White Paper Application

Project Title: Genome sequencing and analysis of additional strains in the *Burkholderia cepacia* complex

Authors:
Primary Investigator Contact:
Name: Steven M. Holland, M.D.
Position: Chief, Laboratory of Clinical Infectious Diseases (LCID)
Institution: NIAID, NIH
Address: Bldg. 10; CRC B3-4141, MSC 1684
City, State: Bethesda, MD
ZIP Code: 20892-1684
Telephone: 301-402-7684
Fax: 301-480-4508
E-mail: smh@nih.gov

1. Executive Summary

The *Burkholderia cepacia* complex (Bcc) is a group of opportunistic pathogens that cause infections in immunocompromised patients. In particular, Bcc causes lung infections in patients with cystic fibrosis (CF) and chronic granulomatous disease (CGD). The most common Bcc species that cause infections in CF and CGD patients are *Burkholderia cenocepacia* (genomovar III) and *Burkholderia multivorans* (genomovar II). Interestingly, the clinical presentation of Bcc infections in CF and CGD are distinctly different (Greenberg et al. 2009. Clin Infect Dis. 48:1577-9).

As a part of the original white paper, JCVI determined the genome sequence of three clinical isolates of *B. multivorans* (released January 22, 2009). These isolates were all from CGD patients. Thus far no genome sequence analysis has been performed on *B. multivorans* strains isolated from CF patients. The goal of the first part of this addendum will be to sequence two additional *B. multivorans* strains: CF1 (ATCC strain BAA-247 from a CF patient) and CF2 (isolated from a CF patient from the NIH). These strains have been compared to the CGD strains as well as an environmental isolate for virulence attributes in a mouse model of infection as well as binding and adherence to cells from CGD and normal individuals (Zelazny et al. 2009. Infect Immun. 77:4337-44).

The genome sequence of *B. cenocepacia* J2315, has recently been published (Holden et al. 2009. J Bacteriol. 191:261-77). This strain is a member of the ET-12 lineage, an epidemic group responsible for numerous deadly outbreaks among CF patients. However due to its inherent antibiotic resistance, J2315 is not often used for genetic manipulations. Other members of the ET-12 lineage, including strains K56-2 and BC7, are more versatile in this regard. In addition, these latter strains have been used in numerous animal infection models and have distinct virulence factor production and pathogenesis compared to J2315. Therefore the second part of this addendum will involve sequencing these additional *B. cenocepacia* strains.

2. Justification

The genus *Burkholderia* contains >30 species, the most pathogenic of which are *B. pseudomallei* and *B. mallei* (both category B select agents), and the members of the Bcc. The Bcc currently includes 17 species (formerly classified as genomovars). Bcc bacteria are opportunistic pathogens in immunocompromised patients, such as those with CF or CGD.

CF is the most common autosomal recessive disease among Caucasians. The defect in the CFTR gene is responsible for the pathology of the disease, however the major cause of death in these patients is due to chronic lung infections with particular pathogens. While *Pseudomonas aeruginosa* is the major cause of morbidity and mortality, members of the Bcc have become more prominent and are of particular interest due to their epidemic nature and their ability to cause "cepacia syndrome", a nectrotizing pneumonia leading to septicemia and often death. Bcc strains in CF are also difficult to treat due to their inherent and acquired antibiotic resistances.

CGD is a primary immunodeficiency that affects the phagocyte NADPH oxidase. The impaired phagocyte killing can lead to recurrent, life-threatening infections by a very specific set of microorganisms, including Bcc. The most common species that cause infections in CF and CGD patients are *B. cenocepacia* (genomovar III) and *B. multivorans* (genomovar II). *B. multivorans* comprise the majority of the Bcc isolates from CGD patients at the NIH (Greenberg et al. 2009. Clin Infect Dis. 48:1577-9).

Whole genome sequencing of *Burkholderia* spp has dramatically increased in the recent years with a large number of strains of *B. pseudomallei* and *B. mallei* sequenced. In comparison, only 4 *B. cenocepacia* have complete genomes (with one more "in progress"). Among *B. multivorans*, 3 strains from CGD patients have been sequenced by JCVI as part of this original proposal, and the only strain that has a complete genome is ATCC 17616, which is not a clinical isolate but an environmental strain. Augmenting the available *Burkhlderia* genome sequences with Bcc from various types of infections through this project will provide information on the core and pan genome of this species and may provide insight in the genes relevant in CF vs. CGD patient populations.

3. Rationale for Strain Selection

In the Holland laboratory at NIH, we have been studying the virulence of Bcc in CGD. With respect to *B. multivorans,* we have compared strains isolated from CF and CGD patients with the environmental strain ATCC 17616. The results of this analysis have been recently published (Zelazny et al. 2009. Infect Immun. 77:4337-44). However it remains unclear whether there are differences in the genes from strains isolated from the environment and from infections in CF and CGD. This project reflects a continuation of the study previously undertaken by the JCVI; the sequence of additional *B. multivorans* strains from CF patients will allow the comparisons between the previously sequenced CGD isolates as well as the published sequence of the environmental isolate ATCC

17616. Thus, we would like to sequence two *B. multivorans* strains; CF1 was obtained from the ATCC and CF2 from the CF clinic at the NIH. The advantage of these strains is that we already have some performed phenotypic and pathogenesis studies on these isolates.

With respect to *B. cenocepacia*, we have found that there is a difference in virulence between the sequenced strain J2315 and the other strains of the ET-12 lineage, K56-2 and BC7 in a mouse model of CGD. We have also noted differences in assays testing the survival of these strains in the presence of human neutrophils. Since these strains have been used by numerous investigators, a comparison between these closely related but phenotypically distinct strains would be valuable to the field to define the genes responsible for these differences in pathogenesis.

Altogether this is an exciting opportunity for the NIH to provide the scientific community with data on select well-characterized *B. multivorans* and *B. cenocepacia* strains. This resource will provide the basis for a more thorough comparison of the genes within these *Burkholderia* species as well as with other members of the genus *Burkholderia*.

4. Approach to Data Production: Data Generation and Data Analysis

We will work with Dr. William Nierman from JCVI, who was the sequencing project leader in the original proposal, to define the approach to be taken for the generation and analysis of the genome sequence data.

5. Community Support and Collaborator Roles:

Steven M. Holland, M.D. (Principal Investigator): Chief, LCID, NIH.

Joanna B. Goldberg, Ph.D. (Co-Investigator): Professor, Department of Microbiology, University of Virginia.

Adrian Zelazny, Ph.D. (Co-Investigator): Clinical Center, NIH.

David E. Greenberg, M.D. (Co-Investigator): Assistant Clinical Investigator, NIH.

There is currently no funding to support this project.

6. Availability & Information of Strains:

The strains include *B. multivorans* CF1 and CF2 and *B. cenocepacia* K56-2 and BC7 and are available in both the Holland and Goldberg laboratories.

7. Compliance Requirements:

The Principal Investigator and Co-Investigators agree to all the stated requirements.