Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Mass Spectrometry 242 (2005) 23–41
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

TIGER: the universal biosensor

Steven A. Hofstadler[a,*], Rangarajan Sampath[a], Lawrence B. Blyn[a], Mark W. Eshoo[a], Thomas A. Hall[a], Yun Jiang[a], Jared J. Drader[a], James C. Hannis[a], Kristin A. Sannes-Lowery[a], Lendell L. Cummins[a], Brian Libby[a], Demetrius J. Walcott[a], Amy Schink[a], Christian Massire[a], Raymond Ranken[a], Jose Gutierrez[a], Sheri Manalili[a], Cristina Ivy[a], Rachael Melton[a], Harold Levene[a], Greg Barrett-Wilt[a], Feng Li[a], Vanessa Zapp[a], Neill White[a], Vivek Samant[a], John A. McNeil[a], Duane Knize[b], David Robbins[b], Karl Rudnick[b], Anjali Desai[b], Emily Moradi[b], David J. Ecker[a]

[a] *Ibis Therapeutics, A Division of Isis Pharmaceuticals, 1891 Rutherford Rd., Carlsbad, CA 92008, USA*
[b] *SAIC, 10260 Campus Point Drive, San Diego, CA 92121, USA*

Received 22 September 2004; accepted 27 September 2004
Available online 22 December 2004

Abstract

In this work, we describe a strategy for the detection and characterization of microorganisms associated with a potential biological warfare attack or a natural outbreak of an emerging infectious disease. This approach, termed TIGER (Triangulation Identification for the Genetic Evaluation of Risks), relies on mass spectrometry-derived base composition signatures obtained from PCR amplification of broadly conserved regions of the microbial genome(s) in a sample. The sample can be derived from air filtration devices, clinical samples, or other sources. Core to this approach are "intelligent PCR primers" that target broadly conserved regions of microbial genomes that flank variable regions. This approach requires that high-performance mass measurements be made on PCR products in the 80–140 bp size range in a high-throughput, robust modality. As will be demonstrated, the concept is equally applicable to bacteria and viruses and could be further applied to fungi and protozoa. In addition to describing the fundamental strategy of this approach, several specific examples of TIGER are presented that illustrate the impact this approach could have on the way biological weapons attacks are detected and the way that the etiologies of infectious diseases are determined. The first example illustrates how any bacterial species might be identified, using *Bacillus anthracis* as the test agent. The second example demonstrates how DNA-genome viruses are identified using five members of *Poxviridae* family, whose members includes Variola virus, the agent responsible for smallpox. The third example demonstrates how RNA-genome viruses are identified using the *Alphaviruses* (VEE, WEE, and EEE) as representative examples. These examples illustrate how the TIGER technology can be applied to create a universal identification strategy for all pathogens, including those that infect humans, livestock, and plants.

Keywords: Pathogen detection; Microbe; Biosensor; Biodefense; Infectious disease

1. Introduction

Mass spectrometry-based methods for analysis of nucleic acids continue to mature as hardware, software, and sample purification protocols grow in sophistication. While there are a number of relatively simple methods one can use to address straightforward analytical questions related to nucleic acids (e.g. slab gels, capillary gel electrophoresis), the recent tendency is to move towards more sophisticated analytical

Abbreviations: CDC, centers for disease control; EEE, eastern equine encephalitis; ESI–MS, electrospray ionization mass spectrometry; FTICR, Fourier transform ion cyclotron resonance; PCR, polymerase chain reaction; TIGER, triangulation identification for the genetic evaluation of risks; TOF, time-of-flight; USDA, United States Department of Agriculture; UTHSC, University of Texas Health Science Center; VEE, Venezuelan equine encephalitis; WEE, western equine encephalitis; bp, base pair; BW, biological weapon; A, adenosine; G, guanosine; C, cytidine; T, thymidine; FWHM, full-width half-maximum; SNPs, single nucleotide polymorphisms

* Corresponding author. Tel.: +1 760 603 2599; fax: +1 760 603 4653.
E-mail address: shofstadler@isisph.com (S.A. Hofstadler).

1387-3806/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2004.09.014

platforms to answer more complex questions. In this work we will introduce the concept of Triangulation Identification for the Genetic Evaluation of Risks (TIGER). TIGER employs high-performance electrospray mass spectrometry (either FTICR or TOF) to derive base compositions of polymerase chain reaction (PCR) products. Core to this approach are "intelligent PCR primers" that target broadly conserved regions of microbial genomes that flank variable regions. This approach requires that high-performance mass measurements be made on PCR products in the 80–140 bp size range in a high-throughput, robust modality. As will be demonstrated, the concept is equally applicable to the detection of bacteria and viruses and could be further applied to fungi and protozoa. In addition to describing the fundamental strategy of this approach, we present several specific examples of TIGER that illustrate the potential impact this approach could have on the way we detect biological weapons attacks and the way we determine the etiology of infectious diseases.

A key problem in determining the cause of a natural infectious outbreak or a bioterrorist attack is the sheer variety of organisms that can cause human disease. According to a recent review [1], there are over 1400 organisms infectious to humans; many of these have the potential to emerge suddenly in a natural epidemic or to be used in a malicious attack by bioterrorists. This number does not include numerous strain variants, bioengineered versions, or pathogens that infect plants or animals. Paradoxically, much of the new technology being developed for detection of biological weapons incorporates a polymerase chain reaction step based upon the use of highly specific primers and probes designed to selectively detect certain pathogenic organisms. Although this approach is appropriate for the most obvious bioterrorist organisms, like smallpox and anthrax, experience has shown that it is very difficult to predict which of hundreds of possible pathogenic organisms might be employed in a terrorist attack. Likewise, naturally emerging human diseases that have caused devastating consequences to public health have come from unexpected families of bacteria, viruses, fungi, or protozoa. Plants and animals also have their natural burden of infectious disease agents and there are equally important biosafety and security concerns for agriculture.

A major conundrum for public health protection, biodefense, and agricultural safety and security is the need to be able to rapidly identify and characterize infectious agents, yet there is no existing technology with the breadth of function to meet this need. In this paper we describe broad-function technology based on mass spectrometric detection that enables the rapid, sensitive, and cost-effective identification of a broad range of infectious microorganisms, including natural human pathogens, bioterrorist agents, and agricultural pathogens. The use of such broad-function technology may be the most practical way to simultaneously survey for all forms of pathogens.

To achieve this objective, we have abandoned the notion of detecting specific target organisms on a "one off" basis, and instead have developed a strategy to identify all the organisms present in the sample without anticipating which might be present. TIGER is based upon the principle that, despite the enormous diversity of microbes, all domains of life on earth share sets of essential common features in the biomolecules encoded in their genomes. We use these common features in an identification strategy that relies on high-performance ESI–MS analysis of broad-range PCR amplification products and base-composition analysis. The base compositions from multiple primer pairs are used to "triangulate" the identity of the organisms present in the sample.

In this paper, we describe the basic principles of TIGER, and provide examples of applications of the technology in environmental surveillance for an aerosol attack with a biological weapon, or for analysis of a human clinical sample. We describe three examples in which the TIGER approach is used to detect and characterize microbes associated with a potential BW attack or naturally occurring pathogen. The first example describes how any bacterial species might be identified, using *Bacillus anthracis* as the test agent. The second example demonstrates how DNA-genome viruses are identified using five members of *Poxviridae* family, whose members includes Variola virus, the agent responsible for smallpox. The third example demonstrates how RNA-genome viruses are identified using the *Alphaviruses (*VEE, WEE, and EEE) as representative examples. These examples illustrate how the TIGER technology can be extended to create a seamless biosensor network for the universal identification of all pathogens.

2. Experimental/materials and methods

2.1. Bacterial genome isolation from air samples

Air samples were collected on a Spincon portable air sampler system, model number PAS540-10 (Camber Corporation, Huntsville, AL) or on dry filter units (DFUs). The airflow on the Spincon unit is approximately 450 L/min and the material is collected directly into a PBS/detergent liquid matrix. For the DFUs, the airflow is approximately 850 L/min and the material is collected onto a 2" polyester fiber filter. Sample times for each collection are indicated in the text. For the DFUs, filters were collected from the units and up to four filters were combined in 20 mL of phosphate buffered saline (pH 7.0) containing 0.1% Tween-20 detergent. The filters and solution were shaken by hand for 30 s prior to storage at 4 °C. Prior to genomic extraction, the filters and solution were shaken again for 30 s and the resulting solution was filtered through a 25 mm 0.2 μm Supor-200 filter (Pall Corporation, Ann Arbor, Michigan). Spincon material was filtered directly by this same method. The resulting Supor-200 filters were then subjected to bead beating by placing the filter in a 1.5 mL tube containing ∼100 μg of 0.7 mm zirconium beads (Biospec Products, Bartlesville, OK) and 350 μL of ATL buffer from Qiagen (Qiagen, Valencia, CA). The beads were shaken on a Retsch 300 MM mixer mill with a frequency of

19 Hz for 10 min and then spun briefly to settle the beads and larger particles. The supernatant, containing the nucleic acid material, was then used as the starting material in a Qiagen DNeasy Tissue Kit isolation on a Qiagen BioRobot 8000 (Qiagen, Valencia, CA) following the manufacturers protocol.

2.2. Bacterial genome isolation from cultures and colonies

For genomic DNA isolation from bacterial cultures, cultures were grown to log phase (OD_{600} of 0.1–0.3). A small amount of this material (generally less than 50 μL) was transferred into 300–350 μL of ATL buffer (Qiagen, Valencia, CA) containing about 100 μg of 0.7 mm zirconium beads (Biospec Products, Bartlesville, OK). The material was bead beaten and isolated using the DNeasy Tissue Kit (Qiagen, Valencia, CA) as described above.

2.3. Viral genome isolation

Viral RNA genomes were isolated from 250 μL of infected cells or culture supernatant spiked with 3 μg of sheared poly A DNA using Trizol or Trizol LS, respectively (Invitrogen Inc., Carlsbad, CA) according to the manufacturer's protocol. For isolation of DNA or RNA genomes from virus-containing samples in an automated fashion, samples were prepared as above for the Qiagen MDx BioRobot and the material isolated using a QiaAmp Virus BioRobot MDx kit as described by the manufacturer. In both cases, the resulting material is ready for either PCR or RT-PCR as appropriate.

2.4. PCR conditions

All PCR reactions were assembled in 50 μL reaction volumes in a 96-well microtiter plate format using a Packard MPII liquid handling robotic platform and M.J. Dyad thermocyclers (MJ research, Waltham, MA). The PCR reaction mix consisted of 4 u of Amplitaq Gold, 1× buffer II (Applied Biosystems, Foster City, CA), 1.5 mM $MgCl_2$, 0.4 M betaine, 800 μM dNTP mix and 250 nM of each primer. The following PCR conditions were used: 95 °C for 10 min followed by eight cycles of 95 °C for 30 s, 48 °C for 30 s, and 72 °C 30 s with the 48 °C annealing temperature increasing 0.9 °C with each of the eight cycles. The PCR was then continued for 37 additional cycles of 95 °C for 15 s, 56 °C for 20 s, and 72 °C 20 s.

2.5. RT-PCR conditions

All RT-PCR reactions were assembled in 50 μL reaction volumes in a 96-well microtiter plate format using a Packard MPII liquid handling robotic platform and M.J. Dyad thermocyclers (MJ research, Waltham, MA). The RT-PCR reaction mix consisted of 4 U of Amplitaq Gold, 1.5× buffer II (Applied Biosystems, Foster City, CA), 1.5 mM $MgCl_2$, 0.4 M betaine, 20 mM sorbitol, 1.2 U Superasin (Ambion Inc, Woodlands TX), 0.4 μg T4 gene 32 protein (Roche Applied Sciences (Indianapolis, IN), 800 μM dNTP mix (Stratagene, La Jolla, CA), 100 ng of poly A RNA, 10 mM DTT, and 250 nM of each primer. The following RT-PCR cycling conditions were used: 45 °C for 45 min followed by 95 °C for 10 min followed by eight cycles of 95 °C for 30 s, 48 °C for 30 s, and 72 °C 30 s with the 48 °C annealing temperature increasing 0.9 °C with each of the eight cycles. The RT-PCR was then continued for 37 additional cycles of 95 °C for 15 s, 56 °C for 20 s, and 72 °C 20 s.

2.6. PCR product purification

Following PCR amplification, amplicon mixtures were rigorously desalted using a protocol based on a weak anion exchange method. A detailed protocol for this approach is published elsewhere [2]. Briefly, amplicons were bound to a weak anion exchange resin directly from the PCR reaction buffer. Unconsumed dNTP's, salts, and other low molecular weight species that might interfere with subsequent ESI–MS analysis were removed by rinsing the resin with solutions containing volatile salts (40 mM NH_4HCO_3) and organic solvents (20% MeOH). Elution of the final purified/desalted PCR products was accomplished by rinsing the resin with an aliquot (typically 25 μL) of a high pH buffer solution (1 M NH_4OH). The final electrospray buffer contained 35% MeOH and 25 mM piperidine/imidizaole [3].

2.7. ESI–FTICR mass spectrometry

The ESI–FTICR mass spectrometer used in this work is based on a Bruker Daltonics (Billerica, MA) Apex II 70e electrospray ionization Fourier transform ion cyclotron resonance mass spectrometer that employs an actively shielded 7 T superconducting magnet. The active shielding constrains the majority of the fringing magnetic field from the superconducting magnet to a relatively small volume. Thus, components that might be adversely affected by stray magnetic fields, such as CRT monitors, robotic components, and other electronics, can operate in close proximity to the FTICR spectrometer. All aspects of pulse sequence control and data acquisition were performed on a 600 MHz Pentium II datastation running Bruker's Xmass software under Windows NT 4.0 operating system. Sample aliquots, typically 15 μL, were extracted directly from 96-well microtiter plates using a CTC HTS PAL autosampler (LEAP Technologies, Carrboro, NC) triggered by the FTICR datastation. Samples were injected directly into a 10 μL sample loop integrated with a fluidics handling system that supplies the 100 μL/h flow rate to the ESI source. Ions were formed via electrospray ionization in a modified Analytica (Branford, CT) source employing an off axis, grounded electrospray probe positioned ca. 1.5 cm from the metalized terminus of a glass desolvation capillary. The atmospheric pressure end of the glass capillary was biased at 6000 V relative to the ESI needle during data acquisition. A counter-current flow of dry N_2 was employed to assist in the desolvation process. Ions were accumulated in an external ion

reservoir comprised of an rf-only hexapole, a skimmer cone, and an auxiliary gate electrode, prior to injection into the trapped ion cell where they were mass analyzed. Ionization duty cycles >99% were achieved by simultaneously accumulating ions in the external ion reservoir during ion detection [4]. Each detection event consisted of 1 M data points digitized over 2.3 s. To improve the signal-to-noise ratio (S/N), 32 scans were co-added for a total data acquisition time of 74 s.

2.8. ESI–TOF mass spectrometry

The ESI–TOF mass spectrometer used in this work is based on a Bruker Daltonics MicroTOF. Ions from the ESI source undergo orthogonal ion extraction and are focused in a reflectron prior to detection. The TOF and FTICR are equipped with the same automated sample handling and fluidics described above. Ions are formed in the standard MicroTOF ESI source that is equipped with the same off-axis sprayer and glass capillary as the FTICR ESI source. Consequently, source conditions were the same as those described above. External ion accumulation was also employed to improve ionization duty cycle during data acquisition. Each detection event on the TOF was comprised of 75,000 data points digitized over 75 μs. To maintain the same 74 s data acquisition time on the TOF as used for FTICR, 980,000 scans were co-added for each mass spectrum. For more rapid analyses (<1 min/spectrum) 660,000 scans were coadded.

The sample delivery scheme allows sample aliquots to be rapidly injected into the electrospray source at high flow rate and subsequently be electrosprayed at a much lower flow rate for improved ESI sensitivity. Prior to injecting a sample, a bolus of buffer was injected at a high flow rate to rinse the transfer line and spray needle to avoid sample contamination/carryover. Following the rinse step, the autosampler injected the next sample and the flow rate was switched to low flow. Following a brief equilibration delay, data acquisition commenced. As spectra were co-added, the autosampler continued rinsing the syringe and picking up buffer to rinse the injector and sample transfer line. In general, two syringe rinses and one injector rinse were required to minimize sample carryover. During a routine screening protocol a new sample mixture was injected every 106 s. More recently we have implemented a fast wash station for the syringe needle which, when combined with shorter acquisition times, facilitates the acquisition of mass spectra at a rate of just under one minute/spectrum.

2.9. Mass-to-charge and amplitude calibration

Mass-to-charge (*x*-axis) calibration is obtained using internal mass standards which are added during the elution step of the desalting protocol. The same peptide-based mass standards are used to calibrate both the FTICR and TOF mass spectrometers.

Amplitude calibration (*y*-axis) is achieved by noting that molecular arrivals in either a frequency bin (FTICR) or time bin (TOF) are Poisson-distributed. Each instrument is calibrated such that individual molecular arrivals (or detections) are counted and scale such that the output distribution is Poisson; this is done by solving for the unique scale factor that causes the mean to equal the variance, drawing upon a fundamental characteristic of the Poisson distribution.

These precise calibrations enable maximum likelihood estimation of molecular abundance at the MS input versus *m*/*z* axis. With use of a PCR calibrant, described in Section 3.4.3, this enables maximum likelihood estimation of input genome abundance for each microbial species present in the input sample.

3. Results and discussion

3.1. Principle of operation

The TIGER process is illustrated in Fig. 1 and is summarized briefly in this section. The process begins (Fig. 1, Step #1) with the extraction of all nucleic acids present in a sample. The resulting nucleic acid material is divided into wells of a micro titer plate that each contain a pair of broad-range primers for PCR. The broad-range primers are designed to amplify a product from a broad group of organisms within a selected domain of microbial life, for example, all bacteria or specific bacterial divisions. The result of the broad-range PCR reactions is typically a mixture of products reflecting the complexity of the original mixture of bacteria or virus present in the starting sample.

The products from the PCR reactions are desalted in a 96-well plate format and sequentially electrosprayed into a mass spectrometer (Fig. 1, Step #2). The spectral signals are processed to determine the masses of each of the PCR products present in each sample with sufficient accuracy that the base composition of adenosines, guanosines, cytidines, and thymidines can be unambiguously deduced, vide infra (Fig. 1, Step #3). Using combined base compositions from multiple PCR reactions, the species of bacteria and their relative concentrations in the starting sample can be determined.

The fundamental distinction between this strategy and traditional (specific) PCR strategies is the nature of the question being asked. Traditional PCR answers the question: “is infectious organism X in my sample?” The TIGER method answers the question: “which infectious organisms are in my sample?” In effect, this is equivalent to running many thousands of specific PCR reactions because the identity of the infectious organism being detected does not need to be anticipated.

What enables this approach is, first, the use of broad-range primers to amplify PCR products from broad groupings of organisms, rather than single organisms, and second, the use of mass spectrometry and base composition determination to analyze the products. Unlike nucleic acid probes or arrays, mass spectrometry does not require anticipation of products analyzed, but simply measures the masses of the nucleic acids present in the sample. The analog signal of mass is converted

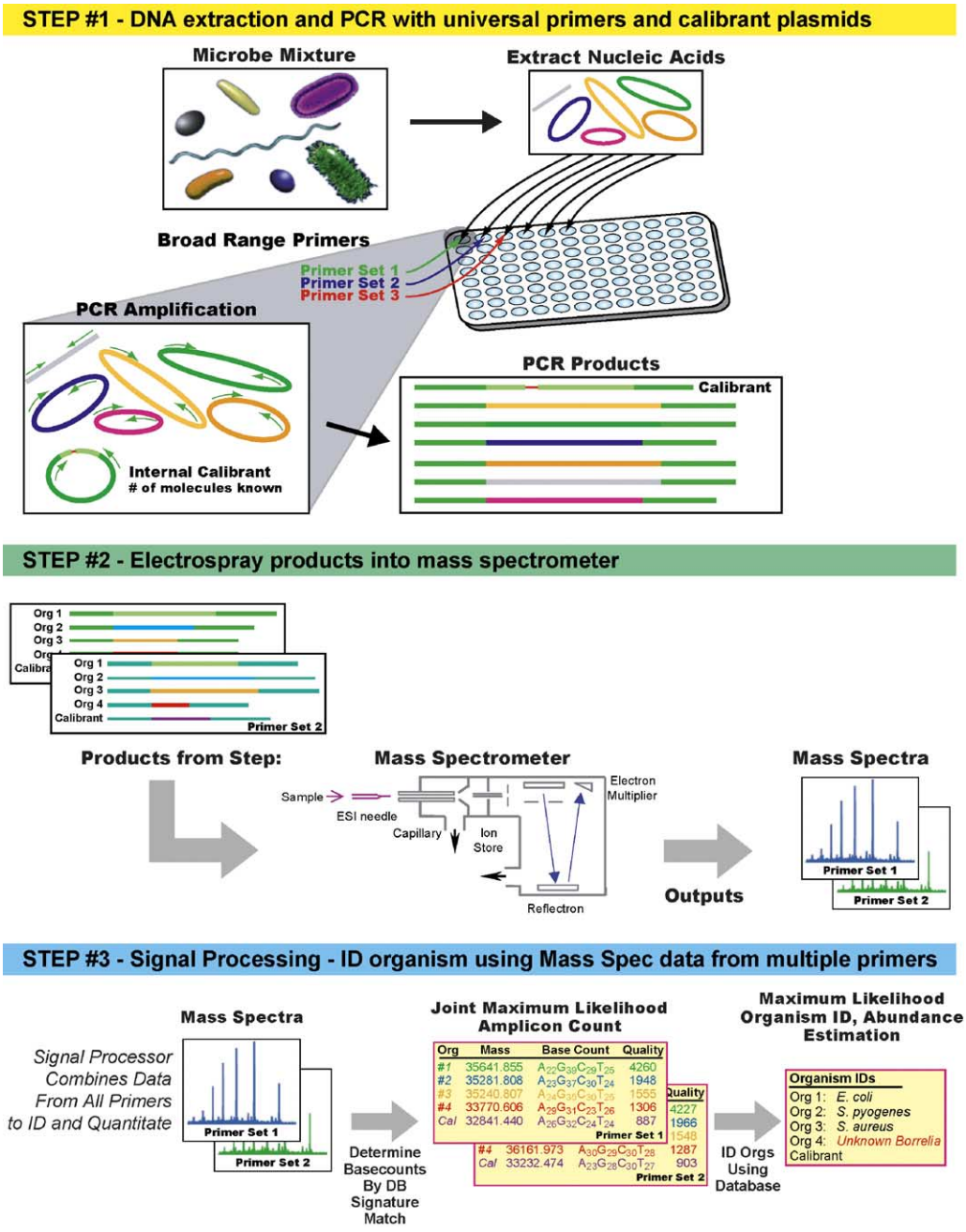

Fig. 1. Step by step TIGER process.

to a digital signal of base composition based upon the accuracy of the mass measurement and the discrete masses associated with different combinations of the four nucleotide bases. Mass spectrometry is remarkably sensitive and can measure the mass and determine the base composition from small quantities of nucleic acids in a complex mixture with a throughput of about one sample per minute. The ability to detect and determine the base composition of a large number of PCR amplicons in a mixed sample enables analysis and identification of broad-range PCR products essentially instantaneously.

3.2. Base composition as a metric of infectious microbes

Base composition is a remarkably useful metric for identification of infectious microbes. A base-composition signature can be thought of as a unique index of a specific gene from a specific organism. Our detection algorithm searches a database that links each sequence for a particular organism to a composition signature so that the presence of the organism can be inferred from the presence of the signature. Base compositions, like sequences, vary slightly from isolate to isolate within species. We have shown that it is possible

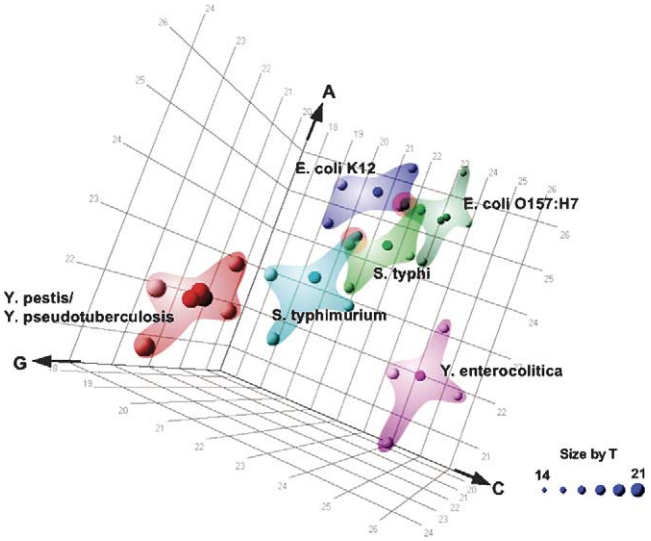

Fig. 2. Base composition cloud for a region of the RNA polymerase B gene from a cluster of enterobacteria. The darker spheres represent the actual base composition of the organisms. The lighter spheres represent the transition changes observed in different isolates of the same organism. Optimal primer design maximizes the separation between different species. Where the clouds overlap are regions that might cause a misclassification. This problem is overcome by selection of additional primer sets that provide orthogonal information resulting in maximum separation of species.

to manage this diversity by building probability "clouds" around the composition constraints for each species. This permits identification of organisms in a fashion similar to sequence analysis, albeit with somewhat lower resolution. A pseudo four-dimensional plot can be used to visualize the concept of base composition clouds (Fig. 2). It is counterintuitive that base composition has sufficient resolving power to distinguish organisms (one might suspect that sequences from different organisms will degenerate to similar overlapping compositions). A rigorous mathematical analysis (Sampath et al., in preparation) has shown, however, that base composition from multiple regions of microbial genomes retains more than sufficient information to identify specific species, provided the target sequences are strategically selected.

It is important to note that, in contrast to probe-based techniques, mass spectrometry-based determination of base composition does not require prior knowledge of the composition in order to make the measurement, only to interpret the results. In this regard, the TIGER strategy is like DNA sequencing and phylogenetic analysis, but at lower resolution. However, the resolution provided by this analysis is more than sufficient to identify the presence or absence of any organism.

3.3. Mass spectrometry of nucleic acids

Mass spectrometric analysis of oligonucleotides would have been considered "heroic" several years ago, but is becoming more and more routine. In fact, in the past 20 years, analysis of nucleic acids has undergone somewhat of a renaissance. In 1981 Macfarlane [5] published a seminal paper in which they acquired a mass spectrum of a 20-mer DNA oligonucleotide using ^{252}Cf desorption and TOF detection. The spectrum, clearly the highest quality mass spectrum of an oligonucleotide acquired at the time the work was published, exhibited a mass resolution of about 25 (FWHM) and boasted a ±5 Da uncertainty (± 800 ppm). In contrast, we recently demonstrated that, after a high-throughput purification protocol, ESI–FTICR spectra of ∼120 bp DNA oligonucleotides could be acquired at a rate of ∼1 spectrum/min with mass measurement errors of ∼1 ppm (internally calibrated) with a concomitant average resolving power exceeding 150,000 (FWHM) [2].

Amplification of specific DNA sequences utilizing the polymerase chain reaction has widespread applications in many scientific disciplines including microbiology, medical research, forensic analysis, and clinical diagnostics [6,7]. Most often, PCR products are "sized" using traditional biochemical techniques such as standard gel electrophoresis using either intercalating dyes or fluorescently labeled primers. The Taqman assay, widely used in a number of PCR-based diagnostic kits, confirms the presence (or absence) of a specific PCR product but provides no direct read out on the size of the amplicon [8]. So called "real-time" PCR devices, which measure the laser induced fluorescence of the PCR product during the amplification cycles, are used to quantify the amplification of DNA from a given DNA template and primer set [9–11]. These methods have limited utility for amplicons of less than 150 bp due to the proportionately high fluorescence background, and do not provide any information with respect to amplicon heterogeneity or exact length.

Compared to these traditional methods, mass spectrometry has several potential advantages as a platform for characterization of PCR products including speed, sensitivity, and mass accuracy [12–16]. Mass spectrometry has been shown to be viable platform for analysis of single nucleotide polymorphisms (SNPs); in recent years MALDI–TOF has emerged as the "gold standard" for SNP counting and is at the core of several large-scale SNP counting projects [17]. Since the exact mass of each of the bases which comprise DNA are known with great accuracy, a high precision mass measurement obtained via mass spectrometry can be used to derive a base composition (or constrained list of base compositions) within the experimentally obtained mass measurement uncertainty [18,19]. This list of possible base composition can be further constrained by taking into account base complementarity (the two strands of the PCR product are by definition complementary thus the number of G's in one strand must equal the number of C's in the other strand) and the known base composition of the primers (e.g. a primer with five T's can only produce an amplicon containing five or more T's).

3.4. Example 1: detecting bacteria

3.4.1. Triangulation principles

TIGER is based upon the principle that, despite the enormous diversity of microbes, all forms of life on earth share sets of essential common features in the biomolecules encoded in their genomes. Bacteria, for example, have highly conserved sequences in a variety of locations on their genome. Most notable is the universally conserved region of the ribosome, but there are also conserved elements in other non-coding RNAs, including RNAse P and the signal recognition particle among others. There are also conserved motifs in essential protein-encoding genes.

The basis of TIGER is the use of these common, conserved features as anchors for broad-range PCR priming to generate amplicons from all organisms in an environmental or clinical sample without prejudice. The trade-off in broad-range priming compared to specific PCR is that PCR is a zero-sum game. The total yield of amplified product has an upper limit value that must be divided among all the targets amplified. It is essential that the technology detects and identifies the signal from the threatening organism in the background of an excess of harmless organisms. While cloning and exhaustively sequencing many colonies can solve this, this cannot be done in a rapid diagnostic device. The strategic breakthrough in TIGER is the use of mass spectrometry to analyze the products of broad-range PCR. Mass spectrometry is remarkably sensitive and can measure the weight and determine the base composition from small quantities of nucleic acids in a complex mixture in a high throughput modality. The ability to detect and determine the base composition of a large number of PCR amplicons in a mixed sample enables analysis and identification of broad-range PCR products essentially instantaneously.

3.4.2. Primers

A schematic of the primer selection and validation process is outlined in Fig. 3. Public bacterial genome sequencing projects have provided extraordinarily valuable data that has led to a basic understanding of both the breadth of diversity and the common features shared by bacteria. Many of the important pathogens, including the organisms of greatest concern as biological weapons agents, have been completely sequenced. This effort has greatly facilitated the design of primers and probes for the detection of bacteria. Using full-length sequences from over 225 bacterial genomes, we have generated alignments of the essential genes that are conserved either broadly across all organisms or within members of specific, related phylogenetic groups. In bacteria, for instance, we have alignments from over 160 housekeeping genes that are present in almost all major bacterial divisions (Fig. 4). These genes have been used for identification of broad diagnostic primers. PCR primer selection and optimization has been largely automated. A number of genes, in addition to 16S rRNA, are targets of “broad-range” primers, thus increasing the redundancy of detection and classification, while minimizing potential missed detections (Fig. 5). Many of these genes are, expectedly, essential to information processing, with more than half being associated with the translational machinery, such as elongation factors, ribosomal proteins, and tRNA synthetases. Other classes of conserved protein-encoding genes include transcription-associated genes such as RNA polymerases and genes associated with DNA replication such as DNA gyrase and DNA polymerase. This combination of broad-range priming with clade-specific priming has been used very successfully in several applications of the TIGER technology, including environmental surveillance for biowarfare threat agents and clinical sample analysis for medically important pathogens described later. The primer

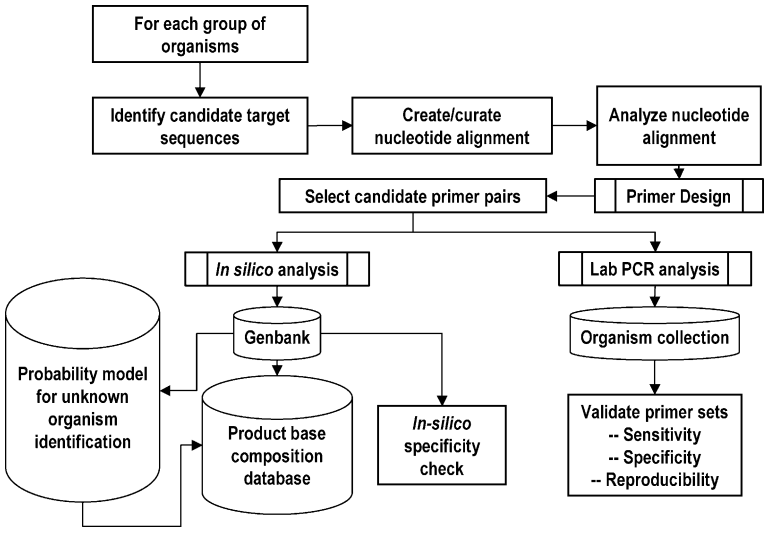

Fig. 3. TIGER primer selection and validation process. Most of the steps in the process have been automated and are routinely used in primer selection and analysis. The choice of group of organisms targeted by primer selection depends on specific applications. Primer validation and tests for sensitivity and specificity are performed on the same instrumentation and using the same analysis conditions as will be done in the final tests, assuring reliable and reproducible results.

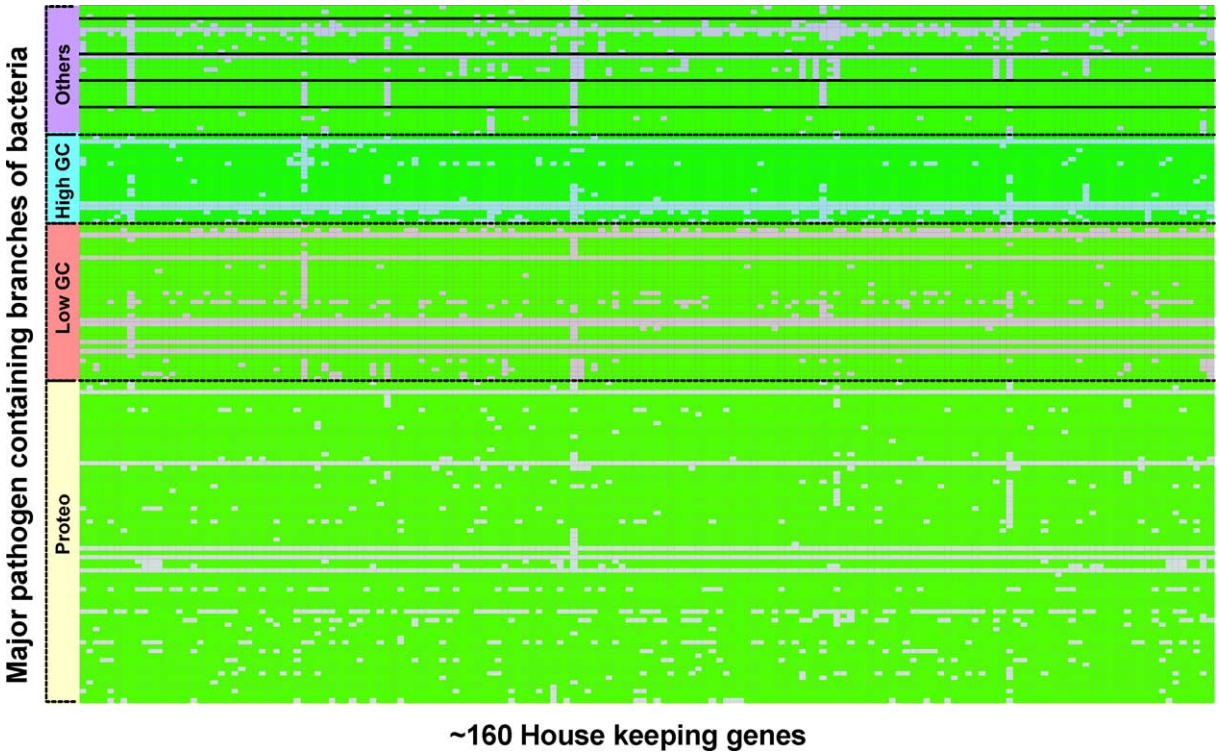

Fig. 4. Heat map of 160 bacterial housekeeping genes (abscissa) and approximately equal number of complete or nearly complete bacterial genomes.

Primer #	Gene target	Bacterial target	Primer specificity
346, 347, 348, 361	16S rDNA	ALL	Universally conserved ribosomal genes
349, 360	23S rDNA		
354	RNA polymerase, β' subunit (rpoC)	Bacteroidetes, Fusobacteria, Spirochaetes, Proteo, Bacilli	Division-wide Housekeeping genes
358	Valyl-tRNA synthetase (valS)	Proteobacteria (γ: Enterobacteria)	
359	RNA polymerase, β subunit (rpoB)	Proteobacteria (γ: Enterobacteria)	
362	RNA polymerase, β subunit (rpoB)	Proteobacteria (α, β)	
363	RNA polymerase, β' subunit (rpoC)	Proteobacteria (β, γ)	
367	Elongation factor EF-Tu (tufB)	Proteobacteria (β)	
356, 449	Ribosomal protein L2 (rplB)	Clostridia. Fusobacteria Bacilli, Proteobacteria (ε)	
352	Protein chain initiation factor (infB)	Bacilli	
355	Spore protein (sspE)	*Bacillus cereus* clade	Clade-specific genes

Fig. 5. Description of primer target genes and their breadth of coverage.

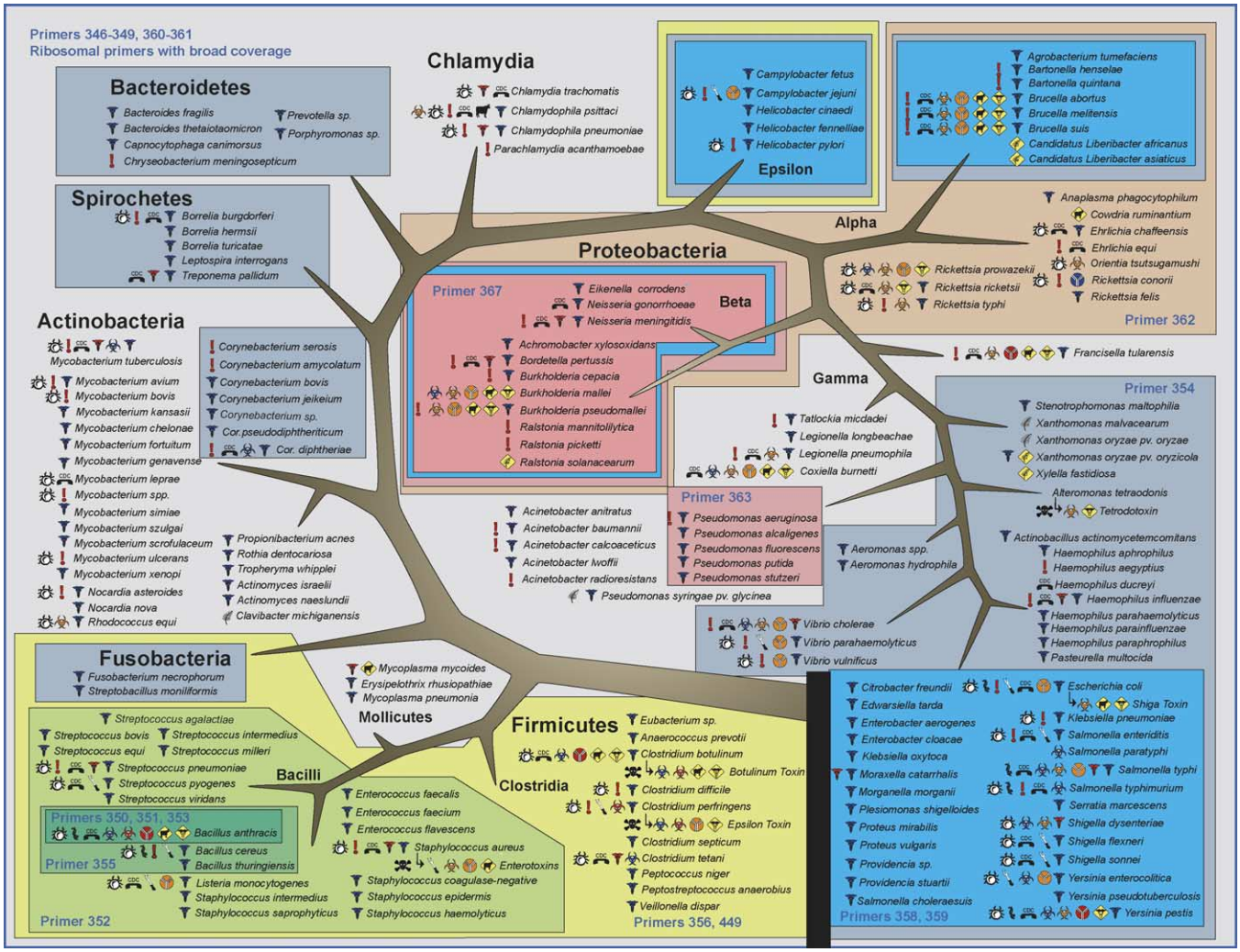

Fig. 6. Graphic representation of broad-range primers and their breadth of coverage.

coverage map of bacterial phylogeny is graphically illustrated in Fig. 6.

We have developed a strategy to survey environmental and clinical samples that allows detection and identification of all bacteria for which we have sequence information (Fig. 7). A set of 12 broad-range PCR primers was used. Six of the 12 primers are "universal" primers, targeted to universally conserved sequences and six are "division-wide" primers, targeted to broad divisions of bacteria (i.e. Bacillus/Clostridia group or gamma-proteobacteria). The division-wide primers have more focused coverage and tend to provide higher species resolution. Using these 12 primers, >98% of all known bacteria can be identified at the species level. Using additional primers, in what we term the *survey/drill-down* strategy, additional information about identified organisms, such as antibiotic susceptibility, virulence, strain type, etc. can be obtained. However, this information is only necessary once an organism is detected, and its nature varies with the organism. Using a recursive *survey/drill-down* strategy, detection, confirmation, and additional information can be provided within hours. Moreover, the drill-down strategy can be focused to identify bioengineering events, such as the insertion of a toxin gene into a bacterial species that does not normally make the toxin. Effectively, the TIGER technology provides a digital barcode in the form of a series of base composition signatures, the combination of which is unique for each known organism. This capability enables real-time infectious disease monitoring across broad geographic locations, which may be essential in a simultaneous disease outbreak or attack in different cities.

3.4.3. Calibrant

A key element of the TIGER system is the ability to quantify the levels of genomic material present in the samples. To accomplish this, we include an internal calibrant in each PCR reaction performed. The calibrant consists of a specially designed nucleic acid sequence that is similar to, but absolutely distinguishable from, any potential sequence we might to detect. Specifically, a calibrant sequence is designed and synthesized for each primer pair tested in the system and

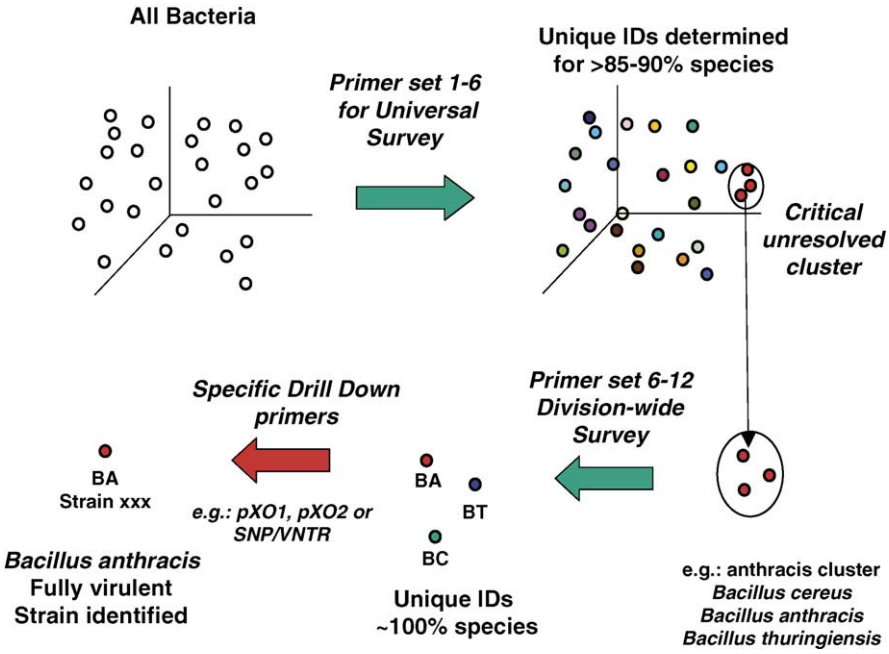

Fig. 7. The survey/drill-down strategy. Twelve pairs of PCR primers are used simultaneously to parse bacteria to near 100% resolution at the species level. Following identification of specific organisms, specific drill-down primers are applied to the sample to determine additional information as desired.

inserted into a plasmid. Initially, alignments of all available sequence information are assembled and a calibrant sequence is designed that is identical to the desired target organism except for a two to five nucleotide deletion internal to the amplified region of the target sequence (Fig. 8A). These calibrant molecules will amplify with an efficiency similar to that of target sample molecules since the sequences are nearly identical. Each reaction contains an amount of calibrant designed to be detected without overwhelming the signal from an organism present at low concentration. Following amplification of the target sample and the calibrant present in the reaction, a direct comparison of the resultant product levels will lead to an estimate of the quantity of the unknown element. Due to the length differences of the calibrant and the target sequences, they will be clearly separated in the mass spectrum (Fig. 8B). The relative abundances of the end products (as measured by the peak heights in the mass spectrum) can be used in conjunction with the starting concentration of the calibrant to get an accurate estimate of the amount of target sequence in the starting sample. These measurements are highly reproducible over a wide range of concentrations (Fig. 8C) and provide improved confidence for the determination of the quantity of the organisms in the sample. The calibrant molecules in each reaction serve several purposes. Foremost is the use as a quantifying element in the reaction. Another is the use of the molecule as an internal positive control. Even in the absence of any input genome, the calibrant should always amplify and be detected. Failure to detect the calibrant in this type of sample indicates a failed PCR reaction, a failed desalting protocol, or a malfunction in the autosampler or mass spectrometer. Calibrant may also not be detected in samples with large quantities of amplifiable DNA that out-compete the calibrant, but in these cases the presence of the product indicates an obviously successful PCR reaction, but quantitation can only be reported as >100–1000 fold the calibrant concentration.

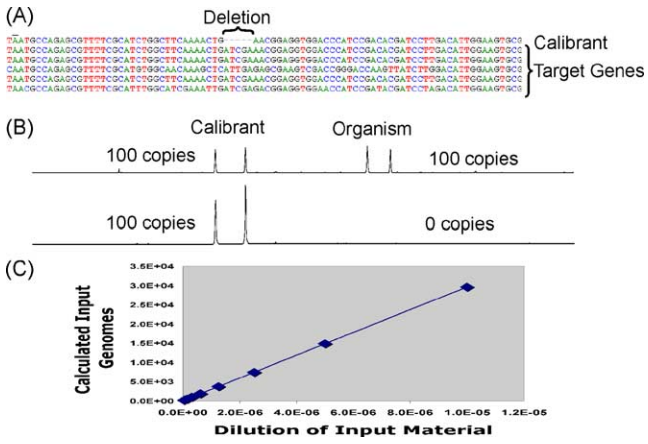

Fig. 8. Using calibrated PCR reactions to quantitate input material levels in unknown samples. (A) Design of calibrant molecules for use in PCR reactions. Calibrant molecules are designed to produce a unique amplicon using alignments of all known potentially amplified species. (B) Use of calibrant molecules in a PCR reaction. PCR reactions containing 100 molecules of calibrant were spiked with either 0 or 100 molecules of input material. The amplitude of the calibrant and the input material can be compared and the quantity of input material in the original sample can be accurately estimated. (C) Linearity of input material quantitation. As described in the text, PCR reactions containing a series of dilutions of unknown material were performed in the presence of 1200 molecules of calibrant. For each dilution, the amount of input material was calculated and plotted vs. the dilution used. The linearity of the plot indicates that this method of quantitation can be used over a wide range of input sample material.

3.5. Example 1: bacteria—B. anthracis

The mass accuracy provided by the FTICR and TOF mass spectrometers limits the base composition of each strand to a finite number of possibilities. As demonstrated by Aaserud et al. [20] and by Muddiman et al. [19], high precision mass measurements can be used to unambiguously derive base compositions of PCR products. For example, Fig. 9 shows mass spectra of PCR products derived from a sample containing *B. anthracis* and 100 copies of the plasmid calibrant amplified using primer pair 356 (see Section 2). Spectra from identical aliquots were analyzed on a 7 T ESI–FTICR mass spectrometer (Fig. 9a) and on the ESI–TOF mass spectrometer (Fig. 9b). As shown in the inset in Fig. 9c, the FTICR provides isotopic resolution with an average resolution of 150,000 (FWHM) while the TOF spectrum is not isotopically resolved. The TOF spectrum exhibits an average resolution of 3500 (FWHM), which approaches the theoretical limit achievable for a non-resolved isotope distribution for a species in this molecular weight regime. The addition of low molecular weight peptides that bracket the *m*/*z* range of interest provide a straightforward internal calibrant for post-calibration. For the FTICR spectra the monoisotopic molecular weights for the complementary strands, derived using an "averageine-like" fitting routine [21], were determined to be 37374.226 and 37231.153 Da with an average mass measurement uncertainty of ∼1.5 ppm (∼0.045 Da).

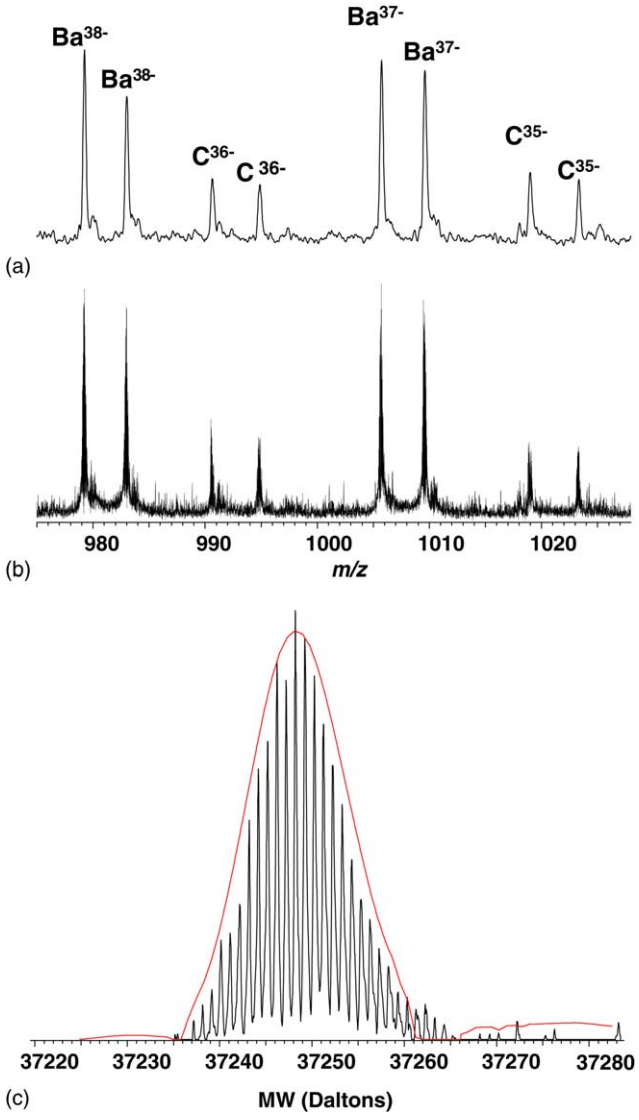

Fig. 9. ESI–MS spectra of PCR products derived from 1500 copies of *Bacillus anthracis* in the presence of 100 copies of the plasmid calibrant using primer 356 (see Table 1). Spectra were acquired using (a) ESI–TOF and, (b) ESI–FTICR mass spectrometers following a rigorous desalting protocol (see text). Accurate mass measurements of the complementary strands allow unambiguous base composition determination of the *Bacillus anthracis* (Ba) amplicon. The internal plasmid calibrant (C) is used to calculate the relative gain of the PCR reaction. As illustrated by the deconvoluted spectrum of one of the amplicon strands in (c), the FTICR spectrum exhibits a resolving power of ∼150,000 (FWHM) while the effective resolution of the non-isotopically resolved TOF spectrum is ∼3500 (FWHM).

Table 2 lists the number of base compositions consistent with determined molecular weights within a range of mass measurement uncertainties from 1 to 100 ppm. A mass measurement of either strand by itself, even at 1 ppm mass measurement error is consistent with more than 100 base compositions for each strand, while a 20 ppm mass measurement error would yield more than 1200 base compositions. Taking into account the fact that the base compositions of the two strands are complementary, the list of putative base compositions can be culled leaving only possibilities in which the base composition of strand 1 is complementary to that of strand 2. Although 1200 base compositions are consistent with the measured masses of each strand at an error of 20 ppm, only one combination of base compositions strand 1 and strand 2 is consistent when complementarity is considered. Importantly, recent advances in ESI–TOF mass spectrometry facilitate the acquisition of relatively high-resolution mass spectra (>10,000 FWHM) with excellent mass accuracy (2–5 ppm for isotopically resolved species with an internal mass standard). As illustrated in Fig. 9b and C, ESI–TOF spectra of PCR products in this molecular weight range are not isotopically resolved, but using a low molecular weight internal standard we are able to acquire high quality mass spectra with mass measurement errors in the 5–10 ppm regime. As illustrated in Table 2, this mass accuracy is more than adequate to derive unambiguous base compositions for amplicons in the size range employed by TIGER. We are currently designing an integrated TIGER system that employs the Bruker MicroTOF detector. Such an integrated system would be significantly reduced in size compared to the existing laboratory-based system and would be amenable to installation in a clinical diagnostics or hospital laboratory.

An additional advantage of the TOF platform for this application is that it is less prone to problems when analyte concentrations are high than is the FTICR platform. At relatively high analyte concentrations (i.e. high ion numbers in the FTICR trapped ion cell) the accuracy of FTICR measurements can be adversely affected leading to poor mass

Table 1
Forward and reverse primer sequences for broad-range PCR primers described in Section 3.4.2 are listed in the following table

Primer #	Forward primer	Reverse primer
346	TAGAACACCGATGGCGAAGGC	TCGTGGACTACCAGGGTATCTA
347	TGGATTAGAGACCCTGGTAGTCC	TGGCCGTACTCCCCAGGCG
348	TTTCGATGCAACGCGAAGAACCT	TACGAGCTGACGACAGCCATG
349	TCTGTTCTTAGTACGAGAGGACC	TTTCGTGCTTAGATGCTTTCAG
352	TTGCTCGTGGTGCACAAGTAACGGATATTA	TTGCTGCTTTCGCATGGTTAATTGCTTCAA
354	TCTGGCAGGTATGCGTGGTCTGATG	TCGCACCGTGGGTTGAGATGAAGTAC
355	TCAAGCAAACGCACAATCAGAAGC	TTGCACGTCTGTTTCAGTTGCAAATTC
356	TGACCTACAGTAAGAGGTTCTGTAATGAACC	TTCCAAGTGCTGGTTTACCCCATGG
358	TCGTGGCGGCGTGGTTATCGA	TCGGTACGAACTGGATGTCGCCGTT
359	TTATCGCTCAGGCGAACTCCAAC	TGCTGGATTCGCCTTTGCTACG
360	TTTAAGTCCCGCAACGAGCGCAA	TTGACGTCATCCCCACCTTCCTC
361	TCTGACACCTGCCCGGTGC	TGACCGTTATAGTTACGGCC
362	TGGGCAGCGTTTCGGCGAAATGGA	TGTCCGACTTGACGGTCAACATTTCCTG
363	TCAGGAGTCGTTCAACTCGATCTACATGAT	TACGCCATCAGGCCACGCAT
367	TCCACACGCCGTTCTTCAACAACT	TGGCATCACCATTTCCTTGTCCTTCG
449	TCCACACGGTGGTGGTGAAGG	TGTGCTGGTTTACCCCATGGAG

measurement accuracy. Alternatively, the ESI–TOF platform is relatively immune to these space charge affects and can effectively analyze a broader range of analyte concentrations. This allows high quality, high signal to noise spectra to be acquired in rapid succession on the TOF platform. Fig. 10 depicts the ESI–TOF spectra of a calibrated *B. anthracis* sample. ESI–TOF spectra were acquired at a rate of ∼1 spectrum/min (50 s of spectral co-adding, 9 s of autosampler overhead). Each spectrum was derived from a different primer pair (Table 1) and contains both plasmid calibrant and *B. anthracis* genome.

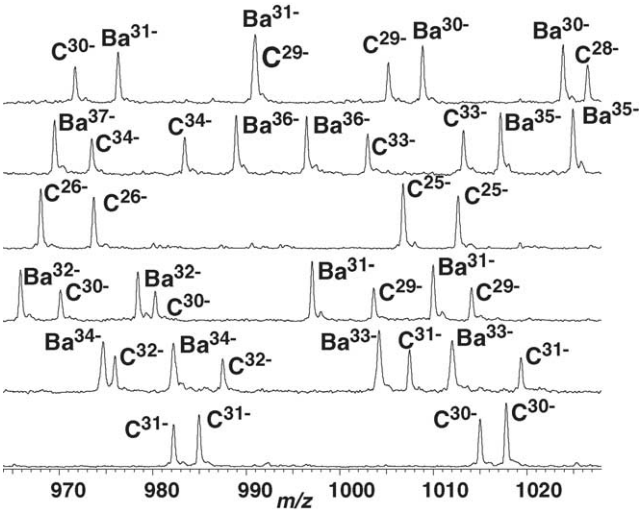

Fig. 10. High throughput ESI–TOF analysis of *Bacillus anthracis* (Ba) present at 1500 copies/well in the presence of plasmid calibrant (C). Spectra were acquired at 1 min intervals in which 50 s of spectral acquisition is preceded by 9 s of autosampler overhead and between-sample buffer rinses. From top to bottom primers employed were 346, 347, 359, 349, 361, and 367 (see Table 1). Note that primers 359 and 367 are not expected to amplify Ba and produce only calibrant amplicons.

3.5.1. TIGER sensitivity

Since the TIGER assay relies on polymerase chain reaction amplification, its inherent sensitivity is similar to other assays that likewise incorporate PCR. As long as at least one genome copy is present in the PCR reaction well, the amplification gain is sufficient to produce measurable product. Sensitivity is limited by material transfer efficiency of all processes upstream of PCR. For TIGER these include filter capture efficiency, recovery efficiency, spore/cell lysis, pipette transfer, and nucleic acid isolation.

We have conducted dilution-to-extinction experiments to verify the TIGER sensitivity and from this test have estimated the combined efficiency of spore/cell lysis, pipette transfer, and nucleic acid isolation. The tests comprised 576 independent PCR reactions and mass spectra: 32 repetitions of each of 6 organism dilution levels for three different primer sets. *B. anthracis* (Sterne 34F2 vaccine strain) spores from Colorado Serum at an initial concentration of 5×10^6 mL^{-1} were used. The organism concentration of the suspension was independently estimated by Picogreen evaluation post-lysis DNA concentration.

The spore suspension was processed by manually filtering, bead beating, and extracting nucleic acids as described

Table 2
Number of base compositions consistent with molecular weight as a function of mass measurement uncertainty

ppm	Strand 1 (33374.266 Da)	Strand 2 (37231.153 Da)	Complementary pairs
1	101	130	1
5	519	631	1
10	933	934	1
20	1321	1214	1
50	3703	3524	20
100	7377	7179	81

Note that at 20 ppm (or better) the base composition of the amplicon pair is constrained to a single, unique base composition ($A_{34}G_{31}C_{29}T_{27}/A_{27}G_{29}C_{31}T_{34}$) due to the fact that the base composition of strand 1 must be complementary to that of strand 2.

in Section 2. Dilutions of the isolated DNA were prepared to be added into the PCR reaction wells to give 1000, 100, 10, 1, 0.1, and 0 spore equivalents per well, as measured in the original suspension. Three different primer pairs that amplified chromosomal protein synthesis gene (sspE) and protein synthesis genes cya and lef were used. Genes cya and lef are contained on the px01 plasmid. Each organism of this Ba strain has one copy of the sspE gene and three copies of the px01 plasmid, and, therefore, three lef and three cya template copies. Six 96-well microtiter plates were prepared with all necessary reagents (e.g. buffer, primers, magnesium salts, dNTPs, etc.) The isolated DNA and polymerase were added just before PCR amplification as described in Section 2. Desalting and FTICR analysis was performed as described above.

The mass spectra for each well were analyzed independently using a maximum-likelihood processor. This processor, referred to as GenX, first makes maximum likelihood estimates of the input to the mass spectrometer for each primer by running matched filters for each base composition aggregate on the input data. This includes the GenX response to a calibrant for each primer. The amplitudes of all base compositions for each primer are calibrated and a final maximum likelihood amplitude estimate per organism is made based upon the multiple single primer estimates. Models of all system noise factor into this two-stage maximum likelihood calculation. The processor reported the number of molecules of each base composition contained in the spectra. The quantity of PCR product corresponding to the appropriate primer set was reported for each well. The quantities of primers remaining in each well are also reported. The amount of PCR product was compared to a relatively low threshold to allow for detection of wells containing only a single copy of the *B. anthracis* genome. This binary statistic is used in the analysis as described below.

When equal aliquots are taken from the DNA stock solutions and distributed to wells for PCR, the actual number of templates deposited in each well varies about a mean value according to a Poisson distribution. The most useful data is produced at levels sufficiently dilute that, by chance, a moderate fraction of the wells contain no genomic templates. Given a starting concentration, c; pre-PCR process efficiency, η; and volume, v; the average number of genomes present in each well is, $\lambda = \eta\, c\, v$. The probability of having exactly m molecules in each well is given by a Poisson probability distribution:

$$P(x = m) = \frac{e^{-\lambda}\lambda^{m}}{m!}, \quad \lambda > 0, m = 0, 1, 2, \ldots$$

Since data analysis distinguishes one or more genome copies per well (expected PCR product measured) from zero copies per well (i.e. no product expected), the above can be reduced to a more appropriate expression giving the probability of having a product, P(product), for a given concentration, λ (specifically, one minus the Poisson probability for m = zero):

$$P(\text{product}) = 1 - e^{-\lambda}$$

The binary data analysis results for each well are simply Bernoulli trials; therefore, the statistical properties for each dilution can be characterized by Binomial probability distributions. Data error bounds can be evaluated as a function of the number of trials and the number of wells exceeding the threshold.

Fig. 11a presents a comparison between the theoretical probability and measured amount of PCR products for the plasmid genes verses the nominal number of organisms per well ($c\,v$). The theoretical probability curves, based on Poisson distributions and considering that there are three copies of the PX01 plasmid per organism, are shown for material transfer efficiencies (η) of 100% (blue) and 50% (green) from the point of sample insertion to the PCR well. The experimental data are plotted for the *lef* and *cya* genes separately. The data points are x/n where x wells have product detected out of n trials. Error bars were evaluated for a 20–80 percentile or a 60% confidence window for each point. Note that the errors are greater for the lower dilutions with fewer detections. The data for 10, 100, and 1000 spore equivalents per well validates the test; when the probability of having no templates was very low, we observed N out of N successful PCR amplifications. The only cases with no product were for the most dilute samples. The dilution to extinction statistical result corresponds very closely, within the 60% confidence interval, to the theoretical curve for 50% efficiency.

Fig. 11b presents a similar plot for the *sspE* gene, with one copy per organism. The theoretical probability and measured frequency of producing the PCR amplicon is compared to the nominal number of organisms per well. Except for the leftmost point (one-tenth organism per well) the data corresponds very closely to the theoretical curve for 50% efficiency. The lowest concentration result falls slightly beyond the standard error bars, corresponding to the 5th percentile. In summary, the combined efficiency of spore/cell lysis, pipette transfer, and nucleic acid isolation is approximately 50% for the TIGER sample processing protocol.

3.6. Example 2: DNA genome viruses—orthopox

Viral identification poses a different kind of challenge than bacterial identification. While we have over 25,000 viral sequences covering important pathogenic viruses in our database, no single gene is essential and conserved across all viral families. Therefore, viral identification is achieved within smaller groups of related viruses, for instance across all members of a particular viral family or genera. The RNA polymerase family of genes, including the retroviral reverse transcriptase (RT) and RNA-dependent RNA polymerases, are present in all single-stranded RNA viruses. We have assembled family-specific alignments of these polymerases and have shown that we can achieve broad priming as well as resolution within these families. We have also collected and

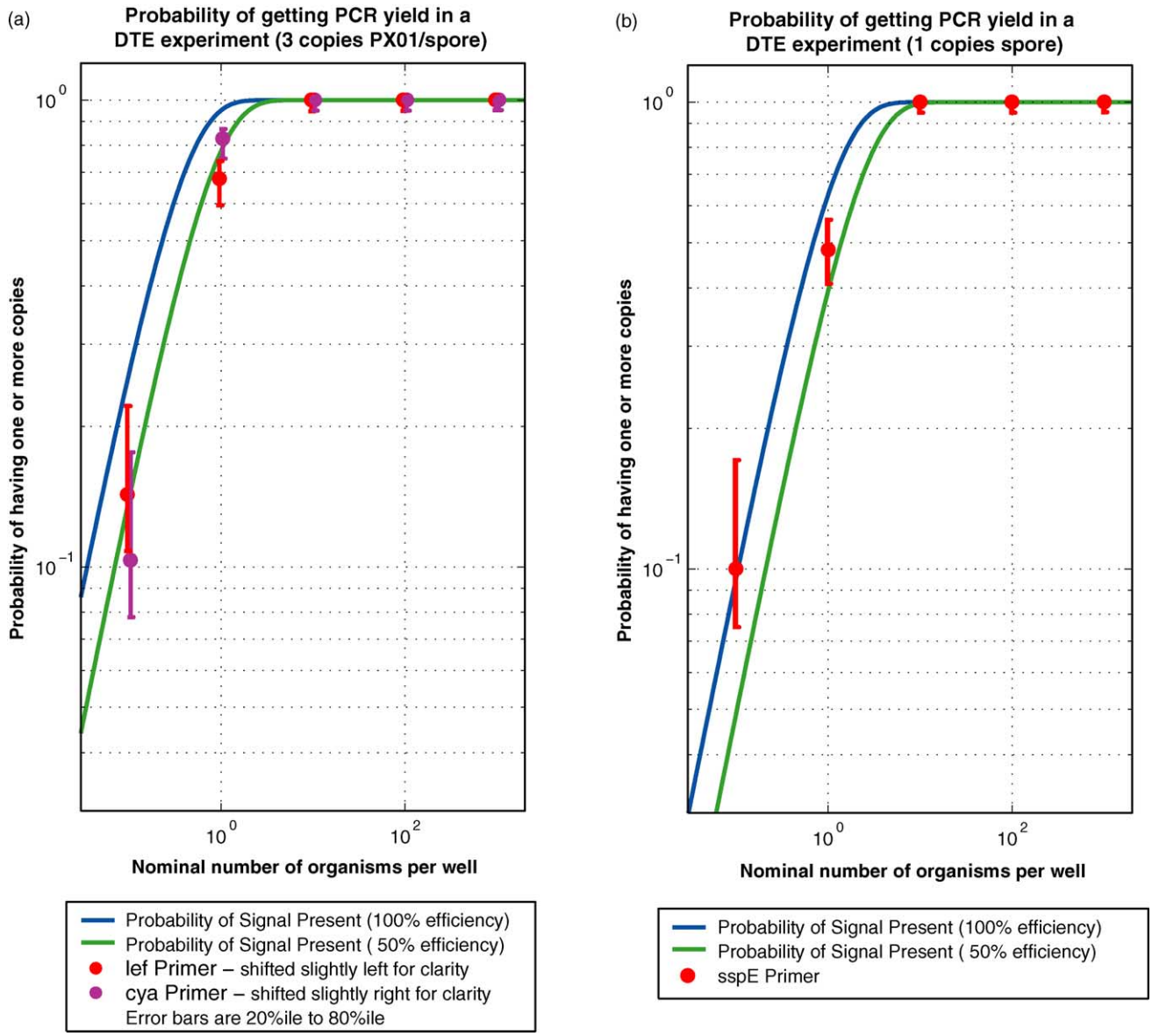

Fig. 11. (a) Theory–data comparison for sensitivity determination using two *B. anthracis* virulence plasmid genes (PX01). (b) Theory–data comparison for a *B. anthracis* chromosomal gene (sspE).

aligned other gene targets that could provide resolution to the level of strain typing. Some specific examples of viral family detections are described in sections below.

Smallpox is caused by a DNA virus, the variola virus, which is a member of the genus Orthopoxvirus. The Orthopoxvirus genus contains several potential biowarfare agents, including smallpox (variola), and monkeypox viruses and other human and animal pathogens such as vaccinia virus, rabbitpox, cowpox, camelpox, and ectromelia virus. Several members of this group have very similar nucleic acid and amino acid sequences (for instance vaccinia and variola share >95% similarity in their nucleotide sequence). A system that can rapidly identify and reliably distinguish these species from one another would have significant impact on both BW detection and clinical diagnostics. Using the TIGER process described above, we have developed multiple PCR primers to essential conserved genes across all members of the Orthopoxvirus genus group that can be used to identify all of the species mentioned above and their variants, and distinguish them from other members of the *Poxviridae* family by base composition analysis.

We analyzed all available sequences for Orthopox viruses (GenBank, Poxvirus DB [22]) to select suitable primer regions for TIGER analysis and identified primer sites on several essential viral genes (DNA polymerase, RNA polymerase, DNA helicase and RNA helicase) that would enable unambiguous detection and identification (Fig. 12). TIGER primers are designed across regions of high conservation, the likelihood of missed detection due to sequence variations at these sites is minimized, and essentially reduced to zero, by

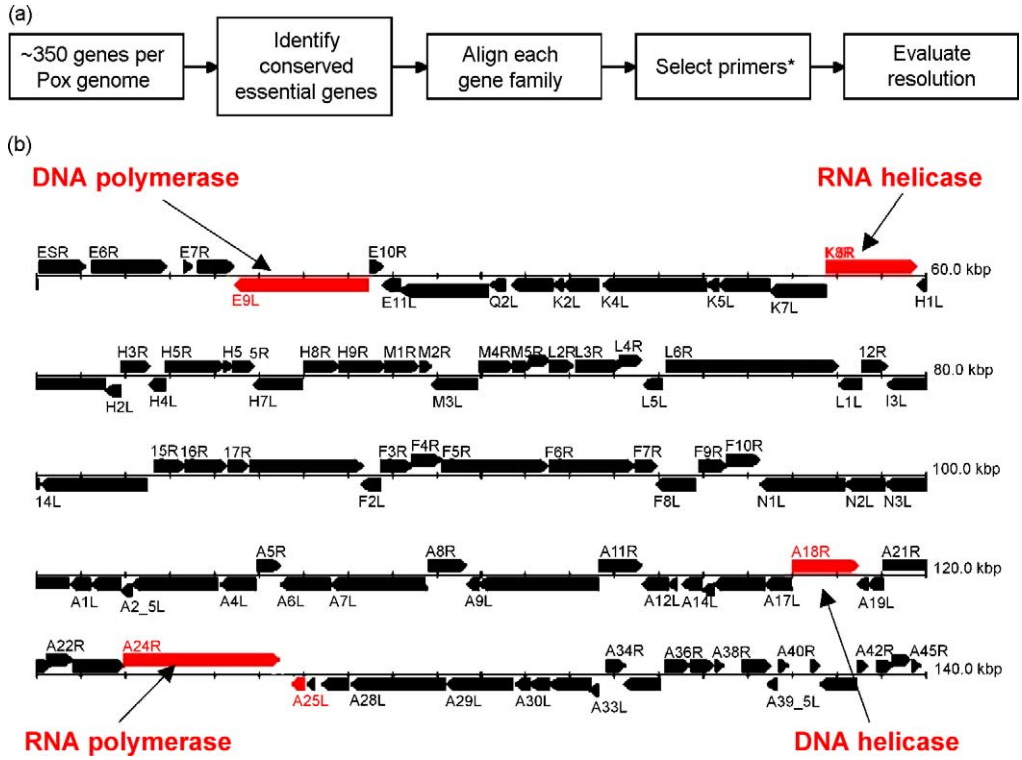

Fig. 12. (a) Orthopoxvirus conserved gene and primer selection strategy and (b) loci.

using multiple targets. The target amplicons in the intervening region between the conserved primers provide the unique signature needed to identify organisms. None of the primers are expected to amplify other viruses or any other known DNA. All the primers were tested against all available sequences from GenBank using electronic PCR (e-PCR) for potential background priming (data not shown).

We obtained DNA from five different Orthopoxvirus species from the laboratory of Dr. Chris Upton at UVIC: monkeypox (MPXV-VR267), cowpox (BR), rabbitpox (Utrecht), vaccinia (WR) and ectromelia (Moscow); we were unable to obtain a variola virus DNA sample. Results from amplification of the DNA polymerase and the DNA helicase genes from each of the available species are shown in Fig. 13a and b. PCR products were generated from each of the test viruses using the above-described primers, desalted, and analyzed by mass spectrometry. Deconvolved spectral signals showing the sense and antisense strands of the PCR products from each sample are shown in Fig. 13c and d. These spectra were processed by an algorithm that converts mass spectrometry signals to base composition data as described previously. All detected masses could be unambiguously mapped to specific base compositions, which were compared to the pre-compiled database of expected products from each of these viruses.

Fig. 13e and f show the deconvoluted base compositions (solid cones) of the experimentally measured spectra in a four-dimensional plot (A, G, C axes, with the T counts represented by the tilt of the cone), overlaid on the expected base count distributions (hollow spheres) of the base compositions from all available Orthopoxvirus species. The TIGER derived base compositions agreed with compositions expected from the sequences in GenBank for all five viruses tested. Vaccinia and ectromelia viruses gave expected products consistent with the database sequence entry in each primer region. In the case of the rabbitpox virus, the sequence of the target region was identical to vaccinia virus in all primer regions selected (these two species are >98% identical across their entire genome, Dr. Upton, personal communication), and were not distinguished by these primers. At the time of primer design, the only strain of monkeypox virus deposited in GenBank was the Zaire 96_I-16 strain. TIGER-determined base compositions for the MPXV-VR267 strain were different from those for the Zaire strain. The experimentally determined based-counts were subsequently validated by comparison to the full genome sequence for the VR267 strain (unpublished data obtained from Dr. Upton). Thus we were able to correctly identify a new variant of a known Orthopoxvirus species with the same technology used for primary detection, without additional primer design or analysis. Finally, while we were analyzing these test species, the whole genome sequence for a new strain of cowpox, the GRI-90 strain, was published [23]. Analysis of several conserved genes across all of the Orthopoxvirus genera revealed that the GRI-90 strain was more closely related to vaccinia strains than it was to the previously known Brighton

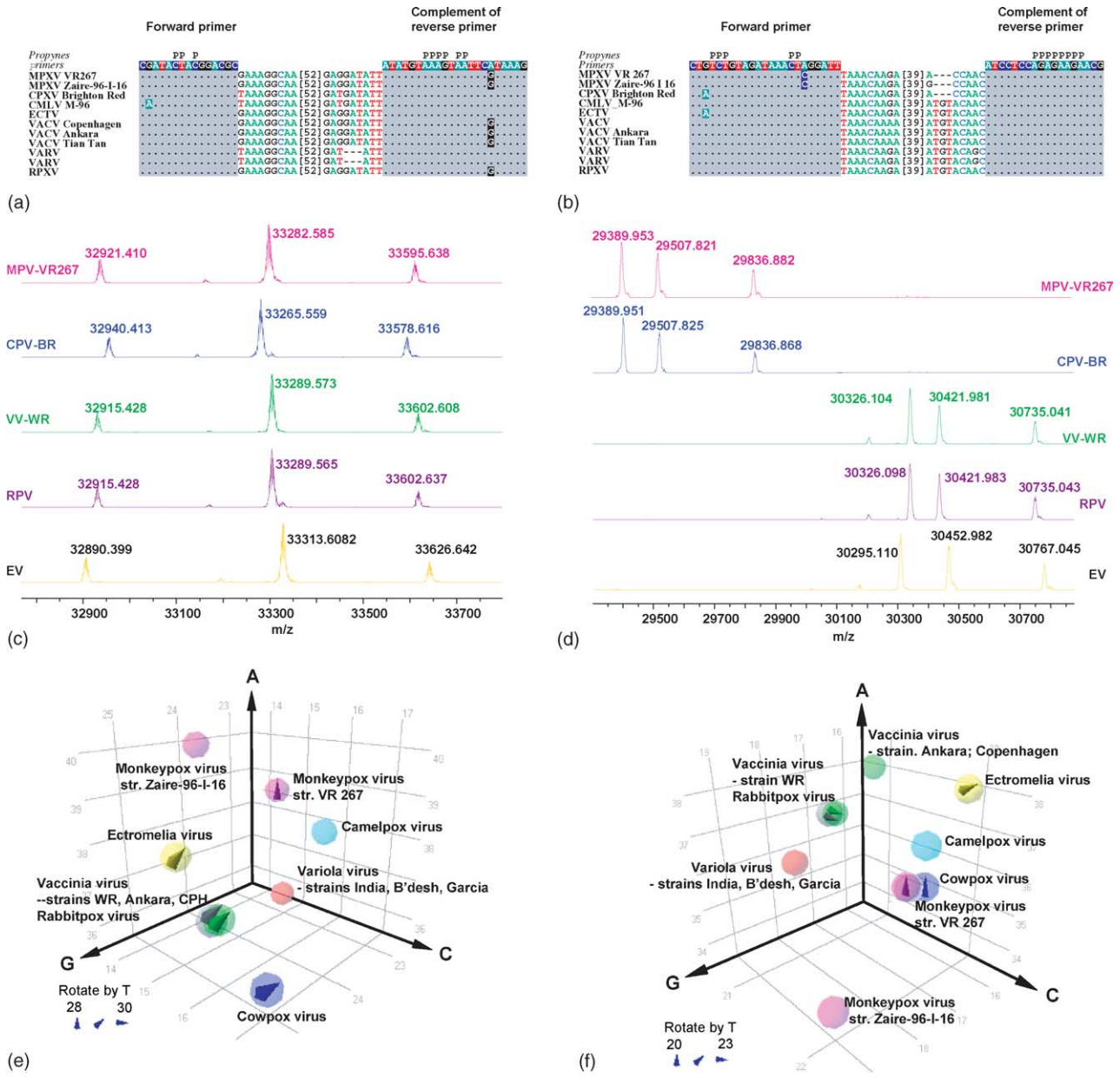

Fig. 13. Orthopoxvirus triangulation primers and TIGER resolution. (a, b) primer regions on two different Orthopoxvirus essential genes, DNA polymerase (E9L) and DNA helicase (A18R). All primers were propynylated at conserved positions (shown with a "P"). (c, d) TIGER MS spectra for the five lab test strains for each primer are shown. Peaks correspond to the sense and anti-sense amplicons generated by PCR. A third peak corresponding to non-templated end-adenylation ($M = 313.053$ Da) of one of the two strands was also often seen. (e, f) 3-D plots of base compositions. Axes represent [A], [C] and [G] counts ([T] counts shown by the rotation of the cone). Distribution of base compositions for all the species in GenBank that were expected to prime with the two primers are represented as a sphere. Each entry here represents a unique base composition in this pseudo four-dimensional space. Multiple isolates of the same species that show different base compositions are shown individually. Experimentally determined base compositions for the test strains are shown as a solid cone projected onto the same plot.

Red (BR) strains of cowpox. The material that was tested in the lab was clearly the BR strain as evidenced by the perfect match to the expected base counts for these in the database.

Table 3 shows the expected base counts of the various Orthopoxvirus species for all primer regions tested. For every species, the experimentally measured signals matched predicted base compositions. The color scheme used in this table groups identical base compositions within a primer region. As is typically the case with TIGER, while a single primer target region does not resolve all species unambiguously, species can be clearly identified and differentiated from one another using the triangulation strategy across multiple loci.

3.7. Example 3: RNA genome viruses—alphaviruses

Viral encephalitis is caused by a number of different viruses from different viral families. Important amongst these

Table 3
Database of Orthopoxvirus species base compositions for each selected TIGER primer region (shown in columns)

			RNA Helicase	DNA Helicase	RNA Polymerase	DNA Polymerase	Thioredoxin-like	DNA Helicase (2)
	TIGER Primer Region ---->		K8R_221_311_P	A18R_100_207_P	A24R_795_878_P	E9L_1119_1222_P	A2.5L_28_127_P	A18R_1348_1445_P
Test Strain	ORGANISM	STRAIN	BaseComp	BaseComp	BaseComp	BaseComp	BaseComp	BaseComp
	Camelpox virus	CMS	A38 G11 C23 T19	A32 G20 C23 T33	A29 G15 C14 T26	A38 G23 C16 T30	A30 G19 C18 T33	A37 G17 C22 T22
	Camelpox virus	M-96(2)	A38 G11 C23 T19	A32 G19 C23 T34	A29 G15 C14 T26	A38 G23 C16 T30	A30 G19 C18 T33	A37 G17 C22 T22
yes	Cowpox virus	Brighton Red(1)	A33 G14 C18 T26	A36 G18 C23 T31	A29 G15 C16 T24	A36 G25 C17 T29	A25 G24 C21 T30	A36 G17 C22 T20
	Cowpox virus	GRI-90(1)	A37 G11 C24 T19	A33 G19 C24 T32	A30 G15 C13 T26	A36 G25 C17 T29	A27 G23 C19 T31	A36 G18 C22 T22
yes	Ectromelia virus	Moscow(1)	A34 G13 C17 T27	A33 G19 C24 T32	A30 G15 C13 T26	A38 G25 C15 T29	A27 G22 C19 T32	A38 G16 C22 T22
yes	Monkeypox virus	VR-267	A34 G14 C18 T25	A33 G20 C22 T33	A29 G15 C15 T25	A39 G24 C16 T28	A28 G20 C21 T34	A36 G17 C22 T20
	Monkeypox virus	Zaire-96-I-16	A34 G14 C18 T25	A33 G20 C22 T33	A28 G16 C15 T25	A40 G24 C14 T29	A28 G20 C21 T34	A34 G19 C22 T20
	Vaccinia virus	Copenhagen(3)	A38 G10 C24 T19	A32 G21 C24 T31	A30 G15 C13 T26	A37 G25 C16 T29	A25 G23 C20 T31	A36 G16 C21 T23
	Vaccinia virus	Copenhagen(4)	A38 G10 C24 T19	A32 G21 C24 T31	A30 G15 C13 T26	A37 G25 C16 T29	A25 G23 C20 T31	A36 G16 C21 T23
	Vaccinia virus	Tian Tan(2)	A36 G12 C24 T19	A32 G21 C24 T31	A30 G15 C13 T26	A37 G25 C16 T29	A27 G22 C19 T31	A36 G16 C21 T23
yes	Vaccinia virus	Western Reserve(1)	A36 G12 C24 T19	A33 G20 C23 T32	A30 G15 C13 T26	A37 G25 C16 T29	A27 G23 C18 T32	A37 G17 C21 T23
	Vaccinia virus	Ankara(1)	A36 G12 C24 T19	A33 G20 C23 T32	A30 G15 C13 T26	A37 G25 C16 T29	A25 G24 C20 T31	A34 G18 C21 T23
yes	Vaccinia virus	Rabbitpox Utrecht(8)	A36 G12 C24 T19	A33 G20 C23 T32	A30 G15 C13 T26	A37 G25 C16 T29	A25 G24 C20 T31	A37 G17 C21 T23
	Variola major virus	Bangladesh-1975(2)	A36 G11 C24 T20	A33 G20 C20 T35	A28 G16 C14 T26	A36 G23 C15 T30	A28 G21 C16 T35	A36 G18 C21 T23
	Variola major virus	India-1967(1)	A36 G11 C24 T20	A33 G20 C20 T35	A28 G16 C14 T26	A36 G23 C15 T30	A28 G21 C16 T35	A36 G18 C21 T23
	Variola minor virus	Garcia-1966(1)	A36 G11 C24 T20	A34 G19 C21 T34	A28 G16 C14 T26	A36 G23 C15 T30	A28 G21 C16 T35	A36 G18 C21 T23

Identical base compositions within a column are grouped by color. All the test isolates (see column 1) showed perfect match to the expected base compositions where prior data was known.

are positive-strand RNA viruses belonging to the genus Alphavirus (Togaviridae family), including the Venezuelan equine encephalitis (VEE), the western equine encephalitis (WEE) and the eastern equine encephalitis (EEE) complex viruses. All of these viruses have the potential to cause severe disease of veterinary and human health consequences. VEE complex viruses represent 13 serologically distinct types belonging to six antigenic subtypes (I–VI), the majority are enzootic and are not (or are only rarely) transmitted to humans [24]. The epidemic or epizootic VEE viruses belong primarily to types IAB and IC and are believed to have emerged from closely related lineages. The EEE virus antigenic complex has four antigenic subtypes (I–IV), spanning the North and South American isolates [25]. Both VEE and EEE are on the CDC and USDA list of select agents. The WEE complex viruses include Old World (Sindbis virus) and New World viruses (Aura, Highlands J, WEE, etc.) [26].

These viruses are extremely diverse at the nucleotide and protein levels and pose a great challenge for most detection and diagnostic techniques. Several species-specific, genus-specific, and multiplex molecular detection methods (RT-PCR) have been described [27–29]. The TIGER broad-priming approach allows the rapid identification of all alphaviruses. We designed several primers that broadly target various members of this family of viruses and used propynylated primers to enhance primer hybridization. Primer pairs targeting conserved sites in the 5'-end of the RNA virus genome (nonstructural protein, nsP1 gene) were designed and tested against a number of different viral isolates from the three major complexes. In Fig. 14a, the primer sequences are shown overlaid on an alignment of viral sequences, with the dots indicating identity to the primer sequence. Propynylated bases are indicated. Theoretical distribution of expected products based on all known Alphavirus sequences in GenBank showed resolution of the various subtypes of VEE, as well as other major members of this genus using TIGER primers (Fig. 14b). None of these primers will produce spurious products from other unrelated sequences (results not shown).

We obtained several Alphavirus isolates of VEE and WEE from Dr. Scott Weaver at the UTHSC, Galveston and a full-length cDNA clone of EEE from Dr. Mike Parker, USAMRIID (Table 4). Five different primer pairs targeting two distinct regions on the Alphavirus genome were used to amplify regions from these isolates. Table 4 shows the TIGER-determined base compositions for these isolates. All the test isolates showed perfect match to the expected base compo-

Table 4
Alphavirus species base compositions for each selected TIGER primer region (shown in columns)

	Region 1			Region 2	
Test Isolates	PP302	PP303	PP304	PP305	PP306
VEE 243937	[23 23 21 19]	[24 22 21 19]	[21 23 23 19]	[25 31 24 24]	[25 31 24 24]
VEE 66637	[23 23 21 19]	[24 22 21 19]	[21 23 23 19]	[26 31 24 23]	[26 31 24 23]
VEE 3908	[23 23 21 19]	[24 22 21 19]	[21 23 23 19]	[25 31 23 25]	[25 31 23 25]
VEE68U201	[23 25 17 21]	[24 24 17 21]	[22 25 19 20]	[24 33 24 23]	[24 33 24 23]
EEE*	[25 24 17 20]	[26 23 17 20]	[23 25 19 19]	[26 31 20 27]	[26 31 21 26]
WEE OR71	[24 25 18 19]	[25 24 18 19]	[22 26 19 19]	[32 28 22 22]	[31 29 22 22]
WEE SD83	[24 25 18 19]	[25 24 18 19]	[22 26 19 19]	[32 28 22 22]	[31 29 22 22]
WEE Fleming	[24 24 18 20]	[25 23 18 20]	[22 25 19 20]	[25 30 22 27]	[26 32 25 21]
WEE ON41	[25 16 27 18]	[25 25 17 19]	[22 27 18 19]	[32 28 22 22]	[31 29 22 22]

EEE was obtained from Mike Parker, USAMRIID. All other isolates were obtained from Scott Weaver, UTHSC. Test isolates highlighted in yellow did not have prior sequence data and represent new measurements.

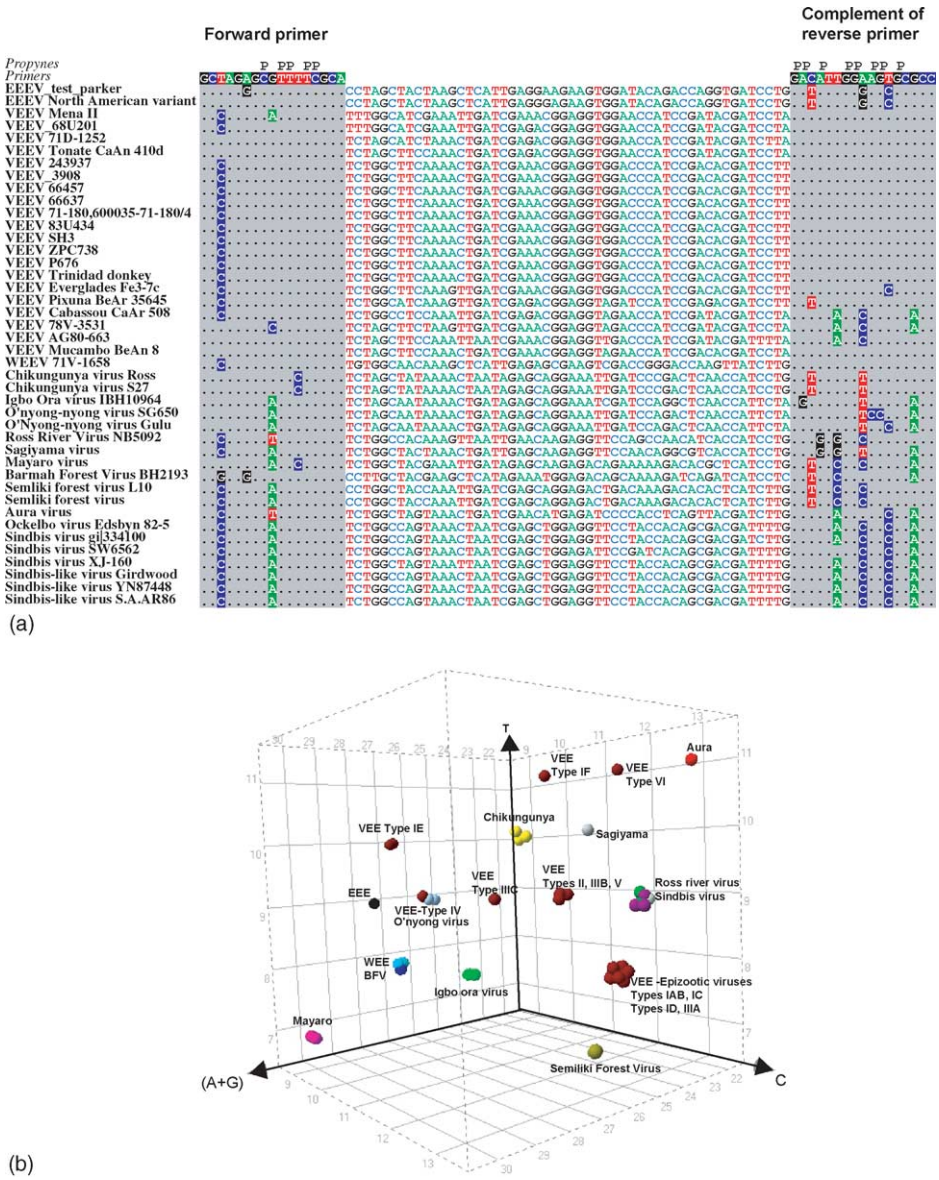

Fig. 14. (a) Alignment of *Alphavirus* sequences showing conservation of PCR primer regions within this viral family, flanking a region of species-specific variations. "Dots" in a column represent homology to the reference sequence above. Any variant is explicitly shown with the varying nucleotide. The consistent pattern of base changes predicted between species shown here is sufficient to resolve these organisms in the high-performance TIGER FTICR. (b) Epidemic, epizootic VEE viruses IAB-IC along with the closely related ID sequences (see text) are clearly distinguishable from enzootic types II–VI and IE–IF.

sitions where prior data was known. Sequences of the three WEE isolates and the EEE isolate shown highlighted in yellow have not been sequenced and represent new measurements. Identical base compositions for different isolates for a particular primer pair are shown grouped by the same color. The four VEE isolates tested include VEE-3908 and VEE-24937 (type IC) VEE-66637 (type ID) and VEE-68U201 (type IE). All of these can be resolved from each other using the primer pairs shown in Table 4. The two Group B WEEs tested here, WEE-OR71 and WEE-SD83 [26], were indistinguishable using the two regions shown here. Additional primer pairs will be required to distinguish these isolates. On the other hand, the two group A WEEs tested, the Fleming and ON41 isolates, were resolved from each other. The EEE test isolate was distinguishable from all other isolates.

4. Conclusions

We have demonstrated that the TIGER strategy can be applied to the detection and identification of a wide variety of bacteria, DNA-genome viruses, and RNA genome-viruses. Broad-range PCR reactions are capable of producing products from groups of organisms, rather than single species, and the information content of each PCR reaction is potentially very high. The mass spectrometer is capable of ana-

lyzing complex PCR products at a rate of approximately one minute per sample. Because the process is performed in an automated, microtiter plate format, it is possible to examine large numbers of samples, making it practical for large-scale analysis of clinical specimens or for environmental surveillance. This approach can be extended to other viral, bacterial, fungal, or protozoal pathogen groups and is a powerful new paradigm for timely identification of a broad range of organisms that cause disease in humans or animals, and for monitoring the progress of epidemics in biological weapons surveillance.

Acknowledgement

This work was funded in part by DARPA under contract MDA972-00-C-0053 and by the CDC under grant # 1R01 CI000099-01 issued to the Ibis Therapeutics Division of Isis Pharmaceuticals. Approved for Public Release, Distribution Unlimited.

References

[1] L.H. Taylor, S.M. Latham, M.E. Woolhouse, Philos. Trans. R. Soc. Lond. B. Biol. Sci. 356 (2001) 983.
[2] Y. Jiang, S.A. Hofstadler, Anal. Biochem. 316 (2003) 50.
[3] M. Greig, R.H. Griffey, Rapid Commun. Mass Spectrom. 9 (1995) 97.
[4] M.W. Senko, C.L. Hendrickson, M.R. Emmett, S.D.H. Shi, A.G. Marshall, J. Am. Soc. Mass Spectrom. 8 (1997) 970.
[5] R.D. Macfarlane, Accounts Chem. Res. 15 (1982) 268.
[6] R.K. Saiki, S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, N. Arnheim, Science 230 (1985) 1350.
[7] K.E. Yoder, O. Sethabutr, D.A. in: D.H. Persing (Ed.), PCR Protoc. Emerging Infect. Dis., ASM Press, Washington, D, 1996, p. 169.
[8] D.H. Gelfand, P.M. Holland, R.K. Saiki, R.M. Watson, Hoffmann-La Roche Inc., USA, US Patent 5,487,972 (1996), 44 pp.
[9] C.A. Heid, J. Stevens, K.J. Livak, P.M. Williams, Genome Res. 6 (1996) 986.
[10] A. Giulietti, L. Overbergh, D. Valckx, B. Decallonne, R. Bouillon, C. Mathieu, Methods, vol. 25, San Diego, CA, United States, 2001, p. 386.
[11] R.G. Kuimelis, K.J. Livak, B. Mullah, A. Andrus, Nucleic Acids Symp Ser. 37 (1997) 255.
[12] M.J. Doktycz, G.B. Hurst, S. Habibi-Goudarzi, S.A. McLuckey, K. Tang, C.H. Chen, M. Uziel, K.B. Jacobson, R.P. Woychik, M.V. Buchanan, Anal. Biochem. 230 (1995) 205.
[13] G.B. Hurst, M.J. Doktycz, A.A. Vass, M.V. Buchanan, Rapid Commun. Mass Spectrom. 10 (1996) 377.
[14] M.T. Krahmer, Y.A. Johnson, J.J. Walters, K.F. Fox, A. Fox, M. Nagpal, Anal. Chem. 71 (1999) 2893.
[15] Y. Naito, K. Ishikawa, Y. Koga, T. Tsuneyoshi, H. Terunuma, Rapid Commun. Mass Spectrom. 9 (1995) 1484.
[16] Y. Naito, K. Ishikawa, Y. Koga, T. Tsuneyoshi, H. Terunuma, R. Arakawa, J. Am. Soc. Mass Spectrom. 8 (1997) 737.
[17] K.H. Buetow, M. Edmonson, R. MacDonald, R. Clifford, P. Yip, J. Kelley, D.P. Little, R. Strausberg, H. Koester, C.R. Cantor, A. Braun, Proc. Natl. Acad. Sci. U.S.A. (2001).
[18] D.J. Aaserud, Z. Guan, D.P. Little, F.W. McLafferty, Int. J. Mass Spectrom. Ion Processes. 167/168 (1997) 705.
[19] D.C. Muddiman, G.A. Anderson, S.A. Hofstadler, R.D. Smith, Anal. Chem. 69 (1997) 1543.
[20] D.J. Aaserud, N.L. Kelleher, D.P. Little, F.W. McLafferty, J. Am. Soc. Mass Spectrom. 7 (1996) 1266.
[21] M.W. Senko, S.C. Beu, F.W. McLafferty, J. Am. Soc. Mass Spectrom. 6 (1995) 229.
[22] C. Upton, S. Slack, A.L. Hunter, A. Ehlers, R.L. Roper, J. Virol. 77 (2003) 7590.
[23] C. Gubser, S. Hue, P. Kellam, G.L. Smith, J. Gen. Virol. 85 (2004) 105.
[24] B. Linssen, R.M. Kinney, P. Aguilar, K.L. Russell, D.M. Watts, O.-R. Kaaden, M. Pfeffer, J. Clin. Microbiol. 38 (2000) 1527.
[25] A.C. Brault, A.M. Powers, C.L. Villarreal, R.N.L. Chavez, M.F. Cachon, L.F.L. Gutierrez, W. Kang, R.B. Tesh, R.E. Shope, S.C. Weaver, Am. J. Trop. Med. Hyg. 6 (1999) 579.
[26] S.C. Weaver, W. Kang, Y. Shirako, T. Rumenapf, E.G. Strauss, J.H. Strauss, J. Virol. 71 (1997) 613.
[27] M. Pfeffer, B. Linssen, M.D. Parker, R.M. Kinney, J. Vet. Med. B 49 (2002) 49.
[28] J.-H. Lee, K. Tennessen, B.G. Lilley, T.R. Unnasch, J. Am. Mosq. Control Assoc. 18 (2002) 26.
[29] M. Pfeffer, B. Proebster, R.M. Kinney, O.-R. Kaaden, Am. J. Trop. Med. Hyg. 57 (1997) 709.