GENOMIC DIVERSITY AND PHYLODYNAMICS OF SARS-COV-2 IN YAVATMAL DISTRICT OF MAHARASHTRA

Ashwin Atkulwar
Department of Zoology, Amolakchand Mahavidyalaya, Yavatmal.
ashuatkulwar@gmail.com

Abstract
In the current study, the phylodynamics and genomic diversity of SARS-CoV-2 virus are studied to gain insight into the genetic diversity and progression of pandemic in Yavatmal District of Maharashtra. Total 39 genomes of the virus deposited to the GISAID virus database from Yavatmal were retrieved and analyzed using highly sophisticated bioinformatic methods. Out of these 39, 27 unique haplotypes were identified from Yavatmal. Pangolin COVID-19 Lineage Assigner tool was used to assign lineages to all 27 genomes. The lineage assignment confirmed the dominance of Omicron variant over Delta. The phylodynamic analysis, and genomic diversity was estimated using Bayesian Skyline Plot Method (BSP) and Lineage Through Time Plot (LTT Plot) as implemented in BEAST v2.2.0 and DnaSPv6 software package. The outcome of the analysis reveals high genetic diversity, while phylodynamic analysis confirmed upward and downward surges in the effective population size (N_e) of virus during first and second wave of COVID in the region. Through this analysis, we suggest continuous and widespread genome sequencing of virus in the region to track the evolution and genomic diversity of SARS-CoV-2.
Keywords: *SARS-CoV-2, Phylodynamics, Genomic diversity, Yavatmal*

Introduction

Throughout human history, viral outbreaks have consistently represented a most common threat to global health, from the Spanish flu epidemic of 1918 to the recent coronavirus (CoV) pandemic that emerged in December 2019 (Patterson and Pyle, 1991; Spreeuwenberg et al., 2018; Biondi-Zoccai et al., 2020). CoVs are the members of family Coronaviridae, characterised by their crown-like structure. They possess a positive-sense ssRNA genome ranging from 26 to 32 kb, with SARS-CoV-2 having a genome size of 29,844 to 29,891 nt (Woo et al., 2010; Lu et al., 2020; Chan et al., 2020). These CoVs has been documented in both livestock and wild species also (Lam et al., 2020; Pensaert et al., 1986; Li et al., 2005; Hu et al., 2015). Their rapid evolutionary rate enables them to adapt and infect a wide range of hosts with varying tissue tropism (Graham and Baric, 2010; Li, 2016). For the first time in the history, the coronavirus outbreak, known as porcine epidemic diarrhoea, was documented in pigs in Europe and Asia in 1971 (Pensaert et al., 1986). The zoonotic transition of CoVs to humans began with the global outbreak of SARS-CoV-1 in China during 2002–2003, followed by the emergence of the Middle East respiratory syndrome (MERS) in 2012 (Hilgenfeld et al., 2013; Gao et al., 2016; Coleman and Frieman, 2013). While the first case of COVID-19 epidemic, i.e. the case of the novel CoV, specifically the new strain SARS-CoV-2, was identified in Wuhan, China, on December 31, 2019, and it spread worldwide by January 2020 (Zhou et al., 2020; Woo et al., 2020). The WHO declared this outbreak a global pandemic on March 11, 2020. This novel CoV strain is primarily recognized for its potential to infect the human respiratory tract, resulting in severe pneumonia in affected individuals. Whole-genome sequencing and phylogenetic analysis indicate that bats may be the original hosts (Lu et al., 2020). However, several studies suggest that the S gene of SARS-CoV-2 shares significant functional domain similarities with isolates from pangolins (Wong et al., 2020; Xiao et al., 2020; Lam et al., 2020). Progress in computational genomics and the availability of genomic data enhance our understanding of the spatial distribution and epidemiology of pandemics. The whole-genome sequencing and annotation of a novel SARS-CoV-2 reference genome enable rapid sequencing and assembly of many SARS-CoV-2 genomes globally.

Pathogen genomics is a valuable resource not only for tracking the current pandemic but also for facilitating effective vaccine development. The genomic information shared by laboratories around the globe offers a basis for identifying the origins and transmission dynamics of the virus in their nations (Zhang and Holmes, 2020). Genomic investigations have already been vital for contact tracing infections and will gain further importance during the second wave of infections after the lockdown is lifted in the ongoing COVID-19 pandemic (Stevens et al., 2017).

Yavatmal is a city and municipal council of Maharashtra state and also the administrative headquarters of Yavatmal District. It is located 90 km away from divisional headquarters Amravati. It is one of the tribal populated districts of Maharashtra. Hence, it is crucial to decipher the genomic studies on SARS-CoV-2 virus from this region. Current research deals with the investigation regarding population dynamics and genomic diversity of SARS-CoV-2 virus from

Yavatmal district. During this in-silico analysis, a total of 27 SARS-CoV-2 whole genome sequences (Haplotypes) of were used tract down the phylodynamics.

Materials and Methods:

The whole-genome sequences of SARS-CoV-2 isolates from Yavatmal, available in GISAID virus database was retrieved. A total of 39 genomes were retrieved and aligned using Clustal Omega (Sievers et al., 2011), while the Lineage Assignment, was carried out in Pangolin COVID-19 Lineage Assigner. Number of haplotypes and haplotype diversity was calculated using DnaSPv6 (Rozas et al., 2017). Out of 39 genomes, 27 unique haplotypes were identified and later utilized for downstream analysis. The Bayesian Skyline Plot (BSP) is one of the robust and extensively used methods to investigate the phylodynamics of species. The Bayesian Skyline Plot (BSP) technique, as applied in BEAST v2.2.0, was used to estimate the effective population size (N_e) for these 27 haplotypes. The analysis was performed using the stamped-date method, which included the HKY nucleotide substitution model, along with 4 gamma category count and coalescent Bayesian skyline tree priors. An MCMC chain length of 10 million steps was executed, with the first 10% discarded as burn-in. A strict molecular with clock rate of $1x10^{-3}$ subs/site/year was implemented. The log file and tree log file were analysed to draw BSP in Tracer v1.7.0. Diversification rate of these SARS-CoV-2 virus was estimated by Lineage Through Time (LTT) plot (Rabosky *et al.,* 2008) in BEAST v2.2.0, with similar Bayesian Skyline Plots (BSP) tree priors.

Result and Discussion:

The genomic diversity calculated confirmed the presence of 27 unique haplotypes wit haplotype diversity 0.9420 in Yavatmal. This haplotype diversity (Hd) within SARS-CoV-2 from Yavatmal was found to be very high. Such a high haplotype diversity within Yavatmal reveals high genomic diversity of SARS-CoV-2. Out of these 39 haplotypes submitted from Yavatmal, a total of 24 Omicron variant was identified, while 11 genomes were belonged to Delta variant, and 2 samples were unidentified. Total 13 different lineages, belongs to Omicron and Delta variants was found in Yavatmal (Figure. 1). Moreover, lineage assignment also confirmed higher prevalence of XBB. 1.16 (Omicron variant) lineages following by B.1.617.2 (Delta variant) (Figure. 1). This dominance of XBB. 1.16 lineage could be due to high virulence Omicron variant than Delta or sequencing biased. Previously published studies also revealed higher transmission rate of Omicron variant in Maharashtra (Atkulwar et al., 2023). The phylodynamic analysis reveals the exponential rise in effective population size (N_e) of virus in Yavatmal from the March 2020, which will continue to rise till December 2020 (Figure. 2 (a)). From January 2021, the effective population size is found to be constant up to March 2022, and then sharply goes downward from June 2022 (Figure. 2 (a)). This sharp decrease in effective population size (N_e) was observed roughly from September 2022 and continued till January 2023 (Figure. 2 (a)). Similar timeframe or trend of SARS-CoV-2 isolates also shaped the diversification and evolution of virus, observed in Lineage Through Time (LTT) plot (Figure. 2 (b)).

Figure. 1. Number of different lineages of SARS-CoV-2 virus belongs to Delta and Omicron variants, identified in the Yavatmal.

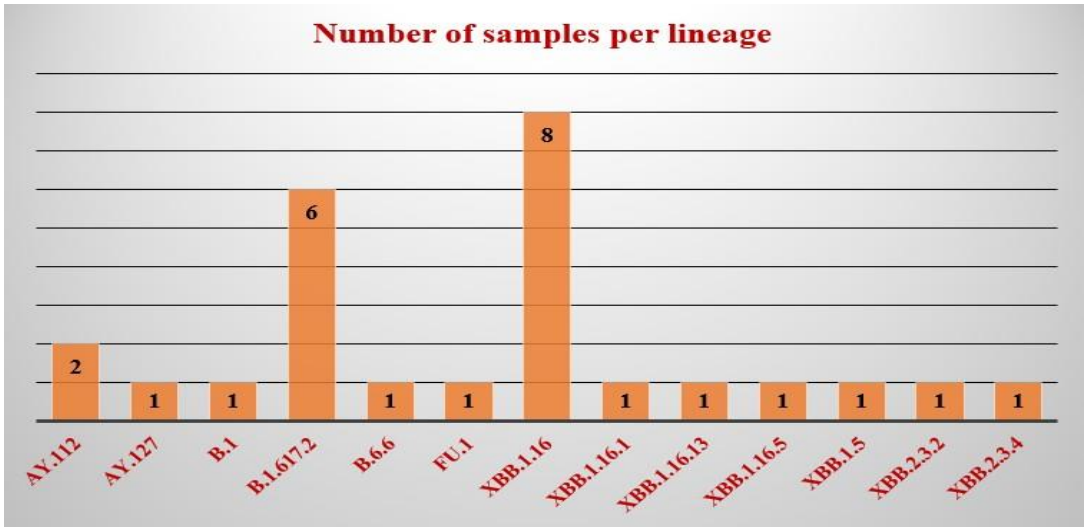
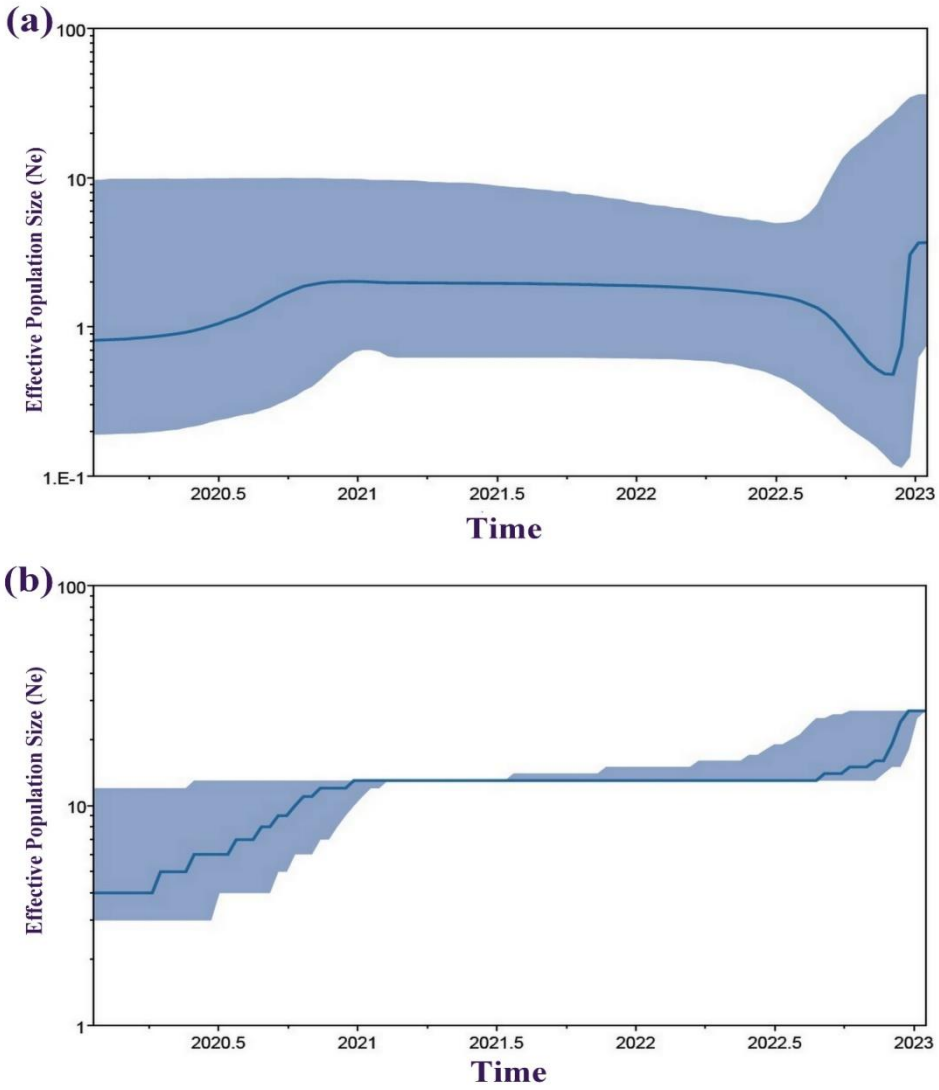


Figure 2. Population dynamics of SARS-CoV-2 in Yavatmal. Bayesian Skyline Plot and Lineage Through Time Plot, (a) & (b) constructed using 27 unique haplotypes, sequenced and deposited from Yavatmal. Bayesian skyline tree priors depict the effective population size on the Y-axis, while the X-axis denotes the timeline in months. N_e: Effective population size.



Conclusion:

This is the first genomic study from Yavatmal District of Maharashtra that cast light on the phylodynamics and genomic diversity of SARS-CoV-2 virus. This study confirmed high genomic diversity of SARS-CoV-2 virus including higher dominance of Omicron variant. Findings of this study also uncover the higher and lower surges in the effective population size of virus with exact timeline during the pandemic. Considering these findings, this study suggests continuous genome sequencing, and full vaccination of the population to counter further outbreaks of such a notorious virus.

Acknowledgments:

I gratefully acknowledge all authors and submitting laboratories of the sequences from GISAID.

References:

1. Atkulwar, A., Rehman, A., Imaan, Y., *et al.* (2023). Analyses of Omicron Genomes from India Reveal BA.2 As a More Transmissible Variant. *Eur. J. Biol. Res*. 10-17.
2. Biondi-Zoccai, G., Landoni, G., Carnevale, R., Cavarretta, E., Sciarretta, S., Frati, G. (2020). SARS-CoV-2 and COVID-19: facing the pandemic together as citizens and cardiovascular practitioners. *Minerva Cardioangiol.* 68(2), 61–64.
3. Chan, J.F., Kok, K.H., Zhu, Z. *et al.* (2020). Genomic characterization of the 2019 novel human-pathogenic corona virus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* 9(1), 221–236.
4. Coleman, C.M., Frieman, M.B. (2013). Emergence of the Middle East respiratory

syndrome coronavirus. *PLoS Pathog.* 9, e1003595

5. Gao, H., Yao, H., Yang, S., Li, L. (2016). From SARS to MERS: evidence and speculation. *Front. Med.* 10, 377–382.
6. Graham, R.L., Baric, R.S. (2010). Recombination, reservoirs, and the modular spike: mechanisms of corona virus cross-species transmission. *J. Virol.* 84(7), 3134–3146.
7. Hilgenfeld, R., Peiris, M. (2013). From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antiviral Res.* 100, 286–295.
8. Hu, B., Ge, X., Wang, L. *et al.* (2015). Bat origin of human corona viruses. *Virol. J.* 12, 221.
9. Lam, T.T., Shum, M.H., Zhu, H. *et al.* (2020). Identifying SARS-CoV-2 related corona viruses in Malayan pangolins. *Nature* 583, 282–285.
10. Lam, T.T., Shum, M.H., Zhu, H. *et al.* (2020). Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. *bioRxiv* doi:10.1101/2020.02.13.945485.
11. Li F. (2016). Structure, function, and evolution of corona virus spike proteins. *Annu. Rev. Virol.* 3(1), 237–261.
12. Li, W., Shi, Z., Yu, M. *et al.* (2005). Bats are natural reservoirs of SARS-like corona viruses. *Science* 310(5748), 676–679.
13. Lu, R., Zhao, X., Li, J., Niu, P. *et al.* (2020). Genomic characterization and epidemiology of 2019 novel corona virus: implications for virus origins and receptor binding. *Lancet* 395(10224), P565–574.
14. Patterson, K.D., Pyle, G.F. (1991). The geography and mortality of the 1918 influenza pandemic. *Bull. Hist. Med.* 65(1), 4–21.
15. Pensaert, M., Callebaut, P., Vergote, J. (1986). Isolation of a porcine respiratory, non-enteric corona virus related to transmissible gastroenteritis. *Vet. Q* 8, 257–261.
16. Rabosky, D.L. and Lovette, I.J. (2008). Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? *Evol*. 62(8):1866-75.
17. Rozas, J., Ferrer-Mata, A., S´anchez-DelBarrio, C. *et al.* (2017). DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Mol. Biol. Evol.* 34, 3299–3302.
18. Sievers, F., Wilm, A., Dineen, D. *et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.
19. Spreeuwenberg, P., Kroneman, M., Paget, J. (2018). Reassessing the Global Mortality Burden of the 1918 Influenza Pandemic. *Am. J. Epidemiol.* 187(12), 2561–2567.
20. Stevens, E.L., Timme, R., Brown, E.W. *et al.* (2017). The public health impact of a publicly available, environmental database of microbial genomes. *Front. Microbiol.* 8, 808.
21. Wong, M.C., Cregeen, S.J.J., Ajami, N.J., Petrosino, J.F. (2020). Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. *bioRxiv* doi:10.1101/2020.02.07.939207.
22. Woo, P.C., Huang, Y., Lau, S.K., Yuen, K.Y. (2010). Corona virus genomics and bioinformatics analysis. *Viruses* 2(8), 1804–1820.
23. Wu, F., Zhao, S., Yu, B. *et al.* (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269.
24. Xiao, K., Zhai, J., Feng, Y. *et al.* (2020). Isolation and characterization of 2019-nCoV-like coronavirus from Malayan pangolins. *bioRxiv* doi:10.1101/2020.02.17.951335.
25. Zhang, Y.Z., Holmes, E.C. (2020). A genomic perspective on the origin and emergence of SARS-CoV-2. *Cell* 181(2), 223–227.
26. Zhou, P., Yang, X., Wang, X. *et al.* (2020). A pneumonia outbreak associated with a new corona virus of probable bat origin. *Nature* 579, 270–273.