

Mitochondrial DNA SNP Detection: Design Issues and the Use of the Mass Spectrometer as an Analysis Platform

Bruce Budowle and Danielle Dickinson

FBI Laboratory, 2501 Investigation Parkway, Quantico, VA 22135

Introduction

DNA typing has become widely accepted for the characterization of forensic biological evidence. The current genetic markers used, i.e., predominately short tandem repeat (STR) loci and to a lesser extent mitochondrial DNA (mtDNA), offer high levels of discrimination. In addition, the polymerase chain reaction (PCR) (1), the cornerstone of forensic DNA typing, provides a sensitivity of detection such that exceedingly small samples can be analyzed. Even with the successes encountered using the PCR and forensically validated genetic markers, there are samples that do not contain sufficient DNA or are too degraded to undergo DNA analysis.

Strategies are being sought to type samples containing very minute amounts of DNA. The approach most widely used which typifies the strategy for attempting to analyze very limited amounts of DNA template is known as low copy number (LCN) typing (2-9). LCN typing, usually carried out with STR loci, is the analysis of any results below the stochastic threshold for normal interpretation. Typing can be achieved by increasing the number of PCR cycles, for example from 28 to 34; by reducing the PCR volume; by reducing salt concentration in the sample before capillary electrophoresis; by use of a formamide with a lower conductivity; by adding more amplified product to the denaturant formamide; and/or by increasing injection time (10). However, at these low levels of template DNA (usually less than 100 pg), stochastic effects can and do occur, resulting in either a substantial imbalance of two alleles at a given heterozygous locus or allelic dropout. Even though the assay may not always be reproducible and suffers from allele drop out, as well as allele drop in, LCN has been used for analysis of forensic evidence and for providing investigative leads (2-4, 7-9).

A more robust approach than LCN typing is the interrogation of smaller DNA target regions. To improve success with STR typing of limited and/or degraded DNA, the PCR primers for the STR loci can be repositioned so they reside closer to the repeat region (11-14). Thus, the PCR amplicon(s) is reduced in length. If the amplicon to be generated is smaller than some of the fragmented DNA template molecules,

STR analysis of the degraded sample may then be possible. This approach has been successful for typing telogen hairs (12) and for typing some of the remains from the World Trade Center disaster (15).

To reduce the amplicon size further another class of genetic markers, known as single nucleotide polymorphisms (SNPs), is being pursued. SNPs are single base pair positions in the DNA that are variable within a population(s). For example, the fourth position in the two sequences in Figure 1 might qualify as a SNP. There has been a substitution of a G with an A in the comparison of sequence 1 to sequence 2. Most SNPs are bi-allelic and are more often the result of transitions, either the polymorphism is a G/A or a C/T. Transversions also occur and include C/A, G/T, C/G, and T/A substitutions. Single base insertions and deletions also can be considered SNPs (Figure 2). About 85% of human variation is due to the presence of SNPs, occurring about 1 in every 100 to 300 bases in the population wide human genome (16-19). The SNP Consortium and the National Center for Biotechnology Information maintain databases of several million potential SNPs in the human genome (20-22). There likely are several million SNPs that can differ between any two unrelated individuals; thus there is an abundant supply of SNPs for forensic analyses. However, to be consistent with traditional definitions of genetic polymorphism, to qualify as a legitimate SNP the least common allele should occur at a frequency of at least 1%. The primary criterion that makes SNPs desirable for forensic analyses is that the amplicon size, in theory, can be reduced to 50-80 base pairs in length. Thus, DNA degraded even beyond the limits required for reduced size STR amplicon typing may become typeable.

SNPs are not likely to replace STRs as the primary method for forensic identification testing. They do not have the power of discrimination of STR loci on a locus-by-locus level. To achieve the resolution of identification of the standard CODIS thirteen STR loci substantially more SNPs will be needed (23). It currently is technologically very demanding to develop a robust multiplex assay containing 50 or more SNP loci. Regardless, SNPs will have a niche in forensic analyses for screening samples for mtDNA sequencing or possibly replacing sequencing, for typing degraded or small quantity human remains from missing persons and mass disasters, for determining the geographical origin of the donor of a sample, for lineage-based studies, and for typing physical attributes (24, 25).

A number of issues need to be considered when developing a SNP assay. These include discrimination power, the target sequence (i.e., the flanking area sequence where the primer binds), the

analytical platform, the ability to quantify components of a mixed sample, the capacity for multiplexing, robustness, ease of assay, cost, and existing expertise. All SNP typing methods include amplification, typically by the PCR, of the region containing the SNP of interest. Then, an analytical method is used to detect the SNP site and allelic state (an exception is real-time PCR (26, 27) where the amplification and typing are performed simultaneously). The analytical portion of many SNP assays is typically based on either hybridization of a probe to amplified product, primer extension chemistry, or primer ligation (24, 28-30). Multiplexing is possible, although some technologies may offer a greater capacity for multiplexing than others. In this paper, issues to consider in developing SNP assays are discussed with a focus on mtDNA because of the greater demands on design. In addition, the value of mass spectrometry as a robust platform for SNP analysis is presented.

SNP Assay Limitations

The high copy number in a cell compared with nuclear DNA and its circular nature enable mtDNA to persist longer than nuclear DNA. Thus, a mtDNA typing result may be obtained when no result is possible by analyzing nuclear DNA. This approach has been very successful for typing bones, teeth and single hair shafts (31-39). However, because of its mode of inheritance, lack of recombination, and genetic diversity, mtDNA will never provide the power of discrimination enjoyed by multiplex STR typing. Yet, typing mtDNA SNPs and other variants may provide the greatest level of sensitivity of detection.

Typing mtDNA in forensics is carried out using the Sanger sequencing method followed by electrophoresis and real-time fluorescent detection (38, 40, 41). Sequencing is the ultimate method in SNP analysis, because all SNPs contained within the sequenced fragment are detected. Although the entire mtDNA genome (16,569 bases in length) exhibits substantial variation, the two hypervariable regions (HV1 and HV2) within the mtDNA control region are often sequenced, because of a high density of SNPs and other variants. However, sequencing is a labor intensive process for most crime laboratories. SNP analysis of mtDNA has been demonstrated using a number of alternative formats. These include a reverse dot blot system of immobilized probes on a nylon support (42,43), mini-sequencing using primer extension chemistry (44) or SNaP Shot (24), pyrosequencing (45), and microarray chips (46).

Unlike sequencing, typing specific SNPs with these methods (excluding pyrosequencing) requires *a priori* knowledge of the SNP position so a probe or primer can be designed to detect the specified SNP. The use of probes and/or primers for post PCR mtDNA SNP analysis has three major limitations. First, the majority of the variation for identification purposes does not reside at “signature” SNP sites within the genome. Even with efficient probe and primer design, most of the variants will not be observed. Therefore, the power of mtDNA SNP-based assays cannot approach the levels of discrimination power afforded by sequencing. Second, with mtDNA and particularly so for regions HV1 and HV2, some variant sites cluster and reside close to one another. Thus, SNPs of interest reside proximal to other polymorphic sites. The proximity of variants (and other SNPs) to the SNP of interest will destabilize hybridization of primers and probes with the amplicon template, because of incomplete complementarity (Figure 3a, 3b). One might consider avoiding such regions and instead search for SNPs residing in highly conserved genomic regions. However, that option is very limited with mtDNA because of the limited size of the genome. Alternatively, primer hybridization can be facilitated by use of redundant primers/probes (Figure 3c) or the primers/probes can be chemically modified to effect hybridization (47-49). The third limitation is that hybridization and the primer extension assay with fluorescently-labeled ddNTPs are not effective methodologies for quantitation. Therefore, resolving the components of a forensic mixture sample or quantifying the contribution of the components of a mtDNA heteroplasmic sample can not be performed reliably.

The aforementioned limitations make it difficult to design robust mtDNA SNP detection assays. To be effective, a methodology should approach the power of discrimination of sequencing by scanning a substantial amount of the variation contained within the amplified fragment, not be hampered by SNP and variant clustering, not require *a priori* knowledge of the potentially forensically informative sites within a sample, be quantitative, be amenable to automation, and, if possible, provide the relevant information without the burden of extensive manual data interpretation (as encumbered with sequencing).

Mass Spectrometry

The mass spectrometer is well-suited as an analytical platform for SNP analysis. It is a highly automatable, highly accurate (over several orders of magnitude) instrument that does not require DNA to

be fluorescently labeled for detection (50-70). To be analyzed in a mass spectrometer compounds must first be converted into gas phase ions. Then, the ions are sorted based on their mass-to-charge ratios (m/z), which are directly correlated with the compounds' molecular weight. Mass spectrometers are composed of three basic parts which are: a sample inlet and ionization source; a mass analyzer; and an ion detector (Figure 4). There are numerous permutations for incorporating these components in a mass spectrometer. The sample, whether in solid, liquid, or gaseous form, is first introduced into the vacuum chamber of the ionization region through a sample inlet. Typical sample inlets include the use of a heated capillary inlet or a sample probe which can be inserted into a vacuum lock in the ionization region of the mass spectrometer (71). The sample, if not already in the gas phase, must then be desorbed, and subsequently ionized, prior to mass analysis. The desorption and ionization step can be coupled together and occur simultaneously or they can be performed separately. A beam of electrons, atoms, ions, or photons from a laser can be used for desorption and/or ionization. Ionization is achieved either by inducing a loss or gain of a charge which typically results from the loss or gain of a proton or electron. The gas phase ions are directed through the use of electrostatic lenses into the mass analyzer, where they are sorted based on their m/z ratio. The ions are detected using an electron multiplier, array detection, or ion imaging, and their m/z spectra are recorded (71).

For DNA analysis, the desorption and ionization process must occur without significantly fragmenting the DNA molecules. Therefore, the two ionization sources that are commonly used are matrix assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). In MALDI, the analyte (i.e., DNA) is allowed to co-crystallize with a matrix compound, usually an organic acid that absorbs well in the UV range. The crystallized sample is then irradiated with a laser which induces the desorption and ionization of both the matrix and analyte. The matrix facilitates the conversion of the DNA to a singly-charged gas phase ion without fragmentation. (71, 72) (Figure 5). With ESI, the analyte is introduced as an aqueous solution through a capillary. The application of high voltage to the capillary causes a Taylor cone to form at the tip. The result is the generation of a fine spray of multiply-charged intact DNA ions in the gas phase which can then be desolvated and analyzed (71, 73) (Figure 6).

The main purpose of a mass analyzer is to separate, or resolve, the ions according to their m/z . To date, the most suitable analyzers for analysis of DNA are time-of-flight (TOF) or Fourier Transform ion

cyclotron resonance mass spectrometer (FT-ICR) mass analyzers. A TOF analyzer functions by measuring the time required for an ion to travel to the detector which is located 1-2m from the source. All ions in the source are accelerated through the same potential difference and thus have the same kinetic energy. Ions then separate as they travel to the detector based on their velocity, Page: 6

with smaller ions traveling faster than larger ions and reaching the detector first. Thus, the analyzer is known as time-of-flight because the mass of the ion is directly proportional to the time it takes the ion to reach the detector (71). The resolution and mass accuracy achievable by current benchtop TOF instruments are 10,000 at m/z 1,000 and 5-10 ppm, respectively. The TOF systems are well-suited for rapid scanning of samples. An FT-ICR mass spectrometer operates based on the fact that an ion will precess in a magnetic field at a frequency, which is inversely proportional to its m/z (71). In an FT-ICR mass spectrometer ions are trapped electrostatically within a cell, which is housed in a constant magnetic field. Cyclotron motion is induced through the use of a radio frequency signal to excite the ions. The different frequencies of the ions in the trap are detected as a time-dependent image current. This is Fourier-transformed to derive the component frequencies of the various ions, which enables determination of the m/z of the ion (71).

Because of the accuracy with which frequency can be measured, FT-ICR MS has the highest mass accuracy (<1ppm) of any mass analyzer available and also has superior resolving power (100,000 at m/z 1000). Both analyzers have detection limits with sensitivities of less than a femtomole of material being routinely possible.

SNP Analyses

MALDI-TOF mass spectrometer SNP detection is exemplified by the MassARRAY Homogenous MassEXTEND™ (hME) assay developed by Sequenom (San Diego, CA) (74-76). The assay is based on primer extension and incorporation of ddNTPs (Figure 7). The primer extension allele products, which typically differ by one nucleotide (but could be no more than the mass differences between two primer extension products with the same number of nucleotides), can be readily distinguished by direct mass analysis with the MALD-TOF mass spectrometer. While highly accurate and automatable, the primer extension system is still susceptible to the hybridization issue of clustered SNPs. Thus, this SNP assay is not likely to be used routinely for mtDNA analysis.

The homogeneous MassCLEAVETM (hMC) assay (Sequenom, San Diego, CA) may be an attractive alternative. The approach has been used for discovery of SNPs (77), mutations (78) and identification of mycobacteria (79). It has been applied recently to mtDNA analysis (80). With this approach two amplicons (ranging 300-700 bp in length) are generated. One amplicon has a T7-promoter tag incorporated into its forward strand; while the other generated amplicon differs only in that the T7-promoter tag is incorporated into the reverse strand. After the PCR, the products are transcribed into RNA. Because modified nucleotides are incorporated into the RNA during transcription, base-specific cleavage can occur during incubation with RNase A (Figure 8). The RNA products are divided into two cleavage reactions (one for cleavage at U (for T) residues and one for cleavage at C residues). Thus, four total reactions are carried out. The masses of the resulting cleavage products are measured by MALDI-TOF mass spectrometry, and a profile is generated. The four cleavage reaction patterns are analyzed using combinatorial algorithms and polymorphisms are identified. The clustering of SNPs has no impact on the assay. Also, with this assay almost all variants can be identified, not just signature SNPs.

Base Composition Analysis

Recently, an ESI-MS method called TIGER (Triangulation Identification for Genetic Evaluation of Risks) was developed by Ibis Therapeutics (Carlsbad, CA) to rapidly identify a broad range of microorganisms and monitor in real time the spread of the etiologic agent of a disease (81-83). The basis of TIGER is similar to that of Multi-Locus Sequence Typing (MLST). MLST analyzes, by sequencing, fragments of five to seven house-keeping genes (although more can be typed if desired) common to many bacteria (84, 85). With proper design, many strains within species are distinguishable. However, TIGER does not sequence the fragments as is done with MLST. Instead, the base composition of each fragment is calculated. Using the known exact masses of the four bases that comprise DNA and the accurately measured mass of a strand of DNA from the mass spectrometry analysis, the base compositions of each strand can be calculated unequivocally within certain limitations (i.e., mass accuracy requirement of 1-25 ppm and a DNA fragment length less than 140 bases) (54, 60). A combination of base composition signatures from several loci yields a strain-specific profile.

The same technology also could prove extremely useful for mtDNA SNP and variant analysis and has been demonstrated (86). Consider a mtDNA fragment that is 100 bases in length that contains a number of population wide variant sites. In lieu of sequencing (which reads all variants in the fragment), the base composition of an amplicon is determined. The result will be that the amplified fragment contains a certain number of A, G, C and T residues. For example in Figure 9, two samples have been typed by ESI-MS and the resultant base composition displayed. The only difference between the samples is that at one and only one position in sample 1 an A has been substituted with a G in sample 2. Thus, the transition results in a difference in base composition at A and G, while the number of C and T residues remains the same. Because of the base composition differences, it can be readily determined that the two samples are different. The caveats with this method are that the position of the SNP (or variant) in the sequence is unknown and that in some cases multiple substitutions can possibly cancel out differences. For example, if in sample 1 above there had been a transition from A to G at the 17th position of the sequence and at position 29 a G to A transition also occurred, then the overall base composition would remain unaltered. Thus, while the ESI-MS method of base composition can identify tremendous variation, it cannot always achieve that of sequencing.

The above limitation of base composition analysis would apply to sequences with clustered SNPs and variants, such as occurs with mtDNA, but not to SNPs residing in population wide invariant regions (which may be found within the nuclear genome). Moreover, in theory and with chemically modified primers, some amplicons could be generated during PCR that consist of only the primers and the SNP site. So if each primer were 25 bases in length and both abut the SNP site of interest, the total amplicon size could be as little as 51 base pairs (Figure 10). Even if another variant exists in the primer binding site (and it does not destabilize primer annealing), that variant is not incorporated into the exponentially increasing amplicons and will not impact base composition analysis.

Base composition analysis with ESI-MS simplifies the assay because no post PCR analytical design is required. No probes or primers are needed to detect the variant; only mass differences are needed to obtain typing results. However, there may be situations where the amplicon size is beyond the current computational limits of 140 bases in length. In such cases, restriction digestion of the amplified fragment to sizes lengths less than 140 nucleotides overcomes the limitation. Moreover, a difference of restriction

fragment patterns between two samples is indicative of the presence of a variant (s) as well. Thus, it would seem that ESI-MS is particularly suited for the analysis of the SNPs and variants contained within the mtDNA.

Indeed, Hall et al (86) have demonstrated that ESI-MS and based composition analysis can be used for assessing variation contained within the mtDNA HV1 and HV2 regions from as little as a single hair shaft. The method enables quantitative analysis of a number of clustered SNPs and other variants contained within a mtDNA amplicon without *a priori* knowledge of specific SNP positions. Of the 2754 different mtDNA sequences in the SWGDAM database (87), about 90% could be resolved by base composition alone. The power of haplotype resolution afforded is more than five-fold that of the currently available allele-specific hybridization SNP assays (43). To overcome the limitation of a fragment needing to be less than 140 nucleotides in length, Hall et al (86) developed a restriction enzyme panel (comprised of RsaI, EaeI, HpaII, HpyCH4IV and PacI) to digest HV1 and HV2 amplicons. So it is feasible to amplify relatively large amplicons and still exploit mass spectrometry and base composition analysis.

While the methodology overcomes the aforementioned limitations of other SNP-based assays, there are additional features worth noting. The typical samples analyzed by mtDNA typing are hair, bone, and teeth. These samples can be cleansed of exogenous DNA, and thus the mtDNA extracted can be considered as arising from a single source. However, mtDNA fragments of similar length and base composition have very similar ionization efficiencies; therefore, the relative signal intensities measured from a mixture of fragments can accurately quantify the relative amounts of the contribution of the components in the sample. This quantitative capability holds promise for deconvolving mtDNA mixtures arising from two different individuals (something not reliably performed when using Sanger sequencing). Consider a scenario in which a sample (for example, from a cigarette butt) is composed of DNA from two individuals. The individuals differ at two SNPs, one residing in HV1 and the other in HV2. If the two contributed DNAs are at different concentrations, for instance 70% and 30%, using mass spectrometry the HV1 and HV2 multiple fragments can be assigned unequivocally to different sources. Thus, mixture analysis, similar to the process for STR mixture deconvolution, can be applied to mtDNA and possibly enable routine analysis of forensic evidence beyond single source samples.

The phenomenon heteroplasmy, i.e., the presence of two or more closely-related mtDNA types within an individual, can occur. Heteroplasmy can manifest itself at single positions or as length variation at homopolymeric cytosine stretches that occur in the HV1 and HV2 regions in a notable portion of the population. Length heteroplasmy creates an out of register electropherogram downstream from the homopolymeric region. Thus, a sequence electropherogram is often uninterpretable beyond the homopolymeric region. However, the ESI-MS was able to characterize HV1 and HV2 heteroplasmic length variants. The different mtDNA length variant ions have different m/z ratios and thus are detected independently. In addition, as stated above for mixture deconvolution, the contributions of amplicons containing heteroplasmic length variants in a sample can be quantified accurately relative to one another.

A major bottleneck with sequence analysis occurs during the interpretation phase, which entails a lengthy manual review process by two scientists. Because of the mass accuracy afforded with the mass spectrometer and accompanying software, most profiles can be interpreted automatically within a few seconds. Thus, in addition to the obvious automation MS technology offers for sample processing, throughput is improved by decreasing data analysis time on the back end. Lastly, cost of analysis can be reduced at least an order of magnitude compared with mtDNA sequence analysis.

Conclusion

Over the past twenty years, there have been remarkable improvements in the capabilities to genetically characterize forensic biological evidence. The use of mass spectrometry to analyze SNPs and less frequent variants in small-sized amplicons is another application to facilitate forensic analyses. Compared with more widely employed SNP analysis methods, mass spectrometry has several potential advantages including: high mass accuracy, quantitation, and automation. There have been attempts to investigate mtDNA SNPs outside HV1 and HV2 (in the coding region) to better resolve the most common haplotypes (for example, in Caucasians - 263G, 315.1C) (88, 89). Yet, these still suffer from requiring *a priori* knowledge of the SNP position and cannot identify the majority of forensically-important variation contained within the coding region. The coding region is much larger than that of the control region making it more demanding to sequence the region(s) of interest. The mass spectrometry approach can readily capture such variation with a more simplified design (86). However, because of spectral overlap, only a few

amplicons may be typed simultaneously. The degree of multiplexing amplicons from various regions of the mtDNA genome still needs to be determined.

There are other technologies that may prove as useful, overcome limitations of multiplexing, or serve in concert with the mass spectrometer. These include: the highly parallel processing DNA hybridization microarray format or DNA chip (46, 90-92) in which a resequencing chip for the entire mtDNA genome has been developed (personal communication P.S. Walsh, Affymetrix, Santa Clara, CA); and pyrosequencing (93-95), a sequencing by synthesis reaction that does not require electrophoresis and does not require a SNP primer that must reside on highly polymorphic regions. Andreasson et al (45) have demonstrated the use of pyrosequencing to type polymorphic sites residing within the mitochondrial control region and the coding region from mock and actual casework samples.

Such tools of the molecular biologist will likely provide robust platforms to expand capabilities for analyzing more challenging samples. In addition, they hold promise for automating the processes and reducing costs.

Disclaimer

This is publication number 05-13 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

Acknowledgments

We would like to thank all our colleagues at Ibis Therapeutics (Carlsbad, CA) for their contributions and development of the base composition method for analyzing mtDNA by mass spectrometry. We also would like to thank Sequenom, Inc. (San Diego, CA) for kindly providing the spectrum in figure 8 and helpful comments.

References

1. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. Enzymatic amplification of beta-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia. *Science* 1985; 230:1350-1354.
2. Barbaro A, Falcone G, Barbaro A. DNA typing from hair shaft. *Prog Forens Genet* 2000; 8:523-525.
3. Findlay I, Frazier R, Taylor A, Urquhart A. Single cell DNA fingerprinting for forensic applications. *Nature* 1997; 389:555-556.
4. Gill P. Application of low copy number DNA profiling. *Croat Med J* 2001; 42:229-232.
5. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forens Sci Int* 2000; 112:17-40.
6. Ladd C, Adamowicz MS, Bourke MT, Scherzinger CA, Lee HC. A systematic analysis of secondary transfer. *J Forens Sci* 1999; 44:1270-1272.
7. van Oorschot RA, Jones MK. DNA fingerprints from fingerprints. *Nature* 1997; 387:767.
8. Wiegand P, Kleiber M. DNA typing of epithelial cells after strangulation. *Int J Leg Med* 1997; 110:181-183.
9. Wiegand P, Trubner K, Kleiber M. STR typing of biological stains on strangulation tools. *Prog Forens Genet* 2000; 8:508-513.
10. Budowle B, Hobson DL, Smerick JB, Smith JAL. Low Copy Number - Consideration and Caution. In: Twelfth International Symposium on Human Identification 2001, Promega Corporation, Madison, Wisconsin, 2001. Available: <http://www.promega.com/geneticidproc/ussymp12proc/default.htm>
11. Butler JM, Shen Y, McCord BR. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J Forensic Sci* 2003; 48:1054-1064.
12. Hellmann A, Rohleder U, Schmitter H, Wittig M. STR typing of human telogen hairs – a new approach. *Int J Leg Med* 2001; 114:269-273.
13. Tsukada K, Takayanagi K, Asamura H, Ota H, Fukushima H. Multiplex short tandem repeat typing in degraded samples using newly designed primers for the TH01, TPOX, CSF1PO, and vWA loci. *Leg Med* 2002; 4:239-245.
14. Wiegand P, Kleiber M. Less is more – length reduction of STR amplicons using redesigned primers. *Int J Leg Med* 2001; 114:285-287.
15. Holland MM, Cave CA, Holland CA, Bille TW. Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center attacks. *Croat Med J* 2003; 44(3):264-272.
16. Cooper DN, Smith BA, Cooke HJ, Niemann S, Schmidtke J. An estimate of unique DNA sequence heterozygosity in the human genome. *Hum Genet* 1985; 69:201-205.
17. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet* 2001; 27:234-236.
18. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, et al. The sequence of the human genome. *Science* 2001; 291:1304-1351.
19. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolshy R, Chandona G, Perkins N, et al. Large scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998; 280:1077-1082.
20. Holden AL. The SNP consortium: summary of a private consortium effort to develop an applied map of the human genome. *Biotechniques* 2002; Suppl June:22-26.
21. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; 409:928-933.
22. www.ncbi.nlm.nih.gov
23. Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B. The utility of STR loci beyond human identification: Implications for the development of new DNA typing systems. *Electrophoresis* 1999; 20:1682-1696.
24. Budowle B, Planz J, Campbell R, Eisenberg A. SNPs and microarray technology in forensic genetics: development and application to mitochondrial DNA. *Forens Sci Rev* 2004; 16:22-36.

25. 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). *Science & Justice* 2004; 44(1):51-53.
26. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction products by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Acad Sci USA* 1991; 88:7276-7280.
27. Lee LG, Connell CR, Bloch W. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nuc Acids Res* 1993; 21:3761-3766.
28. Iannone MA, Taylor JD, Chen J, Li, M-S, Rivers P, Slentz-Kesler KA, Weiner MP. Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. *Cytometry* 2000; 39:131-140.
29. Syvanen AC. From gels to chips: minisequencing primer extension for analysis of point mutation and single nucleotide polymorphisms. *Hum Mut* 1999; 13:1-10.
30. Syvanen AC, Aalto-Setälä K, Harju L, Kontula K, Soderlund HA. Primer-guided nucleotide incorporation assay in genotyping of apolipoprotein E. *Genomics* 1990; 8:684-692.
31. Allen M, Engstrom AS, Myers S, Handt O, Saldeen T, von Haeseler A, Paabo S, Gyllensten U. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. *J Forens Sci* 1998; 43:453-464.
32. Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and mitochondrial DNA: applications, debates, and foundations. *Ann Rev Genomics Hum Genet* 2003; 4:119-141.
33. Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K. Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 1994; 6:130-135.
34. Ginther C, Issel-Tarver L, King MC. Identifying individuals by sequencing mitochondrial DNA from teeth. *Nat Genet* 1992; 2:135-138.
35. Holland MM, Fisher DL, Mitchell LG, Rodriguez WC, Canik JJ, Merrill CR, Weeden VW. Mitochondrial DNA sequence analysis of human skeletal remains: Identification of remains from the Vietnam War. *J Forens Sci* 1993; 38:542-553.
36. Lutz S, Weisser HJ, Heizmann J, Pollak S. MtDNA as a tool for identification of human remains. *Int J Leg Med* 1996; 109:205-209.
37. Pfeiffer H, Huhne J, Ortmann C, Waterkamp K, Brinkmann B. Mitochondrial DNA typing from human axillary, pubic and head hair shafts - success rates and sequence comparisons. *Int J Leg Med* 1999; 112:287-290.
38. Sullivan KM, Hoggood R, Gill P. Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *Int J Leg Med* 1992; 105:83-86.
39. Wilson MR, DiZinno JA, Polansky D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int. J. Leg. Med.* 1995; 108:68-74.
40. Budowle B, Smith JA, Moretti T, DiZinno J. *DNA Typing Protocols: Molecular Biology and Forensic Analysis*, BioTechniques Books, BioForensic Sciences Series, Eaton Publishing, Natick, MA, 2000.
41. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977; 74:5463-5467.
42. Divne AM, Nilsson M, Calloway C, Reynolds R, Erlich H, Allen M. Forensic casework analysis using the HVI/HVII mtDNA linear array assay. *J Forensic Sci.* 2005; 50:548-554.
43. Gabriel MN, Calloway CD, Reynolds RL, Primorac D. Identification of human remains by immobilized sequence-specific oligonucleotide probe analysis of mtDNA hypervariable regions I and II. *Croat Med J* 2003; 44:293-298.
44. Tully G, Sullivan KM, Nixon P, Stones RE, Gill P. Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. *Genomics* 1996; 34:107-113.
45. Andréasson H, Asp A, Alderborn A, Gyllensten U, Allen M. Mitochondrial sequence analysis for forensic identification using pyrosequencing technology; *Biotechniques* 2002; 32:124-133.
46. Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SPA. Accessing genetic information with high-density DNA arrays. *Science* 1996; 274:610-614.

47. Loakes D, Brown DM, Linde S, Hill F. 3-Nitropyrrole and 5-nitroindole as universal bases in primers for DNA sequencing and PCR. *Nucleic Acids Res* 1995; 23:2361-2366.
48. Sala M, Pezo V, Pochet S, Wain-Hobson S. Ambiguous base pairing of the purine analogue 1-(2-deoxy-beta-D-ribofuranosyl)-imidazole-4-carboxamide during PCR. *Nucleic Acids Res* 1996; 24:3302-3306.
49. Van Aerschot A, Rozenski J, Loakes D, Pillet N, Schepers G, Herdewijn P. An acyclic 5-nitroindazole nucleoside analogue as ambiguous nucleoside. *Nucleic Acids Res* 1995; 23:4363-4370.
50. Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J, Little DP, Strausberg R, Koester H, Cantor CR, Braun, A. High-throughput development and characterization of a genome wide collection of gene-based single nucleotide polymorphism markers by chip- based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc Natl Acad Sci USA* 2001; 98:581-584.
51. Doktycz M J, Hurst G B, Habibi-Goudarzi S, McLuckey S A, Tang K, Chen C H, Uziel M, Jacobson K B, Woychik R P, Buchanan MV. Analysis of polymerase chain reaction-amplified DNA products by mass spectrometry using matrix-assisted laser desorption and electrospray: current status. *Anal Biochem* 1995; 230:205-214.
52. Haff LA, Smirnov IP. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Res* 1997; 7:378-388.
53. Hannis JC, Muddiman DC. Accurate characterization of the tyrosine hydroxylase forensic allele 9.3 through development of electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* 1999;13:954-962.
54. Hofstadler SA, Sannes-Lowery KA, Hannis JC. Analysis of nucleic acids by FTICR MS. *Mass Spectrom Rev* 2005; 24:265-285.
55. Huber CG, Oberacher H. Analysis of nucleic acids by on-line liquid chromatography-mass spectrometry. *Mass Spectrom Rev* 2001; 20:310-343.
56. Hurst GB, Doktycz MJ, Vass AA, Buchanan MV. Detection of bacterial DNA polymerase chain reaction products by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 1996;10:377-382.
57. Kirpekar F, Nordhoff E, Larsen LK, Kristiansen K, Roepstorff P, Hillenkamp F. DNA sequence analysis by MALDI mass spectrometry. *Nucleic Acids Res* 1998; 26:2554-2559.
58. Krahmer M T, Johnson YA, Walters J J, Fox KF, Fox A, Nagpal, M. Electrospray quadrupole mass spectrometry analysis of model oligonucleotides and polymerase chain reaction products: determination of base substitutions, nucleotide additions/deletions, and chemical modifications. *Anal Chem* 1999; 71:2893-2900.
59. Meng Z, Simmons-Willis TA, Limbach PA. The use of mass spectrometry in genomics. *Biomol Eng* 2004; 21:1-13.
60. Muddiman DC, Anderson GA, Hofstadler SA, Smith RD. Length and base composition of PCR-amplified nucleic acids using mass measurements from electrospray ionization mass spectrometry. *Anal Chem* 1997; 69:1543-1549.
61. Muddiman DC, Null AP, Hannis JC. Precise mass measurement of a double-stranded 500 base-pair (309 kDa) polymerase chain reaction product by negative ion electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* 1999; 13:1201-1204.
62. Muddiman DC, Wunschel DS, Liu CL, Pasatolic L, Fox KF, Fox A, Anderson GA, Smith RD. Characterization of PCR products from bacilli using electrospray ionization FTICR mass spectrometry. *Anal Chem* 1996; 68:3705-3712.
63. Naito Y, Ishikawa K, Koga Y, Tsuneyoshi T, Terunuma H. Molecular mass measurement of polymerase chain reaction products amplified from human blood DNA by electrospray ionization mass spectrometry. *Rapid Comm Mass Spect* 1995; 9:1484-1486.
64. Nordhoff E, Kirpekar F, Karas M, Cramer R, Hahner S, Hillenkamp F, Kristiansen K, Roepstorff P, Lezius A. Comparison of IR- and UV-matrix-assisted laser desorption/ionization mass spectrometry of oligodeoxynucleotides. *Nucleic Acids Res* 1994; 22:2460-2465.
65. Nordhoff E, Kirpekar F, Roepstorff P. Mass spectrometry of nucleic acids. *Mass Spe rev* 1996; 15:76-138.
66. Oberacher H, Oefner PJ, Holzl G, Premstaller A, Davis K, Huber CG. Re-sequencing of multiple single nucleotide polymorphisms by liquid chromatography-electrospray ionization mass spectrometry. *Nucleic Acids Res* 2002; 30(14):e67.
67. Oberacher H, Wellenzohn B, Huber CG. Comparative sequencing of nucleic acids by liquid chromatography-tandem mass spectrometry. *Anal Chem* 2002; 74:211-218.
68. Ragas JA, Simmons TA, Limbach PA. A comparative study on methods of optimal sample preparation for the analysis of oligonucleotides by matrix-assisted laser desorption/ionization mass spectrometry. *Analyst* 2000; 125:575-581.

69. Sauer S, Lechner D, Berlin K, Lehrach H, Escary JL, Fox N, Gut IG. A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nuc Acids Res* 2000; 28:e13.
70. Wunschel DS, Muddiman DC, Fox KF, Fox A, Smith RD. Heterogeneity in *Bacillus cereus* PCR products detected by ESI-FTICR mass spectrometry. *Anal Chem* 1998; 70:1203-1207.
71. Watson J-T. Introduction to Mass Spectrometry, 3rd Edition, Lippincott-Raven Publishers, Philadelphia, PA, 1977.
72. Karas, M., Hillenkamp, F Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons. *Anal Chem* 1988; 60:2299-3301.
73. Fenn JB, Mann M, Meng CK, Wong SK, Whitehouse CM. ESI of large biomolecules. *Science* 1989; 246: 64-71.
74. Rodi, CP, Darnhofer-Patel B, Stanssens P, Zabeau M, van den Boom D. A strategy for rapid discovery of disease markers using the MassARRAY system. *BioTechniques* 2002; 32:S62-S69.
75. Storm N, Darnhofer-Demar B, van den Boom D, Rodi, CP. MALDI-TOF Mass Spectrometry-Based SNP Genotyping. *Meth Mol Biol* 2002; 212:241-262.
76. New MassARRAY™ Homogenous MassEXTEND™ (hME) Assay Sequenote Bulletin 1021
77. Ehrich M, Correll D, van den Boom D.. SNP discovery using the MassARRAY™ System. Sequenom 2000; Application Note Doc. No. 8876-004.
78. Honisch C, Raghunathan A, Cantor CR, Palsson BO, van den Boom D. High-throughput mutation detection underlying adaptive evolution of *Escherichia coli*-K12. *Genome Res* 2004; 14: 2495–2502 .
79. Lefmann M, Honisch C, Bocker S, Storm N, von Wintzingerode F, Schlotelburg C, Moter A, van den Boom D, Gobel UB. Novel mass spectrometry-based tool for genotypic identification of mycobacteria. *J Clin Microbiol* 2004;42:339-346.
80. Ehrich M, Nelson M, Turner J, Langdown M, and van den Boom D. High-throughput sequence analysis of mitochondrial DNA from major human haplogroups using base-specific cleavage and MALDI TOF mass spectrometry. Presented at the American Society of Human Genetics meeting, Los Angeles, CA, 2004.
81. Ecker DJ, Sampath R, Blyn LB, Eshoo MW, Ivy C, Ecker JA, Libby B, Samant V, Sannes-Lowery KA, Melton RE, Russell K, Freed N, Barrozo C, Wu J, Rudnick K, Desai A, Moradi E, Knize DJ, Robbins DW, Hannis JC, Harrell PM, Massire C, Hall TA, Jiang Y, Ranken R, Drader JJ, White N, McNeil JA, Crooke ST, Hofstadler SA. Rapid identification and strain-typing of respiratory pathogens for epidemic surveillance. *Proc Natl Acad Sci USA* 2005; 102:8012-8017.
82. Hofstadler SA, Sampath R, Blyn LB, Eshoo MW, Hall TA, Jiang Y, Drader JJ, Hannis JC, Sannes-Lowery KA, Cummins LL, Libby B, Walcott DJ, Schink A, Massire C, Ranken R, White N, Samant V, McNeil JA, Knize D, Robbins D, Rudnik K, Desai A, Moradi E, Ecker DJ. TIGER: the universal biosensor. *Int J Mass Spect* 2005; 242:23-41.
83. Sampath R, Hofstadler SA, Blyn LB, Eshoo MW, Hall TA, Massire C, Levene HM, Hannis JC, Harrell PM, Neuman B, Buchmeier MJ, Jiang Y, Ranken R, Drader JJ, Samant V, Griffey RH, McNeil JA, Crooke ST, Ecker DJ. Rapid identification of emerging pathogens: coronavirus. *Emerg Infect Dis* 2005; 11:373-379.
84. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 1998; 95:3140-3145.
85. Spratt BG. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the Internet. *Curr Opin Microbiol* 1999; 2:312-316.
86. Hall TA, Budowle B, Jiang Y, Blyn L, Eshoo M, Sannes-Lowery KA, Sampath R, Drader JJ, Hannis JC, Harrell P, Samant V, White N, Ecker DJ, Hofstadler SA. Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: A novel tool for the identification and differentiation of humans. *Anal Biochem* 2005; 344:53-69.
87. Monson KL, Miller KWP, Wilson MR, DiZinno JA, Budowle B. The mtDNA population database: an integrated software and database resource for forensic comparison. *Forens Sci Comm* April 2002; 4(2): Available: <http://www.fbi.gov/hq/lab/fsc/current/index.htm>.
88. 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 Leg Med* 2004; 118:137-146.
89. 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 Leg Med* 2004; 118:147-157.
90. Fodor SPA, Rava RP, Huang XC, Pease AC, Holmes CP, Adams CL. Multiplexed biochemical assays with biological chips. *Science* 1993; 364:555-556.

91. Fodor SPA, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science* 1991; 251:767-773.
92. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat Genet* 1999; 1 suppl:20-24.
93. Ronaghi M. Pyrosequencing for SNP genotyping. *Methods Mol Biol* 2003; 212:189-195.
94. Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 1996; 242:84-89.
95. Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. *Science* 1998; 281:363:365.

Figure 1. Two sequences that differ at one SNP site. The allele (residing at the fourth position of the sequence) is biallelic for A/G transition.

CAAGCTTAGG -- Sequence 1

CAAACTTAGG -- Sequence 2

Figure 2. Compared with the Reference sequence, the three types of SNPs are displayed: substitution, deletion, and insertion.

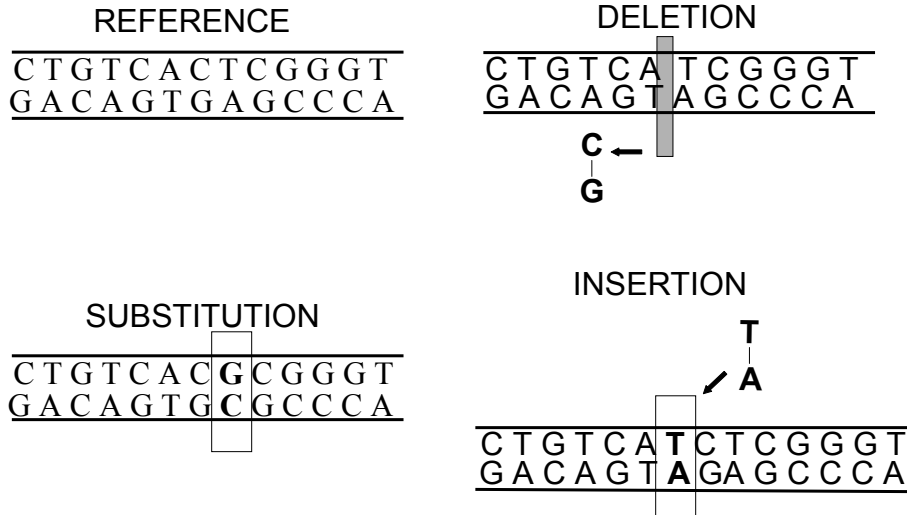


Figure 3c. For amplicon 2* to be detected an additional probe must be used. The number of probes needed is related to the number of SNP sites and alleles per site contained within the amplicon region in the population.

1 - GATCCGTGAA 1 - GATCCGTGAA
 1 - GATCCGTGAA 2* - GATTCATGAA
 2 - GATCCATGAA 1 - GATCCGTGAA
 1 - GATCCGTGAA 2* - GATTCATGAA

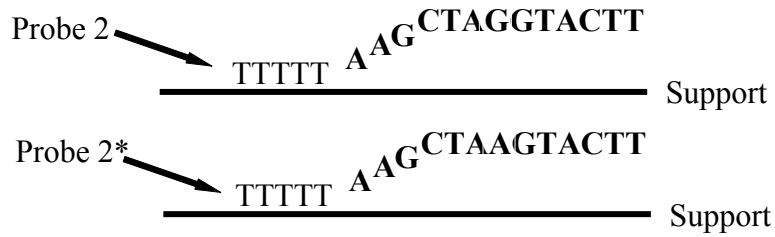


Figure 4. Basic design of a mass spectrometer.

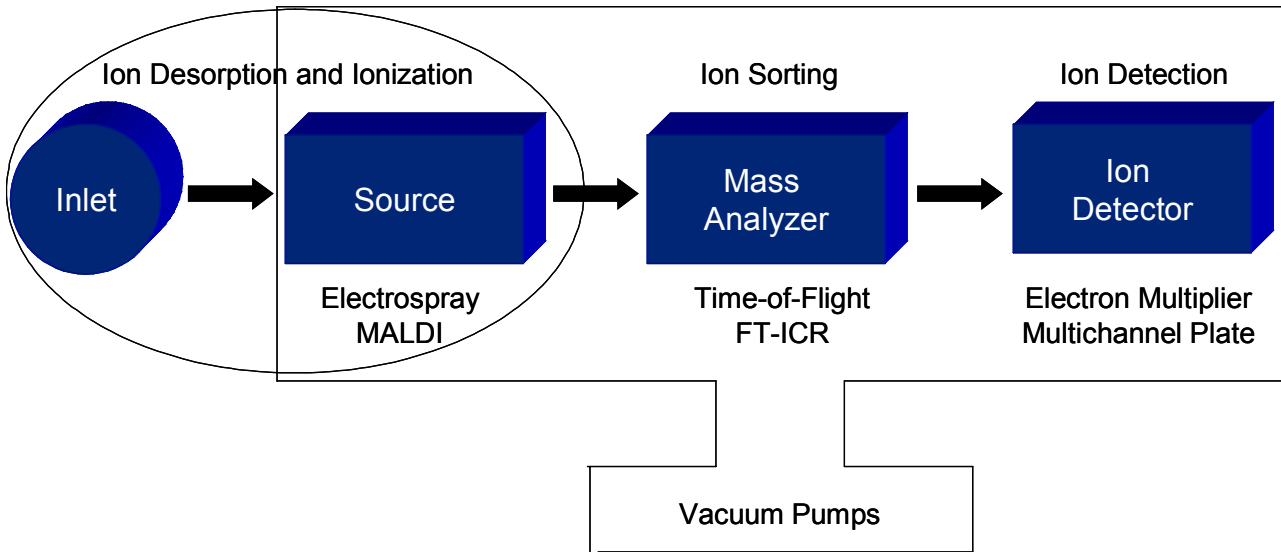


Figure 5. A schematic representation of the ESI process. In ESI, the analyte is introduced as a liquid through a capillary needle. A Taylor cone is formed at the tip of the needle due to the application of a high voltage. The result is a fine mist of tiny, highly charged particles being sprayed into the mass spectrometer.

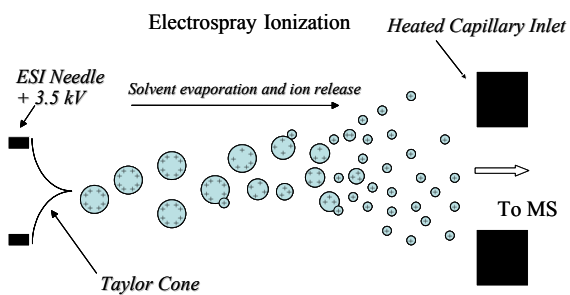


Figure 6. A schematic representation of the MALDI process. In MALDI, analyte molecules (in red) are allowed to co-crystallize with an organic acid matrix (in blue). The irradiation of the matrix using a UV laser allows for the desorption and ionization of the analyte without significant fragmentation.

Matrix Assisted Laser Desorption Ionization (MALDI)

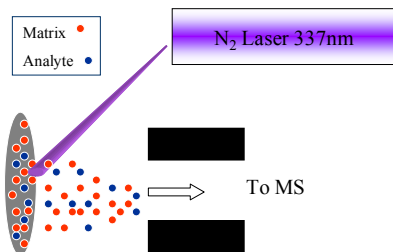


Figure 7. Primer extension assay (or minisequencing).

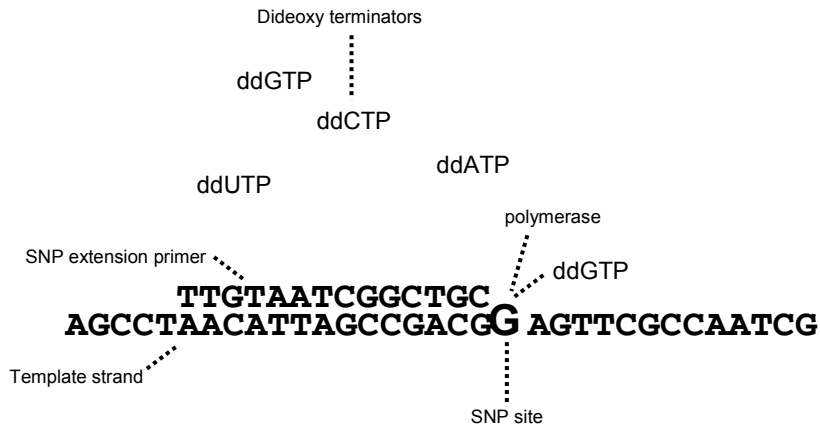


Figure 8. Diagram of protocol for homogeneous MassCLEAVETM assay. The steps in the process are amplification by PCR (incorporation of T7 promoters), transcription (incorporation of modified bases), cleavage of RNA, and mass spectrometry analysis of cleavage fragments. The spectrum kindly provided by Sequenom, Inc. (San Diego, CA).

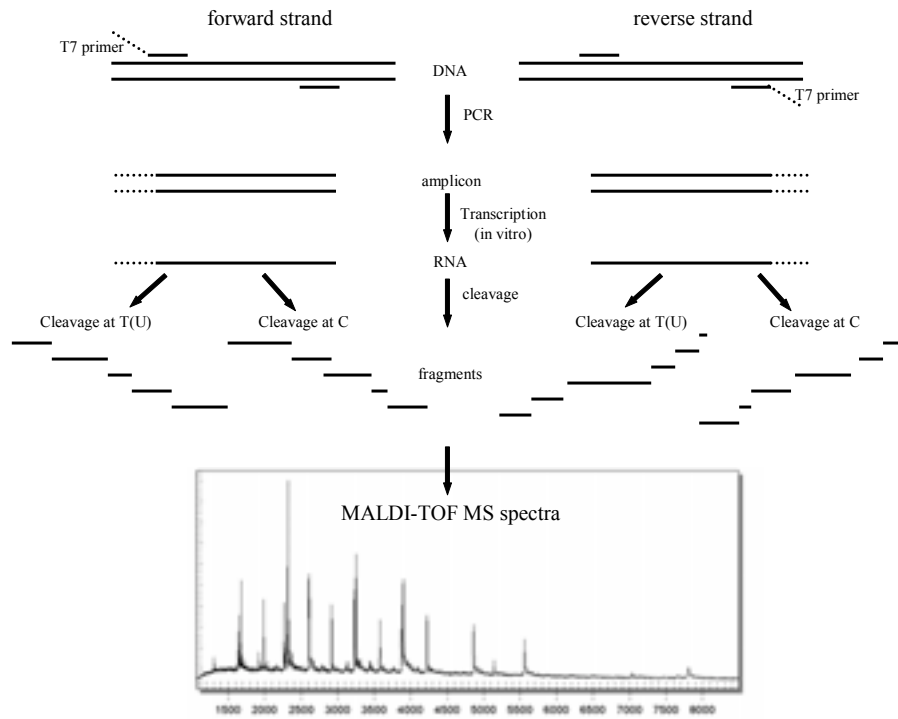


Figure 9. Display of base composition of one strand of DNA duplex for two samples. While the SNP position is unknown, the presence of the SNP and its character state between the two samples are evident.

Sample 1 --- A-24, G-30, C-18, T-28

Sample 2 --- A-23, G-31, C-18, T-28

Figure 10. In theory, an amplicon size could be as short as the length of the primers plus the SNP site. In this case 51 base pairs.

