# Epidemiological and evolutionary dynamics of influenza B virus in coastal Kenya as revealed by genomic analysis of strains sampled over a single season

Festus M. Nyasimi[1,2,#], D. Collins Owuor[1], Joyce M. Ngoi[1], Alexander G. Mwihuri[1], Grieven P. Otieno[1], James R. Otieno[1], George Githinji[1], Joyce U. Nyiro[1], D. James Nokes[1,2,3] and Charles N. Agoti[1,2]

**Author Affiliations**

*1. Epidemiology and Demography Department, Kenya Medical Research Institute (KEMRI) –Wellcome Trust Research Programme, Kilifi, Kenya.*

*2. Department of Public Health, School of Human and Health Sciences, Pwani University, Kilifi, Kenya.*

*3. School of Life Sciences and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, UK.*

**Word count:**

Abstract: 258

Main text: 5284

**[#]Corresponding author:**

Festus Nyasimi (Email: nfestus14@gmail.com)

**Key words:**

Kenya, influenza B, transmission, genomics, evolution, reassortment

**ABSTRACT (Word count: 258)**

The genomic epidemiology of influenza B virus (IBV) remains understudied in Africa despite significance to design of effective local and global control strategies. We undertook surveillance throughout 2016 in coastal Kenya, recruiting individuals presenting with acute respiratory illness at nine outpatient health facilities (any age) or admitted to the Kilifi County Hospital (<5-year-old). Whole genomes were sequenced for a select 111 positives; 94 (84.7%) of B/Victoria lineage and 17 (15.3%) of B/Yamagata lineage. Inter-Lineage reassortment was detected in 10 viruses; nine with B/Yamagata backbone but B/Victoria NA and NP segments and one with a B/Victoria backbone but B/Yamagata PB2, PB1, PA and MP segments. Five phylogenomic clusters were identified among the sequenced viruses; (i) pure B/Victoria clade 1A (n=93, 83.8%), (ii) reassortant B/Victoria clade 1A (n=1, 0.9%), (iii) pure B/Yamagata clade 2 (n=2, 1.8%), (iv) pure B/Yamagata clade 3 (n=6, 5.4%) and (v) reassortant B/Yamagata clade 3 (n=9, 8.1%). Using divergence dates and clustering patterns in the presence of global background sequences, we counted up to 29 independent IBV strain introductions into the study area (~900 $km^2$) in 2016. Local viruses, including the reassortant B/Yamagata strains, clustered closely with viruses from neighbouring Tanzania and Uganda. Our study demonstrated that genomic analysis provides a clearer picture of locally circulating IBV diversity. The high number of IBV introductions highlights the challenge in controlling local influenza epidemics by targeted approaches e.g. sub-population vaccination or patient quarantine. The finding of divergent IBV strains co-circulating within a single season emphasizes why broad immunity vaccines are the most ideal for influenza control in Kenya.

**MAIN TEXT (Word count 5284)**

Human influenza B virus (IBV) is responsible for about 30% of the influenza virus morbidity and mortality during seasonal influenza epidemics (Caini et al. 2018a; Paul Glezen et al. 2013; Seleka et al. 2017). Influenza disease burden is notably highest in low-income countries, majority of which are located in the tropics (Caini et al. 2018a; Sambala et al. 2018). Influenza virus activity in these regions tends to continue throughout the year characterized by a single or multiple epidemic peaks (Caini et al. 2018a; El Guerche-Seblain et al. 2019; Emukule et al. 2019; Hirve and Organization 2015). However, understanding of IBV evolutionary dynamics and molecular epidemiology in these regions, especially in sub-Saharan Africa, remains limited with few genomes available to date for detailed investigations (Langat et al. 2017).

Currently, there are two major IBV lineages co-circulating, B/Yamagata and B/Victoria, which diverged in the early 1970s (Kanegae et al. 1990; McCullers et al. 2004; Rota et al. 1990). Differences between the two lineages are seen in their transmissibility, and genetic and antigenic dynamics (Langat et al. 2017). For instance, B/Victoria lineage viruses have been shown to infect children more commonly than B/Yamagata lineage viruses (Tan et al. 2013; Vijaykrishna et al. 2015; Xu et al. 2015). Further, the B/Victoria lineage viruses display a clear antigenic drift of a single clade in successive years with strong seasonal fluctuations in their incidence, while the B/Yamagata lineage viruses exhibit continuous co-circulation of multiple genetic clades which alternate in their dominance over years (Langat et al. 2017).

Human influenza vaccines that include one or both the two known IBV lineages are currently available. Because of the continuous antigenic evolution inherent in influenza viruses, these vaccines are periodically updated in their antigenic composition (Bedford et al. 2014). Understanding of the prevailing global and local influenza molecular epidemiology is

a key consideration during selection of influenza strains to include in vaccines for upcoming seasons and for understanding observed vaccine effectiveness (Rajao and Perez 2018). However, such information is frequently unavailable for majority of developing countries (Caini et al. 2015; Caini et al. 2018a). Furthermore, most developing countries lack a national influenza vaccination policy (Dawa et al. 2019).

Kenya is a lower middle-income country in East Africa and is currently engaged in the process of formulating its national influenza vaccination policy (Dawa et al. 2019). The country lies on the equator with a climate that varies regionally, mostly between tropical to sub-tropical. An influenza surveillance study conducted between 2012-2016 across 11 sites in Kenya found that ~ 31% of medically-attended influenza cases were of IBV type (Emukule et al. 2019). IBV prevalence among influenza positives fluctuated from year-to-year, with 2016 recording the highest proportion (61%) over the 5-year surveillance period. Further, it was observed that the two IBV lineages alternated in predominance; B/Victoria lineage predominated in the years 2012 and 2016 while B/Yamagata predominated in year 2013, 2014 and 2015 (Emukule et al. 2019).

Influenza genomic analysis is now recognized as instrumental in providing a detailed information on the mutations that could facilitate antigenic escape, antiviral resistance, enhanced virulence and can uncover the transmission history and pathways of locally and globally circulating viruses (Goldstein et al. 2017; Hirve and Organization 2015). Unlike influenza A, no study to-date has examined the genomic epidemiology of IBV in Kenya or elsewhere in Africa to characterise the local phylodynamics and phylogeography in comparison to known global patterns. The extent of inter-connectedness of IBV epidemics that occur locally to those happening regionally and globally is yet to be examined. In this study, we present detailed genomic analysis of the circulating IBV strains over a single year in a rural coastal area of Kenya, their spread, evolutionary dynamics and global context.

## MATERIALS AND METHODS

**Study area and population**. All samples analysed here were collected in health facilities within the Kilifi Health and Demographic Surveillance System (KHDSS) area (Scott et al. 2012), located at the Indian ocean coastal region of Kenya. The study period was over a one year period (January to December 2016) (Nyiro et al. 2018). The KHDSS area spans ~900 $km^2$ and is resident to ~280,000 people as at of 2016. The area has a tropical climate with two rainy seasons; the main rains that usually peak in April and the short rains that usually peak in November (Nokes et al. 2009). Study participants were both resident (majority) and non-resident individuals of any age presenting with acute respiratory illness (ARI) to nine primary outpatient health facilities within the KHDSS area (Nyiro et al. 2018) or inpatients under five years of age admitted with syndromic severe or very severe pneumonia to the Kilifi County Hospital (KCH) (Nokes et al. 2009).

**Study design**. The description of the study design, including selection criteria and case definitions, at KCH and at the included KHDSS outpatient facilities can be found in our previous reports (Nokes et al. 2009; Nyiro et al. 2018). KCH is main referral hospital in Kilifi County providing inpatient care and is located in Kilifi town, the county headquarters. All children meeting the inclusion criteria are eligible for enrolment (around 75% are approached, consented and a sample collected). The outpatient facilities included in the study were: Chasimba, Sokoke, Matsangoni, Ngerenya, Mavueni, Mtondia, Junju, Jaribuni, and Pingilikani, **Figure 1, panel (a)**. The study aimed to collect a sample from the first 15 eligible individuals identified in each clinic per week (7020 samples in total). Roughly equal numbers of samples were collected from each outpatient facility throughout the year (Nyiro et al. 2018), with approximately similar numbers collected per month except for December

which was affected by health worker strike (Ong'ayo et al. 2019). Samples from KCH represented a full year sampling fulfilling the eligibility criteria.

**Ethical approval**. The KEMRI Scientific and Ethics Review Unit (SERU) granted ethical clearance for the study protocol and procedures. Study participants or their parents/caregivers (if aged <18 years) provided a written informed consent to participate in the study before sample collection.

**Sample handling and IBV detection.** Nasopharyngeal (NP) swabs were collected at the outpatient facilities while naso-oropharyngeal (NP/OP) swabs were collected at KCH. The swabs were immediately put into viral transport medium (VTM) and transferred to a cool box with 'ice' blocks before transportation to the Kenya Medical Research Institute (KEMRI) - Wellcome Trust Research Programme laboratories (KWTRP) for long-term storage at -80$^{o}$ C. All samples were screened for 15 different viruses using real-time reverse transcription PCR (RT-PCR) diagnostic assay, as previously described (Hammitt et al. 2011; Onyango et al. 2012). IBV primers and probe targeted a conserved region of the non-structural (NS) segment of IBV genome (Gunson and Carman 2011). IBV positive samples with a considerably high viral load as defined an RT-PCR cycle threshold (Ct) value of <27.0, were selected for the KHDSS outpatient facilities arm, but all IBV positive samples regardless of Ct value from KCH inpatients arm, for whole genome sequencing (WGS), **Figure 1, panel (b)**.

**Nucleic acid extraction and multi-segment reverse transcription PCR**. Nucleic acid was extracted from IBV positive samples using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany). Complete IBV segments were amplified using the Universal IBV-GA2 primer set (Zhou et al. 2014) in a multisegment reverse transcription PCR (M-RT-PCR) utilizing SuperScript III One-Step Kit with Platinum Taq DNA Polymerase High Fidelity (Invitrogen, Carlsbard, CA, USA). The universal primers amplify all eight IBV

genome segments in a single PCR reaction tube (Zhou et al. 2014). Successful amplification was confirmed by running the PCR products and controls on a 2% agarose gel and visualizing on a UV trans-illuminator after staining with RedSafe Nucleic Acid Staining solution (iNtRON Biotechnology Inc., Seoul, South Korea).

**Next generation sequencing**. Amplicons were purified with 1×AMPure XP beads (Beckman Coulter Inc., Brea, CA, USA), quantified with Quant-iT dsDNA High Sensitivity Assay (Invitrogen, Carlsbard, CA, USA), and normalized to 0.2 ng/μL. Indexed paired-end libraries were generated from 2.5 μL of 0.2 ng/uL amplicon pool using Nextera XT sample Preparation Kit (Illumina, San Diego, CA, USA) following the manufacturer's protocol. Amplified libraries were purified using 0.8×AMPure XP beads, quantified using the Quant-iT dsDNA High Sensitivity Assay, and evaluated for fragment size in the Agilent 2100 BioAnalyzer System using the Agilent High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA, USA). Libraries were diluted to 2nM, pooled and denatured, then diluted to 12.5 pM. Sequencing was performed on the Illumina MiSeq using MiSeq v250 cycle kit with 5% PhiX (Illumina, San Diego, CA, USA) spike-in. Sequence assembly was performed using the Iterative Refinement Meta-Assembler (IRMA) default settings: median read Q-score filter of 30, the minimum read length of 125, the frequency threshold for insertion and deletion refinement of 0.25 and 0.60 respectively; Smith-Waterman mismatch penalty of 5; and gap opening penalty of 10 (Shepard et al. 2016).

**Comparison dataset.** Two datasets compiled from the Global Initiative on Sharing All Influenza Data (GISAID) were prepared for comparison with the newly sequenced Kilifi IBV strains. The first dataset comprised reference sequences of B/Victoria and B/Yamagata lineages, and clades within these lineages collected between 1987 and 2020, and IBV strains that were included in trivalent and quadrivalent influenza vaccines that were recommended during the 2015/16 northern hemisphere influenza season

(https://www.who.int/influenza/vaccines/virus/recommendations/2015_16_north/en/). This dataset (n=54) was used to assign lineage and clades to the Kilifi IBV strains and in segment-by-segment evolutionary analyses. The second dataset was a 1207 sample of IBV whole genomes deposited in GISAID database for samples collected between January 2014 and December 2016 across all continents. Details on how these were selected are provided in **S1 Table**. This dataset was used to investigate the global phylogenetic context of the Kilifi IBV strains.

**Phylogenetic Analysis**. The assembled and segment assorted Kilifi nucleotide (nt) sequences were aligned together with the global references using MAFFT v7.245 (Katoh et al. 2013) and visualized using Aliview v1.25 (Larsson 2014). For each of segment dataset, maximum likelihood (ML) phylogenetic trees were inferred using RaxML v.8.2.12 (Stamatakis 2014), based on the best-fit models of nt substitution determined by IQ-TREE v1.5.5 (Nguyen et al. 2015). The individual segment sequences were concatenated to give the full-length genome sequences using SequenceMatrix (Vaidya et al. 2011). For all ML trees, the clustering reliability was evaluated by bootstrap resampling 1000 replicates. The tree topologies across segments were detect reassortment events, using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). Reassortant IBV strains were confirmed computationally using the Graph-incompatibility-based Assortment Finder (GiRaF) tool (Nagarajan and Kingsford 2011).

**Evolutionary analysis**. This was undertaken for the individual segments and the concatenated genomes. The linearity in nt sequence divergence with sampling time for datasets utilized in inferring the time to the Most Recent Common Ancestor (tMRCA) and substitution rates were initially assessed using TempEst v1.5.3 program (Rambaut et al. 2016). Viruses suspected to be reassortants were analyzed separately. Nt substitution rates and the tMCRA were estimated using the Bayesian approach implemented in BEAST v1.10.4

program (Suchard et al. 2018). The log files from the analysis were examined using Tracer v1.7.1 program (http://tree.bio.ed.ac.uk/software/tracer/) and run convergence (defined as estimated sample size of >100 for all sampled parameters) was confirmed before extracting the relevant parameter estimates. All BEAST runs were set to at least 10 million steps with sampling after every 2500 steps. Additional and longer runs were considered if the initial analysis did not show convergence. Maximum clade credibility (MCC) trees were summarized from the tree log file using TreeAnnotator v1.10.4 with a 10% burn-in. The MCC trees were visualized using FigTree v1.4.4 (Suchard et al. 2018).

**Genetic diversity and transmission analysis**. We categorised the diversity observed in the Kilifi sequenced viruses using five measures that reflected the closeness of the viruses in their underlying transmission history and genetic diversity. These levels were defined on the basis of the observed phylogenetic clustering with reference sequences, the concatenated segments phylogeny (that identified reassortants), pairwise nt distances and the estimated time of divergence at the branch nodes. The defined categories (the first two consistent with what has been described in literature) are summarised below:

a) Lineage: Based on phylogenetic clustering of HA segment sequences with B/Yamagata and B/Victoria lineage reference sequences (Arvia et al. 2014).
b) Clade: Based on HA phylogenetic clustering with reference sequences of known clades within B/Yamagata (Clade 1-3) and B/Victoria (Clade 1-6) lineages (Arvia et al. 2014; Tramuto et al. 2016).
c) Phylogenomic cluster: Based on the genome phylogeny clustering. Phylogenomic clusters were assigned to major branches that showed high bootstrap support values (>70%).
d) Epidemiological cluster: Based on the time to the MRCA inferred from the reconstructed genome phylogeny. Epidemiological cluster members belonged to the

same phylogenomic cluster and had a divergence date within a year prior to the start of our surveillance i.e. 2015 onwards.

e) Transmission cluster: These were defined as viruses of the same epidemiological cluster that were independent introductions into the local population. As a conservative estimate, members of different transmission clusters diverged before the start of our local surveillance i.e. a before January 2016.

The potential transmission networks within and between populations visiting the enrolled KHDSS health facilities were inferred in PopART package v1.7.2 using TCS method with an epsilon of zero (Leigh and Bryant 2015). The networks were created for each identified phylogenomic cluster from the concatenated segments alignments.

**Spatial analysis.** We conducted a phylogeographic analysis to assess virus movement between the KHDSS locations and in relation to the rest of the world using methods implemented in BEAST v1.10.4 package. The analysis was implemented with a symmetric discrete trait approach and applied the Bayesian stochastic search variable selection (BSSVS) model (Lemey et al. 2009). To reduce complexity of the MCC, location states were categorized as "non-Kilifi" or the specific health center regions. Phylogeographic inferences were visualized with the spatial phylogenetic reconstruction of evolutionary dynamics using data-driven documents (SPREAD3) v0.9.7.1c (Bielejec et al. 2016). To visualize the geographic migration of the virus over time, a D3 file was generated using SPREAD3 v0.9.7.1c. We used the KHDSS geo.json file. The resulting log files we used to calculate Bayes Factor (BF) values for significant diffusion rates between discrete locations.

**Statistical Analysis**. Numeric variable analyses were conducted using STATA v15.1 (StataCorp. College Station, TX). Mean, median and inter-quartile range (IQR) were used to summarise continuous variables, while proportions were calculated for binary variables. Comparison between means and medians was done using t-test and median test, respectively.

Patient distribution across the age groups and lineage distribution across demographic sub-categories were compared using the Wilcoxon-Mann-Whitney U test and the independent t-test. Demographic and clinical characteristics among IBV positive patients sequenced versus those not sequenced and for the two lineages were compared using Fisher's exact test. Statistical significance (2 tailed) was set at p value of $\leq 0.05$.



## RESULTS

**Demographic characteristics**. Between January and December 2016, a total of 5647 NP swabs were collected from the nine KHDSS outpatient facilities and 574 NP/OP swabs from inpatients at KCH, **Figure 1, panel (b)**. Of these, 4.7% (267/5647) from the outpatient facilities and 1.9% (11/574) of samples from KCH tested IBV positive by real-time RT-PCR. Sequencing attempt on 120/278 selected positives (43.2%, 109 from KHDSS outpatient facilities and 11 from KCH), yielded 111 whole genomes; 101 from KHDSS outpatient facilities and 10 from KCH , **Figure 1, panel (b)**. The demographic and clinical characteristics of the patients from whom genomes were obtained were similar to those for whom genomes were not obtained except for viral loads (as indicated by diagnostic RT-PCR Ct value) and health facility type (inpatient or outpatient), **Table 1**. The patient age among IBV positive patients ranged between 1 month and 85 years (mean: 9.6 years, and median: 6 years). The proportion positive was highest in patients aged between 0-4 years (43.2%) followed by the 5-14 years age group (39.2%). Female patients accounted for most IBV positives (57.2%) and cough was the most common symptom in the IBV positive patients followed by fever and nasal discharge, **Table 1**.

**Seasonality and representativeness of the sequenced samples.** IBV were detected in all months of 2016 in the surveillance although the number of cases fluctuated from month-to-month, **Figure 1, panel (c)**. Notably each KHDSS health facility experienced a

peak incidence at different months of the year but the majority fell between March and August. This coincided with one of the two rainy seasons in region. At the individual health facilities, some months had no IBV detections in the samples analysed. When all the enrolled health facilities were combined, sequence data was available from every month in 2016 except February and December. The fraction of samples from each health facility that were sequenced roughly reflected the total number of the positives that were detected in the specific health facility, **S2 Figure**. At least one sample was sequenced from each of the enrolled health facilities and this enabled our interrogation of IBV transmission between the populations served by the enrolled health facilities.

**The B/Yamagata and B/Victoria lineages co-circulated in the study population**. The HA phylogeny demonstrated that the two known IBV lineages were co-circulating: B/Victoria and B/Yamagata, **Figure 2**, **panel (a)**. Overall, the B/Victoria lineage predominated during this single year (84.7%) with all its sequenced viruses falling into clade 1A, **Figure 2, panel (b)**. The HA of B/Yamagata lineage viruses clustered within two genetically distinct known clades; clade 2 (n=2) and clade 3 (n=15). The B/Yamagata clade 3 viruses were detected in low numbers generally but throughout the year while the B/Yamagata clade 2 viruses were detected only in January and March as shown in **Figure 2, panel (c)**. The demographic and clinical characteristics of patients infected by either B/Yamagata lineage or B/Victoria lineage were not statistically different, **S3 Table.**

**Inter-lineage reassortment in the Kilifi IBV strains.** Segment-specific phylogenies showed a clear separation into the B/Yamagata and B/Victoria lineages for the majority of the Kilifi viruses (n=101, 90.1%) in all the segments; PB2, PB1, PA, HA, NA, NP, M, and NS (**S4 Figure)**. For the remainder viruses (n=10), two inter-lineage reassortment events were suspected that were confirmed in GiRaF analysis, **S5 Table**. The first involving a single virus (B/Kilifi/114/2016/KCH/14-Oct-2016) that had B/Victoria lineage backbone including

the HA segment but its PB2, PB1 PA and MP segments clustered closely to B/Yamagata lineage viruses, **S4 Figure**. Notably, for this virus, its NA and NP segments were distinct from the other B/Victoria lineage viruses. The second reassortment event involved nine B/Yamagata clade 3 viruses in which the NA and NP segments had been acquired from B/Victoria lineage but the backbone (i.e. all other segments) remained of B/Yamagata lineage. The B/Yamagata clade 2 viruses (n=2) maintained a unique and unchanged constellation across all the eight segments.

**Phylogenomic clusters in the Kilifi IBV 2016 epidemic.** The ML phylogeny of the concatenated 8 segments of the Kilifi IBV strains, including reference strains is shown in **Figure 3, panel (a)**. Multiple well-supported clusters were observed with evidence of mixing of virus samples from different health facilities. We assigned the five main branches including Kilifi strains with high bootstrap support (>70%) phylogenomic clusters namely; (i) pure B/Victoria lineage clade 1A viruses (n=93, 83.8%), (ii) reassortant B/Victoria lineage clade 1A virus (n=1, 0.9%), (iii) pure B/Yamagata clade 2 viruses (n=2, 1.8%), (iv) pure B/Yamagata clade 3 viruses (n=6, 5.4%) and (v) B/Yamagata clade 3 reassortant viruses (n=9, 8.1%). The KCH surveillance captured 4 of the 5 of these circulating phylogenomic clusters excepting only the pure B/Yamagata clade 2 viruses. Viruses of this clade were seen only in one KHDSS outpatient facility (Matsangoni, found furthest North in the KHDSS). Most other KHDSS outpatient facilities observed circulation of 1-2 clusters except Mtondia where three phylogenomic clusters were detected, **Figure 3, panel (b**). Mtondia and Matsangoni are located along a Mombasa-Malindi highway, **Figure 1, panel (a)**. Of the 10 months that we obtained genome sequence data, in all except November we detected the pure B/Victoria Clade 1A viruses. The reassortant B/Yamagata clade 3 viruses were the second most persistent phylogenomic cluster with detection in five months of the 10 months, **Figure 3, panel (c).** B/Yamagata clade 2 and 3 pure clusters were observed only at the beginning of

the year while reassorted B/Victoria clade 1A virus was detected only in October and only at the KCH facility.

**Evolutionary dynamics of the Kilifi IBV strains.** A strong linear relationship between root-to-tip genetic distance and sampling date was observed in all assessed phylogenomic clusters ($R^2$ consistently >0.6), **S6 Figure**. The Kilifi genomes fit well in the global continuum of observed diversity of the identified phylogenomic clusters. The time-resolved BEAST phylogenies for the combined non-reassortant viruses, and for main phylogenomic clusters in Kilifi (those with a sample size of >2) are shown **Figure 4.** For the reassorted B/Yamagata clade 3 viruses, the global tMRCA was estimated to be around May 2013 (95% HPD: March 2013 to August 2013), **Figure 4, panel (b)** and **Table 2**. The estimated nt substitution rates and tMRCA for the individual phylogenomic clusters are provided in **Table 2**. For all the individual phylogenomic clusters, tMRCA for the IBV viruses sampled in Kilifi during our surveillance occurred within 2015 except for the pure B/Yamagata clade 3 whose tMRCA occurred in October 2013 (95% HPD interval June 2012 to December 2014), **Figure 4, panel (d)**.

Using the global reference set collected from 1987 to 2020 and the Kilifi strains and the unique Kilifi strains, we estimated the tMRCA and nucleotide substation rate for each of the segment, **Table 3**. The segments arrived at different tMRCA estimates for the included strains. The more recent tMRCA (March 1983) were from the NA and MP segments while the earliest tMRCA (July 1968) was arrived at from analysis of NS segment. As expected, the highest nucleotide substitution rates were observed with the HA segment ($1.97 \times 10^{-3}$) and NA segments ($1.92 \times 10^{-3}$) while the slowest substitution rate was observed with NS segment ($1.24 \times 10^{-3}$).

**Global context of the 2016 Kilifi IBV strains**. We examined this using HA segment analysed by the ML approach. The global comparison dataset included 1207 IBV strains

sampled across 6 continents between 2014 and 2016. The phylogeny showed a clear birfurcation into two major clades corresponding to B/Victoria (n=481) and B/Yamagata (n=743) Lineages (figure not shown). The ML phylogenies for the individual Lineages are shown in **Figure 5.** For both B/Victoria and B/Yamagata, the phylogenies confirmed that multiple distinct strains were in circulation in the Kilifi community in 2016 some of which had extensive local onward transmission e.g. B/Victoria Clade 1A. The viruses clustering closest to the Kilifi IBV strains were commonly those detected in other African countries especially neighbouring Uganda and Tanzania (**S7 Figure**).

By our set criteria (based on clustering with global sequences and divergence dates), we identified a total of 7 epidemiological clusters and 29 transmission clusters from the Kilifi IBV strains, (see **Figure 4 and Figure 5)**. The epidemiological cluster membership size (for Kilifi sequences) ranged from 1 to 93 while transmission cluster membership size varied from 1 to 28, **Figure 4**. The vast majority of the Kilifi transmission clusters (23/29) were within the pure B/Victoria clade 1A.

**Local phylogeography of the detected IBV strains.** The genetic relatedness Kilifi viruses within the same phylogenomic clusters by health facility is shown in **Figure 6 (**for B/Victoria 1A**) and S8 Figure** (for both B/Yamagata pure and reassorted clade 3). In some of the KHDSS facilities, it was clear that a dominant transmission cluster existed e.g. some of the pure Victoria 1A variants for Chasimba, Mavueni and Mtondia while others had no clear dominant transmission cluster, **Figure 6, panel (a)**. Further, the phylogeographic analysis showed that IBV Victoria IA viruses were commonly getting into and out of the KHDSS area through the Matsangoni area, **Figure 6, panel (b) and (c)**. The Matsangoni health facility had connection with 7 of the 9 other health facilities. Although the Yamagata clade 3 clusters had KCH as their link to the rest of the world, the sample size was relatively small and KCH is a referral facility thus a direct link cannot be concluded, **S8 Figure**.

## DISCUSSION

Detailed phylodynamic and transmission studies on influenza in low-income settings in the tropics remain sparse. This is despite these regions bearing a disproportionately large influenza burden (Byarugaba et al. 2016; Katz et al. 2012). Here, through a comprehensive genomic analysis, we show that the year-round circulation of influenza B virus in a coastal region of Kenya (covering ~900 $km^2$) in 2016 (Nyiro et al. 2018) was occasioned by co-circulating IBV clades and viral clusters within both B/Yamagata and B/Victoria lineages. Further, we show that that the epidemic season was instigated by up to 29 independent strain introductions, some of which were inter-lineage reassortants. The genomic analysis recognised extensive local spread of the new IBV strains once introduced accompanied with significant accumulation of nucleotide substitutions.

The B/Victoria lineage predominated the IBV season we observed, being responsible for > 80% of the IBV infections we sequenced. These finding are congruent with a recent IBV report from Kenya of a surveillance study across 10 sites from 2012 to 2016 (Emukule et al. 2019). The authors observed that in the 2016 season, IBV was the predominant influenza type in the enrolled health facilities (just like at this coastal Kenya site, (Nyiro et al. 2018)) and the B/Victoria lineage was the predominant IBV lineage. Here through our detailed genomic analysis, we extend these earlier observations to show that this season was predominated by B/Victoria clade 1A strains and the co-circulating B/Yamagata viruses were of clade 2 and 3, and there were at least two inter-lineage reassortant strains in circulation.

Currently available seasonal influenza vaccines require periodic update to better match circulating influenza strains (Hampson et al. 2017). Both trivalent influenza vaccines that have a representative H3N2, H1N1 and IBV strain (of either B/Victoria or B/Yamagata lineage) and quadrivalent vaccines that have a representative H3N2, H1N1 and both IBV lineages (Victoria and B/Yamagata) are available (Grohskopf et al. 2019). Currently, Kenya

does not have a national influenza vaccination policy (Dawa et al. 2019). In this study, although B/Victoria lineage was predominant, our findings support the notion of deployment of quadrivalent influenza vaccines for optimal vaccine effectiveness. The impact of the additional diversity we observed within B/Yamagata lineage (2 antigenically distinct clades and the emergence of reassortant viruses on the overall vaccine effectiveness require further investigation.

IBV inter-lineage reassortment is well recognised in literature (Dudas et al. 2015; Monamele et al. 2018; Tewawong et al. 2017). For instance, reassortant B/Yamagata lineage viruses with a B/Victoria NA were recently reported in Cameroon (2014 -2017) (Monamele et al. 2018). Here, we identified two inter-lineage reassortment events; (i) B/Yamagata lineage viruses that had acquired NP and NA segments from B/Victoria lineage viruses, and (ii) a B/Victoria lineage virus that had acquired PB2, PB1, PA and MP from B/Yamagata clade 3 viruses. Previous studies observed that IBV reassortant viruses tend to circulate at a low prevalence and do not persist over epidemics (G.-W. Chen et al. 2007; Chi et al. 2005). In the current study, the reassorted B/Yamagata clade 3 appeared to transmit for at least 6 months while the reassortant B/Victoria clade 1A strain had single time point detection. Notably, the latter reassortant was unusual given the co-segregation of PB2, PB1 and PA. Previous analysis noted that PB2, PB1 and HA segments tend to segregate together due to more compatibility (Dudas et al. 2015). A follow up study to investigate the fate, clinical and epidemiological impact of the reassortant strains we observed here will be useful.

Some studies have associated B/Victoria lineage infections with more severe disease compared to B/Yamagata lineage infections while other did not find such relationship (Caini et al. 2018b; Emukule et al. 2019). In the current study, we did not observe a significant difference in lineage distribution between in-patients and outpatients. Further, the genome phylogenies observed interspersing of strains that were found in the inpatients and those

found in mild ARI outpatient cases. These observations suggest that it is host rather than viral factors that are most critical in determining IBV disease severity. Further, we found that IBV infections were most frequent amongst 0-14 year-olds and lineage distribution did not appear to be influenced by age unlike what has been reported in some previous studies (Horthongkham et al. 2016; Korsun et al. 2017; Yang et al. 2018).

The inclusion of regional and global genomes deposited in GISAID significantly improved the power of our phylodynamic analyses and showed that the Kilifi IBV diversity was part of the global continuum. For example we determined that the reassortant Yamagata clade 3 viruses were circulating in several other countries including Uganda, Tanzania, Rwanda, Congo, Nigeria, Cote D'Ivoire, Mali, Burkina Faso, Indonesia, Laos, Bangladesh, Nepal, Singapore, Japan and USA. By tMRCA analysis we found that the reassortment event that resulted in this cluster occurred around October 2012 (95% HDP March 2012 to April 2013). However, we did not find a close relative in the database to the reassortant Victoria clade 1A. Our repeated bioinformatics analysis of the raw short-read data of this sample reproduced the reassorted genome.

The phylogeographic analyses demonstrated IBV migration both into and out of the KHDSS area. The Kilifi IBV genomes seemed to frequently have their close relatives in neighboring Uganda and Tanzania. This observation is one that requires a follow up investigation to test the hypothesis that new influenza epidemics are likely to be seeded from neighboring East African countries than distant countries. Within the KHDSS area, for the phylogenomic cluster that had significant sample size, virus seeding seemed to start from Matsangoni. The area to the north of the KHDSS has two key touristic towns (Watamu and Malindi) and their proximity to Matsangoni might explain the virus entry via matsangoni. Further investigation is required to confirm this hypothesis.

This study had some limitations. First, the sequencing was undertaken for only a single year period. Thus we cannot conclude on the long-term consistency of the observed IBV transmission patterns and the fate of the identified reassortant strains. Second, we sequenced only a fraction of the identified IBV positives (~40%). The prioritised samples were selected on the basis of anticipated probability of successful sequencing inferred from the sample's viral load as indicated by the diagnostic Ct value. Such a strategy ultimately avoided sequencing some samples that may have been critical in reconstructing the true transmission networks and may bias cluster prevalence. However, the demographic and clinical characteristics of the sequenced and not sequenced patients were similar except for their viral load. Third, the KHDSS outpatient facilities surveillance collected a maximum of 15 samples/week/site. This non-exhaustive sampling at the facilities may have introduced bias in the inferred lineage/clade prevalence and transmission networks.

In conclusion, our genomic analysis of IBV confirms that B/Victoria (clade 1A) and B/Yamagata (clades 2 and 3) lineage viruses were in co-circulation together with 2 inter-lineage reassortant variants in coastal Kenya in 2016. The co-circulation of divergent IBV viruses complicates the optimal selection of influenza vaccine strain components for local use. As Kenya formulates her influenza vaccination policy, the choice of broad immunity (Epstein 2018) or more valence vaccines ( e.g. quadrivalent regimen (Dbaibo et al. 2019)) should be considered. Further, this study demonstrates the benefits of analysis of full-length IBV genomes. In addition to providing a clearer understanding of locally circulating viral diversity, a high-resolution tracking of transmission of IBV strains was achievable at a scale impossible with single or few segment analysis. That in a single season up to 29 independent IBV introductions occurred demonstrates the challenge of controlling local influenza epidemics by targeted approaches e.g. sub-population vaccination, patient quarantine or institutional closures as previously observed (Holmes et al. 2011). Future studies should

combine genomic data with various epidemiological data (e.g. host migration, immunity profiles, population densities and social contact patterns) to elucidate patterns of IBV infection and spread in this setting for better-informed control strategies.

**Acknowledgments.** We thank the study participants for providing samples, the field workers for collecting epidemiological data and the clinical samples and members of the Virus Epidemiology and Control (VEC) Research Group at KWTRP for laboratory processing of the samples and stimulating discussions. We thank Dr. John Barnes and Dr. David E. Wentworth of the Centers for Disease Control and Prevention (CDC) Atlanta, U.S.A., for their input in setting up the MiSeq whole-genome sequencing assays in Kenya. We acknowledge the authors, originating and submitting laboratories of the sequences from GISAID™ EpiFlu Database on which this research is based, S9 Table. All submitters of data may be contacted directly via the GISAID website www.gisaid.org.

**Funding.** This study was funded by The Wellcome Trust [102975, 203077]. The authors FMN and CNA were supported by the Initiative to Develop African Research Leaders (IDeAL) through the DELTAS Africa Initiative [DEL-15-003]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) 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 authors and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government. This paper is published with the permission of the Director of KEMRI.

**Data availability**. All data generated or analyzed during this study has been deposited to the VEC Data Repository in Havard Dataverse under the doi: https://doi.org/10.7910/DVN/OE6TS2. The sequence data are deposited on the GISAID database under the accession numbers; EPI_ISL_336258 and EPI_ISL_336282 to EPI_ISL_336395.

**Supporting information.**

**S1 Table.** Global IBV genomic datasets used to place the Kilifi genomes on the global context.

**S2 Figure.** Distribution of the sequenced samples in the Kilifi surveillance.

**S3 Table.** Comparison of the clinical and demographic characteristics the patients who were infected by Yamagata and Victoria lineages

**S4 Figure.** Segment specific ML phylogenetic trees for each of the eight IBV genomic segments.

**S5 Table.** A GiRAF analysis summary table showing the segment-segment report.

**S6 Figure.** Temporal phylogenetic signal in the assigned phylogenomic clusters.

**S7 Figure.** Zoomed in view of the ML global HA phylogeny of IBV strains sampled across the world between 2014 and 2016.

**S8 Figure.** Possible transmission links between the Kilifi Yamagata clade 3 viruses.

**S9 Table.** Acknowlegdgement of submitters of sequences deposited GISAID database that were used in this study.

**REFERENCES**

Agoti, C. N., et al. (2015), 'Successive Respiratory Syncytial Virus Epidemics in Local Populations Arise from Multiple Variant Introductions, Providing Insights into Virus Persistence', *J Virol,* 89 (22), 11630-42.

Agoti, C. N., et al. (2019), 'Genomic analysis of respiratory syncytial virus infections in households and utility in inferring who infects the infant', *Sci Rep,* 9 (1), 10076.

Arvia, R., et al. (2014), 'Molecular markers of influenza B lineages and clades', *Viruses,* 6 (11), 4437-46.

Bedford, Trevor, et al. (2014), 'Integrating influenza antigenic dynamics with molecular evolution', 3, e01914.

Bielejec, F., et al. (2016), 'SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes', *Mol Biol Evol,* 33 (8), 2167-9.

Byarugaba, D. K., et al. (2016), 'Whole-genome analysis of influenza A(H1N1)pdm09 viruses isolated in Uganda from 2009 to 2011', *Influenza Other Respir Viruses,* 10 (6), 486-92.

Caini, S., et al. (2018a), 'Epidemiology of seasonal influenza in the Middle East and North Africa regions, 2010-2016: Circulating influenza A and B viruses and spatial timing of epidemics', *Influenza Other Respir Viruses*.

Caini, S., et al. (2018b), 'Clinical characteristics and severity of influenza infections by virus type, subtype, and lineage: A systematic literature review', *Influenza Other Respir Viruses,* 12 (6), 780-92.

Caini, S., et al. (2015), 'Epidemiological and virological characteristics of influenza B: results of the Global Influenza B Study', *Influenza Other Respir Viruses,* 9 Suppl 1, 3-12.

Chen, Guang-Wu, et al. (2007), 'Multiple genotypes of influenza B viruses cocirculated in Taiwan in 2004 and 2005', *Journal of clinical microbiology,* 45 (5), 1515-22.

Chen, R. and Holmes, E. C. (2008), 'The evolutionary dynamics of human influenza B virus', *J Mol Evol,* 66 (6), 655-63.

Chi, X Sherry, et al. (2005), 'Detection and characterization of new influenza B virus variants in 2002', *Journal of clinical microbiology,* 43 (5), 2345-49.

Dawa, J., et al. (2019), 'Developing a seasonal influenza vaccine recommendation in Kenya: Process and challenges faced by the National Immunization Technical Advisory Group (NITAG)', *Vaccine,* 37 (3), 464-72.

Dbaibo, G., et al. (2019), 'Quadrivalent Influenza Vaccine Prevents Illness and Reduces Healthcare Utilization Across Diverse Geographic Regions During Five Influenza Seasons: A Randomized Clinical Trial', *Pediatr Infect Dis J.*

Dudas, G., et al. (2015), 'Reassortment between influenza B lineages and the emergence of a coadapted PB1-PB2-HA gene complex', *Mol Biol Evol,* 32 (1), 162-72.

ECDC (2019), 'European Centre for Disease Prevention and Control, 2005 - 2019.', <https://ecdc.europa.eu/en/home>, accessed 14 May 2019.

Edgar, R. C. (2010), 'Search and clustering orders of magnitude faster than BLAST', *Bioinformatics,* 26 (19), 2460-1.

El Guerche-Seblain, C., et al. (2019), 'Epidemiology and timing of seasonal influenza epidemics in the Asia-Pacific region, 2010-2017: implications for influenza vaccination programs', *BMC Public Health,* 19 (1), 331.

Emukule, G. O., et al. (2019), 'The Epidemiology and Burden of Influenza B/Victoria and B/Yamagata Lineages in Kenya, 2012-2016', *Open Forum Infect Dis,* 6 (10), ofz421.

Epstein, S. L. (2018), 'Universal Influenza Vaccines: Progress in Achieving Broad Cross-Protection In Vivo', *Am J Epidemiol,* 187 (12), 2603-14.

Goldstein, Emily J., et al. (2017), 'Integrating patient and whole-genome sequencing data to provide insights into the epidemiology of seasonal influenza A (H3N2) viruses', *Microbial Genomics,* 4.

Grohskopf, L. A., et al. (2019), 'Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices - United States, 2019-20 Influenza Season', *MMWR Recomm Rep,* 68 (3), 1-21.

Gunson, R. N. and Carman, W. F. (2011), 'During the summer 2009 outbreak of "swine flu" in Scotland what respiratory pathogens were diagnosed as H1N1/2009?', *BMC Infect Dis,* 11, 192.

Hammitt, L. L., et al. (2011), 'Added value of an oropharyngeal swab in detection of viruses in children hospitalized with lower respiratory tract infection', *J Clin Microbiol,* 49 (6), 2318-20.

Hampson, A., et al. (2017), 'Improving the selection and development of influenza vaccine viruses - Report of a WHO informal consultation on improving influenza vaccine virus selection, Hong Kong SAR, China, 18-20 November 2015', *Vaccine,* 35 (8), 1104-09.

Hirve, Siddhivinayak and Organization, World Health (2015), 'Seasonal influenza vaccine use in low and middle income countries in the tropics and subtropics: a systematic review'.

Holmes, E. C., et al. (2011), 'Extensive geographical mixing of 2009 human H1N1 influenza A virus in a single university community', *J Virol,* 85 (14), 6923-9.

Horthongkham, N., et al. (2016), 'Epidemiological, Clinical and Virological Characteristics of Influenza B Virus from Patients at the Hospital Tertiary Care Units in Bangkok during 2011-2014', *PLoS One,* 11 (7), e0158244.

Kanegae, Y, et al. (1990), 'Evolutionary pattern of the hemagglutinin gene of influenza B viruses isolated in Japan: cocirculating lineages in the same epidemic season', *Journal of virology,* 64 (6), 2860-65.

Katoh, Kazutaka, Standley, Daron M %J Molecular biology, and evolution (2013), 'MAFFT multiple sequence alignment software version 7: improvements in performance and usability', 30 (4), 772-80.

Katz, M. A., et al. (2012), 'Influenza in Africa: uncovering the epidemiology of a long-overlooked disease', *J Infect Dis,* 206 Suppl 1, S1-4.

Korsun, N., et al. (2017), 'Antigenic and genetic characterization of influenza viruses circulating in Bulgaria during the 2015/2016 season', *Infect Genet Evol,* 49, 241-50.

Kumar, S., Stecher, G., and Tamura, K. (2016), 'MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets', *Mol Biol Evol,* 33 (7), 1870-4.

Langat, P., et al. (2017), 'Genome-wide evolutionary dynamics of influenza B viruses on a global scale', *PLoS Pathog,* 13 (12), e1006749.

Larsson, Anders (2014), 'AliView: a fast and lightweight alignment viewer and editor for large datasets', *Bioinformatics,* 30 (22), 3276-78.

Leigh, Jessica W. and Bryant, David (2015), 'popart: full-feature software for haplotype network construction', *Methods in Ecology and Evolution,* 6 (9), 1110-16.

Lemey, P., et al. (2009), 'Bayesian phylogeography finds its roots', *PLoS Comput Biol,* 5 (9), e1000520.

McCullers, J. A., Saito, T., and Iverson, A. R. (2004), 'Multiple genotypes of influenza B virus circulated between 1979 and 2003', *J Virol,* 78 (23), 12817-28.

Monamele, C. G., et al. (2018), 'Genetic characterization of influenza B virus in Cameroon and high frequency of reassortant strains', *J Med Virol,* 90 (12), 1848-55.

Nagarajan, N. and Kingsford, C. (2011), 'GiRaF: robust, computational identification of influenza reassortments via graph mining', *Nucleic Acids Res,* 39 (6), e34.

Nguyen, L. T., et al. (2015), 'IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies', *Mol Biol Evol,* 32 (1), 268-74.

Nokes, D. J., et al. (2009), 'Incidence and severity of respiratory syncytial virus pneumonia in rural Kenyan children identified through hospital surveillance', *Clin Infect Dis,* 49 (9), 1341-9.

Nyiro, J. U., et al. (2018), 'Surveillance of respiratory viruses in the outpatient setting in rural coastal Kenya: baseline epidemiological observations', *Wellcome Open Res,* 3, 89.

Ong'ayo, Gerald, et al. (2019), 'Effect of strikes by health workers on mortality between 2010 and 2016 in Kilifi, Kenya: a population-based cohort analysis', *The Lancet Global Health,* 7 (7), e961-e67.

Onyango, C. O., et al. (2012), 'Influenza surveillance among children with pneumonia admitted to a district hospital in coastal Kenya, 2007-2010', *J Infect Dis,* 206 Suppl 1, S61-7.

Paul Glezen, W., et al. (2013), 'The burden of influenza B: a structured literature review', *Am J Public Health,* 103 (3), e43-51.

Rajao, D. S. and Perez, D. R. (2018), 'Universal Vaccines and Vaccine Platforms to Protect against Influenza Viruses in Humans and Agriculture', *Front Microbiol,* 9, 123.

Rambaut, A., et al. (2016), 'Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen)', *Virus Evol,* 2 (1), vew007.

Rota, Paul A, et al. (1990), 'Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983', *Virology,* 175 (1), 59-68.

Sambala, E. Z., et al. (2018), 'Burden of seasonal influenza in sub-Saharan Africa: a systematic review protocol', *BMJ Open,* 8 (10), e022949.

Scott, J. A., et al. (2012), 'Profile: The Kilifi Health and Demographic Surveillance System (KHDSS)', *Int J Epidemiol,* 41 (3), 650-7.

Seleka, Mpho, et al. (2017), 'Epidemiology of influenza B/Yamagata and B/Victoria lineages in South Africa, 2005-2014', *PloS one,* 12 (5), e0177655.

Shepard, S. S., et al. (2016), 'Viral deep sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler', *BMC Genomics,* 17, 708.

Stamatakis, A. (2014), 'RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies', *Bioinformatics,* 30 (9), 1312-3.

Suchard, M. A., et al. (2018), 'Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10', *Virus Evol,* 4 (1), vey016.

Tan, Y., et al. (2013), 'Differing epidemiological dynamics of influenza B virus lineages in Guangzhou, southern China, 2009-2010', *J Virol,* 87 (22), 12447-56.

Tewawong, N., et al. (2017), 'Evidence for influenza B virus lineage shifts and reassortants circulating in Thailand in 2014-2016', *Infect Genet Evol,* 47, 35-40.

Tramuto, F., et al. (2016), 'The Molecular Epidemiology and Evolutionary Dynamics of Influenza B Virus in Two Italian Regions during 2010-2015: The Experience of Sicily and Liguria', *Int J Mol Sci,* 17 (4), 549.

Vaidya, Gaurav, Lohman, David J, and Meier, Rudolf %J Cladistics (2011), 'SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information', 27 (2), 171-80.

Vijaykrishna, D., et al. (2015), 'The contrasting phylodynamics of human influenza B viruses', *Elife,* 4, e05055.

Xu, C., et al. (2015), 'Comparative Epidemiology of Influenza B Yamagata- and Victoria-Lineage Viruses in Households', *Am J Epidemiol,* 182 (8), 705-13.

Yang, J., et al. (2018), 'Variation in Influenza B Virus Epidemiology by Lineage, China', *Emerg Infect Dis,* 24 (8), 1536-40.

Zhou, B., et al. (2014), 'Universal influenza B virus genomic amplification facilitates sequencing, diagnostics, and reverse genetics', *J Clin Microbiol,* 52 (5), 1330-7.

**Tables**

**Table 1.** Clinical and demographic characteristics of patients who were IBV positive and a comparison of those sequenced versus those whom were not sequenced.

**Table 2.** The evolutionary characteristics of the sequenced IBV strains in the surveillance stratified by phylogenomic cluster.

**Table 3.** Divergence times and rates of nucleotide substitution of gene segments of influenza B virus

**Figure legends**

**Figure 1. Study location, sample laboratory processing and IBV detection in the enrolled health facilities.** Panel (a) a map of the Kilifi Health and Demographic Surveillance System (KHDSS) area showing the spatial distribution of the enrolled KHDSS health facilities. Panel (b) a sample flow gram showing the number of samples and genomes obtained from the Kilifi County Hospital (KCH) and KHDSS outpatient facilities surveillances. Panel (c), a bubble plot showing the number of IBV positives by month and health facility across 2016. The size of the circle is proportional to the number of samples (smallest represent one and largest represent 17 samples).

**Figure 2. Identification of IBV lineages and clades that were in circulation in coastal Kenya in the 2016.** Panel (a) shows a maximum likelihood (ML) phylogenetic tree based on the HA segment of samples we sequenced. From the surveillance, only unique HA sequences are included. The tree includes reference sequences of previously identified clades within B/Yamagata and B/Victoria lineages collected between 1987 and 2020. Kilifi sequences are shown with a red circle, the reference with a black filled circle. Strains included in the 2015/16 influenza vaccine are shown in a green filled circle for B/Victoria and blue filled circle for B/Yamagata. Panel (b) shows the monthly prevalence of the B/Yamagata and B/Victoria lineage viruses across the 12 months the surveillance was undertaken. Panel (c) shows the virus clades that were detected and their frequency across different months of the year 2016.

**Figure 3. Genomic epidemiology of IBV in coastal Kenya.** Panel (a), a ML phylogenetic tree reconstructed from concatenated eight segments of IBV (n=111) from the 9 sampled KHDSS outpatient facilities and KCH. A total of 51 reference sequences (non-Kenya) were

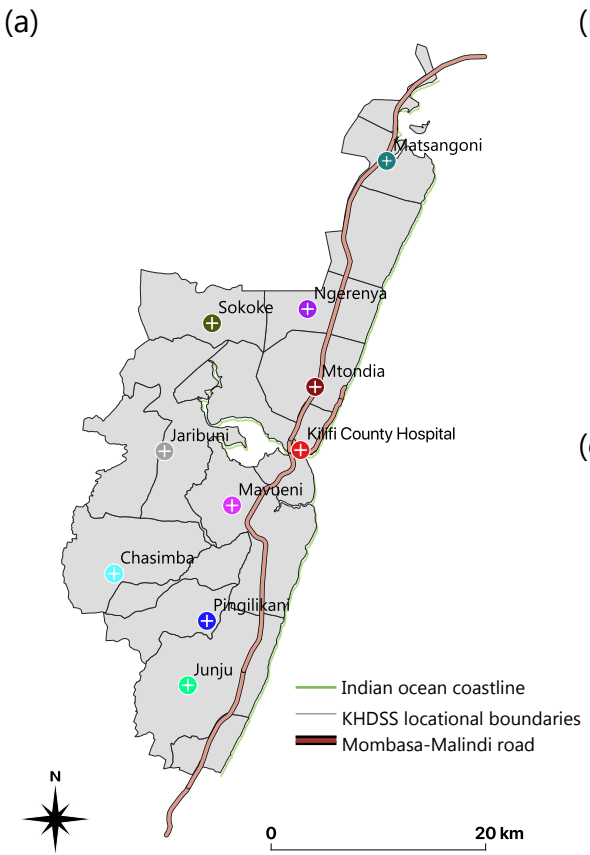
included. The taxa are indicated by filled circles at the tips and colored by health facility of sampling. Panel (b), the location of detection for the defined phylogenomic clusters. Their first three letters abbreviate names of the outpatient health facilities are. Panel C, the monthly prevalence of the defined phylogenomic clusters.

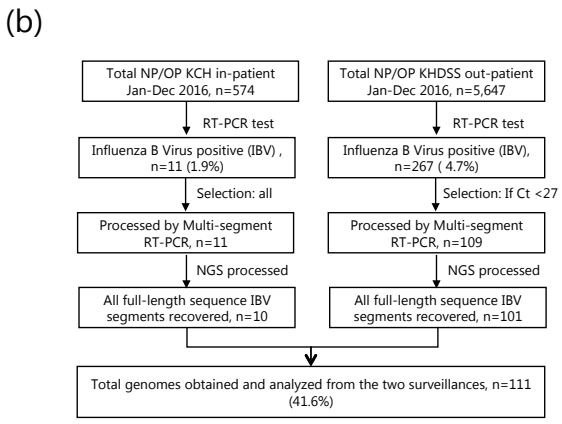
**Figure 4. Time-resolved phylogenies of sub-samples of global genomes combined with Kilifi IBV genomes.** The branches, and the node shapes (size scaled by posterior support of the branch) are coloured by the most probable ancestral location inferred during BSSVS analysis. The threshold for epidemiological clusters (having the MRCA within the year preceding the start of our surveillance (between January and December 2015) is indicated by black dashed line while the threshold of transmission clusters (i.e. occurred first as a single branch at the start of our surveillance in January 2016 with or without onward diversification) is indicated by the orange dashed line. Panel (a) shows combined non-reassortant strains. Panel (b) shows B/Victoria clade 1A viruses. Panel (c) shows reassortant B/Yamagata clade 3 viruses. Panel (d) shows pure B/Yamagata clade 3 viruses.

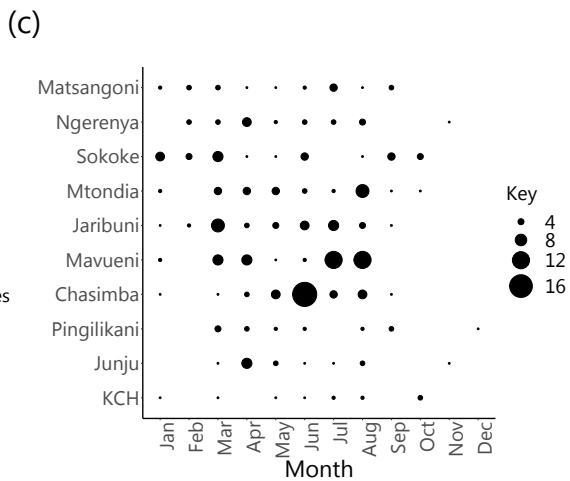
**Figure 5. ML HA phylogeny of IBV strains sampled across the world between 2014 and 2016.** Panel (a), B/Victoria lineage (n=481). Panel (b), B/Yamagata lineage (n=743). The taxa are shown as filled circles and are coloured differently for each geographic origin of the sample by continent. The newly sequenced Kilifi genomes are shown in a bright green colour and the major Kilifi clusters labelled (KLF-1 to KLF-6).

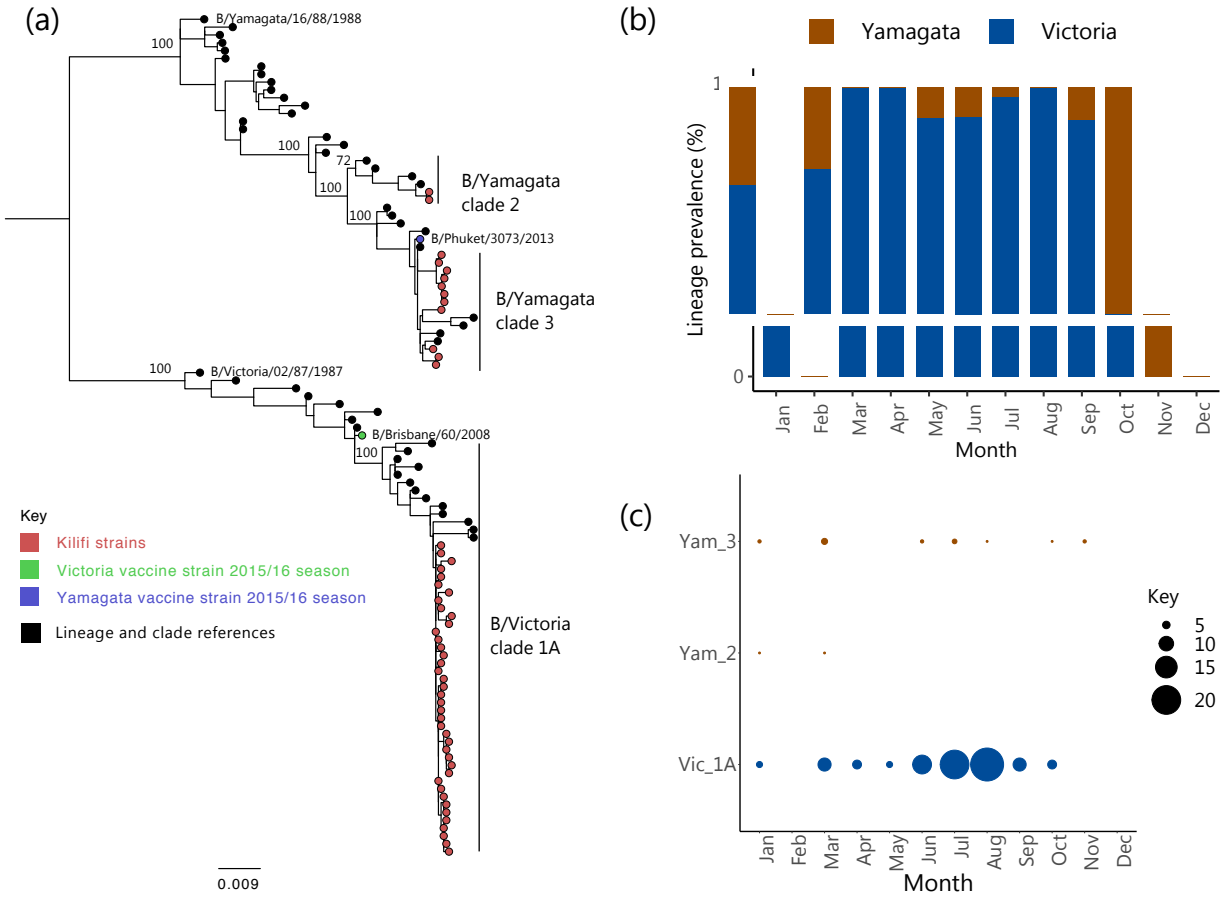
**Figure 6. Possible transmission links between the Kilifi Victoria clade 1A viruses.** Panel (a) shows a TCS POPART network of the 93 newly sequenced pure Victoria clade 1A viruses. The vertices represent the concatenated genome haplotypes. The size of the vertex is

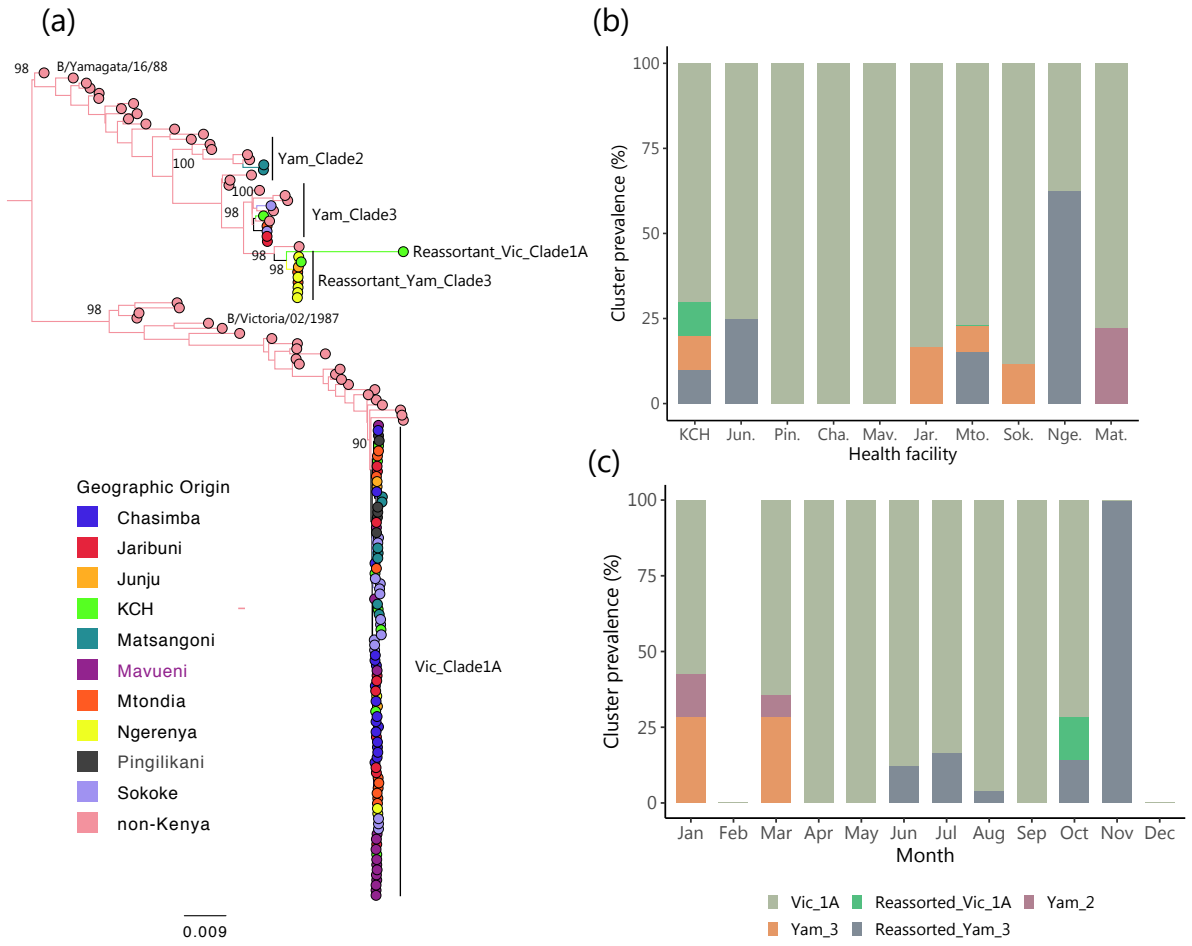
proportional to the number of haplotypes (identical sequences) and is colored by the health facility from which the sequenced sample was collected. The numbers shown on the edges represent the number of nt changes from one vertex (haplotype) to the next. Panel (b) shows the phylogeography the newly sequenced pure Victoria clade 1A in comparison with non-Kilifi genomes (n=116). The lines connecting the health facilities are shown only between location with support of a Bayes factor of > 5. Panel (c) shows the Bayes Factor and posterior probability support for the links shown in panel (b).

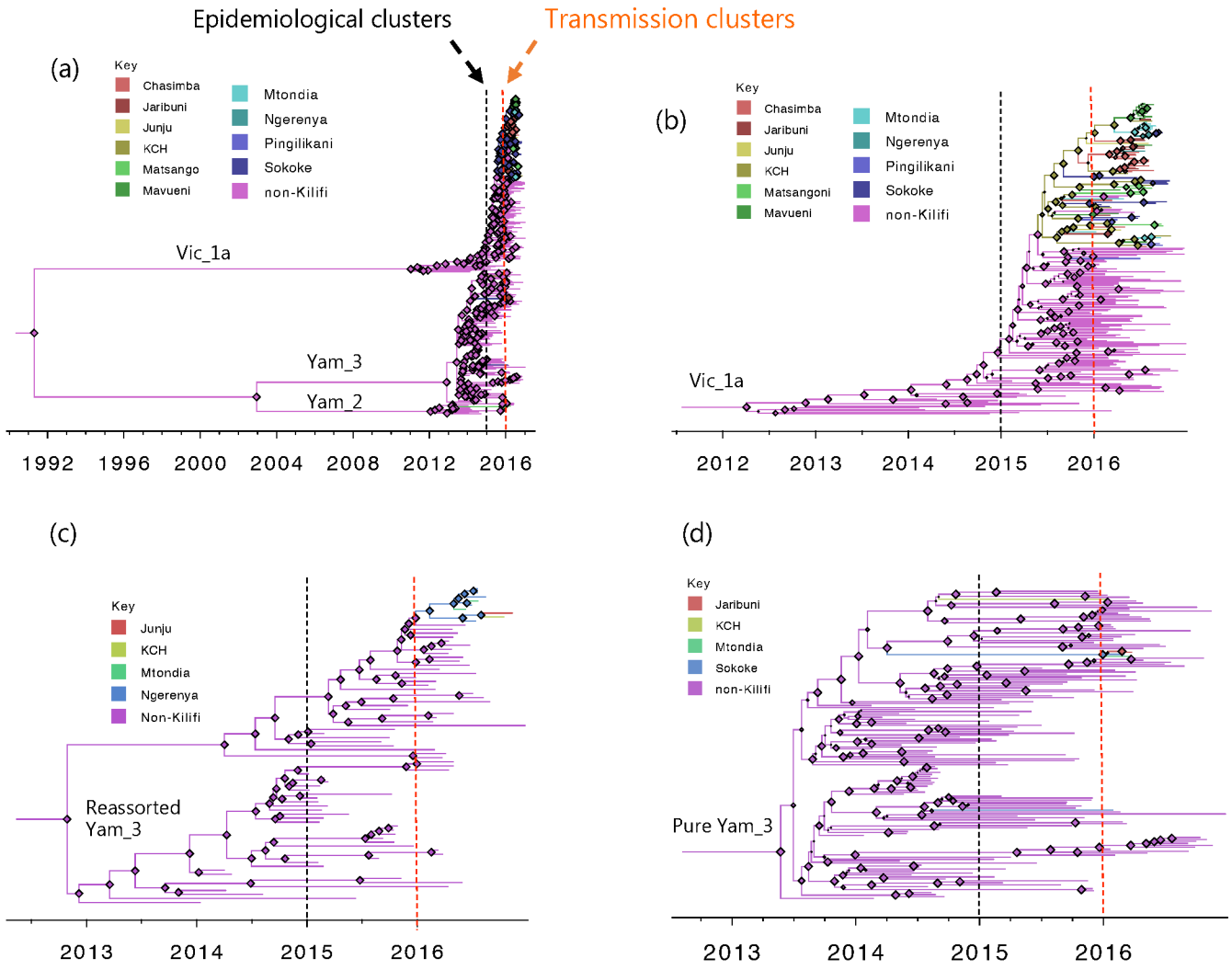


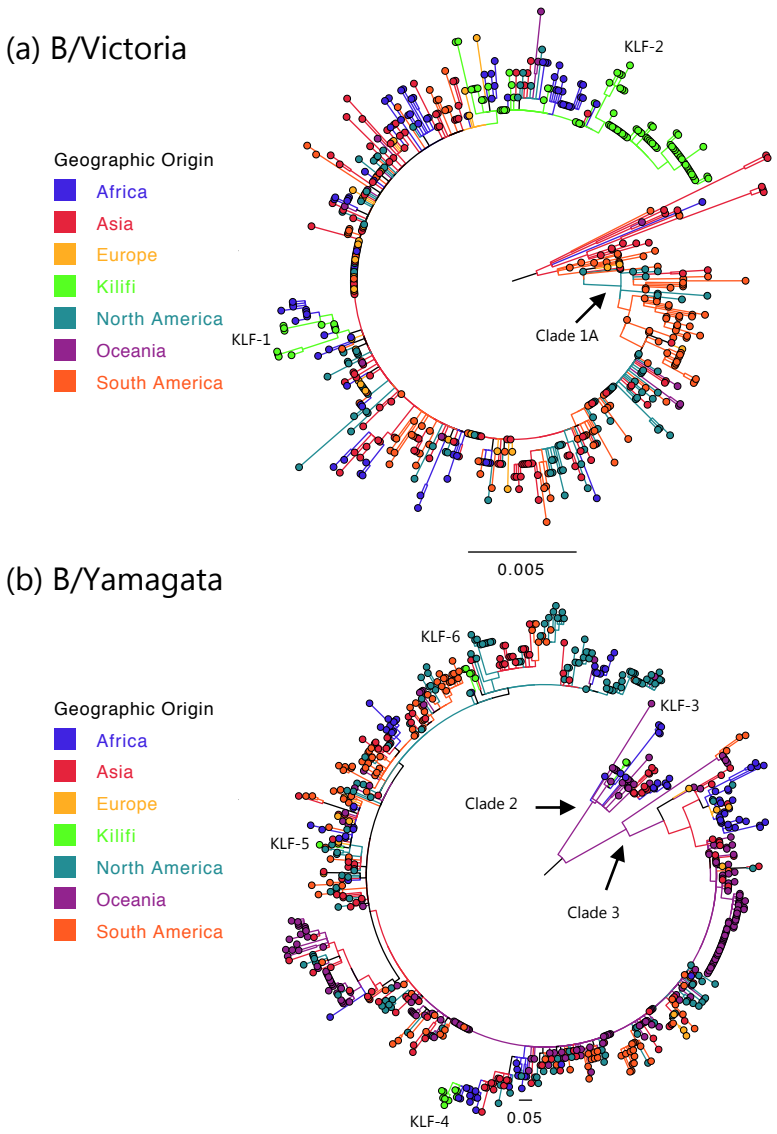



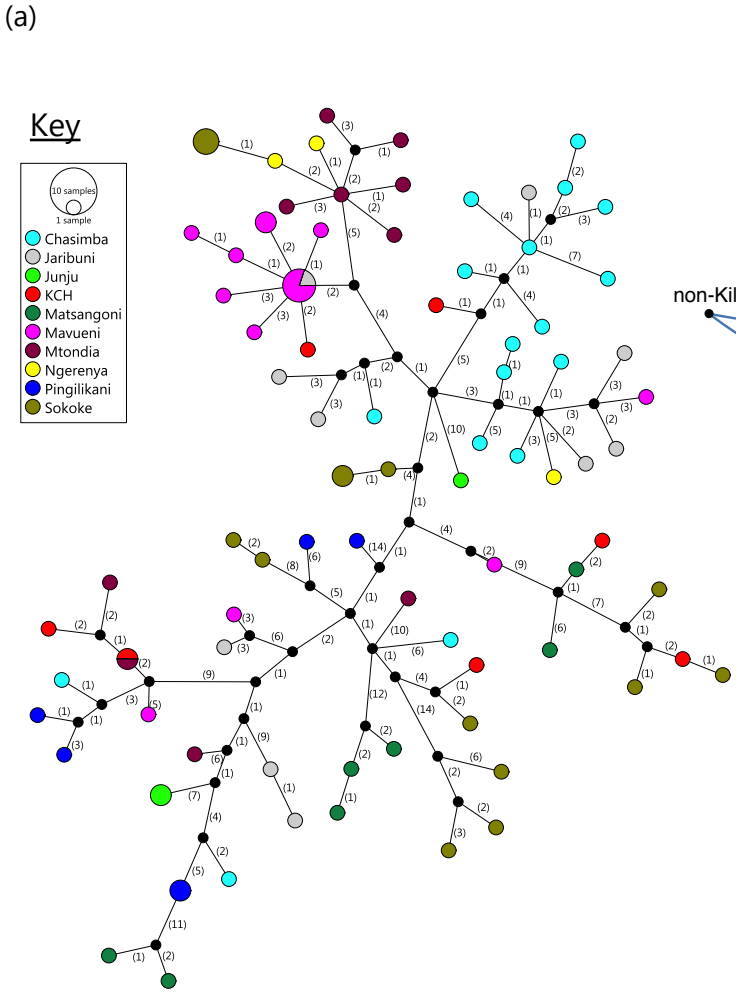



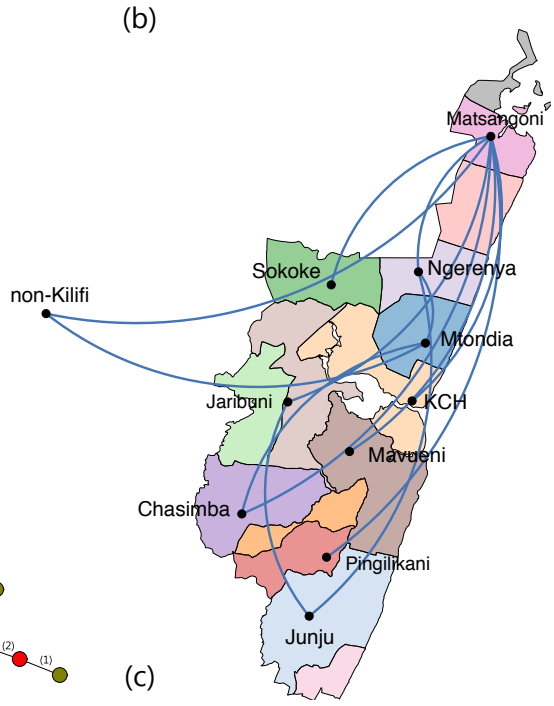











| Locations | Bayes Factor | Posterior |
|---|---|---|
| **Matsangoni <-> Chasimba** | 74582.65 | 0.99 |
| **Ngerenya <-> Junju** | 4657.53 | 0.99 |
| **Junju<-> Jaribuni** | 990.34 | 0.99 |
| **Matsangoni <-> non-Kilifi** | 334.88 | 0.98 |
| **Matsangoni<->Mavueni** | 100.61 | 0.96 |
| **Matsangoni<-> Jaribuni** | 76.05 | 0.94 |
| **Ngerenya <-> Matsangoni** | 68.41 | 0.94 |
| **Matsangoni <-> Pingilikani** | 62.15 | 0.93 |
| **Matsangoni <-> KCH** | 59.98 | 0.93 |
| **Chasimba <-> Mtondia** | 17.95 | 0.81 |
| **Sokoke <-> Matsangoni** | 17.88 | 0.81 |
| **Mtondia <-> non-Kilifi** | 11.28 | 0.73 |
| **Ngerenya <-> Mtondia** | 5.89 | 0.58 |

**Table 1.** Clinical and demographic characteristics of patients who were IBV positive and a comparison of those sequenced versus those not sequenced.

| Characteristic | All positives (n= 278) | Sequenced (n=111) | Not sequenced (n=167) | p value |
|---|---|---|---|---|
| Health facility | | | | 0.001 |
| Inpatient (KCH) | 11(4.0%) | 10 (9.0%) | 1 (1.0%) | |
| Outpatient (KHDSS) | 267 (96.0%) | 101 (91.0%) | 166 (99.0%) | |
| Age (Years) | | | | |
| Mean (SD)[€] | 9.6 (13.3) | 9.7 (15.2) | 9.6 (11.9) | 0.938 |
| Median (IQR) | 6 (2-12) | 6 (2-11) | 5 (2-13) | 0.924 |
| Age class (Years) | | | | 0.203 |
| 0-4 | 120 (43.2%) | 49 (44.2%) | 71 (42.5%) | |
| 5-14 | 109 (39.2%) | 47 (42.3%) | 62 (37.1%) | |
| 15-34 | 32 (11.5%) | 9 (8.1%) | 23 (13.8%) | |
| 35-64 | 13 (4.7%) | 3 (2.7%) | 10 (6.0%) | |
| ≥ 65 | 4 (1.4%) | 3 (2.7%) | 1 (0.6%) | |
| Gender | | | | 0.708 |
| Female | 159 (57.2%) | 65 (58.6%) | 94 (56.3%) | |
| Male | 119 (42.8%) | 46 (41.4%) | 73 (43.7%) | |
| Clinical Symptoms | | | | |
| Fever | 220 (79.1%) | 89 (80.2%) | 131 (78.4%) | 0.727 |
| Cough* | 260 (97.4%) | 98 (97.0%) | 162 (97.6%) | 0.781 |
| Nasal discharge* | 204 (76.4%) | 79 (78.2%) | 125 (75.3%) | 0.586 |
| Breathing difficulty | 28 (10.1%) | 15 (13.5%) | 13 (7.8%) | 0.120 |
| Viral load (Ct value) | | | | |
| Mean (SD)[€] | 27.7 (3.2) | 25.0 (2.3) | 29.5 (2.4) | <0.001 |
| Median (IQR) | 27.6 (25.7-29.6) | 25.2 (23.5-26.5) | 29.2 (28.0-31.0) | <0.001 |

* n = 267, 101 and 166 for the categories; all positives, those sequenced and those not sequenced, respectively. Symptoms data were unavailable for inpatients.

[€]SD stands for Standard of Deviation

**Table 2.** The genomic evolutionary characteristics of the sequenced IBV strains in the surveillance stratified by phylogenomic cluster.

| **Phylogenomic cluster** | **# isolates (Kilifi)** | **tMRCA** | | | **Substitution rate × $10^{-3}$** | | |
|---|---|---|---|---|---|---|---|
| | | Mean | Low 95% HPD | High 95% HPD | Mean | Low 95% HPD | High 95% HPD |
| Pure B/Vic/clade1A | 209 (93) | Mar-2012 | Jul-2011 | Oct-2012 | 1.37 | 1.18 | 1.56 |
| Reassorted B/Vic/clade1A | 1 (1) | - | - | - | - | - | - |
| Pure B/Yam/clade2 | 46 (2) | Oct-2011 | Nov-2010 | Aug-2012 | 1.36 | 1.02 | 1.75 |
| Pure N/Yam/clade3 | 157 (6) | May-2013 | Mar-2013 | Aug-2013 | 1.40 | 1.27 | 1.53 |
| Reassorted B/Yam/clade3 | 79 (9) | Oct-2012 | Mar-2012 | April 2013 | 1.54 | 1.33 | 1.76 |

tMRCA, the Most Recent Common Ancestor; HPD, Highest Posterior Density; Yam for Yamagata, Vic for Victoria, Substitution rate units are nucleotides/site/year.

**Table 3.** Divergence times and rates of nucleotide substitution of gene segments of influenza B virus

| | tMRCA | | | Substitution rate $\times 10^{-3}$ | | |
|---|---|---|---|---|---|---|
| **Segment** | Mean | Low 95% HPD | High 95% HPD | Mean | Low 95% HPD | High 95% HPD |
| **PB2** | Feb-1982 | Nov-1977 | Oct-1985 | 1.53 | 1.29 | 1.78 |
| **PB1** | Jun-1979 | Jan-1976 | Oct-1982 | 1.33 | 1.16 | 1.51 |
| **PA** | Nov-1979 | Jul-1974 | Sep-1984 | 1.46 | 1.28 | 1.66 |
| **HA** | Dec-1980 | May-1976 | Jun-1985 | 1.97 | 1.70 | 2.24 |
| **PA** | Nov-1979 | Jul-1974 | Sep-1984 | 1.46 | 1.28 | 1.66 |
| **NA** | Mar-1983 | Sep-1980 | Aug-1985 | 1.92 | 1.67 | 2.21 |
| **MP** | Mar-1983 | Jun-1970 | Apr-1987 | 1.44 | 1.26 | 1.69 |
| **NS** | Jul-1968 | Sep-1960 | Jul-1975 | 1.24 | 1.00 | 1.48 |

tMRCA, the Most Recent Common Ancestor; HPD, highest posterior density