UNIVERSIDAD SIMÓN BOLÍVAR | INSTITUTO DE POSGRADOS

Análisis molecular y filogenético de SARS-CoV-2 presente en Colombia

Sirly Johanna Rivera Roca

Tesis presentada al Programa de
Postgrado en Genética como
requisito parcial para obtener
el grado de Maestre

Tutor:
Moisés Alberto Arquéz Mendoza

Cotutor:
Dayan Lozano Solano

RESUMEN

El SARS-CoV-2 es un virus que causa la enfermedad del COVID19, la cual apareció en Wuhan, provincia de Hubei en China en diciembre de 2019(1), transmitiéndose a través, de la inhalación de góticas respiratorias, desde una persona infectada a otra(15), siendo una importante amenaza para la salud a nivel mundial, debido a las altas tasas de infección y mortalidad que presenta.
Estudios sugieren la existencia de una recombinación entre coronavirus de animales, que dieron lugar a la existencia del nuevo virus SARS-CoV-2 siendo en esta ocasión eficaz al momento de infectar al hombre(26).Pese a su capacidad mutagénica y recombinogenica, es importante conocer sus rasgos genéticos y los procesos evolutivos en las variantes que existieron en Colombia, con relación a las del resto del mundo ya que no existe información, en nuestro país acerca de ello.
Para poder entender los procesos evolutivos, a los que está sometido este nuevo virus, es importante obtener una filogenia, de las diferentes variantes circulantes y analizar esta variabilidad, entre otras características, tanto en nuestro territorio como en el mundo.
A través de un estudio de variantes del SARS-CoV-2 en Colombia, se analizó la capacidad mutagénica, para comprender los factores como la fidelidad de las enzimas que replican los ácidos nucleicos y permite que el virus se adapte a través de la variabilidad del genoma y la integridad de la información genética, sugiriendo, que el virus, está y seguirá evolucionando.

Lo anterior a través de la realización de un análisis bioinformático utilizando los genes de importancia estructural y funcional del virus, llamados proteínas S, M, N, E y unas proteínas ORF accesorias (3a,6, 7a, 8 y 10) siendo el gen de la proteína S, el más utilizado en estudios de producción de vacunas, contra el virus SARS-CoV-2.

Por lo anterior Se escogieron 460 secuencias en total de 10 cepas de interés y preocupación de los 6 continentes del mundo, de las cuales 68 son pertenecientes a Suramérica, siendo las variantes de Colombia de las cepas delta, gamma, Alpha de la base de datos GenBank, del NCBI(41) y plataforma VIPR(42).

Para la realización de estudio filogenético a través de inferencia bayesiana, se utilizaron programas como MAFFT versión 7(43), para alineamiento de secuencias; ModelTest-NG en XSEDE versión 3.3(44), para la selección del mejor modelo de sustitución; MrBayes versión 3.2(44), para la creación de árboles filogenéticos y gráficos FigTree v1.4.4, para observar los árboles y analizarlo. Lo anterior permitió entender, el desarrollo evolutivo, y mantenimiento de las proteínas del SARS-CoV-2, en cada cepa encontrada en Colombia en relación con el mundo, determinando cepas, que poseen mayores cambios evolutivos, verosimilitudes o que se mantienen a través del tiempo conservadas; determinando de esta forma características genómicas y filogenéticas, de los subtipos de SARS-CoV-2 circulantes.

Llevándonos a la conclusión, que existe una colosal divergencia, en las proteínas S y N, además de una evolución más lenta las proteínas Orf3a y Orf8, la conservación de las proteínas M, Orf6, Orf7a, Orf10 con algunas variantes no resultas debido a evoluciones convergentes, y una conservación en mayor medida de la proteína E.

Objetivos:

Objetivo General

Determinar las características genómicas de los subtipos de SARS-CoV-2 que circulan en Colombia.

Objetivos específicos

Objetivo 01

Compilar información acerca de la estructura, funcionalidad y variaciones en las secuencias genómicas de los subtipos de SARS-CoV-2 circulantes en Colombia y el mundo.

Objetivo 02

Analizar los cambios y relaciones evolutivas, de las variantes de monitoreo del SARS-CoV-2 de Colombia en relación con el mundo, utilizando secuencias representativas, que traduzcan a proteínas, por medio de reconstrucción filogenética, a través de topología de inferencia bayesiana.

Antecedentes:

Epidemias por coronavirus

Se sabe que los coronavirus pueden llegar a infectar mamíferos, aves y peces. Estos son virus de RNA monocatenarios positivos(11) nunca fueron vistos como altamente patógenos, hasta que apareció el brote de síndrome respiratorio severo (SRAS) del 2003, sin embargo, hoy día las epidemias como el síndrome respiratorio de oriente medio (MERS) y el actual SARS-CoV-2 se consideran un desafío para la seguridad sanitaria mundial(1).

Los coronavirus poseen un genoma grande, lo que les permite tener más plasticidad para acomodar y modificar los genes(12), además los virus de ARN presenta una frecuencia de mutaciones relativamente alta, lo que aumenta la virulencia y la formación de nuevas especies(13). En este caso podría ser el resultado de la frecuencia de las mutaciones de algunos genes del SARS-CoV-2 en diferentes regiones geográficas y el cambio en las tasas de mortalidad y los síntomas del COVID-19(10).

Sifuentes“et al”, (1) muestra la forma como se identificó el SARS-CoV-2 en sus inicios, revelando que el 31 de diciembre de 2019 los Centros para el Control y la Prevención de Enfermedades (CDC China) realizaron investigaciones, debido a la frecuencia de personas que aparecieron con neumonía, de las cuales se obtuvieron muestras y en su estudio se encontraron más de 20000 lecturas que mostraron un virus que poseía compatibilidad con el linaje del género beta-coronavirus, denominándolo 2019-nCoV. Tiempo después, 30 de enero de 2020 fue declarado por la OMS como COVID-19 y emergencia de salud pública por el creciente número de casos.

Chitranshi (14) “et al” también nos muestra que los coronavirus pueden causar enfermedades agudas y crónicas, del sistema nervioso central y respiratorio, episodios leves de conjuntivitis folicular, afectación del sentido del olfato y la sensibilidad de las papilas gustativas; y en animales, induce síntomas similares a uveítis anterior, retinitis y neuritis óptica, además de lesiones hiperreflectantes, en las células ganglionares y las capas plexiformes internas de la retina, particularmente alrededor de los haces papiloma culares.

Es importante resaltar que el COVID-19, es causado por el SARS-CoV-2, que se transmite a través de la inhalación de góticas respiratorias, desde una persona infectada(15); su periodo de incubación oscila entre 2 y 14 días. con síntomas aproximadamente al quinto día(16). Posee manifestaciones asintomáticas, neumonía leve o severa, con síntomas como fiebre, tos seca, mialgia, artralgias, anosmia, disgeusia, fatiga y dificultad respiratoria; y en menor medida, diarrea, náuseas y vómitos(17).

Estructura del SARS-CoV-2

se conoce que el SARS-CoV-2 es un virus de ARN monocatenario, con una longitud de 29903 nucleótidos, que codifica 12 péptidos, incluidas dos poli proteínas estrechamente relacionadas, Orf1a y Orf1ab que generan 16 proteínas no estructurales (nsps), responsables de la replicación del genoma viral, la transcripción de ARNm subgenómico, y 4 genes estructurales, como la proteína Spike (S), proteína de la nucleocápside (N), proteína de membrana (M), proteína de envoltura(E) y proteínas accesorias Orf1a, Orf1b, Orf3a, Orf6, Orf8, Orf10(17).

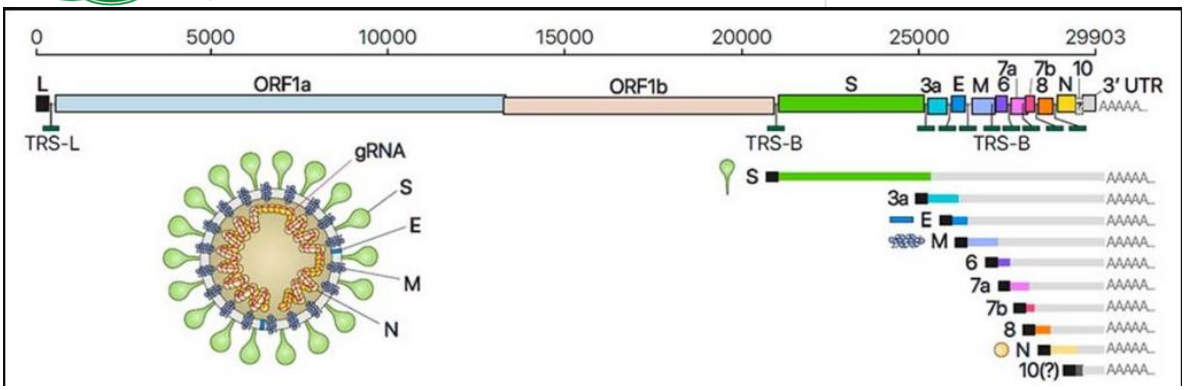


Imagen 1. estructura molecular del Sars-CoV-2. Fuente:(18)

La proteína N de forma helicoidal, recubre al ARN genómico viral y su función es encapsular el material genético, además participa en el ensamblaje, replicación, formación de la envoltura, liberación de la partícula viral, regulación del ciclo celular, e inhibición de respuestas inmunes mediadas por el interferón, por lo que es fundamental en la patogénesis; por lo que es una proteína en estudio, para el desarrollo de antivirales efectivos, en contra de este patógeno respiratorio(19).

La proteína E, es la estructura más externa con funciones estructurales y necesarias, para la maduración y producción del virus. Mientras que la proteína M mantiene la forma y es la responsable del ensamblaje, junto con la proteína N(19).

La proteína ORF8, es una proteína que contiene 121 aminoácidos y es vital para la transmisión y eficacia de la replicación, debido a que actúa regulando negativamente, las moléculas del MHC, importante en la respuesta inmune antiviral del huésped; además de darle a las células infectadas, una posibilidad de subsistencia, por lo que es importante reconocer la variabilidad genética y su evolución, para entender cómo actúan las proteínas, de una mejor forma y así aportar información relevante en estudios de fármacos y antivirales contra el SARS-CoV-2(20).

La proteína ORF6, interviene en la remodelación de la membrana intracelular, que podría influenciar en un incremento de la replicación del virus(21).

ORF10 parece estar involucrado en la inhibición de la inmunidad innata, además, de promover la replicación viral al inducir la mitofagia para degradar MAVS, dándole la facultad al virus de infección y de disminución de respuesta inmunológica(22)

En estudios recientes se demostró, que la proteína ORF7a se une a los monocitos CD14 +, produciendo una disminución de las moléculas HLA-DR / DP / DQ, de modo que disminuye la capacidad de presentación de antígenos, aumentando la capacidad de infección(23)

La proteína accesoria Orf3a, posee funciones de virulencia, inefectividad, actividad del canal iónico, morfogénesis y liberación de virus. Por esta razón debido a las mutaciones que se han producido en el virus es necesario el estudio de esta(24) y de todas las anteriores descritas, para así relacionar, su influencia con la patogenicidad del virus.

La proteína S, se encuentra en la parte externa del virus con una forma particular semejante a una corona, la cual es esencial en el reconocimiento, adhesión y penetración de la célula a infectar(19). Pastrian(17) “et al” en su artículo Bases Genéticas y Moleculares del COVID-19 (SARS-CoV-2). Explica los Mecanismos de Patogénesis y de Respuesta Inmune, dando a conocer que la proteína S se una al receptor de la enzima convertidora de la angiotensina 2 (ACE2) para infectar la célula. Esta proteína se compone de las sub unidades s1 y s2, siendo la s1, la que interacciona con el receptor por medio del dominio de unión del receptor RBD, mientras que la subunidad s2, determina la fusión de la membrana del virus, para que la entrada sea completa; esta debe ser cortada por una enzima proteasa, para ingresar a la célula, vía endocítica y luego liberar su material genético al citoplasma, traduciéndose directamente en las pp1a, y pp1ab, (poliproteínas), que sufrirán proteólisis enzimática para generar las 16 proteínas nsps del complejo RTC, que replica y sintetiza un conjunto de (ARNsg) que codifican, para la producción de las proteínas estructurales y accesorias que conforman al virus, junto con la nucleocápside; y por último estas son ensambladas a nivel del complejo de Golgi, para formar nuevos virus y ser liberadas de la célula infectada.

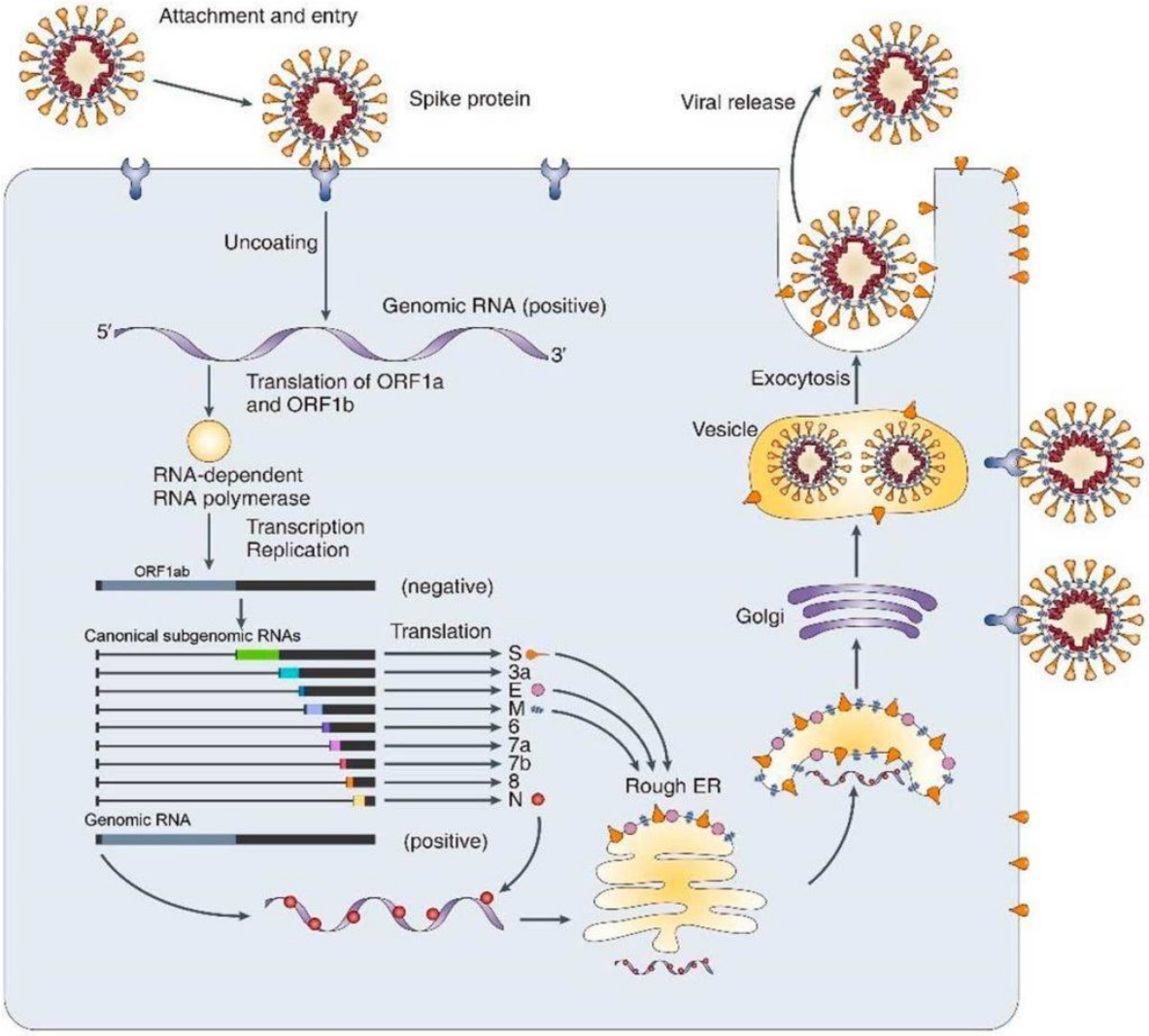


Imagen 1.1 introducción del virus SASR-CoV a la célula. Fuente: (18)

Filogenia de SARS-CoV-2

Los agentes patógenos, causantes de infecciones en especial los virus de RNA, mutan rápidamente, lo que puede llevarlos evolutivamente, a contener una amplia diversidad genética(25). Por lo que es necesario un constante análisis de la dinámica evolutiva de estos organismos, permitiendo hacerles un seguimiento y a su vez, conocer sus patrones de comportamiento.

Es importante resaltar, la divergencia molecular entre SARS-CoV-2 y otros coronavirus, la cual nos revela a través de un análisis filogenético, que la mayoría de ORF en diferentes coronavirus se encuentran conservados, pero difieren en sitios neutrales, demostrando que los linajes L (70%) prevalece más que el S (30%) pero el S está más asociado evolutivamente con los coronavirus de los animales, encontrando una similitud, entre el SARS-CoV-2 más cercana a la rama coronavirus RaTG13, seguido por GD Pangolín SARS-CoV, luego por GX Pangolín SARS-CoV, por el ZC45 y ZXC21, luego por el SARS-CoV humano

dando lugar al Bat coronavirus BM48-31, lo que sugiere la existencia de una recombinación entre los más cercanos, que llevaron a la existencia del nuevo virus SARS-CoV-2 (26). A su vez se muestra que el SARS-CoV-2, es una población homogénea, que evoluciona, aparentemente del RaTG13 de Bat-CoV, diversificado de un ancestro en común evolucionado del pangolín-CoV, siendo los murciélagos reservorios de virus, responsables de contagios zoonóticos del SARS-CoV-2 y el SARS-CoV, dando lugar a coronavirus que evolucionaron a través del tiempo hasta el día de hoy, siendo en esta ocasión eficaces al momento de entrar e infectar a humanos.

Lo anterior, se corrobora en un estudio que analiza la mutabilidad evolutiva del SARS-CoV-2, comparando 9 genomas de origen pangolín-CoV y 3 genomas de origen bat-CoV, encontrando similitudes de 96% con el murciélago-CoV y un 85.98% con el pangolín-CoV, además de 12 mutaciones novedosas repetitivas en genomas virales, de las cuales 4 son únicas de África y 3 son de Suramérica (14), lo que indica que la evolución puede darse debido a únicos cambios en determinadas proteínas, que pueden dar lugar a nuevos genomas de coronavirus; por lo que hace necesario un estudio constante no solo a partir de organismos sino también de las proteínas que lo conforman.

Por otro lado, se conoce que los coronavirus son altamente recombinogénicos y un rasgo ancestral compartido con el COV del murciélago, es la unión al receptor, aunque también se sabe, que es difícil inferir historias evolutivas confiables, por la alta tasa de recombinación que existe entre estos virus, debido a que cada parte del genoma tiene historias diferentes(27). pero si son necesarios para determinar el comportamiento de estos.

Cabe resaltar, que a pesar de las variaciones existentes del SARS-CoV-2, este posee una homología generalmente alta, entre todas las cepas virales del 99,91% en nucleótidos y 99.99 en sus aminoácidos y sus 13 sitios de variabilidad son (1a, 1b, S, 3a, M, 8 y N) sugiriendo mutaciones selectivas(28). Aunque es importante destacar que casi el 80% de las mutaciones recurrentes producen cambios no sinónimos, con respecto a las proteínas, particularmente la proteína S, lo que sugiere una posible adaptación en curso del SARS-CoV-2” (29)(30)(12).

A través de un estudio de variantes del SARS-CoV-2, se analizó la capacidad mutagénica, para comprender los factores como la fidelidad de las enzimas que replican los ácidos nucleicos, el RNA polimerasa que permite que el virus se adapte a través de la variabilidad del genoma y la integridad de la información genética, y este sugiere, que el virus esta y seguirá evolucionando, además sugieren que el SARS-CoV-2 tiene un sesgo de uso de codones relativamente bajo, los cual se determina debido a la selección natural y la presión de mutabilidad, cuyo patrón de uso depende de la ubicación geográfica(31), lo que indica la importancia de conocer la variabilidad y la integridad del genoma de este virus, en relación con su ubicación.

Por lo anterior, en este estudio se utilizan secuencias demarcadas a nivel mundial por continentes, que se identifican a través de variantes y ciudades debido a la relevancia de los factores ambientales, que pueden influir en la evolución de este. Intuyendo que la distribución geo climática y otros factores dieron lugar a

mutaciones descubiertas como únicas y en alta proporción; dando como resultado heterogeneidad a nivel mundial(30), sobre todo si pasan de un país a otro a través del huésped.

Rafiul Islam “et al” revelan diferentes sitios de deleción, en genes del SARS-CoV-2, que codifican las proteínas ORF8 y ORF7a, y un número alto de sustituciones de aminoácidos, además afirma que “los residuos del dominio de unión al receptor (RBD), muestran interacciones cruciales con la enzima convertidora de angiotensina 2 (ACE2) y el anticuerpo neutralizante, de reacción cruzada, se conservaban entre las cepas de virus analizadas” (30). Además, se identificó a través de un estudio filogenético que la variación temporal en la frecuencia de los tipos de coronavirus, fue importante siendo los efectos fundadores que dieron lugar, al subclado A2a, el cual se extendió fácilmente por el mundo, adquiriendo dominancia, en todas las regiones geográficas(32), estos, con una variante no sinónima(D614G), posiblemente la razón por la que puede permanecer e infectar fácilmente las células del huésped(33).

Sureshnee Pillay “et al”, sustenta la importancia de los análisis filogenéticos para rastrear la ruta de transmisión, sobre todo por la evidencia, de que de esta forma, se podría conocer la fuente del brote y proporcionar lecciones para mejorar las estrategias de prevención y control de infecciones”(2), lo cual es necesario debido a su tasa de mutaciones hasta ahora analizadas, lo que podría repercutir en variaciones de alta transmisibilidad o mutagenicidad.

Por otro lado, Leandro N. Jones, cita en su artículo “Cuando la relación entre linajes que coexisten espacialmente es mayor de lo esperado, se dice que, su distribución está estructurada filogenéticamente”(34) y solo hasta determinar qué tan estructurados pueden estar las variantes en relación, una con la otra, se puede determinar, si las diferencias epidemiológicas entre las diferentes regiones se deben o no a diferencias plausibles. Debido a que algunas mutaciones podrían influir en la disminución de la estabilidad de la proteína pico, siendo la mutación R408I relevante, debido a que posee una influencia significativa, sobre el dominio RBD y un efecto de estabilización de la proteína en el genoma de SARS-CoV-2 de diferentes lugares geográficos (35).

Variantes del SARS-CoV-2

Con base en análisis filogenéticos de distintas áreas geográficas, el SARS-CoV-2 se divide en subtipos o cepas. Los dos tipos de SARS-CoV-2 existentes que han sido de interés y preocupación y que se van a utilizar en el análisis de este estudio son: las variantes Mu, Alpha, Kappa, gamma, Épsilon, Beta, Zeta, Eta, iota, Delta a nivel mundial.

Todos los datos registrados hasta ahora, nos arrojan un conocimiento en el ámbito de poblaciones existentes en el mundo, pero a la vez nos indica que las cepas del virus son variables, debido una coyuntura de mutaciones muy uniformes a lo largo de las ramas. Resultados obtenidos a través de análisis de filogenia del SARS-CoV2, identificaron 2 macro haplogrupos principales A y B; el A afecto ampliamente a nivel internacional, mientras que el B se limitó más al continente asiático, con 160 subramas representativas de las originarias en todo el

mundo(32). Dato que podría presentar implicaciones en el diseño de vacunas y diagnóstico del virus.
Actualmente, se han encontrado muchas variantes en el mundo, pero las analizadas en esta investigación fueron catalogadas como de interés y preocupación en algún momento, (Mu, Alpha, Kappa, gamma, Épsilon, Beta, Zeta, Eta, iota, Delta), debido a que presentaron cambios en el genoma que intervinieron, en la transmisibilidad, gravedad de la enfermedad capacidad de escapar del sistema inmune, además de su crecimiento exponencial de los casos detectados o porque además de cumplir con los anteriores criterios, poseían algunas un aumento en la virulencia, cambios en la presentación clínica de la enfermedad, disminución de eficacia de las medidas sociales y de salud pública, siendo estas un riesgo para la salud de la población (36).
La variante gamma P.1 es originaria de Brasil y fue identificada en Japón, en personas que viajaron desde Brasil, lo que contribuyó a su dispersión en el mundo. Esta variante ha incurrido en cambios significativos con respecto a proteínas específicamente, en los genes que codifican la espícula viral, estructura de la superficie del virus importante para la infección y entrada a la célula (37).

Otra variante analizada en este estudio que presenta alta distribución a nivel mundial es la a variante Delta B.1.617.2 que se identificó por primera vez en la India en octubre del 2020 y hoy en día está presente en 92 países incluido Colombia, pero se infiere que en menor medida que en los demás países del mundo, debido a los escasos datos que existe en nuestro territorio. Actualmente variante Delta es catalogada como de interés y preocupación al igual que la variante Beta (variante sudafricana), Alpha (variante británica), así como la Gamma (variante brasileña) cepas introducidas en este estudio.

Por lo anterior, cabe resaltar que gracias a estudios se ha demostrado que vacunas aprobadas por la EMA (Agencia Europea de Medicamentos), como la Pfizer, AstraZeneca por Johnson & Johnson y Moderna, mostraron una respuesta inmune frente a la variante Delta y otras prevalentes, siempre y cuando se reciban las dosis completas. Por lo que a la fecha inferimos la importancia de la vacunación, para frenar la evolución del virus y a su vez, la aparición de nuevas cepas, con capacidad de virulencia alta en el mundo. Además, es necesario resaltar, que esto se encuentra en un constante monitoreo, siendo este exhaustivo, como medida de prevención(38).
La variante beta B.1.351, apareció en Sudáfrica, en mayo de 2020 se identificó que es una variante que afecta comúnmente a los jóvenes sin antecedentes de enfermedades, además presenta similitudes con la variante Alpha, pero presenta mutaciones adicionales en la proteína de pico y genera preocupaciones, debido a que se cree, que puede desarrollar resistencia a las vacunas(39).
La variante Mu B.1.621, apareció por primera vez en Colombia, en enero y hasta hoy se ha informado que se encuentra en 39 países. Según la OMS La variante posee mutaciones con capacidad de escape inmunológico(36).
La variante Alpha B.1.1.7, fue encontrada por primera vez en el Reino Unido, en septiembre de 2020, fue primera cepa catalogada como variante de preocupación,

un 43% y 90% más contagiosa y causante de un número alto de muertes en el mundo según la universidad de Exeter(39).
La variante Iota B.1.526, apareció en estados unidos a fines de noviembre de 2020, y el linaje del 25% de las secuencias fueron en Nueva York, durante febrero de 2021(40);por su parte la variante Eta B.1.525, aprecio por primera vez en Nigeria y se extendió por Europa, EEUU y Canadá Y la variante kappa linaje B.1.617.1, fue detectada por primera vez en India y se asocia al aumento de transmisibilidad y escape inmunológico según CCAES (Centro de Coordinación de Alertas y Emergencias Sanitarias).
La variante Épsilon y Zeta variantes de interés según la OMS hasta el 2020 también se utilizaron en este estudio. Siendo la primera originaria de EEUU, en marzo de 2020 y la segunda originaria de Brasil en abril de 2020.

Materiales y Métodos

Tipo de estudio

Esta investigación, es de tipo descriptiva observacional, porque se mide la evolución y se describe la filogenia dada en las cepas del virus de SARS-CoV-2 reportadas; además de la variabilidad y las verosimilitudes existentes del virus que transita en la localidad, utilizando hipótesis relevantes en este estudio, sin buscar efecto causal de los resultados obtenidos; es transversal porque se utilizaron datos de secuencias adquiridas en un periodo de tiempo y es retrospectivo porque nos ayuda a formular hipótesis a través del análisis de los resultados adquiridos en relación, con estudios previos de la enfermedad, en este caso SARS-CoV-2.
Método de investigación
Representación, adquisición y alineamiento de secuencias
Las secuencias de nucleótidos que fueron utilizadas en este estudio, se obtuvieron de genomas completos del virus SARS-CoV-2, clasificadas por cepas, todo esto a través de la base de datos GenBank, del NCBI(41) http://www.ncbi.nlm.nih.gov/genbank/., adicionalmente se descargaron proteínas de referencia, de uno de los primeros genomas secuenciados, disponible en la plataforma VIPR (42) https://www.viprbrc.org/brc/home.spg?decorator=corona_ncov, cuyo código de acceso es, MT019529. Todas las secuencias fueron guardadas en formato fasta.
En total fueron recolectadas 460, de las cuales 60, son de Suramérica, 202 de Norteamérica, 72 de Europa, 79 de Asia, 11 de Oceanía, 27 de África y 1 una de las primeras secuencias en Wuhan (China). Luego de la recolección se depuraron los datos y se alinearon los diferentes genomas de las cepas del virus por continentes, utilizando el programa MAFFT versión 7(43), posteriormente se eliminaron los nucleótidos encontrados en los extremos del alineamiento, que no codificaron a la proteína de referencia.

Obtención de modelos evolutivos

Para calcular los modelos evolutivos de los genes que codifican las proteínas, S, M, N, E, Orf3a, Orf6, Orf7a, Orf8, Orf10, se analizaron un total de 9 matrices. Cada matriz contenía un único gen y en cada una se encontraban las especies que existen por continente, con un total de 10 cepas evaluadas como cepas de interés o de preocupación. Para cada matriz se calculó el modelo de sustitución de mejor ajuste, por medio del programa ModelTest-NG en XSEDE (44), disponible en la plataforma CIPRÉS SCIENCE GATEWAY(44) versión 3.3 https://www.phylo.org/portal2/login!input.action y se utilizó el criterio de AIC para la selección de los modelos.

Remodelación y análisis filogenético

Los estudio filogenético se reconstruyo a través de la topología por inferencia bayesiana(45), para ello se convirtieron las matrices alineadas en formato fasta, a formatos NEXUS, para que fuese compatible con la herramienta MrBayes en XSEDE 3.2(44), se simuló una Metrópolis-coupled Márkov chain Monte Carlo (MCMCMC) (46) con un número de 10^7 generaciones y una impresión de pantalla cada 1000 generaciones, se utilizó el modelo evolutivo seleccionado por el programa ModelTest-NG en XSEDE(44) para cada gen. Los árboles de IB resultantes, se evaluaron teniendo en cuenta la *probabilidad posterior*, y los filogramas suministrados por el programa MrBayes 3.2(44). Se observaron y editaron, en el programa de gráficos FigTree v1.4.4, en formato *. tree.

Población y Muestra

Criterio de Inclusión

Se escogieron 460 secuencias en total de 10 cepas de interés y preocupación en los 6 continentes del mundo, de las cuales 68 son pertenecientes a Suramérica, siendo 14 de Colombia de las cepas delta, gamma, Alpha y mu.

Criterio de Exclusión.

Se excluyeron genomas que no se encontraban completos y secuencias con caracteres ambiguos.

Resultados y conclusiones

Organización del genoma

En el genoma del SARS-CoV-2, posee una estructura compuesta por 4 proteínas estructurales, la S (espícula), la E (envoltura), la M (membrana) y la N (nucleocápside) de los cuales el gen S es el que contiene mayor cantidad de nucleótidos; además se analizaron una serie de secuencias o proteínas accesorias(ORF3a, ORF7a ORF8, Orf6, orf10), que fueron tomadas de la base de datos VIPR, cuyo número de acceso en la plataforma GENBANK es MN988668.

Colección de las secuencias génicas

Las secuencias analizadas son en total 460 provenientes de las regiones observadas(imagen 1) de las cepas de interés y preocupación Alpha, Mu, Iota, Eta, Zeta, kappa, Delta, Beta, Épsilon, Gamma y una desconocida, pertenecientes

a 40 países encontrados en los 6 continentes del mundo, información representada en (imagen 2 y 3) previamente manipuladas y escogidas, todas de libre acceso y descargadas a través de la base de datos GenBank y VIPR (47).

Tablas y figuras

imagen 2. Distribución de secuencias en el mundo. Fuente: base de datos propia.

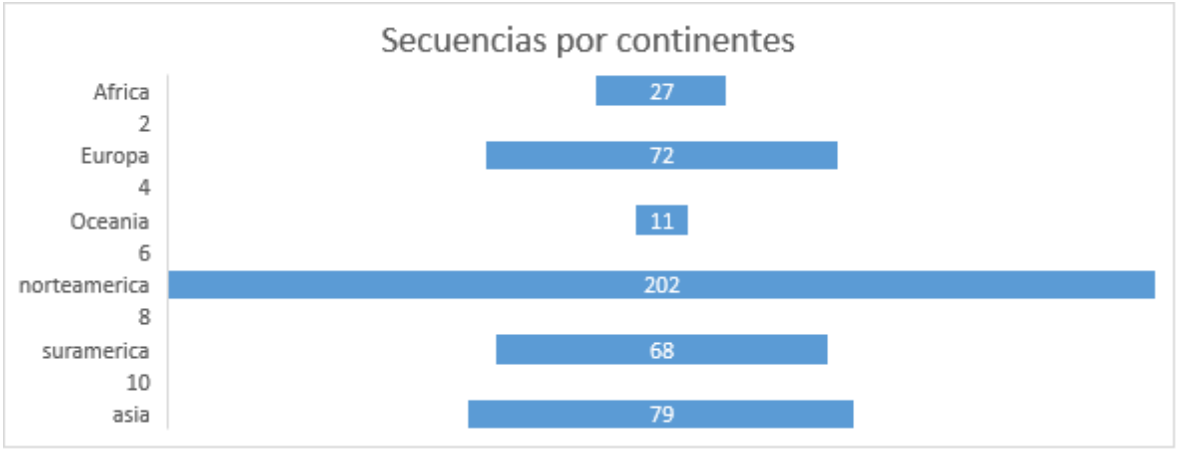


Imagen 3. Distribución de secuencias por continentes. Fuente: base de datos propia.

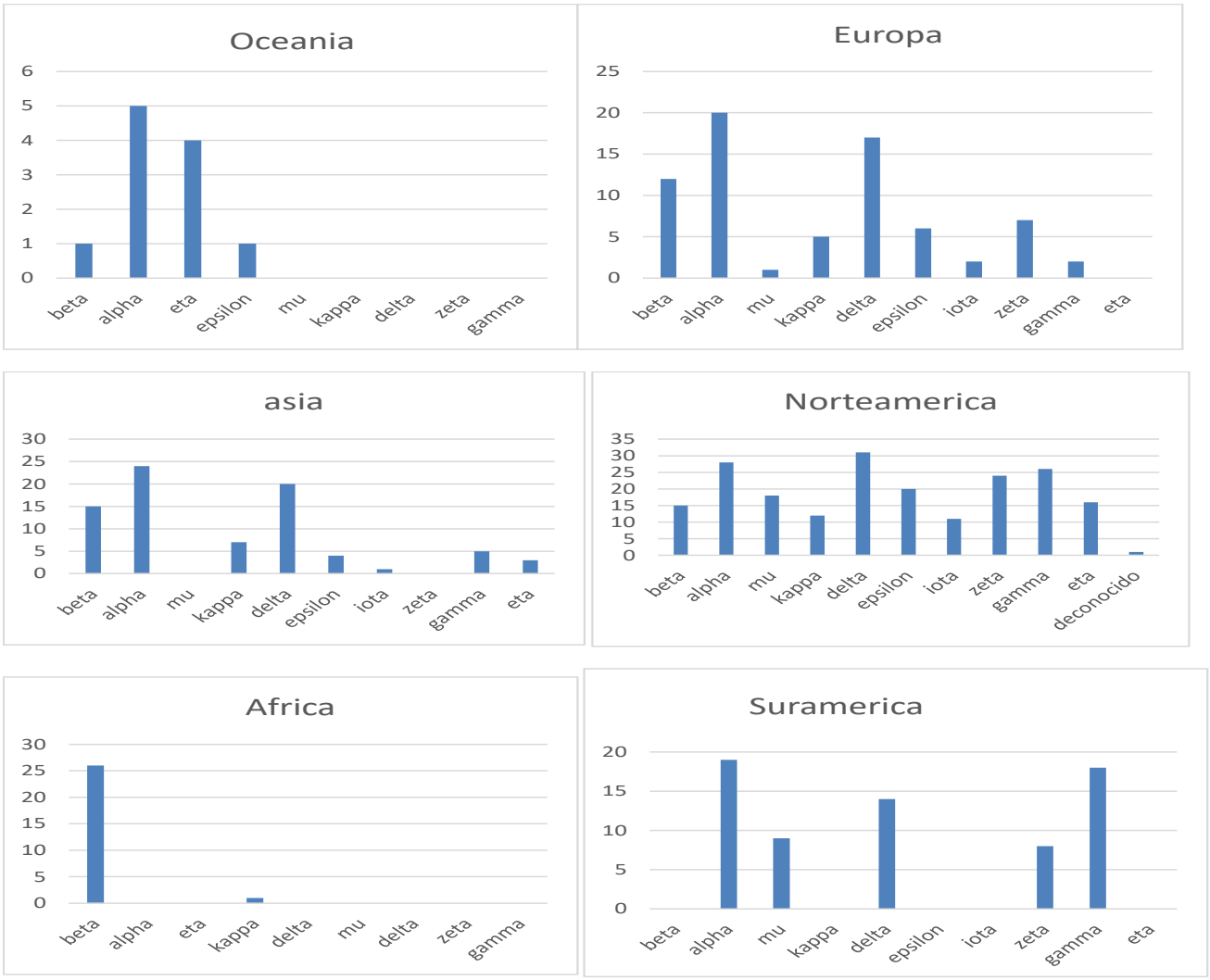


Imagen 4. Distribución de cepas por continentes. Fuente: base de datos propia.

Filogenia de la proteína S de las variantes Iota, Eta, Beta. Épsilon, kappa, Alpha, Mu, Delta, Gamma y Zeta.

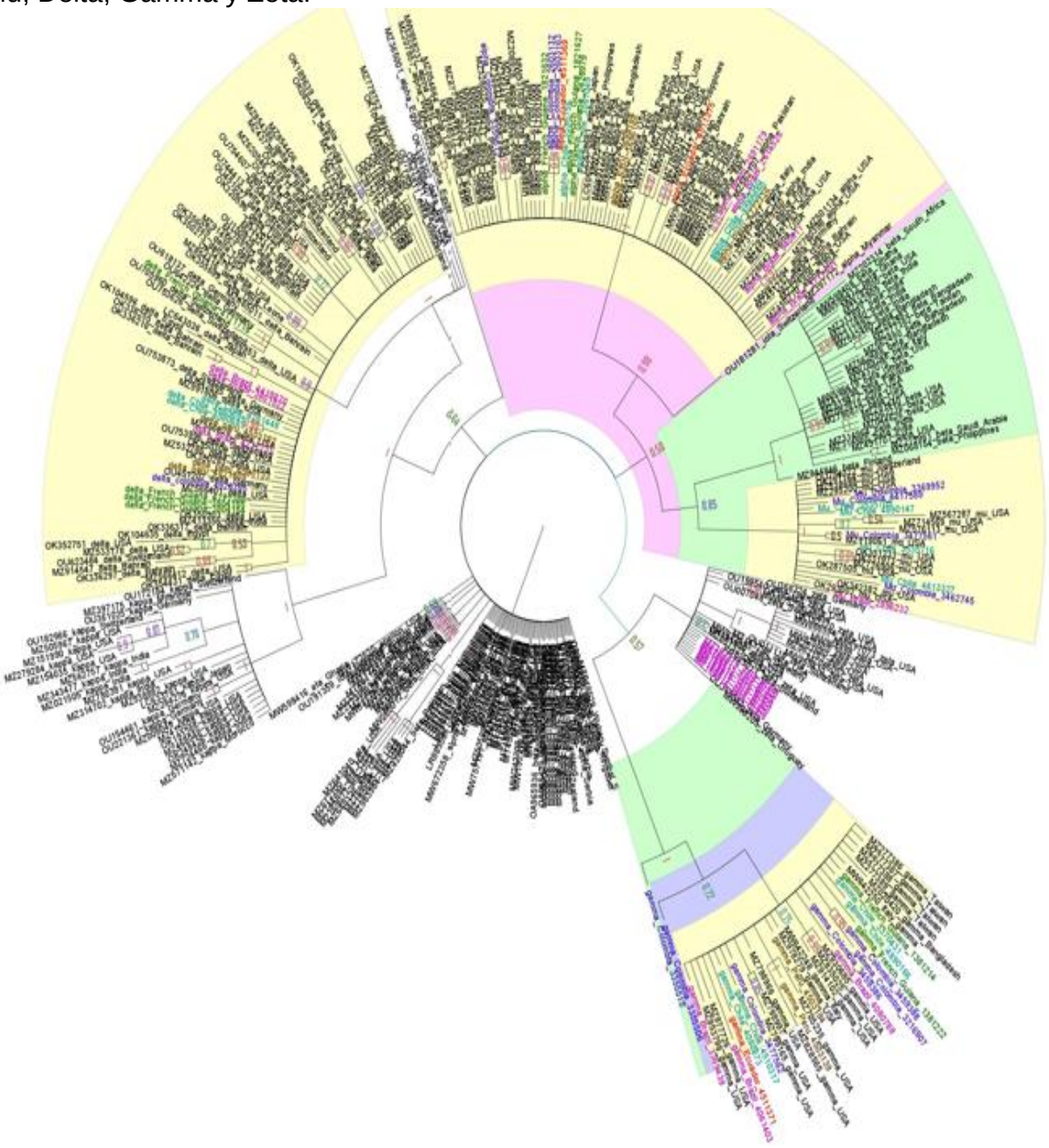


imagen.5Filogenia de la proteína S. Fuente: base de datos propia.
Árbol filogenético obtenido usando el método de Inferencia Bayesiana para el gen S. Los números corresponden a valores de probabilidades posteriores bayesianas. Los nombres en color azul resaltados en color beige corresponden a la ubicación de las cepas de Colombia en relación con las cepas del resto del mundo. Los modelos evolutivos de las posiciones canónicas del gen S fueron, TIM2 y GTR.

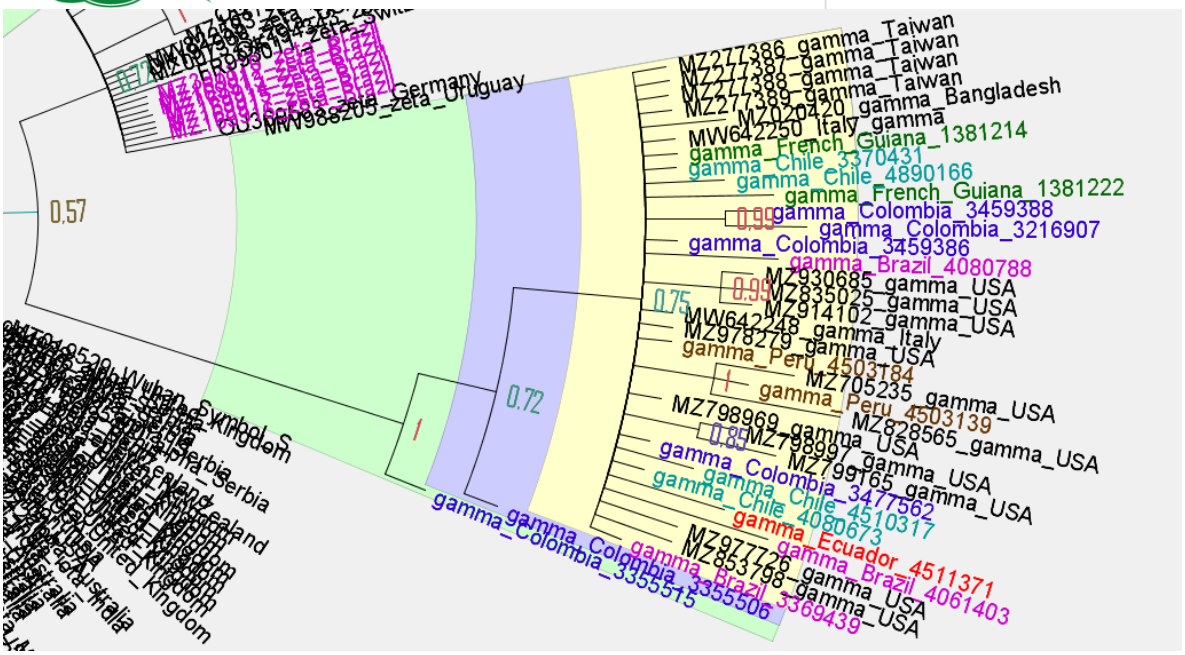


imagen 6 Análisis filogenético de la cepa gamma para la proteína S. Fuente: base de datos propia.

El análisis bayesiano para la proteína S presenta un grupo monofilético, con probabilidades posteriores bayesianas de 0.75, entre las cepas gamma de Colombia, Chile, Ecuador, Brasil, Perú, SA, Italia, Bangladés, Guyana francesa y Taiwán presentan homología entre ellas.
Existen algunas cepas gamma de Colombia con politomías, pero poseen un nodo con una relación posterior de 1 en relación con las demás gamas del mundo.

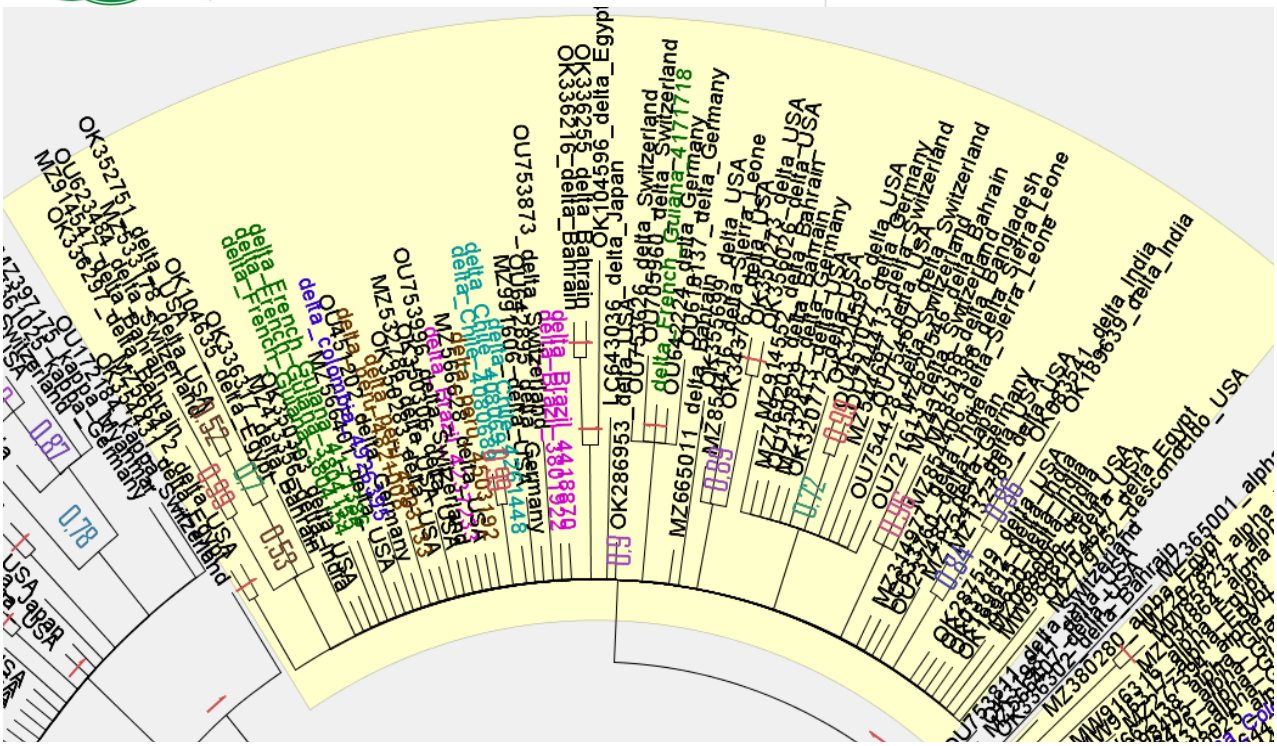


Imagen 7 Análisis filogenético de la cepa delta para la proteína S (base de datos propia).

El análisis bayesiano para la proteína S, presenta un grupo monofilético con probabilidades posteriores bayesianas de 0.9, entre las cepas delta de Colombia, Chile, Perú, Ecuador, Brasil, Suiza, Egipto, Bahréin, Alemania, Japón, India, Perú, USA, Italia Bangladés, Guyana francesa y Taiwán presentando homología entre ellas.

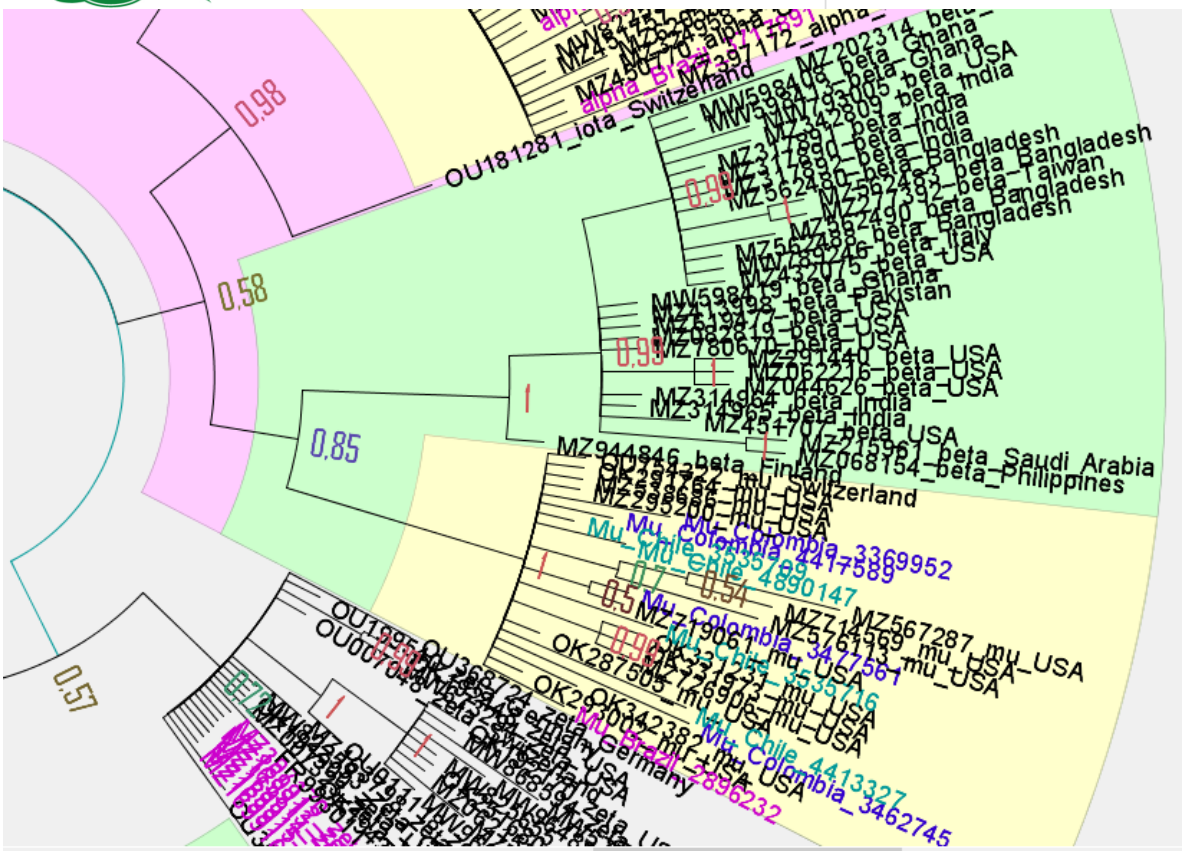


Imagen 8 análisis filogenético de la cepa mu para la proteína S. Fuente: base de datos propia.

El análisis bayesiano para la proteína S, presenta un grupo monofilético con probabilidades posteriores bayesianas de 1 entre las cepas mu de Colombia, Chile, Brasil, Suiza, USA, las cuales presentan homología entre ellas. Las cepas MU de Suramérica y USA; y las variantes beta de países asiáticos y europeos y norteamericanos, proceden de un ancestro en común, con una probabilidad posterior bayesiana de 0.85. Además, existe una relación entre los grupos internos de las variantes mu, beta, iota y Alpha con una probabilidad posterior de 0.58.

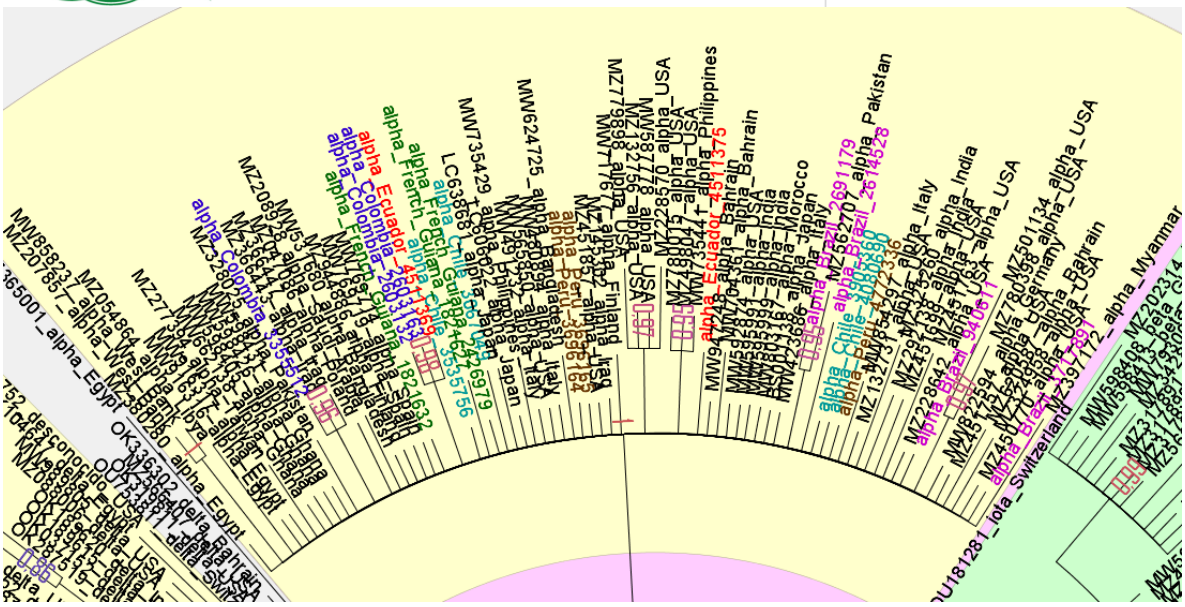


imagen 9 Análisis filogenético de la cepa Alpha para la proteína S. Fuente: base de datos propia.

El análisis bayesiano para la proteína S, presenta un grupo monofilético con probabilidades posteriores bayesianas de 1 entre Las cepas Alpha de Colombia, y las del resto del mundo.

La proteína S, presenta 3 grupos internos y un grupo externo: un grupo interno donde se relacionan las variantes delta, con las variantes Kappa del mundo incluidas las de Colombia, con probabilidades posteriores bayesianas de 0.64. Además, de un grupo parafilético, donde se relacionan los grupos Mu, Alpha, beta donde el nodo que agrupa estas tres variantes posee probabilidad posterior de 0,58 y donde beta está más relacionada a mu, con una probabilidad posterior de 0.85.

También se encuentra, otro grupo interno que relaciona a las variantes zeta y gamma de Colombia y el mundo con una probabilidad posterior de 0.57, la más baja entre los grupos internos, donde las variantes gamas están más relacionadas entre sí con una probabilidad posterior igual a 1.

Existe unas variaciones entre las cepas Alpha debido a que las variantes de Australia, nueva Zelanda, Países Bajos, Reino Unido se encuentran en un grupo externo y no están dentro del grupo de las demás variantes Alpha y están más asociadas a las proteínas de las variantes iota, beta, épsilon y a la proteína S que apareció en Wuhan en sus inicios.

Filogenia de la proteína ORF10 de las variantes Iota, Eta, Beta. Épsilon, kappa, Alpha, Mu, Delta, Gamma y Zeta.

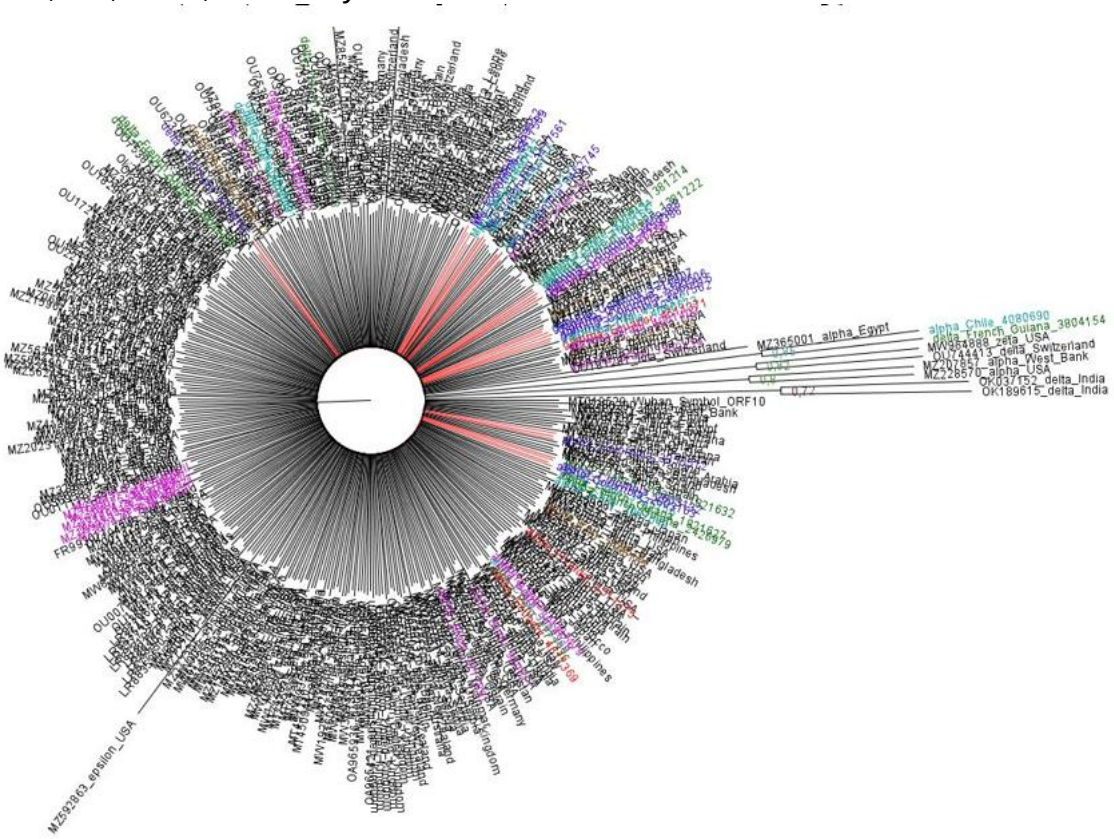


Imagen 10 análisis filogenético para la proteína Orf10. Fuente: base de datos propia.

Árbol filogenético obtenido, usando el método de Inferencia Bayesiana, para el gen ORF10. Los nombres en color azul corresponden a la ubicación de las secuencias de Colombia en relación con las cepas del resto del mundo. Los modelos evolutivos de las posiciones canónicas del gen Orf10 fueron, TIM2, y GTR.

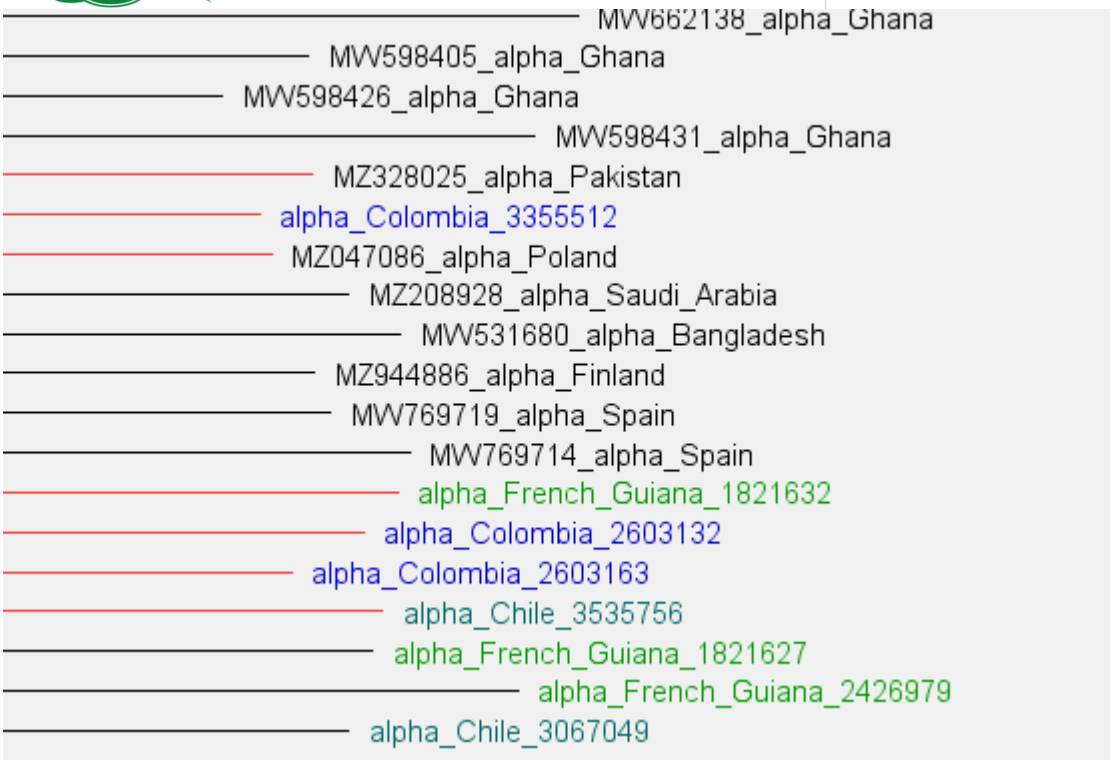


Imagen 11 análisis filogenético de la cepa Alpha para la proteína Orf10. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf10, presenta un grupo monofilético, entre las cepas Alpha de Colombia y el resto del mundo.

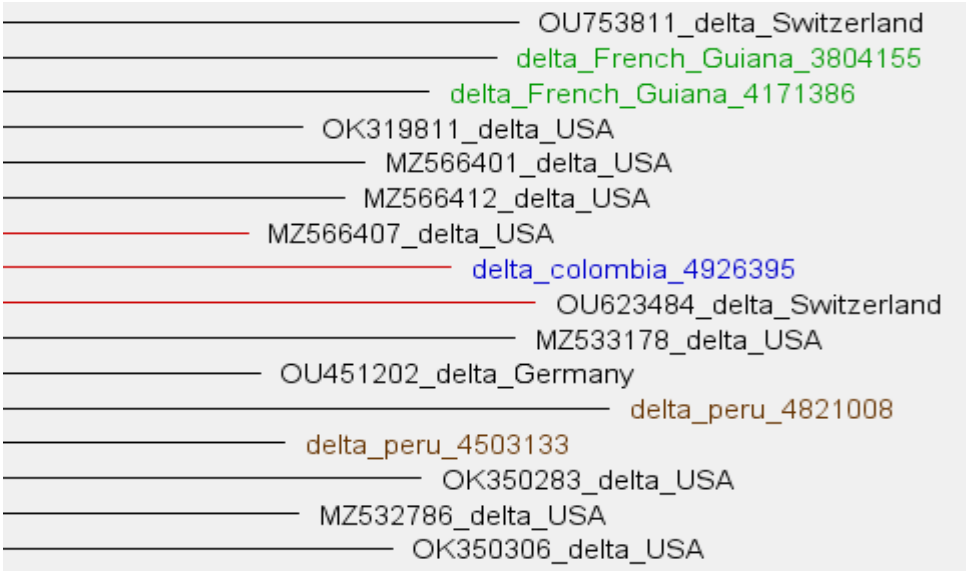


Imagen 12 Análisis filogenético de la cepa delta para la proteína Orf10. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf10, presenta un grupo monofilético, entre las cepas delta de Colombia y el resto del mundo.

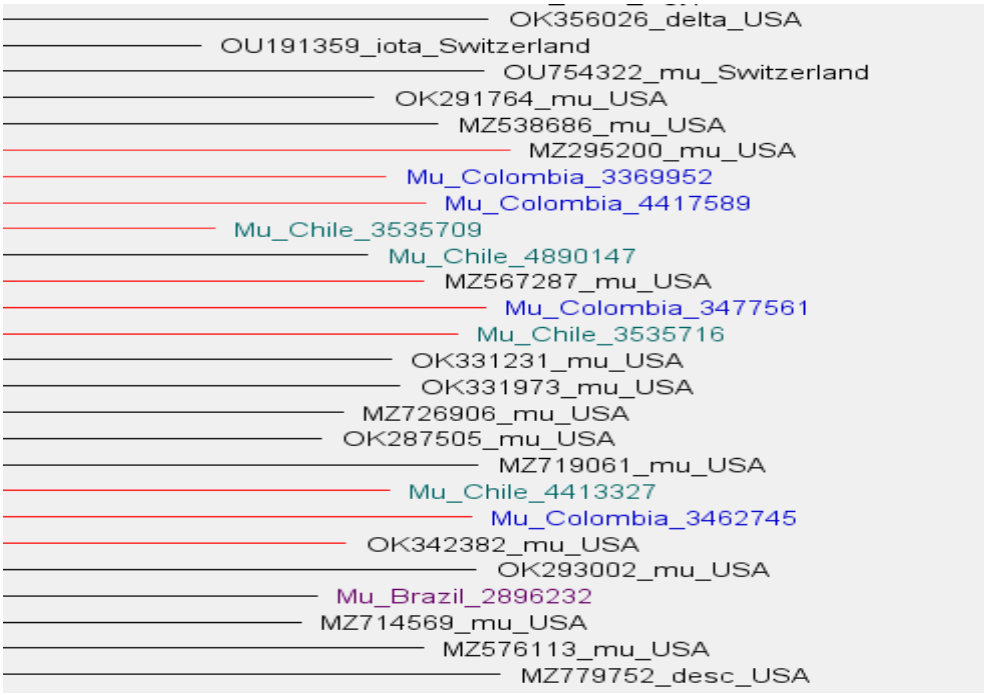


Imagen 13 Análisis filogenético de la cepa Mu para la proteína Orf10. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf10, presenta un grupo monofilético, entre las cepas mu de Colombia y el resto del mundo.

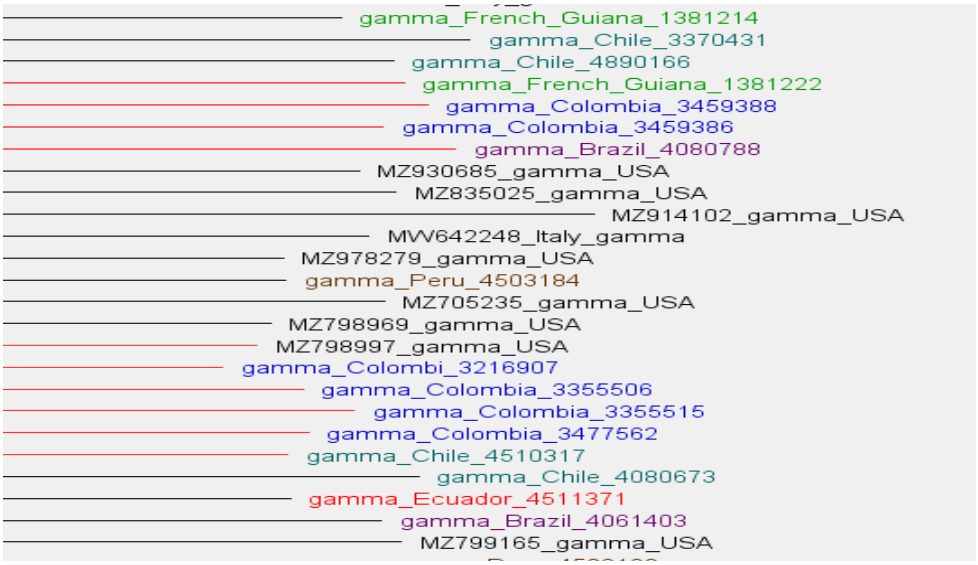


imagen 14 Análisis filogenético de la cepa gamma para la proteína Orf10

El análisis bayesiano para la proteína Orf10, presenta un grupo monofilético, entre las cepas gamma de Colombia y el resto del mundo.

En conclusión la proteína Orf10 del SARS-CoV-2, se encuentra altamente conservada y se ha mantenido por la evolución a pesar de la caracterización de variantes existentes en el mundo, quizás lo que puede relacionarse con una menor importancia de la proteína para la infección del SARS-CoV-2(48).

Filogenia de la proteína ORF8 de las variantes Iota, Eta, Beta. Épsilon, kappa, Alpha, Mu, Delta, Gamma y Zeta.

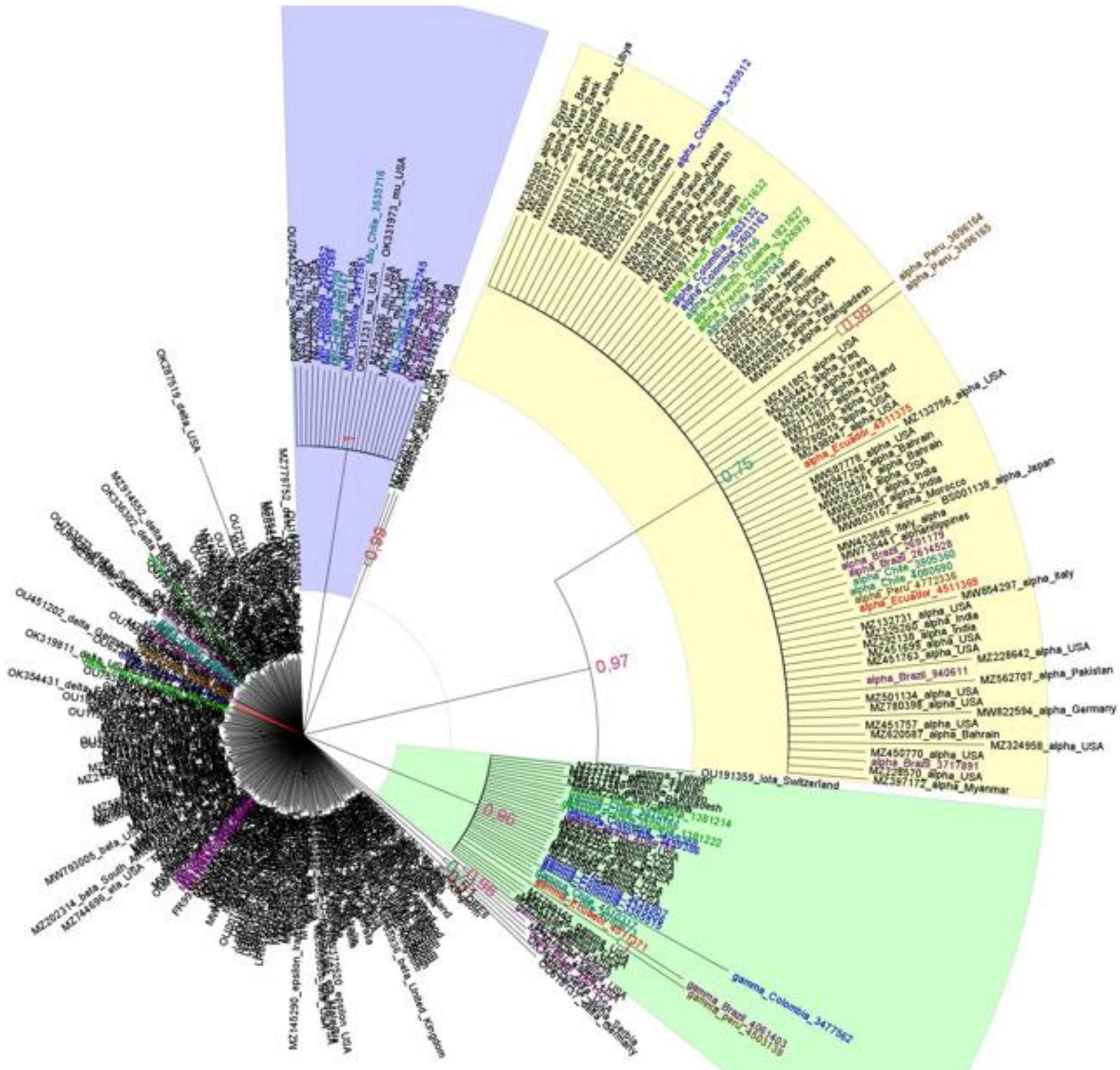


Imagen 15 Análisis filogenético de la cepa delta para la proteína Orf8 (base de datos propia).

Árbol filogenético obtenido, usando el método de Inferencia Bayesiana, para el gen ORF8. Los números corresponden a valores de probabilidades posteriores bayesianas. Los nombres resaltados en color amarillo azul y verde corresponden a la ubicación de las cepas de Colombia en relación con las cepas del resto del mundo.

Los modelos evolutivos de las posiciones canónicas del gen Orf8 fueron, TPM1uf y TVM. Además, se constata lo que dice Rafiul islam acerca de los polimorfismos en algunos sitios del gen, lo que indica, que puede deberse a deleciones y sustituciones en la proteína.

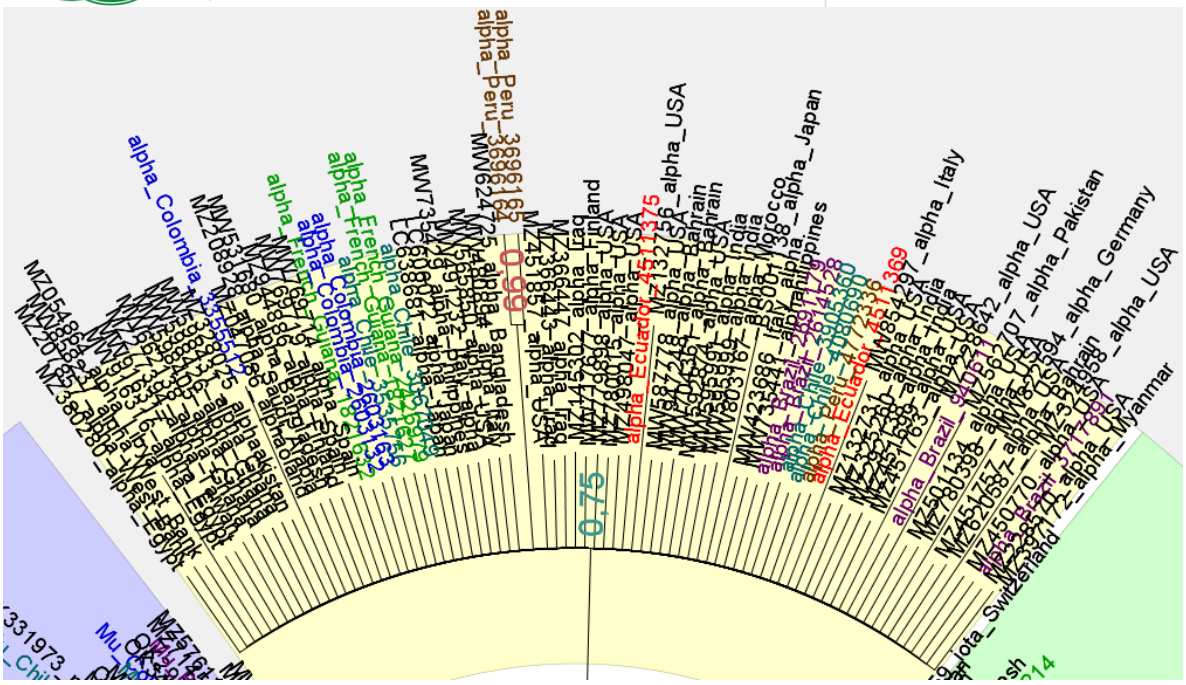


Imagen 16 Análisis filogenético de la cepa Alpha para la proteína Orf8. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf8, presenta un grupo monofilético con probabilidades posteriores bayesianas de 0.75 en las cepas mu de Colombia.
Chile, Brasil, Ecuador, Suiza, Egipto, Bahréin, Alemania, Japón, India, Perú, USA, Italia Bangladés, pilipinas, Pakistán, West Bank, Morrocó, Finlandia Guyana Francesa y Taiwán, Myanmar y USA presentan homología en relación con las secuencias de Colombia.

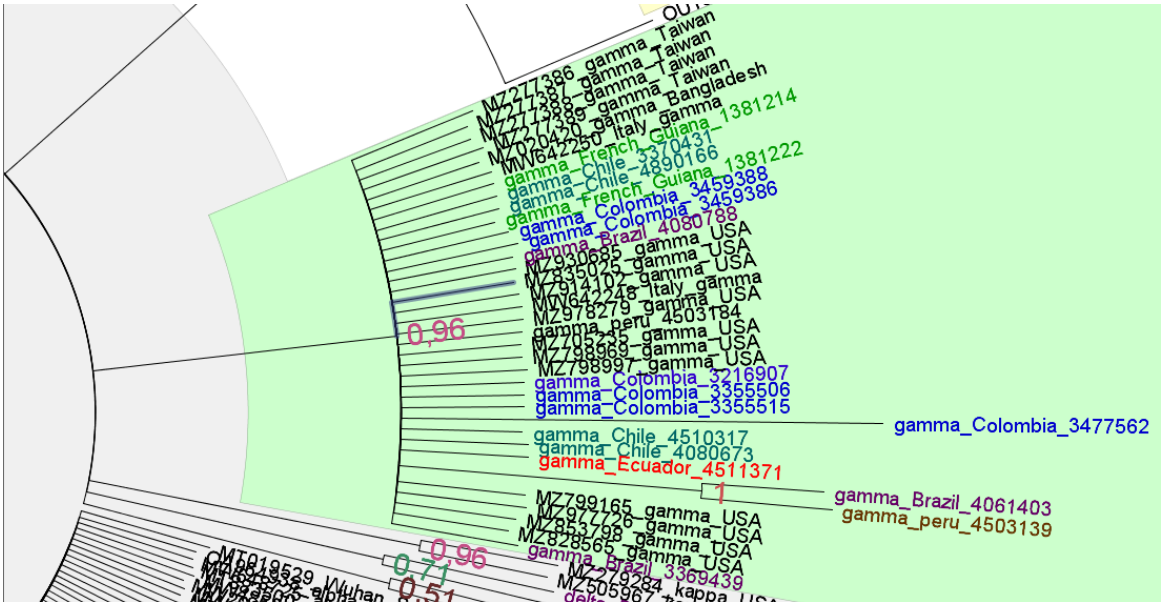


Imagen 17 Análisis filogenético de la cepa gamma para la proteína Orf8. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf8 presenta un grupo monofilético con probabilidades posteriores bayesianas de 0.96 en las cepas gamma.
Chile, Brasil, Ecuador, Suiza, Perú, USA, Italia, Finlandia, Taiwán, y USA presentan homología en relación con las secuencias de Colombia.

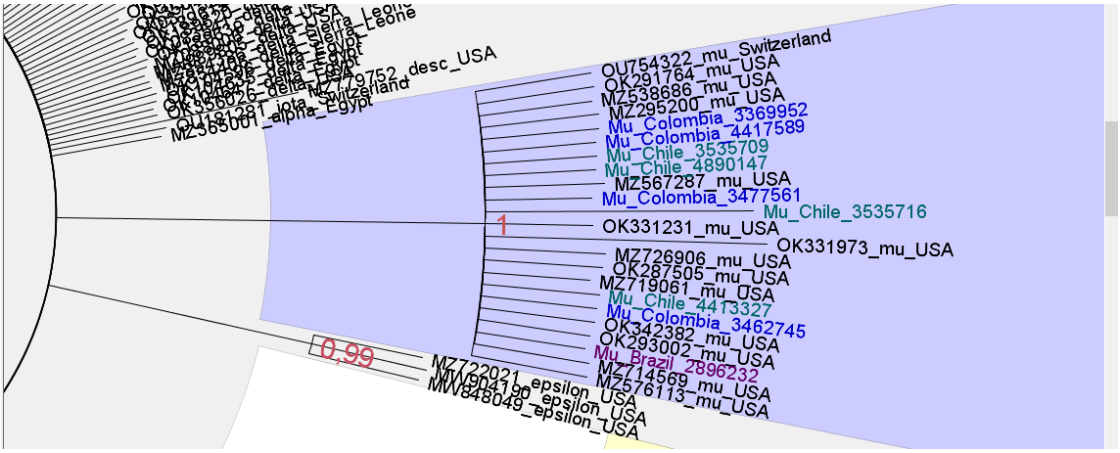


Imagen 18 Análisis filogenético de la cepa Mu para la proteína Orf8 Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf8 presenta un grupo monofilético con probabilidades posteriores bayesianas de 1 en las cepas mu.
Chile, Brasil, Ecuador, Suiza, y USA presentan homología del gen 0rf8, en relación con las secuencias de Colombia.

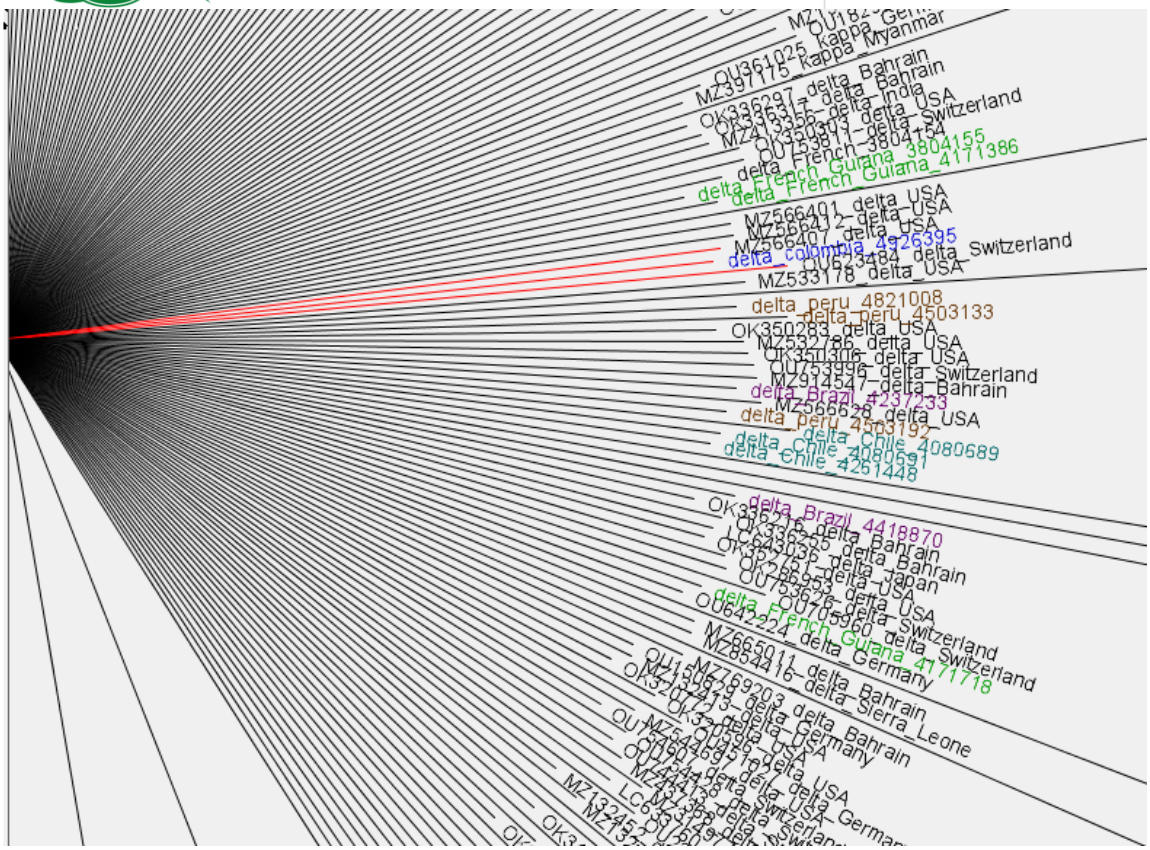


Imagen 19 Análisis filogenético de la cepa gamma para la proteína Orf8 Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf8, presenta un grupo conservado en las cepas delta de Colombia y el resto del mundo.
En conclusión, a través del análisis de inferencia bayesiana, se determinó, que la proteína Orf8, en las diferentes cepas estudiadas poseen 3 grupos internos y un grupo externo; donde un grupo interno, pertenece a las variantes gamma de Colombia y el mundo con una probabilidad posterior de 0.96, otro pertenece a la variante mu de Colombia y el resto del mundo, con una probabilidad posterior a 1 y otro grupo donde se encuentra la variante Alpha, con una probabilidad posterior de 0.97, pero con una variabilidad y conservación en cuanto a la proteína orf8, pertenecientes a las variantes de Australia, Nueva Zelanda, Países Bajos, Reino Unido y Egipto las cuales se encuentran en un grupo externo y no están dentro del grupo de las demás variantes Alpha, encontrándose más asociadas a las proteínas de las variantes delta de Colombia y del mundo y las cepas iota, beta, épsilon y a la primera cepa que apareció en Wuhan en sus inicios.

Filogenia de la proteína ORF7a de las variantes Iota, Eta, Beta. Épsilon, kappa, Alpha, Mu, Delta, Gamma y Zeta.

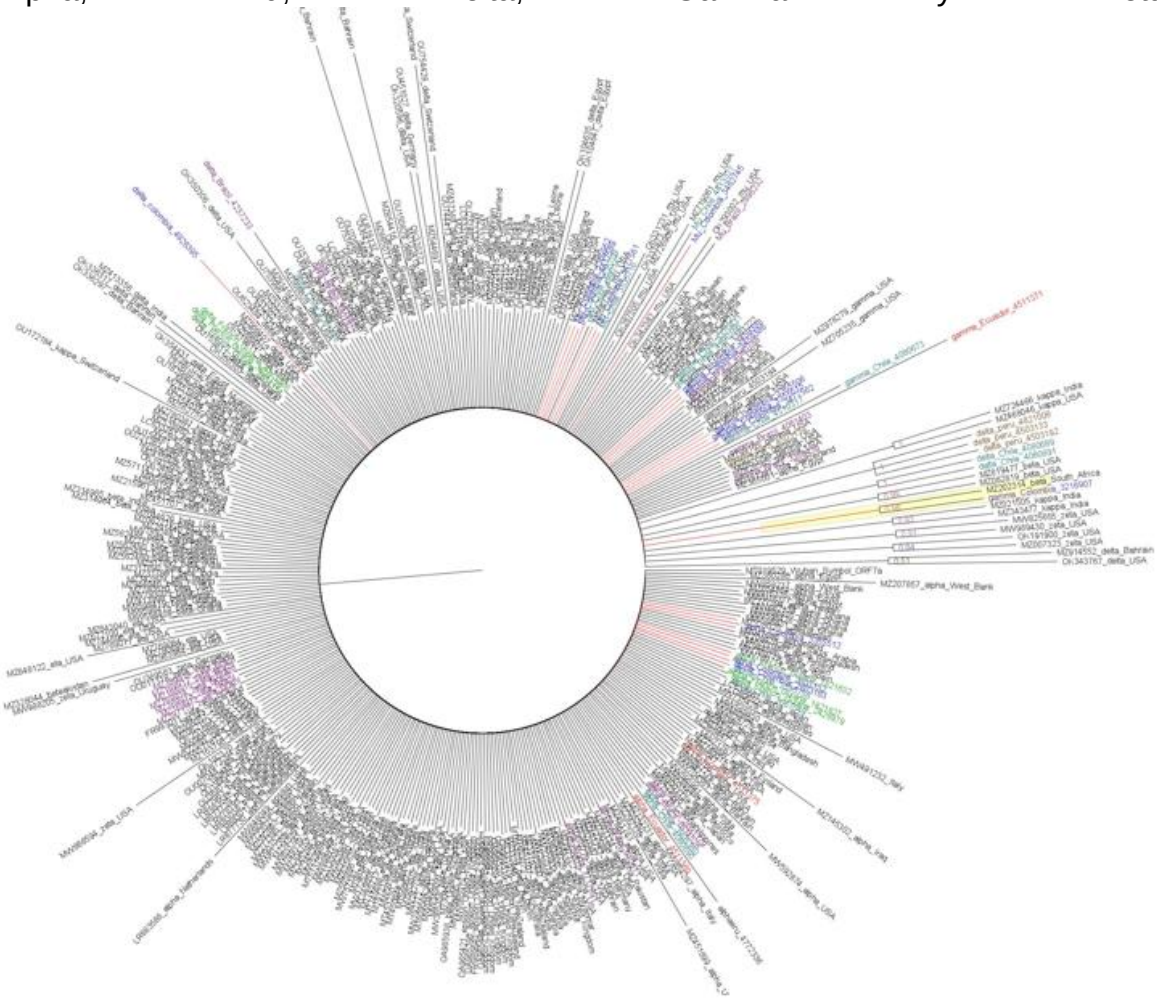


Imagen 20 Análisis filogenético para la proteína Orf7a. Fuente: base de datos propia.

Árbol filogenético obtenido usando el método de Inferencia Bayesiana para el gen ORF7a. Los nombres en color azul corresponden a la ubicación de las cepas de Colombia, en relación con las cepas del resto del mundo. Los modelos evolutivos de las posiciones canónicas del gen Orf7a fueron, TPM1uf y GTR.

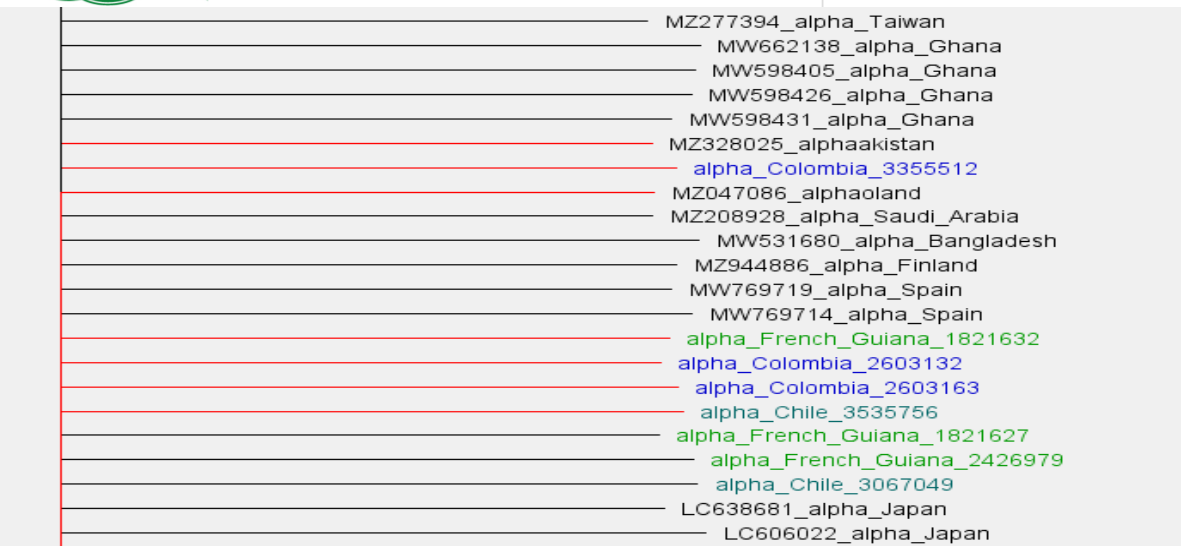


Imagen 21 Análisis filogenético de la cepa Alpha para la proteína Orf7a. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf7a, presenta secuencias biológicas conservadas de las cepas Alpha de Colombia en relación con las del resto del mundo.

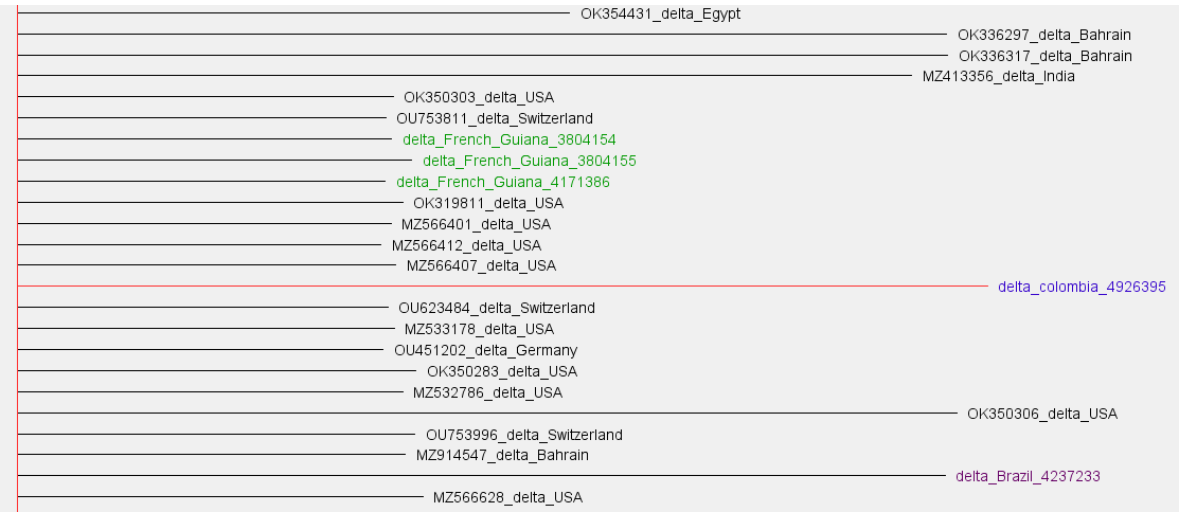


Imagen 22 Análisis filogenético de la cepa delta para la proteína Orf7a. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf7a, presenta secuencias biológicas conservadas de las cepas delta de Colombia en relación con las del resto del mundo.

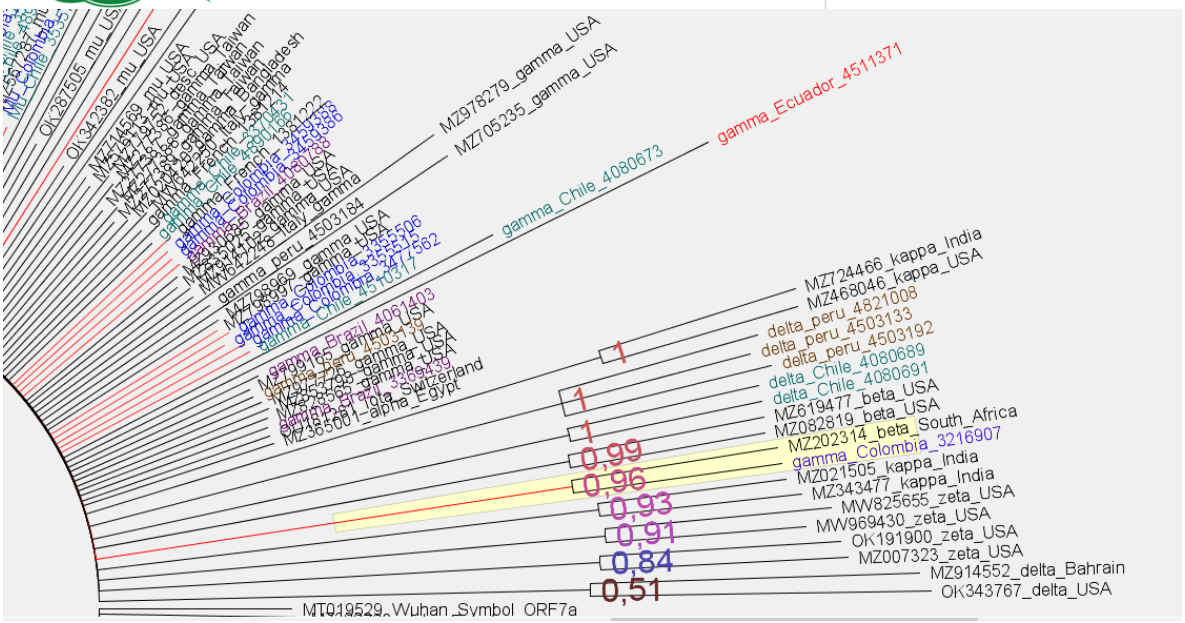


Imagen 23 Análisis filogenético de la cepa gamma para la proteína Orf7a: Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf7a, presenta secuencias biológicas conservadas de las cepas Gamma de Colombia, en relación con el resto del mundo. Todas las cepas gamma, presentan homología en esta proteína en relación con las secuencias de Colombia. Existe una estrecha relación entre las cepas gamma de Colombia y beta de South África con probabilidades posteriores bayesianas de 0.96.

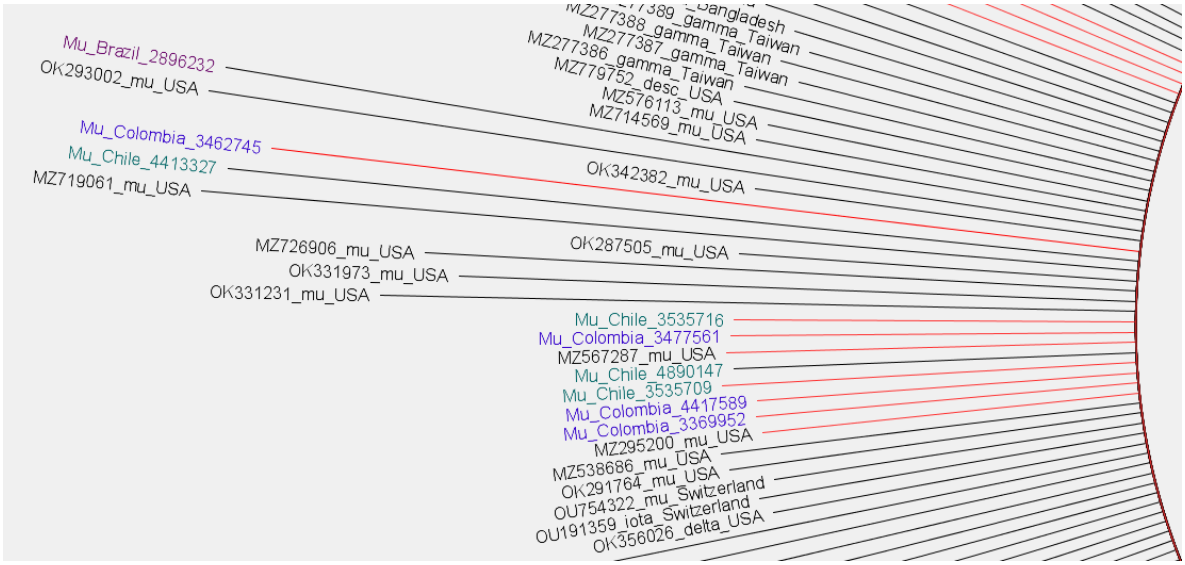


Imagen 24 Análisis filogenético de la cepa Mu para la proteína Orf7a Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf7a, presenta secuencias biológicas conservadas de las cepas Mu de Colombia en relación con las del resto del mundo.

En conclusión, la proteína Orf7a del SARS-CoV-2, se encuentra en mayor medida conservada y se ha mantenido por la evolución a pesar de la caracterización de variantes existentes en el mundo, pero muestra unas bifurcaciones y verosimilitudes entre la variante gama de Colombia y beta South África. De esta proteína se sabe que interacciona con la proteína de transporte ribosómico HEATDR3 y MDN1, reprimiendo el sistema inmunológico, lo que implicaría que cambios importantes en esta proteína, podrían llevar una inhabilitación de sus funciones(49).

Filogenia de la proteína ORF6 de las variantes Iota, Eta, Beta. Épsilon, kappa, Alpha, Mu, Delta, Gamma y Zeta.

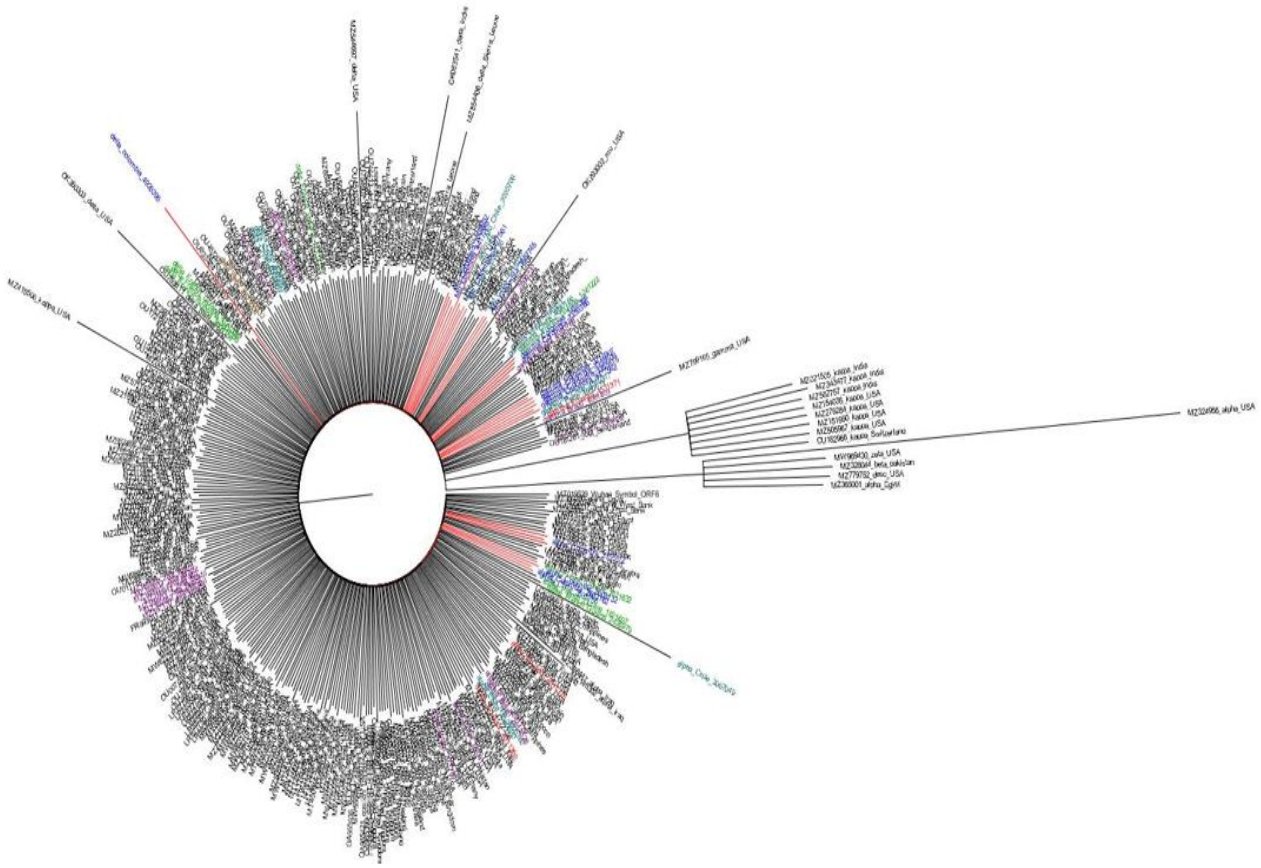


Imagen 25 Análisis filogenético para la proteína Orf6 (base de datos propia).

Árbol filogenético obtenido, usando el método de Inferencia Bayesiana, para el gen ORF6. Los nombres en color azul corresponden a la ubicación de las cepas de Colombia, en relación con las cepas del resto del mundo. Los modelos evolutivos de las posiciones canónicas del gen Orf6 son HKY y TrN.

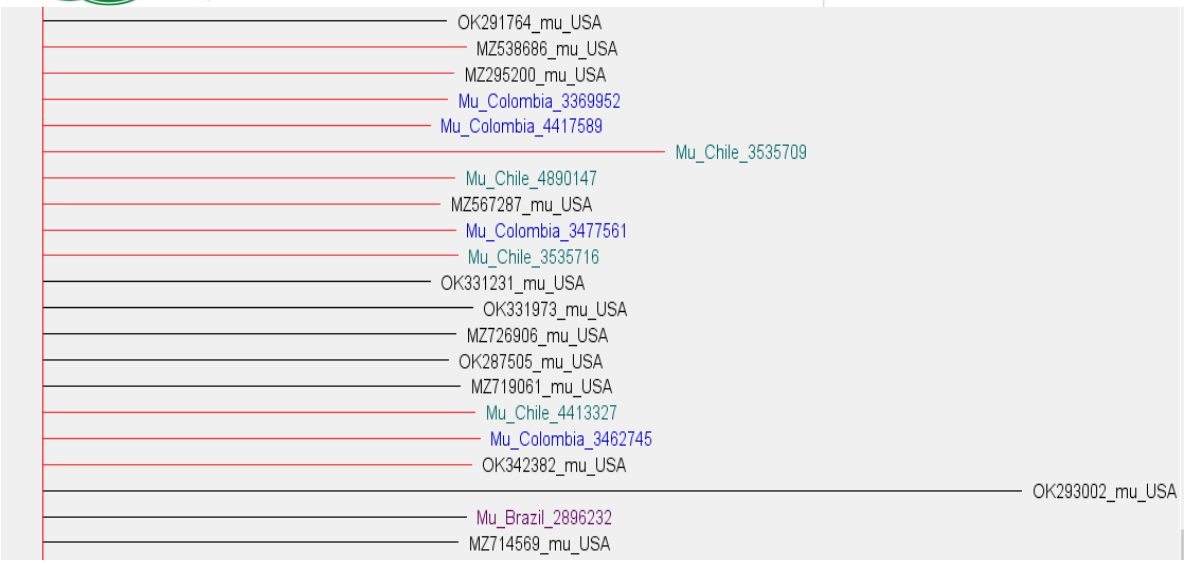


Imagen 26 Análisis filogenético de la cepa Mu para la proteína Orf6. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf6, presenta secuencias biológicas conservadas de las cepas MU de Colombia, en relación con las del resto del mundo.

Imagen 27 Análisis filogenético de la cepa gamma para la proteína Orf6. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf6, presenta secuencias biológicas conservadas de las cepas gamma de Colombia, en relación con las del resto del mundo.

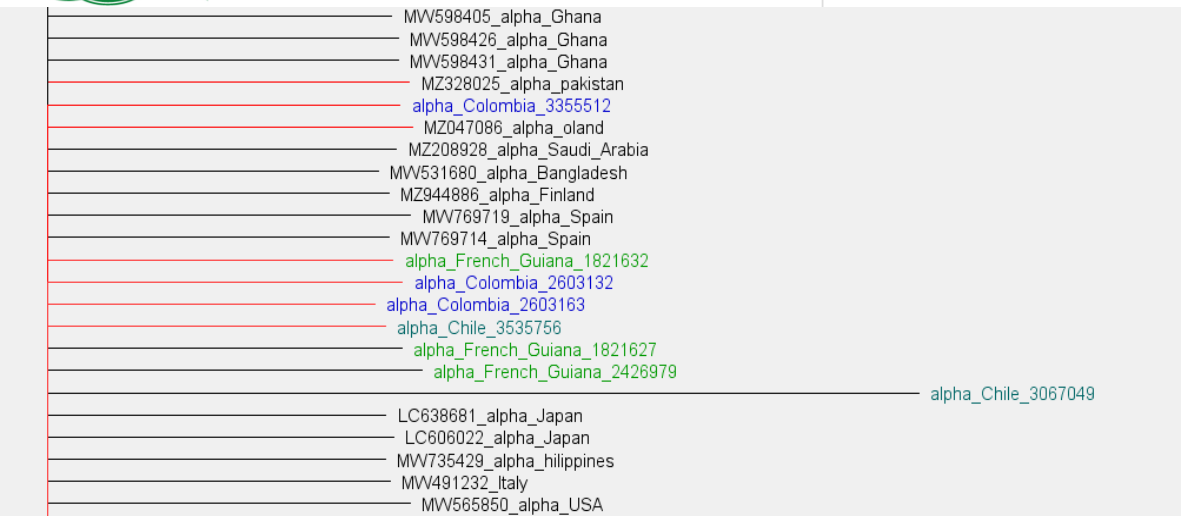


Imagen 28 Análisis filogenético de la cepa Alpha para la proteína Orf6. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf6, presenta secuencias biológicas conservadas de las cepas Alpha de Colombia, en relación con las del resto del mundo.

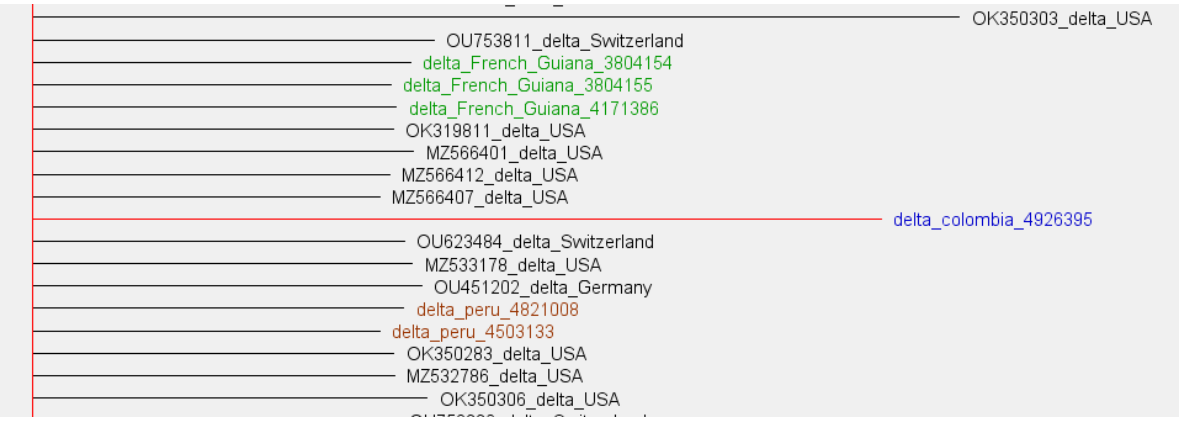


Imagen 29 Análisis filogenético de la cepa delta para la proteína Orf6. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf6, presenta secuencias biológicas conservadas de las cepas Delta de Colombia, en relación con las del resto del mundo.

En conclusión, la proteína Orf6 del SARS-CoV-2, se encuentra muy conservada y se ha mantenido por la evolución, a pesar de la caracterización de variantes existentes en el mundo.

Existen una cladogénesis no resuelta entre variantes kappa de india, USA y Suiza y otra entre las variantes zeta y desconocida de USA, beta de Pakistán y Alpha de Egipto que no están relacionadas con las de Colombia.

Filogenia de la proteína ORF3a de las variantes Iota, Eta, Beta. Épsilon, kappa, Alpha, Mu, Delta, Gamma y Zeta.

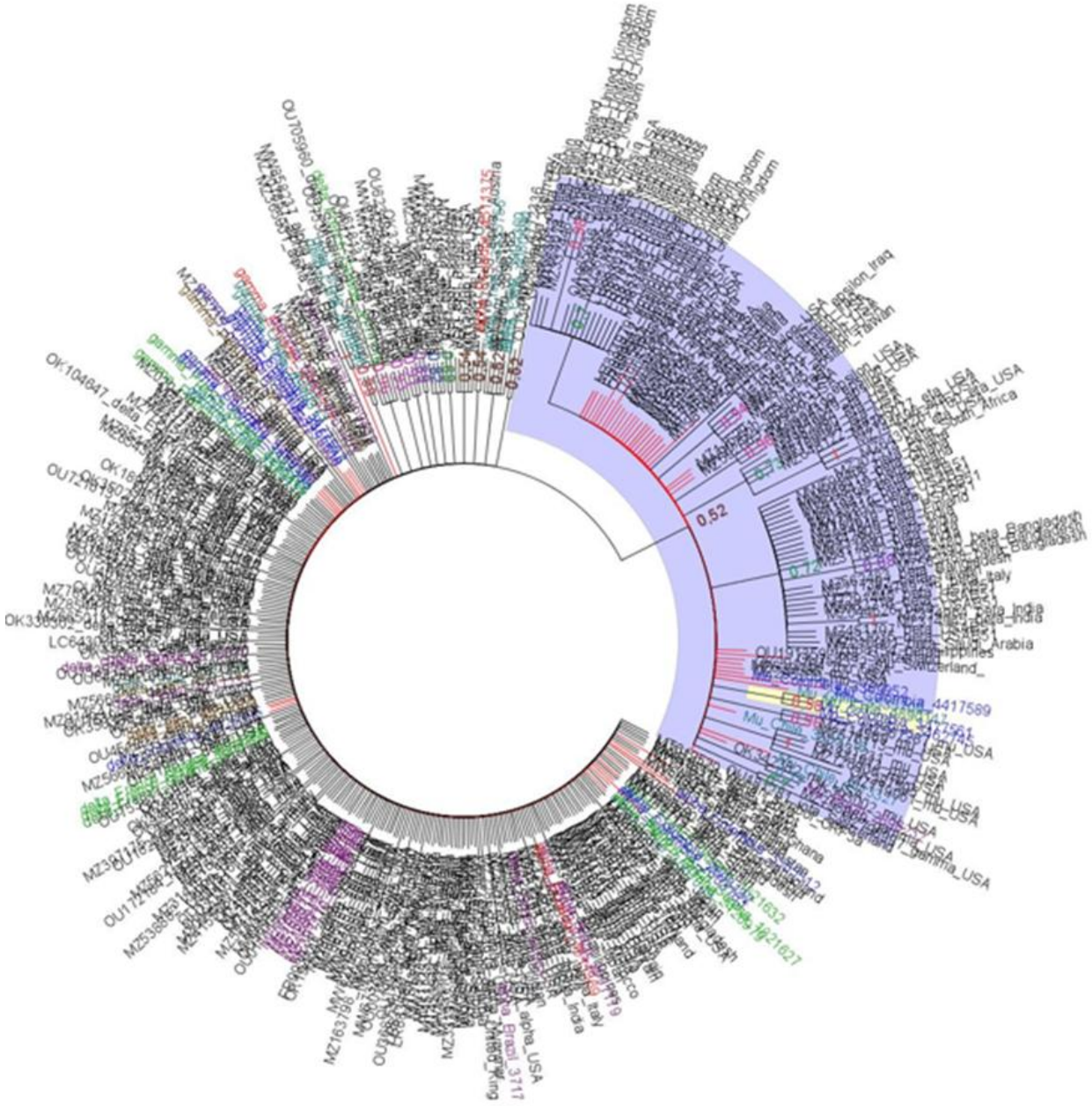


Imagen 30 Análisis filogenético para la proteína Orf3a. Fuente: base de datos propia.

Árbol filogenético obtenido, usando el método de Inferencia Bayesiana, para el gen ORF3a. Los números corresponden a valores de probabilidades posteriores bayesianas. Los grupos resaltados en color en azul y beige corresponden a la ubicación de algunas cepas de Colombia, en relación con las del resto del mundo y los nombres de color azul corresponden a las cepas de Colombia. Los modelos evolutivos de las posiciones canónicas del gen Orf3a fueron TIM2 Y GTR.

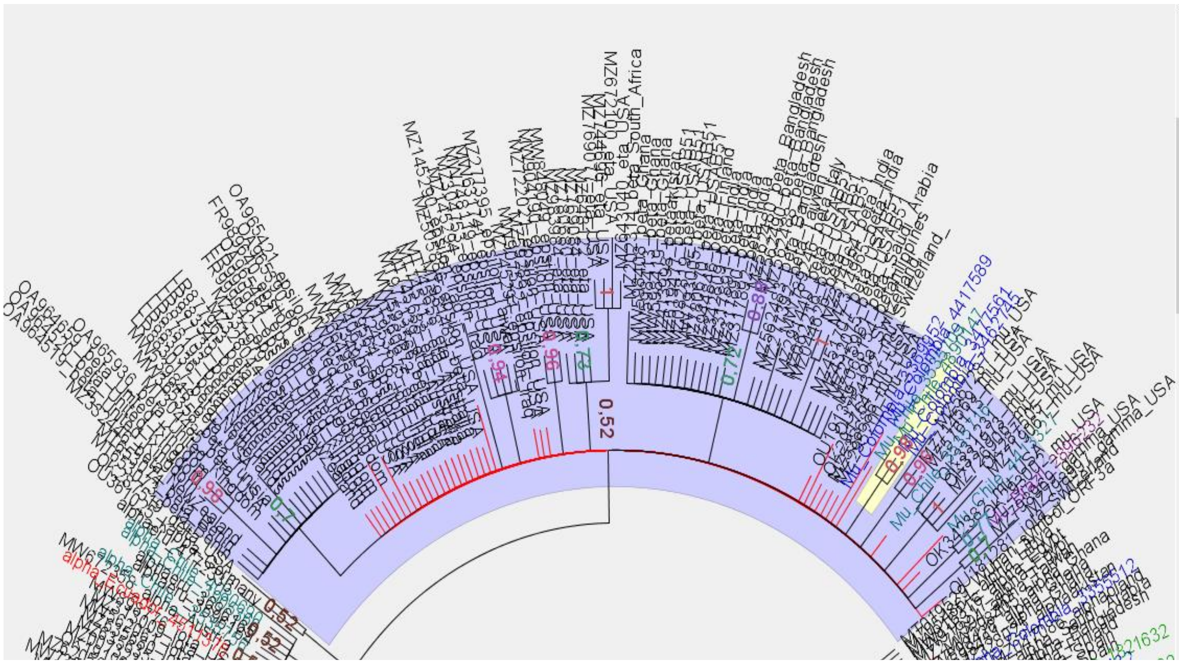


Imagen 31 Análisis filogenético de la cepa Mu para la proteína Orf3a. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf3a, presenta un grupo con probabilidades posteriores bayesianas de 0.52. Las cepas Mu Colombia, presentan homología en la proteína orf3a, en relación con las secuencias Mu del mundo y las cepas, beta, eta, iota y épsilon.

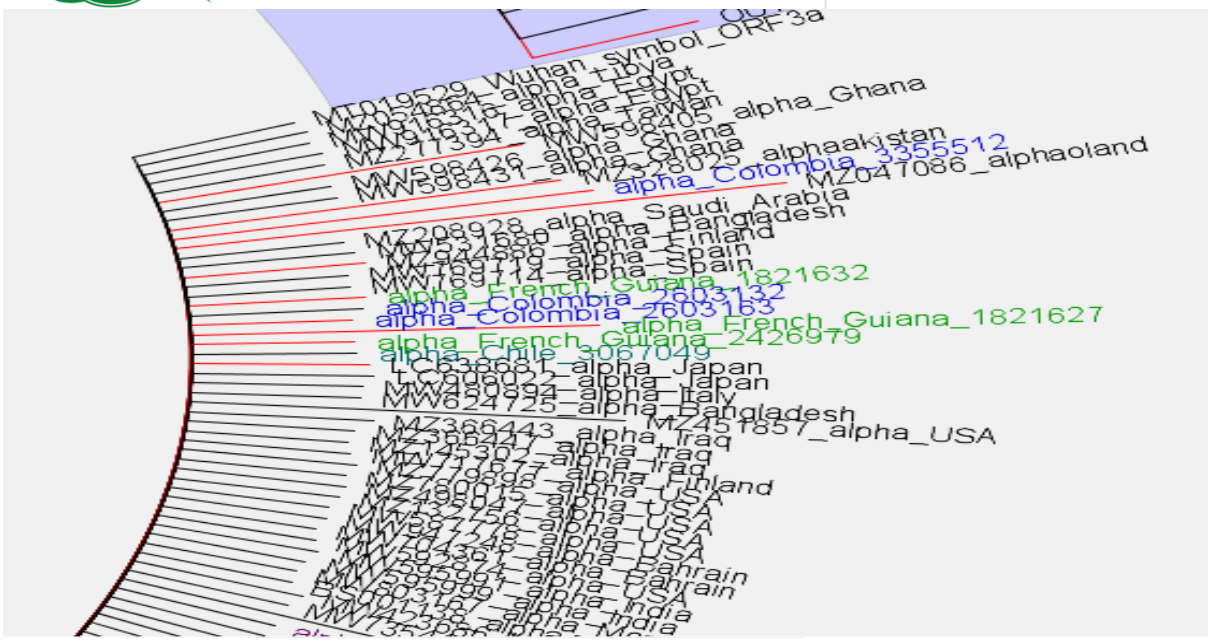
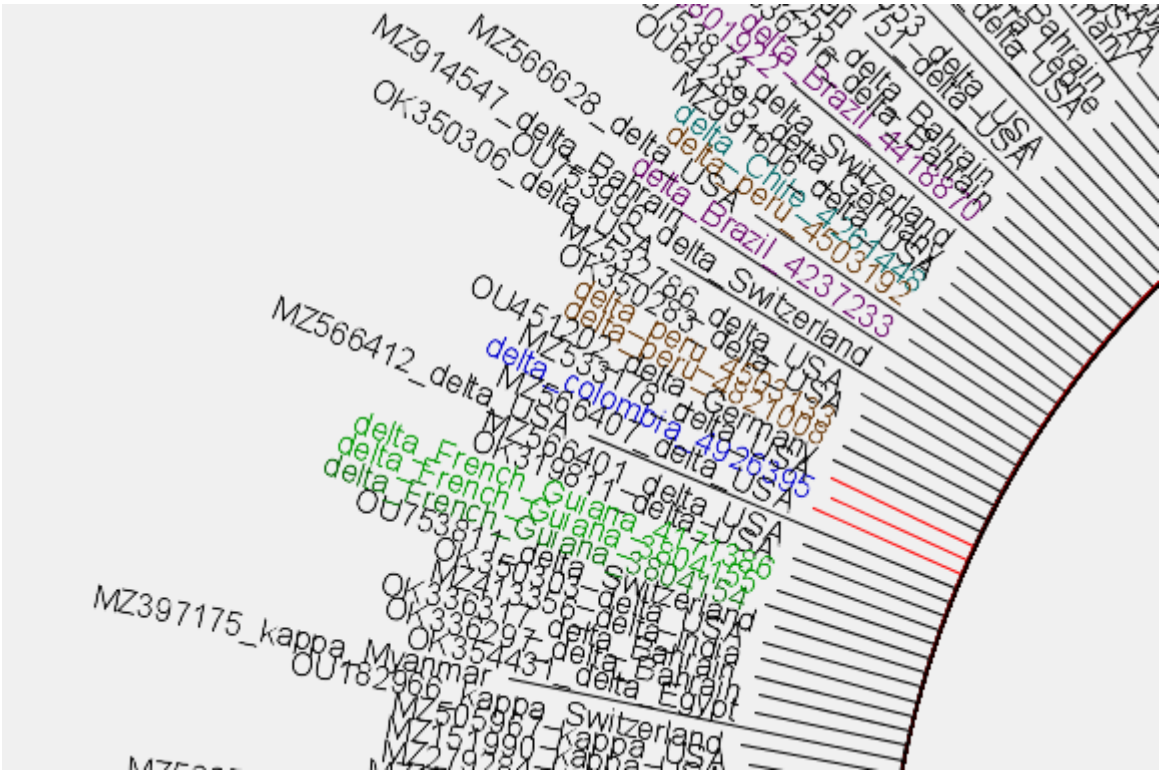


Imagen 32 Análisis filogenético de la cepa Alpha para la proteína Orf3a. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf3a, presenta secuencias biológicas conservadas de las cepas Alpha de Colombia, en relación con las del resto del mundo.



.
Imagen 33 Análisis filogenético de la cepa delta para la proteína Orf3a. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf3a, presenta secuencias biológicas conservadas de la cepa Delta de Colombia, en relación con las del resto del mundo.

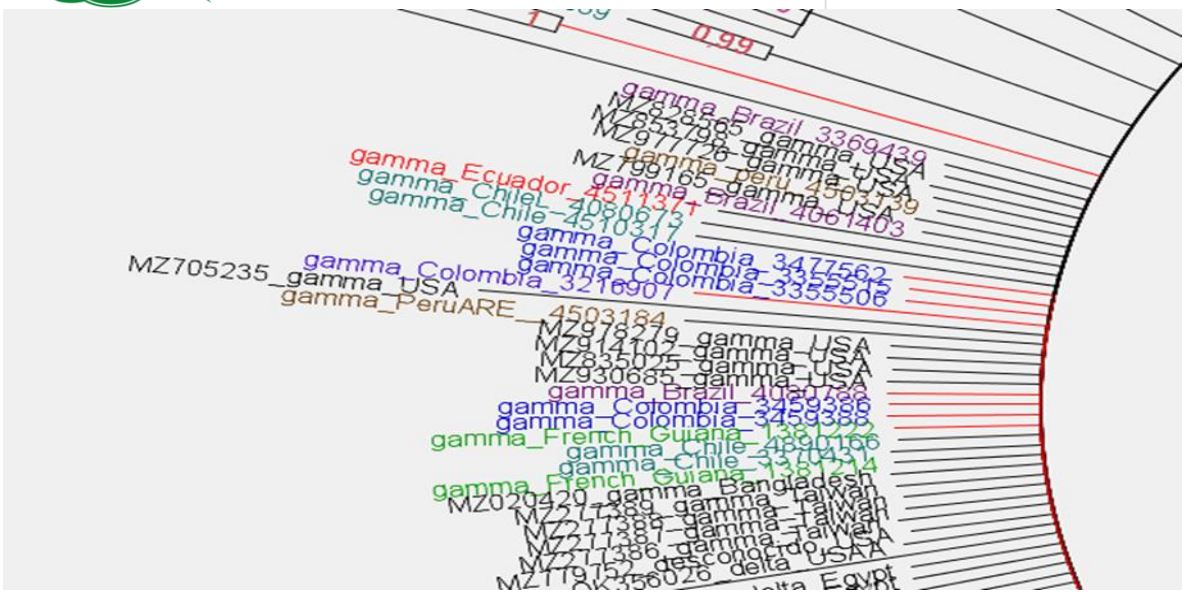


Imagen 34 Análisis filogenético de la cepa gamma para la proteína Orf3a. Fuente: base de datos propia.

El análisis bayesiano para la proteína Orf3a, presenta secuencias biológicas conservadas de las cepas Gamma de Colombia, en relación con las del resto del mundo.

La proteína ORF3a, muestra grupos polifiléticos relacionados en un nodo, donde se encuentran enraizadas las secuencias de Mu, chile y USA y variante iota, épsilon y eta con una probabilidad posterior en su nodo de 0.52, pero se encontró una probabilidad de 0.98 entre las cepas mu de chile y Colombia las cuales se encuentran según el resultado más relacionadas. Existe también, un grupo externo muy conservado que incluye secuencias Alpha, zeta, eta, Kappa, delta, gamma de Colombia y el resto del mundo, lo que indica, que la proteína ORf3a de Colombia, que ha cambiado en el tiempo, ha sido la relacionada a la variante Mu. Lo que indica que este proteína ha evolucionado más lento, de lo que se puede esperar (50).

Filogenia de la proteína M de las variantes Iota, Eta, Beta. Épsilon, kappa, Alpha, Mu, Delta, Gamma y Zeta.

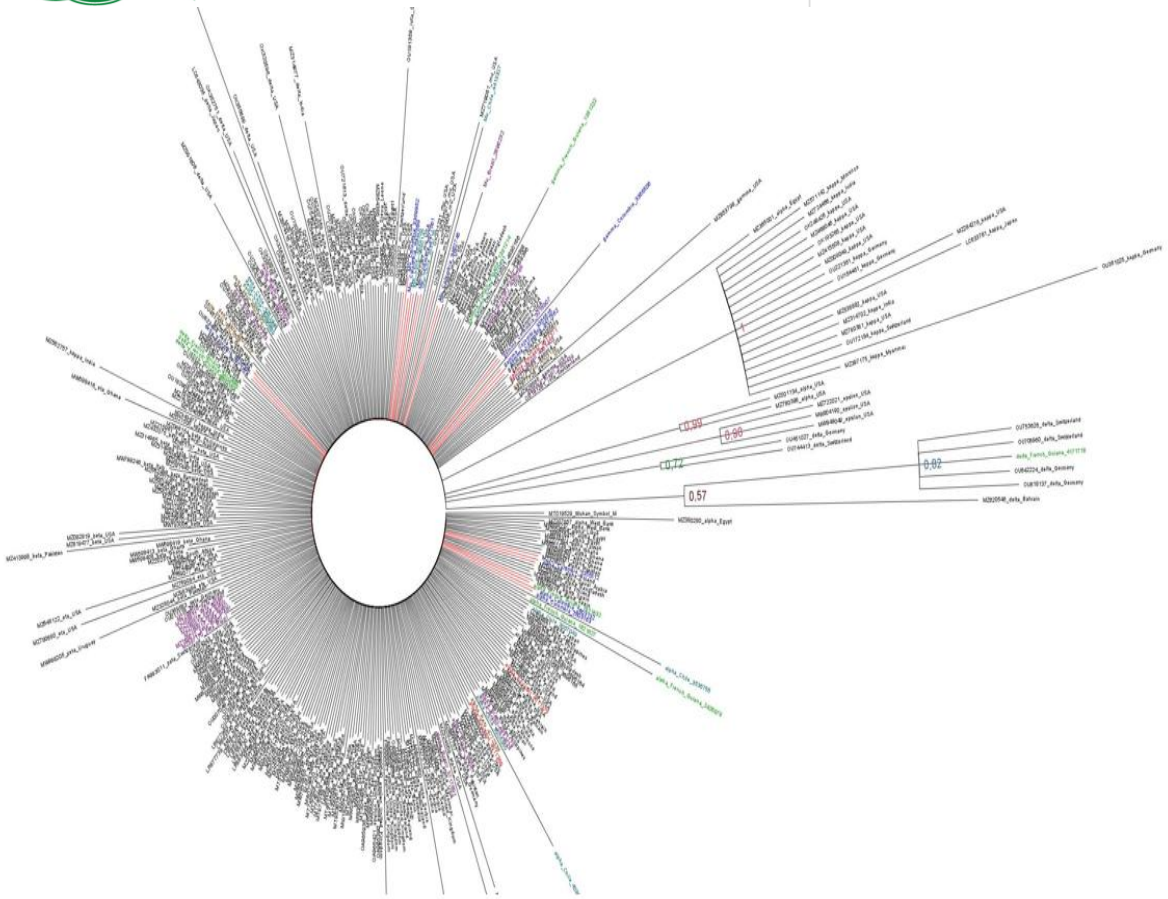


Imagen 35 Análisis filogenético para la proteína M. Fuente: base de datos propia. Árbol filogenético obtenido usando el método de Inferencia Bayesiana para el gen M. Los números corresponden a valores de probabilidades posteriores bayesianas. Los nombres en color azul corresponden a la ubicación de las cepas de Colombia, en relación con las cepas del resto del mundo. Los modelos evolutivos de las posiciones canónicas del gen M son: TPM2uf y TIM2.

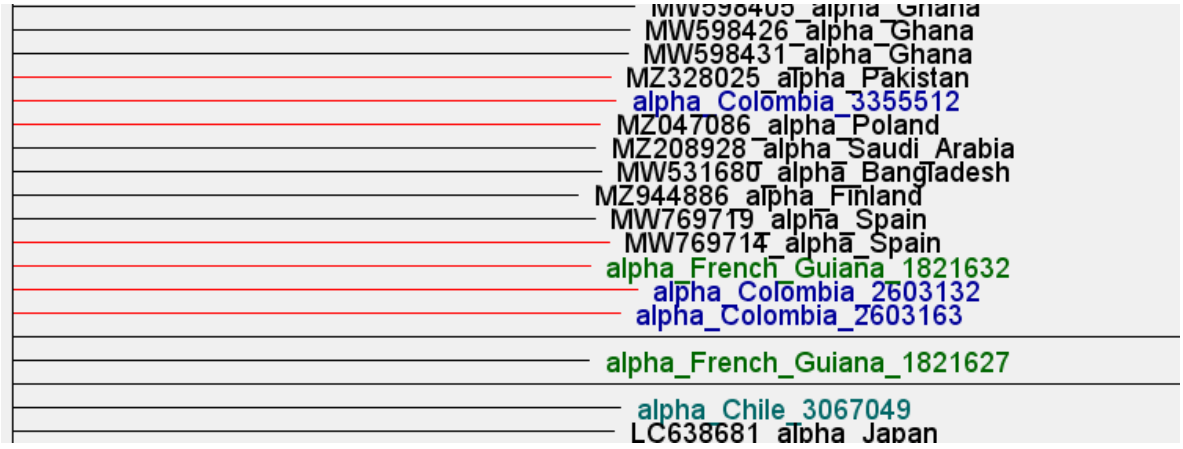


Imagen 36 Análisis filogenético de la cepa Alpha para la proteína M. Fuente: base de datos propia.

El análisis bayesiano para la proteína M, presenta secuencias biológicas conservadas de las cepas Alpha de Colombia, en relación con las del resto del mundo.

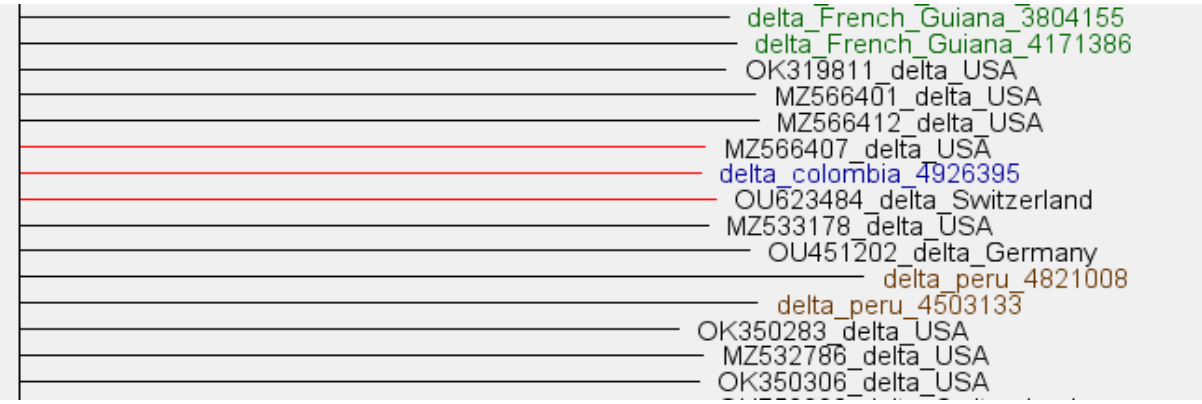


Imagen 37 Análisis filogenético de la cepa delta para la proteína M. Fuente: base de datos propia.

El análisis bayesiano para la proteína M, presenta secuencias biológicas conservadas de las cepas Delta de Colombia, en relación con las del resto del mundo.

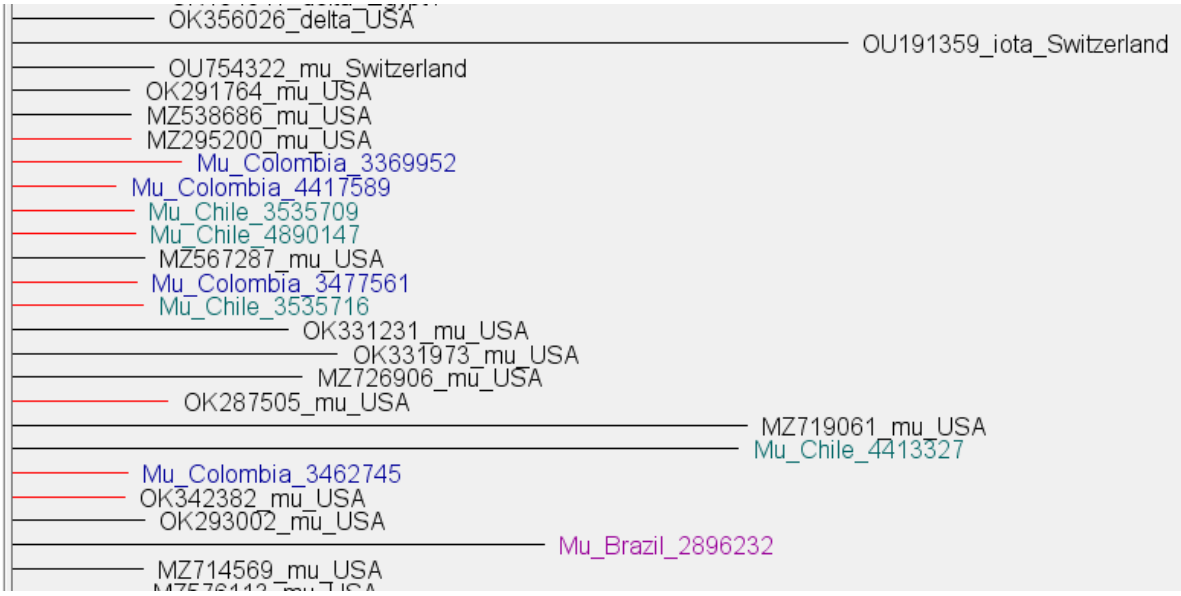


Imagen 38 Análisis filogenético de la cepa Mu para la proteína M. Fuente: base de datos propia.

El análisis bayesiano, para la proteína M, presenta secuencias biológicas conservadas de las cepas Mu de Colombia, en relación con las del resto del mundo.

En conclusión, la proteína M es muy conservada, en cuanto a las cepas que circulan en Colombia. Existen cepas como las europeas y africanas, las cuales poseen la proteína con más cambios evolutivos y se encuentran más alejadas de las cepas de Colombia.

Filogenia de la proteína E de las variantes Iota, Eta, Beta. Épsilon, kappa, Alpha, Mu, Delta, Gamma y Zeta.

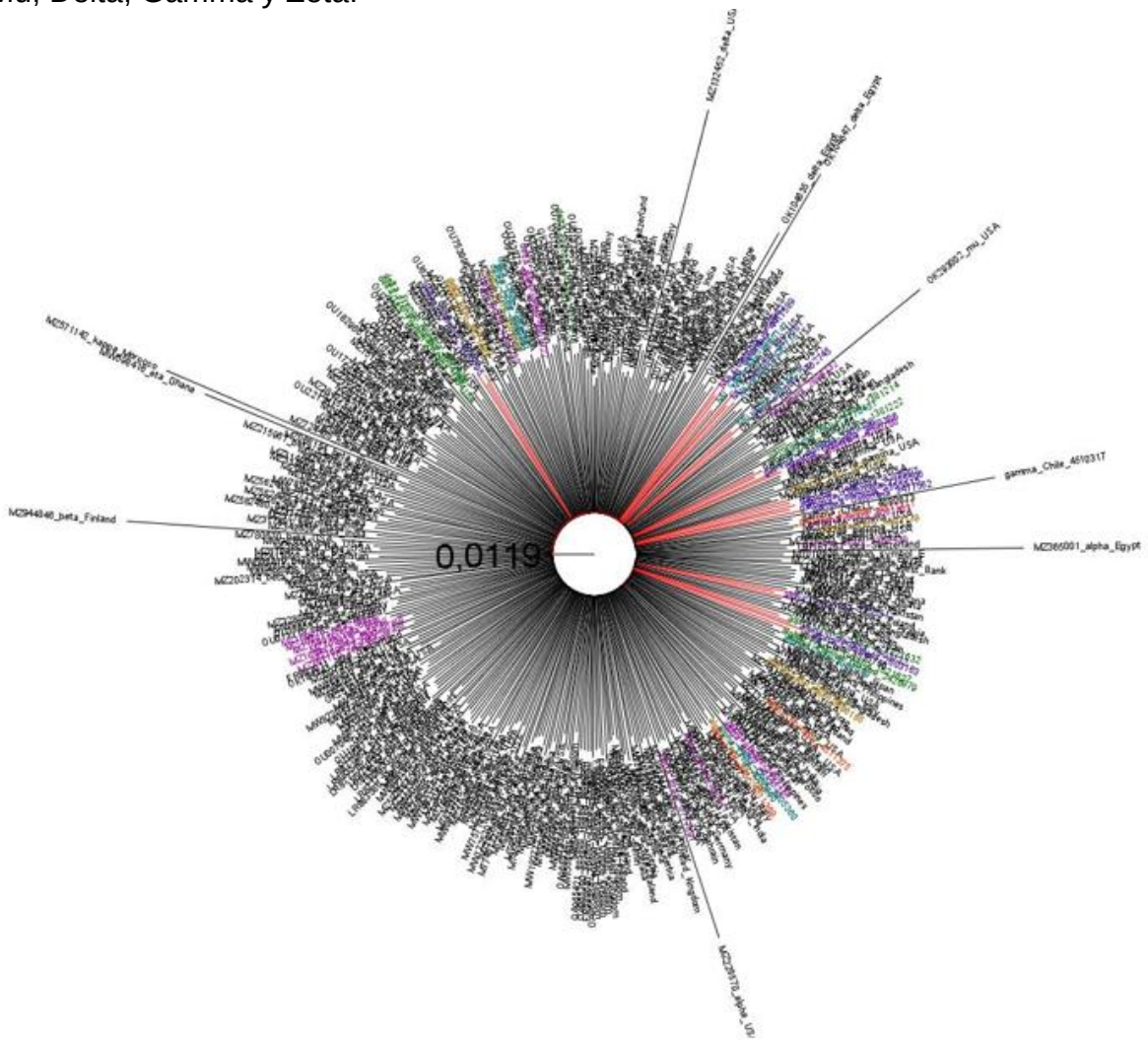


Imagen 39 Análisis filogenético para la proteína E. Fuente: base de datos propia.
Árbol filogenético obtenido, usando el método de Inferencia Bayesiana, para el gen E. Con probabilidades posteriores bayesianas de 0,0119. Las secuencias identificadas en color azul corresponden a la ubicación de las cepas de Colombia, en relación con las cepas del resto del mundo. Los modelos evolutivos de las posiciones canónicas del gen E, TrN y TIM3. Esta proteína se encuentra conservada en todas las cepas del mundo.

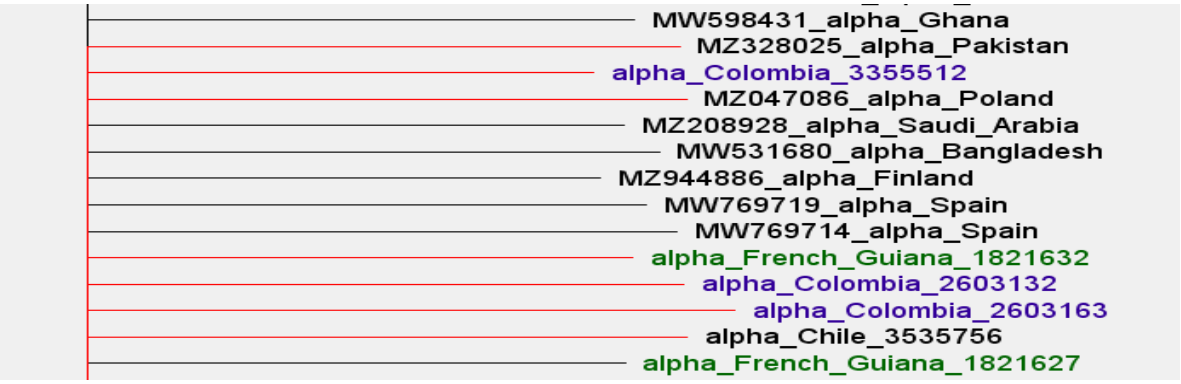


Imagen 40 Análisis filogenético de la cepa Alpha para la proteína E. Fuente: base de datos propia.

El análisis bayesiano para la proteína E, presenta secuencias biológicas conservadas de las cepas Alpha de Colombia, en relación con las del resto del mundo.

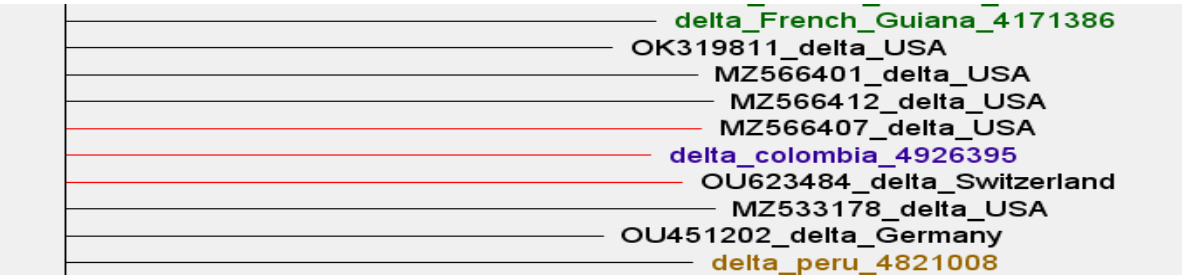


Imagen 41 Análisis filogenético de la cepa Delta para la proteína E. Fuente: base de datos propia.

El análisis bayesiano para la proteína E, presenta secuencias biológicas conservadas de las cepas delta de Colombia, en relación con las del resto del mundo.

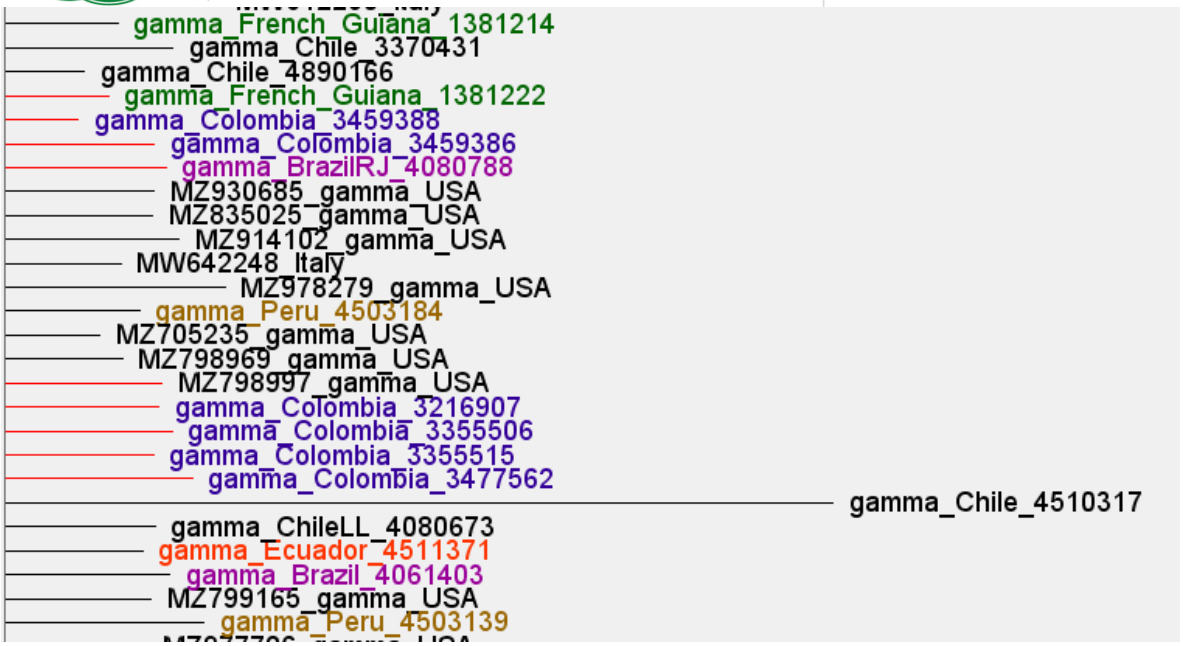


Imagen 42 Análisis filogenético de la cepa gamma para la proteína E. Fuente: base de datos propia.

El análisis bayesiano para la proteína E, presenta secuencias biológicas conservadas de las cepas Gamma de Colombia, en relación con las del resto del mundo.

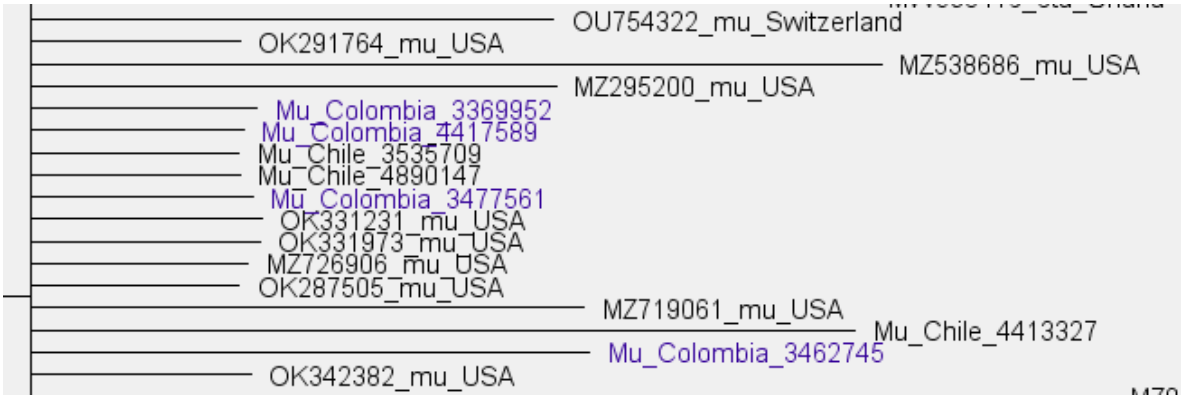


Imagen 43 Análisis filogenético de la cepa Mu para la proteína E. Fuente: base de datos propia.

El análisis bayesiano para la proteína E, presenta secuencias biológicas conservadas de las cepas Mu de Colombia, en relación con las del resto del mundo.

En conclusión, la proteína E es una proteína muy corta y a la vez conservada en su totalidad.

Se desconocen mutaciones importantes en la proteína E y porque ocurren menos errores en esta proteína que en las otras 3 proteínas estructurales hasta el momento.

Filogenia de la proteína N de las variantes Iota, Eta, Beta. Épsilon, kappa, Alpha, Mu, Delta, Gamma y Zeta.

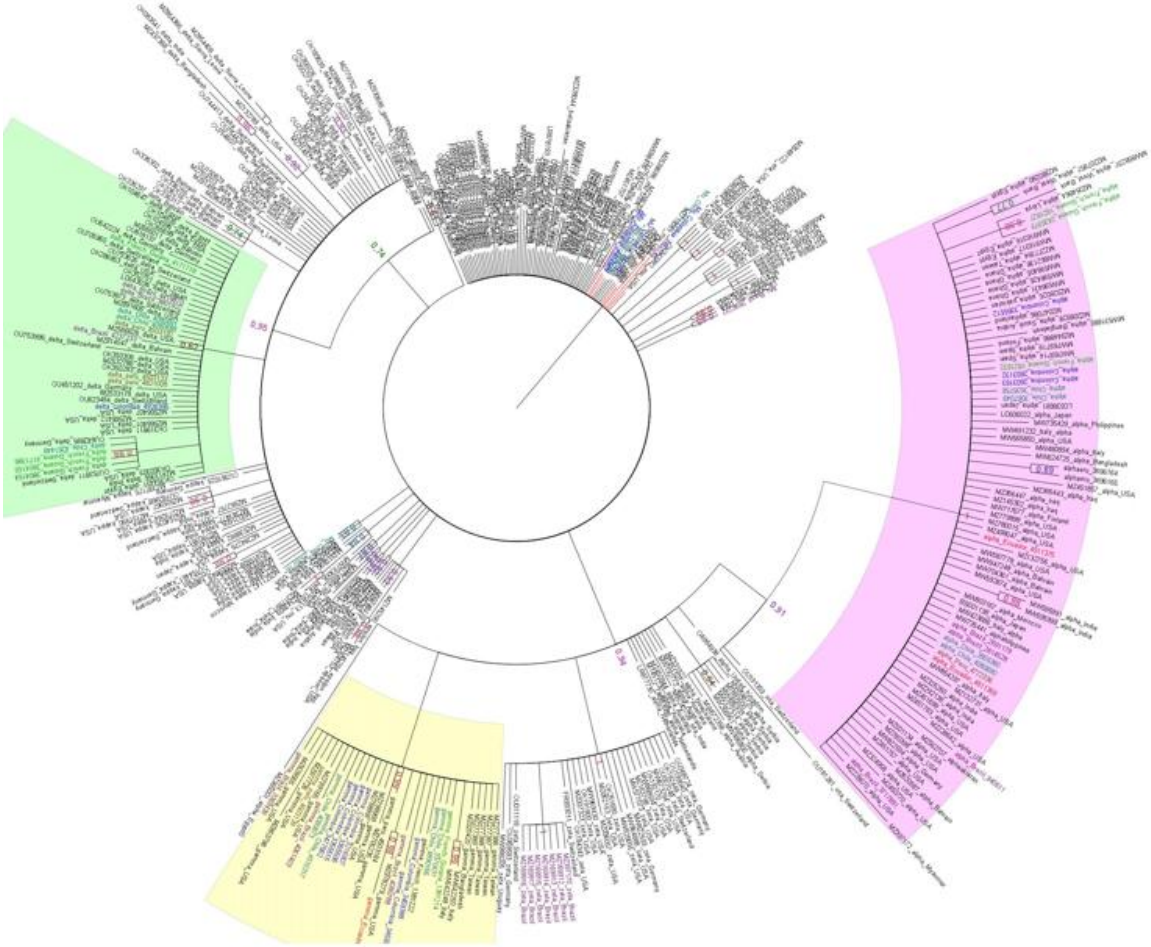


Imagen 44 Análisis filogenético para la proteína N. Fuente: base de datos propia. Árbol filogenético obtenido, usando el método de Inferencia Bayesiana para el gen N. Los números corresponden a valores de probabilidades posteriores bayesianas. Los nombres de color azul son las cepas pertenecientes a Colombia y los nombres resaltados en color beige, corresponden al grupo donde están ubicadas algunas cepas de Colombia, en relación con las cepas del resto del mundo. Los modelos evolutivos de las posiciones canónicas del gen N es GTR.

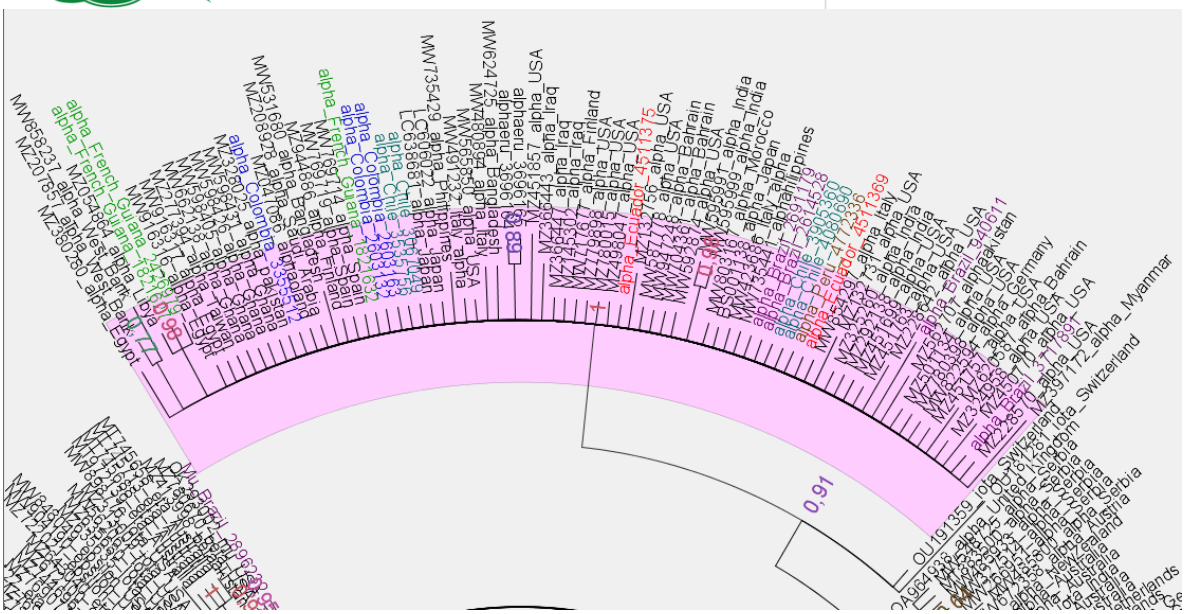


Imagen 45 Análisis filogenético de la cepa Alpha para la proteína N. Fuente: base de datos propia.

El análisis bayesiano para el gen de la proteína N, presenta un grupo polifilético con probabilidades posteriores bayesianas de 0.91. Las cepas Alpha Colombia, presentan homología en la proteína N en relación con las secuencias Alpha del mundo, con excepción de algunas europeas.

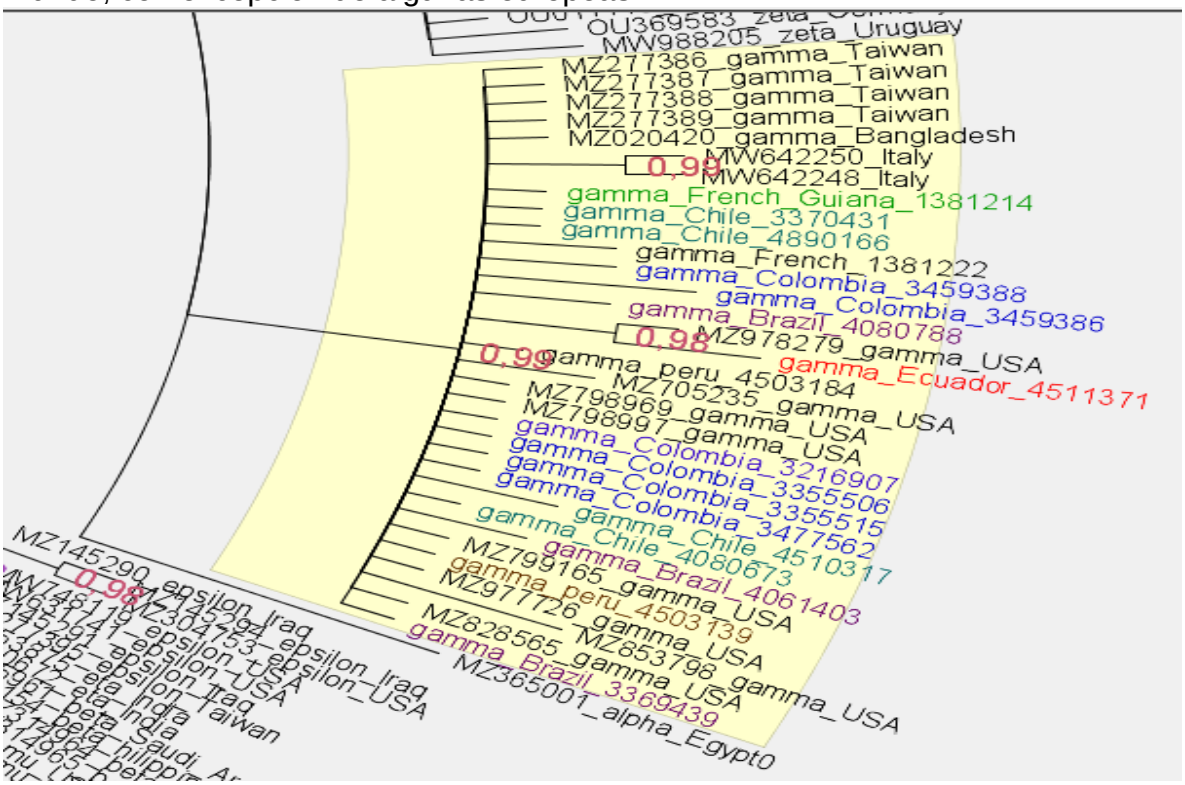


Imagen 46 Análisis filogenético de la cepa gamma para la proteína N. Fuente: base de datos propia.

El análisis bayesiano para el gen de la proteína N, presenta un grupo polifilético con probabilidades posteriores bayesianas de 0.91. Las cepas Gamma Colombia, presentan homología en la proteína N, en relación con las secuencias gamma de Usa, Suramérica, y algunas de Asia y de Europa.

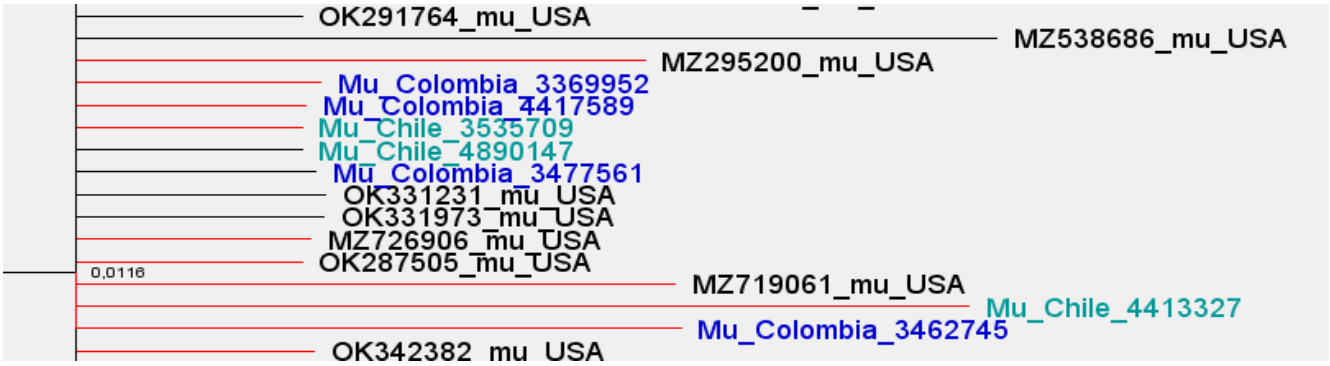


Imagen 47 Análisis filogenético de la cepa Mu para la proteína N. Fuente: base de datos propia.

El análisis bayesiano para la proteína N, presenta secuencias biológicas conservadas de las cepas Mu de Colombia, en relación con las del resto del mundo.

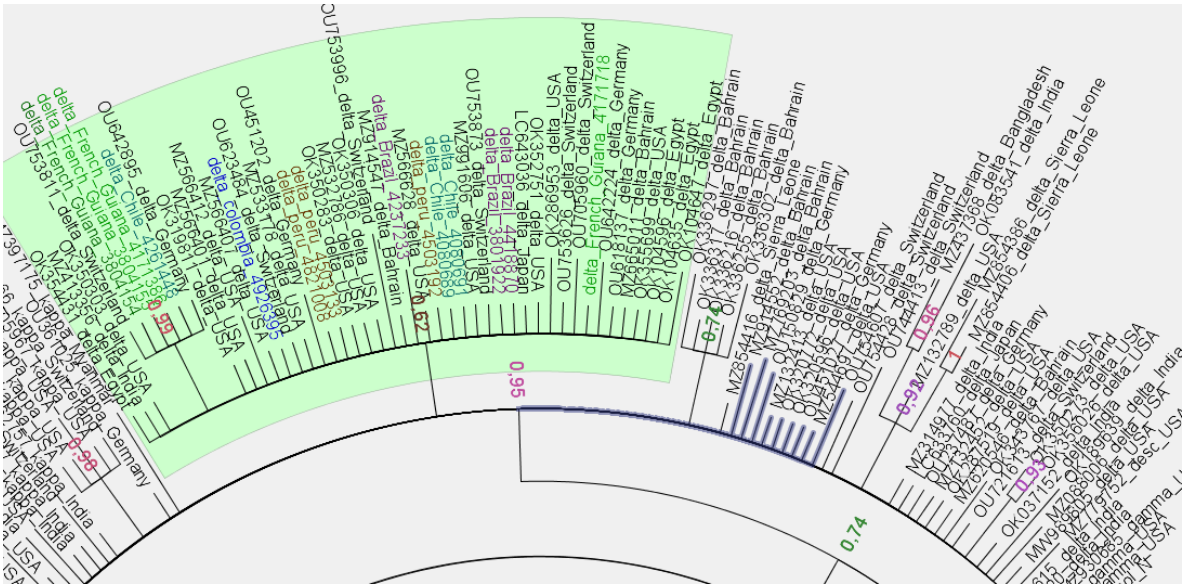


Imagen 48 Análisis filogenético de la cepa delta la para proteína N (base de datos propia).

El análisis bayesiano para la proteína N presenta, un grupo con probabilidades posteriores bayesianas de 0.95, entre las cepas delta de Colombia y Delta del mundo y presenta una probabilidad posterior de 0.62 con la proteína N de Suramérica y algunas de Usa y Europa.

En conclusión, encontramos 3 grupos donde se encuentran las secuencias de Colombia identificamos, diferentes polimorfismos a nivel mundial, lo que determina que esta proteína ha evolucionado, de tal manera que existe relaciones evolutivas entre las cepas gamma de Colombia y el mundo con una probabilidad posterior de 0.99, como también existe verosimilitudes entre las variantes Alpha de Colombia y el resto mundo analizadas en este estudio, con una probabilidad posterior de 1. Existe también, un grupo con probabilidades posteriores de 0.74 entre las cepas delta de Colombia y de resto de países del mundo y la variante kappa de las secuencias analizadas.

Existen variantes de la proteína N conservada y otras similares entre las que se encuentran las cepas Mu de Colombia, chile, USA, Suiza y las cepas épsilon y beta.

Las proteínas Alpha y gama de Colombia poseen homología con una probabilidad posterior de 0,94.

Shaminur “et al” en su artículo Dinámica evolutiva de la proteína nucleocápside del SARS-CoV-2 y sus consecuencias nos indica: que esta proteína implica una evolución continúa, debido a que encontraron 1.034 mutaciones únicas, la mayoría pertenecientes a países europeos, estadounidense y australianos. A la fecha no se registran estudios de mutaciones importantes de la proteína N, en las cepas Mu de Suramérica.

Conclusiones

Los análisis bayesianos, permitieron conocer los cambios evolutivos y la conservación de las proteínas de las proteínas S, M, N, E y las ORF (3a, 6, 7a, 8, 10)

Existe una colosal divergencia, en las proteínas S y N, indicando que son las proteínas menos conservadas, en especial la proteína S, que tiene un alto índice de evolución en el tiempo, y que difiere en mayor medida entre una variante y otra, tanto en Colombia (país del continente suramericano), como en los demás países de los demás continentes del mundo, analizados en este estudio, concluyendo así, que cada variante de Colombia en relación con la misma, en diferentes partes del mundo, se encuentran relacionadas, lo que traería el mismo efecto en cuanto a estudios de efectividad, en las vacunas actuales que usan como ARNm, sitios, que codifican a la proteína S, pero esto, podría cambiar en la medida en que la evolución de esta proteína continúe, y se creen nuevas variantes de interés y preocupación.

Las proteínas Orf3a y Orf8, son proteínas que han evolucionado de forma más lenta que las proteínas S y N y se corrobora en investigaciones mencionadas en este análisis, donde nos revelan, que estas proteínas, poseen mutaciones selectivas(28).

La proteína M, Orf6, Orf7a, Orf10 en Colombia, se encuentran conservadas al igual que las del resto del mundo, aunque existen algunas variantes no resueltas, debido a evoluciones convergentes en algunas secuencias que no pertenecen a Colombia.

Otro resultado importante es que la proteína E, es la proteína más conservada, en el SARS-CoV-2 y la S y la N son las menos conservadas, debido a polimorfismos existentes en ellas, siendo estás probablemente buenos marcadores, para diferenciar variantes.

Palabras clave:
Evolución, filogenética, variabilidad, genoma, mutabilidad.

ABSTRACT

SARS-CoV-2 is a virus that causes COVID19 disease, which appeared in Wuhan, Hubei province in China in December 2019(1), transmitted through the inhalation of respiratory droplets from one infected person to another(15), being an important threat to health worldwide, due to the high infection and mortality rates it presents.
Studies suggest the existence of a recombination between animal coronaviruses, which gave rise to the existence of the new SARS-CoV-2 virus, this time being effective at the time of infecting man(26). Despite its mutagenic and recombinogenic capacity, it is important to know its genetic features and the evolutionary processes in the variants that existed in Colombia, in relation to those of the rest of the world, since there is no information in our country about it.
In order to understand the evolutionary processes to which this new virus is subjected, it is important to obtain a phylogeny of the different circulating variants and to analyze this variability, among other characteristics, both in our territory and in the world.
Through a study of SARS-CoV-2 variants in Colombia, the mutagenic capacity was analyzed to understand factors such as the fidelity of the enzymes that replicate nucleic acids and allow the virus to adapt through the variability of the genome and the integrity of the genetic information, suggesting that the virus is and will continue to evolve.
This was done through a bioinformatic analysis using genes of structural and functional importance of the virus, called S, M, N, E proteins and accessory ORF proteins (3a, 6, 7a, 8 and 10), being the S protein gene the most used in vaccine production studies against the SARS-CoV-2 virus.
Therefore, 460 sequences were chosen in total from 10 strains of interest and concern from the 6 continents of the world, of which 68 belong to South America, being the Colombian variants of the delta, gamma and Alpha strains from the GenBank database of the NCBI(41) and VIPR(42) platform.
For the phylogenetic study through Bayesian inference, programs such as MAFFT version 7(43), for sequence alignment; ModelTest-NG in XSEDE version 3.3(44),

for the selection of the best substitution model; MrBayes version 3.2(44), for the creation of phylogenetic trees and FigTree v1.4.4 graphics, to observe the trees and analyze them, were used. This allowed us to understand the evolutionary development and maintenance of SARS-CoV-2 proteins in each strain found in Colombia in relation to the world, determining strains that have greater evolutionary changes, plausibility or that remain conserved over time; thus determining genomic and phylogenetic characteristics of the circulating SARS-CoV-2 subtypes.
Leading us to the conclusion that there is a colossal divergence in proteins S and N, in addition to a slower evolution of proteins Orf3a and Orf8, the conservation of proteins M, Orf6, Orf7a, Orf10 with some variants not resulting due to convergent evolutions, and a greater conservation of the E protein.

Objectives:
General Objective.
To determine the genomic characteristics of SARS-CoV-2 subtypes circulating in Colombia.

Specific objectives
Objective 01
Compile information about the structure, functionality and variations in the genomic sequences of the SARS-CoV-2 subtypes circulating in Colombia and the world.
Objective 02
To analyze the evolutionary changes and relationships, of the SARS-CoV-2 monitoring variants in Colombia in relation to the world, using representative sequences, translating to proteins, by phylogenetic reconstruction, through Bayesian inference topology**.**

Background:

Coronavirus epidemics
Coronaviruses are known to infect mammals, birds and fish. These are single-stranded RNA positive viruses(11) that were never seen as highly pathogenic until the 2003 outbreak of severe respiratory syndrome (SARS), however, epidemics such as Middle East respiratory syndrome (MERS) and the current SARS-CoV-2 are now considered a challenge to global health security(1).
Coronaviruses have a large genome, which allows them to have more plasticity to accommodate and modify genes(12), and RNA viruses have a relatively high frequency of mutations, which increases virulence and the formation of new species(13). In this case it could be the result of the frequency of mutations of some SARS-CoV-2 genes in different geographical regions and the change in mortality rates and symptoms of COVID-19(10).
Sifuentes "et al", (1) shows how SARS-CoV-2 was identified in its beginnings, revealing that on December 31, 2019 the Centers for Disease Control and Prevention (CDC China) conducted investigations, due to the frequency of people who appeared with pneumonia, from which samples were obtained and in their

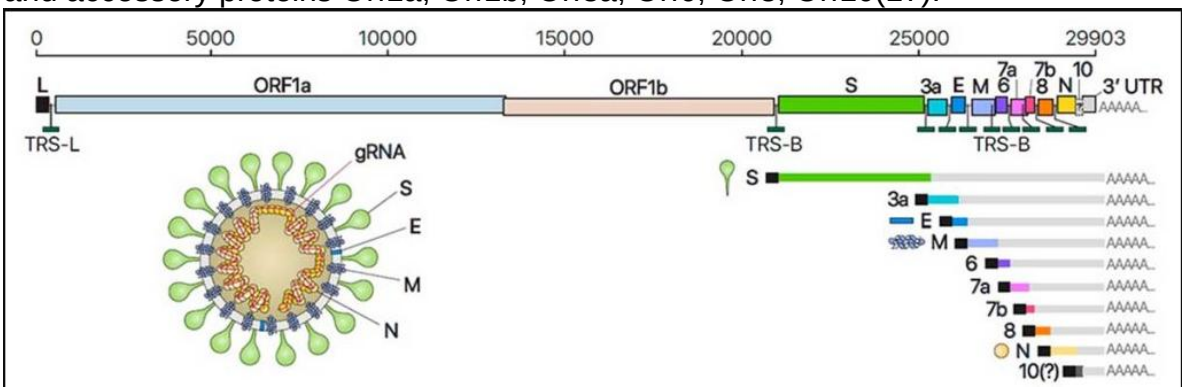
study more than 20000 readings were found that showed a virus that possessed compatibility with the lineage of the genus beta-coronavirus, naming it 2019-nCoV. Some time later, January 30, 2020 was declared by WHO as COVID-19 and public health emergency due to the increasing number of cases.

Chitranshi (14) "et al" also shows us that coronaviruses can cause acute and chronic diseases of the central nervous system and respiratory system, mild episodes of follicular conjunctivitis, impairment of the sense of smell and taste bud sensitivity; and in animals, induce symptoms similar to anterior uveitis, retinitis and optic neuritis, in addition to hyperreflective lesions in the ganglion cells and inner plexiform layers of the retina, particularly around the papilloma cularis bundles.

It is important to highlight that COVID-19 is caused by SARS-CoV-2, which is transmitted through inhalation of respiratory droplets from an infected person(15); its incubation period ranges from 2 to 14 days, with symptoms on approximately the fifth day(16). It has asymptomatic manifestations, mild or severe pneumonia, with symptoms such as fever, dry cough, myalgia, arthralgia, anosmia, dysgeusia, fatigue and respiratory distress; and to a lesser extent, diarrhea, nausea and vomiting(17).

Structure of SARS-CoV-2

SARS-CoV-2 is known to be a single-stranded RNA virus, 29903 nucleotides long, encoding 12 peptides, including two closely related polyproteins, Orf1a and Orf1ab that generate 16 non-structural proteins (nsps), responsible for viral genome replication, subgenomic mRNA transcription, and 4 structural genes, such as Spike protein (S), nucleocapsid protein (N), membrane protein (M), envelope protein (E) and accessory proteins Orf1a, Orf1b, Orf3a, Orf6, Orf8, Orf10(17).



Molecular structure of Sars-CoV-2. Source:(18)

The helical N protein covers the viral genomic RNA and its function is to encapsulate the genetic material. It also participates in the assembly, replication, envelope formation, viral particle release, cell cycle regulation, and inhibition of interferon-mediated immune responses, making it fundamental in pathogenesis; therefore, it is a protein under study for the development of effective antivirals against this respiratory pathogen(19).

The E protein is the most external structure with structural and necessary functions for the maturation and production of the virus. While the M protein maintains the shape and is responsible for assembly, together with the N protein(19).

The ORF8 protein is a protein that contains 121 amino acids and is vital for the transmission and efficiency of replication, because it acts by negatively regulating the MHC molecules, important in the antiviral immune response of the host; in addition to giving the infected cells a chance of survival, so it is important to recognize the genetic variability and its evolution, to understand how the proteins act in a better way and thus provide relevant information in studies of drugs and antivirals against SARS-CoV-2(20).

The ORF6 protein is involved in the remodeling of the intracellular membrane, which could influence an increase in virus replication(21).

ORF10 seems to be involved in the inhibition of innate immunity, in addition to promoting viral replication by inducing mitophagy to degrade MAVS, giving the virus the ability to infect and decrease the immune response(22).

Recent studies have shown that ORF7a protein binds to CD14+ monocytes, producing a decrease in HLA-DR / DP / DQ molecules, thereby decreasing the ability to present antigens, increasing the ability of infection(23).

The accessory protein Orf3a, has functions of virulence, ineffectiveness, ion channel activity, morphogenesis and virus release. For this reason, due to the mutations that have occurred in the virus, it is necessary to study this protein(24) and all the previous ones described, in order to relate its influence with the pathogenicity of the virus.

Protein S is found in the external part of the virus with a particular shape similar to a crown, which is essential in the recognition, adhesion and penetration of the cell to infect(19). Pastrian(17) et al. in their article Genetic and Molecular Basis of COVID-19 (SARS-CoV-2). He explains the Mechanisms of Pathogenesis and Immune Response, showing that protein S binds to the angiotensin-converting enzyme receptor 2 (ACE2) to infect the cell. This protein is composed of the subunits s1 and s2, being s1 the one that interacts with the receptor by means of the RBD receptor binding domain, while the s2 subunit determines the fusion of the virus membrane, so that the entry is complete; This must be cut by a protease enzyme, to enter the cell, via endocytic and then release their genetic material to the cytoplasm, translating directly into pp1a, and pp1ab, (polyproteins), which will undergo enzymatic proteolysis to generate the 16 nsps proteins of the RTC complex, which replicates and synthesizes a set of (sgRNA) that encode for the production of structural and accessory proteins that make up the virus, along with the nucleocapsid; and finally these are assembled at the level of the Golgi complex, to form new viruses and be released from the infected cell.

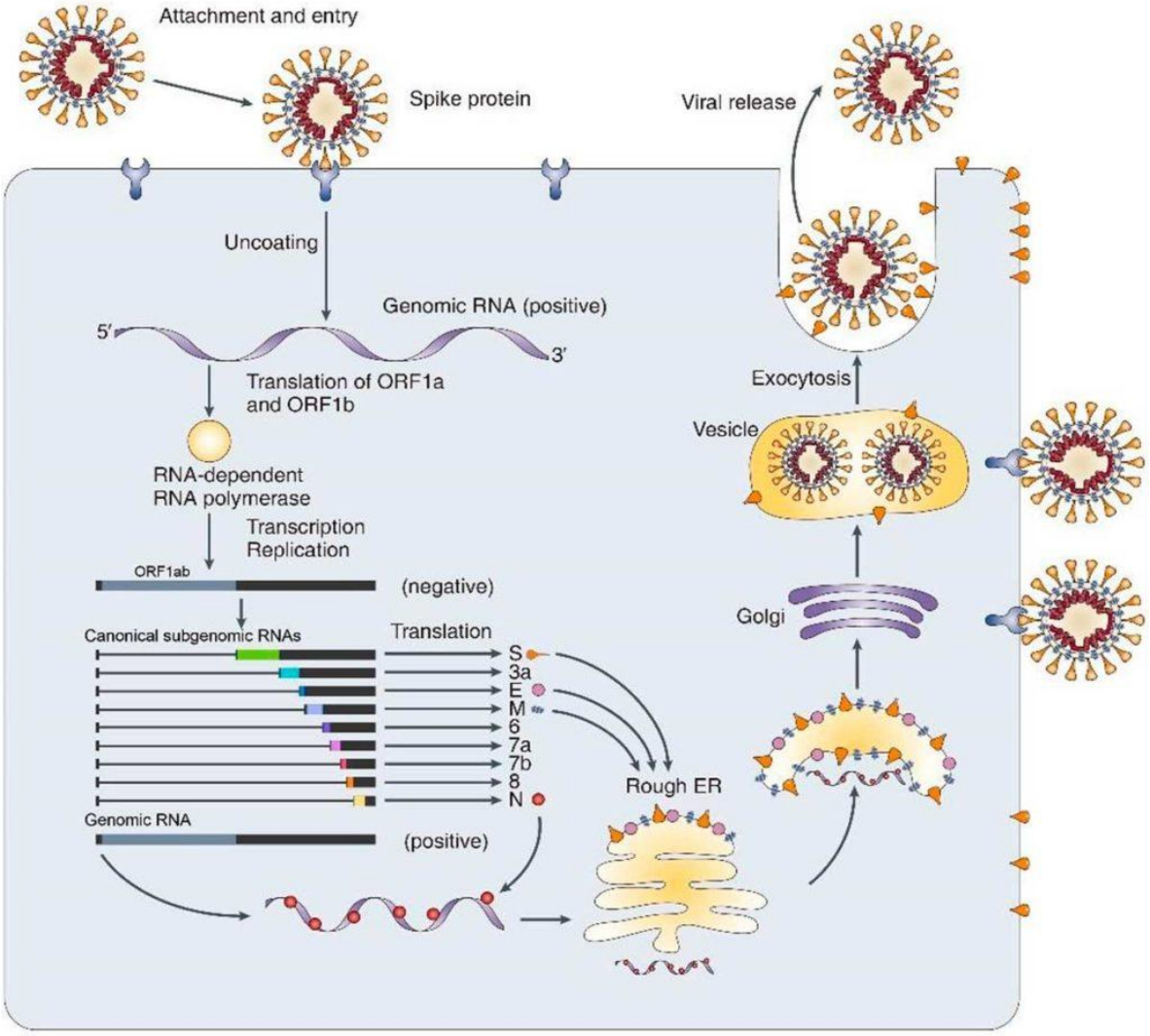


Image 1.1 introduction of SASR-CoV virus into the cell. Source: (18)

Phylogeny of SARS-CoV-2

Pathogens causing infections, especially RNA viruses, mutate rapidly, which can lead them evolutionarily to contain a wide genetic diversity(25). Therefore, a constant analysis of the evolutionary dynamics of these organisms is necessary, allowing them to be monitored and, at the same time, to know their behavior patterns.

It is important to highlight the molecular divergence between SARS-CoV-2 and other coronaviruses, which reveals through a phylogenetic analysis that most of the ORF in different coronaviruses are conserved, but differ in neutral sites, showing that the L lineages (70%) prevail more than the S (30%), but the S is more evolutionarily associated with animal coronaviruses, finding a similarity, between SARS-CoV-2 closest to the RaTG13 coronavirus branch, followed by GD Pangolin SARS-CoV, then by GX Pangolin SARS-CoV, by ZC45 and ZXC21, then by human SARS-CoV giving rise to Bat coronavirus BM48-31, suggesting the

existence of a recombination between the closest ones, which led to the existence of the new SARS-CoV-2 virus (26). At the same time, it is shown that SARS-CoV-2 is a homogeneous population, which apparently evolved from RaTG13 of Bat-CoV, diversified from a common ancestor evolved from pangolin-CoV, being bats virus reservoirs, responsible for zoonotic infections of SARS-CoV-2 and SARS-CoV, giving rise to coronaviruses that evolved through time until today, being this time effective at the moment of entering and infecting humans.

This is corroborated in a study that analyzes the evolutionary mutability of SARS-CoV-2, comparing 9 genomes of pangolin-CoV origin and 3 genomes of bat-CoV origin, finding similarities of 96% with bat-CoV and 85. 98% with the pangolin-CoV, in addition to 12 novel repetitive mutations in viral genomes, 4 of which are unique to Africa and 3 are from South America (14), which indicates that evolution can occur due to unique changes in certain proteins, which can give rise to new coronavirus genomes; therefore, a constant study is necessary not only from organisms but also from the proteins that conform them.

On the other hand, it is known that coronaviruses are highly recombinogenic and an ancestral trait shared with the bat VOC is receptor binding, although it is also known that it is difficult to infer reliable evolutionary histories, due to the high rate of recombination that exists between these viruses, because each part of the genome has different histories(27).

It should be noted that despite the existing variations of SARS-CoV-2, it has a generally high homology among all viral strains of 99.91% in nucleotides and 99.99% in its amino acids and its 13 sites of variability (1a, 1b, S, 3a, M, 8 and N) suggesting selective mutations(28). Although it is important to note that almost 80% of recurrent mutations produce non-synonymous changes, with respect to proteins, particularly the S protein, suggesting a possible ongoing adaptation of SARS-CoV-2" (29)(30)(12).

Through a study of SARS-CoV-2 variants, mutagenic capacity was analyzed to understand factors such as the fidelity of nucleic acid replicating enzymes, the RNA polymerase that allows the virus to adapt through genome variability and the integrity of genetic information, and suggests that the virus is and will continue to evolve, They also suggest that SARS-CoV-2 has a relatively low codon usage bias, which is determined due to natural selection and mutability pressure, whose usage pattern depends on geographic location(31), indicating the importance of knowing the variability and genome integrity of this virus, in relation to its location.

Therefore, in this study, sequences demarcated worldwide by continents are used, which are identified through variants and cities due to the relevance of environmental factors, which can influence its evolution. Intuiting that geo-climatic distribution and other factors resulted in mutations discovered as unique and in high proportion; resulting in heterogeneity worldwide(30), especially if they pass from one country to another through the host.

Rafiul Islam "et al" reveal different deletion sites, in SARS-CoV-2 genes, encoding ORF8 and ORF7a proteins, and a high number of amino acid substitutions, further stating that "receptor binding domain (RBD) residues, showing crucial interactions

with angiotensin-converting enzyme 2 (ACE2) and neutralizing antibody, cross-reacting, were conserved among the virus strains analyzed" (30). In addition, it was identified through a phylogenetic study that the temporal variation in the frequency of coronavirus types was important being the founding effects that gave rise to the A2a subclade, which spread easily around the world, acquiring dominance in all geographical regions(32), these with a non-synonymous variant (D614G), possibly the reason why it can easily remain and infect host cells(33).

Sureshnee Pillay "et al", sustains the importance of phylogenetic analyses to trace the transmission route, especially because of the evidence that in this way, the source of the outbreak could be known and provide lessons to improve infection prevention and control strategies"(2), which is necessary due to its mutation rate so far analyzed, which could have repercussions in high transmissibility or mutagenicity variations.

On the other hand, Leandro N. Jones, cites in his article "When the relationship between spatially coexisting lineages is greater than expected, their distribution is said to be phylogenetically structured"(34) and only until it is determined how structured the variants may be in relation to each other can it be determined whether or not the epidemiological differences between different regions are due to plausible differences. Because some mutations could influence the decrease in peak protein stability, the R408I mutation is relevant because it has a significant influence on the RBD domain and a protein stabilization effect in the SARS-CoV-2 genome from different geographic locations (35).

SARS-CoV-2 variants

Based on phylogenetic analyses from different geographic areas, SARS-CoV-2 is divided into subtypes or strains. The two existing SARS-CoV-2 types that have been of interest and concern and that will be used in the analysis of this study are: Mu, Alpha, Kappa, gamma, Epsilon, Beta, Zeta, Eta, iota, Delta variants worldwide.

All the data recorded so far, give us a knowledge in the scope of existing populations in the world, but at the same time indicate that the strains of the virus are variable, due to a juncture of very uniform mutations along the branches. Results obtained through phylogeny analysis of SARS-CoV2 identified 2 main macro-haplogroups A and B; A affected widely at the international level, while B was more limited to the Asian continent, with 160 sub-branches representative of those originating worldwide(32). This could have implications for the design of vaccines and diagnosis of the virus.

Currently, many variants have been found in the world, but those analyzed in this research were catalogued as of interest and concern at some time (Mu, Alpha, Kappa, gamma, Epsilon, Beta, Zeta, Eta, iota, Delta), because they presented changes in the genome that intervened in the transmissibility, severity of the disease, capacity to escape from the immune system, in addition to their exponential growth of the detected cases or because, besides fulfilling the previous criteria, they showed an increase in virulence, changes in the clinical presentation of the disease, decrease in the efficacy of social and public health measures, being these a risk for the health of the population (36).

The gamma P.1 variant originated in Brazil and was identified in Japan, in people who traveled from Brazil, which contributed to its spread throughout the world. This variant has incurred significant changes with respect to proteins, specifically in the genes that encode the viral spicule, a structure on the surface of the virus that is important for infection and cell entry (37).

Another variant analyzed in this study that presents high distribution worldwide is the Delta B.1.617.2 variant that was identified for the first time in India in October 2020 and today is present in 92 countries including Colombia, but it is inferred that to a lesser extent than in other countries of the world, due to the scarce data that exists in our territory. Currently Delta variant is classified as of interest and concern as well as Beta variant (South African variant), Alpha (British variant) and Gamma (Brazilian variant) strains introduced in this study.

Therefore, it should be noted that studies have shown that vaccines approved by the EMA (European Medicines Agency), such as Pfizer, AstraZeneca by Johnson & Johnson and Moderna, showed an immune response against the Delta variant and other prevalent ones, as long as full doses are received. Therefore, to date we infer the importance of vaccination, to stop the evolution of the virus and in turn, the emergence of new strains with high virulence capacity in the world. In addition, it is necessary to emphasize that this is under constant and exhaustive monitoring as a preventive measure(38).

The beta variant B.1.351, appeared in South Africa in May 2020 and was identified as a variant that commonly affects young people with no history of disease and has similarities to the Alpha variant, but has additional mutations in the peak protein and raises concerns because it is believed that it can develop resistance to vaccines(39).
The Mu B.1.621 variant first appeared in Colombia in January and to date has been reported in 39 countries. According to WHO, the variant has mutations with immune escape capacity(36).
The Alpha B.1.1.7 variant, first found in the United Kingdom in September 2020, was the first strain to be classified as a variant of concern, 43% to 90% more contagious and causing a high number of deaths worldwide according to the University of Exeter(39).
The Iota B.1.526 variant appeared in the United States at the end of November 2020, and the lineage of 25% of the sequences were in New York, during February 2021(40); for its part, the Eta B.1.525 variant appeared for the first time in the United States at the end of November 2020. .525 variant was first detected in Nigeria and spread throughout Europe, the USA and Canada. And the kappa variant, lineage B.1.617.1, was first detected in India and is associated with increased transmissibility and immune escape according to CCAES (Center for Coordination of Alerts and Health Emergencies).
The Epsilon and Zeta variants of interest according to WHO until 2020 were also used in this study. The first originated in the USA in March 2020 and the second originated in Brazil in April 2020.

Objective:
Materials and Methods:
Materials and Methods

Type of study
This research is descriptive and observational, because it measures the evolution and describes the phylogeny of the SARS-CoV-2 virus strains reported, as well as the variability and the existing plausibility of the virus circulating in the locality, using relevant hypotheses in this study, without seeking a causal effect of the results obtained; It is cross-sectional because we used sequence data acquired over a period of time and it is retrospective because it helps us to formulate hypotheses through the analysis of the results acquired in relation to previous studies of the disease, in this case SARS-CoV-2.

Research method
Sequence representation, acquisition and alignment
The nucleotide sequences that were used in this study were obtained from complete genomes of the SARS-CoV-2 virus, classified by strains, through the GenBank database of the NCBI(41) http://www.ncbi.nlm.nih.gov/genbank/. In addition, reference proteins were downloaded from one of the first sequenced genomes available on the VIPR platform (42) https://www.viprbrc.org/brc/home.spg?decorator=corona_ncov, whose access code is MT019529. All sequences were saved in fasta format.
A total of 460 sequences were collected, of which 60 from South America, 202 from North America, 72 from Europe, 79 from Asia, 11 from Oceania, 27 from Africa and 1 of the first sequences in Wuhan (China). After collection, the data were cleaned and the different genomes of the virus strains were aligned by continents, using the MAFFT version 7(43) program, and the nucleotides found at the ends of the alignment, which did not encode the reference protein, were subsequently eliminated.

Obtaining evolutionary models
To calculate the evolutionary models for the protein-coding genes, S, M, N, E, Orf3a, Orf6, Orf7a, Orf8, Orf10, a total of 9 matrices were analyzed. Each matrix contained a single gene and each contained the species that exist per continent, with a total of 10 strains evaluated as strains of interest or concern. For each matrix, the best-fit surrogate model was calculated using the ModelTest-NG program in XSEDE (44), available on the CIPRÉS SCIENCE GATEWAY(44) platform version 3.3 https://www.phylo.org/portal2/login!input.action and the AIC criterion was used for model selection.

Remodeling and phylogenetic analysis
The phylogenetic study was reconstructed through topology by Bayesian inference(45), for which the aligned matrices in fasta format were converted to NEXUS formats, to be compatible with the MrBayes tool in XSEDE 3. 2(44), a

Metropolis-coupled Márkov chain Monte Carlo (MCMCMC) (46) was simulated with a number of 107 generations and a screen print every 1000 generations, the evolutionary model selected by the ModelTest-NG program in XSEDE(44) was used for each gene. The resulting IB trees were evaluated taking into account the posterior probability and phylograms provided by the MrBayes 3.2(44) program. They were observed and edited, in the FigTree v1.4.4 graphics program, in *.tree format.

Population and Sample

Inclusion Criteria

A total of 460 sequences were chosen from 10 strains of interest and concern in the 6 continents of the world, of which 68 belong to South America, being 14 from Colombia of delta, gamma, Alpha and mu strains.

Exclusion Criteria.

Genomes that were not complete and sequences with ambiguous characters were excluded.

Results:

Genome organization

The SARS-CoV-2 genome has a structure composed of 4 structural proteins, the S (spicule), the E (envelope), the M (membrane) and the N (nucleocapsid) of which the S gene contains the most nucleotides; In addition, a series of accessory sequences or proteins (ORF3a, ORF7a ORF8, Orf6, orf10) were analyzed, which were taken from the VIPR database, whose accession number in the GENBANK platform is MN988668.

Collection of gene sequences

The analyzed sequences are in total 460 from the observed regions (image 1) of the strains of interest and concern Alpha, Mu, Iota, Eta, Zeta, kappa, Delta, Beta, Epsilon, Gamma and one unknown, belonging to 40 countries found in the 6 continents of the world, information represented in (image 2 and 3) previously manipulated and chosen, all freely accessible and downloaded through the GenBank and VIPR database (47).

Tables and figures

Image 2. Distribution of sequences in the world. Source: own database.

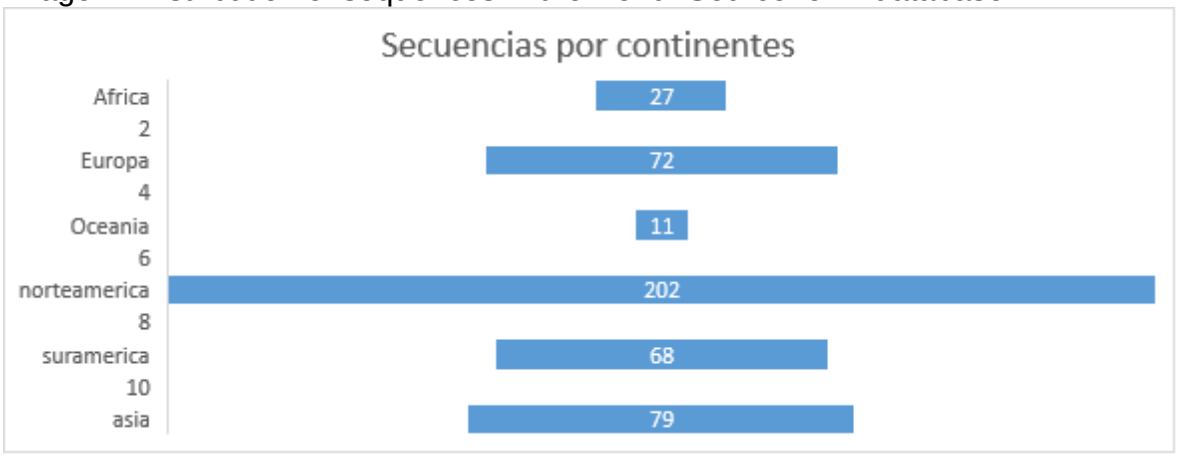


Image 3. Distribution of sequences by continents. Source: own database.

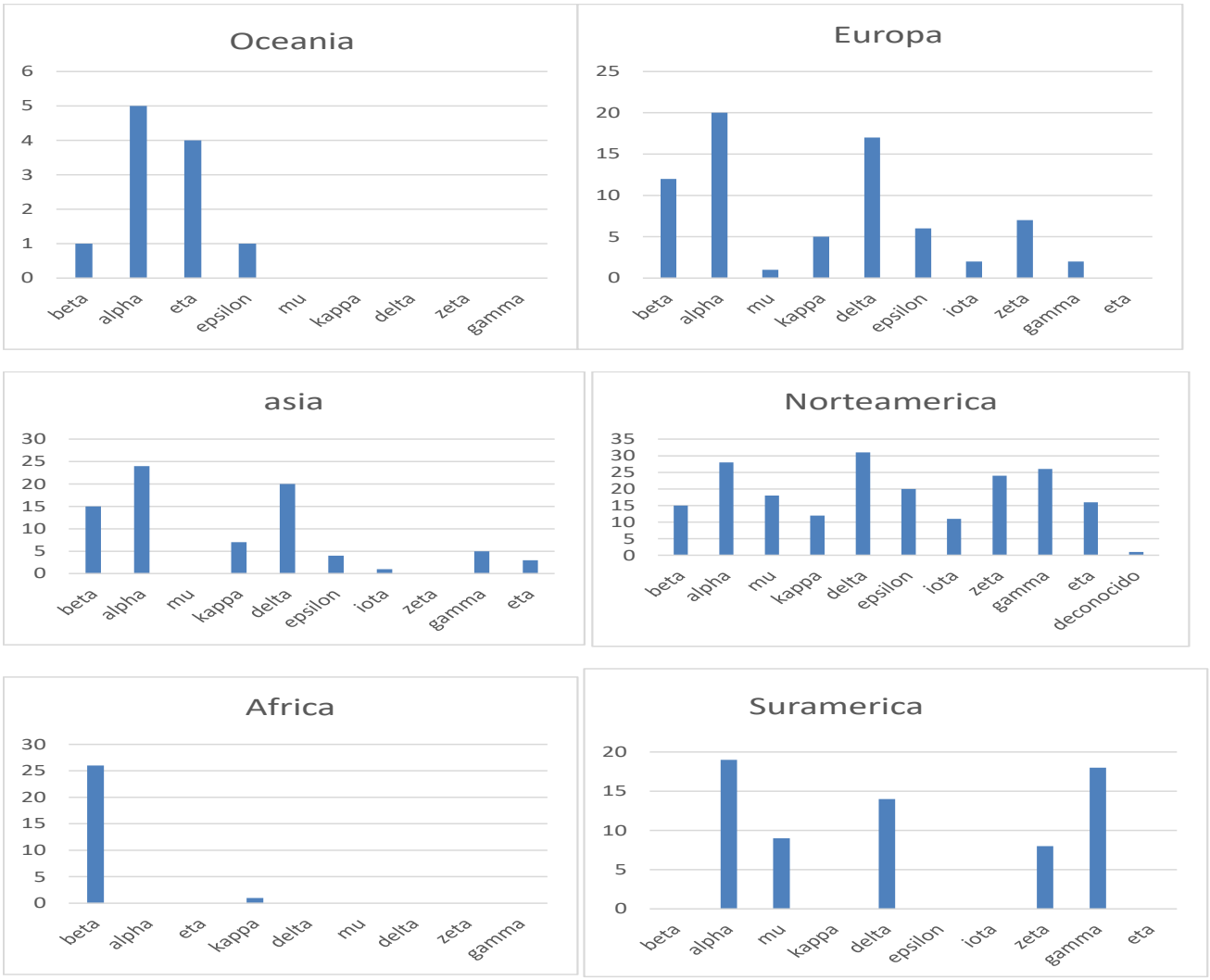


Image 4. Distribution of strains by continent. Source: own database.

Protein S phylogeny of Iota, Eta, Beta, Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta variants. Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta.

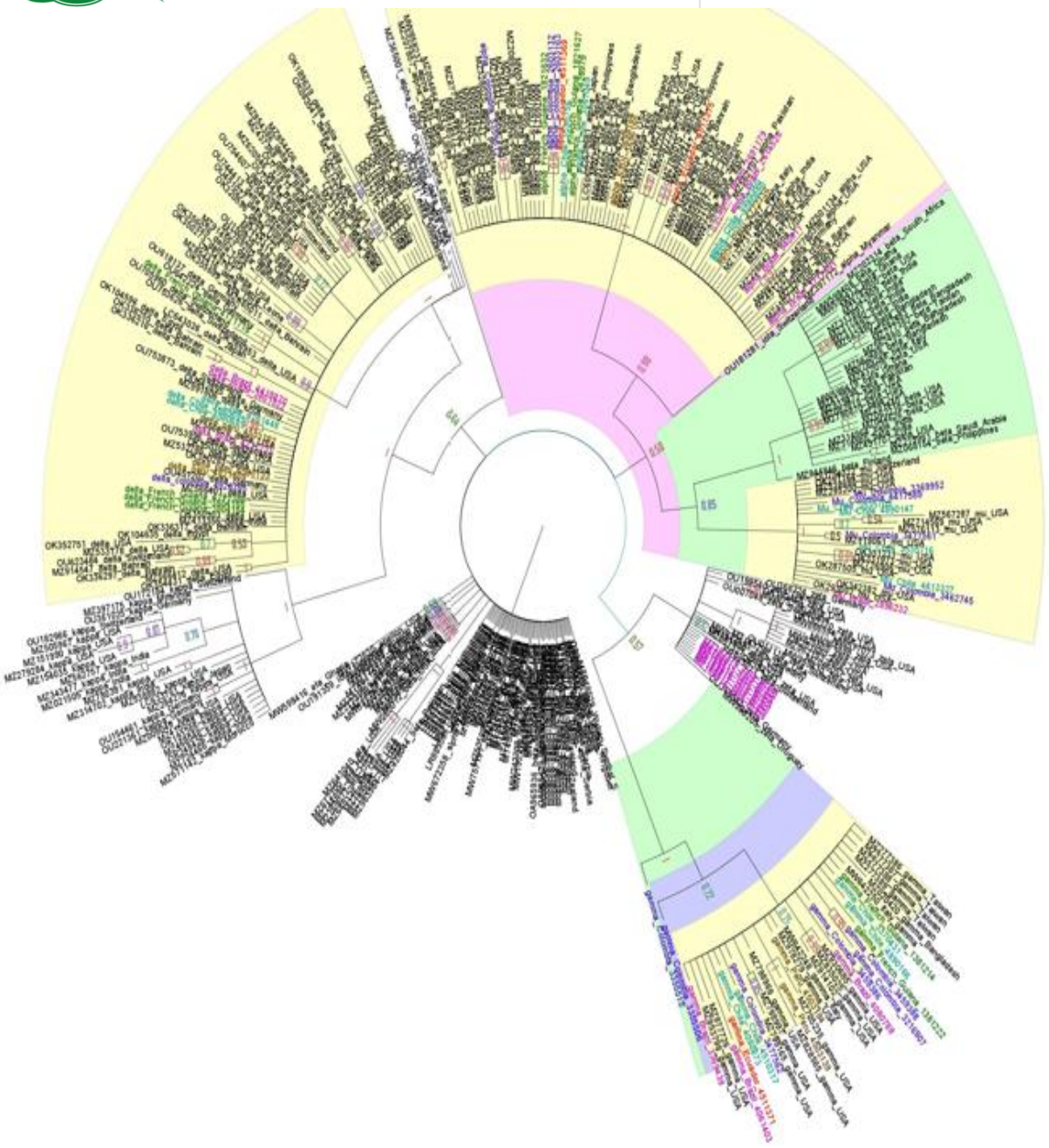


Image 49. Phylogeny of protein S. Source: own database.
Phylogenetic tree obtained using the Bayesian Inference method for gene S. The numbers correspond to Bayesian posterior probability values. Names in blue highlighted in beige correspond to the location of strains from Colombia in relation to strains from the rest of the world. The evolutionary models for the canonical positions of the S gene were TIM2 and GTR.

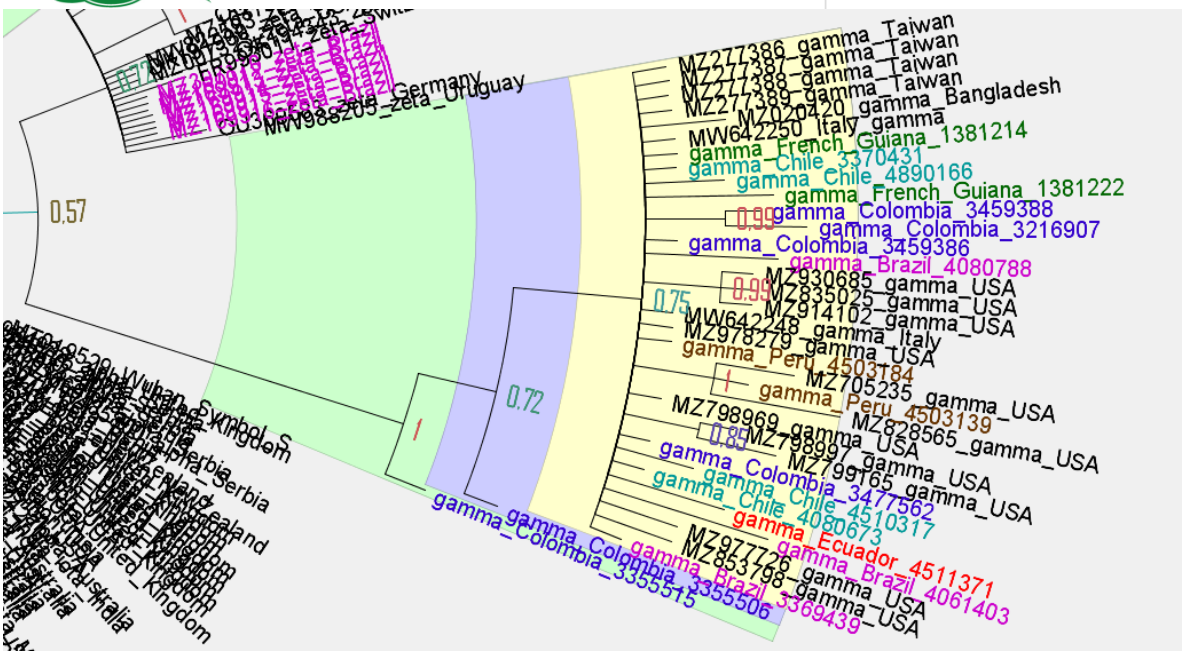


Image 6 Phylogenetic analysis of the gamma strain for protein S. Source: own database.

The Bayesian analysis for protein S presents a monophyletic group, with Bayesian posterior probabilities of 0.75, among the gamma strains from Colombia, Chile, Ecuador, Brazil, Peru, SA, Italy, Bangladesh, French Guyana and Taiwan present homology among them.

There are some gamma strains from Colombia with polytomies, but they have a node with a posterior ratio of 1 relative to the other gamma strains in the world.

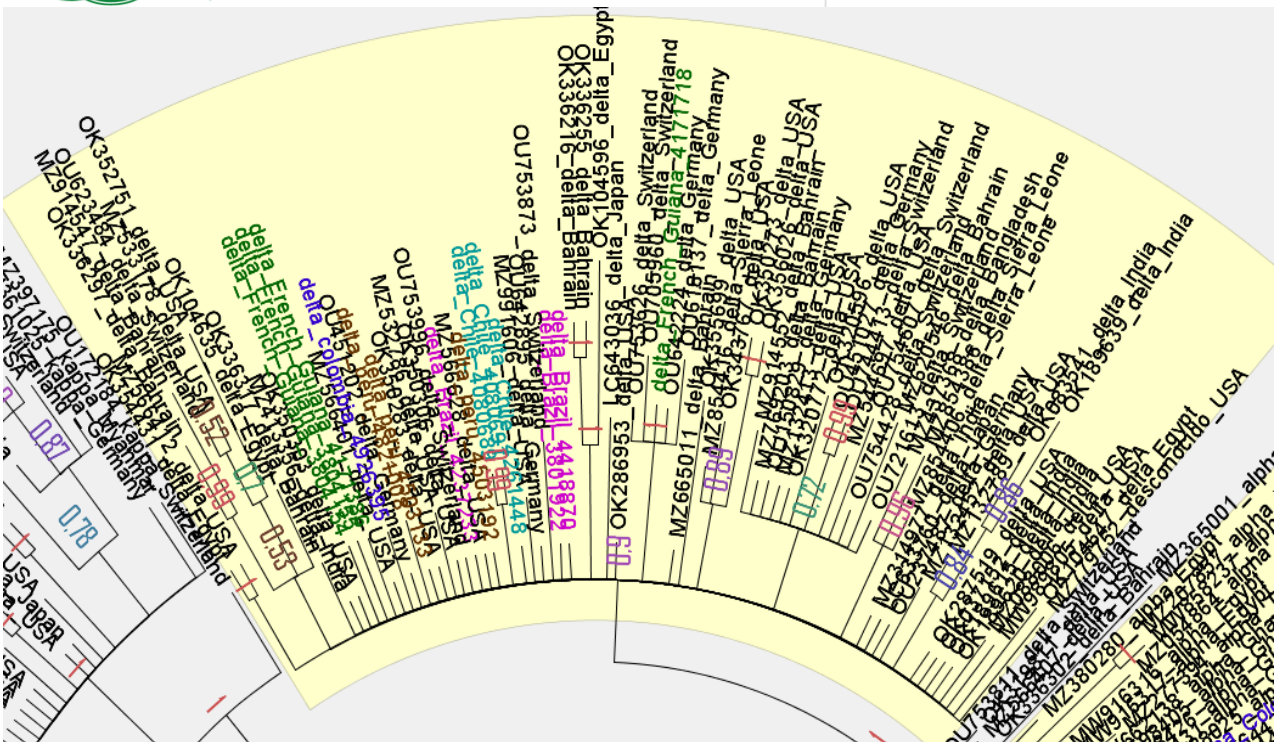


Image 7 Phylogenetic analysis of the delta strain for protein S (own database).

The Bayesian analysis for protein S shows a monophyletic group with Bayesian posterior probabilities of 0.9, among the delta strains from Colombia, Chile, Peru, Ecuador, Brazil, Switzerland, Egypt, Bahrain, Germany, Japan, India, Peru, USA, Italy, Bangladesh, French Guyana and Taiwan showing homology among them.

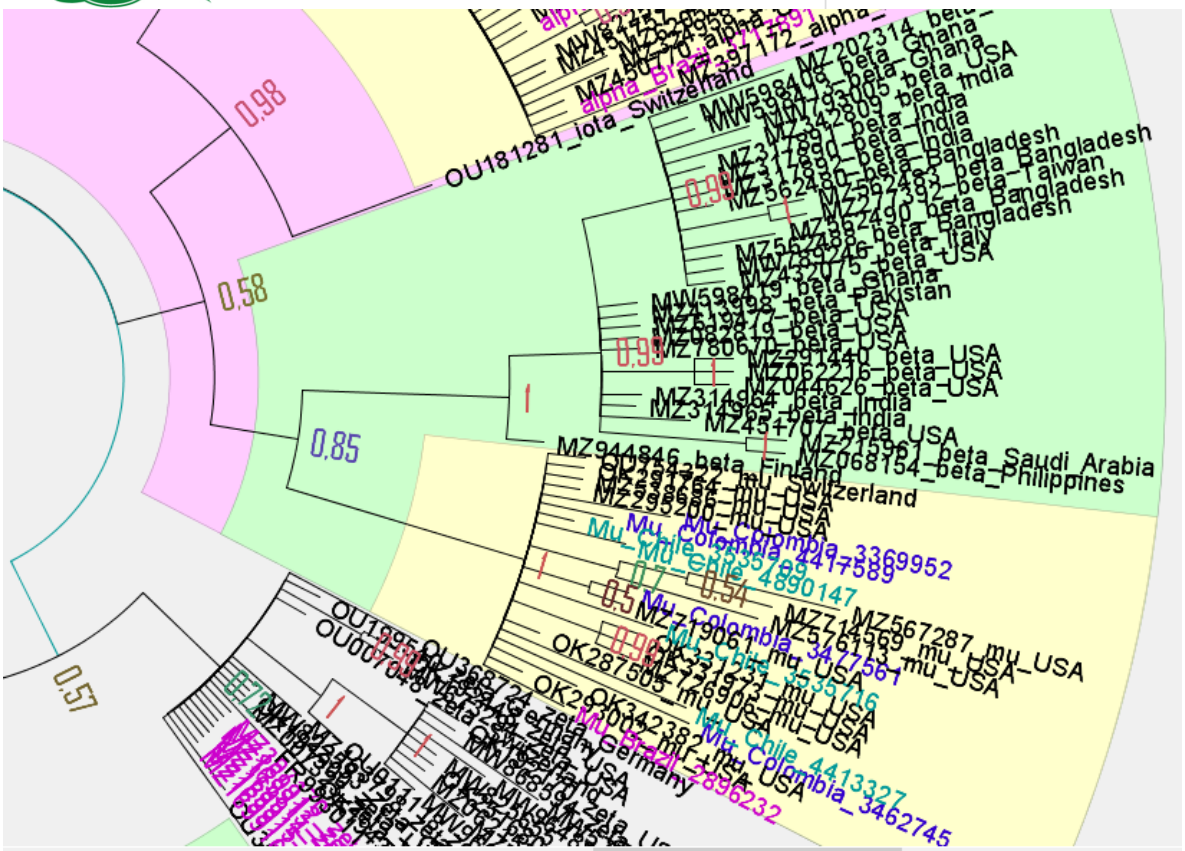


Image 8 phylogenetic analysis of the mu strain for protein S. Source: own database.

The Bayesian analysis for protein S shows a monophyletic group with Bayesian posterior probabilities of 1 among the mu strains from Colombia, Chile, Brazil, Switzerland, USA, which present homology among them. The MU strains from South America and USA; and the beta variants from Asian and European and North American countries, come from a common ancestor, with a Bayesian posterior probability of 0.85. In addition, there is a relationship between the in-groups of the mu, beta, iota and Alpha variants with a posterior probability of 0.58.

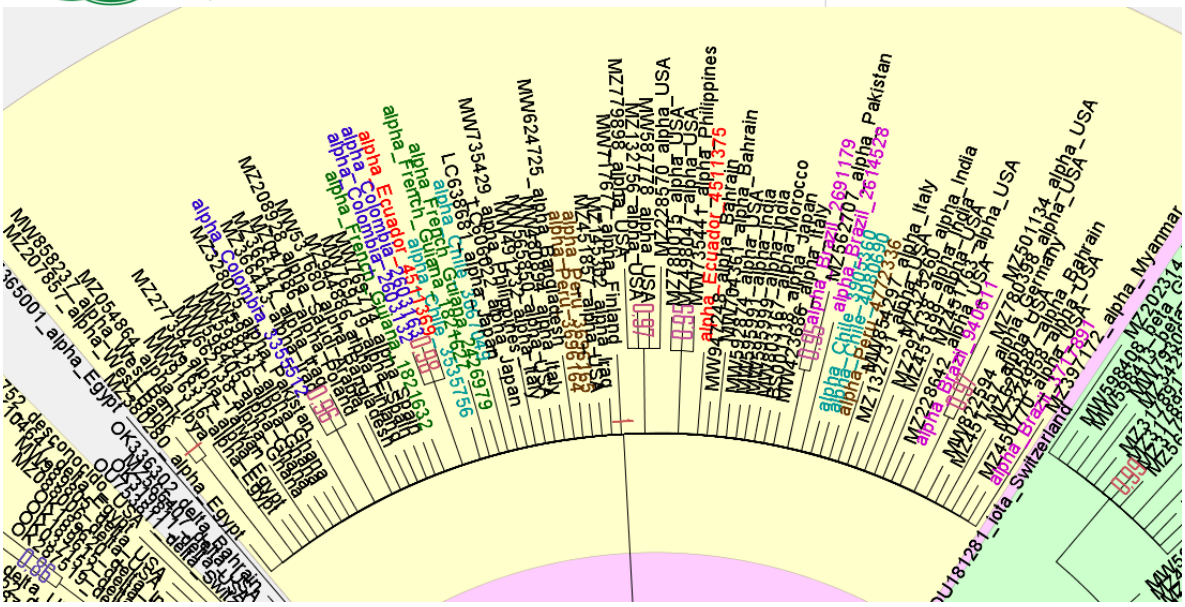


Figure 9 Phylogenetic analysis of the Alpha strain for protein S. Source: own database.

The Bayesian analysis for protein S presents a monophyletic group with Bayesian posterior probabilities of 1 between Alpha strains from Colombia and those from the rest of the world.

Protein S presents 3 internal groups and an external group: an internal group where the delta variants are related to the Kappa variants of the world, including those of Colombia, with Bayesian posterior probabilities of 0.64. In addition, there is a paraphyletic group, where Mu, Alpha and beta groups are related, where the node that groups these three variants has a posterior probability of 0.58 and where beta is more related to mu, with a posterior probability of 0.85.

There is also another internal group that relates the zeta and gamma variants of Colombia and the world with a posterior probability of 0.57, the lowest among the internal groups, where the gamma variants are more related to each other with a posterior probability equal to 1.

There are some variations among the Alpha strains because the variants of Australia, New Zealand, Netherlands, United Kingdom are in an external group and are not within the group of the other Alpha variants and are more associated to the proteins of the iota, beta, epsilon variants and to the S protein that appeared in Wuhan in its beginnings.

ORF10 protein phylogeny of the Iota, Eta, Beta, Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta variants. Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta.

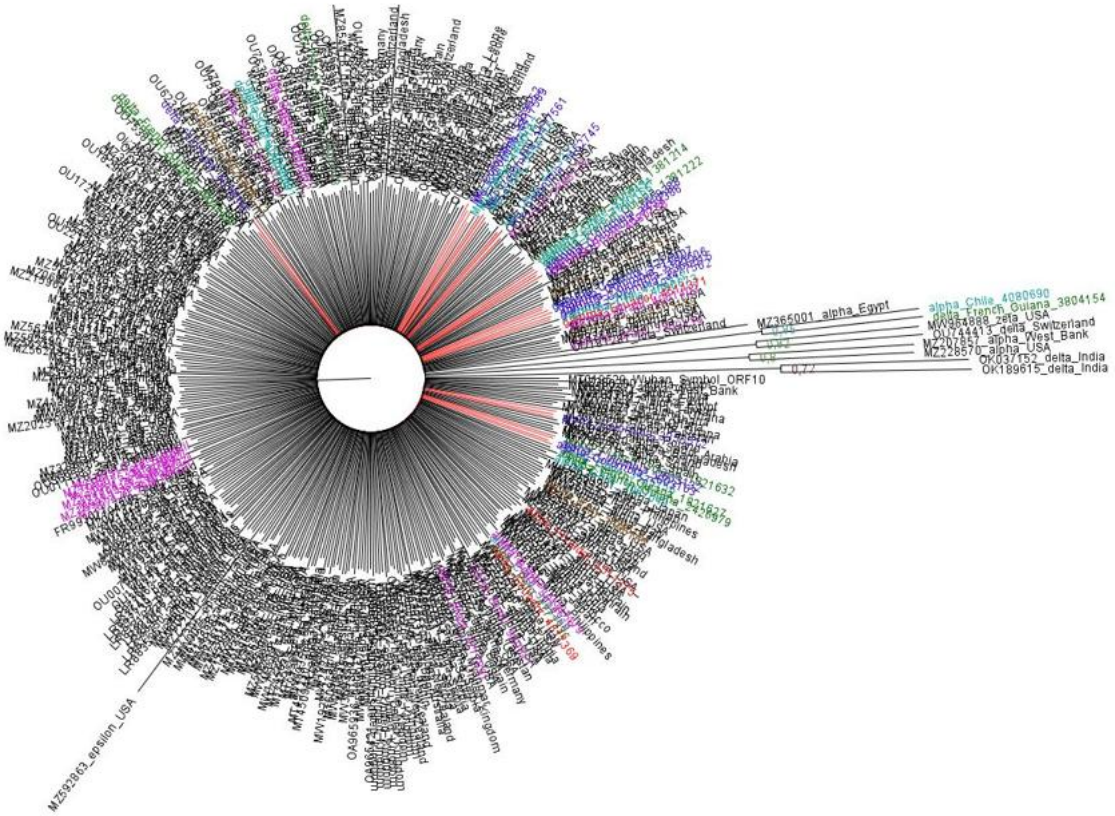


Image 10 phylogenetic analysis for the Orf10 protein. Source: own database.

Phylogenetic tree obtained, using the Bayesian Inference method, for the ORF10 gene. The names in blue correspond to the location of the sequences from Colombia in relation to the strains from the rest of the world. The evolutionary models of the canonical positions of the Orf10 gene were TIM2 and GTR.

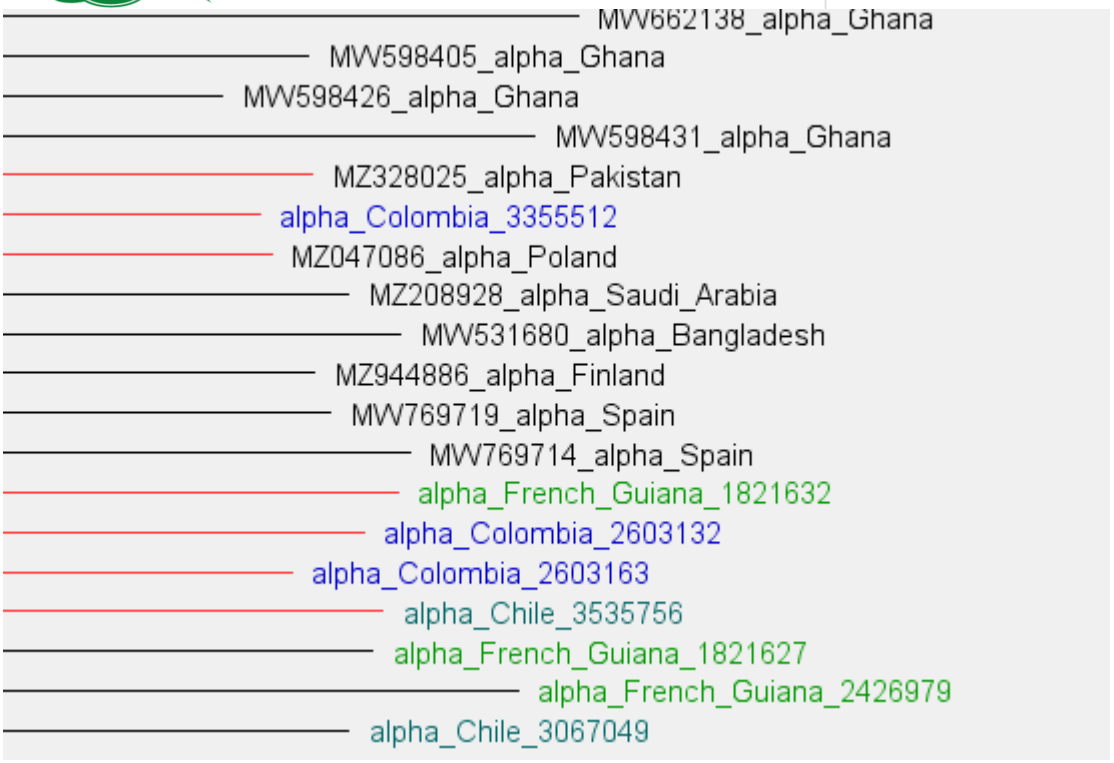


Image 11 phylogenetic analysis of the Alpha strain for the Orf10 protein. Source: own database.

The Bayesian analysis for the Orf10 protein shows a monophyletic group between the Alpha strains from Colombia and the rest of the world.

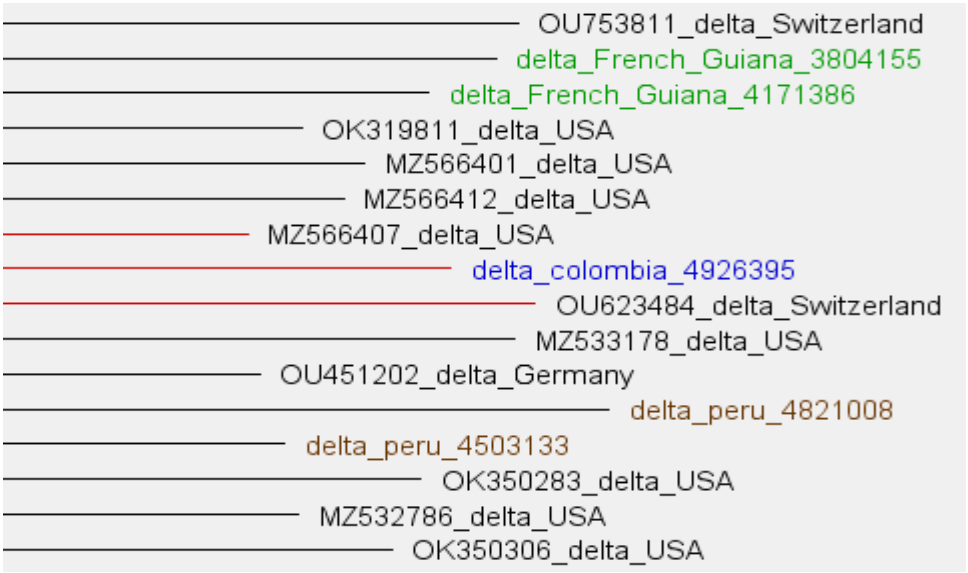


Image 12 Phylogenetic analysis of the delta strain for the Orf10 protein. Source: own database.

The Bayesian analysis for the Orf10 protein presents a monophyletic group between the delta strains from Colombia and the rest of the world.

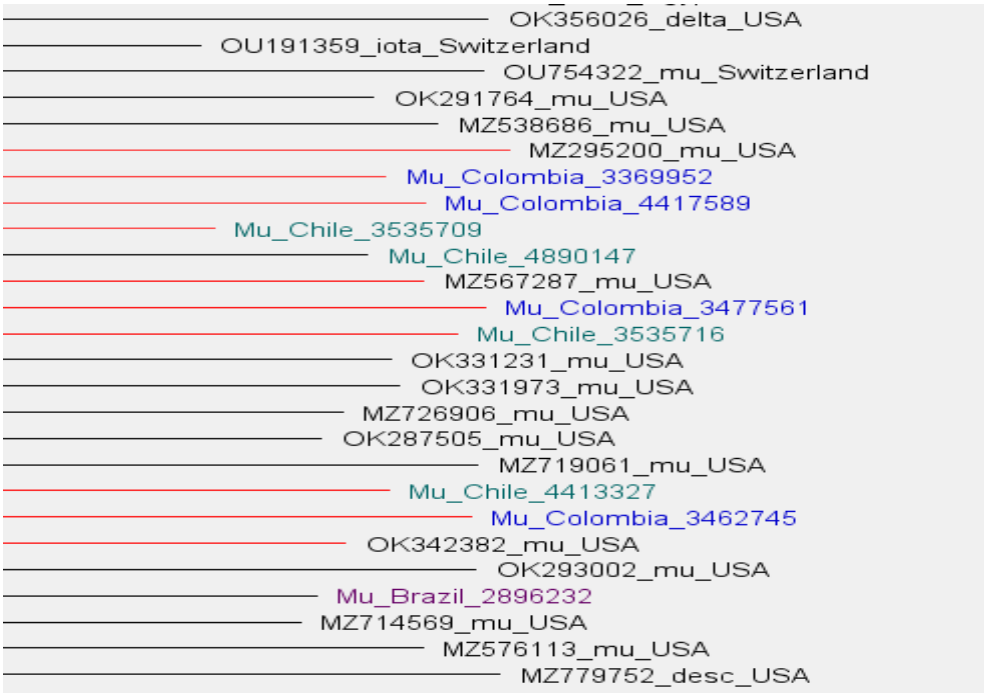


Image 13 Phylogenetic analysis of the Mu strain for the Orf10 protein. Source: own database.

The Bayesian analysis for the Orf10 protein shows a monophyletic group between the Mu strains from Colombia and the rest of the world.

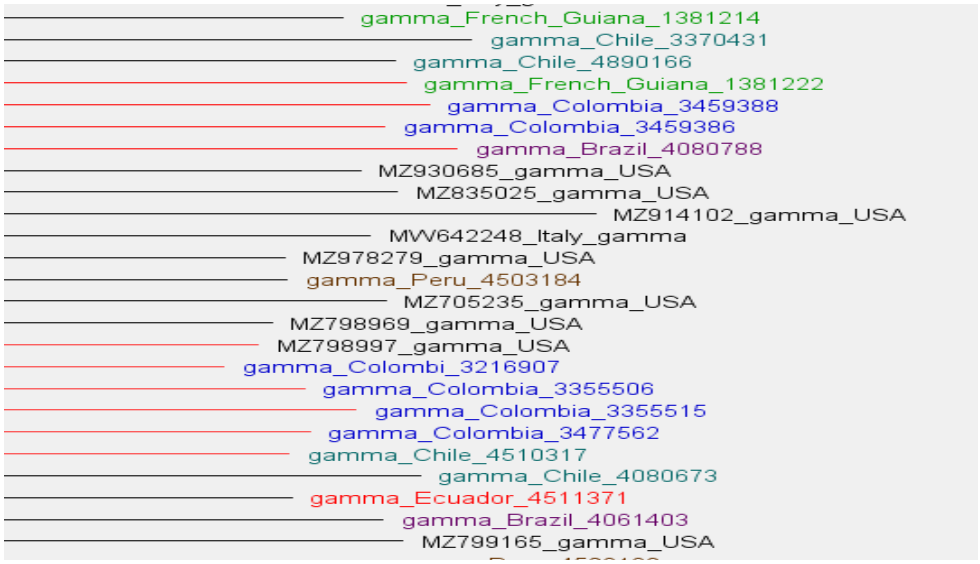


Figure 14 Phylogenetic analysis of the gamma strain for the Orf10 protein.

The Bayesian analysis for the Orf10 protein presents a monophyletic group between the gamma strains from Colombia and the rest of the world.

In conclusion, the Orf10 protein of SARS-CoV-2 is highly conserved and has been maintained by evolution despite the characterization of existing variants in the world, perhaps which may be related to a lesser importance of the protein for SARS-CoV-2 infection(48).

ORF8 protein phylogeny of Iota, Eta, Beta, Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta variants. Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta.

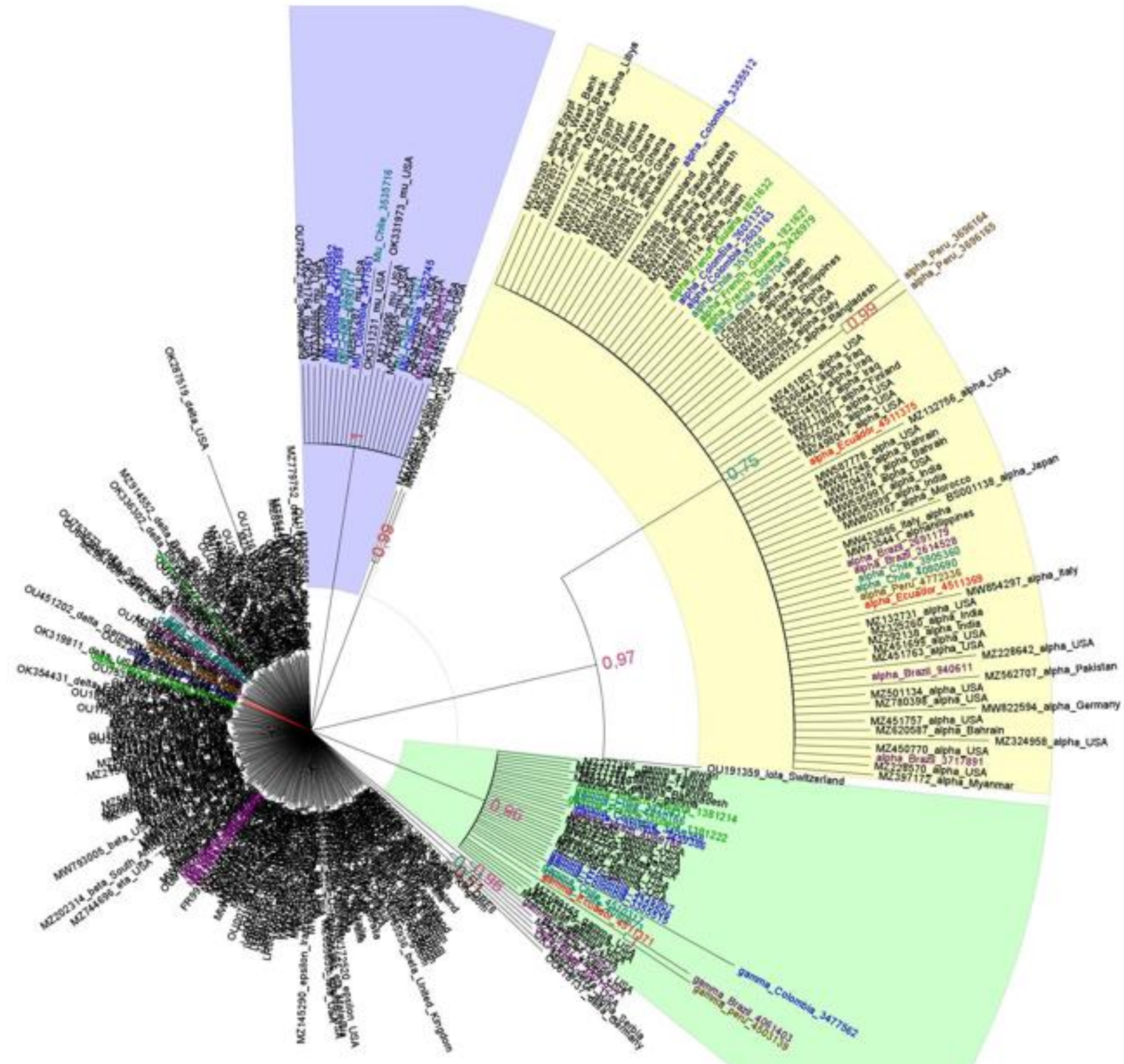


Image 15 Phylogenetic analysis of the delta strain for the Orf8 protein (own database).

Phylogenetic tree obtained, using the Bayesian Inference method, for the ORF8 gene. Numbers correspond to Bayesian posterior probability values. Names highlighted in yellow blue and green correspond to the location of strains from Colombia in relation to strains from the rest of the world.
The evolutionary models of the canonical positions of the Orf8 gene were, TPM1uf and TVM. In addition, what Rafiul islam says about the polymorphisms in some sites of the gene is confirmed, which indicates that it can be due to deletions and substitutions in the protein.

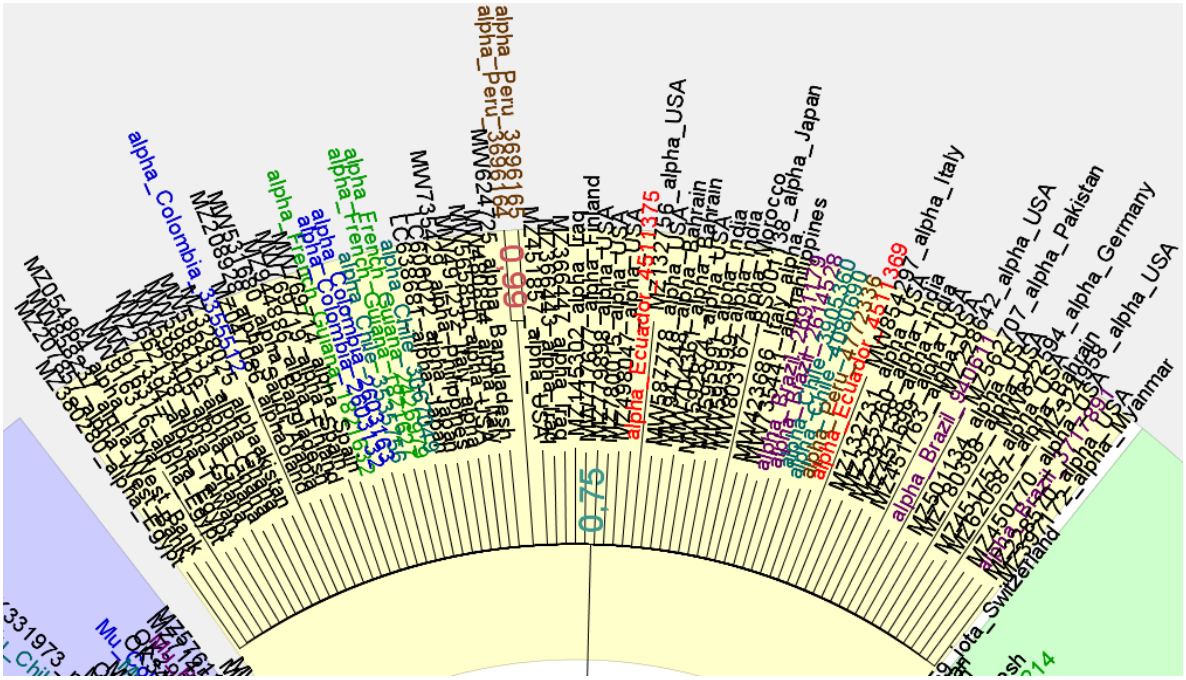


Image 16 Phylogenetic analysis of the Alpha strain for the Orf8 protein. Source: own database.

The Bayesian analysis for the Orf8 protein presents a monophyletic group with Bayesian posterior probabilities of 0.75 in the mu strains from Colombia.
Chile, Brazil, Ecuador, Switzerland, Egypt, Bahrain, Germany, Japan, India, Peru, USA, Italy, Bangladesh, Philippines, Pakistan, West Bank, Morroco, Finland, French Guyana and Taiwan, Myanmar and USA present homology in relation to the Colombian sequences.

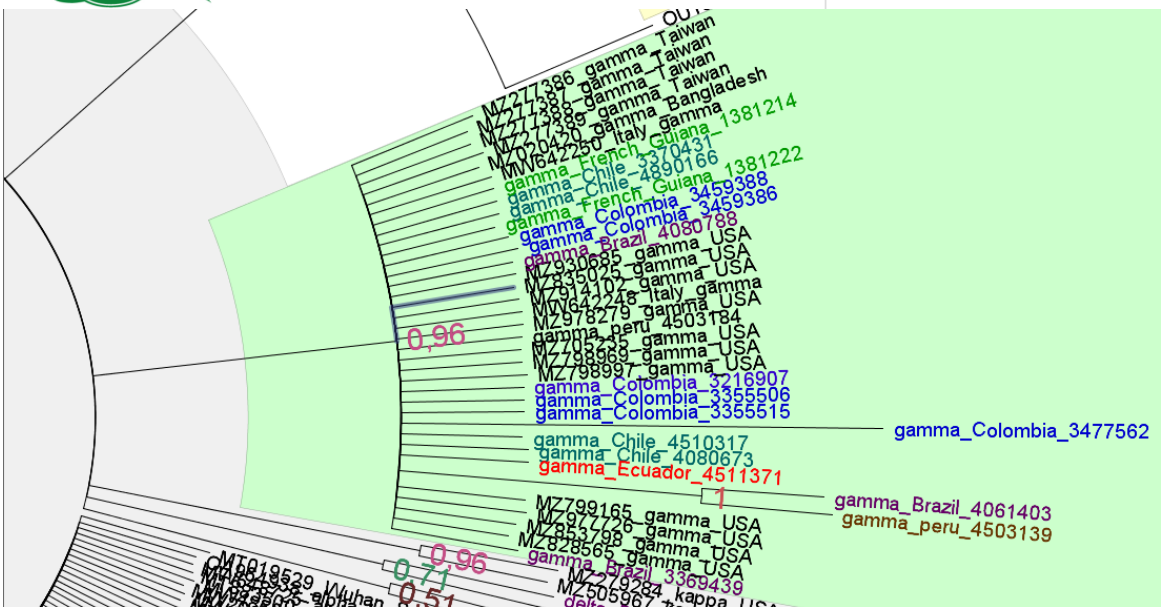


Image 17 Phylogenetic analysis of the gamma strain for the Orf8 protein. Source: own database.

Bayesian analysis for the Orf8 protein presents a monophyletic group with Bayesian posterior probabilities of 0.96 in the gamma strains.
Chile, Brazil, Ecuador, Swizerland, Peru, USA, Italy, Finland, Taiwan, and USA present homology in relation to the Colombian sequences.

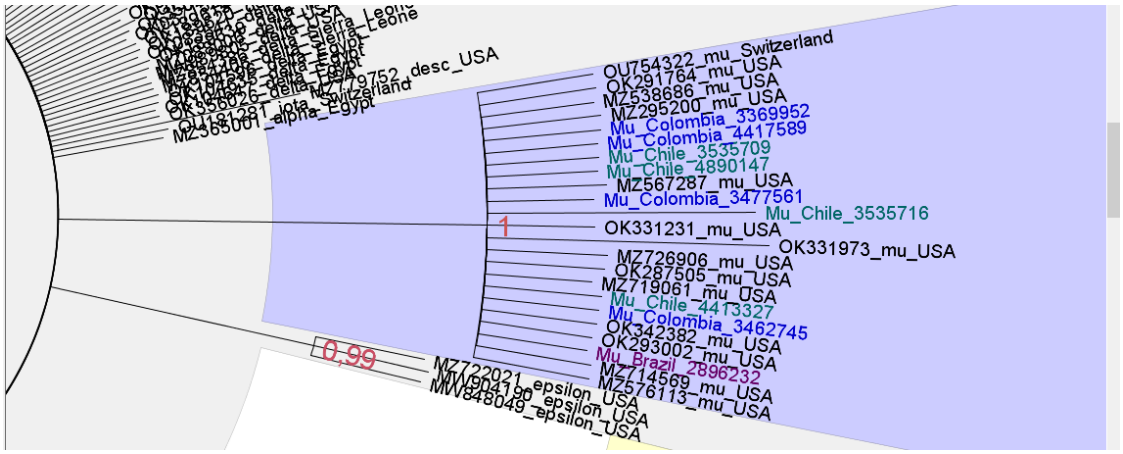


Image 18 Phylogenetic analysis of the Mu strain for the Orf8 protein Source: own database.

The Bayesian analysis for the Orf8 protein presents a monophyletic group with Bayesian posterior probabilities of 1 in the Mu strains.
Chile, Brazil, Ecuador, Switzerland, and USA present homology of the 0rf8 gene, in relation to the Colombian sequences.

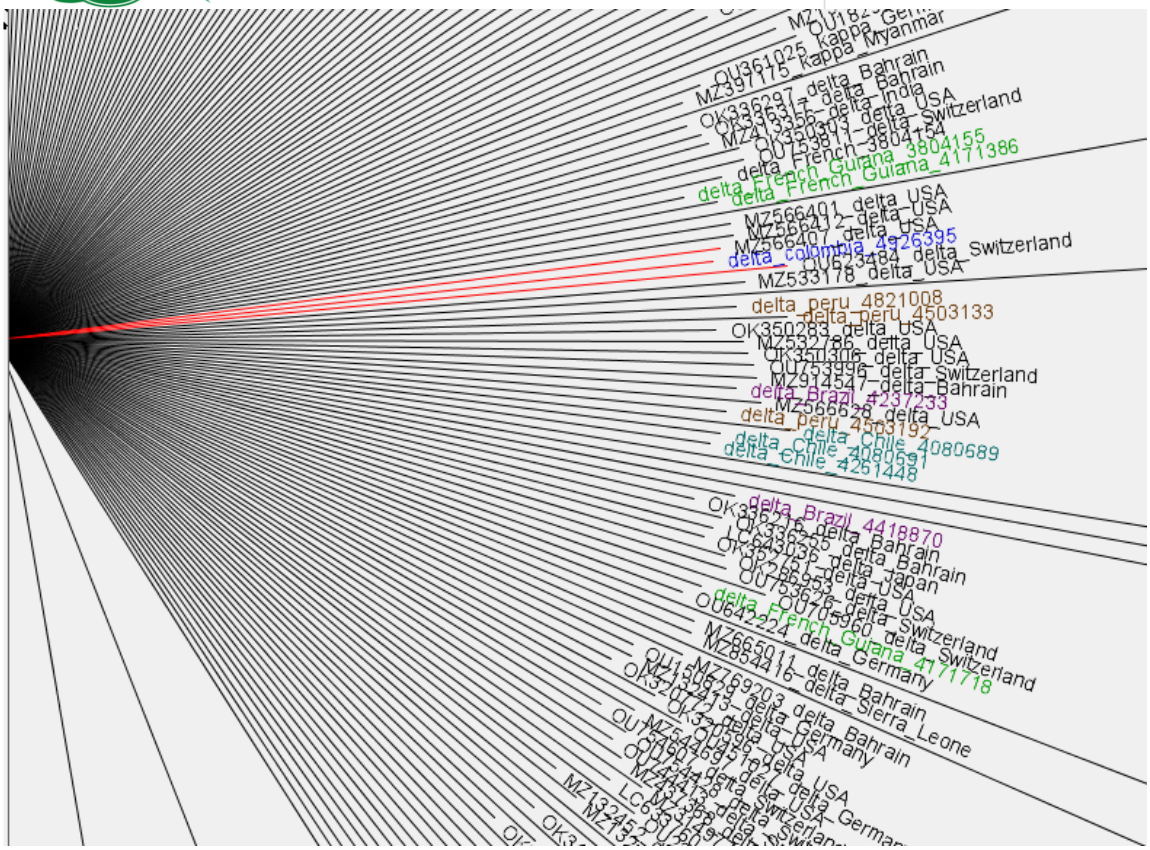


Image 19 Phylogenetic analysis of the gamma strain for the Orf8 protein Source: own database.

The Bayesian analysis for the Orf8 protein shows a conserved group in the delta strains of Colombia and the rest of the world.

In conclusion, through the Bayesian inference analysis, it was determined that the Orf8 protein, in the different strains studied, has 3 internal groups and an external group; where an internal group belongs to the gamma variants of Colombia and the world with a posterior probability of 0.96, another belongs to the mu variant of Colombia and the rest of the world, with a posterior probability of 1 and another group where the Alpha variant is found, with a posterior probability of 0.97, but with a variability and variability of 0. 97, but with a variability and conservation in terms of the orf8 protein, belonging to the variants of Australia, New Zealand, Netherlands, United Kingdom and Egypt which are in an external group and are not within the group of the other Alpha variants, being found more associated to the proteins of the delta variants of Colombia and the world and the iota, beta, epsilon strains and the first strain that appeared in Wuhan in its beginnings.

ORF7a protein phylogeny of Iota, Eta, Beta, Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta variants. Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta.

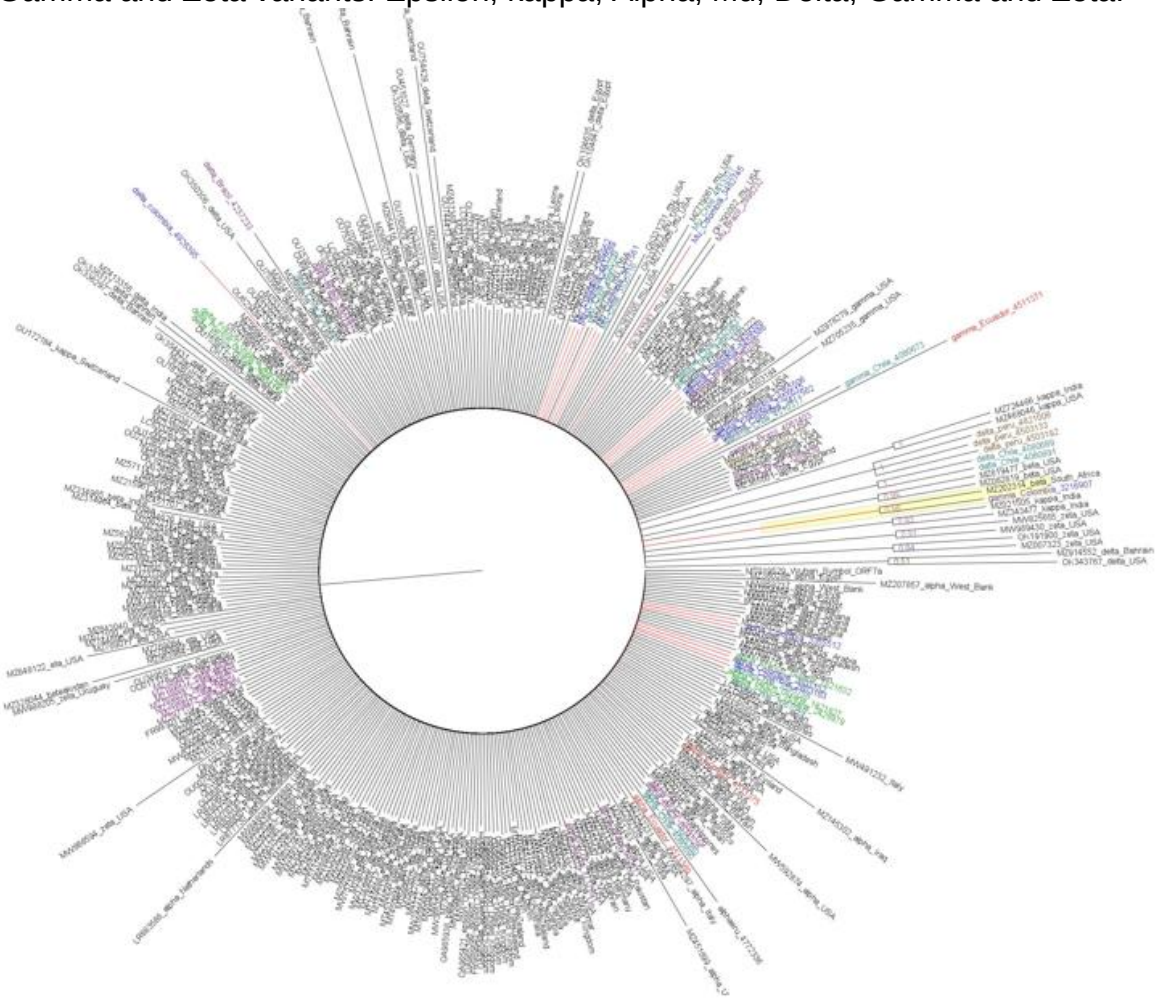


Image 20 Phylogenetic analysis for the Orf7a protein. Source: own database.

Phylogenetic tree obtained using the Bayesian Inference method for the ORF7a gene. The names in blue correspond to the location of the strains from Colombia, in relation to the strains from the rest of the world. The evolutionary models of the canonical positions of the Orf7a gene were TPM1uf and GTR.

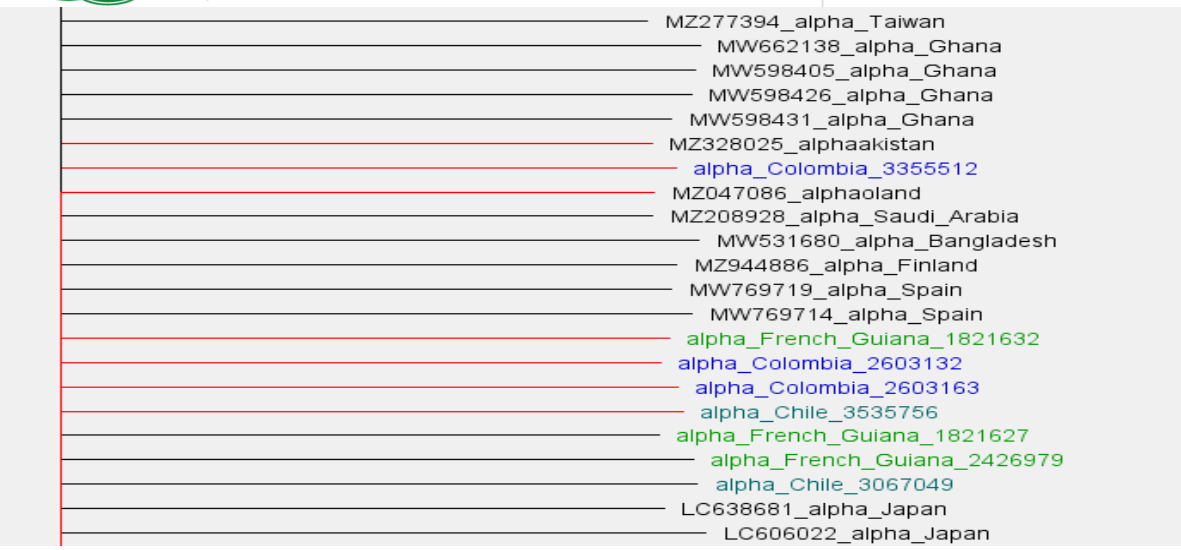


Image 21 Phylogenetic analysis of the Alpha strain for the Orf7a protein. Source: own database.

The Bayesian analysis for the Orf7a protein presents conserved biological sequences of the Alpha strains from Colombia in relation to those from the rest of the world.

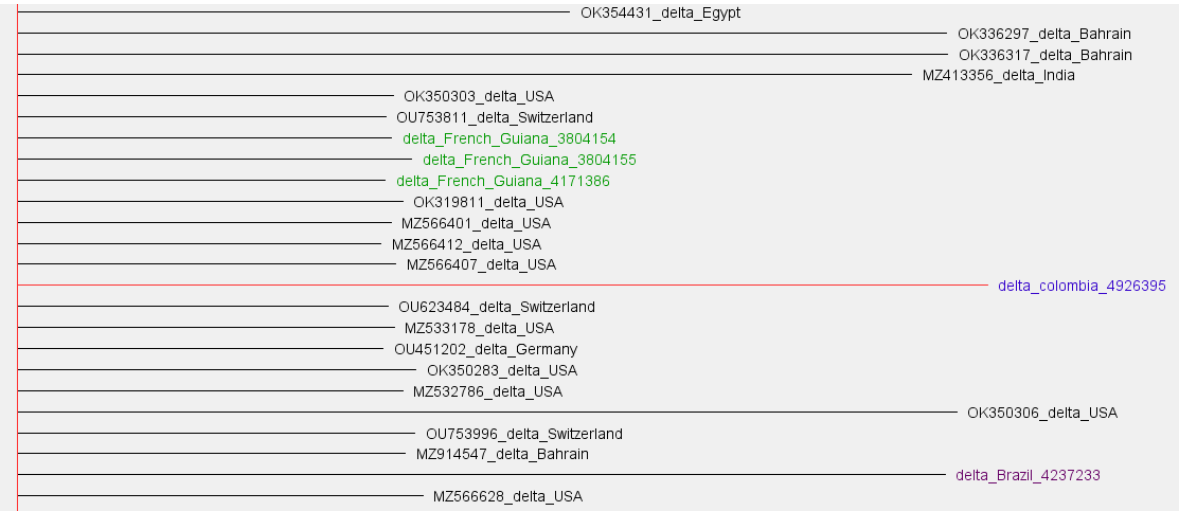


Image 22 Phylogenetic analysis of the delta strain for the Orf7a protein. Source: own database.

The Bayesian analysis for the Orf7a protein presents conserved biological sequences of the delta strains from Colombia in relation to those from the rest of the world.

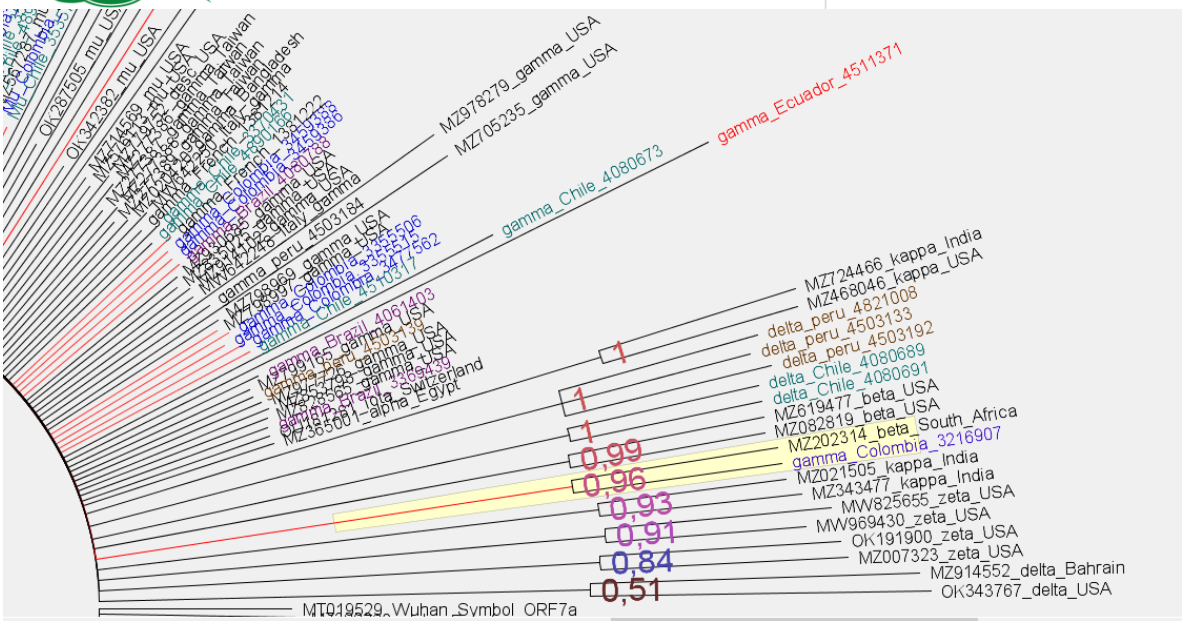


Image 23 Phylogenetic analysis of the gamma strain for the Orf7a protein: Source: own database.

The Bayesian analysis for the Orf7a protein presents conserved biological sequences of the gamma strains from Colombia in relation to the rest of the world. All gamma strains present homology in this protein in relation to the Colombian sequences. There is a close relationship between the gamma strains from Colombia and beta from South Africa with Bayesian posterior probabilities of 0.96.

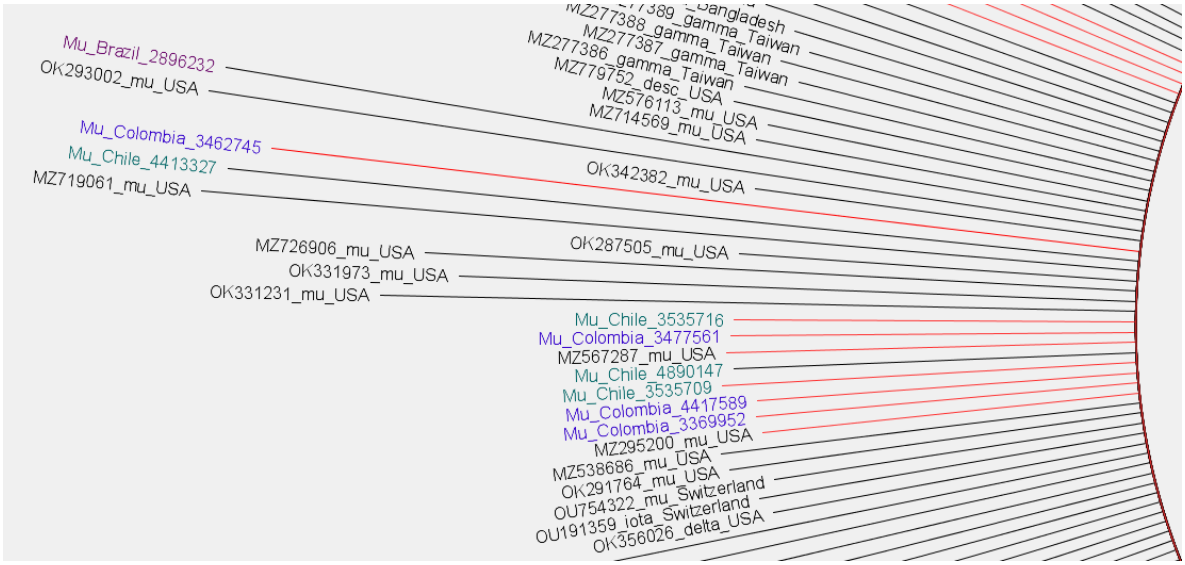


Image 24 Phylogenetic analysis of the Mu strain for the Orf7a protein Source: own database.

The Bayesian analysis for the Orf7a protein shows conserved biological sequences of the Mu strains from Colombia in relation to those from the rest of the world.

In conclusion, the Orf7a protein of SARS-CoV-2, is more conserved and has been maintained by evolution despite the characterization of existing variants in the world, but shows some bifurcations and similarities between the gamma variant of Colombia and beta South Africa. It is known that this protein interacts with the ribosomal transport protein HEATDR3 and MDN1, repressing the immune system, which would imply that important changes in this protein could lead to a disabling of its functions(49).

ORF6 protein phylogeny of the variants Iota, Eta, Beta. Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta variants.

.

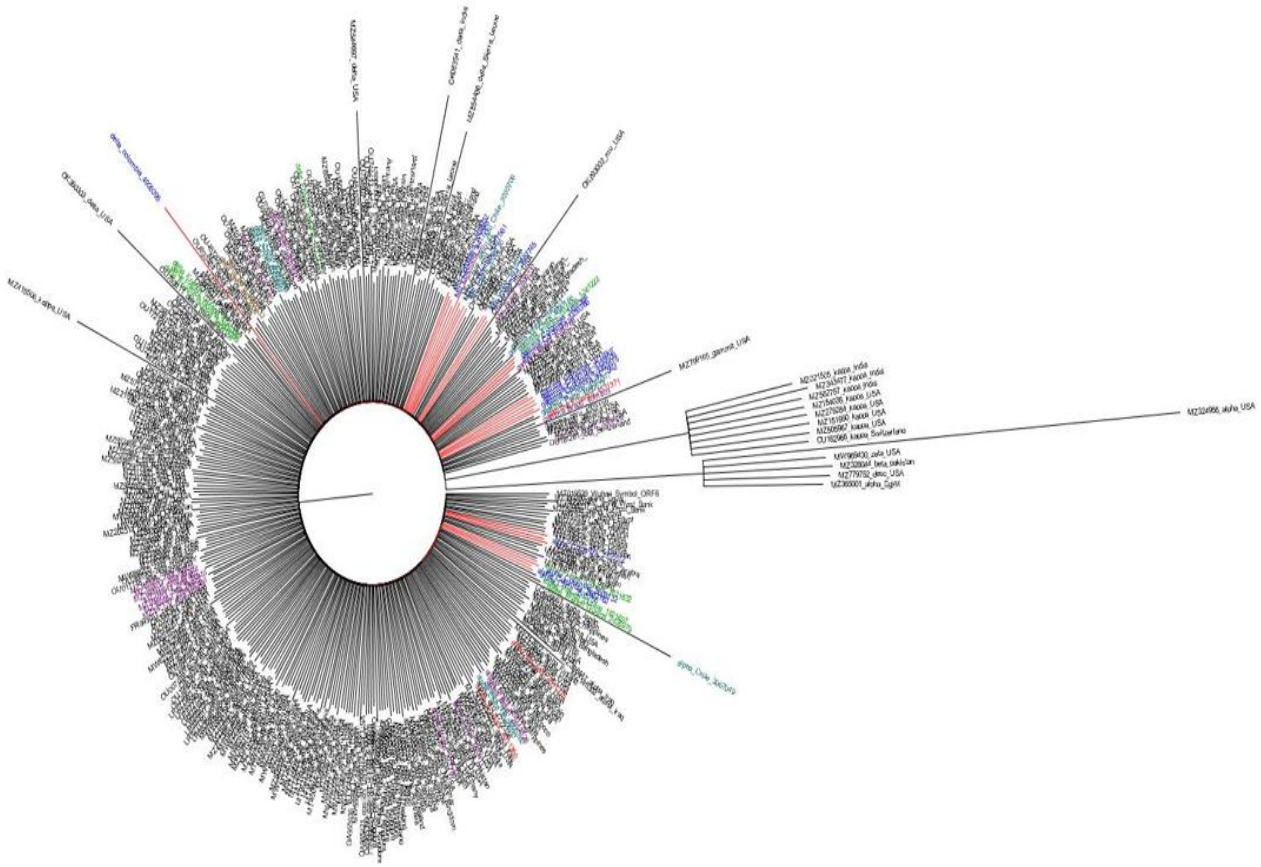


Image 25 Phylogenetic analysis for the Orf6 protein (own database).
Phylogenetic tree obtained, using the Bayesian Inference method, for the ORF6 gene. The names in blue correspond to the location of the strains from Colombia, in relation to the strains from the rest of the world. The evolutionary models of the canonical positions of the Orf6 gene are HKY and TrN.

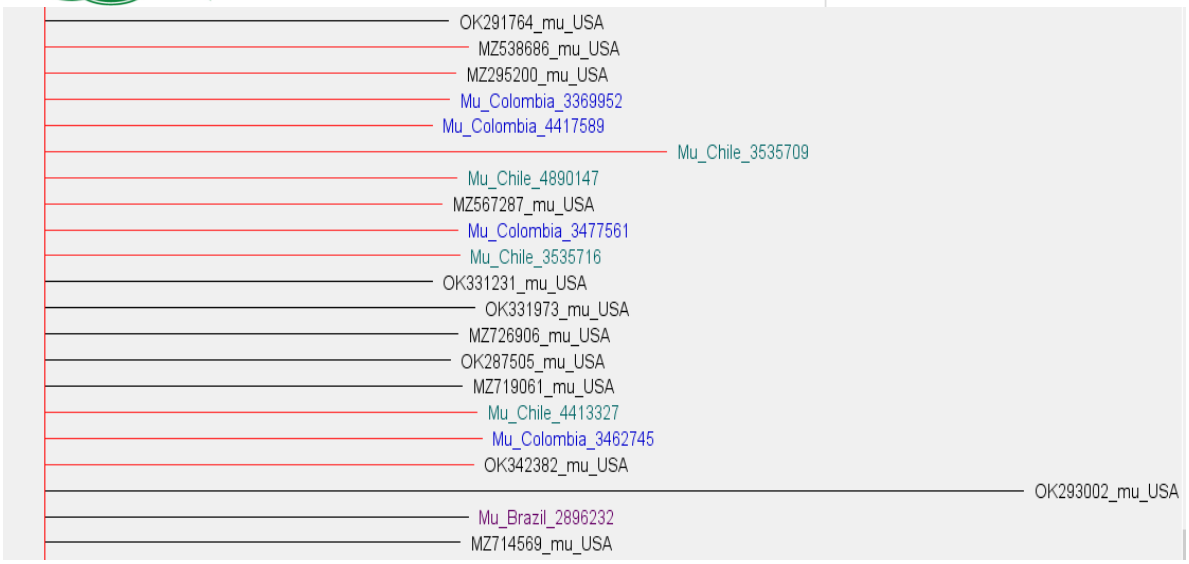


Image 26 Phylogenetic analysis of the Mu strain for the Orf6 protein. Source: own database.

The Bayesian analysis for the Orf6 protein presents conserved biological sequences of the MU strains from Colombia, in relation to those from the rest of the world.

Image 27 Phylogenetic analysis of the gamma strain for the Orf6 protein. Source: own database.

The Bayesian analysis for the Orf6 protein presents conserved biological sequences of the gamma strains from Colombia, in relation to those from the rest of the world.

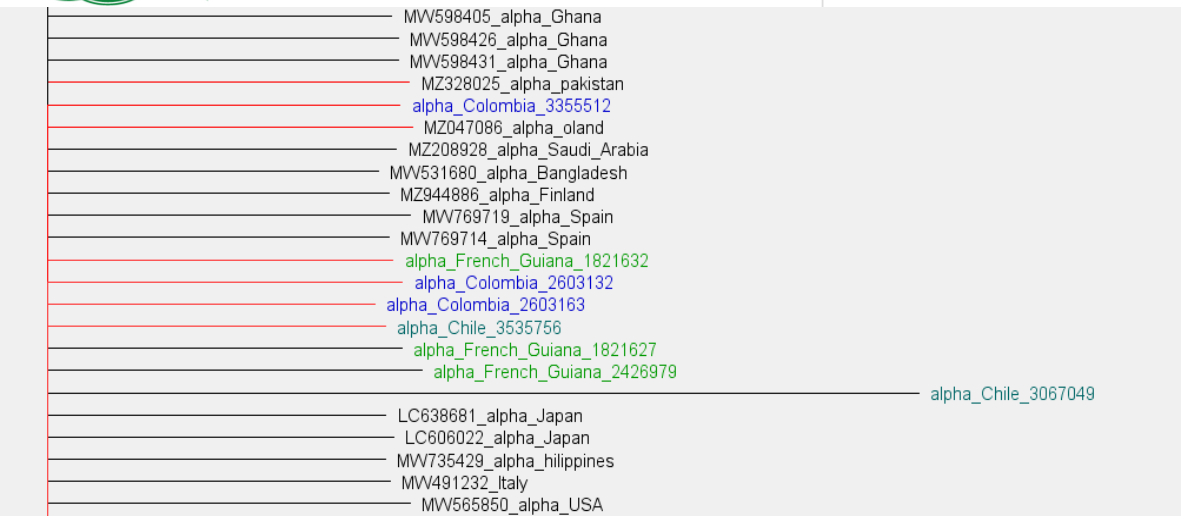


Image 28 Phylogenetic analysis of the Alpha strain for the Orf6 protein. Source: own database.

The Bayesian analysis for the Orf6 protein, presents conserved biological sequences of the Alpha strains from Colombia, in relation to those from the rest of the world.

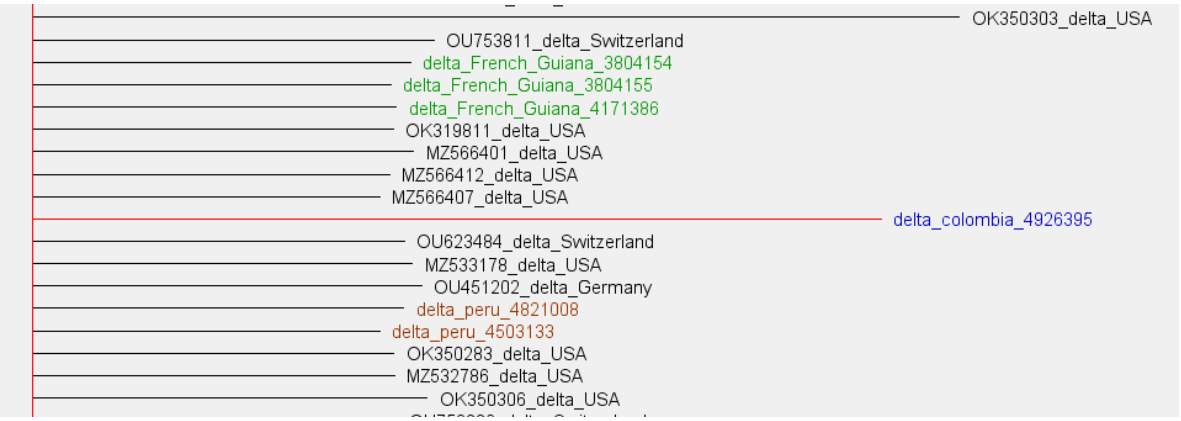


Image 29 Phylogenetic analysis of the delta strain for the Orf6 protein. Source: own database.

The Bayesian analysis for the Orf6 protein shows conserved biological sequences of the Delta strains from Colombia, in relation to those from the rest of the world.
In conclusion, the Orf6 protein of SARS-CoV-2 is highly conserved and has been maintained by evolution, despite the characterization of existing variants in the world.
There is an unresolved cladogenesis between kappa variants from India, USA and Switzerland and another between zeta and unknown variants from USA, beta from Pakistan and Alpha from Egypt that are not related to those from Colombia.

ORF3a protein phylogeny of the variants Iota, Eta, Beta. Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta.

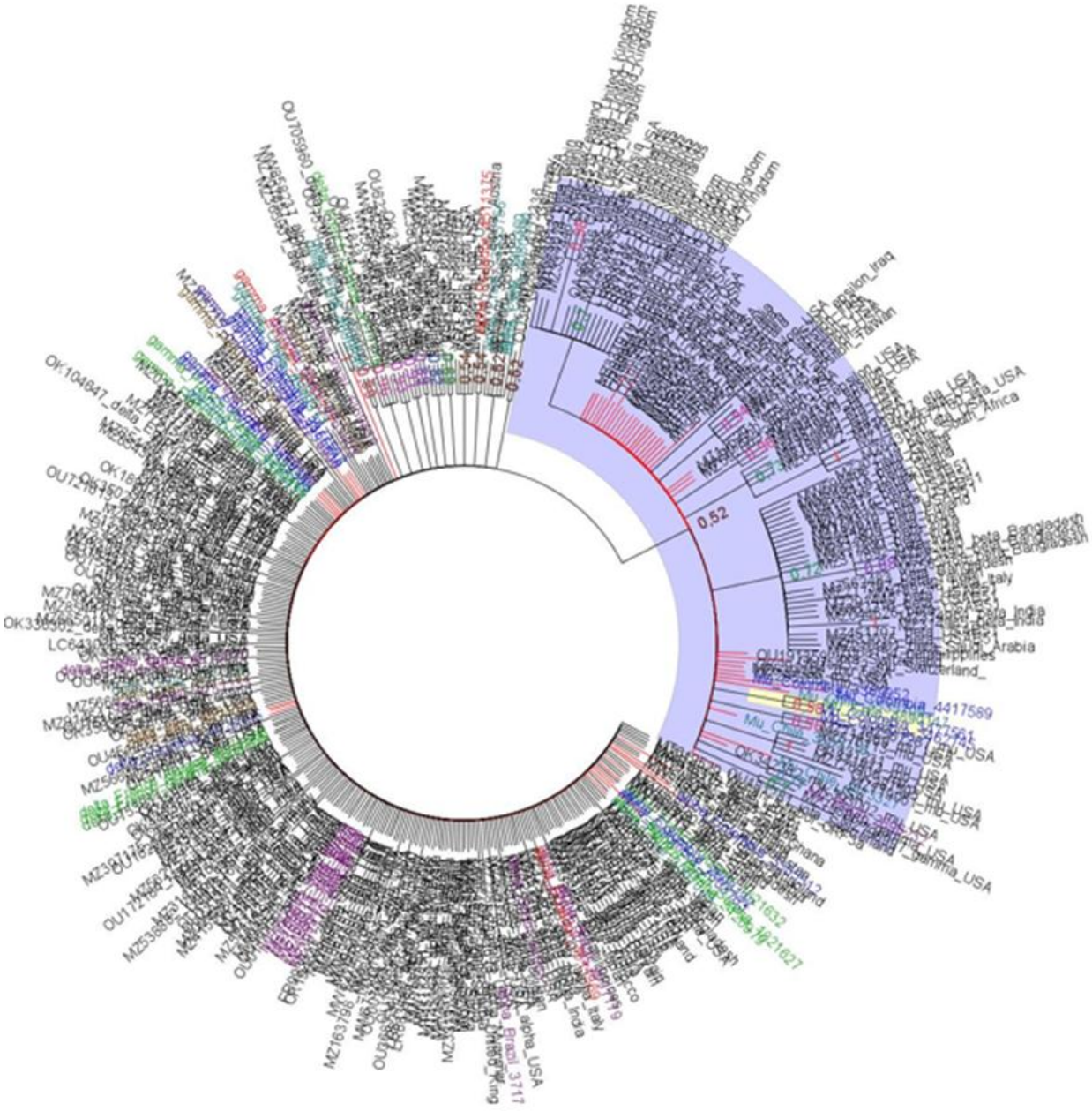


Image 30 Phylogenetic analysis for the Orf3a protein. Source: own database.
Phylogenetic tree obtained, using the Bayesian Inference method, for the ORF3a gene. The numbers correspond to Bayesian posterior probability values. The

groups highlighted in blue and beige correspond to the location of some strains from Colombia, in relation to those from the rest of the world, and the blue names correspond to strains from Colombia. The evolutionary models of the canonical positions of the Orf3a gene were TIM2 and GTR.

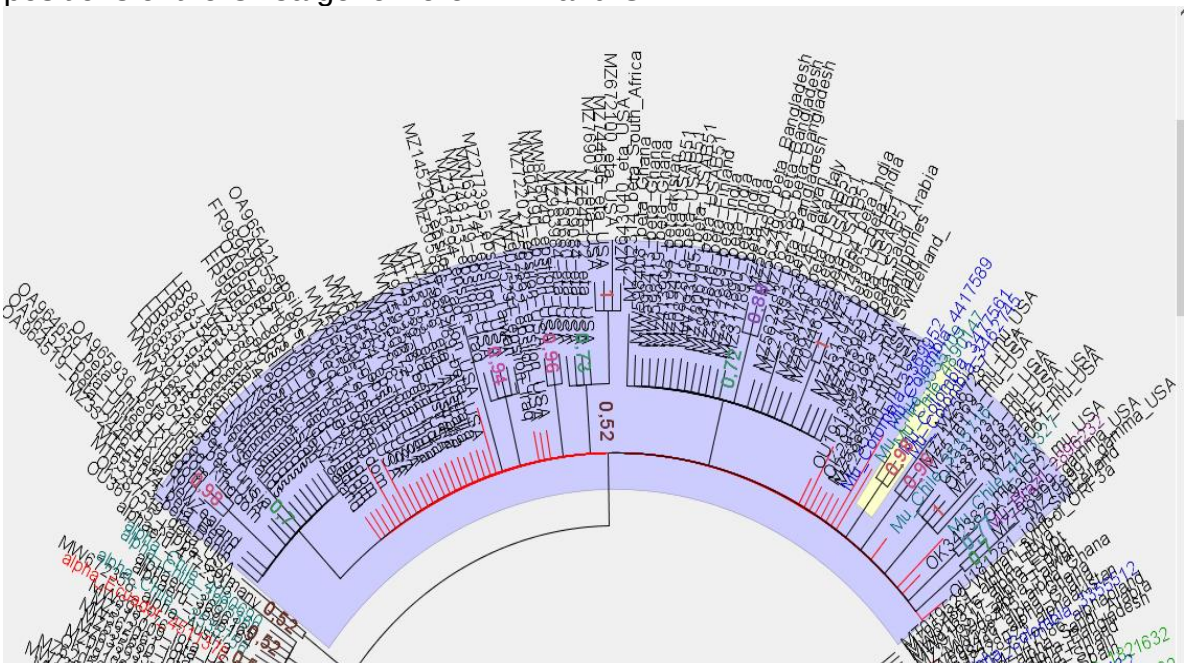


Image 31 Phylogenetic analysis of the Mu strain for the Orf3a protein. Source: own database.

The Bayesian analysis for the Orf3a protein presents a group with Bayesian posterior probabilities of 0.52. The Mu Colombia strains present homology in the orf3a protein, in relation to the Mu sequences of the world and the beta, eta, iota and epsilon strains.

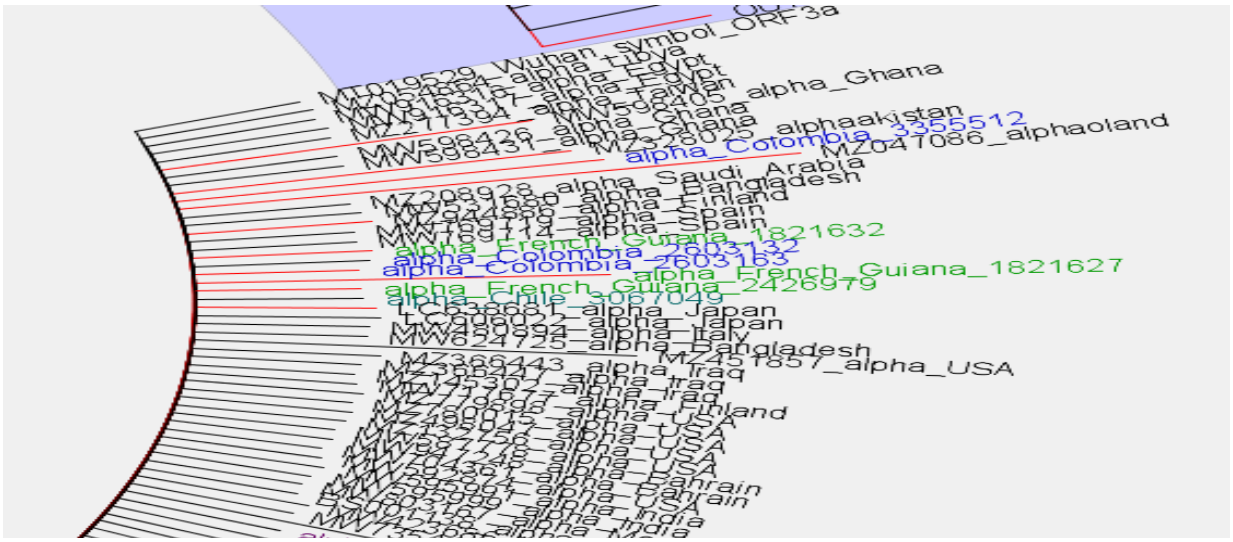
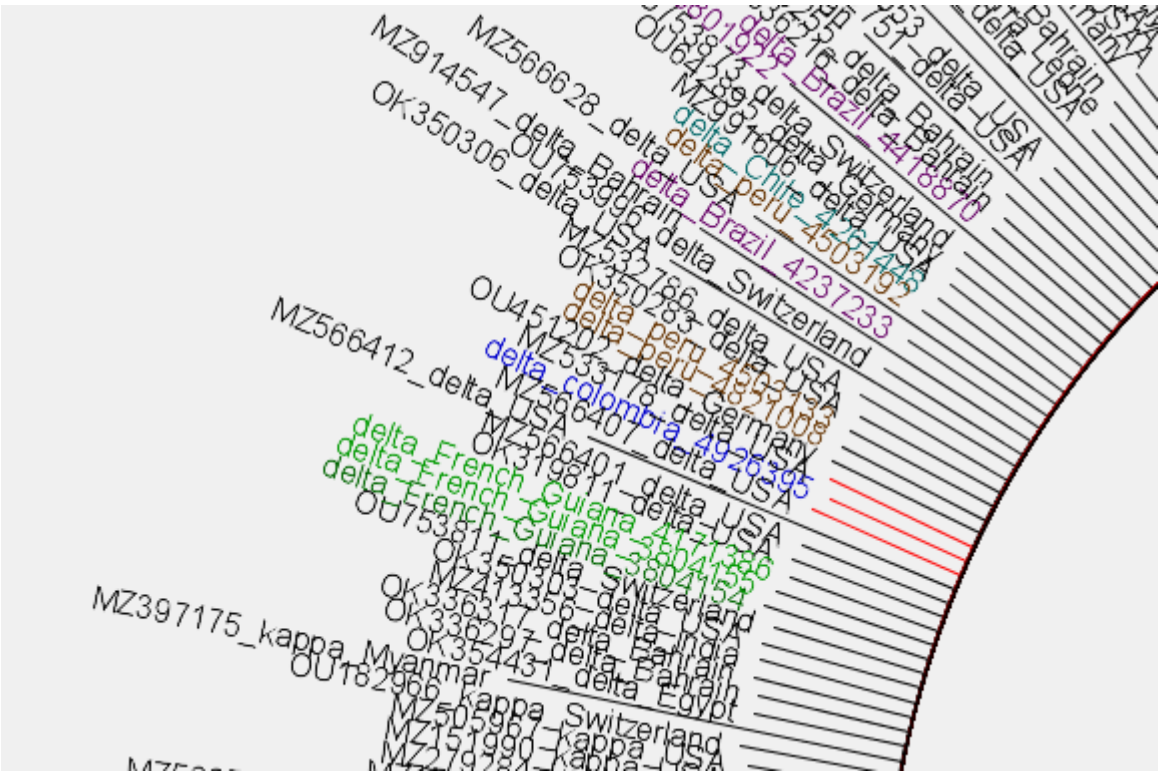


Image 32 Phylogenetic analysis of the Alpha strain for the Orf3a protein. Source: own database.

The Bayesian analysis for the Orf3a protein presents conserved biological sequences of the Alpha strains from Colombia, in relation to those from the rest of the world.



.
Image 33 Phylogenetic analysis of the Delta strain for the Orf3a protein. Source: own database.

The Bayesian analysis for the Orf3a protein, presents conserved biological sequences of the Delta strain from Colombia, in relation to those from the rest of the world.

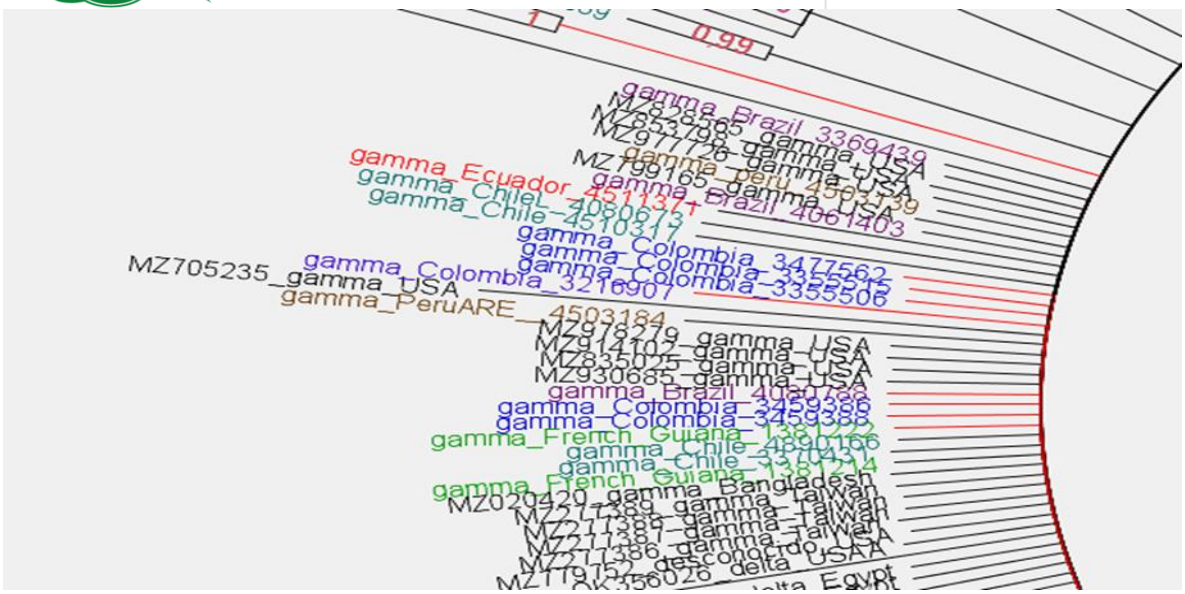


Image 34 Phylogenetic analysis of the gamma strain for the Orf3a protein. Source: own database.

The Bayesian analysis for the Orf3a protein shows conserved biological sequences of the Gamma strains from Colombia in relation to those from the rest of the world. The ORF3a protein shows polyphyletic groups related in a node, where the sequences of Mu, chile and USA and variant iota, epsilon and eta are rooted with a posterior probability in its node of 0.52, but a probability of 0.98 was found between the mu strains of chile and Colombia, which according to the result are more related. There is also a highly conserved outgroup that includes Alpha, zeta, eta, Kappa, delta, gamma sequences from Colombia and the rest of the world, indicating that the ORf3a protein from Colombia, which has changed over time, has been related to the Mu variant. This indicates that this protein has evolved slower than expected (50).

Phylogeny of the M protein of the variants Iota, Eta, Beta. Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta.

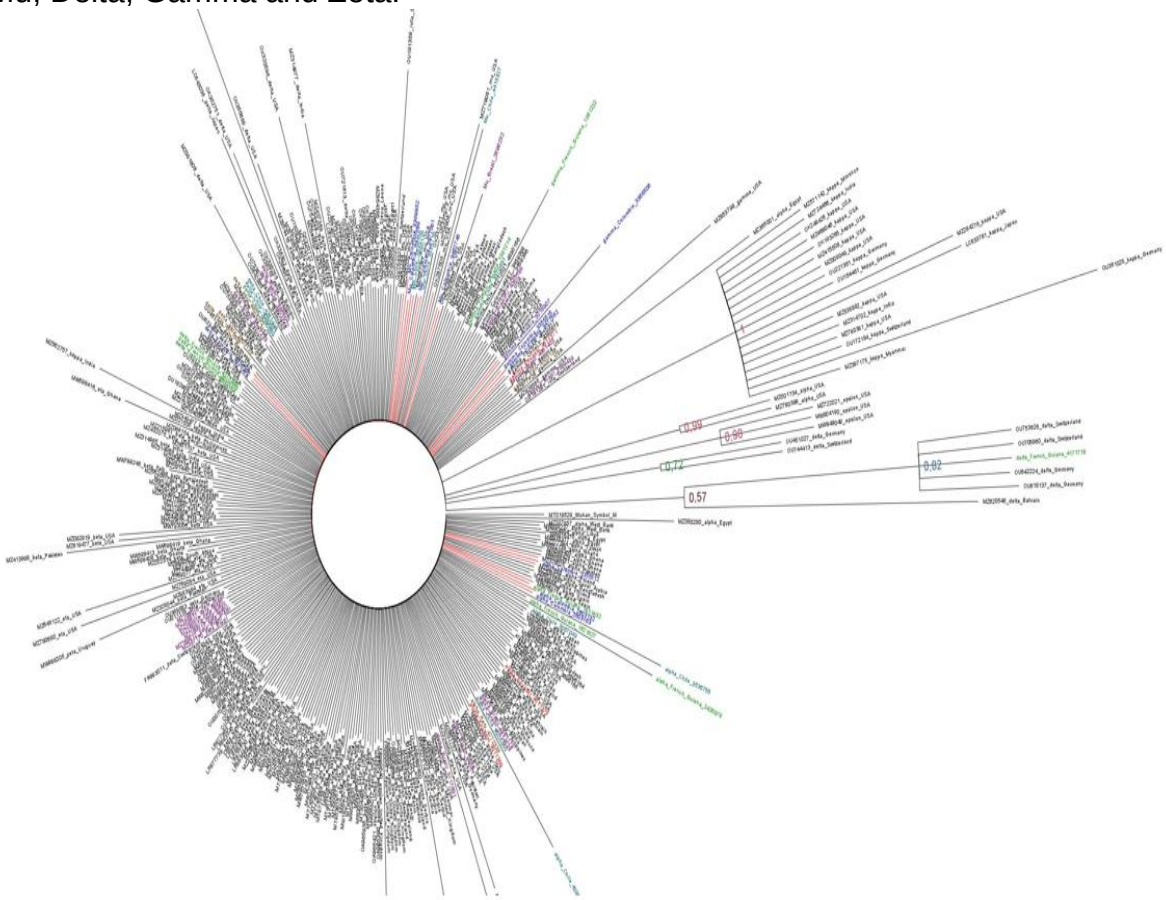


Figure 35 Phylogenetic analysis for protein M. Source: own database.

Phylogenetic tree obtained using the Bayesian Inference method for gene M. The numbers correspond to Bayesian posterior probability values. Names in blue correspond to the location of strains from Colombia, in relation to strains from the rest of the world. The evolutionary models of the canonical positions of the M gene are: TPM2uf and TIM2.

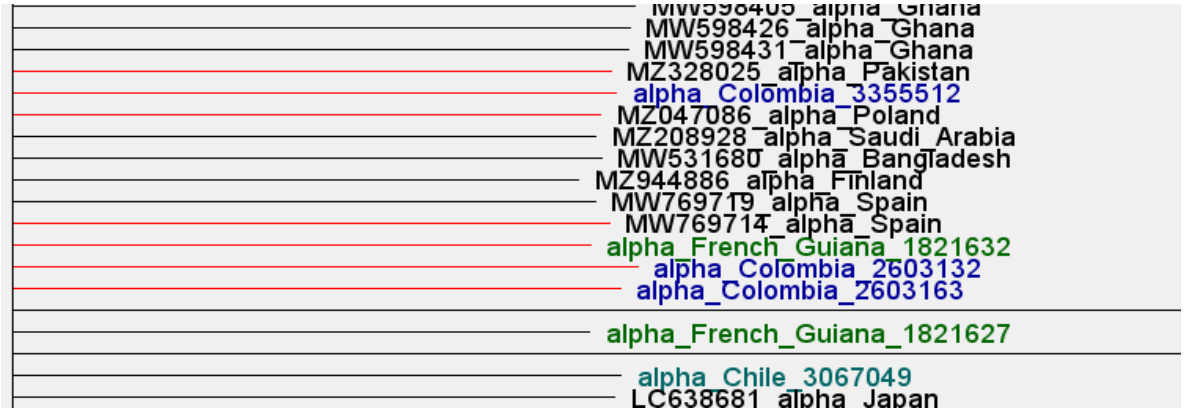


Image 36 Phylogenetic analysis of the Alpha strain for protein M. Source: own database.

The Bayesian analysis for the M protein presents conserved biological sequences of the Alpha strains from Colombia, in relation to those from the rest of the world.

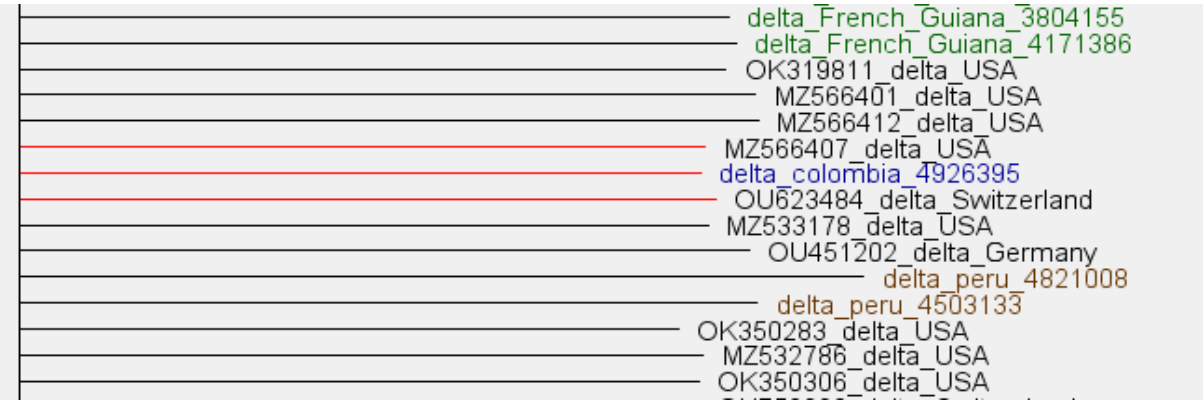


Image 37 Phylogenetic analysis of the Delta strain for protein M. Source: own database.

The Bayesian analysis for the M protein, presents conserved biological sequences of the Delta strains from Colombia, in relation to those from the rest of the world.

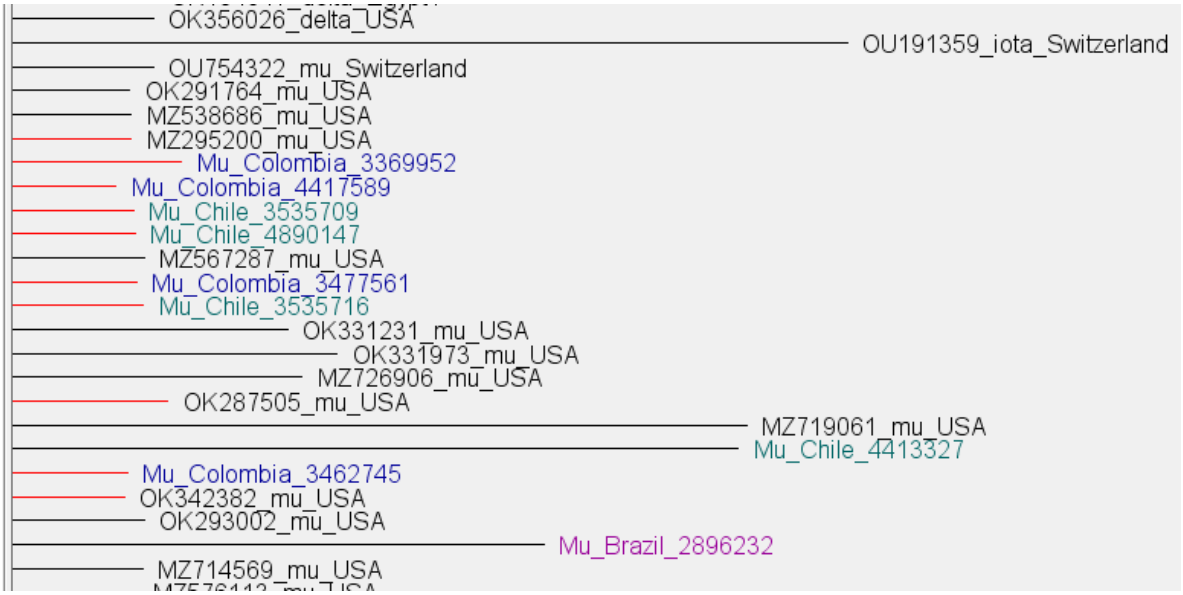


Image 38 Phylogenetic analysis of the Mu strain for protein M. Source: own database.

The Bayesian analysis, for the M protein, presents conserved biological sequences of the Mu strains from Colombia, in relation to those from the rest of the world.
In conclusion, the M protein is highly conserved in the strains circulating in Colombia. There are strains such as the European and African strains, which possess the protein with more evolutionary changes and are more distant from the Colombian strains.

Phylogeny of the E protein of the variants Iota, Eta, Beta. Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta.

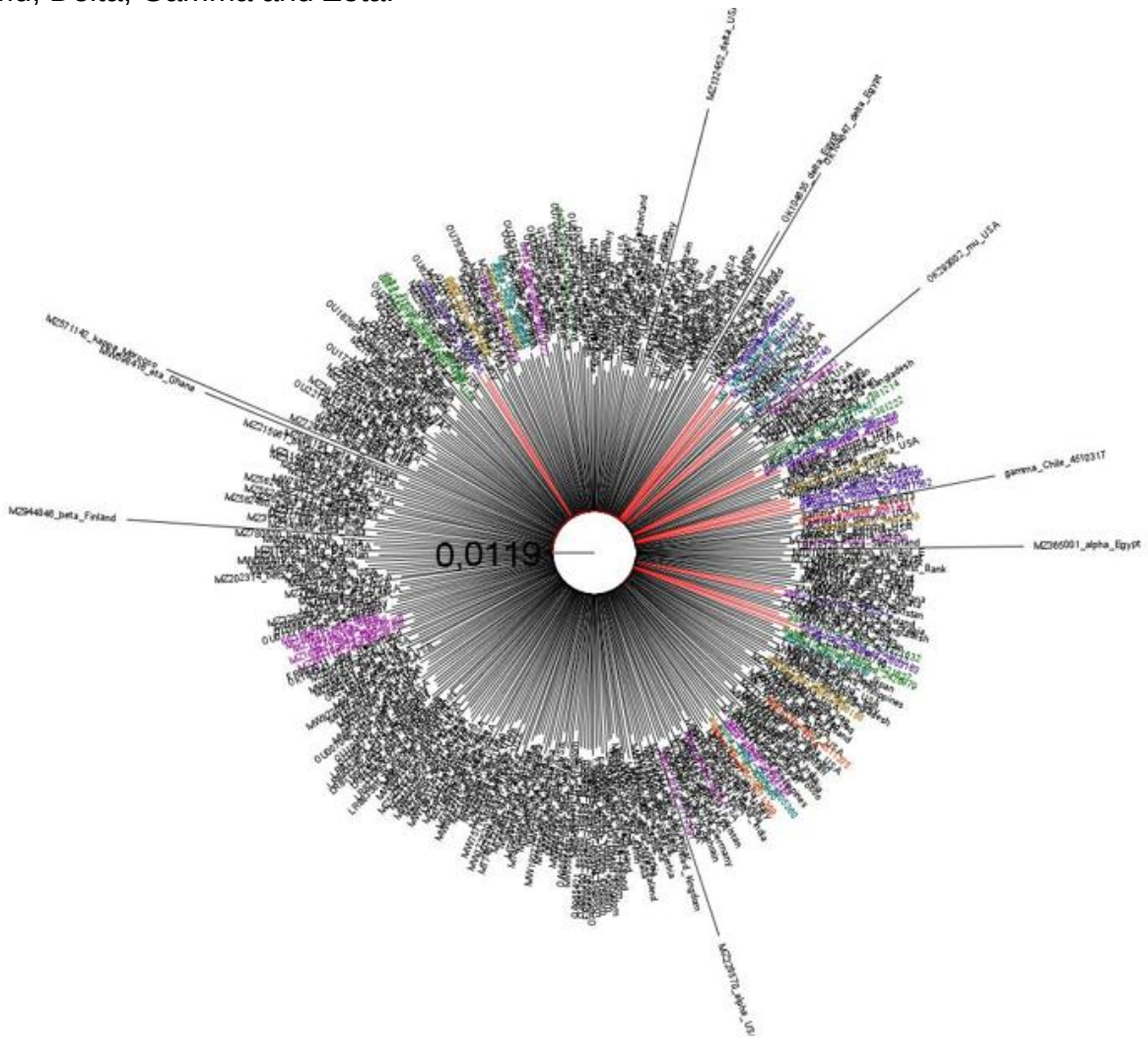


Figure 39 Phylogenetic analysis for protein E. Source: own database.
Phylogenetic tree obtained, using the Bayesian Inference method, for gene E. With Bayesian posterior probabilities of 0.0119. The sequences identified in blue correspond to the location of the strains from Colombia, in relation to the strains from the rest of the world. The evolutionary models of the canonical positions of the E gene, TrN and TIM3. This protein is conserved in all strains worldwide.

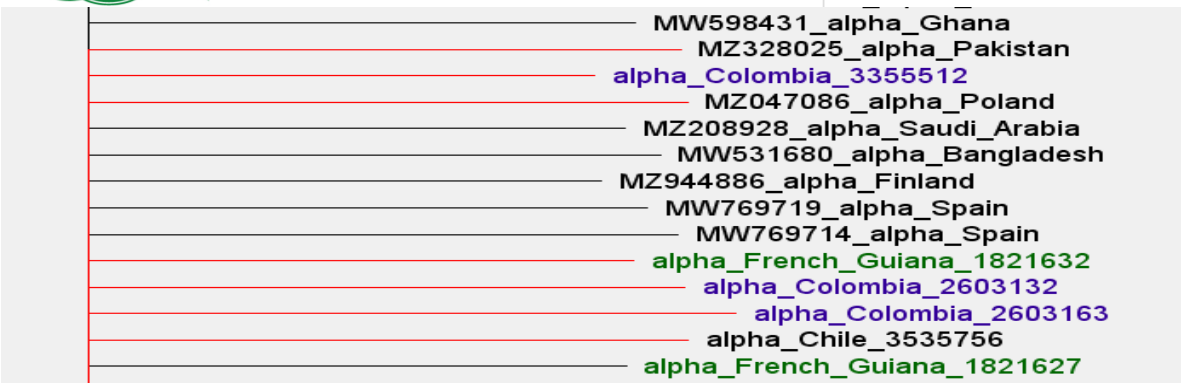


Image 40 Phylogenetic analysis of the Alpha strain for the E protein. Source: own database.

The Bayesian analysis for protein E, presents conserved biological sequences of the Alpha strains from Colombia, in relation to those from the rest of the world.

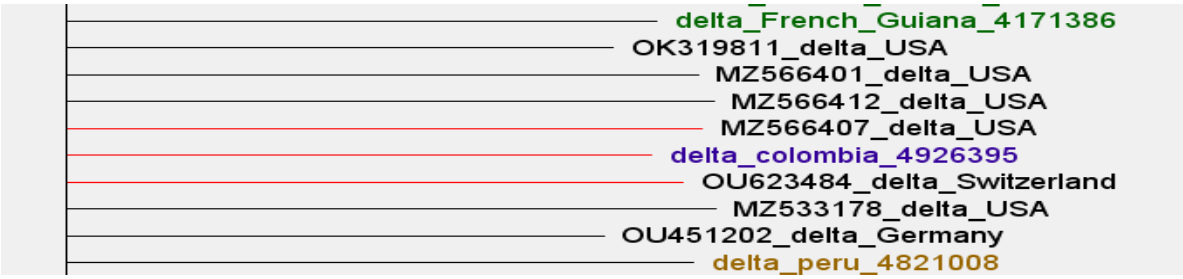


Image 41 Phylogenetic analysis of the Delta strain for the E protein. Source: own database.

The Bayesian analysis for protein E, presents conserved biological sequences of the delta strains from Colombia, in relation to those from the rest of the world.

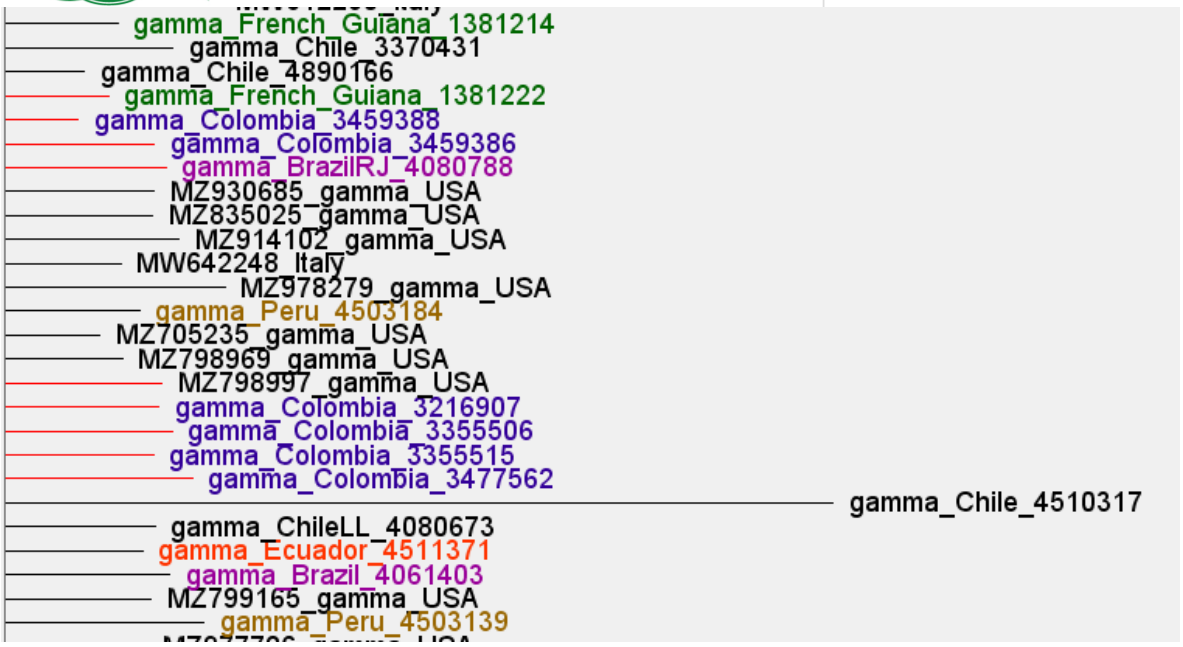


Image 42 Phylogenetic analysis of the gamma strain for the E protein. Source: own database.

The Bayesian analysis for protein E, presents conserved biological sequences of the Gamma strains from Colombia, in relation to those from the rest of the world.

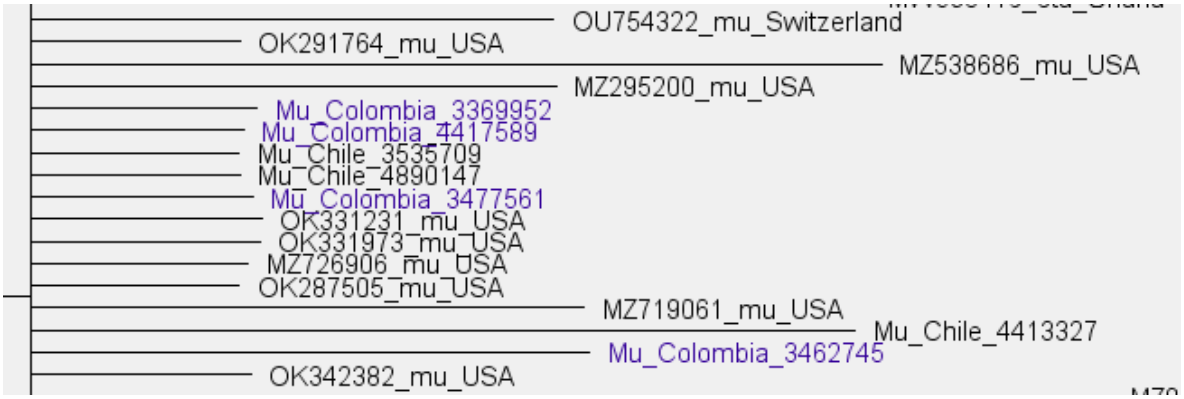


Image 43 Phylogenetic analysis of the Mu strain for the E protein. Source: own database.

The Bayesian analysis for protein E shows conserved biological sequences of Mu strains from Colombia, in relation to those from the rest of the world.
In conclusion, protein E is a very short and at the same time fully conserved protein.
Important mutations in protein E are unknown and why fewer errors occur in this protein than in the other 3 structural proteins so far.

Protein N phylogeny of the variants Iota, Eta, Beta. Epsilon, kappa, Alpha, Mu, Delta, Gamma and Zeta.

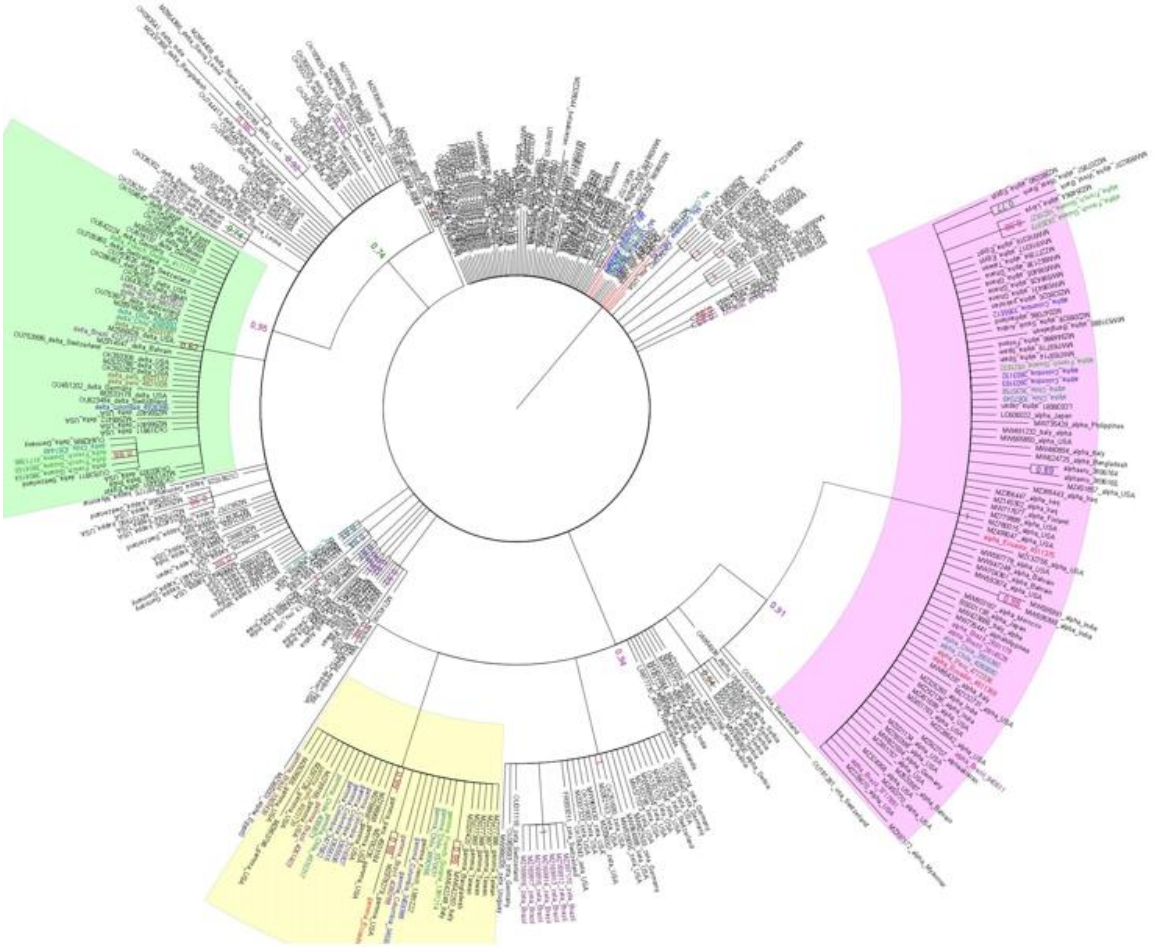


Image 44 Phylogenetic analysis for protein N. Source: own database.

Phylogenetic tree obtained using the Bayesian Inference method for the N gene. The numbers correspond to Bayesian posterior probability values. The names in blue color are the strains belonging to Colombia and the names highlighted in beige color correspond to the group where some strains from Colombia are located, in relation to the strains from the rest of the world. The evolutionary models of the canonical positions of the N gene is GTR.

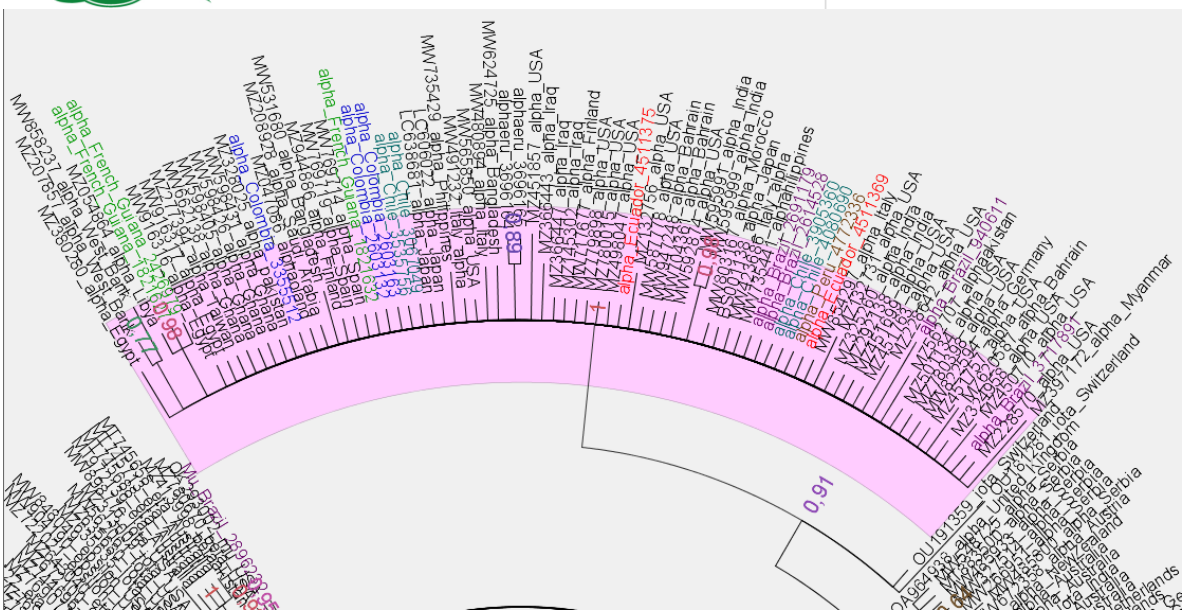


Image 45 Phylogenetic analysis of the Alpha strain for protein N. Source: own database.

The Bayesian analysis for the N protein gene shows a polyphyletic group with Bayesian posterior probabilities of 0.91. The Alpha Colombia strains present homology in the N protein in relation to the Alpha sequences of the world, with the exception of some European strains.

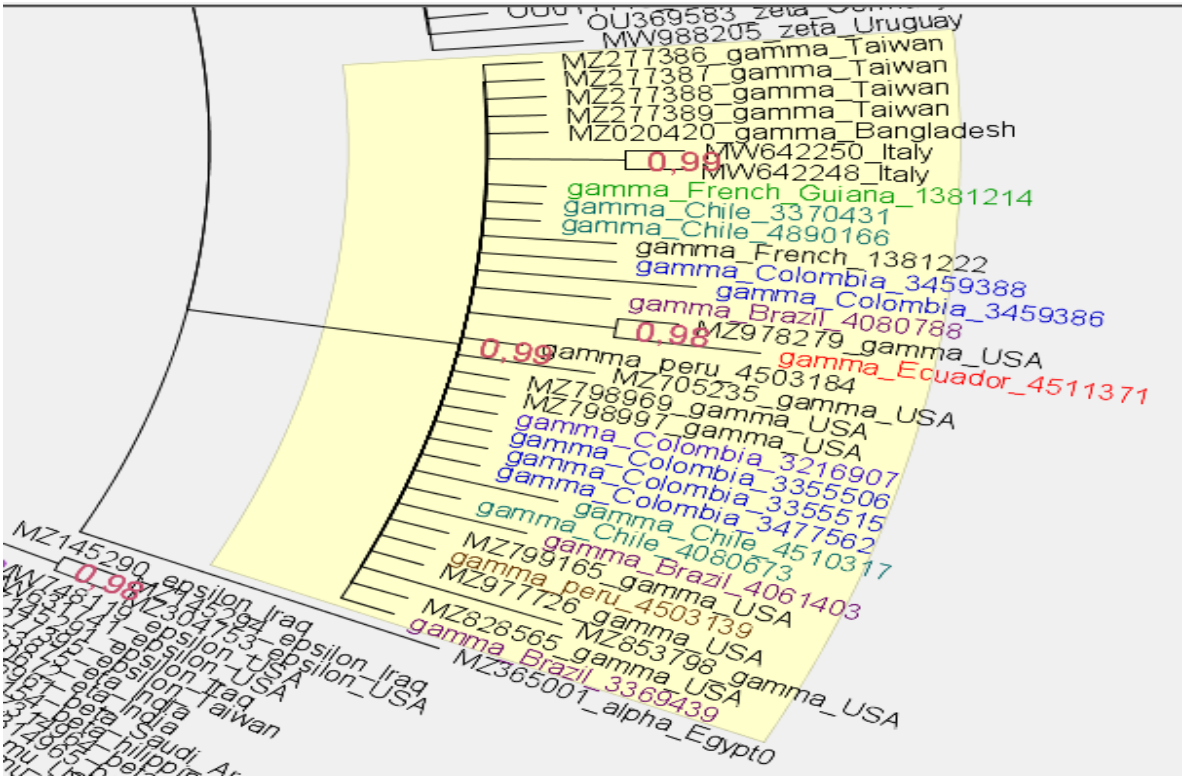


Image 46 Phylogenetic analysis of the gamma strain for protein N. Source: own database.

The Bayesian analysis for the N protein gene shows a polyphyletic group with Bayesian posterior probabilities of 0.91. The Gamma Colombia strains present homology in the N protein, in relation to the gamma sequences of the USA, South America, and some Asian and European strains.

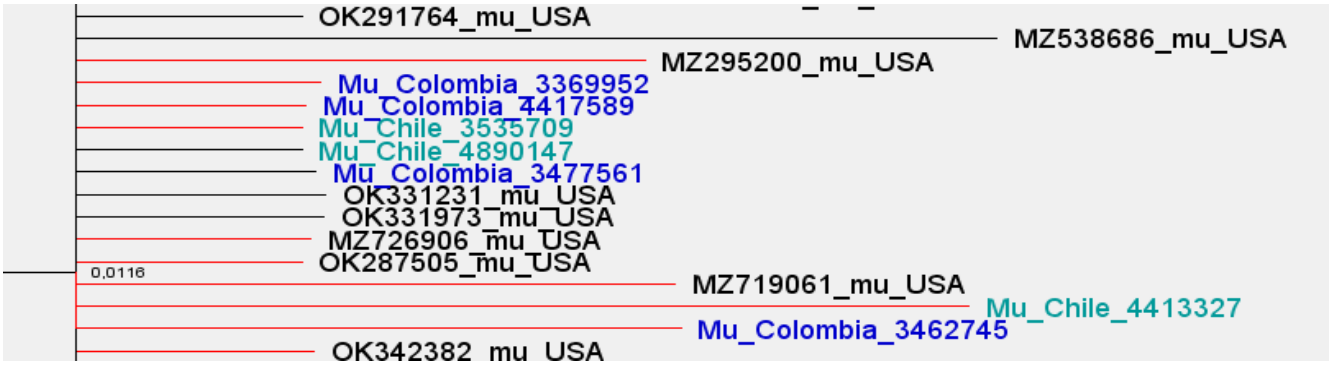


Image 47 Phylogenetic analysis of the Mu strain for protein N. Source: own database.

The Bayesian analysis for the N protein presents conserved biological sequences of Mu strains from Colombia, in relation to those from the rest of the world.

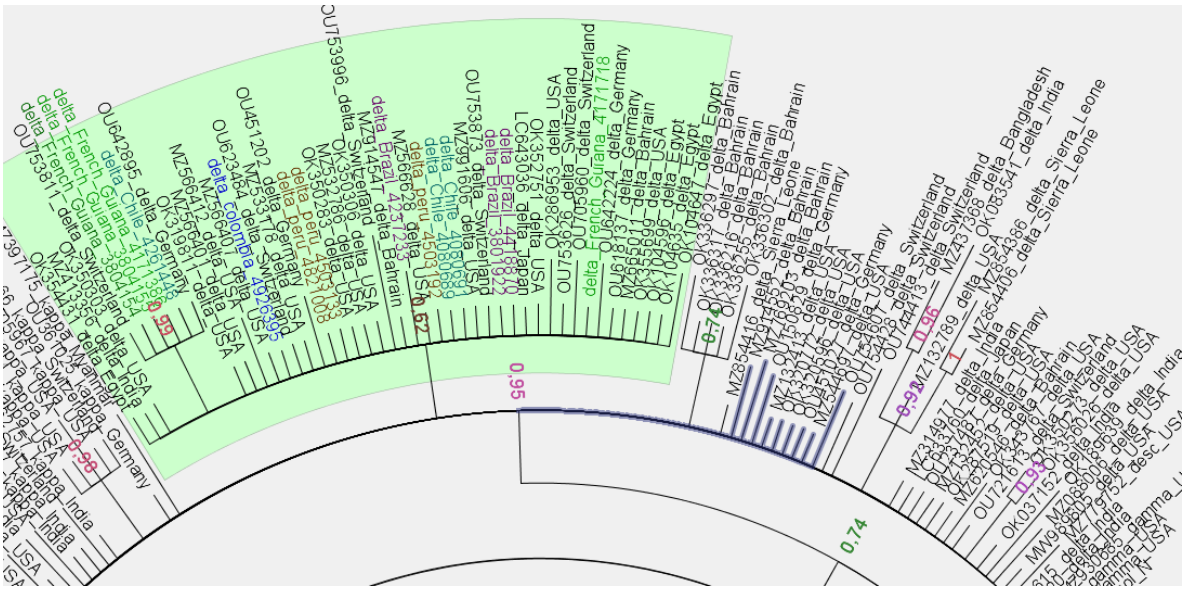


Figure 48 Phylogenetic analysis of the delta la strain for protein N (own database).

The Bayesian analysis for protein N presents a group with Bayesian posterior probabilities of 0.95, between the delta strains of Colombia and Delta of the world and presents a posterior probability of 0.62 with protein N from South America and some from USA and Europe.

In conclusion, we found 3 groups where the sequences of Colombia are identified, different polymorphisms at world level, which determines that this protein has evolved, in such a way that evolutionary relations exist between the gamma strains of Colombia and the world with a posterior probability of 0. There is also a group with posterior probabilities of 0.74 between the delta strains of Colombia and the rest of the countries of the world and the kappa variant of the analyzed sequences.
There are variants of the conserved N protein and other similar ones among which are the Mu strains from Colombia, Chile, USA, Switzerland and the epsilon and beta strains.
Alpha and gamma proteins from Colombia have homology with a posterior probability of 0.94.
Shaminur "et al" in their article Evolutionary dynamics of the SARS-CoV-2 nucleocapsid protein and its consequences indicate that this protein implies a continuous evolution, since they found 1,034 unique mutations, most of them belonging to European, American and Australian countries. To date, there are no studies of important mutations of the N protein in the Mu strains of South America.

Conclusions:

Bayesian analyses allowed us to know the evolutionary changes and protein conservation of proteins S, M, N, E and ORF (3a, 6, 7a, 8, 10).
There is a colossal divergence in proteins S and N, indicating that they are the least conserved proteins, especially protein S, which has a high rate of evolution over time, and which differs to a greater extent from one variant to another, both in Colombia (country of the South American continent) and in the other countries of the other continents of the world, analyzed in this study, thus concluding, This would have the same effect in terms of effectiveness studies in current vaccines that use as mRNA, sites that encode protein S, but this could change as the evolution of this protein continues and new variants of interest and concern are created.
Orf3a and Orf8 are proteins that have evolved more slowly than proteins S and N and this is corroborated by research mentioned in this analysis, which reveals that these proteins have selective mutations(28).

Protein M, Orf6, Orf7a, Orf10 in Colombia are conserved as in the rest of the world, although there are some unresolved variants, due to convergent evolutions in some sequences that do not belong to Colombia.

Another important result is that protein E is the most conserved protein in SARS-CoV-2 and S and N are the least conserved, due to existing polymorphisms in them, being these probably good markers to differentiate variants.

Keywords:
Evolution, phylogenetics, variability, genome, mutability.

REFERENCIAS (colocar a cada artículo el DOI o la URL en caso de no tener DOI)

1. Sifuentes-Rodríguez E, Palacios-Reyes D. Covid-19: The outbreak caused by a new coronavirus. Bol Med Hosp Infant Mex. 2020;77(2):47–53.

2. Pillay S, Giandhari J, Tegally H, Wilkinson E, Chimukangara B, Lessells R, et al. Whole genome sequencing of sars-cov-2: Adapting illumina protocols for quick and accurate outbreak investigation during a pandemic. Genes (Basel). 2020;11(8):1–13.

3. OMS. ABC de las Variantes Causantes del COVID-19 – Pregrados y Posgrados en Bogotá [Internet]. Orientaciones para la vigilancia de las variantes del SARS-CoV-2. 2021 [cited 04 November 2021]. Available at: https://www.konradlorenz.edu.co/noticias/abc-de-las-variantes-causantes-del-covid-19/

4. Revised U.S. Surveillance Case Definition for Severe Acute Respiratory Syndrome (SARS) and Update on SARS Cases—United States and Worldwide, December 2003. JAMA [Internet]. 2004 [cited 11 Oktober 2021];291(2):173. Available at: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5249a2.htm

5. Reina J, Reina N. El coronavirus causante del síndrome respiratorio de Oriente Medio. Med Clin (Barc) [Internet]. 2015 [cited 11 Oktober 2021];145(12):529–31. Available at: https://www.who.int/es/news-room/fact-sheets/detail/middle-east-respiratory-syndrome-coronavirus-(mers-cov)

6. Nuevo coronavirus 2019 [Internet]. Organizacion Mundial de la Salud. 20202 [cited 17 Augustus 2021]. bl 5–8. Available at: https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019?gclid=Cj0KCQjwwY-LBhD6ARIsACvT72P3W38IU_B1qpsbO8kQbHYewpeWkITELinZM48XKuYSexLXjMTvuncaApCiEALw_wcB

7. Sars-cov- E. Orientaciones para la vigilancia de las variantes del. 2021;(1):1–22.

8. Martínez-Flores D, Zepeda-Cervantes J, Cruz-Reséndiz A, Aguirre-Sampieri S, Sampieri A, Vaca L. SARS-CoV-2 Vaccines Based on the Spike Glycoprotein and Implications of New Viral Variants. Front Immunol. 12 Julie 2021;0:2774.

9. Lokman SM, Rasheduzzaman M, Salauddin A, Barua R, Tanzina AY, Rumi MH, et al. Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: A computational biology approach. Infect Genet Evol [Internet]. 2020;84(May):104389. Available at: https://doi.org/10.1016/j.meegid.2020.104389

10. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med [Internet]. 2020;18(1):1–9. Available at: https://doi.org/10.1186/s12967-020-02344-6

11. Kasibhatla SM, Kinikar M, Limaye S, Kale MM, Kulkarni-Kale U. Understanding evolution of SARS-CoV-2: A perspective from analysis of genetic diversity of RdRp gene. J Med Virol [Internet]. 2020;92(10):1932–7. Available at: http://dx.doi.org/10.1002/jmv.25909

12. Woo PCY, Huang Y, Lau SKP, Yuen KY. Coronavirus genomics and bioinformatics analysis. Viruses. 2010;2(8):1805–20.

13. Duffy S. Why are RNA virus mutation rates so damn high? PLoS Biol. 2018;16(8):1–6.

14. Chitranshi N, Gupta VK, Rajput R, Godinez A, Pushpitha K, Shen T, et al. Evolving geographic diversity in SARS-CoV2 and in silico analysis of replicating enzyme 3CLprotargeting repurposed drug candidates. J Transl Med [Internet]. 2020;18(1):1–15. Available at: https://doi.org/10.1186/s12967-020-02448-z

15. Prompetchara E, Ketloy C, Palaga T. Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. Asian Pacific J Allergy Immunol. 2020;38(1):1–9.

16. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun [Internet]. 2020;109(February):102433. Available at: https://doi.org/10.1016/j.jaut.2020.102433

17. Pastrian-Soto G. Bases Genéticas y Moleculares del COVID-19 (SARS-CoV-2). Mecanismos de Patogénesis y de Respuesta Inmune. Int J Odontostomatol. 2020;14(3):331–7.

18. FERNÁNDEZ JOSÉ M. Nuevo mapa genetico del SARS-CoV-2 [Internet]. 2020 [cited 10 Februarie 2022]. Available at: https://www.mendeley.com/reference-manager/library/recently-read/

19. Fernando Benavides-Rosero M. COVID-19 y la pandemia global causada por un nuevo coronavirus COVID-19 and the global pandemic caused by a new coronavirus. [cited 26 Oktober 2021]; Available at: https://doi.org/10.22267/rus.202203.203

20. Hassan SS, Aljabali AAA, Panda PK, Ghosh S, Attrish D, Choudhury PP, et al. A unique view of SARS-CoV-2 through the lens of ORF8 protein. Comput Biol Med [Internet]. 01 Junie 2021 [cited 25 Oktober 2021];133:104380. Available at: /pmc/articles/PMC8049180/

21. Hassan SS, Choudhury PP, Roy B. Rare mutations in the accessory proteins ORF6, ORF7b, and ORF10 of the SARS-CoV-2 genomes. Meta Gene [Internet]. 01 Junie 2021 [cited 25 Oktober 2021];28:100873. Available at: /pmc/articles/PMC7890336/

22. Li X, Hou P, Ma W, Wang X, Wang H, Yu Z, et al. SARS-CoV-2 ORF10 suppresses the antiviral innate immune response by degrading MAVS through mitophagy. Cell Mol Immunol 2021 191 [Internet]. 29 November 2021 [cited 11 Februarie 2022];19(1):67–78. Available at: https://www.nature.com/articles/s41423-021-00807-4

23. Zhou Z, Huang C, Zhou Z, Huang Z, Su L, Kang S, et al. Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14+ monocytes. iScience [Internet]. 19 Maart 2021 [cited 25 Oktober 2021];24(3). Available at: /pmc/articles/PMC7879101/

24. Hassan SS, Attrish D, Ghosh S, Choudhury PP, Roy B. Pathogenic perspective of missense mutations of ORF3a protein of SARS-CoV-2. Virus Res. 15 Julie 2021;300:198441.

25. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet. 2009;10(8):540–50.

26. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev. 2020;7(6):1012–23.

27. Boni MF, Lemey P, Jiang X, Lam TTY, Perry BW, Castoe TA, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol [Internet]. 2020; Available at: http://dx.doi.org/10.1038/s41564-020-0771-4

28. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. J Med Virol. 2020;92(6):667–74.

29. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol [Internet]. 2020;83(May):104351. Available at: https://doi.org/10.1016/j.meegid.2020.104351

30. Islam MR, Hoque MN, Rahman MS, Alam ASMRU, Akther M, Puspo JA, et al. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. Sci Rep [Internet]. 2020;10(1):1–9. Available at: https://doi.org/10.1038/s41598-020-70812-6

31. Hou W. Characterization of codon usage pattern in hepatitis C virus subtype 6xa. Clin Lab. 2020;66(8):1649–51.

32. Gomez-Carballa A, Bello X, Pardo-Seco J, Martinon-Torres F, Salas A.

Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. Genome Res [Internet]. 2020; Available at: http://www.ncbi.nlm.nih.gov/pubmed/32878977

33. Nidhan K Biswas PPM. No Title. Anal RNA Seq 3636 SARS-CoV-2 Collect from 55 Ctries Reveal Sel sweep one virus type. 2020;151(5):450–8.

34. Jones LR, Manrique JM. Quantitative phylogenomic evidence reveals a spatially structured SARS-CoV-2 diversity. Virology [Internet]. 2020;550(August):70–7. Available at: https://doi.org/10.1016/j.virol.2020.08.010

35. Khan MI, Khan ZA, Baig MH, Ahmad I, Farouk AEA, Song YG, et al. Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: An in silico insight. PLoS One [Internet]. 2020;15(9):e0238344. Available at: http://dx.doi.org/10.1371/journal.pone.0238344

36. Organización Mundial de la Salud (OMS). Seguimiento de las variantes del SARS-CoV-2 [Internet]. Organización Mundial de la Salud (OMS). 2021 [cited 22 Januarie 2021]. bl 1–4. Available at: https://www.who.int/es/activities/tracking-SARS-CoV-2-variants

37. BBC News Mundo. Coronavirus: qué se sabe de la variante de “doble mutación” encontrada en India [Internet]. BBC NEWS. 2021 [cited 22 Oktober 2021]. Available at: https://www.bbc.com/mundo/noticias-55662073

38. Cuál es el origen de la variante Delta [Internet]. 2020 [cited 22 Oktober 2021]. Available at: https://www.abc.es/sociedad/abci-origen-variante-delta-nsv-202107081236_noticia.html?ref=https%3A%2F%2Fwww.abc.es%2Fsociedad%2Fabci-origen-variante-delta-nsv-202107081236_noticia.html

39. Aytekin E. Estas son las variantes de la COVID-19 que plantean dificultades en la lucha mundial contra la pandemia [Internet]. ANADOLU AGENCY. 2021 [cited 23 Oktober 2021]. Available at: https://www.aa.com.tr/es/mundo/estas-son-las-variantes-de-la-covid-19-que-plantean-dificultades-en-la-lucha-mundial-contra-la-pandemia/2284715

40. West AP, Barnes CO, Yang Z, Bjorkman PJ. SARS-CoV-2 lineage B.1.526 emerging in the New York region detected by software utility created to query the spike mutational landscape. bioRxiv [Internet]. 23 Februarie 2021 [cited 23 Oktober 2021];2021.02.14.431043. Available at: https://www.biorxiv.org/content/10.1101/2021.02.14.431043v2

41. Centro Integral de Informacion Biotecnologica NCBI. Recursos del SARS-CoV-2 - NCBI [Internet]. 1988 [cited 26 Oktober 2021]. Available at: https://www.ncbi.nlm.nih.gov/sars-cov-2/

42. Recurso de análisis y base de datos de patógenos de virus (ViPR) - Coronaviridae - Herramientas de búsqueda de ViPR [Internet]. [cited 14 Oktober 2021]. Available at: https://www.viprbrc.org/brc/search_landing.spg?decorator=corona_ncov

43. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: Integrated protein sequence and structural alignment. Nucleic Acids Res. 01 Julie 2019;47(W1):W5–10.

44. CIPRES Science Gateway | Hogar [Internet]. "Creando el CIPRES Science Gateway para inferencia de grandes filogenia trees. 2010 [cited 13 Oktober 2021]. Available at: https://www.phylo.org/portal2/home.action

45. Sucar LE. Introduction to bayesian networks and influence diagrams. Decis Theory Model Appl Artif Intell Concepts Solut. 2011;9–32.

46. S C, F L. Bayesian phylogeny analysis via stochastic approximation Monte Carlo. Mol Phylogenet Evol [Internet]. November 2009 [cited 13 Oktober 2021];53(2):394–403. Available at: https://pubmed.ncbi.nlm.nih.gov/19589389/

47. NCBI Virus [Internet]. [cited 15 Oktober 2021]. Available at: https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=SARS-CoV-2, taxid:2697049

48. Mena EL, Donahue CJ, Vaites LP, Li J, Rona G, O'Leary C, et al. ORF10–Cullin-2–ZYG11B complex is not required for SARS-CoV-2 infection. Proc Natl Acad Sci [Internet]. 27 April 2021 [cited 25 Oktober 2021];118(17). Available at: https://www.pnas.org/content/118/17/e2023157118

49. Addetia A, Xie H, Roychoudhury P, Shrestha L, Loprieno M, Huang M-L, et al. Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. medRxiv [Internet]. 09 Junie 2020 [cited 26 Oktober 2021];2020.06.08.20125856. Available at: https://www.medrxiv.org/content/10.1101/2020.06.08.20125856v1

50. Masters PS. Coronavirus genomic RNA packaging [Internet]. Vol 537, Virology. 2019 [cited 26 Oktober 2021]. bl 198–207. Available at: https://DOI: 10.1016/j.virol.2019.08.031.