Programa de pos-graduación en Genética

Tesis de Maestría

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

Sirly Johanna Rivera Roca

Barraquilla (Atl), Colombia

2022

Universidad Simón Bolívar
Programa de pos-graduación en Genética

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

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

Sirly Johanna Rivera Roca

Tutor:

Moisés Alberto Arquéz Mendoza

Cotutor:

Dayan Lozano

Barraquilla (Atl), Colombia

2022

Dedicatoria

A Dios primero que todo, a mi madre Doralis Roca, a mi padre Hichis Rivera, a mi esposo Ademar cardona y a mis hijos Josué Cardona y Taliana cardona, quienes han sido el eje y motor de todos los logros alcanzados en mi vida. También dedico este proyecto a todas aquellas familias, que han perdido de una u otra forma un ser querido, producto SARS-CoV-2.

Agradecimientos

Agradezco, a todas aquellas personas que estuvieron de una u otra manera apoyándome, a lo largo de la realización de este proyecto. Principalmente a mis padres, Doralis Roca y Hichis Rivera, mi esposo y mis hijos quienes han sido el motor de todos mis logros trazados y los que me dieron el valor y la fortaleza para

seguir adelante y cumplir cada una de mis metas.

A mis amigos y compañeros de maestría que de una u otra forma siempre estuvieron ahí para darme ánimos en especial mi compañera y amiga Olga, a mi amigo Roberto, por su apoyo incondicional y desinteresado.

A mi profesora Lourdes Varela, que, aunque ella desconozca que estoy actualmente en este proceso, fue quien influyo en mí, de una manera u otra en el amor y pasión que siento hoy por la genética, a la Dra. Dayan por su apoyo en el proceso continuo de esta investigación, agradezco también al Dr. Moisés Arquéz por haberme enseñado el significado de realizar una buena investigación, apoyándome siempre y enseñándome que yo soy capaz de fomentar nuevos conocimientos que pueden dar valor a la ciencia.

Resumen

El SARS-CoV-2 es un virus que constituye una importante amenaza para la salud a nivel mundial, debido a las altas tasas de infección y mortalidad que presenta. Pese a esto sus rasgos genéticos y los procesos evolutivos a los que está sometido aún son desconocido y es poca la información que existe en nuestro

país. Para poder entender los procesos evolutivos a los que está sometido este nuevo virus, es importante obtener una filogenia de las diferentes variantes circulantes, analizar esta variabilidad entre otras características, tanto en nuestro territorio como en el mundo.

En esta investigación, se realizó un análisis bioinformático de los genes de importancia estructural y funcional del virus, determinando de esta forma características genómicas y filogenéticas, de los subtipos de SARS-CoV-2 circulantes, mediante el método de inferencia bayesiana, adicionalmente se exploraron las relaciones entre las proteínas de algunas cepas y el grado de conservación total de la proteína E en todos los continentes.

Palabras clave:

Evolución, filogenética, variabilidad, genoma, mutabilidad.

Abstract

SARS-CoV-2 is a virus that has caused major health threats in the world. The genetic traits and evolutionary processes of each of its proteins in Colombia are still largely unknown.

In order to have more information on their phylogeny, variability or evolutionary mutability, in the strains that circulate in Colombia, in relation to the strains of the world, a bioinformatic analysis of the genes that translate to structural and accessory proteins was carried out, determining in this way genomic and phylogenetic characteristics of the circulating SARS-CoV-2 subtypes, through Bayesian inference methods and thus be participants in the construction of important bases, for the advancement of new diagnostic and treatment methods, finding relationships between the proteins of some strains and the total conservation of protein E in all continents.

Keywords:

Evolution, phylogenetics, variability, genome, mutability.

Introducción

Anteriormente, los coronavirus no eran considerados posibles virus, que pondrían en riesgo, a la seguridad sanitaria mundial,(1) hasta que aparecieron el MERS, el

SARS-CoV y el SARS-CoV-2 actual pandemia, que ha traído consigo muchas muertes.

Se conoce que esta enfermedad apareció en Wuhan, provincia de Hubei en China(1) y gracias a estudios filogenéticos, se cree, que aparentemente evoluciono, a partir del RaTG13 de Bat-CoV, diversificado de un ancestro en común, evolucionado del pangolín-CoV, siendo los murciélagos reservorios del virus, responsables de contagios zoonóticos del SARS-CoV-2.

Además, se ha comprobado que el virus está conformado por 4 proteínas estructurales llamadas proteínas S, M, N, E y unas proteínas ORF accesorias (1a,1b, 3a,6, 7a, 8 y 10) siendo el gen de la proteína S, el que se utilizada para estudios de producción de vacunas, contra el virus SARS-CoV-2.

Actualmente, existen estudios acerca de este virus, Además de vacunas, que ayudan a neutralizar la infección y la mayoría de estas, están basadas en la proteína S, pero a pesar de ser eficaces se desconoce si debido a tantas mutaciones

producidas en esta proteína disminuya su efectividad con el pasar de los años, por lo que autores como Sureshnee Pillay “et al” sustenta, la importancia de más análisis filogenéticos que nos proporcionen información relevante, para mejorar las estrategias de prevención y control de infecciones(2) además de futuros estudios hacia nuevas diana terapéuticas.

La evolución de este virus ha traído consigo, la aparición diferentes variantes, algunas de ellas catalogadas como variantes de monitoreo, interés o preocupación. Estas variantes, antes mencionadas son monitoreadas a nivel mundial debido a su mayor transmisibilidad. (3).

Actualmente, las más monitoreada es la variante Ómicron según la OMS y no se encuentra analizada en este estudio, debido al tiempo en que fue realizado.

Por lo anterior este estudio filogenético a través de inferencia bayesiana, nos muestra, el desarrollo, entendimiento y mantenimiento de las proteínas del SARS CoV-2, en cada cepa encontrada en Colombia y a su vez, determina cepas, que poseen mayores cambios evolutivos, verosimilitudes o que se mantienen a través del tiempo conservadas

Contenido

Pág.

Resumen V

Abstract.....VI

Introducción..... 7

Problema de Investigación .....
13 1.1 Planteamiento del Problema .....
13

Objetivos 17
1.2 Objetivo General 17
1.3 Objetivos específicos 17

Marco Teórico 18

Diseño Metodológico 32
1.4 Población y Muestra..... 34

Resultados 35

Discusiones 82

Conclusiones y Recomendaciones..... 84
1.5 Conclusiones..... 84
1.6 Recomendaciones..... 85

Referencias Bibliográficas..... 86

Lista de figuras

Pág.

Imagen 1. estructura molecular del Sars-CoV-2..... 20
Imagen 1.1 introducción del virus SASR-CoV a la célula..... 23
imagen 2. distribución de secuencias en el mundo. Fuente: (base de datos propia) 36
Imagen.3Distribución de secuencias por continentes fuente: (base de datos propia) 36
Imagen.4Distribución de cepas por continentes}. Fuente: (base de datos propia).37
imagen.5Filogenia de la proteína S. Fuente:(base de datos propia). 39
imagen 6 Análisis filogenético de la cepa gamma para la proteína S. Fuente: (base de datos propia). 39
Imagen 7 Análisis filogenético de la cepa delta para la proteína S (base de datos propia). 40
Imagen 8 análisis filogenético de la cepa mu para la proteína S. Fuente:(base de datos propia). 41
imagen 9 Análisis filogenético de la cepa Alpha para la proteína S. Fuente: (base de datos propia). ..... 42 Imagen 10 análisis filogenético para la proteína Orf10. Fuente: (base de datos propia). 44
Imagen 11 análisis filogenético de la cepa Alpha para la proteína Orf10. Fuente: (base de datos propia). 45
Imagen 12 Análisis filogenético de la cepa delta para la proteína Orf10. Fuente: (base de datos propia) 46
Imagen 13 Análisis filogenético de la cepa Mu para la proteína Orf10. Fuente: (base de datos propia). ..... 47 imagen 14 Análisis filogenético de la cepa gamma para la proteína Orf10..... 47 Imagen 15 Análisis filogenético de la cepa delta para la proteína Orf8 (base de datos propia). ..... 49 Imagen 16 Análisis filogenético de la cepa Alpha para la proteína Orf8. Fuente: (base de datos propia). ..... 50 Imagen 17 Análisis filogenético de la cepa gamma para la proteína Orf8. Fuente: (base de datos propia). ..... 51 Imagen 18 Análisis filogenético de la cepa Mu para la proteína Orf8 (base de datos propia). 52
Imagen 19 Análisis filogenético de la cepa gamma para la proteína Orf8 (base de datos propia). 53
Imagen 20 Análisis filogenético para la proteína Orf7a. Fuente: (base de datos propia). 55
Imagen 21 Análisis filogenético de la cepa Alpha para la proteína Orf7a. Fuente: (base de datos propia). 56

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

(base de datos propia). 56
Imagen 23 Análisis filogenético de la cepa gamma para la proteína Orf7a: Fuente: (base de datos propia). 57
Imagen 24 Análisis filogenético de la cepa Mu para la proteína Orf7a Fuente: (base de datos propia). 58
Imagen 25 Análisis filogenético para la proteína Orf6 (base de datos propia). 59
Imagen 26 Análisis filogenético de la cepa Mu para la proteína Orf6. Fuente: (base de datos propia). 60
Imagen 27 Análisis filogenético de la cepa gamma para la proteína Orf6. Fuente: (base de datos propia). 60
Imagen 28 Análisis filogenético de la cepa Alpha para la proteína Orf6. Fuente: (base de datos propia). 61
Imagen 29 Análisis filogenético de la cepa delta para la proteína Orf6. Fuente:(base de datos propia). ..... 62 Imagen 30 Análisis filogenético para la proteína Orf3a. Fuente: (base de datos propia). ..... 64 Imagen 31 Análisis filogenético de la cepa Mu para la proteína Orf3a. Fuente: (base de datos propia). 65
Imagen 32 Análisis filogenético de la cepa Alpha para la proteína Orf3a. Fuente: (base de datos propia). 66
Imagen 33 Análisis filogenético de la cepa delta para la proteína Orf3a. Fuente: (base de datos propia). 67
Imagen 34 Análisis filogenético de la cepa gamma para la proteína Orf3a. Fuente: (base de datos propia). 68
Imagen 35 Análisis filogenético para la proteína M. Fuente: (base de datos propia). 69
Imagen 36 Análisis filogenético de la cepa Alpha para la proteína M. Fuente: (base de datos propia). 70
Imagen 37 Análisis filogenético de la cepa delta para la proteína M. Fuente: (base de datos propia). 70
Imagen 38 Análisis filogenético de la cepa Mu para la proteína M. Fuente: (base de datos propia). 71
Imagen 39 Análisis filogenético para la proteína E. Fuente: (base de datos propia). 72
Imagen 40 Análisis filogenético de la cepa Alpha para la proteína E (base de datos propia). 73
Imagen 41 Análisis filogenético de la cepa Delta para la proteína E. Fuente: (base de datos propia). 74
Imagen 42 Análisis filogenético de la cepa gamma para la proteína E. Fuente: (base de datos propia). ..... 74 Imagen 43 Análisis filogenético de la cepa Mu para la proteína E Fuente: (base de datos propia). 75
Imagen 44 Análisis filogenético para la proteína N. Fuente: (base de datos propia).

..... 76

Imagen 45 Análisis filogenético de la cepa Alpha para la proteína N (base de datos propia). 77
Imagen 46 Análisis filogenético de la cepa gamma para la proteína N. Fuente: (base de datos propia).
..... 78 imagen 47
Análisis filogenético de la cepa Mu para la proteína N. Fuente: (base de datos propia). 79
imagen 48 Análisis filogenético de la cepa delta la para proteína N (base de datos propia). 79

1. Problema de Investigación

1.1 Planteamiento del Problema

Anteriormente, los coronavirus no eran considerados altamente patógenos(1), lo que fue cambiando a medida que aparecieron como causantes de síndromes espiratorios severos; en el 2002, con la aparición del SARS-CoV, que dio lugar a un total de 8.098, personas enfermas, de las cuales, 774 murieron hasta julio de 2003(4), en 2012 con el MERS que llevo a un total de 2519 enfermos en todo el mundo de los cuales murieron 866, hasta el 31 de enero de 2020(5) y actualmente el SARS-CoV-2 (COVID-19) con un total de 5.770.023 casos confirmados, incluidas 4.368.986 muertes en el mundo hasta 10 de febrero de 2022(6).

Desde la aparición del COVID-19 en el mundo se vienen presentando diferentes variantes, algunas de ellas catalogadas como variantes de monitoreo, interés o preocupación siendo esta última catalogada de esta manera, debido a la evidencia existente de su mayor transmisibilidad, y probable gravedad de la enfermedad, entre las cuales se encontraron las variantes Alpha y Gamma y delta(7).

El COVID-19 es una enfermedad causada por un coronavirus en humanos vinculado a un mercado de animales vivos en la ciudad de Wuhan, provincia de Hubei en China(1) y catalogada por la OMS como una enfermedad epidémica emergente de tipo pandémico, presentada desde finales de 2019 y de la cual se han confirmado

un total de 207.621.607, incluidas 4.368.986 muertes en el mundo hasta 17 de agosto de 2021(6), siendo esta problemática uno de los principales retos de la salud

y de los científicos de Colombia y el mundo. Hasta el momento se conoce que es un virus conformado por 4 proteínas estructurales, S, M, N, E y proteínas ORF accesorias (orf1a, orf1b, orf3ª, orf6, orf7a, orf8, orf10) de las cuales se conoce que la proteína S, es la utilizada para la producción de vacunas contra el virus.

Actualmente, las vacunas existentes ayudan a neutralizar la infección, pero el problema radica en que han sido vacunas basadas en la glicoproteína S, que a pesar de ser eficaces en la generación de anticuerpos se presume, que con el tiempo, se podría afectar la eficacia, debido a las variantes que han aparecido y pueden aparecer en el tiempo(8), por lo que resulta importante conocer y comprender la evolución de cada una de las proteínas incluida la proteína S, que lleven a futuros estudios hacia nuevas diana terapéuticas y a su vez entender cuáles son las características genómicas, de las secuencias que circulan en Colombia, en relación con las del mundo.

Existen mucha información y estudios acerca de algunas características del virus, pero aún hay muchas cosas por comprender acerca del comportamiento y función de algunas de sus proteínas, su evolución y las implicaciones posibles, debido a cambios evolutivos en su genoma.

Por lo anterior realizaremos un estudio filogenético, a través de inferencia bayesiana, que nos lleve a conocer el desarrollo, entendimiento y mantenimiento del COVID-19 y a su vez determinar cepas que sean de mayor propagación y mayores cambios evolutivos.

1.2 Justificación

Este trabajo, está fundamentado en lo que dice Lokman et al (9)" al afirmar, que es necesario conocer las características únicas de los genes del SARS-CoV-2 a través, de la comprensión del comportamiento y evolución de su genoma, en las diferentes cepas del mundo, para comprender de una mejor manera la biología que los conforman. Teniendo en cuenta la frecuencia de las mutaciones de algunos genes del SARS-CoV-2 que predominan en diferentes regiones geográficas(10) y que han dado lugar a diferentes cepas.

Por lo anterior al encontrar un virus de Novo que atenta contra la salud del ser humano es necesario un seguimiento y un estudio exhaustivo para entenderlo y tratarlo de la manera adecuada. Por esta razón este trabajo permitirá un conocimiento más amplio, acerca del SARS-CoV-2, debido a que pretende indagar, el recorrido evolutivo de este virus en Colombia en relación con mundo, conociendo así, la dinámica espacial del virus, las verosimilitudes y a su vez los diferentes linajes evolutivos existentes, determinando de esta forma las características genómicas de las variantes en Colombia, lo que podría repercutir en futuras investigaciones, hacia el mejoramiento del diagnóstico y dianas terapéuticas.

Por otro lado, a través de este estudio se establecerán variaciones, cambios y relaciones evolutivas y se identificarán posibles genes útiles para el diseño de

primers de detención de variantes en Colombia y los relacionados con estas, en el

mundo y de esta forma brindar información necesaria y relevante que podrían influir en los tratamientos actuales y a futuro.

2. Objetivos

2.1 Objetivo General

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

2.2 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.

3. Marco Teórico

3.1 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).

3.2 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).

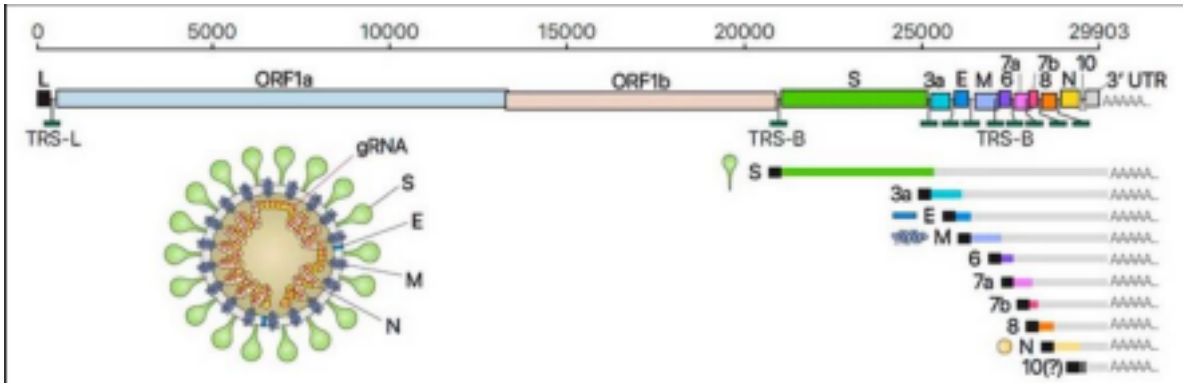

Imagen 1. estructura molecular del Sars-CoV-2. Fuente:(18)

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.

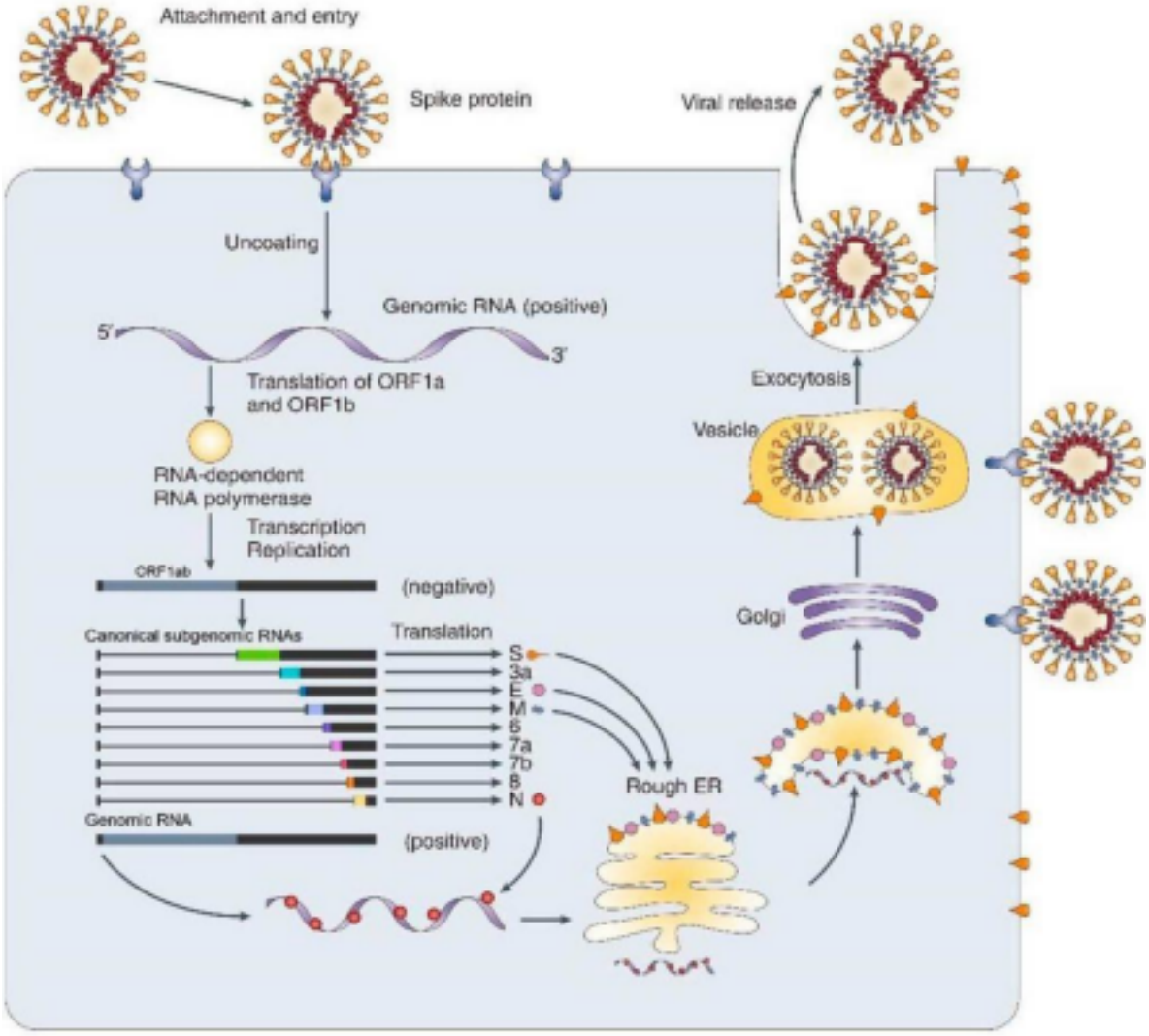

Imagen 1.1 introducción del virus SASR-CoV a la célula. Fuente: (18)

3.3 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 traves del tiempo hasta el dia 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).

3.4 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 traves 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 preocupacion en algun 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 ademas de cumplir con los anteriores criterios, popseian 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.

4. Diseño Metodológico

4.1 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.

4.2 Método de investigación

4.2.1 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.

4.2.2 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.

4.2.3 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.

4.3 Población y Muestra

4.3.1 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.

4.3.2 Criterio de Exclusión.

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

4.4 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.

5. Resultados

5.1 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.

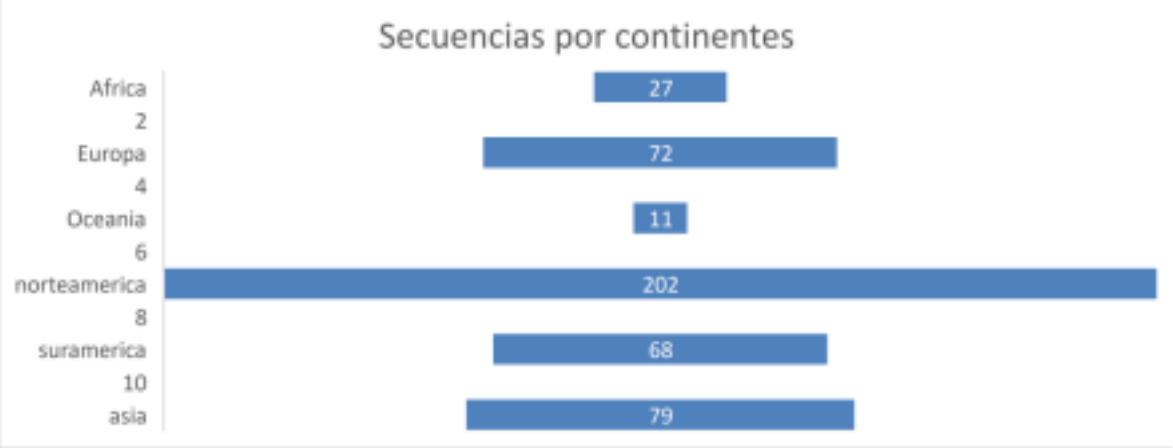

Imagen 3. Distribución de secuencias por continentes. Fuente: base de datos propia.

6
5
4
3
2
1
0

30 25 20 15 10 5
0

30 25 20 15 10 5
0

Oceania

Análisis de las secuencias

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.

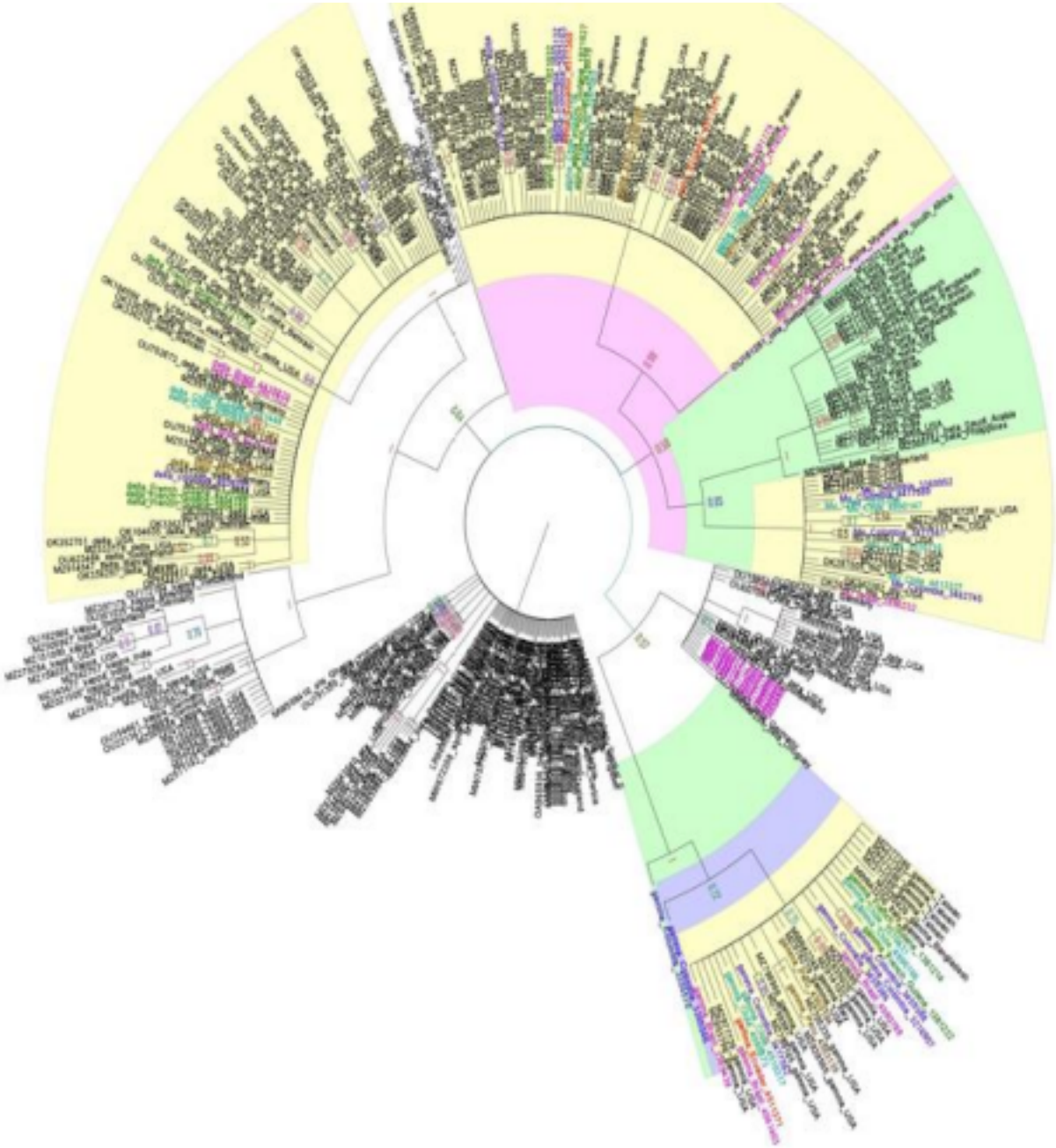

imagen.5Filogenia de la proteína S. Fuente: base de datos propia. Árbol filogenético obtenido usando el método de Inferencia Bayesiana para el gen S. Los números corresponden a valores de probabilidades posteriores bayesianas. Los nombres en color azul resaltados en color beige corresponden a la ubicación

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.

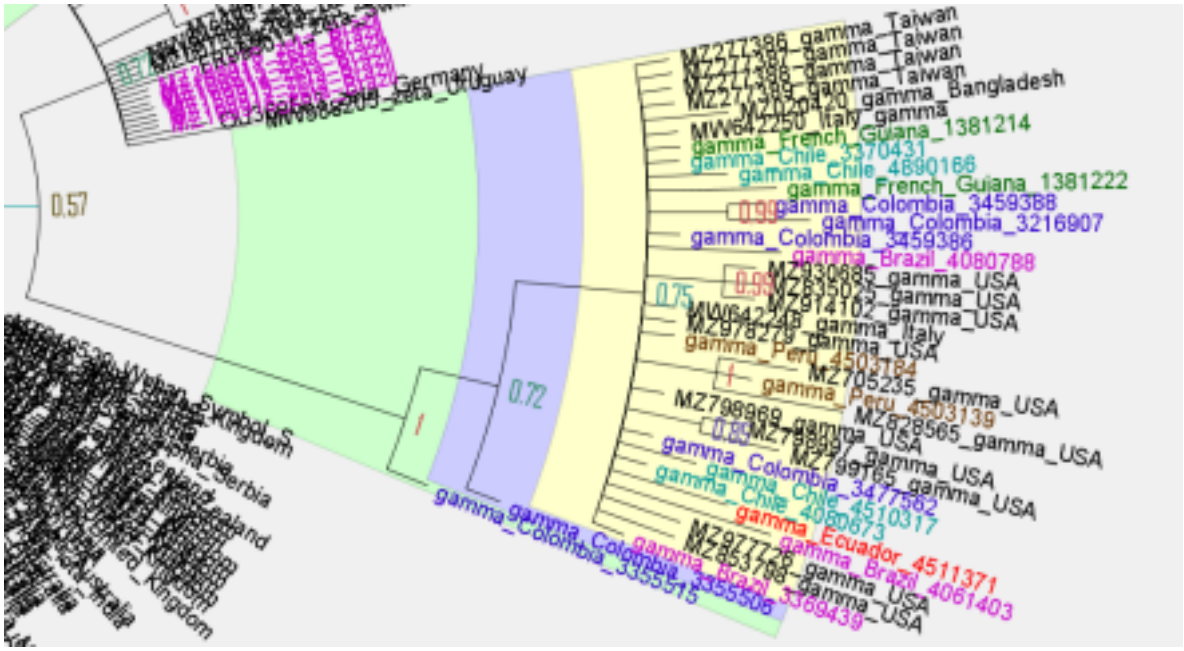

imagen 6 Análisis filogenético de la cepa gamma para la proteína S. Fuente: base de datos propia.

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.

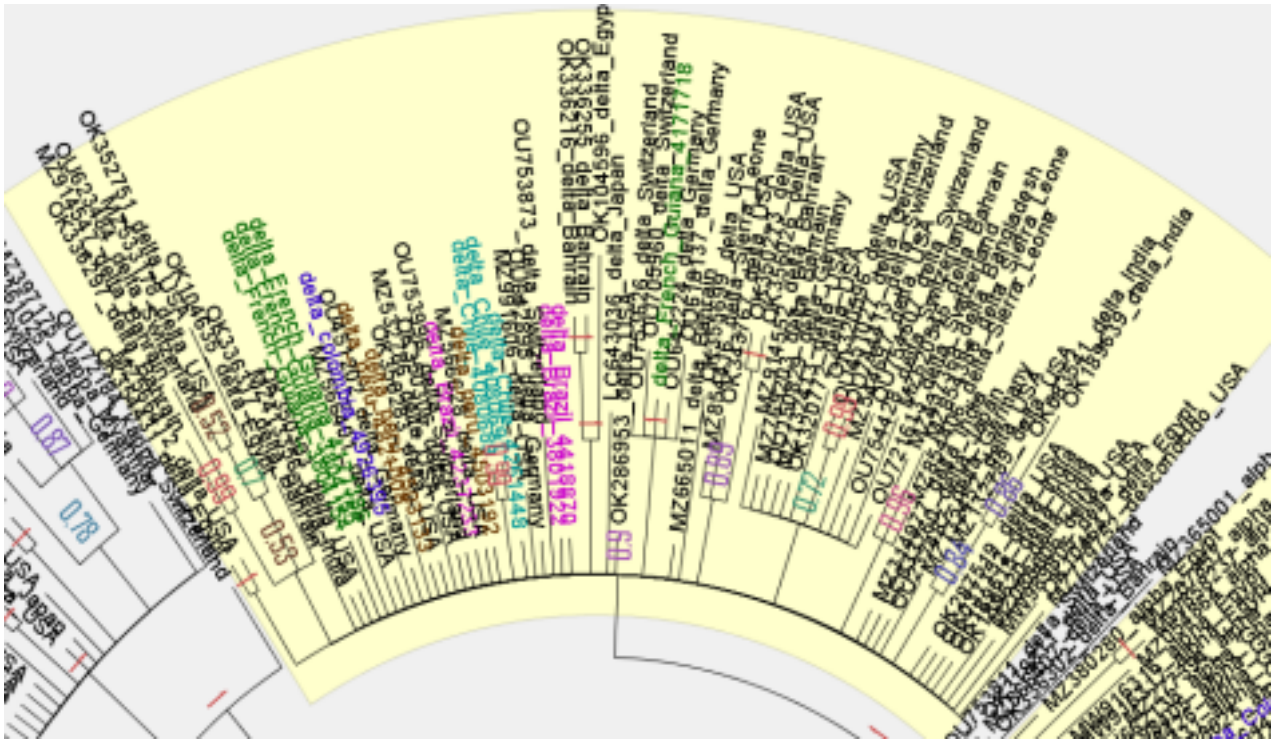

Imagen 7 Análisis filogenético de la cepa delta para la proteína S (base de datos propia).

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.

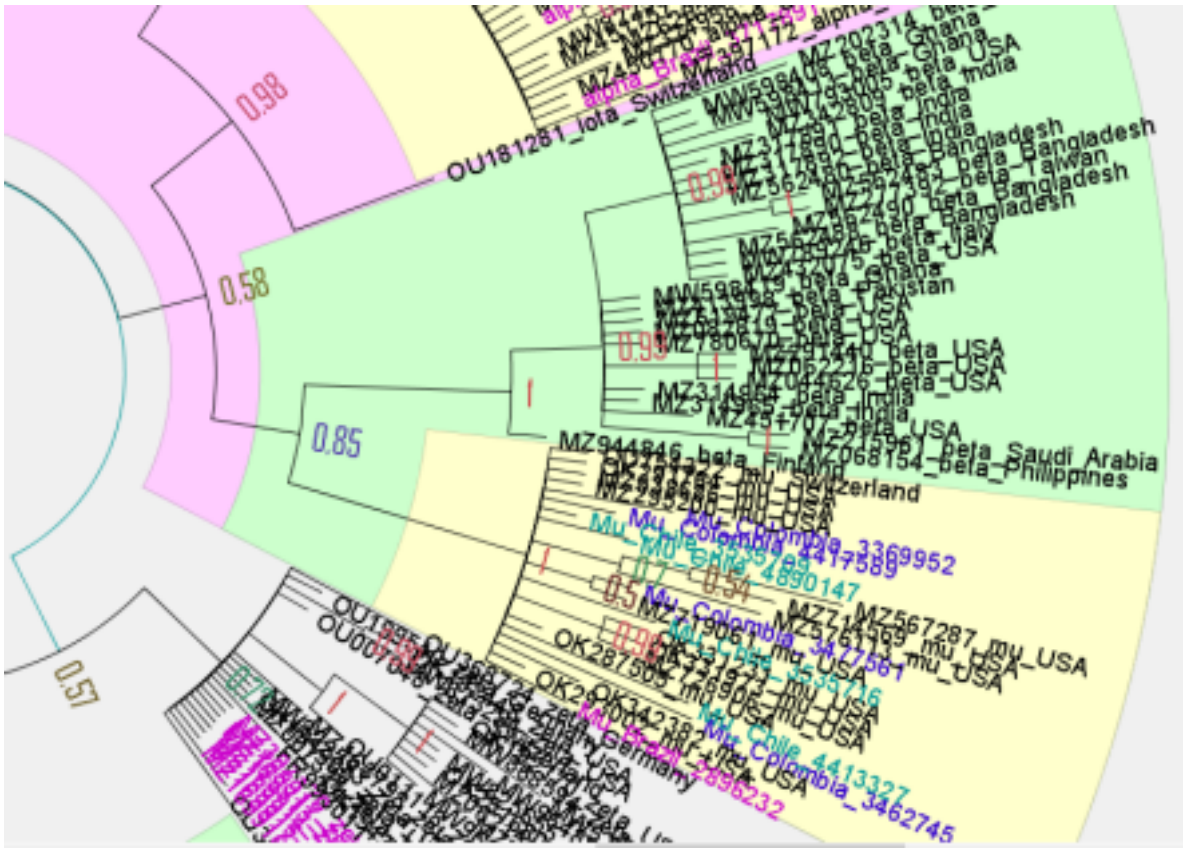

Imagen 8 análisis filogenético de la cepa mu para la proteína S. Fuente: base de datos propia.

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.

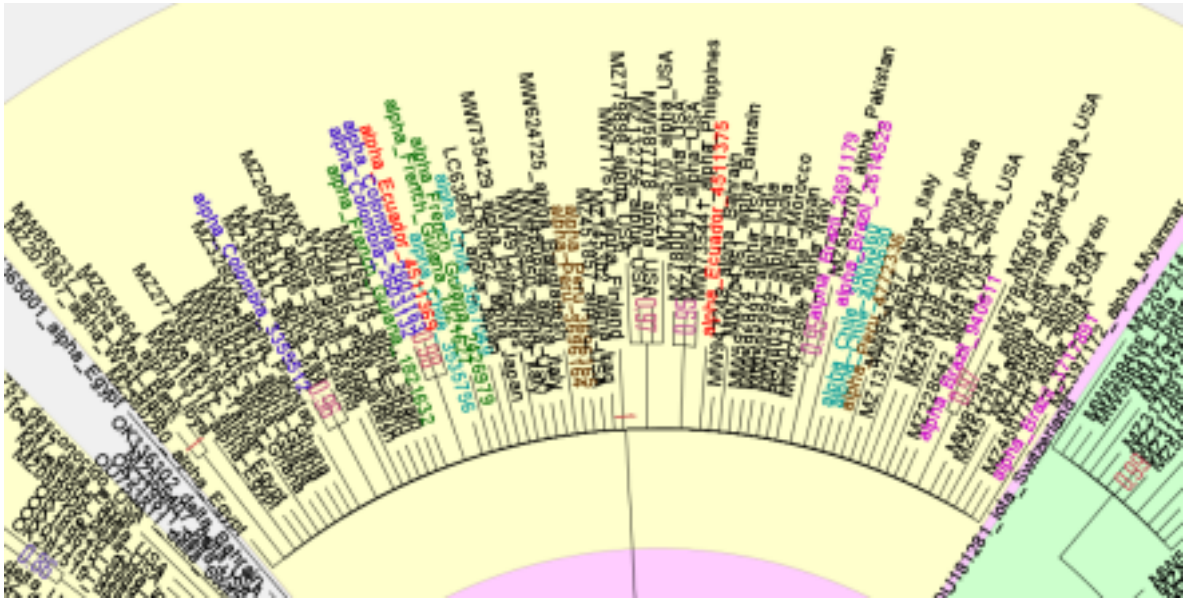

imagen 9 Análisis filogenético de la cepa Alpha para la proteína S. Fuente: base de datos propia.

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.

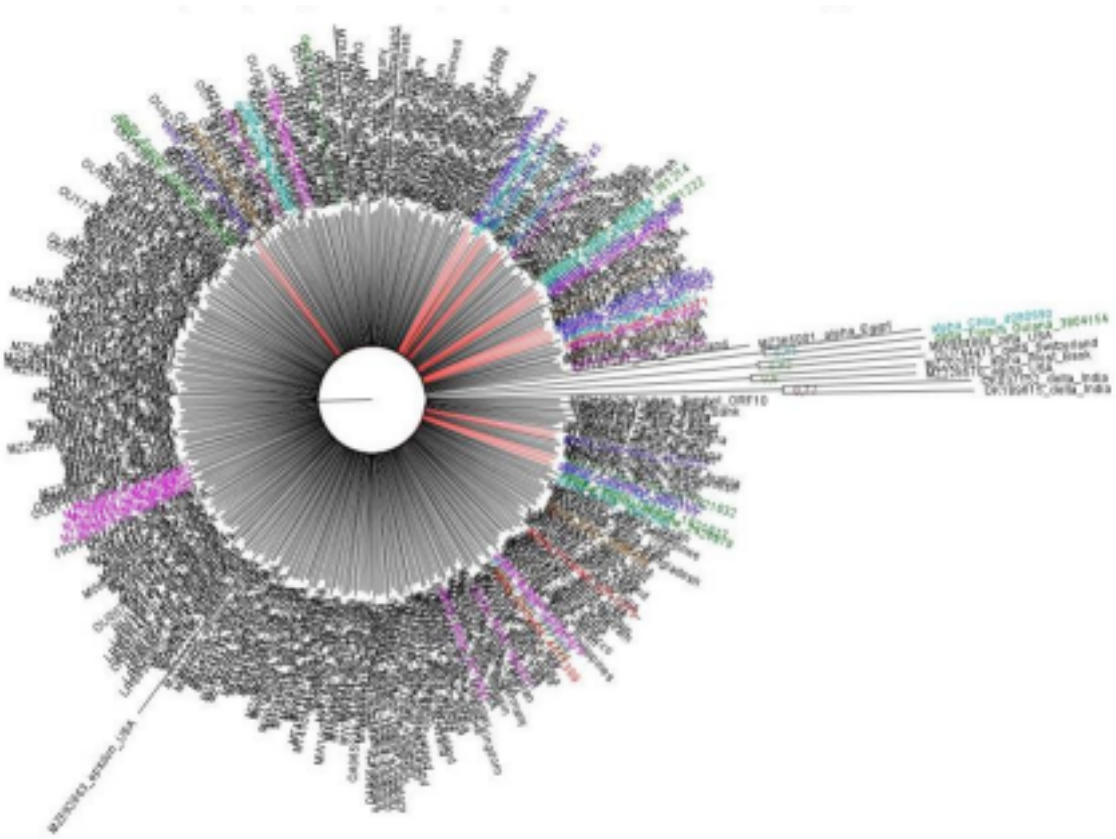

Imagen 10 análisis filogenético para la proteína Orf10. Fuente: base de datos propia.

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.

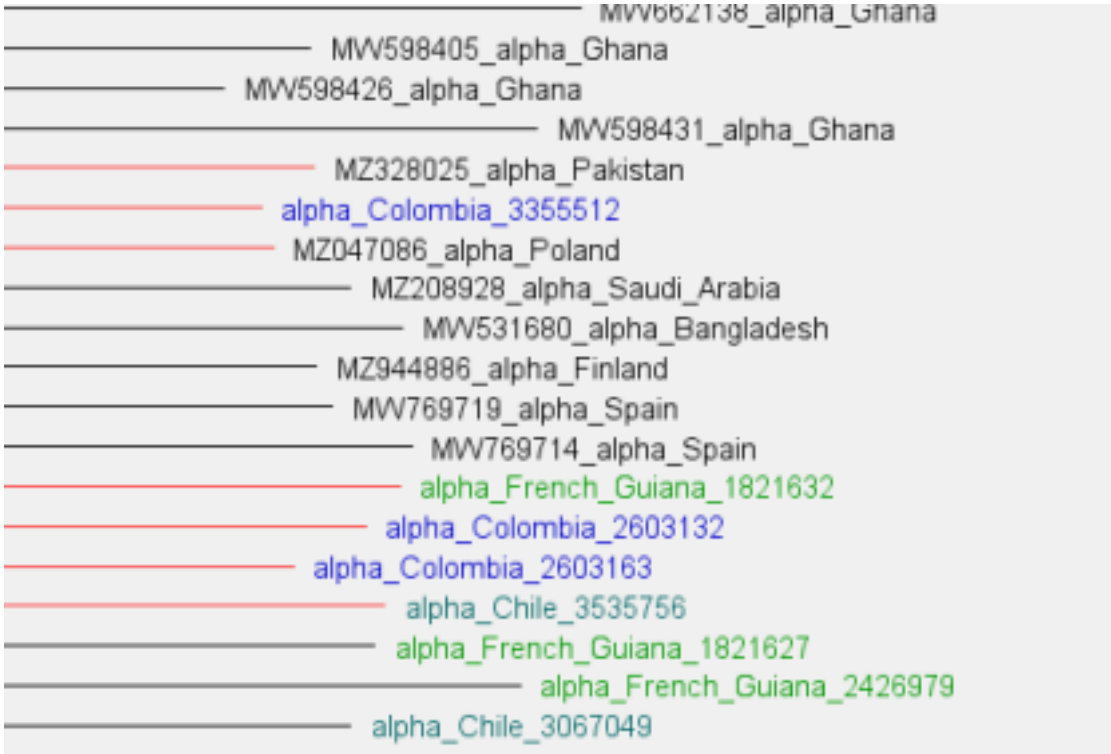

Imagen 11 análisis filogenético de la cepa Alpha para la proteína Orf10. Fuente: base de datos propia.

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

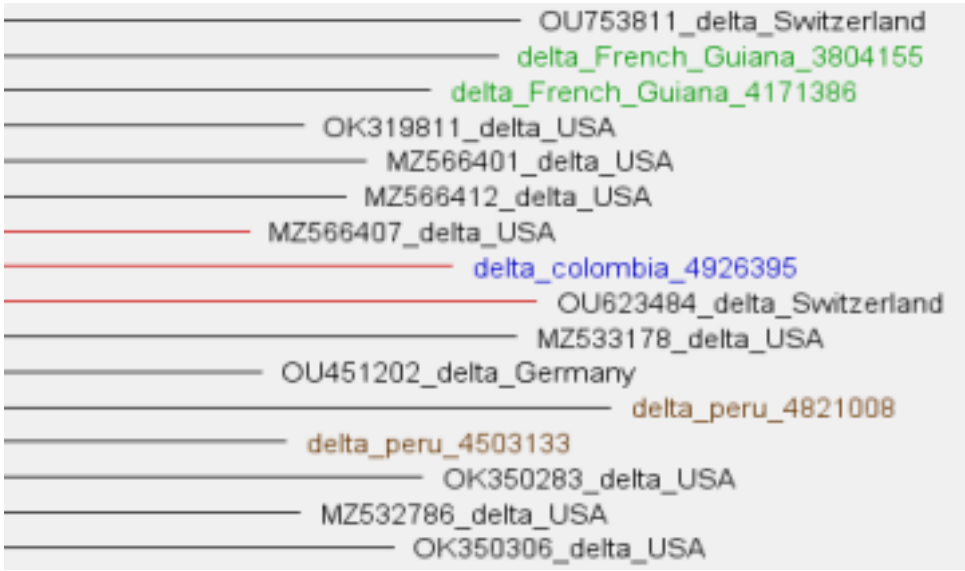

Imagen 12 Análisis filogenético de la cepa delta para la proteína Orf10. Fuente: base de datos propia.

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

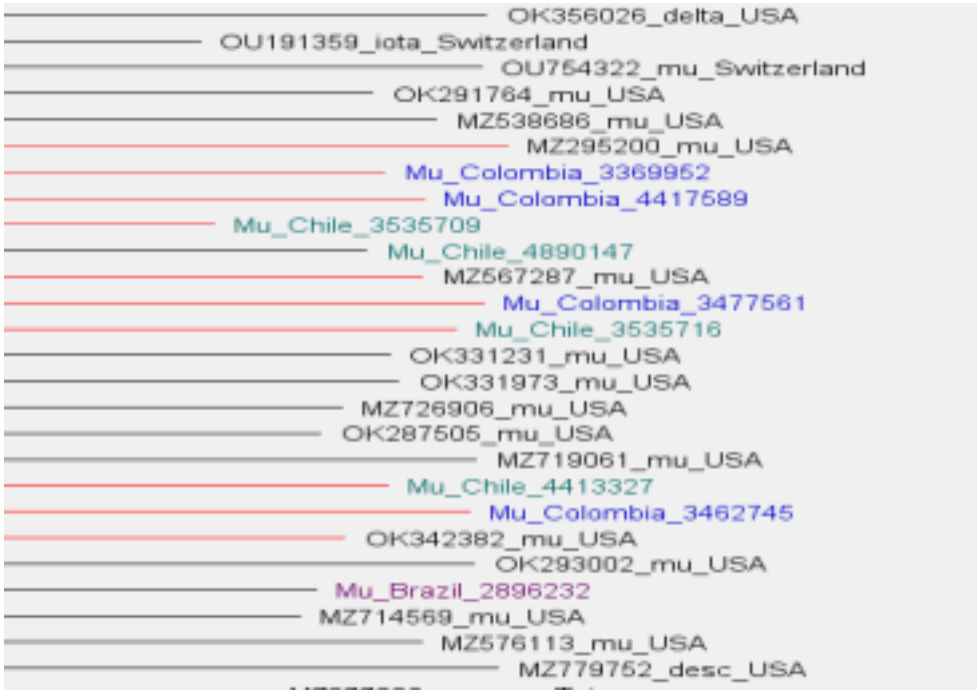

Imagen 13 Análisis filogenético de la cepa Mu para la proteína Orf10. Fuente: base de datos propia.

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

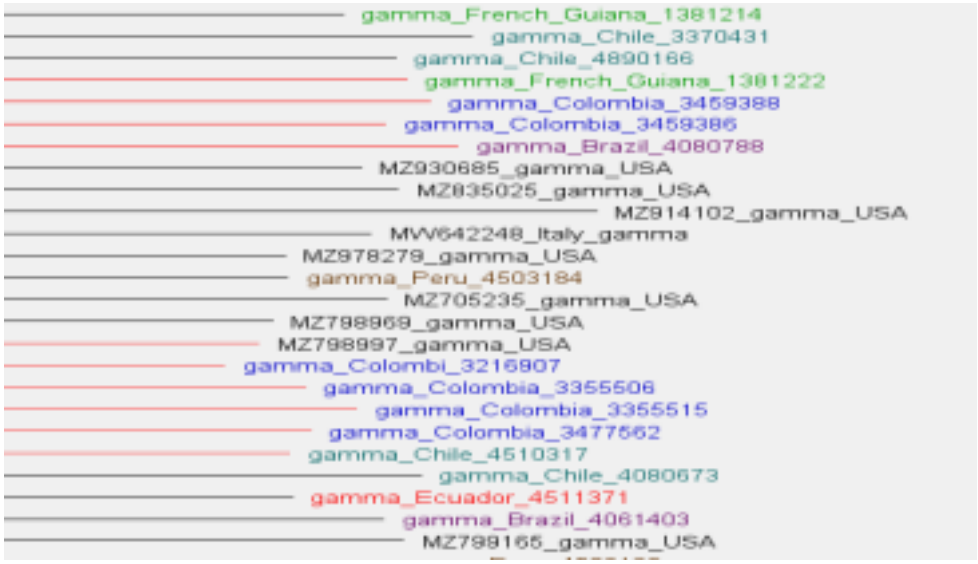

imagen 14 Análisis filogenético de la cepa gamma para la proteína Orf10

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.

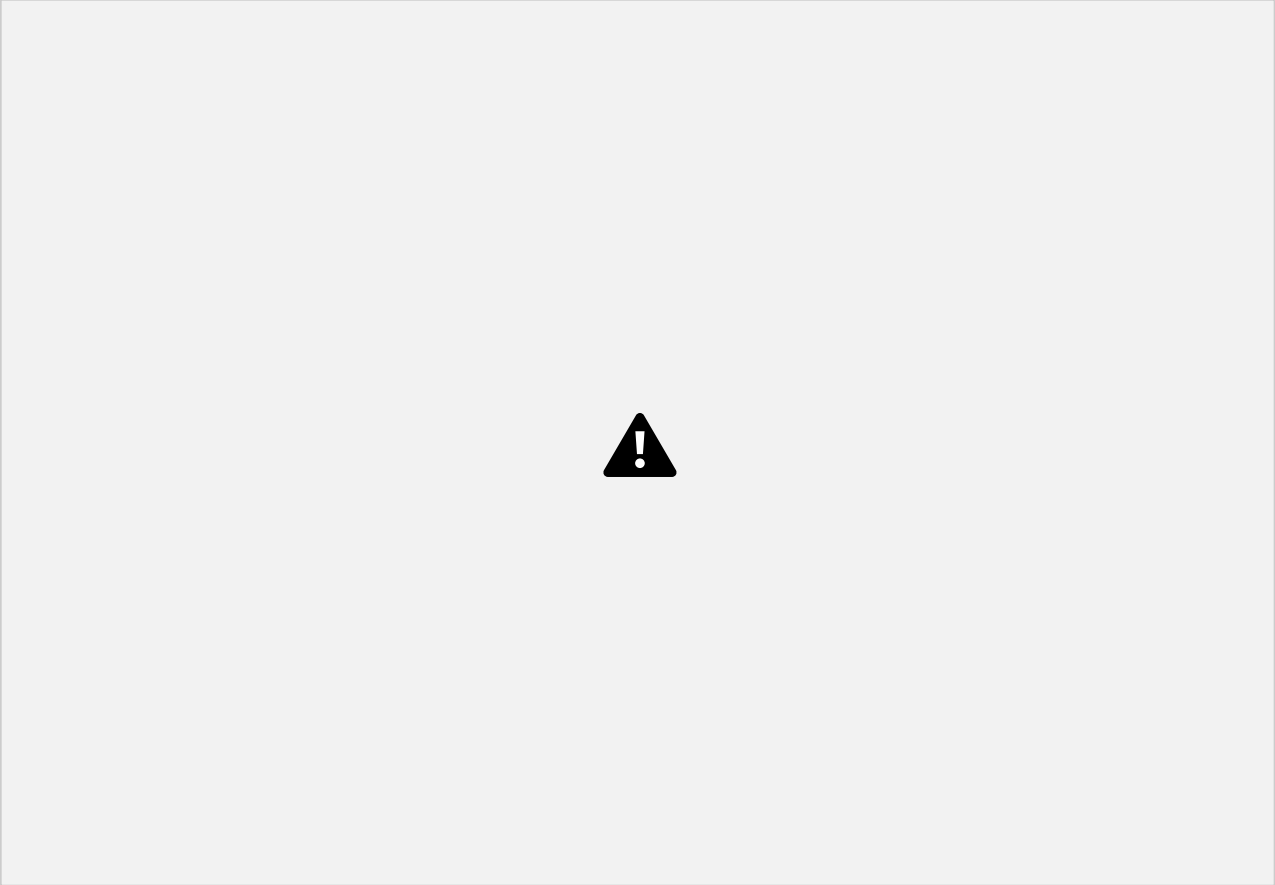

Imagen 25 Análisis filogenético para la proteína Orf6 (base de datos propia). Árbol filogenético obtenido, usando el método de Inferencia Bayesiana, para el gen

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