



Guatemala, 14 de diciembre, 2020

Señor Director Dr. Félix Alan Douglas Aguilar Carrera Director General de Investigación Universidad de San Carlos de Guatemala

Señor Director:

Adjunto a la presente el informe final "Secuenciación genómica de SARS-CoV-2 en pacientes guatemaltecos para el monitoreo de la diversidad y evolución viral como herramienta epidemiológica", coordinado por la MSc Elizabeth Solózano Ortiz y avalado por el Instituto de Investigaciones del Centro Universitario de Zacapa.

Este informe final fue elaborado con base en la guía de presentación de la Dirección General de Investigación, el cual fue revisado su contenido en función del protocolo aprobado, por lo que esta unidad de investigación da la aprobación y aval correspondiente.

Así mismo, el coordinador(a) del proyecto, se compromete a dar seguimiento y cumplir con el proceso de revisión y edición establecido por Digi del **informe final y del manuscrito científico.** El manuscrito científico debe enviarse, por el coordinador(a) del proyecto, para publicación al menos en una revista de acceso abierto (*Open Access*) indexada y arbitrada por expertos en el tema investigado.

Sin otro particular, suscribo atentamente.

"Id y enseñad a todos"

Firma

Elizabeth Solórzano Ortiz MSc.

Coordinadora del proyecto de investigación

Firma y sello

Manuel Barrios Izás, Dr.





Universidad de San Carlos de Guatemala Dirección General de Investigación

Informe final

Secuenciación genómica de SARS-CoV-2 en pacientes guatemaltecos para el monitoreo de la diversidad y evolución viral como herramienta epidemiológica

Equipo de investigación

Elizabeth Solórzano Ortiz

Carmen Lucía Yurrita Obiols

Barbara Beatriz Moguel Rodríguez

Manuel Barrios Izás

Carlos Alberto Montenegro

Guatemala 14 de diciembre 2020

Instituto de Investigaciones, Centro Universitario de Zacapa





DSc. Félix Alan Douglas Aguilar Carrera Director General de Investigación Ing. Agr. MARN Julio Rufino Salazar Coordinador General de Programas Elizabeth Solórzano Ortiz Coordinadora Proyecto Carmen Lucía Yurrita Obiols Barbara Beatriz Moguel Rodríguez Manuel Barrios Izas Carlos Alberto Montenegro Universidad de San Carlos de Guatemala, Dirección General de Investigación, 2020. El contenido de este informe de investigación es responsabilidad exclusiva de sus autores. Esta investigación fue cofinanciada por la Dirección General de Investigación de la Universidad de San Carlos de Guatemala a través de la Partida Presupuestaria 4.8.63.4.41. durante el año 2020





Índice	4
Resumen	6
Palabras clave	7
Abstract and keyword	7
Introducción	8
Planteamiento del problema	9
Preguntas de investigación	10
Delimitación en tiempo y espacio	10
Marco teórico	12
Estado del arte	14
Objetivo general.	15
Objetivos específicos	15
Materiales y métodos	16
Vinculación, difusión y divulgación	24
Resultados	26
Análisis y discusión de resultados:	35
Conclusiones	38
Impacto esperado	39
Referencias	39
Apéndices	42

Índice de Cuadros

Cuadro 1. Cronograma de actividades

Cuadro 2. Resumen de asignación de Clados en el Sistema de GISAID y mutaciones relevantes.

Cuadro 3: Análisis de Linajes en Pangolin 2.0 de acuerdo con la clasificación de Rambaut, A., et al., 2020.





Índice de figuras (fotografías, gráficas, diagramas)

Figura 1. Ruta bioinformática de análisis Primarios de las lecturas obtenidas del proceso de secuenciación. El color fucsia corresponde a cuadros que integran tipos de archivos, y el color azul a procesos y softwares utilizados.

Figura 2. Modelos en 3D de la proteína de Espiga (Spike protein), desarrollada en la aplicación CoVsurver de GISAID.

Figura 3. Filogenia y distribución espacial y temporal de los linajes según Rambaut de las muestras de SARS-CoV-2 evaluadas en Guatemala.

Figura 4. Mapa de distribución de linajes de de SARS-CoV-2 de Pangolin, presentes en Guatemala y 8 países sudamericanos.

Figura 5. Mapa de proyección de linajes de SARS-CoV-2 de Guatemala y Sudamérica construido en el software TreeTom.

Apéndices

Apéndice 1. Aval Bioético.

Apéndice 2. Valores de CT de 90 muestras virales incluidas en la Secuenciación.

Apéndice 3. POE Bioseguridad para el transporte de muestras.

Apéndice 4. Instrumento para registro de datos de los productos de Extracción de ARN viral.

Apéndice 5. Instrumento para registro de datos de los productos de Extracción de ARN viral.

Apéndice 6. Análisis de asignación de clados y análisis de diversidad en GISAID.





Título del proyecto: Secuenciación genómica de SARS-CoV-2 en pacientes guatemaltecos para el monitoreo de la diversidad y evolución viral como herramienta epidemiológica

1. Resumen

Contar con un análisis preciso de las secuencias genómicas del Betacoronavirus SARS-CoV-2, ha sido muy importante para el desarrollo de las estrategias que intentan contener la propagación del virus a nivel mundial, desde el primer paciente detectado en la ciudad de Wuhan, China. La secuenciación de los genomas de SARS-CoV-2 permiten rastrear las rutas de transmisión a nivel local y mundial, así mismo ayudan a determinar la velocidad de cambio del virus a medida que se propaga.

El estudio que se presenta a continuación es el primer proyecto de secuenciación de SARS-COV-2 en Guatemala, el cual nos permitió conocer las variantes virales que circulan en el país, sus rutas de transmisión y la evolución del mismo en los pacientes guatemaltecos. En Guatemala no se conocían las características genómicas de los linajes virales circulantes, debido a que únicamente se habían secuenciado en el extranjero 10 genomas virales aislados de los primeros pacientes guatemaltecos, sin existir algún estudio asociado a estas secuencias.

Este proyecto generó 89 secuencias genómicas de SARS-CoV-2, aisladas de pacientes guatemaltecos y el análisis se complementó con 10 genomas previamente secuenciados por el CDC en Estados Unidos. Posteriormente en este estudio se desarrollaron análisis filogenéticos, filodinámicos y espaciales de los linajes y clados virales identificados, a través de análisis de genómica comparativa utilizando secuencias de SARS-CoV-2 provenientes de 8 países de América del Sur, análisis de la diversidad genética del virus en el país. Además, se generaron 2 modelos tridimensionales de la proteína S observándose una alta incidencia de mutaciones en la proteína de los genomas muestreados en el país, pero que no afectan la región de acoplamiento al receptor ACE2 en las células humanas.

Los análisis de asignación de clados y linajes reflejaron una alta diversidad de los virus SARS-CoV-2 que circulan en la región metropolitana del país, observándose un recambio de clados a lo largo del desarrollo de la pandemia en el país. Además, se pudieron identificar dinámicas de transmisión local. Es importante resaltar que, muestreos más exhaustivos son necesarios para el desarrollo de análisis concluyentes.

Adicionalmente, dentro de los resultados obtenidos podemos resaltar el establecimiento de alianzas y colaboraciones institucionales, relevantes para el proceso de fortalecimiento de la vigilancia epidemiológica con base en información genética, en el país, que aporta información relevante para el desarrollo de estrategias de contención más holísticas.





2. Palabras clave

Covid-19, Secuenciación Genómica, Filodinámica, SARS-CoV-2, diversidad genética.

3. Abstract

Having a precise analysis of the genomic sequences of the Betacoronavirus SARS-CoV-2 has been very important for the development of strategies, trying to contain the spread of the virus worldwide, since the first patient was detected in the city of Wuhan, China. The sequencing of SARS-CoV-2 genomes allows tracing routes of transmission both locally and globally, as well as helps to determine the spread rate of change of the virus.

The present study is the first SARS-COV-2 sequencing project in Guatemala, these results let to know the viral variants that circulate in the country, their transmission routes and its evolution in Guatemalan patients. In Guatemala, the genomic characteristics of the circulating viral lineages were not known, since only 10 viral genomes isolated from the first Guatemalan patients have been sequenced abroad, without any study associated with these sequences.

Among our results 89 SARS-CoV-2 genomic sequences isolated from Guatemalan patients were generated. The analysis was complemented with 10 genomes previously sequenced by the CDC in the United States. Later in this study, phylogenetic, phylodynamic and spatial analysis of the identified viral lineages and clades were developed, through comparative genomics analysis using SARS-CoV-2 sequences from 8 countries in South America, analysis of the genetic diversity of the virus in the country. In addition, 2 three-dimensional models of protein S were generated, observing a high incidence of mutations in the protein of the genomes sampled in the country, those mutations do not affect the ACE2 receptor docking region in human cells.

The analysis of clade and lineage assignment reflected a high diversity of SARS-CoV-2 viruses that circulate in the metropolitan region of the country, observing a change of clades throughout the development of the pandemic in the country. In addition, local transmission dynamics could be identified. It is important to note that a more exhaustive sampling is necessary for the development of conclusive analyzes.

Additionally, within the results obtained, we can highlight the establishment of alliances and institutional collaborations, relevant for the process of strengthening epidemiological surveillance based on genetic information, within the country, providing important information for the development of more holistic containment strategies.





1. keyword

Covid-19, Genomic Sequencing, Phylodinamic, SARS-CoV-2, genetic diversity

4. Introducción

El análisis de las secuencias genómicas del coronavirus SARS-CoV-2, ha sido muy importante para el desarrollo de las estrategias que intentan contener la propagación del virus a nivel mundial. Contar con los genomas virales secuenciados de los primeros pacientes detectados en Wuhan permitió: a) Establecer la identidad del agente etiológico de la enfermedad COVID-19 (por su acrónimo en inglés coronavirus disease 2019), y que se trataba de un nuevo coronavirus, b) El desarrollo de los métodos moleculares de diagnóstico que, actualmente son el estándar, c) La identificación de regiones genómicas asociadas a la agresividad del virus, d) Proponer los mecanismos y origen de la infección y, e) brindar información relevante para el desarrollo de vacunas y tratamientos (Andersen, Rambaut, Lipkin, Holmes, & Garry, 2020; Cui, Li, & Shi, 2019; Lam et al., 2020; Wu et al., 2020; Zhang, Wu, & Zhang, 2020).

Los virus mutan con rapidez, esto debido a características intrínsecas como su rápida proliferación. Durante su acelerado proceso de replicación, se generan errores de copiado en las nuevas secuencias genómicas, es decir mutaciones, que dejan una especie de huella, que permite trazar su historia y rutas de dispersión. La tasa de evolución de los genomas virales, es decir los cambios o mutaciones acumuladas en la composición del material genético a través del tiempo, está relacionada con la diversidad nucleotídica, una mayor tasa mutacional genera mayor diversidad nucleotídica y potencialmente tasas de evolución viral más rápidas. Esta condición tiene incidencia directa en la efectividad de tratamientos, pruebas para diagnóstico y vacunas, estudios actuales ya alertan respecto al hallazgo de mutaciones que podrían tener un impacto en el diagnóstico y en la capacidad infectiva del virus, sin embargo, más estudios son requeridos a nivel mundial para el desarrollo de conclusiones sólidas.

Actualmente, el banco de datos mundial de secuencias genómicas de Influenza (GISAID), se convirtió en el repositorio a nivel mundial de las secuencias de SARS-CoV-2, y cuenta a la fecha (diciembre 2020), con 231,171 genomas completos de SARS-COV-2. Las secuencias de ARN viral provienen de pacientes alrededor del mundo, pero con una clara representación preponderante de genomas de países desarrollados, esto debido a la rápida respuesta de los sistemas nacionales de salud en estos países, y la articulación de los mismos con la academia, lo que permitió la asignación de recursos estatales a la producción científica.

La base de datos GISAID se encuentra en un continuo crecimiento gracias a la información generada por grupos científicos alrededor del mundo. Esta información ha permitido monitorear en tiempo real el desplazamiento del virus, de una forma sin precedentes en la humanidad, así como ha contribuido a la comprensión de los mecanismos de esta nueva enfermedad y al desarrollo de tratamientos y medidas de control. El análisis filogenético de las secuencias virales también ha





permitido rastrear brotes en curso, evaluar la efectividad de medidas de contención e identificar nuevas estrategias de intervención (Dellicour et al., 2019; Gonzalez-Reiche et al., 2020).

En Guatemala y la mayoría de los países centroamericanos aún se desconocen las variantes genéticas del virus que circulan actualmente, el Centro para el Control y Prevención de Enfermedades -CDC- en Estados Unidos, desarrolló hace dos meses la secuenciación de las primeras 10 muestras virales de pacientes nacionales, sin embargo, estas corresponden a tan solo los primeros 15 días de la pandemia en el país, es decir a las primeras introducciones del virus en Guatemala, dichas secuencias fueron descargadas de la base de datos y analizadas en conjunto con las secuencias generadas en este estudio.

Debido a la importancia de conocer la variación genética de SARS-CoV-2 en nuestro país, este proyecto desarrolló: 1) La secuenciación de 89 genomas y para el análisis, se incluyeron 10 genomas secuenciados por CDC, dando un total de 99 genomas de SARS-CoV-2 de pacientes guatemaltecos analizados, 2) análisis filogenéticos de las variantes virales secuenciadas del país 3) Análisis comparativos con secuencias de SARS-CoV-2 de otras regiones del mundo, disponibles en bases de datos libres, 4) Análisis de la diversidad genética y evolución de los linajes virales circulantes en el país y más tarde al generar la publicación en una revista indexada se procederá a 5) depositar las secuencias generadas en bases de datos de acceso libre a nivel mundial, para que científicos alrededor del mundo puedan considerarlas en el desarrollo de tratamientos, estrategias de control y vacunas.

5. Planteamiento del problema

En nueve meses el COVID-19 pasó de una distribución restringida a la ciudad de Wuhan, a una pandemia con 72.253.804 casos en 191 regiones. Los registros continúan en ascenso y el aumento en la velocidad de la proliferación viral suele acompañarse de un incremento en su tasa de evolución, lo que puede afectar su virulencia, infectividad y transmisibilidad. La alta capacidad de expansión de SARS-CoV-2 se debe a características codificadas en su genoma, como su alta afinidad por receptores de células humanas y su capacidad para evadir el sistema inmune (Andersen et al., 2020).

Estudios recientes muestran una alta diversidad en la composición nucleotídica del ARN en SARS-CoV-2 en algunos pacientes (Phan, 2020; Shen et al., 2020), razón por la cual recientemente el virus ha sido clasificado con base en análisis filogenéticos en 8 clados distintos, en la clasificación de GISAID, siendo estos: G, GR, GH, O, L, N, S y V y más de 20 linajes de acuerdo con la clasificación de Rambaut (Rambaut, A., et al., 2020).

Además, estudios de monitoreo de las mutaciones en la proteína S, proteína que permite la unión del virus a los receptores ACE2 de la célula humana, han mostrado cambios en la distribución temporal de una mutación conocida como D614G, dichos estudios muestran la sustitución temporal





y espacial de la variante D614, por la G614 e hipotetizan con base en sus hallazgos que esta cepa es más infectiva, pero no más agresiva que la D614, aunque sus resultados aún se encuentran en un proceso de revisión, y dichos patrones podrían deberse a efectos de fundador, estos alertan respecto a la importancia del monitoreo de los cambios nucleotídicos virales a medida que la enfermedad se expande.

La universidad de Zhejiang por otro lado, también se encuentra desarrollando estudios de la agresividad del virus asociada a sustituciones nucleotídicas en la proteína S, sus resultados aún están en proceso de publicación pero comunicaciones en medios, realizadas por la líder del grupo de investigación, la viróloga Li Lanjuan, advierte sobre diferencias en la agresividad de estos linajes virales, el comportamiento y letalidad de la infección y cómo dichas diferencias podrían explicar el comportamiento tan agresivo de COVID 19 en Europa y Nueva York. Por otro lado, algunos estudios recientes muestran el efecto negativo (disminución) en la sensibilidad de algunos métodos de diagnóstico, que generan algunas de las mutaciones acumuladas en las regiones blanco de hibridación (Álvarez-Díaz et al., 2020).

La secuenciación de los genomas de SARS-CoV-2 es necesaria para rastrear rutas de transmisión a nivel mundial y determinar qué tan rápido está cambiando el virus a medida que se propaga. El conocimiento de las variantes virales que circulan en un país se hará cada vez más indispensable para conocer la efectividad de tratamientos, métodos de diagnóstico y vacunas en cada región. En algunos países como Guatemala aún no se conocen las características genómicas de los linajes virales circulantes, debido a que hasta antes del desarrollo de este proyecto solo se encontraban secuenciados 10 genomas virales, pertenecientes únicamente a introducciones tempranas.

El siguiente proyecto tiene el objetivo de llenar los vacíos de información de los virus que circulan en el país, antes mencionados, a través del desarrollo de la secuenciación genómica, el análisis de la diversidad nucleotídica y evolución, de 89 muestras de ARN viral de SARS-CoV-2, que han circulado en el país desde el 13 de marzo del 2020.

6. Preguntas de investigación

¿Qué linajes y cuál es la diversidad genética del virus SARS-CoV-2 que circula en Guatemala? ¿Qué refleja la distribución espacial de la diversidad y la filogenia de los linajes encontrados, respecto a la epidemiología de la enfermedad en el país?

7. Delimitación en tiempo y espacio

El proyecto dio inicio el 1 de agosto de 2020, sin embargo, desde julio se iniciaron los trámites para la obtención del aval bioético y la obtención de muestras. Inicialmente se tenía





contemplado realizar un solo muestreo correspondiente a los meses de agosto a octubre, sin embargo, la imposibilidad de contar con un alto número de eluidos de ARN que contarán con los criterios de inclusión de muestras para la secuenciación genómica hizo necesario realizar dos muestreos más durante los meses de octubre y noviembre, aunque al inicio esto fue una limitación, permitió enriquecer el análisis en el espacio temporal.

Por otro lado las dificultades de transporte impuestas por la situación que enfrentamos mundialmente, como resultado de la pandemia, retrasó considerablemente los procesos de compra de la mayoría de los reactivos utilizados para la secuenciación, debido a que en Guatemala, solamente se encuentran empresas distribuidoras de los mismo, aunado a esto, muchas de las empresas de manufactura a nivel mundial, generan reactivos e insumos para el diagnóstico molecular de SARS-CoV-2, este aumento en la demanda de productos a nivel mundial, generó un retraso mayor, que desplazó nuestro cronograma en el desarrollo de la secuenciación genómica. Por ello los análisis se generaron en un tiempo récord a nivel mundial.

Actividad/mes	Mes 1 Del 1 al 31 agosto	Mes 2 1-30 de septiembre	Mes 3 1-31 de octubre	Mes 4 1-30 de noviembre	Mes 5 1-14 de diciembre
Obtención de aval bioético	X	X			
Compra de reactivos e Insumos	X	X	X	X	
Obtención de Muestras (eluidos de ARN de muestras retrospectivas)	X	X	X	X	
Preparación de Secuenciación				X	
Secuenciación genómica				X	
Análisis bioinformático				X	X
Análisis Filogenético y de Diversidad nucleotídica				X	X
Desarrollo y entrega de los resultados e informes finales del proyecto a todas las autoridades competentes					X
Liquidación de fondos DIGI					X

Cuadro 1. Cronograma de actividades





8. Marco teórico

COVID-19

El 12 de diciembre del 2019 en la ciudad de Wuhan fue hospitalizado el primer paciente descrito como portador de la enfermedad del coronavirus 2019 (COVID-19), para ese momento el patógeno causante del brote de casos de síndrome respiratorio agudo era desconocido. Después de descartar los patógenos conocidos hasta el momento, se realizó la secuenciación metagenómica de ARN de una muestra de líquido de lavado broncoalveolar de un paciente ingresado el 26 de diciembre al hospital de Wuhan, dicho proceso, permitió la identificación de una nueva cepa de virus ARN de la familia Coronaviridae, que se denominó inicialmente coronavirus 'WH-Human 1'.

El análisis filogenético del genoma viral completo reveló que el virus estaba estrechamente relacionado (grado de similitud en la secuencia de nucleótidos del genoma, de 89.1%) con un grupo de coronavirus similares al SARS (género Betacoronavirus, subgénero sarbecovirus) que se habían encontrado previamente en murciélagos en China (Wu et al., 2020).

En un estudio clínico de los primeros casos registrados en Wuhan, que incluyó 41 pacientes hospitalizados hasta el 2 de enero de 2020, y cuya infección por SARS-CoV-2 fue corroborada por diagnóstico molecular utilizando el método RT-PCR y secuenciación genómica, la mayoría de los pacientes infectados eran hombres (30 [73%] de 41); menos de la mitad tenía enfermedades subyacentes (13 [32%]), incluida diabetes (ocho [20%]), hipertensión (seis [15%]) y enfermedades cardiovasculares (seis [15%]). La mediana de edad fue 49 años. Los síntomas comunes al inicio de la enfermedad fueron fiebre 98%, tos 76% y mialgia o fatiga 44%; síntomas menos comunes fueron la producción de esputo 28%, dolor de cabeza 8%, hemoptisis 5% y diarrea 3%. La disnea se desarrolló en el 55%, tiempo medio desde el inicio de la enfermedad hasta la disnea 8 días. El 63% presentaron linfopenia. Y los 41 pacientes tenían neumonía con hallazgos anormales en la TC de tórax. Las complicaciones incluyeron síndrome de dificultad respiratoria aguda 29%, ARNaemia 15%, lesión cardíaca aguda en el 12% e infección secundaria en el 10%. 13 (32%) pacientes fueron ingresados en una UCI y seis (15%) fallecieron (Wu et al., 2020). En comparación con los pacientes sin UCI, los pacientes con UCI tenían niveles plasmáticos más altos de IL2, IL7, IL10, GSCF, IP10, MCP1, MIP1A y TNFα (Huang et al., 2020).

Una de las principales limitantes en el desarrollo de tratamientos para la COVID 19, se debe a que, durante la expansión geográfica de la enfermedad, se han descrito diversidad de síntomas atípicos, y el aumento exponencial de casos ha complicado el desarrollo de registros sistemáticos, principalmente porque ha colapsado los sistemas de salud, y la alta infectividad ha impedido el desarrollo de autopsias. Actualmente, existen múltiples grupos científicos intentando consolidar la información sin precedentes, generada para la COVID-19, recientemente se publicó un esfuerzo por categorizar las manifestaciones clínicas extrapulmonares, entre estas, complicaciones trombóticas, disfunción y arritmia miocárdicas, síndromes coronarios agudos, lesión renal aguda, síntomas gastrointestinales, lesión hepatocelular, hiperglucemia y cetosis,





enfermedades neurológicas, síntomas oculares y complicaciones dermatológicas (Gupta et al., 2020).

Sin embargo, aun cuando nunca, tantos científicos se encontraron trabajando al unísono en la búsqueda de soluciones para el control de una enfermedad, y con tantos recursos tecnológicos a su disposición, la velocidad de la expansión de la enfermedad debido al nivel de globalización actual, generó una carrera en contravía que solo demuestra que aún muchos esfuerzos deben ser desarrollados por generar más información, que sea útil para el desarrollo de vacunas, tratamientos y estrategias de contención.

SARS-CoV-2

El SARS-CoV-2 es un virus de ARN monocatenario de sentido positivo, no segmentado y envuelto, que se incluye en la subfamilia de Sarbecovirus, Orthocoronaviridae, familia que se encuentra ampliamente distribuida en humanos y otros mamíferos. Su diámetro es de aproximadamente 65-125 nm, y se encuentra provisto de espinas proteicas, las cuales precisamente le confieren la forma de corona a la cual deben su nombre (Astuti & Ysrafil, 2020). SARS-CoV-2 tiene cuatro proteínas estructurales principales: 1. Glicoproteína espiga (S), 2. Glicoproteína de envoltura pequeña (E), 3. Glicoproteína de membrana (M) y 4. Proteína nucleocápside (N), además cuenta con varias proteínas accesorias. La glicoproteína S es una proteína transmembrana con un peso molecular de aproximadamente 150 kDa que se encuentra en la porción externa del virus. La proteína S forma homotrímeros que sobresalen en la superficie viral y facilitan la unión de la envoltura viral a las células huésped por atracción con la enzima convertidora de angiotensina 2 (ACE2) expresada en las células del tracto respiratorio inferior (Astuti & Ysrafil, 2020). Esta glicoproteína es escindida por la proteasa similar a la furina de la célula huésped en 2 subunidades, S1 y S2. La S1 es responsable de la determinación del rango de huésped y del tropismo celular con el dominio de unión al receptor, S2 funciona para mediar la fusión del virus en la transmisión de las células huésped (Astuti & Ysrafil, 2020).

Origen de SARS-CoV-2

Dada la similitud viral en la secuencia genómica con SARS-CoV-1, se manejaron muchas hipótesis del origen del virus, desde hipótesis conspirativas, de generación de un virus a través de métodos de ingeniería genética, hasta accidentes de laboratorio. Actualmente la evidencia científica que aportan los análisis de genómica comparativa y los análisis evolutivos, apuntan a un origen zoonótico, a partir de un ancestro común con coronavirus que se encuentran en murciélagos, y con un eslabón intermedio que se hipotetiza fue el pangolín (Lam et al., 2020; Zhang et al., 2020). Dicho origen zoonótico se plantea como el resultado de actividades humanas, como la ingesta y tráfico ilegal de vida silvestre en la región, y que posteriormente el virus dio un salto a la especie humana, circulando por un tiempo aún no determinado, de forma silenciosa, y que más tarde por eventos de selección natural, se fijó poblacionalmente una nueva variante viral con las características actuales que le capacitaron para generar esta pandemia (Andersen et al., 2020; Lam et al., 2020; Zhang et al., 2020)





Diversidad Genética de SARS-CoV-2

Como se mencionó anteriormente muchos países de Europa y Norteamérica han generado una gran cantidad de secuencias genómicas de los virus que circulan en su región, en menor cantidad dichas secuencias han sido generadas para el sur de américa y para la región centroamericana solamente Costa Rica y Panamá habían desarrollado el proceso hasta ahora, para Guatemala existen 10 genomas secuenciados a través de la tecnología de Oxford Nanopore, sin embargo dicha secuenciación fue generada en los Estados Unidos de Norteamérica, por el personal del CDC. A partir de la información generada en las bases de datos, se ha podido evidenciar la diversidad genética del virus que se ha generado a lo largo de su expansión geográfica y temporal, si bien SARS-CoV-2 no presenta particularmente tasas mayores de mutación a las de otros virus, la rapidez de su expansión ha permitido que estas se acumulen, evidenciándose ya diversidad geográfica (Shen et al., 2020) y también diversidad intra-huésped (Phan, 2020). Aun cuando la información generada en el área de secuenciación genómica es tan grande, su distribución geográfica es heterogénea, y aún existe un camino importante en los estudios que puedan establecer potenciales correlaciones entre las variantes genéticas y la agresividad viral o nivel de infectividad. Aun así, algunos estudios ya brindan evidencia del desplazamiento de variantes virales con mutaciones puntuales en regiones de relevancia como la proteína S (D614G), directamente involucrada con el proceso de ingreso del virus al huésped y que podría tener efectos en su agresividad (Korber et al., 2020) y otros han mostrado cambios en las regiones blanco de algunos métodos de diagnóstico molecular, los cuales podrían generar falsos negativos en el diagnóstico (Álvarez-Díaz et al., 2020). Dichos estudios alertan sobre la

importancia de conocer las cepas virales que se distribuyen en una región, para así acceder a soluciones diagnósticas, de tratamiento y de contención efectivas para cada región.

9. Estado del arte

El virus causante del síndrome respiratorio agudo severo Coronavirus 2 (SARS-CoV-2) identificado por primera vez en la ciudad de Wuhan, China ha tenido una rápida propagación a nivel mundial a pesar de las advertencias y los esfuerzos de agencias de salud por prevenirlo (Sironi, M., et al., 2020), estableciendo el nombre de esta enfermedad como COVID-19 (GISAID 2020). A la fecha se estiman 71.986.171 de casos y 1.609.413 muertes reportadas por COVID a nivel mundial, de este total 129.282 casos han sido reportados para Guatemala (https://coronavirus.jhu.edu/map.html visitado el 13 de dic de 2020), sin embargo, las limitaciones del sistema de salud nacional dejan dudas respecto a la existencia de un subregistro de casos y defunciones en el país. Los análisis filogenéticos han demostrado que SARS-CoV-2 es miembro del subgénero Sarbecovirus (género Betacoronavirus), al igual que SARS-COV (Zhou et al., 2020). Así mismo se observó que MERS-CoV (subgénero Merbecovirus) está más relacionado con el SARS-CoV-2 que con el SARS-CoV (Sironi, M., et al., 2020).

La plataforma GISAID fue creada en 2008 como una alternativa de acceso público científico para el intercambio de datos sobre influenza. Actualmente, GISAID juega un papel esencial en el





intercambio de datos para SARS-CoV-2. En este sitio hasta Universidad de San Carlos de Guatemala

diciembre del 2020 se han registrado 231,171 genomas completos de SARS-COV-2 provenientes de pacientes de todo el mundo (+4,616), excluyendo los de baja cobertura, dejando fuera 245,687 entradas (GISAID 2020). Existen diversos métodos de asignación de clados y linajes para el estudio de la dinámica evolutiva de virus, los dos más relevantes a nivel mundial, durante esta pandemia han sido el método de Rambaut (Rambaut, A., et al., 2020) y el método propio aplicado por GISAID, en este último 8 clados del virus se han identificado, de los cuales cuatro se han reportado para Sudamérica (G, GR, GH, y O). Dentro de las secuencias reportadas en esta base de datos se encuentran 10 secuencias de los primeros pacientes guatemaltecos, reportadas por el CDC de Atlanta, sin embargo, no existe ningún estudio hasta la fecha relacionado a estas. La tasa de generación de secuencias genómicas para el virus no tiene precedentes, por lo que la presentación de una nomenclatura para el virus SARS-COV-2 racional y dinámica, utilizando un marco filogenético para la identificación de los linajes que contribuyen activamente a la propagación fue un aporte muy importante (Rambaut, A., et al., 2020). Esta nomenclatura ayuda a rastrear y comprender los patrones y determinantes de la propagación global del SARS-CoV-2 al centrarse en los linajes de virus activos y aquellos que se propagan a nuevas ubicaciones (Rambaut, A., et al., 2020), por ello, los análisis presentados en este estudio emplean ambas nomenclaturas y métodos de asignación de clados y linajes, permitiendo hacer una inferencia a nivel mundial, acorde a lo establecido y aceptado para unificar resultados en el caso de este virus en particular. Esta asignación de linajes se realizó en el programa Pangolin (Pangolin Phylogenetic Assignment of Named Global Outbreak LINeages).

10. Objetivo general.

Analizar los genomas de los pacientes positivos en Guatemala para SARS-CoV-2, de marzo a noviembre de 2020, para caracterizar el origen y la dinámica de transmisión viral, como herramienta epidemiológica y de monitoreo de la evolución de la enfermedad en el país.

11. Objetivos específicos

- Secuenciar y ensamblar los genomas de SARS-CoV-2 de pacientes en Guatemala.
- Desarrollar un análisis filogenético de las variantes virales de SARS-CoV-2.
- Rastrear rutas de transmisión a partir de análisis filogenéticos virales.
- Evaluar los cambios en la diversidad genética de las variantes virales circulantes en la población.
- Identificación de linajes virales (G, GR, GH, O, L, S, V) circulantes en el país para el establecimiento de riesgo epidemiológico.
- Alimentar las bases de datos de genomas de SARS-CoV-2 para que la información generada se encuentre disponible para investigadores alrededor del mundo, que desarrollan actualmente estudios de prevención y tratamiento del COVID-19.





12. Materiales y métodos

12.1. Consideraciones éticas:

El proyecto fue sometido al Comité de Bioética en Investigación en Salud, COBIINSAUSAC de la Universidad de San Carlos de Guatemala, cumpliendo con todos los requerimientos establecidos para la obtención y uso de una fracción del excedente de las muestras de ARN viral, que fueron previamente tomadas con objetivos de diagnóstico de COVID-19, en el centro hospitalario privado Herrera Llerandi. Posterior a la revisión del comité, se desarrollaron las enmiendas requeridas por el mismo, quienes después de una segunda revisión otorgaron el aval bioético para la ejecución de la investigación (Anexo 1).

Este fue un estudio retrospectivo, debido a la necesidad de analizar el origen de las primeras cepas virales que ingresaron al país, al tratarse de una enfermedad nueva, que en tiempo récord se convirtió en una pandemia, esto representó una imposibilidad de obtener un aval bioético previo al ingreso y expansión de la enfermedad en el país, y es por ello las muestras no contaron con un consentimiento informado, por dicha razón el ARN viral fue entregado al equipo de investigación con un nuevo código, sin identificadores personales y desvinculando a la muestra de cualquier dato sensibles del paciente. Por ello, la única información recopilada y analizada fue la fecha de extracción del ARN viral, la región del país de donde proviene la muestra, la concentración viral durante del proceso de amplificación (medida como valor de carga viral -CT-). Además, tampoco se analizó ni se tuvo acceso al material biológico humano, puesto que no se accedió a la muestra primaria (hisopado nasal) sino únicamente al extracto de ARN viral.

12.2. Enfoque y tipo de investigación: Estudio cuantitativo y observacional básico

Diseño del Estudio: Estudio retrospectivo, descriptivo, analítico.

12.3. Muestreo:

Población: muestras de ARN viral de SARS-CoV-2 obtenidas de pacientes guatemaltecos que accedieron de manera directa, o cuyas muestras fueron referidas de otros centros al Laboratorio Clínico del Hospital Herrera Llerandi, para el diagnóstico de COVID-19, entre mayo y noviembre de 2020.

12.3.1. Muestra: El Laboratorio Clínico del Hospital Herrera Llerandi proporcionó 321 muestras de ARN viral, de muestras de hisopado nasal que fueron positivas para SARS-CoV-2, entre mayo y noviembre de 2020. Las muestras fueron almacenadas en ultracongeladores a -70 °C. Inicialmente se planteó una selección de muestras por conveniencia, con una distribución equitativa en la escala temporal y espacial del país, de acuerdo con el registro de casos por región. Sin embargo, solamente 90 muestras de las 321 muestras cumplieron con los criterios de inclusión para el proceso de Secuenciación Genómica.

12.3.2. Criterios de Inclusión: Se incluyeron dentro del estudio 90 muestras que presentaron CT menores a 32. Los valores de CT fueron obtenidos durante el proceso de diagnóstico por el Hospital





Herrera Llerandi, y más tarde corroborados por el equipo de Universidad de San Carlos de Guatemala investigación del I2QB3 de la Universidad Mariano Gálvez. El valor umbral de CT para la inclusión de muestras fue establecido con base en la carga viral necesaria para obtener suficientes lecturas durante el proceso de secuenciación, que garantizarán valores adecuados de profundidad y cobertura de los genomas.

12.3.3. Transporte de Muestras: Recepción, Transporte, Manipulación y Almacenaje del ARN viral: Utilizando equipo de protección personal (traje completo desechable, guantes de nitrilo, batas desechables, mascarillas, lentes de seguridad, cofias, zapatones y caretas) las investigadoras Elizabeth Solórzano y Carmen Lucía Yurrita, asistieron al Laboratorio Clínico del Hospital Herrera Llerandi para la selección, preparación y transporte de los extractos de ARN. El transporte de las muestras se realizó todos los protocolos de bioseguridad, los cuales se recopilaron en el Procedimiento Operacional Estandarizado -POE- de Bioseguridad para el transporte de muestras (Apéndice 3). Las muestras fueron decodificadas con el objetivo de desvincularlas de todos los datos sensibles y no contenían ningún identificador personal. Se generaron nuevos códigos que se registraron en la tabla contenida en el Apéndice 4, en dicho instrumento se registró la siguiente información que corresponde únicamente a datos del virus:

- Fecha de toma de muestra
- Carga viral (cantidad del virus en CT)
- Origen geográfico.

Los criotubos que contenían la muestras fueron depositados en un sistema de triple embalaje de la marca Nipro, para garantizar el aislamiento de las mismas, y la bioseguridad del transporte, además dicho sistema se colocó en una hielera con hielo seco para garantizar la cadena de frío (Apéndice 3), desde el centro de diagnóstico hasta las instalaciones del Laboratorio de Biología Molecular del Instituto de Investigaciones Químicas, Biológicas, Biomédicas y Biofísicas de la Universidad Mariano Gálvez -I2QB3 UMG-. Una vez en las instalaciones del I2QB3 UMG, las muestras fueron almacenadas en ultracongeladores a -70 °C hasta el procesamiento de estas.

El Laboratorio de I2QB3 es un laboratorio certificado de nivel de bioseguridad 3, este corresponde con un nivel de bioseguridad mayor al recomendado por CDC para el trabajo de involucra ARN genómico de longitud completa. Todos los desechos fueron descartados en bolsas de polipropileno, rojas con la impresión "Riesgo Biológico", autoclavables, las cuales fueron entregadas y tratadas de acuerdo con los protocolos de bioseguridad por la empresa Ecotermo.

12.4. Análisis de Concentración e Integridad del ARN

Para evaluar la concentración del ARN almacenado, se realizaron cuantificaciones de ARN a través de un método fluorométrico utilizando el equipo Quantus® y el kit de cuantificación Quantifluor® de Promega.





12.5. Corroboración del Diagnóstico y valores de CT

Los diferentes métodos de diagnóstico molecular a través de la Reacción en Cadena de la Polimerasa con Transcriptasa Reversa -RT-PCR-, se basan en la detección de regiones conservadas del genoma, en el caso del método de detección utilizado para el diagnóstico de COVID-19 en el Hospital Herrera Llerandi, son los genes E y Rdbp, los que son cuantificados, a través de la hibridación de cebadores específicos para dichas regiones y la visualización de la concentración viral por métodos de PCR en tiempo real.

Después de cierto tiempo de almacenaje, el material genético empieza a degradarse, aún cuando es almacenado en las condiciones ideales, esto debido a la presencia de trazas de algunas ARNasas y otras nucleasas que quedan posterior a los procesos de extracción de los ácidos nucleicos, por dicha razón es necesario corroborar la concentración del ARN de SARS- CoV-2, ya que el método de cuantificación fluorométrico no hace distinción entre el ARN viral y otros ARNs en el medio. Se corroboraron los valores de CT de 274 extractos de ARN, proporcionados por el Laboratorio Clínico del Hospital Herrera Llerandi, a través del método de detección de los genes E y Rdbp, qRT-PCR, utilizando el protocolo de manufactura del kit comercial Genefinder®.

12.6. Secuenciación Genómica

Una vez seleccionadas las 90 muestras con valores de CT iguales o menores a 32 se procedió a la preparación de las librerías.

12.6.1. Preparación de librerías para genotipificación de SARS-CoV-2

Preparación, amplificación y purificación de cDNA

Se utilizaron 10 μ L del SuperScript IV VILOTM Master Mix y 10 μ L de muestra para la síntesis de ADNc utilizando el siguiente programa de termociclador: 25°C x 10 min, 50°C x 10 min, 85°C x 5 min y espera a 4°C (Farr, et al. 2020). La amplificación de ADNc fue llevada a cabo utilizando los cebadores Artic específicos para 2019-nCoV según Primal Scheme 0,015 μ M, Buffer de reacción 5x, dNTPs 10mM y Polimerasa de ADN de alta fidelidad de arranque en caliente Q5 para un volumen final de 20 μ L (Public Health Ontario, 2020). Esto fue realizado por medio del siguiente programa: un ciclo de 50°C x 20 min, un ciclo de 95°C x 5 min, y 45 ciclos x 15 seg y 58°C por 60 seg, seguido de una espera a 4°C. Su purificación se hizo con perlas Ampure XP según el protocolo establecido por Jackson (Jackson, 2016).

Preparación de la biblioteca Tagmento Amplicón ADN

Para la tagmentación se utilizaron 11μL de transposones ligados a perlas (BLT) y 11μL de buffer de tagmentación (TB1) junto con 100–500 ng de ADN por muestra, incubando así a 55 °C x 15 min y una espera a 10°C. Se usaron 10 μL de buffer de parado de tagmentación (TSB) por muestra y se incubó la placa MIDI con las muestras a 55°C x 15 min y luego una espera a 10°C (Public Health Ontario, 2020). Posteriormente, se dejó reposar en una placa magnética por 3 minutos para después remover el sobrenadante. Los tres lavados siguientes se realizaron con 100 μL de buffer de lavado de tagmento (TWB), dejando reposar una placa magnética por 3 minutos





para después remover el sobrenadante. Inmediatamente se posicionó sobre el soporte magnético sin descartar el sobrenadante hasta su utilización (Farr, et al. 2020).

Amplificación de tagmento de ADN

Se mezclaron 22 μ L de mezcla de PCR mejorada (EPM) y 22 μ L de adaptadores de índice (placas), mezclando por vortex y realizando un spindown. De esto se tomaron 40 μ L por muestra y se añadieron 10 μ L de adaptadores de índices. Después de realizar un spindown, se incubó la placa con el siguiente programa: un ciclo de 68°C x 3 min, un ciclo de 98°C x 3 min, 8 ciclos de 98°C x 45 segundos, 62°C x 30 segundos, 68°C x 2 min, un ciclo de 68°C x 1 min y una espera a 10°C (Farr, et al. 2020).

Purificación de librerías

La placa se dejó reposar en una placa magnética por 5 min y se transfirieron $45\mu L$ de sobrenadante a un pozo en una placa MIDI nueva. Luego se agregaron $40~\mu L$ de agua libre de nucleasas, $45~\mu L$ de perlas de purificación de muestra (SPB) a cada pozo, mezclando por pipeteo e incubando a temperatura ambiente por 5 min (Farr, et al. 2020). 125 μL del sobrenadante fueron transferidos a una placa nueva con $15~\mu L$ de SPB y se dejó reposar por 5 minutos a temperatura ambiente. Se posicionó en un soporte magnético hasta que el líquido se tornó claro y se descartó el sobrenadante (Farr, et al. 2020; Public Health Ontario, 2020). Los dos lavados fueron realizados con 200 μL de etanol al 80% fresco, descartando el sobrenadante y dejando secar al aire en la placa magnética x 5 min, para después quitar de la placa magnética, agregar 32 μL de RSB a cada pozo e incubar a temperatura ambiente por dos minutos (Farr, et al. 2020).

Normalización de ADN y desnaturalización de librerías

Luego de cuantificar por el método de preferencia se normalizaron las librerías a 4 nM diluyendo con buffer RSB y se transfirieron 5 μ L a un tubo de 2 ml de biblioteca agrupada de amplicones (PAL), centrifugando brevemente. En un tubo se combinaron 5 μ L de ADN de muestra 4 nM a 5 μ L de NaOH 0.2 N, mezclando por vortex y centrifugando la solución de la muestra a 280 gravedades por 1 minuto. Luego, se incubó por 5 minutos a temperatura ambiente y se agregaron 10 μ L de muestra de ADN 4 nM a 900 μ L de HT1 (buffer de hidrolización) frío (se tendría una biblioteca desnaturalizada de 20 pM en NaOH 1 mM) (Farr, et al. 2020). Para el control de la secuenciación se utilizó el PhiX Control v3 (Ilumina, SF).

Dilución de biblioteca desnaturalizada

Utilizando 356.4 μ L de ADN 20 pM, 3.6 μ L de Phix 5% y 240 μ L de HT1 frío se obtuvo una concentración final de 12 pM, lo cual fue mezclado por inversión. Luego, con 2 μ l de biblioteca PhiX 10 nM y 3 μ l de Tris-Cl 10 mM (pH 8,5) con Tween 20 al 0,1% se diluyó el Phix a 4 nM. En un tubo se añadieron 5 μ l de Phix 4nM con 5 μ l de NaOH 0.2N, mezclando por vortex y centrifugando 280 x g por un minuto (esto se incubó a temperatura ambiente por 5 minutos).





Posteriormente, se combinaron 10 µl de PhiX y 990 µl de HT1 frío para diluir Phix a 20 pM, mezclando por inversión. Finalmente, se combinó la biblioteca con el control de Phix y se almacenó a 4°C hasta ser cargado en el cartucho de reactivo del equipo de secuenciación MiSeq (Farr, et al. 2020).

Almacenaje de Muestras

Durante el tiempo de duración del proyecto, el ARN viral permanecerá guardado en ultracongeladores a -70 C en las instalaciones del Laboratorio de Biología Molecular del -I2QB3 UMG-, con acceso restringido a las investigadoras de este proyecto, esto en caso alguno de los análisis deba ser repetido, una vez finalizado el proyecto, y publicadas las secuencias, se procederá al descarte de los extractos de ARN, de acuerdo con las normas de bioseguridad previamente descritas.

12.7. Análisis bioinformático

12.7.1. Análisis Primario de Secuencias

Antes de proceder a análisis específicos, es necesario el procesamiento inicial de las lecturas obtenidas de la secuenciación, con el fin de generar archivos en formato FASTA, los cuales funcionan como archivos de entrada de diversos métodos de análisis secundarios de las secuencias, este proceso se describe de forma gráfica en la figura X: y los pasos son descritos a detalle a continuación:

12.7.1.1. Filtrado por calidad y recorte de adaptadores

Durante el proceso de secuenciación las lecturas se unen a adaptadores e índices que permiten que la secuenciación se desarrolle, y al mismo tiempo que las lecturas correspondientes a cada muestra sean identificables, estos fragmentos nucleotídicos cortos, deben ser eliminados de las lecturas generadas, puesto que no forman parte de la secuencia de interés, en un proceso llamado recorte (triming), en esta ruta bioinformática se utilizó el software FASTQToolkit de la plataforma de Illumina Basespace, para este fin, y para el que se describe a continuación.

Otro de los primeros pasos de todo análisis de secuencias, es el filtrado de lecturas de baja calidad, medida como la probabilidad de generación de errores en la secuenciación a través del método Phred. Un valor estándar de buena calidad para la secuenciación de tipo Illumina es mayor o igual a Q30, que equivale a una probabilidad de error en la asignación del nucleótido de 1 en 1000, la herramienta FASTQToolkit fue configurada para el filtrado de las secuencias con valores Phred menores a 30.





12.7.1.2. Evaluación de calidad de lecturas

Una vez realizado el proceso de filtrado y recorte de las lecturas generadas en la secuenciación, se procedió a corroborar la calidad de las lecturas filtradas y recortadas utilizando la herramienta Fastqc, la cual genera un segundo proceso de filtrado, en caso hayan quedado lecturas de mala calidad, al mismo tiempo genera dos archivos html para las secuenciaciones de tipo paired end, que generan lecturas de la hebra templada en dirección 3´-5´y 5´-3´. Estos archivos pueden visualizarse en el sitio de Fastqc, y permiten la evaluación tanto del valor Phred como de los indicadores de calidad de las lecturas: Contenido de GC, distribución de kmers y la distribución de las lecturas en el genoma. Como resultado de este análisis se cuenta con archivos en pares de lecturas para cada una de las muestras en formato Fastaq.

12.7.1.3. Mapeo al genoma de referencia

Durante el proceso de secuenciación utilizando el método de amplicones, la secuenciación generada es mucho más específica, que las que utilizan métodos de secuenciación aleatorios, sin embargo, dentro de los archivos de lectura pueden existir lecturas no específicas, por ello es necesario realizar un alineamiento que mapea las lecturas generadas con un genoma de referencia seleccionado. En este caso se utilizó el genoma de referencia para SARS-CoV-2 de la ciudad de Wuhan, depositado en formato fasta con el identificador MN908947.3, y el cual fue descargado desde la base de datos del sitio del Centro Nacional para la Información Biotecnológica -NCBI-.

Para el mapeo con el genoma de referencia se utilizó el software Burrows-Wheeler Aligner (BWA) el cual se utiliza de forma estandarizada en rutas bioinformáticas para el mapeo de secuencias con una baja divergencia contra un genoma de referencia, como es el caso de este estudio (Li H. and Durbin, 2009). El mapeo fue configurado para el ingreso de archivos de secuenciación pareada en formato Fastaq, y el resultado generó archivos binarios de las secuencias que mapearon al genoma de referencia (archivos .bam), índices de los mismos generados con el genoma de referencia (archivos .bai), así como los archivos ordenados en formato binario (archivos .bam.sorted).

12.7.1.4. Llamado, filtrado de variantes y generación de secuencias consenso en formato Fasta

Para este momento se cuenta con archivos de tipo binario que contienen múltiples lecturas de alrededor de 150 pares de bases alineadas al genoma de referencia, con profundidades por arriba de 10X, esto significa que cada sitio nucleotídico fue re-secuenciado al menos 10 veces, el programa mplieup de la herramienta BCFtools (Li, et al., 2009), genera un llamado o búsqueda de variantes dentro de estas lecturas, para cada una de las muestras, y más tarde con el programa bcftools consensus, genera un consenso de dichas variantes para cada sitio nucleotídico de cada genoma, el resultado son los archivos consenso en formato Fasta. Este último paso fue generado en el Cluster de LIIGH/UNAM, debido a la necesidad de alta capacidad computacional para su desarrollo.



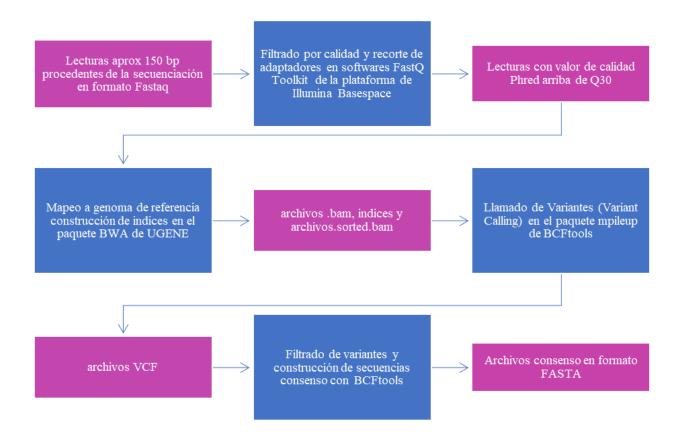


Figura 1. Ruta bioinformática de análisis Primarios de las lecturas obtenidas del proceso de secuenciación. El color fucsia corresponde a cuadros que integran tipos de archivos, y el color azul a procesos y softwares utilizados.

12.7.2. Asignación de linajes, clados y análisis de diversidad

Además de las 89 secuencias generadas en el este proyecto, se procedió a realizar la descarga de las 10 secuencias depositadas en la base de datos de GISAID, así como los metadatos disponibles, este análisis permitió ampliar el lapso de análisis de la diversidad genómica de SARS-CoV-2 en el país, así como comparar, esas primeras introducciones con las secuencias por nosotros generadas.

En total 99 secuencias guatemaltecas fueron ingresadas al sitio en línea del Centro de Vigilancia de Patógenos Genómicos, Pangolin, para proceder a la asignación de linaje a cada una de las secuencias, de acuerdo con el método de Rambaut, 2020.

De la misma forma las 99 secuencias fueron ingresadas a la aplicación CoVsurver mutation app del sitio en línea de GISAID, para la asignación de clados de acuerdo con su método de





clasificación, así como para la detección de mutaciones y otros indicadores de diversidad importantes, y que serán descritos en la sección de resultados.

12.7.3. Alineamiento Múltiple de Secuencias

De manera simultánea a la asignación de clados y linajes, las 99 secuencias fueron alineadas a través de un alineamiento global de tipo iterativo, al genoma de referencia de SARS-CoV-2 previamente descrito, utilizando la versión en línea del software MAFFT (Katoh, Rozewicki, & Yamada, 2018).

Un segundo alineamiento múltiple utilizando la misma configuración, fue desarrollado incluyendo 65 secuencias provenientes de 8 países de América del sur, como una forma de comparación de la diversidad observada en la región.

Ambos alineamientos fueron revisados manualmente y visualizados con la herramienta Jalview 2.11.1.3 (Waterhouse, et al., 2009)

12.7.4. Análisis Filogenéticos

Más tarde se desarrolló un análisis filodinámico de las 99 secuencias alineadas de SARS-CoV-2 de pacientes guatemaltecos utilizando el software IQ-TREE (Nguyen, Schmidt, von Haeseler, y Minh, 2014), el programa se configuró para la búsqueda y asignación del modelo evolutivo que mejor se ajustara a los datos. Para la generación del árbol filogenético se utilizó un método de búsqueda de árboles utilizando algoritmos de Máxima Verosimilitud, para acceder al soporte de las ramas, el proceso de búsqueda se realizó con 10,000 permutaciones de Bootstrap.

Un segundo análisis con las mismas características fue desarrollado para el alineamiento que incluyó las 65 muestras sudamericanas. Ambos árboles fueron explorados inicialmente utilizando la herramienta en línea ITOL v4 (Letunic and Bork, 2019).

12.7.5. Análisis Espaciales

Los análisis espaciales se realizaron utilizando las herramientas en línea Microreact (http://microreact.org) (Argimón et al. 2016) y TreeTom (https://camayal.info/wa/treetom/) (Maya-Lastra, 2020). Estas herramientas permiten visualizar las relaciones filogenéticas de las secuencias en un plano geográfico permitiendo explorar la dinámica espacial de los distintos linajes.

Se realizó un análisis de la distribución geográfica de las secuencias a nivel nacional, y para tener una visión más amplia del comportamiento de los datos se realizó un análisis incluyendo secuencias de libre acceso de 8 países de América del Sur (Argentina, Brasil, Colombia, Chile, Ecuador, Perú, Uruguay y Venezuela).





Las herramientas utilizadas necesitan dos tipos de archivos, uno conteniendo la información geográfica (coordenadas y el código de identificación de las secuencias) en un formato CSV separado por comas y otro conteniendo la información filogenética en un formato Newick. El código de identificación de las secuencias debe ser exactamente el mismo en los dos archivos. Para los análisis a escala nacional se asignaron coordenadas geográficas a nivel de municipio a cada una de las secuencias generadas durante el trabajo. Para los análisis a nivel regional se asignó una misma coordenada de Guatemala a todas las secuencias generadas.

Adicionalmente, se incluyó la siguiente información en el archivo CSV para tener una visualización espacial y temporal más detallada de las secuencias: información del linaje de correspondiente a la nomenclatura de Pangolin (https://cov-lineages.org/pangolin.html) de cada una de las secuencias generadas, otra con un código de color en formato Hex para cada uno de los linajes y una última con la fecha en que fue realizado el examen diagnóstico del paciente.

12.8. Depósito de secuencias genómicas de SARS-CoV-2 en bases de datos de acceso libre

Los archivos de secuencias genómica virales, al igual que todos los proyectos de secuenciación genómica de SARS-CoV-2 a nivel mundial, de forma estandarizada, serán colocados en la base de datos de la iniciativa científica global y fuente primaria de datos genómicos de influenza GISAID (gisaid.org), y que se convirtió en el repositorio mundial para datos genómicos de SARS-CoV-2, además dicha base se conectará con la base de datos Genebank del Instituto Nacional de Salud (NIH). Dichas acciones garantizarán el acceso a investigadores alrededor del mundo que desarrollan estrategias para combatir la enfermedad con base en información genómica. Para poder utilizar la información generada por los proyectos de secuenciación los usuarios deben citar al proyecto según los números de acceso otorgados al momento de subir las secuencias. Dado que, durante el proceso de generación de las secuencias, como se mencionó anteriormente, los datos de secuencias genómicas virales han sido desvinculados de información personal de los pacientes, los únicos datos que se depositaran pertenecen a información del virus. Debido a que las secuencias serán depositadas asociadas a la publicación científica en una revista arbitrada internacional, aun cuando se cuenta con todos los requerimientos necesarios para su depósito, este será concluido hasta el momento de la publicación.

13. Vinculación, difusión y divulgación

Para el desarrollo del proyecto, se realizó una vinculación con la Universidad Mariano Gálvez, a través del Laboratorio de Biología Molecular del Instituto de Investigaciones Químicas, Biológicas, Biomédicas y Biofísicas I2QB3, como resultado de dicha vinculación se integró un equipo de investigación con el grupo de investigadoras del área de genómica del I2QB3, licenciada Suzette Boburg, Licenciada Karen Cifuentes, Licenciada Milagros Gómez, MSc Rebeca Méndez y la MSc. Alejandra Sierra.





El equipo de I2QB3 dirigió la parte de laboratorio de la generación de librerías y secuenciación masiva en la plataforma Miseq de Illumina, así como la corroboración de los valores de CT y calidad del ARN de las muestras.

Debido a que dicho laboratorio es el más completo para secuenciación genómica en Guatemala y a que el nivel de experticia de sus investigadoras en el área de laboratorio de la secuenciación genómica, esta alianza es un resultado importante de este proyecto, ya que permitirá futuras colaboraciones en el área de las ómicas, para investigadores de la Universidad de San Carlos, dado que, hasta el momento la Universidad no cuenta con dicha tecnología.

Alianza con Hospital Herrera Llerandi: Tanto las autoridades del Hospital como el personal del Laboratorio clínico, en especial la licenciada Patricia Díaz encargada del diagnóstico de SARS-CoV-2, fueron unos facilitadores extraordinarios de las muestras necesarias para el desarrollo del proyecto, dicha alianza fue clave para el desarrollo del proyecto, ante la negativa a la participación del monitoreo de la diversidad genética y evolución viral, por las instituciones consultadas en el sistema nacional de salud

A través de la participación en el proyecto de la Dra. Moguel, fue posible realizar una parte de los análisis bioinformáticos en el cluster del Laboratorio Internacional de Investigación sobre el Genoma Humano -LIIGH/UNAM, para la realización de este trabajo se contó con el apoyo de Luis Aguilar, Alejandro De León, Carlos Flores y Jair García del Laboratorio Nacional de Visualización Científica Avanzada". Esta fue otra colaboración clave, debido a la ausencia de estos recursos dentro de la Universidad de San Carlos de Guatemala, y a la necesidad de alta carga computacional, así como al tiempo record en el cual los análisis de este proyecto fueron generados.

Por último, se generó una vinculación con la Universidad del Valle de Guatemala, a través de la participación y apoyo al proyecto, del estudiante de bioquímica y microbiología Alvaro Bautista, quien se encuentra desarrollando sus prácticas profesionales con el proyecto, y quien contribuyó en la elaboración de los Procedimientos Operacionales Estandarizados, revisión de literatura y generación de documentos de registro





14. Resultados

14.1. Análisis de calidad de las secuencias

De las 90 secuencias genómicas, 89 cumplieron con los estándares de calidad, con valores de Phred incluso por arriba de Q37, así como con un número elevado de lecturas por muestra, exceptuando la muestra SC86, razón por la cual, esta muestra fue descartada de los análisis posteriores.

14.2. Asignación de linajes, clados y análisis de diversidad:

14.2.1. Análisis de asignación de clados en GISAID

El análisis en la aplicación CoVsurver mutation analysis of hCoV App, de la plataforma GISAID, asignó un clado de pertenencia a cada una de las muestras guatemaltecas del virus, a partir de un análisis de genómica comparativa con los 231,171 genomas completos de SARS-COV-2 existentes en su base de datos. Además, se desarrollaron análisis de diversidad de las secuencias, los indicadores generados fueron: Número, porcentaje y listado de mutaciones previamente descritas y número, porcentaje y listado de mutaciones se encuentran listados en la Tabla 1 del Apéndice 6, y los resultados más relevantes se listan en el Cuadro 2.

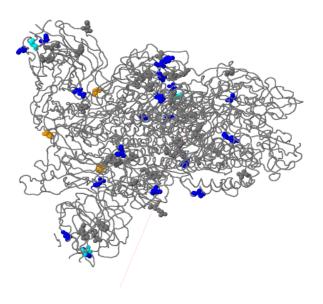
	Numero de Comoneiro	Número de Secuencias que	NO la Mutación
Clada	Numero de Secuencias	presentaron la Mutación	D614G en la
Clado	pertenecientes al clado	D614G en la proteína S	proteína S 1 (Genoma 1234,
			,
			muestreado a
G	26	98	inicios de la
		Número de Secuencias que	
		Presentaron Mutaciones	
GR	27	Unicas	
GH	41	58	
On	11	30	
S	1		
	_		
Otro	4		

Cuadro 2. Resumen de asignación de Clados en el Sistema de GISAID y mutaciones relevantes.

Por último, se generaron dos modelos giratorios en 3D de la proteína S que indican los sitios mutados en el colectivo de secuencias circulantes en el país.



a)



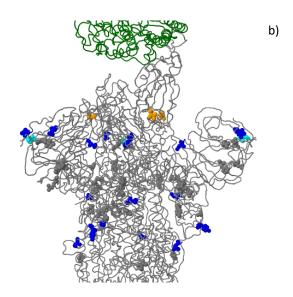


Figura 2. Modelos en 3D de la proteína de Espiga (Spike protein), desarrollada en la aplicación CoVsurver de GISAID. a) Proteína S en conformación hacia abajo PDB: 6acc, EM 3.6 Angstrom con RBD en conformación hacia abajo. b) Proteína S acoplada al receptor celular de las células humanas ACE2 (en verde). Las esferas de colores representan las mutaciones encontradas en la proteína S para las 99 secuencias de Guatemala analizadas. Mutaciones Estructurales: A27V R102T D138Y Y145del(143) M153I D198G L242I Y266F Y313H F318S R319G V367L K558R D614G A647S A653V Q675H(674) T791I S939F Mutaciones no estructurales: L5F(N-term) V6F(N-term) M1229I(C-term) V1264L(C-term).





14.2.2. Asignación de linajes en Pangolin por el método de Rambaut

Utilizando la plataforma en línea de Pangolin se generó la asignación de linaje de acuerdo con el método de Rambaut, para cada asignación se generó el porcentaje de probabilidad de asignación, los valores bajos no indican errores en las secuencias, la presencia de mutaciones únicas disminuye la similitud y homología con los linajes tipo. Además, el análisis brinda información relevante para cada linaje, como los países en los cuales se han presentado con mayor frecuencia, lapsos de reporte, así como el último día en el cual el linaje se reportó en las bases de datos mundiales (Cuadro 3).

Nombre de la secuencia	Linaje de Pangolin	Probabilidad de Asignación	Países en los que el linaje se ha presentado con mayor frecuencia	Fechas en las que los linajes han sido reportados a nivel mundial	Días desde el último reporte de muestreo del linaje
SC104	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC105	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC106	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC107	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC109	B.1.1.68	1	RUS, RU, USA	abr-03, oct-09	62
SC10	B.1	0.54	RU, USA, FRA	ene-10, oct-19	52
SC110	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC111	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC115	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC12	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC137	B.1.1.68	1	RUS, RU, USA	abr-03, oct-09	62
SC139	B.1.107	1	DIN, RU, USA	mar-16, sep-30	71
SC142	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC14	B.1.25	1	AUS, RU, USA	mar-13, agosto-30	102
SC158	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC15	B.1.1.71	1	HOL, RU, AUS	mar-06, jun-16	177
SC166	B.1.1.71	1	HOL, RU, AUS	mar-06, jun-16	177
SC16	B.1.25	1	AUS, RU, USA	mar-13, agosto-30	102
SC17	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC197	B.1.1.71	1	HOL, RU, AUS	mar-06, jun-16	177
SC198	B.1.1.77	1	USA, RU, PER	mar-18, sep-29	72
SC199	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC19	B.1.1.71	1	HOL, RU, AUS	mar-06, jun-16	177
SC201	B.1.1.68	1	RUS, RU, USA	abr-03, oct-09	62
SC202	B.1.5	0.87	RU, USA, ESP	feb-26, oct-17	54
SC204	B.1.1	0.46	RU, USA, PORT	feb-10, oct-18	53
SC205	B.1.1	0.46	RU, USA, PORT	feb-10, oct-18	53
SC207	B.1.1	0.46	RU, USA, PORT	feb-10, oct-18	53
SC209	B.1.1	0.46	RU, USA, PORT	feb-10, oct-18	53



SC213	B.1.1	0.46	RU, USA, PORT	feb-10, oct-18	53
SC218	B.1.5	1	RU, USA, ESP	feb-26, oct-17	54
SC219	B.1.1.68	1	RUS, RU, USA	abr-03, oct-09	62
SC21	B.1.1	0.46	RU, USA, PORT	feb-10, oct-18	53
SC220	B.1.1.71	1	HOL, RU, AUS	mar-06, jun-16	177
SC221	B.1.1.71	1	HOL, RU, AUS	mar-06, jun-16	177
SC248	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC251	B.1.1	0.46	RU, USA, PORT	feb-10, oct-18	53
SC252	B.1.1	0.46	RU, USA, PORT	feb-10, oct-18	53
SC256	B.1.1	0.46	RU, USA, PORT	feb-10, oct-18	53
SC259	B.1.1.71	1	HOL, RU, AUS	mar-06, jun-16	177
SC25	B.1	0.54	RU, USA, FRA	ene-10, oct-19	52
SC264	B.1.76	0.86	RU, NOR	feb-24, oct-07	64
SC267	B.1.107	1	DIN, RU, USA	mar-16, sep-30	71
SC268	B.1.1.71	1	HOL, RU, AUS	mar-06, jun-16	177
SC271	B.1.1	0.46	RU, USA, PORT	feb-10, oct-18	53
SC273	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC275	B.1.5	1	RU, USA, ESP	feb-26, oct-17	54
SC29	B.1.1.71	1	HOL, RU, AUS	mar-06, jun-16	177
SC2	B.1.138	1	AUS	mar-22, mar-26	259
SC34	B.1.2	1	USA, RU	mar-09, oct-12	59
SC36	B.1	0.36	RU, USA, FRA	ene-10, oct-19	52
SC39	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC40	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC41	B.1.1.68	1	RUS, RU, USA	abr-03, oct-09	62
SC43	B.1.1	1	RU, USA, PORT	feb-10, oct-18	53
SC45	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC46	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC47	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC49	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC4	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC52	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC54	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC55	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC56	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC57	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC58	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC59	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC60	B.1.13	1	RU, AUS, CAN	mar-09, oct-05	66
SC61	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC63	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC64	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC65	B.1	0.67	RU, USA, FRA	ene-10, oct-19	52





SC67	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC68	B.1.1	1	RU, USA, PORT	feb-10, oct-18	53
SC69	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC70	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC72	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC73	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC74	B.1.1.71	1	HOL, RU, AUS	mar-06, jun-16	177
SC75	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC76	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC79	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC7	B.1.36	0.89	RU, IND, ARBSA	feb-16, oct-15	56
SC82	B.1.130	1	RU, ARBSA, USA	mar-20, oct-17	54
SC83	B.1	0.5	RU, USA, FRA	ene-10, oct-19	52
SC86	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC87	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC89	B.1	1	RU, USA, FRA	ene-10, oct-19	52
SC95	B.1	1	RU, USA, FRA	ene-10, oct-19	52
hCoV-19/CDC- 1227	B.1	0.54	RU, USA, FRA	ene-10, oct-19	39
hCoV-19/CDC- 1228	B.1	0.54	RU, USA, FRA	ene-10, oct-19	39
hCoV-19/CDC- 1230	B.1	0.54	RU, USA, FRA	ene-10, oct-19	39
hCoV-19/CDC- 1231	B.1	0.54	RU, USA, FRA	ene-10, oct-19	39
hCoV-19/CDC- 1232	B.1	0.54	RU, USA, FRA	ene-10, oct-19	39
hCoV-19/CDC- 1233	B.1	0.54	RU, USA, FRA	ene-10, oct-19	39
hCoV-19/CDC- 1234	A.3	1	USA, RU, AUS	ene-28, jul-01	162
hCoV-19/CDC- 1235	B.1	0.54	RU, USA, FRA	ene-10, oct-19	52
hCoV-19/CDC- 1236	B.1	1	RU, USA, FRA	ene-10, oct-19	52
hCoV-19/CDC- 1229	B.1	0.91	RU, USA, FRA	ene-10, oct-19	39

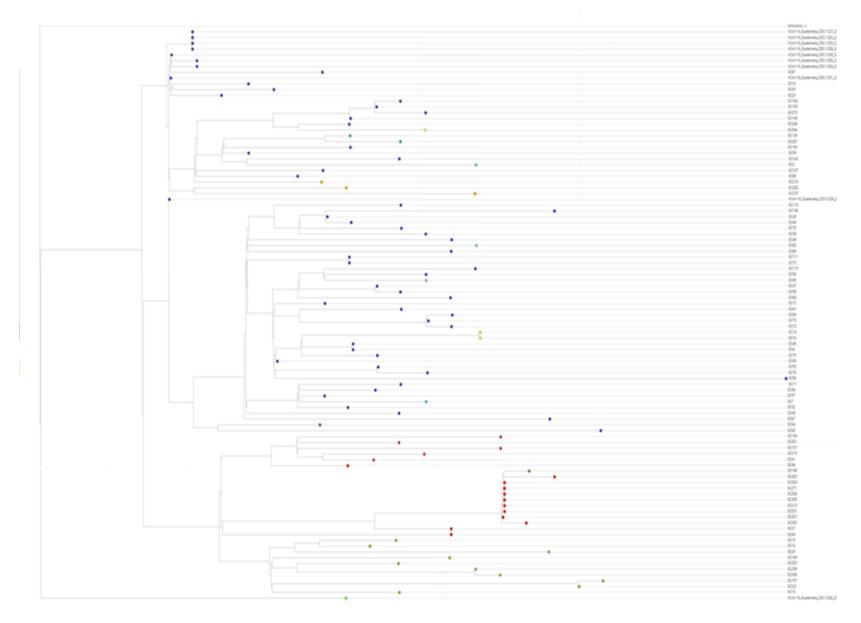
Cuadro 3: Análisis de Linajes en Pangolin 2.0 de acuerdo con la clasificación de Rambaut, A., et al., 2020. RU: Reino Unido, USA: Estados Unidos de Norte América, FRA: Francia, AUS: Australia, ESP: España, RUS: Rusia, PORT: Portugal, NOR: Noruega, DIN: Dinamarca, HOL: Holanda, PER: Perú.

14.3 Análisis Espaciales

14.3.1. Análisis Espacial y Línea de tiempo de aparecimiento de linajes de Rambaut en Guatemala:

Los siguientes gráficos presentan la filogenia de los linajes de Rambaut de las secuencias de SARS-CoV-2 circulantes en Guatemala de marzo a noviembre de 2020 (Figura 3 a). Además, se incluye una línea de tiempo que muestra los linajes introducidos al país, así como el recambio de los mismos a través del desarrollo de la pandemia en el país (Figura 3 b). Por último, la diversidad de linajes es proyectada en el mapa de Guatemala para visualizar la distribución espacial de la diversidad de las muestras evaluadas de SARS-CoV-2 en en el país (Figura 3 c).









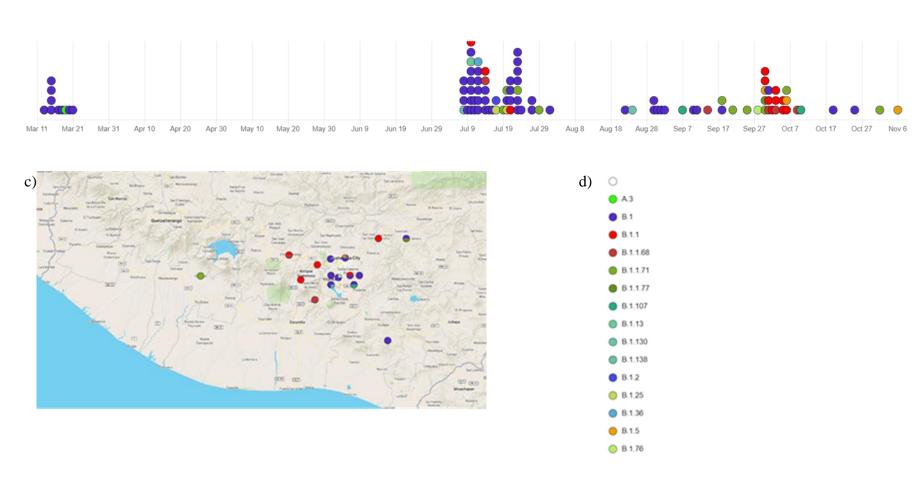


Figura 3. Filogenia y distribución espacial y temporal de los linajes según Rambaut de las muestras de SARS-CoV-2 evaluadas en Guatemala: a) árbol filogenético de Máxima Verosimilitud, construido en IQ-Tree, b) Línea de tiempo de aparecimiento de linajes, c) Mapa de distribución de linajes en Guatemala, y d) código de coloración por linaje (los colores corresponden con la asignación de linajes en Pangolin).





14.3.2. Análisis Espacial de linajes en Guatemala y Sudamérica

Como un método de comparación de la diversidad de linajes de SARS-CoV-2 en Guatemala, se realizó un análisis de la distribución de linajes de forma simultánea en Guatemala y 8 países sudamericanos (Figura 4).



Figura 4. Mapa de distribución de linajes de de SARS-CoV-2 de Pangolin, presentes en Guatemala y 8 países sudamericanos. La proyección del mapa se generó en programa microreact y los colores corresponden con la asignación de linajes en Pangolin





14.3.3. Análisis filodinámico de linajes del virus SARS-CoV-2 en Guatemala y Sudamérica

Los últimos análisis espaciales fueron desarrollados en el software TreeTom y proyectan la correspondencia no solo de los linajes con la procedencia geográfica de la muestra, sino también con los clados obtenidos del análisis filogenético con el método de máxima verosimilitud en IQ-Tree, la proyección permite la visualización no solo de la diversidad sino la homología entre los diferentes linajes de SARS CoV2.

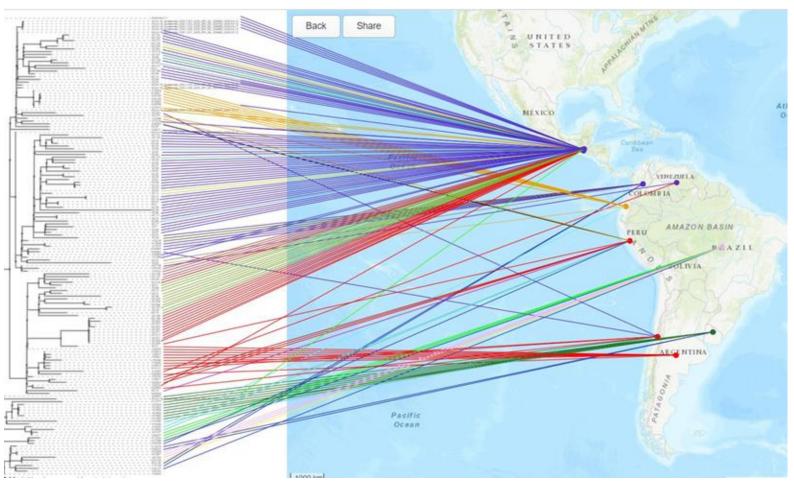


Figura 5. Mapa de proyección de linajes de SARS-CoV-2 de Guatemala y Sudamérica construido en el software TreeTom (los colores corresponden con la asignación de linajes en Pangolin).





15. Análisis y discusión de resultados

15.1. Análisis de calidad de las secuencias

Durante el desarrollo de la pandemia se generaron diversas estrategias y protocolos para la secuenciación y análisis de los genomas de SARS-CoV-2, para este estudio se utilizó el método del panel de secuenciación de ARTIC, el cual hace uso de cebadores específicos que hibridan a lo largo de todo el genoma de SARS-CoV-2, permitiendo la generación de lecturas de 150 pares de bases, que se traslapan hasta cubrir la totalidad de la secuencia del virus. Este método fue seleccionado con el fin de generar secuencias focalizadas, debido a que no se evaluaría de forma simultánea el genoma ni del huésped humano, ni de la microbiota del paciente.

La alta calidad obtenida para las lecturas provenientes de las secuencias, con valores de Phred por encima de 30 y valores de profundidad y cobertura por encima del 95% del genoma, demostró que el protocolo seleccionado fue ideal, para un enfoque de secuenciación específica del virus. Otros protocolos deben ser considerados cuando el objetivo de estudio es la interacción del genoma del virus con el genoma del paciente y su microbiota.

Solamente una de las muestras secuenciadas no presentó un alto número de lecturas, debido a que dicha muestra, la SC86, contaba con una alta carga viral, CT 18, inferimos que la ausencia de suficientes lecturas se debe a la presencia de restos de inhibidores de la PCR, provenientes de pasos previos de limpieza de las librerías de secuenciación.

15.2. Análisis de asignación de clados, linajes y diversidad

Como se mencionó anteriormente la generación de información de genomas totales y parciales, así como su depósito en bases de datos de acceso libre a nivel mundial, ha permitido que los investigadores generen información de la dinámica de evolución y transmisión de las cepas o linajes virales, en tiempo real en un evento sin precedentes. Durante el año 2020 desde el mes de enero cuando China generaba las primeras secuencias del virus, se desarrollaron diversos esfuerzos por generar sistemas de nomenclatura que permitieran además de la clasificación, la comprensión y el rastreo de rutas de transmisión a nivel local y global, dos de los métodos más utilizados y aceptados actualmente, son la clasificación de clados por GISAID y de linajes por Pangolín.

Por lo que en este estudio decidimos evaluar la distribución de linajes y clados utilizando ambos métodos, respecto al método de clasificación de GISAID pudimos observar una alta diversidad en las primeras introducciones del virus al país, registrándose en tan solo los 10 primeros días del desarrollo de la pandemia en el país, 4 de los 8 clados existentes a nivel mundial, sin embargo a lo largo del desarrollo de la enfermedad en meses posteriores, se observó un recambio en la prevalencia de ciertos clados, por otro lado el clado S introducido el 19 de marzo al país, no volvió a muestrearse nunca más, en el universo de muestras de este estudio, lo que podría reflejar un evento de éxito en el aislamiento de dicho paciente puesto que no solo no volvió a presentarse el clado nunca más en las muestras secuenciadas, sino que además el análisis filogenético de esta





misma muestra, utilizando la clasificación de Pangolin, no presenta ramificaciones posteriores (linaje A3 en el sistema de clasificación de Pangolin), lo que evidencia, la ausencia de transmisión de este paciente a otros, al menos en el universo muestreado en este estudio.

Otro hallazgo importante fue el recambio de la prevalencia de ciertos clados, durante las introducciones del mes de marzo se presentó una alta frecuencia de los clados G y O, sin embargo, actualmente los clados de mayor representación en el país son el clado GR y el clado GH, el primero no presenta representación en las secuencias introducidas durante el mes de marzo, sin embargo dado que para los meses de abril y mayo, no se cuenta con secuencias, se desconoce si la alta distribución actual de este clado (GR), corresponde a la proliferación de una introducción no documentada en este estudio, o si se debe a un proceso de evolución convergente de las cepas GR. Respecto al clado GH su alta distribución actual podría indicar el evento opuesto al evidenciado con el clado S, puesto que de una solo introducción del clado GH durante los primeros días de marzo, actualmente este, representa más del 40 por ciento de las secuencias muestreadas de marzo a noviembre del 2020, si el origen de este alto porcentaje proviene de pocas introducción o una sola, este caso sería un ejemplo icónico de malas medidas de aislamiento, que tuvieron como resultado el desarrollo de múltiples infecciones subsecuentes.

Cabe resaltar que la discusión y conclusiones obtenidas de este estudio, provienen de una pequeña fotografía local del área metropolitana de tan solo 99 secuencias, es por ello que más información es necesaria para llegar a conclusiones con un mayor soporte.

Respecto a la asignación de linajes de acuerdo con el método de Rambaut, este, también reveló una alta diversidad de los virus que circulan actualmente en el país, sin embargo, aunque diversos, muchos pertenecientes a clados monofiléticos, lo cual pudo observarse cuando se evaluó la construcción de filogenias de los linajes.

Otra información importante obtenida del análisis de mutaciones y diversidad del sitio de GISAID, es la alta frecuencia de mutaciones previamente descritas, así como de mutaciones únicas y novedosas, cada una de dichas mutaciones merece un análisis posterior profundo, que permita comprender la relevancia de cada una de las centenares de mutaciones detectadas, por otro lado, es importante resaltar las mutaciones relacionadas a la proteína de espiga S (proteína Spike), dado que esta proteína funciona como el sitio de unión de los receptores celulares ACE2 en humanos, es por ello que ha sido el objeto de múltiples estudios, puesto brinda la especificidad del huésped al coronavirus, así como el tropismo tisular. Una de las mutaciones más estudiadas corresponde a la mutación D614G que genera un cambio aminoacídico en la proteína y cuyo cambio no solo ha sido de especial atención porque ha desplazado a la variante inicial, sino porque se le han atribuido una mayor capacidad de infección, y una menor agresividad en los virus que la presentan, en este punto, dentro de las 99 secuencias analizadas solamente 1 no contó con la mutación D614G, de manera interesante pudimos observar que dicha secuencia corresponde justamente con la muestra de la cual no pudimos trazar infecciones subsecuentes (clado S en GISAID y linaje A3 en pangolín), lo cual sería coincidente con las hipótesis establecidas, de que dicha variante confiere una menor capacidad de infección al virus, sin embargo no podemos establecer si se trata de un proceso selectivo, puesto que la alta tasa de introducción de la variante que contenía la mutación D614G al país, podría también darnos fuertes indicios de que la dinámica observada coincide con un efecto de fundador.





Respecto al modelaje en tres dimensiones de la proteína de espiga, se puede observar una alta incidencia de mutaciones en la proteína de los genomas muestreados en el país, sin embargo, además de la relevancia de la mutación D614G, no se encontraron mutaciones en la región de unión de la proteína S al receptor celular ACE2 como se evidencia en la figura 2 b. Y aunque se pueden observar en las imágenes muchas mutaciones en la región cercana al dominio RBD, dominio de unión al receptor, muchas de estas mutaciones, no implican cambios en el marco de lectura, ni cambios en la codificación de aminoácidos que conforman la proteína. Por último, debido a la aparición de la nueva mutación la proteína S de genomas aislados de pacientes europeos A22V, se realizó la búsqueda de esta mutación dentro de los genomas del país, sin embargo esta no pudo ser detectada.

15.3. Análisis Espaciales

Respecto a los análisis espaciales generados en microreact y Treetom, los hallazgos más importantes son la alta diversidad de linajes distribuidos en el país, cuando estos son comparados con secuencias provenientes de América del sur, también la proyección de los clados sobre el mapa de América permite observar, que la distribución de los clados parece tener una estructuración geográfica, que corresponde a clados monofiléticos de distribución limitada, lo que hablaría de que, a pesar de la rápida dispersión de la enfermedad a nivel global, el cierre de fronteras sí parece haber tenido un efecto importante en la limitación posterior de la transmisión de la enfermedad a grandes distancias geográficas (en Guatemala y Sudamérica), así como efectos locales de filodinámica de los linajes virales.

Por otro lado el análisis de escala fina en el país observado en la Figura 3 demuestra procesos de transmisión local provocados por introducciones limitadas, que no fueron contenidas ni aisladas de manera correcta (lo cual puede observarse por pocos casos iniciales que dan origen en el árbol a múltiples infecciones), así estos, fueron de gran relevancia para la expansión de la pandemia en el país, además el recambio observado en la línea de tiempo desarrollada, nos muestra que la falla en contener de manera efectiva, las primeras introducciones (círculos azules en el mes de marzo), provocó un alza importante de casos durante los meses de junio y julio, que parece haber disminuido hasta el mes de septiembre, cuando un nuevo recambio en las distribución de linajes se presentó, y el cual coincide con la reapertura del Aeropuerto Internacional la Aurora, lo cual nos brinda alertas importantes del registro de nuevas introducciones al país, puesto que días después de la apertura del aeropuerto, se observa un claro aumento en la presencia de linajes B.1.1. Además, el análisis filogenético permitió observar también que una sola transmisión durante el mes de septiembre dio origen a 8 infecciones subsecuentes, la ausencia de información de los pacientes impide realizar un trazo epidemiológico con los métodos tradicionales, que permitiría completar el análisis de este brote de transmisión local en la región metropolitana.





16. Conclusiones

- El desarrollo de la secuenciación de 89 genomas de SARS-CoV-2 de pacientes positivos para COVID-19, de junio a noviembre del 2020, y la inclusión en los análisis de 10 genomas más de pacientes nacionales generadas por el CDC en Estados unidos, y que corresponden a las primeras introducciones del virus al país, permitió caracterizar el origen altamente diverso de genomas virales durante los primeros 10 días del inicio de la pandemia en el país, además de permitir evaluar algunas características de la dinámica de transmisión viral, como el recambio temporal de los linajes de mayor presencia en el país, la presencia de transmisión local, casos efectivos y fallidos de contención.
- Se secuenciaron, ensamblaron y mapearon 89 genomas de SARS-CoV-2 de pacientes en Guatemala, de la zona metropolitana.
- El análisis filogenético de las secuencias de SARS-CoV-2 en el país, fue complementado con 65 secuencias provenientes de 8 países sudamericanos, dicho análisis permitió observar la presencia de estructuración geográfica de la diversidad de linajes virales, que agrupaban regiones de américa en clados monofiléticos, lo que nos permite concluir que las medidas de cierre de fronteras fueron efectivas en la dispersión del virus a grandes distancias geográficas.
- A través del análisis de la filodinámica de los linajes virales encontrados en el país, pudimos observar un caso de aislamiento de pacientes, que se hipotetiza exitoso, debido a la ausencia de ramificaciones posteriores en el árbol, sin embargo, esto podría también ser causado por linajes no muestreados.
- Además, los análisis filodinámicos también permitieron la identificación de casos de transmisión local y nuevas introducciones.
- A lo largo de la evolución de la pandemia se ha podido observar un recambio importante en la distribución de los linajes de mayor representación durante el inicio, un pico de infecciones durante los meses de junio y julio, así como un último recambio en el periodo posterior a la reapertura de fronteras áreas y liberación de medias de cuarentena en el país, lo cual alerta respecto, a nuevas introducciones al país, que podrían generar nuevos focos de transmisión local, debido a la relajación en las medidas de monitoreo del sistema nacional de salud.
- De marzo a noviembre del 2020, fue posible identificar la presencia de 5 de los 8 clados existentes a nivel mundial en el sistema de clasificación de GISAID (G, GR, GH, S y O) y de 15 linajes de acuerdo con la clasificación de Rambault, lo que alerta sobre la alta diversidad del virus en el país, y que exige un monitoreo constante y amplio de la diversidad de SARS-CoV-2 durante el desarrollo de la pandemia.
- Las secuencias se encuentran listas para ser depositadas en las bases de datos de acceso libre a nivel mundial para que la información generada se encuentre disponible para investigadores alrededor del mundo, que desarrollan actualmente estudios de prevención y tratamiento de la COVID-19, estas secuencias serán depositadas simultáneamente durante el proceso de publicación del estudio en una revista indexada.





17. Impacto esperado

El desarrollo de este proyecto representa un evento histórico en los análisis de epidemiología molecular para el país, es la primera vez que se desarrolla la secuenciación y análisis de secuencias en el país, de la dinámica de transmisión en tiempo real de una enfermedad infecciosa, por investigadores nacionales, previamente estos procesos, para el estudio de enfermedades como la influenza, necesitaron que las muestras fueran referidas al CDC en Estados Unidos, este proyecto sienta un precedente para el estudio del monitoreo de la diversidad y evolución viral de SARS-CoV-2, y de cualquier otra enfermedad emergente que se presente en el futuro, lo que esperamos permita una respuesta más rápida y eficiente en el futuro, de las instituciones de salud, que les permitan tomar decisiones del manejo de la enfermedad con base en evidencia científica.

Esperamos, además, que la publicación de la información generada en este estudio inste a las autoridades del Sistema Nacional de Salud a tener una mayor apertura al trabajo colaborativo entre la academia y el sector salud.

De manera relevante otro impacto esperado, es que la información generada en este estudio al ser depositada en bases de datos de acceso libre brindara información totalmente novedosa para los diversos estudios científicos que se desarrollan actualmente del virus, con vistas a contener la pandemia que tanto ha afectado al mundo durante el año 2020.

Por último, esperamos que este trabajo colaborativo desarrollado por la Universidad de San Carlos, con el apoyo de la Universidad Mariano Gálvez y el Hospital Herrera Llerandi, continúe y más actores se sumen a esta iniciativa, para el estudio de SARS-CoV-2 y otros agentes infecciosos de importancia médica.

18. Referencias

- 2. Álvarez-Díaz, D. A., Franco-Muñoz, C., Laiton-Donato, K., Usme-Ciro, J. A., Franco-Sierra, N. D., Flórez-Sánchez, A. C., ... Mercado-Reyes, M. (2020). Molecular analysis of several in-house rRT-PCR protocols for SARS-CoV-2 detection in the context of genetic variability of the virus in Colombia. Infection, Genetics and Evolution, 84(June), 104390. https://doi.org/10.1016/j.meegid.2020.104390
- 3. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. Nature Medicine, 89(1), 44–48. https://doi.org/10.1038/s41591-020-0820-9
- 4. Argimón, S., Abudahab, K., Goater, R., Fedosejev, A., Bhai, J., Glasner, C., Feil, E., Holden, M., Yeats, C., Grundmann, H., Spratt, B. & Aanensen, D. (2016). Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microbial Genomics. 2(11): doi: 10.1099/mgen.0.000093.





- 5. Astuti, I., & Ysrafil. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. Diabetes and Metabolic Syndrome: Clinical Research and Reviews, 14(4), 407–412. https://doi.org/10.1016/j.dsx.2020.04.020
- 6. Baker, J. L., y Edlund, A. (2020). Composite Long-and Short-Read Sequencing Delivers a Complete Genome Sequence of B04Sm5, a Reutericyclin-and Mutanocyclin-Producing Strain of Streptococcus mutans. Microbiology Resource Announcements, 9(47).
- 7. BROAD Institute (sf) Genome Analysis Toolkit. Extraído de: https://gatk.broadinstitute.org/hc/en-us
- 8. Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. Nature Reviews Microbiology, 17(3), 181–192. https://doi.org/10.1038/s41579-018-0118-9
- 9. Dellicour, S., Lequime, S., Vrancken, B., Gill, M. S., Bastide, P., Gangavarapu, K., ... Lemey, P. (2019). Phylogeographic and phylodynamic approaches to epidemiological hypothesis testing. BioRxiv, 788059. https://doi.org/10.1101/788059
- 10. Gonzalez-Reiche, A. S., Hernandez, M., Sullivan, M., Ciferri, B., Alshammary, H., Obla, A., ... van Bakel, H. (2020). Introductions and early spread of SARS-CoV-2 in the New York City area Ana. MedRxiv, 1–22.
- 11. GISAID web site https://www.gisaid.org/about-us/imprint-privacy/
- 12. Gupta, A., Madhavan, M. V., Sehgal, K., Nair, N., Mahajan, S., Sehrawat, T. S., ... Landry, D. W. (2020). Extrapulmonary manifestations of COVID-19. Nature Medicine. https://doi.org/10.1038/s41591-020-0968-3
- 13. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet,
- 14. Katoh, K., Rozewicki, J., & Yamada, K. D. (2018). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics, 20(4), 1160–1166. https://doi.org/10.1093/bib/bbx108
- 15. Korber, B., Fischer, W., Gnanakaran, S. G., Yoon, H., Theiler, J., Abfalterer, W., ... Group, S. C.-19 G. (2020). Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. BioRxiv, 2020.04.29.069054. https://doi.org/10.1101/2020.04.29.069054
- 16. Lam, T. T.-Y., Shum, M. H.-H., Zhu, H.-C., Tong, Y.-G., Ni, X.-B., Liao, Y.-S., ... Guan, Y. (2020). Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. Nature, 1–6. https://doi.org/10.1038/s41586-020-2169-0
- 17. Letunic I and Bork P (2019) Nucleic Acids Res doi: 10.1093/nar/gkz239 Interactive Tree Of Life (iTOL) v4: recent updates and new developments.
- 18. Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60. [PMID: 19451168]
- 19. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., y 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England), 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
- 20. Maya-Lastra, C. A. 2020. TreeToM, Simple tool for visualizing phylogenetic trees onto a map. camayal.info/wa/treetom/





- 21. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2014). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution, 32(1), 268–274. doi:10.1093/molbev/msu300
- 22. Phan, T. (2020). Genetic diversity and evolution of SARS-CoV-2. Infection, Genetics and Evolution, 81(February), 104260. https://doi.org/10.1016/j.meegid.2020.104260
- 23. Rambaut, A., Holmes, E.C., O'Toole, Á. et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5, 1403–1407. https://doi.org/10.1038/s41564-020-0770-5
- 24. Shen, Z., Xiao, Y., Kang, L., Ma, W., Shi, L., Zhang, L., ... Li, M. (2020). Genomic diversity of SARS-CoV-2 in Coronavirus Disease 2019 patients. Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America. https://doi.org/10.1093/cid/ciaa203
- 25. Sironi, M., Hasnain, S. E., Rosenthal, B., Phan, T., Luciani, F., Shaw, M. A., Sallum, M. A., Mirhashemi, M. E., Morand, S., González-Candelas, F., & Editors of Infection, Genetics and Evolution (2020). SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases, 84, 104384. https://doi.org/10.1016/j.meegid.2020.104384
- 26. Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M., Barton, G.J (2009), "Jalview version 2: A Multiple Sequence Alignment and Analysis Workbench," Bioinformatics 25 (9) 1189-1191 doi: 10.1093/bioinformatics/btp033
- 27. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., ... Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3
- 28. Zhang, T., Wu, Q., & Zhang, Z. (2020). Probable pangolin origin of 2019-nCoV associated with outbreak of COVID-19. SSRN ELibrary, 30(7), 1–6. https://doi.org/10.1016/j.cub.2020.03.022
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., Zheng, X. S., ... Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 579(7798), 270–273. https://doi.org/10.1038/s41586-020-2012-7





19. Apéndices

Apéndice 1 Aval Bioético para el estudio, otorgado por el Comité de Bioética en Investigación en Salud de La universidad de San Carlos de Guatemala

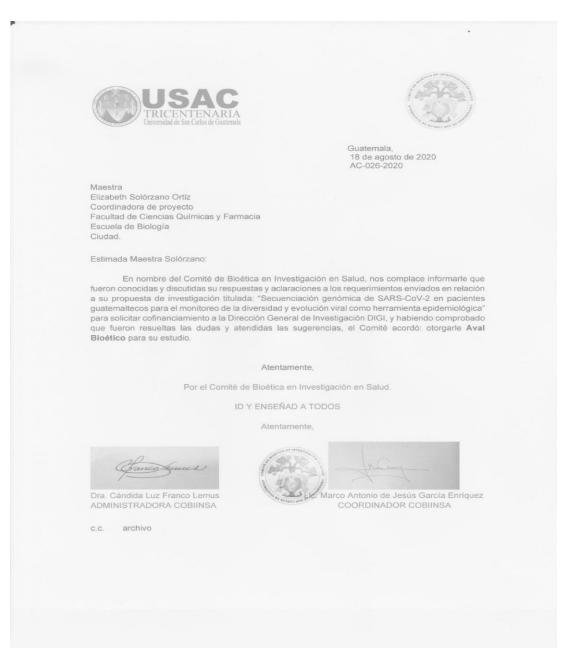






Tabla 1. Valores de CT de 90 muestras virales incluidas en la Secuenciación

Código	Valor de	Código	Valor de	Código	Valor de
SC	Ct	SC	Ct	SC	Ct
SC10	30.0943451	SC221	19.0090389	SC64	27.9459133
SC104	20.0669689	SC248	30.2955246	SC65	23.8827782
SC105	24.515255	SC25	24.3895092	SC67	28.7588177
SC106	25.9150982	SC251	17.1759586	SC68	32.1819115
SC107	23.1585808	SC252	19.6352844	SC69	30.0794411
SC109	29.3324432	SC256	17.6051598	SC7	19.6863613
SC110	20.3407173	SC264	24.0282917	SC70	22.4291267
SC111	20.7036381	SC267	14.7713375	SC72	24.485714
SC115	13.1657429	SC268	18.075531	SC73	25.4043427
SC12	24.9501743	SC271	23.4611664	SC74	26.0513706
SC137	29.9241047	SC273	21.779953	SC75	21.0013981
SC139	20.8949146	SC29	22.7094154	SC76	27.868124
SC14	23.2995358	SC34	30.2088776	SC82	29.0370426
SC142	24.5199623	SC36	31.0995274	SC84	18.2562523
SC15	28.150938	SC39	20.0799656	SC86	18.5747147
SC158	27.9235001	SC4	19.4991608	SC87	22.8969784
SC16	23.9841957	SC40	32.5268822	SC89	29.5288868
SC166	22.8418846	SC41	28.2344913	SC95	23.8900051
SC17	29.1962395	SC43	32.9450264	SC83	26.6974812
SC19	28.5064144	SC45	18.1882401	SC220	29.9439926
SC197	23.91189	SC46	19.7420883	SC259	16.1445274
SC198	18.3077469	SC47	23.3554668	SC79	22.8063145
SC2	25.4126816	SC49	28.8632469	SC275	12.69
SC201	20.5231323	SC52	24.4603786	SC199	22.1568127
SC202	18.8160362	SC54	22.4916821	SC60	25.616785
SC204	25.8874454	SC55	22.9786873	SC61	29.3572578
SC205	17.5283928	SC56	31.5752411	SC63	27.7174892
SC207	24.6536655	SC57	25.2453823	SC218	15.8180323
SC209	25.650753	SC58	24.9531078	SC219	24.750679
SC21	25.3547573	SC59	28.4117527	SC213	25.1529541





POE Bioseguridad para el transporte de muestras

UNIVERSIDAD MARIANO GÁLVEZ DE	PROCEDIMIENTOS TÉCNICOS PT-
GUATEMALA INSTITUTO DE	03/02
INVESTIGACIONES QUÍMICAS,	
BIOLÓGICAS, BIOMÉDICAS Y	
BIOFÍSICAS LABORATORIOS DE	
QUÍMICA Y BIOTECNOLOGÍA	
Procedimiento de Operación Normalizado	Título: Transporte de muestras de ARN
(PON)	de SARS-CoV2
Autor: Álvaro Bautista	Revisor:
Autor: Álvaro Bautista Fecha: 2020/Septiembre/09 Firma: No aplica.	Revisor: Fecha: Firma:
	Fecha: Firma:
Fecha: 2020/Septiembre/09 Firma: No aplica.	Fecha: Firma: Aprobación:
Fecha: 2020/Septiembre/09 Firma: No aplica.	Fecha: Firma: Aprobación: Nombre:
Fecha: 2020/Septiembre/09 Firma: No aplica.	Fecha: Firma: Aprobación: Nombre: Fecha: Firma:

1. TÍTULO

TRANSPORTE DE MUESTRAS DE ARN DE SARS-COV-2

2. INTRODUCCIÓN

El SARS-CoV-2 es un coronavirus de ARN monocatenario en sentido positivo. Su rápida expansión y alto número de casos, le ha clasificado como una pandemia desde marzo del 2020. Este virus presenta una alta tasa de infectividad y en algunos casos ha demostrado ser letal, por ello es necesario contar con altos controles de bioseguridad, durante la manipulación de muestras asociadas al virus.

3. OBJETIVO

Proveer de instrucciones y precauciones para disminuir al máximo posible el riesgo bioinfeccioso para quien transportará y manipulará las muestras genómicas de ARN de SARS-COV-2, así como garantizar la integridad de las muestras durante su transporte.

4. ALCANCE

Este documento está dirigido al transporte de muestras de ARN genómico de SARS-CoV-2. Es por ello que las normas de bioseguridad e indicaciones utilizadas no deben ser aplicadas para muestras en las cuales el virus esté activo. Este documento recoge la información necesaria para la





manipulación y transporte de muestras biológicas en la investigación humana. Si hay más de una mínima probabilidad de que la muestra contenga patógenos, debe enviarse como una sustancia infecciosa de categoría A (UN2814 o UN2900) o una sustancia infecciosa de categoría B (UN3373), según corresponda.

5. EQUIPO Y MATERIALES

- 5.1. Estuche isotérmico para muestras
- 5.2 Traje completo desechable
- 5.3 Guantes de nitrilo
- 5.4 Mascarilla
- 5.5 Lentes de seguridad
- 5.6 Caretas
- 5.7 Hielera
- 5.8 Embalaje marca Nipro
- 5.9 Etiquetas de identificación

6. REACTIVOS

6.1. Hielo seco

7. PROCEDIMIENTO PARA EL TRANSPORTE DE LAS MUESTRAS

7.1. Indicaciones generales

- 7.1.1. A pesar de que las muestras no contienen el patógeno activo, sino únicamente su material genético (no presenta un riesgo directo de transmisión), se tomarán medidas de una sustancia infecciosa categoría B
- 7.1.2 El personal encargado del transporte deberá utilizar equipo de protección personal completo (traje completo desechable, guantes de nitrilo, mascarilla, lentes de seguridad y caretas).
- 7.1.3 Para muestras que caen en categoría B de substancias infecciosas, la cantidad máxima de líquido por empaque primario es de 1 litro y el empaque exterior no puede contener más de 4 litros o 4 kg.
- 7.1.4 El hielo seco libera grandes volúmenes de dióxido de carbono gaseoso, por lo que es necesario una salida del gas para evitar una explosión. Adicionalmente, en el medio de transporte es necesario una correcta ventilación para evitar crear un ambiente deficiente de oxígeno

7.2 Procedimiento detallado

7.2.1 El empaquetamiento debe de contar con lo siguiente (ejemplo en anexo 11.1):





- 7.2.1.1 Un contenedor primario a prueba de derrames con el nombre de los contenidos. La sustancia liofilizada debe de ser transportada en ampollas de vidrio selladas a la llama o frascos de vidrio con tapón de goma con sellos de metal
- 7.2.1.2 Un empaque secundario a prueba de derrames. Se debe colocar material absorbente entre los recipientes primario y secundario para absorber todo el contenido de modo que no llegue líquido al embalaje exterior.
- 7.2.1.3 El embalaje exterior debe de ser lo suficientemente durable para su uso con al menos un lado con dimensiones de 100mm x 100 mm o mayor. Se debe incluir una lista detallada de los contenidos debe estar presente entre el segundo empaque y el embalaje exterior. En el empaque exterior debe de estar marcado para identificar que contiene materiales infecciosos de clase B (anexo 11.2), nombre de quien envía y recibe, el número UN (ver anexo 11.2) y la cantidad neta de cada sustancia.
- 7.2.2 El contenedor con hielo seco debe de poder sostener la fuerza necesaria en el proceso de transporte. Al usar una hielera es necesario que venga identificada con la etiqueta para peligros de clase 9:
- 7.2.2.1 Esta debe de contener el peso en kg del hielo seco (Anexo 11.3)
- 7.2.3 El hielo seco debe de ser posicionado fuera del empaquetamiento secundario (Anexo 11.4)
- 7.2.4 Es necesario asegurar las muestras, para que cuando se sublime el hielo seco, estas no se muevan con libertad dentro de la caja.

8. DESINFECCIÓN DE SUPERFICIES

- 9.1 Para la desinfección de superficies en donde tuvo contacto la caja, es necesario utilizar etanol al 70% o cloro al 5%
- 9.1.1 Cabe notar que al tratarse de áreas en donde el cloro correspondiente al 5% podría causar deterioro o algún daño, es preferible utilizar etanol en aerosol o en pizeta.
- 9.1.2 Si es posible utilizar ambas soluciones es imprescindible no mezclarlas. Es decir, si se usa etanol o cloro, primero limpiar la superficie de la sustancia. El alcohol al mezclarse con hipoclorito de sodio, creará cloroformo y ácido clorhídrico, así como cloroacetona o dicloroacetona. Estos compuestos pueden causar daño al sistema nervioso, pulmones, riñones, hígado, ojos y piel.

9. ALMACENAMIENTO DE LAS MUESTRAS

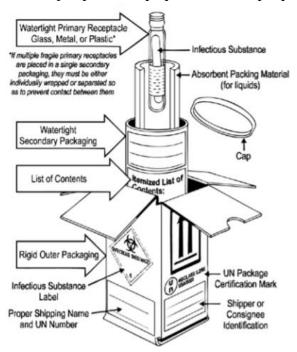
9.1 Almacenaje de las muestras

9.1.1 Las muestras deben de ser almacenadas en ultracongeladores a -70°C hasta su utilización





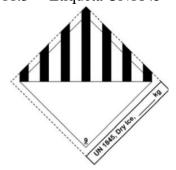
- 9.1.2 Todos los desechos deben de ser descartados en bolas de polipropileno, rojas con la impresión "Riesgo Biológico", autoclavables. Estas tienen que ser entregadas y tratadas de acuerdo con los protocolos de bioseguridad por la empresa Ecotermo.
- 11.2 Ejemplo de empaquetamiento triple para muestras infecciosas de categoría B



11.2 Etiqueta UN3373 para muestras infecciosas de categoría B



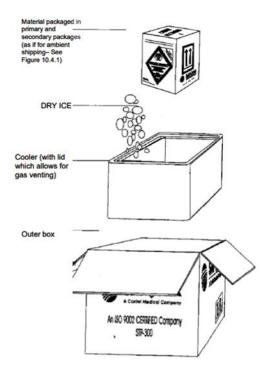
11.3 Etiqueta UN1845



11.4 Empaquetamiento para hielo seco











Instrumento para registro de datos de los productos de Extracción de ARN viral

Código ARN	Origen geográfico	Fecha de toma de Muestra	Valor de Carga Viral, obtenido de la PCR, medido como CT (cycle Treshold)

Apéndice 5

Instrumento para registro de datos de los productos de Extracción de ARN viral, durante el procesamiento y secuenciación

		proc	esamienio y secuencia	icion	
Código ARN	de		Corroboración del Valor de Carga Viral, obtenido de la PCR, medido como CT (cycle Treshold)	calidad del cADN en geles de	Muestra en





Tabla 1. Análisis de asignación de clados y análisis de diversidad en GISAID

Secuencia	Longitud	#Mut	%Mut	#Mut U	%Mut U	#Mut D	%Mut D	Mut U	Mut D	Cla
				U	U	ש	ען			do
hCoV- 19//CDC- 1227/03- 13	29882	2	0.02%	0	0.00%	2	0.02%		(NSP12_P323L,Sp ike_D614G)	Otr o
hCoV- 19//CDC- 1228/03- 15	29882	2	0.02%	0	0.00%	2	0.02%		(NSP12_P323L,Sp ike_D614G)	Otr o
hCoV- 19//CDC- 1230/03- 15	29882	2	0.02%	0	0.00%	2	0.02%		(NSP12_P323L,Sp ike_D614G)	G
hCoV- 19//CDC- 1231/03- 15	29882	2	0.02%	0	0.00%	2	0.02%		(NSP12_P323L,Sp ike_D614G)	G
hCoV- 19//CDC- 1232/03- 17	29882	2	0.02%	0	0.00%	2	0.02%		(NSP12_P323L,Sp ike_D614G)	Otr o
hCoV- 19//CDC- 1233/03- 18	29882	2	0.02%	0	0.00%	2	0.02%		(NSP12_P323L,Sp ike_D614G)	Otr o
hCoV- 19//CDC- 1234/03- 19	29882	8	0.08%	1	0.01%	7	0.07%	(NSP2_G262D)	(NSP1_D75E,NSP 3_P153L,NSP14_F 233L,NSP15_V22 L,NS7a_S81L,NS8 _V62L,NS8_L84S)	S
hCoV- 19//CDC- 1235/03- 20	29882	3	0.03%	0	0.00%	3	0.03%		(NSP6_A136V,NS P12_P323L,Spike_ D614G)	G
hCoV- 19//CDC- 1236/03- 21	29882	3	0.03%	0	0.00%	3	0.03%		(NSP12_P323L,Sp ike_D614G,NS3_ H78Y)	G
hCoV- 19//CDC- 1229/03- 15	29882	4	0.04%	1	0.01%	3	0.03%	(NSP2_N195B)	(NSP12_P323L,Sp ike_D614G,NS3_ Q57H)	GH
SC104	29906	9	0.09%	3	0.03%	6	0.06%	(NSP14_A504S, NSP16_P236L, NS3_L101P)	(NSP3_P1228S,NS P6_L37F,NSP12_P 323L,NSP12_T644 M,Spike_D614G,N _R209I)	G
SC105	29903	9	0.09%	3	0.03%	6	0.06%	(NSP14_A504S, NSP16_P236L, NS3_L101P)	(NSP3_P1228S,NS P6_L37F,NSP12_P 323L,NSP12_T644 M,Spike_D614G,N _R209I)	G





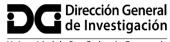
SC106	29906	8	0.08%	2	0.02%	6	0.06%	(NSP16_P236L, NS3_L101P)	(NSP3_P1228S,NS P6_L37F,NSP12_P 323L,NSP12_T644 M,Spike_D614G,N _R209I)	G
SC107	29903	7	0.07%	3	0.03%	4	0.04%	(NSP3_T626I,N SP13_E365D,N S8_G66V)	(NSP3_T1063I,NS P12_P323L,Spike_ D614G,N_R209I)	G
SC109	29903	10	0.10%	1	0.01%	9	0.09%	(N_Q244K)	(NSP4_S481L,NS P5_K90R,NSP9_T 109I,NSP12_P323 L,Spike_D614G,N S7b_A15S,N_G20 4R,N_D377Y,N_R 203K)	GR
SC10	29903	5	0.05%	3	0.03%	2	0.02%	(NSP3_G1440V ,NSP13_T199A, NS7a_R78L)	(NSP12_P323L,Sp ike_D614G)	G
SC110	29878	9	0.09%	1	0.01%	8	0.08%	(NSP16_S2N)	(NSP2_A318V,NS P2_T85I,NSP12_P 323L,NSP14_T31I ,NSP14_P140S,Spi ke_S939F,Spike_D 614G,NS3_Q57H)	GH
SC111	29903	6	0.06%	0	0.00%	6	0.06%		(NSP2_T85I,NSP3 _T1022I,NSP12_P 323L,NSP14_D32 4Y,Spike_D614G, NS3_Q57H)	GH
SC115	29903	7	0.07%	1	0.01%	6	0.06%	(N_R203S)	(NSP2_T85I,NSP1 2_P323L,NSP14_ A119V,Spike_D61 4G,NS3_Q57H,M_ L17F)	GH
SC12	29903	6	0.06%	0	0.00%	6	0.06%		(NSP2_T85I,NSP1 2_P323L,NSP14_ A119V,Spike_D61 4G,NS3_Q57H,N_ R209I)	GH
SC137	29888	17	0.18%	8	0.08%	9	0.09%	(NSP2_L270del, NSP2_N267del, NSP2_D268del, NSP2_Q427R,N SP2_L271del,N SP2_N269del,N SP5_T24I,N_V2 70L)	(NSP4_S481L,NS P12_P323L,NSP15 _M218I,Spike_D6 14G,NS3_Q213K, NS7b_A15S,N_G2 04R,N_D377Y,N_ R203K)	GR
SC139	29903	6	0.06%	2	0.02%	4	0.04%	(NSP6_M58I,N SP12_T20A)	(NSP3_V1243F,N SP12_P323L,Spike _D614G,N_R209I)	G
SC142	29903	9	0.09%	3	0.03%	6	0.06%	(Spike_Y266F,S pike_L242I,Spik e_F318S)	(NSP3_V1243F,N SP6_L37F,NSP12_ P323L,NSP16_R2 16C,Spike_D614G ,N_R209I)	G





Dirección General de Investigación

SC14	29903	9	0.09%	1	0.01%	8	0.08%	(NSP3_A886S)	(NSP2_T85I,NSP3 _T1189I,NSP3_G2 82V,NSP12_P323 L,NSP14_A119V, Spike_D614G,NS3 _Q57H,N_A119S)	GH
SC158	29903	8	0.08%	1	0.01%	7	0.07%	(NSP16_T56I)	(NSP2_T85I,NSP1 2_P323L,NSP13_ R392C,NSP14_P1 40S,Spike_D614G, NS3_Q57H,N_P13 L)	GH
SC15	29903	8	0.08%	2	0.02%	6	0.06%	(NSP3_K836R, NSP12_V330L)	(NSP2_P568L,NS P3_A1736V,NSP1 2_P323L,Spike_D 614G,N_G204R,N _R203K)	GR
SC166	29903	7	0.07%	2	0.02%	5	0.05%	(NSP3_G277E, NSP4_Q9P)	(NSP12_P323L,Sp ike_D614G,N_G20 4R,N_D377Y,N_R 203K)	GR
SC16	29903	9	0.09%	1	0.01%	8	0.08%	(NSP3_A886S)	(NSP2_T85I,NSP3 _T1189I,NSP3_G2 82V,NSP12_P323 L,NSP14_A119V, Spike_D614G,NS3 _Q57H,N_A119S)	GH
SC17	29903	6	0.06%	1	0.01%	5	0.05%	(Spike_R319G)	(NSP2_T85I,NSP3 _T1198I,NSP12_P 323L,Spike_D614 G,NS3_Q57H)	GH
SC197	29903	10	0.10%	2	0.02%	8	0.08%	(NSP14_M153I, NSP15_A217S)	(NSP2_T634I,NSP 3_P109L,NSP3_K 462N,NSP12_P323 L,NSP14_E204D,S pike_D614G,N_G2 04R,N_R203K)	GR
SC198	29903	9	0.09%	1	0.01%	8	0.08%	(NSP3_V418F)	(NSP3_N1785D,N SP12_P323L,NSP1 3_T588I,Spike_D6 14G,N_M210I,N_ G204R,N_R203K, N_A152S)	GR
SC199	29903	8	0.08%	2	0.02%	6	0.06%	(NSP2_V495A, NSP14_P46L)	(NSP3_V1243F,N SP12_P323L,NSP1 5_S287L,Spike_D 614G,NS3_Q185H ,N_R209I)	G
SC19	29903	7	0.07%	1	0.01%	6	0.06%	(NSP3_S403P)	(NSP12_P323L,NS P13_S350L,Spike_ D614G,N_A211V, N_G204R,N_R203 K)	GR
SC201	29903	7	0.07%	0	0.00%	7	0.07%		(NSP4_S481L,NS P12_P323L,Spike_ D614G,NS7b_A15 S,N_G204R,N_D3 77Y,N_R203K)	GR



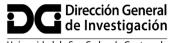


SC202	29903	5	0.05%	2	0.02%	3	0.03%	(NSP3_G171D, Spike_V6F)	(NSP12_P323L,Sp ike_D614G,N_S19 4L)	G
SC204	29903	9	0.09%	1	0.01%	8	0.08%	(NSP3_V418F)	(NSP3_N1785D,N SP12_P323L,NSP1 3_T588I,Spike_D6 14G,N_M210I,N_ G204R,N_R203K, N_A152S)	GR
SC205	29903	10	0.10%	2	0.02%	8	0.08%	(NSP2_R370C, NSP3_V418F)	(NSP3_N1785D,N SP12_P323L,NSP1 3_T588I,Spike_D6 14G,N_M210I,N_ G204R,N_R203K, N_A152S)	GR
SC207	29903	9	0.09%	1	0.01%	8	0.08%	(NSP3_V418F)	(NSP3_N1785D,N SP12_P323L,NSP1 3_T588I,Spike_D6 14G,N_M210I,N_ G204R,N_R203K, N_A152S)	GR
SC209	29903	9	0.09%	1	0.01%	8	0.08%	(NSP3_V418F)	(NSP3_N1785D,N SP12_P323L,NSP1 3_T588I,Spike_D6 14G,N_M210I,N_ G204R,N_R203K, N_A152S)	GR
SC213	29903	9	0.09%	1	0.01%	8	0.08%	(NSP3_V418F)	(NSP3_N1785D,N SP12_P323L,NSP1 3_T588I,Spike_D6 14G,N_M210I,N_ G204R,N_R203K, N_A152S)	GR
SC218	29903	5	0.05%	0	0.00%	5	0.05%		(NSP5_L50F,NSP 12_P323L,Spike_ D614G,NS3_G18 V,N_R209I)	G
SC219	29898	10	0.10%	3	0.03%	7	0.07%	(NSP3_V1583L, Spike_A27V,NS 6_V9F)	(NSP4_S481L,NS P12_P323L,Spike_ D614G,NS7b_A15 S,N_G204R,N_D3 77Y,N_R203K)	GR
SC21	29900	11	0.11%	4	0.04%	7	0.07%	(NSP3_V418F, NSP11_D3G,Sp ike_Y145del,NS 3_G224A)	(NSP12_P323L,NS P13_T588I,Spike_ D614G,N_M210I, N_G204R,N_R203 K,N_A152S)	GR
SC220	29903	7	0.07%	2	0.02%	5	0.05%	(NSP5_M276I, NSP14_M500V)	(NSP3_T1198I,NS P12_P323L,Spike_ D614G,N_G204R, N_R203K)	GR
SC221	29903	9	0.09%	2	0.02%	7	0.07%	(NSP14_M153I, NSP15_A217S)	(NSP3_P109L,NS P3_K462N,NSP12 _P323L,NSP14_E2 04D,Spike_D614G ,N_G204R,N_R20 3K)	GR
SC248	29903	6	0.06%	2	0.02%	4	0.04%	(NSP16_P236L, NS3_L101P)	(NSP6_L37F,NSP 12_P323L,Spike_ D614G,N_R209I)	G





SC251	29903	9	0.09%	1	0.01%	8	0.08%	(NSP3_V418F)	(NSP3_N1785D,N SP12_P323L,NSP1 3_T588I,Spike_D6 14G,N_M210I,N_ G204R,N_R203K, N_A152S)	GR
SC252	29903	10	0.10%	2	0.02%	8	0.08%	(NSP3_G1440V ,NSP3_V418F)	(NSP3_N1785D,N SP12_P323L,NSP1 3_T588I,Spike_D6 14G,N_M210I,N_ G204R,N_R203K, N_A152S)	GR
SC256	29903	9	0.09%	1	0.01%	8	0.08%	(NSP3_V418F)	(NSP3_N1785D,N SP12_P323L,NSP1 3_T588I,Spike_D6 14G,N_M210I,N_ G204R,N_R203K, N_A152S)	GR
SC259	29903	7	0.07%	3	0.03%	4	0.04%	(NSP3_T1056A, NSP15_Q284K, NSP16_S241F)	(NSP3_T1198I,Spi ke_D614G,N_G20 4R,N_R203K)	GR
SC25	29903	2	0.02%	0	0.00%	2	0.02%		(NSP12_P323L,Sp ike_D614G)	G
SC264	29903	8	0.08%	4	0.04%	4	0.04%	(NSP15_S161I, NSP16_P236L, NS3_L101P,NS 3_I35T)	(NSP6_L37F,NSP 12_P323L,Spike_ D614G,N_R209I)	G
SC267	29903	9	0.09%	4	0.04%	5	0.05%	(NSP6_M58I,N SP12_T225I,NS P12_T20A,NS3 _L106P)	(NSP3_V1243F,N SP12_P323L,Spike _D614G,NS3_D15 5Y,N_R209I)	G
SC268	29903	8	0.08%	4	0.04%	4	0.04%	(NSP3_T1056A, NSP9_T35I,NS P15_Q284K,NS P16_S241F)	(NSP3_T1198I,Spi ke_D614G,N_G20 4R,N_R203K)	GR
SC271	29903	9	0.09%	1	0.01%	8	0.08%	(NSP3_V418F)	(NSP3_N1785D,N SP12_P323L,NSP1 3_T588I,Spike_D6 14G,N_M210I,N_ G204R,N_R203K, N_A152S)	GR
SC273	29906	9	0.09%	3	0.03%	6	0.06%	(NSP14_A504S, NSP16_P236L, NS3_L101P)	(NSP3_P1228S,NS P6_L37F,NSP12_P 323L,NSP12_T644 M,Spike_D614G,N _R209I)	G
SC275	29903	8	0.08%	4	0.04%	4	0.04%	(NSP3_M196I, NSP10_T58I,N SP10_A32V,NS P12_L270F)	(NSP3_T428I,NSP 12_P323L,Spike_ D614G,N_S194L)	G
SC29	29903	9	0.09%	4	0.04%	5	0.05%	(NSP12_V330L, NSP12_V472L, NSP13_P238S, NS7a_V29L)	(NSP12_P323L,Sp ike_D614G,NS3_ Q57H,N_G204R,N _R203K)	G
SC2	29903	12	0.12%	6	0.06%	6	0.06%	(NSP3_A886S, NSP3_G1440V, NSP6_W140C, NSP12_A253S, NSP13_G287V, NSP15_R257L)	(NSP3_V1243F,N SP3_G282V,NSP6 _L37F,NSP12_P32 3L,Spike_D614G, N_R209I)	G





SC34	29903	9	0.09%	1	0.01%	8	0.08%	(Spike_A653V)	(NSP2_T85I,NSP1 2_P323L,NSP12_ K59N,Spike_D614 G,NS3_Q57H,NS7 a_T39I,NS8_S24L, N_R209K)	GH
SC36	29903	8	0.08%	1	0.01%	7	0.07%	(NSP12_V700I)	(NSP2_T85I,NSP1 2_P323L,NSP13_ H164Y,NSP14_A1 19V,NSP15_T325I ,Spike_D614G,NS 3_Q57H)	GH
SC39	29903	4	0.04%	0	0.00%	4	0.04%		(NSP3_V1243F,N SP12_P323L,Spike _D614G,N_R209I)	G
SC40	29903	7	0.07%	0	0.00%	7	0.07%		(NSP2_T85I,NSP1 2_P323L,NSP14_ A119V,Spike_D61 4G,NS3_Q185H,N S3_Q57H,N_A119 S)	GH
SC41	29888	9	0.09%	2	0.02%	7	0.07%	(NSP3_A41V,S pike_A27V)	(NSP4_S481L,NS P12_P323L,Spike_ D614G,NS7b_A15 S,N_G204R,N_D3 77Y,N_R203K)	GR
SC43	29898	9	0.10%	3	0.03%	6	0.07%	(NSP3_A1502T, NS3_G224V,M _E167Q)	(NSP12_P323L,Sp ike_D614G,Spike_ Q675H,N_G204R, N_H145Y,N_R203 K)	GR
SC45	29903	7	0.07%	1	0.01%	6	0.06%	(NSP16_T56I)	(NSP2_T85I,NSP1 2_P323L,NSP13_ R392C,NSP14_P1 40S,Spike_D614G, NS3_Q57H)	GH
SC46	29903	7	0.07%	1	0.01%	6	0.06%	(NSP16_T56I)	(NSP2_T85I,NSP1 2_P323L,NSP13_ R392C,NSP14_P1 40S,Spike_D614G, NS3_Q57H)	GH
SC47	29903	5	0.05%	0	0.00%	5	0.05%		(NSP2_T85I,NSP1 2_P323L,NSP14_ A119V,Spike_D61 4G,NS3_Q57H)	GH
SC49	29903	9	0.09%	2	0.02%	7	0.07%	(Spike_V1264L, NS8_S67P)	(NSP2_T85I,NSP1 2_P323L,NSP13_ R392C,NSP13_P4 7L,Spike_D614G, NS3_Q57H,N_D4 01Y)	GH
SC4	29903	7	0.07%	0	0.00%	7	0.07%		(NSP2_T85I,NSP1 2_P323L,NSP14_ A119V,Spike_D61 4G,NS3_Q185H,N S3_Q57H,N_A119 S)	GH





Lagra	20002	~	1 0.050	1 .1	0.010/	l 4	I 0.040/	AIGD4 GOOCE	Augra Togunga	
SC52	29903	5	0.05%	1	0.01%	4	0.04%	(NSP4_S386F)	(NSP2_T85I,NSP1 2_P323L,Spike_D 614G,NS3_Q57H)	GH
SC54	29903	7	0.07%	3	0.03%	4	0.04%	(NSP2_V425I,N SP3_N1178K,N SP10_T111I)	(NSP2_T85I,NSP1 2_P323L,Spike_D 614G,NS3_Q57H)	GH
SC55	29903	6	0.06%	0	0.00%	6	0.06%		(NSP2_T85I,NSP1 2_P323L,NSP14_ A119V,Spike_D61 4G,Spike_M153I, NS3_Q57H)	GH
SC56	29894	49	0.69%	43	0.60%	6	0.08%	(NSP1_K125N, NSP2_I613T,N SP2_H145Y,NS P3_W838stop,N SP3_V802F,NS P3_A831S,NSP 6_T180A,NSP1 2_K41del,NSP1 2_T20L,NSP12_L19V,NSP12_R1 10G,NSP12_R1 10G,NSP12_R1 10G,NSP12_B1 10	(NSP2_T85I,NSP1 2_P323L,NSP14_ A119V,Spike_D61 4G,NS3_Q57H,NS 7b_S31L)	GH

Mut: mutaciones, Mut U: Mutaciones únicas, %: porcentaje en el genoma, Mut D: Mutaciones descritas previamente, #: Número.





Listado de los integrantes del equipo de investigación

Contratados por contraparte y colaboradores

Nombre	Firma
MSc Elizabeth Solórzano Ortiz	Eula -
PhD Carmen Lucía Yurrita Obiols	C+4 te
PhD Barbara Beatriz Moguel Rodríguez	Belloquel
PhD Manuel Barrios Izás	Hum mun for
MSc. Carlos Alberto Montenegro	

Contratados por la Dirección General de Investigación

Nombre	Categoría	Registro de Personal	Pago SI NO		Firma
Elizabeth Solórzano Ortiz	Investigador Titular I	2006097		Х	Eula

Guatemala 14 de diciembre 2020

Elizabeth Solórzano Ortiz

Coordinadora

Proyecto de Investigación

Vo.Bo. DSc. Félix Alan Aguilar Carrera

Director General de Investigación