

Disentangling the Dynamics of VOCs and VOIs during the COVID-19 Pandemic in Peru using Bayesian Phylogenetics

Santiago Justo Arevalo^{*,1,2,3}, Cinthy Jimenez Silva⁴, Carmen Sofia Uribe Calampa¹, Remco Bouckaert⁴, Joao Renato Rebello Pinho^{2,5}

1.- Facultad de Ciencias Biológicas, Universidad Ricardo Palma, Lima – Peru.

2.- Laboratório Clínico do Hospital Israelita Albert Einstein, São Paulo – Brasil

3.- Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo – Brasil

4.- University of Auckland, Auckland, New Zealand

5.- Department of Gastroenterology and Pathology, University of São Paulo School of Medicine, São Paulo, Brazil

*Correspondence to: Santiago Justo Arevalo, santiago.justo@urp.edu.pe

Supporting Information

Contents:

Supplementary Figures

Figure S1. Sampling process to select genomes to infer the date and location of the origin of the Lambda VOC.

Figure S2. Two different samples and different clock models showed the same date and location of the origin of Lambda.

Figure S3. Sampling process to select genomes to infer the date and location of the introduction of the VOCs to Peru.

Figure S4. Comparison between the estimations of the VOCs introductions to Peru for sample 1 and 2.

Figure S5. Delta VOC introduction to Peru.

Figure S6. Alpha VOC was introduced from the United Kingdom to Peru.

Figure S7. Gamma VOC was introduced from Brazil to Peru.

Figure S8. Mu VOC was introduced from the USA to Peru.

Figure S9. Omicron VOC introduction to Peru.

Figure S10. Sampling to estimate the trends of the Effective reproductive number or the Effective population size of the VOCs in Peru.

Figure S11. Sampling to estimate the trends of the Effective reproductive number or the Effective population size of the VOCs in Peru.

Figure S12. Comparison of the parameters estimated for the Birth-Death skyline model using the two different sampling approaches.

Figure S13. Comparison of the parameters estimated for the BICEPS model using the two different sampling approaches.

Figure S14. Adjusted relative prevalence of each variant.

Figure S15. Estimated number of cases and effective population size (N_e) and Effective reproductive number (R_t) calculated from genomes by Bayesian phylogenetics or from the number of cases follow similar trends.

Figure S16. Distribution of VOCs by region in four different stages of the pandemic.

Figure S17. Maximum likelihood analysis of Omicron lineages to identify well-supported Peruvian lineages.

Figure S18. Correlation between number of cases and genomes in Peru.

Supplementary Tables

Table S1. Bayes factor estimation to compare the Strict clock and the Uncorrelated relaxed exponential clock model for the phylogenetic analysis of the origin of Lambda.

Table S2. Classification of Peruvian Cities by Regions.

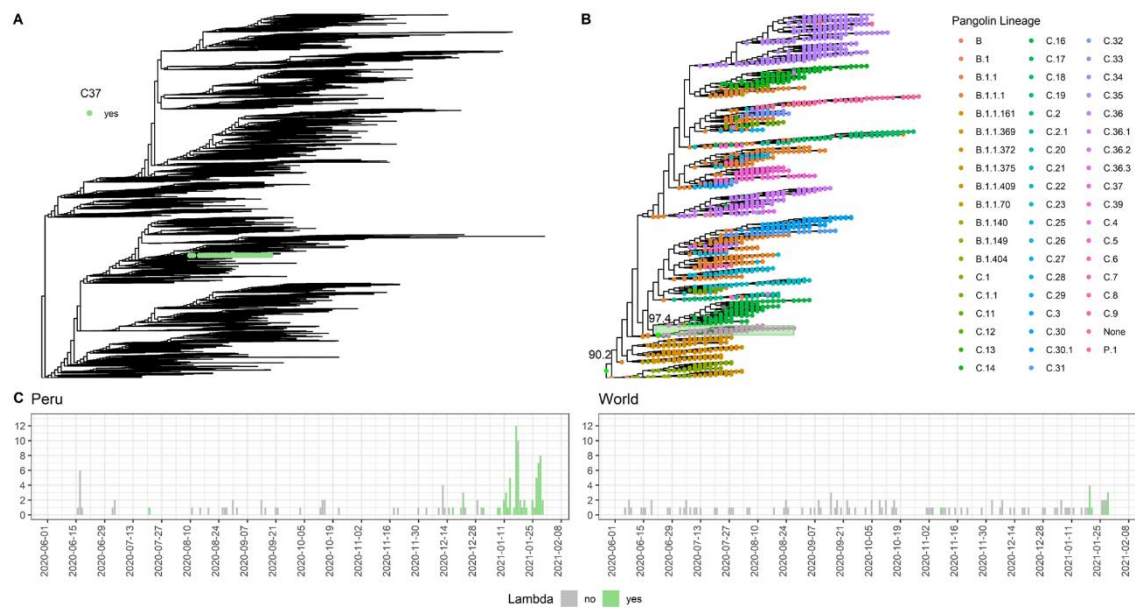


Figure S1. Sampling process to select genomes to infer the date and location of the origin of the Lambda VOC. A) High-quality genomes collected between July 1st, 2020 and January 31th 2021 available in GISAID at January 26th, 2022 were used to obtain a maximum likelihood tree. B) The well-supported node that contains the Lambda clade and other likely related clades was identified and examined. C) A sample of the tree in B that contains all the Lambda genomes plus some genomes of each of the lineages in that tree was collected. We show the number of genomes by date belonging to Lambda (green) or any other lineage (grey) collected in Peru (left) or any other country (world) that are part of the sample that we used to infer the date and location of the Lambda origin.

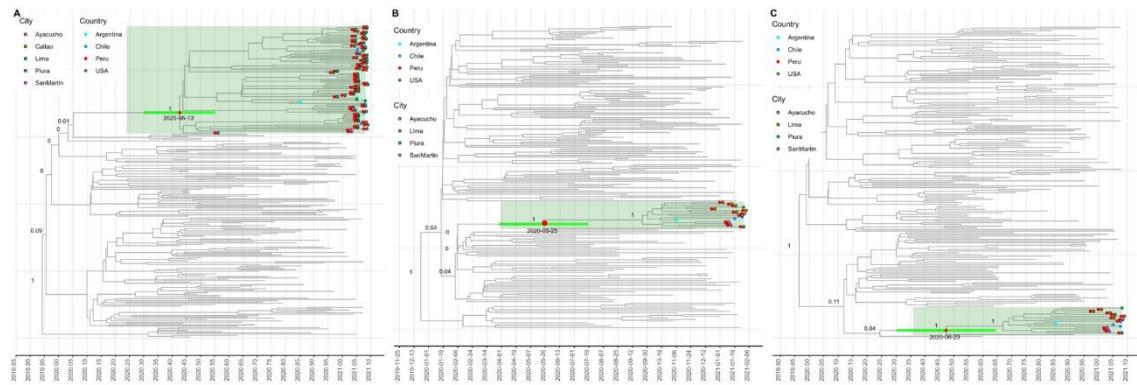


Figure S2. Two different samples and different clock models showed the same date and location of the origin of Lambda. Maximum clade credibility tree of full SARS-CoV-2 genome sequences under a strict clock model with a sample of genomes collected up to September 18th, 2021 (A); under an uncorrelated relaxed exponential clock model with a sample of genomes collected up to January 26th, 2022 (B); or under a strict clock model with a sample of genomes collected up to January 26th, 2022, using three partitions. The Lambda clade is highlighted (green shadow). The green bar represents the high posterior density 95 % of the date of the node that represents the most recent common ancestor of the Lambda clade. This node and the tips belonging to the Lambda are coloured according to its country of origin. Tips from Peru has an additional circle on the right coloured according to the region of collection. Numbers in the nodes represent its posterior probability.

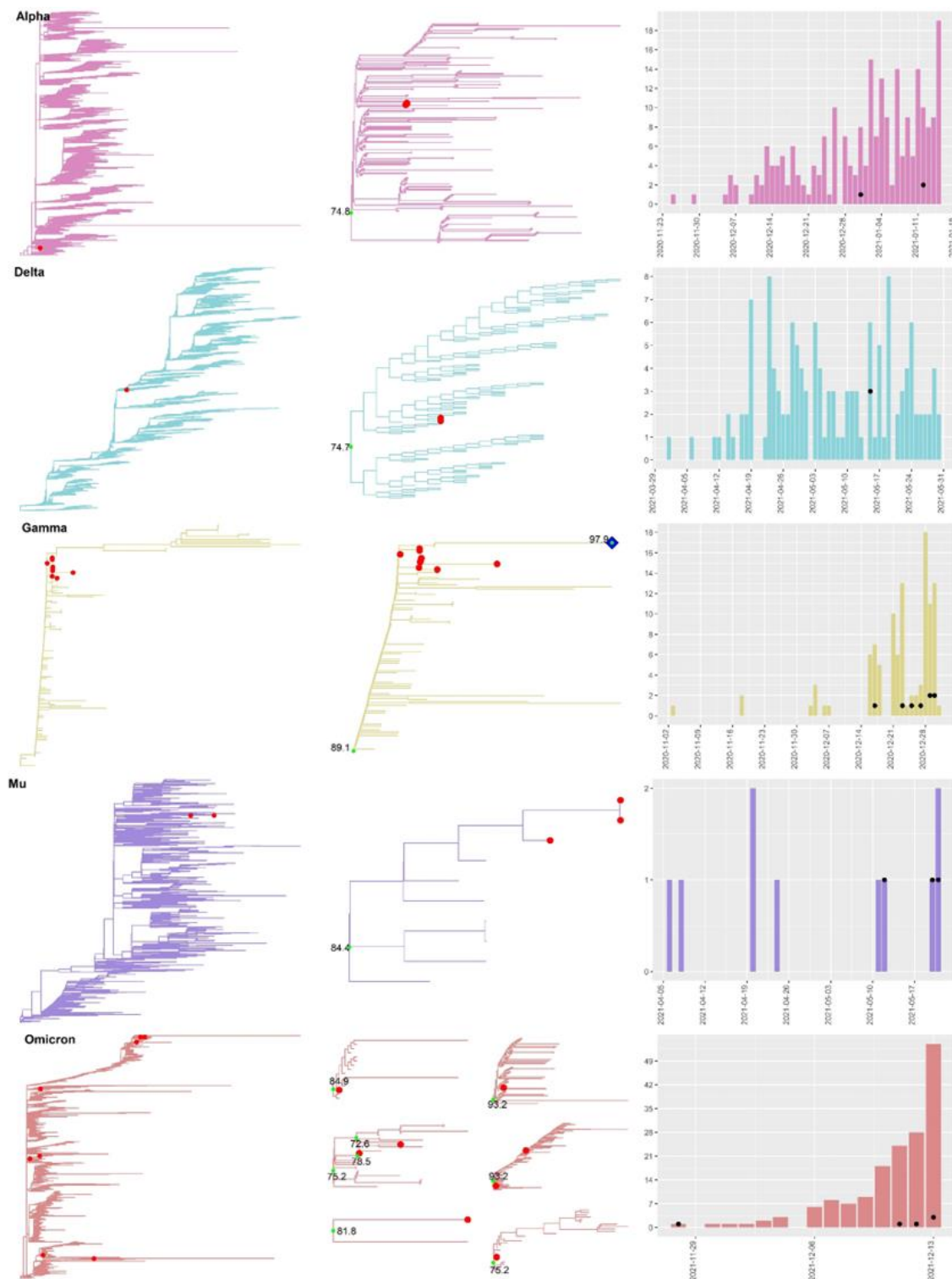


Figure S3. Sampling process to select genomes to infer the date and location of the introduction of the VOCs to Peru. A) High-quality genomes of the VOC of interest collected between 45 days before and 15 days after the collection date of the first genome from Peru were used to reconstruct a maximum likelihood phylogeny. B) The well-supported nodes or nodes that contain Peruvian genomes were identified. C) A sample or all of the genomes in that clade were collected and used to perform a Bayesian phylogeny to determine the introduction of the VOC to Peru. In A and B, Peruvian genome is highlighted with a red circle and the SH-aLRT value of the nodes of the selected clades are shown. In C, Peruvian genomes are highlighted with a black circle. Same procedure was used for the genomes of the VOC of interest collected between 30 days before and 30 days after the collection date of the first genome from Peru.

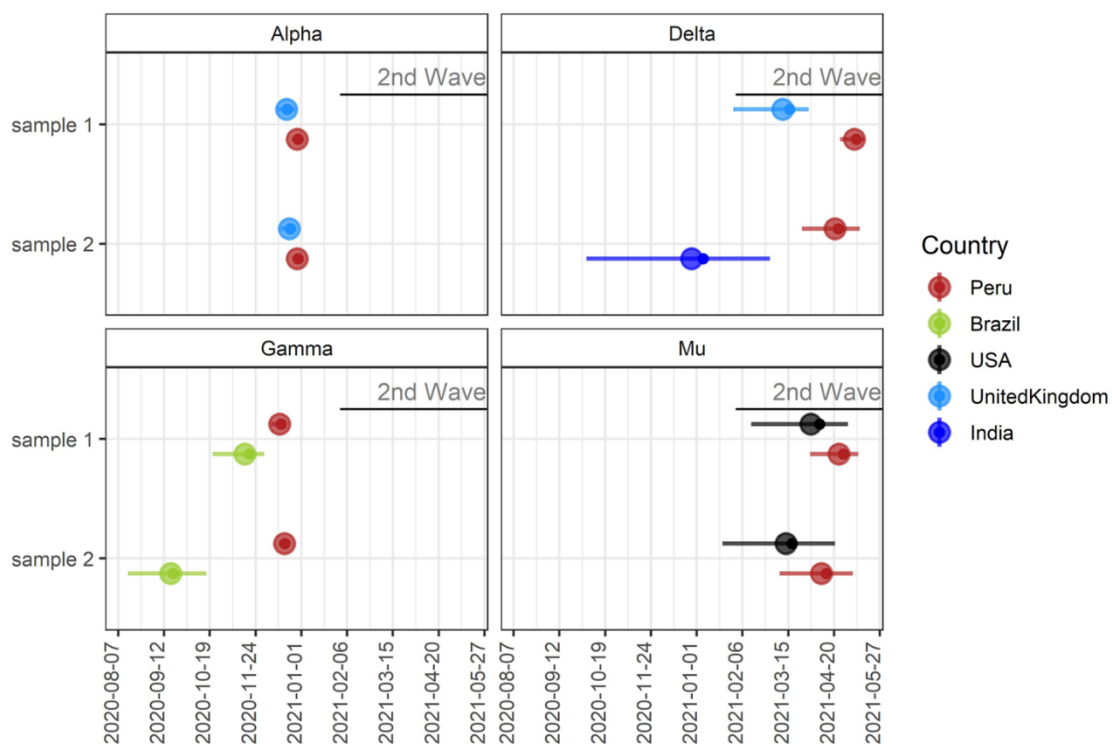


Figure S4. Comparison between the estimations of the VOCs introductions to Peru for sample 1 and 2.
 High posterior density 95 % of the highly supported nodes (posterior ≥ 0.95) that represent the introduction of each of the VOCs to Peru. They are coloured according to the source country.

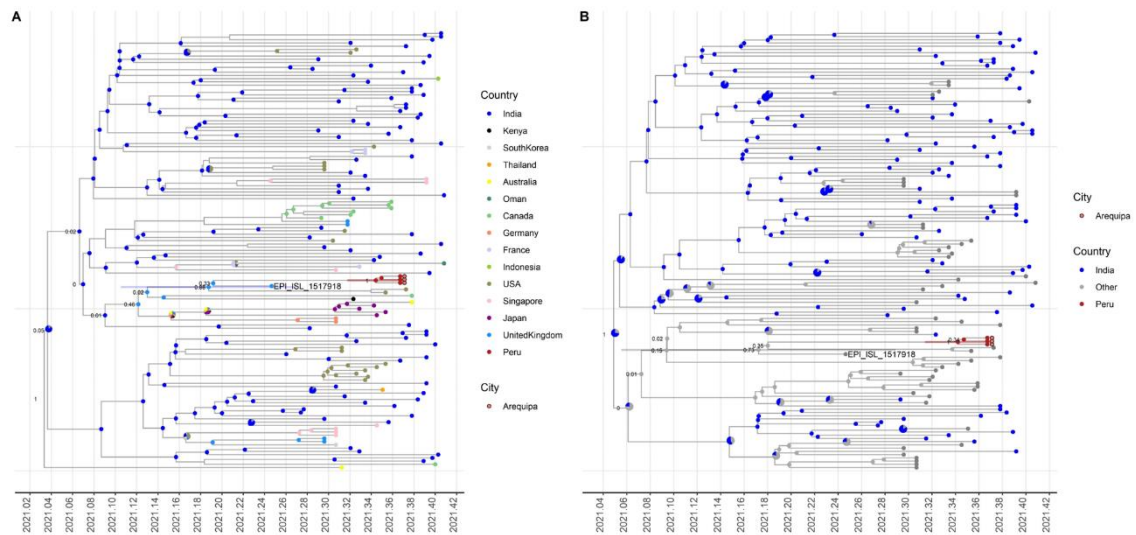


Figure. S5. Delta VOC introduction to Peru. Maximum clade credibility tree of full SARS-CoV-2 genome sequences under a strict clock model showing the introduction of the Delta VOC to Peru. Blue and red bars represent the high posterior density 95 % of the date of the nodes that represent the introduction of the Delta VOC to Peru. Nodes and the tips are coloured according to its country of origin (A) or the location, in the case of the analysis with a different grouping of the discrete states (B). Tips from Peru has an additional circle on the right coloured according to the region of collection. Numbers in the nodes represent its posterior probability.

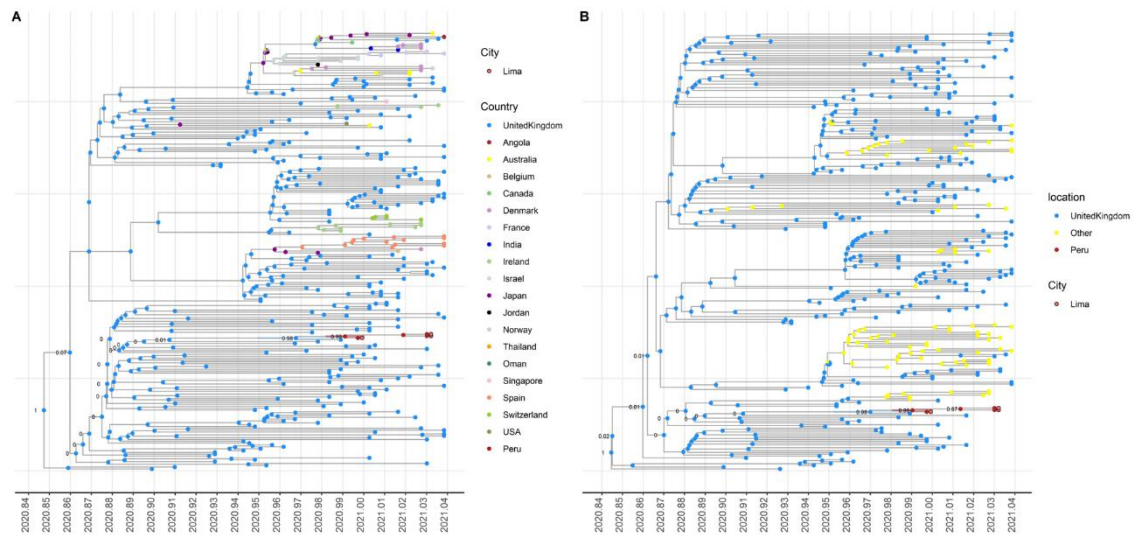


Figure S6. Alpha VOC was introduced from the United Kingdom to Peru. Maximum clade credibility tree of full SARS-CoV-2 genome sequences under a strict clock model showing the introduction of the Alpha VOC to Peru. Cyan and red bars represent the high posterior density 95 % of the date of the nodes that represent the introduction of the Alpha VOC to Peru. Nodes and the tips are coloured according to its country of origin (A) or the location, in the case of the analysis with a different grouping of the discrete states (B). Tips from Peru has an additional circle on the right coloured according to the region of collection. Numbers in the nodes represent its posterior probability.

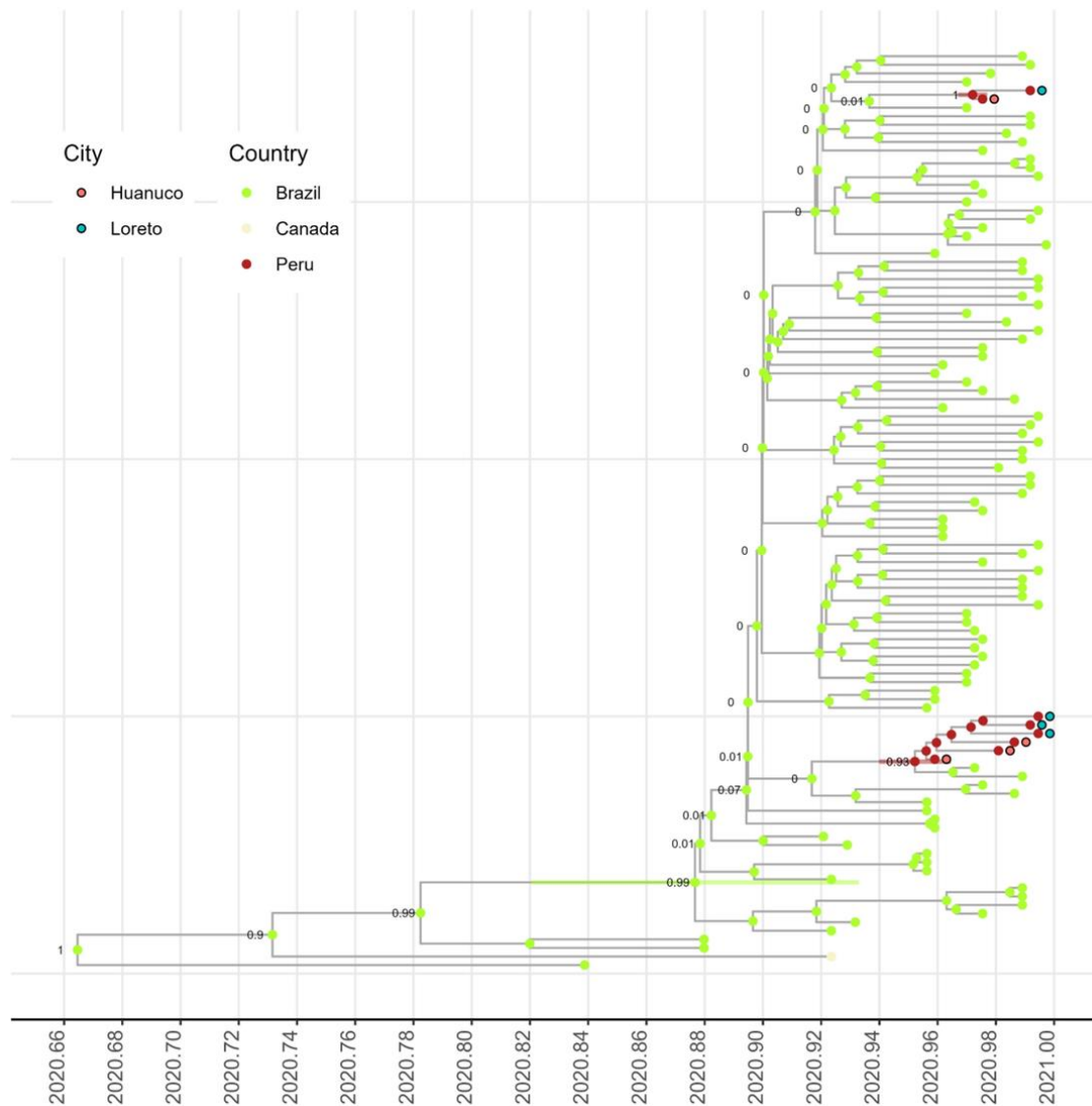


Figure S7. Gamma VOC was introduced from Brazil to Peru. Maximum clade credibility tree of full SARS-CoV-2 genome sequences under a strict clock model showing the introduction of the Gamma VOC to Peru. Green and red bars represent the high posterior density 95 % of the date of the nodes that represent the introduction of the Gamma VOC to Peru. Nodes and the tips are coloured according to its country of origin. Tips from Peru has an additional circle on the right coloured according to the region of collection. Numbers in the nodes represent its posterior probability.

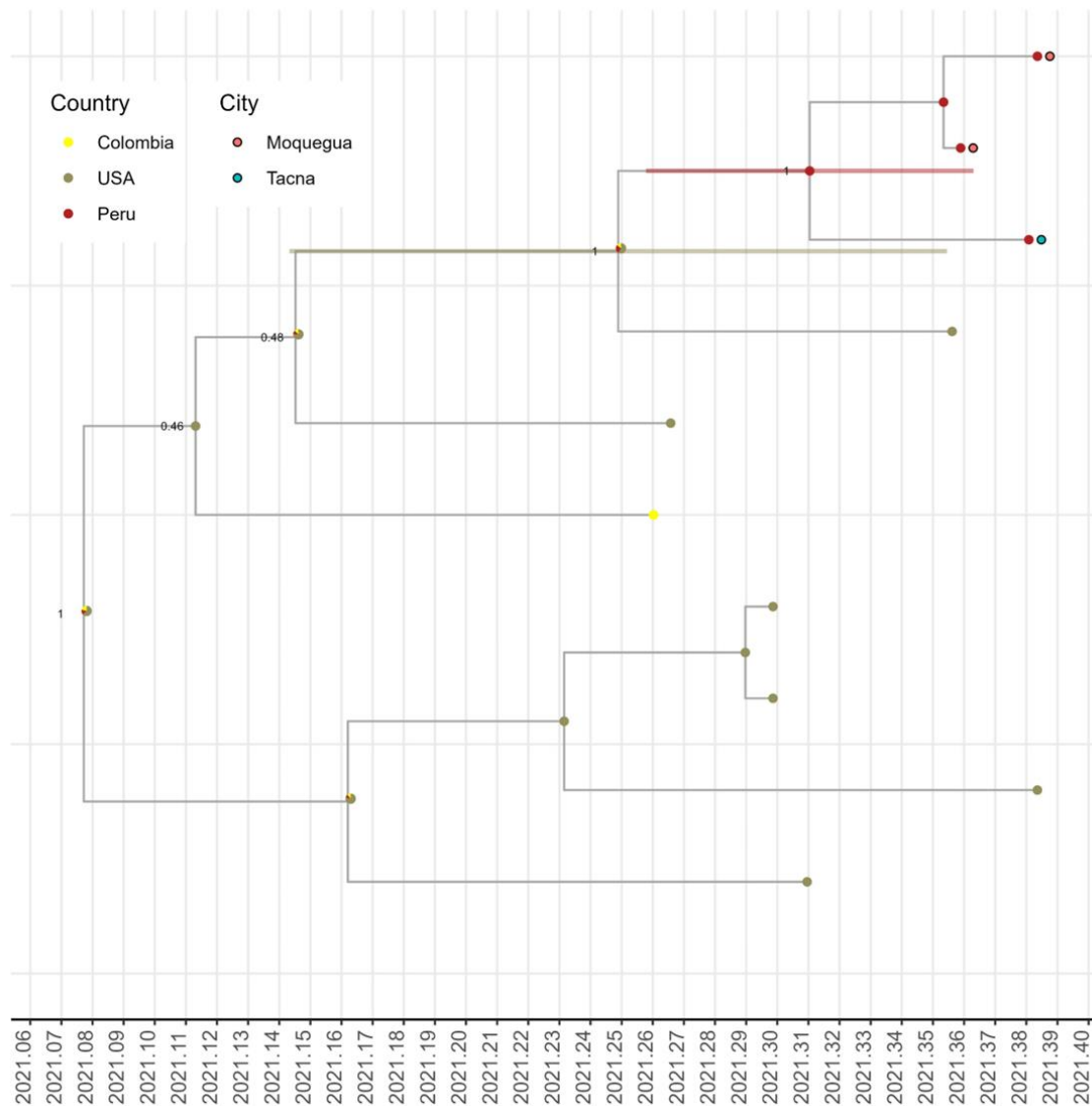


Figure S8. Mu VOC was introduced from the USA to Peru. Maximum clade credibility tree of full SARS-CoV-2 genome sequences under a strict clock model showing the introduction of the Mu VOC to Peru. Brown and red bars represent the high posterior density 95 % of the date of the nodes that represent the introduction of the Mu to Peru. Nodes and the tips are coloured according to its country of origin. Tips from Peru has an additional circle on the right coloured according to the region of collection. Numbers in the nodes represent its posterior probability.

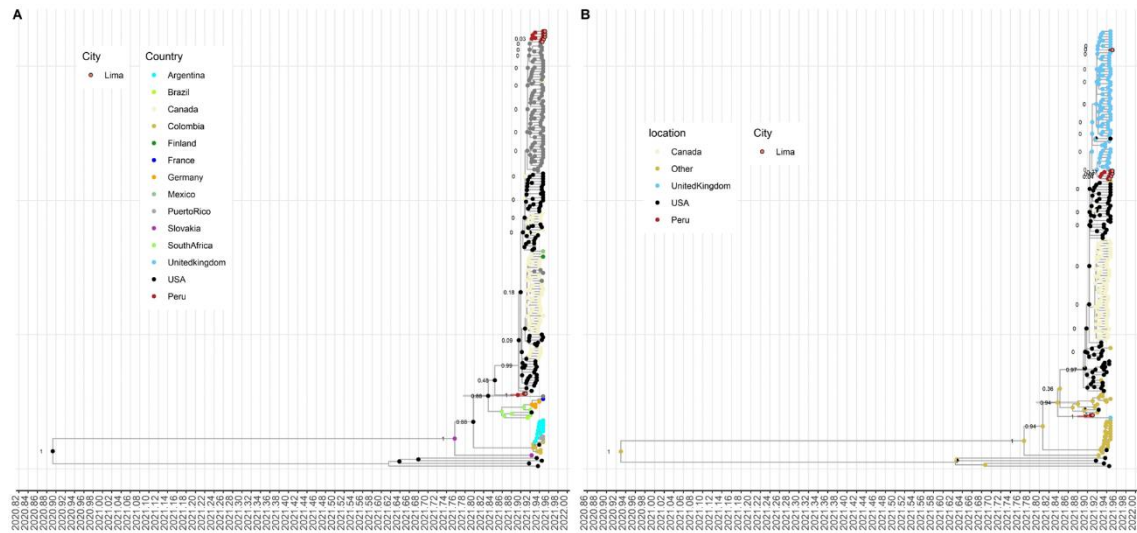


Figure S9. Omicron VOC introduction to Peru. Maximum clade credibility tree of full SARS-CoV-2 genome sequences under a strict clock model showing the introduction of the Omicron VOC to Peru. Grey and red bars represent the high posterior density 95 % of the date of the nodes that represent the introduction of the Omicron to Peru. Nodes and the tips are coloured according to its country of origin (A) or the location, in the case of the analysis with a different grouping of the discrete states (B). Tips from Peru has an additional circle on the right coloured according to the region of collection. Numbers in the nodes represent its posterior probability.

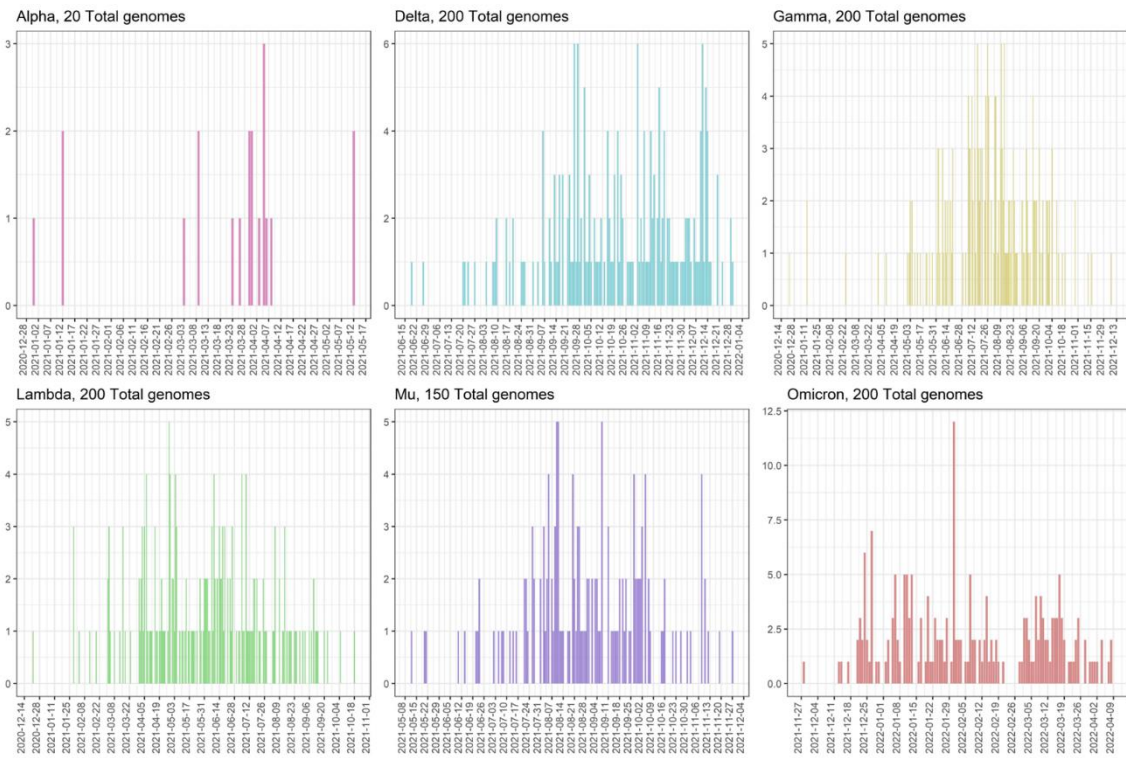


Figure S10. Sampling to estimate the trends of the Effective reproductive number or the Effective population size of the VOCs in Peru. A random sample of the available genomes was considered for this analysis.

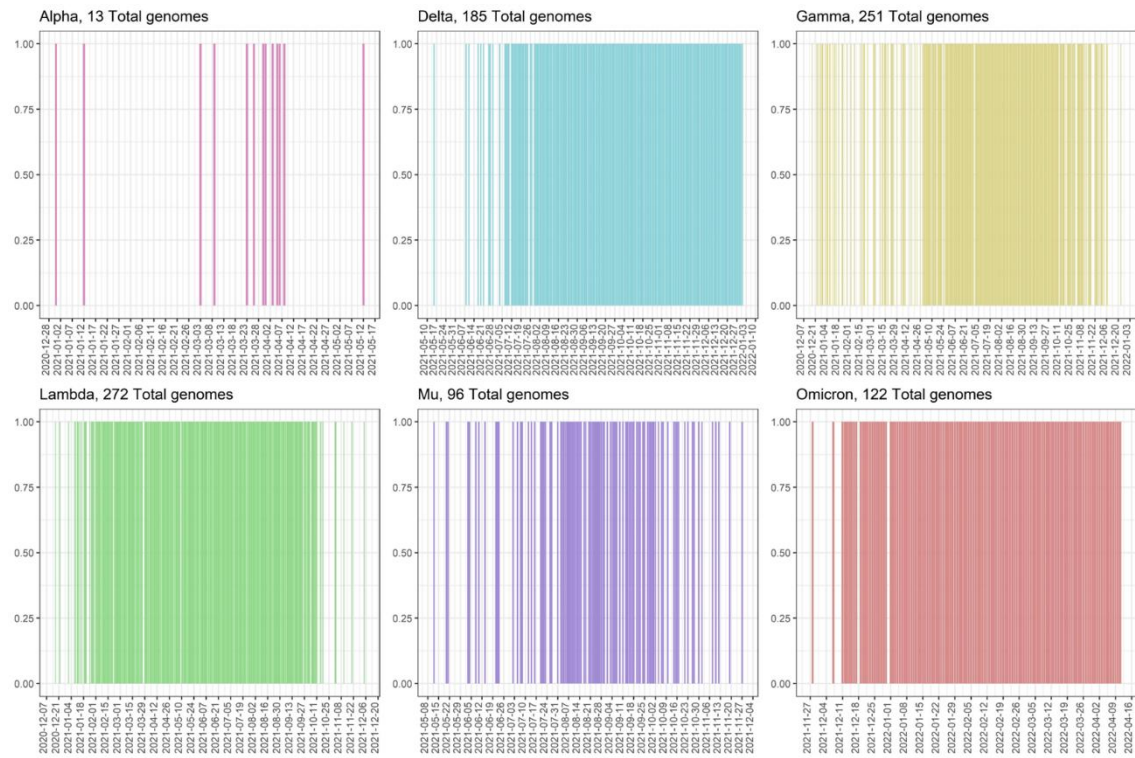


Figure S11. Sampling to estimate the trends of the Effective reproductive number or the Effective population size of the VOCs in Peru. One genome of each date available was randomly sampled to construct this sample.

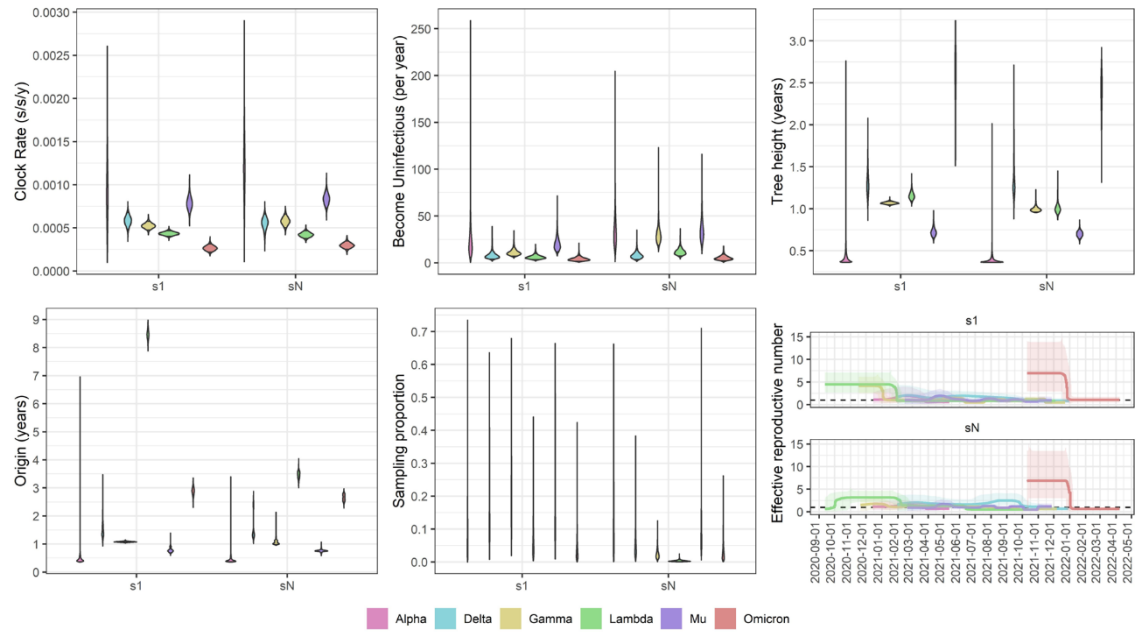


Figure S12. Comparison of the parameters estimated for the Birth-Death skyline model using the two different sampling approaches. sN (a random sample) and s1 (one genome of each date available).

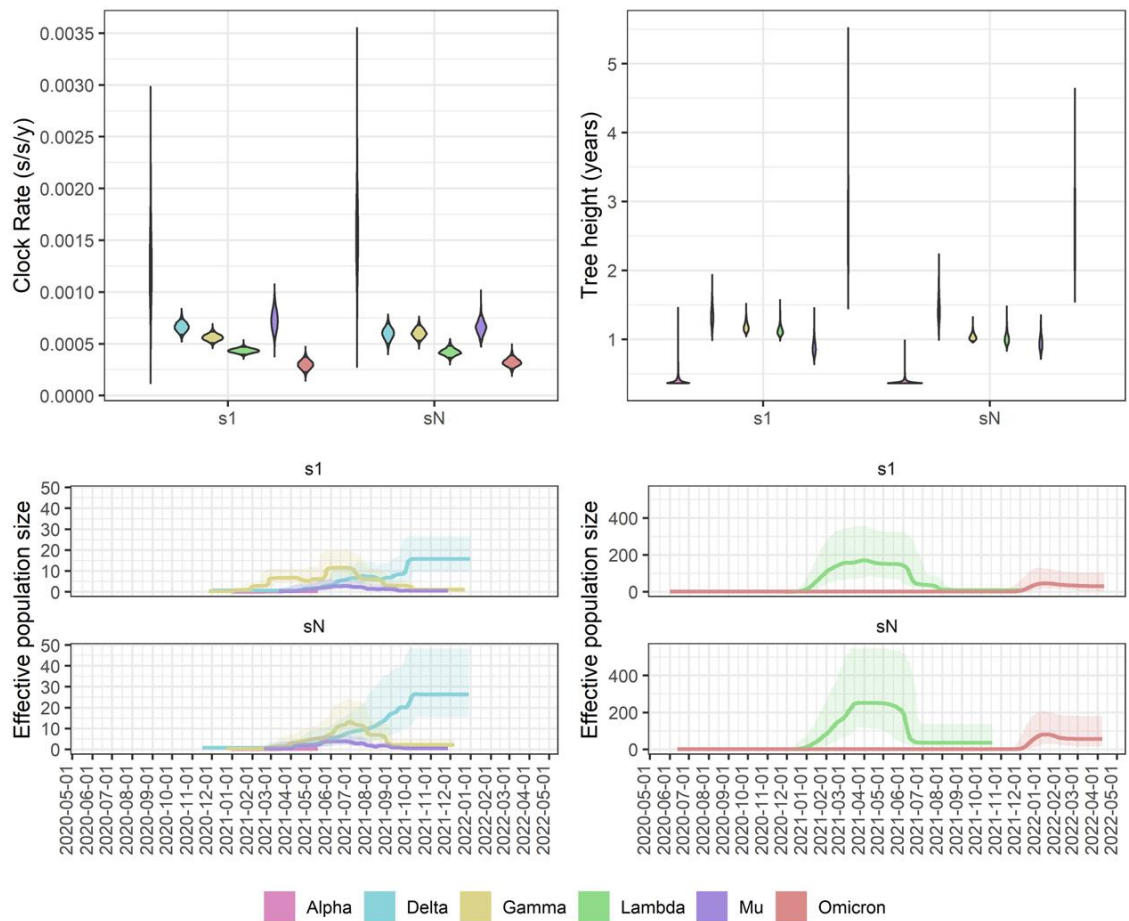


Figure S13. Comparison of the parameters estimated for the BICEPS model using the two different sampling approaches. sN (a random sample) and s1 (one genome of each date available).

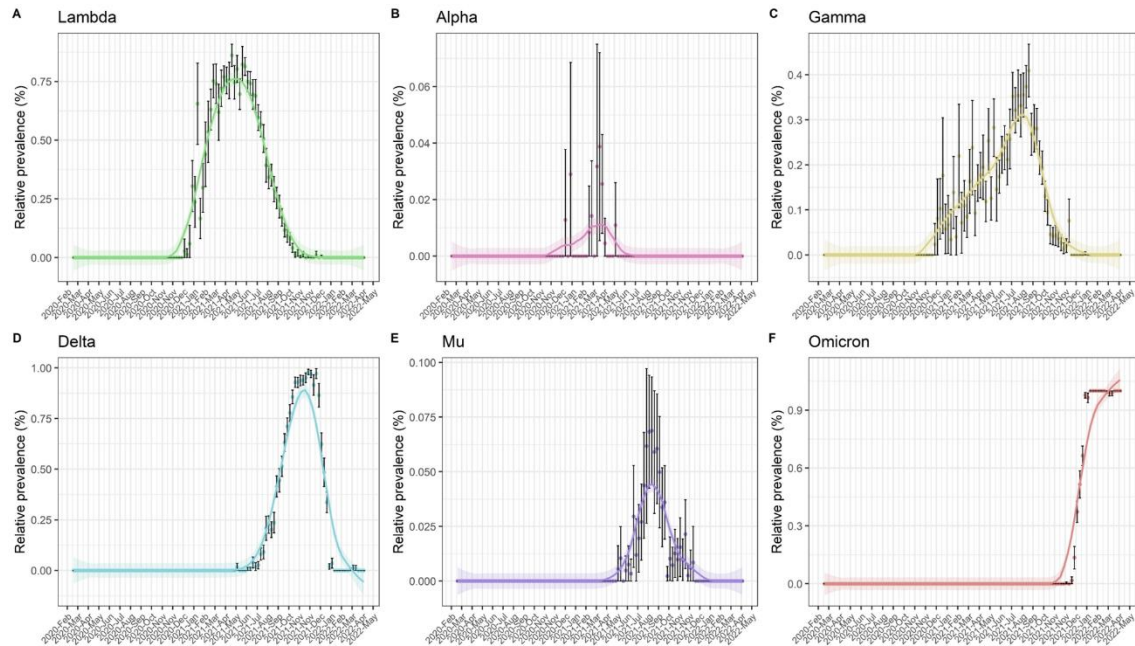


Figure S14. Adjusted relative prevalence of each variant. Each point represents the calculated relative prevalence in each week and the error bars represent the 95-confidence interval of the relative prevalence. A local polynomial regression fitting was done to smooth the relative prevalence by week.

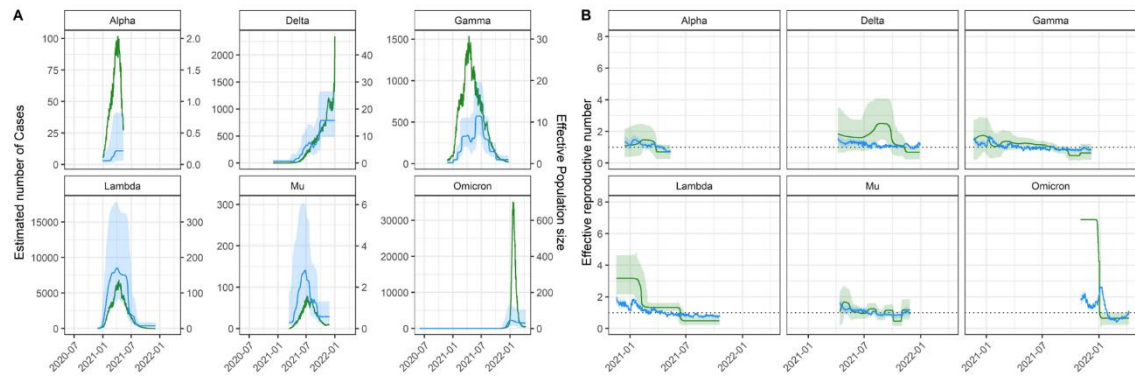


Figure S15. Estimated number of cases and effective population size (N_e) and Effective reproductive number (R_t) calculated from genomes by Bayesian phylogenetics or from the number of cases follow similar trends. A) the number of estimated cases by time is shown together with the estimated N_e from Bayesian phylogenetics. B) Similar trends can be observed when estimations of the R_t from the cases or from the genomes using Bayesian phylogenetics are plotted together.

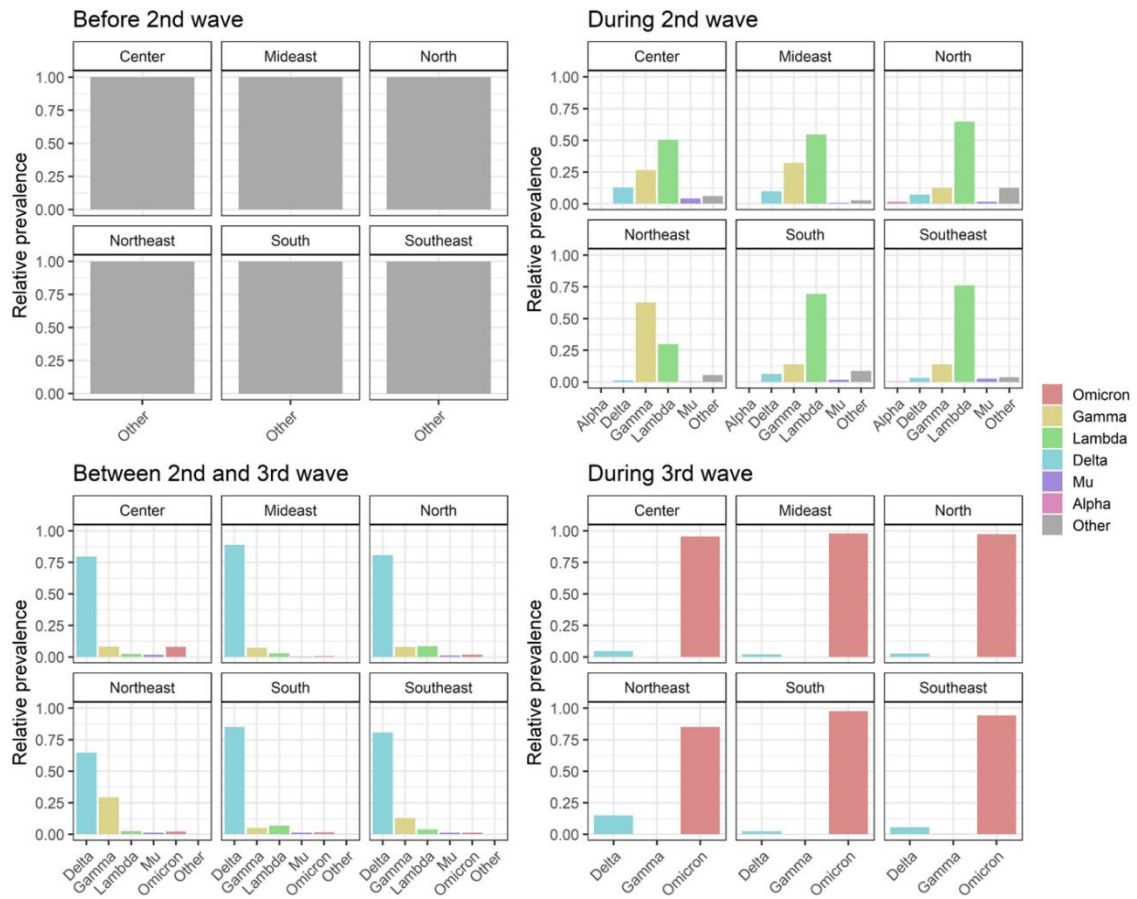


Figure S16. Distribution of VOCs by region in four different stages of the pandemic: Before 2nd wave (upper-left), during 2nd wave (upper-right), before 2nd and 3rd wave (lower-left), and during 3rd wave (lower-right).

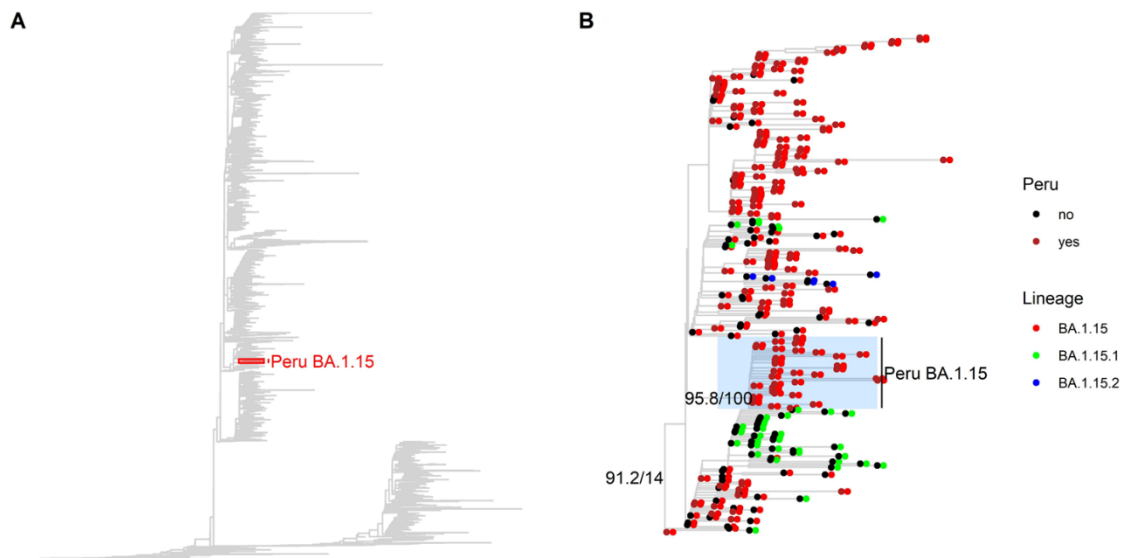


Figure S17. Maximum likelihood analysis of Omicron lineages to identify well-supported Peruvian lineages. A) Maximum likelihood tree of all the Peruvian Omicron genomes plus one genome from each lineage and each country available. In red is highlighted the only well-supported Peruvian monophyletic clade found in this maximum likelihood tree. B) A zoom view of the tree in A showing the well-supported Peruvian monophyletic clade highlighted in a blue shadow. Tip points are colored according to the lineage they belong. The relevant nodes are labelled with their SH-aLRT and UFBoot values.

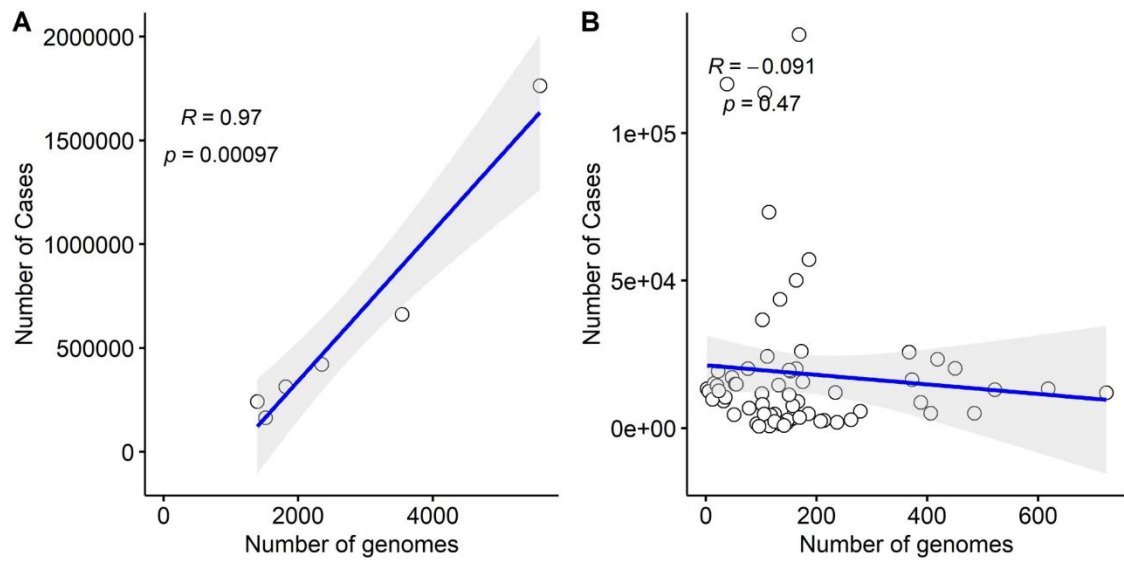


Figure S18. A) Correlation between the number of cases and the number of genomes by region in Peru. B) Absence of correlation between the number of cases and the number of genomes by region by month in Peru.

Table S1. Bayes factor estimation to compare the Strict clock and the Uncorrelated relaxed exponential clock model for the phylogenetic analysis of the origin of Lambda.

Clock Model	Log Marginal Likelihood (Log ML)	Standard deviation	Bayes Factor (Log ML _{SC} – Log ML _{UCEX})
Strict clock (sc)	-50289.806	1,604	-0,825
Uncorrelated Relaxed Exponential (UCEX)	-50288.981	1,595	

Table S2. Classification of Peruvian Cities by Regions.

Cities	Region
Tumbes, Piura, Lambayeque, Cajamarca, Ancash, La Libertad	North
Loreto, Amazonas, San Martin	Northeast
Huánuco, Ucayali, Pasco, Junín	Mideast
Lima, Callao	Center
Ica, Arequipa, Moquegua, Tacna	South
Huancavelica, Ayacucho, Apurímac, Cusco, Puno, Madre de Dios	Southeast