Step I: White Paper Application

Application Guidelines

1. *The application should be submitted electronically per requirements via the web site of any of the NIAID Genomic Sequencing Centers for Infectious Diseases. Include all attachments, if any, to the application.*
2. *There are no submission deadlines; white papers can be submitted at anytime.*
3. *GSC personnel at any of the three Centers can assist / guide you in preparing the white paper.*
4. *Investigators can expect to receive a response within 4-6 weeks after submission.*
5. *Upon approval of the white paper, the NIAID Project Officer will assign the project to a NIAID GSC to develop a management plan in conjunction with the participating scientists.*

White Paper Application

Project Title: Whole Genome Sequence Analysis of Carbapenem-resistant Enterobacteriaceae Isolated from 3 Boston Area Medical Institutions

Authors: Cheryl Murphy, Lisa Cosimi, Michael Feldgarden

Primary Investigator Contact:

Name	Cheryl Murphy
Position	Associate Director, Microbial Genomics
Institution	The Broad Institute
Address	7 Cambridge Center, Cambridge
State	MA
ZIP Code	02142
Telephone	617-714-7893
Fax	617-714-7897
E-Mail	cmurphy@broadinstitute.org

1. Executive Summary *(Please limit to 500 words.)*

Provide an executive summary of the proposal.

One of the most serious problems in global public health today is the increasing incidence of antibiotic resistant bacterial infections. Of particular concern is the emergence of *Enterobacteriaceae* that are resistant to carbapenems, the class of antibiotic considered the last resort for treating increasingly resistant organisms. While uncommon prior to 1992, their prevalence has risen at an alarming rate with reports of 4-11% among clinical isolates in some networks (1, 2). Not surprisingly, infection with these resistant pathogens is associated with increased mortality (3).

Resistance to carbapenems is primarily due to acquisition of carbapenem-hydrolyzing β-lactamases or carbapenemases (4, 5). Among the enterobacteria harboring carbapenemase-encoding genes, *Klebsiella pneumoniae* is the most common (6), having rapidly become endemic in many hospital settings around the world. Outbreaks of carbapenemase-producing *K. pneumoniae* have been reported in at least 36 states in the United States though predominantly in the northeast (1, 7), Brazil (8), Israel (9), Greece (10) and India (11), as well as sporadic detection of isolates in other parts of the world (6).

While a large variety of these enzymes belong to three classes of β-lactamases, the rapid global spread of carbapenem-resistant *Enterobacteriaceae* has largely been attributed to transmission of clonal strains (e.g., ST 258), with horizontal gene transfer of the KPC β-lactamase being the most common. These genes are associated with a Tn*3*-based transposon, Tn*4401*, suggesting that this transposon contributes to the mobilization and dissemination of the *KPC* genes (12-14).

Despite recent analyses of a number of hospital-associated outbreaks, studies that examine and track the genetic relationship of large collections of carbapenem-resistant bacterial isolates from geographically distinct healthcare facilities are lacking. Further, such analyses have used typing methods (Pulse Field Gel Electrophoresis or MultiLocus

Sequence Typing) that lack the level of resolution often required to discriminate individual isolates. Such resolution is usually necessary to inform tracking of sources, transmission patterns, and pathogen evolution (15). A whole genome sequencing approach will increase the understanding of the specific genetic mechanisms underlying phenotypic differences in drug susceptibility in enterobacterial species, while providing the nucleotide-based resolution of strains required to provide real time feedback on the spread of antibiotic resistance in a clinical setting.

Through this effort we will aim to:
1) Understand the epidemiology of carbapenem resistance in our partner hospitals.
2) Better understand mechanisms of resistance as we search for novel β-lactamase genes in the sequenced organisms and study whether the plasmid structure or resistance plasmid gene content vary.
3) Study the mechanism of the spread of resistance (e.g. through clonal spread vs. gene transfer).

In phase I, we will compare the genomic sequences of prospectively sampled carbapenem-resistant enterobacteria in three major medical institutions in Boston, and will monitor changes in sequenced isolates over a six to nine month period within each institution. We will also sequence a number of retrospectively collected isolates of carbapenem-resistant and susceptible strains from each institution for comparative purposes.

In phase II of this analysis we will expand the prospective tracking and sequencing of organisms to include both additional susceptible and non-susceptible species within a facility as well as carbapenem-resistant strains from health care facilities located in other geographic locations to further understand the mechanism of the spread of resistance.

2. Justification

Provide a succinct justification for the sequencing or genotyping study by describing the significance of the problem and providing other relevant background information.

This section is a key evaluation criterion.

1. *State the relevance to infectious disease for the organism(s) to be studied; for example the public health significance, model system etc.*

2. *Are there genome data for organisms in the same phylum / class / family / genus? What is the status of other sequencing / genotyping projects on the same organism including current and past projects of the NIAID GSC? Provide information on other characteristics (genome size, GC content, repetitive DNA, pre-existing arrays etc.) relevant to the proposed study. Have analyses been performed on the raw data already generated/published? If additional strains are proposed for a species, please provide a justification for additional strains?*

3. *If analyses have been conducted, briefly describe utility of the new sequencing or genotyping information with an explanation of how the proposed study to generate additional data will advance diagnostics, therapeutics, epidemiology, vaccines, or*

basic knowledge such as species diversity, evolution, virulence, etc. of the proposed organism to be studied.

Enterobacteriaceae are among the most common human pathogens worldwide, responsible for a substantial number of infections and deaths every year. A major source of community- and hospital-acquired infections, these organisms cause pneumonia, septicemia, meningitis, peritonitis and device-associated infections. Because of their propensity for horizontal gene transfer, transmission of plasmids and insertion elements, often containing antibiotic resistance genes, occurs readily among this family of bacteria. A major consequence of this ease of transfer is the rapidly evolving resistance of these organisms to every class of antibiotic agent in use, posing a formidable challenge to the healthcare system.

Multiple mechanisms exist for the generation of bacterial resistance. A strain may become resistant to an antimicrobial agent when the target of that agent mutates, conferring a reduced affinity or insensitivity to the molecule. Other resistance mechanisms include the acquisition or overexpression of efflux pumps that extrude the antibacterial agent from the cell, development of mutations that limit access of antimicrobial agents to the intracellular target site via downregulation of porin genes, and acquisition of drug inactivation genes.

The most common mechanism for *Klebsiella pneumoniae* and other *Enterobacteriaceae* to acquire resistance to all β-lactam-containing antibiotics, including cephalosporins, penicillins, and monobactams, is through the production of extended spectrum β-lactamases, which hydrolyze the antibiotics (16). The genes encoding ESBLs are generally found on readily transferrable plasmids that additionally carry resistance elements to multiple other antibiotic classes (17). ESBLs now number over 500 distinct enzymes and convey varying degrees of resistance to antimicrobial agents (16). The apparently small fitness cost of their carriage, their rapid transfer among strains and species (14), their linkage to other resistance genes, and ongoing selection with antibiotic exposure, stresses the importance of preventing or delaying their establishment by preventing the intrusion of such resistant clones.

In addition to antibiotic resistance, virulence-associated genes are also not normally tracked by any surveillance system, yet are critical to understanding the rapid spread of resistant bacteria. For example, the epidemic spread of a hyper-virulent *K. pneumoniae* strain is associated with the acquisition and overexpression of novel siderophores (18). Current PFGE or gene specific (PCR) surveillance methods would be unable to detect novel siderophores or novel resistance determinants, whereas high resolution sequencing methods could identify the key differences, as well as provide the basis for future gene expression studies.

A significant epidemiological challenge is our current difficulty in rapidly identifying the spread of novel clones. Currently, tracking is typically done at the species level, with the possible addition of some antibiotic susceptibility data. These data provide insufficient signal to detect an emerging clone, since at the start of an outbreak a new clone, by definition, will be rare, and thus will be viewed as statistical noise in the ongoing, background infection rate. Genomic data, which provide the highest level of resolution possible, would allow the recognition of the initial stages of a clonal outbreak which

otherwise could not be identified by tracking species prevalence alone. Identification of a rapidly spreading clone within and among facilities would trigger a rapid response of containment measures and improved infection control, thus stemming the clonal outbreak.

Genome analysis is the ultimate reference for interpreting evolution, dissemination and recombination among sensitive and resistant clones and subclones. As these technologies interface with the world's microbiology laboratories, the increasing ability to integrate genotypic data with current laboratory descriptions of isolates will ultimately provide a greater definition of the interrelationships among strains and an increased ability to discriminate among the different strains.

A limited number of *Klebsiella pneumoniae* and *Enterobacter cloacae* isolates have been sequenced to date. In the case of *Klebsiella pneumoniae* there are six publically available sequenced strains:
MGH 78578
NTUH-K2044 (20)
ATCC 13884
342 (not a human clinical pathogen)
JH1 (19)
112281 (19).

For *E. cloacae*: ATCC 13047 (The *E. cloacae* ATCC 13047 chromosome and plasmid pECL_A and pECL_B sequences have been deposited in GenBank under accession numbers CP001918, CP001919, and CP001920).

Sequence data on plasmids containing carbapenemase-encoding genes are even more limited.

3. Rationale for Strain Selection

4. *Provide the rationale behind the selection of strains and the number of strains proposed in the study. The focus of the program is on potential agents of bioterrorism or organisms responsible for emerging or re-emerging infectious diseases. Non-select agents or non-pathogenic organisms will be considered when they can provide insight into these scientific areas.*

Retrospectively collected isolates:
19 strains from Massachusetts General Hospital collected by Dr. Mary Jane Ferraro (dates of isolation range from 2007 to 2009 but most were collected in 2009). 13 of 19 had KPC by PCR and 1/19 possessed NDM. All but one were *E. cloacae*.

31 strains from Beth Israel Deaconess Medical Center (James Kirby) of carbapenem-resistant *Klebsiella pneumoniae* (15), *E. coli* (14), and *Enterobacter cloacae* (2) (collected from 2009-2011).

Prospectively collected isolates: Approximately 10 isolates of carbapenem resistant enterobacteriaceae per month expected for 6 to 9 months (60-90 strains) from Massachusetts General Hospital, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center. We will also collect a random sample of susceptible enterobacterial strains from each hospital in parallel.

4a. Approach to Data Production: Data Generation

5. *State the data and resources planned to be generated. (e.g draft genome sequences, finished sequence data, SNPs, DNA/protein arrays generation, clone generation etc.)*

Since we wish to identify the gain of genetic elements that are associated with antibiotic resistance and host adaptation, we will generate draft *de novo* genome sequences for all proposed strains using the Illumina HiSeq platform. Many of these genetic elements contain multiple repetitive regions, such as transposable elements, making assembly and localization of these genes difficult. To overcome this problem, we will use large jumping library insert sizes (~3-7 kb), which will enable us to scaffold over the repeated elements. If a novel plasmid or strain of clinical importance is discovered, we might undertake additional improvement efforts if needed using additional sequencing technology platforms or manually directed finishing.

4b. Approach to Data Production: Data Analysis

6. *Briefly describe the analysis (value-add) envisioned to be performed subsequently by the community and the potential to develop hypotheses driven proposals given the datasets and resources produced by this work.*

With the data generated in this project, we will address the following topics:

1) **Mechanisms of resistance.** While many carbapenem-resistant strains have an identified resistance gene, some lack either KPC genes or NDM-1, as is observed in six isolates from the historical collection of carbapenem-resistant organisms. We will identify novel β-lactamase genes in the sequenced organisms. We will also determine if plasmid structure

or resistance plasmid gene content vary.

2) **The relative effects of clonal spread versus gene transfer.** Carbapenem resistance may be mediated by a number of molecular mechanisms, from transfer of carbapenemase genes to clonal spread. It is not known to what extent the resistance problem is the result of an epidemic strain or strains or whether it is due to promiscuous plasmids. To address this question, we will assess the following:

- Does a single hospital have the same carbapenem-resistant strain (or strains) through time? Is this resistance due to a single resistance mechanism or multiple mechanisms?
- Do different hospitals share the same strains or resistance mechanisms?
- Do different wards or specimen types have the same strains or plasmids?

3) **The epidemiology of carbapenem resistance.** Here, we are interested in how resistant strains, resistance plasmids, and resistance genes have spread through space and time. This approach will focus on fine-scale microvariation, such as SNPs. For instance, we could determine how *K. pneumoniae* ST258 has spread through and evolved within our partner hospitals.

4) Alterations in susceptibility to carbapenems in infecting strains during treatment with carbapenems.
We have observed infecting strains that continued to be cultured from patients while they were undergoing therapy with carbapenems and that later required moderate elevations in the concentration of ertapenem or imipenem, or both, for inhibition in vitro. We will obtain isolates of future examples of these strains, taken from patients at initial stages of antibiotic therapy and later timepoints, to determine whether there are changes in their genomes that might explain this potentially serious problem.

5. Community Support and Collaborator Roles:

7. *Provide evidence of the relevant scientific community's size and depth of interest in the proposed sequencing or genotyping data for this organism or group of organisms. Please provide specific examples.*

8. *List all project collaborators and their roles in the project*

9. *List availability of other funding sources for the project.*

Collaborators on the current project include:
David Hooper (Associate Chief, Division of Infectious Diseases and Chief, Infection Control Unit, Massachusetts General Hospital) will provide strains for this study and relevant epidemiologic data.

Mary Jane Ferraro (Director, Microbiology Laboratories, Massachusetts General Hospital and Professor of Pathology and Medicine, Harvard Medical School) will provide strains for this study.

Thomas O'Brien (Brigham & Women's Hospital). Dr. O'Brien has had extensive experience in tracking the spread of antibiotic resistance, including publishing one of the

seminal studies demonstrating epidemic antibiotic resistance plasmid spread.

James Kirby (Director, Clinical Microbiology, Beth Israel Deaconess Medical Center and Harvard Medical School). Dr. Kirby will provide strains for this study from his longitudinal collection.

Andrew Onderdonk. (Professor of Pathology, Harvard Medical School, Director of Clinical Microbiology, Brigham and Women's Hospital). Dr. Onderdonk has provided large amounts of high quality whole genomic DNA for several projects to investigators at the Broad for their use. His research laboratory also inventories a broad array of strains using a computerized database and can provide both phenotypic characterization and PFGE analysis as preliminary information for high throughput sequencing. The Onderdonk lab will supply high quality total DNA from each of the isolated strains.

Additional members of the team may be added during the course of the work with the general agreement of the current membership.

6. Availability & Information of Strains:

10. *Indicate availability of relevant laboratory strains and clinical isolates. Are the strains/isolates of interest retrospectively collected, prepared and ready to ship?*

Note: If samples are prospectively prepared the GSC can provide protocols and recommendation based on the Centers past experiences. The samples must however meet minimum quality standards as established by the Center for the optimal technology platform (sequencing/ genotyping) to be used in the study.

Approximately 50 of the strains have already been collected at MGH and BIDMC. Dr. Onderdonk at BWH has developed DNA extraction procedures, and has experience shipping DNA to the Broad Institute.

11. *Attach relevant information, if available in an excel spreadsheet for multiple samples: e.g.*
    - *Name*
    - *Identifier*
    - *Material type (DNA/RNA/Strain)*
    - *Genus*
    - *Species*
    - *Specimen / Strain*
    - *Isolation source*
    - *Isolated from*
    - *Select agent status*
    - *International permit requirement*

- *BEIR/ATCC repository accession number*
- *Other public repository location*
- *Other public repository identifier*
- *Sample provider's name*
- *Sample provider's contact*

12. *What supporting metadata and clinical data have been collected or are planned on being collected that could be made available for community use?*

Hospital and Ward
Date of isolation
Specimen type
Antibiotic susceptibility testing

7. Compliance Requirements:

7a. Review NIAID's Reagent, Data & Software Release Policy:

NIAID supports rapid data and reagent release to the scientific community for all sequencing and genotyping projects funded by NIAID GSC. It is expected that projects will adhere to the data and reagent release policy described in the following web sites.

http://www3.niaid.nih.gov/research/resources/mscs/data.htm

http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-013.html

<Each Center to include their website that describes/points to the guidelines>

Once a white paper project is approved, NIAID GSC will develop with the collaborators a detailed data and reagent release plan to be reviewed and approved by NIAID.

Accept ☒ Decline ☐

7b. Public Access to Reagents, Data, Software and Other Materials:

13. *State plans for deposit of starting materials as well as resulting reagents, resources, and datasets in NIAID approved repositories. Sequencing projects will not begin until the strain is deposited into NIAID funded BEI repository (http://www.beiresources.org/). This includes web based forms are completed by the collaborator and received by the NIAID BEI (http://www.beiresources.org/).*

Collaborators have been notified that sequencing will not commence until strain deposition is underway at the NIAID funded BEI repository. All sequences and read files generated under this proposal will be submitted to the Short Read Archive at NCBI/NLM/NIH on a weekly basis. These data will also include information on templates, vectors, and quality values for each sequence.

Genome assemblies will be made available via GenBank and the Broad web site. Assembled contigs and scaffolds will be deposited in the Whole Genome Shotgun (WGS) section of GenBank, http://www.ncbi.nlm.nih.gov/Genbank/wgs.html, within 45 calendar days of completing assemblies. If it is determined that the final assembly can be significantly improved, an updated record will be deposited in the appropriate part of GenBank when complete.

Annotation data will be made available via GenBank and the Broad web site after consistency checks and quality control have been completed by the GSCID and collaborators. Assuming no significant errors are detected during the validation process, annotation data will be released within 45 calendar days of being generated.

Candidate polymorphisms identified by comparison of new genome sequences to a reference will be deposited in dbSNP at NCBI and released through the Broad GSCID's web site within 45 days of data being generated. Prior to public release, polymorphisms will be made available to collaborating scientists for a one-week period for quality control purposes. Candidate polymorphisms will then be released barring any quality issues discovered in the QC process.

7c. Research Compliance Requirements

Upon project approval, NIAID review of relevant IRB/IACUC documentation is required prior to commencement of work. Please contact the GSC Principal Investigator(s) to ensure necessary documentation are filed for / made available for timely start of the project.

Investigator Signature:

Investigator Name: Cheryl Murphy, Lisa Cosimi, Michael Feldgarden
Date: February 28, 2012

References

1. Gupta N, Limbago BM, Patel JB, and Kallen AJ. 2011. Carbapenem-Resistant Enterobacteriaceae: Epidemiology and Prevention. Clin Infect Dis. 53: 60-67.
2. Hidron AI, Edwards JR, Patel J, Horan TC, Sievert DM, Pollock DA, Fridkin SK; National Healthcare Safety Network Team; Participating National Healthcare Safety Network Facilities. 2008. NHSN annual update: antimicrobial-resistant pathogens associated with health care-associated infections: annual summary of data reported to the National Healthcare Safety Network at the centers for disease control and prevention 2006-2007. Infect Cont Hosp Epidemiol 29:996-1011.
3. Schwaber MJ, Klarfeld-Lidji S, Navon-Venezia S, Schwartz D, Leavitt A, Carmeli Y. 2008. Predictors of carbapenem-resistant *Klebsiella pneumoniae* acquisition among hospitalized adults and effect of acquisition on mortality. Antimicrob Agents Chemother 52:1028–33.
4. Queenan AM, Bush K. 2007. Carbapenemases: the versatile b-lactamases. Clin Microbiol Rev 20:440-58.
5. Nordman P, Poirel L. 2002. Emerging carbapenemases in gram-negative aerobes. Clin Microbiol Infect 8:321-31.
6. Nordman P, Cuzon G, and Naas T. 2009. The real threat of KPC carbapenemase-producing bacteria. Lancet Infect Dis 9:228-36.
7. Kitchel B, Rasheed JK, Patel JB, Srinivasan A, Navon-Venezia S, Carmeli Y, Brolund A, and Giske CG. 2009. Molecular epidemiology of KPC-producing Klebsiella pneumoniae isolates in the United States: clonal expansion of multilocus sequence type 258. Antimicrob Agents Chemother. 53:3365-70.
8. Andrade LN, Curiao T, Ferreira JC, Longo JM, Clímaco EC, Martinez R, Bellissimo-Rodrigues F, Basile-Filho A, Evaristo MA, Del Peloso PF, Ribeiro VB, Barth AL, Paula MC, Baquero F, Cantón R, Darini AL, Coque TM. 2011. Dissemination of blaKPC-2 by the spread of Klebsiella pneumoniae clonal complex 258 clones (ST258, ST11, ST437) and plasmids (IncFII, IncN, IncL/M) among Enterobacteriaceae species in Brazil. Antimicrob Agents Chemother 55:3579-83.
9. Leavitt A, Navon-Venezia S, Chmelnitsky I, Schwaber MJ, and Carmeli Y. 2007. Emergence of KPC-2 and KPC-3 in carbapenem-resistant *Klebsiella pneumoniae* strains in an Israeli hospital. Antimicrob Agents Chemother 51:3026-9.
10. Pournaras S, Protonotariou E, Boulgari E, Kristo I, Dimitroulia E, Vitti D, Tsalidou M, Maniatis AN, Tsakris A, and Sofianou D. 2009. Clonal spread of KPC-2 carbapenemase-producing *Klebsiella pneumoniae* strains in Greece. J Antimicrob Chemother 64:348-52.
11. Yong D, Toleman MA, Giske CG, Cho HS, Sundman K, Lee K, and Walsh TR. 2009. Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India. Antimicrob Agents Chemother 53:5046-54.
12. Naas T, Cuzon G, Villegas M-V, Lartigue M-F, Quinn JP, and Nordmann P. 2008. Genetic structures at the origin of acquisition of the β-lactamase bla_{KPC} gene. Antimicrob Agents Chemother 52:1257-63.
13. Cuzon G, Naas T, Truong H, Villegas M-V, Wisell KT, Carmeli Y, Gales AC, Navon-Venezia S, Quinn JP, and Nordmann P. 2010. Worldwide diversity of *Klebsiella pneumoniae* that produces β-lactamase bla_{KPC} gene. Emerg Infect Dis 16:1349-56.
14. Mathers AJ, Cox HL, Kitchel B, Bonatti H, Brassinga AKC, Carroll J, Scheld WM, Hazen KC and Sifri CD. 2011. Molecular dissection of an outbreak of carbapenem-resistant enterobacteriaceae reveals intergenus KPC carbapenemase transmission through a promiscuous plasmid. mBio 2:1-7.
15. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, Fitzgerald M, Godfrey P, Haas BJ, Murphy CI, Russ C, Sykes S, Walker BJ, Wortman JR, Young S, Zeng Q, Abouelleil A, Bochicchio

J, Chauvin S, Desmet T, Gujja S, McCowan C, Montmayeur A, Steelman S, Frimodt-Møller J, Petersen AM, Struve C, Krogfelt KA, Bingen E, Weill FX, Lander ES, Nusbaum C, Birren BW, Hung DT, Hanage WP. 2012. Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci 109:3065-70.
16. Jacoby GA, and Munoz-Price LS. 2005. The new β-lactamases. N Engl J Med. 352:380–91.
17. Sirot J, Chanal C, Petit A, Sirot D, Labia R, and Gerbaud G. 1988. *Klebsiella pneumoniae* and other Enterobacteriaceae producing novel plasmid-mediated β-lactamases markedly active against third generation cephalosporins: epidemiological studies. Rev Infect Dis 10:850-9.
18. Russo TA, Shon AS, Beanan JM, Olson R, MacDonald U, Pomakov AO, and Visitacion MP. 2011. Hypervirulent K. pneumoniae secretes more and more active iron-acquisition molecules than "classical" K. pneumoniae thereby enhancing its virulence. PLoS One 6(10):e26734.
19. Kumar V, Sun P, Vamathevan J, Li Y, Ingraham K, Palmer L, Huang J, and Brown JR. 2011. Comparative genomics of Klebsiella pneumoniae strains with different antibiotic resistance profiles. Antimicrob Agents Chemother 55:4267-76.
20. Wu KM, Li LH, Yan JJ, Tsao N, Liao TL, Tsai HC, Fung CP, Chen HJ, Liu YM, Wang JT, Fang CT, Chang SC, Shu HY, Liu TT, Chen YT, Shiau YR, Lauderdale TL, Su IJ, Kirby R, Tsai SF. 2009. Genome sequencing and comparative analysis of *Klebsiella pneumoniae* NTUH-K2044, a strain causing liver abscess and meningitis. J Bacteriol 191:4492-501.