White Paper Application

Project Title: KPC and *Klebsiella pneumoniae*: virulence and resistance converge

Authors: Robert A. Bonomo and Mark D. Adams

Primary Investigator Contact:

Name	Robert A. Bonomo
Position	Professor
Institution	Case Western Reserve University and Cleveland VAMC
Address	10701 East Blvd., Cleveland
State	OH
ZIP Code	44106
Telephone	216-791-3800x4399, x4801
Fax	
E-Mail	Robert.bonomo@va.gov; rab14@case.edu

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

Klebsiella pneumonia is a contemporary emerging pathogen as more and more multidrug resistant (MDR) strains are being isolated in the clinic. The spectrum of illness associated with *K. pneumoniae* is vast: pneumonia, urinary tract infection, wound, liver abscess, and bloodstream infections. Since the early 2000s, isolates of *K. pneumoniae* resistant to carbapenems and carrying the KPC beta-lactamase (bla_{KPC}) have become much more common. In addition, KPC-carrying *K. pneumoniae* strains tend to be resistant to other antimicrobial agents as well, severely limiting therapeutic options. Colistin is the treatment of "last resort" for carbapenem-resistant *K. pneumoniae*. However, resistance to colistin (Col^R) has been observed in the clinic, signifying that these infections are essentially untreatable.

We propose to use genome sequencing and gene expression analysis to evaluate three aspects of *K. pneumoniae* epidemiology and pathogenesis.

Year 4 Activities

1) Colonizing vs. Invasive isolates. Identify genetic and clinical markers of virulence and transmissibility by comparison of KPC-positive colonizing *K. pneumoniae* isolates to those that have caused an infection (pneumonia, bloodstream infection, etc.) by genome sequencing and gene expression analysis (20 strains each from Cleveland and Detroit).

2) Col^R Matched colistin susceptible and Col^R isolates will be identified from patients that undergo colistin treatment based on *in vitro* susceptibility testing. Genome sequence will be used to find mutations and gene expression analysis to infer altered regulatory programs as have been observed in other Col^R Gram-negative bacteria. (5 pairs of isolates)

The resulting information will be a resource for the *Klebsiella* research community by greatly increasing the number whole genome sequences, by producing gene expression data that will serve as a baseline for future research and an initial snapshot of genome activity, and by illustrating the extent of change in clinical isolates over a two decade time span. This data will also be the basis for development of novel molecular diagnostic assays to test for colistin resistance and other markers that predict likelihood that a colonizing isolate may cause infection.

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.*

K. pneumoniae is an emerging infectious disease, primarily due to the dramatic increase in antimicrobial drug resistance over the past decade [1-3]. The KPC β-lactamase gene confers resistance to all penicillins, cephalosporins, and carbapenems, rending this class of drugs essentially unusable for *Klebsiella* infections.

K. pneumoniae colonization is relatively common, but most colonized patients do not develop an infection. To determine whether there are microbe-specific markers of virulence that distinguish colonizers from pathogens, we will sequence and examine gene expression profiles of contemporaneous isolates from two hospitals.

A critical problem in treatment of *K. pneumoniae* infections is resistance to the "last-line" antibiotic colistin [4]. We will explore the genetic basis of Col^R in *K. pneumoniae*. Because Col^R has been shown to be associated with altered regulation of genes that contribute to the structure and charge of the outer membrane, we will explore stable differences in gene expression between genetically paired Col^R and Col^S isolates.

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?*

Three complete genome sequences of *K. pneumoniae* subsp. *pneumoniae* isolates are available in GenBank (MGH 78578, HS11286, and NTUH-K2044). There are no current or past NIAID GSC projects on *K. pneumoniae*.

The Wellcome Trust Sanger Institute is sequencing 321 geographically diverse *K. pneumoniae* isolates (http://www.sanger.ac.uk/resources/downloads/bacteria/klebsiella-pneumoniae.html). Primary read data is currently available, but not assemblies. Strain information and metadata is not available for these isolates. These strains were selected to represent global diversity rather than to answer specific questions regarding transmission, virulence, or antimicrobial resistance. The data from this broad survey will certainly be valuable in the context of this project by providing a much more complete view of the pan-genome of *K. pneumoniae*. The NIH

Intramural Sequencing Center (NISC) is also sequencing several carbapenem-resistant *K. pneumoniae* isolates. The justification for additional sequencing of these strains is provided in the next section.

K. pneumoniae genomes range between 5 and 6 Mbp and have a high G+C content average 57%. A number of insertion sequences and transposable elements have been described, justifying a paired-end sequencing strategy to obtain the longest range scaffolding during genome assembly.

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.*

See response to #4 below.

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.*

We intend to pursue three inter-related projects:

1) Colonizers vs. Invasive isolates. Rationale: Identify genetic and clinical markers of virulence and transmissibility. Approach: Compare KPC-positive colonizing *K. pneumoniae* isolates to those that are invasive by genome sequencing and gene expression analysis. Sample set: 20 strains from a surveillance study and 20 strains from the same hospital and patient population derived from clinically documented bloodstream infections or pneumonia.

2) Col^R Rationale: Matched colistin susceptible and resistant isolates will be identified from patients that undergo colistin treatment and develop resistance based on *in vitro* susceptibility testing. Approach: Genome sequence will be used to find mutations and gene expression analysis to infer altered regulatory programs as have been observed in other Col^R Gram-negative bacteria. Sample set: 5 Col^S and 5 Col^R isolates from Detroit and Cleveland.

Comparison of colonizers and invasive isolates will, for the first time, shed light on whether strain differences are associated with virulence, or whether the primary factors associated with infection are related to host defenses. We will also explore the genetic basis for Col^R in *K. pneumoniae* by analysis of likely isogenic strains from the same patient before and after colistin therapy. The inclusion of gene expression (transcriptome sequencing) data will further improve genome annotation and represent additional depth of information on gene activity for interpreting

differences in genome content.

These genome sequences and their interpretation will most readily lead to new molecular diagnostic tests to facilitate tracking of outbreaks, identification of antibiotic resistance genes, and virulence determinants. In particular, better understanding of the basis of colistin resistance will assist in monitoring development of resistance in patients. Genes or variants specific to invasion would greatly assist in infection control and surveillance.

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.)*

Draft genome sequences will be generated for each strain. Sequence data will be collected on the Illumina HiSeq platform with paired-end 100 base reads to maximize assembly contiguity. Selected finishing activities will be conducted to clarify differences among isolates in plasmid structure, repeat element junctions, and at positions of gaps in genes that are critical for comparative analysis. Reference-genome standard finishing activities should not be necessary for any of the three projects proposed.

For transcriptome analysis, strand-specific sequencing libraries will be prepared for sequencing on the Illumina HiSeq. For the colistin-resistance project, small RNA libraries will also be prepared, based data from *E. coli* showing small-RNA-mediated regulation of genes involved in colistin resistance [5].

Genome annotations will be performed on each strain independently and then cross-referenced to identify inconsistent annotations across closely related draft genomes for correction. Transcriptome data will be used to annotated operon structure and transcriptional boundaries.

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.*

The information generated through these projects will be a resource for the *Klebsiella* research community by greatly increasing the number of strains with whole genome sequence data, by producing gene expression data that will serve as both a baseline for future research and an initial snapshot of genome activity, and by illustrating the extent of change in clinical isolates over a two decade time span. The data obtained will also be a resource for the clinical microbiology community, providing the basis for development of molecular diagnostic assays to test for colistin resistance and possibly for other markers that predict likelihood that a colonizing isolate may cause infection.

The project will generate resources that will be of general interest the *K. pneumoniae* research community. Project 1 involves identification of candidate virulence genes based on differences in genome content among closely related strains that are colonizers versus those that cause clinical infections. There are likely to be numerous potential virulence factors, each of which will require additional follow-up study. The genome sequencing projects will help to focus attention on these strain differences, but their roles will need to be determined experimentally *in vitro* and in animal models.

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.*

Over 14,000 papers in PubMed address myriad aspects of *K. pneumoniae* biology. Among these are more than 2,500 are related to drug resistance and over 300 related to the KPC carbapenemase. At present, strains possessing bla_{KPC} are threatening the lives of patients though out the world. Despite detailed knowledge of risk factors and the introduction of antibiotic stewardship programs, infection control measures have been slow to arrest the spread of bla_{KPC}.

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

a) **Robert A. Bonomo**, MD, Chief of Medicine, Louis Stokes Cleveland Veterans Affairs Medical Center and Professor of Microbiology and Molecular Biology and of Pharmacology, Case Western Reserve University, Cleveland, OH. Dr. Bonomo will supply strains and associated meta-data for each project and will participate in analysis of genome sequence data and all aspects of interpretation of genomic differences.

b) **David van Duin**, MD, PhD, Department of Infectious Disease, Cleveland Clinic, Cleveland, OH. Dr. van Duin will provide strains and participate in interpretation of clinical data and strain displacement data.

c) **Federico Perez**, MD, Louis Stokes Cleveland Veterans Affairs Medical Center and Assistant Professor of Medicine, Case Western Reserve University, Cleveland, OH. Dr. Perez will assist with epidemiological analysis and strain displacement data.

d) **Michael R. Jacobs** MD, PhD, Professor of Pathology, Case Western Reserve University, Cleveland, OH. Dr. Jacobs will assist with microbiological, epidemiological, and strain displacement data

e) **Dror Marcham**, MD, Researcher, Division of Infectious Diseases, Wayne State University and Detroit Medical Center, Detroit, MI. Dr. Marcham will provide strains and will participate in interpretation of genomic differences associated with colonization vs. infection and colistin resistance.

e) **Keith Kaye**, MD, Professor of Internal Medicine and Infectious Diseases, Wayne State University and Corporate Medical Director, Hospital Epidemiology and Antimicrobial Stewardship, Detroit Medical Center, Detroit, MI. Dr. Kaye will provide strains and will participate in interpretation of genomic differences

associated with colonization vs. infection and colistin resistance.
f) **Mark D. Adams,** PhD, Scientific Director, J. Craig Venter Institute. Dr. Adams will coordinate JCVI activities as project PI and will participate in all aspects of data analysis and interpretation.

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

No other funding is currently available or pending for genome sequence analysis from the participating investigators. Each investigator is prepared to seek funding based on genomics results in each of the targeted areas of study.

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.

All strains are currently available as glycerol stocks in the laboratories of the collaborating investigators (Bonomo, van Duin, Jacobs, Perez, Marcham, and Kaye).

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

See attached

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/LabsAndResources/resources/mscs/data.htm
http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-013.html

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:

12. *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/).*

All strains will be submitted to BEI along with associated metadata related to date and location of isolation and antimicrobial susceptibility profiles.

Raw sequence data will be submitted to the short read archive (SRA) at NCBI. Preliminary sequence assemblies will be submitted as they become available and annotated genomes will be submitted to GenBank as they complete quality review. Transcriptome data will be used to improve the annotation of the genome sequences and will be submitted to GEO to support additional studies of quantitative analysis of gene expression. Data will also be submitted to PATRIC.

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:

Date May 31, 2012

1. Endimiani, A., et al., *Characterization of blaKPC-containing Klebsiella pneumoniae isolates detected in different institutions in the Eastern USA.* J Antimicrob Chemother, 2009. **63**(3): p. 427-37.
2. Gomez-Pinilla, F. and Z. Ying, *Differential effects of exercise and dietary docosahexaenoic acid on molecular systems associated with control of allostasis in the hypothalamus and hippocampus.* Neuroscience, 2010. **168**(1): p. 130-7.
3. Adler, A. and Y. Carmeli, *Dissemination of the Klebsiella pneumoniae carbapenemase in the health care settings: tracking the trails of an elusive offender.* mBio, 2011. **2**(6).
4. Bogdanovich, T., et al., *Colistin-resistant, Klebsiella pneumoniae carbapenemase (KPC)-producing Klebsiella pneumoniae belonging to the international epidemic clone ST258.* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2011. **53**(4): p. 373-6.
5. Moon, K. and S. Gottesman, *A PhoQ/P-regulated small RNA regulates sensitivity of Escherichia coli to antimicrobial peptides.* Molecular microbiology, 2009. **74**(6): p. 1314-30.