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:

Authors:
Primary Investigator Contact:

<table>
<thead>
<tr>
<th>Name</th>
<th>Michael S. Gilmore, Ph.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Sir William Osler Professor of Ophthalmology (Microbiology and Molecular Genetics)</td>
</tr>
<tr>
<td>Institution</td>
<td>Harvard Medical School/Mass. Eye and Ear Infirmary</td>
</tr>
<tr>
<td>Address</td>
<td>243 Charles St.</td>
</tr>
<tr>
<td>State</td>
<td>MA</td>
</tr>
<tr>
<td>ZIP Code</td>
<td>02114</td>
</tr>
<tr>
<td>Telephone</td>
<td>617-573-3845</td>
</tr>
<tr>
<td>Fax</td>
<td></td>
</tr>
<tr>
<td>E-Mail</td>
<td><a href="mailto:michael_gilmore@meei.harvard.edu">michael_gilmore@meei.harvard.edu</a></td>
</tr>
</tbody>
</table>

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

Provide an executive summary of the proposal.

This is an application to support a collaboration with the Broad Institute’s GSCID to allow us to complete the comparative genome analysis for previously sequenced bacteria from 9 vancomycin resistant *S. aureus* cases that have occurred in the US. The strains of interest include the 9 VRSA strains recovered from these infections, as well as the 5 VRE *E. faecalis* strains co-isolated and believed to be the source of the vancomycin resistance determinants. We already have obtained 1) 454 generated sequence information for the VRSA and VRE; 2) Illumina sequence data for the VRSA and VRE; 3) OpGen optical maps for the VRSA and VRE; 4) Merged 454 and Illumina data sets for the VRSA. We are asking for help analyzing these sequence data to determine whether the VRSA strains or VRE donors have unusual traits that predisposed them to move vancomycin into this leading cause of human infection. Moreover, this now more than doubles the data on this most important *S. aureus* lineage (ST5) among hospital strains and second most important lineage for community infection, raising an opportunity to accomplish a more complete analysis of this lineage in comparison to the rest of the species. It’s clear that our team needs help analyzing these data, which for a group like ours represents a significant strain on our genomics knowledge and computational infrastructure. We are proposing to carry out this collaboration with the Broad because the analytic needs are so similar to those of our ongoing Enterococcus project. We are familiar with the Broad’s scientists, analytical tools and output, and already hold frequent face-to-face meetings between the groups. We feel that a relatively small amount of support would be required to complete the analysis and that the most rapid way to do this would be to build on our existing collaboration with the Broad.
2. Justification

Methicillin resistant *S. aureus* (MRSA) are leading causes of life threatening infection in the hospital and in the community. The recent entry of vancomycin resistance into MRSA has now eliminated an important last line of treatment. There now have been well documented cases of transfer of vancomycin resistance from enterococci to MRSA in the United States and additional reports of such infection are also beginning to appear across the globe. Of the transfers of vancomycin resistance from enterococci to *S. aureus* in the USA, the vast majority occurred as independent events in an outbreak in the Detroit metropolitan area. To understand the nature of vancomycin resistance transfer from enterococci to MRSA resulting in VRSA, we have generated high quality sequence data for the first 9 VRSA strains, and 5 of the VRE that were implicated as donors.

An unlikely series of events seems to have led to the transfer of vancomycin resistance from enterococci into staphylococci. The VRSA all possess vancomycin resistance of the VanA phenotype, and in the vast majority of cases, this was transferred on a plasmid carried by an *E. faecalis* donor. The enterococci donors in these cases all appear unique by multi-locus sequence typing (MLST). Despite donor heterogeneity, a striking commonality between the initial well characterized 9 cases is that the *S. aureus* strains were all of the same MLST (i.e. ST5).

In sequencing these isolates we are hoping to get some idea as to why these strain were able to acquire vancomycin resistance. These strains all came from independent de novo transfers of vancomycin resistance from enterococci in mixed VRE/MRSA infections. ST5 represent the most common hospital type of MRSA, and second most common community acquired MRSA, but the genomes sequences of only 2 Japanese strains and 2 strains from New York are available. With the 9 completed VRSA genomes, we will vastly expand available sequence data on this important lineage.

It is important to know why all of these transfers occurred in this lineage. It is the most common, but even at that represents only about 50% of hospital acquired MRSA. It is also important to know why vancomycin resistance was conveyed by *E. faecalis*, when vancomycin resistance is much more common in *E. faecium*. What are the unique properties of ST5's in general, and these strains in particular, that make them the leading cause of hospital MRSA infection, and led to their emergence as VRSA? Do the VRSA strains represent a distinct subset of ST5 that are uniquely susceptible as recipients or where the fitness cost of
vancomycin resistance is minimal compared to other lineages? Are there unique metabolic properties of these ST5 lineages, that may have facilitated or predisposed it to colonizing plantar ulcers in diabetics? Are there special properties of the VRE donors that positioned them at the site of these mixed infections, or facilitated genetic exchange between them and ST5 MRSA? Do they have identical restriction systems or complementary metabolic pathways?

3. Rationale for Strain Selection

The genomes to be analyzed represent all of the well documented VRSA strains that have occurred in the US – VRSA1-9. These strains were obtained through the NIAID NARSA collection. As noted above, all are sequence type 5 (ST5) strains. High quality draft sequences of VRSA 1 – 9 already exist. We also have extensive sequence data on VRE1, 4, 5, 6 and 9, the only VRE donors recovered from mixed infection with the VRSA, and presumed to be the sources for vancomycin resistance.

4a. Approach to Data Production: Data Generation

**Sequence and assembly information**

The first 9 Vancomycin resistant *S. aureus* genomes, and VRE1,4,6 and 9 have been sequenced. Each of these genomes has one 454 run (S. Gill, University of Rochester) and one paired end Illumina run (Tufts Genomic Core, Kip Bodi). With the aid of the Sanger Center, instruction was obtained that allowed the sequence data for the VRSA to be assembled. The Illumina data was assembled first using Velvet and the data combined using Newbler. The assemblies were then run through a program called IMAGE to close some of the gaps computationally. The contigs were then ordered by running them through a program called ABACAS. The contigs were ordered against the genome of a sequenced *S. aureus* strain N315. Unmapped contigs in this case are appended at the end of the assembled genomes. Files of these assemblies are available in fasta format. Using the viewing program, Artemis, .tab files which allow for the viewing of the position of the individual contigs within the sequence are also available. Opgen maps are also on hand and can aid in ensuring that the assemblies are correct. The assembly files (fasta) have been submitted to the Institute for Genomic Science (IGS) at the University of Maryland for annotation. VRE files are available in the FASTA format, but less progress has been made with those.
4b. Approach to Data Production: Data Analysis

The types of analysis which we are currently interested in completing on these genomes to try and address some of these questions include:

1. By comparison to the 15 existing *S. aureus* genomes of all sequence types, analysis of rapid areas of genome drift (e.g., surface proteins)
   - signature sequence analysis (looking for phage, repeats, transposons, integrases)
   - dN/dS ratios
   - identifying orphan genes

2. Calculation of divergence of strains from each other from other sequence type 5's and from other *S. aureus* sequence types
   - SNP analysis
   - OrthoMCL

3. KEGG analysis- is there something specific in the metabolism of these strains that would make them more likely to being found in diabetic patient foot wounds? Survival in mixed infections?

4. The same questions would be asked of the VRE strains relative to the previously completed set of enterococcal genomes in the Broad Institute database.

5. Additional follow up analysis growing out of patterns that emerge in the above comparisons

5. Community Support and Collaborator Roles:
The foundation of these studies was funded by NIAID P01AI083214. This program project is designed to generate new compounds for treating multidrug resistant MRSA, VRE and VRSA infection. This is a Harvard-wide project that involves as subproject investigators David Hooper, MD (MGH); Lefteris Mylonakis, MD, PhD (MGH); Frederick Ausubel, PhD (MGH); Suzanne Walker (PhD) HMS; Roberto Kolter PhD(HMS); Richard Losick, PhD (FAS); and Michael Gilmore, PhD (MEEI). Advisors to the project include Christopher Walsh, PhD (HMS); Stephen Projan, PhD (MedImmune); Jared Silverman, PhD (Cubist), and Nathaneal Gray, PhD (HMS).

The data generated in this work will be validated and placed in GenBank, and will be shared with the NIAID NARSA community through the NARSA website and at the meeting immediately following analysis of the data, prior to publication.

6. Availability & Information of Strains:

1. The strains are available, were obtained from the NIAID NARSA repository, and are largely already sequenced. Limited additional sequence may be helpful if, following assembly, extensive sequence gaps remain.
2. This is mainly a request for bioinformatic analysis of existing data.
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
<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 X Decline □

7b. Public Access to Reagents, Data, Software and Other Materials:
All strains are already available in the NIAID NARSA strain repository. We will make all sequence data available to the community through GenBank and NARSA per standard NIAID supported practices.

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:

[Signature]

Investigator Name: Michael S. Gilmore Date: 01-11-2011