EDGEWOOD CHEMICAL BIOLOGICAL CENTER
U.S. ARMY RESEARCH, DEVELOPMENT AND ENGINEERING COMMAND
Aberdeen Proving Ground, MD 21010-5424

ECBC-TR-1288

DNA SEQUENCING TECHNOLOGIES WITHIN THE CHEMICAL AND BIOLOGICAL DEFENSE ENTERPRISE: HOW TO POSITION THE DEPARTMENT OF DEFENSE TO MAXIMIZE THE USE OF THESE EMERGING TECHNOLOGIES

JUPITR

David L. Hirschberg

COLUMBIA UNIVERSITY
New York, NY 10032-2695

Kelley Betts

LEIDOS, INC.
Gunpowder, MD 21010-0068

Rebecca Hutchins

HUTCHINS CONSULTING, LLC
Stone Mountain, GA 30083-4776

Peter Emanuel

RESEARCH AND TECHNOLOGY DIRECTORATE

July 2015

U.S. ARMY RDECOM

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
XX-07-2015	Final	Mar 2014 - Aug 2014

4. TITLE AND SUBTITLE
DNA Sequencing Technologies within the Chemical and Biological Defense Enterprise: How to Position the Department of Defense to Maximize the Use of These Emerging Technologies—JUPITR

5a. CONTRACT NUMBER
Direct MIPR

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Hirschberg, David L. (Columbia University); Betts, Kelley (Leidos); Hutchins, Rebecca (Hutchins Consulting, LLC); and Emanuel, Peter (ECBC)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Columbia University, Center for Infection and Immunity; 722 W. 168th Street, New York, NY 10032-2695
Leidos, Inc.; P.O. Box 0063, Gunpowder, MD 21010-0063
Hutchins Consulting, LLC; 3976 Carlinswood Way, Stone Mountain, GA 30083-4776
Director, ECBC, ATTN: RDCB-DRB-M, APG, MD 21010-5424

8. PERFORMING ORGANIZATION REPORT NUMBER
ECBC-TR-1288

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Joint Program Executive Office for Chemical and Biological Defense, 5101 Hoadley Road, Aberdeen Proving Ground, MD 21010-5423

10. SPONSOR/MONITOR'S ACRONYM(S)
JPEO-CBD

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

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

13. SUPPLEMENTARY NOTES

14. ABSTRACT:
To generate accurate and actionable disease data that can be integrated into the broader Department of Defense response framework, it is critical to anticipate looming advances in genomic sciences while ensuring that the most advanced molecular diagnostic capabilities are available to chemical and biological defense programs. The recommendations detailed in this report are based on the assessments of the current state-of-the-art technologies for genomic sequencing and on how these technologies mature and evolve. Six recommendations were made for the use and improvement of deoxyribonucleic acid sequencing technologies to better support chemical and biological defense capabilities. This report especially addresses the 2014 time frame of these technologies and their future estimation. This report also provides recommendations to assess new technologies while wisely allocating resources to respond to emerging biological threats that directly affect the health and performance of service members and the risks they face.

15. SUBJECT TERMS
Sequencing
Molecular diagnostic
Web portal
Data analysis
Platform

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Renu B. Rastogi
U	U	U	UU	32	19b. TELEPHONE NUMBER *(include area code)* (410) 436-7545

Blank

PREFACE

The work described in this report was authorized under the Joint United States Forces Korea Portal and Integrated Threat Recognition Advanced Technology Demonstration.* The work was started in March 2014 and completed in August 2014.

The use of either trade or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

This report has been approved for public release.

*The Joint United States Forces Korea Portal and Integrated Threat Recognition Advanced Technology Demonstration is more commonly known by its acronym, JUPITR ATD.

Blank

CONTENTS

FIGURES

TABLES

DNA SEQUENCING TECHNOLOGIES WITHIN THE CHEMICAL AND BIOLOGICAL DEFENSE ENTERPRISE: HOW TO POSITION THE DEPARTMENT OF DEFENSE TO MAXIMIZE THE USE OF THESE EMERGING TECHNOLOGIES JUPITR[i]

1. INTRODUCTION AND BACKGROUND

Deoxyribonucleic acid (DNA) sequencing technology is used to determine the precise order of the four nucleotide bases (adenine, guanine, cytosine, and thymine) that form the DNA of an organism. DNA sequencing technology has now become indispensable for basic biological research, and it is used in numerous applied fields such as diagnostics, biotechnology, forensic biology, biological informatics, and epidemiology. The use of this technology has revealed complete DNA sequences or genomes of many life forms, including human, animal, plant, and microbial species. DNA sequencing is increasingly facilitating the rapid progression of biological and medical research through the identification of known and unknown genes involved in disease. Information gathered through DNA sequencing can be correlated with genotypic and phenotypic data to accurately identify a biological agent, and it provides insight into the development of novel diagnostics, surveillance methods, vaccines, and antimicrobials.

The first DNA sequences were obtained in the early 1970s by academic researchers using laborious methods based on one-dimensional chromatography (i.e., Maxam-Gilbert sequencing) that preceded Sanger sequencing, the system into which fluorescence was eventually incorporated (*1*). With the advent of methods that used automated analysis, DNA sequencing became easier and orders of magnitude faster (*2*). Within the next 20 years, DNA sequencing and the molecular biology that it supports will transition from the limited, expensive, and labor-intensive laboratory-based effort from which it began to a universally routine procedure performed at the bench top, field hospital, or clinic using fast, portable devices operated by minimally trained individuals.

As in the 1960s, when the electronic transistor paved the way for the cheaper, smaller computers and radios we use today, advances in sequencing technology are making DNA sequencing commonplace and affordable. Technological leaps that progressively enable more DNA information to be evaluated are occurring at a pace too swift to anticipate. Therefore, it is difficult to predict which technology will dominate the market within the next 3 years, let alone over the customary decade required for an average military procurement cycle. As hardware improvements become less disruptive (i.e., when performance improvements are introduced, but the underlying technology remains the same), future emphasis will shift from evaluating individual sequencing technologies to advancing supporting applications. The overarching recommendation for the Department of Defense (DoD) Chemical and Biological Defense Program (CBDP) is to develop clear-use scenarios that reflect mission requirements, harmonized reagents, metrics, and software capable of sustaining the U.S. military's current and future sequencing capabilities. A better understanding of how sequencing technology is evolving will allow the military to anticipate and enhance its future mission, leading to technological advantages.

[i] The Joint United States Forces Korea Portal and Integrated Threat Recognition

2. SUMMARY OF RECOMMENDATIONS

The following recommendations were proposed to ensure technological advantages for future U.S. military missions:

(1) Recommendation 1:

Develop clear-use cases for the sequencing technology platform. The CBDP science and technology (S&T) affiliates should develop clear-use cases for sequencing technologies to foster better concept of operations (CONOPS) and tactics, techniques, and procedures (TTPs). The clear-use cases should be used to simplify assay results, which would facilitate the delivery of rapid, actionable information to stakeholders.

(2) Recommendation 2:

Create an open website and portal. The CBDP S&T leadership should develop an open website and portal where sequencing technology performance results and all supporting technologies could be freely compared and discussed.

(3) Recommendation 3:

Establish a sequencing S&T panel of experts. The CBDP S&T leadership should convene an S&T panel of experts chaired by Defense Threat Reduction Agency Joint Science and Technology Office (DTRA JSTO) members to advise DoD leadership of the current state of sequencing technologies. The DTRA JSTO team could develop long-range plans for the evaluation, integration, and implementation of DNA technologies for CBDP stakeholders. The panel of experts should provide a yearly review of the stakeholders' portfolio of research projects with a stated goal of prioritizing research projects and fostering a flexible research environment.

(4) Recommendation 4:

Anticipate rapid advances in sequencing technologies and applications. The CBDP S&T members should create programs that incorporate advances in technologies that read nucleic acids with improved time and cost effectiveness. At the same time, CBDP S&T members should advance peripheral technologies (such as automated sample collection, preparation, loading analysis, and results transmission) that are capable of customizing specific-use applications for stakeholders and customers.

(5) Recommendation 5:

Select a single, commercially available sequencing platform for CBDP performer activities. To realize economies of scale and to tightly focus S&T investment for the service laboratories, the CBDP S&T leadership should select a single sequencing platform that is already in commercial production. The CBDP S&T leadership should also institute guidelines for equipment maintenance so that the latest upgrades, reagents, and software are readily available. The sequencing platform should directly connect with a sequencing data analysis environment, and subsequently, the open website should provide a solution for seamless data transfer, analysis, and reporting.

(6) Recommendation 6:

Evaluate different sample preparation and data analysis technologies. The CBDP S&T leadership should make it a priority to invest in an open, science driven evaluations of sample preparation and sequencing. These should include the development of sample standards that can be safely distributed and that can generate easily evaluated results. Also, as new technologies are released onto the market, multiple sequencing platforms would be included to ensure the single deployed platform is replaced when scientific evidence dictates the need.

3. RECOMMENDATIONS FOR THE FUTURE OF DNA SEQUENCING TECHNOLOGIES FOR THE MILITARY

3.1 Recommendation 1: Develop Clear-Use Cases that Drive Priority Research Projects

CBDP S&T members should develop clear-use cases for sequencing technologies to better develop CONOPS and TTPs. Also, assay results could be simplified to deliver rapid actionable information to stakeholders.

As sequencing technologies become part of the standard toolbox for sample identification and characterization, there will be a need to standardize protocols into clear-use cases so that users understand how and when to use an assay or technology. Users need clear, realistic expectations of the types of results that can be obtained by using a method. Figure 1 depicts potential workflows from nucleic acid material to sequencing data results. As clear-use cases are developed, appropriate technology and methods can be prioritized for research, development, and deployment. Ideally, automated sequencing assays, which take advantage of barcoding to identify known biological organisms, should be developed so that minimal data analyses and interpretation would be required. The exception would be cases in which simpler assays would not suffice; therefore, advanced assays and targeted analyses would need to be performed by highly trained personnel.

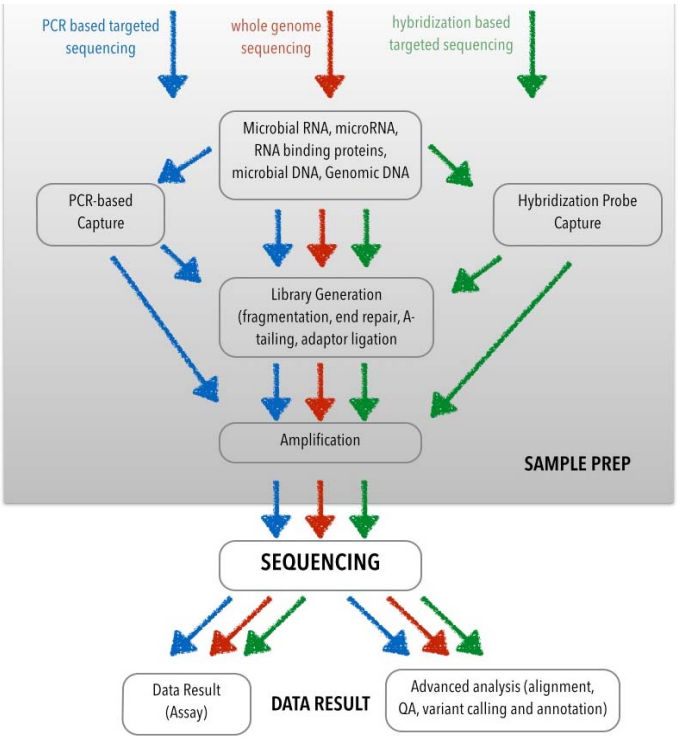

Figure 1. Clear-use cases for sequencing technologies.

The following assays should be developed in support of Recommendation 1:

(1) Assay systems designed to identify and characterize known pathogens: These assay kits could contain tens to hundreds of primers and probes that would interrogate a sample, bind to it, and amplify any regions of interest, if present. These libraries would be sequenced and mapped to provide rapid identification, confirmation, and information regarding virulence and antimicrobial resistance of the pathogen.

(2) Assays that capture novel pathogens for advanced investigation could be characterized by more-sophisticated data analyses, with comparison, alignments, and genome assembly performed by highly trained personnel: Investment in alternate approaches to pathogen detection and identification remains critical to the groups that find themselves at the vanguard of biological defense. Only an estimated 1% of the total potential pathogens in our environment have been identified. Current pathogen analysis algorithms are based only on the information collected to date and on how the organism of interest correlates with disease. Use highly trained individuals to target agnostic assays, which are designed for use on ubiquitous and readily available platforms in conjunction with improved software analysis programs. The assays could be employed to probe, evaluate, and identify new classes of organisms capable of posing a threat, either by themselves or as a co-infection with known organisms. Once validated, the information

and analysis workflow could be uploaded to existing workflows and databases so that all laboratories gain the capability to identify newly discovered pathogens. Novel pathogen discovery could be performed in a central reach-back laboratory with strong communication links to other satellite laboratories with homogenous core equipment, protocols, and cross-training and communications practices similar to all participating laboratories. Because of this standardization, new protocols designed for defined, specific-use cases could be quickly and easily pushed out to relevant laboratories once protocols have been optimized for routine implementation.

(3) Assays to enhance the recovery and sequencing of trace amounts of DNA from pre- and post-detonation improvised explosive devices (IEDs): Recent research in the area of human identification has shown that DNA recovery from IED containers is a viable approach. Trace amounts of DNA found on IEDs were used to identify individuals who had contact with containers and were potentially involved in IED events (*3*).

(4) Assays to enhance the recovery and sequencing of trace and higher levels of DNA evidence from sites and transportation routes of adversarial activity (e.g., IED fabrication lairs, heroin-processing labs, and border portals): The ability to bioprofile illicit drugs, IED construction tools, and transportation routes could provide key insights into how physical assets and the humans who come into contact with them are moved.

(5) A curated relational database or similar system that could provide rapid and accurate correlation of DNA sequence information: This would allow facile exploitation of materials recovered from various sites and sources and provide rapid actionable intelligence. This database should be used to accept data from classified and unclassified sources, and it should also have the capacity to limit data analysis and correlation for only classified purposes.

3.2 Recommendation 2: Create a Web Portal

CBDP S&T leadership should develop a web portal that allows the direct, real-time comparison of sequencing results from several sequencing applications with other nucleic acid assay techniques (i.e., BioFire [BioFire Diagnostics, LLC; Clearwater, UT] , MagPix [Luminex Corporation, Austin, TX], and real-time polymerase chain reaction assays). Additionally, the web portal should be able to accept testing results from orthogonal testing methods that do not use nucleic acid as an input, thereby providing a single source of information regarding a sample. This portal should leverage the flexibility and cost savings of open-source software while establishing standards for the description, exchange, and integration of genomic data.

There is an immediate need to display and discuss results generated on the current set of technologies in real-time for the following reasons:

- Technical problems and unusual results could be rapidly analyzed
- Discordant results could be resolved
- Putative positives could be confirmed

A web portal would build on the existing capability of Pathosphere.org ([Pathosphere.org, Forrest Hill, MD], Figure 2). Pathosphere.org is a bioinformatics framework, government off-the-shelf solution currently used by sequencing laboratories in and outside the contiguous United States to support pathogen detection and assay development needs. Serving as a collaborative hub for communications, software development, and file sharing, the Pathosphere.org open-system architecture is now deployed and available on the Amazon cloud. The site has over 100 users from universities, medical centers, and organizations (e.g., the Centers for Disease Control, Food and Drug Administration, Naval Medical Research Unit No. 6, and the U.S. Army Medical Research Institute of Infectious Diseases).[ii] Pathosphere.org has established a common web front end capable of capturing dynamic and changing sets of analytical features. It can be used to create a web-based dashboard that brings together different types of data from several platforms, along with scientific and technical interpretation and discussion in real time.

[ii] The following organizations are more commonly known by their acronyms: Centers for Disease Control, CDC; Food and Drug Administration, FDA; Naval Medical Research Unit No. 6, NAMRU-6; and the U.S. Army Medical Research Institute of Infectious Diseases, USAMRIID.

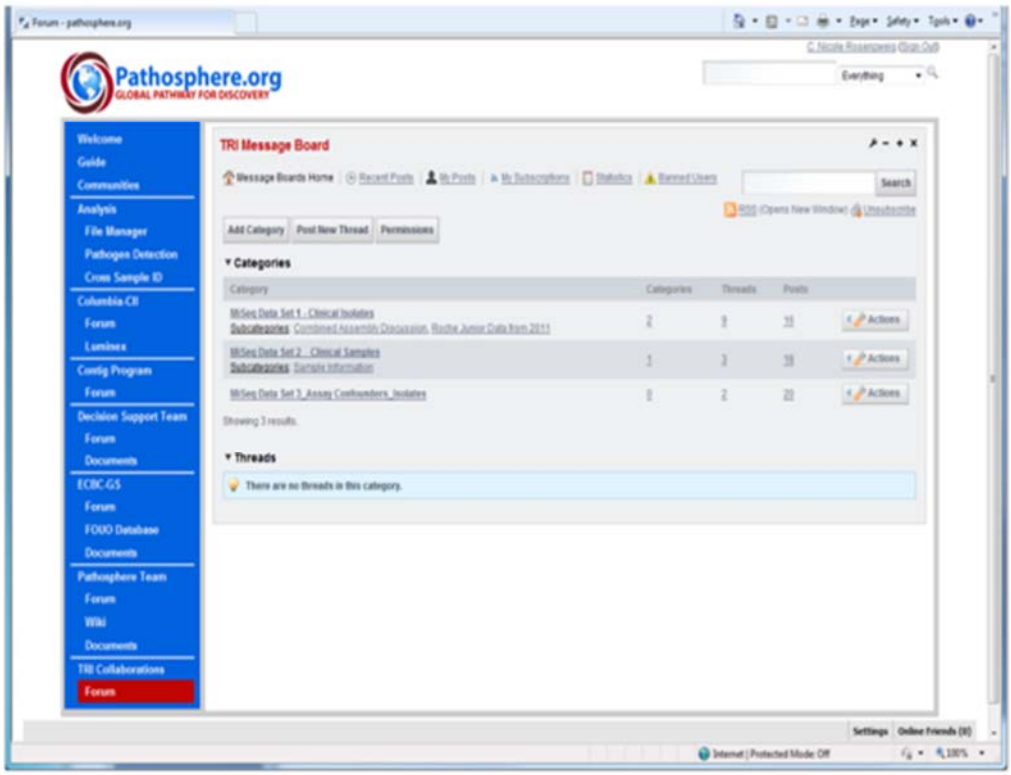

Figure 2. Pathosphere.org snapshot.

3.3 Recommendation 3: Institute a Sequencing S&T Panel of Experts

The CBDP S&T stakeholders should establish an S&T panel of experts comprised of leaders in the field from government institutions, academic world, and industry to provide insight into the current state of sequencing sciences. The CBDP S&T stakeholders should also develop long-range plans for the evaluation, evolution, and implementation of DNA technologies. This panel should include representatives with a variety of expertise in areas such as sample collection, handling, and archiving; method validation and standardization; sequencing technologies, bioinformatics, database development, and other genomic sciences. The field of sequencing is rapidly shifting from evaluation to applications. An expert panel will be necessary to gather, assess, filter, and integrate the deluge of protocols, data, and applications available now and in the future. Individuals directly involved in this process should have a strong understanding of what sequencing technologies can offer as well as their limitations. The panel members' roles will be to not only provide relevant recommendations to the stakeholders but also to anticipate future needs and to guide the research project portfolio to fill those needs. The panel should include

a fixed cadre of trusted experts who have sufficient scientific stature to reach out to their counterparts for focusing on particular or urgent concerns when the need arises.

3.4 Recommendation 4: Anticipate Rapid Advances in Sequencing Technologies and Applications

The CBDP S&T leadership should build programs that could anticipate advances in technologies that read nucleic acids and that advance peripheral technologies (e.g., automated sample collection, preparation, loading analysis, and results transmission) capable of customizing these technologies for specific uses.

The increased demand for low-cost sequencing has significantly advanced the development of next-generation or high-throughput sequencing technologies. These technologies have considerably reduced the cost of large-scale sequencing, and advances continue to be made. In addition to pathogens and vectors, an individual's genome, metagenome, and environmental genome can now be analyzed. Next-generation sequencing technologies also facilitate further advances in actionable decision making and threat prediction in specific geographical areas. However, rapid advances can challenge the validation and use of these constantly evolving technologies and platforms in central and remote laboratories.

Centralized standards and guidelines for evaluation will need to be developed to assist laboratories with the validation of next-generation sequencing methods and platforms, the ongoing monitoring of next-generation sequencing testing to ensure quality results, and the interpretation and reporting of new entities using these technologies. At least three laboratories with highly trained individuals should be designated as exemplar laboratories and serve as high-speed protocol development hubs (Figure 3). The exemplar laboratories should develop working protocols to integrate a number of diverse technologies that could be pushed out to remote laboratories. The remote laboratories should include diverse sets of multiplex nucleic acid and antibody-based assays as well as multiple sequencing platforms that would allow comparison and reporting of direct sample preparation technologies. These laboratories should have the personnel, processes, and procedures in place to rapidly perform the testing, protocol optimization, and data analysis needed to support the deployment of new methods and technologies within a one-year time frame. Stakeholders, in collaboration with the S&T panel of experts, should be able to provide successful research and development data to an exemplar laboratory through the web portal. The exemplar laboratories should then provide the repeated testing and statistical analysis necessary to fully validate and deploy a method.

3.5 Recommendation 5: Select a Single, Commercially Available Sequencing Platform for Service Laboratories

The CBDP S&T stakeholders should select a single, commercially available sequencing platform for the service labs and be able to maintain the equipment so that it has the latest upgrades, reagents, and software. This would allow the military to capitalize on evolving and available technological innovations while leveraging minimized operational costs with economies of scale. Implementation of a single sequencing platform would better meet Warfighter needs by allowing the comparison of different methods, training, and performance levels between

service laboratories. At the time of this report, very few sample-to-analysis assays that can be broadly deployed on existing sequencing platforms were available. Although several diagnostic assays capable of identifying pathogens in complex matrices are currently under development, each distinctive assay requires individualized optimization for every applicable unique sequencing platform, thus increasing time and expense. Use of a single sequencing platform would ensure reliability, normalized distribution, streamlined logistical support and training, minimized validation times across participating laboratories, and standardized data dissemination across the services as a whole.

According to the manufacturer, the chosen platform should be properly maintained with the latest hardware and software upgrades, and all consumables (i.e., reagents, tubes, etc.) should be monitored for availability and expiration. Once a platform is chosen, a secondary and tertiary platform should be selected and placed in exemplar laboratories to allow cross validation and performance comparison. The exemplar laboratories should also provide reach-back support when an alternative sequencing method is necessary. All platforms have varying differences that cause them to perform disparately with diverse samples and conditions. Some situations allow samples run on a PacBio (Pacific Biosciences, Menlo Park, CA) to give a more actionable result than the same samples run on a MiSeq (Illumina, Inc.; San Diego, CA) or on a Ion Torrent (Life Technologies, Thermo Fisher Scientific Corporation; Carlsbad, CA). These samples require very long reads from single ribonucleic acid (RNA) virus quasi-species. Table 1 provides an overview of currently available sequencing platforms and their specifications.

Table 1. Comparison of Sequencing Platforms

System	454 GS Jr. Titaniu	454 FLX+	MiSeq	NextSeq 500	HiSeq 2500	Ion Torrent PGM	Ion Torrent Proton
Sequencing Chemistry	Pyrosequencing	Pyrosequencing	Polymerase-based	Polymerase-based	Polymerase-based	Ion semiconductor	Ion semiconductor
Amplification Approach	Emulsion PCR	Emulsion PCR	Bridge PCR	Bridge PCR	Bridge PCR	Emulsion PCR	Emulsion PCR
Sample Capacity	1 cell/plate	Up to 16 cells /plate	1 lane	1 lane	8 lanes x 2 flowcells (16 total)	1	1
Multiplexing Capacity	Up to 132	Up to 132	Up to 96	Up to 96	Up to 96	Up to 96	Up to 96
Mega / Gigabases per cell or lane	50 MB	650 MB	3.8GB (2x75) 15GB (2x150)	30 GB (1x75) 120 GB (2x150)	144 GB (1 x 36) 1000 GB (2x125)	30-50 MB (1x200) 60-100MB (1x400) 314 300-500 MB (1x200) 600-1GB (1x400) 316 800-1GB (1x200) 1600 - 2GB (1x400) 318	10 GB Proton I Chip 40-80 GB Proton II Chip (in development)
Reads per lane / cell	0.1 M	1 M	150 M (2x75) 600 M (2x150)	150M (2x75) 300M (2x150)	1500M 2000M	450K 314 2.5M 316 4.7M 318	70-80M Proton I Chip 260-320M Proton II Chip (in development)
Time per run from library to raw sequence	10h	10h	20h (2x75) 55h (2x150)	18h (2x75) 30h (2x150)	2 days (1x36) 6 days (2x125)	2.3h (1x200) 3.4h (1x400) 314 3h (1x200) 4.9h (1x400) 316 4.4h (1x200) 7.3h (1x400) 318	4h (Proton I Chip) 6h (Proton I Chip est.)
Read length	400	700	75 or 150	75 or 150	36 or 125	100, 200 or 400	200
Cost per run	$980	$6,200	$825 (2x75) $1450 (2x150)	$2500 (2x75) $4000 (2x150)	$5875 (1x36) $15,000 (2x125)	$349 (1x200) $474 (1x400) 314 $549 (1x200) $674 (1x400) 316 $674 (1x200) $749 (1x400) 318	$1000 (Proton I Chip) $1000 (Proton I Chip est.)
Cost per Megabase	$19.50	$9.54	$0.217 (2x75) $0.096 (2x150)	$0.0833 (1x75) $0.0333 (2x150)	$0.0407 (1x36) $0.0150 (2x125)	$6.98 (1x200) $4.74 (1x400) 314 $1.09 (1x200) $0.67 (1x400) 316 $0.67 (1x200) $0.37 (1x400) 318	$1.00 Proton I Chip $0.25 Proton II Chip (est. @ 40GB)
Cost per Instrument (including ancillary template prep instrumentation from vendor)	$124K	$480K	$99K	$250K	$690K	$78K	$243K

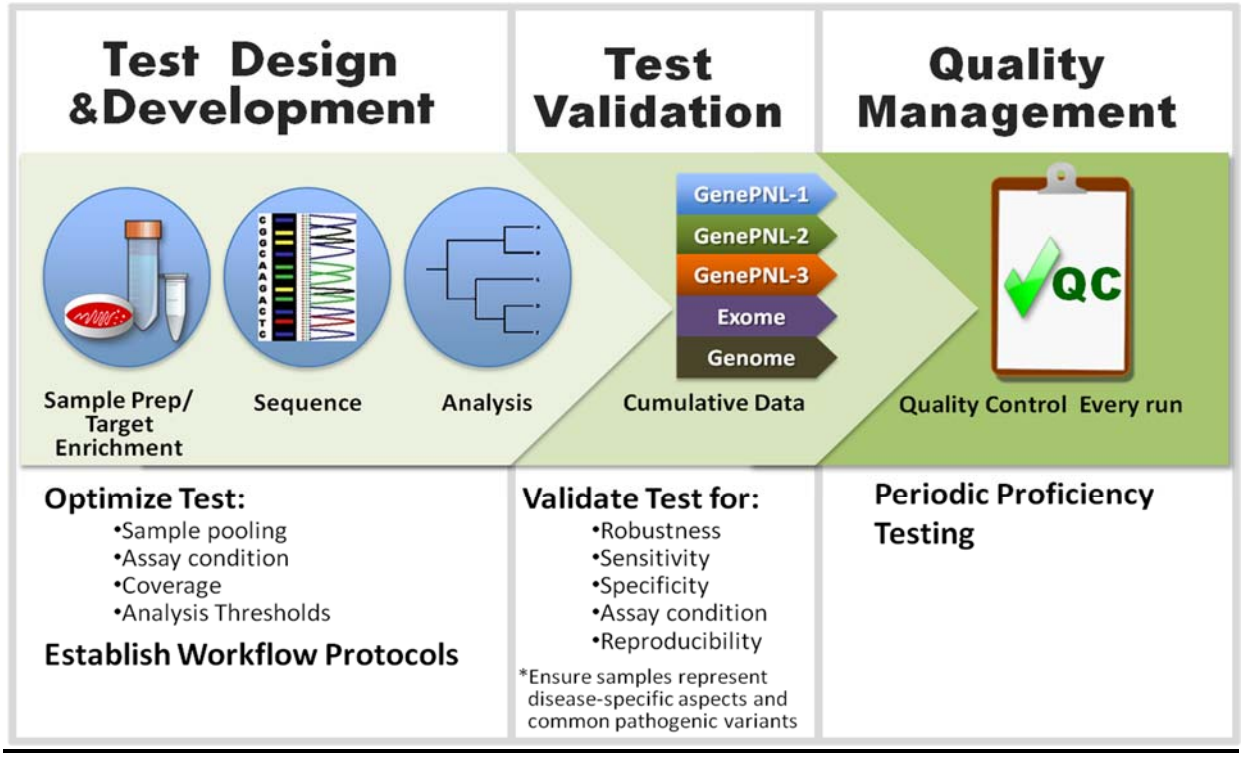

Figure 3. Development and optimization of sequencing technologies.

3.6 Recommendation 6: Evaluate Different Sample Preparation and Data Analysis Technologies

To provide solutions to the specific-use cases identified as priority issues, CBDP S&T members should make it a priority to invest in a comparison of sample preparation and analysis technologies for sequencing platforms. Proper sample preparation technologies that deliver high-quality genetic material are vital for the success of nucleic acid-based pathogen detection. The improved sample preparation techniques have enabled the recovery of very high-quality assemblies from 400,000 year-old Neanderthal samples that correspond with the quality of those collected from a living person (*4*). The current state-of-the-art technology has also allowed for the high-quality sequencing of mammalian samples dating back 700,000 years (*5*). Nucleic acid extraction platforms can be optimized in several areas; however, individual platforms can perform differently in all of these areas. In a recent study focusing on metagenomic samples, the utility of six emerging, commercial sample preparation technologies were evaluated for use within a pipeline and high-throughput sequencing (Table 2).

The following areas of consideration were advocated for Recommendation 6.

(1) Microbial Total Nucleic Acid (TNA) Extraction

TNA extraction is defined as the extraction of microbial RNA and DNA from clinical or environmental samples, without the isolation and expansion of the organism, using plating or tissue culture techniques. In many cases, microbial nucleic acid is found in very small amounts relative to the host's nucleic acid, and many different organisms or strains are also present in the sample. Additionally, microbial nucleic acid is often unstable and partially degraded. Many nucleic acid platforms have been validated and optimized on monocultures of bacteria. Bead size, wash protocols, and reagent stringency can contribute to the amount, size, and quality of the nucleic acid that is collected. This may lead to the preferential recovery of large amounts of host DNA over smaller viral nucleic acids that are generally present at low concentrations in a sample.

(2) Ability to Optimize

The ability to optimize is assessed by how receptive the platform is to modified protocols. Some platforms allow the user to modify protocols, enabling modification of washes, addition of different buffers, and changes in elution volumes and incubation times. For example, this ability allows flexibility in processing blood samples as opposed to processing fecal samples. Some platforms, such as the MagNA Pure compact 96 ([Roche Diagnostics, Indianapolis, IN] data not included in this report), are designed for Clinical Laboratories Improvement Amendment laboratories, and their software does not allow the user to change the parameters of the protocol. Other platforms, such as the Kingfisher, allow a research and development menu that can be used to develop and optimize protocols for a specific sample. The number of available sample preparation kits and the number of protocols accessible to users were also considered.

(3) Sample Throughput

Many labs have variable needs depending on their project size and sample sources. Clinical samples are usually associated with an outbreak or a single case, requiring a processing capability range of many samples down to a single sample. For instance, an environmental experiment or a large-scale clinical study may require the processing of hundreds to thousands of samples. Platforms that excel in this area are able to handle a few to hundreds of samples with no major time losses. In some cases, there is a trade-off with sample scalability, as was found with the Roche platform, which required the increased consumption of reagents.

(4) Sample Yield

Sample yield is the percent of the total possible amount of nucleic acid relevant to the study that is recovered from a given sample. For example, if a blood sample has 500 copies/mL of a negative-strand RNA virus, the percent of nucleic acid recovered would be specific for that particular virus and not for the total nucleic acid found in the sample. Some platforms perform better than others, depending on their sample source (e.g., some platforms use larger volumes for their initial extraction and have lysis buffers not optimized for recovering small amounts of viral nucleic acid). This metric takes into account microbial nucleic acid that has been recovered from complex sources (i.e., tissue, blood, feces, the environment, etc.). In many cases, such as with human biopsy material where large amounts of DNA can be recovered, the platform could still receive a low score if bacterial or viral nucleic acids present in the sample were lost.

(5) Reproducibility

Reproducibility was based on the overall comparison over time of a particular platform to itself or to a similar model used in other laboratories over time. Even if a platform had low yield efficiency, it was still considered of "good utility" as long as enough material was recovered for a measurement to be taken. Additionally, recovery values should not fluctuate because of platform performance variability or quality issues with the reagents. Given that the scope of this report did not allow for detailed comparisons using standardized samples, much of the information used for direct comparisons was gathered and compared.

Small robotic or microfluidic stations that are capable of receiving, extracting, and assessing the quality of recovered nucleic acid or protein; preparing and finally transferring samples to the sequencing platform; and performing the assay will be essential for deploying these technologies successfully. Although different applications will require different input systems, new sample preparation methods will allow for the sequencing of non-coding nucleic acids such as microRNAs and protein-binding RNAs. These will play a critical role in determining how a genome is expressed and in providing critical insight into how microbes invade and usurp the cellular pathways of the eukaryotic cell.

Table 2. Ranking of Nucleic Acid Extraction Platforms

	Technology	Microbial TNA extractrion	Ability to Optimize	Sample Throughput (scale)	Sample Yield	Reproducability	Score	Rank
	Biomerieux EasyMag	95	80	85	95	95	90	1
	Boreal Aurora (SCODA)	60	35	10	85	60	50	6
	Thermoscientific Kingfisher Duo	60	95	95	85	90	85	2
	Roche MagNA Pure Compact	60	60	85	65	80	70	3
	Spriworks TE	60	70	60	65	70	65	4
	Nordiag Arrow	40	65	65	55	75	60	5

4. UP-AND-COMING SEQUENCING TECHNOLOGIES

The up-and-coming sequencing technologies discussed in this section were considered for future use.

4.1 Reagent-Free Sequencing

Technologies, such as the Quantum sequencing platform (Quantum Biosystems [QB], Osaka, Japan) allow for the direct sequencing of single-stranded DNA and RNA (*13*). With no need for labeling or modifications, sequencing occurs directly on silicon devices that are capable of being fabricated on the same production lines as commercial integrated circuits. The QB platform makes use of picoamp level currents and sub-nanometer gaps to detect the conductance of single DNA and RNA molecules and promises to bring about fundamentally new types of sensors allowing for the rapid, inexpensive, and reagent-free high-throughput analysis of whole genome sequences.

4.2 Solid-State Sequencing

Technologies, such as the Oxford Nanopore (Oxford Nanopore Technologies, Oxford, UK) take a single-molecule-sequencing approach. A nanopore can be engineered by drilling a nano-scale opening into a solid-state substrate. As DNA passes through the nanopore, the shift in ionic current is recorded and can be used to infer the DNA sequence (*6*). Several commercial companies are currently investing in nanopore technology, and it is an area of active research in university laboratories. The ability to use an inexpensive and portable sequencing device, such as the Universal Serial Bus-enabled device announced by Oxford Nanopore Technologies, would open up many possibilities for near real-time information gathering during investigations in diverse locations.

4.3 Handling Massive Amounts of Data

Deployed in laboratories across the world, next-generation sequencers are capable of generating thousands or millions of sequences concurrently. The enormous amount of sequence data produced by these advances has spurred the development of new sequence analysis tools that are capable of evaluating these massive amounts of data. As sequencing technologies become more ubiquitous and affordable, the potential amounts of data derived from them are unfathomable. The military should anticipate the need for better analysis tools and evaluate technologies capable of accommodating the tremendous amounts of data produced from the routine and inexpensive sequencing of samples. These technologies should include software that rapidly writes information to sequence and index technologies that can quickly retrieve specific requests for information stored in DNA-based archives. In addition, the technologies would be used to decipher and interpret the sequence data into actionable information.

4.4 DNA as Archival Storage Material

The evaluation of DNA technologies to support or replace modern long-term data storage is not new to the scientific community (*7*), and the military should anticipate and take into account solid-state technologies that encode, store, transmit, and read out non-biological information to and from nucleic acids (Figure 4). When stored correctly, DNA, the molecule responsible for encoding the genetic instructions that direct the form and functioning of every living thing, is stable for millions of years. DNA is an extremely dense and stable molecule. It is the ideal material for archival storage of large amounts of data. However, unlike decades-old electronic data that is no longer easily retrievable by 21st century computing methods, DNA's track record as a data storage media dates back to around 3.5 billion years. The DNA molecule, which has evolved through the millennium to encode, copy, and disseminate information with great fidelity by performing highly-specific pattern matching and sequence recognition via nucleotide base pairing, will continue to serve as nature's preferred data storage material for millennia to come.

Figure 4. DNA as archival storage material.

It has been estimated that all of the present stored digital information in the world ranges somewhere on the order of 1 ZB[iii] (*8*). DNA's high stability gives it the potential storage capacity of a staggering 2.2 PB[iv]/g (approximately 14,000 fifty gigabyte Blu-ray discs) (*9*,*10*). As the technology continues to develop, archival costs for 50+ years of storage will likely drop to competitive levels within the next 10 years (Figure 5). High information density, ability to survive within the natural world, and built-in capacity to replicate stored information make DNA an intriguing, if not powerful, storage material for archival applications.

[iii] ZB: zettabyte
[iv] PB: petabyte

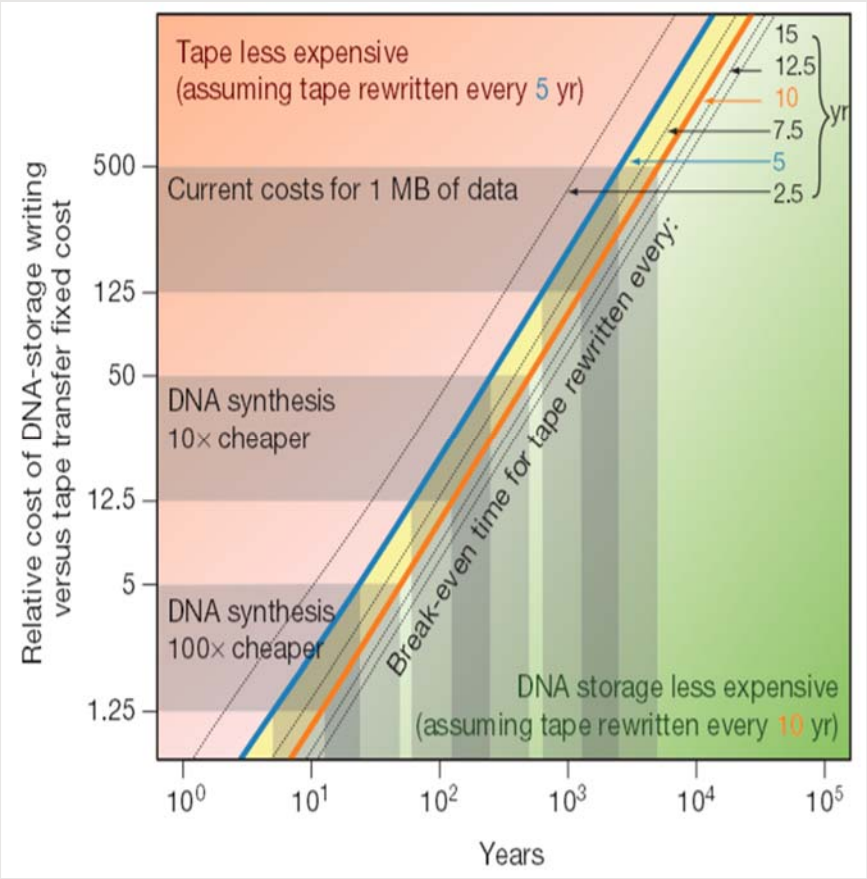

Figure 5. Storage cost for DNA vs tape (*9*).

4.5 3D Printing Technologies for Nucleic Acid Assays, Probes and Reagents

The military should continue to evaluate additive manufacturing or 3D printing technologies that will produce synthetic tissues, organs, vaccines, and diagnostic and surveillance instruments, including nucleic acid probes and reagents. 3D printing has the potential to create tailored, inexpensive assays and sensors that have historically required dedicated manufacturing facilities. Using 3D printing technologies, portable, reconfigurable reactionware; microfluidics; and reagents can be rapidly created onsite from digitally controlled, open-source design software that is downloaded from a tablet or the internet (*11*). Designed by computing big data from the cloud, highly specialized, de novo compounds could be quickly and precisely fabricated as needed in a single step without the usual reagent waste, environmental impact, or specialized manufacturing (*12*). 3D printers (Figure 6) can be used to deposit a selection of reagents (e.g., amino acids, nucleotides, and buffers) in a "plug and play" scenario in which digital biology is transmitted at the speed of light through the cloud and materialized at the end for the user. The advantages of additive manufacturing are boundless and so are the threats it could pose. The majority of commercial 3D printers operate using plastic inks, but these devices can be made to work with reagents of a completely different nature. Because 3D printers can be used to create

vaccines, nucleic acids, and reagents onsite, rogue organizations could hijack this technology to produce biological or chemical weapons in the not-so-distant future.

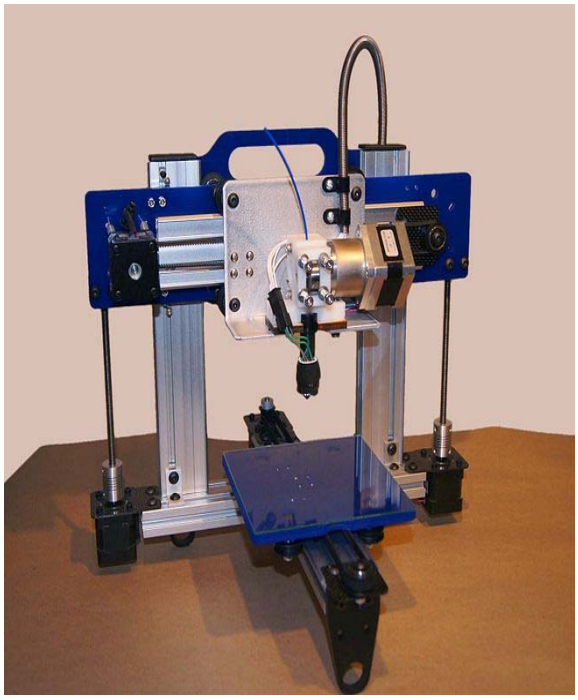

Figure 6. 3D printer.

4.6 Emerging Enhanced and Engineered Pathogens

Solid-state biological technologies that enable the rapid genetic characterization of emerging, enhanced, and engineered pathogens and that recognize biological signatures (e.g., environmental microbiomes or a host response to pathogens or toxins) should be evaluated. The ability to recognize biological signatures is a significant advancement in the disciplines of genomic biosurveillance, epidemiology, and countermeasure development efforts. Also, the ability to rapidly decode the genome is essential to detecting newly acquired or deliberately inserted genetic material. CBDP S&T members should appreciate and anticipate the need to distinguish between these classes of agents by investing in technologies that not only recognize, identify, and characterize a biological threat but also determine whether that threat is a result of intentional manipulation. This capability is critical to decision makers because the information that is garnered guides the development of targeted therapies directed towards treating infections from these agents (i.e., determining antibiotic resistance).

5. FINAL SUMMARY

The DNA sciences of tomorrow are unlikely to resemble the DNA sciences of today. In the way that computers, tablets, and smartphones have stretched the bounds of how we perceive and use communications data, the rapidly evolving science of DNA sequencing will change the way in which we use and leverage DNA data. As this evolving technology continuously changes industry standards through the constant improvement of sequencing products and capabilities, CBDP S&T leadership must anticipate future sequencing technology trends so that new capabilities and processes can be developed. These capabilities and processes should not only meet the ever-changing requirements of the CBDP members but should also provide best value for these technologies while continuing to address and meet stakeholder needs. By establishing an S&T panel of experts and exemplar laboratories equipped to provide high-speed protocol development support, CBDP members will be able to provide their network of service laboratories with state-of-the-art capabilities required to support service members.

REFERENCES

1. Smith, L.M.; Sanders, J.Z.; Kaiser, R.J.; Hughes, P.; Dodd, C.; Connell, C.R.; Heiner, C.; Kent, S.B.; Hood, L.E. Fluorescence Detection in Automated DNA Sequence Analysis. *Nature* **1986**, *321*(6071) 674–679.

2. Luckey, J.A.; Drossman, H.; Kostichka, A.J.; Mead, D.A.; D'Cunha, J.; Norris, T.B. High Speed DNA Sequencing by Capillary Electrophoresis. *Nucleic Acids Res.* **1990**, *18*(15) 4417–4421.

3. Hoffmann, S.G.; Stallworth, S.E.; Foran, D.R. Investigative Studies into the Recovery of DNA from Improvised Explosive Device Containers. *J. Forensic Sci.* **2012**, *57*(3) 602–609.

4. Prufer, K.; Racimo, F.; Patterson, N.; Jay, F.; Sankararaman, S. The Complete Genome Sequence of a Neanderthal from the Altai Mountains. *Nature* **2014**, *505*(7481) 43–49.

5. Lee, J.J. World's Oldest Genome Sequenced from 700,000-Year-Old Horse DNA. *National Geographic*, 26 June 2013.

6. Maitra, R.D.; Kim, J.; Dunbar, W.B. Recent Advances in Nanopore Sequencing. *Electrophoresis* **2012**, *33*(23) 3418–3428.

7. O'Driscoll, A.; Sleator, R.D. Synthetic DNA: The Next Generation of Big Data Storage. *Bioengineered* **2013**, *4*(3) 123–125.

8. Gartz, J.; Reinsel, D. Extracting Value from Chaos. IDC iView, 2011. http://www.informationlifecyclemanagement.net/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf (accessed May 2014).

9. Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E.M.; Sipos, B.; Birney, E. Towards Practical, High-Capacity, Low-Maintenance Information Storage in Synthesized DNA. *Nature* **2013**, *494*(7435) 77–80.

10. Church, G.M.; Gao, Y.; Kosuri, S. Next-Generation Digital Information Storage in DNA. *Science* **2012**, *337*(6102) 1628.

11. Symes, M.D.; Kitson, P.J.; Yan, J.; Richmond, C.J.; Cooper, G.J. Integrated 3D-Printed Reactionware for Chemical Synthesis and Analysis. *Nat. Chem.* **2012**, *4*(5) 349–354.

12. Chen, C.Y.-C. DrugPrinter: Print any Drug Instantly, *Drug Discovery Today* (2014). http://dx.doi.org/10.1016/j.drudis.2014.03.027 (accessed May 2014).

13. Quantum Biosystems Demonstrates First Reads Using Quantum Single Molecule Sequencing, *Nanowerk News*, 28 January 2014. http://www.nanowerk.com/nanotechnology_news/newsid=34157.php#ixzz2zodAjOLH (accessed May 2014).

ACRONYMS AND ABBREVIATIONS

CBDP	Chemical and Biological Defense Program
CDC	Centers for Disease Control
CONOPS	concept of operations
DNA	deoxyribonucleic acid
DoD	Department of Defense
DTRA JSTO	Defense Threat Reduction Agency Joint Science and Technology Office
FDA	Food and Drug Administration
IED	improvised explosive devices
JUPITR	Joint United States Forces Korea Portal and Integrated Threat Recognition
NAMRU-6	Naval Medical Research Unit No. 6
PB	petabyte
QB	Quantum Biosystems
RNA	ribonucleic acid
S&T	science and technology
TNA	total nucleic acid
TTPs	tactics, techniques, and procedures
USAMRIID	U.S. Army Medical Research Institute of Infectious Diseases
ZB	zettabyte

DISTRIBUTION LIST

The following individuals and organizations were provided with one Adobe portable document format (pdf) electronic version of this report:

ECBC BIOSCIENCES
RDCB-DRB
Attn: Emanuel, P.

ECBC BIOSCIENCES
RDCB-DRB
Attn: Chue, C.

Defense Threat Reduction Agency
DTRA/RD-CBD T
Attn: Ward, T.
J9-CBS
Attn: Moore, E.

Department of Homeland Security
DHS ORD CSAC
Attn: Famini, G.

G-3 History Office
U.S. Army RDECOM
Attn: Smart, J.

ECBC Technical Library
RDCB-DRB-BL
Attn: Foppiano, S.
Stein, J.

Office of the Chief Counsel
AMSRD-CC
Attn: Upchurch, V.

Defense Technical Information Center
Attn: DTIC OA

ECBC Rock Island
RDCB-DES
Attn: Lee, K.


EDGEWOOD CHEMICAL BIOLOGICAL CENTER
A U.S. Army RDECOM Laboratory