Phylogenomic analysis of Uganda influenza type-A viruses to assess their relatedness to the vaccine strains and other Africa viruses: a molecular epidemiology study.

Grace Nabakooza, MSc^{1,2,3¶}, David Collins Owuor, PhD⁴, Zaydah R. de Laurent, MSc⁴, Nicholas Owor, BSc⁵, John Timothy Kayiwa, MSc⁵, Daudi Jjingo, PhD^{6,7}, Charles Nyaigoti Agoti, PhD⁴, David James Nokes, PhD^{4,8*}, David Patrick Kateete, PhD^{1,2}, John Mulindwa Kitayimbwa, PhD^{2,3}, Simon David William Frost, DPhil^{9,10*}, Julius Julian Lutwama, PhD^{5*}

1. Department of Immunology and Molecular Biology, Makerere University, Old Mulago Hill Road, P.O Box 7072, Kampala, Uganda.
2. Makerere University/ UVRI Centre of Excellence in Infection and Immunity Research and Training (MUII-Plus), Plot No: 51 -59 Nakiwogo Road, P.O. Box 49, Entebbe, Uganda.
3. Centre for Computational Biology, Uganda Christian University, Plot 67-173, Bishop Tucker Rd, P.O BOX 4, Mukono, Uganda.
4. Department of Virus Epidemiology and Control, KEMRI-Wellcome Trust Research Programme, P.O Box 230-80108, Kilifi, Kenya
5. Department of Arbovirology Emerging & Re-Emerging Infectious Diseases, Uganda Virus Research Institute (UVRI), Plot No: 51 -59, Nakiwogo Road, P.O. Box 49 Entebbe, Uganda.
6. College of Computing & College of Health Sciences, Makerere University, P.O Box 7072, Kampala, Uganda.
7. The African Centers of Excellence in Bioinformatics (ACE), Makerere University, P.O Box 7072, Kampala, Uganda.
8. School of Life Sciences and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry CV47AL, United Kingdom.
9. Microsoft Research, 4820 NE 36th Street, Redmond, Washington, 98052, Redmond, Washington, United States.

25 10. London School of Hygiene & Tropical Medicine (LSHTM), Keppel St, Bloomsbury, London,
26 WC1E 7HT, United Kingdom.

27 ¶Corresponding author: Grace.Nabakooza@mrcuganda.org

28 *Authors contributed equally.

29 ♦Professor.

30

31 **ABSTRACT**

32 **Background:** Genetic characterisation of circulating influenza viruses is essential for vaccine
33 selection and mitigation of viral transmission. The current scantiness of viral genomic data and
34 underutilisation of advanced molecular analysis methods on influenza viruses circulating in Africa
35 has limited their extensive study and representation in the global influenza ecology. We aimed to
36 sequence influenza type-A viruses (IAVs) that previously circulated in Uganda and characterised
37 their genetic relatedness to the vaccine viruses and publicly available Africa IAVs.

38 **Methods:** This was an observational study nested to the Uganda national influenza surveillance
39 programme. We used Next-generation sequencing to locally generate genomes from 116
40 A(H1N1)pdm09 and 118 A(H3N2) viruses collected between 2010 and 2018 from 7 districts across
41 Uganda. A total of 206 hemagglutinin (HA), 207 neuraminidase (NA), and 213 matrix protein (MP)
42 sequences were genetically compared to the WHO-recommended vaccines and other viruses
43 isolated from Africa since 1994. Viral temporal and spatial divergence and circulating genetic clades
44 were characterised using phylogenetic methods.

45 **Findings:** We successfully generated gene sequences for 91·9% (215/234) viruses. Uganda
46 A(H1N1)pdm09 and A(H3N2) virus HA, NA, and MP proteins had 96·36-99·09%, 96·49-99·39%,
47 and 97·48-99·95% amino acid similarity, respectively, to vaccines recommended from 2010 through
48 2020. The local viruses incorporated amino acid substitutions (AAS) in their antigenic, receptor
49 binding, and glycosylation sites each year causing them to antigenically drift away from vaccines.
50 For seasons when vaccine formulations differed, Uganda IAV antigenic sites had 1-2 extra AAS
51 relative to the Southern than Northern hemisphere vaccine viruses. All Uganda IAVs carried the
52 adamantane-resistance marker S31N but not the neuraminidase inhibitor (NAI) resistance markers
53 H274Y and H275Y. However, some A(H1N1)pdm09 viruses had permissive substitutions V234I,
54 N369K, and V241I typical of NAI-resistant viruses. The 2017-2018 A(H1N1)pdm09 viruses belonged
55 to global genetic clade 6B.1, while the A(H3N2) viruses isolated in 2017 belonged to clades 3C.2a

and 3C.3a. Uganda IAVs obtained before 2016 clustered distinctly from other Africa viruses while later viruses mixed with other Africa, especially Kenya and Congo, and global viruses. Several unique viral lineages (bootstrap >90) persisted in Uganda and other countries for 1-3 years.

Interpretation: The study reveals Uganda as part of the global influenza ecology with continuous importation, antigenic drift, and extensive local transmission of IAVs, presenting a potential risk of future outbreaks. For a country with limited health resources and where social distancing is not sustainable, viral prevention by vaccination should be prioritized. The notable viral diversity in Africa is a warning to countries to broaden and incorporate genome analysis in routine surveillance to monitor circulating and detect new viruses. This knowledge can inform virus selection for vaccine production and assist in developing cost-effective virus control strategies.

Funding: This work was supported by the Makerere University-Uganda Virus Research Institute Centre of Excellence for Infection and Immunity Research and Training (MUII). MUII is supported through the Developing Excellence in Leadership, Training and Science (DELTAS) Africa Initiative (Grant no. 107743). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS), 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 (Grant no. 107743) and the UK Government. The work was also funded in part by a Wellcome Trust grant (102975).

74 INTRODUCTION

75 Novel influenza type-A viruses (IAVs) cause human respiratory infections that lead to social
76 lockdowns, economic losses, and millions of deaths¹. Genetic characterisation of IAVs is important
77 to differentiate them from other viruses causing similar clinical symptoms for effective viral control
78 and prevention. Seasonal influenza-related illnesses kill 290000-650000 people globally per year,
79 mostly in sub-Saharan Africa². Influenza accounts for 21·7% and 10·1% of the influenza-like
80 illnesses (ILI) and severe acute respiratory illnesses (SARI) in Africa, respectively, and circulates all-
81 year-round with discernible influenza peaks in North and South Africa³. Uganda's annual epidemics
82 have two major peaks between May and November and usually constitute multiple IAV types and
83 subtypes responsible for 13% and 6% of the ILI and SARI cases, respectively⁴.

84 Vaccination and antiviral treatment are the best ways to prevent and control viral transmission⁵.
85 However, the multi-segmented IAVs continuously mutate, especially in the antigenic surface genes,
86 hemagglutinin (HA) and neuraminidase (NA), causing them to drift from vaccines with possible
87 emergence of drug-resistant viruses⁶. Vaccine formulations for the Northern (NH) and Southern
88 hemisphere (SH) are updated annually to match circulating viruses⁵. Countries select appropriately
89 licensed vaccines based on their viral circulation patterns⁵.

90 Well-sampled and developed collect sufficient data on IAV evolution patterns, drug sensitivity⁶, and
91 circulating and emerging genetic clades⁷ for vaccine selection⁵. However, the high cost of whole-
92 genome sequencing (WGS) and limited phylogenomic analysis capacity restricts the genetic
93 characterisation of IAVs in Africa. We searched Global Health, PubMed, and Web of Science on 25th
94 March 2020 and found only five African studies⁸⁻¹² had sequenced IAV whole genomes (WGs). Two
95 of these sequenced 59 A(H3N2)⁸ and 19 A(H1N1)pdm09⁹ Uganda viruses, sampled in 2008-2009
96 and 2009-2011, respectively, with sequencing conducted in the United States.

97 We aimed to explore the feasibility of using Next-Generation Sequencing (NGS) in a low-income
 98 setting to generate WGs of Uganda IAVs sampled in 2010-2018 and compare the generated
 99 sequences with vaccine viruses and publicly available Africa IAV sequences collected in 1994-2019.
 100 We analysed HA carrying antigenic sites which trigger host immune responses and NA and matrix
 101 protein (MP) genes targeted by antiviral drugs. This work birthed an East African network of
 102 influenza molecular epidemiologists which we hope to expand across Africa.

103 **METHODS**

104 **Design of influenza surveillance and source of swabs**

105 The Uganda Virus Research Institute (UVRI-NIC) National Influenza Centre in Entebbe implements
106 a clinic and hospital-based surveillance at thirteen peri-urban and densely populated sites in seven
107 districts located in Northwest, Western, Central, and Eastern Uganda (sFigure 1, Appendix pp 6)⁴.

108 Nasal and oropharyngeal swabs were collected in 2010-2018 seasons from ILI and SARI patients of
109 all ages with their respective sociodemographic data⁴. The swabs were tested, typed for influenza A
110 and B, and the IAV positives subtyped for seasonal [A(H1N1) and A(H3N2)] and pandemic
111 A(H1N1)pdm09 influenza using the Centers for Disease Control and Prevention's (CDC) real-time
112 reverse-transcription polymerase chain reaction (rRT-PCR) protocols and primers (Atlanta,
113 Georgia)¹³.

114

115 **Swab selection for whole-genome sequencing (WGS)**

116 We aimed to sequence 100 whole genomes (WGs) per subtype [A(H1N1)pdm09 and A(H3N2)] due
117 to financial constraints. Available swabs from 971 IAV patients with real-time PCR cycle threshold
118 (CT_≤35) were randomised, stratified by subtype and year of collection, and selected every first
119 fifteen swabs. Twenty-four percent (234/971) of the swabs [116 A(H1N1)pdm09 and 118 A(H3N2)]
120 were sequenced at a KEMRI-Wellcome Trust Programme collaborating laboratory in Kilifi, Kenya.

121

122 **Viral RNA isolation and amplification**

123 Viral ribonucleic acid (RNA) was extracted from 140µL swab sample using the QIAamp Viral RNA
124 Mini extraction kit and manufacturer's protocol (Qiagen, Hilden, Germany). The extracted RNA was

reverse transcribed and whole genome amplified using the multi-segment real-time polymerase chain reaction (M-RTPCR)¹⁴ and universal IAV Uni/Inf primers at standardised thermocycling conditions (Appendix pp 3)¹¹.

Next-generation sequencing

The M-RTPCR amplicon libraries were prepared using the Nextera XT DNA library preparation kit and protocol (Illumina, San Diego, California, USA), cleaned using the 0.8× AMPure XP beads, quantified on the Qubit 3.0 fluorometer using the dsDNA High sensitivity kit (Invitrogen, Carlsbad, California, USA). Library size distributions were assessed using the Agilent Technology 2100 Bioanalyzer and the High Sensitivity DNA kit (Agilent Technologies, Santa Clara, California, USA). Samples with a broad fragment size spectrum (>250 bp) were normalized manually to 2nM. 5µL per sample library were pooled, denatured using Sodium Hydroxide (NaOH), and diluted to 12.5 pmol. Diluted libraries were spiked with 5% Phi-X control (Illumina, San Diego, CA, USA) and sequenced using the Illumina MiSeq (Illumina Inc., San Diego, California, USA) generating 2x250bp paired reads per sample.

We assessed sequencing efficiency based on the read (coverage depth) and gene segment (genome length) count generated per swab sample.

Sequence quality control and assembly

Raw reads were de-duplicated and decontaminated, and clean reads assembled using the reference-based FLU module of the Iterative Refinement Meta-Assembler (IRMA) v0.6.7¹⁵ at default settings (Appendix pp 4-5). The A(H1N1)pdm09 and A(H3N2) virus assemblies were compared with A/California/7/2009 and A/Perth/16/2009 vaccine viruses, respectively.

148

149 **Genetic characterization of surface and matrix proteins**

150 Gene sequences for SH and NH or both (SNH) vaccine and clade reference viruses per subtype
151 were downloaded from the Global Initiative on Sharing All Influenza Data (GISAID) EpiFlu database
152 (<https://www.gisaid.org/>; accessed on 26th February 2020). Uganda and vaccine virus sequences per
153 subtype were aligned using MUSCLE v3.8.1551¹⁶.

154 Amino acid substitutions (AAS) in the antigenic sites (A, B, C, D, and E) of A(H1N1)pdm09¹⁷ and
155 A(H3N2)¹⁸ virus HA1 proteins were identified manually. AAS in complete HA, NA, and MP protein
156 sequences and their respective functions were identified using the Flusurver tool
157 (<http://flusurver.bii.a-star.edu.sg>; accessed on 30th March 2021).

158

159 **Phylogenetic divergence**

160 Uganda HA, NA, and MP gene sequences per subtype were aligned and maximum likelihood (ML)
161 trees reconstructed using IQtree v1.6.11¹⁹ with a GTR+G4 model and 1000 bootstraps. Trees were
162 rooted using the oldest sequence in the dataset and visualized in ggtree v2.4.1²⁰ and Figtree v1.4.4
163 (<http://tree.bio.ed.ac.uk/software/figtree/>).

164

165 **Phylogenetic clustering and genetic clade classification**

166 Uganda, clade reference, and vaccine virus sequences per subtype were aligned and ML trees
167 reconstructed as above. Phylogenetic clusters were inferred from ML trees using PhyCLIP v2.0²¹
168 and clusters confirmed as genetic clades using the European Centre for Disease Prevention and
169 Control (ECDC) reference-based method⁷. PhyCLIP uses linear integer programming to cluster
170 sequences that may share epidemiological patterns, while the ECDC method classifies a sequence
171 or cluster of sequences to a genetic clade based on AAS shared in their HA1 or HA2 proteins.

For Africa analysis, HA, NA, and MP gene sequences were downloaded from GISAID accessed on 27th February 2020. Sequences with >100 bps shorter or longer than the actual gene size and ambiguous “N” bases were excluded. The remaining A(H1N1)pdm09 (496 H1, 443 N1, and 278 MP) and A(H3N2) (718 H3, 675 N2, and 439 MP) virus sequences were aligned using a codon-aware aligner (<https://github.com/veg/hyphy-analyses/tree/master/codon-msa>) and ML trees reconstructed as above.

178

179 **Statistical analysis**

Patients’ sociodemographic data were compared between the UVRI-NIC and study using chi-square and Wilcoxon t-test in R v3.6.3 (<https://www.r-project.org>).

182

183 **Ethics**

This study was approved by the Makerere University School of Biomedical Sciences Research and Ethics Committee (SBS-REC) (ref:SBS-577) and Uganda National Council of Science and Technology (UNCST) (ref:HS2519).

187

188 **Role of funding source**

The funder of the study played no role in study design, data collection, data analysis, data interpretation, or writing of the report.

190

191 RESULTS

192 Sociodemographic characteristics of sampled patients

193 The UVRI-NIC laboratory tested 18353 patients between 22nd October 2010 and 9th May 2018.
 194 Thirteen-percent (2404/18353) were positive for influenza, 69·88% (1680/2404), 29·62% (712/2404),
 195 and 0·17% (4/2404) had influenza A, B, and A/B co-infection, respectively. Eight (0·33%) viruses
 196 lacked viral type data (sFigure 2, Appendix pp 7). IAV positives included 67·08% (1127/1680)
 197 A(H3N2), 32·2% (541/1680) A(H1N1)pdm09, 0·12% (2/1680) AH1/H3 co-infections, and 0·59%
 198 (10/1680) lacked subtype data.

199 On average, 13 swabs were sampled per subtype per year except 2012, 2016, and 2018 with 2, 1
 200 A(H1N1)pmd09, and no A(H3N2) patient swabs available, respectively (sTable 1, Appendix pp 8).
 201 Three A(H1N1)pdm09 and one A(H3N2)] sampled patients' swabs lacked sociodemographic data.
 202 Of the 230 swabs with data, 65·22% (150/230) and 34·78% (80/230) were from ILI and SARI cases,
 203 respectively. The number of ILI and SARI cases sampled per subtype were comparable (Figure 1A
 204 and 1B). More viruses were sequenced from males 56·09% (129/230) than females 43·16%
 205 (101/230). IAVs from under 5-year-olds, 72% (166/230), were most sequenced and 0·87% (2/230)
 206 from the elderly (≥ 50 years old) (sTable 2, Appendix pp 8). The ILI, SARI, male, and female counts
 207 were independent (chi-square test $p > 0·05$) and patients' mean ages not significantly different
 208 (Wilcoxon test $p > 0·05$) between the sampled and UVRI-NIC per subtype (sFigure 3 and 4, Appendix
 209 pp 9).

210 The majority, 77·35% (181/234) of swabs sampled for WGS were from Central (Kampala and
 211 Entebbe), 11·54% (27/234) from Western (Mbarara and Fort Portal), 5·56% (13/234) Eastern
 212 (Tororo), and 5·13% (12/234) Northwest (Arua and Koboko) [Figure 1C and 1D) as in UVRI-NIC
 213 sTable 3, Appendix pp 10)].

214

215 **Sequencing efficiency**

216 All sampled 234 swabs were analysed and read counts reported as averages.

217 The MiSeq generated 569435 paired reads per sample (data not shown). Following quality control
218 (QC), 266020 (70229-909340) clean reads per sample were put in IRMA, 265868 (70150-908946)
219 passed IRMA's QC, 213809 (1381-908234) matched flu references, and 113164 (777-461172)
220 paired reads were used to build assemblies (Figure 2A).

221 The number of assembled reads decreased with an increase in gene segment size. The shortest,
222 MP and non-structural protein (NS), had 25101 (28-91585) and 19341 (41-79167) assembled reads,
223 respectively. The NA, HA, and nucleoprotein (NP) had 18653 (74-71377), 14466 (35-74550), and
224 14151 (69-74415) reads assembled, respectively. The polymerase subunits: PA, PB2, and PB1 had
225 10465 (31-56243), 7885 (15-47444), and 4497 (8-31354) reads assembled, respectively (Figure
226 2B).

227 We successfully sequenced and assembled viruses from 96.58% (226/234) of the swabs (Figure 3).
228 Eleven WGs (88 segments) with low coverage depth (<100) were excluded remaining with 215
229 viruses. 89.77% (193/215) of these were WGs, spanning 100% and >96.7% nucleotides in the
230 coding sequences (CDS) and complete genome of A/California/7/2009(H1N1pdm09) and
231 A/Perth/16/2009(H3N2) vaccine viruses. The remaining 10.23% (22/215) viruses had partial
232 genomes with complete CDS for 2-7 segments. The overall WG recovery rate was 85.4% (193/226).
233 Two viruses sampled as A(H1N1)pdm09 matched IRMA's A(H3N2) references and were included in
234 the A(H3N2) analysis.

235 The newly-generated sequences were deposited in GISAID under the accessions EPI_ISL498819-
236 EPI_ISL498931 [A(H1N1)pdm09] and EPI_ISL498934-EPI_ISL499037 [A(H3N2)].

237

238 **Antigenic drift in Uganda viruses**

239 Uganda IAV HA1 proteins continuously drifted away from 2010-2020 season vaccines (Table 1).
 240 In 2012, 2013, 2019, and 2020 when formulations differed, Uganda viruses carried 1-2 extra AAS
 241 relative to SH than NH vaccines viruses. Since Uganda's largest part lies north of the equator, the
 242 substitutions described below are relative to NH and SNH vaccines for the sampled 2010-2018
 243 [A(H1N1)pdm09] and 2010-2017 [A(H3N2)] seasons.

244 We observed 16 unique amino acid substitutions (UAAS) across the five antigenic sites amongst the
 245 107 A(H1N1)pdm09 viruses (Table 1A). Ranking from the most variable site, Sa, Ca2, Sb, Ca1, and
 246 Cb had 4, 4, 3, 3, and 2 AAS, respectively. Substitution S164T, S185T, S203T, H138R, and S74R
 247 was most frequent at site Sa, Sb, Ca1, Ca2, and Cb, respectively. Seventy-percent (54/77) and
 248 100% of 2010-2016 viruses carried S185T and S203T, respectively. Ninety-percent (27/30) of the
 249 2017-2018 viruses carried S164T and S74R.

250 There were 93 UAAS across the five antigenic sites amongst the 99 A(H3N2) viruses (Table 1B).
 251 Ranked most variable was site A, D, B, C, and E with 24, 21, 20, 15, and 13 UAAS, respectively.
 252 Substitutions V186G and N145S were the most frequent in 47·47% (47/99) and 41·41% (41/99) of
 253 the all 2010-2017 A(H3N2) viruses. AAS at site A (K144N, N128A, R142G, and N145S), B (Q156H,
 254 V186G, P198S, and P194L), D (T212A, S214I, Y219S, and F219S), and E (K62E) were
 255 incorporated in all viruses analysed in a given season.

256 We observed mutated receptor binding sites (RBS) (A186T, S190V, and D222E) and glycosylation
 257 sites (GS) (A186T and N125D) of A(H1N1)pdm09 viruses previously described^{18,22}. A(H3N2) viruses
 258 had more mutated RBS in the 130-loop (135, 138, and 140), 190-helix (186, 192, 194, and 196),
 259 220-loop (219), and two GS (122 and 144) (Table 1B).

260

261 **Mutational landscape of complete HA, NA, and MP proteins**

262 Uganda HA (H1), NA (N1), and MP proteins had 94, 81, and 21 UAAS, respectively, and were
 263 97·85-99·09%, 98·04-98·97%, and 98·95-99·95% similar to A/California/7/2009,

264 A/Michigan/45/2015, and A/Brisbane/02/2018 vaccines at amino acid level (sTable 4, Appendix pp
265 11-12).

266 N1 proteins lacked the neuraminidase inhibitors (NAIs) resistance substitution H275Y. However,
267 7.55% (8/106) N1 proteins had T362I (1), I117M (2), Y155H (2), and V234I (3) associated with
268 reduced susceptibility to NAIs in vitro²².

269 Uganda HA (H3), NA (N2), and MP proteins had 161, 118, and 31 UAAS and amino acid similarity of
270 96.36-98.56%, 96.49-99.39%, and 97.48-99.86%, respectively, to A/Perth/16/2009,
271 A/Victoria/361/2011, A/Texas/50/2012, A/Switzerland/9715293/2013, A/Hong Kong/4801/2014,
272 A/Singapore/INFIMH-16-0019/2016, A/Switzerland/8060/2017, A/South Australia/34/2019, and
273 A/Kansas/14/2017 vaccines (sTable 5, Appendix pp 13-16).

274 N2 proteins lacked the NAI-resistance H274Y (N2 numbering) but 17.8% (18/101) carried Y155F
275 (2), E/D221D/K/E (15), and H374N (1) associated with reduced NAI susceptibility²².

276 All Uganda A(H1N1)pdm09 and A(H3N2) viruses incorporated adamantane-resistance marker (ARM)
277 S31N. Seven-percent (8/104) of A(H3N2) M2 proteins carried a secondary ARM V27A relative to
278 adamantane-susceptible A/New York/392/2004 virus (sFigure 5, Appendix pp 17).

279

280 **Temporal and spatial divergence of Uganda influenza A viruses**

281 Uganda viruses phylogenetically clustered according to the year of collection with multiple lineages
282 circulating annually (Figure 4). The A(H1N1)pdm09 viruses collected in 2010-2011 and 2012-2018
283 clustered distinctively. Two major lineages co-circulated in 2013 (Figure 4A). The H1 phylogeny
284 showed lineage 1 (shaded area in blue) circulated in 2013-2016, while lineage 2 (shaded area in
285 grey) persisted through 2018. In the N1 and MP phylogenies, lineage 1 and 2 emerged in 2013 and
286 2014, respectively, and co-circulated in 2014-2015. Lineage 2 dominated in 2016 and persisted
287 through 2018. The 2017-2018 A(H1N1)pdm09 viruses monophyletically clustered at the tip in all
288 phylogenies.

289 One dominant A(H3N2) virus lineage (1) (blue) circulated in 2013-2016. Lineage 2 and 3 (pink and
290 grey) emerged in 2015 and 2016, respectively, and co-circulated through 2017 (Figure 4B).

291 Viruses from different sites mixed in all phylogenies (sFigure 6, Appendix pp 18).

292

293 **Viral genetic clades circulating in Uganda**

294 PhyCLIP grouped A(H1N1)pdm09 viruses in eight clusters, seven of which belonged to four ECDC
295 global genetic clades (3, 5, 6, and 7) (Figure 5A). All four clades co-circulated in 2011. Clade 3 had
296 2010-2011 viruses carrying A134T and S183P substitutions. Clade 5 had 2011 viruses with D97N,
297 R205K, I216V, and V249L. One 2011 virus from Kisenyi clustered with A/St.Petersburgh/100/2011
298 in clade 7 with A197T, S143G, and K163I. Clade 6 viruses with AAS D97N, S185T, and S203T
299 dominated since 2012, diverging into 6A, 6B, and 6C. Clade 6B diverged into subclade 6B.1 with
300 S162N, I216T, and S84N dominating in 2017-2018. A novel subclade 6B.1a with T120A emerged in
301 November 2017 and circulated through 2018.

302 PhyCLIP grouped A(H3N2) viruses into 5 clusters belonging to global genetic clades 3 and 7 (Figure
303 5B). Clade 3 and 7 co-circulated in 2010-2012. Most viruses belonged to clade 3 carrying N145S
304 and V223I. Clade 3 diverged into 3B, 3C, 3C.2a, and 3C.3. Subclade 3C.2a diverged into subclades
305 3C.2a1a with T135K (May-June 2017) and 3C.2a1b with K92R and H311Q (May-November 2017).

306

307 **Phylogenetic relatedness of African viruses**

308 We define a group as a cluster with at least three Uganda sequences in the African gene tree.

309 Uganda A(H1N1)pdm09 viruses collected before 2016 clustered uniquely towards the root, while the
310 2017-2018 viruses mixed with others from Eastern, Central, Western, and Southern Africa (sFigure

311 7-9, Appendix pp 19-21). We observed 9 groups in each A(H1N1)pmd09 gene tree. Forty-four
 312 percent (4/9) of the H1 and N1 (2,4, 5, and 6), and MP (2, 4, 5, and 7) were unique to Uganda
 313 (sTable 6, Appendix pp 25).

314 Uganda A(H3N2) viruses collected in 2008-2016 and some before April 2017 clustered uniquely
 315 closest to the root (sFigure 10-12, Appendix pp 22-24). There were 10, 11, and 15 groups in the H3,
 316 N2, and MP gene trees, respectively. Two H3 (1 and 2), one N2 (6), and eight MP (1, 3, 5, 6, 8, 9,
 317 12, and 13) groups had only Uganda viruses (sTable 6, Appendix pp 25). Notably, the Makerere
 318 Walter Reed project (MWRP)⁸ viruses collected in 2008-2009 clustered separately from our new
 319 viruses.

320 If not clustered alone, Uganda viruses interspersed with viruses from Eastern, Central, Western, and
 321 Southern Africa. A(H1N1)pdm09 viruses clustered uniquely with viruses from Seychelles, Burkina
 322 Faso, Sierra Leone, CoteD'ivoire, Mozambique, and Egypt, while A(H3N2) viruses grouped uniquely
 323 with Reunion and Rwanda viruses. Interestingly, five unique H3 lineages co-circulated in 2019; one
 324 Kenyan and four in West Africa (not shown). Group details are described in sTable 6 (Appendix pp
 325 25).

326 Accessions for sequences used in this study are provided in sTable 7 (Appendix pp 26-27).

DISCUSSION

We have demonstrated the feasibility of NGS whole-genome sequencing of IAVs in low-income settings. Analysis of the generated sequences highlighted antigenic drift, multiple introductions, and local transmission of both A(H1N1)pdm09 and A(H3N2) influenza viruses in Uganda. All Uganda viruses carried the M2 adamantane-resistance marker S31N. Africa viruses were genetically similar but several unique viral lineages (with bootstrap >90) were observed in Uganda and other countries persisting for 1-3 years.

We successfully generated 96.58% (226/234) genomes and recovered 85.4% (193/226) WG directly from respiratory swabs. Our WG recovery rate is comparable to the 82-88% reported in developed countries^{14,23} and better than 47.3% and 66% reported in Scotland²⁴ and Mexico²⁵, respectively. 10-20% of frozen swabs fail NGS as the RNA quality gets compromised during collection, storage, and pre-sequencing analysis. The smallest genes, MP and NS, were sequenced at a greater depth and the least for the polymerase segment 2 (PB1) as previously reported in Mexico and France^{23,25}.

Sequence analysis revealed several AAS in the HA, NA, and MP proteins of Uganda IAVs. Notably, antigenic sites Sa and Ca2 for A(H1N1)pdm09, and A, D and B for A(H3N2) viruses varied most. The A(H1N1)pdm09 virus substitutions S164T and S74R, also dominant the 2017-2018 Kenya viruses²⁶. The frequent S203T at site Ca1 has been reported globally but its function is unknown²⁷. We observed RBS D222E, speculated to increase infection severity, in 5.2% (4/77) of the 2010-2011 A(H1N1)pdm09 viruses²⁷. H1 (P100S, S220T, I338V) and N1 (V106I, N248D) earlier detected in 2009-2011 in Uganda viruses against A/California/7/2009⁹ continued to circulate through 2017. All 2010-2013 Uganda A(H3N2) viruses had mutated antigenic sites also observed in Kenya²⁸.

Uganda viruses lacked H274Y or H275Y confirming previous reports of 98% of global IAVs being sensitive to NAIs⁶. NAIs are not publicly available making their intake low in Uganda. However, some N1 proteins carried permissive AAS V234I, N369K, and V241I known to

counteract the effect of H275Y. These AAS enhance the NA surface expression and enzymatic activity, hence increasing viral fitness. Uganda IAVs NA active and catalytic sites²⁹ were highly conserved indicating lack of pharmacological selection pressure.

The A(H1N1)pdm09 viruses formed steeper ladder-like phylogenies than A(H3N2) viruses, indicating directional selection due to immune escape³⁰. Further WG analysis to identify processes such as reassortment and selection pressures driving the observed genetic variations is ongoing.

Phylogenetic analysis also revealed the circulation of multiple viral lineages and global genetic clades in Uganda as observed worldwide^{7,8,11}. New viral lineages and clades were observed in between seasons indicating multiple viral introductions into Uganda per year. Novel A(H3N2) virus subclades 3C.2a1a and 3C. 2a1b were introduced in 2017. Uganda viruses phylogenetically clustered with no geographical signal showing an extensive viral mixing and local transmissions.

Uganda A(H1N1)pdm09 and A(H3N2) viruses collected before 2015 and 2016, respectively, clustered distinctively from most Africa viruses. However, this could be due to insufficient representative viruses sequenced in earlier years across Africa. Viruses collected later interspersed with those from Eastern, Central, Southern, and Western Africa indicative of viral genetic similarity. Also, 2010-2018 Uganda viruses clustered with global viruses indicating that Uganda contributes to the African and global influenza ecology.

The coronavirus disease 2019 (COVID-19) pandemic revealed viral exchange among countries through borders, for example, Kenya and Uganda. Future studies should explore advanced phylodynamic and demographical models to determine key importers and migration rates of IAVs in Uganda.

This is the first and largest study to sequence Uganda A(H1N1)pdm09 and A(H3N2) virus WG locally, and internationally to describe a detailed temporal and spatial genetic diversity

and evolution patterns of Africa IAVs. Our sequences add significantly to publicly available Africa data. However, we used pre-collected data with a biased geographical sampling. Therefore, we did not consider geographical location in the swab randomisation. We sequenced only 13 and 12 viruses from Eastern and Northwest, respectively. Due to financial constraints, we sampled sequenced only 24.1% (234/971) of available viruses, 8.12% (19/234) of which failed quality control and NGS. We did not include clinical data or do phenotypic analysis to assess the effect of the substitutions observed on viral virulence, pathogenicity, and transmissibility.

This study provides a platform for larger future studies and highlights the potential of genomics surveillance to improve viral detection and disease management. Existing African surveillance programmes should prioritize routine sequencing and genome analysis to monitor circulating IAVs. Our findings will inform Uganda's public health, use of NAIs as a prophylactic treatment, and decision to formulate influenza vaccination programs, especially for high-risk groups like children, pregnant mothers, and the elderly.

Authors' contributions

GN proposed the idea and designed the study with supervision from SDWF, JMK, and DPK. NO, JTK and JJL manage the UVRI-NIC influenza surveillance and provided both the samples and patient epidemiological data. DCO, ZRD, CNA, and DJN hosted, provided protocol, and trained GN on laboratory and whole-genome sequencing procedures. GN analysed the data with guidance from SDWF, interpreted results, and wrote the manuscript. SDWF and DJ guided on result presentation. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication, and GN and SDWF verified the underlying data. All authors revised and edited the manuscript's preceding drafts and approved the final manuscript.

403

404 **Data sharing**

405 All sequences generated in this study with their respective metadata were submitted to
406 GISAID EpiFlu (<https://www.gisaid.org/>) under accessions EPIISL498819 - EPIISL498931
407 [A(H1N1pdm09)], and EPIISL498934 - EPIISL499037 [A(H3N2)], and will be available
408 publicly.

409

410 **Acknowledgments**

411 We thank the participants for providing samples. We extend our gratitude to the effort and
412 contribution of the surveillance clinical officers and nurses at the sentinel sites, and the
413 laboratory team at the Department of Arbovirology Emerging & Re-Emerging Infectious
414 Diseases (UVRI) in sample and data collection, processing, archiving, and permitting us to
415 use them. The authors acknowledge the authors, originating and submitting laboratories of
416 the GISAID influenza viral sequences analysed in this study. All data submitters can be
417 contacted through the GISAID EpiFlu platform.

418

419 **Conflict of interest**

420 DJN received a Wellcome Trust Senior Investigator Award (Grant No. 102975).

421

422 **REFERENCES**

- 423 1 Taubenberger JK, Morens DM. Influenza: the once and future pandemic. *Public Health*
424 *Rep* 2010; **125 Suppl 3**: 16–26.
- 425 2 Iuliano AD, Roguski KM, Chang HH, *et al.* Estimates of global seasonal influenza-
426 associated respiratory mortality: a modelling study. *The Lancet* 2018; **391**: 1285–300.
- 427 3 Radin JM, Katz MA, Tempia S, *et al.* Influenza surveillance in 15 countries in Africa, 2006-

2010. *J Infect Dis* 2012; **206 Suppl 1**: S14-21.
- 4 Lutwama JJ, Bakamutumaho B, Kayiwa JT, *et al.* Clinic- and hospital-based sentinel influenza surveillance, Uganda 2007-2010. *J Infect Dis* 2012; **206 Suppl 1**: S87-93.
- 5 WHO. WHO recommendations on the composition of influenza virus vaccines. 2021. <http://www.who.int/influenza/vaccines/virus/recommendations/en/> (accessed April 2, 2021).
- 6 Hussain M, Galvin HD, Haw TY, Nutsford AN, Husain M. Drug resistance in influenza A virus: the epidemiology and management. *Infect Drug Resist* 2017; **10**: 121–34.
- 7 ECDC. Influenza Virus Characterisation Reports, summary Europe. Eur. Cent. Dis. Prev. Control. 2010. <https://www.ecdc.europa.eu/en/seasonal-influenza/surveillance-and-disease-data/influenza-virus-characterisation> (accessed March 19, 2020).
- 8 Byarugaba DK, Ducatez MF, Erima B, *et al.* Molecular Epidemiology of Influenza A/H3N2 Viruses Circulating in Uganda. *PLOS ONE* 2011; **6**: e27803.
- 9 Byarugaba DK, Erima B, Millard M, *et al.* Whole-genome analysis of influenza A(H1N1)pdm09 viruses isolated in Uganda from 2009 to 2011. *Influenza Other Respir Viruses* 2016; **10**: 486–92.
- 10 Gachara G, Symekher S, Otieno M, Magana J, Opot B, Bulimo W. Whole genome characterization of human influenza A(H1N1)pdm09 viruses isolated from Kenya during the 2009 pandemic. *Infect Genet Evol* 2016; **40**: 98–103.
- 11 Owuor DC, Ngoi JM, Otieno JR, *et al.* Genetic characterization of influenza A(H3N2) viruses circulating in coastal Kenya, 2009-2017. *Influenza Other Respir Viruses* 2020; **n/a**. DOI:10.1111/irv.12717.
- 12 Pascalis H, Temmam S, Wilkinson DA, *et al.* Molecular Evolutionary Analysis of pH1N1 2009 Influenza Virus in Reunion Island, South West Indian Ocean Region: A Cohort Study. *PLoS ONE* 2012; **7**. DOI:10.1371/journal.pone.0043742.
- 13 C.D.C. Protocol of real time RTPCR for influenza A(H1N1). The WHO Collaborating Centre for influenza at CDC. Atlanta: USA, 2009.
- 14 Zhou B, Deng Y-M, Barnes JR, *et al.* Multiplex Reverse Transcription-PCR for Simultaneous Surveillance of Influenza A and B Viruses. *J Clin Microbiol* 2017; **55**: 3492–501.
- 15 Shepard SS, Meno S, Bahl J, Wilson MM, Barnes J, Neuhaus E. Viral deep sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler. *BMC Genomics* 2016; **17**: 708.
- 16 Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; **32**: 1792–7.
- 17 Caton AJ, Brownlee GG, Yewdell JW, Gerhard W. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell* 1982; **31**.
- 18 Wiley DC, Skehel JJ. The structure and function of the haemagglutinin membrane glycoprotein of influenza virus. *Ann Rev Biochem* 1987; **56**: 365–94.

- 19Nguyen L-T, 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**: 268–74.
- 20Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017; **8**: 28–36.
- 21Han AX, Parker E, Scholer F, Maurer-Stroh S, Russell CA. Phylogenetic Clustering by Linear Integer Programming (PhyCLIP). *Mol Biol Evol* 2019; **36**: 1580–95.
- 22Nguyen HT, Fry AM, Gubareva LV. Neuraminidase inhibitor resistance in influenza viruses and laboratory testing methods. *Antivir Ther* 2012; **73**.
- 23Simon B, Pichon M, Valette M, *et al.* Whole Genome Sequencing of A(H3N2) Influenza Viruses Reveals Variants Associated with Severity during the 2016–2017 Season. *Viruses* 2019; **11**. DOI:10.3390/v11020108.
- 24Goldstein EJ, Harvey WT, Wilkie GS, *et al.* Integrating patient and whole-genome sequencing data to provide insights into the epidemiology of seasonal influenza A(H3N2) viruses. *Microb Genomics* 2018; **4**: e000137.
- 25Escalera-Zamudio M, Cobián-Güemes AG, Taboada B, *et al.* Efficient whole genome sequencing of influenza A viruses. *bioRxiv* 2019; : 749234.
- 26Opanda S, Bulimo W, Gachara G, Ekuttan C, Amukoye E. Assessing antigenic drift and phylogeny of influenza A (H1N1) pdm09 virus in Kenya using HA1 sub-unit of the hemagglutinin gene. *PloS One* 2020; **15**: e0228029.
- 27Goka EA, Vallety PJ, Mutton KJ, Klapper PE. Mutations associated with severity of the pandemic influenza A(H1N1)pdm09 in humans: a systematic review and meta-analysis of epidemiological evidence. *Arch Virol* 2014; **159**: 3167–83.
- 28Nyang'au EM, Bulimo WD, Mobegi V, Opanda S, Magiri E. Genetic analysis of HA1 domain of influenza A/H3N2 viruses isolated in Kenya during the 2007-2013 seasons reveal significant divergence from WHO-recommended vaccine strains. *Int J Infect Dis IJID Off Publ Int Soc Infect Dis* 2020; **95**: 413–20.
- 29McAuley JL, Gilbertson BP, Trifkovic S, Brown LE, McKimm-Breschkin JL. Influenza Virus Neuraminidase Structure and Functions. *Front Microbiol* 2019; **10**. DOI:10.3389/fmicb.2019.00039.
- 30Volz EM, Koelle K, Bedford T. Viral Phylodynamics. *PLOS Comput Biol* 2013; **9**: e1002947.

Figure 1: Sociodemographic characteristics of patients whose swabs were sampled for whole-genome sequencing (n=230, isolated in 2010-2018) compared to the patients sampled in the general UVRI-NIC surveillance programme in 2007-2018 (N=1800). Cases include influenza-like illnesses (ILI) and severe acute respiratory illnesses (SARI). Panel **A** shows the number of cases, gender, and different age groups (years) infected with A(H1N1)pdm09 viruses. Panel **B** shows the number of cases, gender, and different age groups (years) infected with A(H3N2) viruses. The reported p-values (chi-square and Wilcoxon) were obtained using combined cases, gender, and ages (with no age groups) per subtype. Our sampling depicts similar characteristic trends as in the general UVRI-NIC surveillance EpiInfo data. Two of the 1802 influenza A positives patients in the general UVRI-NIC surveillance dataset were co-infections of AH1/AH3, therefore only the remaining 1800 patients are included in these comparisons. Panels **C** and **D** show, respectively, the number of patients infected with A(H1N1)pdm09 and A(H3N2) viruses whose swabs were sampled and sequenced from each geographical region per year. As of 9th May 2018, the laboratory had no positive case of A(H3N2) virus.

Figure 2: Metrics for clean MiSeq reads used for sequence assembly using IRMA Flu module. Panel **A** shows the number of reads in each read category obtained per viral swab sample. **totalReads** is the number of reads used as input for IRMA. **PassedQC** are the reads that passed IRMA internal quality control (QC). **MatchedFlu** are reads that matched IRMA's flu references. **Assembled_Paired** are the number of paired reads used for the final assemblies, and **Assembled_Unpaired** are the number of unpaired reads used for the final assemblies. Panel **B** shows the number of paired reads used in the final assembly for each of the eight gene segments: polymerase subunits (PB2, PB1 and PA), hemagglutinin (HA), nucleoprotein (NP), neuraminidase (NA), matrix protein (MP) and non-structural protein (NS). Read metrics are for the 215 A(H1N1)pdm09 and A(H3N2) viruses retained after excluding segments with low transverse coverage depth (<100).

Figure 3: Workflow showing how viral samples were excluded before and after sequencing and the rate of whole genome recovery. Eight viruses [2 for subtype A(H1N1)pdm09 and 6 for subtype A(H3N2)] failed quality control (QC) before the sequencing.

Figure 4: Phylogenies showing the temporal divergence of the HA, NA and MP genes of Uganda A(H1N1)pdm09 (A) and A(H3N2) (B) influenza viruses collected from 2010 to 2018. Trees were rooted using the oldest sequence in the Ugandan dataset. Shaded clusters are the two and three major co-circulating lineages observed since 2013 and 2012 for A(H1N1)pdm09 and A(H3N2) viruses, respectively. The third A(H3N2) lineage (with a 2011, 2016, 2017 viruses) disappeared in the MP gene tree.

Figure 5: Genetic clades of influenza A viruses that previously circulated in 2010-2018 in Uganda. Panel **A** shows clusters and genetic clades for 2010 -2018 A(H1N1)pdm09 viruses identified by statistical (PhyCLIP) and reference-based (ECDC) methods, respectively. Genetic clade 6 contains five PhyCLIP clusters (119, 92, 98, 99, and 39) with distinctive detection of 6A, 6B, and 6C. Clades 3, 5 and 7 are comparable to PhyCLIP clusters 15, 25, and 39, respectively. All clade 3, 5, and 7 viruses were collected from Entebbe and Kampala (Central Uganda) and circulated in 2010-2011. Panel **B** shows clusters and genetic clades for 2010-2017 A(H3N2) viruses identified by statistical (PhyCLIP) and reference-based (ECDC) methods, respectively. Genetic clade 3 contains four PhyCLIP clusters (13, 35, 97, and 116). Genetic clade 7 is comparable to PhyCLIP cluster 0. Clade 3 persisted in all the nine years while clade 7 re-emerged in 2017. The tip nodes and tip labels are coloured based on the cluster assigned by PhyCLIP, while labelled genetic clades (indicated by black bars) are those identified based on the amino acid mutations in the HA1 protein (ECDC method). Characteristic substitutions for each clade are indicated on the tree trunk in **bold**.

Table 1: Antigenic drift among Uganda influenza A viruses

Season	Vaccine: Passage, Genetic clade (Hemisphere)	No. of Uganda viruses analysed (year sampled)	Amino acid substitutions	Season	Vaccine: Passage, Genetic clade (Hemisphere)	No. of Uganda viruses analysed (year sampled)	Amino acid substitutions
A	A(H1N1)pdm09 viruses						
2010 - 2016	A/California/7/2009: C3, Clade 1 (SNH)	77 (2010 - 2016)	A73T ^e (1), N125D ^a (1), H138R ^d (24), H138Q ^d (1), K163I ^a (5), K163Q ^a (12), S185T ^b (54), A186T ^b (1), S190V ^b (23), S203T ^c (77), R205K ^c (9), P271S ^c (1), D222E ^d (4)	2020	A/Brisbane/02/2018: E3/E2, Clade 6B.1A1 (SH)	11 (2018)	none
2017-2018	A/Michigan/45/2015: M1, Clade 6B.1 (SNH)	30 (2017- 2018)	S74R ^e (27), K142R ^d (1), S164T ^a (27)		A/Guangdong-Maonan/SWL1536/2019: C1+C2, clade 6B.1A5A (NH)		I185T ^b (11), A187D ^b (11), E189Q ^b (11),
2019	A/Michigan/45/2015: M1, Clade 6B.1 (SH)	11 (2018)	S74R ^e (11), S164T ^a (11)				
	A/Brisbane/02/2018: E3/E2, Clade 6B.1A1(NH)		none				
B	A(H3N2) viruses						
2010-2011	A/Perth/16/2009: MDCKX, A/Perth/16 clade (SNH)	25 (2010 - 2011)	S45N ^c (7), T48I ^c (1), D53N ^c (1), K62E ^e (25), K83E ^e (1), T128A ^a (1), A138S ^a (1), I140K ^a (1), R142G ^a (1), K144N ^a (25), N145S ^a (16), F159S ^b (1), I192V ^b (3), A196T ^b (1), A198S ^b (15), A198P ^b (1), T212A ^d (25), S214I ^d (25), N225D ^d (1), I242V ^d (1), G275S ^c (2), N278K ^c (1), N312S ^c (15)	2016- 2017	A/Hong Kong/4801/2014: E5/E3, clade 3C.2a (SNH)	27 (2016- 2017)	I48T ^c (2), D53N ^c (7), D53E ^c (2), E62K ^e (1), E62G ^e (1), Q75H ^e (2), G78D ^e (2), K83E ^e (1), S91N ^e (1), K92R ^e (5), Y94H ^e (2), S96N ^d (26), N121K ^d (9), N122D ^d (3), S124N ^a (1), T128A ^a (11), T135K ^a (2), A138S ^a (6), I140K ^a (9), R142G ^a (12), R142K ^a (5), S144N ^a (13), S144K ^a (2), S145N ^a (2), S146G ^a (1), H156Q ^b (1), L157S ^b (1), Y159F ^b (7), Y159S ^b (6), K160T ^b (14), N171K ^d (7), P194L ^b (27), A196T ^b (2), Q197R ^b (5), S198P ^b (7), S198A ^b (2), X203T ^{d#} (27), K207Q ^d (1), I214T ^d (3), N225D ^d (7), S262N ^e (1), K278N ^c (2), H311Q ^c (18), S312N ^c (8)
2012	A/Perth/16/2009: MDCKX, A/Perth/16 clade (SH)	16 (2012)	S45N ^c (13), T48I ^c (5), K62E ^e (16), N121K ^d (1), S124N ^a (1), T128A ^a (5), I140K ^a (1), R142G ^a (5), K144N ^a (16), N145S ^a (8), A198S ^b (8), T212A ^d (16), S214I ^d (13), S214V ^d (3), I242V ^d (1), S262N ^e (3), N278K ^c (5), N312S ^c (8)	2018	A/Singapore/INFIMH-16-0019/2016: C1S3/S4, Clade 3C.2a1 (SNH)	13 (2017)	T48I ^c (2), D53N ^c (2), D53E ^c (1), Q75H ^e (1), G78D ^e (2), K83E ^e (1), K92R ^e (3), Y94H ^e (2), K121N ^d (8), N122D ^d (2), T128A ^a (5), T135K ^a (2), A138S ^a (5), I140K ^a (5), G142K ^a (1), G142R ^a (7), S144N ^a (7), S145N ^a (2), S146G ^a (1), H156Q ^b (1), Y159F ^b (2), Y159S ^b (5), K160T ^b (7), K171N ^d (8), A196T ^b (2), Q197R ^b (1), S198P ^b (5), S198A ^b (2), I214T ^d (2), D225N ^d (2), G275S ^c (1), K276R ^c (1), K278N ^c (2), H311Q ^c (10), S312N ^c (7)

	A/Victoria/361/2011: E3/E4, Clade 3C.1 (NH)		N45S^c (3), I48T^c (11), Q156H^b (16), V186G^b (16), S198A^b (8), I214V^d (3), Y219S^d (16), S312N^c (8)	2019	A/Switzerland/8060/2017 : E5/E2, Clade 3C.2a2 (SH)	13 (2017)	I48T^c (2), D53N^c (2), D53E^c (1), Q75H^e (1), G78D^e (2), K83E^e (1), K92R^e (3), Y94H^e (2), S96N^d (13), N121K^d (5), N122D^d (2), T128A^a (5), K131T^a (13), T135K^a (2), A138S^a (5), I140K^a (5), K142G^a (5), K142R^a (7), S144N^a (7), S145N^a (2), S146G^a (1), H156Q^b (1), Y159F^b (2), Y159S^b (5), K160T^b (6), N171K^d (5), P194L^b (13), A196T^b (2), Q197R^b (1), S198P^b (5), S198A^b (2), I214T^d (2), D225N^d (2), Q261R^e (13), G275S^c (1), K276R^c (1), K278N^c (2), H311Q^c (10), S312N^c (7)
2013	A/Victoria/361/2011: E3/E4, Clade 3C.1 (SH)	6 (2013)	D53N^c (1), S91G^e (1), N122S^d (1), S124N^a (1), T128A^a (5), I140K^a (1), R142G^a (6), N145S^a (6), Q156H^b (6), L157S^b (1), V186G^b (6), Q197H^b (1), I214T^d (1), Y219S^d (6), N278K^c (6), V297I^c (1)		A/Kansas/14/2017: Original, Clade 3C.3a (NH)		N91S^e (13), A128T^a (8), S138A^a (8), G142K^a (1), G142R^a (7), K144N^a (7), K144S^a (6), S159F^b (2), S159Y^b (6), S193F^b (13), Q311H^c (3)
	A/Texas/50/2012: E5/E1, Clade 3C.1 (NH)		N128A^a (6), P198S^b (6), F219S^d (6)	2020	A/South Australia/34/2019: E5, Clade 3C.2a1b (SH)	13 (2017)	I48T^c (2), D53E^c (1), D53N^c (2), G62E^e (13), Q75H^e (1), G78D^e (2), K83E^e (1), R92K^e (10), Y94H^e (2), K121N^d (8), N122D^d (2), T128A^a (5), K131T^a (13), T135K^a (2), A138S^a (5), I140K^a (5), G142R^a (7), G142K^a (1), S144N^a (7), S145N^a (2), S146G^a (1), H156Q^b (1), Y159F^b (2), Y159S^b (5), K160T^b (6), K171N^d (8), I186G^b (13), A196T^b (2), Q197R^b (1), S198A^b (2), S198P^b (5), I214T^d (2), F219S^d (13), D225N^d (2), G275S^c (1), K276R^c (1), K278N^c (2), Q311H^c (3), S312N^c (7)
2014	A/Texas/50/2012: E5/E1, Clade 3C.1 (SNH)	11 (2014)	E62K^e (1), N121D^d (1), N128A^a (11), A138S^a (1), I140K^a (6), R142G^a (11), N145S^a (11), F159S^b (1), V186G^b (11), P198S^b (10), F219S^d (11)		A/Hong Kong/2671/2019: E7, Clade 3C.2a1b (NH)		A128T^a (8), K135T^a (11), F137S^a (13), S138A^a (8), I160K^b (7), I160T^b (6), V186G^b (13), S193F^b (13), N225D^d (11), S312N^c (7)
2015	A/Switzerland/9715293/2013: E4/E2, clade 3C.3a (SNH)	14 (2015)	D53N^c (3), V88I^e (1), S91N^e (1), S124R^a (1), S124N^a (1), A128T^a (6), S138A^a (13), R140K^a (6), R140I^a (8), G142K^a (2), G142R^a (4), N144K^a (1), N144S^a (6), L157S^b (1), S159F^b (7), S159Y^b (6), K160T^b (6), M168V^a (2), N171H^d (1), V186G^b (14), Q197R^b (2), I214T^d (1), X219S^d (14), D225N^d (7), S262N^e (1), S265G^e (1), Q311H^c (6)		A/Hong Kong/45/2019: Original, Clade 3C.2a1b (NH)		A128T^a (8), K135T^a (11), F137S^a (13), S138A^a (8), T160K^b (7), S193F^b (13), D225N^d (2),

SH= Southern hemisphere vaccine. NH= Northern hemisphere vaccine. SNH= vaccine shared for both Southern and Northern hemispheres. A(H1N1)pdm09 HA1 antigenic sites are Sa= a, Sb= b, Ca1= c, Ca2= d, Cb= e as described by Caton et al., 1982. A(H3N2) HA1 antigenic sites are A=a, B=b, C=c, D=d, and E=e as described by Wiley and Skehel, 1987. The frequency of an amino acid substitution observed at a given antigenic site in Uganda HA1 proteins is reported as (n). The substitutions in **BOLD** are shared against the SH and NH vaccines of that given year. # = amino acid insertion. Mutations against the NH vaccine strains are colored yellow for visualization purposes only. Recommended vaccines for each influenza season are adopted from WHO (<https://www.who.int/influenza/vaccines/virus/recommendations/en/>)

554 **Table 1: Mutations in the HA1 antigenic, receptor binding, and glycosylation sites of Uganda A(H1N1)pdm09 and A(H3N2) viruses compared to the**
555 **Southern and Northern vaccine viruses.**

QC= quality control
WG= whole genome/s

A**H1** **N1** **MP**

Year

- 2010
- 2011
- 2013
- 2014
- 2015
- 2016
- 2017
- 2018

B**H3** **N2** **MP**

Year

- 2010
- 2011
- 2012
- 2013
- 2014
- 2015
- 2016
- 2017

A

HIGHLIGHT

Vaccine strain

Clade reference

B