Molecular epidemiology of human rhinovirus from one-year surveillance within a school setting in rural coastal Kenya

Martha M. Luka[1,2,#]; Everlyn Kamau[1]; Irene Adema[1]; Patrick K. Munywoki[1]; Grieven P. Otieno[1]; Elijah Gicheru[1]; Alex Gichuki[1]; Nelson Kibinge[1]; Charles N. Agoti[1,2]; D. James Nokes[1,3,##]

1. Epidemiology and Demography Department, KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medicine Research – Coast, Kilifi, Kenya.
2. Department of Public Health, Pwani University, Kilifi, Kenya.
3. School of Life Sciences and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry, United Kingdom.

[*] **Corresponding author:** mluka@kemri-wellcome.org

[##]**Alternate corresponding author**: jnokes@kemri-wellcome.org

Summary points: We describe the molecular epidemiology of human rhinovirus (HRV) within a school setting over one-year in rural coastal Kenya. A high diversity of HRV infections was observed across all classes with evidence of introduction and transmission of 47 different genotypes.

ABSTRACT

Background: Human rhinovirus (HRV) is the most common cause of the common cold but may also lead to more severe respiratory illness in vulnerable populations. The epidemiology and genetic diversity of HRV within a school setting have not been previously described.

Objective: To characterise HRV molecular epidemiology in primary school in a rural location of Kenya.

Methods: Between May 2017 to April 2018, over three school terms, we collected 1859 nasopharyngeal swabs (NPS) from pupils and teachers with symptoms of acute respiratory infection in a public primary school in Kilifi County, coastal Kenya. The samples were tested for HRV using real-time RT-PCR. HRV positive samples were sequenced in the VP4/VP2 coding region for species and genotype classification.

Results: A total of 307 NPS (16.4%) from 164 individuals were HRV positive, and 253 (82.4%) were successfully sequenced. The proportion of HRV in the lower primary classes was higher (19.8%) than upper primary classes (12.2%), *p*-value <0.001. HRV-A was the most common species (134/253, 53.0%), followed by HRV-C (73/253, 28.9%) and HRV-B (46/253, 18.2%). Phylogenetic analysis identified 47 HRV genotypes. The most common genotypes were A2 and B70. Numerous (up to 22 in one school term) genotypes circulated simultaneously, there was no individual re-infection with the same genotype, and no genotype was detected in all three school terms.

Conclusion: HRV was frequently detected among school-going children with mild ARI symptoms, and particularly in the younger age groups (<5-year-olds). Multiple HRV introductions were observed characterised by the considerable genotype diversity.

Keywords: Human rhinovirus, transmission, molecular epidemiology, school-going children, Kenya

INTRODUCTION

Human rhinovirus (HRV) is a common viral respiratory pathogen[1] associated with the common cold [2,3], lower respiratory tract infections [4] and asthma [5]. Although the majority of HRV cases are mild and self-limiting[6], they contribute to substantial economic losses through missed school and workdays [6,7]. HRV is a common reason for prescribing antibiotics [8], potentially contributing to antibiotic resistance. The virus is transmitted via inhalation of contaminated aerosols, e.g. during close contact with infected persons or self-inoculation via touching contaminated surfaces or objects[6,9]. HRV has a median incubation period of two days (range; 1-7 days) and a symptom duration of 7-14 days[10]. Viral shedding can occur up to two weeks (range;1-3 weeks)[11]. Children, the elderly and persons with pre-existing respiratory conditions bear the highest HRV burden[12–14].

HRV is a positive sense, single-stranded RNA virus, classified under the genus *Enterovirus* (family *Picornaviridae)* with a genome approximately 7.2 kb long. It is characterised by high genetic and antigenic diversity[15], frustrating vaccine development efforts. There are 169 HRV genotypes distributed across three species: HRV-A (80), HRV-B (32) and HRV-C (57) genotypes [16]. Molecular classification of HRV into genotypes is based on the nucleotide sequence of either the 5' noncoding region, VP1 or the VP4/2 genome region [17–19].

Children constitute a significant susceptible population that support the transmission and persistence of HRV in populations[20]. In Africa, little has been done to investigate the patterns and mechanisms of transmission of HRV in school setting and to understand the extent to which school settings contribute to HRV transmission in the community. This is despite that school children have the high contact rates compared to other age-groups in the community [2122,23]. Design of effective intervention strategies against HRV will be supported by improved knowledge of transmission dynamics of HRV in different social networks and population structures [6,24]. This study investigated

HRV infections in a school setting in rural coastal Kenya by sequence analysis of the VP4/2 junction to describe the frequency, diversity and temporal occurrence of HRV.

METHODS

Study area and design. The study was conducted in a rural school located within the Kilifi Health and Demographic Surveillance System (KHDSS) in Kenya [25] to characterise the occurrence of respiratory viruses. The study design is described in detail elsewhere [26]. Briefly, the school offers both early childhood development education and primary school education. Pupils and teachers from all classes were enrolled in the study, which took place between May 2017 and April 2018. Pupils were divided into two main groups: the lower primary comprising of day care, kindergarten (KG) levels 1 to 3 and grade one (*n*=5 classes, age range 3-12 years); and the upper primary comprising grades two to eight (*n*=7 classes, age range 7-20 years).

Nasopharyngeal swabs (NPS) were collected a pupil or a teacher had at least one of the following acute respiratory illness (ARI) symptoms: cough, sore throat or runny nose. Students documented the ARI symptoms they experienced in a daily diary. Class teachers recorded symptoms for the lower primary group. A maximum of 8 samples per class was collected from the lower primary group per week, while a maximum of 4 samples per class was collected from the upper primary group per week. A maximum of 3 samples was collected from the teachers per week. We collected more samples from the lower primary due to the perceived critical role of this age-group in childhood infectious diseases and hence the need to reduce the level of uncertainty in the estimated risk in this age-groups. Samples were collected in viral transport media (VTM) and transported in cool boxes to the KEMRI-Wellcome Trust Research Programme laboratory where they were stored at -80°C prior to screening. Sampling was suspended during school holidays: August, November and December 2017 and from 6th April 2018, which marked the end of the study.

Patient consent statement. This article reports on samples collected from two studies: a school[26] and a community ARI surveillance study [27]. An informed written parental consent for persons under the age of eighteen years or individual consent for adults was obtained prior to sample collection. For the school cohort, consent was obtained at the beginning of the study, with new students and those not initially enrolled allowed joining in the second and third terms. In addition, children whose parents consented were asked for individual assent to participate. Ethical approval was provided by the KEMRI-Scientific Ethics Review Unit (KEMRI-SERU #3332 and #3103) and the University of Warwick Biomedical and Scientific Research Ethics Committee (BSREC #REGO_2016-1858 and #REGO-2015-6102).

RNA extraction and rRT-PCR. RNA was extracted from 140 µl of the collected samples using QIAamp® 96 Virus QIAcube HT kit (Qiagen, United Kingdom), according to the manufacturer's instructions. Samples were screened for 15 virus targets[26] - HRV, respiratory syncytial virus (A and B), human coronaviruses (OC43, NL63 and E229), influenza (A, B, and C), parainfluenza (1-4), adenovirus and human metapneumovirus - using an in-house multiplexed real-time reverse-transcription PCR (rRT-PCR) with a QuantiFast® Multiplex RT-PCR kit (Qiagen, United Kingdom) [28–30]. A sample was considered HRV positive if the rRT-PCR cycle threshold (*Ct*) value was <35.

VP4/2 amplification and sequencing. VP4/2 sequencing was used to assign species and genotypes. A genomic region ~549 nucleotides and consisting of a hypervariable region of the 5'UTR, the complete VP4 and partial VP2 gene region was amplified for all HRV positive samples using a One-Step RT-PCR kit (QIAGEN®) as previously described [31,32]. PCR products were purified using the MinElute® PCR purification kit (Qiagen, United Kingdom) and sequenced with the respective forward and reverse primers in a BigDye® terminator version 3.1 (Applied Biosystems, USA) reaction and analysed in an ABI 3130xl genetic analyser.

Sequence analysis. Raw sequence reads were quality-checked, trimmed, edited and assembled to contigs of length 420 nucleotides using Sequencher version 5.4.6 (www.genecodes.com). Alignments were prepared using MAFFT version 7.271 [33]. Full exploratory recombination scans were done using the RDP4 software [34]. IQ-TREE version 1.6.0 [35] was used to estimate the best fit model and infer Maximum Likelihood (ML) trees. Phylogenetic trees were generated with bootstrapping of 1000 iterations. Pairwise nucleotide p-distances were calculated using MEGA version 7.0.21 [36]. Genotype assignment was based on the phylogenetic clustering (bootstrap value >80%) on ML trees and pairwise genetic distances to prototype strains (https://www.picornaviridae.com/enterovirus/prototypes/prototypes.htm) as proposed (10.5% for HRV-A, 9.5% for HRV-B, and 10.5% for HRV-C) [17,18].

Intra-type diversity for genotypes with at least ten sequences was studied by visualisation of the number of nucleotide substitutions to the index sequence for each type. The substitution rate of the VP4/2 coding region in HRV had previously been estimated as 7×10^{-4} - 4×10^{-3} substitutions/site/year [37]. Using the upper evolutionary rate value translated to 1.68 nucleotide substitutions per year across the sequenced 420 nt segment. We, therefore, defined an intra-type variant as a sequence with >2 nucleotide differences from the index sequence. An intra-type variant had to be observed in at least two sequences to increase confidence that this was not the result of sequencing error.

Data Analysis. Data analysis was conducted using STATA version 13 (STATA Corp Texas) and R version 3.6.1 (CRAN R Project). Categorical variables were summarised using frequencies and percentages. HRV proportion for each class and respective 95% confidence intervals were defined as the number of HRV positive samples out of the total number of samples tested per class. The chi-square test for trend was used to check for linear trend in HRV proportion with an increasing hierarchy of classes in the school (from day care to teaching staff). Spearman's rank correlation coefficient was run to determine correlation between age and Ct-value.

Definition of terms. We defined 'persistence' as the continued occurrence of the same genotype within the same school term. Detection of a genotype in a subsequent school term was considered a genotype recurrence. We defined 'frequent' genotypes as those that occurred in at least five samples, from more than two individuals, and further investigated their temporal occurrence and persistence. We defined 'individual HRV re-infection' as the acquisition of a new genotype or the detection of a previously acquired genotype in a subsequent school term. Individual detection with the same genotype in consecutive samples was considered a continuing infection.

Sequence data availability. Sequences generated by this study are available in GenBank under accession numbers MT177659-MT177911.

RESULTS

Baseline characteristics and HRV detection

A cohort of 371 individuals (358 pupils and 13 teachers) was followed up for the development of ARI symptoms between May 2017 and April 2018. The total number of samples collected was 1859, of which 307 (16.5%) tested positive for HRV. Twenty-six of the HRV-positive samples had co-infections with other viruses. Of the 1552 HRV-negative samples, only 115 tested positive for at least one virus target, **Supplementary Table 1**. The HRV positives were collectively from 164 individuals: 160 pupils and four teachers. The mean age of the HRV positive pupils was 9.4 years, with a standard deviation of $\pm$3.9. The most common ARI symptom recorded among the HRV positive cases (single sampling events) was nasal discharge (n=278, 90.6%) followed by cough (n=226, 73.6%) and sore throat (n=62, 20.2%). There was no notable difference between symptoms in HRV-positive and negative cases (*p*-value > 0.05). Only a small proportion of those detected with HRV could identify a household member with ARI-like symptoms (n=43, 14.0%). Among the household members identified, 26/43 (60%) were school-going siblings, with 18 (69%) of them attending the same school and the other eight (31%), a different school, **Table 1**.

HRV circulation was detected throughout the three school terms under observation. Seasonal variations of HRV infections could not be identified due to breaks in sample collection during the school holidays. The lower primary had a higher HRV proportion compared to upper primary (19.8% versus 12.2%), *p*-value <0. 001, **Figure 1**. Spearman's rho indicated no correlation between Ct-values and age (rho= 0.03, *p*-value = 0.615).

HRV diversity

Amplification and sequencing of the VP4/2 genomic region were attempted on all HRV positive samples resulting in 82.4% (253/307) success. The unsuccessful samples either failed to amplify or had poor sequence quality. The resulting sequences were classified into 47 HRV genotypes: 24 HRV-A genotypes, seven HRV-Bs and 16 HRV-Cs. HRV-A was the most common species (134/253, 53.0%), followed by HRV-C (73/253, 28.9%) and HRV-B (46/253, 18.2%)**.** Some sequences violated the previously proposed genotype assignment thresholds, **Supplementary Table 2.** These were included in the analysis and classified with a suffix "-like" to the most similar known genotype (e.g. B48-like). The most frequent genotypes were A2 (24/253, 9.5%), B70 (22/253, 8.7%), A36 (16/253, 6.3%) and B48-like (16/253, 6.3%), **Figure 2**. No recombination events were identified.

Temporal occurrence and spatial clustering patterns

Numerous genotypes circulated simultaneously in the school with 22, 15 and 19 unique genotypes observed in term one, term two and term three, respectively. Nine genotypes recurred during the study period. Of the 22 that occurred in term one, two re-occurred in term two and four in term three; whereas of the 15 observed in term two, three recurred in term three. No genotype was observed across all the three school terms. Four of the recurring genotypes (A13, A59, B48, C3) were detected in the same class, **Figure 3A**.

Twelve genotypes were observed as singletons (i.e. present in a single sample/individual). We observed that frequent genotypes (n=21) circulated averagely for 28 days (median 23 days). Five of the frequent genotypes recurred in a subsequent school term. The longest persisting genotype was B70 (n=22 samples), seen in 81 days. Genotype A2 (n=24 samples) was similarly frequent but persisted for only 16 days. Among the frequent genotypes, none was limited to the upper primary group, whereas A10, A28, A101 and C3 were observed only in the lower primary group. However, no frequent genotype was limited to only one school class, **Figure 4.**

Individual infection patterns

Of the 164 HRV positive individuals, 62 (37.8%) contributed more than one positive sample. Three pupils, all from the lower primary group, presented with the most HRV positive samples (n=8) per person. The 253 successfully sequenced samples were collectively contributed by 144 individuals. Repeat HRV detections (n=109) were a combination of persistent infections (24/109) and re-infections (85/109). The number of genotypes per person ranged from one to five. About two-thirds of the individuals (98/144, 68.1%) had only one HRV genotype across the study period. Overall, the highest HRV diversity per person (5 genotypes) was observed in three pupils, all from the lower primary classes, **Table 2**. Time to re-infection varied greatly (13-307 days), with a median of 77 days. However, no individual was re-infected with the same genotype across the study period **Figure 3**.

Intra-type genetic diversity

Eight genotypes had at least ten samples/sequences, and five of these occurred as a single variant throughout the study. The remaining three, A28, B48 and B70, had more than one variant. A28 had two variants simultaneously observed. One variant was predominantly from KG1 (n=4/5), while the second was heterogenous. For B48, the second variant was observed as a genotype recurrence in a subsequent school term. Genotype B70 had the highest diversity, with three variants observed within one school term. The first B70 variant (13 samples) occurred across lower and upper primary

as well as teaching staff, the second (4 samples) was first observed 49 days after the genotype's overall index sequence (all samples from KG class 1) and the third (5 samples) was first observed 69 days after the genotype index sequence, with 4/5 samples coming from the lower primary. No individual had more than one variant of the same genotype, **Figure 5**. Genotype-specific ML trees of these three genotypes showed variant associated phylogenetic clustering, supported by bootstrap values >70 %. Only two phylogenetic clusters were associated with grade: K1 (n=4/5 samples from KG3) in A28 and K2 (n=4 samples from KG1) in B70, **Supplementary Figure 1**.

HRV among teaching staff

Thirty-two samples were collected from teachers, of which 5 (15.6%) were HRV positive - from 4 individuals. Three HRV genotypes were identified: A2 (n=2), B70 (n=2) and B99 (n=1). The A2 and B70 genotypes were detected in teachers several days after their initial detection in the student population. The B99 sample from a teacher was the only case of this genotype identified during the entire study period suggesting that they acquired the infection from outside the school setting, and no onward transmission was observed.

DISCUSSION

In this coastal Kenya school study, we found that HRV occurs year-round in line studies in this location among symptomatic individuals within the KHDSS area (11%-23%) [31,32,38]. HRV was detected across all age-groups, with the highest proportion in the <5 year-olds and the lowest proportion in older age groups(>=18 years), in agreement with previous findings [20,39,40]. A proportion of the HRV positive children (14%) identified another household member as having ARI-like symptoms, suggesting transmission at the household level that might contribute to transmission at school (or vice versa).

All HRV species were found in circulation throughout the year. HRV-A circulated widely (53%), more than -B (18%) and -C (29%), a contrast to previous reports where HRV-A and HRV-C co-dominated [31,32,38]. However, a similar occurrence of HRV species was reported in the first two years of aboriginal and non-aboriginal Australian children[41]. There was considerable HRV diversity, with almost a third of all known HRV genotypes detected. HRV infections comprised of single genotype occurrences, observed in a single sample/individual, as well as frequent genotypes affecting numerous pupils across several classes. This pattern has been observed elsewhere[42] and it is not clear whether this is due to varying serotype infectivity[43] or other epidemiological factors.

Numerous genotypes co-circulated in every school term, implying that no particular genotype predominates at any given time period. Previous studies have shown that contemporary HRV infections in a given population are characterised by numerous genotype-specific "mini-epidemics"[44] and that up to 30 genotypes circulate simultaneously at a given geographical area[45]. No frequent genotype was limited to one class, suggesting heterogeneous mixing and transmission within the school. However, four frequent genotypes and one variant of B70 were observed only in the lower primary, an indication of social clustering. Genotype recurrence in a subsequent school term was observed in nine genotypes. Although it is not clear whether the study design missed samples between two genotype occurrences, the infrequency of genotype recurrence is possibly a reflection of herd immunity to specific types within the school/local community or a reflection of random introductions into the school/local community. Frequent genotypes in the school persisted for about a month on average. This is a shorter period than earlier observed across the KHDSS (a larger geographical scope) during an outpatient surveillance [38]. This is probably due to increased transmission (steered by high contact rates among school-going children) resulting in a shorter duration epidemic.

The younger age groups exhibited high rhinovirus diversity as they had more HRV re-infections. No individual was re-infected with the same genotype; further evidence of serotype-specific immunity to HRV lasting at least one year [46].

We demonstrate the occurrence of intra-genotype variants and associated phylogenetic clusters, which were either separate rhinovirus introductions or diversification of a single variant after introduction and as a result forming different transmission clusters. This observation highlights the benefit of sequence data over serology to study viral transmission dynamics. The numerous HRV genotypes, sparse sampling of ARIs and the minimal resolution from partial short sequences obtained here did not allow for transmission inferences (due to insufficient within type variation).

An outpatient health facility located within the same location as the school was recruited into an ARI surveillance study [27] from December 2015 to November 2016, five months prior to our school study. This outpatient clinic is within a radius of 4 km from the school. Detailed analysis of molecular epidemiology of HRV for samples collected at this outpatient clinic was reported elsewhere [38] . Although not a primary objective of this study, we compared the diversity of HRV infections between the two study periods. We observed 12 common genotypes in the two studies: A13, A20, A28, A31, A46, A54, A78, A101, B42, C6, C11 and C19. However, only one genotype was frequent in both periods: A101 (**Supplementary Figure 2**). Our comparison of HRV diversity between a school setting and clinical cases in a health facility within the same geographical location and two consecutive seasons showed only one frequent genotype present in both studies. This is an indication that HRV diversity within a community varies widely over time, as previously observed[42]. It is not definite what drives the exchange of common rhinovirus genotypes. The rapid turnover and co-existence of genotypes and variants might be determined by immunologically mediated selection processes or other nonselective epidemiological processes.

Our study had some limitations. First, the dichotomy in the number of samples collected weekly in lower versus upper primary posed a challenge when comparing the two groups. Second, weekly

sampling of only symptomatic persons will likely have resulted in missed HRV infections, impairing the overview of HRV dynamics. In addition, the study failed to successfully amplify and sequence nearly 18% of HRV positive samples. Failure was not correlated with viral load and may have been caused by variability in primer-annealing sites resulting in mismatches. This may have resulted in missed genotypes or sub-variants.

This study provides improved knowledge of the diversity and temporal characteristics of HRV in a school setting, reinforcing the notion that schools are a focal point in understanding HRV transmission in the community. The effect of numerous individuals in close contact enabling HRV transmission is evident. In addition, we see that infections could be linked to transmission events occurring outside the school setting, i.e. household setting. The contemporary inclusion of different population structures (e.g. schools, households, health centres) in studying HRV dynamics will improve understanding of HRV epidemiology in communities. Future studies should focus on whole-genome sequencing to fully elucidate transmission clusters.

Data availability

The replication data and analysis scripts for this manuscript shall be made available from the Harvard Dataverse: (https://dataverse.harvard.edu/dataverse/vec).

Competing interests

The authors declare no competing interests.

Acknowledgements

We thank all the study participants, teachers, School Management, Parent Teacher Association and County Education Management Committees for allowing us to conduct the study within their school. We are grateful to the field study team for participant recruitment and the laboratory staff of the KEMRI-Wellcome Trust Research Programme / Virus Epidemiology and Control research group. This paper is published with the permission of the Director of KEMRI.

Funding

This work was supported by the Wellcome Trust through a Wellcome Senior Investigator Award to the last author (#102975). The first author was supported by the DELTAS Africa Initiative [DEL-15-003]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [107769/Z/10/Z] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

REFERENCES

1. Wang K, Xi W, Yang D, et al. Rhinovirus is associated with severe adult community-acquired pneumonia in China. *J Thorac Dis*. 2017;9(11):4502-4511. doi:10.21037/jtd.2017.10.107

2. Peltola V, Waris M, Osterback R, Susi P, Hyypia T, Ruuskanen O. Clinical effects of rhinovirus infections. *J Clin Virol*. 2008;43(4):411-414. doi:10.1016/j.jcv.2008.08.014

3. Chen J, Hu P, Zhou T, et al. Epidemiology and clinical characteristics of acute respiratory tract infections among hospitalized infants and young children in Chengdu, West China, 2009–2014. *BMC Pediatr*. 2018;18. doi:10.1186/s12887-018-1203-y

4. Launes C, Armero G, Anton A, et al. Molecular epidemiology of severe respiratory disease by human rhinoviruses and enteroviruses at a tertiary paediatric hospital in Barcelona, Spain. *Clin Microbiol Infect*. 2015;21(8):799.e5-7. doi:10.1016/j.cmi.2015.04.021

5. Atkinson TP. Is asthma an infectious disease? New evidence. *Curr Allergy Asthma Rep*. 2013;13(6):702-709. doi:10.1007/s11882-013-0390-8

6. Jacobs SE, Lamson DM, St George K, Walsh TJ. Human rhinoviruses. *Clin Microbiol Rev*. 2013;26(1):135-162. doi:10.1128/CMR.00077-12

7. McLean HQ, Peterson SH, King JP, Meece JK, Belongia EA. School absenteeism among school-aged children with medically attended acute viral respiratory illness during three influenza seasons, 2012-2013 through 2014-2015. *Influ Other Respir Viruses*. 2017;11(3):220-229. doi:10.1111/irv.12440

8. Rotbart HA, Hayden FG. Picornavirus infections: a primer for the practitioner. *Arch Fam Med*. 2000;9(9):913-920.

9. Jennings LC, Dick EC. Transmission and control of rhinovirus colds. *Eur J Epidemiol*. 1987;3(4):327-335.

10. Loeffelholz MJ, Trujillo R, Pyles RB, et al. Duration of rhinovirus shedding in the upper respiratory tract in the first year of life. *Pediatrics*. 2014;134(6):1144-1150. doi:10.1542/peds.2014-2132

11. Stellrecht KA. Chapter 11 - Molecular Testing for Respiratory Viruses. In: Coleman WB, Tsongalis GJBT-DMP, eds. Academic Press; 2017:123-137. doi:https://doi.org/10.1016/B978-0-12-800886-7.00011-X

12. Hung IFN, Zhang AJ, To KKW, et al. Unexpectedly Higher Morbidity and Mortality of Hospitalized Elderly Patients Associated with Rhinovirus Compared with Influenza Virus Respiratory Tract Infection. *Int J Mol Sci*. 2017;18(2). doi:10.3390/ijms18020259

13. Denlinger LC, Sorkness RL, Lee WM, et al. Lower Airway Rhinovirus Burden and the Seasonal Risk of Asthma Exacerbation. *Am J Respir Crit Care Med*. 2011;184(9):1007-1014. doi:10.1164/rccm.201103-0585OC

14. Pierangeli A, Scagnolari C, Selvaggi C, et al. Rhinovirus frequently detected in elderly adults attending an emergency department. *J Med Virol*. 2011;83(11):2043-2047. doi:10.1002/jmv.22205

15. Lewis-Rogers N, Seger J, Adler FR. Human Rhinovirus Diversity and Evolution: How Strange the Change from Major to Minor. *J Virol*. 2017;91(7). doi:10.1128/JVI.01659-16

16. Simmonds P, Gorbalenya AE, Harvala H, et al. Recommendations for the nomenclature of enteroviruses and rhinoviruses. *Arch Virol*. 2020;165(3):793-797. doi:10.1007/s00705-019-04520-6

17. McIntyre CL, Knowles NJ, Simmonds P. Proposals for the classification of human rhinovirus species A, B and C into genotypically assigned types. *J Gen Virol*. 2013;94(Pt 8):1791-1806. doi:10.1099/vir.0.053686-0

18. Simmonds P, McIntyre C, Savolainen-Kopra C, Tapparel C, Mackay IM, Hovi T. Proposals for the classification of human rhinovirus species C into genotypically assigned types. *J Gen Virol*. 2010;91(Pt 10):2409-2419. doi:10.1099/vir.0.023994-0

19. Kiang D, Kalra I, Yagi S, et al. Assay for 5' noncoding region analysis of all human rhinovirus prototype strains. *J Clin Microbiol*. 2008;46(11):3736-3745. doi:10.1128/jcm.00674-08

20. Peltola V, Waris M, Osterback R, Susi P, Ruuskanen O, Hyypia T. Rhinovirus transmission within families with children: incidence of symptomatic and asymptomatic infections. *J Infect Dis*. 2008;197(3):382-389. doi:10.1086/525542

21. Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. *PLoS One*. 2014;9(8):e104786. doi:10.1371/journal.pone.0104786

22. Agoti CN, Munywoki PK, Phan MVT, et al. Transmission patterns and evolution of respiratory syncytial virus in a community outbreak identified by genomic analysis. *Virus Evol*. 2017;3(1):vex006. doi:10.1093/ve/vex006

23. Munywoki PK, Koech DC, Agoti CN, et al. The source of respiratory syncytial virus infection in infants: a household cohort study in rural Kenya. *J Infect Dis*. 2014;209(11):1685-1692. doi:10.1093/infdis/jit828

24. Jefferson T, Del Mar C, Dooley L, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane database Syst Rev*. 2010;(1):CD006207. doi:10.1002/14651858.CD006207.pub3

25. Scott JAG, Bauni E, Moisi JC, et al. Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). *Int J Epidemiol*. 2012;41(3):650-657. doi:10.1093/ije/dys062

26. Adema IW, Kamau E, Uchi Nyiro J, et al. Surveillance of respiratory viruses among children

attending a primary school in rural coastal Kenya [version 1; peer review: awaiting peer review]. *Wellcome Open Res*. 2020;5(63). doi:10.12688/wellcomeopenres.15703.1

27. 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. doi:10.12688/wellcomeopenres.14662.1

28. Berkley JA, Munywoki P, Ngama M, et al. Viral etiology of severe pneumonia among Kenyan infants and children. *Jama*. 2010;303(20):2051-2057. doi:10.1001/jama.2010.675

29. Kamau E, Agoti CN, Lewa CS, et al. Recent sequence variation in probe binding site affected detection of respiratory syncytial virus group B by real-time RT-PCR. *J Clin Virol*. 2017;88:21-25. doi:10.1016/j.jcv.2016.12.011

30. 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*. 2011;49(6):2318 LP - 2320. doi:10.1128/JCM.02605-10

31. 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. doi:10.1002/jmv.23251

32. Kamau E, Onyango CO, Otieno GP, et al. An Intensive, Active Surveillance Reveals Continuous Invasion and High Diversity of Rhinovirus in Households. *J Infect Dis*. 2019;219(7):1049-1057. doi:10.1093/infdis/jiy621

33. 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. doi:10.1093/bioinformatics/btw412

34. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol*. 2015;1(1). doi:10.1093/ve/vev003

35. Nguyen LT. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. 2015;32(1):268-274. doi:10.1093/molbev/msu300

36. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016;33(7):1870-1874. doi:10.1093/molbev/msw054

37. Kiyota N, Kobayashi M, Tsukagoshi H, et al. Genetic analysis of human rhinovirus species A to C detected in patients with acute respiratory infection in Kumamoto prefecture, Japan 2011-2012. *Infect Genet Evol*. 2014;21:90-102. doi:10.1016/j.meegid.2013.10.024

38. Morobe JM, Nyiro JU, Brand S, et al. Human rhinovirus spatial-temporal epidemiology in rural coastal Kenya, 2015-2016, observed through outpatient surveillance. *Wellcome open Res*. 2018;3:128. doi:10.12688/wellcomeopenres.14836.2

39. Kieninger E, Fuchs O, Latzin P, Frey U, Regamey N. Rhinovirus infections in infancy and early childhood. *Eur Respir J*. 2013;41(2):443 LP - 452. doi:10.1183/09031936.00203511

40. Baillie VL, Moore DP, Mathunjwa A, Morailane P, Simoes EAF, Madhi SA. Molecular Subtyping of Human Rhinovirus in Children from Three Sub-Saharan African Countries. *J Clin Microbiol*. 2019;57(9). doi:10.1128/JCM.00723-19

41. Annamalay AA, Khoo S-K, Jacoby P, et al. Prevalence of and risk factors for human rhinovirus infection in healthy aboriginal and non-aboriginal Western Australian children. *Pediatr Infect Dis J*. 2012;31(7):673-679. doi:10.1097/INF.0b013e318256ffc6

42. Monto AS. Epidemiology of viral respiratory infections. *Am J Med*. 2002;112 Suppl:4S-12S. doi:10.1016/s0002-9343(01)01058-0

43. MONTO AS, CAVALLARO JJ. THE TECUMSEH STUDY OF RESPIRATORY ILLNESSIV. PREVALENCE OF RHINOVIRUS SEROTYPES, 1966–196912. *Am J Epidemiol*. 1972;96(5):352-360. doi:10.1093/oxfordjournals.aje.a121466

44. Martin ET, Kuypers J, Chu HY, et al. Heterotypic Infection and Spread of Rhinovirus A, B, and C among Childcare Attendees. *J Infect Dis*. 2018;218(6):848-855. doi:10.1093/infdis/jiy232

45. Mackay IM, Lambert SB, Faux CE, et al. Community-Wide, Contemporaneous Circulation of a Broad Spectrum of Human Rhinoviruses in Healthy Australian Preschool-Aged Children During a 12-Month Period. *J Infect Dis*. 2012;207(9):1433-1441. doi:10.1093/infdis/jis476

46. Barclay WS, al-Nakib W, Higgins PG, Tyrrell DA. The time course of the humoral immune response to rhinovirus infection. *Epidemiol Infect*. 1989;103(3):659-669. doi:10.1017/s095026880003106x

TABLES AND FIGURES

Table legend

Table 1. Baseline characteristics of the HRV positive cases at a rural Kenyan school sampled throughout 3 terms from May 2017 to April 2018.

Table 2. Human rhinovirus detection and genotyping summary among individuals.

Table 1. Baseline characteristics of the HRV positive cases at a rural Kenyan school sampled throughout 3 terms from May 2017 to April 2018.

Characteristic	Categories	No. Pos	% Pos	No. Neg	% Neg	Total
Age (pupils)	<=5yrs	78	22.4	270	77.6	348
	6-10yrs	137	17.6	641	82.4	778
	11-17yrs	82	12.4	578	87.6	660
	>=18yrs	2	5.9	32	94.1	34
	Unspecified	3	42.9	4	57.1	7
Age (teachers)	>=18yrs	5	15.6	27	84.4	32
	Total	307	16.5	1552	83.5	1859
Gender	Male	169	18.8	731	81.2	900
	Female	138	14.4	821	85.6	959
Symptoms in the last two weeks						
Cough	Yes	226	16.6	1137	83.4	1363
	No	81	16.4	413	83.6	494
Nasal Discharge	Yes	278	17.0	1357	83.0	1635
	No	29	13.1	193	86.9	222
Sore throat	Yes	62	16.2	321	83.8	383
	No	245	16.6	1227	83.4	1472

Household (HH) members						
Other persons in HH with symptoms of ARI?	Yes	43	14.2	260	85.8	303
	No	215	17.4	1024	82.7	1239
	Don't know	49	15.5	268	84.5	317
Are they in school? (n=43)	Yes	26	16.1	136	84.0	162
	No	17	12.1	124	87.9	141
	Unspecified	264	17.0	1292	83.0	1556

No. Pos: Number positive

% Pos: Percent positive

No. Neg: Number negative

% Neg: Percent negative

Table 2. Human rhinovirus detection and genotyping summary among individuals.

Distribution of HRV among individuals	**Frequency of HRV positive samples/ person**	**Number of individuals**	**Percentage %**
	One	102	62.2
	Two	31	18.9
	Three	13	7.9
	Four	5	3.0
	Five	3	1.8
	Six	4	2.4
	Seven	3	1.8
	Eight	3	1.8
	Total	164	100
Distribution of sequenced samples among individuals	**Frequency of sequenced samples/ person**	**Number of individuals**	**Percentage %**
	One	96	66.7
	Two	23	16.0
	Three	9	6.3
	Four	5	3.5
	Five	5	3.5
	Six	4	2.8
	Seven	1	0.7

	Eight	1	0.7
	Total	144	100
Diversity of HRV	**Number of unique types/ person**	**Number of individuals**	**Percentage %**
	One	98	68.1
	Two	22	15.3
	Three	12	8.3
	Four	9	6.3
	Five	3	2.1
	Total	144	100

Figure legend

Figure 1. Patterns of HRV infections within the school over one year. (A). Month by month HRV proportion with respective 95% confidence intervals and number of samples tested. (B). Class-specific HRV proportion with respective 95% confidence intervals. The two bars furthest to the right are aggregated proportions of lower primary and upper primary groups.

Figure 2. Phylogenetic analysis and genotype assignment of generated HRV sequences. Species-specific maximum likelihood trees of (A) HRV-A, (B) HRV-B and (C) HRV-C. The tip shapes are colored by the school class of the individual. The scale bars represent nucleotide substitutions per site. The black tips represent the prototype sequence of respective strain. (D) A circular bar plot showing the frequencies of HRV genotypes identified. The bars are colored by HRV species and the tips are labeled by: HRV-type, frequency.

Figure 3. The temporal occurrence and infection patterns of HRV. (A) The temporal occurrence of all HRV genotypes across the one-year study period. (B) Individual infection patterns across the one-year period. The individuals are ordered by date of first HRV infection. (C) Number of days to a subsequent HRV case for individuals who were re-infected during the study period. All individual first infections are dated day 0. The individuals are ordered by number of HRV positive samples, with the individual with most positive samples appearing at the bottom.

Figure 4. Distribution of genotypes across the school. (A) Heatmap showing the distribution of frequent genotypes across the school classes. The intensity of the color correlates to frequency of samples. Color intensity has been scaled to correct for sampling bias between the lower primary and upper primary classes. (B) Distribution of frequent genotypes between the upper and lower primary groups. The sizes of the circles correlate to number of samples and the color to the respective group: either lower or upper primary.

Figure 5. Genetic diversity of HRV genotypes with a frequency of more than ten. From top left A2, A28, A36, B48-like, B70, C_pat19, C3 and C13 (bottom right). Nucleotide substitutions are demonstrated by a colored bar. A substitution to an “A” is indicated by green, “T” by red, “G” by indigo and “C” by blue bars. The sequences are labeled by grade of individual and ordered by date of sample collection with the genotype’s index sequence at the bottom, and acting as a reference.

Figure 1_A

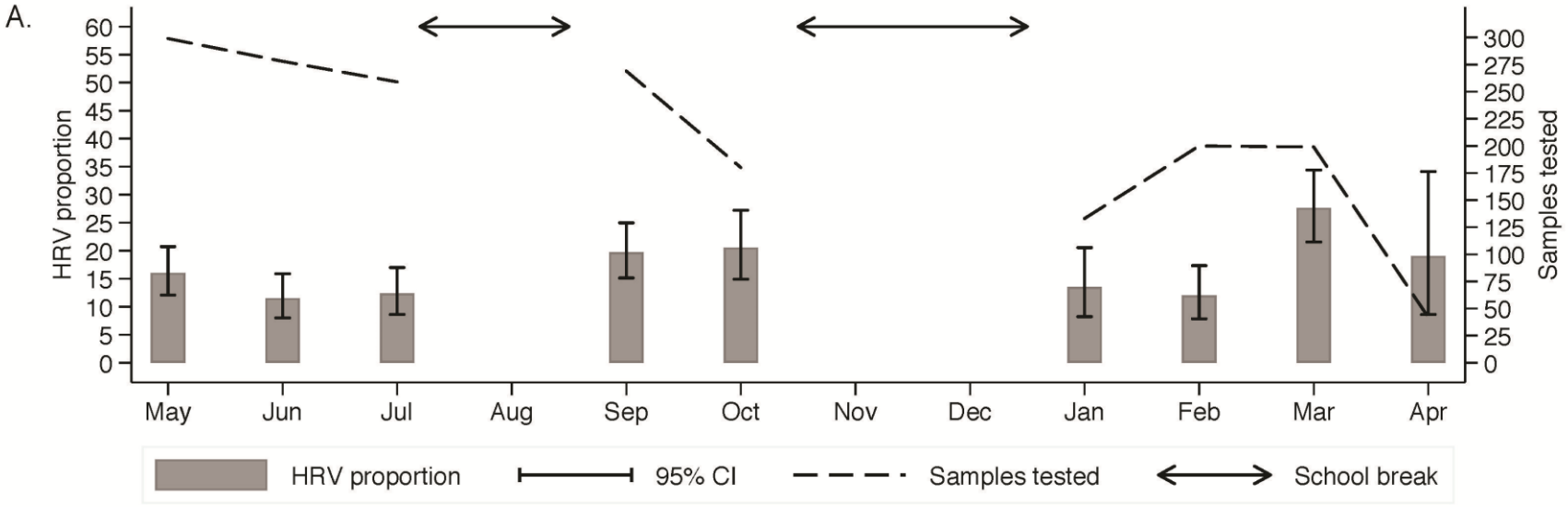

Figure 1_B

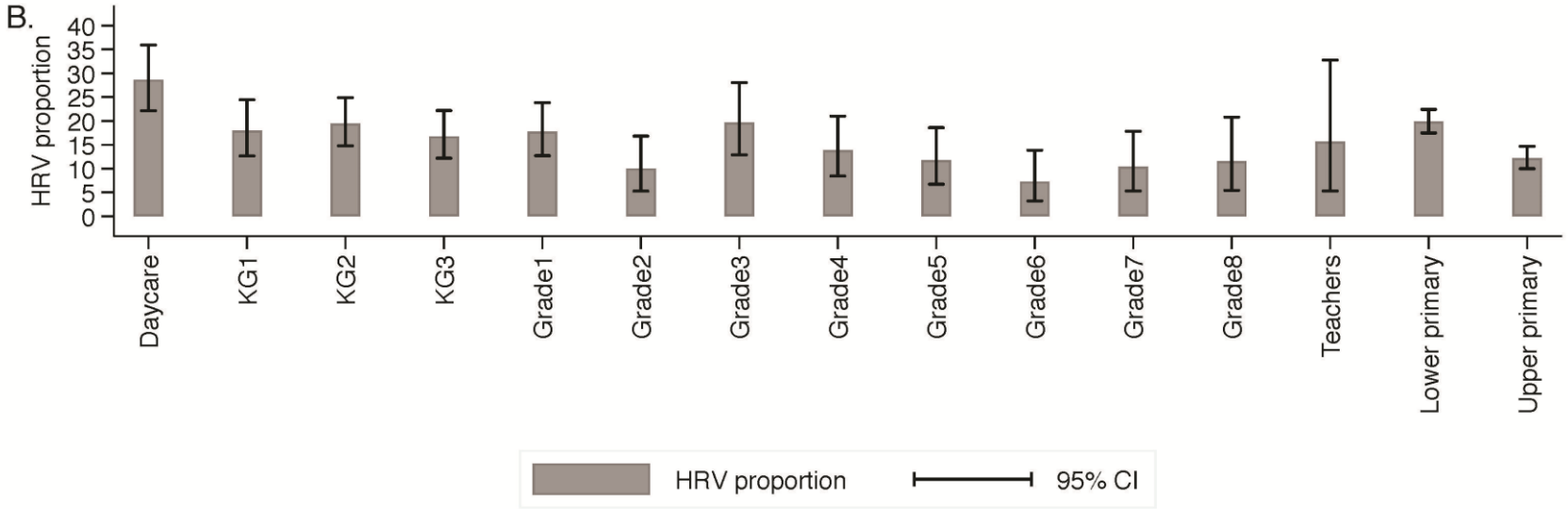

Figure 2_A

Figure 2_B

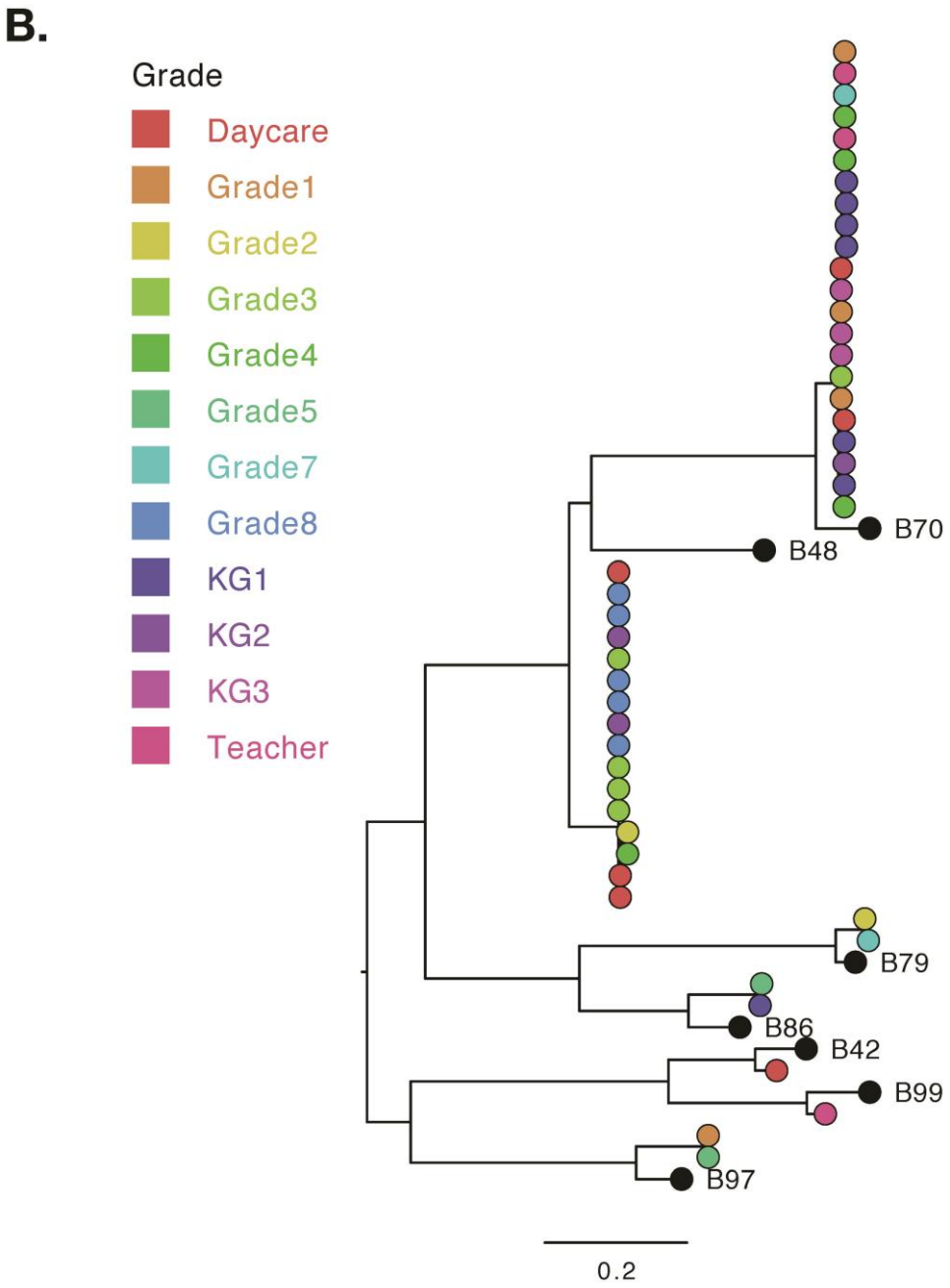

Figure 2_C

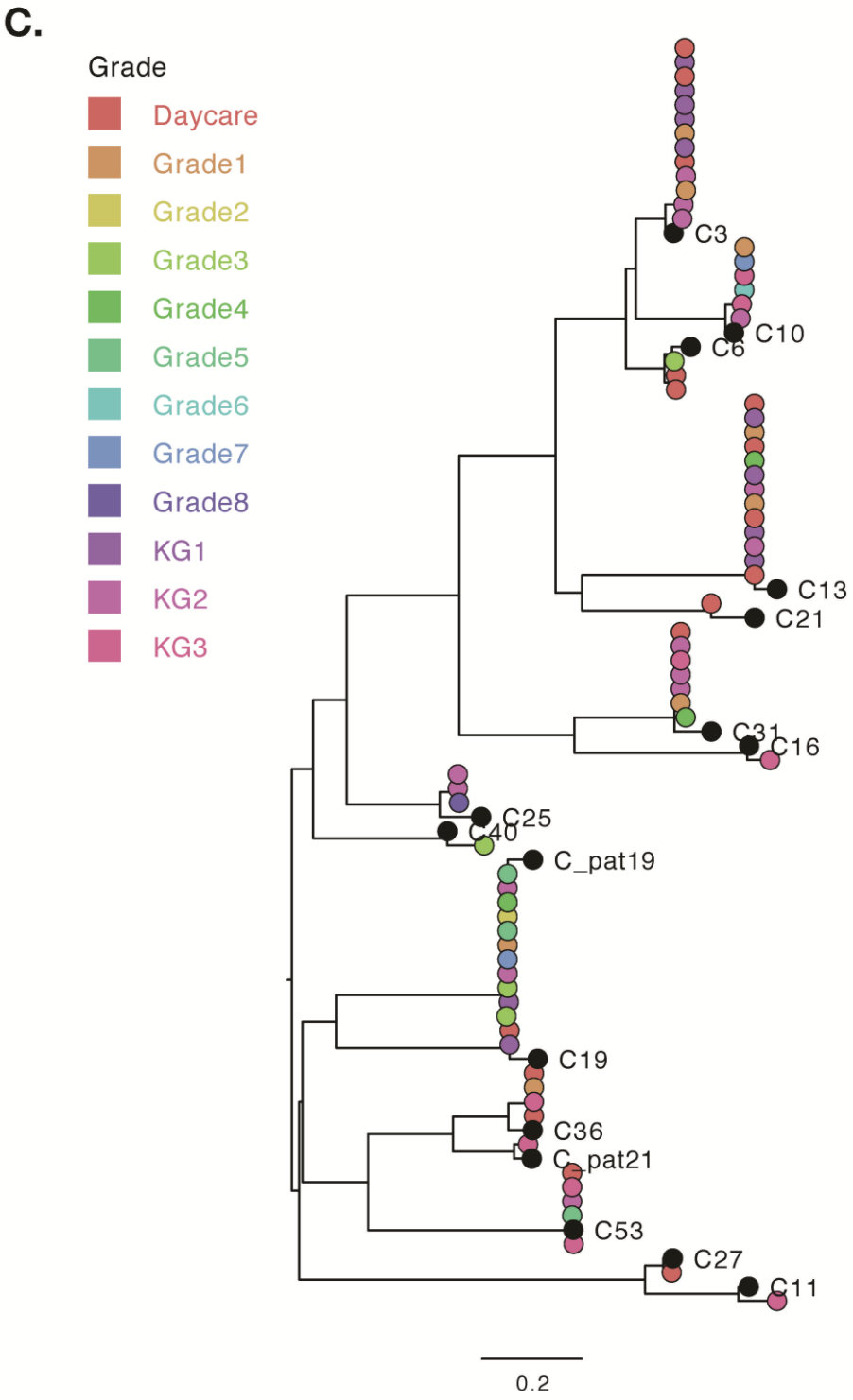

Figure 2_D

D.

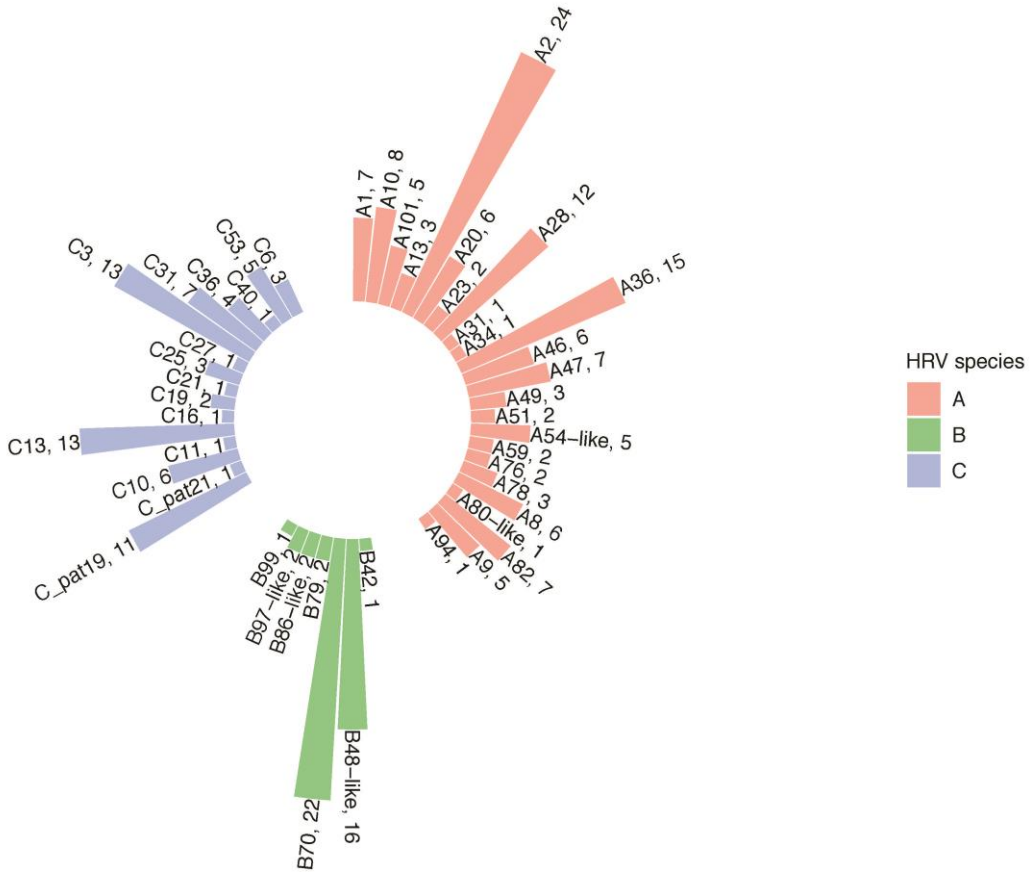

Figure 3_A

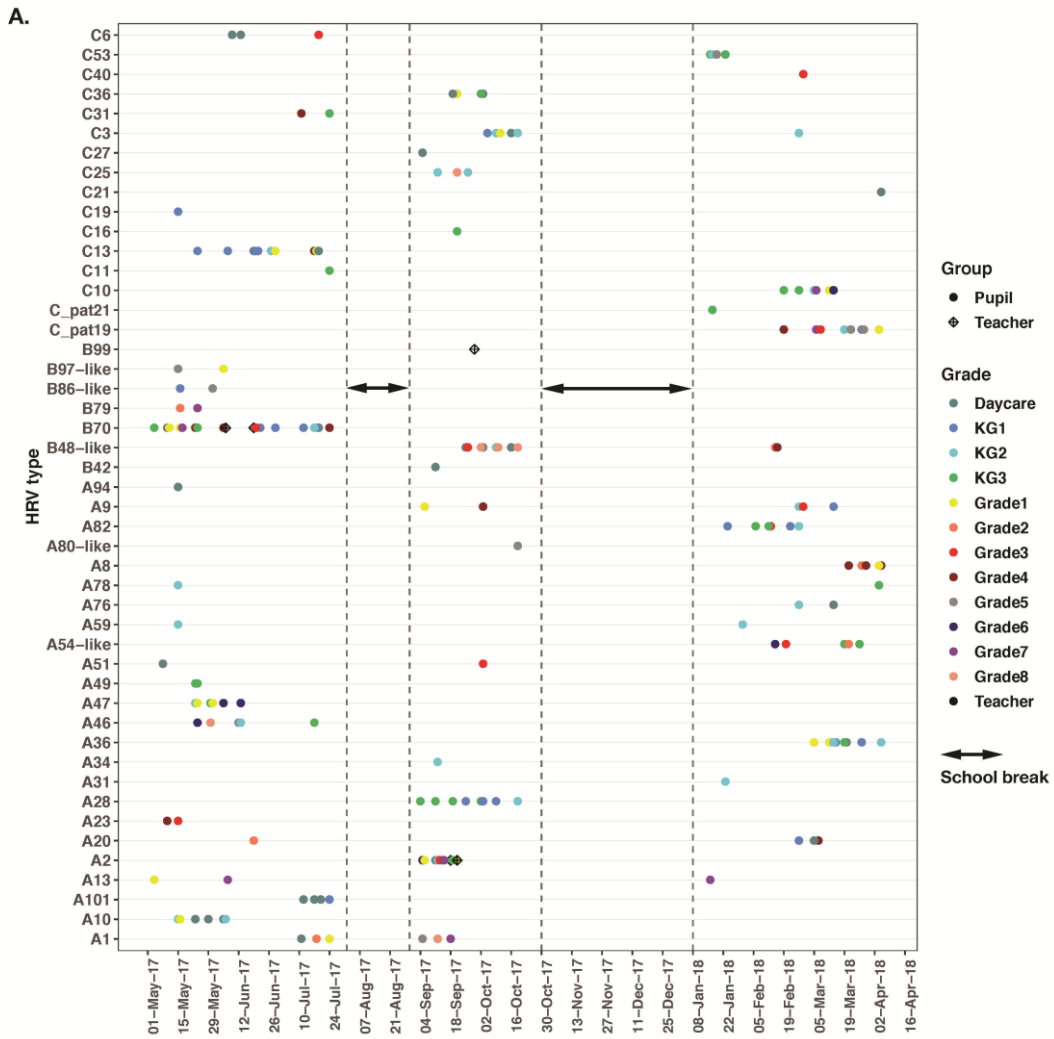

Figure 3_B

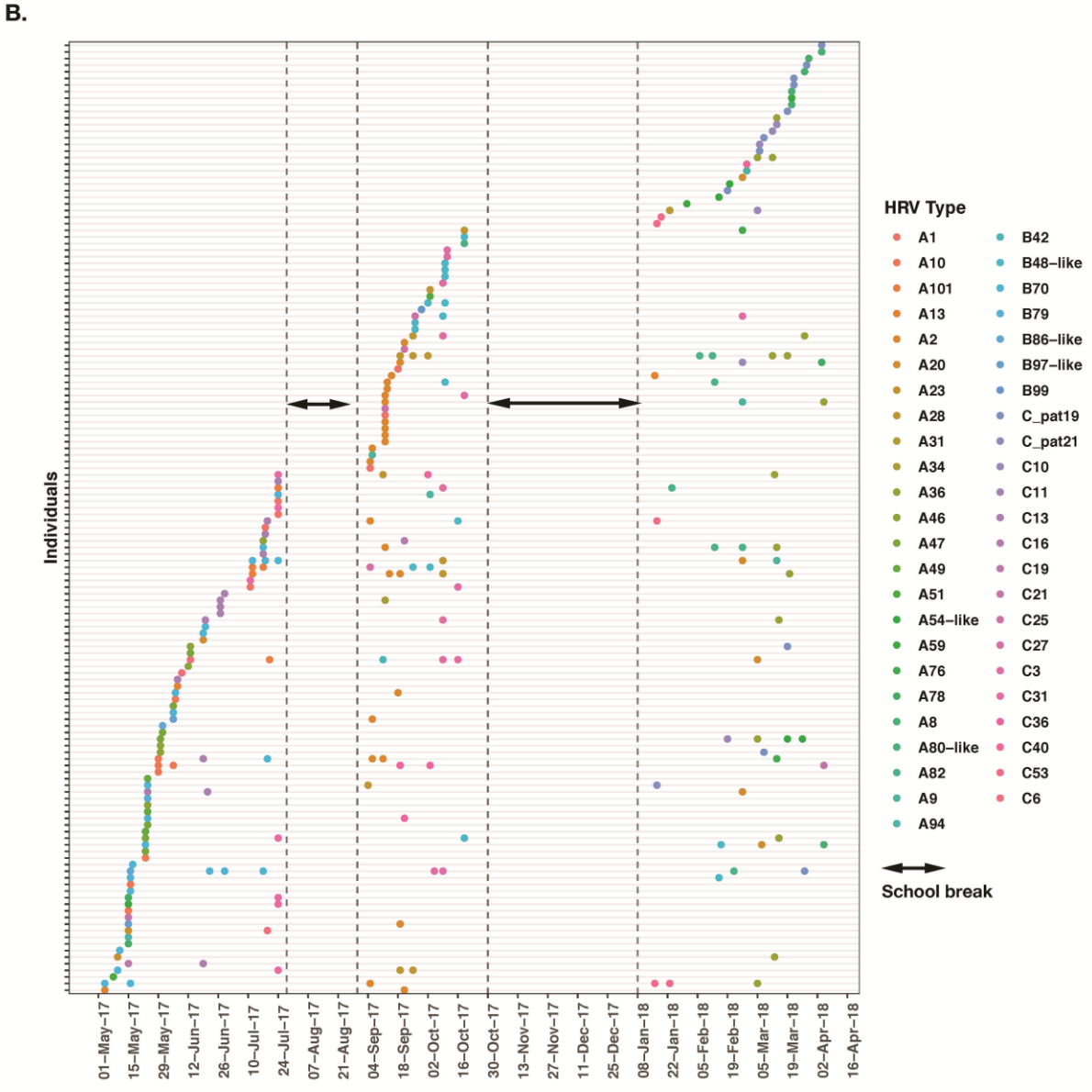

Figure 3_C

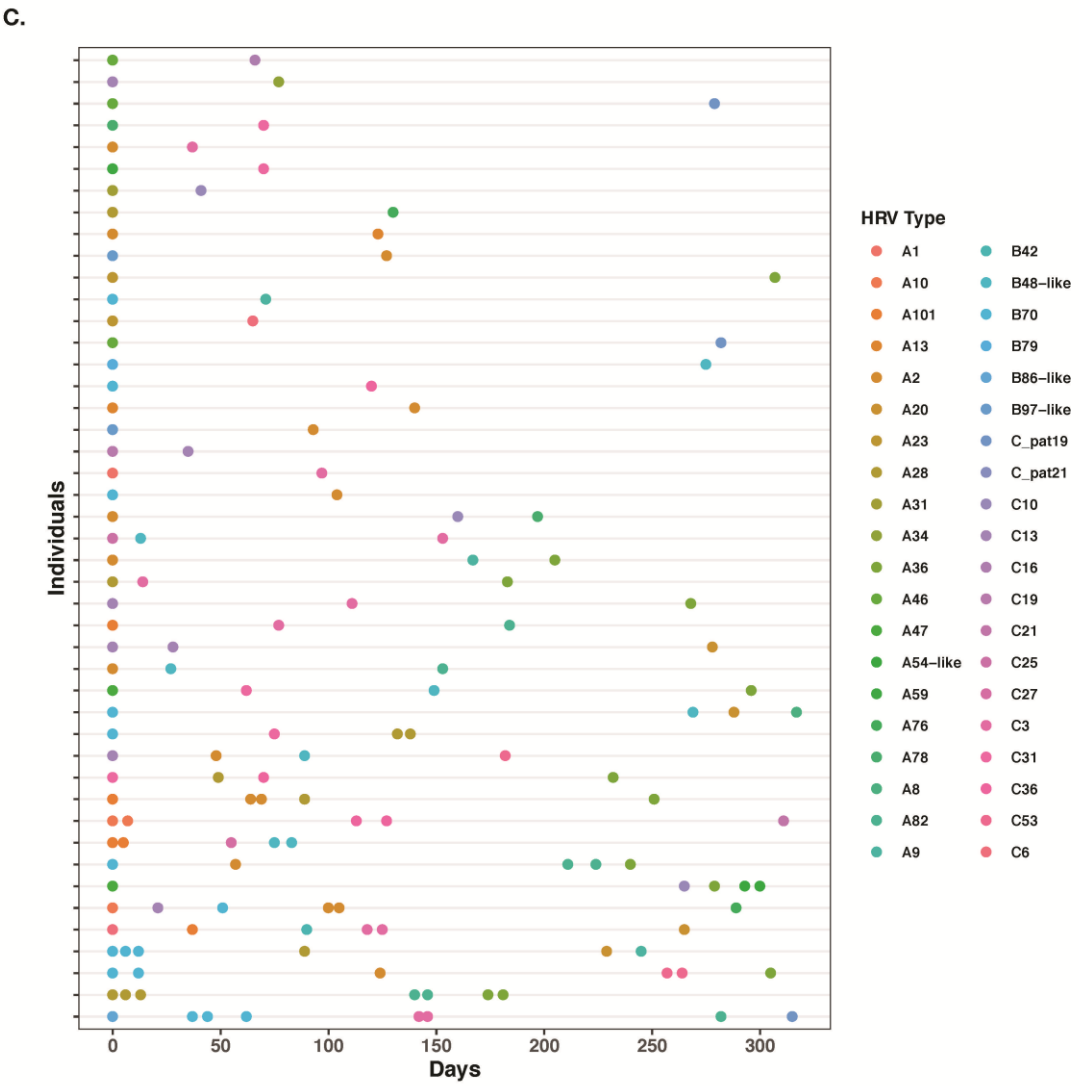

Figure 4_A

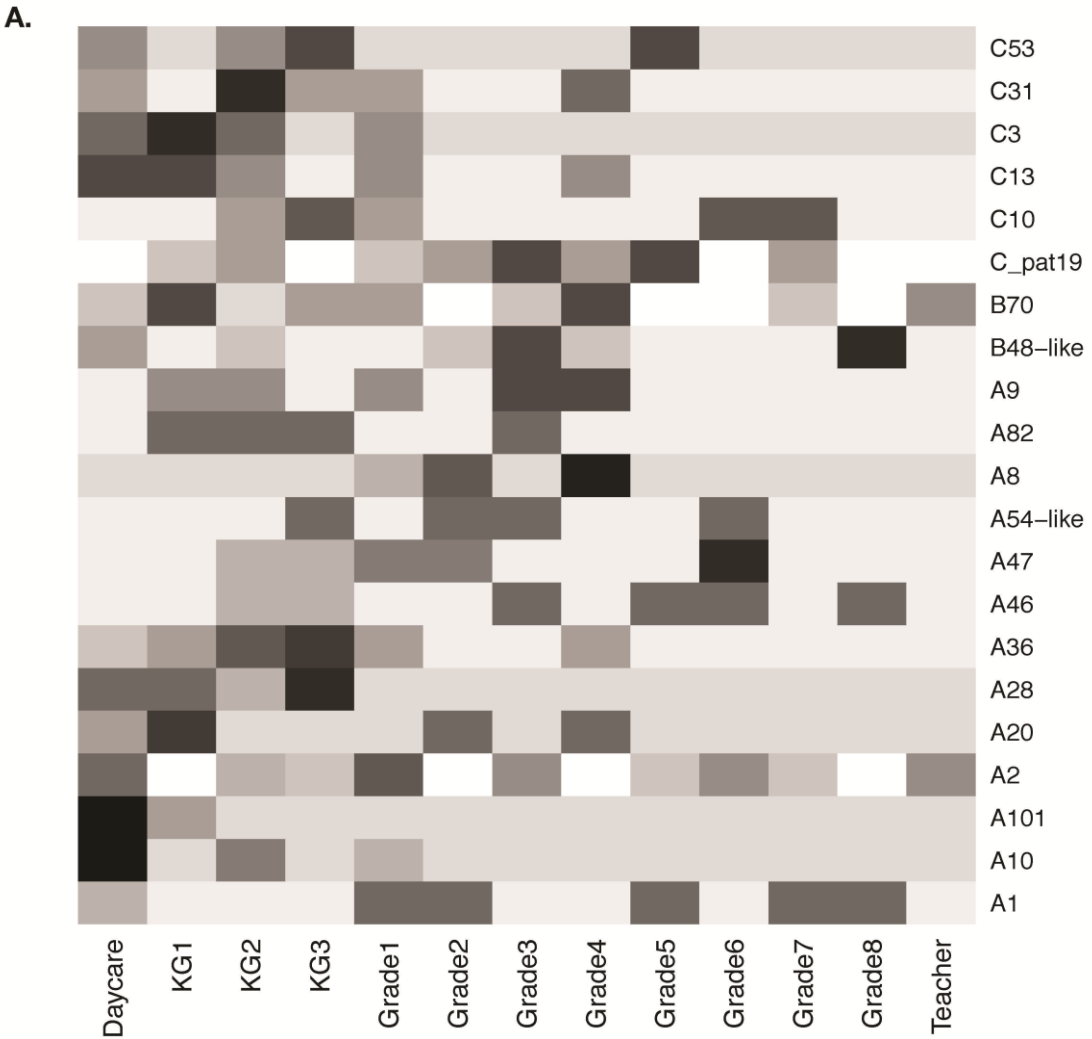

Figure 4_B

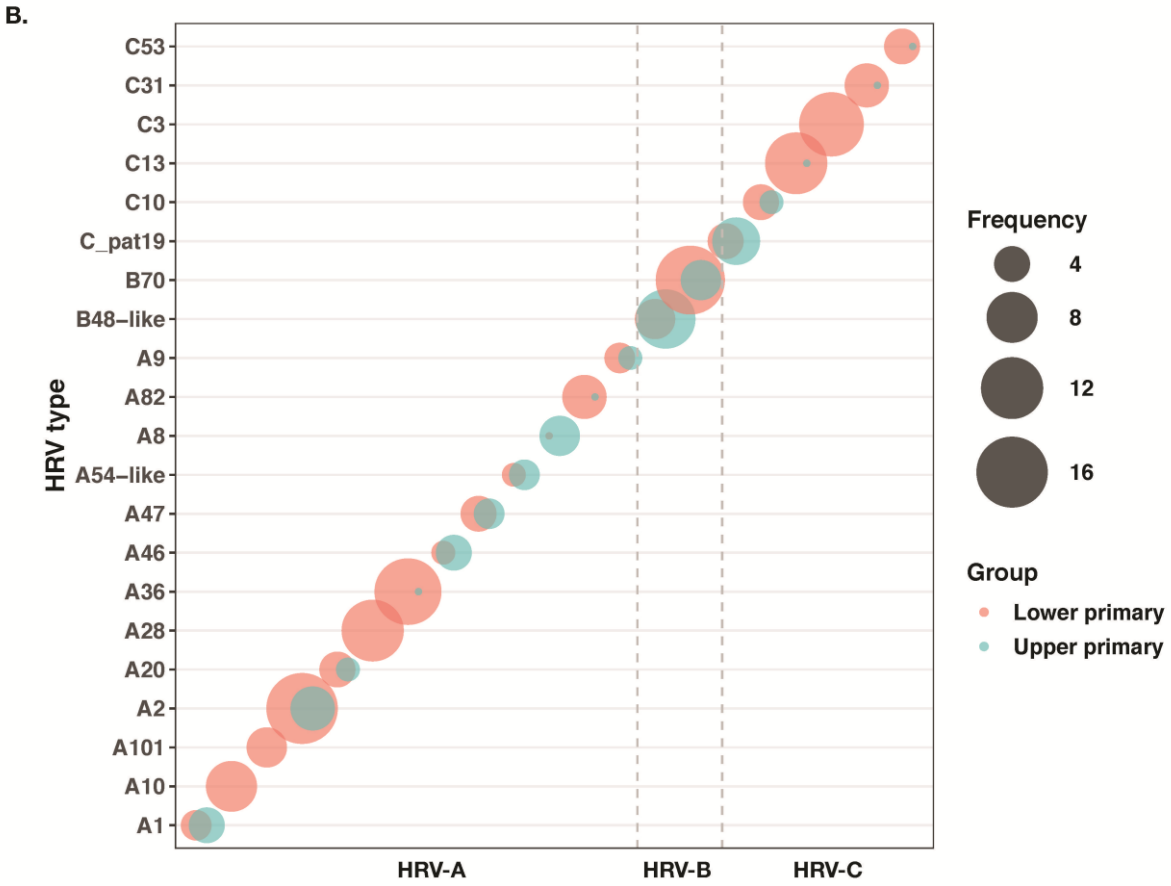

Figure 5