UNIVERSITY of MARYLAND
SCHOOL OF MEDICINE
INSTITUTE FOR GENOME SCIENCES

Sequencing of Rhinovirus Strains from Complex Clinical Samples

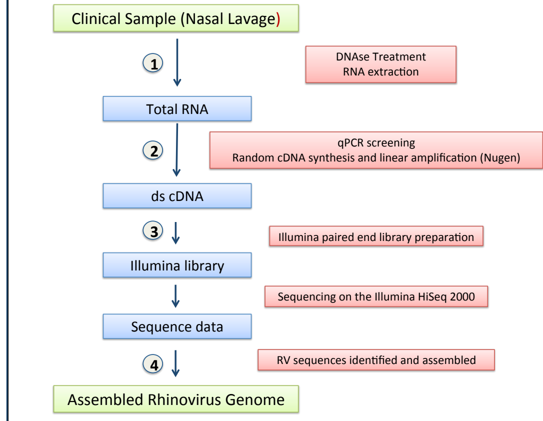
Naomi A. Sengamalay[1], Xuechu Zhao[1], Lisa D. Sadzewicz[1], Al Godinez[1], Erin Hine[1], Qi Su[1], Carrie McCracken[1], Kevin Galens[1], Luke J. Tallon[1], James E. Gern[2], Stephen B. Liggett[3], and Claire M. Fraser[1]

[1]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, [2]University of Wisconsin-Madison School of Medicine & Public Health, Madison, WI, [3]USF Health, University of South Florida, Tampa, FL

Abstract

With more than 150 strains circulating in the human population, the sequencing of human rhinovirus (RV) isolates has been limited to studies in which the genotype of the RV in question is known and a reference sequence is available, or to RVs that were amplified in culture and thus not under the influence of the human environment and with significant titer compared to the native isolates. These limitations make sequencing the rhinovirus genome in clinical samples cost- and time-prohibitive. We have developed a method for sequencing RV strains from complex clinical samples without knowledge of the serotype or amplification of the virus in culture. We have optimized methods to extract RNA from clinical samples, to predict via qPCR which samples are most likely to yield a full rhinovirus genome sequence, and to prepare and sequence samples on the Illumina HiSeq2000. These methods enable large-scale RV genome sequencing from untyped clinical samples and make possible the identification of coinfections and quasi-species as part of the viral swarm of infection.

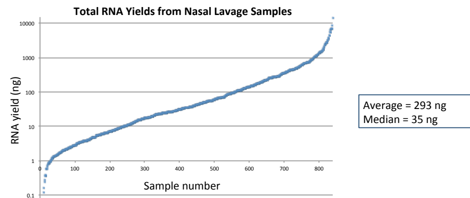
Because the samples were not cultured or subjected to targeted amplification, the majority of the sequence reads in the resulting data sets are of either host or bacterial origin. The identification and extraction of RV reads from this milieu is further complicated by the lack of full-genome reference sequences for modern circulating genotypes. We developed methods for identifying RV reads despite these limitations, and evaluated multiple assembly methods in order to obtain the most complete genomes possible. We have demonstrated that complete draft genome sequences can be obtained from respiratory samples having less than 1% total viral content. As a result of this approach, we have obtained and made publicly available 178 full-length RV genomes from clinical samples, modernizing the sequence dataset available to the enterovirus community. The deep level of sequencing also provides the first opportunity to examine lower prevalence genotypes within the RV swarm. The techniques employed here could be applied to other RNA viruses that have been historically difficult to sequence from native-state samples.



1 RNA Extraction

300 µl of each nasal lavage sample was treated with Dnase I (Ambion, TX) and Viral Nucleic Acid Release Reagent (Express Biotech, MD)

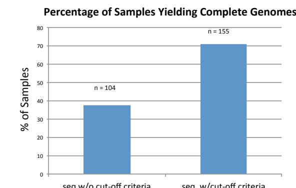
An alkaline lysis extraction-based method was used to isolate RNA (Xpress EZ-RNA 96 Kit (Express Biotech, MD)).



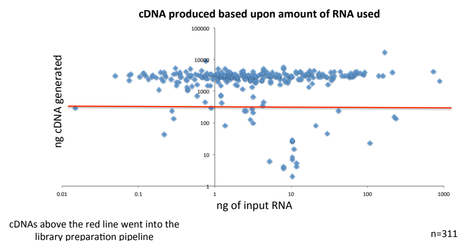
2 RNA assessment and cDNA synthesis

qPCRs were performed to determine the relative levels of Rhinovirus, bacterial 16S, and human RNAse P RNA.

Sequence data from initial specimens, in combination with qPCR results, were used to develop criteria used to determine whether RNA from a particular specimen may produce sufficient rhinovirus reads.



cDNA was synthesized and linearly amplified with the Ovation RNA-seq System (Nugen, CA).



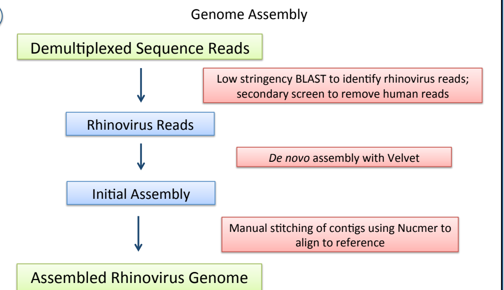
cDNAs above the red line went into the library preparation pipeline

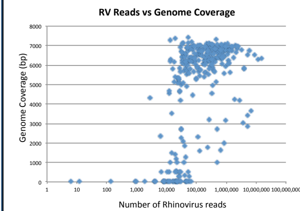
n=311

3 Library Construction and Sequencing

- Indexed paired end libraries for Illumina were made from 250-1500 ng of cDNA
- Libraries were pooled, up to 20 libraries per lane
- Pools were sequenced on 2x100 bp runs on the Illumina HiSeq2000

4 Genome Assembly





- The average % of reads that were mapped to Rhinovirus was 2.65%
- The genotype of each assembly was determined by the VP1 amino acid sequence
- Assemblies that covered at least 85% of the genome were submitted to GenBank

Results and Discussion

Genotype	Count	Genotype	Count	Genotype	Count	Genotype	Count	Genotype	Count	Genotype	Count
RV-A01	4	RV-A34	3	RV-A60	1	RV-B04	1	RV-C02	5	RV-C36	1
RV-A10	4	RV-A36	10	RV-A61	1	RV-B06	12	RV-C03	2	RV-C40	3
RV-A12	1	RV-A38	2	RV-A65	2	RV-B17	1	RV-C04	1	RV-C41	1
RV-A15	1	RV-A40	3	RV-A66	3	RV-B35	3	RV-C05	1	RV-C42	1
RV-A16	7	RV-A43	1	RV-A67	1	RV-B42	3	RV-C06	2	RV-C43	3
RV-A18	2	RV-A44	1	RV-A68	1	RV-B48	1	RV-C07	5	RV-C45	1
RV-A19	2	RV-A47	1	RV-A71	2	RV-B52	1	RV-C08	2	RV-C49	3
RV-A20	4	RV-A49	3	RV-A75	2	RV-B69	2	RV-C15	1		
RV-A21	2	RV-A51	1	RV-A76	3	RV-B70	2	RV-C17	3		
RV-A23	3	RV-A53	1	RV-A80	3	RV-B72	4	RV-C22	1		
RV-A24	4	RV-A55	1	RV-A82	3	RV-B83	1	RV-C25	1		
RV-A28	3	RV-A57	2	RV-A89	2	RV-B84	7	RV-C26	1		
RV-A30	1	RV-A58	1	RV-A101	1	RV-B92	1	RV-C28	1		
RV-A33	2	RV-A59	1	RV-A103	2			RV-C32	2		
RV-A Total: 98						RV-B Total: 39		RV-C Total: 41			

Genotype Representation of Genomes Submitted to GenBank

- Four examples of co-circulating strains were submitted to GenBank. The pairings are: RV-A18/RV-A36, RV-A23/RV-B69, RV-A36/RV-C02, RV-A47/RV-C07
- One strain appears to be a recombination of RV-A76 with RV-A56 (3C and 3D)
- Additional analyses are on-going

We have demonstrated that through deep sequencing of RNA from clinical samples, RNA viruses that are present in the samples at low levels can be detected, and in many cases, their full genomes reconstructed. Most RV-C viruses are not able to be cultured, and with this method, we have been able to contribute 41 complete or nearly-complete RV-C genomes to the GenBank.

Materials and Methods

Clinical Samples

Nasal lavage specimens were collected from two hundred asthmatic and two hundred non-asthmatic children, aged 4-12 years, weekly during fall and spring flu seasons between September 2007 and April 2010. The specimens were stored at -80ºC.

RNA extraction

RNA was extracted from 300 µl of clinical nasal lavage specimen with Xpress EZ-RNA 96 Kit (Express Biotech, MD), following manufacturer’s instruction with modifications. In brief, 300 µl of nasal lavage specimen was pre-treated with 2 µl of DNase I (Ambion) and 100 µl of viral Nucleic Acid Release Reagent (Express Biotech, MD) in 1x DNase I buffer, and incubated at 37ºC for 1 hour. Supernatant of pre-treated sample was further incubated at room temperature for 5 minutes upon addition of 200 µl of cell lysis buffer, supplemented with 2-mercaptoethanol. The samples were then mixed with 300 µl of ethanol and transferred to a filter plate attached to a vacuum manifold. RNA was retained on the filter and washed twice with wash buffer, then eluted in 40 µl of nuclease-free water.

Reverse Transcription and Quantitative Real-time PCR

Each clinical RNA sample was subjected to quantitative real-time PCR to determine the relative quantities of rhinovirus, bacteria, and human RNA. Conversion of 5 µl of clinical RNA into first strand cDNA was performed using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer’s protocol. The resulting cDNA was used for Taqman real-time PCR reactions in a 7900HT Applied Biosystems thermocycler. Real time PCR results were analyzed using the SDS version 2.3 program (Applied Biosystems).

To detect rhinovirus RNA, 2 µl of the cDNA was used in a 10 µl Taqman real-time PCR reaction using a validated human rhinovirus primer/probe mix that target to 5’UTR (a gift from Vanderbilt University. Primers: rhino-fwd: 5’-CY$_{LNA}$**A**GCC$_{LNA}$**T**GCGTGGC-3’; rhino-rev: 5’-GAAACACGGACACCCAAAGTA-3’; probe: 5’-FAM- TCCTCCGGCCCCTGAATGYGGC-BHQ1-3’), under the following cycling conditions: 50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 15 sec, 60°C for 1 min.
To detect the level of bacterial RNA, 2 µl from 1:10 dilution of the cDNA was used in a 10 µl Taqman real-time PCR reaction using a primer/probe set that targets to bacterial 16S ribosomal RNA (Primers: 16S-fwd: 5’-CCT ACG GGD GGC W GCA-3’; 16S-rev: 5’-GGA CTA CHV GGG TMT CTA ATC-3’; probe: 5’-6FAM-CAG CAG CCG CGG TA-MBGNFQ-3’). Cycling conditions are as follows: 50°C for 3 min, 95°C for 5 min, 45 cycles of 95°C for 15 sec, 60°C for 1 min.
For detecting host (human) nucleic acids, 2 µl of the cDNA was used in a 10 µl reaction using Taqman RNase P Control Reagents Kit (Applied Biosystems), with cycling conditions as follows: 50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 15 sec, 60°C for 1 min.

cDNA Synthesis

Clinical RNAs that met certain criteria based upon the qPCR assays were subjected to double-stranded cDNA preparation and amplification using the Ovation RNA-Seq System (NuGen). Briefly, 500pg – 10ng of clinical RNA was reverse transcribed to synthesize the first-strand cDNA by using an RNA/DNA chimeric primer mix. Reverse transcription extends the 3’ DNA end of each primer generating first strand cDNA with a unique RNA sequence at the 5’end. Double-stranded DNA was generated by fragmentation of the RNA template strand using RNA-dependent DNA polymerase. The dsDNA was purified using Agencourt RNAClean XP beads. The DNA was amplified linearly using a SPIA process in which RNase H degrades RNA in DNA/RNA heteroduplex at the 5’-end of the double-stranded cDNA, after which the SPIA primer binds to the cDNA and the polymerase starts replication at the 3’-end of the primer by displacement of the existing forward strand. Finally, random hexamers were used to amplify the second-strand cDNA linearly.

Illumina Library Construction and Sequencing

DNA libraries for Illumina sequencing were prepared from the amplified cDNA samples with the TruSeq DNA Sample Prep Kit (Illumina, San Diego, CA) or with the NEBNext DNA Library Prep Kit for Illumina (NEB, Ipswich, MA), following a variation of the manufacturer’s protocol. The DNA was purified between enzymatic reactions and the size selection of the library was performed with AMPure XT beads (Beckman Coulter Genomics, Danvers, MA). Adapters containing 6 nt index sequences were ligated to the DNA library fragments. The indexed libraries were pooled and sequenced on a 2x100 bp run on the Hiseq 2000 sequencer (Illumina, San Diego, CA).

This project has been funded in whole or part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract number HHSN272200900009C.