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:

Name	Gabriel I Parra
Position	Visiting Fellow
Institution	NIAID, NIH
Address	50 South Drive, BLD 50, Room 6316
State	MD
ZIP Code	20892
Telephone	216-527-7170
Fax	301-480-5031
E-Mail	gabriel_parra@hotmail.com, parrag@niaid.nih.gov

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

Provide an executive summary of the proposal.

Group A rotaviruses are the leading cause of non-bacterial severe diarrhea disease in young children worldwide. Rotavirus genome contains 11 segments of double-stranded RNA, which are enclosed into a triple-layer capsid. The two outermost capsid proteins (VP7 and VP4) elicit neutralizing antibody responses and specify the G and P serotypes, respectively. Although characterization of G- and P-types is considered essential to understand rotavirus epidemiology and to evaluate the efficacy of current vaccines, whole-genome sequencing from human and animal strains have recently become an important tool to track the origin and spread of rotavirus strains. During the last decade, the number of whole-genomes from rotavirus isolated worldwide has increased to over 150; however, most of the data from developing countries is limited to scattered uncommon strains found in humans and animals from different geographical locations. Thus, there is no study that has analyzed the whole-genome from multiple prevalent rotavirus strains consecutively collected in developing countries.

We have a set of samples from South America, longitudinally collected during the 1990s and 2000s, and prior to the implementation of rotavirus vaccines. This collection of samples provides a unique opportunity to study the potential mechanism of evolution used by rotaviruses to persist in a given population and will help to establish a baseline for post-vaccine surveillance in developing countries. With the data collected during this project we expect to: *i.* determine the evolutionary links among strains circulating in different years and different countries, *ii.* identify residues in rotavirus genome that under go through positive selection due to structural constrains or antigenic pressure, *iii.* track the evolutionary pathway followed by minor (uncommon) and major (prevalent) strains, *iv.* corroborate whether the constellations of genes are still maintained in seasons when multiple strains co-circulated at high frequencies, and *v.* determine the frequency of reassortments among major and minor strains.

The data generated from this project will set the baseline for the diversity of rotavirus during the pre-vaccination period in South-America, and will provide valuable information to the post-vaccination period as well as to the worldwide epidemiological dynamics of

rotavirus. Therefore, we expect the data from this project would be of great interest to scientists not only from South-American countries but instead all over the world.

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.*

Group A rotaviruses are the leading cause of non-bacterial severe diarrhea disease in young children worldwide; resulting in approximately half a million deaths per year. More than 90% of the deaths related to rotavirus occur in developing countries (1). It has been estimated that in Latin America and the Caribbean, rotavirus causes approximately 10 million diarrhea episodes, 2 million clinic consultations, 75,000 hospital admissions and 15,000 deaths annually (2).

Rotaviruses spread through the fecal-oral route and can be detected year-round; but seasonal patterns of infection, with the highest incidence during the coolest and driest months of the year, have been reported in temperate zones (3). Improvements in hygiene and sanitation have shown a great impact in reducing diarrheas associated to bacteria and parasites, but not those associated with rotavirus (1). Therefore, mass vaccination is considered the best strategy to decrease the severity of this infection, and protect susceptible populations.

The two outermost capsid proteins (VP7 and VP4) elicit neutralizing antibody responses and specify the G and P serotypes, respectively. Surveillance of G and P serotypes is considered essential to understand rotavirus epidemiology and to evaluate the efficacy of vaccines. Currently, two different rotavirus vaccines have been licensed for use in vaccination programs. Both vaccines have been shown to be highly efficient against any severe rotavirus episode; however, concern has been raised about their effectiveness against uncommon strains (4).

(1) Parashar UD et al. *Emerg Inf Diseases* 2003
(2) de Oliveira LH et al. *Expert Rev Vaccines* 2008
(3) Cook SM et al. *Bull World Health Organ* 1990
(4) Angel et al. *Curr Op Virol* 2012

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?*

The rotavirus genome contains 11 segments of double-stranded RNA (dsRNA), which

encodes for six structural proteins (VP1-VP4, VP6, VP7) and six non-structural proteins (NSP1-NSP6). The genome is approximately 18.5 kb, and the size of the segments span from 0.7 kb to 3.3 kb (5). Similar to other RNA viruses, rotaviruses present a great diversity that is represented by the multiple genotypes described in humans and animals. Based on the sequence diversity of the VP7 and VP4 proteins, 27 G-types and 35 P-types infecting different species have been reported to date (6); however, only few strains (e.g. G1P[8], G2P[4], G3P[8], G4P[8], G9P[8]) have been shown to be predominant in humans (7).

Although G- and P-typing of circulating strains have been essential to monitor rotavirus diversity, whole-genome sequencing (i.e., all 11 segments) analyses from human and animal rotavirus strains have recently become an important tool for tracking the origin and spread of rotavirus strains (6). Despite the increasing number of whole-genomes sequenced for rotavirus over the last decade (>150), most of the data is limited to scattered uncommon strains found in humans and animals from different geographical locations (6). Only one group has analyzed multiple genomes (>60) from common human strains circulating in a community over a defined period, i.e. Washington, DC Hospital during the 1970s and 1980s (8, 9). In these studies it was shown that different strains from G3P[8] and G4P[8] genotypes were introduced into the population with little or no reassortment, suggesting the presence of preferred constellations of genes in individual strains (8). Our groups have monitored rotavirus diversity in Paraguay, Argentina and Uruguay for over 10 years, and we have found that there is an epidemiological link among the strains circulating in the region (10-12). To further understand rotavirus epidemiology in South-America, we conducted whole-genome sequencing of an atypical strain (G12P[8]) that became predominant in Argentina during 2009. We showed that the new G12P[8] strain was newly introduced to Argentina, rather than being a reassortant from G12P[9] strains that were previously shown to circulate at low frequencies in South-America (12). This work constituted the first whole-genome analysis of human rotaviruses from South-America.

(5) Trask et al. *Curr Op Virol* 2012
(6) Matthijnssens and Van Ranst *Curr Op Virol* 2012
(7) Santos and Hoshino *Rev Med Virol* 2005
(8) McDonald et al. *PLoS Pat* 2010
(9) McDonald et al. *Inf Gen Evol* 2011
(10) Parra et al. *J Med Virol* 2005
(11) Parra et al. *J Clin Virol* 2007
(12) Stupka et al. *J Clin Virol* 2012

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.*

We recently published a whole-genome analysis of an emergent strain (G12P[8]) that become predominant in Argentina during 2009. We showed that the new G12P[8] strain possessed a constellation of genes different from the G12P[9] strains that were previously shown to circulate at low frequencies in South-America (12, 13). Although we were able to

predict that the emergent G12P[8] strains were newly introduced to Argentina, we have no data to compare the genome constellation with common strains previously circulating, and therefore making it difficult to track the relationship with other circulating strains. Thus, we expect that the whole-genome sequence of the strains proposed in this work will set the baseline for future epidemiology and evolutionary analyses conducted regionally (i.e. South-America) as well as to provide valuable information to gain insights on rotavirus worldwide epidemiological dynamics and adaptation to humans.

(13) Martinez et al. *Arch Virol* 2010

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 have a large set of characterized samples from South America, collected during the 1990s and 2000s and prior to the implementation of rotavirus vaccines, which provides a unique opportunity to study the potential mechanism of evolution used by rotaviruses to persist in a given population and will establish a baseline for post-vaccine surveillance in developing countries. The rationale to include the samples collected in three different countries is as follows:

1. Paraguay: During a surveillance of rotavirus diversity in Paraguay from 1998 to 2009 we shown that the peaks of incidence correlated with single prevalent strains (e.g., G4P[8], G9P[8], G1P[8], or G2P[4]) that shifted from season to season (or year-by-year). These replacements in the prevalent strains coincided with the emergence of new strains that showed a wider spectrum on the age of infected children (14). Interestingly, the phylogenetic analyses of the VP7 gene from strains isolated from neighboring countries (e.g. Paraguay, Argentina, Uruguay) have shown the simultaneous introduction of new strains, suggesting a common epidemiological source (10, 15, 16). Thus, even though we were able to show changes of the prevalent strains by characterizing their G and P-types, it is still not clear whether the yearly changes were due to the emergence of novel rotavirus strains into the population or were minor strains that by reassortment (or point mutations) resulted in "new" advantageous characteristics, such as major antigenic shifts or better viral protein interactions. Thus, most of the strains (~85/100) included in this work will represent a set of samples from each of the seasons.
2. Argentina and Uruguay: During 1997 and 1999 a high prevalence of G4P[8] strains were detected in Argentina, which was shown to correlate with the emergence of three different strains. One of those was a prevalent strain in Paraguay and, thus, we expect to gain insights in the evolution and cross-border transmission of strains with the inclusion of G4P[8] samples from Argentina and Uruguay (10, 15). In addition, G4P[6] strains were detected in Paraguay and Argentina and the whole-genome analysis will help us to confirm the possible animal origin of these strains detected in humans (16), and ultimately to understand the mechanisms used by rotavirus to evolve and persist in humans.

(14) Parra *J Med Virol* 2010
(15) Berois et al. *J Med Virol* 2003
(16) Stupka et al. *Inf Gen Evol* 2010

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.)*

The data generated in this work will provide the first whole-genome sequence analyses of prevalent rotavirus strains longitudinally collected in a developing country, as well as provide baseline information to further track the evolution of emergent rotavirus strains worldwide.

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.*

This data collected in this study will be used to: *i.* determine the evolutionary links among strains circulating in different years and different countries, *ii.* identify residues in the rotavirus genome that under go through positive pressure due to structural constrains or antigenic pressure, *iii.* track the origin of uncommon or minor strains, as is the case of the emergence of uncommon G4 or G12 strains, *iv.* corroborate whether the constellations of genes are still maintained in seasons when multiple strains co-circulated at high frequencies, and *v.* determine the frequency of reassortments among major and minor strains. We will use computational analyses to determine the phylodynamics of the strains as well as to map the amino acid substitutions in the solved crystals structures from viral proteins.

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.*

Rotavirus diversity has been shown to be higher in developing countries; and concern has been raised about how the massive vaccination against rotavirus can change the frequencies of the strains circulating (4, 5). The data generated from this project will set the baseline for the diversity of rotavirus during the pre-vaccination period in South-America, and will provide valuable information to the post-vaccination period as well as to the worldwide epidemiological dynamics of rotavirus. Thus, we expect the data from this project would be of great interest to scientists not only from South-American countries but instead worldwide.

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

1. **Dr. Gabriel I Parra,** Molecular Biology Department-IICS, National University of Paraguay, Asuncion, Paraguay (See present address in PI contact).

Role: Dr. Parra will lead the study; coordinate sample collection, data analyses and manuscript drafting.

2. **Dr. Juan Stupka**, Viral Gastroenteritis Laboratory, Malbran Institute, Buenos Aires, Argentina.

Role: Dr. Stupka will provide will provide stool samples with rotavirus G4P[6] and G4P[8] from different locations from Argentina and responsible in the data analyses and manuscript drafting.

3. **B Sc. Magaly Martinez (Ph D student)**, Molecular Biology Department-IICS, National University of Paraguay, Asuncion, Paraguay.

Role: BSc. Martinez will be responsible to provide stool samples from Paraguay and responsible in the data analyses and manuscript drafting.

4. **Dr. Juan Arbiza**, Virology Section, National University of Uruguay, Montevideo, Uruguay.

Role: Dr. Arbiza will provide stool samples with rotavirus G4P[8] from Uruguay and responsible in the manuscript drafting.

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

Currently, we have not projects to fund the whole-genome sequence of these samples. Using internal funding the Argentinean group was able to sequence the whole-genome from seven strains (3 G12P[8], 4 G3P[8]) not directly linked to the set of samples from this proposal (see reference 12). We expect that the data from this project will help us to present future projects to continue the surveillance of rotavirus in the post-vaccine period and to gain insights in the mechanism of rotavirus persistence in humans.

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.

We have a large collection of rotavirus-positive samples, which were collected from 1998 to 2009, genotyped, and stored at -20°C for posterior analysis. Viral dsRNA from those samples will be extracted, and its quality will be tested in electrophoresis gels before being shipped for whole genome sequencing.

Although we have selected 135 samples from our collection, we request funding to sequence 100 genomes of rotavirus detected in South America. The distribution of samples will be:

1. 80 samples representing the predominant strains from each epidemiological season from the Paraguayan collection (expected to sequence ≥6 strains from each season).
2. 10 samples representing atypical G4P[8] and G4P[6] strains collected in Argentina and Uruguay.
3. 10 samples representing atypical G12P[9] and G3P[9] strains.

11. *Attach relevant information, if available in an excel spreadsheet for multiple samples: e.g*
    - *Name:* **Rotavirus**
    - *Identifier*: **Samples coded with identifier, locations and date of collection**
    - *Material type (DNA/RNA/Strain):* **double-stranded RNA**
    - *Genus:* **Rotavirus**
    - *Species:* **Group A rotavirus**
    - *Specimen / Strain*
    - *Isolation source:* **Human feces**
    - *Isolated from:* **Human**
    - *Select agent status:* **Not a selected agent**
    - *International permit requirement:* **Yes**
    - *BEIR/ATCC repository accession number:*
    - *Other public repository location*
    - *Other public repository identifier*
    - *Sample provider's name* **Gabriel I Parra**
    - *Sample provider's contact* **gabriel_parra@hotmail.com; parrag@niaid.nih.gov**

Besides of the common information for each sample (listed above), in the excel file is added the G- and P-type as well as the country of origin and date of collection.

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

Date and Location, Gender, Age. The samples were collected based on the presence of acute diarrhea. No other clinical data have been collected for this set of 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://www.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:

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/).

The stool samples for this project are unique and limited in quantity and/or non-renewable. After collecting the data from this project, we will make attempts to cell-culture-adapt selected viral strains; however, it will be difficult to offer these samples for large distribution. Aliquots of the original stool material could be provided for few samples (when enough material is available) for the NIAID repository if required. The sequence data from these studies will be promptly analyzed for publication, and immediately uploaded into the GenBank database for 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: Gabriel I Parra **Date: 9/7/2012**