Published in final edited form as:

Electrophoresis. 2010 December ; 31(23-24): 3881–3888. doi:10.1002/elps.201000379.

Cost-effective interrogation of single-nucleotide polymorphisms using the mismatch amplification mutation assay and capillary electrophoresis

Erin P. Price^{1,†}, Molly A. Matthews^{1,†}, Jodi A. Beaudry¹, Jonathan L. Allred¹, James M. Schupp^{1,2}, Dawn N. Birdsell¹, Talima Pearson¹, and Paul Keim^{1,2,*}

¹Northern Arizona University, Center for Microbial Genetics and Genomics, Flagstaff, Arizona, United States

²Translational Genomics Research Institute, Phoenix, Arizona, United States

Abstract

The ability to characterize SNPs is an important aspect of many clinical diagnostic, genetic and evolutionary studies. Here, we designed a multiplexed SNP genotyping method to survey a large number of phylogenetically informative SNPs within the genome of the bacterium *Bacillus anthracis*. This novel method, capillary electrophoresis universal-tail mismatch amplification mutation assay (CUMA), allows for PCR multiplexing and automatic scoring of SNP genotypes, thus providing a rapid, economical, and higher-throughput alternative to more expensive SNP genotyping techniques. CUMA delivered accurate *B. anthracis* SNP genotyping results and when multiplexed, saved reagent costs by more than 80% compared with TaqMan real-time PCR. When real-time PCR technology and instrumentation is unavailable or the reagents are cost-prohibitive, CUMA is a powerful alternative for SNP genotyping.

Keywords

Bacillus anthracis; SNP; CUMA; genotyping; capillary electrophoresis

1. Introduction

The prevalence of DNA fingerprinting has risen dramatically in recent years due to the increase in specificity and versatility of nucleic acid-based methods. These molecular techniques allow for accurate and rapid genetic characterization of any biological organism, including phylogenetic inquiries that range from paternity testing to criminal investigations, such as in the 2001 *Bacillus anthracis* letter attacks [1]. Commonly, the determination of a species' phylogeny is accomplished with the recognition and interrogation of stable genetic mutations between two closely related organisms [2,3]. SNPs are an attractive target for phylogenetic analyses due to their evolutionary stability and abundance. Several methods exist for interrogating SNPs, including matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry [4], DNA microarrays [5], ligation [6], electrophoresis [7], probe hydrolysis (i.e. TaqMan) [8,9], high-resolution melt analysis [10] and primer extension [11]. However, these methods suffer one or more disadvantages that

*Corresponding author: Paul Keim, Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff AZ United States 86011-4073, paul.keim@nau.edu, Ph: +1 928-523-1078, Fax: +1 928-523-4015.

†Authors contributed equally to this work.

Conflict of Interest Statement: The authors have declared no conflict of interest.

range from high assay or equipment costs, technical difficulty or laboriousness, subjectivity in allele interpretation, inability to multiplex, poor interlaboratory adaptability or lack of resolution. One SNP genotyping method, TaqMan minor groove binder (MGB) probe hydrolysis [8], has gained popularity in recent years due to the specificity, ease and rapidity of this method, and the ubiquity of the real-time PCR platform in many public health and research laboratories. Despite these advantages, TaqMan quickly becomes time-consuming and cost-prohibitive when screening larger quantities of SNPs [2]. In addition, not all laboratories have access to real-time PCR instrumentation. Therefore, there is a need to develop a new method that addresses these inherent issues in light of the desire for large-scale SNP analyses.

We developed a SNP genotyping technique that combines a primer extension method, the mismatch amplification mutation assay (MAMA) [12-14], with a cost-effective universal-tail (UT) labeling system [15,16], and amplicon detection using capillary electrophoresis (CE). This method is termed capillary electrophoresis UT MAMA (CUMA). MAMA has been employed for both biallelic genotyping and allele frequency estimation [13,14,17-19]. Use of a UT fluorescent labeling system reduces analysis costs by eliminating the dependency on specifically labeled allele-specific probes and primers, while CE analysis enables multiplexing and fine-scale amplicon discrimination [16,19,20] on a common, versatile instrument [21].

Here, we validate the utility of CUMA for SNP genotyping using the bacterium *Bacillus anthracis* as our model organism, by converting a set of previously published *B. anthracis* canonical SNP (canSNP) TaqMan assays [2,3,22] into CUMA assays. In addition to the published *B. anthracis* SNP assays, we developed three CUMA assays for recently identified SNPs that further resolve the phylogenetic relationships. We demonstrate that CUMA, a multiplexed universal amplicon labeling PCR step coupled with CE, enables multiplexed SNP locus analyses capable of accommodating up to 40 assays. CUMA provides a cost-effective alternative to TaqMan MGB probes in circumstances where many SNP loci require interrogation and is a robust technique, even when crudely extracted DNA templates are used.

2. Materials and Methods

2.1 CUMA design

CUMA is based on the mismatch amplification mutation assay (MAMA) [12] approach, which takes advantage of the differential priming efficiency of allele-specific primers and the inherent inability of *Taq* polymerase to efficiently extend primers with 3' mismatches [23]. Allele-specific CUMA primers (termed 'universal tail (UT)-MAMA primers') were designed with two features; first, the SNP was positioned at the 3' ultimate end of both primers and an additional mismatch was engineered at the penultimate position to increase allele specificity and PCR stringency [14,24]; second, the UT-MAMA primers possessed a UT sequence at their 5' end to enable incorporation of this sequence into the PCR amplicons (Figure 1). The UT oligonucleotides were conjugated with one of four spectrally distinct fluorophores (5' to 3': 6FAM-ACCCAAGTGAATGGAGC, VIC-ACGCACTTGACTTGTCTTC, NED-ATCGACTGTGTTAGGTCAC, PET-CTGTCCTTACCTCAATCTC) (Applied Biosystems, Foster City, CA) [15,16]. All other primers were obtained from IDT (Coralville, IA) and are listed in Supporting Information 1. The UT-MAMA and consensus primers were designed with the web-based software Primer3 (<http://frodo.wi.mit.edu/primer3/>) and optimized with the primer-probe test tool in Primer Express v2.3 (Applied Biosystems) to minimize secondary structure interference. The primers were selected and designed in such a manner that annealing temperatures differed by no more than 2°C, with calculated annealing temperatures between 55-60°C. To avoid

amplicon migration overlap during CE, where possible each set of assay-specific MAMA primers was designed to generate amplicons that differed in length by at least two nucleotides. Assays sharing a dye set (e.g. FAM/VIC) were spaced by at least 10bp, and alternate dye sets by 5bp. Assay sets were designed such that all *in silico* amplicon sizes were between 110-300bp, allowing up to ~40 SNP assays to be analyzed in a single electrophoretic injection.

2.2 Genomic DNA and SNPs used in this study

The samples used to validate CUMA were part of a genetically diverse panel comprising 26 previously described *B. anthracis* isolates [25] (A0034 and A0402 were not used in the current study). Most of these samples have been extensively genotyped at the 20 canSNPs [3]. The diversity panel was further augmented by the addition of the *B. anthracis* strain A3783, collected in 1956, as this strain falls within the underrepresented C Branch clade [26] and contains both *B. anthracis* pX01 and pX02 virulence plasmids [27]. Near-neighbor species *Bacillus thuringiensis* HD1011 and *Bacillus cereus* FRI-42 were included in the diversity panel to test the performance of the *B. anthracis*-specific *plcR* assay using CUMA [22]. Twenty phylogenetically informative canSNPs for *B. anthracis* were converted to CUMA in the current study. The first canSNP differentiates *B. anthracis* from non-pathogenic *Bacillus* species based upon a polymorphism within the *plcR* gene [22], a further sixteen SNPs define important major phylogenetic clades within *B. anthracis* [3], and three SNPs differentiate the Ames lineage (Ames was used in the 2001 United States bioterrorist attacks) from other *B. anthracis* strains [28] (Supporting Information 1). The majority of these SNPs have previously been validated using TaqMan, enabling a direct comparison of CUMA accuracy.

2.3 PCR Parameters

Genomic DNA was generated for CUMA using a simple heat-lysis method [29]. DNA templates were approximated using real-time PCR cycle threshold values for the *plcR* assay and normalized to between 200pg/μL and 1ng/μL. This normalization improved first-pass success rates by minimizing reruns caused by either low or saturated signal intensity in the CE analysis (results not shown). Each 10 μL singleplexed or multiplexed PCR consisted of 1X PCR Buffer, 2mM MgCl₂, 0.2mM dNTPs (Invitrogen, Carlsbad, CA), 0.2U iTaq DNA Polymerase (Bio-Rad, Hercules, CA), and molecular grade water (GIBCO, Carlsbad, CA). Depending on the SNP assay, uneven amplicon fluorescence favoring one allele state over the other required the standard 0.2 μM primer concentration to be skewed for particular assays (Table 1). This skew equalized robustness between alleles, resulting in more consistent fluorescent signal between the alleles. All PCRs were performed in 96-well plates (Bio-Rad) using an MJ Research PTC-200 thermocycler. Cycling conditions were identical for all reactions and consisted of an initial hot-start of 95°C for 5 minutes, followed by 40 cycles of 30 sec at 95°C, 30 sec at 63°C, and 30 sec at 72°, with a final extension at 72°C for 5 minutes. Annealing temperatures optimal for differential allele amplification were between 61-65°C for all assays; therefore, 63°C was chosen for all multiplex reactions.

2.4 Capillary Electrophoresis

All amplicons from separate PCR mastermixes were pooled and diluted in molecular grade water (GIBCO) to a final dilution of 1:100. The CE injection mix per well included 1 μL of pooled diluted amplicons, 12 μL HiDi formamide and 1 μL of 1:5 diluted GeneScan 600 LIZ Size Standard (Applied Biosystems). Samples were denatured at 95°C for four minutes and rapidly cooled to approximately 4°C prior to CE analysis. The amplicons were analyzed on an ABI PRISM 3100 Genetic Analyzer using POP4 polymer, a 36cm array, and the default GeneScan36_POP4Default module (oven temperature of 60°C, run voltage of 15kV and injection time of 22 s; Applied Biosystems). We modified the run time from 1500 to

1800 s to capture longer reads. Amplicon sizes were verified across the *Bacillus* spp. diversity panel, which made it possible to view the different allele states for each assay. Following singleplex optimization, assays were pooled based upon dye set and amplicon length to facilitate multiplexing (Table 1).

2.5 Data Analysis

Electrophoretic data were analyzed with the GeneMapper program SNaPShot according to the *GeneMapper Installation and Administration* v4.0 and *SNaPShot* manuals (Applied BioSystems). The automated binning feature was used to select the appropriate peaks, and reduced the amount of non-specific PCR artifacts and primer dimer calls. Bins were created with approximately ± 2 bp buffer ranges to accommodate the inherent intra- and inter-instrument variability of CE analyzers. The necessity for recognizing two amplicons per locus due to potential non-specific amplification of the alternate UT-MAMA primer required an altered analysis method similar to that of peak recognition for heterozygotes. The analysis method parameters involved an advanced peak detection algorithm that incorporated light smoothing, a minimum peak half-width of 3 points, a polynomial degree of 2, and a peak window size of 19 points. The large polynomial degree increases peak detection sensitivity, light smoothing assists peak size optimization and reduces baseline noise, and a small peak window size allows detection of single base-pair size differences without interference from peak shoulder noise. Peak quality preferences included a heterozygous minimum peak height of 200, and a maximum of one expected allele.

2.6 Cost Comparison between CUMA and TaqMan MGB probes

Product list prices were used to determine the cost differential between CUMA and TaqMan MGB probe reagents as of May 2010 (Table 2). Several reaction scenarios were evaluated based on whether reactions were carried out as singleplex, 5-plex, or 20-plex PCR. All necessary dilutions and reagents were considered. Cost per data point and reaction were derived assuming a full 96-well plate of samples.

3. Results and Discussion

We describe a novel SNP interrogation technique, CUMA, which enables rapid, high throughput, accurate, multiplexed and inexpensive SNP genotyping on a common CE detection platform. CUMA exploits the premise of the MAMA system for SNP interrogation, which is based on the inefficiency of *Taq* polymerase to extend primers containing nucleotide mismatches at their 3' end [23,24]. The utility of MAMA for SNP interrogation is well-documented, and includes TaqMAMA [13], agarose gel MAMA [18,30], and real-time PCR MAMA [17]. However, these methods either lack sensitivity and precision, or cannot be multiplexed, reducing the high-throughput capacity of these techniques [18]. Other CE-based SNP interrogation methods have recently emerged due to the versatility of this instrument [31-37]. However, none of these CE methods combine the advantages of throughput, rapidity, large multiplex capacity and use of commercially available consumables. CUMA, on the other hand, is adaptable to a broad multiplex format, involves minimal processing steps and provides accurate results using generic CE instrumentation and reagents, providing an attractive alternative to other CE and MAMA SNP interrogation methods.

In addition to MAMA, CUMA takes advantage of a UT approach, which drastically minimizes assay cost irrespective of how many SNP loci are interrogated. UT oligonucleotides are differentially labeled with four fluorescent dyes conjugated to the 5' nucleotide (Figure 1), and have been previously used to differentiate VNTRs in *B. anthracis* and *Burkholderia pseudomallei* [15,16]. Within a singleplex CUMA assay are five

oligonucleotides; two UT-MAMA primers, two corresponding fluorescently labeled UTs and a consensus primer. The UT-MAMA primers, each targeting one polymorphism, facilitate amplification in the presence of matched template. The deliberate incorporation of a penultimate mismatch within the UT-MAMA primers ensures that the alternate UT-MAMA primer does not prime efficiently, thereby minimizing non-specific amplification. UT-MAMA primers are engineered with a 5' tail that is identical to one of the four 5' fluorescently labeled UT oligonucleotides. Depending on which UT-MAMA primer anneals to the DNA target, the corresponding UT sequence is integrated into the PCR products during early rounds of amplification and acts as a hybridization site for the complementary labeled UT primer. This UT primer is subsequently integrated into late PCR products resulting in the detection of amplicons with the correct SNP state corresponding to the incorporated fluorophore (Figure 2). Following PCR, CUMA amplicons are size separated using a CE instrument that allows both amplicon size and fluorophore signal to be determined. Therefore, the UT system can be used across multiple assays, eliminating the requirement for assay-specific fluorescently labeled primers and probes and providing a high-throughput screening platform that is able to screen a large numbers of SNPs per injection at minimal cost.

To assess the performance of CUMA, we tested this method against a diverse panel of 25 *B. anthracis* isolates and two near-neighbor *Bacillus* sp. using the singleplex CUMA format. The CUMA SNP genotyping system enabled rapid and accurate genotyping of these isolates against 17 previously characterized *B. anthracis* canSNPs [3] providing 100% agreement with previously published TaqMan data [3,13,28]. Additionally, the CUMA system validated three newly discovered SNPs (ABr005, ABr010 and CBr001) on the previously published phylogenetic construct of *B. anthracis* [3]. The ABr005 SNP separates all A branch *B. anthracis* isolates from the B and C branches, the ABr010 SNP separates Western North America, Ames and Australia 94 lineages from the Vollum clade, and the CBr001 separates the newly identified C branch from A and B branches [3,25,38]. A *B. anthracis* C-branch isolate (A3783), which contains both pX01 and pX02 virulence plasmids, was accurately genotyped with CUMA, with the results being in phylogenetic agreement with the previously validated TaqMan MGB probe and VNTR data for this lineage [3,26,29]. The near-neighbor strains, as expected, were 'ancestral' for the *plcR* assay, indicating that they were not *B. anthracis*.

Once it was demonstrated that CUMA accomplished highly differential allele-specific amplification in a singleplex reaction, this method was tested in multiplexed format. We were routinely able to multiplex four to six loci in a single PCR without compromising PCR robustness (Table 1; Figure 3). We were also able to pool all 20 loci into a single injection without a loss in electrophoretic accuracy (Figure 3II). PCR multiplexing was made possible by segregating assays into mixes based on dye set and amplicon size. Separation reduced competition for reagents between smaller and larger amplicons, and decreased the interference created by extraneous UT oligonucleotides. PCR multiplexing allowed for a drastic reduction in time and money and showed little to no effect on genotyping accuracy or efficiency, with CUMA providing identical SNP genotyping data to TaqMan probes. Specifically, the multiplexed PCRs yielded a 97% data conversion rate compared with singleplexed counterparts; in other words, 97% of multiplexed reactions yielded the expected detectable amplicon when compared with the singleplexed reactions. Importantly, in all instances, the correct allele and fragment sizes were consistent between the multiplexed and singleplexed reactions, indicating that the multiplexed format did not interfere with accuracy or locus migration. Using the multiplexed CUMA format, a considerable cost-reduction was obtained. A 5-plex CUMA assay was calculated to cost \$95 per 96-well plate (46 samples plus positive and negative controls), whereas the TaqMan equivalent was \$329. Further multiplexing of PCR amplicons prior to CE injection, as

performed with the 20-plex CUMA format (Figure 3II), was calculated at \$178 per 96-well plate as opposed to TaqMan probes, which at \$1318 per 96-well plate yields a cost differential of 86% between the two assay formats (Table 2). In some instances, CUMA assay redesign was required due to overly robust amplification of the alternate allele, which required a change of the penultimate nucleotide. However, the costs associated with assay redesign were substantially lower using the UT primer system compared with TaqMan probes (Table 2) or labeled locus-specific primers. For example, reordering of all CUMA primers would cost \$25-30/assay, whereas a redesign of both TaqMan probes would cost ~\$550/assay, a differential of 97%. Therefore, CUMA not only provides an inexpensive alternative to more expensive SNP interrogation techniques, but has the additional advantage of assay flexibility in cases where optimization is necessary.

Despite its many advantages, there are some caveats to the CUMA approach. First, in a small number of assays, one allele exhibited amplification for both UT-MAMA primers, resulting in a dual fluorescent signal. As bacteria are haploid, a heterozygote genotype should not occur if the DNA template is truly homogeneous. However, in some instances it was found that alteration of the mismatched penultimate nucleotide increased allele-specific stringency and resulted in a single genotype (results not shown). In other instances, this alteration was ineffective at suppressing a dual fluorescent signal for one allele (e.g. the *plcR* assay, FAM (blue) allele; see Figure 3II and Supporting Information 2); however, the alternate allele always only yielded a single peak (e.g. *plcR* assay, VIC (green) allele), thus providing the necessary allelic discrimination. This circumstance demonstrates the importance of running control samples that represent both allelic states to ensure that the correct allele is called. For diploid genomes, running controls for both homozygous and heterozygous genotypes is particularly important to prevent incorrect interpretation of a dual fluorescent signal. Second, the UT component of the UT-MAMA primers can result in unwanted secondary structure that can inhibit PCR signal and result in primer-dimer peaks in the CE analysis. We found that this phenomenon can be minimized by testing secondary structure of primers *in silico* with the inclusion of the UT sequence during assay design and adjusting primer positions accordingly. Third, not all assays performed with the same level of robustness, particularly in the multiplexed format, sometimes resulting in some assays that out-competed others. To minimize the impact of differential PCR efficiency, it was necessary to alter the stoichiometry of primers within a multiplex reaction or modify the combination of assays to ensure that all assays amplified with relatively equal efficiency. Fourth, primer dimers and nonspecific amplification can result in spurious or erroneous electrophoretic peaks, and it is therefore important to ensure that the PCRs are optimized such that PCR artifacts are minimized. Lastly, we found that this method was not suitable for use in reproducible or accurate determination of allele frequency for mixed target analysis. Despite these disadvantages, we were able to overcome many issues with minor PCR optimization and highly specific assay design. Additionally, while the CUMA primer sets presented here have been shown to be robust, our subsequent CUMA work has shown that increasing the annealing temperature of the consensus primer to ~5°C above that of the MAMA primers enhanced assay efficiency and should be considered when designing new assays.

4. Concluding Remarks

We describe a novel and cost effective system for surveying 20 phylogenetically informative SNPs in *Bacillus anthracis* quickly, economically, and accurately, using the common and multifunctional CE instrument. However, this technology is not limited to *B. anthracis*. Future applications of CUMA include highly cost-effective verification of newly discovered SNPs from whole genome sequencing data for *B. anthracis* or for other species, or the screening of large panels of DNA samples across targeted sets of novel SNP assays for

clinical and genetic studies. We have demonstrated that multiplexed CUMA costs substantially less than TaqMan real-time PCR, without compromising genotyping accuracy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the U.S. Department of Homeland Security S&T CB Division Bioforensics R&D Program. Use of products/names does not constitute endorsement by DHS or USG. EPP was supported in part by the NIH NIAID U01 AI175568 grant.

References

1. Keim P, Pearson T, Okinaka R. *Anal Chem*. 2008; 80:4791–4799. [PubMed: 18609763]
2. Keim P, Van Ert MN, Pearson T, Vogler AJ, et al. *Infect Genet Evol*. 2004; 4:205–213. [PubMed: 15450200]
3. Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, et al. *PLoS ONE*. 2007; 2:e461. [PubMed: 17520020]
4. Ross PL, Lee K, Belgrader P. *Anal Chem*. 1997; 69:4197–4202. [PubMed: 9337595]
5. Sapolsky RJ, Hsie L, Berno A, Ghandour G, et al. *Genet Anal*. 1999; 14:187–192. [PubMed: 10084113]
6. Tobler AR, Short S, Andersen MR, Paner TM, et al. *J Biomol Tech*. 2005; 16:398–406. [PubMed: 16522862]
7. Mishima K, Takarada T, Maeda M. *Anal Sci*. 2005; 21:25–29. [PubMed: 15675511]
8. Kutyavin IV, Afonina IA, Mills A, Gorn VV, et al. *Nucleic Acids Res*. 2000; 28:655–661. [PubMed: 10606668]
9. Livak KJ. *Genet Anal*. 1999; 14:143–149. [PubMed: 10084106]
10. Krypuy M, Ahmed AA, Etemadmoghadam D, Hyland SJ, et al. *BMC Cancer*. 2007; 7:168. [PubMed: 17764544]
11. Germer S, Higuchi R. *Genome Res*. 1999; 9:72–78. [PubMed: 9927486]
12. Cha RS, Zarbl H, Keohavong P, Thilly WG. *PCR Methods Appl*. 1992; 2:14–20. [PubMed: 1490171]
13. Easterday WR, Van Ert MN, Zanecki S, Keim P. *Biotechniques*. 2005; 38:731–735. [PubMed: 15945372]
14. Li B, Kadura I, Fu DJ, Watson DE. *Genomics*. 2004; 83:311–320. [PubMed: 14706460]
15. Jay, Z. Department of Biological Sciences. Northern Arizona University; Flagstaff: 2005.
16. U'Ren JM, Schupp JM, Pearson T, Hornstra H, et al. *BMC Microbiol*. 2007; 7:23. [PubMed: 17397553]
17. Mattarucchi E, Marsoni M, Binelli G, Passi A, et al. *J Biochem Mol Biol*. 2005; 38:555–562. [PubMed: 16202235]
18. Bui M, Liu Z. *Plant Methods*. 2009; 5:1. [PubMed: 19133153]
19. Mizuno N, Kitayama T, Fujii K, Nakahara H, et al. *Forensic Sci Int Genet*. 2010; 4:73–79. [PubMed: 20129464]
20. Yin BC, Wang XF, Ye BC. *Anal Biochem*. 2009; 387:221–229. [PubMed: 19344657]
21. Forster RE, Hert DG, Chiesl TN, Fredlake CP, Barron AE. *Electrophoresis*. 2009; 30:2014–2024. [PubMed: 19582705]
22. Easterday WR, Van Ert MN, Simonson TS, Wagner DM, et al. *J Clin Microbiol*. 2005; 43:1995–1997. [PubMed: 15815042]
23. Huang MM, Arnheim N, Goodman MF. *Nucleic Acids Res*. 1992; 20:4567–4573. [PubMed: 1408758]
24. Rhodes RB, Lewis K, Shultz J, Huber S, et al. *Mol Diagn*. 2001; 6:55–61. [PubMed: 11257212]

25. Pearson T, Busch JD, Ravel J, Read TD, et al. *Proc Natl Acad Sci U S A*. 2004; 101:13536–13541. [PubMed: 15347815]
26. Sue D, Marston CK, Hoffmaster AR, Wilkins PP. *J Clin Microbiol*. 2007; 45:1777–1782. [PubMed: 17392445]
27. Okinaka R, Cloud K, Hampton O, Hoffmaster A, et al. *J Appl Microbiol*. 1999; 87:261–262. [PubMed: 10475962]
28. Van Ert MN, Easterday WR, Simonson TS, U'Ren JM, et al. *J Clin Microbiol*. 2007; 45:47–53. [PubMed: 17093023]
29. Keim P, Price LB, Klevytska AM, Smith KL, et al. *J Bacteriol*. 2000; 182:2928–2936. [PubMed: 10781564]
30. Luo HR, Tu GC, Zhang YP. *Clin Chem Lab Med*. 2001; 39:1195–1197. [PubMed: 11798074]
31. Asari M, Watanabe S, Matsubara K, Shiono H, Shimizu K. *Anal Biochem*. 2009; 386:85–90. [PubMed: 19070585]
32. Chen YL, Chang YS, Chang JG, Wu SM. *Anal Bioanal Chem*. 2009; 394:1291–1297. [PubMed: 19015838]
33. Leader BT, Frye JG, Hu J, Fedorka-Cray PJ, Boyle DS. *J Clin Microbiol*. 2009; 47:1290–1299. [PubMed: 19261787]
34. Wang CC, Chang JG, Jong YJ, Wu SM. *Electrophoresis*. 2009; 30:1102–1110. [PubMed: 19373809]
35. Dixon LA, Murray CM, Archer EJ, Dobbins AE, et al. *Forensic Sci Int*. 2005; 154:62–77. [PubMed: 16182950]
36. Arakawa H, Watanabe K, Kashiwazaki H, Maeda M. *Biomed Chromatogr*. 2002; 16:41–46. [PubMed: 11816010]
37. Gomez-Llorente C, Antunez A, Blanco S, Suarez A, et al. *Eur J Haematol*. 2004; 72:121–129. [PubMed: 14962249]
38. Kenefic LJ, Pearson T, Okinaka RT, Schupp JM, et al. *PLoS One*. 2009; 4:e4813. [PubMed: 19283072]

Abbreviations

MAMA	mismatch amplification mutation assay
UT	universal tail
CUMA	capillary electrophoresis universal tail mismatch amplification mutation assay
canSNP	canonical SNP

Figure 1. UT-MAMA primers and the UT approach

SNP interrogation is achieved by the use of two MAMA primers, in which the 3' ultimate nucleotide directly overlaps the SNP [14,24]. Each MAMA primer is specific to a different polymorphism at the SNP of interest. The MAMA primers are engineered with two further features; the first, a penultimate mismatch, which acts to destabilize non-complementary primer-template complexes, thus ensuring only the 3' matched primer will extend and be detected at end-point; the second, a UT site that enables one of four fluorescently labeled UTs (FAM, VIC, PET, NED) can be used [15,16], which enables greater multiplexing than a single dye format.

Figure 2. Principle of CUMA

(I and II) Representation of a haploid template (e.g. *Bacillus anthracis*) encountering both UT-MAMA primers. (Ia) One of the two UT-MAMA primers binding with its corresponding matched template. (Ib) Polymerase extending from the 3' matched UT-MAMA primer. (Ic) Second PCR cycle, indicating complementary strand amplification from the amplicon made in (Ib). (Id) UT-FAM annealing with newly synthesized complementary amplicon; polymerase binding and elongation from the UT. (Ie) At PCR endpoint. The labeled amplicon generated from the 3' matched UT MAMA primer greatly outnumbers the molecules generated by the mismatched UT-MAMA oligonucleotide. (If) Electrophoretic image of the labeled amplicon from Ie. Orange peaks indicate internal control LIZ600-labeled ladder rungs. (IIa) Binding of the mismatched UT-MAMA primer with the haploid template. The penultimate nucleotide mispairing is shown. (IIb) Due to mispairing at the engineered mismatch, the 3' end of the primer is greatly destabilized and cannot efficiently anneal, leaving no attachment point for the polymerase, thus resulting in little to no molecules for the VIC UT label to bind. As a consequence, there is no detectable amplification of the mismatched UT-MAMA primer and typically only a single peak (corresponding to the FAM label) is detected on the electrophoregram (IIf).

Figure 3. Multiplexed CUMA PCR electropherogram

(I) Description of the principles from Figure 2 in multiplexed format. An example PCR mastermix targeting four SNP loci for one DNA template is shown. In general, only those UT-MAMA primers that contain a direct match with the DNA template will amplify. The result is an electropherogram containing four individual fluorescence peaks that correspond to four different SNP loci. (II) An electrophoretic image of the 20-plex *B. anthracis* SNP system for strain A0463 (A.Br.008/009 ‘Trans Eurasian’ (TEA) clade [38]); all four PCR mastermixes containing 20 *Bacillus anthracis* SNP loci (see Table 1 for mastermix details) are pooled prior to CE and run as a single injection. The result is similar to the electropherogram depicted in Figure 3A, but contains 20 individual fluorescence peaks that correspond to 20 different SNP loci, minus sizing standard peaks.

Table 1
CUMA primer concentrations and mastermixes for singleplex and multiplex PCR of 20 phylogenetic *Bacillus anthracis* SNPs

	MasterMix 1					MasterMix 2					MasterMix 3					MasterMix 4				
SNP assay ^a	<i>pICR</i>	A.Br001	A.Br005	A.Br009	B.Br001	A.Br002	A.Br004	A/B Br001	B.Br003	PS-1	A.Br007	A.Br003	A.Br010	B.Br002	C.Br001	A.Br006	A.Br008	B.Br004	Br1-31	PS-52
Consensus primer	0.4	0.25	0.25	0.5	0.25	0.4	0.4	0.4	0.5	0.4	0.25	0.25	0.25	0.25	0.5	0.4	0.4	0.4	0.4	0.5
UT-MAMA primer, Allele 1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.1
UT-MAMA primer, Allele 2	0.1	0.1	0.1	0.1	0.1	0.1	0.06	0.1	0.1	0.06	0.1	0.06	0.1	0.06	0.1	0.06	0.06	0.06	0.06	0.1
UT primer, Allele 1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
UT primer, Allele 2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2

^aSee Supporting Information 1 for assay details.

All primer concentrations are in μM. Allele 1 refers to the 'derived' SNP state and Allele 2 refers to the 'ancestral' SNP state.

Table 2
Cost comparison of CUMA and TaqMan MGB probes for SNP genotyping

	Real-time PCR (AB7900HT)	CUMA (ABI PRISM 3100)
Initial start-up cost (single assay):	Price* \$3,996.47	Price* \$2,271.01
SINGLEPLEX		
DNA polymerase	\$1.62	\$15.44
Universal tails (UTs)	N/A	\$3.50
Unlabeled primers	\$0.09	\$0.14
TaqMan MGB probes	\$17.28	N/A
TaqMan universal PCR mastermix	\$38.10	N/A
PCR consumables	\$9.06	\$8.11
Electrophoresis consumables	N/A	\$66.95
Total:	\$66.15	\$94.14
FIVE LOCI		
DNA polymerase	\$6.80	\$15.44
Universal tails (UTs)	N/A	\$3.50
Unlabeled primers	\$0.46	\$0.68
TaqMan MGB probes	\$86.40	N/A
TaqMan universal PCR mastermix	\$190.50	N/A
PCR consumables	\$45.30	\$8.11
Electrophoresis consumables	N/A	\$66.95
Total:	\$329.46	\$94.68
TWENTY LOCI		
DNA polymerase	\$27.20	\$61.76
Universal tails (UTs)	N/A	\$14.00
Unlabeled primers	\$1.80	\$2.70
TaqMan MGB probes	\$345.60	N/A
TaqMan universal PCR mastermix	\$762.00	N/A
PCR consumables	\$181.20	\$32.44
Electrophoresis consumables	N/A	\$66.95
Total:	\$1,317.80	\$177.85
Redesign costs (per locus)		
Unlabelled primers	\$11.60	\$25-30
TaqMan MGB probes	\$540.00	N/A
Total	\$551.60	\$25-30

* using 10uL PCR volumes. Cost is based on genotyping 46 samples and includes four wells (in duplicate) dedicated to positive and negative controls. Total costs assume 96-well format. All costs current as of May 2010 and inclusive of tax.