

NRL/6920/FR--2021/1

Silent Guardian at NRL October 2004–March 2005

DAVID A. STENGER

*Branch Head
Center for Biomolecular Science & Engineering*

CHRIS TAITT
LISA C. SHRIVER-LAKE

*Laboratory for Biomolecular Dynamics
Center for Biomolecular Science & Engineering*

ANTHONY P. MALANOSKI

*Laboratory for Biosensors and Biomaterials
Center for Biomolecular Science & Engineering*

FRANCES S. LIGLER
JOEL M. SCHNUR
ANNE W. KUSTERBECK

*Center for Biomolecular Science & Engineering
Retired NRL Employee*

STEVE K. POLLACK
*Carbon, Inc. Life Sciences
Redwood, CA*

July 29, 2021

REPORT DOCUMENTATION PAGE					Form Approved OMB No. 0704-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.						
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.						
1. REPORT DATE (DD-MM-YYYY) 29-07-2021		2. REPORT TYPE NRL Formal Report		3. DATES COVERED (From - To) October 2004–March 2005		
4. TITLE AND SUBTITLE Silent Guardian at NRL October 2004–March 2005				5a. CONTRACT NUMBER MIPR5ANRL000222/MIPR5ANRLPSG16		
				5b. GRANT NUMBER		
				5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) David S. Stenger, Chris R. Taitt, Lisa C. Shriver-Lake, Anthony P. Malanoski, Frances S. Ligler*, Anne W. Kusterbeck*, Joel M. Schnur, and Steve K. Pollack**				5d. PROJECT NUMBER		
				5e. TASK NUMBER		
				5f. WORK UNIT NUMBER 8794		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Research Laboratory 4555 Overlook Avenue, SW Washington, D.C. 20375-5320				8. PERFORMING ORGANIZATION REPORT NUMBER NRL/6920/FR--2021/1		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of the Secretary of Defense Nuclear, Chemical and Biological Programs 100 S Washington Blvd, Arlington, VA 22202				10. SPONSOR/MONITOR'S ACRONYM(S) OSD		
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT In Nov 2004, NRL and the USAF Surgeon General accelerated the development of an emergent resequencing microarray system per direction of the Deputy Assistant to the Secretary of Defense, Nuclear, Chemical and Biological Programs. This early demonstration project was called Silent Guardian (SGD). During SGD approximately 5,000 RPM assays were performed using patient samples from six MTFs and spiked normal nasal wash, demonstrating: (1) a capability to establish a full production, high throughput CLIA quality analysis laboratory using high density resequencing microarrays in under 2 months, (2) a capability to integrate disparate data systems for collection of clinical data, tracking analysis and reporting the result to geographically dispersed PIs, (3) that hundreds of diverse genetic sequences for common respiratory pathogens causing illness in the general population could be obtained and unequivocally identified in approximately 18-19 hours, and (4) that the system was capable of sensitive detection and discrimination of a panel of biothreat agents and discriminated from nearly identical non-pathogenic organisms.						
15. SUBJECT TERMS Advanced diagnostics Biosurveillance Epidemic Outbreak Surveillance resequencing arrays						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT	b. ABSTRACT	c. THIS PAGE			Lisa C Shriver-Lake	
U	U	U	UU	81	19b. TELEPHONE NUMBER (Include area code) (202) 404-6045	

Reset

This page intentionally left blank

TABLE OF CONTENTS

EXECUTIVE SUMMARY	E-1
CHAPTER 1: PROJECT OVERVIEW	1
Introduction	1
Results	2
<i>Achievement of Full Operational Capability (FOC).....</i>	<i>2</i>
<i>Common Respiratory Pathogen Detection in Clinical Specimens</i>	<i>2</i>
<i>BT Agent Detection in Clinical Specimens</i>	<i>3</i>
<i>Detection and Discrimination of BT Agents and Near Neighbors</i>	<i>3</i>
<i>Common Respiratory Pathogen Detection in BT Spiked Experiments</i>	<i>4</i>
Conclusions and Recommendations	4
CHAPTER 2: BACKGROUND	5
Existing Technology	5
DNA Microarrays	5
EOS Program	6
RPM v.1 Microarray Development	7
CHAPTER 3: SYSTEM DEVELOPMENT.....	9
Background	9
Biochemistry.....	9
Bioinformatics and Information Technology	10
<i>Operating Principles</i>	<i>12</i>
Project Assembly.....	17
<i>Staffing.....</i>	<i>17</i>
<i>Analysis of Protocol and Identification of Personnel Requirements.....</i>	<i>17</i>
<i>Protocol Training.....</i>	<i>18</i>
Procurement of Equipment and Reagents.....	18
Facilities.....	19
Computer Hardware.....	20
Data Storage and Backup for SGD.....	21
Xserve BLAST Server.....	21
LIMS Integration with COHORT	21
Safety	24
Quality Assurance / Quality Control.....	24
Troubleshooting and Establishment of Final SGD Protocol.....	27
CHAPTER 4: PATIENT RESULTS.....	29
Overview	29
Background	30
Experimental Design and Methods.....	31
<i>Aliquoting of Patient Specimens</i>	<i>31</i>
<i>RPM v.1 Sample Processing and Analysis</i>	<i>31</i>

<i>Microbiological Laboratory Analysis</i>	<i>31</i>
Results and Discussion	32
<i>Number of Patients Tested.....</i>	<i>32</i>
<i>Microbiological Laboratory Analyses.....</i>	<i>34</i>
<i>Overall Microarray Results versus Culture.....</i>	<i>34</i>
<i>Influenza A Detection</i>	<i>37</i>
<i>Influenza B Detection</i>	<i>39</i>
<i>Influenza A/B Tile Data</i>	<i>39</i>
<i>Human Coronavirus Detection</i>	<i>40</i>
<i>Other RNA Viruses</i>	<i>40</i>
<i>Adenovirus Serotypes and Possible Co-Infections.....</i>	<i>40</i>
<i>Antibiotic Resistance Markers</i>	<i>41</i>
<i>Attempt to Correlate Patient Data with Culture Data.....</i>	<i>41</i>
<i>BT Agents in Clinical Specimens</i>	<i>42</i>
<i>Blind B. anthracis Spike Tests.....</i>	<i>42</i>
Summary and Conclusions	42
 CHAPTER 5: DETECTION AND DISCRIMINATION OF BIOTHREAT AGENTS AND NEAR NEIGHBORS	 44
Overview	44
Background	44
Experimental Design and Methods.....	47
Results and Discussion	48
<i>Detection of B. anthracis Sequences</i>	<i>48</i>
<i>Detection of F. tularensis Sequences.....</i>	<i>53</i>
<i>Detection of Y. pestis Sequences.....</i>	<i>56</i>
<i>Detection of Vaccinia virus (smallpox simulant) sequences.....</i>	<i>58</i>
<i>Unanticipated results</i>	<i>59</i>
Summary and Conclusions	61
 CHAPTER 6: CONCLUSIONS, LESSONS LEARNED, RECOMMENDATIONS.....	 63
 ADDENDUM	 66
 ACKNOWLEDGEMENTS.....	 66
 REFERENCES	 68
 APPENDIX A: MEMORANDUM STARTING SILENT GUARDIAN.....	 70
 APPENDIX B: STANDARD OPERATING PROTOCOLS	 71
 APPENDIX C: PRIMER LISTS.....	 72
 APPENDIX D: STAFF AT NRL DURING SGD.....	 73
 APPENDIX E: POSTER DESCRIBING NEW AND IMPROVED PROTOCOL.....	 75

ACRONYMS

ACTD	Advanced Concept Technology Demonstration
AF	Air Force
AFIOH	Air Force Institute for Operational Health, Brooks-City Base, TX
AFMS	Air Force Medical Service
AF/SG	Air Force Surgeon General
ATL	Acquisition, Technology and Logistics
BLAST	Basic Local Alignment Search Tool
BMT	Basic Military Trainee
BSR	Biological Safety Review
BT	Biothreat
CAP	College of American Pathologists
CBMSE	Center for Bio/Molecular Science and Engineering
CDC	Centers for Disease Control
cfu	Colony Forming Units
CLIA	Clinical Laboratory Improvement Amendments
COHORT	Composite Occupational Health and Operational Risk Tracking
CPE	Cytopathic Effect
DATSD	Deputy Assistant to the Security of Defense
DoD-CLIP	DoD Clinical Laboratory Improvement Program
DSB	Defense Sciences Board
EOS	Epidemic Outbreak Surveillance
FDA	United States Food and Drug Administration
FF BAT	Air Force Biological Augmentation Team
FOC	Full Operational Capability
FRI	Febrile respiratory illness
GCOS	GeneChip® Operating System
GDAS	GeneChip® Data Analysis System
HQ USAF/SGR	AF/SG Modernization Directorate
HSBS	Health Surveillance Biodefense Systems
IRB	Institutional Review Board
IT	Information Technology
JPEO-CBD	Joint Program Executive Office for Chemical and Biological Defense
LIMS	Laboratory Information Management System
MDW	Military District Washington
MTF	Medical Treatment Facility
NA	Nucleic Acid
NCBI	National Center for Biotechnology Information
NHRC	Naval Health Research Center, San Diego, CA
NIH	National Institutes of Health
NRL	Naval Research Laboratory, Washington, DC
NRL-BSC	NRL Biological Safety Committee
PCR	Polymerase Chain Reaction
pfu	Plaque Forming Unit
PI	Principle Investigator
PIN	Patient Identification Number
PPE	Personal Protective Equipment

QC/QA	Quality Control/Quality Assurance
REPI	Respiratory Pathogens Identifier
RPM	Respiratory Pathogen Microarray
SGD	Silent Guardian Demonstration Project
S/N	Signal-to-Noise
SOP	Standard Operating Procedure
USN	United States Navy

EXECUTIVE SUMMARY

In November 2004, the Naval Research Laboratory (NRL) and the Air Force Surgeon General (AF/SG) accelerated the development of an emergent resequencing microarray system per the direction of the Deputy Assistant to the Secretary of Defense, Nuclear, Chemical and Biological Programs. The AF/SG was tasked with implementing this capability as a demonstration from December 1, 2004, to March 31, 2005, in the Military District Washington (MDW). This early demonstration project was called Silent Guardian (SGD). The AF/SG Modernization Directorate (HQ USAF/SGR) managed the overall project, including institutional review board (IRB) applications, clinical sample collection, and reporting to the military treatment facilities (MTFs) and DoD chain of command. NRL was tasked to provide laboratory space and personnel to process collected samples; to provide appropriate procedural protocols for military/civilian laboratory personnel, as developed for the program; to train laboratory personnel; and to oversee the initial operation. The first challenge was the rapid transition of a laboratory bench-top protocol capable of processing several samples per day into a high-throughput production process capable of processing up to 450 samples per day. On 9 December 2005, clinical samples began arriving from MTFs located throughout the MDW. These clinical samples were analyzed and the results reported to the Air Force Medical Service (AFMS) Composite Occupational Health and Operational Risk Tracking (COHORT) electronic surveillance system. NRL hosted more than 40 active duty AF and USN personnel and provided technical support for those personnel to implement and operate a central laboratory. The laboratory staff, jointly managed by medical specialists of the Air Force Institute for Operational Health (AFIOH) and NRL Code 6900 scientists, established laboratory practices consistent with those prescribed by the College of American Pathologists (CAP).

Using the Respiratory Pathogen Microarray (RPM), NRL analyzed 598 samples from individuals presenting with febrile respiratory illness and temperatures greater than 100.4°F. The results from the RPM random amplification protocol were reported to COHORT during SGD. Aliquots from 450 specimens were shipped to the Naval Health Research Center (NHRC) and analyzed using culture and Polymerase Chain Reaction (PCR) methods. During SGD, approximately 5,000 RPM assays were performed using patient samples from six MTFs and spiked normal nasal wash, demonstrating: (1) a capability to establish a full production, high throughput Clinical Laboratory Improvement Amendments (CLIA) quality analysis laboratory using high density microarrays in under 2 months, (2) a capability to integrate disparate data systems for collection of clinical data, tracking analysis and reporting the result to geographically dispersed principal investigators (PIs), (3) that a highly multiplexed, genomics-based molecular diagnostic assay could be implemented by active duty military personnel, (4) that hundreds of diverse genetic sequences for common respiratory pathogens causing illness in the general population could be obtained and unequivocally identified in approximately 18-19 hours, and (5) that the system was capable of sensitive detection and discrimination of a panel of biothreat (BT) agents, including *Bacillus anthracis*, a smallpox virus simulant, and pathogens of operational significance (respiratory pathogens such as influenza virus, adenovirus) and discriminated from nearly identical non-pathogenic organisms.

Numerous areas for improvement of the biochemistry and bioinformatics used in the project were identified. These included: (1) an obvious need for protocol simplification and automation of manual processing steps, (2) adjustment of assay sensitivity for pathogen titers varying by 6-8 orders of magnitude, and (3) algorithms and user interfaces that allow less technically initiated operators to interpret genetic sequence information. Despite many technical and logistical hurdles, the overall feasibility of resequencing microarray-based respiratory pathogen identification was clearly demonstrated. **This was the very first demonstration of a capability for rapidly (< 24 h) obtaining sequence data for a broad range of DNA and RNA pathogens in clinical specimens collected from**

the general population. The Silent Guardian project points to a clear pathway for the maturation and validation of this protocol and related technologies in the Epidemic Outbreak Surveillance (EOS) Advanced Concept Technology Demonstration (ACTD).

SILENT GUARDIAN AT NRL OCTOBER 2004-MARCH 2005

CHAPTER 1: PROJECT OVERVIEW

Introduction

Technologies for the Silent Guardian Demonstration (SGD) Project originated in the Epidemic Outbreak Surveillance (EOS) Program sponsored by the Air Force Surgeon General's Modernization Directorate (HQ USAF SGR).[1] The version 1 Respiratory Pathogen Microarray (RPM v.1) used in the project was a custom Affymetrix resequencing microarray designed by the Naval Research Laboratory (NRL). The output of the RPM v.1 chip was genetic sequence, read directly from "tiles" on the array corresponding to pathogen types. The sequence output from the chip was compared directly against exponentially growing sequence databases (public and private) to allow extremely high confidence pathogen identification and discrimination of closely related organisms.

The RPM v.1 was designed in April 2003 for proof-of-concept detection of common respiratory pathogens (esp. adenovirus, influenza A/B viruses, and *Streptococcus pyogenes*) encountered in the basic military trainee (BMT) population at Lackland AFB, San Antonio, TX. It was not originally intended for use in a more heterogeneous urban population exhibiting a broader range of respiratory pathogens. In addition to tiles specific for common respiratory pathogens, the RPM v.1 chips were designed with a few tiles specific for several CDC Category A biothreat (BT) pathogens: *Bacillus anthracis* (3 tiles), variola major virus (2 tiles), Ebola virus (1 tile), Lassa virus (1 tile), *Francisella tularensis* (2 tiles), and *Yersinia pestis* (2 tiles).

Prior to the SGD project, all pathogen tiles on the RPM v.1 except those for West Nile virus were determined to be correct by resequencing the corresponding purified genomic targets. In the summer of 2004, human adenovirus and influenza A virus, the model DNA and RNA pathogens, respectively, were detectable in nasal wash specimens with an overall agreement compared to culture of approximately ~85% (See Chapter 4, Influenza A detection for description of calculation) without reliance on specific PCR primers to amplify targets prior to chip hybridization. However, two separate protocols were used for detection of DNA and RNA pathogens, and there was no information collected yet on the clinical sensitivity and specificity of a fully multiplexed RPM assay for anything other than the model pathogens. The assay was not optimized for speed, as a few samples at a time were typically processed over a period of 24–36 hours.

In the late summer of 2004, the increasing probability of a Military District Washington (MDW) demonstration project using the RPM v.1 chip mandated the combination of the DNA and RNA sample processing protocols and reduction of the individual assay performance time to 14–16 hours in order to satisfy a requirement of performing 300 or more RPM assays per day during the Full Operational Capability (FOC) period. With verbal instructions to initiate the demonstration in October 2004, an entire laboratory infrastructure with supporting Laboratory Information Management System (LIMS), Quality Control / Quality Assurance (QA/QC) measures, biosafety, and contingency plans (in the event of a BT-positive assay) had to be designed and implemented in a period of six weeks. Written instructions were received November 2005 from the Assistant to the Secretary of Defense for Nuclear, Chemical and Biological Programs (Appendix A).

Results

Achievement of Full Operational Capability (FOC)

For representative clinical specimens, measured turnaround times (defined as the time from initiation of the analysis at NRL until the results were reported to Composite Occupational Health and Operational Risk Tracking or COHORT) for the period of 25 January to 7 February averaged 18.9 hours. As a demonstration of FOC, 300 samples were accepted for analysis on two separate days and all samples processed within a 24-hr period. Since this number of patient specimens was never obtained in a 24-hour period, the operational capacity was used to perform re-runs of patient specimens using different sample processing methods as well as experiments in which normal nasal wash samples were spiked with BT agents or near neighbors.

Common Respiratory Pathogen Detection in Clinical Specimens

Influenza A virus (Flu A) was the most commonly identified respiratory pathogen by all methods. Thirty-day culture methods revealed that 184 of the 450 (41% of total) samples sent for culture were positive for Flu A. A little less than half this number—79 (18% of total) were reported to COHORT as positive for Flu A when using the standard SGD protocol (random amplification without DNA subtraction followed by cutoff of BLAST results with expect values of $< 10^{-9}$). Therefore, the unused fractions of the processed RNA and DNA aliquots for 464 of the patient samples were amplified using multiplex PCR (UberMix) instead of random amplification. Also, the sequence calls made by the RPM in these and the previous experiments (using the standard algorithm) were re-evaluated using “relaxed” conditions in which targets were identified by associating the top sequence similarity ranking (uniquely high BLAST bit score with corresponding low expect value) with single pathogen genome database record(s) instead of satisfying an arbitrary BLAST cut-off expect value of $1.0e^{-9}$. For all pathogens, this required that the primer sequences be removed from the FASTA sequence output file (a standard genetic sequence format text file) prior to BLAST analysis. In all cases, the criterion for a positive by RPM analysis was that sequences from one or more of the pathogen tiles on the RPM would return a BLAST result that was uniquely identifiable with records for a pathogen species.

There were 351 “matched” sets of experiments. A “matched” set contained data for culture, random amplification, and multiplexed amplification as analyzed with “relaxed” conditions for each sample. Of the 351 specimens, 30-day culture detected Flu A in 147 (42%) of the specimens. The RPM v.1 detected Flu A in 73 (21%) with random amplification and 88 (25%) with multiplex (UberMix) to produce 114 unique positive results for 32% of the total “matched” sets and 76% of the number of 30-day culture positives in the “matched” set. Thus, two uncharacterized and compromised RPM amplification protocols were combined during the project to provide for approximately the same level of Flu A detection as predicted for random amplification without interference of DNA background on the RPM prior to SGD. In all cases, Flu A sequences were immediately observed to be consistent with H3N2 database records, and sequences were screened for H5 (avian influenza virus) prior to the release of Flu A-positive results.

For all samples analyzed by culture and PCR (450 patient samples), in addition to Flu A, standard microbiological analysis detected several cases of influenza B virus, adenovirus, Type II parainfluenza virus, *S. pyogenes*, *S. pneumoniae*, and *Haemophilus influenzae*. The RPM v.1 did not contain resequencing tiles for Type II Parainfluenza virus or *H. influenzae*. The 598 patient specimens analyzed with the standard SGD protocol (random amplification relaxed results) also returned positives for influenza B virus, coronavirus 229E; rhinovirus, and respiratory syncytial virus (RSV) B. The 464 patient aliquots analyzed with the UberMix protocol (multiplex PCR amplification relaxed results) returned positive results for Influenza B virus, Adenovirus, Coronavirus 229E, *Neisseria meningitidis*, *S. pyogenes*, *S. pneumoniae*, RSV A, RSV B, and erythromycin-resistance genes in addition to influenza

A virus. The antibiotic resistance genes had a high correlation with *S. pneumoniae* and *N. meningitidis* detection.

BT Agent Detection in Clinical Specimens

Of the 598 valid patient specimens analyzed at NRL, no BT agent sequences were observed using either the standard protocol (random amplification) or UberMix (multiplex PCR amplification). The application of these approaches to nasal washes spiked with BT agents (Chapter 5) showed that the UberMix achieved highly discriminatory detection of BT agents at sensitivities similar to those of optimized specific PCR reactions. Between 25 January and 7 February 2005, 11 blind spikes of *B. anthracis* started the SGD process at NRL as patient samples, and all 11 were detected using the RPM assay. Using the SOP, each of these RPM results triggered positive confirmatory PCR assays by the US Air Force (AF) Biological Augmentation Teams (FF BAT) at NRL.[2, 3]

Detection and Discrimination of BT Agents and Near Neighbors

The relatively small number of patient samples (Chapter 4) obtained in SGD afforded an excellent opportunity to perform the first systematic assessment of the ability of the RPM v.1 to: (a) detect pathogens across large concentration ranges and (b) withstand challenge from a battery of genetically similar organisms. In this initial assessment of the RPM v.1 for BT agent detection, we observed that the SGD standard protocol [4], which used random amplification without prior DNA subtraction (due to reagent failure – see Chapter 2), was found to have inadequate sensitivity for most BT agents of interest. To enhance assay sensitivity over that observed with random amplification, two multiplex PCR primer mixes specific to the following were also used (Appendix B): 1) BT agents, adenovirus, and influenza virus only (BTA mix), or 2) all targets on the RPM v.1 chip (UberMix). Due to time constraints, neither of these multiplex approaches was optimized during the demonstration project.

A total of more than 2100 spiked samples and blank controls were tested using random amplification, BTA mix, or UberMix. Using these specific primer mixes, we detected *B. anthracis* sequences from Sterne strain vegetative cells at the lowest concentration tested - 20 colony forming units (cfu) and in Ames spores at concentrations as low as 6×10^4 cfu per 150 μ L sample. *F. tularensis* specific sequences were detected in two strains of *F. tularensis* at concentrations as low as 10 cfu while in Schu4 strain they were detected at 10^3 cfu. *Y. pestis* sequences in samples from two sources were detected at 10-300 cfu, depending on the source. Sequences for Vaccinia, a simulant for variola major virus, were detected at the lowest concentration of vaccinia tested (10^5 plaque forming units (pfu))—but the true limit of detection will not be determined until additional lower concentrations are tested. Targets were identified by associating the top sequence similarity ranking (uniquely high BLAST bit score with corresponding low expect value) with single pathogen genome database records.

Of equal value, we observed experimental anomalies that will drive refinements to the design and processing of subsequent resequencing microarrays. For example, *cve* (*Y. pestis*) and *fopA* (*F. tularensis*) sequences allowed for extremely sensitive detection but were also identified unexpectedly in some samples (1/892 for random amplification, 55/1010 for BTA mix, and 6/284 for UberMix) that were not intentionally spiked with those targets. Similarly, a plasmid marker (pX01 or pX02) corresponding to a *B. anthracis* database record was unexpectedly found in 5/2186 samples. There was high agreement between these misidentified samples and the presence of a high concentration of the misidentified target in an adjacent tube ($P < 0.001$). When combined with the high complexity of the processing procedures, these statistics strongly support the hypothesis that many of these misidentifications occurred as a result of handling errors. Simplification, and preferably automation, of the sample processing procedures should greatly improve the reproducibility and reliability of future demonstrations. Furthermore, optimization of amplicon length and sequence regions tiled should not only improve sensitivity, but also discriminatory capabilities of future versions of the RPM chip. The results of this initial BT agent assessment are of

obvious benefit to the refinement of chip design, primers for random and multiplex amplification, and automation.

Common Respiratory Pathogen Detection in BT Spiked Experiments

In more than 2100 experiments conducted with BT agents and near neighbors spiked into healthy nasal wash, no influenza A/B was detected.

Conclusions and Recommendations

This early feasibility demonstration of the capability to provide real-time surveillance using a microarray in the Silent Guardian Project highlights the promise of the EOS concept as well as pinpoints the specific issues that require engineering solutions in order to achieve the ACTD (Accelerated Concept Technology Demonstration) objectives. A clear pathway exists for the optimization and validation of the resequencing microarray development in support of the EOS ACTD: (I) simplification and improvement of the RPM protocol used in SGD protocol for use with newly designed RPMs, (II) development of a fully automated sample preparation device that will allow a 6-hour RPM assay at a clinical point-of-care, and (III) transfer of developed technologies to validation partners and transition support as appropriate into Level 5 and Level 3 medical applications. The successful demonstrations of the technology during the ACTD will be in compliance with the submission of certified laboratory data to the U.S. Food and Drug Administration (FDA) and licensing of patented technology for transition to a commercial partner and eventual vendor to the U.S. Government. Moreover, the technology infrastructure put into place during SGD and the EOS ACTD will be supportive of future methods for acquisition and analysis of pathogen genomic sequence data.

CHAPTER 2: BACKGROUND

Existing Technology

The current “gold standard” assays for pathogen identification are culture (requiring days to weeks) and, more increasingly, polymerase chain reaction (PCR). In reality, it is difficult to achieve consensus on what constitutes “gold” standards for many pathogens, and therefore these complementary methods are henceforth referred to as “reference assays”. In PCR, DNA targets or cDNA (made from RNA targets) are amplified with specific primer pairs. For higher diagnostic capacity, multiplex PCR (amplifying more than one target in one reaction) can be used to identify multiple pathogens or multiple targets per microorganism. This assay format has been shown to be a valuable method for pathogen detection. Recently, real-time PCR (using fluorescent signals for real-time monitoring of PCR products) has shown great potential for pathogen detection.[5] Up to 4 different amplicons can be detected simultaneously in one solution by using fluorescent labels of different colors that are detected with respective optical detectors.

The major advantages of specific PCR are: (1) high sensitivity and specificity, albeit complicated by a number of potential confounders, and (2) speed. The process requires only an extraction of the total nucleic acids (all RNA and DNA) from a sample, followed by application of specific PCR primers and thermal cycling. Two important limitations are: (1) dependence on known, unique sequences for detection, and (2) scalability—each PCR reaction requires a set of specific primers and a means for detection, especially a dedicated optical detector for each reaction (fluorescent color) in real-time multiplex PCR. Thus multiplexing large numbers of PCR reactions is generally not feasible.

DNA Microarrays

DNA microarrays are comprised of a high number of DNA probes immobilized on a solid support in a geometrically defined manner. Microarrays use immobilized DNA “probes” to capture “target” pathogen nucleic acids from a solution. Virtually all microarrays intended for the present purpose use DNA probe molecules ranging in length from 20 to 70 nucleotide bases (20-mers to 70-mers, respectively). Short (20- to 25-mer) probes are intolerant of a single base pair mismatch with the complementary target sequence (i.e., little or no hybridization occurs) which allows detection of single-point mutation. In comparison, a 70-mer can tolerate upwards of 10 base pair mismatches along the length of the target sequence, depending on hybridization conditions. A greater target capture capability is afforded by the 70-mers but greater specificity is achieved with the 25-mer. The NRL designed Affymetrix resequencing array used a high density of 25-mers to allow target sequences to be read from the chip.

With the currently available technologies, microarrays still require that the targets be amplified prior to attempting to hybridize them to the immobilized probes. When specific PCR is used to amplify the targets, a very high signal-to-noise (S/N) optical signal is obtained from each respective hybridization site on the array. In combination with a resequencing array, the sensitivity and specificity can be *greater* than with a specific PCR assay alone. Thus, a resequencing array provides a *digital*, not an analog readout. It is dependent only on a S/N ratio, not on an absolute optical intensity (above a baseline minimum). This means that amplicons that are in too few in quantity to be detected can still satisfy the resequencing “logic function”, even at barely detectable absolute signal intensities, provided that background “noise” on the array is low enough. This is usually the case when only a single specific target is amplified. Moreover, the increased sensitivity relative to specific PCR alone does not come with an increased risk of false positives, because the sequence output of the array can be compared against all known nucleotide sequences using available algorithms (e.g., BLAST), allowing a numerical evaluation of the chip sequence output against near-neighbors.

Multiplex PCR can also be used to amplify sufficient quantities of targets for microarray analysis. NRL has an internal program that uses multiplexed PCR to amplify 50-60 waterborne pathogen target sequences simultaneously for subsequent analysis by 70-mer DNA microarrays. For microarray applications, multiplexed PCR has considerable advantages. A sizeable number of targets can be presented to the array simultaneously and very high sensitivity is obtained in minimal time. Multiplexed PCR is under development as part of environmental monitoring initiatives sponsored by the Joint Program Executive Office for Chemical and Biological Defense (JPEO-CBD).

The disadvantage of multiplex PCR is that the amplification reactants can interfere with one another. Thus, there are a large number of potential interferences, making systematic validation of a highly multiplexed PCR-microarray nontrivial. In addition, in terms of cross-contamination, the number of possibilities is multiplied. The multiplex reaction used for the 70-mer array at NRL has been systematically optimized over a period of 30 months. Once the assay is fixed, it is not highly amenable to changes (i.e., addition of new primer pairs) because any potential complications caused by cross-reactions in the PCR, possibly affecting sensitivity and specificity for all targets, must be examined for each added primer pair.

Finally, random amplification methods can be used to prepare target sequences for microarray analysis. Unlike specific PCR, *all* DNA molecules in the complex mixture are usually amplified, each to a much lower extent than if any one target would be amplified by specific PCR. Thus, we expect to see at least 3 orders of magnitude fewer amplicons for each pathogen target than would be expected from specific PCR. However, random amplification is particularly useful for rapidly mutating or unknown pathogens.

When the products from a random amplification are applied to a microarray, a large amount of “noise” can appear on the array. This is because the large numbers of different amplicons hybridize in an unpredictable manner with the probes on the array. In the case of the Affymetrix resequencing arrays, this background noise does not result in spurious base calls across contiguous stretches of sequence. However, the noise can interfere with the ability of small amounts of pathogen target amplicons to produce a hybridization pattern that is visible above the background.

Major factors that affect the sensitivity and time to perform a microarray assay include:

1. The starting concentration and purity of the pathogen target.
2. The method used to amplify the target prior to contact with the array.
3. The abundance of the amplified target molecules relative to other molecules that can bind to the array.
4. The time allowed for target capture by the DNA probes on the array.
5. The labeling method for amplified targets and the detection method on the array surface.
6. Level of discrimination provided by the software for image analysis and comparison to the gene database.

EOS Program

The Summer 2000 Defense Sciences Board (DSB) on Homeland Defense first recommended that the Defense Department capitalize on the increasing availability of genomic information and technology by widely deploying (DoD TRICARE) a massively-parallel assay (i.e. “Zebra Chip” microarray) as a means to simultaneously detect a large number of pathogens, including BT agents, in point-of-care clinical settings. The DSB argued that the Zebra Chip could provide the earliest detection of BT agents among persons presenting with “flu-like” symptoms, a common feature of both common pathogens and BT agents during early infection stages.

In the spring of 2001, the Deputy Assistant to the Secretary of Defense, Acquisition, Technology and Logistics DATSD (ATL) commissioned the Health Surveillance Biodefense Systems (HSBS) panel, which convened from April–December of 2001 in order to assess the overall feasibility of the DSB proposals. The HSBS final report recommended the trial assessment of DNA microarray technology for pathogen genotyping and host genome response in an active duty military cohort (e.g., a basic military trainee (BMT) population). A significant opinion offered by physicians on the HSBS medical sub-panel was that the BT surveillance capability would need to be embedded in a reimbursable (FDA-approved) assay for routine infectious diagnostics in order to achieve widespread use.

In the mid 1990's, NRL initiated a basic research (6.1) program to develop microarrays using the surface modification techniques pioneered at NRL's Center for Bio/Molecular Science and Engineering (CBMSE) in the late 1980's.[6] In 2001, NRL started a base 6.2 program to exploit the results with this 6.1 research. Progress in this NRL-funded 6.2 program in CBMSE coupled with the initiatives discussed above led the AF/SG and NRL to team up for the development and assessment of microarray technologies.

In 2002, the Epidemic Outbreak Surveillance (EOS) program was initiated by HQ USAF SGR in order to perform the trial assessment of the DNA microarray technologies recommended by the HSBS panel. This proof-of-concept effort was initially funded by the Defense Threat Reduction Agency (DTRA). In keeping with the theme of an eventual FDA-approved device, a Version 1.0 Respiratory Pathogen Microarray (RPM v.1) was designed for detection of common upper respiratory pathogens (esp. adenovirus, influenza A/B viruses, and *S. pyogenes*) encountered in the BMT population at Lackland AFB in San Antonio, TX.

RPM v.1 Microarray Development

From June 2003 until December 2003, the basic functionality of the EOS RPM v.1 microarray was determined, primarily using pathogen-specific PCR or reverse-transcriptase PCR (RT/PCR) in complex mixtures or total amplification of purified pathogen DNA using the GenomiPhi™ kit (Amersham Biosciences, Piscataway, NJ). This confirmed that pathogen probe sequences on the array were correct, and suggested the development of software that would scan the base calls made by the arrays for sequences suitable for automated BLAST searching against existing public databases (e.g., GenBank).

In February 2004, the EOS Operations Board decided that a major, non-incremental goal of the program relative to diagnostics was to maximize the development of non-biased (i.e., “random”) amplification strategies for as many pathogen targets as possible from a nasal wash clinical specimen. Success could enable the use of a single set of common reagents that could be combined with the flexible content of resequencing microarrays. Only an array-specific set of references would be needed for the analysis software to allow “plug-and-play” of any type of resequencing microarray. This would address a significant logistical problem faced by the DoD relating to the development, acquisition, validation, supply and distribution of “critical reagents” (in this case, specific PCR primers) necessary to support detection of a constantly growing number of pathogen threats.

The most significant technical issue facing a non-biased amplification strategy (total genome amplification, random PCR, etc.) from a nasal wash specimen is the abundance of the pathogen genome targets relative to the background nucleic acids (RNA and DNA from both the human host and the commensal microorganisms) of the nasopharyngeal tract. During infection, the abundance of the pathogen increases relative to the commensal microorganisms, but there is an accompanying increase in the background level of human DNA and RNA from lysed epithelial cells.

The RPM v.1 was designed in April 2003 for proof-of-concept detection of common respiratory pathogens (esp. adenovirus, influenza A/B viruses, and *S. pyogenes*) encountered in the BMT population at Lackland AFB in San Antonio, TX (Lin et al., 2006). It was not originally intended for use in a more heterogeneous urban population exhibiting a broader range of respiratory pathogens. In addition to tiles specific for common respiratory pathogens, the RPM v.1 chips were designed with tiles specific for several Centers for Disease Control (CDC) Category A biothreat pathogens: *B. anthracis* (3 tiles), variola major virus (2 tiles), Ebola virus (1 tile), Lassa virus (1 tile), *F. tularensis* (2 tiles), and *Y. pestis* (2 tiles).

During the summer and fall of 2003, the EOS team began to using nasal wash specimens from Lackland AFB that were positive for adenovirus type 4 and determined that the lower limit of detection for adenovirus using total genome amplification was approximately 10^3 genome copies per microliter of nasal wash. This is considerably less sensitive than specific PCR would be for the same pathogen. However, most specimens were positive; indicating that high amounts of adenovirus can be present in nasal wash specimens.

From June 2003 until September 2004, separate processing pathways existed for RPM detection of RNA and DNA pathogens. For development of the RNA processing pathway, spiked and clinical (nasal wash and throat swab) influenza A positive specimens were processed using different methods (random, universal, and multiplexed PCR). It was shown in early 2004 that in order to achieve acceptable levels of sensitivity with random amplification approaches, human messenger and ribosomal RNA had to be subtracted using the MICROBEnrich™ kit (Ambion, Inc., Austin, TX). A preferred random amplification strategy was adapted from a previously described method (Wang et al., 2002). This was found to allow positive H3N2 and H1N1 Flu A detection in 16/19 different culture-positive Flu A nasal wash and throat swab specimens at concentrations as low as 0.45-3.75 pfu/150 μ L sample aliquot, detecting Flu A in 6/8 specimens in the low concentration range following subtraction of human RNA.

Prior to the SGD demonstration, all pathogen tiles on the RPM v.1 were determined to be correct by resequencing of the corresponding purified genomic targets except those for West Nile Virus. In the summer of 2004, Human adenovirus and influenza A virus, the model DNA and RNA pathogens, respectively, were detectable in nasal wash specimens with an overall agreement compared to culture of ~85% (See Chapter 4, influenza A detection for description of calculation) without reliance on specific PCR primers to amplify targets prior to chip hybridization. However, two separate protocols were used for detection of DNA and RNA pathogens, and there was no information collected yet on the sensitivity and specificity of a fully multiplexed RPM assay for pathogens other than the model pathogens. The assay was not optimized for speed, as a few samples at a time were typically processed over a period of 24–36 hours.

CHAPTER 3: SYSTEM DEVELOPMENT

Background

In November 2004, the Assistant to the Secretary of Defense for Nuclear, Chemical and Biological Programs directed the NRL to accelerate the development of an emergent resequencing microarray system for deployment in the MDW (Appendix A). The demonstration required rapid transition of the laboratory bench-top protocol for individual clinical samples into an “assembly-line process suitable for implementation.” Demonstration implementation involved rapid acquisition of equipment and reagents, recruitment of skilled management and technical personnel, facility location and set up, protocol development, and training of all staff. After these tasks were completed, patient specimens provided by six military treatment facilities (MTF) located throughout the MDW would be transferred to NRL for analysis and the results reported to the Air Force Medical Service (AFMS) COHORT electronic surveillance system. The Air Force Surgeon General’s Modernization Directorate would manage the overall demonstration, including institutional review board applications, collection of all specimens and direct reporting to the DoD chain of command and the Laboratory Response Network. NRL would host more than 40 active duty AF and USN personnel and provide technical support for those personnel to implement and operate a central laboratory capable of processing more than 300 clinical specimens per day using the RPM v.1 system. The laboratory team, jointly managed by personnel from the Air Force Institute for Operational Health (AFIOH) and the NRL Center for Bio/Molecular Science & Engineering (Code 6900) would establish laboratory practices consistent with those prescribed by the College of American Pathologists (CAP). Meeting these objectives imposed major new requirements for biochemical processing of clinical samples, bioinformatics, and supporting information technology (IT) that would need to be satisfied in periods ranging from several weeks to two months.

Biochemistry

In September 2004, the RPM v.1 protocol did not have integrated RNA and DNA target processing pathways for random amplification from clinical specimens. Furthermore, the DNA process, using total genome amplification, was unacceptably long (16 hours just for the GenomiPhi™ amplification). A concentrated effort was made in the following weeks to shorten this time by increasing the reactants and varying the temperature, but it could only be reduced to 8 hours. This was unacceptable from the standpoint of the projected production schedule.

In October 2004, as the likelihood of the SGD increased, EOS technicians from San Antonio and San Diego came to NRL for two weeks in order to arrive at a consensus recommendation for the demonstration protocol. One strategy tested was the use of multiplex PCR for all of the pathogen targets assayed by the RPM v.1. Each of the RPM probe tiles included specific PCR primer sites within or adjacent to them. This approach had the potential to be highly sensitive and would remove many technical steps (background subtraction) and hours from the length of the protocol.

The multiplex PCR amplification approach, however, had four important drawbacks. First, the multiplex was not optimized. That is, the primers were largely taken from the open literature and were not designed to function together under a single condition. Thus, some of the primers might not work at all, and there was significant potential for primers to interfere with each other via the formation of primer-dimers. Second, there was a much greater chance for cross-contamination between samples or from trace contamination with nucleic acids from a biological threat agent (e.g. *B. anthracis*, *Y. pestis*) from a different molecular lab at NRL. Because the production line would be new and untested, it was considered especially vulnerable to cross-contamination. It was also possible that an unoptimized primer set could themselves hybridize to the tiles on the chip leading to reduced signal-to-noise. Finally, there was no consensus on how to report results with little or no positive predictive value. In pilot experiments

during October 2004, *N. meningitides* and *S. pneumoniae*, which colonize the nasopharynx, were detected in a number of “healthy” nasal wash specimens.

Ultimately, a protocol was recommended for SGD that was different the previous version for DNA amplification (Fig. 1). This protocol was adapted from the lab of Joseph DeRisi at University of California San Francisco (http://derisilab.ucsf.edu/pdfs/Round_A_B_C.pdf) and randomly incorporated a single, fixed-sequence PCR primer binding site into all pathogen DNA or cDNA molecules, allowing them all to be amplified in a subsequent step using a conventional PCR thermal cycling protocol. Thus the initial basic SGD protocol used “random amplification” and did not include any specific PCR primers in this step.

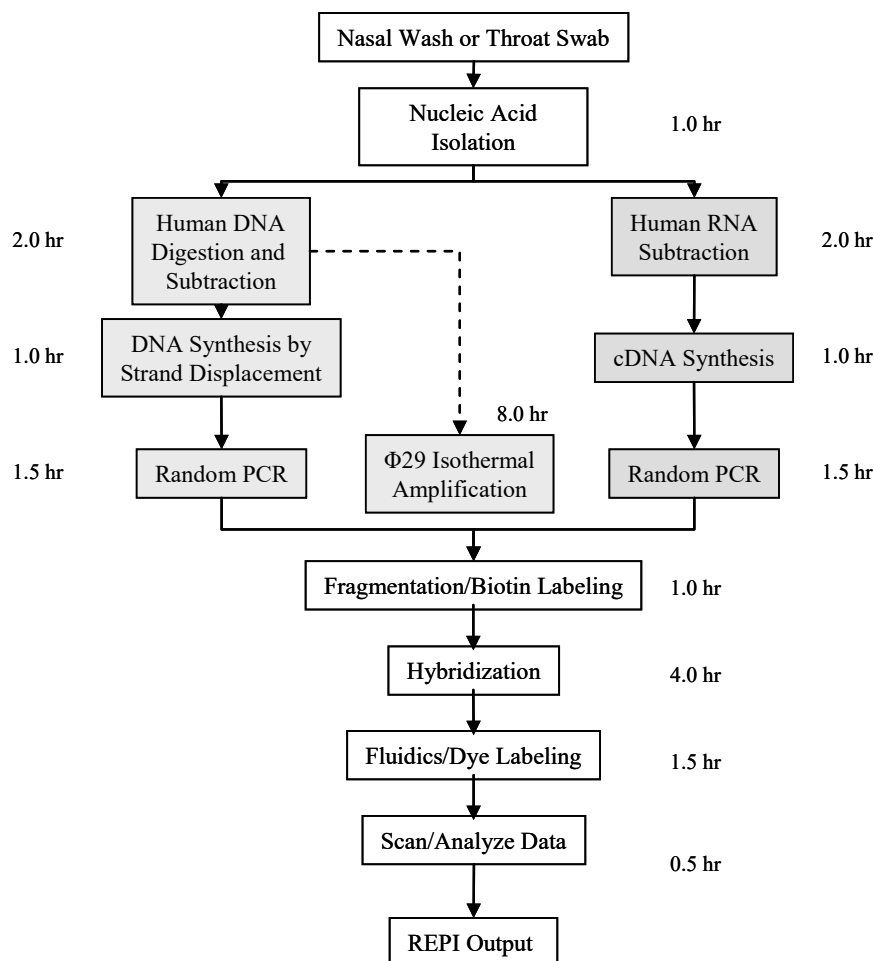


Fig. 1—Recommended nucleic acid processing protocol in SGD

Bioinformatics and Information Technology

SGD was characterized by a need for rapid design, acquisition, validation and deployment of complex, interacting data/hardware systems in an 8-10 week timeframe. Prior to the SGD project, the data analysis procedure for one sample utilized several unconnected pieces of software and hardware and significant expertise at each stage of the procedure. At the close of the project, expert intervention was required only to confirm positive calls in a patient sample. To achieve this, a number of software systems were modified and others designed and created. Hardware to support the target operating conditions was

reagent batch numbers, exact equipment used, user notes, etc.) not originally captured in LIMS were captured in this database. These data were held in a Microsoft MSDE database, which is a compacted version of Microsoft SQL Server 2000®. Ultimately the common patient identifier number linked the information contained in the MSDE and LIMS databases. This number was used as a key to combine data from the disparate databases using standard database manipulation tools.

GCOS/GDAS (GeneChip® Operating System/GeneChip® Data Analysis System) is the system software for analyzing the Affymetrix GeneChip® System. GCOS maintained a database of raw GeneChip® data, experimental protocols for preparing the GeneChip® for a nucleic acid sample, and controlled the Affymetrix hardware, the scanners and fluidic stations. The GCOS program examined the raw GeneChip® data file (.DAT) and performed grid alignment on the chips to generate the intensity in each and every cell of each tile (.CEL file). GDAS used an implementation of the ABACUS algorithm (Cutler, et al. 2001) to extract DNA sequences for the various “tiles” on the GeneChip® from the intensity file (.CEL) generated by GCOS. In addition, these programs had low level “hooks” to allow programmatic control of their operation as was done in the project (see Python, below). A Microsoft SQL 2000 database housed the Affymetrix data.

REPI (Respiratory Pathogens Identifier) was a Java-based application written at NRL that filtered gene sequences generated by the Affymetrix GDAS program for efficient searching on genetic information databases. REPI prescreened genetic data from the GDAS analysis, bypassing those “tiles” where there was insufficient length of continuous base calls to give a useful search. It then ran BLAST for the sequences that met its internal criteria and collected the resulting output. Besides making REPI run in an automated fashion, it was also modified to format the BLAST output into a more readable form as well as distilling out the more reliable species and strain level information in the results. This reliability was determined by the BLAST expect value, a measure of the probability that database record(s) returned by the algorithm occurs by chance. The lower the E-value, the more unlikely it was that a search results was a false positive and so the more easily it was incorporated into an automated program. An arbitrary cutoff of $1.0e^{-9}$ was arbitrarily selected for SGD.

BLAST is a public domain gene database-searching program. The program will find the gene sequences in a database that are the closest match to a submitted sequence. The database used in this study was taken from NCBI (National Center for Biotechnology Information) public database with the addition of sequences not yet placed in the public database but available to researchers at NRL.

Python is a public domain, high level, object-oriented programming language used to code “glue” software written at NRL. Python is an interpreted language, allowing for a rapid development cycle. Libraries exist for interaction with the Windows operating system and for inter-process communication using the “COM” interface. It is these latter capabilities that were utilized in SGD.

ACCESS, the Microsoft database application is used to provide easy access to the underlying Oracle database for “data mining”.

COHORT is a database application, maintained in the AF/Surgeon General’s office to provide SGD with a patient identification number, the clinic where their sample was generated and the time/date of sample acquisition. LIMS provided to COHORT the final results of the SGD analysis for each of these patient samples.

Operating Principles

Initially, a sample entered the system when logged in at a field location by collection staff at the clinics. This created a patient record in COHORT. COHORT published a random patient identification

number (PIN) which identified only the clinic to the LIMS system via a table at the SGD site. The LIMS system did not have the ability to access the patient information in COHORT, preserving the patient's privacy. This table also recorded the collection time and date for the sample. The LIMS system, at regular intervals, examined this table and upon receipt of a new patient record, created a set of LIMS identifiers for the sample as a whole, for the primary process sample and its DNA and RNA sub-fractions, as well as a number of other sub-samples for archiving and for submission for culture. At this time, LIMS also reserved storage locations for all these fractions.

Upon physical receipt of the sample to Lab A at NRL, a technician interacted with LIMS through the local client workstation and recorded sample volume information as well as the actual time of arrival at the SGD site. If a sample arrived prior to the COHORT notification to LIMS, the operator entered all the relevant data. In the case of the process fraction and its two nucleic acid fractions, the locations reflected the sample progress through the physical processes in Labs A-D. The workstations in Lab A and all other laboratories were equipped with a bar code reader and bar code printer. Upon receipt of the samples, bar code labels were printed for the process fractions as well as the fractions designated for storage.

After nucleic acid extraction in Lab A, the operator logged the sample out of the lab via the workstation and bar-code reader. Upon sample receipt in Lab B, a technician in Lab B logged the sample DNA and RNA fractions into the lab via the Lab B workstation. Multiple samples were assembled into a batch via LIMS. The batches were provided with an aggregate label for handling and transfer to Lab C1. The composition of the batch was maintained within LIMS. When transferred out of Lab B to Lab C1, the nucleic acid batches were subjected to PCR for amplification of the genetic material.

Following amplification, a Lab C operator moved the samples from Lab C1 to Lab C and the batch was logged in via the LIMS workstation's bar-code reader. As the separate RNA and DNA fractions of each sample were recombined in Lab C, the LIMS system created new barcode labels for the individual samples within the batch. After recombination, the amplified nucleic acids were applied to a RPM v.1 chip for hybridization. At this point, the operator used the LIMS station in Lab C to associate ("marry") the nucleic acid mixture for each sample with a specific chip, reading both the samples and the chips barcodes into LIMS.

In the next stage, LIMS created an entry in an Oracle database view, indicating that a sample would be processed to create genetic sequence data. Residing on the GCOS server was two Python services (resident programs running in the background). The first, **concordance**, read from the above-mentioned view every 60 seconds and when a new "married" sample was created, the program communicated with GCOS to create a GCOS experiment and sample. Then the operator processed the sample through its final stages, washing and staining the hybridized chip on the Affymetrix® fluidics stations and then transferring the chip to the scanner autoloaders. Both workstations used the chip barcode to identify the sample to GCOS with minimal operator intervention. At the same time as it created a GCOS experiment, **concordance** created a table of active samples. A second script, **movethrough**, monitored the table and when new entries occurred, it monitored the creation of a completed image from the scanner for each sample. When this process was complete, **movethrough** instructed GDAS to perform a genetic sequence determination. The results were converted to a standard genetic sequence format text file (FASTA format) which **movethrough** used to trigger REPI to do the genetic database searches. These results were compiled by **movethrough** and placed in appropriate shared file folders on the GCOS server. Finally, LIMS monitored this shared folder on the GCOS server, and as new files were encountered, data were entered into the LIMS system. At this stage, the sample was ready for authorization and release. An operator was notified via beeper that positive samples were pending release in the LIMS workstation. The operator would then enter the system via password to reject or release the data.

The first step was to track the sample and open the data file. Figure 3 depicts the sample data inventory by completion date. A green flask beside the LIMS number implied that there were no pathogens present and that the two control regions, based on internal plant gene control sequences, were present. A red flask could be due to either the presence of a positive pathogen or the absence of one or both of the control sequences. Details were obtained by opening the file in a separate window.

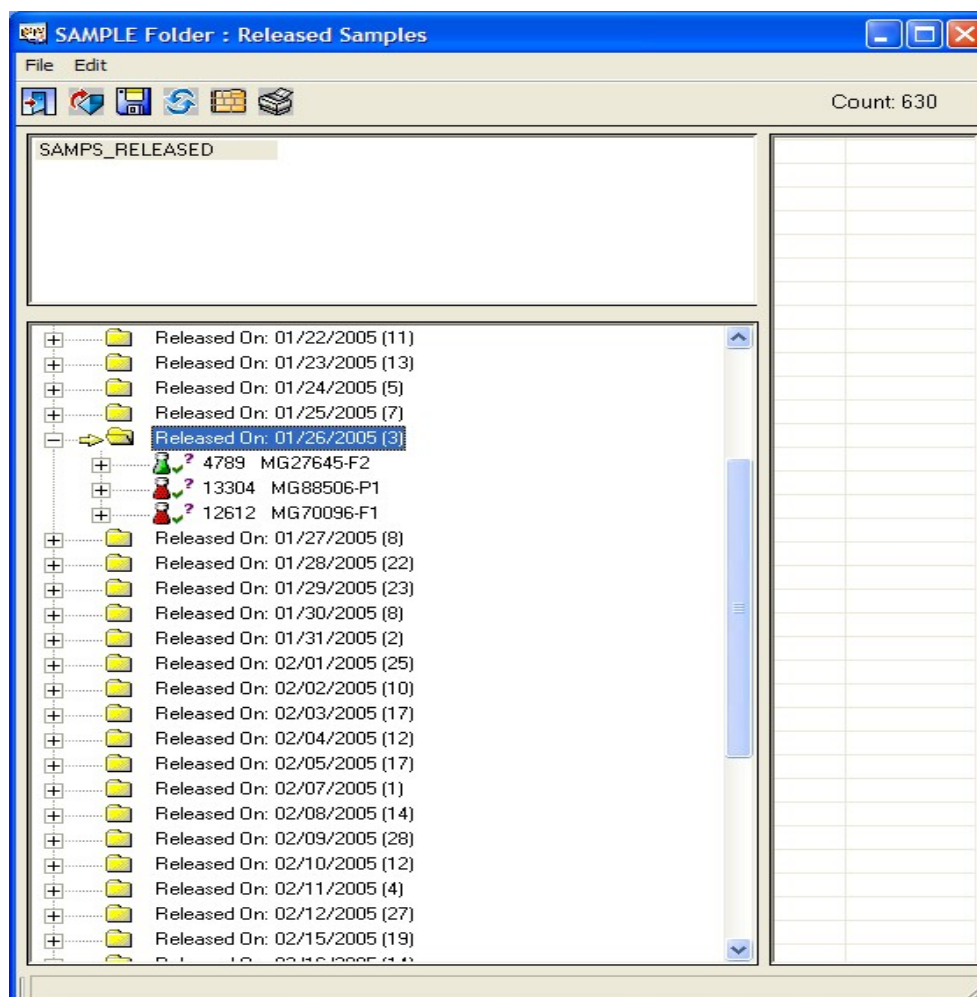


Fig. 3— LIMS sample 13304 (PIN MG88506) shows a red flask versus the green flask for sample 4789 (PIN MG27645). Both were released on 1/26/2005.

In the individual files, three “layers” of pathogen information were presented to the operator. The first layer of output was “species-level” only: (e.g. influenza A, influenza B, adenovirus, *S. pyogenes*, *B. anthracis*, *F. tularensis*) (Fig. 4). This was the same level of information as reported to COHORT. In the SGD protocol, the simple rule for a positive at this (“species”) level was that if any one of the multiple tile regions for a species produced a positive (i.e., produced a subsequence having an expect value of 10^{-9} or less), then the result was POSITIVE at the species level. This allowed a fully autonomous “first layer” of detail to be established. Time did not allow for the display of further information within this layer. In practice, the species identification based on the tile identity was usually, but not always correct. In one case, for example, a Flu B pathogen sequence was clearly identified using a sequence read on a Flu A tile.

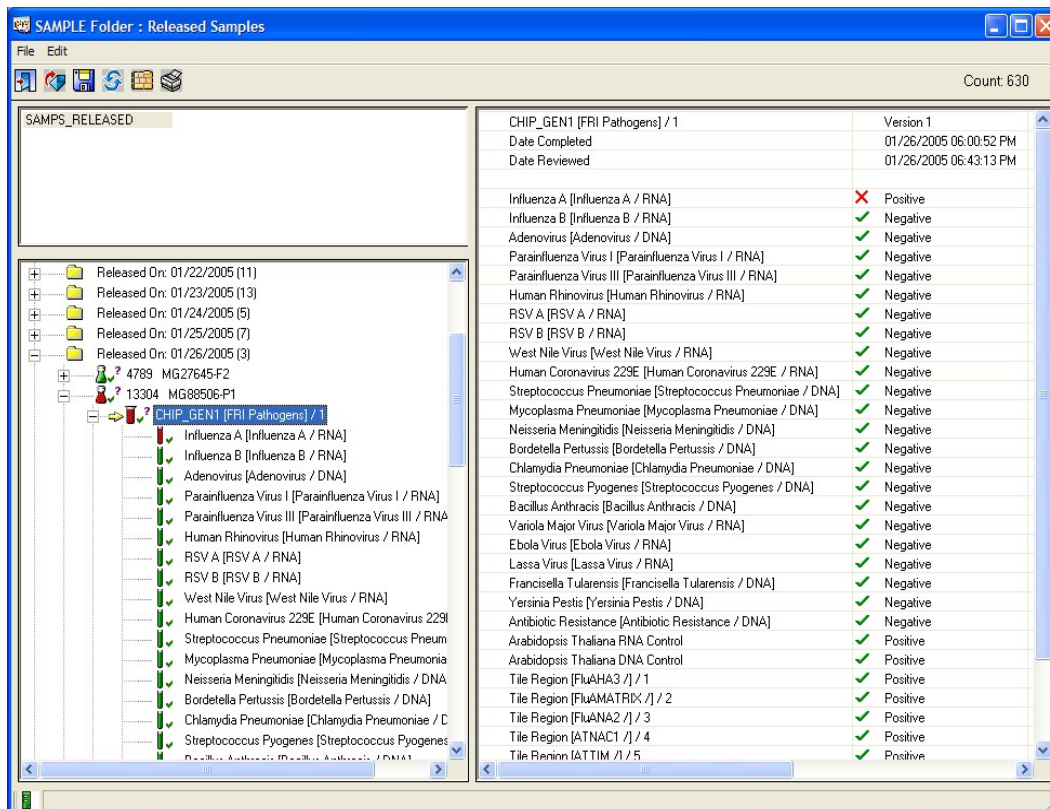


Fig. 4— LIMS results drill-down for sample in Fig. 3. Note red square for Influenza A pathogen.

The second layer of output would be the "serotype/strain level" and could be viewed using a prescribed protocol by authorized individuals. This layer displayed the actual tile regions of the array and whether a uniquely high BLAST score could distinguish between different database records or if multiple database records were returned with the same BLAST bit score.

The third layer available was the raw REPI output (Fig. 5), which allowed a knowledgeable operator to visually inspect for database records of interest. Finally, a Lab D LIMS operator could simultaneously display the raw chip data and base sequence calls made by GDAS (Fig. 6).

internal controls. In practice, since Room D operators used this notification to come to Room D to perform the release inquiry tasks, pathogen-positive samples tended to be released faster than pathogen-negative samples.

Using both LIMS and GCOS/GDAS, the validity of the LIMS data was confirmed. Once authorized and released, the LIMS software published the final results to an Oracle database view which was monitored via COHORT. When new data appeared, COHORT retrieved and published them electronically to the relevant Principle Investigators (PIs) at the MTFs.

At each transition point mentioned above, the time/date was recorded to allow for evaluation of sample turnaround time. Any sample movement into or out of storage was also traced in the LIMS system. All data are stored in relational databases and/or human readable text files for later analysis.

Project Assembly

Staffing

Once officially tasked with performance of the demonstration, NRL senior management worked with AF to develop an appropriate staffing plan. The project staff was composed of senior management (AF officers, AF staffing specialists and NRL management), the on-site production managers, and the laboratory performers. Navy and AF staff were integrated at all operational levels, and NRL staff was available during all shifts of operation (Fig. 7).

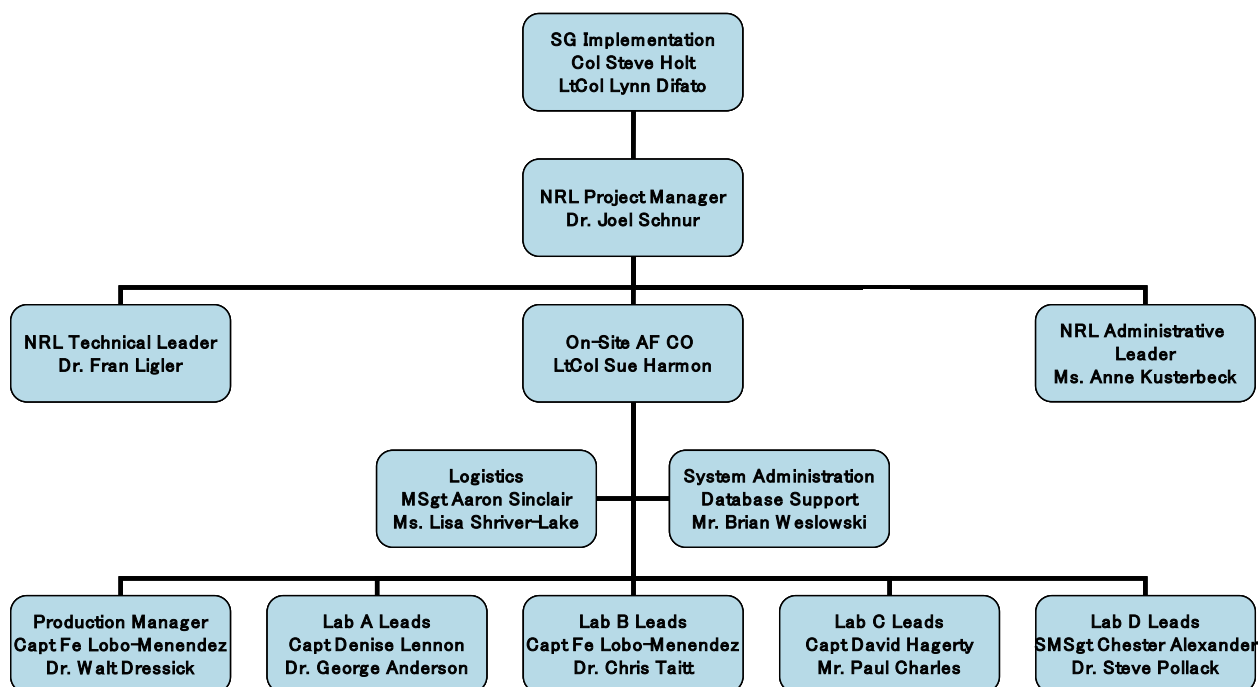


Fig. 7—Silent Guardian Management

Analysis of Protocol and Identification of Personnel Requirements

Identification of appropriate knowledge, skills and ability required for the SGD was first done by the EOS team and NRL, in conjunction with input from the AF SGR Management Team and staffing specialists. The key skills required for SGD were a basic knowledge of molecular biology and solid

biochemical laboratory expertise, with an emphasis on the ability to adhere to detailed protocols, pipette small volumes accurately, and follow good laboratory practices, and clinical laboratory expertise. Three groups of people met these criteria: 1) Experienced scientists at NRL and USN reservists with extensive laboratory research backgrounds who were familiar with molecular biology, 2) AF medical officers, civil service laboratorians, and enlisted technicians from AFIOH and select AFMS MTFs, and 3) experienced clinical laboratorians, applied research scientists, and preventive medicine and public health specialists from AFIOH and NHRC. The third group primarily provided reach-back, specimen referral, and operational and assessment subject matter expertise support regarding laboratory operations and sample analyses.

Prior to the start of SGD, specimen analysis consisted of a single technician running 1–20 samples over 2 days from start to finish. In order to process the projected 300 specimens/day (with a surge capacity to 450 samples/day) during FOC, an assembly line type of operation was established at NRL. Also, based on the requirements for 24/7 operations, 3 complete shifts were needed, with approximately 9–12 people per shift. The initial staffing plans assumed that non-EOS NRL personnel would comprise 1/3 of the total staff, while the remaining personnel would come from an outside contractor and/or existing AF laboratory staff. As plans progressed in the pre-FOC period, however, it became clear that there were significant difficulties involved in obtaining outside contractors. Consequently, additional AF personnel were identified to fill positions on the remaining two laboratory teams, as well as to provide staffing for logistics and handling of military personnel. The laboratory workers needed for only operations during the peak weeks of January were recruited from the NRL Chemistry Division and Naval reservists available to work for that month. In all, a total of 52 people participated in the SGD at the NRL site, with all the associated in-processing, transport and training (staff list, Appendix D).

Protocol Training

Training of all laboratory personnel involved several steps. First, based on their experience and expertise with the process, the NRL EOS Team trained the first set of NRL staff during a 2-week period in mid-November 2004. Both laboratory personnel and the management team were given detailed instructions on every step of the protocol, as well as hands-on training. Next, the AF Team Alpha, which arrived at NRL 29 November 2004, was given an overview of all protocols followed by 4 days of lab-specific training for each step. Finally, Team Alpha and NRL staff trained the second AF team, Team Bravo, when they arrived 6 December 2004. In general, the ratio of trainers to trainees was 1:2 for the NRL group and 1:4 for the NRL/AF sessions.

Procurement of Equipment and Reagents

The first increment of funds was accepted by NRL 21 October 2004. Those funds were used to buy reagents to process 600 chips and part of the equipment needed for the project. The goal was for the laboratory to be up and running within 2 weeks of receiving initial funds, with all reagents and equipment to be received by the first week in January. During the initial procurements, it became clear that a few items had long lead times. These included the Revco -80°C freezers, PCR reaction chambers, and the Tetrad thermal cycler. NRL's contracting and small purchase branches worked with the SGD to place orders rapidly and obtain rapid, if not overnight delivery of all purchases. Normal procedures for receipt were modified to have all procurements delivered directly to the building if possible; otherwise NRL's receiving department delivered them with 2 hours. Since the freezers and refrigerators needed for reagent storage were ordered concurrently with the reagents, storage of the reagents became difficult as orders came in faster than the freezer/refrigerators. This was not an issue for long. All labs were minimally operational within 3 weeks, with one complete processing line for the chips in place.

The second funding increment, including funds to purchase reagents for 5000 chips and most of the remaining equipment, was received the beginning of Nov 2004. At this time, the remaining seven Revco -80°C freezers were ordered. During this time, many reagent companies were being depleted of their stocks due to the large quantities of reagents being ordered. The companies were informed that orders for similar quantities would be placed within 2–3 weeks. For a few companies, this was a problem since the holiday season was approaching. NRL worked directly with the managers at these companies to work out compromises. In most cases, the equipment was ordered directly from the manufacturer, as they had the largest inventory. The small laboratory supplies were ordered from general supply companies. A significant amount of time was devoted to searching supplier web sites for availability of needed supplies. In some cases, identical items were purchased from different suppliers to fulfill the quantities needed for this project.

A third funding increment was received mid-November and included funds for the remaining reagents and 5000 chips. Several ordering problems occurred during this time due to companies having limited or no staff during Thanksgiving and a few orders were lost. NRL staff verified orders and delivery dates during this time frame and worked out issues arising with companies closing down for several weeks in December. At this time it became clear that a large amount of storage space was required to house all the reagents and chips; an inventory room for all non-perishable supplies and equipment was located and used for the SGD duration.

Early logistics were handled by NRL staff, while work during FOC was done using a combination of AF staff familiar with logistics, an administrative support contractor, and NRL staff familiar with NRL procurement.

Facilities

The highest priority for SGD laboratory location and set-up was to provide adequate separation of microarray steps and to minimize the potential for cross-contamination during the processing of patient samples. Consideration was also given to power requirements, limiting access in the first step due to unknown pathogenicity of patient samples, and physical space requirements for instruments and personnel during FOC. As seen in Fig. 8, three primary laboratories (Labs A, B, and C) were selected in two separate, but adjacent buildings in such a way that all labs had independent ventilation. An additional space, Lab D, was used for data storage and analysis. This configuration ensured, in addition to independent ventilation, a unilateral workflow, also critical to minimizing workplace cross-contamination. Finally, office spaces and computer workstations were provided for personnel.

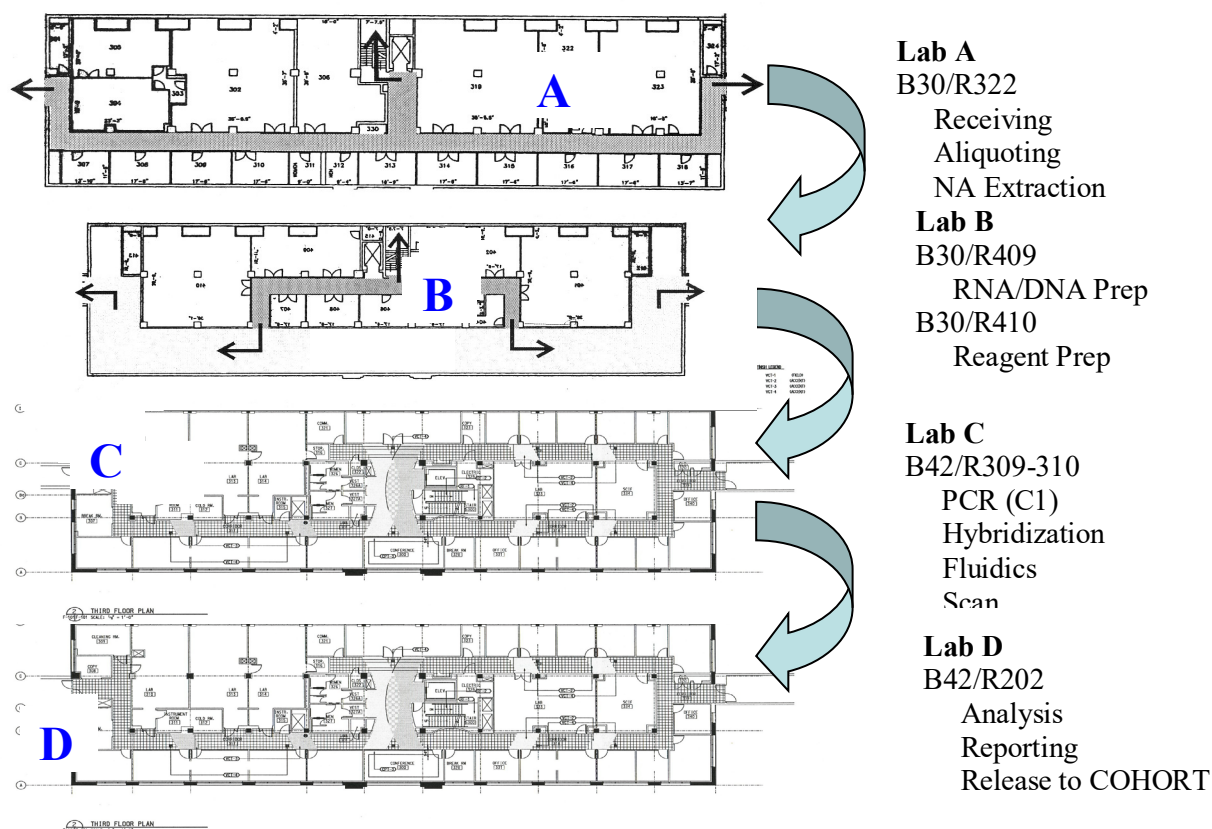


Fig. 8—Location of facilities at NRL

Minimal laboratory modifications were required to prepare NRL facilities for the SGD. Initially, power requirements were calculated based on the equipment needed in each lab as well as the refrigerators and freezers required for sample and reagent storage. The NRL Public Works Division then performed electrical upgrades, including installation of instrument-specific outlets and new connections to emergency power generators to be used in the event of power outages. Additional network cabling was installed throughout all laboratories to provide connectivity. A final decontamination was done on all laboratory spaces after renovations were complete.

Computer Hardware

The SGD project depended on real-time data capture and analysis. The hardware behind this effort was selected to minimize network traffic between servers and maximize the processing power per dollar spent. The network architecture was very important in moving the raw data between the Affymetrix scanners collecting image data and the server processing these images. The connections between the workstations and the server were physically isolated from the rest of the CBMS-EDC domain to reduce network traffic, and all cables and switches between the two were gigabit-enabled to ensure maximum throughput of data. Since the building that housed the system was already wired with category 6 cable and Ethernet jacks, only one building switch had to be replaced.

The servers used for the LIMS database were VisionMan servers, which are custom-made for TigerDirect. The servers were picked for the processing power needed, and all internal storage was purchased separately. This setup allowed maximum flexibility for the server configuration, and both servers with all associated storage were purchased for the same price as a comparable single Dell server.

The two LIMS servers were exact mirrors of each other, with real-time duplication of the log files and database from one server to the other using standard Oracle 10g tools. Six drives in each server were set up in the following configuration: 2 200GB hard drives held the operating system and Oracle installation. These drives were constantly mirrored using a hardware RAID card. The Oracle log files were placed on a separate 200GB hard drive. The database was placed on 3 200GB hard drives, which used software RAID 5 to ensure data protection in case of a single drive failure. The two servers were placed in different buildings for disaster protection.

All of the LIMS workstations Pentium III computers were obtained from either the National Institutes of Health or NRL. Barcode printers and scanners were purchased through Tracking Solutions, Inc. (Springboro, OH), a small company that deals exclusively in this area. Tracking Solutions recommended which printers and scanners to use, and found a label stock based on the requirements. All products were tested prior to purchasing large quantities to ensure they worked in the SGD system. In the future, we recommend the use of higher resolution bar code readers. Some operators had difficulty getting the readers to read the barcodes into LIMS and therefore resorted to manually typing in the numbers—a potential source of error.

Data Storage and Backup for SGD

The data for the SGD were stored in three major places. The first storage site was a Windows 2003 server in the CBMSEDC domain named CBMSE5.nrl.navy.mil, which was backed up to a SCSI RAID 5 array, and also to an AIT-2 tape drive. The LIMS program files and IT documentation were stored on a shared drive on this server. The second site was a VisionMan server running Windows Server 2003 with multiple RAID arrays, which housed the LIMS Oracle database. The operating system and Oracle programs were placed on a mirrored array, and the database was placed on a RAID 5 array. The entire server was mirrored to a second identical server in another building at NRL. The last site was an Apple Xserve RAID array, which consisted of fourteen 400GB drives placed in two mirrored RAID 5 arrays with separate power supplies and controller cards for each array. The Affymetrix raw data and SQL 2000 database were hosted on this array, along with the REPI data and all data analysis spreadsheets.

Xserve BLAST Server

The BLAST searches performed on patient samples during SGD were run on an Apple Xserve G5 cluster. The hardware consisted of five Apple G5 servers, each with dual G5 processors. The REPI program could multi-thread the samples processing, so 10 patient samples could be BLAST searched at one time. The genome database was housed on an Apple Xserve RAID array, and portions of it could be called into memory and searched against. The large amount of memory in the cluster allowed fast searches without any bottlenecks from disk searches.

LIMS Integration with COHORT

When SGD started, the working assumption was the demonstration would be managing an average of 300 specimens per day with a surge capacity of 450 samples. Several reagents were time-sensitive and needed to be mixed within hours of their use or they would degrade. In order to help the lab prepare for incoming samples, the goal was to be able to preview the specimen information after collection at the MTF, before the specimen arrived on-site at NRL.

Each patient encounter at the MTF resulted in a completed hard copy questionnaire that included site, PIN number and date and time of collection. MTF personnel entered the questionnaire information into the COHORT system via a secure socket 128-bit encrypted website. The server for COHORT was located in Falls Church at Skyline 3, and the LIMS database was located at the NRL in Washington, DC. The challenge was transferring the information from the COHORT system to the LIMS system in near real time to allow the laboratory to prepare for arriving samples. In order to accomplish this objective, a workstation placed at NRL housed a JAVA script (herein referred to as “Agent”) that was capable of viewing a table on the COHORT database that contained PIN and date/time of specimen collection. The Agent extracted the information from the COHORT database and pushed the data to the LIMS system so the data could be presented to the laboratory personnel. To accomplish transfer of the sample information, a table was created on COHORT and in the LIMS system. The greatest challenges were ensuring consistent data structure between the two sites and getting permission from the network security personnel at both sites to open a port to allow the information to transfer from the two sites. Figure 9 represents the overall workflow requirements for the LIMS system that was provided to the Labware® developers. It outlines the overall workflow and demonstrates how the data needed to be entered at the collection site. Once the data was entered at the collection site, it was automatically transferred to the LIMS system at NRL.

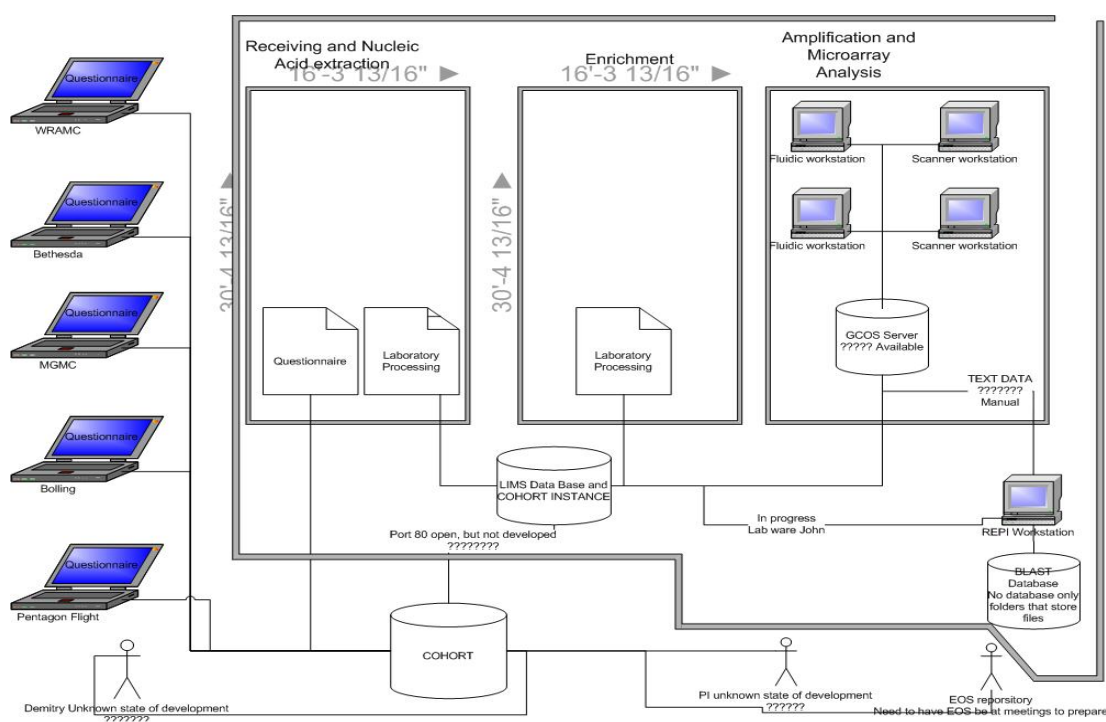


Fig. 9—Computer hardware and network diagram for SG

Two LIMS developers from Labware®, Inc. were resident on site for 6 weeks to evaluate the requirements and develop solutions for the LIMS. The major tracking requirement was that there would be three different rooms for laboratory processing and one room for data analysis and release of results. A rough diagram was provided to the developers to convey the overall information flow. The developers created an interface using crystal reports that mapped to each of the rooms and functions (Fig. 10).

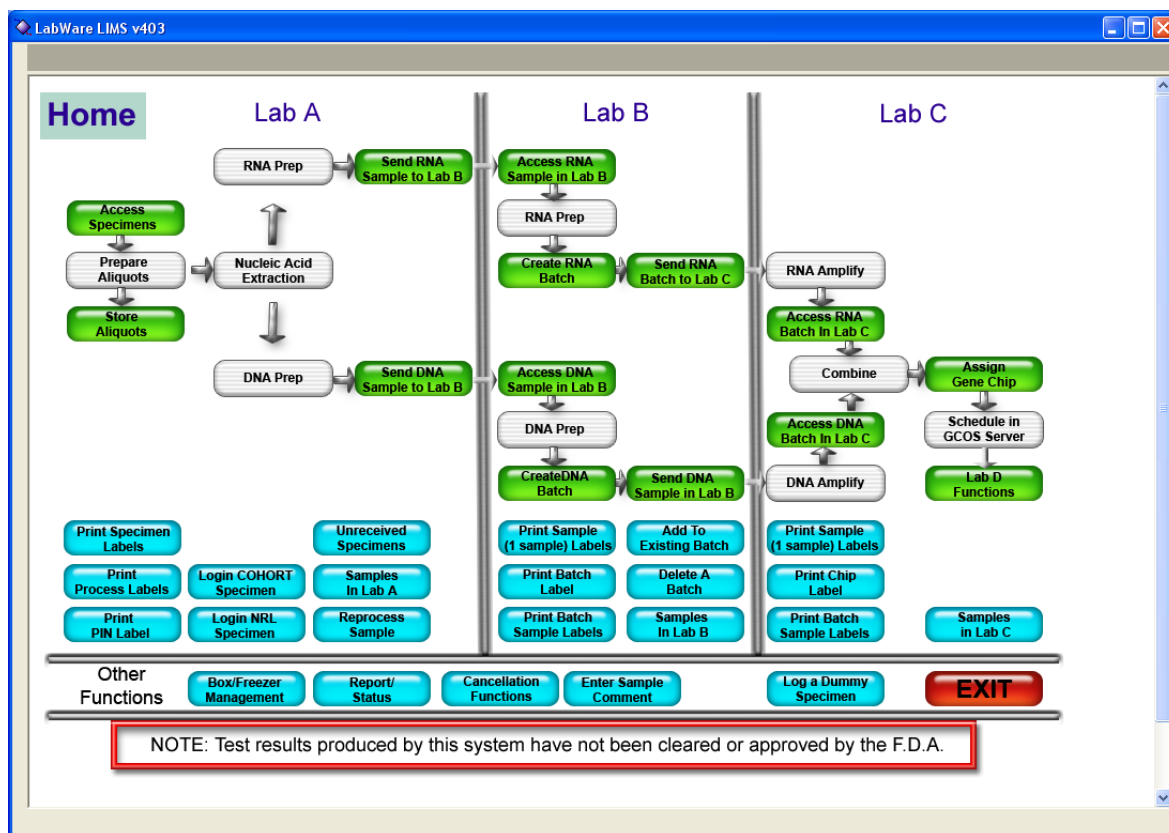


Fig. 10—LIMS Workspace View 1

To track specimens during processing and for storage, a solution was adopted from the EOS program that was currently in operations at San Antonio, Texas. The Zebra 110XiIII plus was used to print and scan the labels. The MTFs were provided pre-printed labels that included the pins and site information. Lab A, B and C each had a barcode scanner to track the logistics of the sample through the lab. The significant challenges were establishing the content of the barcode that was compatible with the tubes and that the barcode would adhere after freeze/thaw cycles. In addition, the Affymetrix chip analysis results had to be transferred back to the LIMS in association with a sample bar code.

FASTA files were manually generated and exported from the Affymetrix GDAS software prior to the SGD project. In order to handle the projected throughput, a GDAS script was written that would automatically convert the GCOS DAT file into a FASTA file with base calls. The REPI software then was capable of uploading the file and determining the pathogen present on the chip which was determined by an Expect value of less than or equal to 10^{-9} for a given tile region for the pathogen. This was a departure from the original operating procedure secondary to challenges with inconsistent pathogen names in GenBank and the limited amount of time to integrate a text search tool.

Prior to of SGD, results were only returned locally in San Antonio to a single PI. The results had never been returned to six geographically dispersed MTF. As part of the demonstration, the capability was to be able to return the results faster than culture to each of the Principal Investigators so they could have a near real-time surveillance based on pathogen identification.

Safety

A project-specific, biological safety review (BSR) was submitted to the NRL Biological Safety Committee (NRL-BSC) for approval prior to initiation of SGD. The BSR described the following: project title, principal investigator, personnel, type of research and materials, equipment requirements, experimental methods and procedures, nature of the research, previously-approved BSRs (previous relevant research), personal protective equipment, engineering controls, and the suggested biosafety level for the project. Once SGD was approved, the divisional safety representative consulted with the NRL BSC chair, the NRL Safety Branch, and the NRL Medical Clinic regarding safety matters for the SGD project to ensure that the project was conducted in compliance with the approved BSR and all applicable federal, Navy, and NRL policies. The project was also approved by NRL's Institutional Review Board for working with human samples.

In compliance with 29CFR1910.1450 and 29CFR1910.1030, prior to working in the laboratories, SGD personnel were provided safety (bloodborne pathogens, laboratory, and general) training. Training sessions informed personnel of: the contents and availability of the relevant safety policies (federal, Navy, NRL, and Code 6900), methods of compliance, personnel and responsibilities, communication of hazards (applicable definitions, information regarding exposure (signs, symptoms, routes, prevention, and response) to hazardous materials, communication of hazards (standard, warning labels and signs), common work practices, and standard operating procedures (SOPs) for the SGD labs.

A master safety folder was prepared, which contained print copies of all relevant safety documents (BSR, safety SOPs - NRL and SGD-specific), training rosters, training presentations, NRL's Exposure Control Plan, NRL instructions, DoN/DoD policies, and federal policies). SGD personnel were informed of these documents and were provided access to all safety documents for the duration of the project. In compliance with 29CFR1910.1030, hepatitis B vaccine was offered to all SGD personnel. Personnel in Room A were all vaccinated for influenza prior to the start of SGD. More than half of the personnel in other rooms were also vaccinated for influenza.

SGD laboratories were designed to meet the requirements of the project's biosafety level (BSL), as identified by the NRL-BSC chair on the SGD BSR. The project operated under a BSL-2 level; SGD laboratories were modified with appropriate engineering controls and SGD SOPs were written to ensure compliance with CDC/NIH guidelines for working at BSL-2.

Routine inspections of SGD laboratories were conducted to ensure that: appropriate Personal Protective Equipment (PPE) was available and used properly, engineering controls were functioning and used properly, and SGD personnel were performing tasks according to specific, safety SOPs, as designated during training and posted in SGD laboratories, to ensure compliance with regulations.

Quality Assurance / Quality Control

In mid-November, an advance team from AFIOH arrived and began to audit the SGD lab layout and experimental protocols. Discussions in that period led to the specification that each sample to be tested would have both RNA and DNA internal control spiked in Lab A prior to nucleic acid extraction. Since the RPM v.1 carried two *Arabidopsis thaliana* (plant) tiles, we adapted one target for DNA (ATNAC1 tile) the other for RNA (ATTIM tile) by transcribing the ATTIM DNA gene to RNA. These controls were spiked into the nasal wash aliquot being used for RPM analysis prior to the nucleic acid extraction phase. The appearance of both the RNA and DNA controls on the RPM would confirm that the entire nucleic acid isolation and processing had worked properly.

A rigid set of criteria was imposed in order for a result to be released to COHORT (Fig. 11) and thereto back to the PIs. In order for a pathogen negative result to be released, both *A. thaliana* controls were required to survive the human nucleic acid subtraction protocol and be amplified properly, producing hybridization signals on their respective tiles. The absence of either of these in the final microarray result constituted a rejection of that microarray assay unless an RNA virus or DNA pathogen (largely bacteria) was detected. The latter accommodation was included in the case the pathogen was in excess over the control and diminished the amount of PCR reactants available to the respective control type. Thus any sample producing a positive pathogen response or two positive controls was released to COHORT. If the sample was negative for both pathogen and at least one control, a duplicate aliquot was thawed and reanalyzed.

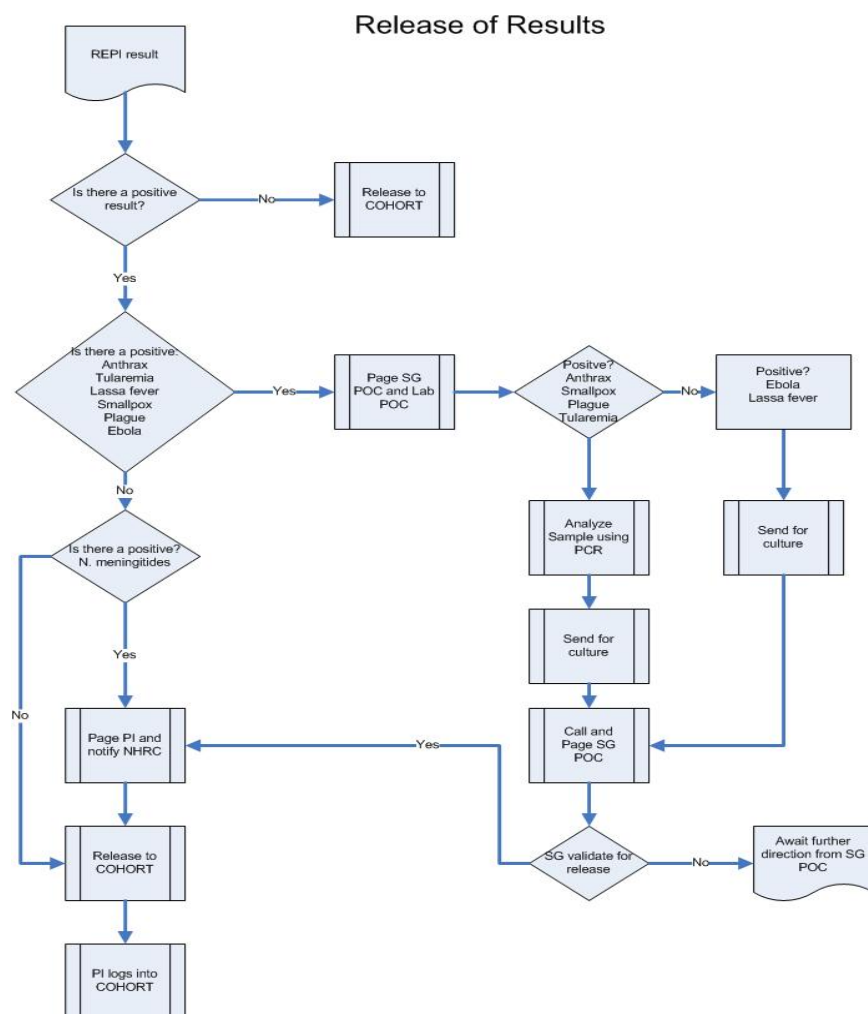


Fig. 11—Lab D criteria and protocols for release of NRL results to COHORT

At the direction of the AF/SG, final SGD protocols and procedures were put in place to meet standards established by the DoD Clinical Laboratory Improvement Program (DoD-CLIP) for moderate and high complexity laboratories, as well as the Clinical Laboratory Improvement Amendments of 1988 (CLIA'88) as implemented in DoD by a Memorandum of Agreement between DoD and the Department of Health and Human Services. This operating instruction specifically covers patient test management, quality control, proficiency testing, and personnel competency. Technical and analytical performance

standards set by CAP and Federal agencies (i.e., FDA) were also incorporated into the program guidelines, with the intent that lessons learned in SGD would be used to guide a future application of the technology for FDA approval and employment into the clinical laboratory setting. QA/QC details are provided in the complete protocol package.[4] The major areas of concern were:

- Specimen integrity and chain of custody. During the SGD demonstration, all incoming patient specimens were labeled with a PIN barcode, as soon as they arrived at NRL. The PIN barcode was entered into LIMS system in Lab A, which in turn assigned unique identification numbers for storage, shipment and processing as the specimen moved through Lab B and Lab C. Specimen integrity or the specimen's suitability for testing was noted and every attempt was made to correct problems prior to discarding a specimen.
- Protocols/Operating instructions were made available in all technical areas. All personnel in each laboratory were required to read and understand protocols/operating instructions to include all authorized changes. Operating and maintenance procedures for reagents and each piece of equipment were available in the appropriate work area.
- Reagents and materials were prepared daily and were labeled with expiration dates, initials of the preparer, contents and concentration. Reagents purchased from commercial sources were visually inspected upon arrival for overall appearance. Manufacturer's labels had the manufacturer's name and lot number, date of expiration, contents and concentration of reagent, chemical hazards, date received and handling/storage requirements. Reagents were not used beyond the expiration date.
- Overview of Laboratory QC Procedures
 - Laboratory A was responsible for recording the temperature inside the cooler when specimens were received. Temperatures were recorded on the chain of custody form that arrived with each sample.
 - Laboratory B delivered RNA aliquots obtained after a reverse transcription reaction and DNA aliquots obtained after a Sequenase reaction to be tested on the BioRad MyiQ-cycler to detect the presence of *A. thaliana* genes. The presence of *Arabidopsis* indicated the successful execution of all steps performed in room B as well as the proper nucleic acid extraction in room A.
 - Laboratory C1. PCR amplification results from the Bio-Rad MyiQ-Cycler were downloaded and stored for independent review and QC.
 - Laboratory C2 technicians reviewed scans to ensure AFFX-TaqIQ-EX, Oligo B2 (on-chip control) was present.
 - Laboratory D personnel ensured DNA and RNA spiked controls in each sample were present before the release of results.
 - All Laboratories were responsible for recording temperatures of various pieces of equipment (e.g., refrigerators, freezers, heat blocks) for their laboratory

A Quality Assurance Committee met bi-weekly to review policies, oversee procedure implementation and assist the Test and Evaluation Assessment done by AFMESA. Participants included the AF Project Site Leader and NRL equivalent; Project Manager and NRL equivalent; Quality Assurance Coordinator, and the laboratory managers for each lab.

Affymetrix has built-in Standard Controls, Process Controls and User Defined Controls. The protocol used Affymetrix Standard and Process controls. In the design of the microarray, the EOS team built in *Arabidopsis* RNA/DNA control sequence information on the microarray, which provided the capability of including internal controls for monitoring the entire analytical process.

Troubleshooting and Establishment of Final SGD Protocol

In the second week of November 2004, the NRL SGD team that was trained in the EOS lab was given 20 healthy nasal wash samples spiked with RNA and DNA controls and processed those samples unsupervised on the MDW production line. They achieved positive results on approximately 80% and 60% of the DNA and RNA controls, respectively (~50% chance of a releasable result). The negatives were attributed to loss of the controls in the technically challenging background subtraction steps in Lab B or a failure to amplify in Lab C1.

The first AF team (Alpha) arrived in the third week of November and attempted to reproduce the NRL data following one week of training. Initially, they fared less well than the NRL team but a second attempt following further training (didactic and hands-on instruction) showed them to have a similar of success in executing the protocol.

In the first two weeks of December, a second AF team (Bravo) was trained, more equipment arrived, the LIMS system was brought into daily use, and the total number of specimens that could be processed in a single day increased. Some clinical specimens began to be collected at the MTFs – these were tested and found to be negative for RNA viruses but spurious results with DNA background subtraction (loss of DNA controls in some batches) did not allow release of results.

By mid-December, the increased experience of AF personnel allowed them to consistently achieve ~90% success rates on the RNA internal controls. However, both NRL and AF staff experienced a sudden and complete loss of the ability to obtain DNA control signals on the chip. This forced the initiation of a complete audit of the SGD production equipment, reagents, and protocols by EOS personnel during the Christmas break.

The protocol audit continued with the return of the AF personnel after Christmas. By then, AF Capts Denise Lennon, Fe Lobo-Menendez, and David Hagerty were expert in carrying out the functions of Labs A, B, and C, respectively. Working with NRL staff, they established a record system that tracked the metadata not captured in LIMS. The SGD and EOS personnel worked diligently over a period of two weeks to systematically eliminate sources of error and DNA target loss, finally discovering that the DNA targets were being completely lost by non-specific adsorption on the streptavidin-coated magnetic beads. While a fractional loss of DNA target in this step would be considered normal, we had not previously encountered a complete loss onto the bead surface in over six months of routine use prior to SGD. The nonspecific adsorption of DNA to the streptavidin magnetic beads corresponded to the use of the new lot of this reagent that was acquired for the SGD project. Later, the vendor of these beads confirmed that the beads were of suboptimal quality and offered a complete replacement. However, in early January 2005, a replacement of the faulty lot was not possible. Three additional days during the first week of January were directed at attempting to modify the DNA subtraction step using protein and DNA blocking molecules, without success.

In the second week of January, the EOS scientists advocated pilot experiments with omission of the DNA subtraction step in order to enable the release of the results obtained for RNA pathogens, which were predicted to be much more likely with the increasing appearance of influenza viruses in the DC metro area. These experiments, using human adenovirus type 4 (DNA) and FluMist® vaccine (influenza A and B mixture) spiked into nasal wash in Lab A, demonstrated that the background noise on the chip was increased but manageable, allowing 100% detection of high titer adenovirus and FluMist® viruses with approximately 90% of assays satisfying the minimum release criteria (both RNA and DNA controls positive).

On 12 January 2005, the protocol was modified so that DNA subtraction was omitted and 160 samples were run, inclusive of the 109 SGD clinical samples that had been backlogged until a protocol

was approved. Of the SGD samples, 9 (8%) were influenza A (H3N2)-positive, 1 was human coronavirus 229E-positive, 89 (81%) were pathogen-negative, and 10 (9%) were not releasable (lacking RNA and/or DNA control without corresponding pathogen observed) and required repeat analyses. Influenza A was only observed using the original protocol in samples collected after January 4, 2005.

In subsequent operations, several opportunities for improvement in the established practices were identified:

- During most of the month of January, LIMS did not record the time of day in which each sample entered each laboratory. This shortfall was corrected in February, but individual lab turnaround times for the 300-sample throughput experiments were not easily accessible.
- The tubes used for the i-Cycler PCR instrument were too small to accommodate the bar code labels. Thus the samples were placed in rows of attached tubes with batch labels only by staff in Lab B, with samples processed by the DNA pathway in one row and by the RNA pathway in another row. After the PCR step, the samples were removed from the PCR machines by Lab C operators and the DNA and RNA pathway samples combined for hybridization onto the chip. In two instances, the operators in Lab C recognized that the samples were not matching up correctly and restarted the patient analyses using the labeled aliquots stored at the end of the Lab B process. These were patient analyses run during relatively low intensity periods and the analysis was only delayed by about 5 hours while the PCR step was redone. However, the same problem could have occurred during the high intensity operations associated with spiked sample experiments and mismatches would be much more difficult, if not impossible to observe. A method for eliminating sample mix-up from future protocols needs to be identified (most likely by combining the DNA and RNA processing pathways much earlier in the protocol); it will not be a problem once the sample processing is automated.
- A software bug was found in the REPI program which called the BLAST engine when the process was switched from a single-processor server to the multi-processor server. The operations were not multi-threaded correctly thus creating errors in pathogen identification which were obviously absurd and easily caught in the Lab D analyses. This was a problem easily amenable to solution and was corrected.

CHAPTER 4: PATIENT RESULTS

Overview

This study was conducted between December 2004 and February 2005 employing patients recruited from the Emergency Departments and clinic settings at six Military Treatment Facilities (MTFs) in the Military District Washington (MDW). The participating facilities were: Malcolm Grow USAF Medical Center, Andrews Air Force Base, MD; Bolling Air Force Base Medical Clinic, Washington DC; Pentagon Flight Clinic, Arlington, VA; DiLorenzo Clinic-Pentagon, Arlington, VA; Walter Reed Army Medical Center, Washington DC; and National Naval Medical Center, Bethesda, MD. The collection teams covered all standard clinic times and typically recruited patients in the Emergency Departments from 0600-2400. Active duty military, retired military, and their families are eligible for care at MTFs. The demographics of this group have previously been shown to have a broad age, gender, and geographic distribution that can be generalized to the Washington DC region. Patients presenting to the participating MTFs who were 18 years of age or older, had a documented or self-reported measured temperature >100.4 degrees Fahrenheit, and had any flu-related symptom(s) were eligible for the study. For each patient, clinical signs and symptoms were recorded and nasal wash and/or throat swab specimen were collected. The results of the signs and symptoms were compared to the culture and microarray results at the conclusion of the study. As expected, clinical signs or symptoms showed no statistical difference among the different pathogens. This result reinforces the requirement for a diagnostic test to assist in differentiating between pathogens causing “flu-like” symptoms and possible exposure to biothreat agents.

A total of 565 patients (598 samples, including 33 throat swabs) met all study criteria and were tested using the standard SGD (random amplification) protocol with human DNA subtraction omitted. The microarray test results for patient samples were reviewed as soon as each analysis was complete, then reported to COHORT immediately. From the total number of samples, 450 were sent to Naval Health Research Center (NHRC) in San Diego, CA for culture and polymerase chain reaction (PCR) characterizations. Influenza virus was the most commonly identified respiratory pathogen by all methods. Of the 598 samples reported to COHORT, 102 (17%) were identified as positive for influenza when using the SGD standard protocol. Culture methods (30–42 days) performed on the subset of 450 samples revealed that 213 (47%) were positive for influenza by culture. Additional pathogens were identified by culture or PCR in low incidence.

Further experiments were done to examine the microarray performance with modified protocols. First, 464 replicates of the 598 patient samples were amplified by multiplex PCR using a specific primer cocktail (UberMix) rather than unbiased amplification with random primers. A second set of experiments involved re-analysis of sequence calls previously made by the REPI with the standard algorithm. There-analyses used a “relaxed” algorithm in which targets were identified by associating the top sequence similarity ranking (uniquely high BLAST bit score with corresponding low expect value) with single pathogen genome database records instead of satisfying an arbitrary BLAST cut-off expect value of 10^{-9} . There were 351 “matched” results, that is, samples for which there was data for culture, random amplification, and multiplexed amplification as analyzed with the relaxed algorithm. Of the 351 matched experiments, thirty-day cultures detected Flu A in 147 (42%) of the specimens. Used in combination with the relaxed algorithm, the random amplification (standard SGD protocol) and multiplexed (UberMix) amplification protocols resulted in a total of 114 unique positive results for 32% of the total “matched” sets; 90 of these agreed with the culture results (30-to 42-day). Overall, modifications to the protocol allowed increased detection of a variety of additional respiratory pathogens without false detection of BT agent genetic sequences.

Background

The Version 1.0 Respiratory Pathogen Microarray (RPM v.1) was designed for proof-of-concept detection of common respiratory pathogens (especially adenovirus, influenza A/B viruses, and *Streptococcus pyogenes*) encountered in the basic military trainee (BMT) population at Lackland AFB in San Antonio, TX. Approximately one third of the 29.7 kb of sequences tiled on the RPM v.1 were dedicated to three human adenovirus serotypes which are unlikely causes of respiratory illness in adults in a typical urban population, but have been frequently observed with military trainee populations. However, the RPM v.1 had very representative coverage of recent influenza A and B gene sequences for hemagglutinin (HA; 1, 3, and 5), neuraminidase (N; 1, 2), and matrix (M) proteins. Target sequences for the remaining common respiratory pathogens were obtained from the open literature and were composed of much shorter sequences than those for adenovirus and influenza A and B viruses. A complete description of the RPM v.1 can be found in Lin et al. [7]

In previous studies, all non-BT pathogen tiles on the RPM v.1 except those for West Nile virus and human coronaviruses were determined to be correct by resequencing of the corresponding purified genomic targets. However, in the summer of 2004, the RPM and its attendant DNA and RNA pathogen nucleic acid protocols were not optimized for speed and there was no information collected on the sensitivity and specificity of a fully multiplexed RPM assay for pathogens other than influenza A and human adenovirus type 4 (HAdV4). The SGD demonstration afforded the first opportunity to adapt the protocol for clinical or military operations and perform a systematic assessment of the RPM for detection of common respiratory pathogens in an urban environment while providing the utility of BT agent detection during the winter of 2004-2005.

From June 2003 until September 2004, separate processing pathways existed for RPM detection of RNA and DNA pathogens. For development of the RNA processing pathway, spiked and clinical (nasal wash and throat swab) influenza A (+) specimens were processed using different methods (random, universal, and multiplexed PCR). In early 2004, it was shown that to achieve acceptable levels of sensitivity with random amplification approaches, human messenger and ribosomal RNA had to be subtracted using the MICROBEnrich™ kit (Ambion; Austin, TX) first. A preferred random amplification strategy was adapted from a previously described method.[8, 9] This allowed positive H3N2 and H1N1 Flu A detection in 16/19 different culture-positive Flu A nasal wash and throat swab specimens at concentrations as low as 0.45-3.75 pfu/150 µL sample aliquot. Flu A was detected in 6/8 specimens in this concentration range following subtraction of human RNA. This approach was incorporated into the final recommended SGD protocol (Chapter 2).

To satisfy an operational requirement of processing 300 or more samples per day, SGD necessitated a partially combined RNA and DNA protocol. Because of the resultant time limit for protocol performance and schedule for finalizing the protocol, the combined DNA and RNA pathogen processing protocol was frozen and implemented for training in November, 2004. However, in mid-December 2004, a manufacturing failure of the streptavidin-coated magnetic beads used for binding and removal of captured human DNA prior to random amplification forced a decision on whether to finalize the protocol without human DNA subtraction or to delay full operational capability (FOC). After preliminary testing in the first week of January 2005, a final protocol without DNA subtraction was approved for use in FOC. In February 2005, an alternative strategy was developed and tested using a non-optimized multiplex PCR amplification approach for replicate sample testing.

Experimental Design and Methods

Aliquoting of Patient Specimens

Patient specimens meeting the criteria approved by individual service's Institutional Review Boards (IRB's) began to arrive at NRL from all six MTFs on 9 December 2004. Each nasal wash specimen having a volume of ≥ 2 mL was diluted with sterile normal saline to a final volume of 4.0 mL. The diluted specimen was then split into four culture aliquots (2 for bacterial and 2 for viral culture, each containing 0.75 mL of nasal wash and 0.75 mL of transport media) and six 150 μ L specimen aliquots. Samples from the first weeks (9 Dec 2004–12 Jan 2005) were stored in a -80° C freezer until all SGD protocols had been finalized (11 Jan 2005). From that point on, previously stored samples were processed by 14 January and all new incoming patient samples were processed as they arrived at NRL, with few exceptions. During non-FOC periods, samples received after 2000 hr were aliquoted and nucleic acid extractions were frozen in a -80° C freezer. Subsequent processing of the DNA/RNA fractions from these samples was done at the start of the next scheduled morning shift.

RPM v.1 Sample Processing and Analysis

Each 150 μ L nasal wash aliquot for microarray processing was subjected to total nucleic acid extraction (Lab A), separation of pathogen RNA from human RNA (human DNA not removed due to reagent failure) (Lab B), PCR amplification using the standard SGD (random amplification) protocol (Lab C1), and chip analysis (Lab C). Prior to amplification (Lab C1), each sample was split; half of each sample was carried through the rest of the analysis (Lab C1, Lab C2) using the standard (random amplification) protocol, whereas the remaining split was frozen for later analysis using a different (multiplex PCR) primer mix. Use of these “splits” controlled for variations in extraction and preparation efficiencies and enabled direct comparisons between results obtained with the different primer mixes.

As described earlier (Chapter 2), random amplification of mixed populations of DNA and RNA can result in large numbers of different amplicons, without amplification bias in favor of any specific target. The inability to perform a satisfactory DNA subtraction step resulted in an increase in background noise on the RPM, thereby decreasing the ability of the SGD chip to detect relevant but low abundance sequences. To enhance sensitivity, specific primers were added to the PCR mixes, such that sequences specific to the tiled regions on the SGD RPM v.1 chip would be selectively amplified. This “UberMix” primer mixture (130 different primers) was used to amplify the remaining “split” of processed nucleic acids from Lab B in February 2005.

Microarray sequence data was analyzed using either the standard protocol (BLAST subjected to the 10^{-9} expect value and used to report to COHORT), or the “relaxed” algorithm. In the latter approach, BLAST returns were not subjected to the 10^{-9} expect value threshold prescribed in the standard SGD protocol. Rather, the database record having the highest BLAST bit score was referenced to a species look-up table.

Microbiological Laboratory Analysis

The Naval Health Research Center (NHRC) in San Diego, CA was tasked as the lead reference laboratory for the processing of samples collected during SGD using standard clinical assays. To define appropriate conditions for collection of specimens, transport media, aliquoting procedures, sample storage and shipping of the collected samples, extensive communication took place between the NHRC Respiratory Disease Laboratory and SGD personnel in the weeks leading up to FOC. For confirmatory NHRC testing, preference was given to individuals enrolled with complete sample sets of 4 aliquots, although some were received without adequate aliquots for complete testing.

It was recognized that culture with preservation and storage of the isolate was the preferred reference assay methodology. Previous work has shown the limitations of culture for certain pathogens (e.g., Coronavirus) for which PCR is a better method. Therefore, it was agreed that for some hard-to-culture pathogens of interest (Coronavirus, *Mycoplasma pneumoniae*, *Chlamydia pneumoniae*, and *Bordetella pertussis*), molecular detection (PCR) techniques would be utilized and treated as the reference assay for comparison to SGD analyses.

Viral laboratory analyses were performed by FOCUS Diagnostics, Inc. (Cypress, CA), a CAP-certified laboratory, or by AFIOH. Patient samples were stored in CVM media and subsequently shipped frozen to NHRC. Viral samples were also sent to FOCUS Diagnostics for culture processing. Viral processing included: inoculation of sample, monitoring for distinguishing cytopathic effect (CPE) for given pathogen (7-15 days of monitoring), sub-passage as necessary and confirmatory testing as needed (i.e. direct monoclonal immunofluorescence assay (DFA), indirect immunofluorescence assay (IFA) or PCR). Positive culture isolates were frozen and archived for shipment back to NHRC.

Bacterial analyses were performed at NHRC Respiratory Disease Lab. Patient samples were stored at -80°C in tryptic soy broth with 30% glycerol at NRL, and subsequently shipped frozen to NHRC. Samples were then analyzed for the presence of bacterial pathogens utilizing classic culture techniques. Blood agar plates were inoculated with a given sample for the detection of *S. pyogenes* and *S. pneumoniae*. Chocolate agar plates were inoculated for the detection of *H. influenzae* and *N. meningitides*. All plates were incubated at 35-37°C with CO₂ for 24-48 hours and then checked for growth. Bordet gengou agar plates were inoculated for the detection of *B. pertussis*. These plates were incubated at 35-37°C for 7-9 days and checked daily for growth. Both SP4 agar plates and broth were inoculated for the detection of *M. pneumoniae*. For all samples, media was incubated at 35-37°C for 42 days and checked frequently for growth. For all culture methods, if growth resembled colonies appropriate to the pathogen sought, confirmatory tests (i.e., susceptibility disks, biochemical testing, etc.) were conducted to ascertain the identity of the suspected colony.

Additional PCR tests performed after all samples were received were aimed at detection of *C. pneumoniae*/*C. psittaci* DNA, coronavirus OC43 RNA and coronavirus 229E RNA, *B. pertussis* DNA, and *M. pneumoniae* DNA. PCR for these targets also allowed detection of a coronavirus in 33 specimens and *M. pneumoniae* in a single specimen. Weekly reports were provided to NHRC and then compiled with bacteriology laboratory testing.

Results and Discussion

Number of Patients Tested

For SGD, the inclusion and exclusion criteria for entry into the study were set as follows: *Any subject who presented to medical treatment facility with a documented or self-measured temperature >100.4°F, >18 years old and febrile respiratory illness symptoms was eligible for voluntary consent to the study.* The data from COHORT ICD-9 codes for previous years indicated that the projected range of samples ranged from 15 to 450 individuals presenting to the MTFs in the MDW per day. These numbers were used to help predict the number of persons that could be expected in SGD. In fact, SGD sample numbers rarely exceeded 20/day (Fig. 12). The discrepancy between the original projected ICD-9 code numbers and actual results was likely due to inaccuracies in medical coding at the MTFs, as well as lack of coding for fever or respiratory symptoms. The lower-than-expected number of patients was also attributed to a mild-moderate flu season in the MDW during the study [10]. Figure 12 shows the distribution of samples from the MTF's from 9 December 2004 to 25 February 2005, while Figure 13 shows the total number of samples per site. A total of 605 unique questionnaires were entered into the COHORT system for the specimens that were collected, resulting in 598 valid samples for analysis. Seven of these were removed (consent issues (4), specimen leak (1), controls (1), quantity not

sufficient (1)), as summarized in Table 1. No discernable trend or pattern was observed for sample distribution during the limited period of SGD.

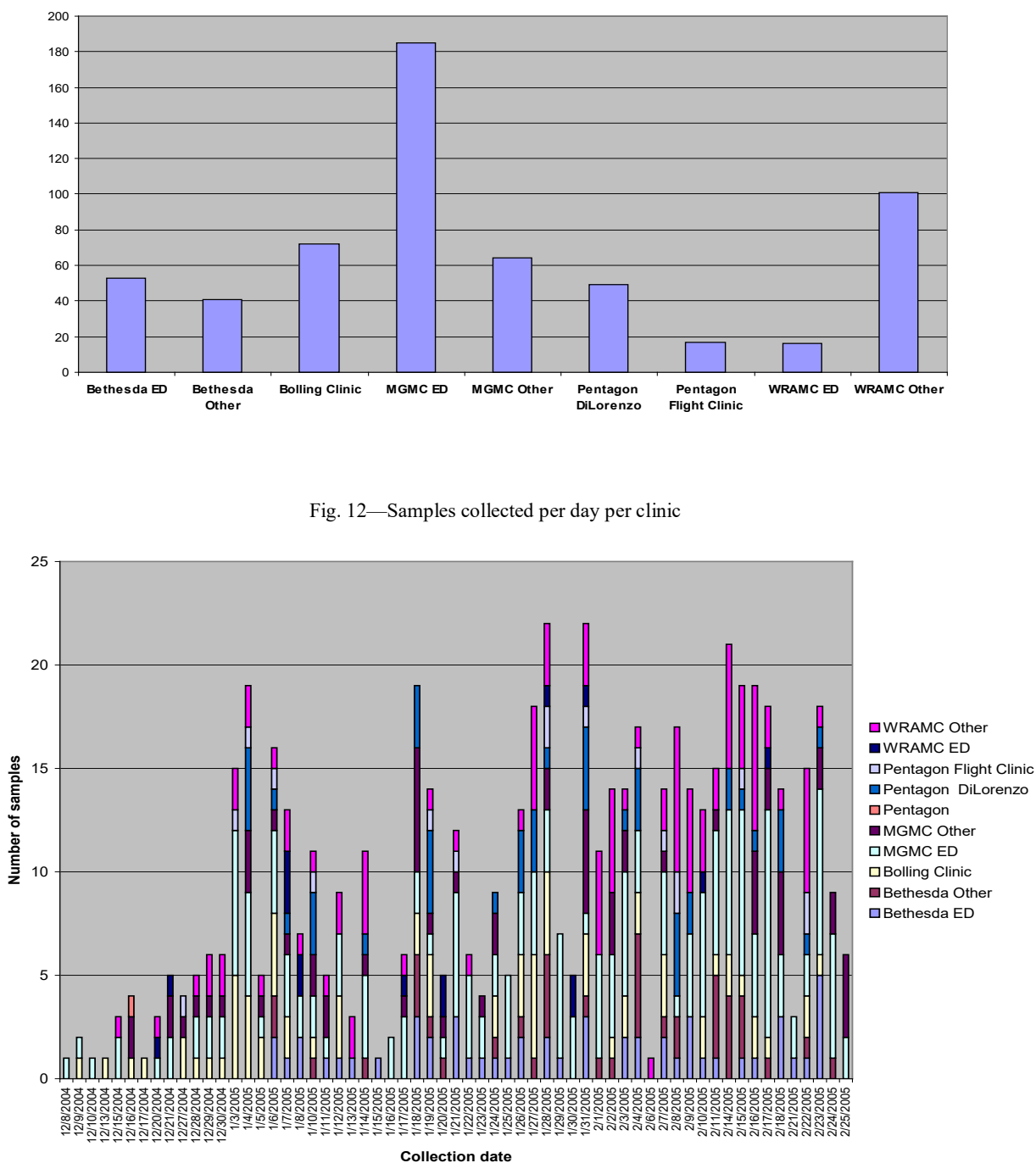


Table 1—Patient Samples

Number of samples	Released to COHORT	Equivocal result in COHORT	Cancelled not processed	Used in Analysis	Comment
589	Y	N	N	Y	
9	Y	Y	N	Y	Report only in COHORT (manual interpretation)
4	Y	N	N	N	Incomplete consent forms
1	N	N	Y	N	Cancelled secondary to leakage of transport to lab
1	N	N	Y	N	Quantity of material collected not sufficient
1	N	N	Y	N	Controls did not work
Total	602	9	3	598	

Microbiological Laboratory Analyses

Table 2 summarizes the results obtained from NHRC testing (culture and PCR). All results for each patient were reported as testing was completed. All cultured viruses and bacteria have been archived at the NHRC Respiratory Disease Laboratory in San Diego, CA.

Table 2—Summary of Results for Pathogens by Culture and PCR (n=450)

PATHOGEN	POSITIVES, n (%)	NEGATIVES, n (%)
<i>C. pneumoniae</i>	0 (0.0%)	450 (100%)
Adenovirus	2 (0.5%)	448 (99.5%)
Coronavirus (non-SARS)	33 (7.3%) ^a	417 (92.7%)
Human metapneumo virus (HMPV)	0 (0.0%)	450 (100%)
Influenza virus	213 (47.3%) ^b	237 (52.7%)
Parainfluenza virus (Type 2)	2 (0.5%) ^c	448 (99.5%)
Respiratory Syncytial virus	0 (0.0%)	450 (100%)
Rhinovirus	0 (0.0%)	450 (100%)
<i>S. pyogenes</i>	9 (2.0%)	441 (98.0%) ^d
<i>S. pneumoniae</i>	10 (2.2%)	440 (97.8%) ^d
<i>H. influenzae</i>	7 (1.6%)	443 (98.4%) ^d
<i>B. pertussis</i>	0 (0.0%)	450 (100%) ^d
<i>M. pneumoniae</i>	1 (0.2%) ^e	449 (99.8%) ^f
<i>N. meningitides</i>	0 (0.0%)	450 (100%) ^d

^a Coronavirus detection was made by PCR only, there were no culture positives: OC43 RNA (n=21), 229E RNA (n=12)

^b Influenza virus types: FLUA (n=184), FLU B (n=29)

^c Both culture isolates were parainfluenza virus type 2.

^d There were 11 samples that were unavailable for culture testing

^e *M. pneumoniae* detection was made by PCR only; there were no culture positives.

^f There were 47 samples unavailable for culture testing, but all PINs were tested by PCR detection methods

Overall Microarray Results versus Culture

The number of positives reported to COHORT during the demonstration was lower than the number observed by culture. This was not known until the completion of the demonstration, since sample collection ended on February 25, 2005 and the thirty-day culture results did not become available until the

end of March 2005. If these results had been available on a next day basis (versus 30–42 days), it is possible that protocol changes could have been made. However, since the results were not available until after SGD had been completed, only the culture data could be used for final assessments. Positive results sent to COHORT were obtained using the standard protocol with no human DNA subtraction and automated BLAST cutoff values of 10^{-9} (which later proved too restrictive). No bacterial pathogens were detected in nasal wash specimens using the standard SGD protocol.

Table 3 shows the collective number of positive results obtained using culture, random PCR (standard protocol as reported to COHORT), using the “relaxed criteria” for sequence calls in the standard protocols and multiplex PCR (UberMix). As seen in this table, culture successfully detected influenza A, influenza B, *S. pneumoniae* and *S. pyogenes* (Group A Strep), and some adenovirus from nasal wash specimens. Coronavirus is very difficult to culture, so the lack of observed culture positives for that species is not surprising. The coronavirus results shown in the table were obtained by PCR only.

Table 3—Positive Results Obtained for All Methods: Reference Assays, Random PCR (Standard Protocol as Reported to COHORT, Random PCR with Relaxed Criteria and Multiplexed PCR (UberMix, Relaxed Criteria)

Positives for	Culture/PCR	COHORT	“Relaxed”	UberMix
Total	450	598	598	464
Influenza A	184	91	108	104
Influenza B	29	11	13	15
Adenovirus	2	0	0	17
Coronavirus	33	4	7	2
Rhinovirus	0	1	3	0
<i>N. meningitides</i>	0	0	0	38
<i>S. pneumoniae</i>	10	0	0	21
<i>S. pyogenes</i>	9	0	0	3
Respiratory syncytial virus A	0	0	0	1
Respiratory syncytial virus B	0	0	1	4
Erythromycin resistance	NA	NA	0	19

The number of positive samples using the “relaxed” column of Table 3 were obtained by re-analyzing each chip data (the experiment was not repeated) using a modified algorithm for calling a positive. This modified algorithm did not use the (arbitrarily defined) BLAST expect score cutoff of 10^{-9} but instead relied on the use of pathogen database algorithms provided by the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>) which maintains the database (GenBank) used for BLAST analysis. Thus, any BLAST result that returned a database record having a unique (high bit score / low expect value) numerical combination could be identified by finding it in a list of records assigned to a particular pathogen by NCBI. If two or more pathogens had the same numerical combination, the method returned all records with those values—indicating ambiguity between different species. This was a far superior method than that used in the standard software protocol used to report to COHORT because it did not penalize species for having a large number of redundant database records (e.g. influenza A) and did not suffer from the lack of a naming convention in the GenBank database. Use of this algorithm alone increased the number of positive influenza A results by 19% but this was not enough to completely overcome the increased noise due to random amplification of human DNA in samples processed without subtraction. The use of the relaxed algorithm on the same raw data used for COHORT reporting also gave improvements in the number of positive microarray results for influenza B, coronavirus, and rhinovirus.

In addition to influenza A virus, the 450 cultured specimens returned positive results for influenza B virus (29 specimens; 6.4%), adenovirus (2 specimens; 0.4%), Type II parainfluenza virus (2 specimens; 1.5%), *S. pyogenes* (9 specimens; 2.0%), *S. pneumoniae* (10 specimens; 2.2%), and *H. influenzae* (7

specimens; 1.5%). The RPM v.1 did not contain resequencing tiles for Type II parainfluenza virus, or *H. influenzae*, both of which were detected by culture.

The 598 patient specimens analyzed with the standard RPM protocol (random amplification) returned positive results for influenza B virus (13; 2.1%), Coronavirus 229E (7; 1.2%); Rhinovirus (3; 0.5%), and RSV B (1; 0.2%). The 464 patient aliquots analyzed with the UberMix protocol (multiplex PCR amplification) returned positive results for influenza B virus (15; 3.2%), adenovirus (17; 3.7%), Coronavirus 229E (2; 0.4%), *N. meningitides* (38; 8.2%), *S. pyogenes* (3; 0.65%), *S. pneumoniae* (21; 4.5%), RSV A (1; 0.2%), RSV B (4; 0.8%), and one or more erythromycin-resistant bacterial plasmid genes (19; 4.1%) having a high correlation with *S. pneumoniae* occurrence. Antibiotic resistance cannot be determined with the culture/PCR tests performed.

As expected, the use of multiplexed primers (UberMix) increased both the number and types of pathogens detected relative to the standard protocol. Especially noticeable was the large increase in calls for adenovirus, and the bacterial pathogens *N. meningitides* and *S. pneumoniae*, both of which are known to colonize the nasal passage in healthy individuals, and the increased detection of *S. pyogenes*, RSV A, and RSV B. Only this amplification technique gave detection of erythromycin resistance gene markers (plasmid conferred) in 19 individuals.

Representative RPM images are shown in Fig. 14 for (A) standard SGD protocol, and (B) multiplexed PCR amplification (B). The use of multiplexed primers resulted in the appearance of high intensity signals corresponding to primer binding sites within each of the respective tile regions (B), whereas this was not observed when random amplification was used (A). In order to eliminate potential artifacts, binding sites within the respective tile regions that matched prime sequences were marked as N (typically 20-25 bases) prior to REPI and BLAST analysis. This process was applied to each experiment conducted using multiplexed PCR described in this chapter and Chapter 5.

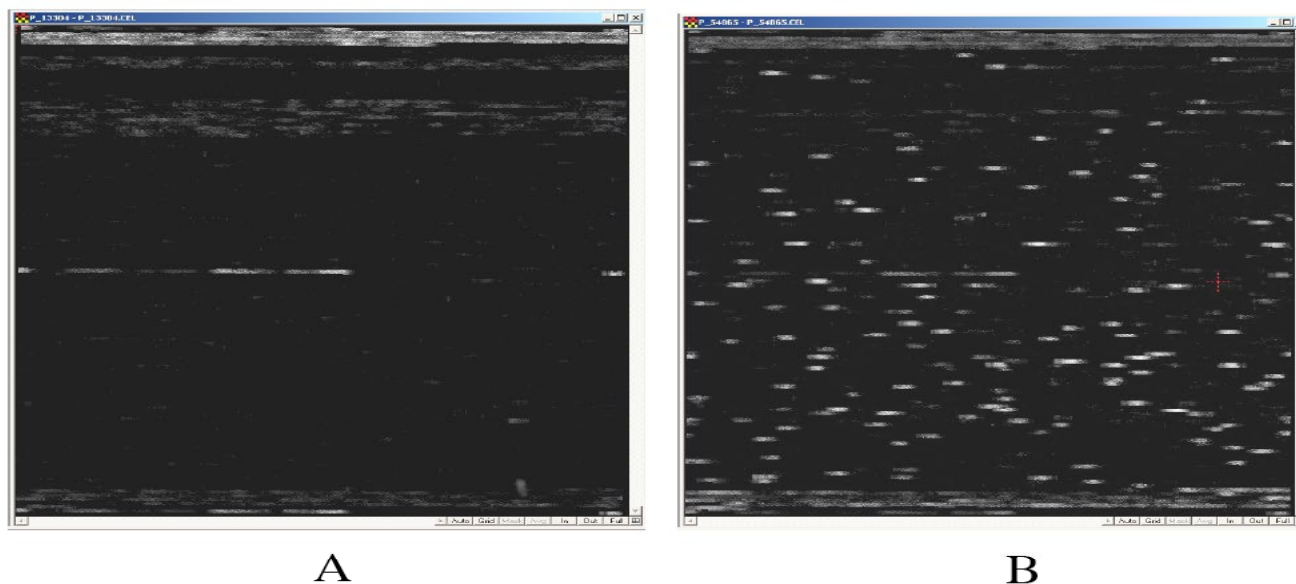


Fig. 14—Images of RPM v.1.Dat for the same clinical specimen (MG88506) that was positive for Influenza A using the standard protocol (A) and UberMix PCR protocol (B)

Another potential problem not considered when the chip was designed only for use with random primers was false positives arising from primers binding to the internal sequences of any one of the tiles. When the primers used were carefully evaluated, several primer sequences with potential to bind non-

specifically to internal tile sequences (particularly in adenovirus tiles) were identified. These regions of the tiles were eliminated from consideration in the BLAST analysis.

During the last two weeks of SGD, throat swabs were collected in addition to nasal washes. Based on difficulties in getting consent to perform both procedures on a single patient and the low numbers of protocol-qualified patients presenting to the MTFs, only 33 throat swab samples were obtained. Of these samples, the culture reported 6 samples as positive for Flu A and 4 for Flu B. RPM results showed no positives for any pathogen. This data was not sufficient to determine the utility of throat swabs, but suggests that the RPM maybe capable of detecting certain pathogens using throat swabs. A more complete study of collection and sample processing is required and will be part of future research.

Influenza A Detection

Influenza A virus (Flu A) was the most commonly identified respiratory pathogen by all methods. Culture methods (30–42 days) revealed that 184 (41%) of the 450 samples (nasal wash and throat cultures) sent for culture were positive for Flu A. Of the 417 nasal wash samples cultured, 79 (19%) were reported to COHORT as positive for Flu A when using the standard protocol (random amplification without DNA subtraction followed by cutoff of BLAST results with expect values of $>1.0e^{-9}$). For nasal wash samples only, 30-day culture methods revealed that 178 (43%) of the 417 samples were positive for Flu A; concordance between culture and RPM results was observed with 61 samples. An additional 18 were reported to COHORT as positive for Flu A by RPM v.1, but were reported as negative by culture, possibly due to loss of viability of virus in samples during shipment.

For standard comparisons of RPM results to those obtained with culture, simple ratios may be sufficient. However, draft FDA guidelines for reporting on diagnostic tests suggest statistical evaluations should be made based on agreement with a standard reference method (<http://www.fda.gov/cdrh/osb/guidance/1428.pdf>). If that reference method is imperfect, the FDA recommends using the simplest measurement of overall percent agreement, defined as the proportion of samples where both the new test and the imperfect standard agree. Formulas for calculations of agreement are as follows:

$$\text{overall percent agreement} = 100\% \times (a+d)/(a+b+c+d)$$

$$\text{agreement of new test with imperfect standard-positive} = 100\% \times a/(a+c)$$

$$\text{agreement of new test with imperfect standard-negative} = 100\% \times d/(b+d)$$

where a = samples positive by both methods, b = RPM positive /culture or PCR negative, c = RPM negative/culture or PCR positive and d = negative by both methods. Data for detection of influenza A by both the RPM and the standard reference method is shown in Table 4, using the standard 2×2 table format. Based on these values, the overall percent agreement for SG samples tested by both methods was 70%. Additional determinations of agreement for positive and negative samples showed 33% positive agreement and 93% negative agreement.

Table 4—Agreement of Reference Assay and RPM v.1 for Influenza A

Genus Level	Culture or PCR Positive	Culture or PCR Negative	Grand Total
RPM positive	61 (a)	18 (b)	79
RPM negative	117 (c)	221 (d)	338
Total	178 (a+c)	239 (b+d)	417

* Note: the 33 throat swab samples obtained during the last two weeks of SGD were not included in these calculations

To examine methods for improving the sensitivity of the RPM assay, the unused fraction of processed RNA and DNA aliquots (from Lab B) for 464 of the patient samples were amplified using UberMix. Also, the sequence calls made by the RPM in these and the previous experiments (using the standard algorithm) were re-analyzed using “relaxed” conditions, in which targets were identified by associating the top sequence similarity ranking (uniquely high BLAST bit score with corresponding low expect value) with single pathogen genome database record(s) instead of satisfying an arbitrary BLAST cut-off expect value of $1.0e^{-9}$. In all cases, the criterion for a positive by RPM analysis was that sequences from one or more of the pathogen tiles on the RPM would return a BLAST result that was uniquely identifiable with records for a pathogen species. When similar calculations were done using the “relaxed” data, FDA analysis method indicated 70% overall percent agreement with 66% agreement for positive analyses and 92% agreement for negative analyses, suggesting that improvements in the RPM assays and analytical software will be required to meet the clinical thresholds in the future.

Figure 15 shows the cumulative culture and RPM results obtained for 351 “matched” experiments for which there was data for culture, random amplification, and multiplexed amplification as analyzed with “relaxed” conditions. Of the 351 specimens, 30-day culture detected Flu A in 147 (42%) of the specimens. The RPM detected Flu A in 73 (21%) with random amplification and 88 (25%) with multiplex (UberMix) to produce 114 unique positive results for 32% of total and 76% the number of 30-day culture positives. Using the FDA formula, the comparison between the RPM v.1 (random and multiplexed) results and the reference assays shows 77% overall agreement with 61% positive agreement and 88% negative agreement. Thus, two uncharacterized and compromised RPM amplification protocols were combined during the demonstration to provide for approximately the same level of Flu A detection as was anticipated by the random amplification alone without interference of DNA background on the RPM. In all cases, influenza A sequences were immediately observed to be consistent with type H3N2 database records, and H5 (avian influenza virus) was ruled out prior to the release of Flu A positive results

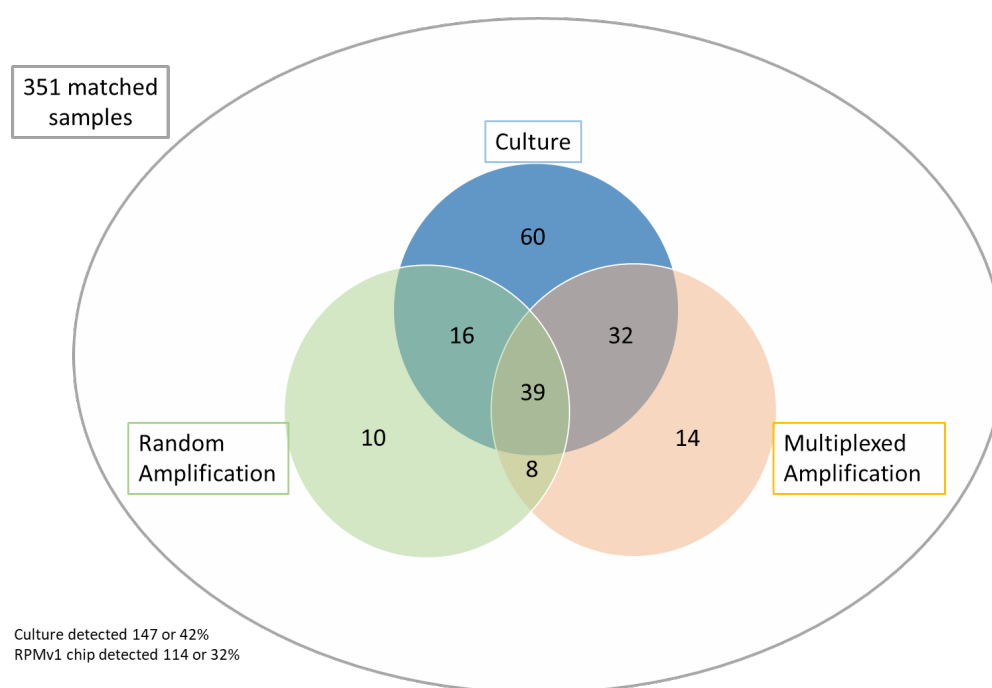


Fig. 15—Cumulative Influenza A virus culture and RPM v.1 results

Influenza B Detection

Figure 16 shows the cumulative culture and RPM results obtained for 351 “matched” experiments for which there was data for culture, random amplification, and multiplexed amplification as analyzed with “relaxed” conditions. Of the 351 specimens, 30-day culture detected Flu B in 17 (4.8%) of the specimens. The RPM detected influenza B (Flu B) in 11 (3.1%) with random amplification and 12 (3.4%) with multiplex (UberMix) to produce 14 unique positive results for 4% of total and 83% of the number of 30-day culture positives. Using the FDA formula to compare RPM v.1 (generic and multiplexed) and reference assay results yields 77% overall agreement with 61% agreement for positive results and 88% agreement for negative results.

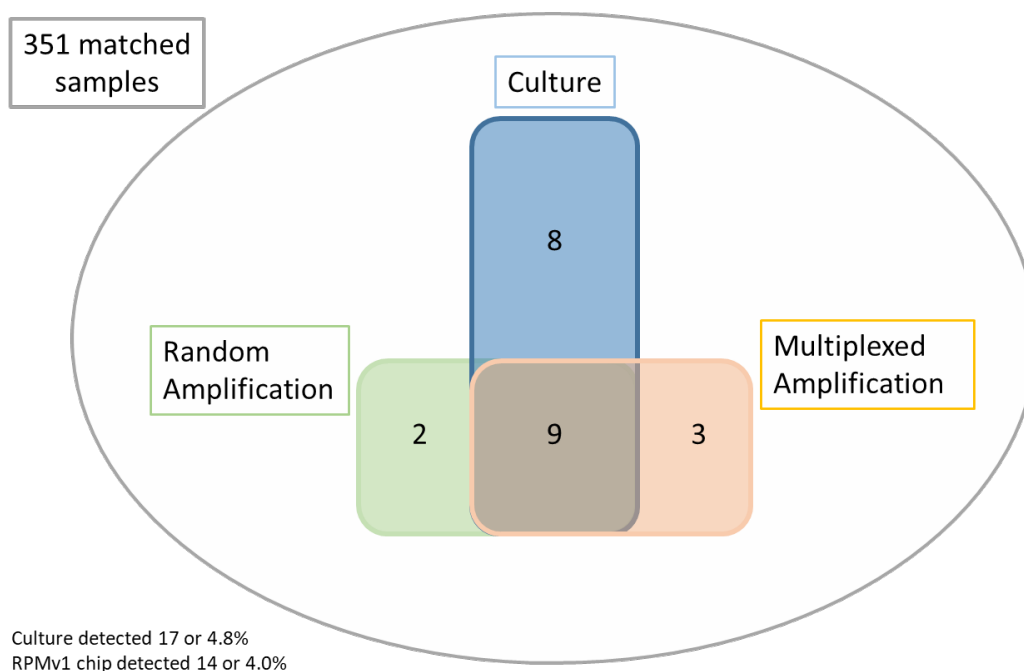


Fig. 16—Cumulative influenza B virus culture and RPM results obtained for 351 "matched" experiments

Influenza A/B Tile Data

Table 5 shows the total number of all influenza A-positive tiles detected on the RPM v.1 during SGD. With one exception for NA1, the only tiles which detected Flu A sequences were for H3 and N2. This observation is consistent with the reported serotyping of influenza A by the Centers for Disease Control for the 2004–2005 Flu season.⁽¹⁰⁾ No sequences from the Flu A HA5 or HA1 (characteristic of Avian influenza) tiles led to significant BLAST results. Table 5 also shows the total number of all influenza B-positive tiles detected on the RPM v.1 during the SGD demonstration. The Flu B tiles detected were highly dependent on the amplification method used.

Table 5—RPM v.1 Tiles Detecting Influenza Sequences

Influenza Tile Type	Random	UberMix
FLU A Tiles		
FluAHA1	0	0
FluAHA3	92	84
FluAHA5	0	0
FluANA1	1	0
FluANA2	89	78
FluAMATRIX	0	0
FLU B Tiles		
FluBHA	1	8
FluBNA	11	3
FluBMATRIX	12	10

Human Coronavirus Detection

Human coronavirus was detected in a small number of samples when using random PCR, but not by culture or using multiplex PCR (UberMix). For coronaviruses this was not unexpected, as they are known to be difficult to culture. However, specific PCR assays for coronavirus yielded 33 positive results for human coronaviruses (229E and OC43). Each of these coronavirus types were represented by a single tile on the RPM.

Other RNA Viruses

The RPM also detected several other RNA viral pathogens including RSVA and RSVB. There was one RSVA-positive sample identified with UberMix relaxed conditions. For RSVB, there was one random relaxed and four UberMix relaxed. None of these samples resulted in positive culture results.

Adenovirus Serotypes and Possible Co-Infections

Human adenovirus was detected in 17 specimens, all using UberMix amplification. Of the 17 specimens that were adenovirus-positive by UberMix, a subset of 12 were cultured, returning one positive adenovirus result. Seven (58%) of the 12 culture results were positive for either influenza A (six samples) or *S. pyogenes* (1 sample); two additional culture-negative samples were positive for influenza A by RPM. The serotypes identified for the 17 adenovirus-positive RPM results are shown in Table 6.

Table 6—Adenovirus Serotypes Determined by RPM and the Presence of other Pathogens as Determined by Culture and RPM

PIN	RPM Serotype	Culture Result	Other RPM
WR97949	Ad5 or Ad1	negative	negative
BO97731	Ad1, Ad2, Ad5 or Ad6	Adenovirus isolated	negative
BO91912	Ad1 or Ad5	N/A	negative
DL13970	Ad1	N/A	negative
BO051266	Ad5	N/A	negative
DL42894	Ad5	Influenza A	negative
MG69217	Ad1 or Ad5	negative	negative
MG89144	Ad5	<i>S. pyogenes</i>	negative
BO00274	Ad1 or Ad2	negative	negative
DL32587	Ad1 or Ad5	Influenza A	Influenza A (Uber)
MG81021	Ad1, Ad5 or Ad6	Influenza A	Influenza A (Uber)
NN17501	Ad1 or Ad5	N/A	Influenza A (Uber)
DL42894	Ad3	Influenza A	negative
WR82344	Ad5	Influenza A	negative

MG20799	Ad5	Influenza A	Influenza A (Uber)
NN04079	Ad1 or Ad5	N/A	negative
MG10889	Ad1	negative	Influenza A (Uber)

Antibiotic Resistance Markers

Three different macrolide-resistance genes were present on the RPM. Two of these, SPYERMB and SPYMEFAE, were detected in 19 SGD specimens when using the UberMix amplification protocol, as shown in Table 7. In 14 of these samples, the RPM also detected *S. pneumoniae* and/or *N. meningitides*. Under antibiotic selective pressure, bacteria have the ability to develop and exchange antibiotic resistance genes. [11] The presence of these antibiotic resistance markers in samples positive for *S. pneumoniae* and/or *N. meningitides* suggested that these two commensal bacteria might be erythromycin resistant, though we did not attempt to perform antimicrobial susceptibility testing on any isolates. In future applications of the microarray technology, this information may be helpful when selecting the appropriate antimicrobial therapy for infection.

Table 7—RPM Detection of Antibiotic Resistance Genes and other Bacterial Pathogens using the UberMix Amplification Protocol. Throat culture samples are denoted with a *. SP-*S. pneumoniae*, NM-*N. meningitides*

PIN	SPYERMB	SPYERMTR	SPYMEFAE	other 1	other 2
MG46951	0	0	1	SP	NM
MG43310	1	0	0		NM
MG62447	1	0	0		
WR90049	1	0	0		NM
NN89191	1	0	1		NM
MG55311	1	0	1	SP	NM
MG54424	1	0	1	SP	NM
WR80276	1	0	0		
MG27403	1	0	1	SP	NM
NN89191	1	0	1	SP	
NN47063	1	0	0		
MG00844	1	0	1	SP	NM
MG51394	1	0	0	SP	NM
MG254TC*	1	0	0	SP	
WR83183	1	0	0		
MG725TC*	1	0	1		NM
MG271TC*	1	0	1	SP	NM
NN42801	1	0	1	SP	
MG485TC*	1	0	0		

Attempt to Correlate Patient Data with Culture Data

The culture data was also evaluated to determine if any clinical signs or symptoms could be used to differentiate between the pathogens. There were no statistically significant differences found between any of the subsets of data based on the pathogen. Figure 17 is representative of those results specifically for patient temperature. This was an expected outcome. Even though there is limited data from pathogens other than Influenza, it is consistent with the clinical findings that it is difficult to ascertain the type of pathogen based on symptoms or signs. Hierarchical clustering was also performed to determine if any of the signs or symptoms correlated with the pathogen culture results. No combination of signs or symptoms clustered based on pathogen identified or not identified.

BT Agents in Clinical Specimens

Of the 598 patient specimens tested (nasal wash and throat swabs), no positives for BT agents were observed using either the standard protocol (random amplification) or UberMix (multiplex PCR amplification). However, positive detection of various agents in nasal washes spiked with BT agents (below, and Chapter 5) showed that the non-optimized UberMix allowed highly discriminatory detection of BT agents at sensitivities similar to those of optimized specific PCR reactions.

Blind *B. anthracis* Spike Tests

To demonstrate the detection capability for BW agents during SGD and test the notification system, 11 test samples containing spikes of killed *B. anthracis* (BA), (1 Ames strain and 10 Sterne vaccine strain) were run between 25 January and 8 February 2005. An initial stock solution of 4×10^7 cfu inactivated Ames spores were put through the RAPIDS bead beater, then spiked into pooled normal nasal wash. Samples were put into the normal patient specimen jars with a special LIMS barcode that appeared to be from an MTF. The samples were given to laboratory personnel responsible for accessing patient samples in Lab A, with no indication that they were not clinical samples, that is, each was processed in normal mode as a blind sample. Once the samples were processed, scanned and analyzed, the results were sent to the GCOS server, which immediately sent a page to the designated senior staff member at NRL for that shift. The staff member notified the laboratory production team lead, who in turn initiated confirmatory testing with the Ruggedized Advanced Pathogen Identification Device™ (RAPID, Idaho Technologies, Inc., Salt Lake City, UT) protocol to verify microarray results. Samples that tested positive for BA in both tests were reported directly to the on-duty AF SGD officer. As shown in Table 8, all 11 of the BA-spiked, blinded samples were identified correctly by the microarray chip (indicated by +) and verified by RAPID. Notification of a positive result for a BA spike was received AF SGD senior staff within 10 minutes of sample analysis. Turnaround times for sample analysis ranged from 17.4 hours to 21.3 hours.

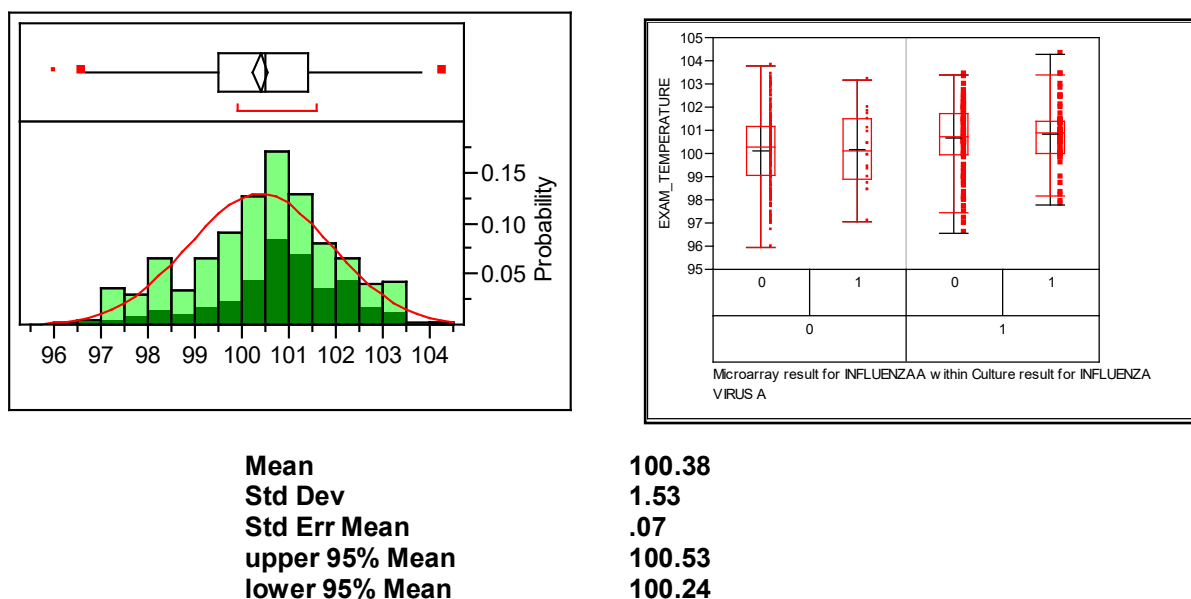


Fig. 17—Distribution of temperature among culture sampled population. The darker green bars indicate Flu A positive by culture. (n=423, 450 total minus 17 deleted due to incomplete vital signs)

Table 8—*B. anthracis* Spiked Samples Run during SGD

PIN #	LIMS #	Date Tested	LIMS Results	BAT Results	Turn Around Time (Hrs)
BO15476	13268	1/25/05	+	+	21:26
BO27854	16662	2/1/05	+	+	18:22
BO31892	17563	2/7/05	+	+	18:14
BO34441	17052	2/3/05	+	+	19:31
BO47428	17078	2/3/05	+	+	19:02
BO57493	17065	2/3/05	+	+	18:49
BO80539	16792	2/2/05	+	+	17:39
BO87048	17550	2/7/05	+	+	17:54
MG29960	17091	2/3/05	+	+	19:16
MG50275	17039	2/3/05	+	+	18:56
MG94867	17537	2/7/05	+	+	18:08

Although the RPM and its supporting sample processing protocols were in an early developmental status at the onset of SGD, a combination of real-time adjustments and improvements enabled a first-ever multiple pathogen surveillance based on genetic sequence. We demonstrated that it was possible to transition a research protocol to a large-scale operation including six military treatment facilities and to train and equip approximately 40 laboratory personnel without experience with the Affymetrix Gene Chip system in 6 weeks. This was the first formal assessment of the technology for broad-spectrum identification of both RNA- and DNA-based pathogens. The operational issues observed in SGD with applying the resequencing DNA microarray technology to multiple pathogen identification are all being addressed. Near-term improvements in RPM content, sample processing, data analysis, and target amplification have already been implemented. A newly designed RPM chip v.2.5, coupled with simpler, optimized protocols and automation (robotics) will provide a highly sensitive system to detect common respiratory pathogens with a high degree of concordance with reference assay microbiological methods requiring days to weeks to perform. This robust platform should have a turnaround time of less than 24 hours from the time of sample receipt to results reporting. Further improvements should lead to significantly shorter results turnaround times (perhaps as much as a factor of 4), FDA approval as a diagnostic system, and pathogen detection at sufficient sensitivity for clinical use.

CHAPTER 5: DETECTION AND DISCRIMINATION OF BIOTHREAT AGENTS AND NEAR NEIGHBORS

Overview

Prior to the SGD, all BT pathogen tiles on the RPM v.1 were determined to be correct by resequencing of the corresponding purified genomic targets. However, no investigation or validation of the RPM or its attendant protocols had been performed related to sensitivity and specificity for sequences from BT targets. The relatively small number of patient samples obtained in the SGD demonstration (see Chapter 4) afforded an excellent opportunity to perform the first systematic assessment of the ability of the RPM v.1 to: (a) detect sequences of pathogens across large concentration ranges, (b) withstand challenge from a battery of genetically similar organisms, (c) simultaneously screen for the presence of inactivated BT agents and operationally significant pathogens in clinical matrices, and (d) assess utility for use in a clinical laboratory setting with routine and surge testing.

In this initial assessment of the RPM v.1 for BT agent detection, we observed that the SGD standard protocol, which used random amplification without prior DNA subtraction, had inadequate sensitivity for most BT agents of interest. To enhance assay sensitivity over that observed with random amplification, two multiplex PCR primer mixes specific to the following targets were also used: 1) BT agents +adenovirus + influenza (A/B) only (BTA mix), or 2) all tiles on the RPM V.1 chip (UberMix) (Appendix B). Due to time constraints, neither of these multiplex approaches was optimized prior to use.

A total of more than 2100 spiked samples and blank controls were tested using random amplification, BTA mix, or UberMix. Using specific primer mixes, we detected *B. anthracis*-specific sequences in inactivated Sterne strain vegetative cells at the lowest concentration tested - 20 cfu and in killed Ames spores at concentrations as low as 6×10^4 cfu per sample. Sequences specific for *F. tularensis* were observed from inactivated *F. tularensis* LVS and biovar novicada strains at concentrations as low as 10 cfu while inactivated Schu4 cells were detected at 10^3 cfu. Plaque sequences were detected in samples spiked with inactivated *Y. pestis* from two sources at 10-300 cfu, depending on the source. Sequences for Vaccinia virus, a simulant for Variola major virus, were detected and identified as Vaccinia at the lowest concentration tested (10^5 pfu) but the true limit of detection will not be determined until additional lower concentrations are tested. Targets were identified by associating the top sequence similarity ranking (uniquely high BLAST bit score with corresponding low expect value) with single pathogen genome database record(s).

Of equal value, we observed experimental anomalies that will drive refinements to the design and processing of subsequent resequencing microarrays. Most notably, *cve* (*Y. pestis*) and *fopA* (*F. tularensis*) tiles allowed for extremely sensitive sequence detection but were also identified unexpectedly in some samples that were not intentionally spiked with those targets (47/1010 for BTA mix, and 6/284 for UberMix). These *fopA*- and *cve*-positive samples accounted for 85% of the total number of misidentifications. A statistically significant proportion of these unexpected results (30 of 62, $P < 0.01$, chi-square corrected for continuity) could be positively correlated to the presence of high concentrations of the misidentified target in an adjacent tube during processing. The results of this initial BT agent assessment are of obvious benefit to the refinement of chip design, selection of primers for random and multiplex amplification, and automation.

These initial results suggest that a suitably modified RPM resequencing microarray, coupled with straightforward modifications to amplification protocols and sample preparations, should provide an excellent technology for the simultaneous detection of naturally occurring and BT pathogens in a clinical environment.

Background

In addition to tiles specific for common respiratory pathogens, the RPM v.1 chips were designed with tiles specific for several CDC Category A BT pathogens: *B. anthracis* (3 tiles), variola major virus (2 tiles), Ebola virus (1 tile), Lassa fever virus (1 tile), *F. tularensis* (2 tiles), and *Y. pestis* (2 tiles). Although particulates are rapidly cleared from the nasal cavity, nasal swabs were the preferred diagnostic sample for initial exposure to aerosolized infectious or intoxicating agents at that time. [12] To determine whether the RPM V.1 chips could be used to detect sequences specific to these BT agents in such clinical samples, nasal wash samples were spiked with varying concentrations of inactivated agents and simulants, then analyzed using the SGD random amplification protocol, as well as two modified amplification procedures.

As described in earlier chapters, random amplification of DNA and RNA species results in large numbers of different amplicons, without large amplification bias in favor of a specific target. The inability to perform a satisfactory DNA subtraction step resulted in increased background noise on the RPM v.1, thereby potentially decreasing the ability of the SGD chip to detect relevant but low abundance sequences. Analyses of spiked nasal washes were performed using random amplification as a baseline for comparison with additional amplification procedures. Specific primers were added to the PCR mixes, such that sequences specific to the tiled regions on the SGD RPM v.1 chip would be selectively amplified. Two mixes of primers were used: UberMix, containing primers specific for *all* tiled regions of the chip (130 different primers); and BT Agent mix (BTA mix, Table 9), which contained degenerate as well as specific primers for adenovirus and influenza, and specific primers for the six BT agents represented on the RPM v.1 chip. The potential for false-positive results from use of specific primers was addressed using closely related species, with similar or identical nucleic acid sequences. These “near neighbors”, as well as several non-relevant species, were also used to determine the ability of the proposed technology to discriminate between different species and strains. Furthermore, several of the test organisms were obtained from different sources to determine the effect of inactivation protocols and source material (Table 10).

Table 9—Components of BTA PCR Primer Mixes

Target	Tiled Region (marker)	Direction of primer	Primer Sequence
<i>B. anthracis</i>	<i>capB</i>	Forward	5'-CTG ACC AAT CTA AGC CTG C
		Reverse	5'-TCG TTT CTC CAA TCG CAA T
	<i>pag</i>	Forward	5'-ATC CTA GTG ATC CAT TAG AAA CGA C
		Reverse	5'-CTT CTC TAT GAG CCT CCT TAA CTA CTG
	<i>rpoB</i>	Forward	5'-CCA CCA ACA GTA GAA AAT GCC
		Reverse	5'-AAA TTT CAC CAG TTT CTG GAT CT
<i>F. tularensis</i>	<i>fopA</i>	Forward	5'-AAC AAT GGC ACC TAG TAA TAT TTC TGG
		Reverse	5'-CCA CCA AAG AAC CAT GTT AAA CC
	<i>tul4</i>	Forward	5'-GCT GTA TCA TCA TTT AAT AAA CTG CTG
		Reverse	5'-TTG GGA AGC TTG TAT CAT GGC ACT
<i>Y. pestis</i>	<i>cafI</i>	Forward	5'-CAG TTC CGT TAT CGC CAT TGC
		Reverse	5'-TAT TGG TTA GAT ACG GTT ACG GT
	<i>cve</i>	Forward	5'-GGA TAA CGT TGC AGC AGC TTC G
		Reverse	5'-CCT TCG CCA CCT TCA CCT GC
Variola	<i>crmB</i>	Forward	5'-GCT TCC AGA TTA TGT GAT AGC AAG ACT A
		Reverse	5'-TCC GGA TAC TCC GTA TCC TAT TCC
	J7R (HA)	Forward	5'-TAA ATC ATT GAC TGC TAA
		Reverse	5'-GTA GAT GGT TCA TTA TCA TTG TG
Ebola virus		Forward	5'-ATC GGA ACT TTT CTT TCT CAT T
		Reverse	5'-ATG TGG TGG ATT ATA ATA ATC ACT GAC ATG
Lassa virus	GPC	Forward	5'-ACC GGG GAT CCT AGG CAT TT
		Reverse	5'-ATA TAA TGA TGA CTG TTG TTC TTT GTG CA
Adenovirus	E1A Fiber Hexon		All primers listed in Appendix D (UberMix) for generic and specific detection of adenovirus
FluA, FluB			All primers listed in Appendix D (UberMix) for generic and specific detection of Flu A and Flu B
<i>A. thaliana</i> (positive controls)	DNA control	Forward	5'-CAA CAT GGG AAG CTG TTT TG
		Reverse	5'-CAA GCA CAC GTT ATT TCC CC
	RNA control	Forward	5'-TCA AAT CCT CGT TGA CAG AC
		Reverse	5'-CTG TTG CCT CCA TTG ACA GA

Table 10—Target Bacteria and Viruses Spiked into Pooled Nasal Washes

Target	Source	Relevant genotype	Method of inactivation
<i>B. anthracis</i> Ames (spores)	DoD Critical Reagent Program (CRP) lot # AGD0000511	pXO1 ⁺ ; pXO2 ⁺	4.1 Mrad/125 min.
<i>B. anthracis</i> Sterne (vegetative cell prep)	Nav. Med. Res. Cmd. Bethesda, MD	pXO1 ⁺ ; pXO2 ⁻	0.01% Na Azide
<i>B. cereus</i> PT2	Tetracore, Gaithersburg, MD	pXO1 ⁻ ; pXO2 ⁻	Epicure™ enzyme digestion
<i>B. cereus</i> 03BB102	Tetracore	pXO1 ⁺ (truncated, <i>pagA</i> ⁺); pXO2 ⁺ (truncated, <i>capA</i> ⁺) ¹	Epicure™ enzyme digestion
<i>B. cereus</i> ATCC 14579	Tetracore	Type strain	Epicure™ enzyme digestion
<i>B. globigii</i> (spore)	Dugway Proving Ground, Dugway, UT		Live spores ²
<i>B. thuringiensis</i> Al Hakam	Tetracore		Epicure™ enzyme digestion
<i>B. thuringiensis</i> var. Kurstaki spores	Abbott Laboratories, (Lot # 04-355-V9)		Live spores ² (Dipel Tech Powder)
<i>E. coli</i> ATCC 35218 ³	American Type Culture Collection, Manassas, VA		0.01% Na Azide
<i>F. tularensis</i> Schu4	CRP lot # AGD0000056	pOM1 ⁻ ; pNFL10 ⁻	4.1 Mrads/125 min
<i>F. tularensis</i> LVS	Tetracore	pOM1 ⁺ ; pNFL10 ⁺	Epicure™ enzyme digestion
<i>F. tularensis</i> biovar novicida	Tetracore	pOM1 ⁺ ; pNFL10 ⁺	Epicure™ enzyme digestion
<i>Vaccinia</i> Lister Elstree	CRP lot #AGD0000119		4.1 Mrads/125 min
<i>Y. pestis</i> CO-92	CRP lot #AGD0000520	pMT1 ⁺ ; pCD1 ⁺ ; pPCP1 ⁺	4.1 Mrad/125 min.
<i>Y. pestis</i> CO-92	Tetracore	pMT1 ⁺ ; pCD1 ⁺ ; pPCP1 ⁺	Epicure™ enzyme digestion
<i>Yersinia</i> sp. MX018	Tetracore	pPCP1 ⁺	Epicure™ enzyme digestion
<i>Y. ruckeri</i>	Tetracore		Epicure™ enzyme digestion

¹Although the original source identified a truncated pXO2 plasmid and *capB*⁺ genotype in this strain, personnel at Tetracore were unable to detect this plasmid or *capB* by PCR (Dr. William Nelson, personal communication)

²Disrupted using “bead-beater” module of BAT protocol for RT-PCR validation

³Grown overnight in Luria Broth and collected by centrifugation. After two washes in PBS, cells were resuspended in PBS to a concentration of approximately 2.3×10⁹ cells/ml

Experimental Design and Methods

Nasal washes from healthy informed volunteers were pooled and spiked with known amounts of inactivated bacterial and viral analytes (Table 10); two non-hazardous *Bacillus* spp. strains that can be handled under biosafety level-1—*B. globigii* and *B. thuringiensis* Kurstaki—were live preparations. After spiking, each sample was processed using the same procedures as the patient samples including total nucleic acid extraction (Lab A), separation of pathogen RNA from human RNA components with inclusion of primer sequences (Lab B), PCR amplification (Lab C1), and chip analysis (Lab C2) using the standard SGD (random amplification) protocol. Prior to amplification (Lab C1), each sample was split; half of each sample was carried through Lab C1 and Lab C2 using the standard (random amplification) protocol. The remaining split was frozen for later analysis using a different (multiplex PCR) primer mix. Because variations in extraction and preparation efficiencies were controlled, use of these “splits” enabled direct comparisons between results obtained with the different PCR approaches and primer mixes. As an additional control, a subset of samples was prepared in duplicate and then extracted, purified, and

analyzed on different days using the same primer mix. All concentrations shown are for an initial specimen volume of 150 μ L.

Several *Bacillus* species (*B. anthracis* Ames, *B. globigii*, *B. thuringiensis*) were provided in spore form and were diluted 2-fold in lysis buffer and disrupted in the “bead beater” module of the BAT protocol for real-time PCR validation. [3] Disrupted samples were then spiked into aliquots of the pooled nasal wash.

Purified nucleic acids from several targets were also spiked into total nucleic acid mixtures previously purified in Lab A (Table 11). These last controls assessed the effect of nasal wash on purification of nucleic acids from spiked bacteria. All sample and data manipulations were performed blind; the sample identities were not provided to any personnel until after a final determination of positive/negative. In each case, the data were analyzed using only the “relaxed” algorithm. That is, BLAST returns were not subjected to the $1.0e^{-9}$ expect value threshold prescribed in the standard SGD protocol. Rather, the database record having the highest BLAST bit score was referenced to a species look-up table, as described in the previous chapter.

Table 11—Purified Nucleic Acids from BT Agents and Near Neighbors, Spiked into Total Nucleic Acids Prepared from Nasal Washes

Species	Source	Relevant genotype
<i>F. tularensis</i> biovar novicida	Tetracore	pOM1 ⁺ ; pNFL10 ⁺
<i>Y. pestis</i> CO-92	Tetracore	pMT1 ⁺ , pCD1 ⁺ , pPCP1 ⁺
<i>Y. pseudotuberculosis</i>	Tetracore	pPCP1 ⁺ ; pCD1 ⁺
<i>Yersinia</i> spp. MX018	Tetracore	pPCP1 ⁺

Results and Discussion

Detection of *B. anthracis* Sequences

The RPM V.1 chips contained three tiles designed to detect sequences from *B. anthracis*: a *capB*-specific tile to detect pXO2, a *pag*-specific tile to detect the protective antigen gene on pXO1, and an *rpoB*-specific tile for the *rpoB* marker on the chromosome. Thus, the three *B. anthracis* tiles were designed to distinguish between sequences from *Bacillus* species and strains based on the presence of the plasmids and the sequence of the tiled *rpoB* region.

Table 12 shows the overall results for the detection of *B. anthracis* sequences, based on the detection of one or more of the three marker sequences. A significant proportion of samples spiked with Ames spores and Sterne cells were positively identified as containing *B. anthracis* sequences at 10^6 and 2000 cfu, respectively using the standard SGD protocol. The difference in sensitivities may have been related to a variety of differences, ranging from differences in numbers of non-viable cells to a lack of optimization of the Ames spore disruption using the mini-bead beater prior to the project. [3] In all cases, however, inclusion of specific primers for the *B. anthracis* markers improved assay sensitivity. Samples containing *B. anthracis* Sterne were positively identified as containing *B. anthracis* sequences when only 2×10^2 cfu were present when amplified with BTA mix.

Table 12—Summary of Results with Spiked (Inactivated) *B. anthracis* Ames and Sterne

Sample spike (cfu)	Random Amplification	BTA mix	UberMix
<i>B. anthracis</i> Ames*			
3×10 ²	0/5	0/10	0/4
1×10 ³	0/4	0/9	0/5
3×10 ³	0/4	1/8	1/5
1×10 ⁴	0/4	0/9	0/5
3×10 ⁴	1/4	2/8	0/5
6×10 ⁴	0/4	2/2	-
1-3×10 ⁵	0/15	15/20	9/10
6×10 ⁵	1/5	4/4	-
1-3×10 ⁶	6/15	18/19	8/9
6-8×10 ⁶	8/9	14/15	4/4
2×10 ⁷	1/1	1/1	-
<i>B. anthracis</i> Sterne			
2 × 10 ¹	1/5	1/4	Not done
2 × 10 ²	0/5	5/5	
7 × 10 ²	3/5	3/4	
2 × 10 ³	4/5	4/4	
7 × 10 ³	5/5	4/4	
2 × 10 ⁴	5/5	5/5	
7 × 10 ⁴	5/5	3/3	
2 × 10 ⁵	2/2	2/2	

*pooled data from multiple series

As expected, samples spiked with Ames strain *B. anthracis* were positive for all three-marker sequences; this strain contains both pXO1 and pXO2 (Table 13). In each case where the chromosomal marker sequence was detected on the resequencing array, a species-level sequence identification was made; however, the pathogenic nature of *B. anthracis* Ames was further confirmed with each positive detection of *pag* and *capB* sequences. None of the samples spiked with Sterne strain *B. anthracis* showed positive calls for the sequence from *capB*, which is consistent with the lack of pXO2 in this strain; however, the chromosomal marker (*rpoB*) and pXO1 marker sequences were detected and were identified as *Bacillus* sp. and *B. anthracis* sequences, respectively. None of the three markers tested (*capB*, *pag*, *rpoB*) proved to be more sensitive than the others when low concentrations of *B. anthracis* were analyzed. These results are somewhat surprising, since high copy numbers for pXO1 (marker = *capB*) have been documented for *B. anthracis* Ames. [13, 14]

Table 13—Detection of *B. anthracis* Sequences on Specific Tiles from Spiked *B. anthracis* Ames and Sterne

Sample spike (cfu)	Random Amplification			BTA mix			UberMix		
Marker	<i>capB</i>	<i>pag</i>	<i>rpoB</i>	<i>capB</i>	<i>pag</i>	<i>rpoB</i>	<i>capB</i>	<i>pag</i>	<i>rpoB</i>
<i>B. anthracis</i> Ames*									
3×10^2	0/5	0/5	0/5	0/10	0/10	0/10	0/4	0/4	0/4
1×10^3	0/4	0/4	0/4	0/9	0/9	0/9	0/5	0/5	0/5
3×10^3	0/4	0/4	0/4	1/8	2/8	1/8	1/5	1/5	1/5
1×10^4	0/4	0/4	0/4	0/9	0/9	0/9	0/5	0/5	0/5
3×10^4	1/4	0/4	0/4	1/8	2/8	1/8	0/5	0/5	0/5
6×10^4	0/4	0/5	0/5	2/2	2/2	2/2	-	-	-
$1-3 \times 10^5$	0/15	0/15	0/15	14/20	14/20	14/20	5/10	9/10	9/10
6×10^5	1/5	1/5	0/5	3/4	4/4	4/4	-	-	-
$1-3 \times 10^6$	5/15	3/15	3/15	18/19	14/19	17/19	8/9	8/9	7/9
$6-8 \times 10^6$	7/9	6/9	4/9	14/15	14/15	14/15	4/4	4/4	4/4
2×10^7	1/1	1/1	1/1	1/1	1/1	1/1	-	-	-
<i>B. anthracis</i> Sterne									
20	0/5	1/5	1/5	0/4	1/4	1/4	Not done		
2×10^2	0/5	0/5	0/5	0/5	5/5	5/5			
7×10^2	0/5	3/5	1/5	0/4	3/4	3/4			
2×10^3	0/5	4/5	3/5	0/4	4/4	3/4			
7×10^3	0/5	5/5	5/5	0/4	4/4	4/4			
2×10^4	0/5	5/5	4/5	0/5	5/5	5/5			
7×10^4	0/5	4/5	5/5	0/3	3/3	2/3			
2×10^5	0/2	2/2	2/2	0/2	2/2	2/2			

*pooled data from multiple series

Several “near neighbors” of *B. anthracis* were tested for cross-reactivity and the ability of the SGD protocols and chip to detect the presence of related sequences. Only random amplification and BTA mix-based amplification were used for analysis of these samples; UberMix was not used. The near neighbors included *B. globigii* and *B. thuringiensis*, which are both common Department of Defense simulants for gram-positive spore formers. “Anthrax-like” *B. cereus* isolates, some containing fragments of *B. anthracis* plasmids pXO1 and pXO2, were also tested. These species and strains were specifically included in this study, as a high degree of homology between *B. anthracis*, *B. thuringiensis*, and *B. cereus* has been observed. [15, 16]

Of the 93 samples spiked with *B. globigii* or *B. cereus* (183 total analyses), three samples returned positive BLAST results that were positive for *B. anthracis* sequences based on database record retrieval (Table 14). All three were spiked with a high concentration of *B. cereus* 03BB102, a clinical isolate taken from a patient with anthrax-like respiratory illness. This strain possesses truncated forms of both pXO1 and pXO2 and therefore, it is possible that the high concentration required for positive detection of the plasmid sequences, 10^8 cfu, was above the threshold for detection of these potentially truncated sequences. Also, no chromosomal (*rpoB*) markers for *B. anthracis* were detected. A single positive sample spiked with *B. cereus* PT2 that returned a BLAST result did not allow distinction (equivalent BLAST bit score and expect values) between *B. anthracis*, *B. cereus*, or *B. thuringiensis* sequences and hence would not be flagged as a positive result for *B. anthracis* sequences.

Table 14—Detection of *B. anthracis* Sequences (Specific Tiles and Overall Calls) from Spiked *B. cereus* Strains and *B. globigii*

Sample spike (cfu)	Random Amplification			BTA mix			Overall Call	
Marker	<i>capB</i>	<i>pag</i>	<i>rpoB</i>	<i>capB</i>	<i>pag</i>	<i>rpoB</i>	Random	BTA mix
<i>B. cereus</i> 14579								
10	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ²	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ³	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁴	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2
10 ⁵	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2
10 ⁶	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁷	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁸	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2
<i>B. cereus</i> PT2								
10	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ²	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ³	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁴	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁵	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁶	0/2	0/2	0/2	0/3	0/3	0/3	0/2	0/3
10 ⁷	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁸	0/2	0/2	0/2	0/3	0/3	1/3*	0/2	1/3*
<i>B. cereus</i> 03BB102								
10	0/2	0/2	0/2	0/3	0/3	0/3	0/2	0/3
10 ²	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ³	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁴	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁵	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁶	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁷	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁸	0/2	0/2	0/2	3/3	1/3	0/3	0/2	3/3
<i>B. globigii</i>								
1.5 × 10 ⁵	0/5	0/5	0/5	0/4	0/4	0/4	0/5	0/4
5 × 10 ⁵	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4
1.5 × 10 ⁶	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5
5 × 10 ⁶	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5
1.5 × 10 ⁷	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5
5 × 10 ⁷	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1

*BLAST results showed this sample contained either *B. anthracis*, *B. cereus*, or *B. thuringiensis*; therefore, species-specific identification could not be made.

Two strains of *B. thuringiensis* were also tested as Near Neighbors (Table 15). The chromosomal *rpoB* sequence was deemed positive for a number of samples containing *B. thuringiensis* Al Hakam. However, in all of these samples, BLAST results showed best fit for *B. thuringiensis* sequences or could not determine the species of *Bacillus* to which the sequence was attributable. In the latter case, a best-fit analysis showed the sample sequence was from one of three species: *B. anthracis*, *B. cereus*, or *B. thuringiensis*. Furthermore, in samples containing *B. thuringiensis* Kurstaki where *rpoB* was the only target sequence detected, no unequivocal species identification could be made. These results were not surprising - between 92 and 99% pairwise identity has been observed in 9 different chromosomal markers between this strain and *B. anthracis*. [17, 18] A longer tile sequence on future resequencing arrays, or tiling of a more variable region of the gene, should allow a finer degree of resolution between sequences from such closely related pathogens.

Table 15—Detection of sequences on *B. anthracis* tiles (specific and overall calls) from spiked *B. thuringiensis* strains

Sample spike(cfu)	Random Amplification			BTA mix			Overall Call	
Marker	<i>capB</i>	<i>pag</i>	<i>rpoB</i>	<i>capB</i>	<i>pag</i>	<i>rpoB</i>	Random	BTA mix
<i>B. thuringiensis</i> Kurstaki								
4×10^4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4
1.2×10^5	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4
4×10^5	0/5	1/5	0/5	0/4	0/4	0/4	1/5	0/4
1.2×10^6	0/5	0/5	0/5	0/5	0/5	4/5*	0/5	4/5*
4×10^6	0/5	0/5	0/5	0/5	0/5	3/5*	0/5	3/5*
1.2×10^7	0/5	0/5	4/5*	1/5	1/5	5/5*	4/5*	5/5*
4×10^7	0/2	0/2	2/2*	0/2	0/2	2/2*	2/2*	2/2*
<i>B. thuringiensis</i> Al Hakam crude cell prep								
10	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10^2	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10^3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10^4	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10^5	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10^6	0/2	0/2	0/2	0/3	0/3	0/3	0/2	0/3
10^7	0/2	0/2	0/2	0/3	0/3	0/3	0/2	0/3
10^8	0/1	0/1	0/1	0/3	0/3	1/3*	0/1	1/3*
<i>B. thuringiensis</i> Al Hakam purified DNA								
10	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2
10^2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2
10^3	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2
10^4	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2
10^5	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
10^6	0/2	0/2	0/2	0/2	0/2	2/2*	0/2	2/2*
10^7	0/3	0/3	3/3*	0/3	0/3	3/3*	3/3*	3/3*
10^8	0/3	0/3	3/3*	0/3	0/3	3/3*	3/3*	3/3*

*BLAST results showed best-fit for *B. thuringiensis* alone, or *B. anthracis*, *B. cereus*, or *B. thuringiensis*; therefore, species-specific identification could not be made.

Two samples spiked with *B. thuringiensis* var Kurstaki showed positive results for at least one plasmid marker sequence found in *B. anthracis*; several additional samples also had inconclusive results from the *rpoB* tile. Although substantial conservation of both pXO1 and pXO2 sequences has been observed among *B. thuringiensis* strains [19, 20], neither *pagA* nor *capB* is among the homologous sequences shared. Therefore, positive results for the plasmid genes in these two samples cannot be explained at this time. However, the 13 other samples returning ambiguous *rpoB* calls would not be flagged as containing *B. anthracis*.

Purified DNA from *B. thuringiensis* Al Hakam was spiked into total nucleic acid extract, the product of Lab A, and carried through the rest of the protocol in parallel with crude cell preparations spiked into nasal wash previously extracted in Lab A. These samples were included to examine the potential losses resulting from procedures performed in Lab A. As discussed above, samples were identified as containing *rpoB* sequences from either *B. thuringiensis*, *B. cereus*, or *B. anthracis*. Detection of these genus-specific sequences was more sensitive in the samples with DNA spiked after Lab A (versus crude cell prep spiked into pooled nasal wash), indicating some losses occurred in Lab A or in prior processing steps. It is further possible that a sub-population of spores in the crude cell preparation was present, effectively lowering the number of genome equivalents available for extraction and analysis.

Detection of *F. tularensis* Sequences

The RPM V.1 chip had two tiles specific for chromosomal markers of *F. tularensis*: *fopA*, which encodes an outer membrane protein, and *tul4*, which encodes a 17 kDa outer membrane lipoprotein. These markers have been used in other PCR-based detection assays and have shown high specificity for *F. tularensis*. [15, 21-25] The latter marker is conserved among all tested strains of *F. tularensis*, and can be used to distinguish between *F. tularensis* and *F. philomiragia*. [25, 26]

Table 16 shows concentration-dependent detection of one or both *F. tularensis* target sequences for different strains of the pathogen spiked into nasal wash. We were unable to detect *F. tularensis* sequences at concentrations below 10^6 cfu using random amplification. However, inclusion of specific primers in the PCR reaction (BTA primer mix, UberMix) resulted in detection of *F. tularensis* sequences at much lower levels in all strains tested.

Table 16— Summary of results from spiked *F. tularensis* Schu4 and LVS

Sample spike(cfu)	Random Amplification	BTA mix	UberMix
<i>F. tularensis</i> Schu4*			
1 × 10 ²	0/5	0/6	0/4
3 × 10 ²	0/5	0/7	1/5
1 × 10 ³	0/5	4/9	0/5
3 × 10 ³	0/4	4/5	1/5
1 × 10 ⁴	0/4	6/7	2/5
3 × 10 ⁴	0/5	8/9	4/5
1 × 10 ⁵	0/8	5/11	2/5
3 × 10 ⁵	0/10	6/11	1/5
1 × 10 ⁶	0/10	10/12	3/5
2-3 × 10 ⁶	0/9	12/13	2/5
1 × 10 ⁷	0/2	2/2	-
<i>F. tularensis</i> LVS			
10	0/3	3/3	Not done
10 ²	0/2	2/2	
10 ³	0/3	3/3	
10 ⁴	0/3	3/3	
10 ⁵	0/3	3/3	
10 ⁶	2/2	3/3	
10 ⁷	3/3	3/3	
10 ⁸	3/3	2/3	
<i>F. tularensis</i> var. novicida crude cell prep			
10	0/2	3/3	Not done
10 ²	0/3	2/3	
10 ³	0/3	2/3	
10 ⁴	0/2	3/3	
10 ⁵	0/3	3/3	
10 ⁶	3/3	3/3	
10 ⁷	3/3	3/3	
10 ⁸	3/3	2/3	
<i>F. tularensis</i> var. novicida purified DNA			
10	0/2	1/2	Not done
10 ²	0/2	2/2	
10 ³	0/2	1/2	
10 ⁴	0/2	2/2	
10 ⁵	0/3	2/3	
10 ⁶	2/3	3/3	
10 ⁷	3/3	3/3	
10 ⁸	3/3	3/3	

*pooled data from multiple series

Table 17 shows results taken from the two *F. tularensis*-specific tiles: *fopA* and *tul4*. Although both tiles are specific to chromosomal markers, the *fopA* tile detected *F. tularensis* sequences at lower levels of spiked target, independent of primer mix or strain. The difference in sensitivities varied from approximately 10-fold (LVS, novicida) to as much as 100-fold (Schu4). This difference may be due to both the length of the tile and amplification efficiency of the primers. A similar phenomenon was observed with *Y. pestis* (cf. Table 19). Inactivated *F. philomiragia*, the other species in this genus and an opportunistic pathogen, was spiked into pooled nasal wash samples as a “near neighbor” species (Table 18). Although 20 of the 23 *F. philomiragia* samples analyzed were negative for *F. tularensis* sequences, several samples had positive results in the *fopA* tile when amplified with BTA mix.

Table 16—Detection of *F. tularensis* Sequences on Specific Tiles from Spiked *F. tularensis* Schu4 and LVS

Sample spike (cfu)	Random Amplification		BTA mix		UberMix	
Marker	<i>fopA</i>	<i>tul4</i>	<i>fopA</i>	<i>tul4</i>	<i>fopA</i>	<i>tul4</i>
<i>F. tularensis</i> Schu4						
1 × 10 ²	0/5	0/5	0/6	0/6	0/4	0/4
3 × 10 ²	0/5	0/5	0/7	0/7	1/5	0/5
1 × 10 ³	0/5	0/5	4/9	0/9	0/5	0/5
3 × 10 ³	0/4	0/4	4/5	0/5	1/5	0/5
1 × 10 ⁴	0/4	0/4	6/7	0/7	2/5	0/5
3 × 10 ⁴	0/5	0/5	8/9	0/9	4/5	0/5
1 × 10 ⁵	0/8	0/8	5/11	0/11	2/5	0/5
3 × 10 ⁵	0/10	0/10	6/11	0/11	1/5	0/5
1 × 10 ⁶	0/10	0/10	10/12	0/12	3/5	0/5
2-3 × 10 ⁶	0/9	0/9	12/13	4/13	2/5	1/5
1 × 10 ⁷	0/2	0/2	2/2	1/2	-	-
<i>F. tularensis</i> LVS						
10	0/3	0/3	3/3	0/3	Not done	
10 ²	0/2	0/2	2/2	0/3		
10 ³	0/3	0/3	3/3	0/3		
10 ⁴	0/3	0/3	3/3	0/3		
10 ⁵	0/3	0/3	3/3	1/3		
10 ⁶	2/2	0/2	3/3	3/3		
10 ⁷	3/3	3/3	3/3	2/3		
10 ⁸	3/3	3/3	2/3	3/3		
<i>F. tularensis</i> biovar novicida						
10	0/2	0/2	3/3	0/3	Not done	
10 ²	0/3	0/3	2/3	1/3		
10 ³	0/3	0/3	2/3	0/3		
10 ⁴	0/2	0/2	3/3	1/3		
10 ⁵	0/3	0/3	3/3	1/3		
10 ⁶	3/3	0/3	3/3	1/3		
10 ⁷	3/3	2/3	3/3	2/3		
10 ⁸	3/3	3/3	2/3	1/3		
<i>F. tularensis</i> biovar novicida purified DNA						
10	0/2	0/2	1/2	0/2	Not done	
10 ²	0/2	0/2	2/2	0/2		
10 ³	0/2	0/2	1/2	0/2		
10 ⁴	0/2	0/2	2/2	0/2		
10 ⁵	0/3	0/3	2/3	1/3		
10 ⁶	2/3	0/3	3/3	2/3		
10 ⁷	3/3	1/3	3/3	3/3		
10 ⁸	3/3	3/3	1/3	3/3		

*pooled data from multiple experiments

Table 17—Detection of *F. tularensis* Sequences (Specific Tiles and Overall Calls) from Spiked *F. philomiragia*

Sample spike (cfu)	Random Amplification		BTA mix	
Overall detection (based on 1 positive tile)				
10	0/3		1/3	
10 ²	0/2		0/2	
10 ³	0/3		0/3	
10 ⁴	0/3		1/3	
10 ⁵	0/3		0/3	
10 ⁶	0/3		0/3	
10 ⁷	0/3		0/3	
10 ⁸	0/3		0/3	
Detection of sequences on individual tiles				
Marker	<i>fopA</i>	<i>tul4</i>	<i>fopA</i>	<i>tul4</i>
10	0/3	0/3	1/3	0/3
10 ²	0/2	0/2	0/2	0/2
10 ³	0/3	0/3	0/3	0/3
10 ⁴	0/3	0/3	1/3	0/3
10 ⁵	0/3	0/3	1/3	0/3
10 ⁶	0/3	0/3	0/3	0/3
10 ⁷	0/3	0/3	0/3	0/3
10 ⁸	0/3	0/3	0/3	0/3

Table 18—Summary of Results from Samples Spiked with *Y. pestis* CO-92

Concentration(cfu)	Random Amplification	BTA mix	UberMix
<i>Y. pestis</i> (CRP)*			
3 × 10 ²	0/4	3/9	2/4
1 × 10 ³	0/4	1/6	1/4
3 × 10 ³	0/5	3/8	2/5
1 × 10 ⁴	0/5	5/8	4/5
3 × 10 ⁴	0/4	6/7	5/5
1 × 10 ⁵	0/5	7/7	4/4
3-4 × 10 ⁵	0/9	11/12	3/5
1 × 10 ⁶	0/9	8/9	5/5
3-4 × 10 ⁶	0/9	6/7	4/5
7-8 × 10 ⁶	1/5	12/12	5/5
1 × 10 ⁷	1/4	4/4	
4 × 10 ⁷	2/3	3/3	
<i>Y. pestis</i> (Tetracore)			
10	0/3	3/3	Not done
10 ²	0/3	3/3	
10 ³	0/3	3/3	
10 ⁴	0/3	3/3	
10 ⁵	2/2	2/2	
10 ⁶	1/2	2/2	
10 ⁷	2/2	2/2	
10 ⁸	2/2	2/2	
<i>Y. pestis</i> purified DNA			
10	0/2	0/2	Not done
10 ²	0/2	0/2	
10 ³	0/2	2/2	
10 ⁴	0/2	2/2	
10 ⁵	0/3	3/3	
10 ⁶	0/3	3/3	
10 ⁷	2/3	3/3	
10 ⁸	3/3	3/3	

*Data for random amplification and BTA mix amplification were pooled from several experiments.

Pre-purified DNA from *F. tularensis* was also spiked into total nucleic acid extracts from Lab A, in a similar experiment as that performed with *B. thuringiensis* Al Hakam; the purpose was again to show whether the purification/extraction procedures in Lab A significantly affected DNA recoveries. This experiment indicated that *F. tularensis* sequences could be detected in crude cell preparations at similar or lower levels than in samples spiked with purified DNA (data not shown). The contrast between the *F. tularensis* results and those obtained with *B. thuringiensis* Al Hakam further support the hypothesis that preparation of the crude *B. thuringiensis* cell preps (and the potential for a sub-population of unbroken spores) may be responsible for lower sensitivity for sequence detection in assays of potential spore-formers.

Detection of Y. pestis Sequences

The RPM V.1 chip had two tiles designed for detection of *Y. pestis* sequences: a *cafI*-specific tile to detect the presence of plasmid pMT1, and a tile specific for a 5' non-coding chromosomal region *cve*. The *cafI* gene encodes F1 antigen, a secreted, capsular antigen used for immunization and diagnostic purposes, and should be detected only in *Yersinia* strains possessing pMT1, while the *cve* tile should be able to detect sequences from all *Yersinia* species.

Preparations of *Y. pestis* CO-92 from different sources were spiked into nasal wash and analyzed. Table 19 shows positives based on detection of one or more *Y. pestis* target sequences. *Y. pestis* obtained from CRP was inactivated by gamma irradiation, whereas samples from Tetracore were processed using Epicure™ extraction kits; all samples from both sources were certified non-viable before shipment. Using random amplification, *Y. pestis* sequences could be detected at significantly lower concentrations with Tetracore crude cell preparations (10^5 cfu) versus preps from CRP (high 10^6 - 10^7 cfu). However, when either BTA mix or UberMix was used during amplification, the lowest tested concentrations of target resulted in positive detection of *Y. pestis* sequences; these sample spikes corresponded to 300 cfu and 10 cfu for *Y. pestis* from CRP and Tetracore, respectively. A higher percentage of positive sequence hits was observed at lower spike levels (25-40% at the same concentrations) with UberMix than with BTA mix (13-33% at 300-3000 cfu). We are currently determining whether this difference is statistically significant.

Tiles for *cve* allowed detection of more sequences at lower target concentrations than the tiles specific for *cafI* (Table 20). This trend was observed in all samples containing *Y. pestis* cells or DNA, although it was particularly noticeable when *Y. pestis* was obtained from CRP. It should be noted that *cafI* is a plasmid-encoded gene, whereas *cve* is a chromosomal mark

Table 19—Detection of *Y. pestis* Sequences on Specific Tiles from Spiked *Y. pestis* CO-92

Sample spike(cfu)	Random Amplification		BTA mix		UberMix	
Marker	<i>cafI</i>	<i>cve</i>	<i>cafI</i>	<i>cve</i>	<i>cafI</i>	<i>cve</i>
<i>Y. pestis</i> (CRP)*						
3 × 10 ²	0/4	0/4	1/9	3/9	0/4	2/4
1 × 10 ³	0/4	0/4	0/6	1/6	0/4	1/4
3 × 10 ³	0/5	0/5	0/8	3/8	0/5	2/5
1 × 10 ⁴	0/5	0/5	1/8	5/8	0/5	4/5
3 × 10 ⁴	0/4	0/4	0/7	6/7	0/5	5/5
1 × 10 ⁵	0/5	0/5	0/7	7/7	3/4	4/4
3-4 × 10 ⁵	0/9	0/9	4/12	11/12	1/5	3/5
1 × 10 ⁶	0/9	0/9	8/9	8/9	3/5	5/5
3-4 × 10 ⁶	0/9	0/9	4/7	6/7	4/5	4/5
7-8 × 10 ⁶	1/5	0/5	12/12	12/12	5/5	4/5
1 × 10 ⁷	1/4	0/4	4/4	4/4		
4 × 10 ⁷	2/3	0/3	3/3	3/3		
<i>Y. pestis</i> crude cell prep (Tetracore)						
10	0/3	0/3	2/3	3/3	Not done	
10 ²	0/3	0/3	2/3	3/3		
10 ³	0/3	0/3	3/3	3/3		
10 ⁴	0/3	0/3	3/3	3/3		
10 ⁵	2/2	0/2	1/2	2/2		
10 ⁶	1/2	0/2	2/2	2/2		
10 ⁷	2/2	0/2	2/2	2/2		
10 ⁸	2/2	1/2	2/2	2/2		
<i>Y. pestis</i> purified DNA						
10	0/2	0/2	0/2	0/2	Not done	
10 ²	0/2	0/2	0/2	0/2		
10 ³	0/2	0/2	1/2	2/2		
10 ⁴	0/2	0/2	2/2	2/2		
10 ⁵	0/3	0/3	2/3	3/3		
10 ⁶	0/3	0/3	3/3	3/3		
10 ⁷	2/3	0/3	3/3	3/3		
10 ⁸	3/3	3/3	3/3	3/3		

*Data pooled from several experiments

Experiments assessing the effect of extraction on assay sensitivity demonstrated that *Y. pestis* sequences could be detected at lower cfus in crude cell preps than purified DNA spiked into total nucleic acid extract. We do not currently have an explanation for these results, as one might expect some loss in nucleic acid content during extraction in Lab A.

Several “near neighbors” of *Y. pestis* were analyzed using random amplification and BTA mix. These species included *Yersinia* sp. MX018 and *Y. ruckeri*, as well as purified DNA from *Yersinia* sp. MX018 (to assess extraction effects) and *Y. pseudotuberculosis* (Table 21). As none of these species harbor pMT1, no positive calls for *cafI* were made. However, a positive call for *cve* was made in a single sample containing *Yersinia* sp. MX018 purified DNA.

Table 21—Detection of *Y. pestis* Sequences (Specific and Overall Results) from Related *Yersinia* Species

Sample Spike (cfu)	Random Amplification		BTA mix		Overall Call	
Marker	<i>cafI</i>	<i>cve</i>	<i>cafI</i>	<i>cve</i>	Random	BTA mix
<i>Y. pseudotuberculosis</i> purified DNA						
10	0/2	0/2	0/2	0/2	0/2	0/2
10 ²	0/2	0/2	0/2	0/2	0/2	0/2
10 ³	0/2	0/2	0/2	0/2	0/2	0/2
10 ⁴	0/2	0/2	0/2	0/2	0/2	0/2
10 ⁵	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁶	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁷	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁸	0/3	0/3	0/3	0/3	0/3	0/3
<i>Yersinia</i> sp. MX018 crude cell prep						
10	0/3	0/3	0/3	0/3	0/3	0/3
10 ²	0/3	0/3	0/3	0/3	0/3	0/3
10 ³	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁴	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁵	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁶	0/1	0/1	0/2	0/2	0/1	0/2
10 ⁷	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁸	0/3	0/3	0/3	0/3	0/3	0/3
<i>Yersinia</i> sp. MX018 purified DNA						
10	0/2	0/2	0/2	0/2	0/2	0/2
10 ²	0/2	0/2	0/2	0/2	0/2	0/2
10 ³	0/2	0/2	0/2	0/2	0/2	0/2
10 ⁴	0/1	0/1	0/2	0/2	0/1	0/2
10 ⁵	0/3	0/3	0/3	1/3	0/3	1/3
10 ⁶	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁷	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁸	0/3	0/3	0/3	0/3	0/3	0/3
<i>Y. ruckeri</i>						
10	0/3	0/3	0/3	0/3	0/3	0/3
10 ²	0/3	0/3	0/3	0/3	0/3	0/3
10 ³	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁴	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁵	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁶	0/2	0/2	0/3	0/3	0/2	0/3
10 ⁷	0/3	0/3	0/3	0/3	0/3	0/3
10 ⁸	0/2	0/2	0/2	0/2	0/2	0/2

Detection of Vaccinia virus (smallpox simulant) sequences

The RPM v.1 chip had two tiles to detect variola major virus: one to detect the *crmB* gene encoding a homolog to tumor necrosis factor receptor; and the *J7R* gene encoding viral hemagglutinin; this viral hemagglutinin distinguishes orthopoxviruses from other poxvirus genera and several real-time PCR assays for orthopox viruses have been developed using this gene. [27-30] The *crmB* gene is found in many orthopox viruses, but is truncated in Lister strains of vaccinia, such as that used in this study. However, the RPM v.1 tile (and its associated primers) was designed to allow detection of the truncated *crmB* variant as well as the full-length *crmB*.

Relatively high concentrations (5×10^6 plaque-forming units, pfu) were required for positive detection of vaccinia sequences on the RPM v.1 chip using random amplification (Table 22). However, when using BTA mix, we were able to detect vaccinia in all samples spiked with 5×10^5 pfu, and it is likely that these sequences can be detected when much lower spikes of this strain are used. We have not yet determined the lower limit of positive calls. The tile for *crmB* proved to be more effective for detection of vaccinia sequences than *J7R*. As observed with both *Y. pestis* and *F. tularensis*, the more effective tile for detection, *crmB*, proved to be shorter (291 bp) than *J7R* (510 bp). In all but two cases (independent of primer mix), BLAST returned data indicating camelpox virus M-96 as best fit, followed by vaccinia strain WR; in the remaining two cases, the best-fit match was vaccinia. Thus, no positive results would have resulted in a variola major database record identification, illustrating the high discriminatory power of a resequencing microarray.

Table 20—Detection of Vaccinia-specific sequences using Variola major virus tile sequences (specific tiles and overall calls) from spiked Vaccinia virus

Concentration(pfu)	Random Amplification		BTA mix		Overall Call	
Marker	<i>crmB</i>	<i>J7R</i>	<i>crmB</i>	<i>J7R</i>	Random	BTA mix
5×10^5	0/5	0/5	4/4	0/4	0/5	4/4
1.6×10^6	0/5	0/5	4/4	0/4	0/5	4/4
5×10^6	4/5	0/5	4/4	0/4	4/5	4/4
1.6×10^7	4/4	1/4	4/4	0/4	4/4	4/4
5×10^7	3/3	3/3	2/2	0/2	3/3	2/2

Unanticipated results

A total of 892 spiked nasal wash samples were randomly amplified and analyzed using the v.1 RPM chip as originally described in the Silent Guardian SOPs. These samples included 374 samples spiked with inactivated BT agents (or their nucleic acids) and 518 samples spiked with “near neighbor” species, potentially interfering species, or buffer (negative control samples). Only a single sample showed a positive result for one or more genomic sequences of targets that were not intentionally spiked into the samples; samples for which definitive, species-specific sequence identification could not be made because of identical BLAST bit score / expect values (e.g., where *rpoB* results indicated either *B. anthracis*, *B. cereus*, or *B. thuringiensis*) were not included in this count. The single misidentification was due to detection of the *B. anthracis pag* gene sequence in a sample spiked with *B. thuringiensis* Kurstaki.

A total of 1010 spiked nasal wash samples were amplified using BTA mix and analyzed using the v.1 RPM chip. These samples included 460 samples spiked with inactivated BT agents and 550 samples spiked with “near neighbor” species, potentially interfering species, or buffer (negative control samples). Of these, 55 unexpectedly showed positive detection for BT (or BT-like) sequences on tiles for genomic targets not intentionally spiked into the samples. Positive results for pX01 or pX02 plasmid marker sequences (13 samples) were encountered in 5 samples, with two additional samples giving positive results for *B. anthracis rpoB*. *Y. pestis*-specific sequences were unexpectedly detected in 7 samples; five samples were positive for *cve* alone, and two positive for both *Y. pestis* markers. The rest of the presumptive false-positives (forty samples) were due to a positive call on the *fopA* tile.

A total of 284 samples were amplified using UberMix and analyzed on the v.1 RPM chips. Of these, 6 unexpectedly showed positive detection for BT (or BT-like) sequences on tiles for which the intended target not intentionally spiked into the samples. All of these were attributable to a positive for *F. tularensis fopA* or a combination of *Y. pestis cve* and *cafI*.

All unanticipated detections of BT target sequences were encountered during BT spike-in experiments and were not observed in any of the patient specimens using the standard (random) and

UberMix amplification protocols and analyzed using “non-relaxed” conditions (see Chapter 4). Furthermore, a statistically significant proportion of these unexpected results (30 of 62, $P < 0.01$, χ^2 , corrected for continuity) could be positively correlated to the presence of high concentrations of the misidentified target in an adjacent tube during processing. All of these were encountered during demonstration of production conditions, rather than validation conditions.

We considered the following factors that might lead to the unanticipated results described above:

1. Errors caused by target hybridization to the resequencing microarray or the Affymetrix base-calling algorithm.
2. Errors caused by REPI and/or the BLAST analysis methods.
3. False assumptions regarding the purity or composition of the near-neighbors.
4. High assay sensitivity leading to detection of unexpected target sequences.
5. Cross-contamination of samples during processing due to intrinsic drawbacks of the SGD production line.
6. Operator error caused by mishandling or mislabeling.

Considering these individually:

1. The first possibility, related to the accuracy of the Affymetrix chip and base calling algorithm, is of very low probability. As shown before, the ABACUS algorithm used for the Affymetrix base calling algorithm is extremely conservative, allowing accuracies of >99.9999% (Cutler et al., 2001). In control experiments, we have observed 100% concordance between RPM v.1 base calls and conventional DNA sequencing.
2. The second factor is more of a possibility. Because detection and identification of sequences are based on the achievement of a uniquely high BLAST bit score / expect value combination, the method can be biased against a small number of database records. While a well-known pathogen may have a complete genome sequence available, a near neighbor might not be represented in a database for the same genomic target, eliminating the possibility of a fine-scale discrimination between the well-known pathogen and the near neighbor.
3. All of the inactivated pathogens and near-neighbors were not characterized by prior RPM experiments. Previous work in this lab has demonstrated that microarrays routinely detect previously unknown contaminants or that trace contamination can occur at the aliquot stage. Furthermore, personnel involved with preparation of the sample spikes have indicated significant potential for contamination and sample mix-ups before the samples were ever received by SGD personnel; use of hand-labeled tubes (versus bar-coded tubes) further compounded the potential for misrepresentation of sample contents.
4. The majority of unexpected target detections were observed with detection of sequences on either the *fopA* or *cve* tiles. As mentioned in previous sections, these two markers were highly effective at detecting low concentrations of sequences from their intended target analytes, *F. tularensis* and *Y. pestis*. The amplicons from these markers are shorter and may be amplified more efficiently than longer ones. The ability to produce shorter PCR fragments that hybridize more quickly and more tightly to the shorter tiles may prove to be a double-edged sword. The same qualities that endow these primers and tiles with greatly improved sensitivities in assays using specific primers (shorter sequences, ability to hybridize faster and more tightly) may also cause a significant increase in both sensitivity and resultant detection on the RPM v.1. It is possible that a specific PCR assay alone would not detect sequences from the same contaminant at these low levels.

5. At present, a large number of manual processing steps are required in the SGD protocol. Tubes containing samples were opened at least 30 times prior to PCR amplification and at least 13 times after amplification, each time providing a potential opportunity for contamination. Use of strip tubes and associated strip caps, while allowing for parallel amplification of 30 samples, further enhanced the potential for cross-contamination both before and after PCR. Analysis of the BT-spiked samples under production, rather than validation, conditions compounded the likelihood of cross-contamination and operator error. These two possibilities are the most likely sources of error during the BT study, but will be difficult to pinpoint in retrospect. Attribution of errors to these two sources will probably occur deductively by systematic elimination of possibilities 1–4. Moreover, there was a statistically significant correlation of misidentified samples with adjacent tubes containing high concentrations of target ($P < 0.01$). This correlation further supports the hypothesis that cross-contamination from neighboring tubes during the many processing steps may have been, at least in part, responsible for misidentified samples.

Summary and Conclusions

The RPM v.1 was designed in 2003 as a proof-of-concept microarray, primarily for the detection of sequences from common respiratory pathogens encountered in the basic military training environment. A minimal amount of tile sequences were allocated to detection of six BT agent targets. The sequences of these targets, and their corresponding PCR primers, were taken from the open literature and used primarily to test for cross-hybridization with common respiratory pathogens. No formal technical assessments, such as the Silent Guardian Demonstration Project, were part of the intended range of RPM v.1 experiments. With no time to design a fully optimized set of BT agent tiles on the array or optimize random and multiplexed PCR protocols, the SGD demonstration was initiated with a single random protocol with a compromised (due to DNA subtraction reagent failure) human DNA subtraction pathway.

Although the compromised random amplification protocol still allowed for detection of sequences from common viral pathogens in clinical samples (Chapter 4), the sensitivity was inadequate for detection of sequences from low levels of bacterial pathogens and BT agents. Using an opportunity to perform an initial assessment of the RPM v.1 for BT detection, the sensitivity of the system was increased by combining individual primer pairs to create two non-optimized multiplex mixtures (BTA and UberMix). The revised amplification protocols were then systematically challenged with BT agents across large concentration ranges and batteries of genetic near neighbors.

Therefore, it is quite encouraging that the first formal assessment of this technology for the present purpose produced results comparable to the most sensitive single BT PCR assays when using the BTA or UberMix primer mixes. In addition to the sensitivity of the assays, direct sequence reads of the targets are provided, allowing an additional level of discrimination as well as a numerical estimate of the probability of a false target identification from direct sequence similarity comparisons – a feature not available using any other available BT detection technology.

A total of more than 2100 analyses were performed on spiked samples and blank controls using random amplification, BTA mix, or UberMix. Target sequences were identified by associating the top sequence similarity ranking (uniquely high BLAST bit score with corresponding low expect value) with single pathogen genome database record(s). Using these specific primer mixes, we detected *B. anthracis* sequences from Sterne strain vegetative cells at the lowest number of cells tested (20 cfu) and from Ames spores at levels as low as 6×10^4 cfu per sample. *F. tularensis* sequences were also detected in two strains at cell numbers as low as 10 cfu per sample in LVS and *F. tularensis* biovar novicida and 10^3 cfu per sample in Schu4. *Y. pestis* sequences were detected in samples containing 10-300 cfu of *Y. pestis* target, depending on the source. Sequences specific for vaccinia virus, a simulant for variola major virus,

were detected at the lowest concentration tested (10^5 pfu) – while allowing for fine scale genomic discrimination between vaccinia and variola virus.

A straightforward pathway exists for the rational design of optimized RPM chip, target amplification, and analysis. A newly designed v.2.5 RPM chip containing validated target sequences coupled with simpler, optimized protocols and robotics will likely provide a highly sensitive device to detect sequences specific to BT agents in the clinical environment with an extremely low probability of misidentified species.

CHAPTER 6: CONCLUSIONS, LESSONS LEARNED, RECOMMENDATIONS

The Silent Guardian Demonstration successfully showed its potential as a biosurveillance system in the Military District Washington during January and February 2005 for both naturally occurring respiratory pathogens and biothreat species. **Multiple pathogens were identified in clinical samples from the general population in less than 24 hours.** Identifications of naturally occurring pathogens were made at the **strain**, as well as species, level in many cases. All influenza-positive samples were scrutinized for strain identification; no bird flu was detected. None of the sequences from six bioterrorist agents included in the tiles on the RPM1 chip were detected in patient samples, although *B. anthracis* sequences from all spiked (blinded) samples were immediately detected. Similarly, no influenza A and/or B was detected in over 2100 BT challenge experiments when inactivated agents were spiked into “healthy” nasal wash. The results produced a high level of confidence in the reliability of the EOS resequencing approach for biosurveillance in both military and civilian populations.

Many of the lessons learned during SGD can be attributed to the unique requirements of the demonstration and could be eliminated with more time for planning. For problems related to the technology itself, all technical shortcomings identified in SGD will be addressed (“lessons learned”), leading to a revolutionary approach to medical diagnostics and biosurveillance. We anticipate a dramatic reduction in time and personnel needed to achieve a 6-hour analysis time using a single operator. Given the demonstrated utility of the RPM v.1 chip in SGD, it may make sense to begin developing protocols for environmental sampling (air, water, and soil) and then assessing utility for non-clinical operations.

Lessons learned during SGD and recommendations for improvement include the following:

1. *Standards for comparison.* Culture was used as the standard for comparison, yet samples had to be frozen and shipped to California for culture. Despite the use of bacterial and viral transport media, some loss of viability was inevitable. Furthermore, the SGD protocol can sequence DNA from non-viable pathogens, leading to a potential lack of correlation between the culture results and the EOS chip results.

Recommendation: Perform culture close enough to the location of the EOS analysis that no freezing of samples is required for transport. Employ accepted PCR methods as well as culture with resequencing of recovered isolates for comparison.

2. *Complexity of the processing steps.* The protocol included complex biochemistry and extensive manipulations of microliter quantities, leading to possible errors in chemistry, loss of material, and operator error even by the most skilled of operators. Furthermore, the high intensity of operation as hundreds of samples per day were processed decreased the opportunity for operators to detect errors at the point in the protocol where samples were transferred to tubes labeled only in batches (to fit in the PCR machine) and recombined into labeled tubes thereafter. The production conditions used for analysis of BT-spiked samples, however, were presumably not the same as used with patient samples, and will be significantly modified for official validation of these protocols and this technology.

Recommendation: Combine DNA and RNA processing pathways if a method of doing so can be developed. Automate as many steps in the protocol as possible.

3. *LIMS.* The LIMS system, though vital to SGD, was developed in a fraction of the time normally needed. As anticipated, several functions or operations that could have helped were not available or simply could not be developed with the time and money available. In particular, minimal provisions were made to incorporate comments or metadata into the LIMS system.

Recommendation: Adequate time should be allowed for development, testing, and implementation of a tracking and reporting system.

4. *Sensitivity.* The sensitivity using the SGD random amplification protocol was not as high as is desirable. Removal of human DNA was inoperable due to reagent problems, leading to a higher background on the chip and decreased signal/background.

Recommendations: Improve the process for signal/noise discrimination; develop a better procedure for separating human and pathogen nucleic acids and minimize any other sources of noise. Develop a PCR protocol that uses random amplification for rapidly mutating pathogens and specific amplification for pathogens with distinctive, highly conserved genes. (Possibly decrease the sensitivity for commensals in relevant fluids to avoid competition for PCR reagents.)

5. *Time.* The turnaround time of 18–24 hours was slower than desired. Hybridization times were reduced from 18 hours in the standard Affymetrix protocol to 4 hours in the SGD protocol, which may have reduced sensitivity by as much as a factor of 2. The PCR step was run in 4 hours. Manual manipulations were very time consuming throughout the procedure, especially as the number of samples per batch increased.

Recommendations: Better systems for mixing during hybridization may increase hybridization efficiency and reduce time. Automated PCR in microfluidic formats could decrease the amplification time to 30 minutes. Automation and protocol improvements are required to reduce operational time. Faster software would expedite sequence analysis.

6. *Space.* The SGD protocol required five “wet” laboratories and two computer hardware spaces.

Recommendation: The system needs to be reduced in size and complexity without adding a risk of sample cross-contamination.

7. *Cost.* The SGD project was very expensive. The analysis effort at NRL alone required \$8-9M to provide the capacity to analyze 10,000 samples over a 3-month period. Cost included equipment, supplies, facilities, travel for Navy reservists, and NRL staff salaries, but not AF salaries.

Recommendation: Reduce personnel numbers, operator time, supplies, and chip costs by automation and use of smaller (higher density or fewer pathogens) chips, reduction of the number of chips on-hand and/or improved (lengthened) chip shelf life.

8. *Training.* Most operators were already skilled in molecular biology and were up to the task. The training was extensive, though after the initial in-brief, most laboratory personnel were trained only on protocols for one specific lab. The higher trainer: trainee ratio for the AF Bravo team also meant that the focus was primarily on learning the protocols step-by-step-rather than the overall concepts of the RPM assays. As SGD progressed, AF laboratory personnel felt that explanations for changes and the “big picture” were not communicated clearly on several occasions. Cross training in the different labs, similar to what occurred during the last two weeks of SGD, should have been done from the beginning. Also, a more gradual increase in workload with improved checkpoints to identify process variables and bottlenecks should be implemented as opposed to infrequent “high sample load” scenarios (i.e., implementation of a Quality Management System/QSM). Finally, more training from the manufacturers, including Affymetrix, would have aided troubleshooting.

Recommendation: Simplify and automate the protocols as much as possible to reduce eliminate training requirements.

9. *Reagent quality control.* Measures taken to document the preparation and use of various reagents improved over the course of the Silent Guardian project. However, insufficient personnel and reagents were available to pre-test the large numbers of different reagents and solutions required for all steps in sample processing on a daily basis. Furthermore, the rapid rate at which reagents were used precluded post-chip QC determinations; this problem was further exacerbated by the short shelf-life of a number of reagents. Due to the daily log sheets which were integrated into the SGD protocols, we were able to determine the sources of specific errors in several, but not all, cases.

Recommendation: Obtain rapid, specific QC protocols for each reagent solution or optimally for groups of reagents. By batching groups of reagents (e.g., all reagents used in the first three steps in RNA isolation), the quality of each group of solutions could be determined in a timely fashion, allowing production to halt before additional samples were processed. For future operations, develop stabilized, pre-aliquoted reagents.

10. *Project communication.* Given the constraints of the project, it was often difficult to provide complete review of all operations on a timely basis, in large part due to the 24-hour, 3-shift workday. During FOC, daily briefings were held by senior management to review project status and plan for future work. However, at the laboratory level, plans and information were not always adequately communicated, resulting in operator mistakes or confusion.

Recommendation: Daily electronic status reports and specific taskings distributed to all team members could be used to minimize communication problems in multi-operator procedures. Automation and potential for single operator operation will eliminate the problem.

11. *Set-up and maintenance logistics.* As described in the report, many reagents, supplies, instruments and people were involved at every stage of the project, while preparation and planning was minimal due to the short initiation time and the time of the year. Not only was it the very beginning of the new federal fiscal year, but the major holiday season of the calendar year. In some cases, the manufacturers/vendors required considerable lead times before equipment and/or large quantities of reagents were available. Predictably, problems were encountered in obtaining and storing sufficient quantities of all materials need for large-scale production. Storage for perishable reagents was also lacking and specialized equipment was often not readily available on short notice. Utilization rates and supply streams required for each laboratory were largely unknown until mid-way through the Silent Guardian project.

Recommendation: Provide more lead time and logistical support personnel in advance, thus allowing for more careful planning, set up, operation, and maintenance of required supplies/supply levels.

ADDENDUM

Following completion of the Silent Guardian Demonstration Project and this report, several of the recommendations suggested in Chapter 6 were evaluated. Improved protocols were developed and a paper describing the new protocols has been submitted. [7] All the Silent Guardian samples previously assessed by culture and the standard protocol were retested using newer, more efficient and far more sensitive protocols. Using the modified protocols, assay times have been further reduced to 8.5 hours. The data for these repeat analyses are shown in Lin et.al. [7] The concordance between culture and the results from the new protocols is 98%. The concordance most likely is not exactly 100% due to new protocols that are now more sensitive than the combination of enhanced RT-PCR and culture methods used as reference standards. We are working hard to find alternative ways to confirm that the RPM v.1 positives that were negative by the complementary techniques do not represent false-positives.

ACKNOWLEDGEMENTS

Finally, Silent Guardian was not only a unique demonstration of the capability of the joint services to bring research to operation in record time, but it exemplified the amazing synergy possible when individuals with widely diverse capabilities are committed to a common endeavor. The NRL Silent Guardian team owes its success to:

The Office of the Secretary of Defense: Secretary Klein for initiative and leadership and Secretary Klaus Schafer for the vision of what the EOS chip could do and the support of both SGD and EOS.

The Air Force Surgeon General LtGen Peach Taylor and his Modernization Directorate (Col Peter Demitry, Col Steve Holt, and LtCol Lynn Difato) for commitment and organization,

Joint Program Executive Office: Chemical and Biological Defense (BG(P) Stephen V. Reeves, Mike Walter, Kristin Korte, David Cullin, Curt Wilhide, CDR Charlie Cutshall, and Linda Greer) for value added support,

Defense Threat Reduction Agency, Joint Science and Technology Office, Chemical Biological Medical: Dr. Charles Gallaway and LtCol Patricia Reilly for additional support,

NRL: Code 6900 scientists and administrators, Codes 6000, 6100, 1001, and 1000 for active support, R & D Services, Contracting (Kevin King), Supply, and Security for their timely assistance,

AFIOH, Brooks AFB, Tyndall AFB clinical staff, especially LtCol Sue Harmon and MSgt Chester Alexander III,

NHRC and CDR Kevin Russell for coordination of all culture and PCR reference assays,

AFRI, Navy SG, Navy Reserves for providing technical capability,

The PIs and collection staff at the six MTFs in the Washington DC area for the sample collection.
Capt Erik Von Rosenvinge, Malcolm Grow USAF Medical Center, Andrews Air Force Base, MD;
LtCol J Karen Klingenger, Bolling Air Force Base Medical Clinic, Washington DC;
Dr. John Baxter, Pentagon Flight Clinic, Arlington, VA;
LtCol Mary Shepard, DiLorenzo Clinic-Pentagon, Arlington, VA;
LTC Glenn Wortmann, Walter Reed Army Medical Center, Washington DC;

CAPT Sybil Tasker, National Naval Medical Center, Bethesda, MD

Each of these participants had a direct effect on the Silent Guardian project and was critical to its success.

We also want to thank Col Debra Neimeyer, LtCol Lynn DiFato, and CDR Michael Zwick for their extensive review of the report and helpful suggestions.

REFERENCES

1. E. H. Hanson, R. K. Rowley, B. K. Agan, C. Tibbetts, D. M. Niemeyer, in *Preparedness Against Bioterrorism and Re-emerging Infectious Diseases*, J. Kocik, J. Janiak, M. Negut, Eds. (IOS Press, 2004).
2. M. B. Heller *et al.*, Laboratory response to anthrax bioterrorism, New York City, 2001. *Emerging Infectious Diseases* **8**, 1096-1102 (2002).
3. D. M. Niemeyer, *Laboratory BioDefense Guide*. (Air Force Force Protection Battlelab, Lackland Air Force Base, TX, 2002).
4. B. C. Lin *et al.*, Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. *Genome Research* **16**, 527-535 (2006).
5. E. M. Elnifro, A. M. Ashshi, R. J. Cooper, P. E. Klapper, Multiplex PCR: Optimization and application in diagnostic virology. *Clinical Microbiology Reviews* **13**, 559-+ (2000).
6. C. S. Dulcey *et al.*, Deep UV photochemistry of chemisorbed monolayers - Patterned coplanar molecular assemblies. *Science* **252**, 551-554 (1991).
7. B. Lin *et al.*, Application of broad-spectrum, sequence-based pathogen identification in an urban population. *Plos One* **2**, (2007).
8. D. Wang *et al.*, Microarray-based detection and genotyping of viral pathogens. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 15687-15692 (2002).
9. D. Wang *et al.*, Viral discovery and sequence recovery using DNA microarrays. *Plos Biology* **1**, 257-260 (2003).
10. CDC. (2005), vol. 2019.
11. V. Perreten *et al.*, Microarray-based detection of 90 antibiotic resistance genes of gram-positive bacteria. *Journal of Clinical Microbiology* **43**, 2291-2302 (2005).
12. D. R. Franz, in *Director's Series on Proliferation*, K. C. Bailey, Ed. (1994), pp. 55-65.
13. T. D. Read *et al.*, Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* **296**, 2028-2033 (2002).
14. P. R. Coker *et al.*, *Bacillus anthracis* virulence in guinea pigs vaccinated with anthrax vaccine adsorbed is linked to plasmid quantities and clonality. *Journal of Clinical Microbiology* **41**, 1212-1218 (2003).
15. J. Ellis, P. C. F. Oyston, M. Green, R. W. Titball, Tularemia. *Clinical Microbiology Reviews* **15**, 631-+ (2002).
16. H. Roloff, P. Glockner, K. Mistele, R. Böhm, The taxonomic relationship between *B. anthracis* and the *B. cereus* group, investigated by DNA-DNA hybridization and DNA amplification fingerprinting (DAF). *Salisbury Medical Bulletin* **87**, 38-39 (1996).
17. E. Helgason *et al.*, *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* - One species on the basis of genetic evidence. *Applied and Environmental Microbiology* **66**, 2627-2630 (2000).
18. Y. A. Qi *et al.*, Utilization of the *rpoB* gene as a specific chromosomal marker for real-time PCR detection of *Bacillus anthracis*. *Applied and Environmental Microbiology* **67**, 3720-3727 (2001).
19. J. Pannucci, R. T. Okinaka, R. Sabin, C. R. Kuske, *Bacillus anthracis* pXO1 plasmid sequence conservation among closely related bacterial species. *Journal of Bacteriology* **184**, 134-141 (2002).
20. J. Pannucci *et al.*, DNA sequence conservation between the *Bacillus anthracis* pXO2 plasmid and genomic sequence from closely related bacteria. *Bmc Genomics* **3**, (2002).

21. M. Fulop, D. Leslie, R. Titball, A rapid, highly sensitive method for the detection of *Francisella tularensis* in clinical samples using the polymerase chain reaction. *American Journal of Tropical Medicine and Hygiene* **54**, 364-366 (1996).
22. R. Grunow *et al.*, Detection of *Francisella tularensis* in biological specimens using a capture enzyme-linked immunosorbent assay, an immunochromatographic handheld assay, and a PCR. *Clinical and Diagnostic Laboratory Immunology* **7**, 86-90 (2000).
23. Z. Junhui *et al.*, Detection of *Francisella tularensis* by the polymerase chain reaction. *Journal of Medical Microbiology* **45**, 477-482 (1996).
24. A. Sjostedt, U. Eriksson, L. Berglund, A. Tarnvik, Detection of *Francisella tularensis* in ulcers of patients with tularemia by PCR. *Journal of Clinical Microbiology* **35**, 1045-1048 (1997).
25. J. L. Versage, D. D. M. Severin, M. C. Chu, J. M. Petersen, Development of a multitarget real-time TaqMan PCR assay for enhanced detection of *Francisella tularensis* in complex specimens. *Journal of Clinical Microbiology* **41**, 5492-5499 (2003).
26. A. Johansson *et al.*, Evaluation of PCR-based methods for discrimination of *Francisella* species and subspecies and development of a specific PCR that distinguishes the two major subspecies of *Francisella tularensis*. *Journal of Clinical Microbiology* **38**, 4180-4185 (2000).
27. M. S. Ibrahim *et al.*, Real-time PCR assay to detect smallpox virus. *Journal of Clinical Microbiology* **41**, 3835-3839 (2003).
28. S. L. Ropp, Q. Jin, J. C. Knight, R. F. Massung, J. J. Esposito, PCR Strategy for Identification and Differentiation of Smallpox and Other Orthopoxvirus. *Journal of Clinical Microbiology* **33**, 2069-2076 (1995).
29. D. A. Kulesh *et al.*, Smallpox and pan-orthopox virus detection by real-time 3'-minor groove binder TaqMan assays on the Roche LightCycler and the cepheid smart cycler platforms. *Journal of Clinical Microbiology* **42**, 601-609 (2004).
30. M. J. Espy *et al.*, Detection of smallpox virus DNA by LightCycler PCR. *Journal of Clinical Microbiology* **40**, 1985-1988 (2002).

APPENDIX A: MEMORANDUM STARTING SILENT GUARDIAN



NUCLEAR AND CHEMICAL
AND BIOLOGICAL DEFENSE
PROGRAMS

ASSISTANT TO THE SECRETARY OF DEFENSE
3050 DEFENSE PENTAGON
WASHINGTON, DC 20301-3050

NOV 10 2004

MEMORANDUM FOR SECRETARY OF THE ARMY
SECRETARY OF THE NAVY
SECRETARY OF THE AIR FORCE
THE SURGEON GENERAL OF THE ARMY
THE SURGEON GENERAL OF THE NAVY
THE SURGEON GENERAL OF THE AIR FORCE
JOINT PROGRAM EXECUTIVE OFFICE FOR CHEMICAL
AND BIOLOGICAL DEFENSE
NAVAL RESEARCH LABORATORY

SUBJECT: National Capital Region Microarray Biodefense Detection (MBD)

The Department of Defense is implementing a biodefense detection project in the National Capital Region (NCR). This project will implement a demonstration of microarray detection technology which is currently under development as an Advanced Concept Technology Development supported jointly by my office, the Deputy Under Secretary of Defense for Advanced Systems & Concepts and the U.S. Air Force Surgeon General. This demonstration project will take place from December 1, 2004, to March 31, 2005. I request the following organizations provide support to this effort as outlined below.

The Air Force Surgeon General is tasked with implementing this capability from December 1, 2004, to March 31, 2005, at the NCR, and to coordinate this effort with the Surgeons General of the Army and Navy. Funding will be provided by my office for the purpose of accelerating the demonstration of this ACTD technology.

The Joint Program Executive Office for Chemical and Biological Defense will assist through consultation, financial management, contracting, and operational evaluation.

The Naval Research Laboratory (NRL) will provide laboratory space for equipment and personnel to process collected samples. NRL will also provide appropriate procedural protocols for government/civilian laboratory personnel, as developed for the program. NRL will train laboratory personnel, as needed, and oversee their initial operation.

My point of contact is Colonel Joseph M. Palma, M.D., 703-697-5116.

Dale Klein

ENCLOSURE (7)

APPENDIX B: STANDARD OPERATING PROTOCOLS

At this time, the standard operating protocols are proprietary. If you are interested obtaining further information about the protocols, please contact Dr. David Stenger at the address below. A non-disclosure agreement may need to be signed.

Dr. David Stenger
Center for Bio/Molecular Science and Engineering
Naval Research Laboratory
4555 Overlook Ave., SW
Washington, DC 20375-5320
david.stenger@nrl.navy.mil

APPENDIX C: PRIMER LISTS

At this time, the UberMix primer mix is proprietary. If you are interested obtaining further information about this primer mix, please contact Dr. David Stenger at the address below. A non-disclosure agreement may need to be signed.

Dr. David Stenger
Laboratory Head, Center for Bio/Molecular Science and Engineering
Naval Research Laboratory
4555 Overlook Ave., SW
Washington, DC 20375-5320
dstenger@cbmse.nrl.navy.mil

APPENDIX D: STAFF AT NRL DURING SGD

Points of Contact (Name and Home-base association location)

Dr. Joel M. Schnur (Silent Guardian's NRL Program Manager) (retired)
 Director, Center for Bio/Molecular Science and Engineering
 Naval Research Laboratory
 Washington, DC 20375-5320

LtCol Sue Harmon, USAF (Silent Guardian's On-site AF Commanding Officer) (retired)
 Air Force Institute for Operational Health (AFIOH)
 Brooks City-Base, TX 87235

Dr. Frances S. Ligler (Silent Guardian's NRL Technical Leader) (retired)
 USN Senior Scientist for Biosensors and Biomaterials
 Naval Research Laboratory
 Washington, DC 20375-5320

Ms. Anne Kusterbeck (Silent Guardian's NRL Administrative Leader)
 Laboratory Head, Center for Bio/Molecular Science and Engineering
 Naval Research Laboratory
 Washington, DC 20375-5320

Staff at NRL (Home-base assignment location)

SMSgt Chester Alexander, USAF, Brooks-City Base, TX
 TSgt Angel Allison, USAF, Brooks-City Base, TX
 Dr. George Anderson, NRL-Code 6900, Washington, DC
 SSgt Nicanor Barboza, USAF, Hickman Air Force Base, HI
 Mr. Jason Bongard, George Mason University, Fairfax, VA
 Mr. Paul Charles, NRL-Code 6900, Washington, DC
 SSgt Patrick Chernay, USAF, Brooks-City Base, TX
 SSgt Sarah Cooper, USAF, Brooks-City Base, TX
 TSgt Annette Davis, USAF, Brooks-City Base, TX
 Capt Wilfred Delacruz, USAF, PhD, Travis Air Force Base, CA
 Mr. Michael Dinderman, NRL-Code 6900, Washington, DC
 Dr. Walter Dressick, NRL-Code 6900, Washington, DC
 SSgt Tiffany Finch, USAF, Brooks-City Base, TX
 Ms. Gloria Gover, Brooks-City Base, TX
 Capt David Haggerty, USAF, Brooks-City Base, TX
 LT Shelly Hakspiel, USN, NRL-Code 6100, Washington, DC
 CAPT(s) Billy Hardas, USNR, Ph.D.
 SSgt Symone Johnson, USAF, Brooks-City Base, TX
 SSgt Lureca Jones, USAF, Brooks-City Base, TX
 Capt Morris Jones, USAF, Travis Air Force Base, CA
 Dr. Joanne Jones-Meehan, NRL-Code 6100, Washington, DC
 SSgt Anna Kramer, USAF, US Air Force Academy, CO
 SSgt Valerie Lee, USAF, Luke Air Force Base, AZ
 Capt Denise Lennon, USAF, Brooks-City Base, TX
 Dr. Tomasz Leski, NRC post-doc, NRL-Code 6100, Washington, DC

Dr. Jinny Liu, NRL-Code 6900, Washington, DC
Capt Fe Lobo-Menendez, USAF, Brooks-City Base, TX
SSgt(s) Anthony Lowman, USAF, Brooks-City Base, TX
SRA Stephanie Lowman, USAF, Brooks-City Base, TX
ENS Kathleen Mandell, USNR, Ph.D.
A1C Mirian Martinez, USAF, Brooks-City Base, TX
LT James McCardle, USN, NRL-Code 6900, Washington, DC, Ph.D.
Dr. Miriam Ngundi, NRC post-doc, NRL-Code 6900, Washington, DC
MSgt(s) Kelbey Norton, USAF, Pope Air Force Base, NC
TSgt Traci Oaks, USAF, F. E. Warren Air Force Base, WY
CDR Buchan A. Oyoyo, USN, Ph.D., NRL-Code 6100, Washington, DC
Dr. Steve Pollack, NRL-Code 6900, Washington, DC
Dr. Bradley Ringeisen, NRL-Code 6100, Washington, DC
Dr. Robb K. Rowley, MD, USAF Surgeon General's Office, Washington, DC
SRA Brian Sanchez, USAF, Vandenberg Air Force Base, CA
SSgt Joseph Schuerger, USAF, Brooks-City Base, TX
Ms. Kara Shaffer, NRL-Code 6900, Washington, DC
Ms. Lisa Shriver-Lake, NRL-Code 6900, Washington, DC
MSgt Aaron Sinclair, USAF, Brooks-City Base, TX
SSgt Crystal Stokely, USAF, Brooks-City Base, TX
Dr. William Straube, Geo-Centers, Inc., Lanham, MD
CDR William Suiter, USNR, Ph.D.
SSgt Rhonda Sullivan, USAF, Brooks-City Base, TX
Dr. Chris Taitt, NRL-Code 6900, Washington, DC
SSgt Rodney Tyler, USAF, Brooks-City Base, TX
Mr. Brian Weslowski, NRL-Code 6900, Washington, DC
Dr. Brandy White, NRC post-doc, NRL-Code 6900, Washington, DC

APPENDIX E: POSTER DESCRIBING NEW AND IMPROVED PROTOCOL



The Silent Guardian Project (Y-018)

F. S. Ligler, J. M. Schnur, A. W. Kusterbeck, C. R. Taitt, L. Shriver-Lake, Z. Wang, B. Lin, D. A. Stenger

Center for Bio/Molecular Science & Engineering, Code 6910, Naval Research Laboratory, Washington, DC



Summary

Background: Emerging infections have an enormous impact on public health, e.g. the severe acute respiratory syndrome (SARS) outbreaks, the emergent avian (H5N1) influenza, and the deadly anthrax attacks of 2001 are a few recent examples that highlight the threat of naturally occurring, or intentionally used biological agents. Recent advances in microarray technologies offer rapid screening capabilities for pathogen detection that could prove effective in the prompt identification of bioterrorism events. We demonstrated the ability of DNA microarray for broad-spectrum pathogen detection by performing medical surveillance of the Washington DC metropolitan region from December 2004 to February 2005-The Silent Guardian Demonstration Project (SGDP).

Methods: A respiratory pathogen microarray v.1 (RPM v.1) was designed to identify 20 common respiratory pathogens, including viral and bacterial organisms, as well as potential bioterrorism agents. Nasal wash specimens were collected from patients presenting with febrile respiratory illness at military treatment facilities. Pathogen target genes were amplified using a highly multiplexed reverse transcriptase polymerase chain reaction [RT/PCR] protocol. RPM v.1 were then used to obtain direct genetic sequence from the amplified gene targets. The pathogen was identified at species, serotype, and strain levels by determining genetic similarity to genome database records. Additional aliquots for each specimen were used for reference assays, including viral/microbial culture, and conventional and/or real-time RT/PCR assays.

Findings: Compared with reference assays, RPM v.1 assay showed high concordance (> 98% overall agreement) in 424 fully characterized specimens. This approach not only detected pathogens, including co-infections, with high sensitivity and specificity, but also produced correct species and strain-level identification via direct sequence reads in an improved assay time (8.5 hours).

Interpretation: Our results suggest that this microarray-based approach is useful for both diagnostics and surveillance purposes.

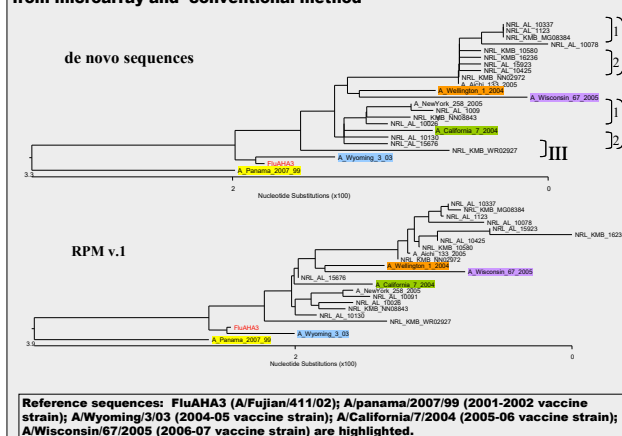
Pathogens identified-Overall microarray result vs. reference methods

Organism	Culture (+)	RPM v.1 (+)	Ref ⁺ (+), RPM v.1 (-)	Ref ⁻ (-), RPM v.1 (+)
Adenovirus	2	9	0	1
Coronavirus	28*	24	5	2
Influenza A	176	269	1	7
Influenza B	28	46	1	6
PIV 1	N.D.	2	0	0
PIV 2	2	N.D.	-	0
PIV3	N.D.	1	0	1
Rhinovirus	N.D.	2	1	1
M. pneumoniae	N.D.	4	2	2
N. meningitides	N.D.	13	6	5
S. pneumoniae	8	41	0	3
S. pyogenes	9	13	6	0
Negative	176	58	0	4

Evaluation of the detection efficiency for influenza A and B viruses in clinical samples

	Influenza A		Influenza B	
	Ref ⁺ +	Ref ⁺ -	Ref ⁺ +	Ref ⁺ -
RPM v.1 +	262	7	40	6
RPM v.1 -	1	154	1	377
Sensitivity	99%		98%	
Specificity	96%		98%	
Overall agreement	98%		98%	

Phylogenetic analysis for hemagglutinin using sequences from microarray and conventional method



Reference sequences: FluA/HA3 (A/Fujian/411/02); A/panama/2007/99 (2001-2002 vaccine strain); A/Wyoming/3/03 (2004-05 vaccine strain); A/California/7/2004 (2005-06 vaccine strain); A/Wisconsin/67/2005 (2006-07 vaccine strain) are highlighted.

Nucleotides difference in hemagglutinin (HA) genes identified by RPM v.1 from all 250 influenza A/H3N2 isolates.

Position aa/nt*	A/Fujian/411/02*	A/Wyoming/3/03	A/Wellington/1/04	A/California/7/04
8/25	Val/GTT	Val/GTT (1)	Val/GTC (71)	Val/GTT (166)
28/83	His/CAT	His/CAT (1)	His/CAT (56)	His/CAT (166)
39/116	Ile/ATA	Val/GTA (1)	Ile/ATA (71)	Ile/ATA (167)
81/244	Arg/AGG	Arg/AGG (1)	Arg/AGA (64)	Arg/AGG (161)
100/299	Ala/GCT	Thr/ACT (0)	Thr/ACT (33)	Thr/ACT (90)
117/352	Lys/AAA	Asn/AAC (1)	Asn/AAC (62)	Asn/AAC (144)
131/393	Tyr/TAC	Phe/TTC (1)	Phe/TTC (65)	Phe/TTC (159)
149/446	Leu/TTG	Leu/CTG (1)	Leu/CTG (39)	Leu/TTG (144)
161/483	Ser/AGT	Asn/AAT (1)	Asn/AAT (67)	Asn/AAT (152)
188/564	Asn/AAT	Asn/AAT (1)	Ser/AGT (67)	Asn/AAT (144)
195/584	Val/GTA	Val/GTA (1)	Ile/ATA (59)	Val/GTA (160)
198/593	Val/GTC	Ile/ATC (1)	Val/GTC (61)	Ile/ATC (43)
199/596	Ser/TCC	Ser/TCC (1)	Pro/CCC (66)	Pro/CCC (5)

*Amino acid (aa) and nucleotide (nt) positions correspond to those of HA3 tiled prototype sequence on RPMv.1, amino acid and its corresponding codon at each position was separated by "/>. Mismatched nucleotides identified by the tiled prototype sequence are underlined.
 *HA3 tiled prototype sequence was derived from A/Fujian/411/02 strain. A/Wyoming/3/03, A/Wellington/1/04 and A/California/7/04 represent three group of strains identified from all Silent Guardian specimen, respectively.
 *The number in each parenthesis indicates number of nucleotides at SNP position unambiguously identified by RPMv.1.

Co-infections identified

Organism	Culture (+)	RPM v.1 (+)
Adenovirus + Influenza A	N.D.	3
Coronavirus + Influenza A	4*	8
Coronavirus + N. meningitides	N.D.	1
Influenza A + Influenza B	N.D.	5
Influenza + PIV	N.D.	2
Influenza A + B. anthracis	N.D.	1
Influenza A + H. influenzae	2	N.D.
Influenza A + M. pneumoniae	N.D.	2
Influenza A + N. meningitides	N.D.	7
Influenza A + S. pneumoniae	1	16
Influenza B + Adenovirus	N.D.	1
Influenza B + Coronavirus	N.D.	1
Influenza B + F. tularensis	N.D.	1
Influenza B + S. pneumoniae	N.D.	7
Influenza B + S. pyogenes	N.D.	3
N. meningitides + S. pneumoniae	N.D.	2
Influenza A + N. meningitides + S. pneumoniae	N.D.	3
Influenza B + N. meningitides + S. pneumoniae	N.D.	1
Influenza A + Influenza B + S. pneumoniae	N.D.	1
Influenza A + Influenza B + Coronavirus + M. pneumoniae + S. pneumoniae	N.D.	1
Rhinovirus + PIV + B. pertussis + N. meningitides + S. pneumoniae	N.D.	1

Conclusion

- SGDP, a first-ever multiple pathogen surveillance based on genetic sequence. This was the first formal assessment of the technology for broad-spectrum identification of both RNA- and DNA-based pathogens.
- Multiple pathogens were identified in clinical samples from the general population in less than 24 hours. Identifications were made at the strain, as well as species level.
- All Influenza-positive samples were scrutinized for strain identification; no bird flu was detected. From phylogenetic analysis, these samples consisted of three major groups of influenza A/H3N2 virus strains. All groups were evolved from A/Fujian/411/02 strain.
- None of the sequences from six bioterrorist agents included in the tiles on the RPM1 chip were detected in patient samples. The results produced a high level of confidence in the reliability of the re-sequencing approach for biosurveillance in both military and civilian populations.
- Simpler, optimized protocols and automation will provide a highly sensitive system to detect common respiratory pathogens with a high degree of concordance with reference assay microbiological methods requiring days to weeks to perform.

Acknowledgements

The NRL Silent Guardian team owes its success to The Office of the Secretary of Defense: Secretary Klein and Secretary Klaus Schaffer; The Air Force Surgeon General LIGen Peach Taylor and his Modernization Directorate (Col Peter Demity, Col Steve Holt, and LtCol Lynn Difato); Joint Program Executive Office: Chemical and Biological Defense (Mike Walters, Kristen Korte, David Cullen, Curt Withide, Charlie Cuthbert, and Linda Greer); NRL: Code 6900 scientists and administrators, Code 6000, 6100, 1001, and 1000 for active support, R & D Services, Contracting (Kevin King), Supply, and Security for their timely assistance; AFIOH, Brooks AFB, Tyndall AFB clinical staff, especially LtCol Sue Harmon and MSgt Chester Alexander III; NHRD and CDC Kevin Russell for coordination of all culture and PCR reference assays; AFRL, Navy SG, Navy reserves for providing technical capacity, Capt Erik Von Rosenberg, LtCol J Karen Klingenberg and Dr John Baxter, LtCol Mary Shepard, LTC Glenn Wortmann, CAPT Sybil Tasker, and their staffs for sample collection in the Washington DC area. These participants were critical to the success of the Silent Guardian project.