A new MALDI-TOF based mini-sequencing assay for genotyping of SNPs

Xiyuan Sun, H. Ding, K. Hung and Baochuan Guo

INTRODUCTION

Currently, there is a critical need to develop high-throughput, low-cost, accurate methods for genotyping of single nucleotide polymorphisms (SNPs) (1). The matrix-assisted-laser-desorption-ionization time-of-flight (MALDI-TOF) mass spectrometric based technique represents a new approach to SNP genotyping, followed by mini-sequencing in the presence of three ddNTPs and the fourth nucleotide in the deoxy form. In this way, the primary is extended by only one base from one allele, while it is typically extended by two bases from another allele. The products are then analyzed using MALDI-TOF mass spectrometry. The genotype of the SNP site is identified based on the number of nucleotides added. This assay has been examined using both synthetic and genomic DNA samples. In addition, multiplexed assays were successfully performed to genotype four SNP sites in a single tube. The main aspect of this assay is that it can overcome the key problems associated with the currently used mini-sequencing methods. First, it significantly reduces the stringent high-resolution and extensive desalting requirements that are essential to the pinpoint assay. Second, it avoids the long extension problem associated with the PROBE assay.

MATERIALS AND METHODS

A MALDI-TOF based mini-sequencing approach typically consists of four major steps. They are: (i) extraction of genomic DNA from blood or tissue samples; (ii) PCR amplification of
specific fragments of genomic DNA containing the SNP site(s); (iii) mini-sequencing of PCR products; and (iv) analysis of mini-sequencing products using MALDI-TOF, respectively. In addition, PCR products are treated prior to mini-sequencing to remove the unreacted PCR primers and dNTPs, and the mini-sequencing products will be desalted.

Materials
Shrimp alkaline phosphatase, Exonuclease I and ThermoSequenase were purchased from Amersham Pharmacia Biotech (Piscataway, NJ). Taq DNA polymerase was obtained from Promega (Madison, WI). The synthetic DNA templates and primers used in PCR and mini-sequencing were obtained from either Sigma Genosys (Woodlands, TX) or Research Genetics (Huntsville, AL).

Templates and PCR amplification
The model SNP systems used in this work are synthetic DNA containing the variation site at the second base of codon 12 of k-ras. The human genomic DNA sample was used for the genotyping of other SNP sites. PCR was performed using a thermal cycler (PCR-2400, PE Applied Biosystems, Foster City, CA). The 20 µl mixture solutions were prepared with the buffer provided by the manufacturer. Thereafter, the PCR products were treated with alkaline phosphatase and exonuclease I to destroy residual dNTPs and PCR primers. One unit of alkaline phosphatase and exonuclease I were added to the PCR tube, respectively. Then, the tube was incubated at 37°C for 30 min, followed by heating the tube at 80°C for 15 min to deactivate alkaline phosphatase and exonuclease I.

Mini-sequencing and ethanol precipitation desalting
Mini-sequencing was typically performed in 20 µl using the same thermal cycler. The mixture solutions were prepared with the buffer provided by the manufacturer. 5 pmol of primers were used for all mini-sequencing reactions. Thereafter, 2 µl of the ammonium acetate solution (10 M) was added to the mini-sequencing reaction tube. After incubation of several minutes at room temperature, 4–6 volumes of ethanol were added, followed by incubating the tube in a freezer at ~20°C for ~30 min. Then, the tube was centrifuged, followed by removing the supernatant solution. The precipitated DNA was washed with 70% cooled ethanol solution. Finally, 5 µl of water was added to re-dissolve the precipitated DNA.

MALDI-TOF analysis
The MALDI sample was prepared by mixing 0.5 µl of the purified mini-sequencing products and 0.5 µl matrix (saturated 3-hydroxycolinic acid in a 1:1:2 mixture of water, acetonitrile and 0.1 M ammonium citrate). The sample was dried and then analyzed using MALDI-TOF (Dynamo, Thermo Bioanalysis, Santa Fe, NM). The negative ion mode was used to collect all spectra.

RESULTS AND DISCUSSION
A biallelic SNP marker yields three different genotypes. For example, an A/G marker could produce the genotype of A homozygote, G homozygote and A/G heterozygote, respectively. Figure 1 displays a schematic representation of the VSET method for genotyping of a model SNP (A/G) system.
sequence of k-ras gene (GenBank ID: M54968). The SNP site is the second base of codon 12, at which the nucleotide was A, C, G and T, respectively. The mini-sequencing primer is 5′-tcaaggcacttgctacgcca-3′, as shown in Figure 1. Heterozygous samples were created by mixing equimolar two oligonucleotides that were different only at the second base of codon 12. Figure 2a–c displays the result of typing three different genotypes arising from A and G targets in the presence of ddATP, ddCTP, ddGTP and dTTP. As expected, the homozygous G target yields one-base extension products by the addition of ddC, while the homozygous A target generated two-base extension products by the addition of dT and ddC. In contrast, the heterozygous A/G target produced both one and two-base extension products.

Figure 2d–f display the result of typing three different genotypes of the C and T target samples, in the presence of ddTTP, ddGTP, ddATP and dGTP. As expected, the homozygous T samples produced only one-base extension products by the addition of ddA, while the homozygous C sample yielded only two-base extension products by adding dG and ddC. In contrast, both one- and two-base extension products were produced if the heterozygous C/T target samples were used, as seen from Figure 2f. Comparing the expected extension patterns with Figure 2, it was seen that all six different targets were correctly identified.

It should be noted that the minor peaks labeled with an asterisk in Figure 2a and d were due to unterminated products. In other words, the primers were extended by the addition of a dNTP, but they were not completely terminated by a ddNTP. As seen from Figure 2c and f, two extension products yielded from a heterozygote sample would have similar peak intensities and thereby, the genotype (homozygote or heterozygote) of a sample can be easily identified simply by comparing the peak intensity of products. For example, because the intensity of the peak labeled with an asterisk in Figure 2a was significantly weaker than that of another extended products, we can easily, therefore, conclude that the product corresponding to this
Minor peak was not terminated by a ddNTP and that the sample typed was homozygous.

Next, we examined the feasibility of VSET in genotyping of SNPs with double-stranded DNA amplified from PCR as the DNA templates. Figure 3 displays the result of typing three different genotypes associating with an A/T model system. Single-strand synthetic model templates (either homozygous or heterozygous targets of A/T) were first amplified by PCR, then mini-sequencing was performed in the presence of ddCTP, ddTTP, ddGTP and dATP, respectively. As a result, the A target is expected to produce one-base extension products, while the T target yields two-base extension products. The A/T heterozygous sample produces both one- and two-base extension products. As seen from Figure 3, the primers were clearly extended according to the genotype and three genotypes of A/T were unambiguously identified.

It should be noted that the primers are typically extended by either one or two bases in VSET. If the template sequence contains a run of the nucleotide that is the same as the base in the SNP site, and it immediately follows the SNP site, the primer may be extended by more than two bases. In this case, several measures can be taken. First, the combination of ddNTPs and dNTP should be alternated and this change will limit extension to either one or two bases. Second, a new primer can be designed to target another strand. Third, the same reaction conditions are utilized and the primer is allowed to extend by three or four bases as long as the extension products will not lead to the overlapping problem.

Figure 2. (Previous page and above) MALDI-TOF mass spectra of the results arising from genotyping the model systems: (a) A target; (b) G target; (c) A/G target; (d) C target; (e) T target; (f) C/T target. Note that the minor peaks labeled with an asterisk were due to unterminated products.
After the feasibility study, we evaluated the performance of VSET using human genomic DNA samples which were kindly provided by Professor Chakravarti of Case Western Reserve University. A total of five different SNP sites were examined and their ID/SNP sites in GenBank were AJ243297/17274, AJ243297/20297, AJ243297/20547, AJ243297/25187 and AJ243297/29366, respectively. We found that VSET was able to correctly identify the genotype of all the samples we have examined thus far. In addition, we have also successfully used VSET to identify the G → A mutation of the Factor V Leiden gene in 27 people. This part of our study will be the subject of future publication.

For comparison, both PINPOINT and PROBE were also evaluated using these SNP samples. The results of genotyping an individual of A/T heterozygote (AJ243297/29366) using VSET, PINPOINT and PROBE were displayed in Figure 4a, b and c, respectively. A 540 bp DNA fragment containing this SNP site was first amplified with PCR, followed by mini-sequencing. All three mini-sequencing methods utilized the same mini-sequencing primer 5'-taagaaaggttgcatgatt-3'. The partial sequence of the template containing the SNP site is 5'-tcttgagaatgattttttttttaaatcatgcaacctttcctta-3', the base labeled by the bold letter is the SNP site.

For VSET, mini-sequencing was performed in the presence of ddATP, ddCTP, ddGTP and dTTP. For PINPOINT, mini-sequencing was performed in the presence of all four ddNTPs. For PROBE, mini-sequencing was performed in the presence of dATP, dCTP, dGTP and ddTTP. All other conditions and the DNA sample used were also the same. Two extension products were generated from PINPOINT, where one allele led to one-base extension by adding ddA, while another resulted in one-base extension by adding ddT. It was seen from Figure 4b that these two extension products were poorly resolved and unassignable due to the small mass difference of 9 Da between
ddA and ddT. In contrast, the A/T genotype was easily and unambiguously identified based on Figure 4a since two extension products obtained from VSET differs by one base.

In PROBE, the mini-sequencing primers were extended by 1 and 10 bases, respectively, for this A/T heterozygous sample. It was seen from Figure 4c that the peak corresponding to 10-base extension products was barely seen and that its intensity was more than seven times weaker than that corresponding to one-base extension products. The weaker ion intensity for long extension products was also seen in another work (12). In contrast, two extension products obtained from VSET yield comparable ion intensities. The weaker intensity for long extension products can be caused by two factors. First, the extension efficiency may be poorer for longer extension. Second, MALDI-TOF tends to have poorer detection sensitivity for larger oligonucleotides (26). One may argue that an increase of the polymerase concentration would improve extension, but this significantly increases costs and reduces the specificity of mini-sequencing.

Finally, we studied the feasibility of using VSET for multiplex genotyping. Four SNP sites (AJ243297/17247, AJ243297/20297, AJ243297/20547 and AJ243297/25187) were examined. For AJ243297/20297 and AJ243297/25187, the sense primers were designed to target the anti-sense strand templates, while for AJ243297/19274 and AJ24397/20254, the anti-sense primers were designed to target the sense strand templates. As a result, the targeted bases at all four SNP sites was G/A. In this experiment, the PCR amplification was first performed separately for each sample, then these double-stranded PCR products were pooled together for mini-sequencing. In addition, mini-sequencing was performed in 30 µl.

Figure 4. MALDI-TOF mass spectra of the results arising from genotyping an individual of A/T heterozygote using (a) VSET; (b) PINPOINT; and (c) PROBE, respectively.
A combination of ddATP, ddCTP, ddGTP and dTTP was used in mini-sequencing. The sequence of primers and their expected extension products are listed in Table 1. As will be discussed later, the ideal size of the primers for multiplexed genotyping is 15–27 bases in length. Therefore, the four primers used in this test were 15, 20, 24 and 27 bases in length, respectively. The portion of the primer sequence underlined in Table 1 acted as mass tags and the remaining sequences targeted the templates. Figure 5 displays the result of the 4-fold multiplexed VSET assay. It was seen that all primers were

<table>
<thead>
<tr>
<th>Primers</th>
<th>GenBank ID/ SNP site</th>
<th>Mass of primer (Da)</th>
<th>Sequence (5′→3′)</th>
<th>Genotype</th>
<th>Expected extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>AJ243297/19274</td>
<td>4772</td>
<td>AGACTCCCTGCCCCA</td>
<td>A/G</td>
<td>A (dTddG) G (ddC)</td>
</tr>
<tr>
<td>P2</td>
<td>AJ243297/20297</td>
<td>5923</td>
<td>TTTACACCCCTGACCCCA</td>
<td>G</td>
<td>G (ddC)</td>
</tr>
<tr>
<td>P3</td>
<td>AJ243297/20547</td>
<td>7124</td>
<td>CTGTAGGAAAAATAGCAGGAT</td>
<td>A/G</td>
<td>A (dTddG) G (ddC)</td>
</tr>
<tr>
<td>P4</td>
<td>AJ243297/25187</td>
<td>8500</td>
<td>TTTTTTTTTCTGACGAGTGCCCA</td>
<td>A/G</td>
<td>A (dTddG) G (ddC)</td>
</tr>
</tbody>
</table>

**Figure 5.** MALDI-TOF mass spectra of the results arising from genotyping four different SNP sites in a single tube. Note that the peaks labeled with an asterisk were due to the impurities present with the primers.
extended according to the expected extension pattern and therefore the genotypes of all four SNP sites were unambiguously identified based on Figure 5. The minor peaks labeled with an asterisk in Figure 5 were due to impurities present in the primers provided by the manufacturers. It should be noted that these primers were not purified when they were ordered and that they can be removed using HPLC and PAGE purification.

In this work, we reported an alternative MALDI-TOF based mini-sequencing method for typing SNPs. Compared with PINPOINT, VSET greatly relaxes the stringent mass resolution requirements since the bases at a SNP site are identified based on the number of nucleotides added. As a result, the salt effect on the assignment of extended products is also minimized and therefore, simple and cost-effective desalting methods can be utilized. In this work, mini-sequencing products were exclusively purified using ethanol precipitation. Compared with other methods, ethanol purification costs virtually nothing, therefore, significantly reduces the costs associated with DNA purification.

Compared with PROBE, VSET leads to shorter extension. Extension can be terminated by three of the four ddNTPs in VSET, while extension can only be terminated by one ddNTP in PROBE. Therefore, VSET statistically produces much smaller extension products. As shown in Figure 5, long extension could lead to a major problem for multiplex genotyping since the extension efficiency varies significantly with the SNP sites and samples. In other words, when multiple SNP sites were genotyped in a single tube, some of the SNP samples will be well extended, while others may be poorly extended. As a result, long extension products may not be observable if extension is poor and, therefore, a heterozygous genotype could be mistakenly assigned as homozygote. Other problems associated with long extension include overlapping of long extension products with the products from other SNP sites; overlapping of unterminated extension products with other extension products; and overlapping double-charged and dimer ions with singly-charged monomer ions.

Finally, we discuss the capability of VSET for multiplex genotyping of SNPs. The peaks of doubly-charged or dimer ions of a given oligonucleotide should not overlap with the peaks of singly-charged monomer ions of other oligonucleotides if the oligonucleotides are 15–29 bases in length. In addition, extension efficiency and specificity with the primers of these sizes are fairly good. Hence, 15–27 bases may be the ideal lengths for the primers of multiplexed VSET assays. Because each primer and its extension products generally cover three bases, five different sets (15, 18, 21, 24 and 27mers) of oligonucleotides can be used as the primers. The primers of different size can be created with a 3′ portion of the sequence complementary to the target and a 5′ portion which is not complementary to the target (3). The 5′ portion of the sequence serves as mass tag. In addition, three primers can be created from each set of oligonucleotides as long as the masses of the unextended and extended primers are sufficiently different (the average mass difference is ~100 Da). Therefore as many as 15 primers (3 × 5) can be used to genotype 15 SNP sites in a single tube. This throughput is comparable to that achieved with the PINPOINT method (3–5). Currently, we are working to apply VSET to high-level multiplex genotyping of SNPs.

ACKNOWLEDGEMENTS
The authors would like to thank Prof. A. Chakravarti of Case Western Reserve University for kindly providing the human genomic DNA samples and helpful discussion during the course of this work. The financial support by HG01437 and HG01815 is acknowledged.

REFERENCES