White Paper Application

Project Title: Norovirus Genomics and its Application to Vaccine Development

Authors:

Primary Investigator Contact:

Name	Kim Y. Green
Position Sr.	Investigator
Institution Caliciv	iruses Section/NIAID/NIH
Address	50 South Drive-MSC8026 Bethesda, Maryland 20892
State	Maryland
ZIP Code	20892
Telephone	
Fax	
E-Mail	kgreen@niaid.nih.gov

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

Noroviruses are a major cause of acute, epidemic gastroenteritis. Extensive genetic and antigenic diversity exists among circulating noroviruses, but the molecular mechanisms responsible for the generation of this diversity are poorly understood. In order to facilitate development of norovirus vaccines, the goal of this proposal is to establish a genomics approach to the analysis of norovirus diversity and evolution. Two parameters of norovirus evolution will be examined:

1. Norovirus evolution in populations: The genomic sequences of noroviruses belonging to diverse genotypes obtained over a period of several decades will be determined in order to establish the phylodynamic features of different norovirus genotypes.
2. Norovirus evolution in individuals: The genomic sequences of noroviruses collected from the same individual over the course of a single infection will be examined in order to define sites in the viral genome subject to positive selective pressure within a single host. A comparison of human norovirus evolution in immunocompetent and immunocompromised individuals as well as in chimpanzees will be conducted.

Taken together, this information should give insight into the regions of the norovirus genome associated with strain specificity and host adaptation. An assessment of the role of antibody selective pressure in the viral capsid protein (if identified) will be of special interest in designing and evaluating vaccines.

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.*

   Noroviruses cause acute gastroenteritis in all age groups. There are an estimated 23

million norovirus infections in the U.S. alone each year, and outbreaks are common in closed settings such as nursing homes, schools, institutions, hospitals, ships, and military facilities. Furthermore, noroviruses are now recognized as the second leading cause (after rotaviruses) of severe, life-threatening gastroenteritis in infants and young children. Noroviruses can cause prolonged shedding and diarrhea in immunocompromised individuals that lasts for months, complicating patient care. Noroviruses are on the NIAID Group B pathogens list due to their high infectivity, and potential to cause disruption if introduced deliberately into a common water or food supply system. There are no vaccines or antiviral drugs for these viruses, and the development of control strategies is an important public health goal.

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?*

   Noroviruses are classified in the genus ***Norovirus*** of the virus family ***Caliciviridae***. The caliciviruses are a large and diverse family of single-stranded positive-sense RNA viruses. Norovirus genomics has lagged behind other viruses due to the tremendous genetic diversity among strains and technical difficulties such as the absence of a cell culture system in which to amplify virus. Complete genomic sequences have been determined for certain prototype strains by the analysis of viral RNA extracted from stool material, but many genotypes have not been fully sequenced. The norovirus RNA genome is approximately 7.6 kb in length and polyadenylated at the 3'-end. The norovirus sequencing database consists predominantly of partial genomic sequences such as the diagnostic polymerase region and the capsid gene. The 5'-end of the genome is relatively conserved. The available norovirus sequence database would likely enable assembly of a genomic contig for most uncharacterized or emerging norovirus 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.*

   The available norovirus genomics database consists predominantly of partial sequences obtained from viruses associated with individual cases of gastroenteritis. There has been no systematic attempt to scale-up sequence analysis of large numbers of norovirus genomes that could yield insight into evolution and host selective pressure. In fact, most diagnostic assays have been directed toward small fragments of the genome that would miss the detection of a virus that had undergone recombination. The noroviruses have a single major capsid protein that is thought to bear the protective epitopes and bind to the host cell receptor (s). This is a key site where evolution in the host will be monitored throughout these proposed studies, and the role of antibody pressure will be explored. We are

interested also in the no nstructural proteins, which hijack the cellular machinery in order to replicate the RNA genom e. Ma pping sites involved in adaptation to the host may give insight into protein function and potential drug targets.

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.*

The norovirus strains targeted for this study are from the following sources:

1. Archival collections of norovirus outbreak specimens generated and maintained by the Laboratory of Infectious Diseases (LID), NIAID
2. Adult volunteer stool specim ens obtained through collaboration with Ligocyte, Inc. (curren tly evalu ating noroviru s vaccine candidates) and Drs. Estes and Atmar at the Baylor College of Medicine
3. Chimpanzee animal studies (animal study LID 15)
4. Immunocompromised individuals in the NIH Clinical Center (IRB study protocol pending)
5. Infants and young child ren admitted to a pediatric hospital (CHOP) with acute gastroenteritis through collaboration with Dr. Sheila Nolan

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.)*

It is anticipated that this work will provide the following data:

1. First full-length genomic sequences for several distinct norovirus strains and genotypes
2. First full-length genomic sequences of sequentially-collected noroviruses from an individual immunocompetent host (human or chimpanzee) or an immunocompromised host
3. Identification of residues in norovirus genome undergoing positive selective pressure within a single host

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.*

An expansion of the public norovirus sequence database would provide a valuable tool for researchers. These data will undoubtedly be utilized in basic research of noroviruses (e.g., prediction of conserved functional dom ains in proteins, addressing an tigenic diversity in vaccine formulations, m apping conserved RNA signals, and defining host cell ad aptation) as well as global epidem iologic surveillance studies of e merging norovirus strains (e.g., tracking new strains, developm ent of m ore sensitive an d broadl y-reactive diagnostic assays, and defining “hot-spots” of recom bination between genom es). Som e i mportant

questions around which new hypotheses could be formed include:

1. How diverse are the noroviruses in regard to genogroups and genotypes?
2. Which genotypes are most prevalent in human disease, and why ? Does the nucleotide substitution rate influence the epidemiology of the predominant genotypes, such as GII.4?
3. Is there an animal reservoir, and do zoonotic infections occur?
4. How does evolution of the virus contribute to fitness in the host? Can this information be used to develop a permissive cell culture system?
5. How often does recombination between two norovirus genomes occur, and by what mechanism?
6. Does herd immunity exist and do antibodies influence the emergence of new strains? Can a virus "escape" antibody pressure?
7. Do humans and chimpanzees exert similar selective pressures on viral replication in the gut? Can the chimpanzee animal model be improved by selection of a "chimpanzee-adapted" virus?
8. How does recognition of histo-blood group antigens influence norovirus pathogenesis?
9. Can tracking of norovirus strains globally by sequence analysis of complete genomes improve disease control measures?

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.*

Noroviruses have a global distribution and increasing numbers of public health diagnostic laboratories are testing for their association with acute gastroenteritis. The genetic diversity of these viruses has created a demand for broadly-reactive PCR primers and probes. A larger sequence database would enable the design of such primers. The Center for Disease Control and Prevention (CDC) in the United States and the European Foodborne Diseases Network are examples of agencies that maintain databases for members to track the emergence of new norovirus strains, and several other countries (continents) are tracking noroviruses as well. Newly-obtained sequences are often sequestered or delayed for some time prior to release. This project would provide new sequences in a timely manner to researchers.

Norovirus vaccine candidates are under investigation, but their efficacy must be evaluated in human volunteers. Little information has been generated relating to the evolution of a norovirus strain within an individual host, and it is unclear whether antibody pressure plays a role in escape from immunity. Furthermore, the extent of genetic and antigenic diversity and the epidemiologic role of individual genotypes require additional clarification.

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

Dr. Kim Green
Role: Dr. Green will lead the study and coordinate sample collection and data distribution to the collaborators.

Dr. Stanislav Sosnovtsev and Dr. Karin Bok, NIAID

Role: Drs. Sosnovtsev and Bok are members of the Caliciviruses Section of the NIAID DIR and will be responsible for optimizing the technical aspects of the project and in the analysis of data.

Dr. Mary Estes and Dr. Robert Atmar, Baylor College of Medicine
Role: Drs. Estes and Atmar have conducted numerous human challenge studies with Norwalk virus, the prototype norovirus strain. They will provide stools from adult volunteers shedding Norwalk virus for analysis of the evolution of Norwalk virus over time in an individual immunocompetent host.

Dr. Charles Richardson, Ligocyte, Inc.
Role: Ligocyte, Inc. is nearing the completion of Phase II clinical trials in the evaluation of norovirus vaccine candidates for efficacy. Stools from individuals receiving placebo or vaccine "failures"(if identified) will be provided for sequence analysis.

Dr. Robert Purcell, Laboratory of Infectious Diseases, NIAID
Role: Dr. Purcell leads the chimpanzee research program in the intramural research program, LID, NIAID. His expertise is essential to the design and completion of norovirus challenge studies in chimpanzees.

Dr. Sheila Nolan, Children's Hospital of Philadelphia (CHOP)
Role: Dr. Nolan is an infectious diseases specialist and will coordinate the identification and collection of stool specimens from infants and young children with severe norovirus gastroenteritis at CHOP. The analysis of these noroviruses will yield insight into the characteristics of epidemiologically important strains for inclusion in a norovirus vaccine.

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

The project will be supported, in part, by the DIR under Project Z01AI000897, "The Noroviruses Associated with Epidemic Gastroenteritis"

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. Attach relevant information, if available in an excel spreadsheet for multiple samples: e.g

We request funding to sequence the 7.6 kb genome of 130 full-length norovirus genomes, distributed among the collaborator specimens as follows:

AVAILABLE NOW:
1. "Archival" noroviruses: 30 strains representing different genotypes or variants of genotypes
2. Sequentially-collected noroviruses from normal adult volunteers: 30 strains (10 volunteers with an early, middle, and late-shedding time point)

3. Chimpanzee study: 30 strains

AVAILABLE IN NEAR FUTURE:
4. Vaccine study placebo recipients: 10 strains
5. Vaccine study "failures": 10 strains (assuming possible vaccine failure at low doses of vaccine)
6. Children's Hospital study: 10 strains
7. Immunocompromised patients: 10 strains (5 norovirus-infected patients, with one current and one future time point stool collection)

- Name: Norovirus
- Identifier: Coded identifier and date collected
- Material type (DNA/RNA/Strain): Single-stranded viral RNA
- Genus: Norovirus
- Species: Norwalk virus
- Specimen / Strain: Named for location of outbreak (infection) and year
- Isolation source: Stool material
- Isolated from: Humans or chimpanzees
- Select agent status: Not a select agent
- International permit requirement: No
- BEIR/ATCC repository accession number: To be determined
- Other public repository location
- Other public repository identifier
- Sample provider's name: Kim Y. Green
- Sample provider's contact: kgreen@niaid.nih.gov (301 594 1665)

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

The clinical data from patients with norovirus infection have been or will be obtained under approved IRB protocols. Clinical data, as relevant, will be provided in the published results.

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/).*

The starting materials in this project are human or chimpanzee stool samples that are limited in quantity and non-renewable. Presently, there is no cell culture system for propagating the noroviruses in these stools. Our plan is to provide high-quality purified RNA extracted from virus-positive stool specimens to the designated facility for deep sequencing analysis. We propose that the original stool specimens remain in the laboratory of each collaborator in the event that additional RNA must be extracted for analysis. Aliquots of the original stool material without personal identifiers can be provided to the NIAID repository if required. However, it will be difficult- if not impossible- to offer these specimens for distribution because they are non-renewable. This likely accounts for the unavailability of noroviruses on the BEI website.

The sequencing data from these studies will be analyzed for publication, but the data can be released immediately into a public database, once vetted for accuracy by Dr. Green and the corresponding collaborators. Publications will include NIH investigators, insuring that public ations are available in PubMed Central for rapid and free public access.

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: Kim Y. Green **Date 6/21/2010**