# NAVAL HEALTH RESEARCH CENTER

# *RAPID IDENTIFICATION AND STRAIN-TYPING OF RESPIRATORY PATHOGENS FOR EPIDEMIC SURVEILLANCE*


***D. J. Ecker***

| | |
|---|---|
| ***R. Sampath*** | ***J. C. Hannis*** |
| ***L. B. Blyn*** | ***P. M. Harrell*** |
| ***V. Samant*** | ***C. Massire*** |
| ***K. Russell*** | ***T. A. Hall*** |
| ***N. Freed*** | ***Y. Jiang*** |
| ***C. Barrozo*** | ***R. Ranken*** |
| ***J. Wu*** | ***J. J. Drader*** |
| ***K. Rudnick*** | ***N. White*** |
| ***A. Desai*** | ***J. A. McNeil*** |
| ***E. Moradi*** | ***S. T. Crooke*** |
| ***M. W. Eshoo*** | ***S. A. Hofstadler*** |



***Report No. 03-19***





**NAVAL HEALTH RESEARCH CENTER**
**P. O. BOX 85122**
**SAN DIEGO, CA 92186-5122**

**BUREAU OF MEDICINE AND SURGERY (M2)**
**2300 E ST. NW**
**WASHINGTON, DC 20372-5300**

# Rapid Identification and Strain-Typing of Respiratory Pathogens for Epidemic Surveillance

David J. Ecker*†

Rangarajan Sampath†
Lawrence B. Blyn†
Vivek Samant†
Kevin Russell[¥]
Nikki Freed[¥]
Chris Barrozo[¥]
Jianguo Wu[¥,]
Karl Rudnick[§]
Anjali Desai[§]
Emily Moradi[§]
Mark W. Eshoo†
James C. Hannis†
Patina M. Harrell†
Christian Massire†
Thomas A. Hall†
Yun Jiang†
Raymond Ranken†
Jared J. Drader†
Neill White†
John A. McNeil
Stanley T. Crooke†
Steven A. Hofstadler†

[†]Ibis Therapeutics, a division of Isis Pharmaceuticals
1891 Rutherford Rd., Carlsbad, CA 92008

[§]SAIC
10260 Campus Point Drive, San Diego, CA 92121

[¥]Naval Health Research Center
P.O. Box 85122, San Diego, CA, 92186-5122

Report No. 03-19 supported by Work Unit No. 6609. The views expressed in this article are those of the authors and do not reflect the official policy or position of the U.S. Navy, Department of Defense, or the U.S. Government. Approved for public release; distribution unlimited. This research has been conducted in compliance with all applicable federal regulations governing the protection of human subjects in research..

Epidemic respiratory infections are responsible for extensive morbidity and mortality within both military and civilian populations. We describe a methodology to examine respiratory samples that simultaneously identifies broad groups of bacteria. The process uses electrospray ionization mass spectrometry and base composition analysis of broad-range PCR amplification products. The base composition analyses from a small set of broad-range primer pairs are used to "triangulate" the identity of pathogenic organisms present in the sample. Once a species has been identified, the rapid recursive use of species-specific primers to housekeeping genes allows strain-typing. This strategy was used to examine samples from military recruits sickened in a recent Group A streptococcal (GAS) pneumonia outbreak (MMWR 52, 6, p106-109, 2003). The strain-typing results were essentially identical to those obtained using classic *emm* typing and Multi Locus Sequence Typing. This method allows real-time evaluation of patient samples and will make possible more rapid and appropriate treatment of patients in an ongoing epidemic, regardless of the etiology, in a time frame not previously achievable.

**Background**

Despite the prevalence of epidemic respiratory infections and their associated widespread morbidity and mortality, the molecular underpinnings of these conditions remain poorly understood. Epidemic respiratory infections can be caused by a wide variety of bacteria, including several species of *Streptococcus*, *Haemophilus influenzae*, *Staphylococcus aureus*, *Neisseria meningitidis*, or viruses such as influenza, adeno, rhino, and corona (1, 2). While various culture methods, molecular techniques, and serologic diagnostic tests exist, the causative organism is often never determined. Laboratory tests are generally limited to bacterial culture or a molecular test for a single viral or bacterial agent, and results are seldom available rapidly enough to influence intervention efforts. Ability to track epidemic outbreaks that are dispersing geographically is often limited to monitoring of vague disease classifications (ICD-9 codes) that lack specific laboratory diagnoses.

Group A streptococci (GAS), or *Streptococcus pyogenes*, is one of the most important causes of respiratory infections because of its prevalence and ability to cause severe disease with complications, such as acute rheumatic fever and acute glomerulonephritis (3). GAS also causes infections of the skin (impetigo) and, in rare cases, invasive disease such as necrotizing fascititis (flesh-eating bacteria) and toxic shock syndrome. GAS outbreaks are enhanced by the crowded conditions and close physical contact that occurs in civilian schools, correctional facilities, and military training barracks. Much of our current knowledge of the epidemiology of GAS was obtained from pioneering studies in the 1950s in military environments (4-6). The disease is spread by direct person-to-person contact via droplets or nasal secretions. The development in the 1930s of the Lancefield classification of strains into distinct serogroups based on the antigenic properties of the cell surface M-protein was an important step in understanding GAS infections; the surface M-protein plays a critical role as a primary virulence factor (7, 8).

However, after many decades of study, the underlying microbial ecology and natural selection that favors enhanced virulence and explosive GAS outbreaks is still poorly understood. The ability to simultaneously identify GAS and other bacteria and viruses in patient samples would greatly facilitate our understanding of respiratory epidemics. It is also essential to be able to follow the spread of virulent strains of GAS in populations and to distinguish virulent strains from less virulent or avirulent streptococci that colonize the nose and throat of asymptomatic individuals at a frequency ranging from 5-20% of the population (3). Molecular methods have been developed to type GAS based upon the sequence of the *emm* gene that encodes the M-protein virulence factor (9-11). Using this molecular classification, over 150 different *emm* types have been defined and correlated with phenotypic properties of thousands of GAS isolates (http://www.cdc.gov/ncidod/biotech/strep/strepindex.html) (12). Recently, a strategy known as Multi Locus Sequence Typing (MLST) was developed to follow the molecular epidemiology of GAS (13). In MLST, internal fragments of seven housekeeping genes are amplified, sequenced, and compared with a database of previously studied isolates (http://test.mlst.net/). The results from MLST are highly

concordant with several other typing methods (13). Polymorphisms in other genes have also been correlated with M-protein type virulence properties (14-17).

While these methods provide detailed analysis of isolated GAS strains, each of these techniques requires culture, colony isolation, PCR amplification, and sequencing, and thus these methods are inherently limited. Culture and selection of isolated colonies of *S. pyogenes* on selective blood agar plates requires a minimum of 24 hr, and PCR amplification followed by sequencing is a slow and labor-intensive process. We now report a new technique that rapidly identifies the presence of multiple respiratory microorganisms simultaneously. In the case of a GAS outbreak, the organism can be identified and its *emm* type determined directly from throat swabs within 6 hr. These attributes allow strain tracking of an ongoing, geographically dispersed epidemic on a larger scale than ever before achievable. As an example, we describe characterization of a GAS outbreak at a military training camp (18) and analyze its potential spread to other military facilities.

**Materials and Methods**

**Genome preparation and PCR***:* Genomic materials from culture samples or swabs were prepared using the DNeasy 96 Tissue Kit (Qiagen, Valencia, CA) using manufacturer's procedures. PCR reactions were performed using Platinum Taq (Invitrogen, Carlsbad, CA) and Hotstart PFU Turbo (Stratagene, La Jolla, CA ) polymerases. Cycling conditions consisting of an initial 2 min at 95°C followed by 45 cycles of 20 s at 95°C, 15 s at 58°C, and 15 s at 72°C. Broad-range PCR primers were designed to conserved regions of bacterial ribosomal RNAs (16 and 23S) and the gene encoding DNA-dependent RNA polymerase, B' subunit (rpoC) (Table 1). The allelic profile of a GAS strain by MLST can be obtained by sequencing the internal fragments of seven housekeeping genes. The nucleotide sequences for these genes, from 212 isolates of GAS (78 distinct *emm* types), are available at the Web site: http://www.mlst.net. This corresponds to one hundred different allelic profiles referred to by Enright et al. as ST1-ST100 (13). For each profile, we created a virtual transcript for each allelic profile by concatenating sequences from each of the seven genes. Primers were designed using these sequences and were constrained to be within each gene loci. Twenty-four primer pairs were designed and tested against GAS strain 700294. A final subset of six primer pairs (Table 1) was chosen based on a theoretical calculation of minimal number of primer pairs that maximized resolution of between *emm* types.

**Mass spectrometry and base composition analysis:** Following amplification, 15 uL aliquots of each PCR were desalted and purified using a weak anion exchange protocol as described in detail elsewhere (19). Accurate-mass (+ 1 ppm) high-resolution (M/ M > 100,000 FWHM) mass spectra were acquired for each sample using high throughput electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI-FTICR) protocols described previously for nucleic acid analysis (20). For each sample approximately 1.5 uL of analyte solution was consumed during the 74-s spectral acquisition. Raw mass spectra were postcalibrated with an internal mass standard and deconvolved to monoisotopic molecular masses. Unambiguous base compositions were derived from the exact mass measurements of the complementary single stranded oligonucleotides (21).

**Microbiology:** GAS isolates were identified from swabs on the basis of colony morphology and beta-hemolysis on blood agar plates, Gram stain characteristics, susceptibility to bacitracin, and positive latex agglutination reactivity with group A-specific antiserum.

**Sequencing:** Bacterial genomic DNA samples of all isolates were extracted from freshly grown GAS strains using QIAamp DNA Blood Mini Kit (Qiagen). Group A streptococcal cells were analyzed using *emm* gene -specific PCR as previously described (9, 22). Homology searches on DNA sequences were conducted against known *emm* sequences (http://www.cdc.gov/ncidod/biotech/infotech_hp.html). MLST analysis was performed as previously described (13).

## Results and Discussion

### Example of broad surveillance, identification, and rapid strain-typing of bacterial pathogens

In managing epidemic outbreaks of respiratory disease, it would be valuable to analyze clinical samples for a broad range of bacterial pathogens, and then, based upon the organism identified, rapidly determine additional organism-specific details useful for making treatment or quarantine decisions. We have developed a method to achieve this and tested it on samples obtained from an outbreak of *S. pyogenes* in a military training camp. The first step, called universal survey, employs PCR primers that allow amplification and identification of many different pathogens. Based upon what is found in the universal survey, an organism-specific drill-down step is immediately employed to obtain the desired additional information (Figure 1). For example, after identification of *S. pyogenes* in a throat swab, one might want to determine its *emm* type, since this information is useful both for treatment considerations and epidemic surveillance. Initially, we examined isolated colonies from throat culture samples, but subsequently analyzed throat swabs directly without the culture step. The latter path can be completed within 6-12 hr after sample acquisition, providing information rapidly enough to be useful in managing an ongoing epidemic.

The experimental methodology is based upon analysis of PCR amplicons using ESI-FTICR and deciphering the base compositions of the amplicons from highly accurate mass measurements (Figure 2) (19, 23). ESI-FTICR is a platform that has been used for automated high throughput drug screening (20, 24, 25), where samples containing a complex mixture of PCR amplicons can be analyzed at a rate of one per minute (19), and the results can be immediately used to direct a predetermined drill-down path in an automated fashion. In the current effort, throat swabs or culture sample were analyzed by extraction of total nucleic acids and surveyed for pathogenic respiratory bacteria using a small set of broad-range PCR primers. Upon identification of *S. pyogenes*, the same nucleic acid extract was reanalyzed using a set of primers specific for *S. pyogenes* designed to decipher its *emm* type.

The universal survey primers were chosen by a computational analysis of sequence alignments of the ribosomal operons and 160 broadly conserved protein-encoding housekeeping genes of all sequenced bacteria (26, 27). The target locations of the primers were selected based upon (a) the ability to broadly amplify all bacteria, but not eukaryotes or archaea; (b) the ability to maximally distinguish bacterial species from each other by the information content of the base composition across all primer pairs used; and (c) an upper limit on amplicon size (currently <= 140 base pairs) where ionization and accurate deconvolution can be achieved.

Four pairs of broad-range primer pairs were selected that include regions of 16S and 23S rRNA, and the gene encoding DNA-dependent RNA polymerase, B' subunit (rpoC) (Figure 1B, red labels). While there was no special consideration of *S. pyogenes* in the selection of the universal survey primers, analysis of genomic sequences shows that the base compositions of these regions distinguished Streptococcus pyogenes from other respiratory pathogens and normal flora, including closely related species of streptococci, staphylococci, and bacilli (Figure 3). While any single primer might have an overlap of base compositions for two or more organisms (see, for example, the rpoC compositions for *Bacillus anthracis* and *Stapyhlocococcus aureus*), combining information across all primer pairs provided unique organism-specific signature.

Fifty-one GAS isolates were taken from healthy recruits and hospitalized patients in December 2002, during the peak of the military training camp outbreak. Twenty-seven additional isolates from previous infections ascribed to GAS were also examined. The data obtained from the 51 epidemic samples are shown in Figure 3. This plot shows the base composition analysis for the four universal survey primer pairs using a different symbol for each primer pair. The upper panel shows the expected distribution of based compositions from multiple isolates of respiratory pathogens derived from sequences in Genbank

(color-coded by species and color-matched lines connecting different sequenced strains). The lower panel shows the observed base compositions for the epidemic samples. All the base compositions were consistent with those from the four completely sequenced strains of *S. pyogenes* (28-33). Thus, the epidemic samples were clearly identified as GAS. At the outset, it was not clear how many universal primers would be needed to identify *S. pyogenes* in the background of normal throat flora. Surprisingly, the experimental results showed that as few as two primer pairs are sufficient for the initial surveillance of bacterial pathogens in respiratory samples.

**Determining strain-specific signatures and *emm* typing**

In order to obtain strain-specific information about the epidemic, we designed a strategy to generate strain-specific signatures and simultaneously correlate with *emm* types. In classic MLST analysis, internal fragments of seven housekeeping genes (Figure 2B, blue labels) are amplified and sequenced (13). Since our method of analysis provides base composition data rather than sequence, the challenge was to identify the target regions that provide the highest resolution of species and least ambiguous *emm* classification. We constructed an alignment of concatenated alleles of the seven housekeeping genes from each of 212 previously *emm*-typed strains (13). From this alignment, we determined the number and location of the primer pairs that would maximize strain discrimination using base composition data.

An initial set of 24 primer pairs were selected that amplify regions that covered over 97% of the nucleotide variation in the alignment. We then analyzed these primers to determine how much strain discrimination could be achieved by base composition analysis of different subsets of the primers. For a given subset, the measure of discrimination performance was defined as the ratio of the number unique *emm* types over the total number of *emm* types represented in the alignment. The upper bound of performance when all 24 pairs are used is approximately 97% (see supplemental data for details of performance calculations). Performance calculations for different possible combinations of primer subsets showed an inflection point at six pairs, where 89% of the *emm* types could be discriminated. This degree of resolution is sufficient for many applications, such as real-time tracking of an epidemic strain. However, if complete *emm* typing is required, additional primers selected to specifically resolve the encountered ambiguities can be applied.

The results of the base composition analysis with six primer pairs, 5'-emm gene sequencing, and the MLST gene sequencing methods for samples from the epidemic, archived samples and follow-up epidemiology studies are compared in Table 2. The ambiguities of *emm* type assignment from base composition analysis from six primer pairs are shown without further resolution by additional recursive analysis steps. Although not all samples were completely resolved to a unique *emm* type using six primer pairs, base compositions showed the correct identification (either uniquely, or as a member of a small set) as 5'-*emm* gene sequencing or full the MLST sequencing method. Of the 51 samples taken during the peak of the epidemic (Table 2, first 3 rows, highlighted in yellow), all but three had identical compositions and corresponded to *emm* 3. The three outliers, all from healthy individuals, probably represent nonepidemic strains harbored by asymptomatic carriers. Archived samples (Table 2, next nine rows) from previous infections at other training facilities, showed a much greater heterogeneity of composition signatures and *emm* types, as would be expected in the absence of a specific epidemic, where a clonal expansion of a single virulent strain is occurring.

**Postepidemic surveillance of GAS at other military facilities**

The November/December 2002 epidemic was caused by a virulent *emm* 3 strain. Following this epidemic, it was considered important to survey GAS outbreaks at other military facilities populated by military recruits who had completed their training, and who might have carried the epidemic strain to

these locations. Culture samples from GAS-positive patients were obtained from other military bases following the epidemic. These samples were analyzed by base composition analysis and by *emm*-gene sequencing. The results (Table 2) showed concordance between base composition analysis and *emm*-gene sequencing. One or two samples from each location had an *emm* 3. However, the distribution of GAS types at these locations showed a pattern significantly different from the original epidemic, suggesting that the epidemic strain was not dominating the population of GAS at other locations. In support of this, hospitalized pneumonia morbidity was not seen at other military locations.

**Culture-free analysis of swabs by direct PCR**

To determine whether GAS could be identified directly from swabs without culturing cells, 8 throat swabs from individuals showing respiratory symptoms were obtained. Five of the eight patients tested positive for GAS by culture (Table 2). All culture-positive swabs were also GAS-positive by base composition analysis, while the 3 culture-negative samples did not produce identifiable amplicons for any of the primers. To test the sensitivity directly on swabs, we performed a limiting dilution experiment using *S. pyogenes* spiked onto either dry swabs or onto swabs with normal flora backgrounds from healthy volunteers. Each of the *emm* typing primer pairs gave strong, single base composition detections at a lower level averaging approximately 14 CFU per PCR for both pure swabs and for the swabs containing normal respiratory flora. The direct swab material was also examined using the two 16S broad-range primers. Four of the five culture-positive samples showed base compositions consistent with *S. pyogenes*, while the three culture-negative samples were negative for *S. pyogenes* by broad-range priming. The broad-range primers also showed evidence of other throat flora with identified base compositions consistent with species of *Pseudomonas, Moraxella, Corynebacteria, Acidophilus, Haemophilus, Actinobacillus, Clostridium, Pasteurella, and Ralstonia*. Remarkably, two of the culture-positive samples showed such an intense signal for *S. pyogenes* that no other organism could be reliably identified in these samples by mass spectrometry analysis of broad-range PCR products (which has a dynamic range of approximately 1,000), suggesting that the pathogenic *S. pyogenes* titer was numerically overwhelming in these patients with respect to other background throat flora.

**Conclusions**

We have developed a strategy to survey respiratory samples for the presence of many different pathogenic agents simultaneously, and then to determine additional valuable pathogen-specific information. Using a relatively small set of universal survey primers targeted to broadly conserved regions of bacterial genomes, PCR amplicons are generated. The amplicons are analyzed by ESI-FTICR mass spectrometry, and the identity of pathogens present is determined by triangulating the base composition data from multiple primer pairs. In order to track and understand an epidemic, additional information is needed that is different for each pathogen. In the case of *S. pyogenes*, the *emm* type can be used both to track the spread of a virulent strain in a population and to correlate with previously determined phenotypic information, such as the respiratory virulence or tendency to be invasive.

In the example detailed here, the survey primers identified *S. pyogenes* and the drill-down primers determined the *emm* type. Our method gave results in concordance with the established methods of *emm* typing in a fraction of the time. Previous methods used to determine the *emm* type of a strain of *S. pyogenes* require bacterial culture, selection of isolated colonies, PCR amplification, and sequencing of the *emm* gene, or other genes that correlate with *emm* type. Culture typically requires a minimum of 24-hr delay and PCR amplification followed by sequencing is impractical to perform on a large number of samples in the time needed track an ongoing epidemic. Had a different organism been identified, a different drill-down question could have been answered using an appropriate set of primers. The advantage of this recursive capability is that it can be automated, and a great deal of information can be obtained in a short time. In the current method, the first sample is analyzed in the mass spectrometer within 6 hr of sample collection, and additional samples are analyzed in the mass spectrometer at a rate of one per minute thereafter. Prompt analysis of patient samples using the “survey drill-down” strategy will

allow medical personnel to take immediate action in treatment and/or isolation of patients necessary to halt an epidemic.

**Acknowledgments**

The authors gratefully acknowledge the Defense Advanced Research Projects Agency for financial support for this technology under the TIGER program. We also recognize the efforts of Dr. Bradley Hale and Dr. Nancy Crum of the Department of Infectious Diseases, Naval Medical Center San Diego, for their initial collection of the GAS samples used in this project.

To whom correspondence should be addressed: e-mail: decker@isisph.com

## References


1. Gwaltney, J. M. (1995) in *Principles and Practice of Infectious Diseases*, eds. Mandell, G. L., Bennett, J. E. & Dolin, R. (Churchill Livingstone, New York), Vol. 1, pp. 566-572.
2. Smith, H. & Sweet, C. (2002) in *Polymicrobial Diseases*, eds. Brogen, K. A. & Guthmiller, J. M. (ASM Press, Washington, DC), pp. 201-212.
3. Bisno, A. L. (1995) in *Principles and Practice of Infectious Diseases*, eds. Mandell, G. L., Bennett, J. E. & Dolin, R. (Churchill Livingston, New York), Vol. 2, pp. 1786-1799.
4. Rammelkamp, C. H. (1955) *Harvey Lectures* 51, 113-142.
5. Wannamaker, L. W. (1953) in *Streptococcal infections*, ed. McCarty, M. (Columbia University Press, New York), pp. 157-175.
6. Rammelkamp, C. H., Denny, F. W. & Wannamaker, L. W. (1952) in *Rheumatic Fever*, ed. Thomas, L. (University of Minnesota Press, Minneapolis), pp. 72-89.
7. Lancefield, R. C. (1933) *J . Exp. Med.* 57, 571-595.
8. Lancefield, R. C. (1962) *J. Immunol.* 89, 307-313.
9. Beall, B., Facklam, R. & Thompson, T. (1996) *J. Clin. Micro.* 34, 953-958.
10. Beall, B., Facklam, R., Hoenes, T. & Schwartz, B. (1997) *J. Clin. Micro.* 35, 1231-1235.
11. Facklam, R., Beall, B., Efstratiou, A., Fischetti, V., Johnson, D., Kaplan, E., Kriz, P., Lovgren, M., Martin, D., Schwartz, B., et al. (1999) *Emerg. Infect. Dis.* 5, 247-253.
12. Facklam, R., Martin, D. R., Lovgren, M., Johnson, D. R., Efstratiou, A., Thompson, T. A., Gowan, S., Kriz, P., Tyrrell, G. J., Kaplan, E. & Beall, B. (2002) *Clin. Infect Dis.* 34, 28-38.
13. Enright, M. C., Spratt, B. G., Kalia, A., Cross, J. H. & Bessen, D. E. (2001) *Infection and Immunity* 69, 2416-2427.
14. Reid, S. D., Green, N. M., Buss, J. K., Lei, B. & Musser, J. M. (2001) *Proc. Natl. Acad. Sci. USA* 98, 7552-7557.
15. Reid, S. D., Green, N. M., Sylva, G. L., Voyich, J. M., Stenseth, E. T., DeLeo, F. R., Palzkill, T., Low, D. E., Hill, H. R. & Musser, J. M. (2002) *J. Bact.* 184, 6316-6324.
16. Tyrrell, G. J., Lovgren, M., Forwick, B., Hoe, N. P., Musser, J. M. & Talbot, J. A. (2002) *J. Clin. Micro.* 40, 4466-4471.
17. Hoe, N. P., Nakashima, K., Lukomski, S., Grigsby, D., Liu, M., Kordari, P., Dou, S.-J., Pan, X., Vuopio-Varkila, J., Salmelinna, S., et al. (1999) *Nature Medicine* 5, 924-929.
18. Crum, N. F., Hale, B. R., Bradshaw, D. A., Malone, J. D., Chun, H. M., Gill, W. M., Norton, D., Lewis, C. T., Truett, A. A., Beadler, C., et al. (2003) *MMWR Morb. Mortal. Wkly. Rep.* 52, 106-109.
19. Jiang, Y. & Hofstadler, S. A. (2003) *Anal. Biochem.* 316, 50-57.
20. Sannes-Lowery, K. A., Drader, J. J., Griffey, R. H. & Hofstadler, S. A. (2000) *TrAC, Trends Anal. Chem.* 19, 481-491.
21. Muddiman, D. C., Anderson, G. A., Hofstadler, S. A. & Smith, R. D. (1997) *Anal. Chem.* 69, 1543-1549.

22. Facklam, R., Beall, B., Efstratiou, A., Fischetti, V., Kaplan, E., Kriz, P., Lovgren, M., Martin, D., Schwartz, B., Totolian, A., et al. (1999) *Emerg. Infect. Dis.* 5, 247-253.

23. Muddiman, D. C., Anderson, G. A., Hofstadler, S. A. & Smith, R. D. (1997) *Anal. Chem.* 69, 1543-1549.

24. Hofstadler, S. A., Sannes-Lowery, K. A., Crooke, S. T., Ecker, D. J., Sasmor, H., Manalili, S. & Griffey, R. H. (1999) *Anal. Chem.* 71, 3436-3440.

25. Griffey, R. H., Hofstadler, S. A., Sannes-Lowery, K. A., Ecker, D. J. & Crooke, S. T. (1999) *Proc. Natl. Acad. Sci. USA* 96, 10129-10133.

26. Mushegian, A. R. & Koonin, E. V. (1996) *Proc. Natl. Acad. Sci. U S A* 93, 10268-10273.

27. Maidak, B. L., Cole, J. R., Lilburn, T. G., Parker, C. T., Saxman, P. R., Farris, R. J., Garrity, G. M., Olsen, G. J., Schmidt, T. M. & Tiedje, J. M. (2001) *Nuc. Acids Res.* 29, 173-174.

28. Ferretti, J. J., McShan, W. M., Ajdic, D., Savic, D. J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A. N., Kenton, S., Lai, H. S., et al. (2001) *Proc. Natl. Acad. Sci. USA* 98, 4658-4663.

29. Tettelin, H., Masignani, V., Cieslewica, M. J., Eisen, J. A., Peterson, S., Wessels, M. R., Paulsen, I. T., Nelson, K. E., Margarit, I., Read, T. D., et al. (2002) *Proc. Natl. Acad. Sci. USA* 99, 12391-12396.

30. Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J., et al. (2001) *Science* 293, 498-506.

31. Smoot, J. C., Barbian, K. D., Van Gompel, J. J., Smoot, L. M., Chaussee, M. S., Sylva, G. L., Sturdevant, D. E., Ricklefs, S. M., Porcella, S. F., Parkins, L. D., et al. (2002) *Proc. Natl. Acad. Sci. USA* 99, 4668-4673.

32. Beres, S. B., Sylva, G. L., Barbian, K. D., Lei, B., Hoff, J. S., Mammarella, N. D., Liu, M.-Y., Smoot, J. C., Porcella, S. F., Parkins, L. D., et al. (2002) *Proc. Natl. Acad. Sci. USA* 99, 10078-10083.

33. Ajdic, D., McShan, W. M., McLaughlin, R. E., Savic, G., Chang, J., Carson, M. B., Primeaux, C., Tian, R., Kenton, S., Jia, H., et al. (2002) *Proc. Natl. Acad. Sci. USA* 99, 14434-14439.

To whom correspondence should be addressed: e-mail: decker@isisph.com

--- figures not provided with e-copy---

**Figure 1.**

A. Overall process for simultaneous GAS detection and emm typing. Genetic material is extracted from throat swabs or cultured samples and amplified using universal survey primers. The PCR products are analyzed by mass spectrometry to identify important pathogens. If GAS is detected, the emm typing primers are immediately used to reexamine the extract in order to provide more detailed information on the strain. If a different primary pathogen were discovered by the survey primer analysis, a different recursive analysis would be performed appropriate to the information desired for that specific pathogen.

B. Target genes for broad detection and emm typing of *Streptococcus pyogenes*. Genes used for broad priming and initial detection and identification of all bacteria present are shown in red. Upon identification of *Streptococcus pyogenes*, a second set of primers targeting the genes used in Multi Locus Sequence Typing (MLST), shown in blue, were used to assign the strain a base composition code and infer emm-type. The emm1 gene is shown in green.

**Figure 2.**

Deconvoluted ESI-FTICR spectra of the PCR products produced by the *gtr* primer for samples corresponding to emm types 3 (red) and 6 (blue), respectively. Accurate mass measurements were obtained by using an internal mass standard and post calibrating each spectrum; the experimental mass measurement uncertainty on each strand is ± 0.035 Daltons (1 ppm). Unambiguous base compositions of the amplicons were determined by calculating all putative base compositions of each stand within the measured mass (and measured mass uncertainty).In all cases there was only one base composition within 25 ppm. Note that the measured mass difference of 15.985 Da between the strands shown on the left is in

excellent agreement with the theoretical mass difference of 15.994 Da expected for an A to G substitution.

--- figure not provided with e-copy---

**Figure 3.**

Base composition analysis for broad respiratory pathogens using four primer pairs. The symbol shape corresponds to the primer pair used and the colors represent different organisms. Note that *Streptococcus pyogenes* is shown in blue. A.)The distribution of base compositions for representative respiratory pathogens. B.)The results for 51 epidemic samples colored in red with the respiratory pathogen background in gray.

--- figure not provided with e-copy---

# REPORT DOCUMENTATION PAGE



**1. Report Date (DD MM YY)**
8 April 2003

**2. Report Type**
New

**3. DATES COVERED (from - to)**
2003

**4. TITLE AND SUBTITLE**
Rapid Identification and Strain-Typing of Respiratory Pathogens for Epidemic Surveillance

**5a. Contract Number**:
**5b. Grant Number**:
**5c. Program Element**:
**5d. Project Number:**
**5e. Task Number:**
**5f. Work Unit Number:** 6609
**5g. IRB Protocol Number:** 32287

**6. AUTHORS**
David J. Ecker, Rangarajan Sampath, Lawrence B. Blyn, Vivek Samant, Kevin Russell, Nikki Freed, Chris Barrozo ,Jianguo Wu, Karl Rudnick, Anjali Desai, Emily Moradi ,Mark W. Eshoo, James C. Hannis, Patina M. Harrell, Christian Massire, Thomas A. Hall, Yun Jiang, Raymond Ranken, Jared J. Drader, Neill White, John A. McNeil, Stanley T. Crooke, Steven A. Hofstadler.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

**9. PERFORMING ORGANIZATION REPORT NUMBER**
Report No. 03-19

**8. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)**
Chief, Bureau of Medicine and Surgery
Code M2
2300 E St NW
Washington DC 20372-5300

**10. Sponsor/Monitor's Acronyms(s)**
**BUMED**

**11. Sponsor/Monitor's Report Number(s)**

**12 DISTRIBUTION/AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited.

**13. SUPPLEMENTARY NOTES**

14. ABSTRACT (maximum 200 words) Epidemic respiratory infections are responsible for extensive morbidity and mortality within both military and civilian populations. We describe a methodology to examine respiratory samples that simultaneously identifies broad groups of bacteria. The process uses electrospray ionization mass spectrometry and base composition analysis of broad-range PCR amplification products. The base composition analyses from a small set of broad-range primer pairs are used to "triangulate" the identity of pathogenic organisms present in the sample. Once a species has been identified, the rapid recursive use of species-specific primers to housekeeping genes allows strain-typing. This strategy was used to examine samples from military recruits sickened in a recent Group A streptococcal (GAS) pneumonia outbreak (MMWR 52, 6, p106-109, 2003). The strain-typing results were essentially identical to those obtained using classic *emm* typing and Multi Locus Sequence Typing. This method allows real-time evaluation of patient samples and will make possible more rapid and appropriate treatment of patients in an ongoing epidemic, regardless of the etiology, in a time frame not previously achievable.

**15. SUBJECT TERMS**
respiratory infections, bacteria, military recruits, Group A Strteptococcal,

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b.ABSTRACT** | **c. THIS PAGE** | | | Commanding Officer |
| UNCL | UNCL | UNCL | UNCL | **12** | **19b. TELEPHONE NUMBER (INCLUDING AREA CODE)** COMM/DSN: (619) 553-8429 |