EVOLUTIONARY DYNAMICS AND GEOGRAPHIC DISPERSAL OF BETA CORONAVIRUSES IN AFRICAN BATS

Babatunde O Motayo[1,2], Olukunle O Oluwasemowo[2] , Paul A Akinduti[3]

1. Department of Medical Microbioogy and Parasitology, Federal Medical Center, Abeokuta, Nigeria.
2. Department of Virology, College of Medicine, University of Ibadan, Ibadan, Nigeria.
3. Department of of Biological Science, Covenant University, Otta, Nigeria.

Address for correspondence, babatundemotayo@yahoo.com; niyiakinduti@gmail.com.

RUNNING TITLE: EVOLUTION OF CORONAVIRUSES IN AFRICAN BATS

ABSTRACT

Bats have been shown to serve as reserviour host of various viral agents including coronaviruses. They have also been associated with the novel coronavirus SARS-CoV-2. This has made them an all important agent for CoV evolution and transmission. Our objective in this study was to investigate the dispersal, phylogenomics and evolution of betacoronaviruses (β-CoV) among African bats. Three data sets were generated from sequences of the partial RNA depedent RNA polymerase (RdRP) gene of Bat coronaviruses from seven African, three Asian, and four European countries. Phylogenetic trees was infered using MEGA Vs 7, Phylogeography and Evolutionary analysis were carried out, using Markov Chain Monte Carlo (MCMC) implemented in BEAST, and visualised with SPREAD3. Results showed that majority of the African strains fell within lineage D. Evololutionary rate of the African Bt-β-CoV was 1.301×10^{-3}, HPD ($1.064 \times 10^{-3} – 1.434 \times 10^{-3}$), the MCC tree of the African strains showed two clades, lineage D, and lineage C. The TMRCA for lineage D strains was 1972, 95% HPD (1951-1990), and 2010, 95% HPD (2006-2014) for the South African strains. There were inter-continental dispersal of Bt-CoV from China and Hong Kong into Central and Southern Africa. In conclusion we advocate for broader studies of whole genome sequences of BtCoV to further understand the drivers for their emergence and zoonotic spillovers into human population.

Keystwords: Phylogeography, Evolutionary dynamics, BtCoV, Africa

INTRODUCTION

Coranaviruses are a large group of enveloped, positive sense, single stranded RNA viruses belonging to the order Nidovirales and family Coronaviridea (Masters and Pearlman, 2013). The submilycoronavirinae contains four genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus (Masters and Pearlman, 2013). They generally infect animals such as mammals and birds, they are the causative agents of many pathogenic diseases such as transmissible gastroenteritis of swine (TGEV), infectious bronchitis virus (IBD), mouse hepatitis virus (MHV), and bovine coronavirus (BCoV) (Saif, 2004). Coronaviruses have been reported as early as 1930 (Masters and Pearlman, 2013) but the earliest report of human coronaviruses was in the year 1960, where two strains namely HuCoV 229-E and HuCoV OC43 were described (Hamre and Procknow, 1966; McIntoch et al, 1967). Generally betacoranviruses have been observed to cause very severe disease in man and are largely known to be zoonotic. It was not until after the advent of the severe acute respiratory syndrome (SARS) outbreak in Hong Kong and parts of China in 2003, that studies into the zoonotic origin of the incriminated pathogen SARS CoV revealed that the Chinese rinolophid bats maintained a genetically related strain of the SARS CoV, (Lau et al, 2005; Li et al, 2005). This finding sparked up interest in bat CoV research. Another coronavirus termed middle east severe respiratory syndrome virus (MERS CoV) was reported in the Arabian peninsular in 2012 (Zaki et al, 2012). Genetically similar strain to the MERS CoV were also isolated from Pipistellus, Tyloncteris and Neoromica bats (Anan et al, 2010).

In Africa large scale surveillance studies have identified diverse strains of coranviruses circulating among African bats from Kenya, Ghana and Nigeria (Drexler et al, 2014). Studies have alsoprovided evidence that the human coronavirus HGoV-229E originated from African bat coronaviruses (Cornman et al, 2015). Also the rich fauna and biodiversity in Africa has made it a hotspot for emerging viral diseases. It is also inhabited by a diversity of bats which have been identified to serve as reservoir of high consequence zoonotic diseases such as Marburg hemorrhagic fever and Rabies (Anan et al, 2013). Recently a novel coronavirus SARS CoC-2 was identified to be the cause of a gobal pandemic which originally broke out in Wuhan, Hubei province, China (Zhou et al, 2020). Some studies have also postulated that the SARS CoV-2 probably spilled over into human population through a zoonotic event involving Chinese SARS-related BtCoV (Zhou et al, 2020; Li et al, 2020).

The recent evidence of African bats as a potential reservoir host for several betacoronaviruses gave rise to the conceptualization of this study which aims to investigate the spatial dispersal,phylogenomics and evolution of betacoronaviruses among African bats.

TABLE 1: DISTRIBUTION OF β-CoV INFECTED BAT SPECIES ACORDING TO COUNTRY AND β-CoV LINEAGE ASSIGNMENT

Species name	**Country(Number)**	**Lineage**
Elidion hevium	Nigeria(6), Kenya(6)	D
Neormicia	South Africa(27)	C
Scotophilus leucogaster	Cameroun(2)	D
Epomophorus gambianus	Cameroun(4)	D
Rousettus aegypticaus	Rwanda(3)	D
Epomophorus labiatus	Rwanda(1)	D
Micropteropus pusillus	Cameroun(9), DRC(12)	D
Epomops franquenti	DRC(3), Cameroun(1)	D

MATERIALS AND METHODS

Data Collection

We searched for partial or complete gene codding regions for the RNA depedent RNA polymerase sequences (RdRP) of African Bat coronaviruses from the Virus Pathogen resource database http://www.viprbic.org . Three data sets were generated for this study, the first contained only African batbetacoronaviruses coronaviruses from seven countries, namely Nigeria, Kenya, Cameroun, Democrtic Republic of Congo (DRC), Rwanda, Madagascar and South Africa (n=91). The second dataset contained African bat coronaviruses and Bat coronaviruses from China, Hong Kong and Philippines in Asia, and France, Spain, Italy and Luxumburg in Europe (n=203). For the second dataset, countries which had sequences fewer than five were excluded. Finally, a third data set included all RdRP sequences selected in the second dataset and representative African human CoV OC43 and HKU1, MERS CoV from Africa, and alpha coronaviruses from Africa (n=216). Information such as country of origin, Host species, and date of collection were combined with the sequence information for the purpose of accurate phylogenetic determination. The final data stets had information from seven African countries, four European countries and three Asian countries. All the data used in this study can be assessed in Supplementary Table 1.

Phylogenetic analysis

Cluster analysis was carried out with the online program CD-HIT version 4.6 for identifying sequence similarity to reduce duplication of data. Sequences were aligned using clustal W version 2.1 using default settings. Phylogenetic trees were constructed in MEGA 7.0 software www.megasoftwre.net using the maximum likelihood method with a general time reversible GTR with a gama distributed rate variation (T_4) and a p- distance model with 1000 bootstrap resampling. The final trees were then visualized in fig tree http://tree.bio.ed.ac.uk/software/figtree/.

Discrete Phylogeographic analysis

Aligned sequences were analyzed for evidence of sufficient temporal clock signal using TempEst version 1.5 (Rambault et al, 2016). The relationship between root-to-tip divergence and sampling dates supported the use of molecular clock analysis in this study. Phylogenetic trees were generated by Bayasian inference through Markov chain Monte Carlo (MCMC), implemented in BEAST version 1.10.4 (Suchard et al, 2018). We partitioned the coding genes into first+second and third codon positions and applied a separate Hasegawa-Kishino-Yano (HKY+G) substitution model with gamma-distributed rate heterogeneity among sites to each partition (Hasegawa et al, 1985). Two clock models were initially evaluated strict and relaxed molecular clock, with four differenttree priors, constant population size, exponential population size, Bayesian Skyride plot and Gausian Markov Random Field Skyride plot. Each selected model was run for an initial 30, 000, 000 states. Models were compared using Bayes factor with marginal likelihood estimated using the path sampling and stepping stone methods implemented in BEAST version 1.10.4 (Scuchard et al, 2018). The relaxed clock with Gausian Markov Random Field Skyride plot (GMRF) coalescent prior was selected for the final analysis. The MCMC chain was set at 100, 000, 000 states with10% as burn in. Results were visualized using Tracer version 1.8. (http://tree.bio.ed.ac.uk/software/tracer/), all effective sampling size ESS values were >200 indicating sufficient sampling. Bayesian skyride analysis was carried out to visualize the epidemic evolutionary history using Tracer v 1.8. (http://tree.bio.ed.ac.uk/software/tracer/).To reconstruct the ancestral-state phylogeographic transmission across countries and hosts, we used the discrete-trait model implemented in BEAST version 1.10.4 (Suchard et al, 2018).The Bayesian stochastic search variable selection (BSSVS) approach (Lemmy et al, 2009) was used to explore the most important historical dispersal routes for the spread BtCoV across their countries of origin, as well as the most probable host-species transition.The spatiotemporal viral diffusion was then visualized using the Spatial Phylogenetic Reconstruction of Evolutionary Dynamics SPREAD3 software (Bielejec et al, 2016).

RESULTS AND DISCUSION

We analyzed betacoronavirus sequences from seven African countries distributed among eight bat species as shown in Table 1. The most abundant bat species sampled in this study was M. pussilus, and Cameroon had the highest distribution of bat species sampled in this study. This result does not necessarily represent the true picture of bat species diversity in Africa, as some countries lack sequence information for bats due lack of surveillance. Few studies have been carried in Africa on coronaviruses among bats leaving a huge gap in epidemiologic information regarding beta BtCoV(β-BtCoV) in Africa.

Phylogenetic analysis of β-BtCoV sequences revealed that majority of the African strains fell within lineage D consisting of strains from Cameroon, DRC, Kenya, Madagascar and Nigeria. This observation is in agreement with a previous report which identified the existence of a single BtCoV clade circulating around Africa (Leopardi et al, 2016). However, it was identified that isolates consisting largely of strains isolated among Neomorican South African bats clustering within lineage C together with strains from Italy and Spain (Figure 1). The species-specific phylogenetic clustering observed among the Neomorican bats suggests limited cross-species β-CoV transmission among these species of bats in Africa, although larger epidemiological studies among these species of bats will shed more light as to the cause of this observed trend in Africa. Studies have shown that β coronaviruses of lineage C such as MERS CoV are capable of both intra-species transmission and inter host transmission (Min et al, 2016). In this study there seems to be inter-species transmission among the lineage D betacoronaviruses, allowing for potential recombination and rapid evolution of this lineage, as previously reported (Tao et al, 2017). This type of observation was also reported in earlier study of an inter-species transmission events of alpha CoV HKU10 in bat species of different orders (Lau et al, 2012). Resultsfor eoot to tip divergence showed the data set had a positive temporal signal (Figure 2) with the correlation coefficient = 0.0286 and R^2 = 0.0818.

The MCC tree of the African strains shows clearly the two major clades the lineage D, and the South African linieage C (Figure 3).The TMRCA for lineage D strains dating back to 1972, 95% HPD(1951-1990), and the TMRCA for lineage C strains 2010 95% HPD (2006-2014). This shows that lineage C is relatively recent and probably evolved from the existing lineage D strains. Evololutionary rate of the African bat β-CoV was set at 1.301×10^{-3}, HPD (1.064×10^{-3} – 1.434×10^{-3}), this is slightly higher that the evolutionary rate for the ongoing SARS CoV-2 which has been estimated to have an evolutionary rate of $8.0 \times 10\text{-}^{4}$ (www.nextstrain.org/ncov/global). This is also slightly higher than the evolutionary rate reported for the partial RdRP gene of HuCoV OC43 of 1.06×10^{-4} (Vijen et al, 2006). A similar topology was also observed for the MCC tree of the second data set which included β-BtCoV from Asia and Europe, with a MRCA of 1892 (HPD 1850-1926) for all the strains, with the African strains showing the consistent TMRCA as described above (Figure 4). The African lineage D strains seemed to emerge from their parental strain at around the year 1960 (HPD, 1944-1976) this observation supports the hypothesis that lineage D viruses could have been circulating in Africa long before they were first isolated. Whereas for the South African lineage C strains they seemed to emerge from their parental lineage around the year 1983 (Figure 4), indicating a more recent introduction into Africa, however studies have dated their origin based on partial RdrP gene to as back as 1859 (Lau et al, 2013).

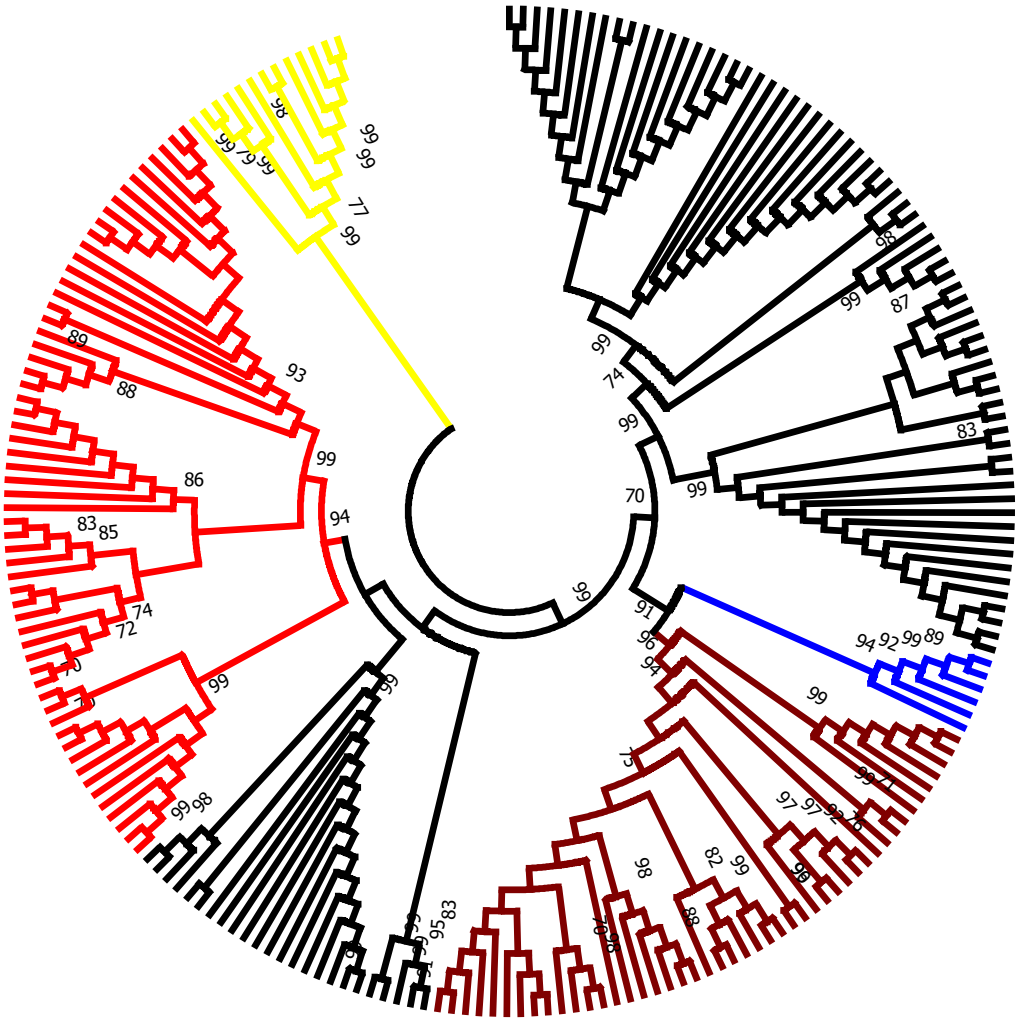
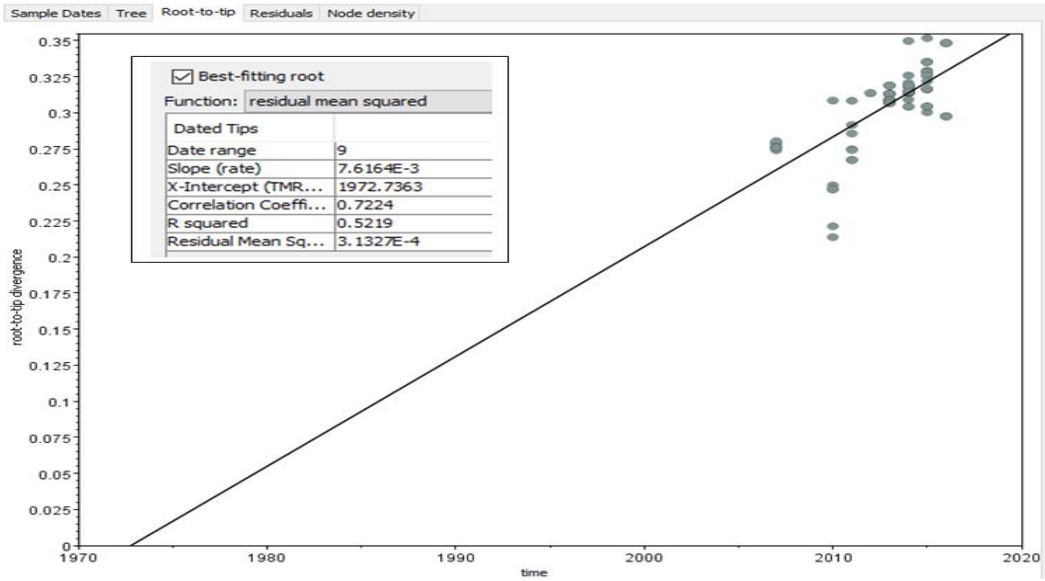
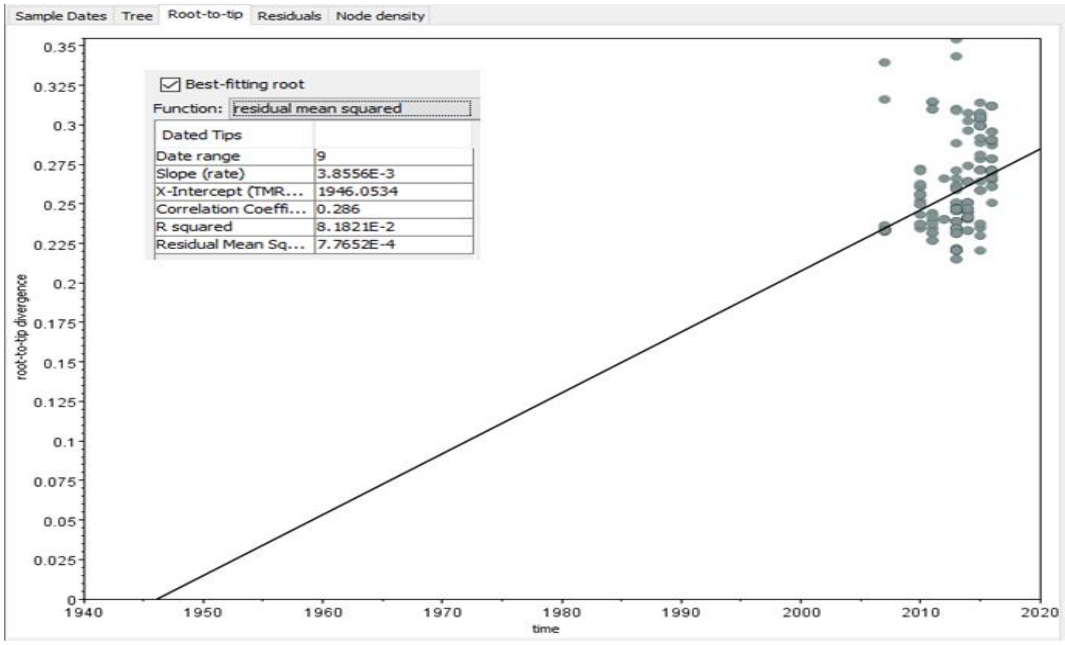


Figure 1. Maximum Likelyhood Phylogenetic tree, of Bat β-CoV from this study (n=219). Yellow colored clade indicate alpha CoV from bats, blue colured lines indicate Human CoV isolated in Africa. The clade in red represents lineage D Bat β-CoV, while the brown clolured clade represents lineage C Bat β-CoV. Bootstrap values above 60% are indicated at the nodes of the tree.



A



B

Figure 2. Root to tip regression analysis plots of the genetic distances against the year of isolation for the Bt-CoV sequences analyzed in this study. Figure 2a represents Afr-Bt-CoV, while 2b represents sequences of African, Asian and European origins.

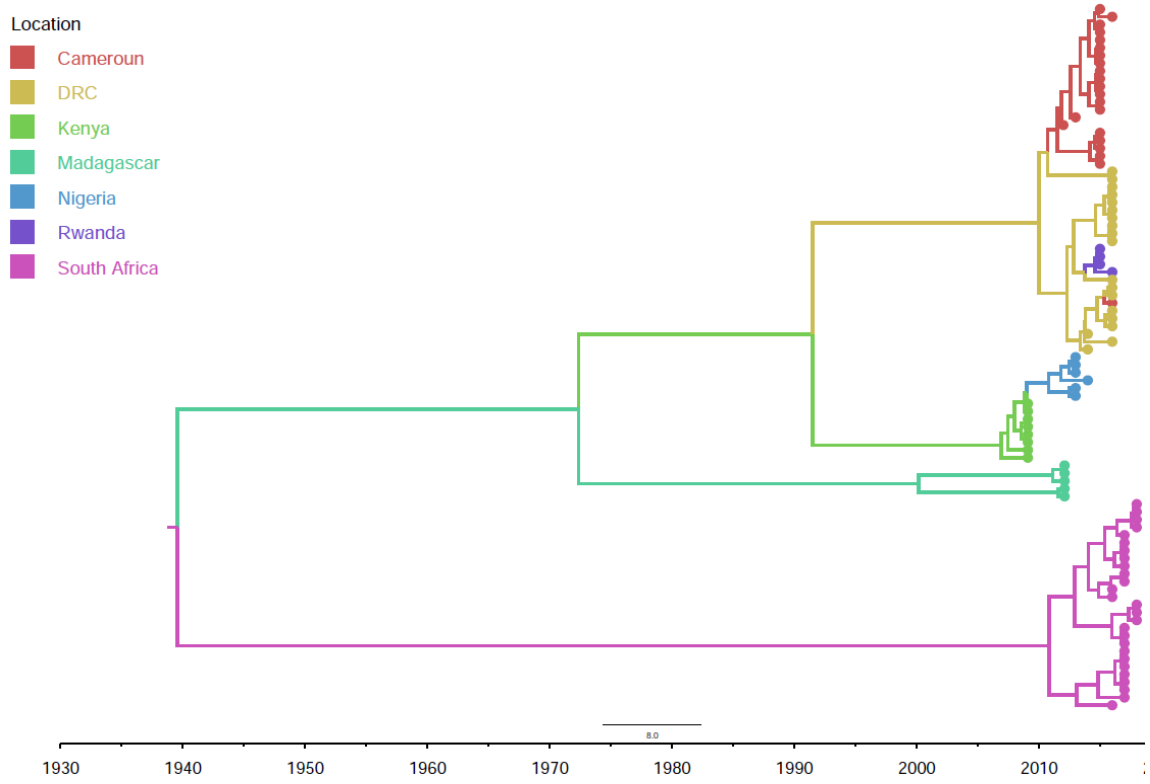


Figure 3. Time scaled Maximum clade credibility tree (MCC) with Bayesian phylogeographic reconstruction of partial RdrP gene region of African Bat β-CoV. Branch lengths are proportional to time in years, while branch colours are indicative of country of origin as shown in the legend.

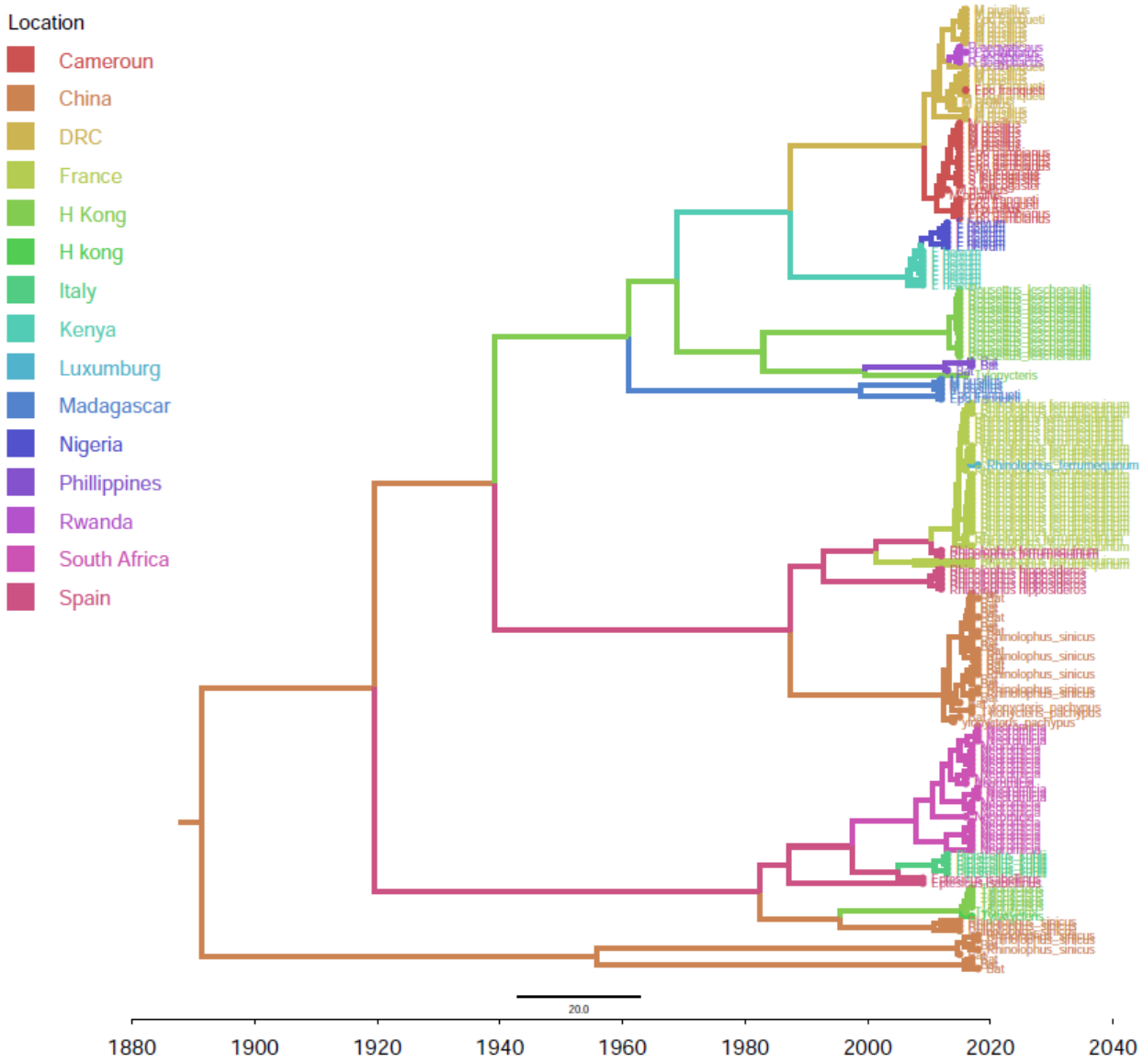


Figure 4. Time scaled Maximum clade credibility tree (MCC) with Bayesian phylogeographic reconstruction of partial RdrP gene region of Bat β-CoV. Branch lengths are proportional to time in years, while branch colours are indicative of country of origin as shown in the legend. Names of bat species of CoV isolation are shown in the tip labels

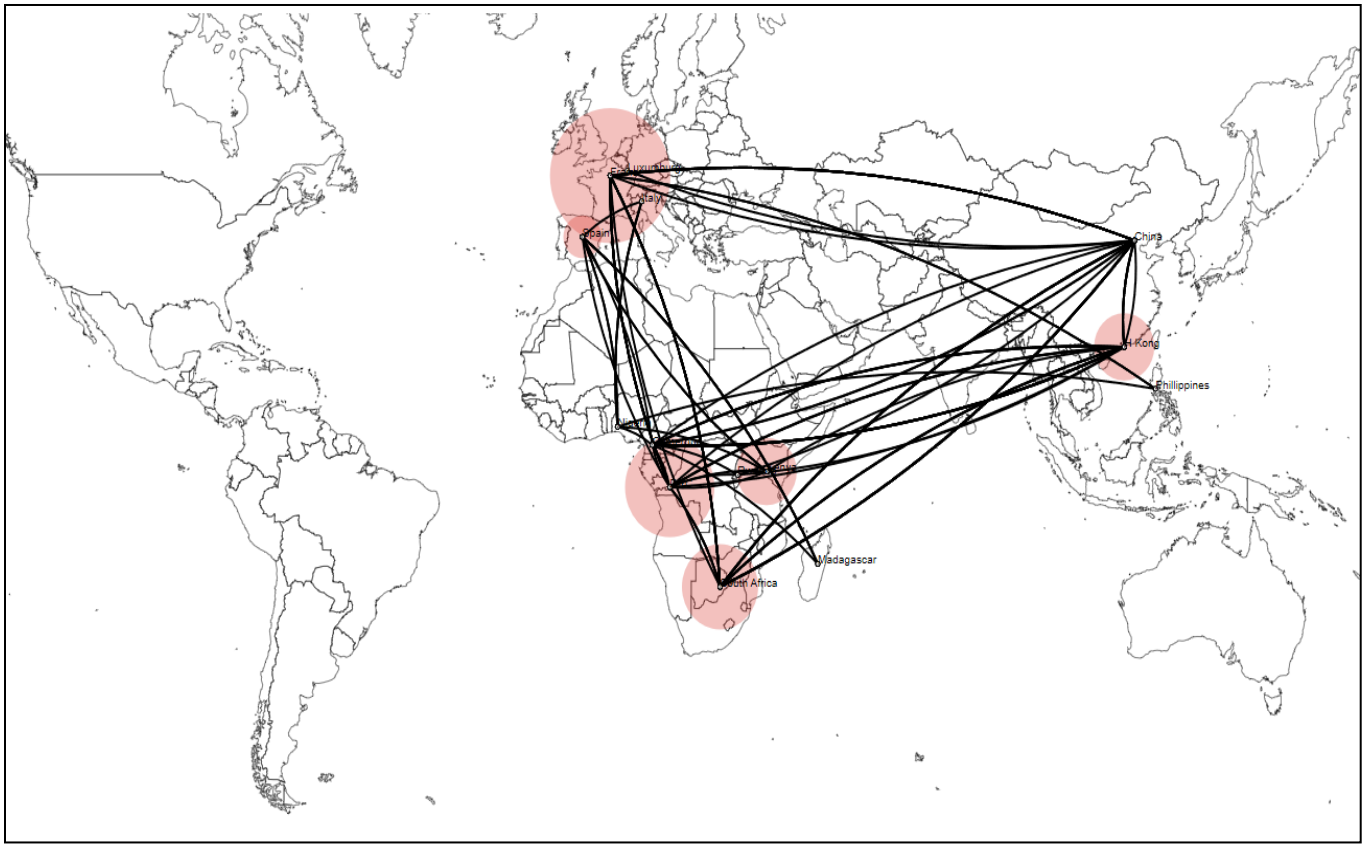


Figure 5. Spaciotemporal diffusion of Bat β-CoV, the names of the countires of isolation are written in map. The diamerter of the pink circles is proportional to the number of maximum clade credibility branches (MCC). Concave lines show a clockwise spread movement between two points, while the convex bending lines depit anticlockwise movement between the points.

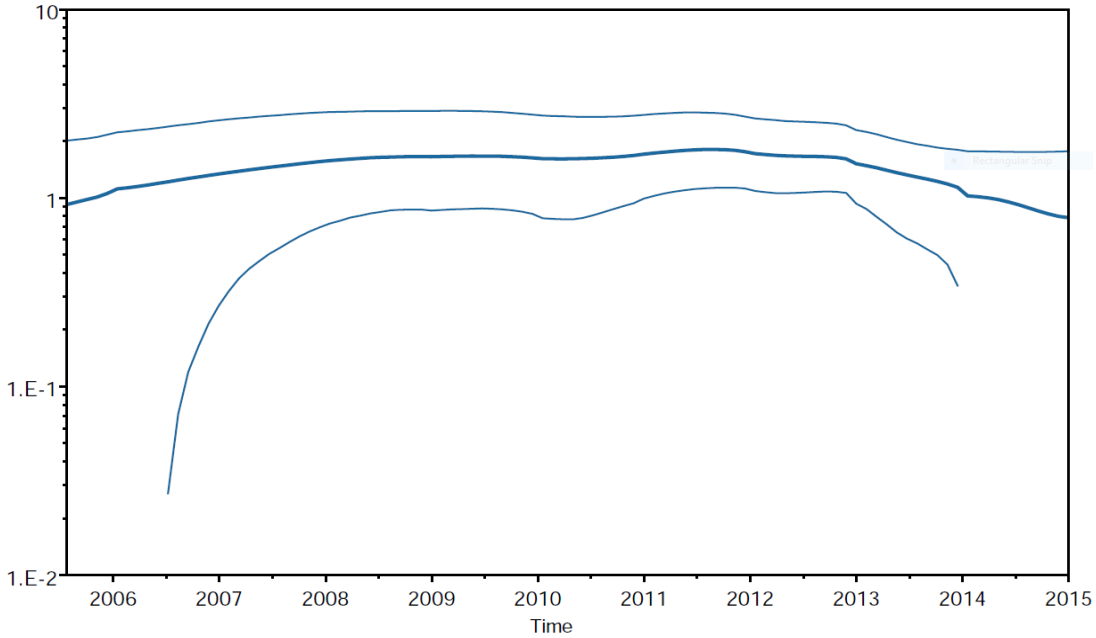


Figure 6. Bayesian Skygrid plot, showing effective virus population size estimates over time of Afr-β-CoV. The solis blue line represents the median population size, while the upper and lower blue lines represent the 95% high posterior density intervals (95% HPD) for the population.

Phylogeographic dispersal of the Bat β-CoV revealed numerous inter-continental spread events from China and Honk Kong into Cental Africa (DRC and Kenya), Cameroun in West Africa, and South Africa Figure 5. There were also dispersal of these viruses between France and some African countries such as Cameroun, South Africa and the DRC. There was also dispersal from Spain into Kenya, South Africa and Madagascar. The only dispersal event from Italy into Africa was into Nigeria. The dispersal observed between the Bt CoV isolated in Africa and the ones in Asia and Europe show evidence of long distance transmission event although a significant limitation to this was our inability to demonstrate bat migration patterns between these countries. Studies have shown the potential for African fruit bats to migrate long distances covering thousands of Kilometers, for instance a study using satellite telemertry in Zambia showed that *Elodiun hevium* is capable of covering thousands of kilometers during migration (Richter and Cummings, 2008). Another study showed that African bats were capable of migration exceeding 2000km (Ossa et al, 2012). Intra-continental dispersal events were observed between Cameroun, DRC and South Africa, as well as direct dispersal from Cameroon into Madagascar. The AfrBtCoV strains displayed steady state population demography as depicted by their Bayesian

Skyline plot (Figure 6). The population demography reported in this study might not represent the true picture of the virus population, as the dataset utilized in this study is limited by its size and might not represent the true demographic population of BtCoV in Africa.

We have presented data on the phylodiversity and evolutionary dynamics of Afr-β-CoV and their possible dispersal across the continent. We have shown the circulation of two distinct lineages B and C among African bat species and the possibility of inter-species transmission. A major limitation was our inability to analyze spike protein sequence data of these viruses as this would have shed more light on their evolution in relation to infectivity and transmission. We have however shown the importance of molecular surveillance of viruses with zoonotic potential such as coronaviruses. We advocate for broader trans-continental studies involving full genome sequences of BtCoV to further understand the drivers for their emergence and zoonotic spillovers into human population.

AUTHORS CONTRIBUTION

B.O.M and O.O.O conceptualized the study, B.OM, O.O.O, and B.A.O collected data, B.O.M conducted the formal analysis, B.O.M, and O.O.O software, B.O.M wrote the first draft, P.A.A contributed in intellectual discuss all authors read and approved the manuscript.

REFERENCES

Annan A, Baldwin HJ, Corman VM, Klose SM, Owusu M, Nkrumah EE,Badu EK, Anti P, Agbenyega O, Meyer B, Oppong S, Sarkodie YA, KalkoEK, Lina PH, Godlevska EV, Reusken C, Seebens A, Gloza-Rausch F, ValloP, Tschapka M, Drosten C, Drexler JF. 2013. Human betacoronavirus 2cEMC/2012-related viruses in bats, Ghana and Europe. Emerg Infect Dis 19:456–459.

Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P (2016) SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. Mol Biol Evol 33:2167–2169.

Corman VM, Baldwin HJ, Tateno AF, Zerbinati RM, Annan A, Owusu M,Nkrumah EE, Maganga GD, Oppong S, Adu-Sarkodie Y, Vallo P, da SilvaFilho LV, Leroy EM, Thiel V, van der Hoek L, Poon LL, Tschapka M,Drosten C, Drexler JF. 2015. Evidence for an ancestral association ofhuman coronavirus 229E with bats. J Virol 89:11858 –11870.

Drexler JF, Corman VM, Drosten C. 2014. Ecology, evolution and classificationof bat coronaviruses in the aftermath of SARS. Antiviral Res101:45–56.

Hamre, D., Procknow, J.J., 1966. A new virus isolated from the human respiratory tract. Proc. Soc. Exp. Biol. Med. 121, 190–193.

Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Tsoi, H.W., Wong, B.H., Wong, S.S., Leung, S.Y., Chan, K.H., Yuen, K.Y., 2005. Severe acute respiratory syndrome coronaviruslike virus in Chinese horseshoe bats. Proc. Nat. Acad. Sci. USA 102, 14040–14045.

Lau SK, Li KS, Tsang AK, Shek CT, Wang M, Choi GKY, Guo R, WongBH, Poon RW, Lam CS, Wang SY, Fan RY, Chan KH, Zheng BJ, WooPC, Yuen KY.2012. Recent transmission of a novel alphacoronaivurs, batcoronavirus HKU10, from Leschenault's rousettes to Pomona leaf-nosedbats: first evidence of interspecies transmission of coronavirus betweenbats of different suborders. J. Virol.86:11906–11918.

Lau SK, Li KS, Tsang AK, Lam CS, Ahmed S, Chen H, Chan KH, Woo PC, Yuen KY. Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus. J Virol.87(15):8638-50.2013.

Leopardi S, Oluwayelu D, Meseko C, Marciano S, Tassoni L, Bakarey S, Monne I, Cattoli G, De Benedicts P. (2016). The close genetic relationship of lineage D Betacoronavirus from Nigerian and Kenyan straw-colored fruit bats (Eidolon helvum) is consistent with the existence of a single epidemiological unit across sub-Saharan Africa. Virus Genes.

Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian Phylogeography Finds Its Roots. PLoS Comput Biol. 2009;5:e1000520. doi:10.1371/journal.pcbi.1000520.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B.T., Zhang, S.,Wang, L.F., 2005. Bats are natural reservoirs of SARS-like coronaviruses. Science(New York, NY) 310, 676–679.

Li X, Zai J, Zhao Q, Nie Q, Li Y, Foley BT, Chaillon A. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. J Med Virol. 2020 Feb 27. doi: 10.1002/jmv.25731.

Masters P.S and Perlman S. Chapter 28, Coronaviridea; In Knipe D. Howley P, Fields Virology 6th edition. Lippincot Williams and Wilkins. 2013.

McIntosh, K., Dees, J.H., Becker, W.B., Kapikian, A.Z., Chanock, R.M., 1967. Recovery in tracheal organ cultures of novel viruses from patients with respiratorydisease. Proc. Nat. Acad. Sci. USA 57, 933–940.

Min, J., Cella, E., Ciccozzi, M. *et al.* The global spread of Middle East respiratory syndrome: an analysis fusing traditional epidemiological tracing and molecular phylodynamics. *glob health res policy* **1,** 14 (2016). https://doi.org/10.1186/s41256-016-0014-7.

Ossa G, Kramer-Schadt S, Peel AJ, Scharf AK, Voigt CC (2012) The Movement Ecology of the Straw-Colored Fruit Bat,Eidolon helvum, in Sub-SaharanAfrica Assessed by Stable Isotope Ratios. PLoS ONE 7(9): e45729. doi:10.1371/journal.pone.0045729.

Rambaut, A.; Lam, T.T.; Carvalho, L.M.; Pybus, O.G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. **2016**, 2, vew007.

Richter HV, Cumming GS (2008) First application of satellite telemetry to track African straw-coloured fruit bat migration. J Zool 275: 172–176.

Saif, L.J., 2004. Animal coronaviruses: what can they teach us about the severe acute respiratory syndrome? Rev. Sci. Tech. 23, 643–660.

Suchard MA, Lemey P, Baele G et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evolution. 2018. 4, vey016. DOI:10.1093/ve/vey016.

Tao Y, Shi M, Chommanard C, QueenK, Zhang J, Markotter W, Kuzmin IV, Holmes EC, Tong S. 2017. Surveillance of bat coronavirusesin Kenya identifies relatives of human coronaviruses NL63 and 229E and theirrecombination history. J Virol91:e01953-16.

Vijgen L, Keyaerts E, Lemey P, Maes P, Van Reeth K, Nauwynck H, Pensaert M, Van Ranst M. Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. J Virol. 2006;80(14):7270–7274.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D., Fouchier, R.A., 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N.Engl. J. Med. 367, 1814–1820.

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin.Nature. 2020 Mar;579(7798):270-273.