

Chapter 7

Molecular Genotyping of Microbes by Multilocus PCR and Mass Spectrometry: A New Tool for Hospital Infection Control and Public Health Surveillance

David J. Ecker, Christian Massire, Lawrence B. Blyn, Steven A. Hofstadler, James C. Hannis, Mark W. Eshoo, Thomas A. Hall, and Rangarajan Sampath

Abstract

We describe a new technology for the molecular genotyping of microbes using a platform known commercially as the Ibis T5000. The technology couples multilocus polymerase chain reaction (PCR) to electrospray ionization/mass spectrometry (PCR/ESI-MS) and was developed to provide rapid, high-throughput, and precise digital analysis of either isolated colonies or original patient specimens on a platform suitable for use in hospital or reference diagnostic laboratories or public health settings. The PCR/ESI-MS method measures digital molecular signatures from microbes, enabling real-time epidemiological surveillance and outbreak investigation. This technology will facilitate understanding of the pathways by which infectious organisms spread and will enable appropriate interventions on a time frame not previously achievable.

Key words: Molecular genotyping, mass spectrometry, PCR/ESI-MS, base composition, Ibis T5000.

1. Introduction

Multilocus polymerase chain reaction (PCR) followed by electrospray ionization mass spectrometry (PCR/ESI-MS) is the analysis of PCR amplicons using ESI-MS. The technique was initially developed for the identification of microbes, including previously unknown or unculturable organisms, in original patient specimens or environmental surveillance samples in which multiple microbes may be present (1–3).

In brief, multiple pairs of primers are used to amplify carefully selected regions of pathogen genomes; the primer target sites are broadly conserved, but the amplified region carries information on the microbe's identity in its nucleotide base composition. Regions of this nature appear in the DNA that encodes ribosomal RNA and in housekeeping genes that encode essential proteins. Following PCR amplification, a fully automated ESI-MS analysis is performed. The mass spectrometer effectively weighs the PCR amplicons, or mixture of amplicons, with sufficient mass accuracy that the composition of A, G, C, and T can be deduced for each amplicon present. The base compositions are compared to a database of calculated base compositions derived from the sequences of known organisms to determine the identities of the microorganisms present. In the event that there is no match of the measured base composition with a sequence in the database, the nearest neighbor organism is identified. Thus, analysis by the PCR/ESI-MS method provides information that enables identification of a broad range of microbes in a sample without having to anticipate what microbes might be present. The identities of microbes in a mixed population are determined because the primers amplify the nucleic acids from all organisms in the sample simultaneously, and the mass spectrometer analyzes and reports on multiple peaks in the same spectrum.

2. High-Resolution Molecular Genotyping by Multilocus PCR and Mass Spectrometry

The Ibis T5000 technology was initially developed for broad bacterial and viral detection and identification; however, PCR/ESI-MS is also a very powerful tool for high-resolution molecular genotyping of microbes. Applications of the technology can be thought of in an hourglass model as illustrated in **Fig. 1**. The upper portion of the hourglass depicts identification of microbes, generally bacteria and viruses, present in an unknown sample at the species level as described. The utility of PCR/ESI-MS has been demonstrated for broad bacterial surveillance (2) and for identification of virus families, including coronaviruses (4), influenza viruses (5), adenoviruses (6), alphaviruses (7), and orthopoxviruses (3). The bottom portion of the hourglass in **Fig. 1** refers to assays developed on the PCR/ESI-MS platform that are specific for a particular species; these assays reveal molecular details such as the presence of virulence factors, antibiotic or antiviral drug resistance, or high-resolution molecular signatures that distinguish closely related subspecies. These high-resolution molecular analyses require separate assays that investigate important questions unique to a particular microbe. For example, for *Staphylococcus aureus*, it is important to determine the presence

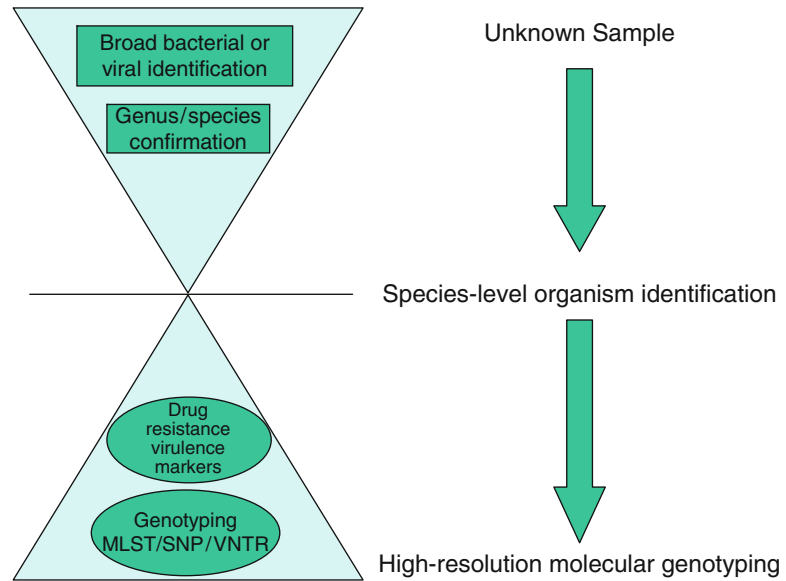


Fig. 1. Hourglass model for applications of PCR/ESI-MS. In broad surveillance mode (top of hourglass), the technology can be used to answer the question, Which organisms are in my sample? The pinch point is identification of the species, which is where most molecular methods are focused. The lower portion of the hourglass is the “drill-down” mode of PCR/ESI-MS. In this mode, species-specific primers yield high-resolution details that distinguish strain types and identify virulence and drug resistance markers. (See Color Plates)

or absence of certain virulence factors, mobile genetic elements, or mutations in housekeeping genes that mediate drug resistance. For understanding the genetic lineage of microbes, the PCR/ESI-MS method follows the general principles of multilocus sequence typing (MLST).

MLST is a high-resolution molecular tool for discriminating closely related bacterial subspecies (8) (see **Chapter 11** in this book). In this method, the data are digital and portable, facilitating comparison among laboratories worldwide. However, conventional MLST requires isolation of pure colonies of the target microbe followed by multiple PCR reactions and sequencing of each amplicon. While sequencing technology has become much more facile in recent years, it is still not practical to use conventional MLST in a clinical laboratory setting. Clinical and public health laboratories require simple, automated analytical methods that match their throughput needs and cost limitations. In contrast to conventional MLST, multilocus PCR/ESI-MS provides an automated, high-throughput alternative that approaches the resolution of sequence-based, conventional MLST and can be implemented in a clinical laboratory at very low per-sample costs.

The multilocus PCR/ESI-MS strategy is graphically depicted in **Fig. 2**. The same set of housekeeping genes used for conventional

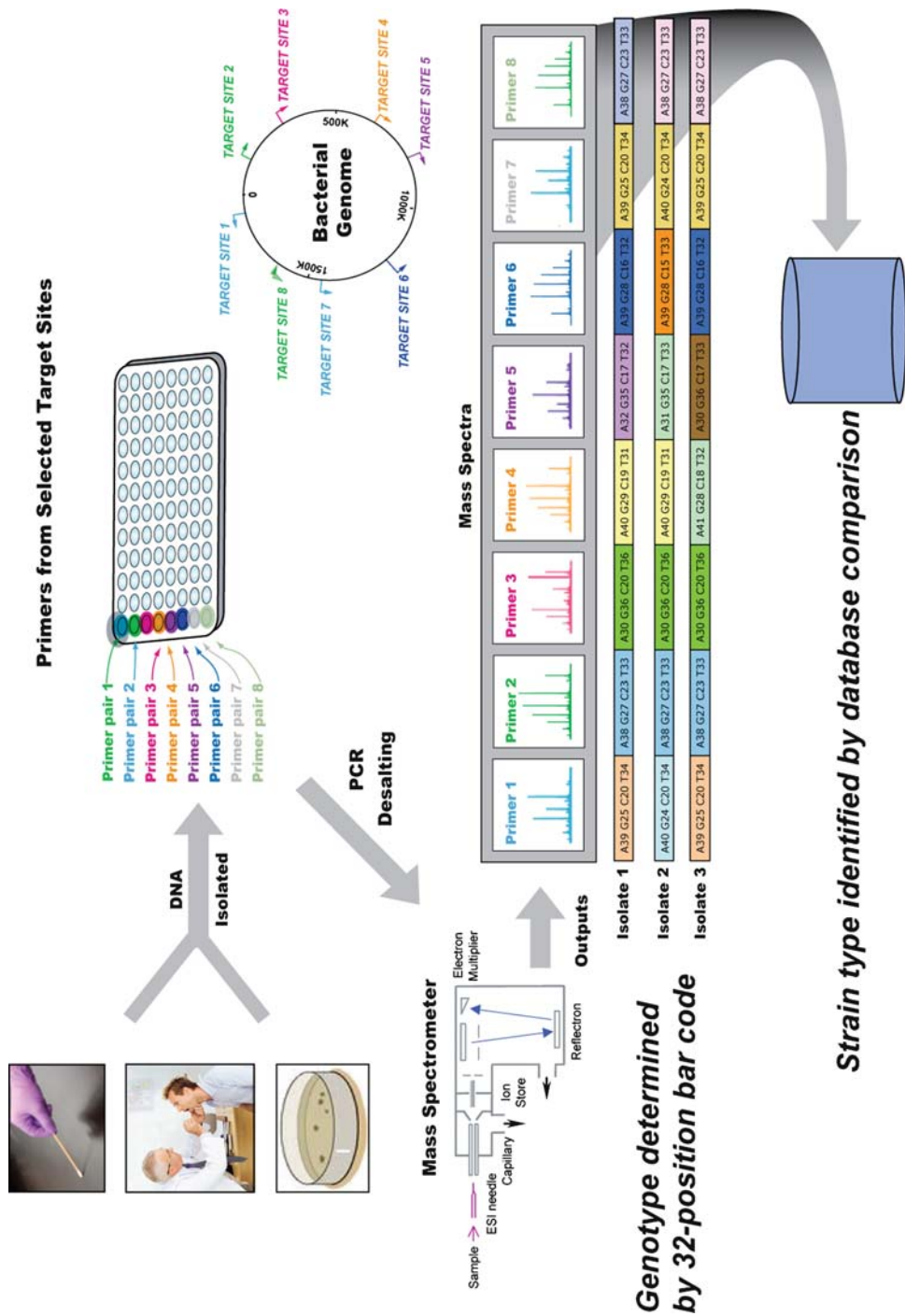


Fig. 2. Flow scheme for high-resolution genotyping of microbes by PCR/ESI-MS. A set of species-specific primer pairs are designed to distinguish strain types. Each primer pair generates a four-position bar code comprised of the AGCT count for each amplicon. Taken together, a set of eight primer pairs produces a 32-position bar code, which provides strain differentiation. The method works on isolated colonies or original patient or environmental samples. Mixed populations of microbes can be analyzed because multiple peaks in the mass spectrum can simultaneously be analyzed. (See *Color Plates*)

MLST are analyzed to identify the regions that contain the highest information content in their base compositions, and sets of primer pairs are designed to these regions. Typically, 100–150 nucleotide regions are selected for amplification. The information values of the amplicons are evaluated until an optimal set of primer pairs is identified. Each primer pair is assigned to a position in a 96-well plate such that a sample is amplified by eight pairs of primers and analyzed by MS. Each of the primer pairs produces an amplicon that results in a spectral signal and base composition, or four-position A, G, C, T signature (but since here amplicons are generally of constant length, each base composition signature actually contains only three independent variables). Base compositions from each of the eight primer sets result in a 24-dimensional digital signature that can be compared to calculated base composition signatures generated from an MLST database.

The ability to distinguish MLST alleles by PCR followed by MS is, at first glance, counterintuitively high. Molecular biologists generally think in terms of the sequence of the nucleotides as the signature of a microbe. But, while the potential number of distinct sequences within any given MLST locus is astronomical (4^x , where x is the number of nucleotides showing mutations), the number of actual, biologically relevant sequences is typically much more manageable: First, only a fraction (10–20%) of the positions within MLST loci show variation. Second, most of these sites do not display the full range of possible mutations, but merely transitions. Third, only a fraction of these sites is simultaneously mutated. Thus, only 50 to 100 alleles, differentiated by specific sets of mutations, are typically reported in MLST databases for a single locus.

This level of resolution can be approached by base composition analysis. Any single mutation that separates one allele from another can be identified by MS analysis since even a single-nucleotide substitution results in spectral signals that can be identified as distinct masses and compositions. There are 12 possible types of single mutations ($A \rightarrow G$, $A \rightarrow C$, $A \rightarrow T$, $G \rightarrow A$, $G \rightarrow C$, $G \rightarrow T$, $C \rightarrow A$, $C \rightarrow G$, $C \rightarrow T$, $T \rightarrow A$, $T \rightarrow G$, and $T \rightarrow C$), and all result in masses that are unique (*see Fig. 3*). As additional mutations occur, the resulting space of possible base compositions grows accordingly, following a third-degree polynomial expression (**Fig. 3**). Of course, not all possible base compositions are actually generated with a given set of alleles, and each allele does not necessarily generate its own distinct base composition. The most common way for two alleles to share the same base composition is to differ from each other by one of the six self-cancelling pairs of single-nucleotide polymorphisms (SNPs) (e.g., $A \rightarrow G$ and $G \rightarrow A$ or $A \rightarrow C$ and $C \rightarrow A$). If three SNPs are involved, retrieving the same base composition involves one of the eight possible “triangular” mutation patterns (e.g., $A \rightarrow G$, $G \rightarrow C$,

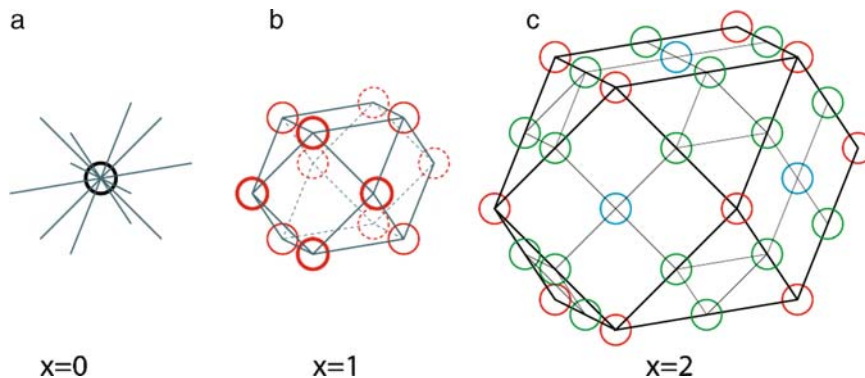


Fig. 3. Representation of the base composition space that is covered after x mutations. **(a)** The original base composition space (black sphere) can be affected by 12 distinct mutations (gray lines). **(b)** After one mutation, the 12 resulting base compositions define a hollow, cuboctahedron-shaped shell. Any of these 12 base composition spaces can be similarly affected by 12 additional mutations; in each case, only 1 mutation will revert back to the original base composition ($x = 0$), whereas 4 mutations would yield an adjacent base composition (within the same shell), and 7 mutations would yield base compositions located in the next $x = 2$ shell. **(c)** Base composition space for $x = 2$ mutations. For clarity, only the front-facing base compositions of the outer shell are represented. Equivalent positions are similarly colored. With each subsequent x_{th} mutation, an additional shell of $N(x) = 10x^2 + 2$ new, distinct base composition space is added. The total number of base compositions that can be reached after x mutations follows a third-degree polynomial progression: $N_{BC}(x) = (2x + 1)(5x(x + 1) + 3)/3$. (See Color Plates)

C \rightarrow A), whereas with four SNPs “quadrangular” mutation patterns are possible (e.g., A \rightarrow G, G \rightarrow C, C \rightarrow T, T \rightarrow A). As is apparent in **Table 1**, the occurrence of such combinations decreases as the number of SNPs increases, meaning that base compositions naturally tend to be more diverse as the number of SNPs increases in the allele set. In practice, a typical PCR/ESI-MS amplicon of a MLST gene carries from two to six mutations, which is enough to observe a number of distinct base compositions in the same order of magnitude (about 70% on average) as the number of alleles that are distinguished by sequence within the same locus.

The practical utility of MS analysis of PCR amplicons to distinguish MLST alleles was determined by examination of multiple sequence alignments from housekeeping genes from *Acinetobacter baumannii*, *S. aureus*, *Pseudomonas aeruginosa*, and *Streptococcus pyogenes* (see **Fig. 4**). In all cases, a single primer pair targeted to a single allele excluded more than 60% of sequence types on average, and amplification of four loci resulted in elimination of more than 95% of all sequence types on average. Thus, by using six to eight primer pairs it is possible to resolve different isolates of these microbes at a level that is more than sufficient for establishing clonality in an outbreak investigation.

An important advantage of multilocus PCR/ESI-MS is that nucleic acid does not need to be isolated from pure colonies of the target microbe. Patient specimens have been successfully

Table 1
Fraction of Single-Nucleotide Polymorphisms (SNP) Combinations Silent in Mass Spectroscopic (MS) Analysis of Polymerase Chain Reaction Amplicons

Number of SNPs	Number of unique SNP combinations	Number of SNP combinations yielding no net base composition change	Fraction (%) of SNP combinations silent in MS analysis ^a
1	12	–	0.00
2	78	6	7.69
3	364	8	2.20
4	1,365	27	1.98
5	4,368	48	1.10
6	12,376	112	0.90
7	31,824	192	0.60
8	75,582	378	0.50
9	167,960	624	0.37
10	352,716	1,092	0.31

^aEvolution of the fraction of SNP combinations silent by MS (fourth column) is given as a function of the number of SNPs (first column). Since the number N of unique SNP combinations (second column) is insensitive to the SNP order, it follows that the relation $N = (x + n - 1)!/x!(n - 1)!$, where n is the number of possible single changes (12 here), and x is the total number of single changes made. The number of SNP combinations yielding no net change (third column) was determined empirically.

analyzed using this technology without culture (2). As eliminating the culture step can save 1 or 2 d, multilocus PCR/ESI-MS can be used to track an epidemic on a time frame not previously achievable. Samples that contain more than one strain type in a mixture can also be analyzed because multiple amplicons are individually identified in the mass spectrum. The peak heights for each of the amplicons in the mixture can be used to determine the relative ratios of microbes in the sample, provided that the low abundance microbe represents at least 2–5% of the microbial population. The fact that some clinical samples have mixed populations of strain types is often missed when a culture step is used, as bias can be introduced by culture conditions, and multiple colonies from the same sample are not always analyzed.

For bacterial pathogens that have emerged in relatively recent history, the numbers of mutations found in housekeeping genes are limited, and genetic markers that evolve at faster evolutionary clock speeds are necessary to establish clonality. For these

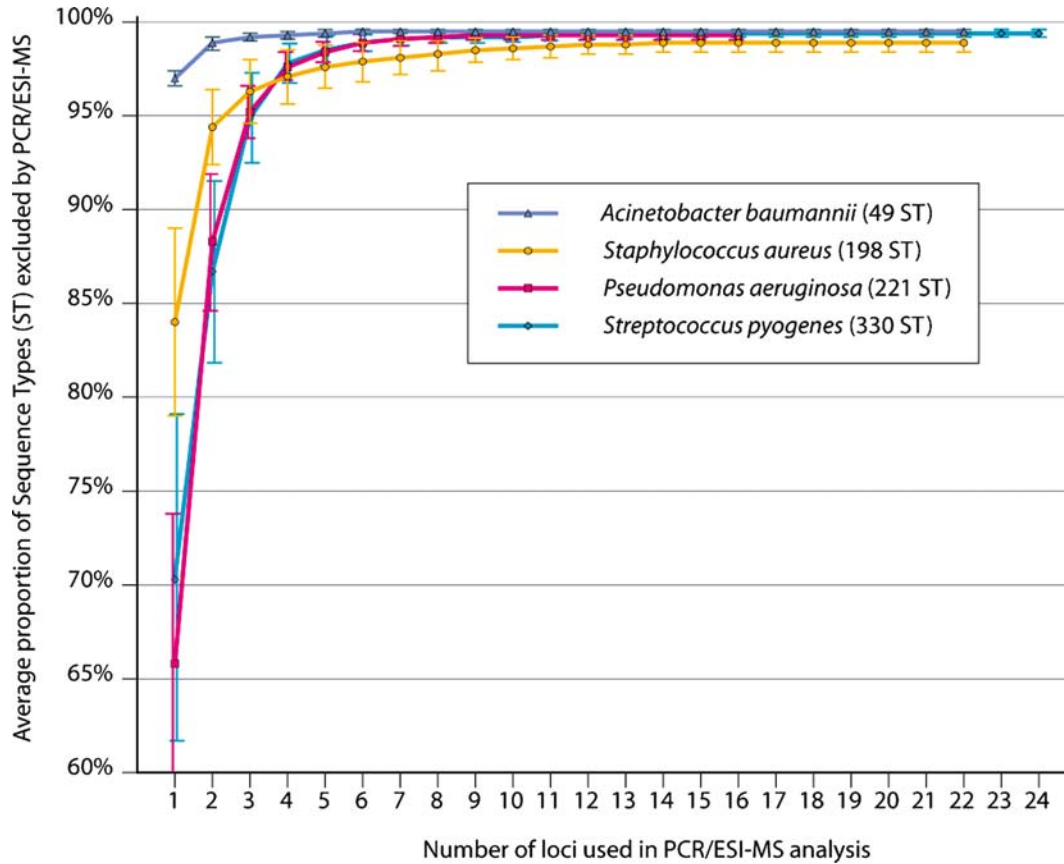


Fig. 4. Comparative resolution of PCR/ESI-MS typing schemes for four organisms. For each of the four organisms, a set of unique sequence types (STs) was first assembled using six or seven genes commonly used for MLST. These reference alignments were then used for the design of 16 to 24 primer pairs for PCR/ESI-MS analysis. The resolution provided by PCR/ESI-MS analysis was evaluated as follows: Starting with the primer pair providing the best sequence resolution, amplicon base compositions were determined for each of the sequence types. Comparison of these base composition signatures defined the number of sequence types that were incompatible with any particular type at this particular locus. The average proportion of sequence types excluded by PCR/ESI-MS with their corresponding standard deviations (vertical lines) was plotted versus the number of loci used in the analysis. This process was repeated, using base composition signatures extended by one additional locus at a time, to yield the full curves shown. (See Color Plates)

organisms, short repeated elements known as variable number of tandem repeats (VNTR) have proven to be useful markers (9). These elements vary in the number of repeats of short strings of nucleotides. Examples of organisms for which VNTR elements have been used to establish clonality are *Bacillus anthracis* (10), *Francisella tularensis* (9), and *Mycobacterium tuberculosis* (11). VNTR analysis can be conducted using PCR/ESI-MS simply by designing primers that bracket the VNTR. The base composition of the amplicon is used to precisely calculate the number of repeats as well as any single-nucleotide variations that may appear within the repeat, providing greater resolving power than the repeat count that is obtained from gel analysis. VNTR, SNP, and

MLST analyses can be combined into a single assay with PCR/ESI-MS, if simultaneous analysis of genetic biomarkers with a range of clock speeds is desired, simply by bracketing the appropriate target region on the microbial genome with PCR primers and assembling the primer set in 96-well plate configuration as shown in **Fig. 2**.

3. Examples of Applications of Multilocus PCR and Mass Spectrometry

3.1. Streptococcus pyogenes Epidemic Analysis

For high-resolution strain genotyping of *S. pyogenes*, a strategy was designed to generate strain-specific signatures like those provided by MLST (2). Primer pairs were designed to the *S. pyogenes* MLST gene targets that correlate with the *emm* classification. To identify target regions that provided the highest resolution of species and least ambiguous *emm* classification by base composition analysis, we constructed an alignment of concatenated alleles of the seven MLST housekeeping genes from each of 212 previously *emm*-typed strains (12) and determined the number and location of the primer pairs that would maximize strain discrimination. An initial set of 24 primer pairs was selected that would amplify regions covering over 97% of the known nucleotide variations in the MLST sequencing targets. We then determined how much strain discrimination could be achieved from a smaller set of primers. Calculations showed that six pairs of primers allowed discrimination at the individual *emm*-type level of about 75% of all the *emm* types listed by Enright et al. (12), while the remaining 25% clustered into groups of two or more *emm* types. This degree of resolution is sufficient for applications such as tracking the clonal expansion of a particular strain type during a specific epidemic.

We used this method to genotype *S. pyogenes* in patient samples taken at a military training camp during one of the most severe outbreaks of pneumonia associated with group A *Streptococcus* (GAS) in the United States since 1968 (13). Throat swabs were taken from both healthy and hospitalized recruits and plated for selection of putative GAS colonies. A second set of 15 original patient specimens was taken during the height of this disease outbreak. The third set consisted of historical samples from disease outbreaks at this and other military training facilities during previous years. The fourth set of samples was collected from five geographically separated military facilities in the continental United States in the winter immediately following the severe outbreak.

Colonies isolated from GAS-selective media from all four collection periods were analyzed with the six GAS genotyping primers. The results of the base composition analysis with genotyping primer pairs for samples from all four collection periods

were compared to results from 5'-*emm* gene sequencing and the MLST gene sequencing methods in **Table 2**. When only these six primer pairs were used, some of the samples could not be resolved to a unique *emm* type. However, base composition analysis showed identification consistent with (either uniquely or as a member of a small set) 5'-*emm* gene sequencing or the MLST sequencing method. These data showed that the GAS genotypes found during the epidemic were remarkably homogeneous (*see Fig. 5*), as would be expected for a clonal expansion during an outbreak in which the same genotype was being passed from person to person. In contrast, surveillance samples taken at diverse military bases showed a heterogeneous pattern reflecting a normal disease season in the absence of a major outbreak. This study demonstrated the power of PCR/ESI-MS in a real epidemic setting.

3.2. Acinetobacter baumannii Epidemic Analysis

Acinetobacter baumannii is often associated with hospital-acquired infections, and *Acinetobacter* also has a history of association with war-wound infections. During the Vietnam War, *A. baumannii* was the most common gram-negative bacteria recovered from traumatic injuries to extremities (14). This is because *Acinetobacter* naturally occurs in the soil. During blast injuries, wounds frequently become inoculated with soil organisms, leading to infections that later occur in the hospital. Over a 2-yr period from 2002 to 2004, military health officials identified 102 patients with blood cultures that grew *A. baumannii* from Landstuhl Regional Medical Center in Germany and from Walter Reed Army Medical Center (WRAMC) in the United States. In both facilities, the number of patients with *A. baumannii* bloodstream infections in 2003 and 2004 significantly exceeded those reported in previous years, suggesting nosocomial transmission.

Understanding the fundamental mechanisms underlying *Acinetobacter* infections, including the original sources of the infecting organisms, their clonality, and geographical spread, is important for the development of appropriate infection control measures. Genotyping allows investigation of clonal spread and can be used to identify the source of the original infection. We developed a high-throughput genotyping method for *Acinetobacter* using PCR/ESI-MS (15). At the time the method was developed, there was no MLST database for *Acinetobacter*, so we used *Moraxella catarrhalis* (the most closely related organism that had an MLST database) as a model to select the housekeeping genes for sequencing of *A. baumannii* isolates and to identify regions diverse enough to distinguish between strains by PCR/ESI-MS. We sequenced regions of six housekeeping genes (*trpE*, *adk*, *efp*, *mutY*, *fumC*, *ppa*) from 267 *Acinetobacter* isolates and designed eight PCR primer target sites covering about 1,700 nucleotides overall.

Table 2
Base composition signatures for *Streptococcus pyogenes* and correlations with *emm* types

No. of instances	Location	Base compositions of target gene regions										Strain (<i>emm</i> -type)	
		<i>murl</i>	<i>mutS</i>	<i>xpt</i>	<i>yqjL</i>	<i>gki</i>	<i>gtr</i>	Mass spec-trometry	Gene sequencing				
75	San Diego MCRD	A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A32 G35 C17 T32	A39 G28 C16 T32	3	3				
2		A40 G24 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A31 G35 C17 T33	A39 G28 C15 T33	6	6				
1		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A41 G28 C18 T32	A30 G36 C17 T33	A39 G28 C16 T32	28	28				
6	Archives	A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A32 G35 C17 T32	A39 G28 C16 T32	3	3				
3		A40 G24 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A30 G36 C20 T30	A39 G28 C15 T33	5, 58	5				
6		A40 G24 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A31 G35 C17 T33	A39 G28 C15 T33	6	6				
1	San Diego MCRD	A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A30 G36 C20 T30	A39 G28 C16 T32	11	11				
3		A40 G24 C20 T34	A38 G26 C24 T33	A30 G36 C19 T37	A40 G29 C19 T31	A31 G35 C17 T33	A39 G28 C15 T33	12	12				
1		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A31 G35 C17 T33	A38 G29 C15 T33	22	22				
3	San Diego MCRD	A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A30 G36 C17 T33	A38 G29 C15 T33	25, 75	75				
4		A40 G24 C20 T34	A38 G26 C24 T33	A30 G36 C20 T36	A41 G28 C19 T31	A30 G36 C18 T32	A39 G28 C15 T33	44/61, 82, 9	44/61				
2		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C19 T37	A40 G29 C19 T31	A32 G35 C17 T32	A39 G28 C16 T32	53, 91	91				
1	Wood	A39 G25 C20 T34	A38 G27 C24 T32	A30 G36 C20 T36	A40 G29 C19 T31	A30 G36 C17 T33	A39 G28 C15 T33	2	2				
2		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A32 G35 C17 T32	A39 G28 C16 T32	3	3				
1		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C19 T37	A41 G28 C19 T31	A31 G35 C17 T33	A39 G28 C15 T33	4	4				
1	Wood	A40 G24 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A31 G35 C17 T33	A39 G28 C15 T33	6	6				
11		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A30 G36 C17 T33	A39 G28 C15 T33	25 or 75	75				
1		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C19 T37	A40 G29 C19 T31	A30 G36 C17 T33	A39 G28 C15 T33	25, 75, 33, 75	75				
1	Leonard	A40 G24 C20 T34	A38 G26 C24 T33	A30 G36 C20 T36	A41 G28 C19 T31	A30 G36 C18 T32	A39 G28 C15 T33	34, 4, 52, 84	84				
2		A40 G24 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A30 G36 C17 T33	A39 G28 C15 T33	44/61 or 82	44/61				
2		A40 G24 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A30 G36 C20 T30	A39 G28 C15 T33	5 or 58	5				
3	Sierra	A40 G24 C20 T34	A38 G27 C23 T33	A30 G36 C19 T37	A40 G29 C19 T31	A30 G36 C18 T32	A39 G28 C15 T33	1	1				
2		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A32 G35 C17 T33	A39 G28 C16 T32	3	3				
1		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C19 T37	A41 G28 C19 T31	A31 G35 C17 T33	A39 G28 C15 T33	4	4				
1	Sierra	A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A41 G28 C18 T32	A30 G36 C17 T33	A39 G28 C16 T32	28	28				

(continued)

Table 2
(continued)

No. of instances	Location	Base compositions of target gene regions										Strain (emm-type)		
		<i>murl</i>	<i>mutS</i>	<i>xpt</i>	<i>yqjL</i>	<i>gki</i>	<i>gtr</i>	Mass spec- trometry	Gene sequencing					
1		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A32 G35 C17 T32	A39 G28 C16 T32	3						3
1		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C19 T37	A41 G28 C19 T31	A31 G35 C17 T33	A39 G28 C15 T33	4						4
3		A40 G24 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A31 G35 C17 T33	A39 G28 C15 T33	6						6
1		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A30 G36 C20 T30	A39 G28 C16 T32	11						11
1		A40 G24 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A41 G28 C19 T31	A30 G36 C19 T31	A39 G28 C15 T33	13						94
1		A40 G24 C20 T34	A38 G26 C24 T33	A30 G36 C20 T36	A41 G28 C19 T31	A30 G36 C18 T32	A39 G28 C15 T33	44/61 or 82						82
1		A40 G24 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A40 G29 C19 T31	A30 G36 C20 T30	A39 G28 C15 T33	5 or 58						58
1		A39 G25 C20 T34	A38 G27 C23 T33	A30 G36 C20 T36	A41 G28 C19 T31	A30 G36 C18 T32	A39 G28 C15 T33	78 or 89						89

¹MCRD: Marine Corps Recruit Depot.

²NHRC: Naval Health Research Center.

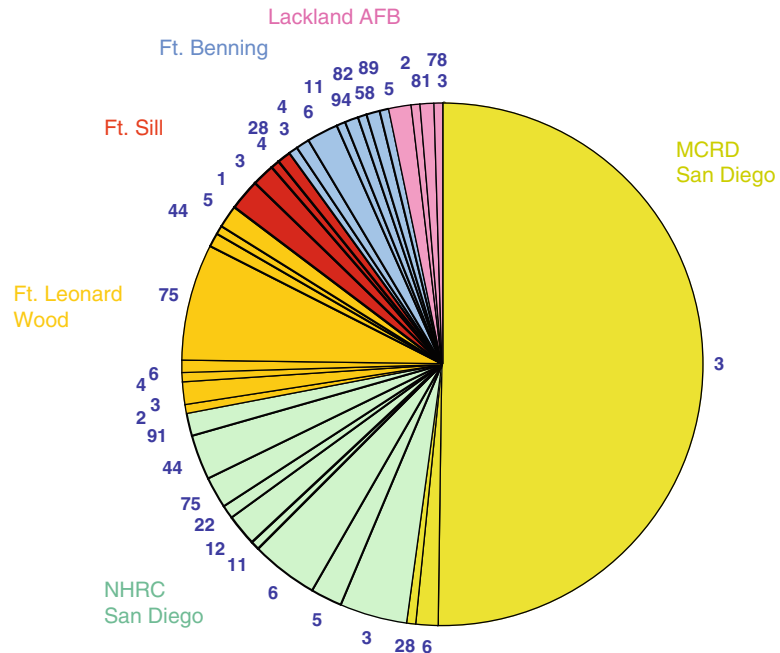


Fig. 5. Pie chart illustrating data given in Table 2. The area of each slice of pie is proportional to the number of instances of each *Streptococcus pyogenes emm* type. The colors indicate various military locations. MCRD: Marine Corps Recruit Depot; NHRC: Naval Health Research Center; AFB: Air Force Base. (See Color Plates)

Using this set of primers, isolates were analyzed from infected and colonized soldiers and civilians involved in an outbreak in the military health care system associated with the conflict in Iraq, from previously characterized outbreaks in European hospitals, and from culture collections. The goal of this study was to identify the reason for the increased nosocomial *Acinetobacter* infections observed during this period. Twenty-seven isolates from the outbreak in the military personnel were found to have genotypes representing different *Acinetobacter* species, including 8 representatives of *Acinetobacter* sp. 13TU and 13 representatives of *Acinetobacter* sp. 3. However, most of the isolates from the Iraqi conflict were *A. baumannii* (189 of 216 isolates). Among these, 111 isolates had genotypes identical or very similar to those associated with well-characterized *A. baumannii* isolates from European hospitals (Table 3). This observation suggested a second mode for the origin of *A. baumannii* infections: contamination with European strains that had developed multidrug resistance and properties that favored hospital transmission. Remarkably, isolates from WRAMC showed genotypes from all three major clones I, II, and III obtained from the European hospital collection (16,17), suggesting that the U.S. service personnel were exposed to a diverse set of European strain types.

Table 3
Genotypes, Number of Isolates and Correlation
with European Multidrug-Resistant *Acinetobacter* Clones

Organism	Genotype	Number of isolates	Clone type
<i>A. baumannii</i>	1, 10, 11, 47	65	II
	15, 16, 45, 46	24	I
	14	22	III
	(30 types)	78	
<i>Acinetobacter</i> sp. 13TU	(4 types)	8	
<i>Acinetobacter</i> sp. 3	(9 types)	13	
Other species	(6 types)	6	

A follow-up study was conducted by PCR/ESI-MS of *A. baumannii* isolates collected from wounded soldiers returning from the Iraqi conflict during 2006–2007 (18). The distribution of genotypes obtained during this period was remarkably similar to those observed in samples collected during 2003–2004, suggesting a stable reservoir of strain types that continued to infect U.S. service personnel wounded in the war. This composition of genotypes was significantly different from the nosocomial strains identified at nonmilitary U.S. hospitals, dispelling the hypothesis that repatriated soldiers infected with *Acinetobacter* were having an impact on U.S. nonmilitary hospital infections.

3.3. Virus Identification and Genotyping

The PCR/ESI-MS technology is also useful for identifying viruses and for tracking the spread of viral infections through a population. Despite higher mutation rates and greater sequence variability than bacteria, conserved primer target sites can be identified that enable priming of entire genera or even complete viral families. RNA-dependent RNA polymerase is a housekeeping gene common to all RNA viruses that provides several target site opportunities for developing primers that amplify multiple species within a virus family. This strategy is powerful because a single PCR reaction analyzed by MS can be used to detect and identify tens to hundreds of related viral species. The inherently high mutation rate of viruses results in base composition differences that provide a high-resolution molecular signature of viral subtypes. Generally, at least two sets of primer pairs are targeted to different regions of the viral genome for each virus group, and potential misclassification is avoided because two regions taken

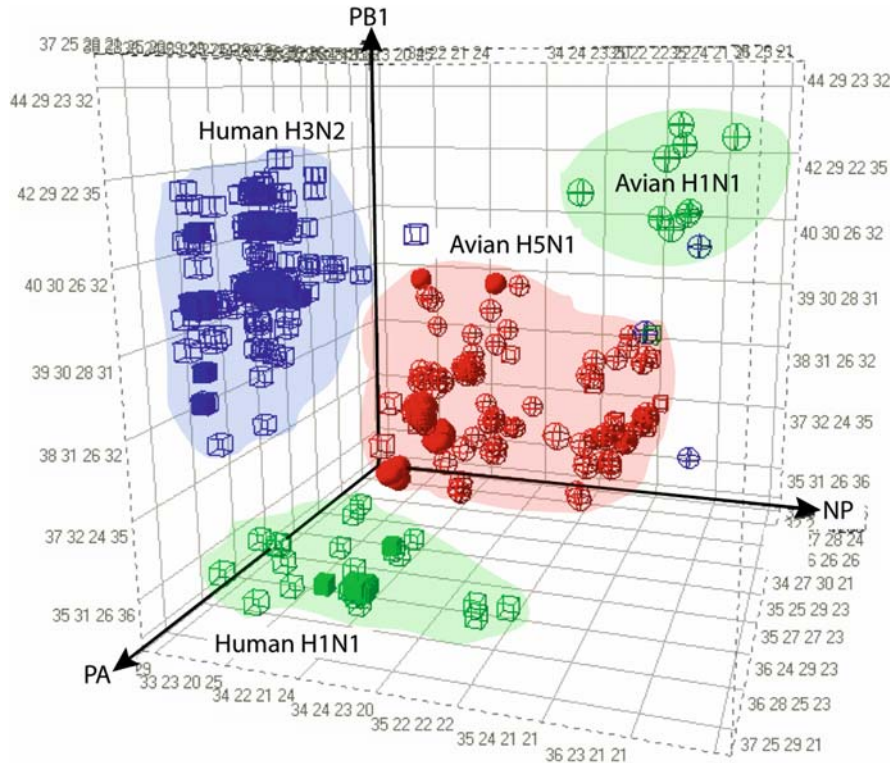


Fig. 6. Distribution of base compositions for influenza A viruses using three primer pairs. Hollow symbols represent calculated base compositions derived from sequences in GenBank, and solid symbols represent actual samples analyzed by PCR/ESI-MS. Red symbols, H5N1; green symbols, H1N1; blue symbols, H3N2. Cubes indicate human samples, and spheres indicate avian samples. (See Color Plates)

together provide unambiguous speciation and subtype determination. For example, we used PCR/ESI-MS to identify and sub-speciate over 50 types of adenoviruses (6). This strategy has also been used effectively for detection and strain typing of influenza viruses (5), alphaviruses (7), coronaviruses (4), and orthopoxviruses (3).

Base composition signatures provide a multidimensional fingerprint of the genomes of various viruses and can be used to determine clusters of related species/subtypes. One such representation (see Fig. 6) shows base composition data derived from the primer pairs targeted to PA, PB1, and NP gene segments of influenza A viruses. Human H3N2 and H1N1 viruses clustered independently from each other and from the avian/human H5N1 and H1N1 viruses. Thus, although mutations occur rapidly in viruses, base composition of certain regions can be used to cluster viruses into groups that are clearly distinguishable.

4. Conclusion

The Ibis T5000 PCR/ESI-MS technology couples PCR to ESI-MS and provides rapid, high-throughput, precise digital analysis of the microbes present in either isolated colonies or original patient specimens. The platform is suitable for use in hospital or reference diagnostic laboratories and other public health settings due to ease of use, high throughput, and affordability. The PCR/ESI-MS method measures digital molecular signatures from microbes, enabling real-time epidemiological surveillance and outbreak investigation. The method facilitates understanding of the pathways by which infectious organisms spread and enables appropriate interventions on a time frame not previously achievable.

References

- Ecker, D. J., Drader, J., Gutierrez, J., Gutierrez, A., Hannis, J., Schink, A., et al. (2006). The Ibis T5000 universal biosensor—an automated platform for pathogen identification and strain typing. *J. Assoc. Lab. Automat.* **11**, 341–351.
- Ecker, D. J., Sampath, R., Blyn, L. B., Eshoo, M. W., Ivy, C., Ecker, J. A., et al. (2005). Rapid identification and strain-typing of respiratory pathogens for epidemic surveillance. *Proc. Natl. Acad. Sci. U S A.* **102**, 8012–8017.
- Hofstadler, S. A., Sampath, R., Blyn, L. B., Eshoo, M. W., Hall, T. A., Jiang, Y., et al. (2005). TIGER: the universal biosensor. *Int. J. Mass. Spectrom.* **242**, 23–41.
- Sampath, R., Hofstadler, S. A., Blyn, L., Eshoo, M. W., Hall, T. A., Massire, C., et al. (2005). Rapid identification of emerging pathogens: coronavirus. *Emerg. Infect. Dis.* **11**, 373–379.
- Sampath, R., Russell, K. L., Massire, C., Eshoo, M. W., Harpin, V., Blyn, L. B., et al. (2007). Global surveillance of emerging influenza virus genotypes by mass spectrometry. *PLoS ONE* **2**, e489.
- Blyn, L. B., Hall, T. A., Libby, B., Ranken, R., Sampath, R., Rudnick, K., et al. (2008). Rapid detection and molecular serotyping of adenovirus by the use of PCR followed by electrospray ionization mass spectrometry. *J. Clin. Microbiol.* **46**, 644–651.
- Eshoo, M. W., Whitehouse, C. A., Zoll, S. T., Massire, C., Pennella, T. T., Blyn, L. B., et al. (2007). Direct broad-range detection of alpha-viruses in mosquito extracts. *Virology* **368**, 286–295.
- Urwin, R., and Maiden, M. C. J. (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol.* **11**, 479–487.
- Farlow, J., Smith, K. L., Wong, J., Abrams, M., Lytle, M., and Keim, P. (2001). *Francisella tularensis* strain typing using multiple-locus, variable-number tandem repeat analysis. *J. Clin. Microbiol.* **39**, 3186–3192.
- Van Ert, M. N., Easterday, W. R., Huynh, L. Y., Okinaka, R. T., Hugh-Jones, M. E., Ravel, J., et al. (2007). Global genetic population structure of *Bacillus anthracis*. *PLoS ONE* **2**, e461.
- Skuce, R. A., McCorry, T. P., McCarroll, J. F., Roring, S. M., Scott, A. N., Brittain, D., et al. (2002). Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. *Microbiology* **148**, 519–528.
- Enright, M. C., Spratt, B. G., Kalia, A., Cross, J. H., and Bessen, D. E. (2001). Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between *emm* type and clone. *Infect. Immun.* **69**, 2416–2427.
- Crum, N. F., Hale, B. R., Bradshaw, D. A., Malone, J. D., Chun, H. M., Gill, W. M., et al. (2003). Outbreak of group A streptococcal pneumonia among marine corps recruits—California, November 1–December 20, 2002. *MMWR Morb. Mortal. Wkly Rep.* **52**, 106–109.
- Tong, M. J. (1972). Septic complications of war wounds. *J. Am. Med. Assoc.* **219**, 1044–1047.

15. Ecker, J. A., Massire, C., Hall, T. A., Ranken, R., Pennella, T. T., Agasino Ivy, C., et al. (2006). Identification of *Acinetobacter* species and genotyping of *Acinetobacter baumannii* by multilocus PCR and mass spectrometry. *J. Clin. Microbiol.* **44**, 2921–2932.
16. van Dessel, H., Dijkshoorn, L., van der Reijden, T., Bakker, N., Paauw, A., van den Broek, P., et al. (2004). Identification of a new geographically widespread multiresistant *Acinetobacter baumannii* clone from European hospitals. *Res. Microbiol.* **155**, 105–112.
17. Dijkshoorn, L., Aucken, H., Gerner-Smidt, P., Janssen, P., Kaufmann, M. E., Garaizer, J., et al. (1996). Comparison of outbreak and nonoutbreak *Acinetobacter baumannii* strains by genotypic and phenotypic methods. *J. Clin. Microbiol.* **34**, 1519–1525.
18. Wortmann, G., Weintrob, A., Barber, M., Scott, P., Zoll, S. T., Eshoo, M. W., et al. (2008). Genotypic evolution of *Acinetobacter baumannii* strains in an outbreak associated with war trauma. *Infect. Control Hosp. Epidemiol.* **29**, 553–555.