Whole genome sequencing of two human rhinovirus A types (A101 and A15) detected in Kenya, 2016-2018 [version 1; peer review: 1 approved]

Martha M. Luka1,2, Everlyn Kamau1, Zaydah R. de Laurent1, John Mwita Morobe1, Leonard K. Alii3, D. James Nokes1,4, Charles N. Agoti1,5

1Epidemiology and Demography Department, KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi, Kenya
2Department of Biochemistry and Biotechnology, Pwani University, Kilifi, Kenya
3Department of Mathematics and Computer Science, Pwani University, Kilifi, Kenya
4School of Life Sciences and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry, UK
5Department of Public Health, Pwani University, Kilifi, Kenya

Abstract
Background: Virus genome sequencing is increasingly utilized in epidemiological surveillance. Genomic data allows comprehensive evaluation of underlying viral diversity and epidemiology to inform control. For human rhinovirus (HRV), genomic amplification and sequencing is challenging due to numerous types, high genetic diversity and inadequate reference sequences.

Methods: We developed a tiled amplicon type-specific protocol for genome amplification and sequencing on the Illumina MiSeq platform of two HRV types, A15 and A101. We then assessed added value in analyzing whole genomes relative to the VP4/2 region only in the investigation of HRV molecular epidemiology within the community in Kilifi, coastal Kenya.

Results: We processed 73 samples collected between 2016-2018, and 48 yielded at least 70% HRV genome coverage. These included all A101 samples (n=10) and 38 (60.3%) A15 samples. Phylogenetic analysis revealed that the Kilifi A101 sequences interspersed with global A101 genomes available in GenBank collected between 1999-2016. On the other hand, our A15 sequences formed a monophyletic group separate from the global genomes collected in 2008 and 2019. Improved phylogenetic resolution was observed with the genome phylogenies compared to the VP4/2 phylogenies.

Conclusions: We present a type-specific full genome sequencing approach for obtaining HRV genomic data and characterizing infections.
Keywords
human rhinovirus, whole-genome, sequencing, phylogenetics
Introduction

Genomic surveillance of respiratory viruses is important for (i) developing molecular diagnostics\(^2\), (ii) investigating transmission and evolution\(^4\), and (iii) development of vaccines and therapeutic drugs\(^1\). Human rhinovirus (HRV) is the most common cause of upper respiratory infections\(^6\) and is also occasionally associated with lower respiratory infections\(^8\). It is a highly diverse virus, with over 160 distinct types identified globally\(^9\). This diversity presents a challenge in developing a sequencing protocol that works well across the different HRV types\(^10\). Most HRV molecular epidemiology studies utilize partial genome sequences, which offer lower resolution in identifying epidemiologically linked infections\(^11\).

Viral genome sequencing can take one of the two approaches available: a targeted/enrichment approach or an aagnostic/metagenomics approach\(^12\). Theoretically, the viral metagenomics approach is an unbiased way to obtain all viral genetic content in a sample as it does not require prior knowledge of their sequences\(^13\). However, this approach requires high viral titers to succeed\(^14\), which are not always available in a clinical sample\(^15\). Furthermore, most clinical samples especially those relevant to HRV sequencing are dominated by host and bacterial nucleic acids\(^16\). Nonetheless, this challenge can be overcome by target enrichment, for example, polymerase chain reaction (PCR) to bulk up for the target virus before sequencing\(^15\)-\(^17\). We describe a target enrichment sequencing approach of two HRV types, A15 and A101, using type-specific primers and compare the phylogenetic inferences between partial and whole genome sequences.

Methods

Study population

The study utilized nasopharyngeal swabs collected during two previous studies in Kilifi County, coastal Kenya: outpatient surveillance of nine health dispensaries within the Kilifi Health and Demographic Surveillance System (KHDSS) between January and December 2016\(^6\) and primary school surveillance between May 2017 and April 2018\(^8\). The school was situated in Junju, a location within the KHDSS. All samples were collected from symptomatic individuals (mild symptoms of acute respiratory tract infection) of varied age (one month - 49 years) and archived at -80°C.

Study design

Samples were screened for HRV and typed as previously described\(^4\). A cycle threshold (Ct) value < 35.0 was used to define positives. HRV positives underwent VP4/2 sequencing to characterize the diversity, spatial and temporal occurrence of HRV in the two settings. Comparison of the HRV diversity within the school in Junju and the health dispensary located in Junju revealed 12 common types, and the most frequent common type was HRV-A101 (n=5 in each setting)\(^6\),\(^11\).

For this study, we purposively selected two types from the two studies: the most frequent type observed in the KHDSS health dispensary surveillance (A15, n=63) and the most frequent common type identified at the Junju health dispensary and school (A101, n=10) for whole genome sequencing (WGS).

Ethics statement

Sample collection was undertaken following an informed written consent provided by parents or guardians for persons <18 years or by participating individuals if aged >17 years. Children whose parents consented were also asked for individual assent to participate. The study protocols were reviewed and approved by the University of Warwick Biomedical and Scientific Research Ethics Committee (BSREC #REGO-2016-1858 and #REGO-2015-6102) and the KEMRI-Scientific Ethics Review Unit (KEMRI-SERU #3332 and #3103).

Primer design

We retrieved nine type A101 genomes and three type A15 genomes, all >6000 nt long from GenBank\(^18\) on 30th September 2019. Geneious Prime 2019.2.1 (https://www.geneious.com) was used to design eight overlapping primer pairs across the ~7.2 kb HRV genome. The primers targeted eight amplicons 0.9–1.6 kb in size, with overlaps varying in size from 300 to 800 bases, Table 1.

RNA extraction, reverse transcription and PCR

Viral RNA was extracted from 140 µl sample using QIAamp Viral RNA kit (Qiagen, USA) as per the manufacturer’s recommendations. Reverse transcription was carried out using random hexamers and the Superscript III First-Strand Synthesis System (Invitrogen, United Kingdom). Genome-wide amplification using HRV-specific primers was done using the Q5 High-Fidelity 2X Master Mix (New England Biolabs, United Kingdom). PCR success was assessed by electrophoresing the products on a 1% agarose gel. Once suitable PCR conditions per amplicon were established, a duplex PCR of non-consecutive amplicons of similar conditions was set up (Protocol doi - dx.doi.org/10.17504/protocols.io.buknxnux).

Sequencing

PCR products were purified with 1X AMPure XP beads (Beckman Coulter Inc.), quantified with Qubit dsDNA High Sensitivity Assay (Invitrogen, United Kingdom), pooled per sample and normalized to 0.2 ng/µL. Sequencing libraries were prepared using the Nextera XT Sample Preparation Kit (Illumina, CA) and sequencing performed on Illumina MiSeq platform (200 bp×2) per sample.

Sequence assembly

The raw reads were quality checked using FastQC v0.11.9 and trimmed (Phred score >30) using Trimomatic v0.39\(^19\). HRV reads were identified by mapping to the respective reference strains (https://www.picornaviridae.com/sg3/enterovirus/rv-a/rv-a_segs.htm) and subsequently assembled into contigs using SPAdes v3.12.0. The contigs were checked for completeness and assembled to a consensus sequence using Sequencher v5.4.6 (www.genecodes.com). We defined sequencing success as obtaining HRV reads covering at least 70% of the genome (>5040 bases). Sequencing depth was visualized using the deepTools\(^20\) package.
<table>
<thead>
<tr>
<th>Name</th>
<th>Target type</th>
<th>Amplicon</th>
<th>Start</th>
<th>Length</th>
<th>%GC</th>
<th>Tm (°C)</th>
<th>Hairpin Tm (°C)</th>
<th>Self-dimer Tm (°C)</th>
<th>Pair dimer Tm (°C)</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>95 F_a101</td>
<td>A101</td>
<td>1</td>
<td>95</td>
<td>23</td>
<td>39.1</td>
<td>58</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>ACCCCAAATGTACTTGAAGCA</td>
</tr>
<tr>
<td>716 R_a101</td>
<td>A101</td>
<td>1</td>
<td>1,334</td>
<td>22</td>
<td>45.5</td>
<td>60</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>TCATCATGTTGGTGTGTTGAGT</td>
</tr>
<tr>
<td>726 F_a101</td>
<td>A101</td>
<td>2</td>
<td>726</td>
<td>20</td>
<td>50</td>
<td>60</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>AGCATAAGTGGAGCGGTCAA</td>
</tr>
<tr>
<td>1,215 R_a101</td>
<td>A101</td>
<td>2</td>
<td>1,833</td>
<td>22</td>
<td>54.5</td>
<td>62</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>GACCCCAACGAACCTGCATA</td>
</tr>
<tr>
<td>827 F_a101</td>
<td>A101</td>
<td>3</td>
<td>1,445</td>
<td>22</td>
<td>36.4</td>
<td>56</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>ACCTGTGTTGTTGGTGCTGTT</td>
</tr>
<tr>
<td>726 F_a101</td>
<td>A101</td>
<td>4</td>
<td>2,142</td>
<td>22</td>
<td>40.9</td>
<td>56</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>TACCACCTGATACATACATCA</td>
</tr>
<tr>
<td>1,524 F_a101</td>
<td>A101</td>
<td>5</td>
<td>2,468</td>
<td>20</td>
<td>50</td>
<td>60</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>ACCCGAGAACACGAGCAGA</td>
</tr>
<tr>
<td>5,248 R_a101</td>
<td>A101</td>
<td>6</td>
<td>3,164</td>
<td>23</td>
<td>39.1</td>
<td>58</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>TCTGGTTGTTGGTGCTGTT</td>
</tr>
<tr>
<td>1,524 F_a101</td>
<td>A101</td>
<td>7</td>
<td>3,464</td>
<td>24</td>
<td>37.5</td>
<td>57</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>ATCCCCCTGAAACCTC CCCA</td>
</tr>
<tr>
<td>5,505 R_a101</td>
<td>A101</td>
<td>1</td>
<td>5,623</td>
<td>24</td>
<td>33.3</td>
<td>56</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>TAGATACCAACACTCTCTCAAC</td>
</tr>
<tr>
<td>2,446 F_a101</td>
<td>A101</td>
<td>8</td>
<td>6,168</td>
<td>21</td>
<td>38.1</td>
<td>55</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>ACCAAATGATACCTTCCCAA</td>
</tr>
<tr>
<td>5,550 R_a101</td>
<td>A101</td>
<td>1</td>
<td>7,001</td>
<td>24</td>
<td>33.3</td>
<td>56</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>TCCATCATTTTCTGCTG TAAATC</td>
</tr>
<tr>
<td>A15_1F</td>
<td>A15</td>
<td>1</td>
<td>21</td>
<td>20</td>
<td>55</td>
<td>61</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>ATCCCCCTGAAACCTCCC A</td>
</tr>
<tr>
<td>A15_1R</td>
<td>A15</td>
<td>1</td>
<td>1,251</td>
<td>20</td>
<td>55</td>
<td>60</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>CCAAGGGTGAGGTTGACCA</td>
</tr>
<tr>
<td>5,34 F_a15_22</td>
<td>A15</td>
<td>2</td>
<td>621</td>
<td>21</td>
<td>52.4</td>
<td>60</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>CCATGGGCGCTCAAGATCTCA</td>
</tr>
<tr>
<td>1,889 R_a15_22</td>
<td>A15</td>
<td>2</td>
<td>1,988</td>
<td>24</td>
<td>33.3</td>
<td>57</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>CACAAAAACATGAACTCTGAT</td>
</tr>
<tr>
<td>1,391 F_a15_22</td>
<td>A15</td>
<td>3</td>
<td>1,478</td>
<td>21</td>
<td>42.9</td>
<td>56</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>AGCATAACCACTGAGCTTG</td>
</tr>
<tr>
<td>2,848 R_a15_22</td>
<td>A15</td>
<td>3</td>
<td>2,947</td>
<td>23</td>
<td>34.8</td>
<td>56</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>TCCATCGATCTCATAACACA</td>
</tr>
<tr>
<td>2,417 F_a15_22</td>
<td>A15</td>
<td>4</td>
<td>2,516</td>
<td>22</td>
<td>40.9</td>
<td>55</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>TCACAGATCTGAGATGAGTA</td>
</tr>
<tr>
<td>3,464 R_a15_22</td>
<td>A15</td>
<td>4</td>
<td>3,563</td>
<td>20</td>
<td>45</td>
<td>55</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>CCTACACACTGTGGTGCAC</td>
</tr>
<tr>
<td>2,900 F_a15_22</td>
<td>A15</td>
<td>5</td>
<td>2,999</td>
<td>24</td>
<td>33.3</td>
<td>54</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>CTATGTGCTAAAGATGCTCAGA</td>
</tr>
<tr>
<td>4,352 R_a15_22</td>
<td>A15</td>
<td>5</td>
<td>4,451</td>
<td>23</td>
<td>39.1</td>
<td>58</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>ACCAGGATTTTGGCAT AAATGTC</td>
</tr>
<tr>
<td>3,576 F_a15_22</td>
<td>A15</td>
<td>6</td>
<td>3,675</td>
<td>21</td>
<td>38.1</td>
<td>55</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>TCGTAGGCGGTTGTTGAAATA</td>
</tr>
<tr>
<td>4,991 R_a15_22</td>
<td>A15</td>
<td>6</td>
<td>5,090</td>
<td>24</td>
<td>33.3</td>
<td>55</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>CAAATAATGCGCTACTA CTA</td>
</tr>
<tr>
<td>4,385 F_a15_22</td>
<td>A15</td>
<td>7</td>
<td>4,484</td>
<td>23</td>
<td>43.5</td>
<td>58</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>TCAAGTGTAACCTTTATC CCC</td>
</tr>
<tr>
<td>5,943 R_a15_22</td>
<td>A15</td>
<td>7</td>
<td>6,042</td>
<td>23</td>
<td>43.5</td>
<td>59</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>GTCACAAACACACTTCCCAA</td>
</tr>
<tr>
<td>A15_8F</td>
<td>A15</td>
<td>8</td>
<td>5,972</td>
<td>20</td>
<td>55</td>
<td>60</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>ACYCTTGAYATTGRC ACG</td>
</tr>
<tr>
<td>A15_8R</td>
<td>A15</td>
<td>8</td>
<td>7,029</td>
<td>20</td>
<td>55</td>
<td>60</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>CTACACACTGCAATCCC TCC</td>
</tr>
<tr>
<td>5,656 F_a15_22</td>
<td>A15</td>
<td>8</td>
<td>5,659</td>
<td>21</td>
<td>42.9</td>
<td>55</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>CATTCTGTTGGTG TGAATGG</td>
</tr>
<tr>
<td>7,076 R_a15_22</td>
<td>A15</td>
<td>8</td>
<td>7,076</td>
<td>20</td>
<td>55</td>
<td>60</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>AAGCCGGGATATA C AGTGC</td>
</tr>
</tbody>
</table>

Tm - melting temperature

Start - Genome position (of the whole genome template used) where the primer sequence starts

GenBank sequences used in primer design were accession numbers: MN306051.1, DQ473493.1 and JN541268.1 for A15 and; KY460514.1, GQ415052.1, KY369891.1, KY189315.1, KY369897.1, KY369892.1, KY369889.1, JQ245965.1 and GQ415051.1 for A101.
Sequence analysis
Sequences were aligned using MAFFT v7.271. Recombination scans were done using RDP5 and visualized on SimPlot. Nucleotide substitutions across the genomes were visualized using a python script to examine genetic diversity across the genome. POPART was used to construct haplotype networks using the Minimum Spanning Network model. The best-fitting model and maximum likelihood trees were inferred using IQ-TREE, v1.6.0. Branch support for phylogenetic trees was assessed using bootstrapping of 1000 iterations. MegaX was used to calculate mean pairwise distances, and the respective standard errors were assessed using 100 iterations.

Bayesian phylogeny was used to create time-structured phylogenetic trees using BEAST v1.10.4. BEAST was run with 200 million MCMC steps using the best fitting substitution model and a coalescent-based relaxed clock framework. The output was assessed for convergence using Tracer v1.7.1. Maximum clade credibility (MCC) trees were identified using TreeAnnotator v1.10.4 after removal of 10% burn-in. The trees were then visualised in FigTree v1.4.4 and branching posterior probabilities were noted.

Statistical analysis
Statistical analysis was undertaken using R version 3.6.1 (R Core, 2021). The Shapiro–Wilk test was used to check for the normality of the data. The T-test was then used to compare Ct-values of successfully sequenced versus failed samples.

Results
Whole genome sequencing
We successfully sequenced all 10 (100%) A101 and 38 of 63 (60.3%) A15 samples. Cycle threshold (Ct) values ranged from 20.2 – 34.7, with a median of 28.4 for A15 and 30.2 for A101. The failed 25 samples did not have a significantly higher median Ct-value than those successfully sequenced based on the T-test: 28.3 (IQR = 4.0) versus 29.2 (IQR = 3.7), respectively (p= 0.21), Figure 1A and B. Besides, samples that failed

![Figure 1. Summary statistics of sequenced samples.](image)
sequencing did not have unique phylogenetic clustering patterns based on their previously generated VP4/2 sequences. Sequencing depth was comparable across Ct-values, with the mean depth coverage per genome ranging from 351 - 13356 reads per base pair, Figure 1C and D.

Phylogenetic analysis identified interspersion of local A101 sequences with global sequences (n=9) collected between the years 1999–2016. However, A15 local genomes clustered separately from global sequences (n=3), Figure 2. The global A15 genomes were collected in the years 2008 (n=2) and 2019 (n=1).

The ends of the 5’ untranslated region (UTR) and 3’ UTR were not amplified due to lack of suitable primers. Genetic diversity was observed across the entire genome and not within

**Figure 2.** Maximum-likelihood phylogenetic trees of local (generated) and global (A) A15 and (B) A101 sequences. The tips are coloured by origin, indicating the global sequences used in primer design. The scale bar represents nucleotide substitutions per site.
a particular genomic region for both types, as shown in Figure 3.

**Phylogenetic resolution**

We compared the phylogenetic bifurcation patterns and statistical uncertainty of VP4/2 and WGS for the two types. The depiction of sister taxa was comparable across the two trees. However, WGS resolved phylogenetic polytomies/unresolved branches observed previously in the VP4/2 phylogenies. For example, using VP4/2 sequences, all viruses collected in the school formed one polytomy, which was now fully bifurcated using WGS. Similarly, for A15, the four polytomies observed on VP4/2 phylogeny were well resolved using WGS, effectively distinguishing one sample from the other. Although the mean pairwise distances across VP4/2 and WGS were close, the standard error of pairwise distance calculations was notably less (about a tenth) in WGS than VP4/2, Figure 4.

Overall higher branching posterior probabilities in Bayesian phylogenetic trees were observed using WGS than VP4/2.

---

**Figure 3.** Genetic diversity across the genome for (A) A15 and (B) A101. The known HRV strains (https://www.picornaviridae.com/sg3/enterovirus/rv-a/rv-a_seq.htm) are used as the reference. A substitution to “A” is indicated by green, “C” by blue, “G” by indigo and “T” by red bars. Gray contiguous bars indicate unknown/unsequenced bases.
**Figure 4.** Maximum-likelihood phylogenetic trees of local (A) A15 and (B) A101 VP4/2 and whole genome sequences. The tips are coloured by site of origin. The scale bar represents nucleotide substitutions per site while node labels indicate bootstrap value. (C) Mean pairwise distances and respective standard errors of VP4/2 and whole genome sequences.
sequences. In VP4/2 Bayesian trees, 17.5% of A15 and 64.7% of A101 nodes had a posterior probability greater than 0.7 compared to 64.3% and 88.2% nodes in WGS trees, respectively, as illustrated in Figure 5.

The improved resolution was further depicted by haplotype networks that displayed notably more alleles using WGS than VP4/2, e.g., in HRV-A101, school sequences that were considered a single allele using VP4/2 sequences resolved into five alleles when using whole-genome sequences, Figure 6. Identical samples at the VP4/2 region had a median of 3 nt changes for A101 and 5 nt changes for A15 across the whole genome.

Recombination analysis
Recombination scans identified breakpoints within the VP3 of one A101 sequence (KEN_Rhinovirus_7018), with both parents belonging to A101 type (p-value < 1.922E-2), Figure 6C.

---

**Figure 5.** Bayesian phylogenetic trees of local and global (A) HRV-A15 and (B) HRV-A101 VP4/2 and whole genomes. The branches are coloured by site of origin. Node labels indicate branching posterior probabilities.
Figure 6. Haplotype networks displaying sequence variation of VP4/2 and whole-genome sequences of (A) A15 and (B) A101. Numbers along the edges indicate the nucleotide substitutions. The alleles are coloured by study site. (C) Recombination scan of recombinant sequence KEN_Rhinovirus_7018 compared to its major and minor parents and the A101 prototype sequence, GQ415051.1. A putative recombinant region was identified within the VP3.
Recombination within HRV structural regions has been shown to be rare and sporadic\textsuperscript{30}.

**Discussion**

This study presents a type-specific whole genome sequencing protocol for two HRV types. A101 had a higher success (100\%) than A15 (60.3\%). We attribute the higher A101 sequencing success to the higher number of sequences (n=9) that were available for primer design, which captured more intra-type variation, compared to A15 (n=3). Having more genomes contributing to the consensus sequence used in primer design increased genetic variation, and subsequently, the likelihood that the local and contemporaneous diversity was captured.

Whole-genome sequencing provided greater phylogenetic resolution and less statistical uncertainty to partial sequencing. Polytomies are a product of inadequate data and are, therefore, a potential source of bias. They also result in reduced statistical power due to increased uncertainty. The loss of terminal phylogenetic resolution may result in two opposing predictions: the underestimation (due to unresolved taxa) or overestimation (due to increased total tree length) of diversity\textsuperscript{31}. Due to the short size of VP4/2 (~420nt), insufficient data results in reduced phylogenetic resolution and increased uncertainty evidenced by a higher standard error in phylogenetic distance. Unresolved phylogenies are a challenge in epidemiology as one cannot distinguish infections from one individual to another for transmission inference.

Posterior probabilities summarize the uncertainties about a parameter and indicate confidence in the evidence\textsuperscript{32}. High posterior probabilities indicate high confidence, and the reverse is also true. Whole genomes consistently provided higher confidence across the two genotypes assessed in this study.

With pathogen sequencing now an established tool to track viral infections\textsuperscript{3,12}, it is crucial to compare the resolution of different sequence analysis. As the huge antigenic diversity of HRV continues to pose a challenge in vaccine development, efforts should be directed towards understanding and mitigating transmission. Our study shows that HRV WGS is better suited for transmission inference to the commonly used VP4/2 sequences.

The sequencing approach we developed has some limitations. First, it requires prior genotyping of the HRV positive samples, commonly done by VP4/2 or VP1 sequencing. It is therefore unsuitable for sequencing new or highly divergent types due to the requirement of matching primer sequencing. Second, having to create primer sets for each type is cumbersome and relies on adequate number of pre-existing genomes to design conserved primers. In addition, an amplicon-based target enrichment does not work well for low complexity regions such as the 5’UTR and 3’ UTR. Notwithstanding, the new method can successfully enrich for human rhinovirus in archived samples of varying virus titers. It can also effectively capture intra-type recombinant regions enabling detailed study of viral dynamics.

**Conclusions**

With HRV being the most common respiratory virus, it is surprising that we have such few publicly available whole genomes to allow detailed intra-type analysis. We describe a new protocol for the whole genome sequencing of two HRV types and enrich the public database of HRV genomes. The protocol can be adapted for other HRV types. Our study also shows that WGS is more informative than VP4/2 sequencing in studying HRV dynamics as it maximizes resolution and reduces phylogenetic uncertainty.

**Data availability**

Accession number: GenBank, MW713746-MW713793

Accession number: BioProject, PRJNA701406

Root URL: https://identifiers.org/bioproject:PRJNA701406


This project contains the following underlying data:

- This is a replication dataset for the manuscript titled: "Whole genome sequencing of two human rhinovirus A types (A101 and A15) detected in Kenya, 2016–2018." The dataset contains contains Cycle threshold (Ct) values, and read/sequencing depth.

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

**Acknowledgements**

We thank all the study participants, the field study team and the laboratory staff of the Virus Epidemiology and Control research group at the KEMRI-Wellcome Trust Research Programme. This paper was published with the permission of the Director of KEMRI.
References


29. Rambut A, Drummond AJ: FigTree version 1.4. 0. 2012.


Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 17 August 2021

https://doi.org/10.21956/wellcomeopenres.18656.r45240

© 2021 Nyaga M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Martín Munene Nyaga

South African Medical Research Council/Diarroheal Pathogens Research Unit, Faculty of Health Sciences, Sefako Makgatho Health Sciences University, Pretoria, South Africa

Overall comments:
Luka et al. (2021) present a type-specific full genome sequencing approach for obtaining genomic data of human rhinovirus. The authors have done an excellent job of addressing a research gap in rhinovirus amplification and sequencing challenges caused by a high genetic diversity and relatively inadequate reference sequences. The study is well designed, with detailed analysis and well-reported conclusions.

Abstract:
The abstract concisely summarizes the research article.

Introduction:
The introduction is well written. The authors have provided a sufficient background and articulated the purpose of the research work in a concise manner.

Methodology:
The methodology section is well designed.

Results:
The result findings are well described. The data analysis is very thorough.

Discussion and conclusion:
The discussion part is well-written and the authors have done a good job of describing the study's shortcomings. It is commendable to see that, despite the limitations, the study's findings corroborate the type-specific sequencing approach being promoted. A question that may need more discussion on is: what could have caused the sequencing failure of 25 samples with a median ct value that was not significantly higher than those successfully sequenced? The conclusion wraps up the study nicely.
Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Virology (eneteric and respiratory )

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.