



## RESEARCH ARTICLE

**REVISED** Whole genome sequencing of two human rhinovirus A types (A101 and A15) detected in Kenya, 2016-2018 [version 2; peer review: 2 approved]

Martha M. Luka <sup>1,2</sup>, Everlyn Kamau<sup>1</sup>, Zaydah R. de Laurent <sup>1</sup>,  
John Mwita Morobe <sup>1</sup>, Leonard K. Alii<sup>3</sup>, D. James Nokes <sup>1,4</sup>, Charles N. Agoti <sup>1,5</sup>

<sup>1</sup>Epidemiology and Demography Department, KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi, Kenya

<sup>2</sup>Department of Biochemistry and Biotechnology, Pwani University, Kilifi, Kenya

<sup>3</sup>Department of Mathematics and Computer Science, Pwani University, Kilifi, Kenya

<sup>4</sup>School of Life Sciences and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry, UK

<sup>5</sup>Department of Public Health, Pwani University, Kilifi, Kenya

**V2** First published: 08 Jul 2021, 6:178  
<https://doi.org/10.12688/wellcomeopenres.16911.1>  
Latest published: 23 Sep 2021, 6:178  
<https://doi.org/10.12688/wellcomeopenres.16911.2>

**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 nasopharyngeal swabs 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. An 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.

**Open Peer Review**

Reviewer Status

Invited Reviewers

1

2

**version 2**

(revision)

23 Sep 2021

**version 1**

08 Jul 2021



report



report

1. **Martin Munene Nyaga** , Sefako  
Makgatho Health Sciences University,  
Pretoria, South Africa

2. **Yewande Nejo** , Bowen University, Iwo,  
Nigeria

Any reports and responses or comments on the article can be found at the end of the article.

## Keywords

human rhinovirus, whole-genome, sequencing, phylogenetics

**Corresponding author:** Martha M. Luka ([mawia.martha@gmail.com](mailto:mawia.martha@gmail.com))

**Author roles:** **Luka MM:** Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kamau E:** Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation; **de Laurent ZR:** Methodology, Resources, Writing – Original Draft Preparation; **Morobe JM:** Methodology, Writing – Original Draft Preparation; **Alii LK:** Supervision; **Nokes DJ:** Conceptualization, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Agoti CN:** Conceptualization, Methodology, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

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

**Grant information:** This work was supported by the Wellcome Trust through a Wellcome Trust Senior Investigator Award to DJN (#102975). MML was supported by the Fogarty International Center (#U2RTW010677) of the National Institutes of Health (NIH) and DELTAS Africa Initiative (#DEL-15-003) of the African Academy of Sciences (AAS). The content is the authors' responsibility and does not necessarily represent the official views of the Wellcome Trust, NIH, nor AAS.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2021 Luka MM *et al.* This is an open access article 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.

**How to cite this article:** Luka MM, Kamau E, de Laurent ZR *et al.* **Whole genome sequencing of two human rhinovirus A types (A101 and A15) detected in Kenya, 2016-2018 [version 2; peer review: 2 approved]** Wellcome Open Research 2021, 6:178 <https://doi.org/10.12688/wellcomeopenres.16911.2>

**First published:** 08 Jul 2021, 6:178 <https://doi.org/10.12688/wellcomeopenres.16911.1>

**REVISED Amendments from Version 1**

Revisions have been made to the manuscript to address reviewers' comments. The changes in the new version are listed below.

Abstract:

We specify the type of samples used in the Abstract section

Methods:

We mention the type of research approaches used in the study.

We restructure the sentences to better explain the rationale behind choosing two rhinovirus types for this study.

Discussion:

We address the reviewers' concerns on (i) possible reasons for failed sequencing of 25 samples, and (ii) implications of not sequencing the extreme sections of the 5' and 3' non-coding regions.

Funding:

We have listed all funding agencies that supported this work, which were missed out in the previous version.

Other minor changes include the addition of suitable references to support the changes above.

**Any further responses from the reviewers can be found at the end of the article**

## Introduction

Genomic surveillance of respiratory viruses is important for (i) developing molecular diagnostics<sup>1,2</sup>, (ii) investigating transmission and evolution<sup>3,4</sup>, and (iii) development of vaccines and therapeutic drugs<sup>5</sup>. Human rhinovirus (HRV) is the most common cause of upper respiratory infections<sup>6,7</sup> and is also occasionally associated with lower respiratory infections<sup>8</sup>. It is a highly diverse virus, with over 160 distinct types identified globally<sup>9</sup>. This diversity presents a challenge in developing a sequencing protocol that works well across the different HRV types<sup>10</sup>. Most HRV molecular epidemiology studies utilize partial genome sequences, which offer lower resolution in identifying epidemiologically linked infections<sup>11</sup>.

Viral genome sequencing can take one of the two approaches available: a targeted/enrichment approach or an agnostic/metagenomics approach<sup>12</sup>. 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<sup>13</sup>. However, this approach requires high viral titers to succeed<sup>14</sup>, which are not always available in a clinical sample<sup>15</sup>. Furthermore, most clinical samples especially those relevant to HRV sequencing are dominated by host and bacterial nucleic acids<sup>13</sup>. Nonetheless, this challenge can be overcome by target enrichment, for example, polymerase chain reaction (PCR) to bulk up for the target virus before sequencing<sup>15-17</sup>. 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. We used a combination of genomic, method-development, descriptive and retrospective approaches to achieve this.

## 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<sup>7</sup> and primary school surveillance between May 2017 and April 2018<sup>6</sup>. 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<sup>8</sup>. 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. The most frequent type observed in the KHDSS health dispensary surveillance was A15 (n=63). Comparison of the HRV diversity within the school (located in Junju) and the Junju outpatient clinic revealed 12 common types, and the most frequent common type was HRV-A101 (n=10, with a frequency of n=5 in each setting)<sup>9,11</sup>.

For this study, we purposively selected two types from the two studies: A15 from the KHDSS surveillance and A101 from both the Junju health dispensary and school for whole genome sequencing (WGS). These two types were selected due to their high frequency of occurrence at the various scales of observation studied.

### 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<sup>18</sup> on 30<sup>th</sup> 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

**Table 1. Type-specific primers for the whole-genome amplification of two human rhinovirus types-A15 and A101.**

Name	Target type	Amplicon	Start	Length	%GC	Tm (°C)	Hairpin Tm (°C)	Self-dimer Tm (°C)	Pair dimer Tm (°C)	Sequence
95 F_a101	A101	1	95	23	39.1	58	None	None	6.9	ACCCCAAATGTAAGCTTAGAAGCA
716 R_a101	A101	1	1,334	22	45.5	60	None	None	None	TCATCAGTGGGTTGTTGTGAGT
726 F_a101	A101	2	726	20	50	60	None	None	None	AGCATCAAGTGGAGCGTCAA
1,215R_a101	A101	2	1,833	22	54.5	62	None	None	None	GACACCCACACGAACTGCATAC
827 F_a101	A101	3	1,445	22	36.4	56	None	None	None	ATGCTGTTCTATGGATTCAAT
2,468R_a101	A101	3	2,468	20	50	60	None	None	None	TCTGGTTGTGTTTGGCTGGT
1,524F_a101	A101	4	2,142	22	40.9	56	None	None	None	TACCACACCTGATACATACTCA
3,516R_a101	A101	4	3,516	20	55	60	None	8.8	3.2	TCCACAATCTCCAGGTGCAC
2,546F_a101	A101	5	3,164	23	39.1	56	None	None	None	TACCTACAAGAACAGACCTTACT
3,901R_a101	A101	5	4,519	22	40.9	55	None	None	None	GTTTCCCTTTGTCTGGTAAATC
4,102F_a101	A101	6	4,102	20	50	60	None	None	None	ACCCAGAAACAGCAGCAAGA
5,248R_a101	A101	6	5,248	23	39.1	58	None	None	None	ACCCTGTGAACCTTCCATTACAT
4,306F_a101	A101	7	4,924	24	37.5	57	None	None	None	AAATCAGTTAGGAATCCAGATGTC
5,905R_a101	A101	7	6,523	24	33.3	56	None	None	None	TAGAATTACACAACCTTCTAACCA
5,550F_a101	A101	8	6,168	21	38.1	55	None	None	None	ACCAATGATCACTTCTCTCAA
6,383R_a101	A101	8	7,001	24	33.3	56	None	None	None	TGGTCATATTTGTCTTTTCCACTA
A15_1F	A15	1	21	20	55	61	None	None	None	ATCCACCTGAACCTCCCAA
A15_1R	A15	1	1,251	20	55	60	None	None	None	CCAGCCGTGACATTACCTYT
534F_a15_22	A15	2	621	21	52.4	60	None	2.9	None	CCATGGGCGCTCAAGTATCTA
1,889R_a15_22	A15	2	1,988	24	33.3	57	None	None	None	CACAAAACATGAACTGAATCGTA
1,391F_a15_22	A15	3	1,478	21	42.9	56	None	None	None	AGACATAACAAGTGGAGCTTG
2,848R_a15_22	A15	3	2,947	23	34.8	56	None	None	None	TCCATCGTATCCATCATAAAACA
2,417F_a15_22	A15	4	2,516	22	40.9	55	None	None	None	TCACAGACTAGAGATGAGATGA
3,464R_a15_22	A15	4	3,563	20	45	55	None	None	None	CTATCACACCATGTTTGCAC
2,900F_a15_22	A15	5	2,999	24	33.3	54	None	4.3	None	CTATGTTCAAGAAATAGTCACTGAA
4,352R_a15_22	A15	5	4,451	23	39.1	58	None	3	None	CACCAGGATTTTGCATAATGTCA
3,576F_a15_22	A15	6	3,675	21	38.1	55	None	None	None	TTGGTGACGGGTTTGTAAATA
4,991R_a15_22	A15	6	5,090	24	33.3	55	None	None	None	CAATATAATGCCTGTACTACTGA
4,385F_a15_22	A15	7	4,484	23	43.5	58	None	None	None	TCAAGTGTAACTTTATCCCTCC
5,943R_a15_22	A15	7	6,042	23	43.5	59	None	None	None	GTTCCAAACACACTATCTCTCAA
A15_8F	A15	8	5,972	20	55	60	None	None	None	ACYCTTGAYATTGRCCACGC
A15_8R	A15	8	7,029	20	55	60	None	None	None	CTCACACTGCGAATCCCCTT
5,560F_a15_22	A15	8	5,659	21	42.9	55	None	None	None	CATTATGTTGGTGGTAATGG
7,076R_a15_22	A15	8	7,076	20	55	60	None	None	None	AAGGCGGGATATACAGTGCG

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.

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.bukxnuxn](https://doi.org/10.17504/protocols.io.bukxnuxn)).

### 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/uL. Sequencing libraries were prepared using the Nextera XT Sample Preparation Kit (Illumina, CA) and sequencing performed on Illumina MiSeq platform (200bp×2) per sample.

### Sequence assembly

The raw reads were quality checked using FastQC v0.11.9 and trimmed (Phred score >30) using Trimmomatic v0.39<sup>19</sup>. HRV reads were identified by mapping to the respective reference strains ([https://www.picornaviridae.com/sg3/enterovirus/rv-a/rv-a\\_seqs.htm](https://www.picornaviridae.com/sg3/enterovirus/rv-a/rv-a_seqs.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](http://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<sup>20</sup> package.

### Sequence analysis

Sequences were aligned using MAFFT v7.271<sup>21</sup>. Recombination scans were done using RDP5<sup>22</sup> and visualized on SimPlot<sup>23</sup>. Nucleotide substitutions across the genomes were visualized using a python script to examine genetic diversity across the genome. POPART<sup>24</sup> 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<sup>25</sup>. Branch support for phylogenetic trees was assessed using bootstrapping of 1000 iterations. MegaX<sup>26</sup> 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<sup>27</sup>. BEAST was run with 200 million MCMC steps using the best fitting substitution model and a coalescent-based relaxed clock framework<sup>28</sup>. 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<sup>29</sup> 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 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 interspersions 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).

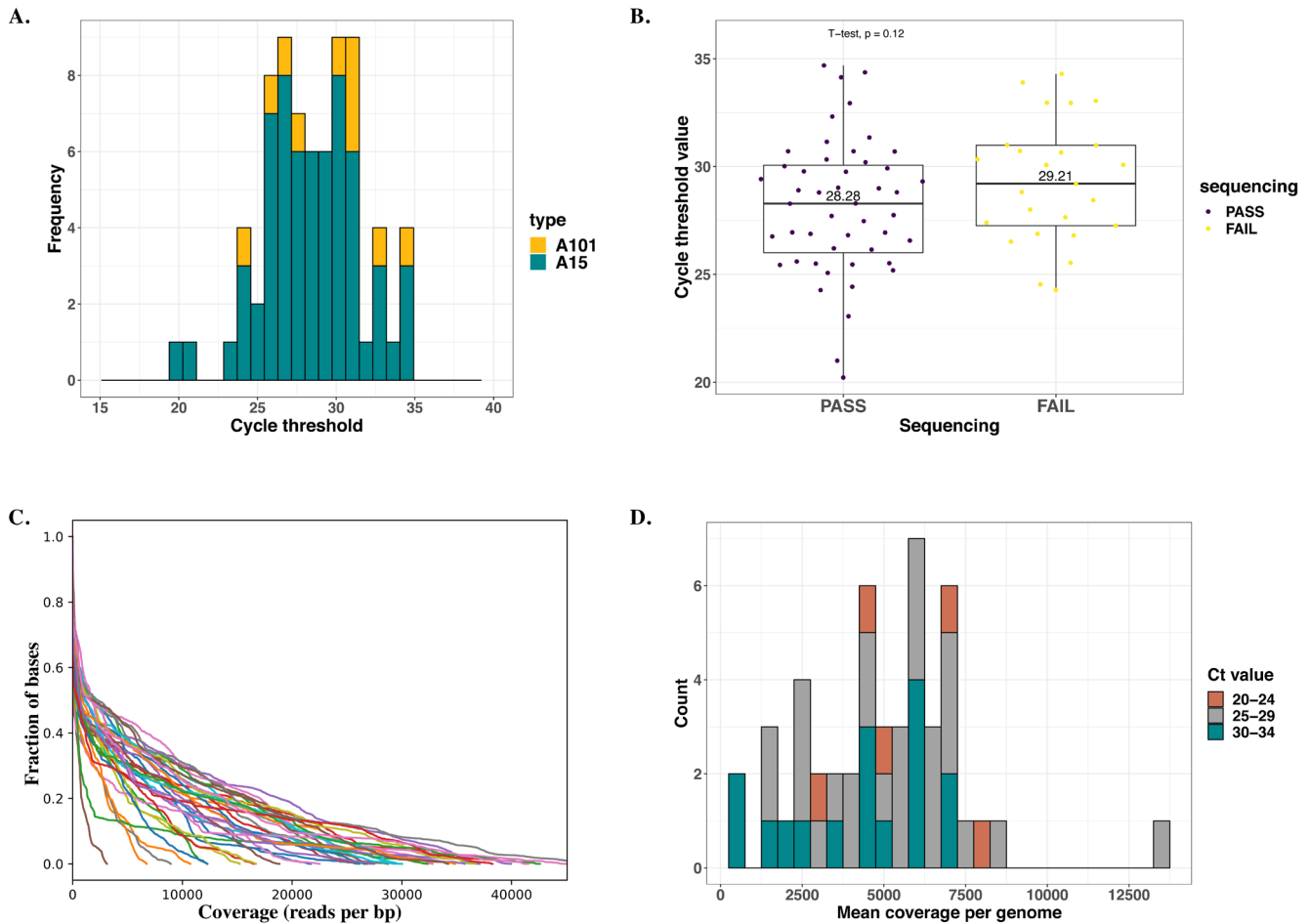
The ends of the 5' untranslated region (UTR) and 3' UTR were not amplified due to lack of suitable primers. Genetic diversity was observed across the entire genome and not within a particular genomic region for both types, as shown in [Figure 3](#).

### 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 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



**Figure 1. Summary statistics of sequenced samples.** (A) Distribution of cycle threshold (Ct) values across all samples selected for sequencing. Bars are colored by HRV type. (B) Dispersion of Ct-values across samples successfully sequenced and those that failed. (C). Read depth (per base pair) distribution per (successfully) sequenced sample. Each line represents a genome/sample. (D). Distribution of mean coverage per base pair per genome across successfully sequenced samples. The bars are colored by Ct-value group.

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](#). Recombination within HRV structural regions has been shown to be rare and sporadic<sup>30</sup>.

#### Discussion

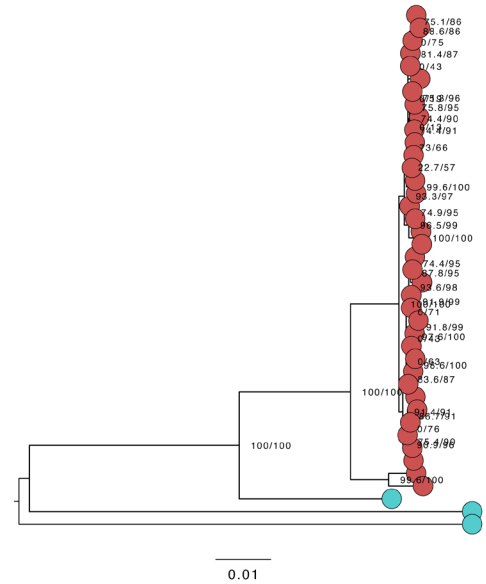
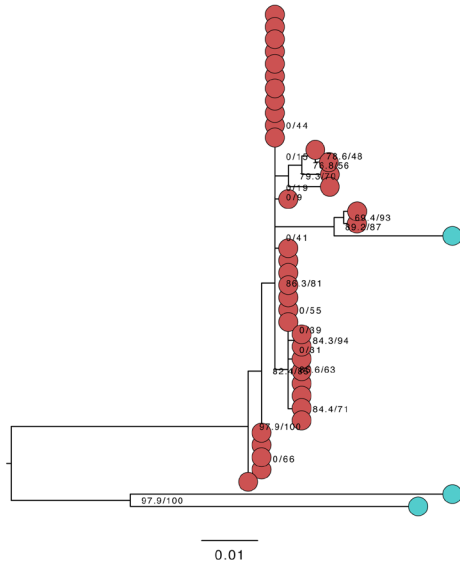
This study presents a type-specific whole genome sequencing protocol for two HRV types. A101 had a higher success (100%) than A15 (60.3%). We attribute the higher A101 sequencing success to the higher number of sequences ( $n=9$ ) that were available for primer design, which captured more intra-type 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. While it's not clear what the cause of sequencing failure of the 25 samples was, we speculate that either (i) their genomic diversity was not captured in primer design resulting in primer



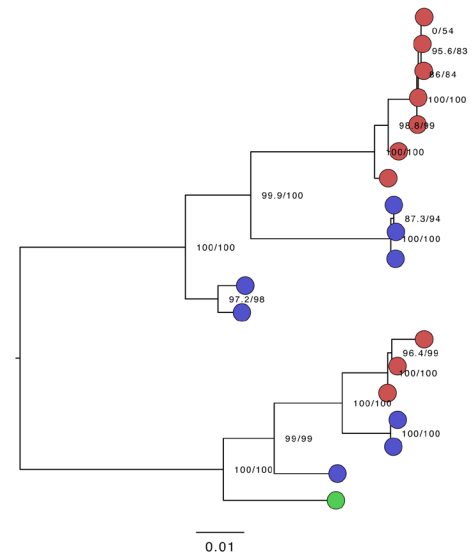
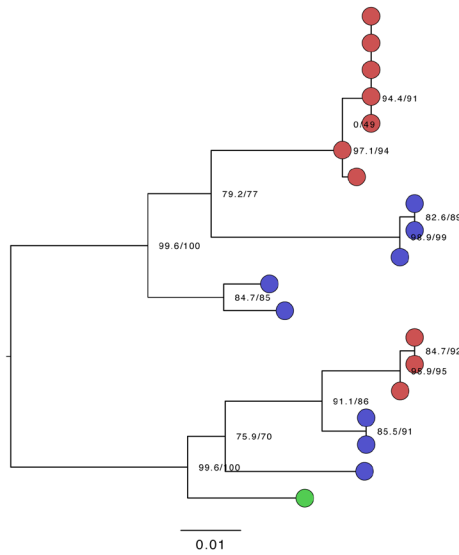
VP4/2 sequences

Whole-genome sequences

A. HRV-A15



B. HRV-A101

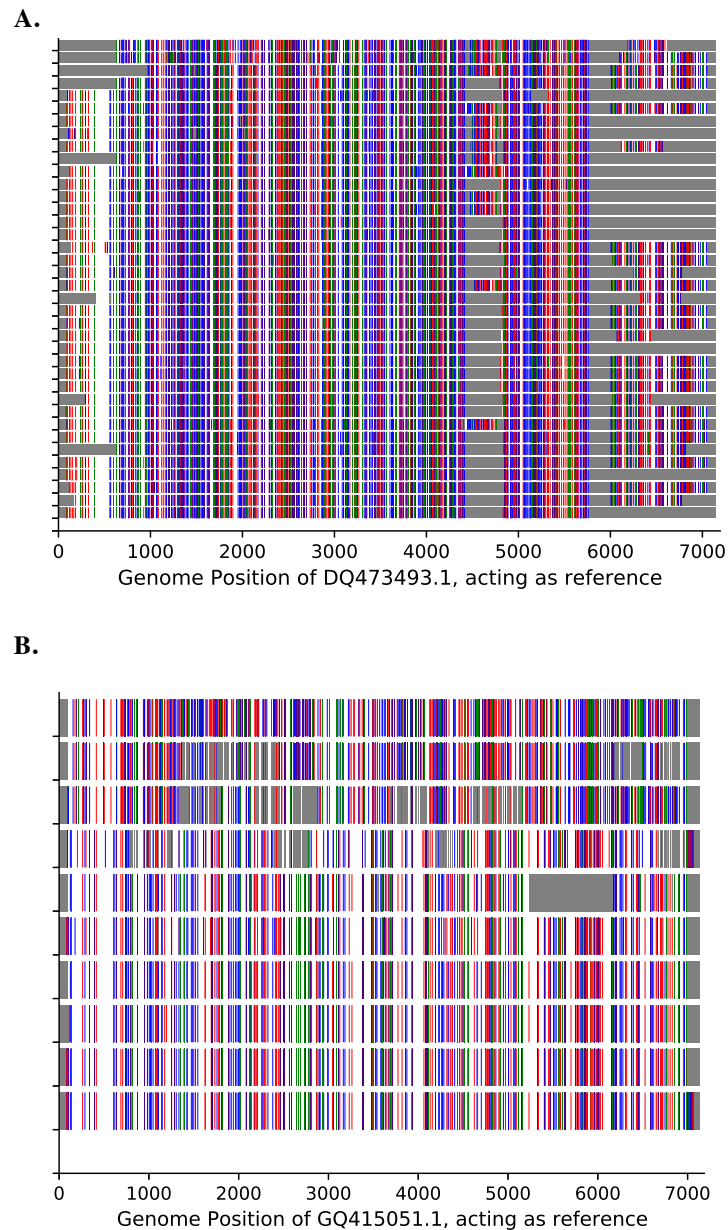


**Figure 2.** Maximum-likelihood phylogenetic trees of local (generated) and global (A) A15 and (B) A101 sequences. The tips are coloured by origin, indicating the global sequences used in primer design. The scale bar represents nucleotide substitutions per site.

mismatches, (ii) the sample quality had deteriorated over time or (iii) there was presence of PCR inhibitors/ nuclease enzymes in the sample.

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<sup>31</sup>. Due to the



**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\\_seqs.htm](https://www.picornaviridae.com/sg3/enterovirus/rv-a/rv-a_seqs.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.

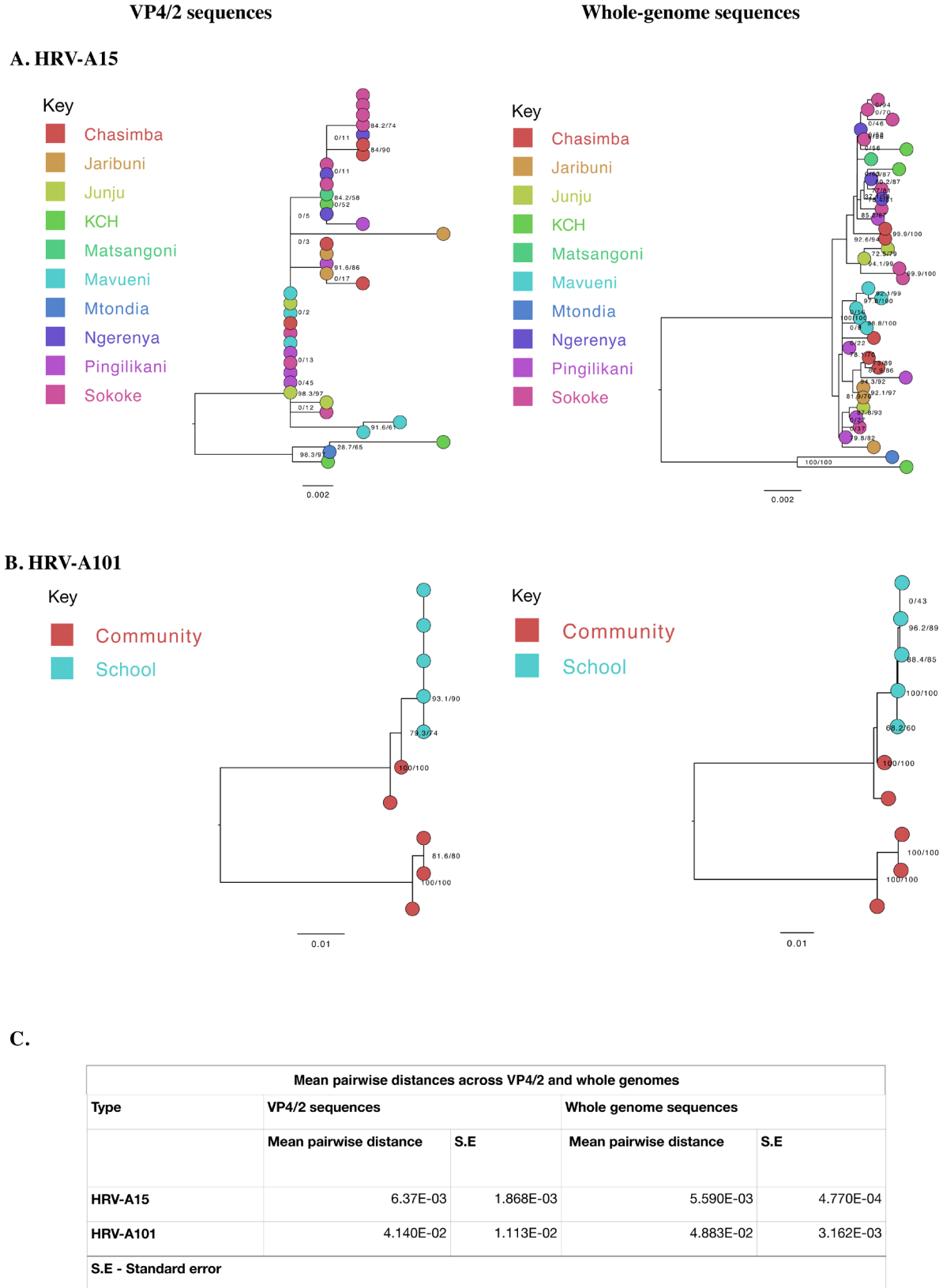
short size of VP4/2 (~420nt), insufficient data results in reduced phylogenetic resolution and increased uncertainty evidenced by a higher standard error in phylogenetic distance. Unresolved phylogenies are a challenge in epidemiology as one cannot distinguish infections from one individual to another for transmission inference.

Posterior probabilities summarize the uncertainties about a parameter and indicate confidence in the evidence<sup>32</sup>. High

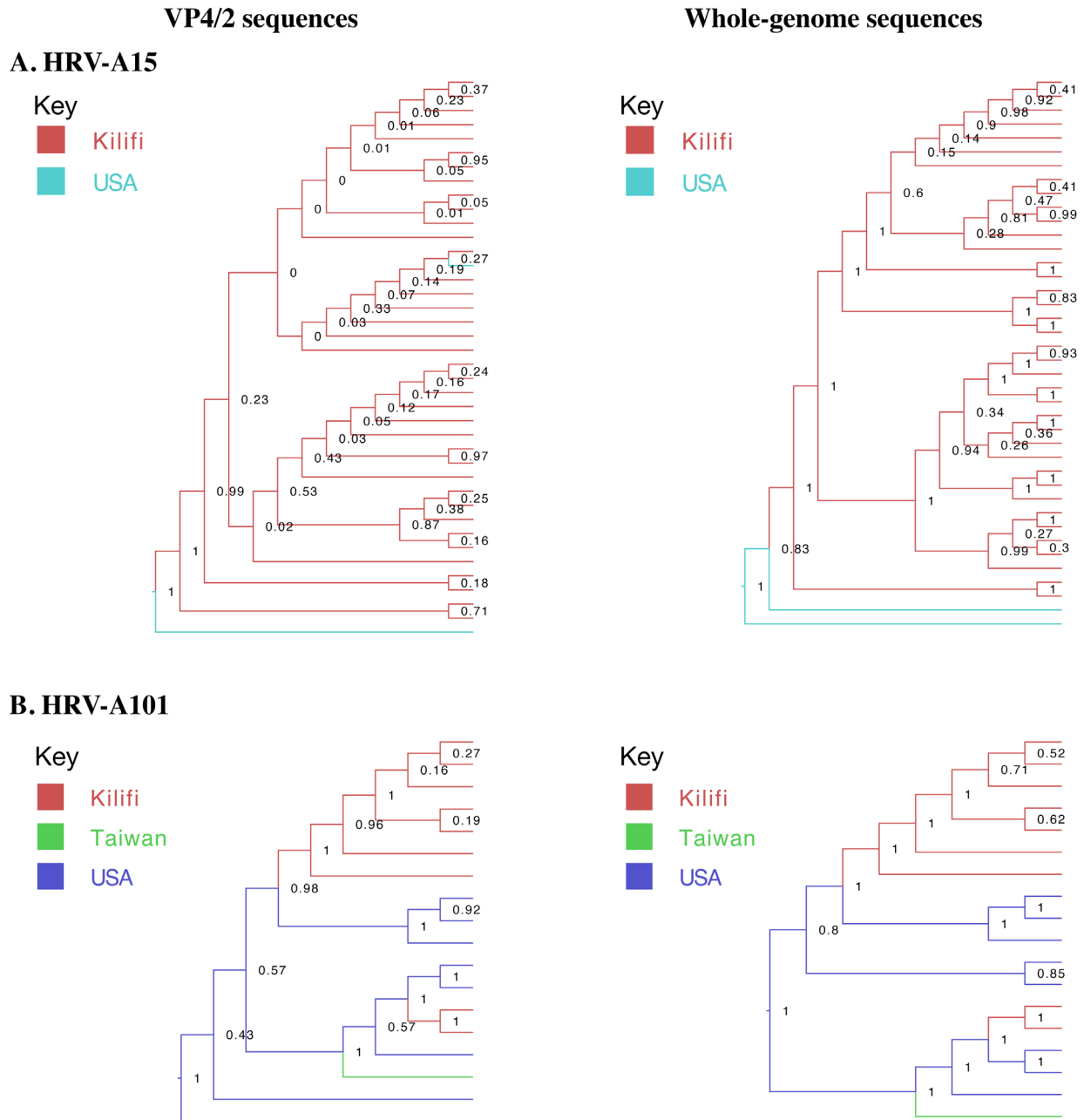
posterior probabilities indicate high confidence, and the reverse is also true. Whole genomes consistently provided higher confidence across the two genotypes assessed in this study.

With pathogen sequencing now an established tool to track viral infections<sup>2,12</sup>, 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





**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.

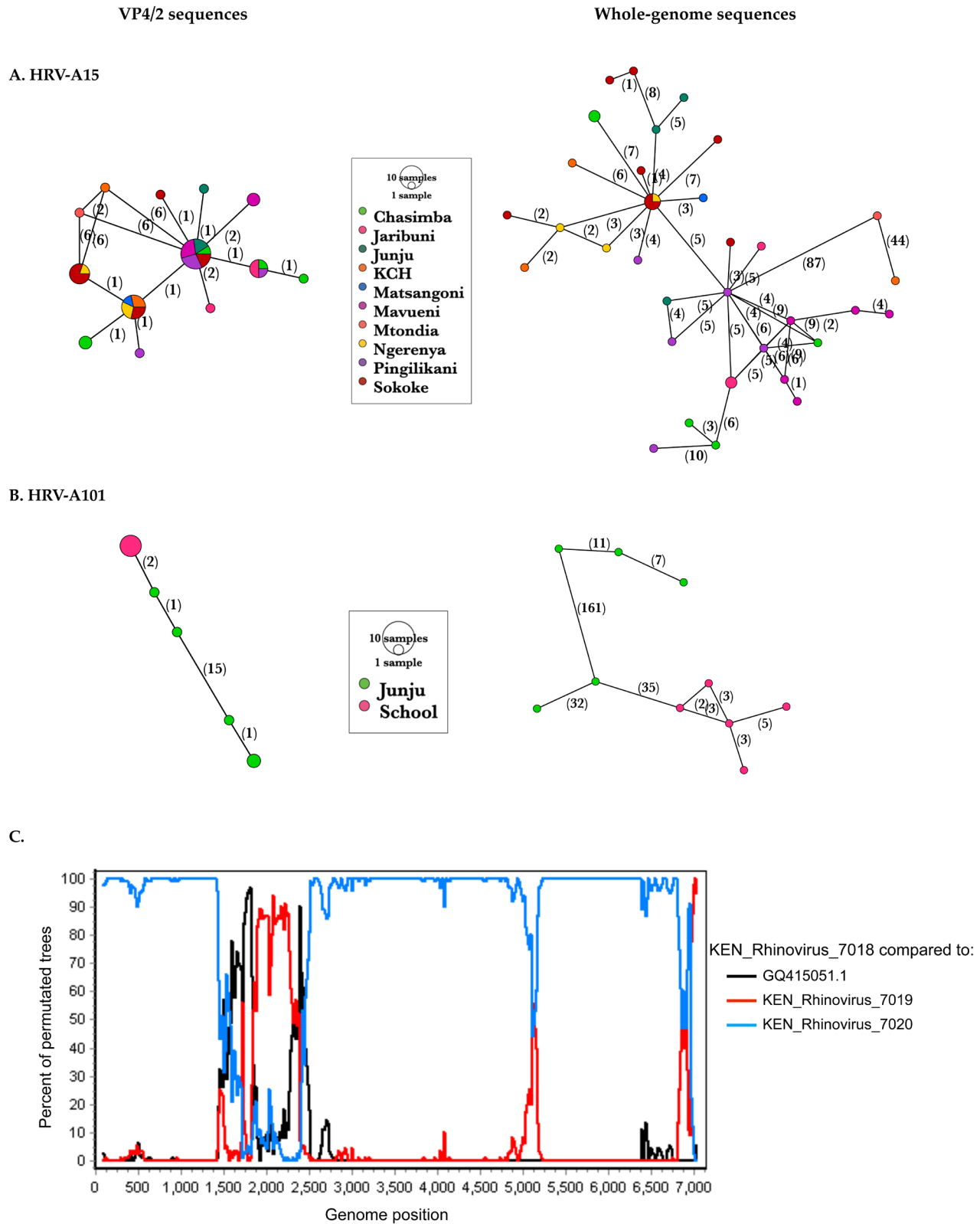


**Figure 5.** Bayesian phylogenetic trees of local and global (A) HRV-A15 and (B) HRV-A101 VP4/2 and whole genomes. The branches are coloured by site of origin. Node labels indicate branching posterior probabilities.

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. Although the 5'UTR alone does not offer adequate resolution to confidently distinguish HRV types<sup>33</sup>, it is speculated to be a hotspot for recombination<sup>30,33</sup>. Not sequencing these extreme regions may therefore result in missing out on important evolutionary/phylogenetic signal. Notwithstanding,



**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.

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>

Accession number URL: <https://identifiers.org/bioproject:PRJNA701406>

Harvard Dataverse. Replication Data for: Whole genome sequencing of two human rhinovirus A types (A101 and A15) detected in Kenya, 2016–2018. DOI: <https://doi.org/10.7910/DVN/QGXZLI><sup>34</sup>

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 Cycle threshold (Ct) values, and read/sequencing depth.

Data are available under the terms of the [Creative Commons Attribution 4.0 International license \(CC-BY 4.0\)](https://creativecommons.org/licenses/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

- Agoti CN, Kiyuka PK, Kamau E, *et al.*: **Human Rhinovirus B and C Genomes from Rural Coastal Kenya.** *Genome Announc.* 2016; **4**(4): e00751–16. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Meredith LW, Hamilton WL, Warne B, *et al.*: **Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study.** *Lancet Infect Dis.* 2020; **20**(11): 1263–1271. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Otieno JR, Kamau EM, Oketch JW, *et al.*: **Whole genome analysis of local Kenyan and global sequences unravels the epidemiological and molecular evolutionary dynamics of RSV genotype ON1 strains.** *Virus Evol.* 2018; **4**(2): vey027. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Agoti CN, Phan MVT, Munywoki PK, *et al.*: **Genomic analysis of respiratory syncytial virus infections in households and utility in inferring who infects the infant.** *Sci Rep.* 2019; **9**(1): 10076. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Thanh Le T, Andreadakis Z, Kumar A, *et al.*: **The COVID-19 vaccine development landscape.** *Nat Rev Drug Discov.* 2020; **19**(5): 305–306. [PubMed Abstract](#) | [Publisher Full Text](#)
- Adema IW, Kamau E, Nyiro JU, *et al.*: **Surveillance of respiratory viruses among children attending a primary school in rural coastal Kenya [version 2; peer review: 2 approved].** *Wellcome Open Res.* 2020; **5**: 63. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nyiro JU, Munywoki P, Kamau E, *et al.*: **Surveillance of respiratory viruses in the outpatient setting in rural coastal Kenya: baseline epidemiological observations [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2018; **3**: 89. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Onyango CO, Welch SR, Munywoki PK, *et al.*: **Molecular epidemiology of human rhinovirus infections in Kilifi, coastal Kenya.** *J Med Virol.* 2012; **84**(5): 823–831. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Morobe JM, Nyiro JU, Brand S, *et al.*: **Human rhinovirus spatial-temporal epidemiology in rural coastal Kenya, 2015–2016, observed through outpatient surveillance [version 2; peer review: 2 approved].** *Wellcome Open Res.* 2019; **3**: 128. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tapparel C, Junier T, Gerlach D, *et al.*: **New complete genome sequences of human rhinoviruses shed light on their phylogeny and genomic features.** *BMC Genomics.* 2007; **8**: 224. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Luka MM, Kamau E, Adema I, *et al.*: **Molecular epidemiology of human rhinovirus from one-year surveillance within a school setting in rural coastal Kenya.** *medRxiv.* 2020; 2020.03.09.20033019. [Publisher Full Text](#)
- Houldcroft CJ, Beale MA, Breuer J: **Clinical and biological insights from viral genome sequencing.** *Nat Rev Microbiol.* 2017; **15**(3): 183–192. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gu W, Miller S, Chiu CY: **Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection.** *Annu Rev Pathol.* 2019; **14**: 319–338. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Greninger AL, Naccache SN, Federman S, *et al.*: **Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis.** *Genome Med.* 2015; **7**: 99. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mertes F, Elsharawy A, Sauer S, *et al.*: **Targeted enrichment of genomic DNA regions for next-generation sequencing.** *Brief Funct Genomics.* 2011; **10**(6): 374–386. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kamaraj US, Tan JH, Xin Mei O, *et al.*: **Application of a targeted-enrichment methodology for full-genome sequencing of Dengue 1–4, Chikungunya and Zika viruses directly from patient samples.** *PLoS Negl Trop Dis.* 2019; **13**(4): e0007184. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hasan MR, Rawat A, Tang P, *et al.*: **Depletion of Human DNA in Spiked Clinical Specimens for Improvement of Sensitivity of Pathogen Detection by Next-Generation Sequencing.** *J Clin Microbiol.* 2016; **54**(4): 919–927. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sayers EW, Agarwala R, Bolton EE, *et al.*: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res.* 2019; **47**(D1): D23–D28. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics.* 2014; **30**(15): 2114–2120. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ramirez F, Ryan DP, Grüning B, *et al.*: **deepTools2: a next generation web server**

- for deep-sequencing data analysis. *Nucleic Acids Res.* 2016; **44**(W1): W160–5.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Yamada KD, Tomii K, Katoh, K: **Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees.** *Bioinformatics.* 2016; **32**(21): 3246–3251.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  22. Martin DP, Varsani A, Roumagnac P, *et al.*: **RDP5: A computer program for analysing recombination in and removing signals of recombination from, nucleotide sequence datasets.** *Virus Evol.* 2020; **7**(1): veaa087.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  23. Lole KS, Bollinger RC, Paranjape RS, *et al.*: **Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination.** *J Virol.* 1999; **73**(1): 152–160.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  24. Leigh JW, Bryant D: **POPART: Full-feature software for haplotype network construction.** *Methods Ecol Evol.* 2015; **6**(9): 1110–1116.  
[Publisher Full Text](#)
  25. Nguyen LT, Schmidt HA, von Haeseler A, *et al.*: **IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.** *Mol Biol Evol.* 2015; **32**(1): 268–274.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  26. Kumar S, Stecher G, Li M, *et al.*: **MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.** *Mol Biol Evol.* 2018; **35**(6): 1547–1549.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  27. Drummond AJ, Suchard MA, Xie D, *et al.*: **Bayesian phylogenetics with BEAUti and the BEAST 1.7.** *Mol Biol Evol.* 2012; **29**(8): 1969–1973.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  28. Drummond AJ, Ho SYW, Phillips MJ, *et al.*: **Relaxed phylogenetics and dating with confidence.** *PLoS Biol.* 2006; **4**(5): e88.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  29. Rambaut A, Drummond AJ: **FigTree version 1.4. 0.** 2012.
  30. McIntyre CL, Savolainen-Kopra C, Hovi T, *et al.*: **Recombination in the evolution of human rhinovirus genomes.** *Arch Virol.* 2013; **158**(7): 1497–1515.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  31. Swenson NG: **Phylogenetic Resolution and Quantifying the Phylogenetic Diversity and Dispersion of Communities.** *PLoS One.* 2009; **4**(2): e4390.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  32. Llewelyn H: **Replacing P-values with frequentist posterior probabilities of replication-When possible parameter values must have uniform marginal prior probabilities.** *PLoS One.* 2019; **14**(2): e0212302.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  33. Savolainen-Kopra C, Blomqvist S, Smura T, *et al.*: **5' noncoding region alone does not unequivocally determine genetic type of human rhinovirus strains.** *J Clin Microbiol.* United States; 2009; **47**(4): 1278–80.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  34. Luka MM, Kamau E, de Laurent ZR, *et al.*: **Replication Data for: Whole genome sequencing of two human rhinovirus A types (A101 and A15) detected in Kenya, 2016-2018.** 2021.  
<http://www.doi.org/10.7910/DVN/QGXZLI>

# Open Peer Review

Current Peer Review Status:  

---

## Version 1

Reviewer Report 31 August 2021

<https://doi.org/10.21956/wellcomeopenres.18656.r45239>

© 2021 Nejo Y. 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.



**Yewande Nejo** 

Microbiology Programme, College of Agriculture, Engineering and Science, Bowen University, Iwo, Nigeria

### Overall comments:

This is an important and well executed study by Luka *et al.* (2021). It provides detailed informative data on the type-specific whole genome sequencing method designed to detect human rhinovirus. The study is able to present target enrichment sequencing approach which addresses the hurdles in genome amplification and sequencing of human rhinovirus posed by the high genetic diversity of the virus and inadequate information of reference sequences. The study is properly designed; however, it will require few additions.

### Abstract:

The abstract is a good summary of the research article but please indicate in the abstract the type of samples that were collected.

### Introduction:

The introduction provides information that is significant to the research and is well referenced. The importance and purpose of the study is well explained and justifiable.

### Methodology:

The methodology segment is well designed and all analyses were carried out with details and sound scientific merit. The author should however specify in this section the type of research carried out.

### Results:

The results are well stated and illustrated in a comprehensive manner. The data is meticulously analyzed

### Discussion and Conclusion:

The results of this study are well discussed and the data supports the conclusion. The limitations

of the study are well stated. However, the reasons for selecting only two types of HRV for whole genome sequencing out of 12 types were not properly indicated. Can the author also explain what could be responsible for the sequencing failure of 25 samples? What could be the implications of not sequencing 5'UTR and 3' UTR regions in the study of viral dynamics?

**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?**

Yes

**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

**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.**

Author Response 21 Sep 2021

**Martha Luka**, KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi, Kenya

Thank you for taking the time to review our article. Please see below our responses to the issues raised.

**Abstract**

**Please indicate in the abstract the type of samples that were collected.**

Nasopharyngeal swabs. (This has now been specified in the abstract)

**Methodology**

**The author should specify in this section the type of research carried out.**

We used a combination of genomic, method-development, descriptive and retrospective approaches to develop a new laboratory protocol for the whole-genome sequencing of human rhinovirus and compare the phylogenetic inferences between partial and whole



genome sequences

### **Discussion and Conclusion**

#### **The reasons for selecting only two types of HRV for whole genome sequencing out of 12 types were not properly indicated.**

The current study was nested within a larger epidemiological study to study pathways of respiratory virus disease within different scales which included: (i) outpatient surveillance of nine health dispensaries within the Kilifi Health and Demographic Surveillance System (KHDSS)[1] and (ii) a primary school surveillance[2]. The most frequent type observed in the KHDSS health dispensary surveillance was A15 (n=63). The second social scale, the school, was situated in Junju, a location within the KHDSS. Comparison of the HRV diversity within the school and the Junju outpatient clinic revealed 12 common types[3], and the most frequent common type was A101 (n=5 in each setting). These two types were identified as of interest due to their high frequency of occurrence at the various scales of observation studied.

#### **Can the author also explain what could be responsible for the sequencing failure of 25 samples?**

While it's not clear what the cause of sequencing failure of the 25 samples was, we speculate that either (i) their genomic diversity was not captured in primer design resulting in primer mismatches, (ii) the sample quality had deteriorated over time or (iii) there was presence of PCR inhibitors/ nuclease enzymes in the sample.

#### **What could be the implications of not sequencing 5'UTR and 3' UTR regions in the study of viral dynamics?**

Analysis of the 5'UTR across HRV shows high conservation of certain regions[4], making it a suitable target for HRV detection assays[5]. Contrastingly, the ~50 bp long 3'UTR is not universally conserved across HRV species[4]. Although the 5'UTR alone has not been proved to offer adequate resolution to confidently distinguish HRV types[6], it is speculated to be a hotspot for recombination[7]. Not sequencing these extreme regions may therefore result in missing out on important evolutionary/phylogenetic signal.

### **References**

1. Nyiro JU, Munywoki P, Kamau E, et al. Surveillance of respiratory viruses in the outpatient setting in rural coastal Kenya: baseline epidemiological observations. Wellcome open Res. 2018;3:89.
2. Adema IW, Kamau E, Uchi Nyiro J, Otieno GP, Lewa C, Munywoki PK, Nokes DJ. Surveillance of respiratory viruses among children attending a primary school in rural coastal Kenya [version 1; peer review: awaiting peer review]. Wellcome Open Res [Internet]. 2020;5(63). Available from: <https://wellcomeopenresearch.org/articles/5-63/v1>
3. Luka MM, Kamau E, Adema I, et al. Molecular epidemiology of human rhinovirus from one-year surveillance within a school setting in rural coastal Kenya. medRxiv [Internet]. 2020 Jan 1;2020.03.09.20033019. Available from: <http://medrxiv.org/content/early/2020/03/19/2020.03.09.20033019.abstract>
4. Kloc A, Rai DK, Rieder E. The Roles of Picornavirus Untranslated Regions in Infection and Innate Immunity. Front Microbiol. 2018;9:485.

5. Hammitt LL, Kazungu S, Welch S, et al. Added Value of an Oropharyngeal Swab in Detection of Viruses in Children Hospitalized with Lower Respiratory Tract Infection. *J Clin Microbiol* [Internet]. 2011 Jun 1;49(6):2318 LP – 2320. Available from: <http://jcm.asm.org/content/49/6/2318.abstract>
6. Savolainen-Kopra C, Blomqvist S, Smura T, et al. 5' noncoding region alone does not unequivocally determine genetic type of human rhinovirus strains. Vol. 47, *Journal of clinical microbiology*. United States; 2009. p. 1278–80.
7. McIntyre CL, Savolainen-Kopra C, Hovi T, Simmonds P. Recombination in the evolution of human rhinovirus genomes. *Arch Virol*. 2013;158(7):1497–515.

**Competing Interests:** No competing interests.

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.



**Martin Munene Nyaga** 

South African Medical Research Council/Diarrhoeal 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 (enteric 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.**

Author Response 20 Sep 2021

**Martha Luka**, KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi, Kenya

Thank you for taking the time to review our manuscript. Please find below our response (in *italics*) to the issue raised .

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?

*While it's not clear what the cause of sequencing failure of the 25 samples was, we speculate that either (i) their genomic diversity was not captured in primer design resulting in primer mismatches, (ii) the sample quality had deteriorated over time or (iii) there was presence of PCR inhibitors/ nuclease enzymes in the sample.*

**Competing Interests:** No competing interests

