Sequencing of Rhinovirus Strains from Complex Clinical Samples


Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD; University of Wisconsin-Madison School of Medicine & Public Health, Madison, WI; YSM 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 changes compared to the native isolates. These limitations make sequencing the rhinovirus genome in clinical samples time- and cost-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 which samples are most likely to yield a full rhinovirus genome sequence, and to prepare and sequence samples on the Illumina HSeq2000. 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 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 genome 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, modernising 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 technique employed here could be applied to other RNA viruses that have been historically difficult to sequence from native state samples.

Materials and Methods

Clinical samples
Nasal lavage samples were collected from two hundred and two hundred and two children, aged 0-12 years, weekly during fall and spring flu seasons between September 2007 and April 2010. The specimens were stored at -40° C.

RNA extraction
RNA was extracted from 500 µl of clinical nasal lavage specimens with Xprem EZ-RNA Kit (Express Biotech, MD). Following manufacturer's instructions with modifications. In brief, 500 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 a 1:1 buffer, and incubated at 37°C for 1 hour. Supernatant of pre-treated sample was furtherly incubated at room temperature for 5 minutes upon addition of 20 µl of RQ I (Qiagen) buffer, supplemented with 2-nucleoprep. The samples were then mixed with 800 µl of ethanol and transferred to a filter plate attached to a sorbus manifold. RNA was retained on the filter and washed twice with wash buffer, then eluted in 25 µl of RNase-free water.

qPCR screening Random cDNA synthesis and linear amplification (Nagen)
For each clinical RNA sample was subjected to quantitative real-time PCR to determine the relative quantities of rhinovirus, bacteria, and human RNA. Expression of 1 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 via TaqMan Applied Biosystems hspxc. Receiver RNA samples were amplified using the USB version 2.0 protocol (Applied Biosystems).

To detect rhinovirus RNA, 2 µl of the cDNA was used in a 30 µl TaqMan reverse PCR reaction using a validated human rhinovirus primer (Taqman Gene Index, Applied Biosystems). The probe (5'-FAM-GAGATCTGCTTCTGGCTGGT-C3') was designed to target a conserved 36-bp region of the viral genome. The cycling conditions were 95°C for 1 minute, 45 cycles at 90°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute. The resulting TaqMan PCR reactions were performed using the ABI 7300 (Applied Biosystems).

To detect 3' terminal redundancy in the viral genome were amplified using primer sets specific to the 3' terminal redundancy in the viral genome (Sidney). The primer sets were 5'-FAM-AGAGATCTGCTTCTGGCTGGT-C3' and 5'-FAM-AGAGATCTGCTTCTGGCTGGT-C3'. The cycling conditions were 95°C for 1 minute, 45 cycles at 90°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute. The resulting TaqMan PCR reactions were performed using the ABI 7300 (Applied Biosystems).

Dilution of clinical RNA was performed by tenfold dilution in RNase-free water. 2 µl of the cDNA was used in a 10 µl reaction using Taqman M slide Control Reagents (Applied Biosystems), with 1 cycle at 95°C for 2 min, 95°C for 2 cycles, 50 cycles at 95°C for 10 sec, 60°C for 10 sec. The cycling conditions were 95°C for 1 min, 45 cycles at 90°C for 1 min, 55°C for 1 min, 72°C for 1 min. The resulting TaqMan PCR reactions were performed using the ABI 7300 (Applied Biosystems). The conditions with fluorescence was 95°C for 1 min, 45 cycles at 90°C for 1 min, 55°C for 1 min, 72°C for 1 min.