Whole genome characterization of human influenza A(H1N1)pdm09 viruses isolated from Kenya during the 2009 pandemic

ARTICLE *in* INFECTION, GENETICS AND EVOLUTION: JOURNAL OF MOLECULAR EPIDEMIOLOGY AND EVOLUTIONARY GENETICS IN INFECTIOUS DISEASES · FEBRUARY 2016

Impact Factor: 3.02 · DOI: 10.1016/j.meegid.2016.02.029

6 AUTHORS, INCLUDING:

George Gachara

University of Venda

9 PUBLICATIONS **25** CITATIONS

SEE PROFILE

Japheth Mukuru Magana

Kenya Medical Research Institute

7 PUBLICATIONS **55** CITATIONS

SEE PROFILE

Wallace D Bulimo

US Army Medical Research Unit Kenya

50 PUBLICATIONS **211** CITATIONS

SEE PROFILE

Infection, Genetics and Evolution 40 (2016) 98–103

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Whole genome characterization of human influenza A(H1N1)pdm09 viruses isolated from Kenya during the 2009 pandemic

George Gachara [a,*], Samuel Symekher [b], Michael Otieno [a], Japheth Magana [b], Benjamin Opot [c], Wallace Bulimo [c]

[a] Department of Medical Laboratory Sciences, Kenyatta University, Kenya
[b] Centre for Virus Research, KEMRI, Kenya
[c] United States Army Medical Research Unit—Kenya, Kenya

ARTICLE INFO

Article history:
Received 9 December 2015
Received in revised form 16 February 2016
Accepted 23 February 2016
Available online 26 February 2016

Keywords:
Influenza A(H1N1)pdm09
Whole genome
Sequencing
Kenya

ABSTRACT

An influenza pandemic caused by a novel influenza virus A(H1N1)pdm09 spread worldwide in 2009 and is estimated to have caused between 151,700 and 575,400 deaths globally. While whole genome data on new virus enables a deeper insight in the pathogenesis, epidemiology, and drug sensitivities of the circulating viruses, there are relatively limited complete genetic sequences available for this virus from African countries. We describe herein the full genome analysis of influenza A(H1N1)pdm09 viruses isolated in Kenya between June 2009 and August 2010. A total of 40 influenza A(H1N1)pdm09 viruses isolated during the pandemic were selected. The segments from each isolate were amplified and directly sequenced. The resulting sequences of individual gene segments were concatenated and used for subsequent analysis. These were used to infer phylogenetic relationships and also to reconstruct the time of most recent ancestor, time of introduction into the country, rates of substitution and to estimate a time-resolved phylogeny. The Kenyan complete genome sequences clustered with globally distributed clade 2 and clade 7 sequences but local clade 2 viruses did not circulate beyond the introductory foci while clade 7 viruses disseminated country wide. The time of the most recent common ancestor was estimated between April and June 2009, and distinct clusters circulated during the pandemic. The complete genome had an estimated rate of nucleotide substitution of 4.9×10^{-3} substitutions/site/year and greater diversity in surface expressed proteins was observed. We show that two clades of influenza A(H1N1)pdm09 virus were introduced into Kenya from the UK and the pandemic was sustained as a result of importations. Several closely related but distinct clusters co-circulated locally during the peak pandemic phase but only one cluster dominated in the late phase of the pandemic suggesting that it possessed greater adaptability.

1. Introduction

Influenza A virus (Family *Orthomyxoviridae*, Genus *Influenzavirus A*) is the most diverse and epidemiologically significant of all influenza viruses (Gatherer, 2009). The virus is a negative sense, single-stranded RNA virus that infects a large variety of hosts, including aquatic birds, humans and swine. The 13.5-kb influenza A virus genome consists of eight segments, ranging from 890 to 2341 nucleotides (nt) in length, encoding up to 13 proteins (Vasin et al., 2014). A remarkable feature of influenza viruses is their inclination to undergo antigenic variation through antigenic drift and antigenic shift (Shoham, 2011). Antigenic drift consists of relatively minor mutational alterations in the antigenicity of the viral surface glycoproteins and occurs continuously as a result of selection pressure from host immunity (Shoham, 2011). Antigenic shift on the other hand, occurs by genetic reassortment of the eight gene segments. This can result in the appearance of a novel virus against which the human population has little or no immunity and which if transmitted efficiently from human to human may lead to a pandemic (Taubenberger and Morens, 2010).

In early April 2009, a new influenza A/H1N1 virus emerged among humans (CDC, 2009b). The virus was found to be antigenically and genetically unrelated to human seasonal viruses and genetically related to viruses known to circulate in swine. Phylogenetic analysis of the virus showed that it contained a constellation of gene segments from human, avian and swine origins. Due to its sustained human-to-human transmission, the WHO declared this virus as the first worldwide pandemic of the 21st Century (Smith et al., 2009). In its first year of circulation, the virus (named influenza A(H1N1)pdm09) is estimated to have caused between 151,700 and 575,400 deaths globally (Viboud and Simonsen, 2012). Kenya reported her first case of infection with this virus on 29th June 2009, in a British medical student visiting the country on an education service program (CDC, 2009a). After this introduction, the virus spread throughout the country with infections peaking in September–October 2009 and later became the dominant circulating influenza virus (Majanja et al., 2013). However, no laboratory confirmed deaths were documented in the country.

* Corresponding author.

http://dx.doi.org/10.1016/j.meegid.2016.02.029
1567-1348/

The bulk of public data on influenza virus genome comprises short fragments from the hemagglutinin (HA) or neuraminidase (NA) segments of the genome, which encode the two main surface proteins and which, it is widely believed, are the source of most of the antigenic variation in the virus (Ghedin et al., 2005). However, knowledge about the complete genome constellation of pandemic influenza A viruses is valuable for monitoring and understanding their evolutionary dynamics. Currently, very few complete genome sequences of influenza A(H1N1)pdm09 viruses from Africa are publicly available and no studies of these are available. The present study aimed to study the complete genome of influenza A(H1N1)pdm09 viruses isolated in Kenya during the pandemic period.

2. Materials and methods

2.1. Ethical considerations

The samples in this study were collected as part of the global influenza pandemic response, routine diagnostic treatment by hospital physicians and were also provided by sentinel hospitals and medical facilities around the country. Ethical clearance to carry out this particular protocol (Protocol #1753) was sought and approval granted by both the Scientific Steering Committee and Ethical Review Committee of the Kenya Medical Research Institute (KEMRI).

2.2. Sample collection

Between July 2009 and August 2010, 2519 samples from influenza-like-illness (ILI) patients who met the WHO case definition for the pandemic H1N1 infection were collected. Duplicate nasopharyngeal samples were collected using a flocked Dacron swab and placed in cryovials containing 1 ml of in house virus transport medium and stored in a liquid nitrogen dry shipper within 8 h of collection. All samples were transported weekly from the surveillance sites to the National Influenza Center (NIC) laboratory located within the Kenya Medical Research Institute (KEMRI) in Nairobi, Kenya, maintaining the cold chain throughout.

2.3. RNA extraction

RNA extraction from clinical samples and isolates was performed using the QIAamp® viral RNA mini kit (Qiagen, Germany) according to the manufacturer's instructions.

2.4. Virus detection and isolation

One-step multiplex real-time RT PCR assays were performed to detect and subtype influenza viruses based on Taqman probes as previously described (WHO, 2009). All real-time PCR-positive samples were inoculated onto MDCK cells and cultured with serum-free Dulbecco's Modified Eagles Medium (Sigma- Aldrich Co Ltd., UK) pretreated with 2.0 μg/ml of TPCK-trypsin. They were then incubated at 33 °C in an atmosphere of 5% CO_2 and observed daily for 10 days for visual cytopathic effect by using an inverted microscope. Positive isolates were identified by hemagglutination inhibition test (HI) using 1% guinea pig red blood cells and post-infection rabbit antisera obtained from the WHO influenza collaboration centre, Melbourne, Australia.

2.5. Gene amplification and sequencing

RT-PCR was performed using the Superscript III One-Step RT-PCR system (Invitrogen Corporation, CA, USA) to amplify each of the eight gene segments. Complete non-structural (NS), matrix (M) and nucleoprotein (NP) genes were amplified as one fragment while NA and HA were amplified as two overlapping fragments. The larger (polymerase) genes namely PB2, PB1, and PA were amplified as four overlapping fragments. The primer combinations used in this study are found in Supplementary Table 1. Briefly, 3 μl RNA, 12.5 μl 2X reaction mix, 0.5 μl of each primer, 1 μl of Superscript III RT/Platinum Taq mix and 7.5 μl of RNase free water were mixed for a total volume of 25 μl per reaction. Reverse transcription involved incubation at 50 °C for 30 min followed by inactivation of the RT enzyme at 94 °C for 2 min. PCR amplification included an initial denaturation at 94 °C for 45 s, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s and strand extension at 68 °C for 60 s. This was followed by end-filling incubation at 68 °C for 7 min. Upon electrophoresis in a 1.5% agarose gel, the amplicons were visualized on a UV trans-illuminator after staining with ethidium bromide. Amplicons were purified by treating with shrimp alkaline phosphatase exonuclease I (ExoSapI) (US Biological, Swampscott, MA, USA) and sequenced directly on a 3500 XL Genetic Analyzer (Applied Biosystems, CA, USA). Contiguous nucleotide sequences (contigs) assembly was carried out using the contig assembly program (CAP) of DNA Baser Sequence Assembler v3 (Heracle BioSoft SRL Romania, http://www.DnaBaser).

2.6. Sequence characterization, phylogenetics, and evolutionary rate analyses

The resulting nucleotide sequences were translated and aligned using Clustal W program implemented in MEGA 5.10 (Tamura et al., 2011), manually inspected and trimmed to include the coding regions only. The relative amino acid frequency in the analysis of the genome signatures for each gene was carried out using WebLogo (Crooks et al., 2004). To test for reassortment, each gene segment was analyzed using the FluGenome web tool (Lu et al., 2007). The coding regions of all segments for each isolate were then concatenated in the following sequence; PB2, PB1, PA, HA, NP, NA, M and NS using the SequenceMatrix software (Gaurav et al., 2011). The concatenated alignments were used for all subsequent analyses. Extensive phylogenetic analysis based on concatenated whole genome sequences of representative influenza A(H1N1)pdm09 viruses previously sampled globally revealed seven distinct clades named 1–7 in circulation early in the pandemic (Nelson et al., 2009). Global influenza A(H1N1)pdm09 virus sequences representing these clades were downloaded from the NCBI influenza virus database for comparative purposes.

Preliminary phylogenetic reconstruction was carried out utilizing the Bayesian method of tree inference using the MrBayes program version 3.1.2 (Ronquist and Huelsenbeck, 2003). The Bayesian tree was inferred by running a Markov-chain Monte Carlo (MCMC) algorithm for 10 million generations, sampling at every 200th generation with a burn in setting of 10% of generations. The GTR + G + I model (general time-reversible model with gamma distributed rates of variation among sites and a proportion of invariable sites) was found to be the best-fit model for our dataset. The resulting phylogenetic tree was visualized using FigTree, version 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

The temporal phylogeny and rates of substitution of the pandemic H1N1/09 virus sequences were estimated by using a Bayesian Markov chain Monte Carlo (MCMC) approach as implemented in the BEAST package v1.7.4 (Drummond and Rambaut, 2007). The data were analyzed under a coalescent exponential-growth model as a prior on the tree, the HKY + Γ model of nucleotide substitution and an uncorrelated exponential relaxed clock model. The analysis was run for 50 million states with sampling every 10,000 states to ensure adequate sample size of all analysis parameters. Convergence was assessed on the basis of the effective sampling size (ESS) after a 10% burn-in using Tracer software version 1.5 (Rambaut and Drummond, 2009). The resulting tree with the maximum product of posterior probabilities (maximum clade credibility tree) for analyzing the MCMC data set was annotated by TreeAnotator incorporated in the BEAST package and visualized using FigTree, version 1.3.1.

The sequences from this study have been deposited in the Global Initiative on Sharing All Influenza Data (GISAID) database under the

following isolate identities; EPI_ISL_139648, EPI_ISL_139667, EPI_ISL_139680, EPI_ISL_139706–139723, EPI_ISL_140395–140413.

3. Results

3.1. Viral isolation, amplification and sequencing

In total, 369 out of 2519 samples (14.6%) collected during the study period from July 2009 to August 2010 were positive for influenza A(H1N1)pdm09 using real time RT-PCR. Out of these samples, 55 were successfully cultured after the first passage. These were filtered to 40 isolates by selecting only two isolates collected from a site in the same week to control for possible founder effects.

3.2. Phylogenetic characterization

The phylogenetic tree of the concatenated genome which also included viral genomes isolated from elsewhere whose clade identities are known is shown in Fig. 1. The results showed that 2 of the 4 earliest Kenyan isolates from June/July 2009 clustered with global clade 2 viruses while the other 2 clustered with clade 7 viruses. Throughout the pandemic, clade 7 viruses dominated local transmission and no clade 2 viruses were observed beyond the initial introductions in the country. Of the seven clades of pandemic H1N1 virus that has been shown to circulate globally, the Kenyan isolates belong to only clades 2 and 7. The other globally circulating influenza A(H1N1)pdm09 clades namely 1, 3, 4, 5 and 6 were not detected in the Kenyan population. The concatenated genomes of the Kenyan isolates were also observed to be interspersed with genomes of viruses from other countries. Assessment of gene segment lineage did not detect reassortment in the local viruses.

3.3. Overall profile of amino acid variations observed

Many amino acid signatures were found in local isolates some of which showed temporal patterns. Early in the pandemic, several defining mutations became fixed namely PA (P224S), HA, (P100S, S220T and I338V), NP (V100I), NA (V106I and N248D), and NS (I123V). Majority of the amino acid substitutions observed during the peak of the pandemic did not persist into the late phase of the pandemic. During the late phase of the pandemic, several mutations dominated including PA (N321K and I330V), HA (D114N, S202T, E391K and S468N), NA (N369K), M1 (V80I) and M2 (F91L).These are summarized in Fig. 2. At the same time all the local viruses had D239 (D222 in alternative numbering), 31N and 275H in their HA, M2 and NA proteins respectively.

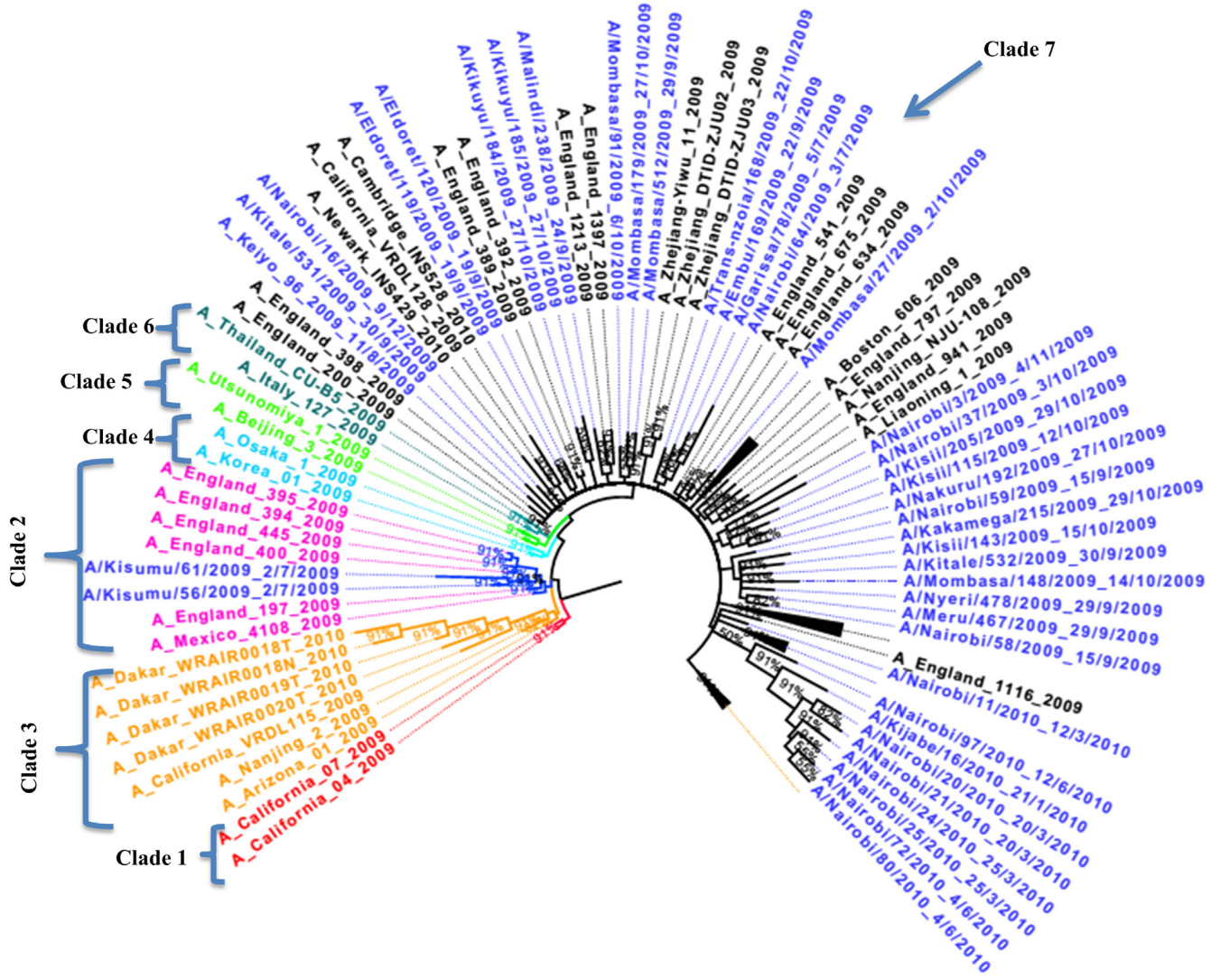

Fig. 1. Bayesian phylogenetic tree of the concatenated influenza A(H1N1)pdm09 genome of Kenyan and selected global isolates. The 40 Kenyan isolates characterized in this study are highlighted in blue at the taxa and are also indicated with their date of collection. The global influenza A(H1N1)pdm09 virus clades are indicated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

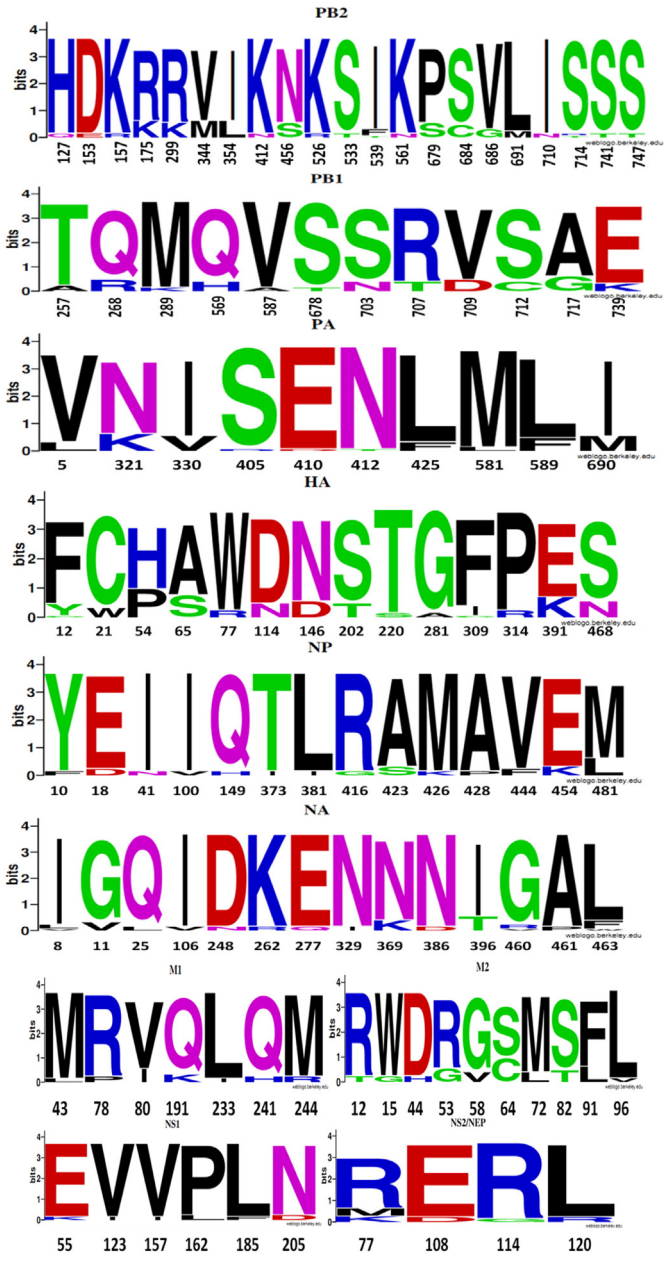

Fig. 2. Genome signatures in amino acid residues of influenza A(H1N1)pdm09 virus proteins isolated in Kenya during the 2009 pandemic. The identity of the protein is indicated on each weblogo. The relative frequency of the corresponding amino acid at a given position is proportional to the residue height. Residue positions are indicated on the x-axis and are based on the nucleotide positions of each gene. The sequence logos were generated using WebLogo 2.8.2.

3.4. Evolutionary analysis

We sought to identify the Kenyan-specific circulating clusters and the divergence times of the 40 sampled viruses which in our opinion represent the entire pandemic period. The resulting temporal phylogeny of the complete genome of Kenyan influenza A(H1N1)pdm09 viruses is shown in Fig. 3. This phylogeny demonstrates the existence of multiple clusters during the early and peak pandemic phase. The time of the most recent common ancestor (tMRCA) of each cluster, corresponding to the date at which they emerged was also estimated. The study identified six such Kenyan circulating clusters each containing 2 to 8 Kenyan isolates. Clusters were considered if they had >90% posterior probability support. These were then named to reflect their placement within global clades 1 to 7, defined previously (Nelson et al., 2009); hence, KENA-GC2 indicates Kenyan cluster A which falls within global clade 2.

The viruses isolated during the early phase of the pandemic clustered in two different clusters namely KENA-GC2 and KENB-GC7. KENA-GC2 viruses were characterized by PB2 526R, PB1 587 A, PA 581 L, HA 220S, NP 373I, NA 329I and NS 123I while the KENB-GC7 cluster was characterized by HA 54P and HA 65S. KENA-GC2 which was made up of only two samples did not spread beyond the early pandemic period. On the other hand, KENB-GC7 continued to circulate into the peak pandemic phase. During the peak phase of the pandemic, four clusters namely KENB-GC7, KENC-GC7, KEND-GC7 and KENE-GC7 were co-circulating. The amino acid signatures for these clusters are, PB2 175 K (KENC-GC7), HA 54P (KEND-GC7) and PA 425F (KENE-GC7). One cluster, KENF-GC7 which was well supported (pp = 100%) dominated in the late phase of the pandemic. This cluster was characterized by PA 321K.

The mean time of the most recent common ancestor (tMRCA) of the earliest Kenyan influenza A(H1N1)pdm09 viruses was estimated to be 19th April 2009, with a 95% credibility interval of between December 2008 and July 2009. The tMRCAs and the most probable months in the calendar time scale of the main clusters are shown in Table 1. The complete genome of the Kenyan influenza A(H1N1)pdm09 viruses had an estimated rate of nucleotide substitution of 4.9×10^{-3} substitutions/site/year (95% confidence interval [CI], 2.6×10^{-3}–7.2×10^{-3}).

4. Discussion

Using global isolates whose clade identities are known, the phylogenetic analysis of the concatenated genome showed that the initial introductions of the influenza A(H1N1)09pdm virus in the country belonged to two different global clades namely 2 and 7 (Fig. 1). Besides differences in geographical distribution, these two clades differ genetically in the mutations they harbor relative to the prototype A/California/7/09 strain. Clade 2 viruses have the M581L and T373I mutations in the PA and NP proteins respectively while clade 7 viruses lack these mutations but possess the mutations S220T in their HA, V100I in their NP, V106I and N248D in their NA and I123V in their NS1 (Nelson et al., 2009). In this study, clade 2 viruses were not isolated beyond the initial introduction cases. The absence of clade 2 viruses beyond the introduction cases is worth noting and this may mean that this particular clade did not acquire adequate fitness for sustenance locally.

On the other hand, clade 7 viruses were isolated throughout the pandemic period in Kenya. This dominance of circulation of clade 7 viruses throughout the pandemic period and the disappearance of all other clades was observed in other studies in the United Kingdom, Argentina, India and in the United States of America (Baillie et al., 2012; Barrero et al., 2011; Mullick et al., 2011; Nelson et al., 2011). While clade 7 viruses have dominated globally, they have also been the most diverse and their genetic diversity may have arisen from reassortment among clades considering that they co-circulated with all the other clades. However, no reassortment was detected among the local viruses. This predominance of clade 7 viruses has been hypothesized to be due to its fitness and adaptability and more efficient human transmissibility (Potdar et al., 2010). However, this requires further investigations.

It has been observed that the main strategy of influenza viruses to outrun the immune system of the host is by maintaining a high mutation rate (Fitch et al., 1991). As a consequence, the virus has a very high evolutionary rate. This study thus estimated the evolutionary rate of the Kenyan isolates. The complete genome of the local viruses had an estimated rate of nucleotide substitution of 4.9×10^{-3} substitutions/site/year. In 2009 Smith analyzed evolutionary rates during the early phase of the pandemic and found the evolutionary rate for the complete genome to be 3.66×10^{-3} with a range of $0.61–6.58 \times 10^{-3}$ (Smith et al., 2009). Therefore, the observed rate of substitution for

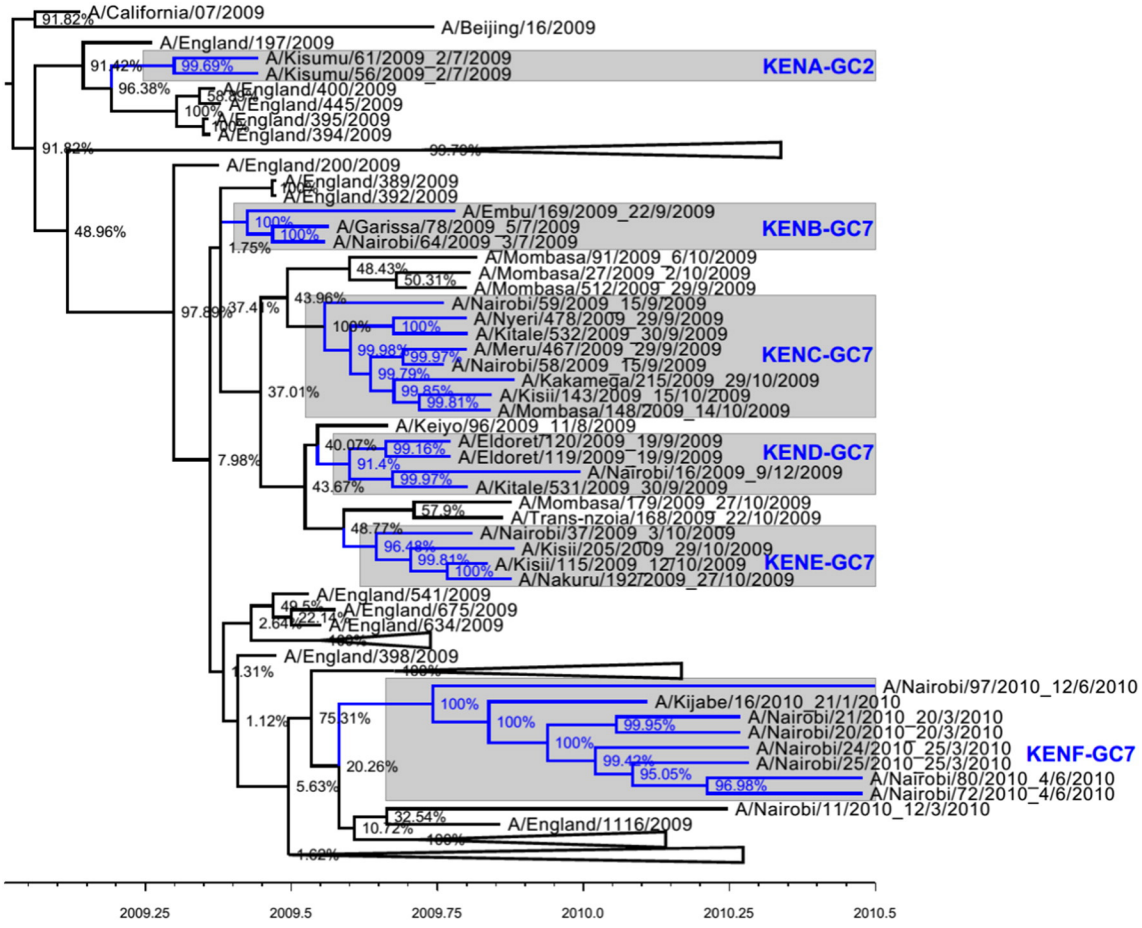

Fig. 3. Simplified time scaled maximum clade credibility tree showing distribution of the circulating Kenyan clusters of influenza A(H1N1)pdm09. Identified Kenyan clusters are highlighted and named to reflect their placement within global clades. The numbers on the internal nodes indicate posterior probabilities.

the complete genome for the Kenyan viruses is similar to that estimated during the early phase of the pandemic.

Regarding the local virus genome signatures, all isolates were able to synthesize a truncated (11 amino acids) PB1-F2 protein and seemed to have a normal rate of viral replication and transmission to mammalian host cells by possessing 627E in the PB2 protein. The HA-D239G (D222G in alternative numbering) mutation which has been implicated in severe outcomes in some studies was not detected in the local viruses. The markers associated with antiviral susceptibility indicated that local isolates were resistant to adamantanes but susceptible to neuraminidase inhibitors. This observation had been made early in the pandemic (Dawood et al., 2009). A temporal examination of the genome signatures suggests that the amino acid substitutions that were observed in the early phase of the pandemic may have been crucial for efficient replication and transmission. Overall, the surface expressed gene segments (HA and NA), the transmembrane expressed segment (M) and the NS gene segment had the highest genetic diversity compared to internally expressed segments (PB2, PB1, PA, and NP). This observation has also been shown previously among seasonal influenza viruses (Rambaut et al., 2008).

Table 1
Divergence times of Kenyan clusters.

Cluster	No. of isolates	TMRCA	TMRCA	
			Low 95% HPD*	High 95% HPD
KENA-GC2	2	19th April 2009	27th Dec 2008	6th July 2009
KENB-GC7	3	4th June 2009	16th April 2009	29th June 2009
KENC-GC7	8	8th August 2009	19th May 2009	2nd October 2009
KEND-GC7	4	7th August 2009	10th June 2009	17th Sept 2009
KENE-GC7	4	4th August 2009	10th June 2009	2nd October 2009
KENF-GC7	8	2nd October 2009	20th June 2009	10th Dec 2009
Root	40	28th Feb 2009	8th Oct 2008	19th May 2009

* Highest posterior density.

To identify possible temporal or spatial discrete lineages in influenza A(H1N1)09pdm viruses circulating in Kenya, we conducted a maximum clade credibility tree analysis. This was in an attempt to seek molecular insights into the trajectory of the influenza A(H1N1)09pdm virus following its introduction into the country. The chronological phylogeny for the Kenyan viruses revealed that six distinct viral clusters circulated throughout the pandemic period. Using contact tracing methods, it had been shown that influenza A(H1N1)09pdm virus was introduced in Kenya at three points; the first 2 points were in Western Kenya while the third was in Nairobi and that all these initial introductions were from the UK (CDC, 2009a). Using molecular data, this study shows that the two introductions in Western Kenya were due to both clades 2 and clade 7 viruses while the introduction in Nairobi was due to a clade 7 virus. In the UK, the first laboratory confirmed cases were detected in April 2009 and the first wave of infections peaked in June - July 2009 (Baillie et al., 2012).

The tMRCA estimate of the tree root dating back to February 2009 is in line with the majority of the previous estimations, which place the origin of the pandemic H1N1 strain to January 2009 with intervals of credibility between late 2008 and March 2009 (Fraser et al., 2009; Shiino et al., 2010; Smith et al., 2009). In agreement with the contact tracing findings, the tMRCA's of the earliest clusters identified in Kenya were estimated to be between April and June 2009. It is worth noting that the tMRCA of the Kenyan clade 2 viruses overlap with that of the UK clade 2 viruses (Baillie et al., 2012). At the same time, both local clade 2 and

7 viruses were closely associated with UK isolates as shown in the time scaled MCC tree and not with isolates from USA, Asia or Africa. Molecular data presented here therefore confirms previous classical epidemiological work while at the same time shedding light on the genetic characteristics of the viruses introduced into the country.

There are two possible shortcomings in this study. First, the analysis in this study only involved the coding region of the gene segments. Although the non-coding regions are considered to be conserved, mutations that affect viral replication may occur and this information may have been lost in the current study. Secondly, since this study did not allow for follow up of the patients, it is not possible to associate any of the observed mutations with virulence. However, the strength of this study lies in the fact that it is among the first of its kind in Africa that employs whole genome sequencing in an attempt to understand the transmission and evolutionary dynamics of influenza A(H1N1)09pdm virus in the region.

In conclusion, on the basis of these observations, we confirm that two clades of influenza A(H1N1)pdm09 virus were introduced into Kenya from the UK and that the pandemic was sustained as a result of multiple importations into the country. Several closely related but distinct clusters co-circulated locally during the peak pandemic phase but only one cluster dominated in the late phase of the pandemic suggesting greater adaptability and fitness in this cluster.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.meegid.2016.02.029.

Acknowledgments

We wish to acknowledge the Director of Kenya Medical Research Institute (KEMRI) and Director of Center for Virus Research (CVR) for facilitating this study. We also wish to thank CDC-Kenya for kindly donating primer sets used in the laboratory assays and the National Influenza Centre staff for their involvement in the influenza pandemic response in Kenya.

References

Baillie, G.J., Galiano, M., Agapow, P.M., Myers, R., Chiam, R., Gall, A., Palser, A.L., Watson, S.J., Hedge, J., Underwood, A., Platt, S., McLean, E., Pebody, R.G., Rambaut, A., Green, J., Daniels, R., Pybus, O.G., Kellam, P., Zambon, M., 2012. Evolutionary dynamics of local pandemic H1N1/2009 influenza virus lineages revealed by whole-genome analysis. J. Virol. 86, 11–18.

Barrero, P.R., Viegas, M., Valinotto, L.E., Mistchenko, A.S., 2011. Genetic and phylogenetic analyses of influenza A H1N1pdm virus in Buenos Aires, Argentina. J. Virol. 85, 1058–1066.

CDC, 2009a. Introduction and transmission of 2009 pandemic influenza A (H1N1) virus—Kenya, June–July 2009. MMWR Morb. Mortal. Wkly Rep. 58, 1143–1146.

CDC, 2009b. Swine influenza A (H1N1) infection in two children-Southern California, March–April 2009. MMWR Morb. Mortal. Wkly Rep. 58, 400–402.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.

Dawood, F.S., Jain, S., Finelli, L., Shaw, M.W., Lindstrom, S., Garten, R.J., Gubareva, L.V., Xu, X., Bridges, C.B., Uyeki, T.M., 2009. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N. Engl. J. Med. 360, 2605–2615.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Fitch, W.M., Leiter, J.M., Li, X.Q., Palese, P., 1991. Positive Darwinian evolution in human influenza A viruses. Proc. Natl. Acad. Sci. U. S. A. 88, 4270–4274.

Fraser, C., Donnelly, C.A., Cauchemez, S., Hanage, W.P., Van Kerkhove, M.D., Hollingsworth, T.D., Griffin, J., Baggaley, R.F., Jenkins, H.E., Lyons, E.J., Jombart, T., Hinsley, W.R., Grassly, N.C., Balloux, F., Ghani, A.C., Ferguson, N.M., Rambaut, A., Pybus, O.G., Lopez-Gatell, H., Alpuche-Aranda, C.M., Chapela, I.B., Zavala, E.P., Guevara, D.M., Checchi, F., Garcia, E., Hugonnet, S., Roth, C., 2009. Pandemic potential of a strain of influenza A (H1N1): early findings. Science 324, 1557–1561.

Gatherer, D., 2009. The 2009 H1N1 influenza outbreak in its historical context. J. Clin. Virol. 45, 174–178.

Gaurav, V., David, J.L., Rudolf, M., 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27, 171–180.

Ghedin, E., Sengamalay, N.A., Shumway, M., Zaborsky, J., Feldblyum, T., Subbu, V., Spiro, D.J., Sitz, J., Koo, H., Bolotov, P., Dernovoy, D., Tatusova, T., Bao, Y., St George, K., Taylor, J., Lipman, D.J., Fraser, C.M., Taubenberger, J.K., Salzberg, S.L., 2005. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature 437, 1162–1166.

Lu, G., Rowley, T., Garten, R., Donis, R.O., 2007. FluGenome: a web tool for genotyping influenza A virus. Nucleic Acids Res. 35, W275–W279.

Majanja, J., Njoroge, R.N., Achilla, R., Wurapa, E.K., Wadegu, M., Mukunzi, S., Mwangi, J., Njiri, J., Gachara, G., Bulimo, W., 2013. Impact of influenza A(H1N1)pdm09 virus on circulation dynamics of seasonal influenza strains in Kenya. Am.J.Trop. Med. Hyg.

Mullick, J., Cherian, S.S., Potdar, V.A., Chadha, M.S., Mishra, A.C., 2011. Evolutionary dynamics of the influenza A pandemic (H1N1) 2009 virus with emphasis on Indian isolates: evidence for adaptive evolution in the HA gene. Infect. Genet. Evol. 11, 997–1005.

Nelson, M., Spiro, D., Wentworth, D., Beck, E., Fan, J., Ghedin, E., Halpin, R., Bera, J., Hine, E., Proudfoot, K., Stockwell, T., Lin, X., Griesemer, S., Kumar, S., Bose, M., Viboud, C., Holmes, E., Henrickson, K., 2009. The early diversification of influenza A/H1N1pdm. PLoS Curr. 1, RRN1126.

Nelson, M.I., Tan, Y., Ghedin, E., Wentworth, D.E., St George, K., Edelman, L., Beck, E.T., Fan, J., Lam, T.T., Kumar, S., Spiro, D.J., Simonsen, L., Viboud, C., Holmes, E.C., Henrickson, K.J., Musser, J.M., 2011. Phylogeography of the spring and fall waves of the H1N1/09 pandemic influenza virus in the United States. J. Virol. 85, 828–834.

Potdar, V.A., Chadha, M.S., Jadhav, S.M., Mullick, J., Cherian, S.S., Mishra, A.C., 2010. Genetic characterization of the influenza A pandemic (H1N1) 2009 virus isolates from India. PLoS One 5, e9693.

Rambaut, A., Drummond, J., 2009. MCMC Trace Analysis Tool Version v1.5.0. University of Oxford, UKhttp://beast.bio.ed.ac.uk/.

Rambaut, A., Pybus, O.G., Nelson, M.I., Viboud, C., Taubenberger, J.K., Holmes, E.C., 2008. The genomic and epidemiological dynamics of human influenza A virus. Nature 453, 615–619.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Shiino, T., Okabe, N., Yasui, Y., Sunagawa, T., Ujike, M., Obuchi, M., Kishida, N., Xu, H., Takashita, E., Anraku, A., Ito, R., Doi, T., Ejima, M., Sugawara, H., Horikawa, H., Yamazaki, S., Kato, Y., Oguchi, A., Fujita, N., Odagiri, T., Tashiro, M., Watanabe, H., 2010. Molecular evolutionary analysis of the influenza A(H1N1)pdm, May–September, 2009: temporal and spatial spreading profile of the viruses in Japan. PLoS One 5, e11057.

Shoham, D., 2011. The modes of evolutionary emergence of primal and late pandemic influenza virus strains from viral reservoir in animals: an interdisciplinary analysis. Influenza Res. Treat. 2011.

Smith, G.J., Vijaykrishna, D., Bahl, J., Lycett, S.J., Worobey, M., Pybus, O.G., Ma, S.K., Cheung, C.L., Raghwani, J., Bhatt, S., Peiris, J.S., Guan, Y., Rambaut, A., 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature 459, 1122–1125.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Taubenberger, J.K., Morens, D.M., 2010. Influenza: the once and future pandemic. Public Health Rep. 125 (Suppl. 3), 16–26.

Vasin, A.V., Temkina, O.A., Egorov, V.V., Klotchenko, S.A., Plotnikova, M.A., Kiselev, O.I., 2014. Molecular mechanisms enhancing the proteome of influenza A viruses: an overview of recently discovered proteins. Virus Res. 185, 53–63.

Viboud, C., Simonsen, L., 2012. Global mortality of 2009 pandemic influenza A H1N1. Lancet Infect. dis. 12, 651–653.

WHO, 2009. CDC Protocol of Realtime RT-PCR for Influenza A (H1N1). Available at http://www.who.int/csr/resources/publications/swineflu/realtimeptpcr/en/index.html (Accessed on 05.03.2015).