Comparative Analysis of Spatial-temporal Patterns of Human Metapneumovirus and Respiratory Syncytial Virus in Africa Using Genetic Data, 2011-2014

John W Oketch (jwo5@leicester.ac.uk)
KEMRI-Wellcome Trust Research Programme: Centre for Geographic Medicine Research Coast
https://orcid.org/0000-0002-8020-0787

Everlyn Kamau
Oxford University

James R Otieno
KEMRI-Wellcome Trust Research Programme: Centre for Geographic Medicine Research Coast

Anthony Mwema
KEMRI-Wellcome Trust Research Programme: Centre for Geographic Medicine Research Coast

Clement Lewa
KEMRI-Wellcome Trust Research Programme: Centre for Geographic Medicine Research Coast

Everlyne Isoe
Pwani University School of Pure and Applied Sciences

James D Nokes
KEMRI-Wellcome Trust Research Programme: Centre for Geographic Medicine Research Coast

Charles N Agoti
KEMRI-Wellcome Trust Research Programme: Centre for Geographic Medicine Research Coast

Research

Keywords: Human metapneumovirus, respiratory syncytial virus, Phylogeographic analysis, spatial-temporal

Posted Date: December 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-135759/v1

Abstract

Background: Human metapneumovirus (HMPV) and respiratory syncytial virus (RSV) are leading causes of viral severe acute respiratory illnesses in childhood. Both the two viruses belong to the *Pneumoviridae* family and show overlapping clinical, epidemiological and transmission features. However, it is unknown whether these two viruses have similar geographic spread patterns which may provide insight into designing and evaluating their epidemic control measures.

Methods: We conducted comparative phylogenetic and phylogeographic analyses to explore the spatial-temporal patterns of HMPV and RSV across Africa using 232 HMPV and 842 RSV attachment (G) glycoprotein gene sequences obtained from 5 countries (The Gambia, Zambia, Mali, South Africa, and Kenya) between August 2011 and January 2014.

Results: Phylogeographic analyses found frequently similar patterns of spread of RSV and HMPV. Viral sequences commonly clustered by region, i.e., West Africa (Mali, Gambia), East Africa (Kenya) and Southern Africa (Zambia, South Africa), and similar genotype dominance patterns were observed between neighbouring countries. Both HMPV and RSV country epidemics were characterized by co-circulation of multiple genotypes. Sequences from different African sub-regions (East, West and Southern Africa) fell into separate clusters interspersed with sequences from other countries globally.

Conclusion: The spatial clustering patterns of viral sequences and genotype dominance patterns observed in our analysis suggests strong regional links and predominant local transmission. The geographical clustering further suggests independent introduction of HMPV and RSV variants in Africa from the global pool, and local regional diversification.

Introduction

Human metapneumovirus (HMPV) and respiratory syncytial virus (RSV) are leading viral respiratory pathogens that cause seasonal epidemics of acute respiratory tract illness and are responsible for a significant fraction of childhood pneumonia (1). A recent multi-country study, PERCH (Pneumonia Etiology Research for Child Health), in Africa and Asia reported RSV as the leading cause of pneumonia in children aged under five years accounting for at least 31% of the aetiological distribution (2). In the same study, HMPV accounted for 5% of the aetiological distribution. The current study presents a molecular-epidemiological analysis of samples collected by the PERCH study from the five African counties, i.e., Kenya, South Africa, Zambia, Mali and Gambia.

RSV and HMPV infections present with overlapping clinical and epidemiological profiles (3, 4). Following infection with either virus, the clinical presentation can range from asymptomatic infection to mild upper respiratory tract illness to severe lower respiratory tract disease. Further, these clinical features are also observed with several other respiratory viruses e.g. influenza and human coronaviruses (3, 5). Both HMPV and RSV infect persons across all ages but severe disease is majorly limited to infants and young [illegible] elderly, immunocompromised and persons with

cardiopulmonary co-morbidities)(4, 6, 7). Re-infection with these viruses occurs throughout life probably due to incomplete immunity that wanes over time combined with ongoing antigenic variation in viral epitopes that may support immune escape. (5, 8, 9). RSV and HMPV seasonal distributions are observed to frequently overlap. In temperate climatic regions they tend to peak in cold seasons while in the tropics the association of peaks transmission months and weather patterns is less clearly defined(10). Transmission of these two viruses is primarily via direct inhalation of infected droplets or indirect via fomites (contaminated objects or surfaces) contacts (11, 12).

The two viruses belong to *Pneumoviridae* family and share a number of genomic features (12). HMPV genome is about 13 kb encoding eight genes (3N– P– M– F– $M2$– SH– G– $L5$) (13) while RSV genome is about 15 kb encoding ten genes (3`NS1-NS2-N-P-M-SH-G-F-M2-L5) (14) thus HMPV has a different gene order and lacks non-structural proteins NS1 and NS2 (15). For both RSV and HMPV, the attachment glycoprotein (G) gene is the most genetically variable region across their entire genomes and is commonly used to discriminate genetic variants (16, 17). HMPV is classified into two groups, A and B, which are further classified into four subgroups, A1, A2 (group A) and B1 and B2 (group B) (13, 18). Subgroup A2 is the most genetically diverse and is further divided into sub-lineages A2a, A2b and A2c (18, 19). Similarly, RSV is classified into two groups (A and B) that are both antigenically and genetically distinct (20). The two groups are further divided into multiple genotypes based on nucleotide differences within the RSV G gene (21). Epidemiological studies have shown that multiple HMPV and RSV subtype/genotypes can co-circulate during epidemics both locally and globally, implying fast and widespread dispersal HMPV and RSV variants once they arise (22, 23). The dominant subgroup/genotype can also vary based on year and location (23, 24).

It is unclear whether HMPV and RSV share geographic spread patterns. Although this can be investigated using sequence crucial data, such data is scarce and there is strongly asynchronous sampling in time and space, especially in Africa (25). As a result, the origins and interconnectedness of RSV and HMPV epidemics across many global locations including Africa is not well understood. Integrating pathogen sequence data with other data e.g. spatial-temporal data allows reconstruction of transmission histories necessary for tracing of epidemiological linkages especially when there is limited case surveillance and tracing (26, 27). Both HMPV and RSV are undergoing continuous genetic sequence evolution leading to occasional emergence of novel genotypes (28–30) thus understanding their geographic spread could help inform interventions in future epidemics. Here, we report comparative phylogenetic analysis of HMPV and RSV sequence data collected between 2011–2014 across five African countries (Kenya, Mali, Gambia, South Africa and Zambia) located in different sub-regions (East, West and South). Our study provides an initial view of RSV and HMPV phylogeography across Africa detailing their overall spatial-temporal transmission patterns within the continent in relation to the rest of the world.

Materials And Methods

Study samples. The analyzed HMPV and RSV positive samples were identified during the Pneumonia study (2)(31, 32) conducted between August 2011 and

January 2014 from 5 African countries (The Gambia, Zambia, Mali, South Africa and Kenya) (Table 1). A single hospital site, backed by well-defined catchment areas of known population size, was selected in each country (31). Site characteristics for each country are reported in (31, 32). Cases (hospital admissions) and controls (persons attending outpatient facilities for mild illness or vaccination) were selected within the defined catchment areas. Cases included children aged between 28 days and 59 months with severe or very severe pneumonia (31, 32). Controls were randomly enrolled regardless of the respiratory symptoms and matched to cases by location and age group (1 to < 6 months, 6 to < 12 months, 12 to < 24 months, and 24–59 months of age) (2, 32). Written informed consent was obtained from the parent or a guardian.

The present study was approved by the KEMRI-Scientific and Ethical Review Unit (SERU# 3443) and the PERCH Committee (http://www.jhsph.edu/ivac/resources/perch-background-and-methods/).

Laboratory methods. Viral RNA was extracted using QIAamp Viral RNA Minikit (Qiagen, Germany) following the manufacturer's instructions. Reverse transcription and PCR amplification of the HMPV and RSV G genes followed protocols that have been reported elsewhere (23, 33). The amplified fragments were sequenced in both forward and reverse strands using the BigDye Terminator v1.3 chemistry on ABI 3130xl. The sequenced fragments were edited and assembled using Sequencher v5.4.6 (Gene Codes Corporation). The GenBank accession numbers of the sequences generated and the contemporaneous sequences retrieved from GenBank are listed in Additional file 1.

Sequence analysis. Sequences were aligned using MAFFT v7.407 (34) and manually curated in AliView v1.26 (35). Pairwise genetic distances were calculated in MEGA v7. 0.2. software (36) under the Maximum Composite Likelihood method to assess the genetic diversity between sequences within the groups.

Phylogenetic and phylogeographic analysis. The best fitting nucleotide substitution and site heterogeneity models were determined using ModelFinder (37) in IQ-TREE v1.6.11 (38). Phylogenetic trees were constructed using the Maximum Likelihood (ML) approach in IQ-TREE v1.6.11. Branch support was evaluated by bootstrapping. HMPV and RSV subgroups/genotypes were confirmed if sequences clustered with reference sequences of HMPV (18, 19, 39) and RSV (29, 40, 41) genotypes.

Phylogeographic analyses were done in BEAST v1.10.4 (42). First, preliminary analysis was done to test for temporal signal and to identify the best coalescent demographic model. The temporal signal in the sequence data i.e. a root-to-tip divergence of genetic distance against the year of sampling was assessed using TempEst software v1.5.3 (43). Four coalescent tree priors i.e. constant population size, exponential growth, Bayesian skyline plot and Bayesian Gaussian Markov Random Field (GMRF) skyride plot were tested under an uncorrelated lognormal relaxed molecular clock, and path sampling and stepping-stone analyses carried out to estimate marginal likelihoods (44). The marginal likelihood measures the average fit of a model to the data (42). Lower marginal likelihoods indicated weak evidence against the competing model. The Markov Chain Monte Carlo (MCMC) chains convergence (effective sample size

[ESS] > 200) were evaluated in TRACER v1.7.1 (45). The best combination of uncorrelated lognormal relaxed molecular and demographic models was selected for subsequent analysis.

A discrete trait representing geographical location was assigned to each sequence: Western Africa (Mali, Gambia), Eastern Africa (Kenya) and Southern Africa (South Africa and Zambia). To attain high spatial resolution, the country of sampling was also assigned to sequences. Viral dispersal patterns between locations were inferred using the Bayesian symmetric discrete trait evolution model with Bayesian stochastic search variable selection (BSSVS) procedure, implemented in BEAST v1.10.4. software. The symmetric diffusion model infers ancestral reconstruction using the standard continuous-time Markov chain (CTMC), in which the transition rates between locations are reversible (26). MCMC chains were run for at least 200 million generations sampling every 20,000 steps.

Contemporaneous sequences with known collection date were retrieved from GenBank **(Additional file 1)** for global phylogeographic analysis. Due to the scarcity of HMPV G gene sequences, 714 sequences collected from 20 countries between 2000 and 2018 were included. For RSV, sequences collected a year before (2010) and after (2015) our study were analysed to place our data into immediate context. A total of 1810 sequences from 28 different countries were retrieved **(Additional file 1)**. Both the continent and the country of sampling were assigned sequences as discrete traits. The global phylogeographic analysis was carried out under the symmetric diffusion model. The BEAST trees were summarised using Tree annotator v2.6.0 (46) after the removal of 10% burn-in. Maximum clade credibility trees (MCC) were visualized in FigTree v1.4.4. (http://tree.bio.ed.ac.uk/software/figtree/). Significant migration events between discrete locations were determined using the Bayes factor (BF) (44) and summarized using SpreaD3 v0.9.7.1 software (47) after discarding 10% burn-in. BF ≥ 1000 indicated very strong support, 10 ≤ BF ≤ 1000 strong support, and 3 ≤ BF ≤ 10 supported viral migration pathway.

Results

HMPV subtyping and subgroup temporal patterns. In total, 232 HMPV G gene sequences were obtained of which 44% (102/232) belonged to subgroup A2 and further clustered into sub-lineages A2a 18% (18/102), A2b 34% (35/102) and A2c 49% (48/102) **(Additional file 2)**. There were no subgroup A1 viruses. Among the sequenced HPMV strains, 56% (130/232) belonged to HMPV group B, of which 82% (107/130) and 18% (23/130) were subgroup B1 and B2, respectively. Multiple subgroups co-circulated in each country (Fig. 1, **panel a**). Notably, A2a viruses were only identified in South Africa and Zambia. HMPV subgroup temporal patterns in Mali mirrored those in The Gambia (Fig. 1, panel a).

HMPV Intra-country genetic diversity. Only subgroup B1 viruses were detected in high frequencies in all the five countries and were analysed for intra-country diversity (Table 2). ML trees were reconstructed independently for each country. At least two well supported (bootstrap value > 95%) phylogenetic clades were observed in each country **(Additional file 3)**. Sequences from different within-country sampling locations were mixed within the phylogenetic clusters suggesting rapid spread of HMPV variants within each country. Sequences from cases and controls were mixed within the clades (**Additional file 3**).

HMPV spatial origins and dispersal patterns in Africa. B1 sequences clustered into three major phylogenetic clades, numbered b1C1 to b1C3 (Fig. 2, panel b). Clade b1C2 contained sequences predominantly from South Africa. Sequences from the same geographical region, i.e. West Africa (Mali and Gambia), East Africa (Kenya) and Southern Africa (South Africa and Zambia) closely clustered together (Fig. 2, panel b). On the global MCC tree the three clades (b1C1, b1C2 and b1C3) were placed into two major clades alongside global sequences (Fig. 2, panel a). Clades b1C1 and b1C2 fell into the same clade interspersed with global sequences, and appeared discretely from clade b1C3, suggesting that at least two distinct B1 variants were in circulation (Fig. 2, panel a). The two variants reflect the genetic clusters that were observed on country-specific ML phylogenies above **(Additional file 3)**. Clade b1C1 and b1C2 clustered closely with sequences from Spain and Canada and Malaysia. Clade b1C3 clustered closely with sequences from Malaysia.

Consistent with B1 MCC phylogenies of A2b, B2 and A2c African sequences showed at least two circulating variants for each subgroup (Fig. 3, **Additional files 4 and 5, respectively**). Sequences from South Africa and Zambia clustered together. Similarly, sequences from Gambia and Mali clustered more closely among themselves, indicating an epidemiological linkage between neighbouring countries and separate introductions of HMPV variants in Africa. For A2b, sequences clustered into three well supported (posterior probability > 95%) major clades, numbered A2bC1 to A2bC3 (Fig. 3, panel b). On the global phylogeny, the three clades clustered separately interspersed with global sequences (Fig. 3, panel a). Clade A2bC2 and A2bC3 were exclusively made of Kenyan sequences and clustered closely with sequences from Canada (Fig. 3, panel a). Clade A2bC1 contained sequences from South Africa and Zambia and clustered closely with sequences from Peru (Fig. 3, panel a). African B2 viruses clustered at least into two major clades (B2C1 and B2C1) and were placed separately in context of other global B2 sequences suggesting multiple introductions into Africa **(Additional file 4)**. The two clades (B2C1 and B2C1) clustered closely with sequences from Malaysia **(Additional file 4)**. In the A2c time resolved phylogeny, African sequences clustered into two major well supported clades (numbered A2cC1 to A2cC2), see Additional file 5. Clade A2cC1 contained only sequences Zambia and South Africa. On the global A2c MCC tree, the two clades were placed in the same clade and were interspersed with sequences from Spain and Malaysia **(Additional file 5)**. For A2a viruses, African sequences were placed into a single monophyletic clade indicating a single introduction **(Additional file 6)**. Notably, A2a sequences were only detected in Zambia and South Africa and clustered closely with sequences from Peru. On the B1 (Fig. 2) and A2c **(Additional file 5)** phylogenies, although sequences from Africa were interspersed with global sequences, they mostly clustered together. Of note, 81% (178/228) of B1 and 71% of A2c (165/232) sequences were from Africa and Asia, making it difficult to assess viral introductions from unsampled locations. It is also possible that B1 and A2c viruses could have been prevalent to Africa and Asia.

Our phylogeographic analysis indicated global movement of HMPV variants. Very strong (BF > 1000, posterior probability > 95%) and strongly supported (BF > 10, posterior probability > 95%) migration pathways were indicated between different regions globally **(Additional file 7)**. Only migration pathways with BF > 10 and posterior probability > 70% are shown **(Additional file 7)**. Strong connections between

RSV subtyping and subgroup temporal patterns. Based on the G gene phylogeny **(Additional file 2)**, there were 509/842 (60%) RSVA and 333/842 (40%) RSVB sequences. All RSV B sequences belonged to the genotype BA. Among RSV A, 32% (163/509) were genotype ON1, and 68% (346/509) were genotype GA2. Similar to HMPV, multiple RSV genetic groups co-circulated within epidemics (Fig. 1, panel b). Similar genotype dominance patterns were observed between Mali and Gambia, South Africa and Zambia, and were all different from Kenya (Fig. 1, panel b).

RSV intra country diversity. To assess within-country genetic diversity, RSV BA genotype sequences were selected (Table 2). ML trees were reconstructed for each of the five African countries. From the country-specific phylogenies, sequences from the different within-country sampling locations were mixed within the phylogenetic clusters suggesting rapid spread movement of RSV variants within each country **(Additional file 8)**. Similarly, the RSV G gene sequences did not cluster by case or control status of the sampled individuals.

RSV spatial patterns and Origins in Africa. RSV phylogeographic analysis revealed markedly similar spatial patterns to those of HMPV. On the continental scale (Africa), geographical clustering was evident, and multiple variants of each RSV genotype were detected **(**Figs. 4 and 5**)**. The inferred continental migration pathways indicated very strongly supported links between neighbouring countries (BF > 1000, posterior probability > 95%) i.e., between The Gambia and Mali, and between South Africa and Zambia **(Additional file 9)**. We further explored the RSV spatial patterns globally to elucidate on the viral introductions into Africa. Only RSV genotype ON1 were analysed. African sequences fell into two major clades (numbered C1 to C2, Fig. 6) interspersed with global sequences suggesting at least two distinct variants of RSV ON1 circulated in each of the five African countries. Although the clades C1 and C2 were interspersed with global sequences, high sequence similarity (99%) was observed among them indicating widespread movement of similar variants globally. Of the two African clades (Fig. 6), clade C1 clustered closely with sequences from USA and clade C2 clustered closely to sequences from Spain.

The ON1 global phylogeographic reconstructions indicated viral movements globally although significant links (BF > 100, posterior probability > 95%) were inferred between only a number of pair of regions **(Additional file 7)**. Only locations with migration links of BF > 100 and posterior probability > 70% are reported. Strong connections were still evident between Gambia and Mali, South Africa and Zambia suggesting stronger epidemiological links between close African regions.

Discussion

Our comparative analysis revealed markedly similar patterns of spread of HMPV and RSV in Africa. Geographical clustering of sequences by sub-region was evident suggesting high sequence relatedness between neighbouring countries and separate introductions of HMPV and RSV variants into continental Africa. This clustering indicates predominant local transmission and a common source of introduction between neighbouring countries. Within each country, sequences from the different catchment areas were mixed within the phylogenetic clusters, which indicates that following an introduction into the country,

there is a rapid movement of HMPV and RSV variants between locations in a relatively short time followed by local diversification. However, we cannot ignore the fact that only a single site was sampled in each country. Therefore, we may not have characterised all locally circulating strains. At least two distinct variants of the various genetic groups were observed in each country, indicating multiple importations from the global pool. These results are not unique to only HMPV and RSV as similar findings have been reported by early reports on SARS-CoV-2 transmission in Kenya, Uganda and South Africa (48, 49).

HMPV and RSV epidemics were characterised by co-circulation of multiple genotypes. Genotype circulation patterns were similar between neighboring African countries (South Africa and Zambia, and Mali and The Gambia), indicative of the epidemiological linkage between neighbouring African countries and the independent introduction of multiple HMPV and RSV variants into Africa sub-regions from the global pool. South Africa and Zambia HMPV genotype patterns were characterised by a unique circulation of HMPV A2a viruses, which were not detected in the other study sites. On the global phylogenies, clustering of HMPV sequences by African sub-region was still evident, suggestive of a common source of virus introduction between neighbouring African countries and wide spread movement of similar HMPV variants. Notably, on the global HMPV A2c and RSV ON1 MCC trees, African sequences clustered more closely together and could also indicate common sources of introduction into Africa and further local diversification.

Previous studies of HMPV (50) and RSV (51) done in Argentina; reveal the two viruses' dispersal patterns occur both locally and globally. Similar findings have been reported for influenza viruses in Asia (52) and the USA (53). Air travel has been shown to be the dominant determinant of influenza H3N2 and H1N1 viruses on the global scale (53, 54). However, on smaller geographic scales, factors such demography, other forms of mobility, geographical proximity, etc. can be significant predictors of spatial spread (51) (53). The spatial diffusion pathways of HMPV and RSV revealed strong connections between countries in the same African sub-region and weak links between distant locations. Overall, the patterns of spread of HMPV and RSV observed in this study may reflect underlying host mobility patterns. In particular, Africa experiences separate introduction of HMPV and RSV variants from the global pool influenced by human mobility patterns. Following a virus introduction, there is an establishment of a local epidemic in countries proximal to each other due to more interactions, associated with predominant migration between neighbouring countries (55), as a result of environmental and socioeconomic factors such as distribution of ethnic groups, colonial and regional trade ties (55). Recent reports on the role of long-distance truck drivers from neighbouring countries on the spread of COVID-19 in Uganda underscores these links between neighbouring countries (49). We acknowledge that due to biased sampling, we did not assess possible introductions from unsampled locations. More analysis will be required to test the contribution of human mobility and other potential predictors on the spatial spread to explore the patterns further.

Strongly supported links were identified between Africa and other countries globally (Table 3). These links ction of HMPV and RSV variants into Africa. Canada, Peru,

Malaysia, USA and Spain were determined as the most probable ancestral locations to African strains. However, due to disproportionate sampling, our ancestral location estimates could be biased because the discrete trait analysis is inherently biased by the sampling intensities of locations (56) (57). It was also difficult to pinpoint the source-sinks for HMPV and RSV epidemics. To pin-point the source-sinks, more representative sampling will be required globally. On the MCC phylogenies, the basal branches were occupied by sequences sampled outside Africa. This shows ancestral strains may reside in regions outside Africa or from unsampled locations. Future analysis of other RSV genotypes (GA2 and BA) that were detected in this study and not analysed will help to validate our inferences on the potential sources of introductions of HMPV and RSV variants into Africa.

Although our analysis was based on a modest sample size (HMPV n = 232 and RSV n = 842 sequences), this did not hinder our ability to assess sequence relatedness and infer spatial-temporal spread of HMPV and RSV in Africa. Also, sequences were collected simultaneously over two years and allowed exploration of the spatial patterns to assess possible epidemiological linkages between Kenya, Mali, Gambia, South Africa, and Zambia. Conversely, we did not assess possible epidemiological links from unsampled locations in Africa. Future studies across different countries in different Africa sub-regions (East, West, South, Central and North) will be necessary for tracing transmission patterns of HMPV and RSV in Africa. Genetic clusters containing similar sequences, especially within-country clusters, will require whole-genome sequencing for increased resolution and detailed transmission studies.

Conclusions

In conclusion, our study provides the first contemporaneous HMPV and RSV sequences across 5 African countries, acting as a significant reference for future work. HMPV and RSV molecular epidemiological patterns were consistent across the study locations in the continent. Multiple strains can co-circulate, and distinct strains can circulate in different Africa sub-regions at the same time. The occurrence of strong regional links suggested that local, tailored public health intervention measures should be considered. By comparing the strain epidemiology geographic patterns of HMPV and RSV across Africa, and our study illuminates on the spread characteristics of two seasonally recurring respiratory viruses.

Abbreviations

HMPV	Human metapneumovirus
RSV	Respiratory syncytial virus
ALRTI	Acute lower respiratory tract infection
tMRCA	Time to the most recent common ancestor
PERCH	Pneumonia Etiology Research for Child Health
ESS	Effective sample size
KML	Keyhole markup language
BSSVS	Bayesian Stochastic Search Variable Selection

Declarations

Ethics approval and consent to participate

The Kenya Medical Research Institute Scientific and Ethics Review Unit (SERU) approved the study. A written informed consent was obtained from participant's parent or guardian.

Consent for publication

Not applicable.

Availability of data and materials

The replication data set for this manuscript are available from the Harvard Dataverse under the DOI: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/POHLE3. Data access can be requested from the KEMRI-Wellcome Trust Research Programme, Data Governance Committee (dgc@kemri-wellcome.org). Publicly accessible data are included in this published article [Additional file 1].

Competing interests

The authors declare that they have no competing interest.

Funding

The current study was supported by the Wellcome Trust [grant 102975]. PERCH study was supported by grant 48968 from The Bill & Melinda Gates Foundation to the International Vaccine Access Center, Department of International Health, Johns Hopkins Bloomberg School of Public Health.

Authors' contributions

DJN and CNA: conceived and designed the study. DJN, CNA, EI and JRO supervised the work and gave technical guidance. JOW, CL and AM: performed lab work. JOW, JRO, EK: participated in data curation, sequence and Phylogenetic analyses. JOW, EK, DJN and CAN: wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank PERCH team at Johns Hopkins for permission to make use of samples (http://www.jhsph.edu/research/centers-and-institutes/ivac/projects/perch/). We thank the PERCH Study Group for sample collection and for allowing access to the samples. We are grateful to study participants for providing samples and data. We also thank all the staff that were involved in participant recruitment and laboratory processing of samples. Many thanks to Eastern Africa Network for Bioinformatics Training (EANBIT) and Initiative to Develop African Research Leaders (IDeAL) for their collaborative sponsorship. This work is published with permission of director KEMRI.

References

1. Shi T, McLean K, Campbell H, Nair H. Aetiological role of common respiratory viruses in acute lower respiratory infections in children under five years: a systematic review and meta–analysis. Journal of global health. 2015 Jun;5(1)
2. O'Brien KL, Baggett HC, Brooks WA, Feikin DR, Hammitt LL, Higdon MM, et al. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. Lancet. 2019;394(10200):757–79.
3. Moe N, Krokstad S, Stenseng IH, Christensen A, Skanke LH, Risnes KR, et al. Comparing human metapneumovirus and respiratory syncytial virus: Viral codetections, genotypes and risk factors for severe disease. PLoS One. 2017;12(1):1–19.
4. Schildgen V, van den Hoogen B, Fouchier R, Tripp RA, Alvarez R, Manoha C, et al. Human metapneumovirus: Lessons learned over the first decade. Clin Microbiol Rev. 2011;24(4):734–54.
5. Williams J, Shafagati N. Human metapneumovirus - what we know now. F1000Research. 2018;7(0):1–11.
6. Shi T, Arnott A, Semogas I, Falsey AR, Openshaw P, Wedzicha JA, Campbell H, Nair H. The etiological role of common respiratory viruses in acute respiratory infections in older adults: a systematic review and meta-analysis. The Journal of infectious diseases. 2020 Nov 1;222(Supplement_7):S563-9.
7. Shi T, McAllister DA, O'Brien KL, Simoes EAF, Madhi SA, Gessner BD, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. Lancet. 2017;390(10098):946–58.
8. Glezen WP, Taber LH, Frank AL, Kasel JA. Risk of primary infection and reinfection with respiratory syncytial virus. American journal of diseases of children. 1986 Jun 1;140(6):543-6.

9. Pelletier G, Déry P, Abed Y, Boivin G. Respiratory tract reinfections by the new human Metapneumovirus in an immunocompromised child. Emerg Infect Dis. 2002;8(9):976–8.
10. Li Y, Reeves RM, Wang X, Bassat Q, Brooks WA, Cohen C, Moore DP, Nunes M, Rath B, Campbell H, Nair H. Global patterns in monthly activity of influenza virus, respiratory syncytial virus, parainfluenza virus, and metapneumovirus: a systematic analysis. The Lancet Global Health. 2019 Aug 1;7(8):e1031-45.
11. Hall CB, Douglas RG, Schnabel KC, Geiman JM. Infectivity of respiratory syncytial virus by various routes of inoculation. Infect Immun. 1981;33(3):779–83.
12. Kahn JS. Epidemiology of human metapneumovirus. Clin Microbiol Rev. 2006;19(3):546–57.
13. Kim J Il, Park S, Lee I, Park KS, Kwak EJ, Moon KM, et al. Genome-wide analysis of human metapneumovirus evolution. PLoS One [Internet]. 2016;11(4):1–19. Available from: http://dx.doi.org/10.1371/journal.pone.0152962
14. Agoti CN, , Otieno JR, Munywoki PK, Mwihuri AG, Cane PA, Nokes DJ, et al. Local Evolutionary Patterns of Human Respiratory Syncytial Virus Derived from Whole-Genome Sequencing. J Virol. 2015;89(7):3444–54.
15. Van Den Hoogen BG, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Analysis of the genomic sequence of a human metapneumovirus. Virology. 2002;295(1):119–32.
16. Bastien N, Liu L, Ward D, Taylor T, Li Y. Genetic variability of the G glycoprotein gene of human metapneumovirus. J Clin Microbiol. 2004;42(8):3532–7.
17. Johnson PR, Spriggs MK, Olmsted RA, Collins PL. The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins. Proceedings of the National Academy of Sciences. 1987 Aug 1;84(16):5625-9.
18. Huck B, Scharf G, Neumann-Haefelin D, Puppe W, Weigl J, Falcone V. Novel human metapneumovirus sublineage. Emerg Infect Dis. 2006;12(1):147–50.
19. Jagušić M, Slović A, Ljubin-Sternak S, Mlinarić-Galinović G, Forčić D. Genetic diversity of human metapneumovirus in hospitalized children with acute respiratory infections in Croatia. J Med Virol. 2017;89(11):1885–93.
20. Melero JA, García-Barreno B, Martínez I, Pringle CR, Cane PA. Antigenic structure, evolution and immunobiology of human respiratory syncytial virus attachment (G) protein. Journal of General Virology. 1997.
21. Goya S, Galiano M, Nauwelaers I, Trento A, Openshaw PJ, Mistchenko AS, Zambon M, Viegas M. Toward unified molecular surveillance of RSV: A proposal for genotype definition. Influenza and Other Respiratory Viruses. 2020 May;14(3):274-85.
22. Agoti CN, Otieno JR, Ngama M, Mwihuri AG, Medley GF, Cane PA, et al. Successive Respiratory Syncytial Virus Epidemics in Local Populations Arise from Multiple Variant Introductions, Providing Insights into Virus Persistence. J Virol. 2015;89(22):11630–42.
23. Oketch JW, Kamau E, Otieno GP, Otieno JR, Agoti CN, Nokes DJ. Human metapneumovirus ersistence identified through surveillance of pediatric

pneumonia hospital admissions in coastal Kenya, 2007–2016. BMC Infect Dis. 2019;19(1):1–13.

24. Sullender WM. Respiratory syncytial virus genetic and antigenic diversity. Clin Microbiol Rev. 2000;13(1):1–15.
25. Bose ME, He J, Shrivastava S, Nelson MI, Bera J, Halpin RA, et al. Sequencing and analysis of globally obtained human respiratory syncytial virus a and B genomes. PLoS One. 2015;10(3):1–22.
26. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009;5(9).
27. Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, Park DJ, Ladner JT, Arias A, Asogun D, Bielejec F. Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature. 2017 Apr;544(7650):309-15.
28. Saikusa M, Nao N, Kawakami C, Usuku S, Sasao T, Toyozawa T, Takeda M, Okubo I. A novel 111-nucleotide duplication in the G gene of human metapneumovirus. Microbiology and immunology. 2017 Nov;61(11):507-12.
29. Eshaghi A, Duvvuri VR, Lai R, Nadarajah JT, Li A, Patel SN, Low DE, Gubbay JB. Genetic variability of human respiratory syncytial virus A strains circulating in Ontario: a novel genotype with a 72 nucleotide G gene duplication. PloS one. 2012 Mar 28;7(3):e32807.
30. Trento A, Galiano M, Videla C, Carballal G, García-Barreno B, Melero JA, Palomo C. Major changes in the G protein of human respiratory syncytial virus isolates introduced by a duplication of 60 nucleotides. Journal of General Virology. 2003 Nov 1;84(11):3115-20
31. Levine OS, O'Brien KL, Deloria-Knoll M, Murdoch DR, Feikin DR, DeLuca AN, Driscoll AJ, Baggett HC, Brooks WA, Howie SR, Kotloff KL. The Pneumonia Etiology Research for Child Health Project: a 21st century childhood pneumonia etiology study. Clinical infectious diseases. 2012 Apr 1;54(suppl_2):S93-101.
32. Deloria-Knoll M, Feikin DR, Scott JA, O'Brien KL, DeLuca AN, Driscoll AJ, Levine OS, Pneumonia Methods Working Group. Identification and selection of cases and controls in the Pneumonia Etiology Research for Child Health project. Clinical infectious diseases. 2012 Apr 1;54(suppl_2):S117-23.
33. Agoti CN, Mwihuri AG, Sande CJ, Onyango CO, Medley GF, Cane PA, et al. Genetic relatedness of infecting and reinfecting respiratory syncytial virus strains identified in a birth cohort from rural Kenya. J Infect Dis. 2012;206(10):1532–41.
34. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
35. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014 Nov 15;30(22):3276-8.
36. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular biology and evolution. 2016 Mar 22;33(7):187
37. Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermiin LS. ModelFinder: fast model mates. Nature methods. 2017 Jun;14(6):587-9.

38. Nguyen, Tung, Lam, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
39. de Graaf M, Osterhaus ADME, Fouchier RAM, Holmes EC. Evolutionary dynamics of human and avian metapneumoviruses. J Gen Virol. 2008;89(12):2933–42.
40. Peret TC, Hall CB, Schnabel KC, Golub JA, Anderson LJ. Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community. Journal of General Virology. 1998 Sep 1;79(9):2221-9.
41. Venter M, Madhi SA, Tiemessen CT, Schoub BD. Genetic diversity and molecular epidemiology of respiratory syncytial virus over four consecutive seasons in South Africa: identification of new subgroup A and B genotypesThe GenBank accession numbers of the sequences reported in this paper are AF348802–AF348826. Journal of General Virology. 2001 Sep 1;82(9):2117-24.
42. Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Systematic biology. 2011 Mar 1;60(2):150-60.
43. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus evolution. 2016 Jan 1;2(1):vew007.
44. Lartillot N, Philippe H. Computing Bayes factors using thermodynamic integration. Systematic biology. 2006 Apr 1;55(2):195-207.
45. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic biology. 2018 Sep;67(5):901.
46. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution. 2012 Aug 1;29(8):1969-73.
47. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. Molecular biology and evolution. 2016 Apr 23;33(8):2167-9.
48. Giandhari J, Pillay S, Wilkinson E, Tegally H, Sinayskiy I, Schuld M, Lourenço J, Chimukangara B, Lessells RJ, Moosa Y, Gazy I. Early transmission of SARS-CoV-2 in South Africa: An epidemiological and phylogenetic report. medRxiv. 2020 Jan 1.
49. Bajunirwe F, Izudi J, Asiimwe S. Long-distance truck drivers and the increasing risk of COVID-19 spread in Uganda. International Journal of Infectious Diseases. 2020 Sep 1;98:191-3.
50. Velez Rueda AJ, Mistchenko AS, Viegas M. Phylogenetic and Phylodynamic Analyses of Human Metapneumovirus in Buenos Aires (Argentina) for a Three-Year Period (2009-2011). PLoS One. 2013;8(4).
51. Rojo GL, Goya S, Orellana M, Sancilio A, Perez AR, Montali C, García C, Sosa L, Musto A, Alvarez D, Castello A. Unravelling respiratory syncytial virus outbreaks in Buenos Aires, Argentina: Molecular basis of the spatio-temporal transmission. Virology. 2017 Aug 1;508:118-26.
52. Zar Htwe KT, Dapat C, Shobugawa Y, Odagiri T, Hibino A, Kondo H, et al. Phylogeographic analysis of human influenza A and B viruses in Myanmar, 2010–2015. PLoS One. 2019;14(1):2010–5.

53. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. PLoS Pathog. 2014;10(2).
54. Su YC, Bahl J, Joseph U, Butt KM, Peck HA, Koay ES, Oon LL, Barr IG, Vijaykrishna D, Smith GJ. Phylodynamics of H1N1/2009 influenza reveals the transition from host adaptation to immune-driven selection. Nature communications. 2015 Aug 6;6(1):1-3.
55. Flahaux ML, De Haas H. African migration: trends, patterns, drivers. Comparative Migration Studies. 2016 Dec 1;4(1):1
56. De Maio N, Wu CH, O'Reilly KM, Wilson D. New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation. PLoS Genet. 2015;11(8):1–22.
57. Baele G, Suchard MA, Rambaut A, Lemey P. Emerging concepts of data integration in pathogen phylodynamics. Systematic biology. 2017 Jan 1;66(1):e47-65.

Tables

Table 1: Virus positive by site and Number sequenced

A) B)

		HMPV						RSV					
			Cases		Controls				Cases		Controls		
Site	**Enrollment date**	**No. of samples**	No. of cases	No. sequenced	No. of controls	No. sequenced	Total sequenced	**No. of samples**	No. of cases	No. sequenced	No. of controls	No. sequenced	Total sequenced
Gambia	November 2011 - October 2013	**46**	37	**32**	9	**9**	**41**	**117**	113	**97**	4	**2**	**99**
Kenya	August 2011 - November 2013	**62**	50	**50**	13	**8**	**58**	**263**	251	**251**	12	**12**	**263**
Mali	January 2012 - January 2014	**46**	39	**34**	7	**6**	**40**	**182**	154	**138**	28	**20**	**158**
South Africa	August 2011 - August 2013	**77**	55	**44**	22	**14**	**58**	**260**	232	**208**	28	**22**	**230**
Zambia	October 2011 - October 2013	**47**	39	**30**	8	**5**	**35**	**112**	94	**82**	18	**10**	**92**
	Totals	**278**	**220**	**190**	**59**	**42**	**232**	**934**	**844**	**776**	**90**	**66**	**842**

Abbreviations: HMPV, human metapneumovirus; RSV, respiratory syncytial virus.

Total number of HMPV and RSV positive samples collected between August 2011 and January 2014 from 5 African countries. Panel A: Total number of HMPV sequences stratified by cases and controls, and total sequenced. Panel B: Total number of RSV sequences stratified by cases and controls, and total sequenced.

Table 2: HMPV and RSV subgroup detection patterns

HMPV subgroup detection

A)

Country	A2a	A2b	A2c	B1	B2	Total
Kenya	0	12	9	21	16	58
Gambia	0	0	12	27	2	41
Mali	0	1	7	32	0	40
South Africa	6	16	17	15	4	58
Zambia	12	6	4	12	1	35
Total	**18**	**35**	**49**	**107**	**23**	**232**

RSV subgroup detection

B)

Country	RSVA_ON1	RSVA_GA2	RSVB_BA	Total
Kenya	114	42	107	263
Gambia	2	8	89	99
Mali	5	47	106	158
South Africa	13	188	29	230
Zambia	29	61	2	92
Total	**163**	**346**	**333**	**842**

Abbreviations: HMPV, human metapneumovirus; RSV, respiratory syncytial virus.

Total number of HMPV and RSV sequences obtained from samples collected between August 2011 and January 2014 from the 5 African countries. Panel A: Total number of HMPV sequences obtained by HMPV subgroup for each study site. Panel B: Total number of RSV sequences obtained by RSV subgroup for each study site.

Table 3: Inferred locations of tMRCA of African sequences

	Origin				
A)	**HMPV subgroup**				
Region	**A2a**	**A2b**	**A2c**	**B1**	**B2**
West Africa	Not detected	*Canada (North America)	#Spain (Europe) Malaysia (Asia)	**Malaysia (Asia)	*Malaysia (Asia)
East Africa	Not detected	**Canada (North America)	# Spain (Europe) Malaysia (Asia)	**Malaysia (Asia)	*Malaysia, Nepal (Asia)
Southern Africa	*Peru (Southern America)	*Peru (South America)	#Spain (Europe) Malaysia (Asia)	*Malaysia (Asia)	#Canada (North America) *Malaysia (Asia)
B)	**RSV subgroup**				
Region	**ON1**		**GA2**		**BA**
West Africa	#Spain (Europe) USA (North America)		Not determined		Not determined
East Africa	#Spain(Europe) USA (North America)		Not determined		Not determined
Southern Africa	#Spain(Europe) USA (North America)		Not determined		Not determined

Abbreviations: HMPV, human metapneumovirus; RSV, respiratory syncytial virus; tMRCA, time to the most recent common ancestor.

Location of contemporaneous sequences that clustered with African sequences within the same clades, summarized from our phylogeographic analysis for each HMPV and RSV subgroup stratified by African sub-region. The support for the tMRCA leading to African clades was indicated as follows:

***Very strong support: BF ≥ 1000, Posterior probability ≥95%

**Strong support: 100 ≤ BF ≤ 1000

*Supported: 3 ≤ BF ≤ 100

Not supported but shared the tMRCA with African clades

Figures

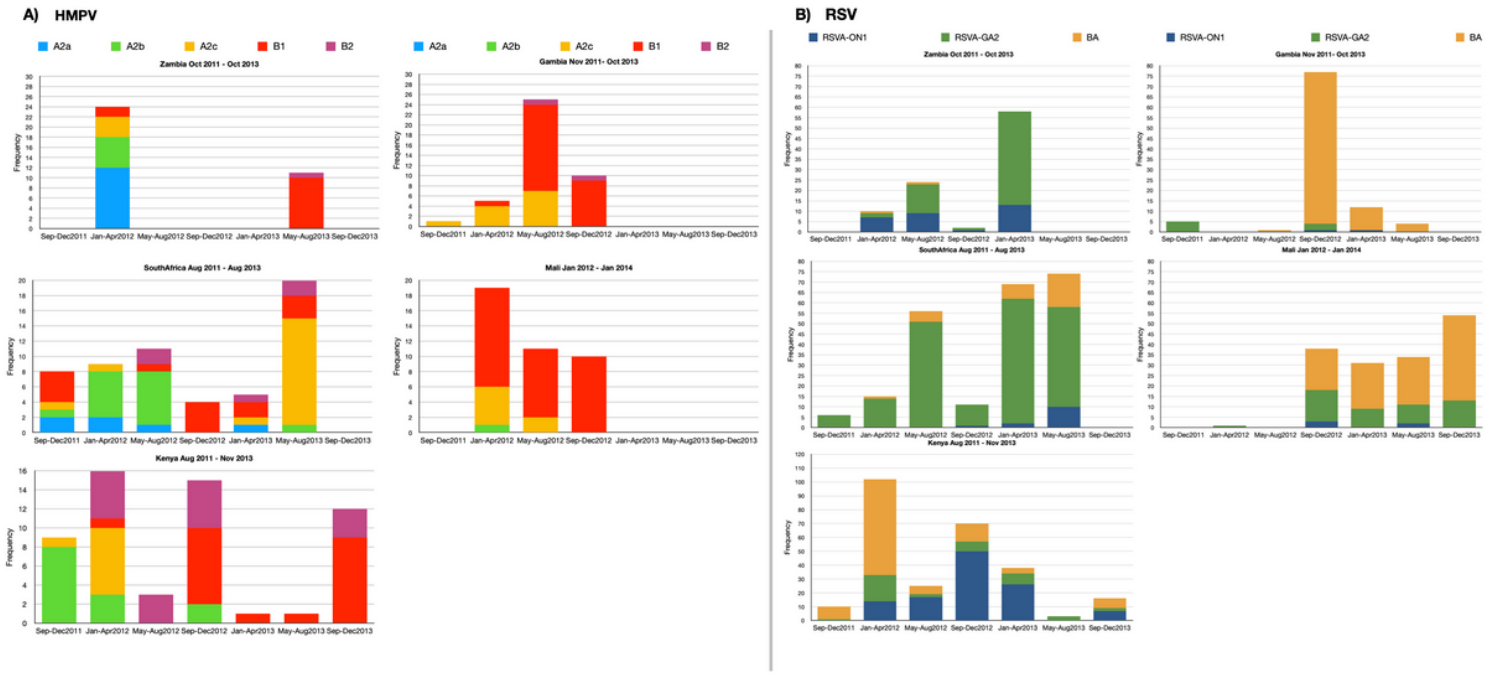

Figure 1

HMPV and RSV subgroup prevalence and temporal patterns derived from G gene sequence data collected from Kenya, Mali, Gambia, South Africa and Zambia. Panel a, HMPV temporal patterns. Panel b, RSV temporal patterns.

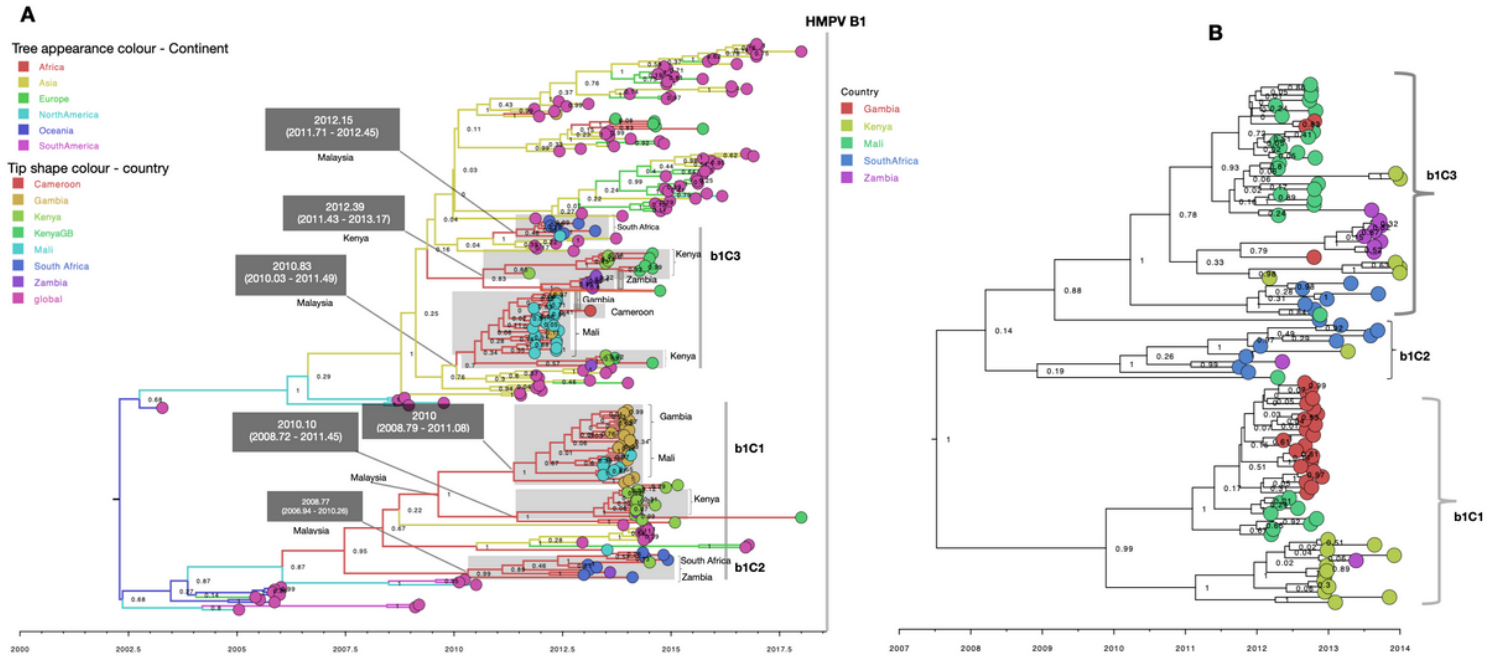

Figure 2

Panel A; Temporal scaled maximum clade credibility (MCC) tree constructed using HMPV B1 G gene sequences obtained from Africa and GenBank collected between 2000 to 2018. Branches are coloured according to the most probable location as inferred using symmetric discrete phylogeographic diffusion

model. Geographic locations considered are shown in the figure key. Sequences from Kenya, Mali, Gambia, South Africa and Zambia collected beyond the study period are indicated with a suffix GB. Posterior probabilities are shown next to nodes. Clades containing African sequences falling in monophyletic clades are highlighted in grey boxes. For each clade, the mean estimated time of the most recent common ancestor (tMRCA) and respective 95% Bayesian credible intervals are shown in a black box alongside the most probable location leading to each clade. Panel B; MCC tree of B1 sequences collected from Africa.

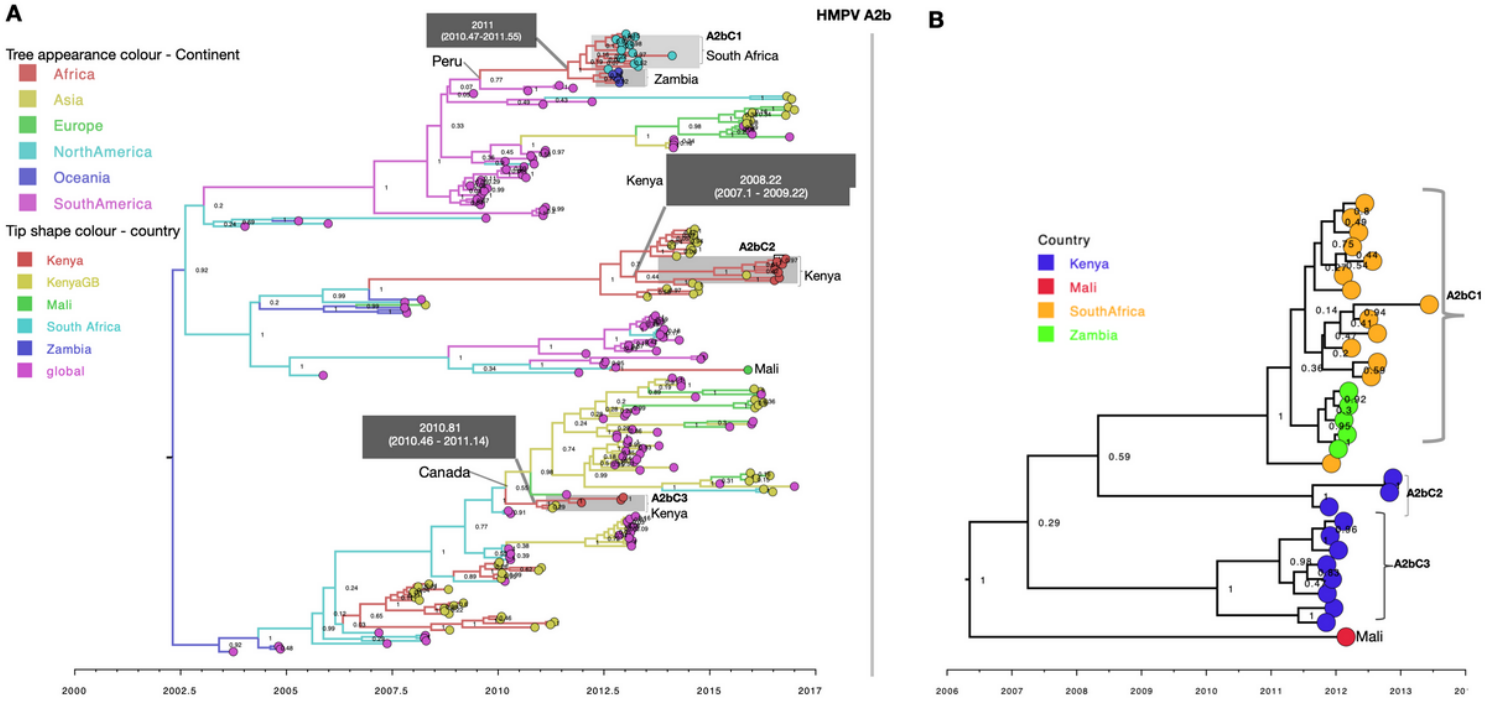

Figure 3

Panel A; Temporal scaled maximum clade credibility (MCC) tree constructed using HMPV A2b G gene sequences obtained from Africa and GenBank collected between 2000 to 2018. Branches are coloured according to the most probable location as inferred using symmetric discrete phylogeographic diffusion model. Geographic locations considered are shown in the figure key. Sequences from Kenya, Mali, Gambia, South Africa and Zambia collected beyond the study period are indicated with a suffix GB. Posterior probabilities are shown next to nodes. Clades containing African sequences falling in monophyletic clades are highlighted in grey boxes. For each clade, the mean estimated time of the most recent common ancestor (tMRCA) and respective 95% Bayesian credible intervals are shown in a black box alongside the most probable location leading to each clade. Panel B; MCC tree of A2b sequences collected from Africa.

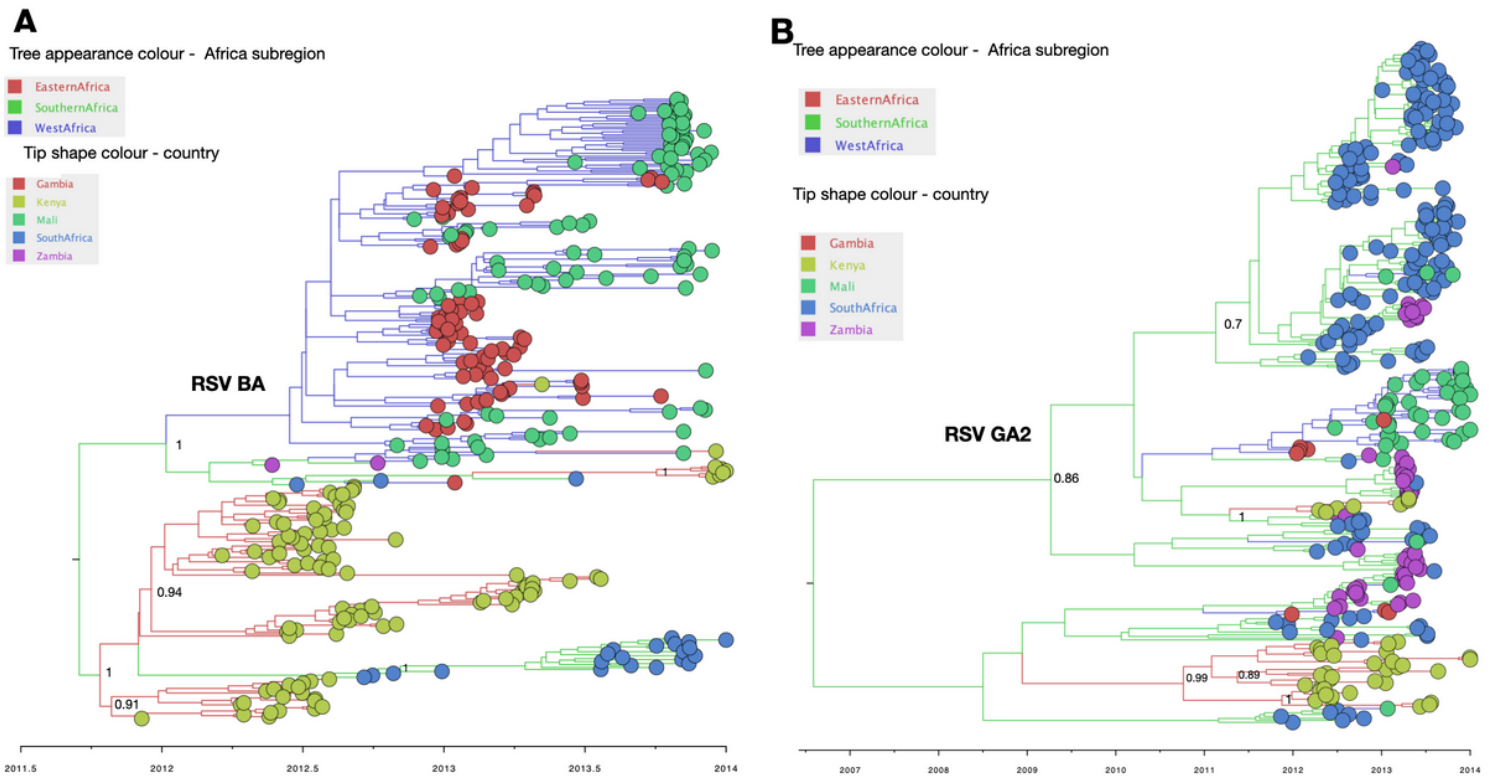

Figure 4

Temporal scaled maximum clade credibility trees constructed using RSV BA (panel a), RSV and RSV GA2 (panel b) G gene sequences obtained from Kenya, Mali, Gambia, South Africa and Zambia. Branches are coloured according to the most probable location as inferred using symmetric discrete phylogeographic diffusion model. The tips were coloured according to the country of sampling Posterior probabilities support for each node are indicated next to the nodes.

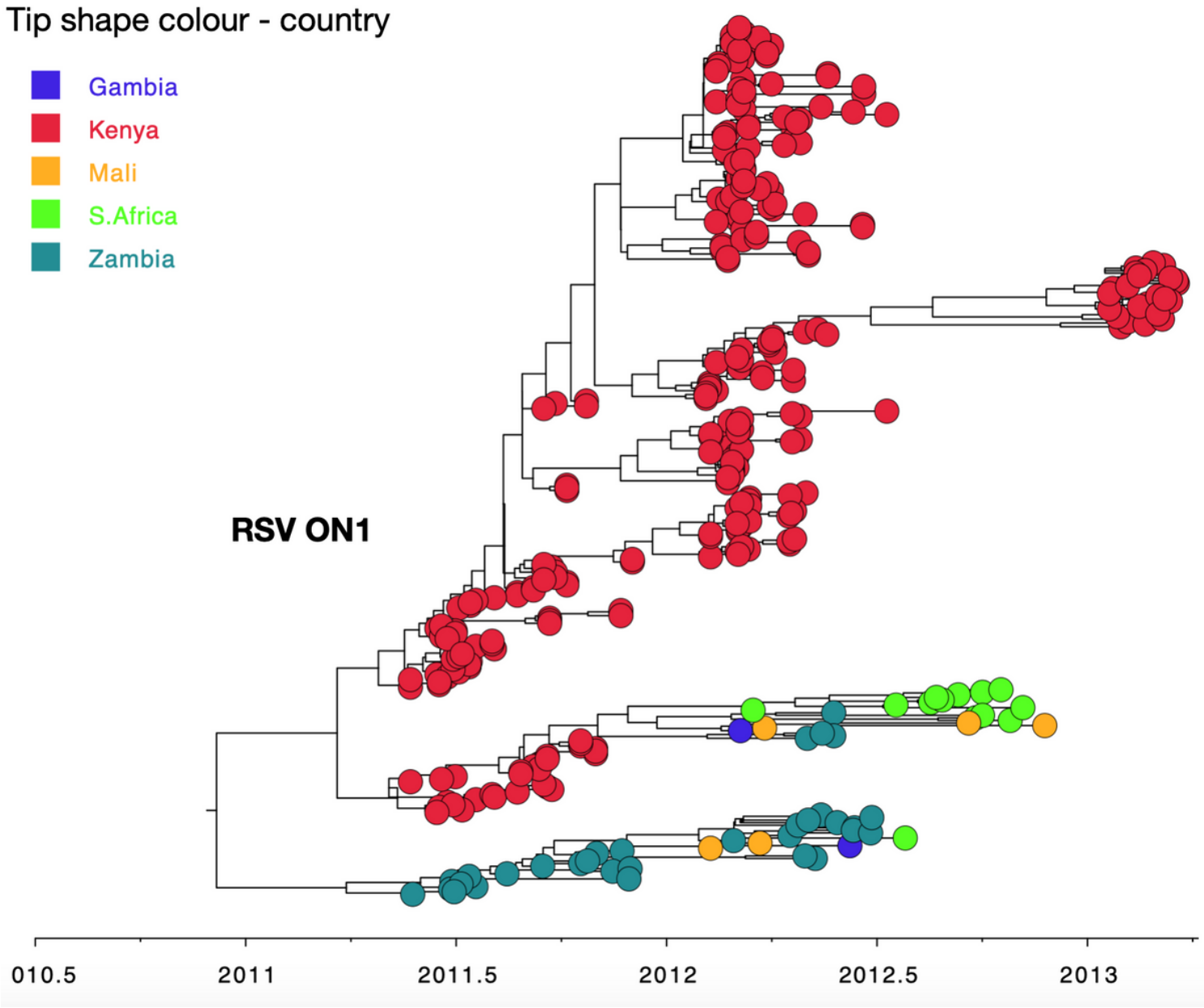

Figure 5

Temporal scaled maximum clade credibility trees constructed using RSV ON1 G gene sequences obtained from Kenya, Mali, Gambia, South Africa and Zambia. The tips were coloured according to the country of sampling.

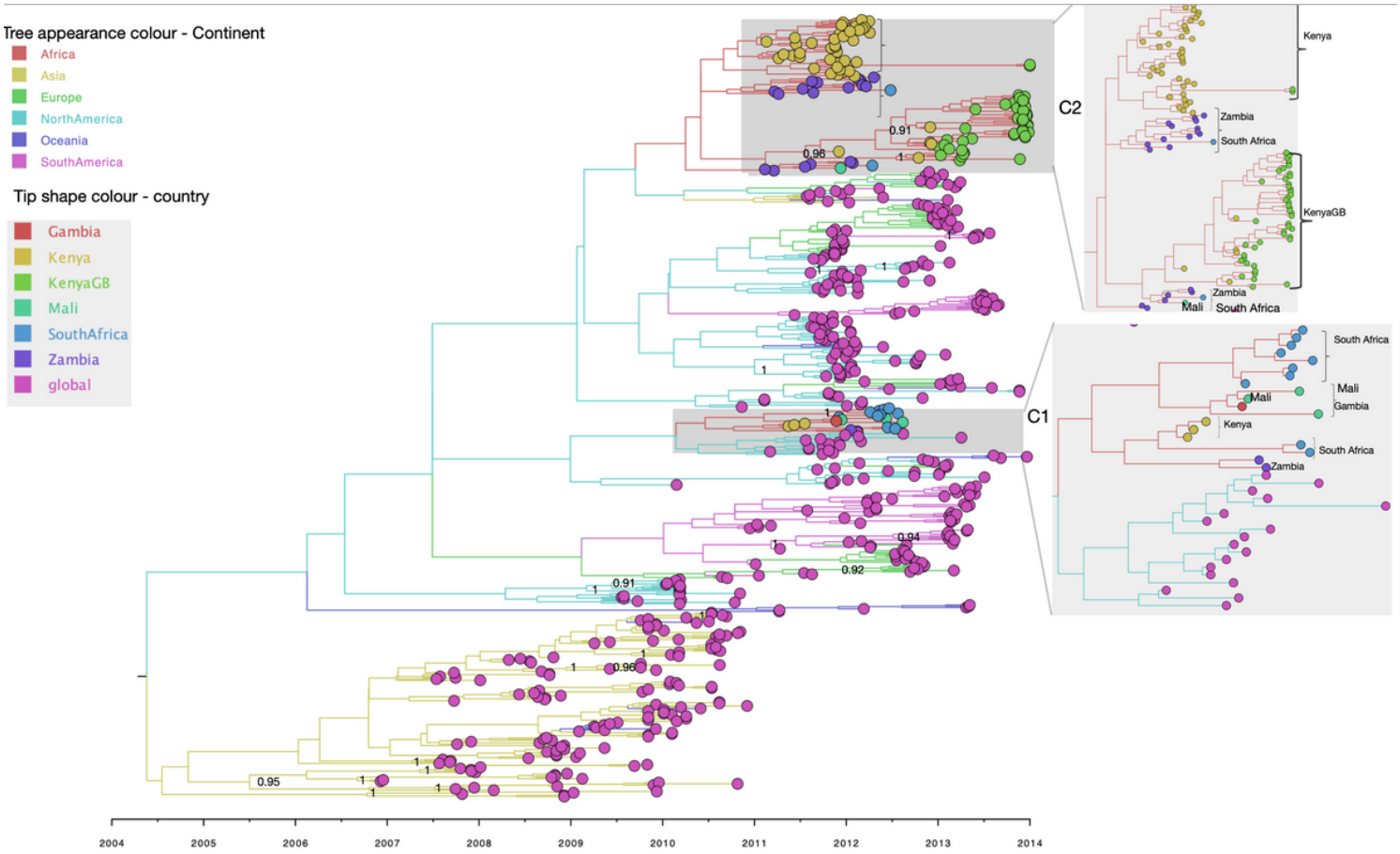

Figure 6

Temporal scaled maximum clade credibility tree constructed using RSV ON1 G gene sequences obtained from Africa and GenBank collected between 2010 to 2015. Branches are coloured according to the most probable location as inferred using symmetric discrete phylogeographic diffusion model. Geographic locations considered are shown in the figure key. Sequences from Kenya, Mali, Gambia, South Africa and Zambia collected beyond the study period are indicated with a suffix GB. Posterior probabilities are shown next to nodes. Clades containing African sequences falling in monophyletic clades are highlighted in grey boxes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.xls
- Additionalfile2.pdf
- Additionalfile3.pdf
- Additionalfile4.png
- Additionalfile5.png

- Additionalfile6.png
- Additionalfile7.png
- Additionalfile8.pdf
- Additionalfile9.png