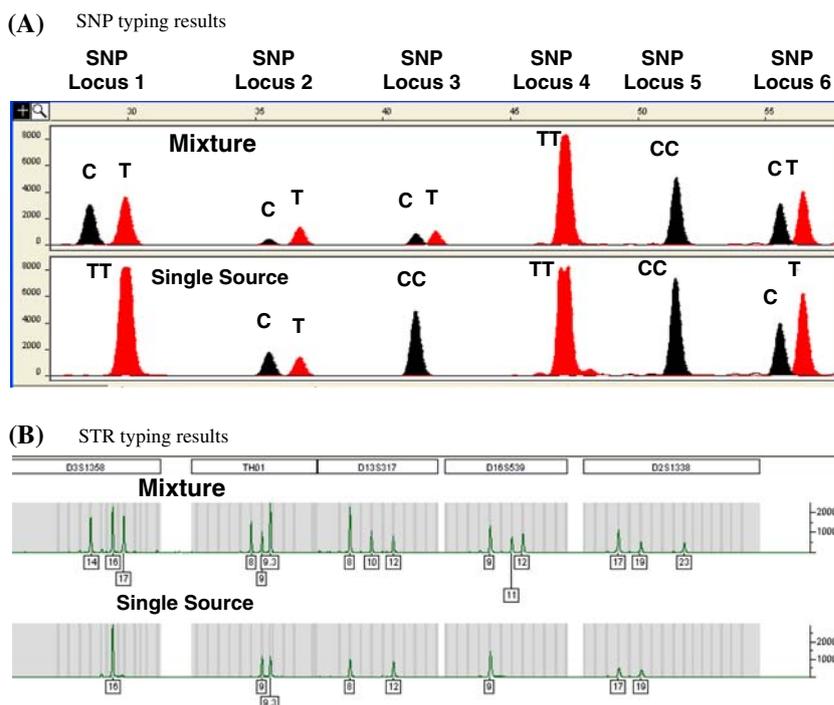
exist for SNP typing [8, 9, 21, 22]. The various chemistries involved are far from being standardized. Without consensus throughout the community on the SNP markers to be employed or the detection platform(s) to be utilized, SNP typing cannot gain a foothold over the dominance of the widely used STR typing systems.

### Likely future role of SNPs

However, SNPs may play a useful role in niche applications such as mitochondrial DNA (mtDNA), Y-SNPs, ancestry informative markers (AIMs), predicting phenotypic traits, and other potential forensic case work applications. Coding region SNPs can fulfill a useful role for separating common HV1/HV2 mitochondrial DNA types [23, 24], and assays have been developed to reliably examine mtDNA coding region SNP variation [25, 26]. While Y-SNPs have limited utility for individualizing a sample, they may, depending on the population(s) of interest, be helpful in aiding estimations of ethnic origin [27, 28].

Predicting ancestry [29, 30] or phenotypic characteristics, such as red hair color [31] or eye color [32], is another role that SNPs may play in the future where the amount of sample is not limited. In these cases, SNP typing could help provide investigators with information about a perpetrator based solely on the biological evidence left at the crime

**Fig. 2** Mixture detection on the same samples with (A) SNPs and (B) STRs. The top panel in each section contains a mixture possessing as one of its components the same DNA that is shown as a single source in the bottom panel. Six different SNP loci (from a total of 70 SNPs evaluated; see [20]) are shown in (A). Note that only SNP locus 2 in the top panel of (A) has an allele imbalance suggesting that a possible second contributor is present. On the other hand in the top panel of (B), all 5 STR loci shown from the green dye channel of the Identifiler kit contain three alleles making mixture detection much easier



scene. Research in these areas is on-going and not yet ready for routine use. Thus, it is important to stress that while we do not think autosomal SNPs will replace autosomal STRs anytime soon, the utility of various SNP assays should still be evaluated.

It will be valuable to understand which SNP markers may be useful when/if the technology platforms catch up to the level acceptable for forensic usage. Since SNP markers are abundant, a larger pool of potential markers is available for testing. The majority of SNPs are bi-allelic so a search for ‘most informative’ loci is not required (conversely a useful STR should have multiple well-populated allelic states). For human identification purposes, loci with low  $F_{ST}$  values and a heterozygosity of >40% are generally adequate for forensic usage [15]. The abundance of SNPs means that ‘poor’ markers can be thrown out since the information content is essentially equivalent between loci (this would be done in the initial selection phase of a SNP panel).

It is important to keep in mind that an infrastructure exists that supports STR typing. Databases with core STR loci contain millions of profiles and are still being populated. More than 10 years of expertise in STR typing is present in many forensic laboratories. A question that should be asked is, “Is it worth the time and cost to convert over to a new marker/technology?” The main proposed benefits of SNPs are the potential for use on degraded samples and possible rapid—high throughout analysis. However, it appears that miniSTRs complete well in the arena of degraded samples [13], and a high throughput

platform/chemistry for SNPs that is robust enough for forensic use (at least as good as STRs) has not been fully developed.

**Summary**

The overall points made in this article are summed up in Table 1. While automation of SNP detection has improved significantly in recent years enabling millions of SNPs to be examined on hundreds of samples in a relatively short period of time (e.g., 7), due to the large number of loci that must be co-amplified and the inability to easily decipher mixtures, we do not feel that SNPs stand on the horizon as future markers (i.e., replacing STRs) for widespread use in forensic DNA testing. That being said, there are and will likely be important

**Table 1** A simple summary comparing the characteristics of STRs and SNPs

Characteristics	STRs/miniSTRs	SNPs
Success with degraded DNA	+	+
Power of discrimination (without significant multiplexing)	+	-
Mixture detection/interpretation	+	-
<i>Other applications:</i>		
Ethnicity estimation, physical traits, etc.	-	+
Commercial testing kits available	+	-
Consistent platform/assay for analysis	+	-

niche roles for SNP typing. We agree with Gill et al. [33] that autosomal SNPs will likely supplement STR results rather than supplant them due to the immense investment already made in national DNA databases.

### Educational message

1. The two primary advantages for SNPs include (a) potential ability to work well on degraded DNA because a small target region can be amplified and (b) lower mutation rates compared to STRs, which could aid kinship testing.
2. Significant disadvantages for SNPs include needing 40–60 loci to obtain equivalent match probabilities as 13–15 STRs commonly used today and the greater difficulty with mixture interpretation due to a limited number of alleles compared to multi-allelic STR markers.
3. SNPs do have a potential future role in aiding investigators with predicting ancestry or phenotypic characteristics although research is still on-going in these areas.

**Acknowledgements** The assistance of Amy Decker with performing SNP typing and Becky Hill with performing STR and miniSTR typing in our laboratory is gratefully acknowledged. This work was funded in part by the National Institute of Justice through interagency agreement 2003-IJ-R-029 with the NIST Office of Law Enforcement Standards.

### References

1. Butler JM. Forensic DNA typing: Biology, technology, and genetics of STR markers. 2nd ed. New York: Elsevier; 2005.
2. Gill P. Role of short tandem repeat DNA in forensic casework in the UK – past, present, and future perspectives. *BioTechniques* 2002;32:366–72.
3. Butler JM. Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci* 2006;51:253–65.
4. National Institute of Justice. The future of forensic DNA testing: predictions of the research and development working group. Nov 2000, Washington, DC; available at <http://www.ojp.usdoj.gov/nij/pubs-sum/183697.htm>
5. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1229–320.
6. Brookes AJ. The essence of SNPs. *Gene* 1999;234:177–86.
7. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. Whole-genome patterns of common DNA variation in three human populations. *Science* 2005;307:1072–9.
8. Budowle B. SNP typing strategies. *Forensic Sci Int* 2004;S139–42.
9. Sobrino B, Brion M, Carracedo A. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 2005;154:181–94.
10. Chakraborty R, Stivers DN, Su Y, Budowle B. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis* 1999;20:1682–96.
11. Krawczak M. Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis* 1999;20:1676–81.
12. Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med* 2001;114:204–10.
13. Dixon LA, Dobbins AE, Pulker HK, Butler JM, Vallone PM, Coble MD, Parson W, Berger B, Grubwieser P, Mogensen HS, Morling N, Nielsen K, Sanchez JJ, Petkovski E, Carracedo A, Sanchez-Diz P, Ramos-Luis E, Brion M, Irwin JA, Just RS, Loreille O, Parsons TJ, Syndercombe-Court, Schmitter H, Stradmann-Bellinghausen B, Bender K, Gill P. Analysis of artificially degraded DNA using STRs and SNPs—results of a collaborative European (EDNAP) exercise. *Forensic Sci Int* 2006;164(1):33–44.
14. Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 2006;27:1713–24.
15. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SL, Karoma NJ, Kungulilo S, Kim JJ, Lu RB, Odunsi A, Okonofua F, Parnas J, Schulz LO, Zhukova OV, Kidd JR. Developing a SNP panel for forensic identification of individuals. *Forensic Sci Int* 2006;164(1):20–32.
16. Butler JM, Shen Y, McCord BR. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J Forensic Sci* 2003;48:1054–64.
17. Coble MD, Butler JM. Characterization of new miniSTR loci to aid analysis of degraded DNA. *J Forensic Sci* 2005;50:43–53.
18. Biesecker LG, Bailey-Wilson JE, Ballantyne J, Baum H, Bieber FR, Brenner C, Budowle B, Butler JM, Carmody G, Conneally PM, Duceman B, Eisenberg A, Forman L, Kidd KK, LeClair B, Niezgoda S, Parsons T, Pugh E, Shaler R, Sherry ST, Sozer A, Walsh A. DNA identifications after the 9/11 World Trade Center attack. *Science* 2005;310:1122–3.
19. Amorim A, Pereira L. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Sci Int* 2005;150:17–21.
20. Vallone PM, Decker AE, Butler JM. Allele frequencies for 70 autosomal SNP loci with U.S. Caucasian, African-American, and Hispanic samples. *Forensic Sci Int* 2005;149:279–86.
21. Gut IG. Automation in genotyping of single nucleotide polymorphisms. *Hum Mutat* 2001;17:475–92.
22. Syvanen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2001;2:930–42.
23. Coble MD, Just RS, O’Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, Parsons TJ. Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. *Int J Legal Med* 2004;118:137–46.
24. Coble MD, Vallone PM, Just RS, Diegoli TM, Smith BC, Parsons TJ. Effective strategies for forensic analysis in the mitochondrial DNA coding region. *Int J Legal Med* 2006;120:27–32.
25. Brandstatter A, Parsons TJ, Parson W. Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med* 2003;117(5):291–8.
26. Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ. A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. *Int J Legal Med* 2004;118:147–57.
27. Vallone PM, Butler JM. Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension. *J Forensic Sci* 2004;49:723–32.

28. Wetton JH, Tsang KW, Khan H. Inferring the population of origin of DNA evidence within the UK by allele-specific hybridization of Y-SNPs. *Forensic Sci Int* 2005;152:45–53.
29. Shriver MD, Kittles RA. Genetic ancestry and the search for personalized genetic histories. *Nat Rev Genet* 2004;5(8):611–8.
30. Frudakis T, Venkateswarlu K, Thomas MJ, Gaskin Z, Ginjupalli S, Gunturi S, Ponnuswamy V, Natarajan S, Nachimuthu PK. A classifier for the SNP-based inference of ancestry. *J Forensic Sci*. 2003;48(4):771–82.
31. Grimes EA, Noake PJ, Dixon L, Urquhart A. Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Sci Int* 2001;122:124–9.
32. Sturm RA, Frudakis TN. Eye colour: portals into pigmentation genes and ancestry. *Trends Genet* 2004;20(8):327–32.
33. Gill P, Werrett DJ, Budowle B, Guerrieri R. 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 (SWGDM). *Sci Justice* 2004;44:51–3.