



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO  
INSTITUTO DE CIENCIAS DE LA SALUD  
**MAESTRÍA EN CIENCIAS BIOMÉDICAS Y DE LA SALUD**

**TESIS**

**Epidemiología genómica del SARS-CoV-2 en el  
Estado de Hidalgo, México**

Para obtener el título de

**Maestro en Ciencias Biomédicas y de la  
Salud**

**PRESENTA**

I.B Ángel Alberto Quintana Zúñiga

**Director (a)**

Dra. Jeannett Alejandra Izquierdo Vega

**Codirector (a)**

Dr. José Alberto Díaz Quiñonez

**Comité tutorial**

Dr. José Ernesto Ramírez González  
Dr. Marco Antonio Becerril Flores  
Dr. Manuel Sánchez Gutiérrez

Pachuca de Soto, Hidalgo, septiembre, 2024

**Mtra. Ojuky del Rocío Islas Maldonado**  
**Directora de Administración Escolar**  
**Presente.**

El Comité Tutorial de la **TESIS** del programa educativo de posgrado titulada "**Epidemiología genómica del SARS-CoV-2 en el Estado de Hidalgo, México**", realizado por el sustentante **Ángel Alberto Quintana Zúñiga** con número de cuenta 314719 perteneciente al programa de **Maestría en Ciencias Biomédicas y de la Salud**, una vez que ha revisado, analizado y evaluado el documento recepcional de acuerdo a lo estipulado en el Artículo 110 del Reglamento de Estudios de Posgrado, tiene a bien extender la presente:

**AUTORIZACIÓN DE IMPRESIÓN**


Por lo que la sustentante deberá cumplir los requisitos del Reglamento de Estudios de Posgrado y con lo establecido en el proceso de grado vigente.

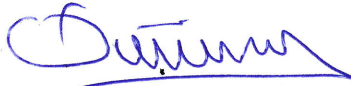
**Atentamente**

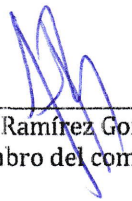
**"Amor, Orden y Progreso"**

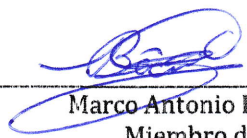
**Pachuca, Hidalgo a 06 de septiembre de 2024**

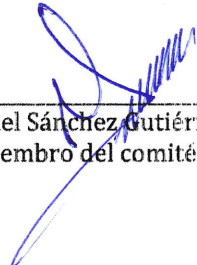
El Comité Tutorial

  
 Jeannett Alejandra Izquierdo Vega  
 Directora

  
 José Alberto Díaz Quiñones  
 Co director

  
 Ernesto Ramírez González  
 Miembro del comité

  
 Marco Antonio Becerril Flores  
 Miembro del comité

  
 Manuel Sánchez Gutiérrez  
 Miembro del comité

Circuito ex-Hacienda La Concepción s/n  
 Carretera Pachuca Actopan, San Agustín  
 Tlaxiaca, Hidalgo, México. C.P. 42160  
 Teléfono: 52 (771) 71 720 00 Ext. 41557 y  
 41556  
 medicina@uaeh.edu.mx

## **Agradecimientos institucionales**

Durante el desarrollo de este estudio, se contó con una beca de manutención otorgada por el Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT), número de beca 822647 y número de CVU 1233480. Agradezco al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT) por el apoyo brindado estos dos años a través de la beca de manutención.

Agradezco a la Unidad de Desarrollo Tecnológico e Investigación Molecular (UDTI) perteneciente al Instituto de Diagnóstico y Referencia Epidemiológicos “Dr. Manuel Martínez Báez” (InDRE), ubicado en la Ciudad de México, por su apoyo en la estancia de investigación realizada durante el periodo de septiembre a diciembre de 2023, bajo la tutela del Dr. José Ernesto Ramírez González y el Mtro. Helios Cárdenas Hernández para el análisis y presentación de la información epidemiológica y genómica concerniente al SARS-CoV-2.

Agradezco a mi Comité Tutorial, mismo que se encuentra conformado por el Dr. Marco Antonio Becerril Flores, el Dr. Manuel Sánchez Gutiérrez, el Dr. José Ernesto Ramírez González, la Dra. Jeannett Alejandra Izquierdo Vega y el Dr. José Alberto Díaz Quiñonez por el seguimiento dado al avance y redacción de esta tesis durante el periodo que duró la maestría.

Agradezco al Dr. José Alberto Díaz Quiñonez su aceptación, confianza y apoyo brindado durante este tiempo para la correcta realización del presente proyecto de investigación.

Agradezco al Dr. José Ernesto Ramírez González por la confianza y apoyo brindados para la realización y durante la estancia en el InDRE.

Este estudio fue aprobado por el Comité de Ética e Investigación del Instituto de Ciencias de la Salud de la Universidad Autónoma del Estado de Hidalgo bajo el número de folio ICSa 155/2023.

## **Agradecimientos**

Agradezco a mis padres, por siempre confiar en mí y siempre apoyarme. Los amo con todo mi ser y por siempre.

A sus parejas, no solo por apoyar a mis padres, sino también a mi hermano y a mi de una u otra forma.

A mi hermano Abraham, simplemente por dejarme crecer a su lado y que a pesar de todo sepa que lo amo.

A Rodo, Rubén, Ele, Caro, Eve, Lu, Isaac por permitirme llamarlos mis amigos y por el simple hecho de estar.

Todos ustedes son mucho más de lo que yo merezco, gracias.

A todos los profesores que se tomaron un poco de su tiempo para guiarme hasta aquí.

A mi yo del pasado por haber creído en nosotros y atreverse a llevar a cabo esta travesía.

A mi yo del futuro, para que pueda mirar hacia atrás y recuerde que siempre se es capaz de llegar a donde queremos.

A la vida y a Dios por permitirme llegar hasta este momento y por todos aquellos que aún nos falten...

## Índice general

Índice general .....	3
Índice de tablas .....	6
Abstract .....	11
I. Introducción.....	12
II. Antecedentes.....	14
2.1 Generalidades .....	14
2.1.1 Virus y su clasificación.....	14
2.1.2 Coronavirus del síndrome respiratorio agudo grave (SARS-CoV) .....	15
2.1.3 Coronavirus del síndrome respiratorio de Oriente Medio (MERS-CoV) .....	15
2.1.4 Coronavirus de tipo 2 del síndrome respiratorio agudo grave (SARS-CoV-2) .....	15
2.1.5 Enfermedad causada por coronavirus 2019 (COVID-19) .....	15
2.1.6 Diagnóstico para COVID-19 .....	16
2.1.7 Epidemia .....	17
2.1.8 Emergencia de Salud Pública de importancia internacional.....	17
2.1.9 Pandemia .....	17
2.1.10 Epidemiología molecular .....	17
2.1.11 Genómica del virus.....	18
2.1.12 Mutaciones del virus .....	18
2.1.13 Variantes en Salud Pública.....	19
2.1.14 Árboles filogenómicos.....	21
2.1.15 Vectores de infección / transmisión .....	21
2.1.16 Poblaciones inmunológicamente susceptibles .....	22
2.2 Antecedentes del problema.....	22
2.2.1 Descubrimiento de los HCoV .....	22
2.2.2 Brote del SARS-CoV-2 .....	23
2.2.3 Pandemia por COVID-19 (11/marzo/2020-05/mayo/2023).....	23
2.2.4 Postpandemia por COVID-19.....	24
2.3 Planteamiento del problema .....	25
III. Justificación.....	26

<b>IV.</b>	<b>Hipótesis</b> .....	27
<b>V.</b>	<b>Objetivos</b> .....	27
5.1	<b>Objetivo general</b> .....	27
5.2	<b>Objetivos específicos</b> .....	27
<b>VI.</b>	<b>Materiales y métodos</b> .....	27
6.1	<b>Contexto de la investigación</b> .....	27
6.2	<b>Tipo y diseño de estudio</b> .....	28
6.3	<b>Selección de la población, criterios de inclusión y exclusión</b> .....	29
6.4	<b>Tamaño de muestra y muestreo</b> .....	30
6.5	<b>Diagrama de diseño experimental</b> .....	30
6.6	<b>Definición de variables</b> .....	30
6.7	<b>Análisis estadístico</b> .....	30
<b>VII.</b>	<b>Resultados</b> .....	30
<b>VIII.</b>	<b>Discusión</b> .....	54
<b>IX.</b>	<b>Conclusiones</b> .....	58
<b>X.</b>	<b>Recomendaciones</b> .....	59
<b>XI.</b>	<b>Referencias</b> .....	60
<b>XII.</b>	<b>Anexos</b> .....	65
	Anexo 1. Operacionalización de variables .....	65
	Anexo 2. Uso de Nextstrain .....	71

## Índice de figuras

Figura 1. Diagrama de diseño experimental utilizado en este proyecto.....	30
Figura 2. Clasificaciones de tipos de casos derivada de la base de datos, se observa que casi el 50% corresponde a casos confirmados por laboratorio.....	32
Figura 3. Estadísticos para la variable Edad. Se observan diversos valores atípicos en la edad de ambos sexos.....	34
Figura 4. Distribución de los casos positivos por grupo etario. De mayor a menor: 30-39, 18-29, 40-49, 60 y más, 50-59 y 0-17 años. ....	35
Figura 5. Grupos etarios separados por sexo del caso.....	36
Figura 6. Comorbilidades reportadas por los casos confirmados de COVID-19. ....	37
Figura 7. Síntomas reportados por los casos confirmados. Síntomas populares como disgeusia y anosmia se presentaron en un bajo porcentaje de la población.....	38
Figura 8. Acumulación de casos positivos y negativos en el estado de Hidalgo. Se observan 6 distintos picos, correspondientes a cada una de las olas reportadas hasta finales de 2023.....	39
Figura 9. Casos confirmados en el estado de Hidalgo. Se observan 6 distintos picos, correspondientes a cada una de las olas reportadas hasta finales de 2023. ....	39
Figura 10. Defunciones confirmadas por COVID-19 en el estado de Hidalgo hasta finales de 2023. ....	40
Figura 11. Visualización parcial en Clustal del alineamiento múltiple de las 674 secuencias de SARS-CoV-2 del estado de Hidalgo. ....	41
Figura 12. Metadatos de una secuencia del estado de Hidalgo, fuente GISAID.....	42
Figura 13. A) Distribución de las distintas variantes del SARS-CoV-2 en el estado de Hidalgo a través del tiempo. B) Número de genomas reportados por el estado de Hidalgo por semana epidemiológica (GISAID).....	43
Figura 14. Árbol filogenómico de las secuencias de SARS-CoV-2 en el estado de Hidalgo, México. Construido con 662 genomas recolectados entre dic 2019-feb 2024, de estas, 634 pertenecen al estado de Hidalgo (amarillo) y 28 como referencia de todo el mundo (azul).....	44
Figura 15. Vista de las 43 secuencias correspondientes a la variante B1.1.519 (visibles a través del clado 20B y la mutación S:478K) y de las posibles fechas en las que comenzó su circulación en el estado de Hidalgo. ....	45

Figura 16. Vista de las 43 secuencias correspondientes a la variante alfa y de las posibles fechas en las que comenzó su circulación en el estado de Hidalgo. ....	46
Figura 17. Vista de las 183 secuencias correspondientes a la variante delta y de las posibles fechas en las que comenzó su circulación en el estado. ....	47
Figura 18. Vista de las 51 secuencias correspondientes a la variable gamma y de las posibles fechas en las que comenzó su circulación en el estado.....	48
Figura 19. Vista de las 296 secuencias correspondientes a la variante ómicron y de las posibles fechas en las que comenzó su circulación en el estado.....	49
Figura 20. Muestra de la base de metadatos utilizada para la construcción del árbol filogenómico, formato tsv.....	50
Figura 21. Metadatos observados en el árbol filogenómico construido.....	50
Figura 22. Proporción de los sexos para los casos confirmados correspondientes a las variantes Delta y Ómicron .....	51
Figura 23. Grupos etarios para la variante delta. ....	51
Figura 24. Grupos etarios para la variante ómicron.....	52

## Índice de tablas

Tabla 1. Tipos de variantes del SARS-CoV-2 .....	20
Tabla 2. Observaciones de la base de datos de COVID-19 (48).....	31
Tabla 3. Casos y defunciones por región operativa del Estado.....	33
Tabla 4. Sexo contra clasificación de los casos. ....	33
Tabla 5. Casos y defunciones por región operativa correspondientes a la variante delta. ....	52
Tabla 6. Casos y defunciones por región operativa correspondientes a la variante ómicron. ....	53
Tabla 7. Operacionalización de variables.....	65



## Abreviaturas

+ssRNA: Virus de ARN monocatenario positivo (del inglés *positive sense single-stranded RNA viruses*).

-ssRNA: Virus de ARN monocatenario negativo (del inglés *negative sense single-stranded RNA viruses*).

CLIA: Inmunoensayo de quimioluminiscencia (del inglés *Chemiluminescent Immunoassays*).

COVID-19: Enfermedad por coronavirus 2019 (del inglés *Coronavirus disease 2019*).

DGE: Dirección General de Epidemiología.

DNA: Ácido desoxirribonucleico (del inglés *deoxyribonucleic Acid*).

DNA-RT: Virus de ADN de transcripción inversa (del inglés *reverse transcribed DNA virus*).

dsDNA: Virus de ADN bicatenario (del inglés *double-stranded DNA viruses*).

dsRNA: Virus de ARN bicatenario (del inglés *double-stranded RNA viruses*).

ELISA: Ensayo por inmunoabsorción ligado a enzimas (del inglés *Enzyme-Linked immunosorbent Assay*).

GISAID: Iniciativa global para compartir los datos de los virus gripales (del inglés *Global Initiative on Sharing All Influenza Data*).

HCoV: Coronavirus humanos (del inglés *Human Coronaviruses*).

ICSa: Instituto de Ciencias de la Salud.

IHR: Reglamento sanitario internacional (del inglés *International Health Regulations*).

InDRE: Instituto de Diagnóstico y Referencia Epidemiológicos "Dr. Manuel Martínez Báez".

LFIA: Inmunoensayo de flujo lateral (del inglés *Lateral flow immunoassay*).

MAFFT: Alineamiento múltiple con uso de la transformada rápida de Fourier (del inglés *Multiple Alignment using Fast Fourier Transform*).

MERS-CoV: Coronavirus del síndrome respiratorio de oriente medio (del inglés *Middle East respiratory syndrome coronavirus*).

NSPS: Proteínas no estructurales (del inglés *Non-Structural Proteins*).

OMS: Organización Mundial de la Salud.

ORF: Marco abierto de lectura (del inglés *Open Reading Frame*).

PHEIC: Emergencia de salud pública de importancia internacional (del inglés *Public Health Emergency of International Concern*).

POC: Ensayo en punto de atención (del inglés *Point of Care*).

RNA: Ácido ribonucleico (del inglés *ribonucleic Acid*).

RNA-RT: Virus de ARN de transcripción inversa (del inglés *reverse transcribed RNA virus*).

RNLSP: Red Nacional de Laboratorios de Salud Pública.

RT-PCR: Reacción en cadena de la polimerasa de transcripción inversa (del inglés *Reverse transcription polymerase chain reaction*).

SALUD: Secretaría de Salud.

SARS-CoV: Coronavirus del síndrome respiratorio agudo grave (del inglés *Severe acute respiratory syndrome coronavirus*).

SARS-CoV-2: Coronavirus de tipo 2 del síndrome respiratorio agudo grave (del inglés *Severe acute respiratory syndrome coronavirus 2*).

SINAVE: Sistema Nacional de Vigilancia Epidémica.

ssDNA: Virus de ADN monocatenario (del inglés *single-stranded DNA viruses*).

UAEH: Universidad Autónoma del Estado de Hidalgo.

VOC: Variantes de preocupación (del inglés *Variants of Concern*).

VOI: Variantes de interés (del inglés *Variants of Interest*).

VUM: Variante bajo monitoreo (del inglés *Variant Under Monitoring*).

## Resumen

Desde su aparición a finales de 2019, el SARS-CoV-2 se ha esparcido a todo el mundo. La Organización Mundial de la Salud (OMS) declaró una pandemia por el virus el 11 de marzo del 2020 que concluyó el 5 de mayo de 2023. El virus ha tenido una infinidad de mutaciones a través del tiempo, modificando su comportamiento dentro de distintas poblaciones y dificultando el desarrollo de sistemas de diagnóstico, vacunas y tratamientos. Dichas mutaciones generan variantes del SARS-CoV-2 que se clasifican como variantes de interés y de preocupación para la salud pública.

Este trabajo es un estudio observacional del tipo transversal, que busca describir la epidemiología y la genómica del SARS-CoV-2 en el estado de Hidalgo, México.

Se describió la pandemia de COVID-19 en el estado de Hidalgo en base a la información publicada por la Dirección General de Epidemiología (DGE) desde 2020 hasta 2023, considerando las variables de sexo, edad, municipio (agrupados en 12 regiones operativas), sintomatologías y comorbilidades, realizando pruebas de asociación entre éstas y la clasificación de los casos (positivo y negativo), mismos que se representan en una curva epidémica. Se construyeron árboles filogenómicos con las secuencias reportadas en el estado, publicadas en GISAID (del inglés *Global Initiative on Sharing All Influenza Data*) por el Instituto de Diagnóstico y Referencia Epidemiológicos “Dr. Manuel Martínez Báez” (InDRE) y se estudió su distribución a través del tiempo.

El comportamiento epidemiológico de la pandemia en el estado fue muy similar a la media nacional; más del 50% de los casos confirmados fueron mujeres, la mediana de edad fue de 39 años, la región de Pachuca de Soto fue la que reporta el mayor porcentaje de casos positivos, y el porcentaje de mortalidad en el estado fue de 6.5%, siendo este mayor al nacional. La distribución temporal de las variantes del virus SARS-CoV-2 demuestra su presencia en momentos tardíos o tempranos a su identificación inicial en el estado de Hidalgo.

La combinación de ambos tipos de análisis permitió identificar el comportamiento del virus y sus distintas variantes en una población específica a través del tiempo. Este trabajo evidencia la necesidad de establecer una vigilancia permanente, mejorando los protocolos para la obtención de datos genómicos y captura de metadatos, con la finalidad de facilitar estudios posteriores y establecer medidas de prevención y control adecuadas.

Palabras clave: SARS-CoV-2, COVID-19, Hidalgo, México, variantes, pandemia.

## **Abstract**

Since its emergence in late 2019, SARS-CoV-2 has spread worldwide. The World Health Organization (WHO) declared a pandemic for the virus on March 11, 2020, ending on May 5, 2023. The virus has undergone a myriad of mutations over time, modifying its behavior within different populations and hindering the development of diagnostic systems, vaccines, and treatments. These mutations are classified as variants of public health concern. This work is a cross-sectional observational study that seeks to describe the epidemiology and genomics of SARS-CoV-2 in the State of Hidalgo, Mexico.

The COVID-19 pandemic in the State of Hidalgo was described based on information published by the General Directorate of Epidemiology (DGE) from 2020 to 2023, considering the variables of sex, age, municipality (grouped into 12 operational regions), symptomatology, and comorbidities, performing association tests between these, and the classification of cases (positive and negative), which are represented in an epidemic curve. Phylogenomic trees were constructed with the sequences reported in the state, published in GISAID (Global Initiative on Sharing All Influenza Data) by the Instituto de Diagnóstico y Referencia Epidemiológicos "Dr. Manuel Martínez Báez" (InDRE), and their distribution over time was studied.

The epidemiological behavior of the pandemic in the state was very similar to the national average; more than 50% of the confirmed cases were women, the median age was 39 years, the region of Pachuca de Soto reported the highest percentage of positive cases, and the mortality rate in the state was 6.5%, higher than the national rate. The temporal distribution of the SARS-CoV-2 virus variants demonstrates its presence in the late or early stages of its initial identification.

The combination of both types of analysis made it possible to identify the virus's behavior and its different variants in a specific population over time. This work highlights the need to establish permanent surveillance and improve protocols for obtaining genomic data and capturing metadata to facilitate subsequent studies and establish appropriate prevention and control measures.

Key words: SARS-CoV-2, COVID-19, Hidalgo, México, variants, pandemic.

## I. Introducción

El coronavirus del Síndrome Respiratorio Agudo Grave 2 (SARS-CoV-2 por sus siglas en Inglés) es un virus miembro de la familia Coronaviridae junto al SARS-CoV y el MERS-CoV (responsables de las epidemias por SARS en 2003 y MERS en 2012, respectivamente)(1–3). Es el encargado de causar la enfermedad conocida como COVID-19 (del inglés *Coronavirus disease 2019*) y la posterior pandemia por COVID-19 declarada desde el 11 de marzo de 2020 y hasta el 5 de mayo de 2023 (4,5). Al 9 de agosto de 2023, la Organización Mundial de la Salud (OMS) ha reportado más de 760 millones de casos y 6.9 millones de defunciones en todo el mundo a causa de esta enfermedad, aunque se cree que estos números pueden ser mucho mayores (6).

El SARS-CoV-2 es un virus de RNA monocatenario de sentido positivo, cuyos viriones miden entre 50-200 nm. Su genoma se encuentra compuesto, en promedio, por 29 kilobases (kb), mismo que codifica para 16 proteínas no estructurales, 9 proteínas accesorias y 4 proteínas estructurales. Cada una de estas proteínas cumple con un rol importante en la infección de células y la replicación del virus, sobre todo la proteína S (7,8) . Como la mayoría de los virus de RNA, el SARS-CoV-2 evoluciona de forma acelerada debido a la aparición de mutaciones en su genoma viral (9). Tales mutaciones se atribuyen a la velocidad de propagación del virus en grandes poblaciones. Debido a esto, han surgido un gran número de variantes al virus original (conocidas como variantes de interés, de preocupación y bajo monitoreo) en todo el mundo, que han dificultado el desarrollo de vacunas y tratamientos efectivos para la enfermedad.(10)

En México se reportan más de 7 millones de casos y 300 mil muertes por COVID-19, de acuerdo al último reporte técnico oficial, publicado el 2 de agosto de 2023.

El estudio se realizó a través del análisis de bases de datos de la DGE correspondientes a COVID-19 desde 2020 hasta 2023, mediante el uso de herramientas como R studio, tomando las variables de sexo, edad, municipio, sintomatologías y comorbilidades de los casos reportados y realizando pruebas de asociación entre éstas y la clasificación del caso (positivo y negativo). Se

construyeron árboles filogenómicos, mediante el uso del software de libre acceso Nextstrain, de más de 600 secuencias (correspondientes al estado de Hidalgo, México) disponibles en la plataforma GISAID, mismos que fueron secuenciados y publicados por el InDRE. Posteriormente, se analizó su distribución a través del tiempo.

## **II. Antecedentes**

### **2.1 Generalidades**

#### **2.1.1 Virus y su clasificación**

Los virus son organismos acelulares, cuyo genoma está constituido de DNA o RNA, los cuales están obligados a replicarse dentro de una célula hospedera; con un diámetro de entre 20 y 300 nanómetros (nm), únicamente son observables a través del uso de microscopios electrónicos; careciendo de actividades metabólicas propias, utilizan los mecanismos celulares del huésped para la síntesis de sus componentes virales y posterior ensamble en partículas conocidas como viriones, los cuales permiten la transferencia del material a otras células para la diseminación del virus. (11)

Para su clasificación, David Baltimore diseñó un sistema basándose en las rutas de expresión de los virus, siendo a la fecha aceptadas 7 distintas clasificaciones (12), las cuales se muestran a continuación:

- Virus de DNA bicatenario (dsDNA): Encapsulan un dsDNA y, al igual que las células, utilizan la ruta clásica de transcripción de la información.
- Virus de DNA monocatenario (ssDNA): Encapsulan un ssDNA, el cual es replicado y expresado a través de un dsDNA intermedio.
- Virus de DNA bicatenario (dsRNA): Encapsulan un dsRNA que debe ser transcrito para la producción de un mRNA (mensajero).
- Virus de RNA monocatenario positivo (+ssRNA): Empaquetan viriones de ssRNA de la misma polaridad que el mRNA, sintetizando las proteínas virales a manera de que el genoma pueda ser traducido directamente.
- Virus de RNA monocatenario negativo (-ssRNA): Encapsula un ssRNA complementario del mRNA, el cual es transcrito para la producción del último.
- Virus de RNA de transcripción inversa (RNA): Virus que empaquetan un RNA de carga positiva, replicado a través de un DNA intermedio.
- Virus de DNA de transcripción inversa (DNA-RT): Encapsulan un dsDNA y se replican a través de un ARN intermedio.



### **2.1.2 Coronavirus del síndrome respiratorio agudo grave (SARS-CoV)**

Virus (+ssRNA) identificado a principios del nuevo milenio (2002), responsable de causar la enfermedad conocida como síndrome respiratorio agudo grave (SARS) con una tasa de letalidad del 10.8%, y la subsecuente epidemia por SARS que duró desde finales de 2002 a mediados de 2003; su identificación se dio a partir del estudio de casos de neumonías inusuales en la provincia china de Guangdong (Cantón). Se ha identificado que las civetas vendidas en el mismo fueron los principales reservorios del virus y que su consumo como carne exótica fue la causa principal de la transmisión hacia los humanos. (2)

### **2.1.3 Coronavirus del síndrome respiratorio de Oriente Medio (MERS-CoV)**

Virus (+ssRNA) identificado a mediados de 2012 en Arabia Saudita después de estudiar el caso de un hombre de 60 años de edad que presentaba cuadros de neumonía aguda y falla renal. Debido al lugar de identificación, la enfermedad causada por el virus recibió el nombre de síndrome respiratorio agudo de Oriente Medio (MERS), siendo identificado al camello dromedario como el principal reservorio del virus, y el contacto cercano con ellos como la fuente de transmisión a los humanos. Hasta agosto de 2023 se han registrado 937 muertes de los 2605 casos confirmados en todo el mundo, teniendo una tasa de letalidad del 36%, convirtiéndose en uno de los virus más devastadores conocidos a la fecha.(3)

### **2.1.4 Coronavirus de tipo 2 del síndrome respiratorio agudo grave (SARS-CoV-2)**

Virus (+ssRNA) originario de China, identificado en 2019 después de analizar casos de pacientes con neumonías atípicas a finales del mismo año en la ciudad de Wuhan. El mercado marino de Huanan ha sido reconocido como epicentro del brote. (13) Mamíferos como los pangolines y los perros mapache han sido identificados como posibles reservorios del virus; la venta, consumo y contacto cercano de éstos ha sido la causa de la transmisión a los humanos. (1)

### **2.1.5 Enfermedad causada por coronavirus 2019 (COVID-19)**

Enfermedad causada por la infección del SARS-CoV-2, cuyo espectro clínico puede ir desde una infección asintomática hasta complicaciones que pueden poner en riesgo la vida del afectado. La definición operacional de caso sospechoso para

COVID-19, dada por la OMS, establece los siguientes criterios: Cualquier persona que sufra de la aparición repentina de fiebre y tos, junto a la aparición de 3 o más síntomas que se presentan a continuación: debilidad general/fatiga, cefalea, mialgia, dolor de garganta, resfriado nasal, disnea, anorexia/náuseas/vómitos, diarrea, estado mental alterado. (14)

En algunos casos, la enfermedad puede llegar a afectar las vías respiratorias inferiores, trayendo consigo complicaciones como dificultad respiratoria aguda, falla multiorgánica y en el peor de los casos la muerte, lo cual hace necesaria la atención hospitalaria, en especial de cuidados intensivos y ventilación mecánica. (15)

La enfermedad puede agravarse gracias a la influencia de comorbilidades del paciente. Éstas son enfermedades que afectan a un individuo al mismo tiempo que otra enfermedad aparece en él, como enfermedades cardiovasculares, diabetes, enfermedades respiratorias crónicas, hipertensión, cáncer y obesidad. (16)

#### **2.1.6 Diagnóstico para COVID-19**

Debido a que la enfermedad presenta sintomatologías extremadamente similares con otras enfermedades, tales como un resfriado común o influenza, la confirmación se tiene que dar a través del uso de métodos de laboratorio (especialmente para la detección de casos asintomáticos) especializados en la detección del virus o alguno de sus componentes (proteínas) (17,18). Entre estas técnicas tenemos:

- Reacción en cadena de la polimerasa de transcripción inversa (RT-PCR, por sus siglas en inglés): Prueba estándar de oro y por ende la más utilizada actualmente; apoyándose principalmente de muestras tomadas con un hisopo de la mucosa nasal y secreciones oculares. A pesar de ser la técnica más utilizada, tiene la desventaja de necesitar de infraestructura especializada y varios días para la obtención de resultados.
- Pruebas serológicas: Involucra el análisis de muestras de suero sanguíneo o cualquier otro fluido biológico, que permita identificar la presencia de biomarcadores tales como antígenos y anticuerpos; además de su uso de diagnóstico, generalmente se utilizan para monitorear la evolución de la enfermedad. La principal prueba utilizada es el enzimoimmunoanálisis de adsorción (ELISA, por sus siglas en inglés). Algunas otras pruebas serológicas

utilizadas son: Inmunoensayo de quimioluminiscencia (CLIA, por sus siglas en inglés) e Inmunoensayo de flujo lateral (LFIA, por sus siglas en inglés).

- Pruebas serológicas a base de nanopartículas: Desarrollándose rápidamente en años recientes debido, principalmente, al uso de instrumentación sencilla; las nanopartículas de oro son las utilizadas en este tipo de ensayos, analizando una pequeña muestra de suero biológico en busca de anticuerpos, dando resultados en tan solo 15 minutos.

### **2.1.7 Epidemia**

Una epidemia se define como la propagación rápida y anormal de una enfermedad infecciosa en números superiores a los esperados con anterioridad. (19)

### **2.1.8 Emergencia de Salud Pública de importancia internacional**

Una emergencia de Salud Pública de importancia internacional (PHEIC, por sus siglas en inglés) es una declaración oficial, emitida por la OMS, que define a un evento extraordinario cuya magnitud pone en riesgo la Salud Pública de varios estados a través de la propagación de enfermedades y cuyo manejo requiera de un apoyo coordinado internacional, por lo cual los estados tienen deberes legales para responder de manera adecuada a una PHEIC. Desde la implementación del reglamento sanitario internacional (IHR, por sus siglas en inglés) en 2005, las PHEIC se han declarado de manera oficial en 7 ocasiones. (20,21)

### **2.1.9 Pandemia**

La OMS define a las pandemias como “epidemias que afectan a una gran proporción del mundo a causa de un virus y/o bacterias, que tiene la capacidad de mantener una infección sostenida de persona a persona, debido a que la población humana cuenta con poca o incluso ninguna inmunidad en contra del agente patógeno; impredecibles pero recurrentes, son capaces de traer consecuencias sanitarias, económicas y sociales en todo el mundo”. (22)

### **2.1.10 Epidemiología molecular**

La epidemiología molecular se describe como la aplicación de métodos moleculares (como por ejemplo la proteómica y la metabolómica) para el desarrollo de análisis estadísticos que a su vez se apoyan de herramientas epidemiológicas sobre datos biológicos. El término fue descrito por primera vez entre las décadas de los 1970 y

1980 a partir de la unión de 3 áreas de estudio: epidemiología del cáncer, epidemiología ambiental y epidemiología de las enfermedades infecciosas con el objetivo principal de identificar los factores que determinan la transmisión, manifestación y la progresión de una enfermedad infecciosa. Usando como base los análisis basados en ácidos nucleicos, es capaz de monitorear brotes de infección, así como patrones y cadenas de transmisión dentro de la población. (23,24)

#### **2.1.11 Genómica del virus**

El genoma del SARS-CoV-2 está conformado por una cadena de 29,903 nucleótidos de largo, presentando una similitud de un 79% y 50% con el SARS-CoV y el MERS-CoV, respectivamente. Aproximadamente el 70% de este genoma codifica en un marco abierto de lectura (ORF, por sus siglas en inglés), conocido como ORF1ab, el cual está involucrado con el proceso de replicación viral y la patogénesis del virus y codifica para 16 proteínas no estructurales (nsps 1-16), 9 proteínas accesorias (ORFs 3a, 3b, 6, 7a, 7b, 8, 9a, 9b, y 10) y 4 proteínas estructurales [membrana (M), nucleocápside (N), envoltura (E) y espícula o spike (S)]. De todas las anteriores, la proteína S es la protagonista principal en el proceso de infección celular y como blanco de la respuesta inmune protectora. (7,8)

#### **2.1.12 Mutaciones del virus**

Los virus poseen una tasa de evolución muy elevada (más que cualquier otro organismo), esta tasa de evolución se ha atribuido a varias causas: grandes poblaciones, cortos tiempos de generación y altas tasas de mutaciones (siendo estas últimas las más importantes). Las mutaciones son errores producidos durante la replicación del genoma viral y que se usan como estimadores de la cantidad de diversidad genética generada. La tasa de mutación genómica es otro indicador importante, siendo el producto de la tasa de mutación por sitio de nucleótido y el tamaño del genoma y funciona para determinar el número promedio de mutaciones comparándolo con el genoma original. Los virus de ARN (como en el caso del SARS-CoV-2) son los que tienen mayores tasas de mutaciones. (9)

### **2.1.13 Variantes en Salud Pública**

Las variantes aparecen a través de los cambios en los nucleótidos que se dan de forma natural en el genoma durante la replicación. Estos cambios genómicos brindan al virus una amplia gama de ventajas con respecto a su replicación, transmisión y evasión inmune. (25)

Para 2021 se habían reportado más de 10,000 variaciones genéticas; esto se debe a su rápida propagación en personas de distintas edades, composiciones genéticas y predisposiciones médicas, las cuales le dan al virus un ambiente ideal para el desarrollo de poblaciones heterogéneas. (26)

La OMS ha clasificado las cepas variantes del SARS-CoV-2 en tres importantes grupos: variante de preocupación (VOC, del inglés *Variant of Concern*), variante de interés (VOI, del inglés *Variant of Interest*) y variantes bajo monitoreo (VUM, del inglés *Variant Under Monitoring*). Un gran número de variantes han surgido a partir de la cepa inicial aislada en Wuhan en 2019.

Las VUM son definidas como aquellas variantes que se sospecha que pueden presentar cambios genéticos que afecten sus características. Las VOI son aquellas que cumplen la definición de VUM y que, además, se sabe que sus cambios afectan sus características, como pueden ser: virulencia, transmisibilidad, patogenicidad, detectabilidad, etc. Las VOC son definidas como aquellas que cumplen la definición de VOI y que cumplen con al menos una de las siguientes características: Cambio perjudicial en la gravedad de la enfermedad clínica; o cambio en la epidemiología de la COVID-19 que causa un impacto sustancial en la capacidad de los sistemas de salud para brindar atención a pacientes con COVID-19 u otras enfermedades y, por lo tanto, requerir importantes intervenciones de salud pública; o disminución significativa de la eficacia de las vacunas disponibles para la prevención de enfermedad grave. (10)

Actualmente tenemos cinco variantes que han sido clasificadas como VOC, 17 variantes clasificadas como VOI y cinco variantes como VUM (a finales de 2023). (9,27,28)

**Tabla 1. Tipos de variantes del SARS-CoV-2**

<b>VOC</b>	<ol style="list-style-type: none"><li>1. Alfa (<math>\alpha</math>): Detectada a finales de septiembre de 2020, convirtiéndose en la cepa predominante en Reino Unido.</li><li>2. Beta (<math>\beta</math>): Detectada en octubre de 2020, se convirtió en la cepa predominante durante la segunda ola de contagios en Sudáfrica.</li><li>3. Gamma (<math>\gamma</math>): Detectada en brasileños que visitaban Japón en enero de 2021, responsable de la reinfección en Manaus.</li><li>4. Delta (<math>\delta</math>): Detectada en diciembre de 2020, responsable de la segunda ola de contagios en la India.</li><li>5. Ómicron (<math>\omicron</math>): Detectada en noviembre de 2021 en Sudáfrica, actualmente es la cepa dominante en el mundo.</li></ol>
<b>VOI</b>	<ol style="list-style-type: none"><li>1. Épsilon (<math>\epsilon</math>): Identificada en California, Estados Unidos.</li><li>2. Zeta (<math>\zeta</math>): Identificada en Brasil.</li><li>3. Eta (<math>\eta</math>): Identificada en Nigeria y Reino Unido.</li><li>4. Theta (<math>\theta</math>): Identificada en las Filipinas.</li><li>5. Iota (<math>\iota</math>): Identificada en Nueva York, Estados Unidos.</li><li>6. Kappa (<math>\kappa</math>): Identificada en la India.</li><li>7. Lambda (<math>\lambda</math>): Identificada en el Perú.</li><li>8. Mu (<math>\mu</math>): Identificada en Colombia.</li><li>9. Delta Plus (<math>\delta+</math>, B.1.617.2.1): Identificada en India</li><li>10. B.1.1.519: Identificada en México</li><li>11. XBB.1.5 (sublinaje de Ómicron)</li><li>12. XBB.1.16 (sublinaje de Ómicron)</li><li>13. EG.5 (sublinaje de Ómicron)</li><li>14. BA.2.86(sublinaje de Ómicron)</li><li>15. N.1. (sublinaje de Ómicron)</li><li>16. BA.2.86 (sublinaje de Ómicron)</li><li>17. JN.1 (sublinaje de Ómicron)</li></ol>

## VUM

1. XBB.1.9.1 (sublinaje de Ómicron)
2. XBB.1.9.2 (sublinaje de Ómicron)
3. XBB.2.3 (sublinaje de Ómicron)
4. DV.7 (sublinaje de Ómicron)
5. XBB (sublinaje de Ómicron)

### 2.1.14 Árboles filogenómicos

Se trata de diagramas jerárquicos que describen las relaciones que existen entre organismos. Estos se basan en una hipótesis de ascendencia común entre dichos organismos, agrupando a aquellos que presenten alguna mutación en común en su código genético (29). Los árboles filogenómicos se componen de:

- Puntas u hojas, representan el genoma del organismo. Se puede representar únicamente como una rama terminal, un círculo, o colocando el nombre de la muestra.
- Nodos internos: Representan ancestros comunes hipotéticos, los cuales no se observan de manera directa.
- Ramas: Forman conexiones entre los nodos y las puntas dentro del árbol, representan las relaciones existentes entre ancestros y descendientes. Pueden ser externas si conectan con un solo nodo o internas si conectan con dos.

### 2.1.15 Vectores de infección / transmisión

Los vectores de infección y/o transmisión son organismos vivos que son capaces de transmitir enfermedades (conocidas como enfermedades transmitidas por vector) entre animales y humanos, o de humano a humano. Una gran cantidad de estos vectores son insectos voladores (como los mosquitos), los cuales ingieren durante su proceso de alimentación al microorganismo causante de la enfermedad (principalmente algún virus) presente en un huésped ya infectado, el cual puede ser tanto humano como animal; posteriormente este mismo microorganismo es “inyectado” en otro huésped en un nuevo ciclo de alimentación.(30)

Por otra parte, las enfermedades zoonóticas se definen como toda aquella enfermedad transmitida de forma natural desde animales vertebrados hacia los humanos. Esto último a través del contacto directo, la ingesta de los mismos, tomas

de agua o el ambiente en sí, lo que vuelve a estas enfermedades tan impredecibles y con capacidad de provocar alertas globales. Podemos mencionar que igualmente, los humanos podemos transmitir enfermedades a los animales.(31)

### **2.1.16 Poblaciones inmunológicamente susceptibles**

En el combate a las enfermedades, cada persona puede reaccionar de forma diferente debido a una amplia gama de factores que repercuten directamente con la capacidad del sistema inmune del individuo, como, por ejemplo: edad, sexo, microbiota, condiciones climáticas, condiciones de salud, hábitos alimenticios y composición genética del mismo. El conjunto de todos estos factores es conocido como susceptibilidad del huésped y es el responsable de que se presente una enfermedad asintomática o una que atente contra la vida del mismo. (32)

## **2.2 Antecedentes del problema**

### **2.2.1 Descubrimiento de los HCoV**

El descubrimiento de los primeros coronavirus humanos (HCoV) se dio en la década de 1960, identificados como agentes causales de enfermedades leves en el sistema respiratorio, debido al estudio de pacientes con infecciones leves en las vías respiratorias superiores. Estos primeros HCoV fueron nombrados como: HCoV-229E y HCoV-OV43. (33)

Desde entonces se han descubierto otros HCoVs en el mundo, como aquellos identificados en el 2004 en los Países Bajos y en Hong Kong. El primero fue aislado de una muestra tomada de un niño de 7 meses de edad que presentaba síntomas de fiebre; este virus fue nombrado como HCoV-NL63. El segundo fue aislado de un hombre de 71 años que se encontraba hospitalizado a causa de neumonía y bronquiolitis; el virus fue nombrado como HCoV-HKU1. En la actualidad, todos los coronavirus se encuentran dispersos por todo el mundo. (34)

Estos 4 HCoVs identificados inicialmente se caracterizan por la producción de enfermedades con un cuadro clínico que va de leve a moderado y son los principales responsables de las enfermedades de temporada que afectan las vías respiratorias en todo el mundo (principalmente durante épocas de invierno). Debido a lo anterior, estos virus son pobremente descritos en la literatura, a pesar de haberse descubierto hace ya tiempo. (35)



En 2002, se identificó la primera enfermedad infecciosa del siglo XXI, siendo igualmente la primera causada por un HCoV. El síndrome respiratorio agudo grave (SARS), provocado por la infección del conocido SARS-CoV, desencadenó la epidemia por SARS entre 2003 y 2004. (36)

A mediados de 2012, fue identificado un sexto HCoV que provoca una enfermedad infecciosa, el MERS-CoV, causante del síndrome respiratorio de Medio Oriente (MERS), provocando el subsecuente brote de MERS en el mismo año. A diferencia de los primeros HCoVs identificados, que causan enfermedades moderadas, estos dos últimos HCoV han sido estudiados a detalle y se han realizado grandes avances, desde el 2003, en el entendimiento de los mecanismos de infección y en el desarrollo de sistemas de diagnóstico y tratamiento. (37)

### **2.2.2 Brote del SARS-CoV-2**

El SARS-CoV-2 apareció en diciembre de 2019 en China, siendo identificado como un nuevo virus hasta enero de 2020. El primer caso sospechoso se reportó el día 8 de diciembre de 2019 y la primera muerte por la nueva enfermedad se reportó el 9 de enero de 2020. El día 13 de enero de 2020 se reporta en Tailandia el primer caso de COVID-19 fuera de China. Para el día 31 de enero de 2020, la OMS declara al COVID-19 como una emergencia de Salud Pública de importancia internacional, siendo esta la sexta vez que se declara este estado en el mundo. El primer caso en México se identificó el 28 de febrero y en el estado de Hidalgo el 19 de marzo. Finalmente, el día 11 de marzo de 2020 el director general de la OMS, declara el brote de la nueva enfermedad como una pandemia después de reportarse 118,319 casos confirmados por laboratorio y 4,292 muertes en más de 198 países de las 6 regiones de la OMS. (5,38–41)

### **2.2.3 Pandemia por COVID-19 (11/marzo/2020-05/mayo/2023)**

Desde la declaración del COVID-19 como pandemia el 11 de marzo de 2020, la enfermedad se propagó por el resto del mundo. Al 16 de marzo de 2021 ya se habían reportado más de 119,791,453 casos y 2,652,966 muertes (4). Las campañas globales de vacunación comenzaron en diciembre de 2020, encabezadas por la vacuna de la farmacéutica Pfizer/BioNTech (la primera aprobada para su uso de emergencia, en diciembre de 2020). Esto logró reducir las

tasas de infección y los síntomas graves que conllevan a su vez a una menor tasa de mortalidad, incluso con aquellas variantes del virus reportadas a la fecha. (42,43) Para México, la campaña de vacunación comenzó el 24 de diciembre de 2020 en la Ciudad de México, el estado de México y el estado de Querétaro; iniciando únicamente con el personal de sector salud y escalando progresivamente hacia la población en general, siendo María Irene Ramírez (jefa del área de enfermería de la Unidad de Terapia Intensiva del Hospital General “Rubén Leñero”) la primera mexicana en vacunarse en contra de la enfermedad. En la misma fecha, SALUD reportaba 1,362,564 casos confirmados y 121,172 muertes por COVID-19 en territorio nacional, mientras que en el estado de Hidalgo se reportaron 23,208 casos confirmados a COVID-19 y 3,186 defunciones desde el comienzo de la emergencia sanitaria en la entidad. (44,45)

Finalmente, el 5 de mayo de 2023 y tras más de 3 años de pandemia, la OMS declaró que la COVID-19 ya no representaba una emergencia de Salud Pública de importancia internacional, terminando así con la pandemia de COVID-19. En esa misma fecha se contabilizaban 764,416,156 casos confirmados y 6,914,914 defunciones por COVID-19, de igual manera se habían administrado 13,300 millones de dosis de vacunas contra la COVID-19 en todo el mundo. (46,47) En México se reportaron 7,587,643 casos y 333,913 defunciones por COVID-19, de los cuales 129,724 casos y 8,399 defunciones corresponden al estado de Hidalgo.(46)

#### **2.2.4 Postpandemia por COVID-19**

Al 2 de agosto de 2023, la Organización Mundial de la Salud ha reportado más de 760 millones de casos y 6.9 millones de defunciones en todo el mundo a causa de esta enfermedad, aunque se cree que estos números pueden ser mucho mayores. En México se reportan más de 7 millones de casos y 300 mil muertes por COVID-19, correspondiendo al estado de Hidalgo 130,638 casos y 8,419 defunciones, esto de acuerdo al último reporte técnico oficial publicado. (6)

### **2.3 Planteamiento del problema**

Aunque oficialmente la pandemia ha terminado, el SARS-CoV-2 sigue circulando entre la población, habiendo infectado a más de 760 millones de personas para agosto de 2023, con una tasa de mortalidad global del 0.91%. En México, el virus ha infectado a más de 7 millones de personas y presenta una mortalidad del 4.3%, cifra alarmante si se compara con los datos globales. En el estado de Hidalgo se han infectado poco más de 130 mil personas y se presenta una mortalidad aún más alarmante, del 6.6% (46). Es importante conocer el comportamiento del virus en la población ya que permitirá establecer medidas de prevención y control.

Desde su aparición, la cepa inicial del virus ha sufrido un gran número de mutaciones, que le ha permitido alterar su genoma y desarrollar variantes de sí misma, presentándose hasta la fecha cinco variantes de preocupación (tabla I): alfa, beta, gamma, delta y ómicron (21). La variante delta fue responsable de la tercera oleada de COVID-19 en México y la variante ómicron la responsable de la cuarta. A su vez, esta última ha terminado por desplazar por completo a las demás variantes (a través de sus sublinajes) convirtiéndose en la cepa dominante en todo el mundo hasta 2024 (24). Al día de hoy, no hay evidencia publicada de estudios que permitan observar el comportamiento de la pandemia en las entidades federativas del país de manera individual.

La sintomatología descrita en reportes clínicos de los casos y en la definición operacional de casos sospechosos para COVID-19 se comparte con la de muchas otras enfermedades respiratorias causadas por virus, como es el caso de la influenza. Los agentes etiológicos de estas enfermedades se encuentran presentes de manera simultánea en el ambiente una vez concluida la emergencia de salud y únicamente pueden ser identificados por técnicas de diagnóstico molecular (7). Las mutaciones en los virus les permiten adquirir mejoras en sus procesos de replicación, transmisión y evasión inmune. (21)

La información epidemiológica en México se ha obtenido a través del conjunto de datos suministrados por el Sistema Nacional de Vigilancia Epidemiológica (SINAVE), el Instituto de Diagnóstico y Referencia Epidemiológicos “Dr. Manuel Martínez Báez” (InDRE) y la Red Nacional de Laboratorios de Salud Pública

(RNLS). La combinación de esta información con la información genómica disponible puede ser útil para identificar el comportamiento de la enfermedad y de la población en un periodo de tiempo determinado, lo que permite el desarrollo de métodos que puedan facilitar la toma de decisiones para la prevención y la contención de futuros brotes.

### **III. Justificación**

La COVID-19 es un padecimiento causado por la infección del SARS-CoV-2, el cual puede ser confirmado únicamente por RT-PCR, considerada el estándar de oro. El virus, descubierto en China a finales de 2019, se esparció rápidamente por todo el mundo gracias al proceso de globalización sin precedentes que vivimos hoy en día. El virus fue identificado en México a finales de febrero de 2020. En marzo de 2020, la OMS declaró a la nueva enfermedad como una pandemia, tras haberse identificado en casi 200 países pertenecientes a las 6 regiones de la organización; esta misma fue concluida en mayo de 2023 al considerar que la enfermedad ya no representaba una emergencia de Salud Pública de importancia internacional.

Desde su aparición inicial, el virus ha sufrido un sinnúmero de mutaciones (eventos naturales en el proceso evolutivo de los virus), de las cuales se han originado nuevas variantes distintas al virus original. Estas variantes han desarrollado nuevas capacidades (ejemplo de ello es una mayor velocidad de propagación), que les permiten disminuir la efectividad de todas las medidas de contención utilizadas hasta ahora (como, por ejemplo, las vacunas); esto las ha convertido en objeto de interés y sobre todo de preocupación para la salud pública.

Este proyecto busca describir la epidemiología del COVID-19 en el estado de Hidalgo desde su introducción, así como comparar las secuencias genómicas de las diversas variantes reportadas en el estado de Hidalgo. Lo anterior permitirá deducir el momento en el que una cepa de interés circule en la población y cómo podrían comportarse en una población inmunológicamente susceptible. De igual forma, el estudio permitirá el desarrollo de nuevos métodos de búsqueda y contención de casos de infección y el seguimiento de la dinámica de la infección, a través de la obtención de estimadores descriptivos de la pandemia en el Estado.

Finalmente, la combinación de los datos de la epidemiología clásica y genómica permitirá reconocer los diferentes cambios en el comportamiento del brote a lo largo de la pandemia.

#### **IV. Hipótesis**

H1. El comportamiento de la pandemia dentro del Estado de Hidalgo fue similar al comportamiento medio de la misma dentro del país.

H0. El comportamiento de la pandemia dentro del Estado de Hidalgo fue distinto al comportamiento medio de la misma dentro del país.

#### **V. Objetivos**

##### **5.1 Objetivo general**

Describir la epidemiología y la genómica de la pandemia de COVID-19 en el Estado de Hidalgo, México, utilizando estimadores clásicos y moleculares.

##### **5.2 Objetivos específicos**

- Describir la epidemiología de COVID-19 en el estado de Hidalgo, desde su comienzo en 2020, considerando variables como sexo, edad, municipio, sintomatologías y comorbilidades.
- Desarrollar curvas epidémicas y de mortalidad de los casos confirmados para el Estado de Hidalgo.
- Analizar el comportamiento de la pandemia por SARS-CoV-2 a través de la construcción de árboles filogenómicos de las diferentes variantes del virus en el Estado de Hidalgo.
- Estudiar la distribución de variantes genómicas del SARS-CoV-2 en el Estado de Hidalgo.

#### **VI. Materiales y métodos**

##### **6.1 Contexto de la investigación**

El trabajo se llevó a cabo como proyecto de investigación perteneciente a la Maestría en Ciencias Biomédicas y de la Salud de la Universidad Autónoma del Estado de Hidalgo (UAEH). La recolección de los datos se realizó en el InDRE, ubicado en la Ciudad de México. El análisis de los datos se realizará tanto en el

InDRE como en el Instituto de Ciencias de la Salud (ICSa), UAEH, ubicado en el municipio de San Agustín Tlaxiaca, en el estado de Hidalgo, México, durante el periodo de enero de 2023 a febrero de 2024.

## **6.2 Tipo y diseño de estudio**

Estudio observacional del tipo transversal.

El estudio se realizó de acuerdo con los siguientes puntos:

1. Se recolectó toda la información relacionada con los casos confirmados de COVID-19 en el estado de Hidalgo, México. Se recolectaron los genomas secuenciados mediante RT-PCR y publicados por el InDRE en el sistema GISAID (del inglés *Global Initiative on Sharing All Influenza Data*), correspondientes al estado de Hidalgo.
2. La información fue depurada basándose en los criterios de inclusión, exclusión y eliminación que se muestran más adelante (sección 6.3) mediante el uso de la herramienta ofimática de R Studio.
3. Se describió la caracterización clínica de los casos confirmados y se realizaron pruebas de asociación entre las distintas variables.
4. Se construyó una curva epidémica de los casos, así como una curva de mortalidad con ayuda de la herramienta ofimática R Studio.

La curva epidémica se refiere a una representación gráfica, en un eje de coordenadas, de las frecuencias (diarias, semanales, mensuales, anuales, etc.) de la enfermedad. Usualmente con una distribución asimétrica, el eje horizontal representa el tiempo, mientras que el eje vertical representa las frecuencias.

5. Se construyeron árboles filogenómicos a partir de las secuencias del SARS-CoV-2 correspondientes al estado de Hidalgo disponibles en GISAID, mediante el uso del software Nextstrain. En la sección 6.7 se describe la función del programa y en el anexo 2 se muestra con detalle los pasos realizados para la construcción del árbol filogenético.
- Los árboles filogenómicos o filogenéticos son esquemas que permiten visualizar las relaciones ancestrales, en este caso, del virus del SARS-CoV-2 de un grupo de individuos muestreados. Existen diversas maneras de

presentar un árbol filogenómico. Esto permite observar una aproximación de la red de transmisión.

- R Studio: Se trata de un entorno de desarrollo integrado, de código abierto, para el lenguaje de programación R, que facilita la ejecución del código. R está enfocado en el análisis estadístico computacional y la producción de gráficos (última versión de R disponible: 4.3.2 “Eye Holes”).
  - Nextstrain: Iniciativa y software de código abierto que permite el alineamiento de secuencias a través de MAFFT (del inglés *Multiple Alignment using Fast Fourier Transform*) y la construcción de árboles filogenómicos basados en el método de máxima verosimilitud. Constituido por diversas partes que permiten, en conjunto visualizaciones, canales de análisis, gestión de datos y más relacionado con la bioinformática.
6. Se describió la distribución temporal de las variantes del SARS-CoV-2 en el Estado.

### **6.3 Selección de la población, criterios de inclusión y exclusión**

- Todo aquel caso que haya sido confirmado, a través de un estudio de laboratorio, como positivo o negativo a la infección por SARS-CoV-2 y reportado a través del SINAVE y la DGE a partir del primer caso confirmado. Se utilizaron las secuencias genómicas disponibles de casos confirmados para SARS-CoV-2 mediante RT-PCR en el Estado de Hidalgo, México, depositadas por el InDRE en las bases de datos de GISAID.
- Criterios de inclusión: Todo caso confirmado por laboratorio, que cumpla con la definición operacional de caso sospechoso de la OMS y que se encuentre registrado en las bases de datos de la Dirección General de Epidemiología (DGE).
- Criterios de exclusión: Todo caso que haya sido confirmado a través de métodos que no sean por laboratorio. Aquellos casos que no se encuentren reportados en el SINAVE.

## 6.4 Tamaño de muestra y muestreo

Muestreo por conveniencia. Se utilizó la totalidad de los casos confirmados como positivo y negativo en el estado de Hidalgo observados en la base de datos de la DGE e InDRE, alcanzando una N de 248,392 casos.

## 6.5 Diagrama de diseño experimental

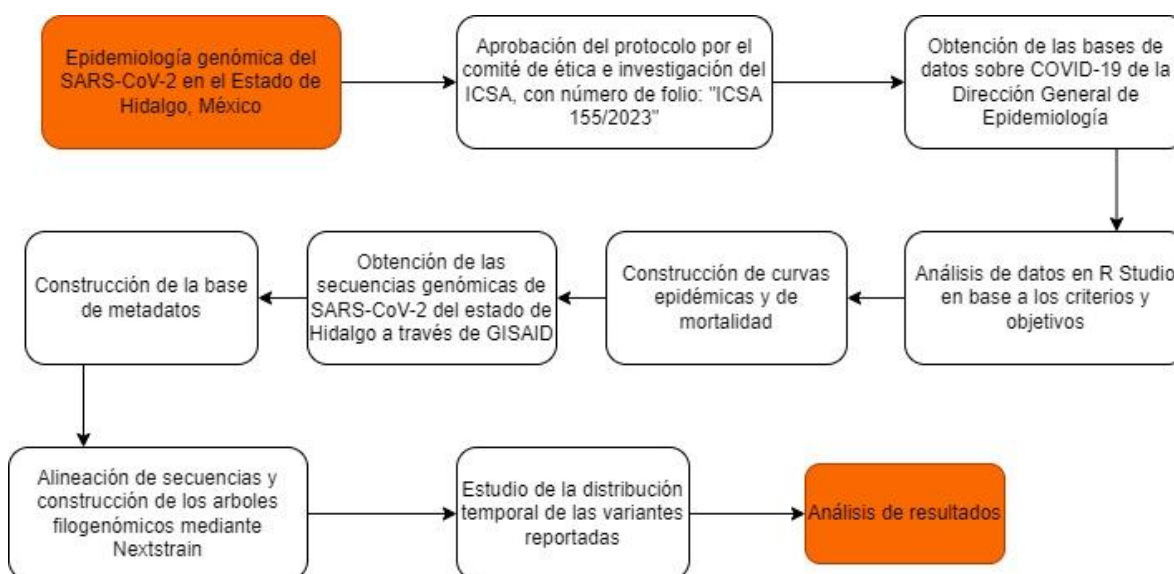


Figura 1. Diagrama de diseño experimental utilizado en este proyecto.

## 6.6 Definición de variables

Las variables utilizadas fueron recolectadas a partir de la información oficial disponible en el SINAVE. (Ver Anexo 1)

## 6.7 Análisis estadístico

Se realizaron pruebas de asociación (prueba de Chi cuadrada) para identificar la existencia de dependencias / independencias entre las distintas variables.

## VII. Resultados

### Descripción epidemiológica

Se realizó un análisis exploratorio a la base de datos relacionadas a COVID-19 disponible para su descarga en el sitio oficial de la Dirección General de Epidemiología (DGE) mediante el uso de R Studio, enfocándonos en las variables de interés mencionadas previamente en la metodología. Actualmente en México, la DGE dispone de datos sobre los casos asociados a COVID-19 desde el año 2020



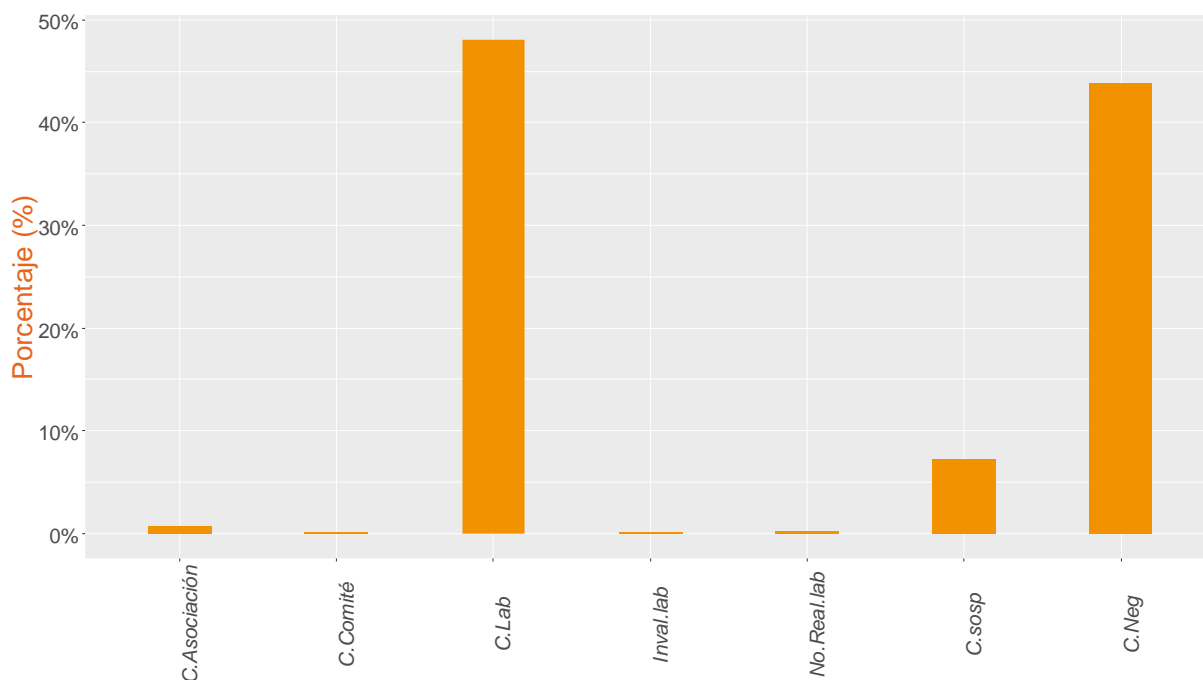
hasta 2023 (48). Contando con un total de 20,367,516 observaciones correspondientes a todo México a lo largo de todo este periodo (tabla 2).

**Tabla 2. Observaciones de la base de datos de COVID-19 (48).**

<b>Año</b>	<b>Observaciones Totales (%)</b>	<b>Observaciones del Estado (% nacional)</b>	<b>Confirmados del Estado (% del estado)</b>	<b>Negativos del Estado (% del estado)</b>
<b>2020</b>	3,868,396	49,834 (1.29%)	25,837 (51.8%)	18,666 (37.45%)
<b>2021</b>	8,830,345	92,950 (1.05%)	37,119 (39.93%)	51,127 (55.14%)
<b>2022</b>	6,451,944	101,048 (1.57%)	57,588 (57%)	34,906 (34.54%)
<b>2023</b>	1,216,831	26,545(2.1%)	9,371 (37.63%)	13,778(52.31%)
<b>Total</b>	20,367,516	270,377(1.32%)	129,915(48.05%)	118,477(43.82%)

Información obtenida de la base de datos de la DGE

Las observaciones dentro de la base de datos se clasifican en 7 grupos: Confirmado por asociación, confirmado por dictaminación, confirmado por laboratorio, invalido por laboratorio, no realizado por laboratorio, caso sospechoso y caso negativo (ver Anexo 1). El porcentaje total de casos correspondiente a cada una de las clasificaciones mencionadas se muestra en la figura 2. Para este proyecto nos enfocamos únicamente en los casos confirmados por laboratorio y los casos negativos.



Clasificación de las observaciones

Figura 2. Clasificaciones de tipos de casos derivada de la base de datos, se observa que casi el 50% corresponde a casos confirmados por laboratorio.

A través del análisis se observó que el total de observaciones correspondientes al estado de Hidalgo fue de 270,377 contribuyendo así con el 1.32% del total de observaciones nacionales. De estas, 129,915 registros son casos positivos (48%) y 118,477 son casos negativos (43.8%). El 8.2% restante corresponde a observaciones que se encuentran fuera de estas 2 clasificaciones.

El Estado de Hidalgo se compone de 12 regiones operativas que integran a 84 municipios, de acuerdo a características culturales, económicas y geográficas comunes. La distribución de casos dentro de estas regiones se muestra en la tabla 3. Se observa que la región con el mayor número de casos confirmados es la región de Pachuca de Soto, representando el 26.5% del total de casos en el estado, mientras que la región con el menor número de casos es la región de Jacala de Ledezma, representando únicamente el 0.4% del total de casos. De igual manera se presentan las defunciones de los casos confirmados por región operativa. Se observa que la región de Pachuca de Soto es la región con el mayor número de defunciones, siendo un total de 1,713 (20.1%).

**Tabla 3. Casos y defunciones por región operativa del Estado.**

<b>Región</b>	<b>Casos (Porcentaje %)</b>	<b>Defunciones (Porcentaje %)</b>
<b>Jacala de Ledezma</b>	529 (0.4%)	74 (0.9%)
<b>Huichapan</b>	3,071 (2.4%)	226 (2.7%)
<b>Apan</b>	10,531 (8.1%)	547 (6.4%)
<b>Zacualtipán de Ángeles</b>	2,826 (2.2%)	226 (2.7%)
<b>Ixmiquilpan</b>	4,021 (3.1%)	477 (5.6%)
<b>Actopan</b>	6,066 ( 4.7%)	768 (9%)
<b>Tizayuca</b>	12,652 (9.7%)	730 (8.6%)
<b>Mineral de la Reforma</b>	17,616 (13.6%)	734 (8.6%)
<b>Huejutla de Reyes</b>	3,603 (2.8%)	381 (4.5%)
<b>Pachuca de Soto</b>	34,447 (26.5%)	1,713 (20.1%%)
<b>Tulancingo de Bravo</b>	16,643 (12.8%)	1,150 (13.5%)
<b>Tula de Allende</b>	17,910 (13.8%)	1,494 (17.5%)
<b>Total</b>	129,915 (100%)	8,520 (100 %)
<i>Pearson's Chi-squared test Chi^2 = 662.408 d.f. = 12 p &lt;0.0000000000000002</i>		

La distribución por sexo del total de observaciones se muestra en la tabla 4, se observa una positividad a COVID-19 muy similar tanto para hombres como para mujeres. Sin embargo, estas últimas presentan una ligera positividad por encima de la de los varones, con un 52% del total de casos confirmados en el estado.

**Tabla 4. Sexo contra clasificación de los casos.**

Sexo	Clasificación		
	Positivo	Negativo	Total
Mujer	67,726	65,108	132,834
Hombre	62,189	53,369	115,558
<b>Total</b>	<b>129,915</b>	<b>118,477</b>	<b>248,392</b>

Pearson's Chi-squared: Chi<sup>2</sup> = 198.5087 d.f. = 1 p <0.0000000000000002

La figura 3 muestra los estadísticos para la variable de edad (en años cumplidos), agrupada por el sexo del caso. En ella se observa una media de edad en general de 40.8 años, una mediana edad de alrededor de 38 años para las mujeres y de 37 años para los hombres. El primer cuartil se ubica en los 28 años y el tercer cuartil en los 51 años. El rango de edad va de los 0 años hasta los 111 años.

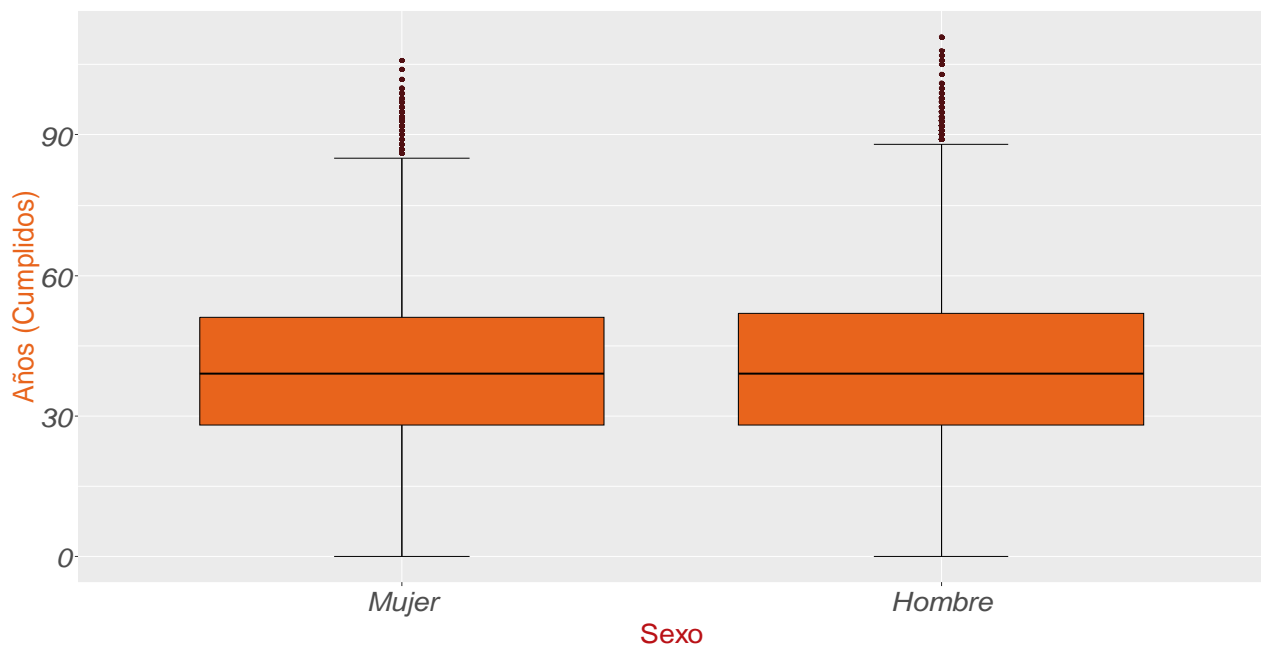
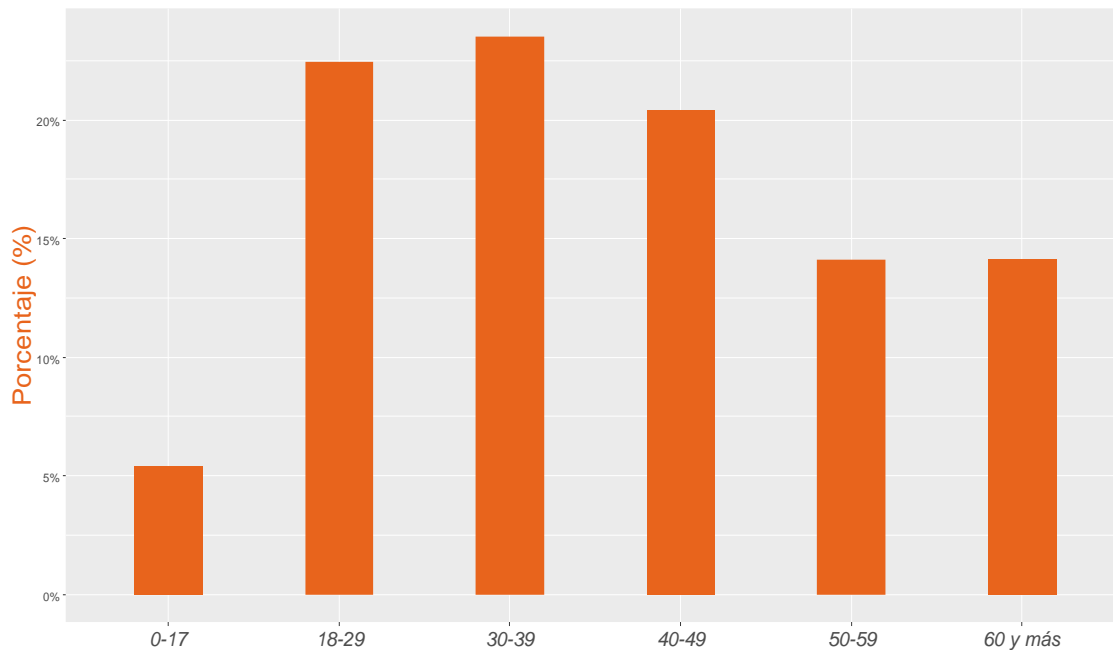


Figura 3. Estadísticos para la variable Edad. Se observan diversos valores atípicos en la edad de ambos sexos.

La edad de los casos fue agrupada, tomando como base el formato presentado en los reportes de situaciones nacionales sobre COVID-19 presentados por la DGE (0-17, 18-29, 30-39, 40-49, 50-59, 60 y más años) y se muestran en la figura 4. Se observa que el grupo de 30-39 años es el más afectado, con alrededor del 27% del

total de los casos confirmados en el Estado. Por el contrario, el grupo de 0-17 años es el grupo menos afectado, representando poco más del 5% del total de casos confirmados.



*Figura 4. Distribución de los casos positivos por grupo etario. De mayor a menor: 30-39, 18-29, 40-49, 60 y más, 50-59 y 0-17 años.*

Al estratificar los grupos etarios por el sexo del caso, se obtiene la figura 5, en la cual podemos apreciar que las mujeres de los grupos de 18 hasta 59 años son ligeramente más afectadas que los hombres. Por el contrario, dentro de los grupos de 0 a 17 y 60 y más, los hombres se ven más afectados con respecto a las mujeres.

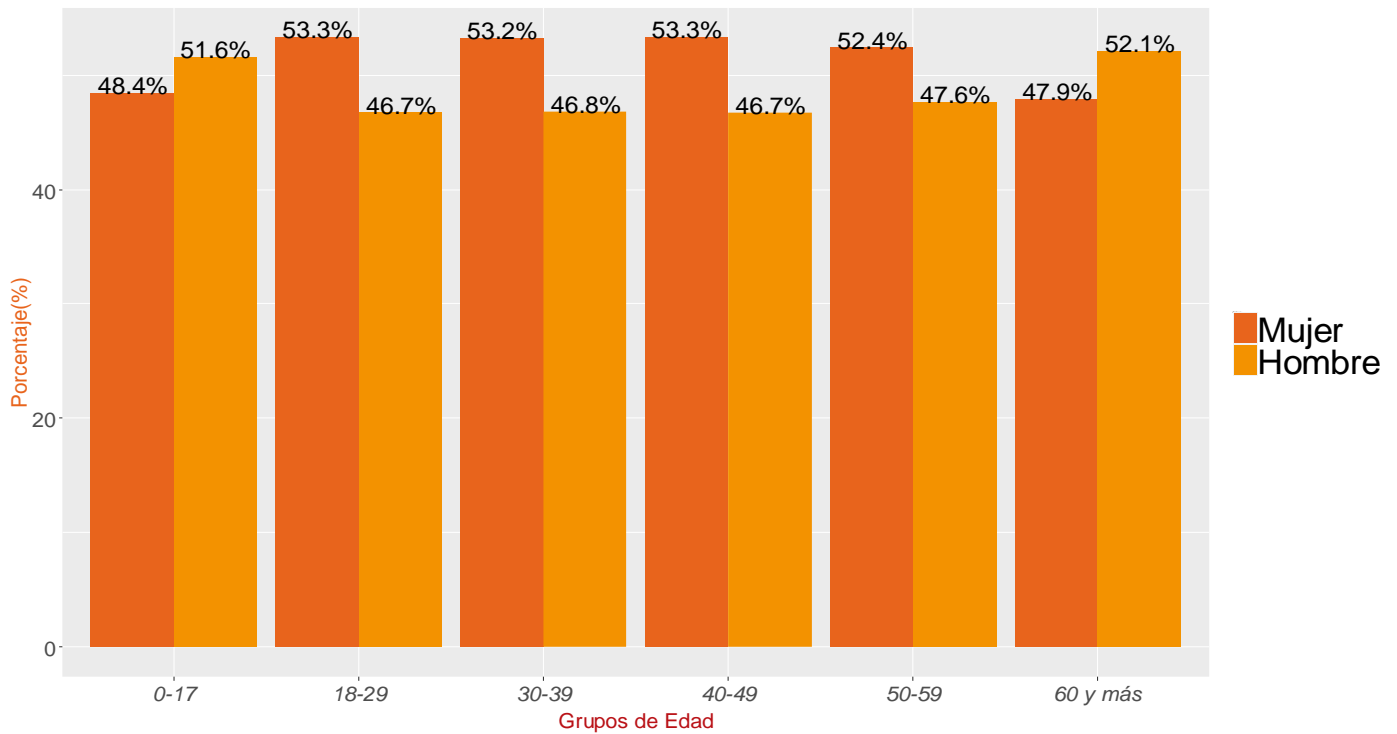


Figura 5. Grupos etarios separados por sexo del caso.

Las bases de datos contienen un conjunto de variables individuales correspondientes a las comorbilidades reportadas por el paciente al momento de la toma de datos de muestra. Para fines prácticos del trabajo se consideraron únicamente las principales 5 comorbilidades reportadas. Las comorbilidades se muestran de manera descendente en la figura 6. La hipertensión es la comorbilidad con mayor presencia entre los casos confirmados en el Estado (alrededor del 13%).

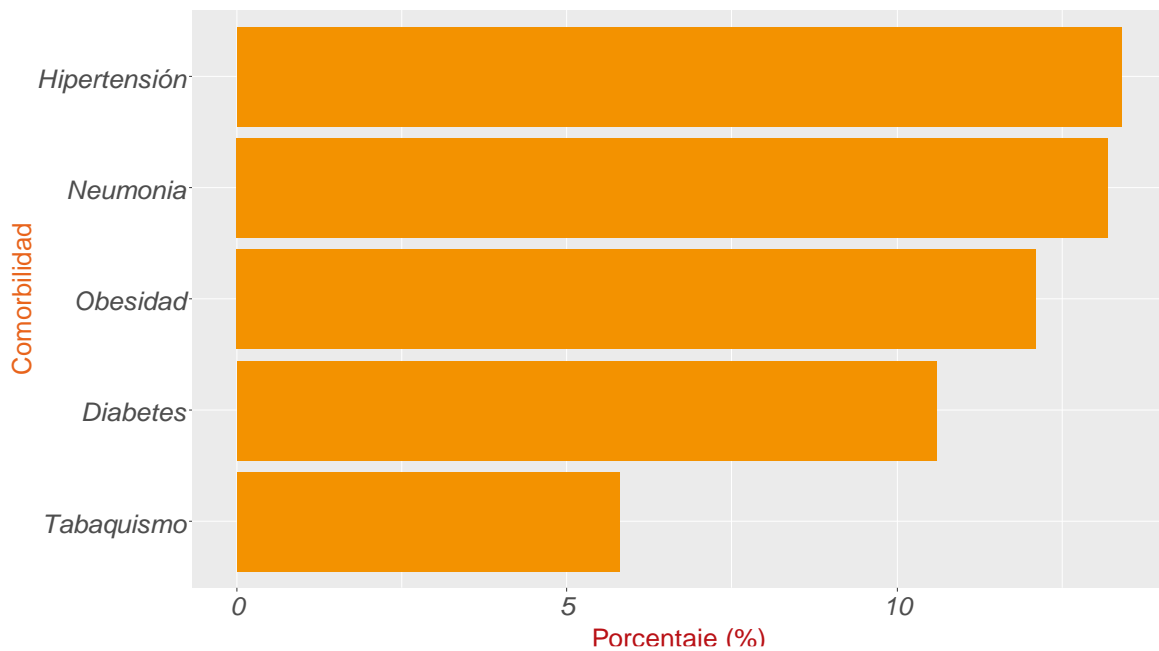


Figura 6. Comorbilidades reportadas por los casos confirmados de COVID-19.

De igual manera que con las comorbilidades, para las sintomatologías se consideraron únicamente los 8 principales síntomas reportados por los casos al momento de la toma de muestra. La figura 7 presenta estos síntomas de manera descendente. La tos es el principal síntoma de la enfermedad en el estado (con una presencia de alrededor del 82%).

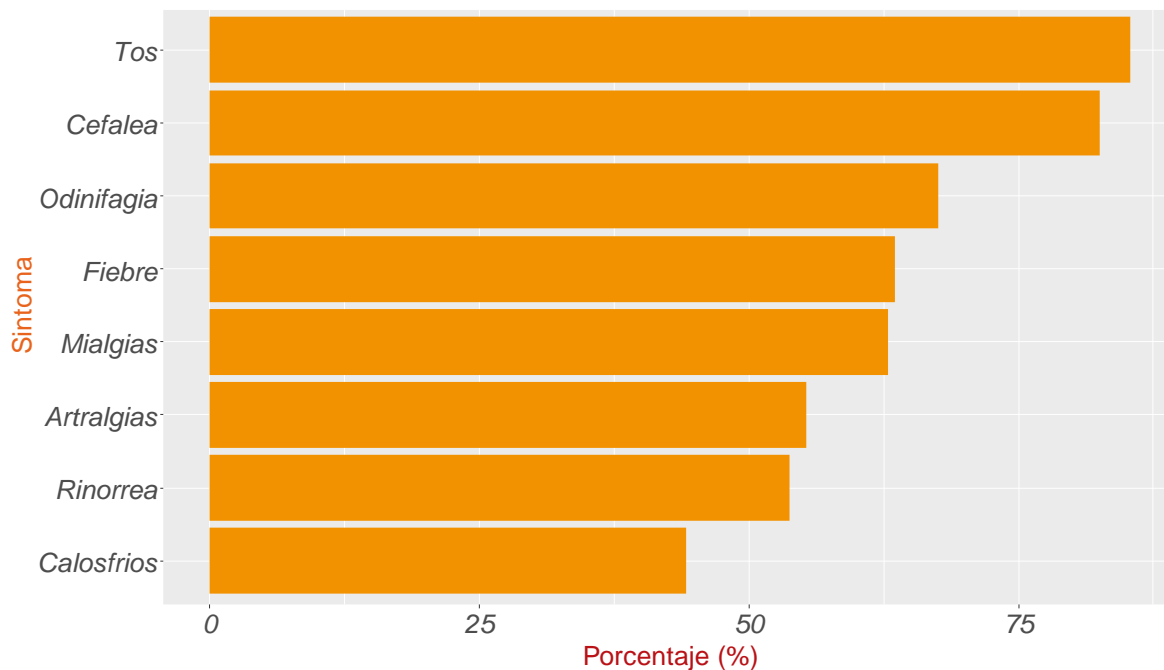


Figura 7. Síntomas reportados por los casos confirmados. Síntomas populares como disgeusia y anosmia se presentaron en un bajo porcentaje de la población

La curva epidémica se construyó tomando en cuenta la fecha de ingreso, convirtiendo el dato capturado en la base a un formato “dd-mm-yyyy”, para posteriormente obtener la fecha con su correspondiente semana epidemiológica. Se construyeron 2 gráficos, la primera considera únicamente los casos positivos (figura 8) y la segunda combina casos confirmados y negativos (figura 9). Se observa que los primeros casos en el estado fueron reportados en la semana epidemiológica 12 de 2020. De igual forma, se observa que el pico máximo de casos confirmados se dio en el repunte número 4, el cual ocurrió entre las semanas epidemiológicas 50 de 2021 y 10 de 2022.



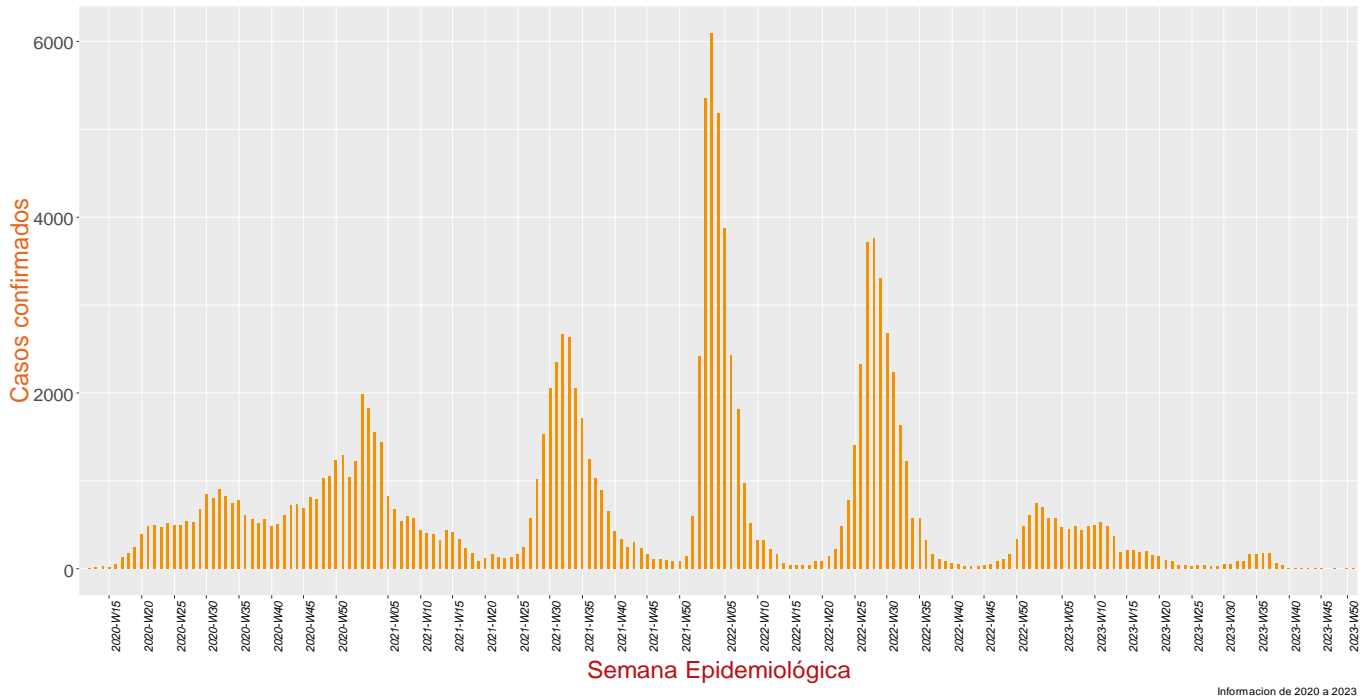


Figura 9. Casos confirmados en el estado de Hidalgo. Se observan 6 distintos picos, correspondientes a cada una de las olas reportadas hasta finales de 2023.

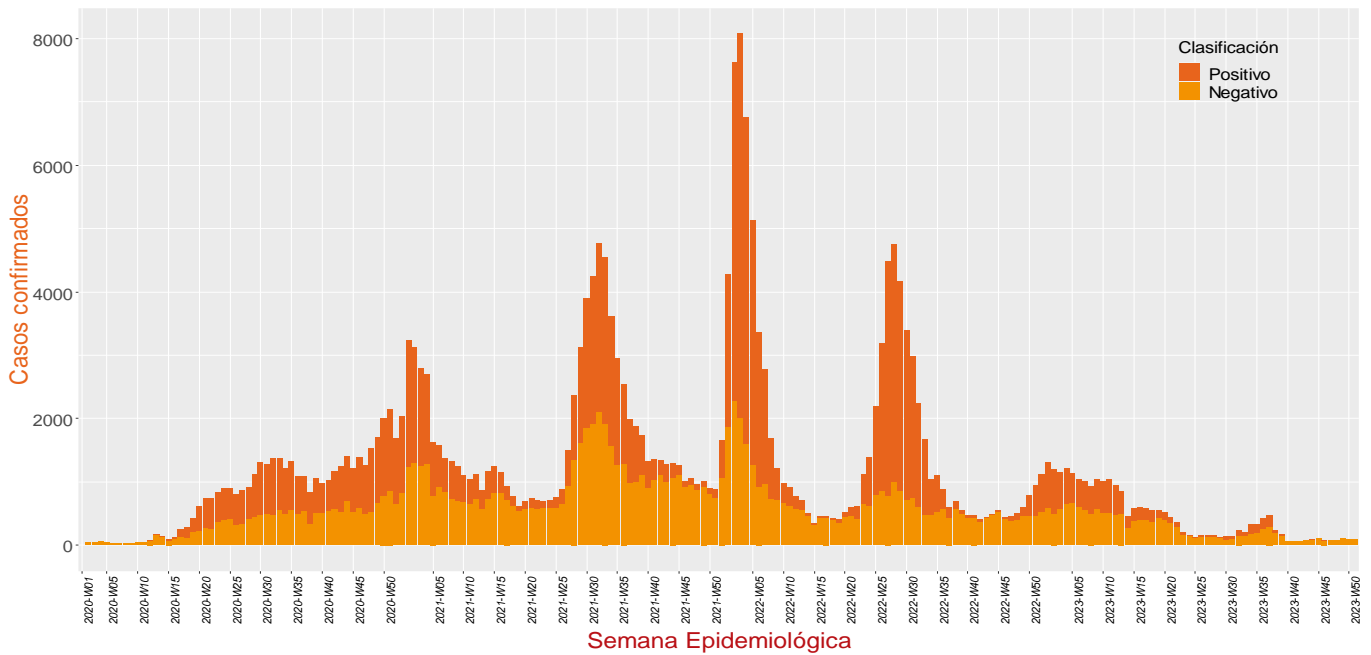
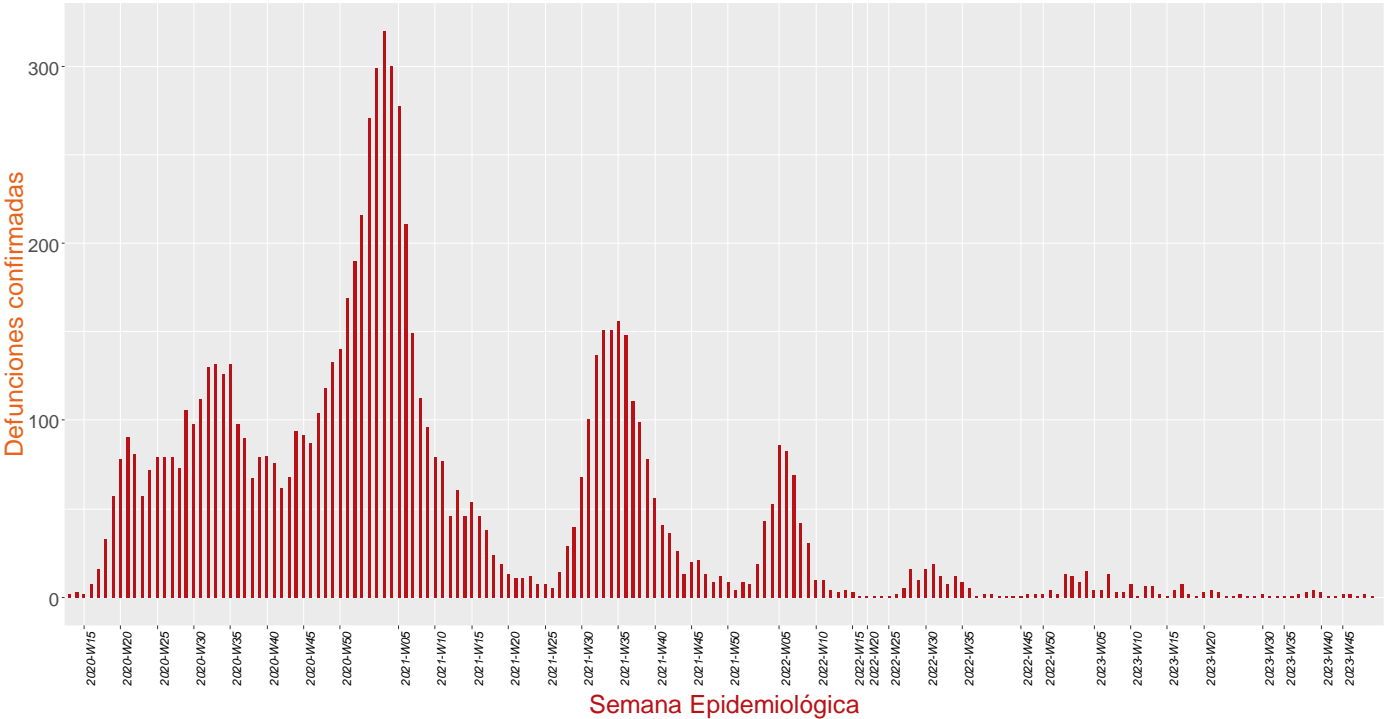


Figura 8. Acumulación de casos positivos y negativos en el estado de Hidalgo. Se observan 6 distintos picos, correspondientes a cada una de las olas reportadas hasta finales de 2023.

De igual manera, se construyó una curva de mortalidad únicamente de los casos confirmados, tomando la fecha de defunción como base y obteniendo de ella su semana epidemiológica correspondiente. Dicha curva se muestra en la Figura 10. Dentro de la curva se observa que las primeras defunciones en el Estado se dieron alrededor de la semana epidemiológica 15 de 2020. El pico máximo de defunciones se dio entre las semanas epidemiológicas 50 de 2020 y 5 de 2021. El total de defunciones en el Estado asciende a 8,520 defunciones (63% hombres), lo que representa una mortalidad del 6.55%.



Informacion de 2020 a 2023

Figura 10. Defunciones confirmadas por COVID-19 en el estado de Hidalgo hasta finales de 2023.

## Epidemiología genómica

En la plataforma GISAID se encontraron y descargaron 674 secuencias genómicas correspondientes a SARS-CoV-2 y al estado de Hidalgo desde 2020 hasta 2023, las cuales fueron secuenciadas y reportadas por el Instituto de Diagnóstico y Referencia Epidemiológicos (InDRE).

La construcción del árbol filogenómico con las secuencias de SARS-CoV-2 se realizó a través del software Nextstrain, el cual crea un árbol en base al método de máxima verosimilitud. Se realizó una alineación de secuencias (figura 11) y la construcción de los árboles filogenómicos.

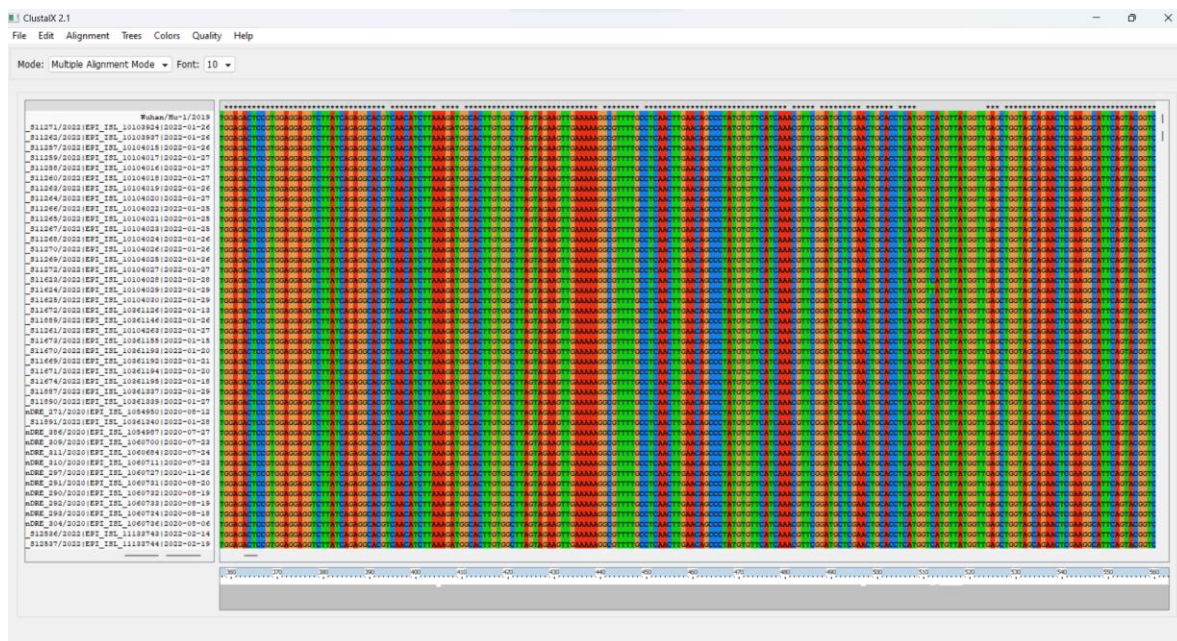


Figura 11. Visualización parcial en Clustal del alineamiento múltiple de las 674 secuencias de SARS-CoV-2 del estado de Hidalgo.

Para la construcción de los árboles se requiere, además de las secuencias genómicas, de una serie de metadatos. Estos metadatos incluyen información muy variada y de gran importancia, en donde queda registrada la edad y el sexo del caso, la fecha de recolección, el laboratorio de origen y de publicación, el país, la región, el linaje del virus, las mutaciones del virus y los autores responsables de su secuenciación y publicación (figura 12). Tales datos son recolectados al momento de la toma de muestra del caso seleccionado y complementados por el laboratorio de publicación.

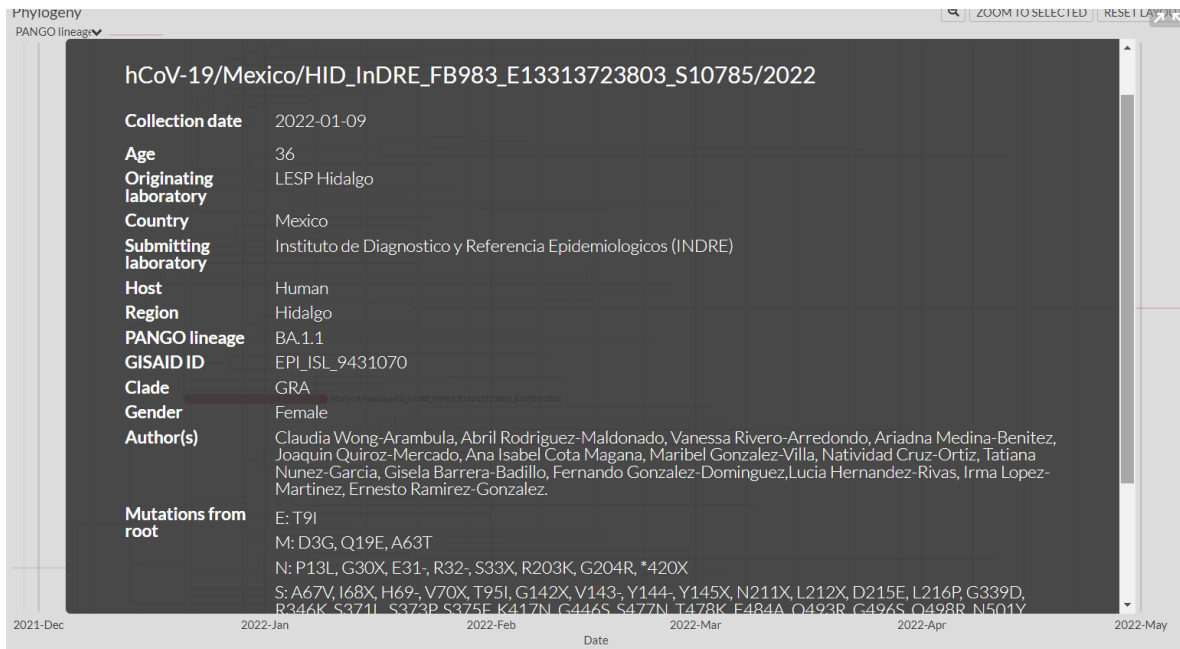


Figura 12. Metadatos de una secuencia del estado de Hidalgo, fuente GISAID.

En la figura 13, se observa la distribución temporal de las distintas variantes del SARS-CoV-2 (figura 13A), así como el número de genomas reportados por el estado de Hidalgo desde 2020 hasta finales de 2023, por semana epidemiológica (figura 13B). Dentro de la figura se aprecia una baja cantidad de genomas reportados en Hidalgo durante el periodo inicial de la emergencia sanitaria, los cuales fueron aumentando a partir de la semana epidemiológica 1 de 2021 y alcanzando su pico máximo en la semana epidemiológica 1 de 2022 (175 genomas reportados), para después disminuir nuevamente. La variante B.1.1.519 (reportada por primera vez en México) fue la primera variante identificada en el estado, alrededor de la semana epidemiológica 50 de 2020, desplazándose rápidamente con la aparición de las variantes alfa y gamma en el estado, aproximadamente en la semana 12 de 2021 y la semana 17 de 2021 respectivamente. La variante delta se identificó a partir de la semana 25 de 2021 y se convirtió en la variante predominante en el estado hasta la aparición de ómicron en la semana 50 del mismo año. A partir de ese momento todas las secuencias corresponden a la variante ómicron y sus sublinajes. La variante beta no fue identificada en el Estado.

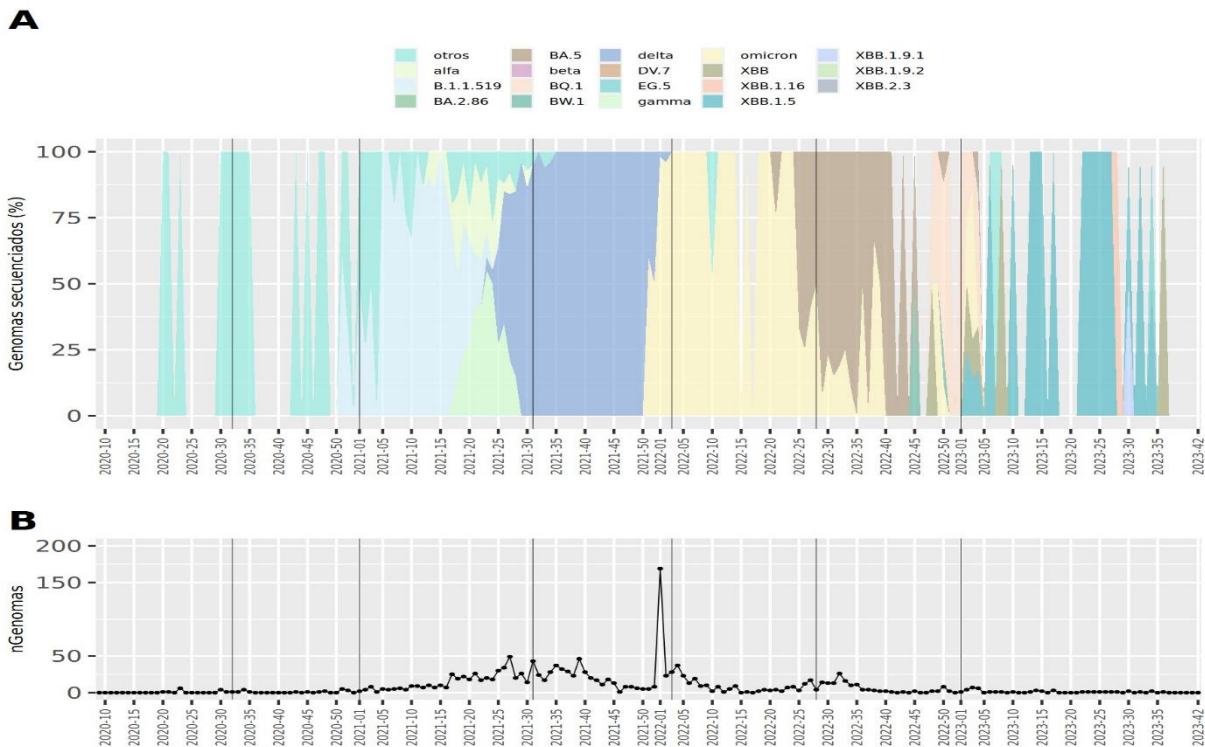


Figura 13. A) Distribución de las distintas variantes del SARS-CoV-2 en el Estado de Hidalgo a través del tiempo. B) Número de genomas reportados por el Estado de Hidalgo por semana epidemiológica (GISAID).

El árbol resultante se muestra en la figura 14, durante la construcción del árbol, el programa de Nextstrain filtra y elimina las secuencias genómicas con metadatos incompletos y con información genómica de baja calidad, por lo que el número de secuencias se redujo de 674 a 633. Las secuencias en color amarillo corresponden al estado de Hidalgo, mientras que aquellas en azul son secuencias de referencia de distintas partes del mundo, las cuales comparten similitudes genéticas entre sí y las secuencias del estado de Hidalgo, lo que lleva a un árbol con un total de 662 secuencias en su construcción. En la figura 14 se observa una similitud genética entre las secuencias reportadas por el estado y aquellas reportadas en países como pueden ser los Estados Unidos, sobre todo a lo que la variante ómicron se refiere.

## Hidalgo-specific genomic surveillance build

Data updated 2024-04-26.

Showing 662 of 662 genomes sampled between Dec 2019 and Feb 2024.

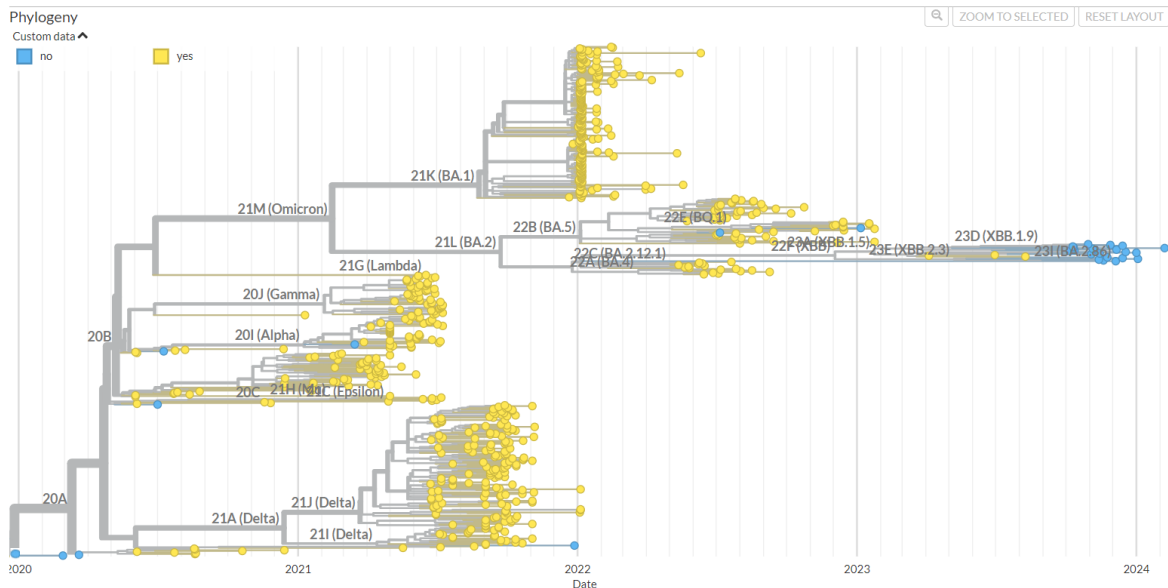


Figura 14. Árbol filogenómico de las secuencias de SARS-CoV-2 en el Estado de Hidalgo, México. Construido con 662 genomas recolectados entre dic 2019-feb 2024, de estas, 633 pertenecen al estado de Hidalgo (amarillo) y 29 como referencia de todo el mundo (azul).

Dentro del árbol podemos observar que de las 662 secuencias 43 corresponden a la variante B.1.1.519 (6.19%) (figura 15), 43 corresponden a la variante alfa (6.19 %) (figura 16), 183 corresponden a la variante delta (27.64 %) (figura 17), 51 a la variante gamma (7.7 %) (figura 18) y 296 a la variante ómicron y sus sublinajes (44.71 %) (figura 19). Las restantes 46 secuencias corresponden a las variantes de interés (Épsilon, Zeta, Eta, Theta, Iota, Kappa, Lambda, Mu) y a secuencias anteriores al surgimiento de las variantes del SARS-CoV-2 (7.57 %).

La variante B.1.1.519 comenzó a circular en el estado alrededor del 28 de mayo de 2020 y el 10 de diciembre de 2020. En la figura 15 se observa un *close up* del árbol filogenómico enfocado exclusivamente a las secuencias correspondientes a la variante B.1.1.519, recolectadas entre diciembre de 2020 y junio de 2021. A diferencia de las demás variantes, B.1.1.519 no se puede identificar a simple vista dentro del árbol filogenómico, por lo que se filtra la información del mismo en base al clado de Nextstrain (20B) al que pertenece de la variante y a la mutación de la proteína S que caracteriza a dicha variante (S:478K). No se identificaron cercanías genómicas con secuencias de otras regiones fuera del Estado de Hidalgo.

## Hidalgo-specific genomic surveillance build

Data updated 2024-04-26.

Showing 43 of 662 genomes sampled between Dec 2020 and Jun 2021. Filtered to 20B (64) S:478K

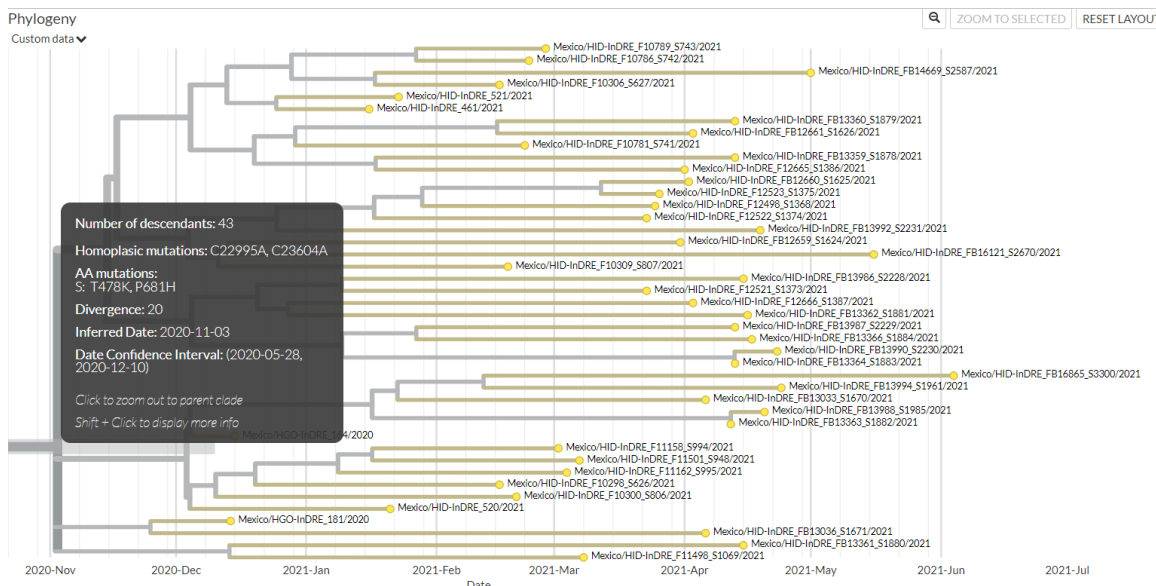


Figura 15. Vista de las 43 secuencias correspondientes a la variante B1.1519 (visibles a través del clado 20B y la mutación S:478K) y de las posibles fechas en las que comenzó su circulación en el Estado de Hidalgo.

La variante alfa comenzó a circular en el estado de Hidalgo alrededor del 7 de diciembre de 2020 y el 21 de febrero de 2021, en la figura 16 se observa un *close up* del árbol filogenómico enfocado exclusivamente a las secuencias correspondientes a la variante alfa, recolectadas entre abril y julio de 2021. Se observa una cercanía genómica entre secuencias recolectadas en mayo de 2021 en el estado con secuencias recolectadas en Camboya en marzo del mismo año.

## Hidalgo-specific genomic surveillance build

Data updated 2024-04-13.

Showing 43 of 662 genomes sampled between Mar 2021 and Jul 2021.

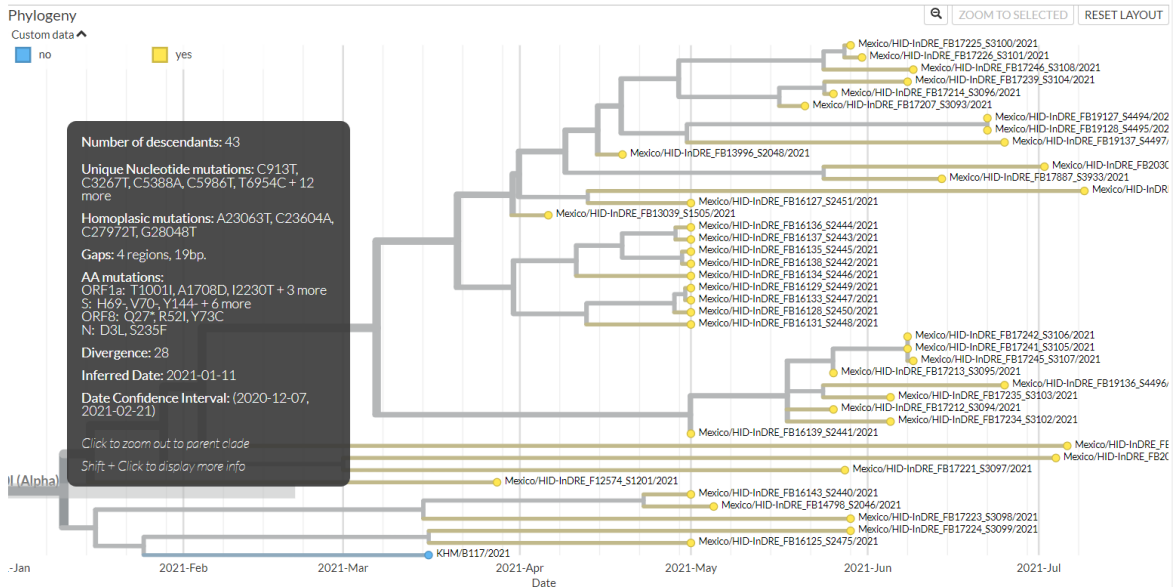


Figura 16. Vista de las 43 secuencias correspondientes a la variante alfa y de las posibles fechas en las que comenzó su circulación en el Estado de Hidalgo.

La variante delta comenzó a circular en el estado de Hidalgo alrededor del 14 de julio de 2020 y el 13 de marzo de 2021, en la figura 17 se observa un *close up* del árbol filogenómico enfocado exclusivamente a las secuencias correspondientes a la variante delta, recolectadas entre junio de 2021 y enero de 2022. Existe cercanía genómica entre una secuencia del Estado recolectada en septiembre de 2021 y una secuencia reportada en Canadá en diciembre del mismo año.



## Hidalgo-specific genomic surveillance build

Data updated 2024-04-13.

Showing 183 of 662 genomes sampled between Jun 2021 and Jan 2022.

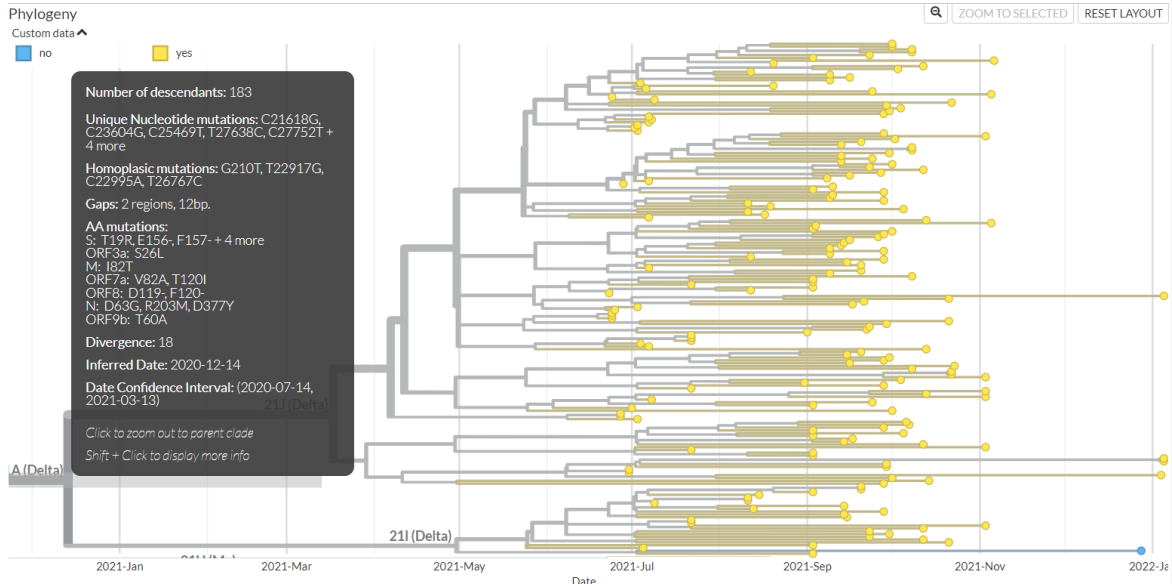


Figura 17. Vista de las 183 secuencias correspondientes a la variante delta y de las posibles fechas en las que comenzó su circulación en el estado.

La variante gamma empezó a circular en la región entre el 8 de agosto de 2020 y el 30 de marzo de 2021, la figura 18 muestra un *close up* del árbol filogenómico para enfocarse únicamente a todas las secuencias correspondientes a la variante gamma, recolectadas entre mayo y julio de 2021. No se identificaron cercanías genómicas con secuencias de otras regiones fuera del Estado de Hidalgo.

## Hidalgo-specific genomic surveillance build

Data updated 2024-04-13.

Showing 51 of 662 genomes sampled between May 2021 and Jul 2021.

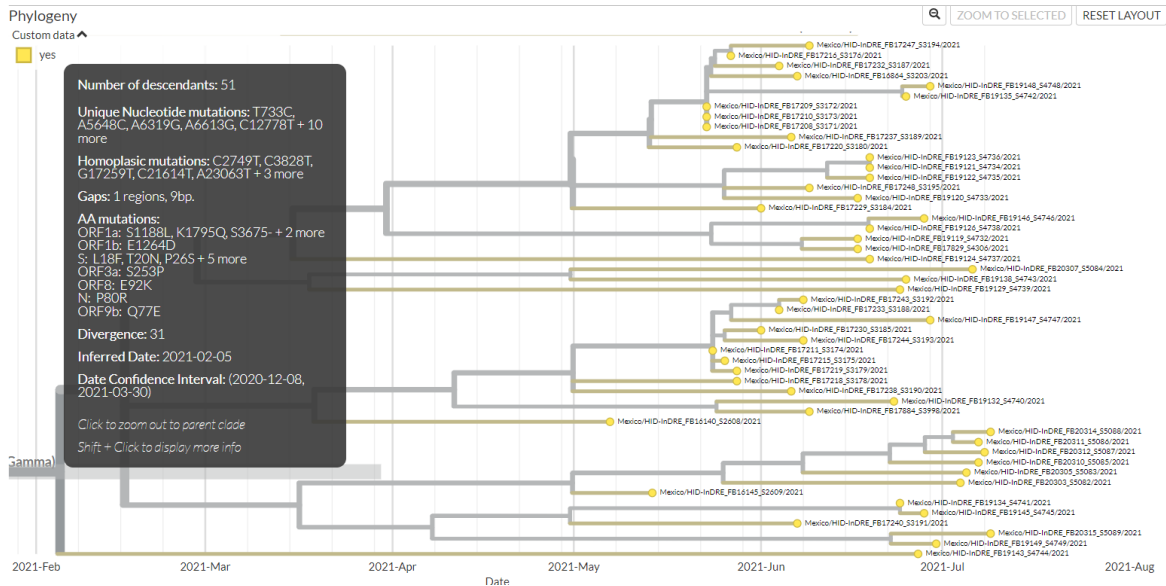


Figura 18. Vista de las 51 secuencias correspondientes a la variable gamma y de las posibles fechas en las que comenzó su circulación en el Estado.

De acuerdo a la figura 19, la variante ómicron comenzó a circular en el estado de Hidalgo alrededor del 29 de octubre de 2020 y el 15 de septiembre de 2021, las secuencias fueron recolectadas entre diciembre de 2021 y agosto de 2023. Se observan cercanías genómicas en múltiples secuencias del estado reportadas desde junio de 2022 y correspondientes al sublinaje BQ.1 con diversas secuencias correspondientes a estados en los Estados Unidos entre julio de 2022 y enero de 2023. De igual manera, las secuencias correspondientes al sublinaje XBB del estado de Hidalgo muestran cercanías genómicas con secuencias de diversos lugares como el Perú, Estados Unidos, Vietnam y Japón, desde julio de 2023 hasta principios de 2024.

## Hidalgo-specific genomic surveillance build

Data updated 2024-04-13.

Showing 296 of 662 genomes sampled between Dec 2021 and Feb 2024.

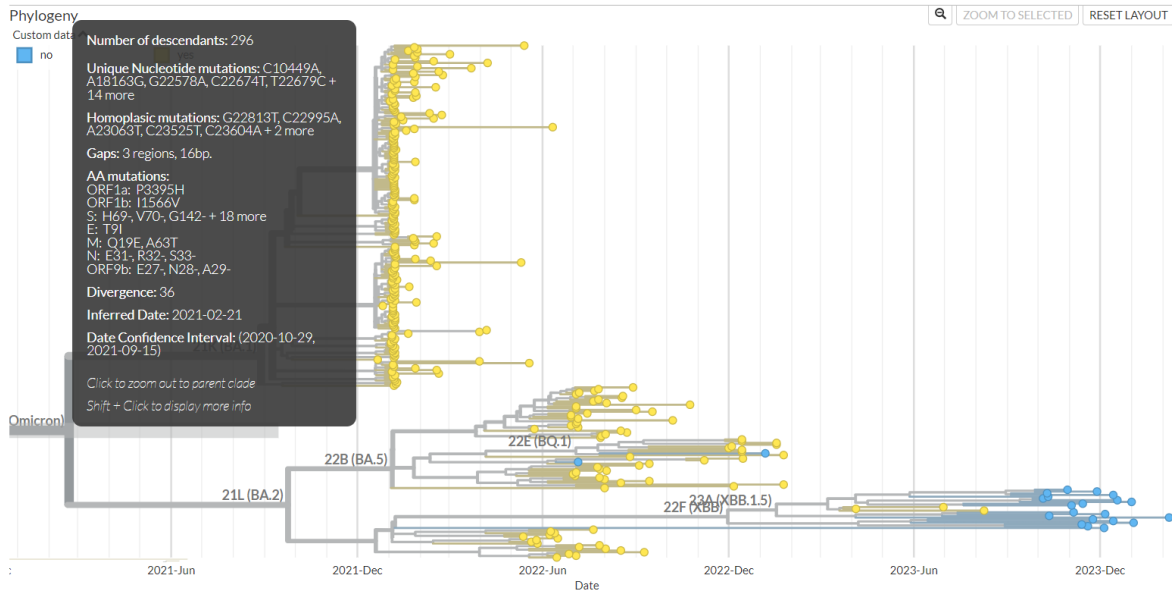


Figura 19. Vista de las 296 secuencias correspondientes a la variante ómicron y de las posibles fechas en las que comenzó su circulación en el Estado.

A diferencia de lo que se muestra en la figura 12, para la construcción del árbol filogenómico se utilizó una nueva base de metadatos, la cual contiene únicamente los datos mínimos necesarios para la correcta construcción del árbol, eliminando de esta manera cualquier dato sensible. Estos son: nombre de la muestra, virus al cual pertenece, clave de identificación en GISAID, clave de identificación en GenBank (no aplica), fecha de recolección y región. En la figura 20 se pueden observar los primeros registros de esta nueva base de metadatos.

Strain	Virus	gi	aid	epi_isl	genbank_accession	date	region			
Wuhan-Hu-1/2019	SARS-CoV-2	EPI_ISL_402125	?			2019-12-26	Asia			
Mexico/HID_InDRE_FB17244_S1939/2021	SARS-CoV-2	EPI_ISL_26746680	?			2021-06-08	Mexico			
Mexico/HID_InDRE_FB905_E13113596050_S10715/2022	SARS-CoV-2	EPI_ISL_9431000	?			2022-01-05	Mexico			
Mexico/HID_InDRE_FB17239_S10104/2021	SARS-CoV-2	EPI_ISL_2663371	?			2021-06-08	Mexico			
Mexico/HID_InDRE_FB43912_S7931/2021	SARS-CoV-2	EPI_ISL_5428294	?			2021-09-24	Mexico			
Mexico/HID_InDRE_FB37623_S6984/2021	SARS-CoV-2	EPI_ISL_4602942	?			2021-09-03	Mexico			
Mexico/HID_InDRE_S18/2021	SARS-CoV-2	EPI_ISL_1168588	?			2021-01-19	Mexico			
Mexico/HID_InDRE_FB21210_E13319002707_S14917/2022	SARS-CoV-2	EPI_ISL_15255730	?			2022-08-17	Mexico			
Mexico/HID_InDRE_FB928_E1331359396_S10737/2022	SARS-CoV-2	EPI_ISL_9431022	?			2022-01-05	Mexico			
Mexico/HID_InDRE_20170208	SARS-CoV-2	EPI_ISL_1060731	?			2020-08-20	Mexico			
Mexico/HID_InDRE_IBT-65/2020	SARS-CoV-2	EPI_ISL_1301569	?			2020-06-02	Mexico			
Mexico/HID_InDRE_FB9220_E13314831142_S11271/2022	SARS-CoV-2	EPI_ISL_10103924	?			2022-01-26	Mexico			
Mexico/HID_InDRE_FB9211_E13314857467_S11262/2022	SARS-CoV-2	EPI_ISL_10103937	?			2022-01-26	Mexico			
Mexico/HID_InDRE_FB9205_E13314876657_S11257/2022	SARS-CoV-2	EPI_ISL_10104015	?			2022-01-26	Mexico			
Mexico/HID_InDRE_FB9207_E13114872177_S11258/2022	SARS-CoV-2	EPI_ISL_10104016	?			2022-01-27	Mexico			
Mexico/HID_InDRE_FB9208_E13314884576_S11259/2022	SARS-CoV-2	EPI_ISL_10104017	?			2022-01-27	Mexico			
Mexico/HID_InDRE_FB9209_E13314882181_S11260/2022	SARS-CoV-2	EPI_ISL_10104018	?			2022-01-27	Mexico			
Mexico/HID_InDRE_FB9212_E13314857148_S11263/2022	SARS-CoV-2	EPI_ISL_10104019	?			2022-01-26	Mexico			
Mexico/HID_InDRE_FB9213_E13314931730_S11264/2022	SARS-CoV-2	EPI_ISL_10104020	?			2022-01-27	Mexico			
Mexico/HID_InDRE_FB9214_E13314785798_S11265/2022	SARS-CoV-2	EPI_ISL_10104021	?			2022-01-25	Mexico			
Mexico/HID_InDRE_FB9215_E13114794226_S11266/2022	SARS-CoV-2	EPI_ISL_10104022	?			2022-01-25	Mexico			
Mexico/HID_InDRE_FB9216_E13114777798_S11267/2022	SARS-CoV-2	EPI_ISL_10104023	?			2022-01-25	Mexico			
Mexico/HID_InDRE_FB9217_E13314810133_S11268/2022	SARS-CoV-2	EPI_ISL_10104024	?			2022-01-26	Mexico			
Mexico/HID_InDRE_FB9218_E13314819201_S11269/2022	SARS-CoV-2	EPI_ISL_10104025	?			2022-01-26	Mexico			
Mexico/HID_InDRE_FB9219_E13314839702_S11270/2022	SARS-CoV-2	EPI_ISL_10104026	?			2022-01-26	Mexico			
Mexico/HID_InDRE_FB9221_E13314852261_S11272/2022	SARS-CoV-2	EPI_ISL_10104027	?			2022-01-27	Mexico			
Mexico/HID_InDRE_FB9201_E13314947125_S11623/2022	SARS-CoV-2	EPI_ISL_10104028	?			2022-01-28	Mexico			
Mexico/HID_InDRE_FB9202_E13314939926_S11624/2022	SARS-CoV-2	EPI_ISL_10104029	?			2022-01-29	Mexico			
Mexico/HID_InDRE_FB9203_E13115087326_S11625/2022	SARS-CoV-2	EPI_ISL_10104030	?			2022-01-29	Mexico			
Mexico/HID_InDRE_FB9210_E13314921926_S11261/2022	SARS-CoV-2	EPI_ISL_10104023	?			2022-01-27	Mexico			
Mexico/HID_InDRE_FB6416_E13314044208_S11672/2022	SARS-CoV-2	EPI_ISL_10361126	?			2022-01-13	Mexico			
Mexico/HID_InDRE_FB9241_E13314812270_S11889/2022	SARS-CoV-2	EPI_ISL_10361146	?			2022-01-26	Mexico			
Mexico/HID_InDRE_FB6417_E13314178307_S11673/2022	SARS-CoV-2	EPI_ISL_10361155	?			2022-01-15	Mexico			
Mexico/HID_InDRE_FB6407_E131014561953_S11669/2022	SARS-CoV-2	EPI_ISL_10361192	?			2022-01-21	Mexico			
Mexico/HID_InDRE_FB6413_E13314470645_S11670/2022	SARS-CoV-2	EPI_ISL_10361193	?			2022-01-20	Mexico			

Figura 20. Muestra de la base de metadatos utilizada para la construcción del árbol filogenómico, formato tsv.

Con la información anterior, se pueden visualizar dentro del árbol filogenómico la fecha de recolección de la muestra, su correspondiente semana epidemiológica, el linaje al que pertenece, y los distintos cambios que se presentan en su genoma (figura 21).

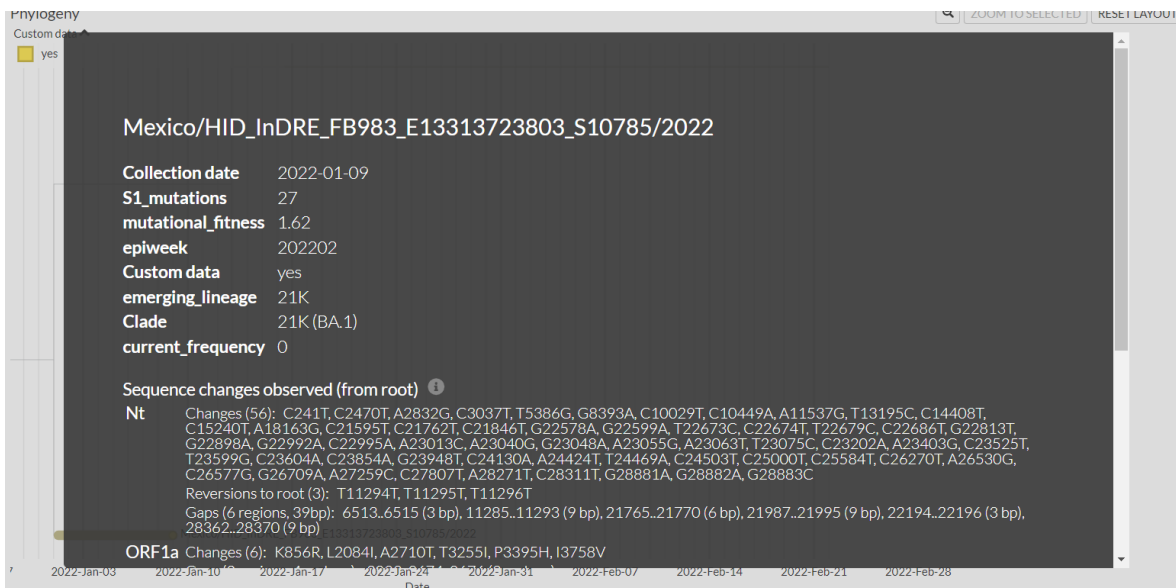


Figura 21. Metadatos observados en el árbol filogenómico construido.

La figura 22A muestra el sexo de los casos confirmados únicamente para la variante delta y la figura 22B para los casos de ómicron y sus sublinajes. Considerando la información revelada en la figura 13. Para la variante Delta ambos sexos se vieron

afectados por igual, en el caso de ómicron, las mujeres fueron más afectadas que los hombres, con un 56.6% del total de los casos confirmados.

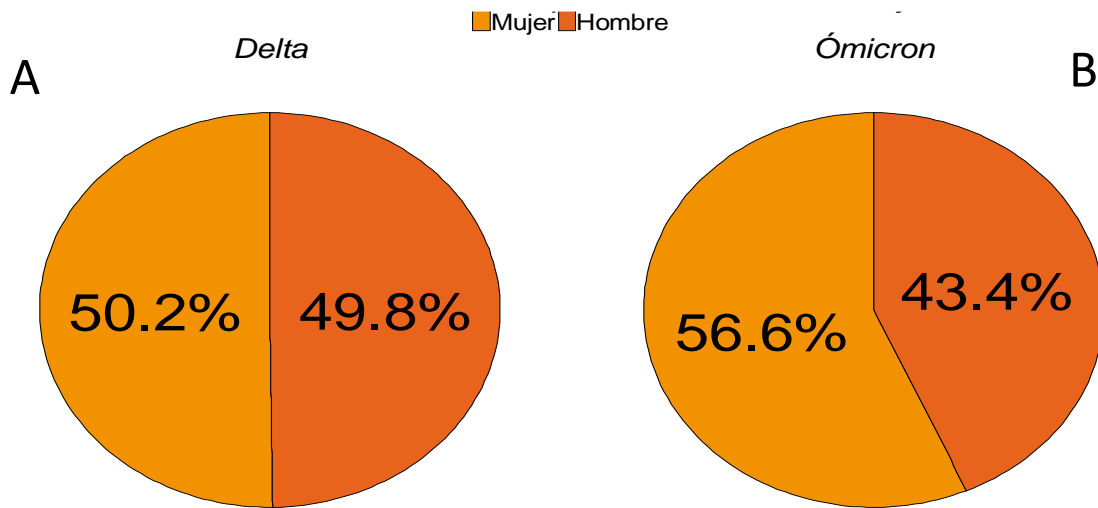


Figura 22. Proporción de los sexos para los casos confirmados correspondientes a las variantes Delta y Ómicron

En la figura 23 se observan los grupos etarios de los casos correspondientes a la variante delta. El grupo más afectado por esta variante del SARS-CoV-2 es el grupo de 18 a 29 años, con el 25% del total de casos, seguido por el grupo de 30-39 años ( $\pm$  20%) y 40 a 49 años ( $\pm$  17.5%).

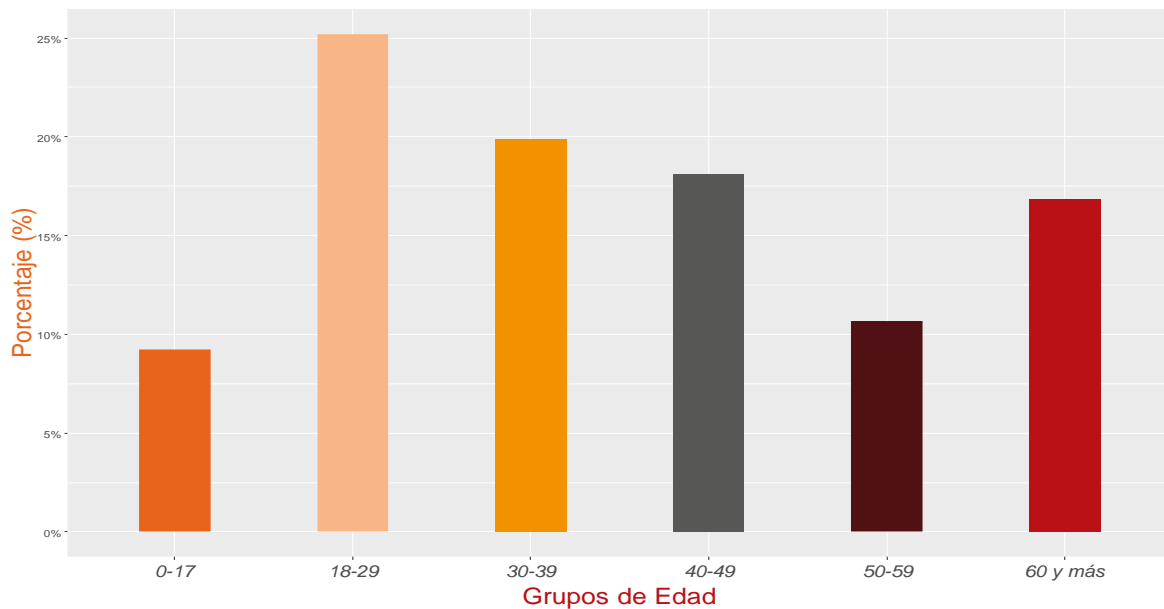


Figura 23. Grupos etarios para la variante delta.

La figura 24 muestra nuevamente los grupos etarios, esta vez asociados con la variante ómicron en el estado de Hidalgo. El grupo de edad más afectado por la variante ómicron es el grupo de 18 a 29 años, con alrededor del 23.5% de los casos, seguido del grupo de 30 a 39 años con el 23.3% y 40 a 49 años con el 20.4 %.

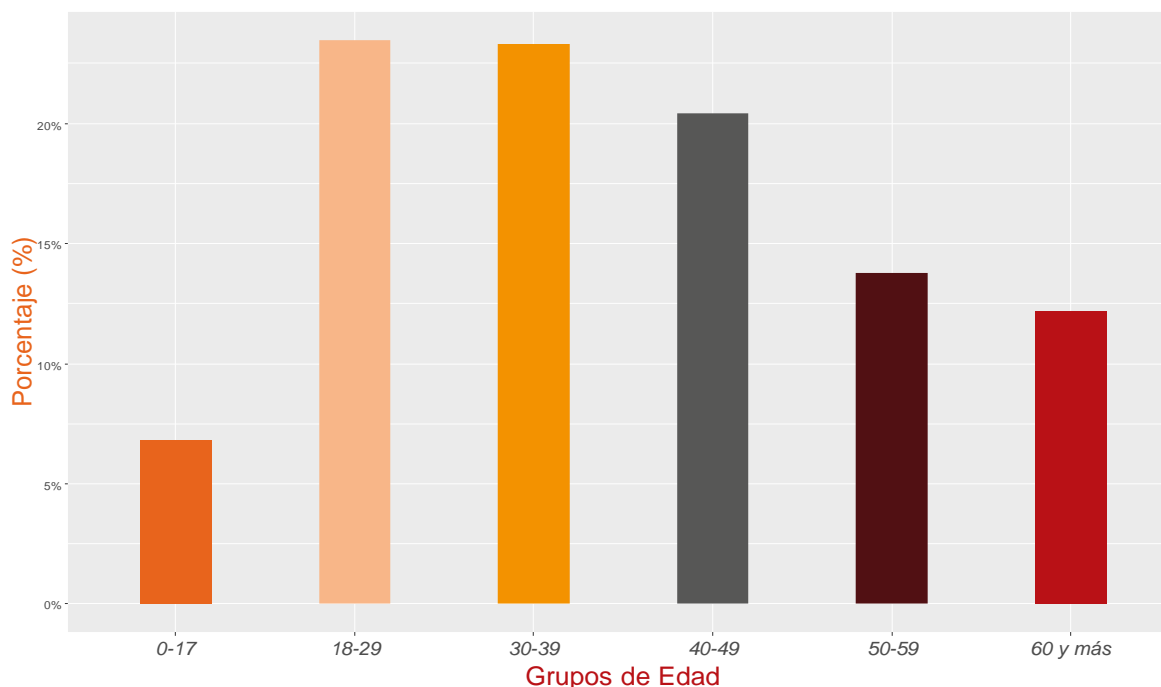


Figura 24. Grupos etarios para la variante ómicron.

En la tabla 5 se observan las regiones operativas del Estado de Hidalgo y los casos correspondientes a la variante delta.

**Tabla 5. Casos y defunciones por región operativa correspondientes a la variante delta.**

Región	Casos (Porcentaje %)	Defunciones (Porcentaje %)
Jacala de Ledezma	69 (0.9%)	11 (1.7%)
Huichapan	348 (4.6%)	27 (4.2%)
Apan	632 (8.4%)	49 (7.6%)
Zacualtipán de Ángeles	409 (5.5%)	34 (5.2%)

<b>Ixmiquilpan</b>	396 (5.3%)	50 (7.7%)
<b>Actopan</b>	457 (6.1%)	54 (8.3%)
<b>Tizayuca</b>	529 (7.1%)	25 (3.9%)
<b>Mineral de la Reforma</b>	763 (10.2%)	46 (7.1%)
<b>Huejutla de Reyes</b>	496 (6.6%)	57 (8.8%)
<b>Pachuca de Soto</b>	1171 (15.6%)	74 (11.4%)
<b>Tulancingo de Bravo</b>	1052 (14%)	123 (19%)
<b>Tula de Allende</b>	1181 (15.7%)	98 (15.1%)
<b>Total</b>	7503 (100%)	648 (100%)

En la tabla 6 se observan las regiones operativas del estado de Hidalgo y los casos correspondientes a la variante ómicron.

**Tabla 6. Casos y defunciones por región operativa correspondientes a la variante ómicron.**

<b>Región</b>	<b>Casos (Porcentaje %)</b>	<b>Defunciones (Porcentaje %)</b>
<b>Jacala de Ledezma</b>	124 (0.3%)	9 (1.7%)
<b>Huichapan</b>	744 (1.6%)	22 (4.3%)
<b>Apan</b>	3,967 (8.6%)	35 (6.8%)
<b>Zacualtipán de Ángeles</b>	741 (1.6%)	17 (3.3%)
<b>Ixmiquilpan</b>	951 (2.1%)	48 (9.3%)
<b>Actopan</b>	1,578 (3.4%)	49 (9.5%)
<b>Tizayuca</b>	4,548 (9.8%)	31 (6%)
<b>Mineral de la Reforma</b>	7,353 (15.9%)	37 (7.2%)
<b>Huejutla de Reyes</b>	741 (1.6%)	24 (4.6%)

<b>Pachuca de Soto</b>	14,902 (32.2%)	96 (18.6%)
<b>Tulancingo de Bravo</b>	6,052 (13.1%)	72 (13.9%)
<b>Tula de Allende</b>	4,594 (9.9%)	77 (14.9%)
<b>Total</b>	46,293 (100%)	517 (100%)

## VIII. Discusión

### **Panorama epidemiológico de la pandemia de COVID-19 en el estado de Hidalgo.**

Los primeros casos confirmados de COVID-19 en el estado de Hidalgo fueron reportados por el InDRE el 19 de marzo de 2020 (49) mientras que el primer caso registrado en las bases de datos de la DGE data del 27 de febrero de 2020, la diferencia de fechas puede ser atribuida al tiempo que tardaron en procesar las muestras de los posibles casos. De igual manera, coinciden con las fechas en las que comenzó el aislamiento preventivo que establecieron los gobiernos a nivel mundial con el fin de parar la transmisión de la enfermedad. (50)

El grupo de edad que registró más casos confirmados fue el rango de 30 a 39 años, en este grupo se encuentran aquellas personas en edad productiva laboralmente y, por ende, aquellas que se vieron en la necesidad constante de viajar fuera de sus hogares, exponiéndose en mayor medida al riesgo de contagio (51). La mediana de edad nacional de los casos confirmados es de 38 años (6), en este trabajo para el caso del estado de Hidalgo se ubicó la mediana general en los 39 años.

Asimismo, la región más afectada fue aquella en donde se ubica la capital del estado de Hidalgo y es la ciudad más poblada, Pachuca de Soto (población de 314,331 habitantes en el 2020) (52). El alto registro de casos confirmados (34,447 infectados) se puede atribuir a la alta densidad poblacional de la región, a la gran actividad económica que conlleva el ser la capital del estado y al alto flujo de personas hacia la región y fuera de la misma. Situación que se asemeja con la Ciudad de México, (capital del país) y su número extremadamente alto de casos confirmados (6,53). Esto se puede observar indistintamente para toda región en la que se encuentre la capital, regional o nacional, en cualquier parte del mundo.



Con respecto a los principales síntomas que se observaron en el estudio fueron predominantes en la población de Hidalgo tos, cefalea, odinofagia, fiebre, mialgias, artralgias, rinorrea y calosfríos, mientras que aquellos reportados por distintas literaturas y definiciones operacionales incluyen fatigas, náuseas, vómitos, diarreas, tos, fiebre, disgeusia y anosmia (14). Este cambio puede atribuirse a la susceptibilidad de la población hidalguense en la que se encuentran factores geográficos, alimenticios y genéticos únicos de la región. Así como también al cambio en el comportamiento infeccioso del virus con el pasar del tiempo (aparición e introducción de variantes del mismo).

Los factores de riesgo asociados a la infección y a la mortalidad de la enfermedad incluyen enfermedades crónicas (comorbilidades) como pueden ser diabetes, hipertensión, cáncer, obesidad y muchas otras. La prevalencia de estas enfermedades en el país se ha reportado por la ENSANUT (Encuesta Nacional de Salud y Nutrición) de 2022. La prevalencia de obesidad y sobrepeso nacional es del 75.2%, la hipertensión es del 47.8%, la diabetes es del 18.3%, el consumo de tabaco es del 19.5%, daño renal es del 22.8%, el hipercolesterolemia es del 30.6% y de enfermedades cardiovasculares es del 4.9% (54–57). Los casos confirmados en el estado de Hidalgo presentan hipertensión, neumonía, obesidad, diabetes y tabaquismo como principales comorbilidades.

El periodo en el que se observa un incremento sostenido en el conteo de casos confirmados corresponde a aquellos periodos vacacionales de mediados y finales de año, sobre todo durante 2021 y 2022 cuando las medidas de aislamiento social comenzaron a ser retiradas (58,59). Estos periodos se caracterizan por el flujo constante de personas a través y fuera del estado, impulsados por el deseo de convivencia con familia y/o amigos, sin las medidas de cuidado necesarias. El mismo panorama sucede a nivel nacional.

Las defunciones en el estado de Hidalgo por COVID-19 (8,520 hasta 2023) alcanzan el pico máximo a finales de 2020, coincidiendo con las festividades de navidad y año nuevo y con uno de los picos de realce de casos confirmados en el estado debido al flujo de personas que viajaban a través del estado. El número de defunciones tendió a disminuir a partir de estas fechas. La mortalidad registrada en

el estado de Hidalgo fue de 6.55% en el periodo de estudio. Este valor fue superior al reportado en México por la Secretaría de Salud (334,472 defunciones, mortalidad del 4.31%) y global (6.9 millones de defunciones, mortalidad del 0.91%) (14). El alto porcentaje de mortalidad en el estado de Hidalgo podría estar relacionado con la presencia de factores de riesgo en la población, el uso de distintos tratamientos y la saturación de los servicios de salud en el territorio nacional. (60)

La disminución de las defunciones asociadas a la COVID-19 puede atribuirse al comienzo de la campaña de vacunación en el estado, a finales de 2020 (44,45). Aunque no se logró detener la red de transmisión, la vacunación en la población logró la prevención de las complicaciones en la enfermedad y, por ende, la disminución en el número de defunciones a partir del comienzo de la campaña, lo cual se ve reflejado en la disminución de la mortalidad en la figura 10. La persistencia en el contagio de la enfermedad se puede atribuir directamente al surgimiento de las variantes genómicas del SARS-CoV-2 (10).

### **Vigilancia genómica y su impacto dentro de la epidemiología.**

Aunque el primer caso confirmado de COVID-19 en el estado de Hidalgo se dio a conocer en marzo de 2020 (49), las primeras secuencias recolectadas del SARS-CoV-2 por el InDRE se dieron hasta el 2 de junio de 2020. La información registrada en los metadatos no contempla el municipio de recolección de la misma, por lo que es imposible identificar la zona de origen de la muestra. Lo anterior indica una falta de vigilancia genómica en las etapas tempranas de la emergencia sanitaria en el estado de Hidalgo y, a su vez, una vigilancia genómica deficiente e incompleta.

La tasa de mutación del virus hace que se presente algún cambio en su genoma cada 2 semanas, lo que permitió que la variante de interés B.1.1.519 apareciera por primera vez en México y se reportara en Hidalgo a finales de 2020 (SE 50-2020) (61), coincidiendo con el pico máximo de defunciones reportadas en el estado en las mismas fechas. También permitió que la primera variante de preocupación (alfa) apareciera en septiembre de 2020 en el Reino Unido (25,26) y posteriormente fuera identificada en el estado de Hidalgo aproximadamente a finales abril de 2021 (SE 15-2021).

La información que se obtiene del árbol filogenómico generado en este estudio permite observar el comportamiento de transmisión del virus en la población del estado de Hidalgo. Las similitudes de distintas muestras genómicas del estado de Hidalgo y muestras reportadas en lugares como Estados Unidos, Dinamarca, Camboya, Perú e incluso Japón (identificadas por la longitud de los brazos que separan dichas muestras) indican diferentes eventos de introducción en las distintas zonas en los mismos periodos de tiempo (29), lo que a su vez indica un desplazamiento directo o indirecto hacia o desde estas regiones con el estado de Hidalgo, sobre todo en periodos vacacionales. (62)

La ausencia de la variante de preocupación beta (Sudáfrica, octubre 2020), puede ser atribuida a su incapacidad de infección en la población hidalguense, ya sea por cuestiones genéticas o por un contacto cero con aquellas personas con antecedentes de viajes hacia o desde zonas donde sí fue claramente identificada.

Mediante la incorporación de la información genética en el estudio epidemiológico, es posible observar el comportamiento característico de las distintas variantes del virus SARS-CoV-2 en la población del estado de Hidalgo (29). Como se mostró en la descripción epidemiológica, se detalló la emergencia sanitaria en el estado de Hidalgo, enfocándonos en las variables mencionadas en la sección 5.2 pero, de manera general, sin considerar los cambios que pudieran presentar las variantes del SARS-CoV-2 en la población. En la figura 13 se observa la identificación de las variantes alfa, gamma, delta y ómicron en el estado. Únicamente se puede realizar la interpretación para las variantes delta y ómicron, debido a que son las únicas variantes que se reportaron en un periodo de tiempo determinado. De acuerdo a lo observado en la figura 13, las semanas del 35 al 50 de 2021 corresponden a la variante delta y de la semana 5 de 2022 en adelante a la variante ómicron y sus sublinajes.

El impacto de la combinación de la información propia de una investigación epidemiológica tradicional y de la información obtenida del análisis del agente patógeno causal (emergente o no), va desde el creciente conocimiento de los agentes, las tendencias y factores que repercuten en la dispersión de cualquier brote, hasta la toma de decisiones tanto en materia de salud como en política para

su control. La disciplina, ahora conocida como epidemiología genómica, tendrá grandes papeles en el futuro y seguirá desarrollándose de la mano de los avances tecnológicos vistos en técnicas de diagnóstico, secuenciación de siguiente generación, análisis de datos y toda aquella que facilite el seguimiento y control de enfermedades emergentes e incluso reemergentes en todo el mundo. La colaboración entre distintas disciplinas, administrativos de salud y gobiernos de todo el mundo es necesaria para la correcta práctica de la epidemiología genómica, por lo que promover la colaboración de todos es indispensable.

## **IX. Conclusiones**

La mortalidad por COVID-19 en el Estado de Hidalgo fue superior a la mortalidad nacional y global. La positividad en las mujeres fue superior a la de los hombres, pero la mortalidad fue mayor en hombres. Los adultos jóvenes se vieron más afectados. La sintomatología principal reportada por los casos en el estado difiere de las principales reportadas por la literatura.

Bajo nivel de comorbilidades en comparación a lo reportado a nivel nacional (ENSANUT). Las pruebas de chi cuadrada indicaron independencia entre las variables de sexo contra resultado (positivo/negativo a COVID-19) y región operativa contra resultado. En general, el Estado de Hidalgo presentó un comportamiento epidemiológico similar al presentado a nivel nacional. La vigilancia genómica en el Estado de Hidalgo fue pobre e insuficiente durante la emergencia sanitaria.

El presente estudio generó conocimiento al evidenciar las variantes B.1.1.519 (semana 50 de 2020), alfa (semana 15 de 2021), delta (semana 25 de 2021), gamma (semana 20 de 2021) y ómicron (semana 50 de 2021) circulando en el estado de Hidalgo. También se muestra que las fechas de aparición de la variante B.1.1.519, coincidieron con el pico máximo de defunciones en el estado.

Existieron diversos eventos de propagación con las distintas variantes que circularon en el Estado de Hidalgo durante la emergencia sanitaria.

## **X. Recomendaciones**

- Este estudio consigue mostrar la valía de estudios retrospectivos.
- Se busca la mejora de los sistemas de vigilancia en el nivel estatal mediante la incorporación de las tecnologías emergentes (secuenciación de siguiente generación, análisis bioinformático, inteligencia artificial, big data, cloud computing, etc.)

## XI. Referencias

1. Tanya Lewis. What New Evidence from the Wuhan Market Tells Us about COVID's Origins | Scientific American [Internet]. 2023 [cited 2024 Feb 14]. Available from: <https://www.scientificamerican.com/article/what-new-evidence-from-the-wuhan-market-tells-us-about-covids-origins1/>
2. Kirtipal N, Bharadwaj S, Kang SG. From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses. *Infect. Genet. Evol.* 2020;85:104502.
3. OMS. WHO EMRO | MERS outbreaks | MERS-CoV | Health topics [Internet]. 2023 [cited 2024 Feb 14]. Available from: <https://www.emro.who.int/health-topics/mers-cov/mers-outbreaks.html>
4. OMS. Weekly operational update on COVID-19 - 16 March 2021 [Internet]. 2021 [cited 2022 Aug 29]. Available from: <https://www.who.int/publications/m/item/weekly-operational-update-on-covid-19---16-march-2021>
5. OMS. Situation Report-51. 2020 [cited 2022 Aug 29]; Available from: [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57\\_10](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10)
6. SALUD. Coronavirus COVID19 Informe Técnico Semanal | Secretaría de Salud | Gobierno | gob.mx [Internet]. 2023 [cited 2024 Feb 15]. Available from: <https://www.gob.mx/salud/documentos/coronavirus-covid19-informe-tecnico-semanal>
7. Bakhiet M, Taurin S. SARS-CoV-2: Targeted managements and vaccine development. *Cytokine. Growth. Factor. Rev.* 2021;58:16–29.
8. Deshpande R, Li W, Li T, Fanning K V., Clemens Z, Nyunoya T, et al. SARS-CoV-2 Accessory Protein Orf7b Induces Lung Injury via c-Myc Mediated Apoptosis and Ferroptosis. *Int. J. Mol. Sci.* 2024;25(2).
9. Peck KM, Lauring AS. Complexities of Viral Mutation Rates. *J. Virol.* 2018;92(14):e01031-17.
10. OMS. Updated working definitions and primary actions for SARSCoV2 variants [Internet]. 2023 [cited 2024 Feb 14]. Available from: <https://www.who.int/publications/m/item/updated-working-definitions-and-primary-actions-for--sars-cov-2-variants>
11. Leroy H, Han M, Woottum M, Bracq L, Bouchet J, Xie M, et al. Virus-Mediated Cell-Cell Fusion. *Int. J. Mol. Sci.* 2020;21(24):1–28.
12. Koonin E V., Krupovic M, Agol VI. The Baltimore Classification of Viruses 50 Years Later: How Does It Stand in the Light of Virus Evolution?. *Microbiol. Mol. Biol Rev.* 2021;85(3):e0005321.
13. Díaz-Quiñonez JA. Emergencia del coronavirus SARS-CoV2 en China y la respuesta en México Emergence of novel coronavirus SARS-CoV2 in China and the response in Mexico. *Gac. Med. Mex.* 2020;156(2):91–3.
14. OMS. Vigilancia de salud pública en relación con la COVID-19: orientaciones provisionales, 14 de febrero de 2022 [Internet]. 2022 [cited 2022 Sep 1]. Available from: <https://apps.who.int/iris/handle/10665/352396>
15. Safiabadi Tali SH, LeBlanc JJ, Sadiq Z, Oyewunmi OD, Camargo C, Nikpour B, et al. Tools and Techniques for Severe Acute Respiratory Syndrome

- Coronavirus 2 (SARS-CoV-2)/COVID-19 Detection. *Clin. Microbiol. Rev.* 2021;34(3):e00228-20.
16. Kang Y, Lu S, Zhong R, You J, Chen J, Li L, et al. The immune inflammation factors associated with disease severity and poor prognosis in patients with COVID-19: A retrospective cohort study. *Heliyon.* 2023;10(1):e23583.
  17. Majumder J, Minko T. Recent Developments on Therapeutic and Diagnostic Approaches for COVID-19. *AAPS. J.* 2021;23(1):14.
  18. Rai P, Kumar BK, Deekshit VK, Karunasagar I, Karunasagar I. Detection technologies and recent developments in the diagnosis of COVID-19 infection. *Appl. Microbiol. Biotechnol.* 2021;105(2):441–55.
  19. Botero-Rodríguez F, Franco O, Carlos. Glosario para una pandemia: el ABC de los conceptos sobre el coronavirus. *Biomédica.* 2020;40(Suppl 2):16–26.
  20. Wilder-Smith A, Osman S. Public health emergencies of international concern: a historic overview. *J. Travel. Med.* 2020;27(8):taaa227.
  21. OMS. WHO Director-General declares the ongoing monkeypox outbreak a Public Health Emergency of International Concern [Internet]. 2022 [cited 2024 Feb 15]. Available from: <https://web.archive.org/web/20220725054644/https://www.who.int/europe/news/item/23-07-2022-who-director-general-declares-the-ongoing-monkeypox-outbreak-a-public-health-event-of-international-concern>
  22. OMS. Influenza (Avian and other zoonotic) [Internet]. 2018 [cited 2022 Sep 1]. Available from: [https://www.who.int/news-room/fact-sheets/detail/influenza-\(avian-and-other-zoonotic\)?gclid=Cj0KCQjw08aYBhDIARIsAA\\_gb0fxftsEupPJ2PV2EedkIZ-DfBgyWShWqwBY\\_EwZjAwK\\_4Wi30FlnSUaAhX1EALw\\_wcB](https://www.who.int/news-room/fact-sheets/detail/influenza-(avian-and-other-zoonotic)?gclid=Cj0KCQjw08aYBhDIARIsAA_gb0fxftsEupPJ2PV2EedkIZ-DfBgyWShWqwBY_EwZjAwK_4Wi30FlnSUaAhX1EALw_wcB)
  23. Tümmler B. Molecular epidemiology in current times. *Environ. Microbiol.* 2020;22(12):4909–18.
  24. Hägg S, Belsky DW, Cohen AA. Developments in molecular epidemiology of aging. *Emerg. Top. Life. Sci.* 2019;3(4):411–21.
  25. Mistry P, Barmania F, Mellet J, Peta K, Strydom A, Viljoen IM, et al. SARS-CoV-2 Variants, Vaccines, and Host Immunity. *Front. Immunol.* 2022;12:809244.
  26. Majumdar P, Niyogi S. SARS-CoV-2 mutations: the biological trackway towards viral fitness. *Epidemiol. Infect.* 2021;149:e110.
  27. OPS. Actualización sobre la evolución y circulación de sublinajes del SARS-CoV-2 - OPS/OMS | Organización Panamericana de la Salud [Internet]. 2024 [cited 2024 Feb 15]. Available from: <https://www.paho.org/es/documentos/actualizacion-sobre-evolucion-circulacion-sublinajes-sars-cov-2>
  28. OMS. Seguimiento de las variantes del SARS-CoV-2 [Internet]. 2021 [cited 2024 Aug 29]. Available from: <https://www.who.int/es/activities/tracking-SARS-CoV-2-variants/>
  29. Allison Black, Gytis Dudas. An applied genomic epidemiological handbook. [Internet]. 2023 [cited 2024 Apr 10]. Available from: <https://alliblk.github.io/genepi-book/index.html>
  30. Wang X, Chen Y, Martcheva M, Rong L. Asymptotic analysis of a vector-borne disease model with the age of infection. *J. Biol. Dyn.* 2020;14(1):332–67.

31. Elsohaby I, Villa L. Zoonotic diseases: understanding the risks and mitigating the threats. *BMC. Vet. Res.* 2023 Dec 1;19(1).
32. Abu Toamih Atamni H, Nashef A, Iraqi FA. The Collaborative Cross mouse model for dissecting genetic susceptibility to infectious diseases. *Mamm. Genome.* 2018;29(7–8):471–87.
33. Pillaiyar T, Meenakshisundaram S, Manickam M. Recent discovery and development of inhibitors targeting coronaviruses. *Drug. Discov. Today.* 2020;25(4):668–88.
34. Ye ZW, Yuan S, Yuen KS, Fung SY, Chan CP, Jin DY. Zoonotic origins of human coronaviruses. *Int. J. Biol. Sci.* 2020;16(10):1686–97.
35. Saib I, Aleisa S, Ardah H, Mahmoud E, Alharbi AO, Alsaedy A, et al. Non-SARS Non-MERS Human Coronaviruses: Clinical Characteristics and Outcome. *Pathogens.* 2021;10(12):1549.
36. Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. The spike protein of SARS-CoV--a target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* 2009;7(3):226–36.
37. De Wit E, Van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* 2016;14(8):523–34.
38. Leao JC, Gusmao TP de L, Zarzar AM, Leao Filho JC, Barkokebas Santos de Faria A, Morais Silva IH, et al. Coronaviridae-Old friends, new enemy! *Oral. Dis.* 2022;28 Suppl 1(Suppl 1):858–66.
39. Boldog P, Tekeli T, Vizi Z, Dénes A, Bartha FA, Röst G. Risk Assessment of Novel Coronavirus COVID-19 Outbreaks Outside China. *J. Clin. Med.* 2020;9(2):571.
40. Jee Y. WHO International Health Regulations Emergency Committee for the COVID-19 outbreak. *Epidemiol. Health.* 2020;42:e2020013.
41. SALUD. Aviso Epidemiológico Enfermedad COVID-19 por SARS-CoV-2 (actualización) [Internet]. 2020 [cited 2022 Aug 31]. Available from: <https://www.gob.mx/salud/documentos/aviso-epidemiologico-enfermedad-covid-19-por-sars-cov-2-actualizacion>
42. OMS. La OMS publica su primera validación para uso en emergencias de una vacuna contra la COVID-19 y hace hincapié en la necesidad de un acceso mundial equitativo [Internet]. 2020 [cited 2022 Aug 31]. Available from: <https://www.who.int/es/news/item/31-12-2020-who-issues-its-first-emergency-use-validation-for-a-covid-19-vaccine-and-emphasizes-need-for-equitable-global-access>
43. Vitiello A, Ferrara F, Troiano V, La Porta R. COVID-19 vaccines and decreased transmission of SARS-CoV-2. *Inflammopharmacology.* 2021;29(5):1357–60.
44. SALUD. 266. Arranca vacunación contra COVID-19 en México [Internet]. 2020 [cited 2022 Aug 31]. Available from: <https://www.gob.mx/salud/prensa/266-arranca-vacunacion-contra-covid-19-en-mexico?idiom=es>
45. SALUD. Informe Técnico Diario COVID-19 MÉXICO 24/12/2020 [Internet]. 2020 [cited 2022 Aug 31]. Available from: [https://www.gob.mx/cms/uploads/attachment/file/602346/Comunicado\\_Tecnico\\_Diario\\_COVID-19\\_2020.12.24.pdf](https://www.gob.mx/cms/uploads/attachment/file/602346/Comunicado_Tecnico_Diario_COVID-19_2020.12.24.pdf)



46. SALUD. Informe Técnico Semanal COVID-19 MÉXICO 02/05/2023 [Internet]. 2023 [cited 2024 Feb 14]. Available from: [https://www.gob.mx/cms/uploads/attachment/file/822055/Informe\\_Tecnico\\_Semanal\\_COVID-19\\_2023.05.02.pdf](https://www.gob.mx/cms/uploads/attachment/file/822055/Informe_Tecnico_Semanal_COVID-19_2023.05.02.pdf)
47. OMS. Declaración acerca de la decimoquinta reunión del Comité de Emergencias del Reglamento Sanitario Internacional (2005) sobre la pandemia de enfermedad por coronavirus (COVID-19) [Internet]. 2023 [cited 2024 Feb 14]. Available from: [https://www.who.int/es/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-coronavirus-disease-\(covid-19\)-pandemic](https://www.who.int/es/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic)
48. Dirección General de Epidemiología. Datos Abiertos Dirección General de Epidemiología | Secretaría de Salud | Gobierno | gob.mx [Internet]. 2023 [cited 2024 Aug 29]. Available from: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>
49. Grupo Milenio. Coronavirus en Hidalgo. Se registran primeros dos casos de Covid-19 - Grupo Milenio [Internet]. 2020 [cited 2024 Apr 4]. Available from: <https://www.milenio.com/ciencia-y-salud/coronavirus-en-hidalgo-se-registran-primeros-dos-casos-de-covid-19>
50. Suárez V, Suarez Quezada M, Oros Ruiz S, Ronquillo De Jesús E. Epidemiología de COVID-19 en México: del 27 de febrero al 30 de abril de 2020. *Rev. Clin. Esp.* 2020;220(8):463.
51. Mora Salas M, Urbina Cortés G, Mora Salas M, Urbina Cortés G. Las juventudes populares mexicanas frente a la Covid-19: estigmas, apremios y prácticas de prevención. *Ultima. década.* 2021;29(56):104–48.
52. INEGI. EN HIDALGO SOMOS 3 082 841 HABITANTES: CENSO DE POBLACIÓN Y VIVIENDA 2020 [Internet]. 2021 [cited 2024 Apr 4]. Available from: [https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodem/ResultCenso2020\\_Hgo.pdf](https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodem/ResultCenso2020_Hgo.pdf)
53. Klein B, Hartle H, Shrestha M, Zenteno AC, Barros Sierra Cordera D, Nicolás-Carlock JR, et al. Spatial scales of COVID-19 transmission in Mexico. *PNAS. Nexus.* 2024;1–13.
54. Campos-Nonato I, Galván-Valencia O, Hernández-Barrera L, Oviedo-Solís C, Barquera S. Prevalencia de obesidad y factores de riesgo asociados en adultos mexicanos: resultados de la Ensanut 2022. *Salud. Publica. Mex.* 2023;65:s238–47.
55. Basto-Abreu A, López-Olmedo N, Rojas-Martínez R, Aguilar-Salinas CA, Moreno-Banda GL, Carnalla M, et al. Prevalencia de prediabetes y diabetes en México: Ensanut 2022. *Salud. Publica. Mex.* 2023;65:s163–8.
56. Escamilla-Núñez MC, Castro-Porras L, Romero-Martínez M, Zárate-Rojas E, Rojas-Martínez R. Detección, diagnóstico previo y tratamiento de enfermedades crónicas no transmisibles en adultos mexicanos. *Ensanut 2022. Salud. Publica. Mex.* 2023;65:s153–62.
57. Barrera-Núñez DA, López-Olmedo N, Zavala-Arciniega L, Barrientos-Gutiérrez I, Reynales-Shigematsu LM. Consumo de tabaco y uso de cigarro

- electrónico en adolescentes y adultos mexicanos. *Ensanut Continua* 2022. *Salud. Publica. Mex.* 2023;65:s65–74.
58. El País. México comienza a relajar las medidas contra la covid-19 | EL PAÍS México [Internet]. 2022 [cited 2024 Apr 10]. Available from: <https://elpais.com/mexico/2022-04-26/mexico-comienza-a-relajar-las-medidas-contra-la-covid-19.html>
  59. SALUD. Comité de Nueva Normalidad actualiza lineamientos sobre COVID-19 para continuidad de actividades económicas | Secretaría de Salud | Gobierno | gob.mx [Internet]. 2022 [cited 2024 Apr 10]. Available from: <https://www.gob.mx/salud/prensa/comite-de-nueva-normalidad-actualiza-lineamientos-sobre-covid-19-para-continuidad-de-actividades-economicas>
  60. Ordoñez-González I, Basurto MA. La atención primaria a la salud durante la pandemia COVID-19 en México. *Rev. Med. Inst. Mex. Seguro. Soc.* 2023;61(4):509.
  61. Rodríguez-Maldonado AP, Vázquez-Pérez JA, Cedro-Tanda A, Taboada B, Boukadida C, Wong-Arámbula C, et al. Emergence and spread of the potential variant of interest (VOI) B.1.1.519 of SARS-CoV-2 predominantly present in Mexico. *Arch. Virol.* 2021;166(11):3173–7.
  62. El Heraldo. Justifican incremento de casos de Covid-19 a movilidad por vacaciones [Internet]. 2023 [cited 2024 May 2]. Available from: <https://elheraldodesaltillo.mx/2023/08/06/justifican-incremento-de-casos-de-covid-19-a-movilidad-por-vacaciones/>

## XII. Anexos

### Formatos

#### Anexo 1. Operacionalización de variables

Tabla 7. Operacionalización de variables				
Variable	Descripción	Categorías	Definición	Tipo de variable
Sexo	El caso es del sexo: Femenino o masculino.	0= Masculino 1= Femenino	Es la característica biológica que permite clasificar a los seres humanos en hombres y mujeres.	Cualitativa nominal dicotómica.
Edad	Indica la edad del caso. Está expresada en años.	De 1 año, 2,3,4,5,6... etc.	Tiempo que ha vivido una persona.	Cuantitativa discreta.
Municipio	Es el municipio de residencia del caso, donde adquirió COVID-19.	Municipio: 001 Acatlán 002 Acaxochitlán 003 Actopan 004 Agua Blanca de Iturbide 005 Ajacuba 006 Alfajayucan 007 Almoloya 008 Apan 009 El Arenal 010 Atitalaquia 011 Atlapexco 012 Atotonilco el Grande 013 Atotonilco de Tula	Lugar en que se reside o se vive.	Cualitativa nominal politómica.

		014 Calnali 015 Cardonal 016 Cuauhtepc de Hinojosa 017 Chapantongo 018 Chapulhuacán 019 Chilcuautila 020 Eloxochitlán 021 Emiliano Zapata 022 Epazoyucan 023 Francisco I. Madero 024 Huasca de Ocampo 025 Huautla 026 Huazalingo 027 Huehuetla 028 Huejutla de Reyes 029 Huichapan029 Huichapan 030 Ixmiquilpan 031 Jacala de Ledezma 032 Jaltocán 033 Juárez Hidalgo 034 Lolotla 035 Metepec 036 San Agustín Metzquititlán		
--	--	--	--	--

		037 Metztlán		
		038 Mineral del Chico		
		039 Mineral del Monte		
		040 La Misión		
		041 Mixquiahuala de Juárez		
		042 Molango de Escamilla		
		043 Nicolás Flores		
		044 Nopala de Villagrán		
		045 Omitlán de Juárez		
		046 San Felipe Orizatlán		
		047 Pacula		
		048 Pachuca de Soto		
		049 Pisaflores		
		050 Progreso de Obregón		
		051 Mineral de la Reforma		
		052 San Agustín Tlaxiaca		
		053 San Bartolo Tutotepec		
		054 San Salvador		
		055 Santiago de Anaya		

		056 Santiago Tulantepec de Lugo Guerrero 057 Singuilucan 058 Tasquillo 059 Tecozautla 060 Tenango de Doria 061 Tepeapulco 062 Tepehuacán de Guerrero 063 Tepeji del Río de Ocampo 064 Tepetitlán 065 Tetepango 066 Villa de Tezontepec 067 Tezontepec de Aldama 068 Tiangustengo 069 Tizayuca 070 Tlahuelilpan 071 Tlahuiltepa 072 Tlanalapa 073 Tlanchinol 074 Tlaxcoapan 075 Tolcayuca 076 Tula de Allende 077 Tulancingo de Bravo 078 Xochiatipan 079 Xochicoatlán		
--	--	--	--	--

		080 Yahualica 081 Zacualtipán de Ángeles 082 Zapotlán de Juárez 083 Zempoala 084 Zimapán		
Sintomatología	Principales síntomas que presentó el caso de COVID-19 (Sintomatología extraída del estudio de caso).	<ul style="list-style-type: none"> <li>• Tos</li> <li>• Fiebre</li> <li>• Dolor de cabeza</li> <li>• Dificultad para respirar (dato de alarma)</li> <li>• Dolor en las articulaciones</li> <li>• Dolor muscular</li> <li>• Dolor de garganta</li> <li>• Escurrimiento nasal</li> <li>• Conjuntivitis (enrojecimiento, picazón y ojos llorosos)</li> <li>• Dolor en el pecho</li> </ul>	Es la referencia subjetiva que da un enfermo de la percepción que reconoce como anómala o causada por un estado patológico o una enfermedad.	Cualitativa nominal politómica.
Comorbilidades	Comorbilidades extraídas del estudio de caso.	<ul style="list-style-type: none"> <li>• Obesidad</li> <li>• Diabetes</li> <li>• Hipertensión</li> <li>• Cáncer</li> <li>• Asma</li> <li>• Enfermedades cardíacas</li> </ul>	Coexistencia de dos o más enfermedades en un mismo individuo, generalmente relacionadas.	Cualitativa nominal dicotómica.

		• Enfermedades Crónicas		
Variante	Cepa del virus detectada en los casos reportados.	1: Alfa ( $\alpha$ ) 2: Beta ( $\beta$ ) 3: Gamma ( $\gamma$ ) 4: Delta ( $\delta$ ) 5: Ómicron ( $\omicron$ )	Cepa del virus que presenta modificaciones genómicas con respecto a la cepa original.	Cualitativa Nominal dicotómica.
Fecha de introducción	Indica la fecha en la que se introdujo una cepa del virus en la población.	Fecha (dd/mm/aaaa)	Fecha en la que se identificó una cepa.	Cuantitativa discreta.
Variante de interés (VOI)	Clasificación dada a la cepa en el caso reportado.	0=No 1=Si	Variantes del SARS-CoV-2 que presentan cambios genómicos que pudieran afectar las características del mismo.	Cualitativa, nominal, dicotómica.
Variante de preocupación (VOC)	Clasificación dada a la cepa en el caso reportado.	0=No 1=Si	Variante que cumple con la definición de VOI y que además se haya demostrado aumento en la transmisibilidad y virulencia.	Cualitativa, nominal, dicotómica.
Técnica de diagnóstico	Método de estudio realizado por laboratorio para confirmar diagnóstico por el InDRE.	1= RT-PCR	Prueba de laboratorio complementaria para confirmar o descartar una enfermedad.	Cualitativa Nominal dicotómica.



## Anexo 2. Uso de Nextstrain

# Run using custom data

This tutorial builds on [the previous tutorial](#). You will learn how to run the workflow with your own genomic data, using the reference data from the previous tutorial as the genetic context for these new data. Although you will download data from GISAID for this tutorial, you can replace these data with your own local sequences and metadata in your future analyses.

### Table of Contents

- [Prerequisites](#)
- [Setup](#)
- [Curate data from GISAID](#)
- [Run the workflow](#)
  - [Break down the command](#)
- [Visualize the results](#)

### Prerequisites

1. [Run using example data](#). This tutorial sets up the command line environment used in the following tutorial.
2. [Register for a GISAID account](#), if you do not have one yet. However, registration may take a few days. Follow [alternative data preparation methods](#) in place of [Curate data from GISAID](#), if you wish to continue the following tutorial in the meantime.

### Setup

If you are not already there, change directory to the ncov directory:

```
cd ncov
```

### Curate data from GISAID

We will retrieve 10 sequences from GISAID's EpiCoV database.

1. Navigate to [GISAID](#) and select **Login**.



2. Login to your GISAID account.



## Platform Login

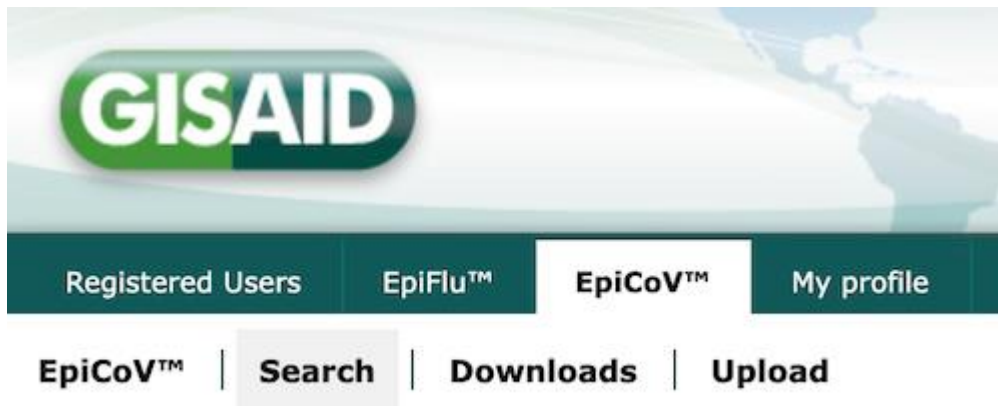
Username

Password

Login

[Forgot your password or username?](#)

3. In the top left navigation bar, select **EpiCoV** then **Search**.



4. Filter to sequences that pass the following criteria:
  1. Has a complete genome
  2. Has high coverage
  3. Has an exact collection date

**Search** Reset filters

EPI\_ISL ID  Virus name   Complete **1**

Location  Host   High coverage **2**

Collection  to  Submission  to   Low coverage excluded **3**

Clade  Lineage  Substitutions  Variant   With patient status **3**

Collection date complete **3**

Under investigation

<input type="checkbox"/>	Virus name	Passage date	Accession ID	Collection date	Submission date	Length	Host	Location	Originating
<input checked="" type="checkbox"/>	hCoV-19/USA/CO-CDC-MMB14798743/2021	Original	EPI_ISL_11683608	2022-03-21	2022-04-01	29,747	Human	North America / US	Mako Medical
<input checked="" type="checkbox"/>	hCoV-19/USA/SC-CDC-MMB14795956/2021	Original	EPI_ISL_11683568	2022-03-21	2022-04-01	29,747	Human	North America / US	Mako Medical
<input checked="" type="checkbox"/>	hCoV-19/Panama/GMI-PAM18357/2021	Original	EPI_ISL_11681860	2021-05-31	2022-04-01	29,894	Human	North America / Panama	Gorgas Medical
<input checked="" type="checkbox"/>	hCoV-19/Panama/GMI-PAM17995/2021	Original	EPI_ISL_11681859	2021-05-28	2022-04-01	29,884	Human	North America / Panama	Gorgas Medical
<input checked="" type="checkbox"/>	hCoV-19/Panama/GMI-PAM17936/2021	Original	EPI_ISL_11681858	2021-05-28	2022-04-01	29,884	Human	North America / Panama	Gorgas Medical
<input checked="" type="checkbox"/>	hCoV-19/Panama/GMI-PAM17703/2021	Original	EPI_ISL_11681857	2021-05-28	2022-04-01	29,903	Human	North America / Panama	Gorgas Medical
<input checked="" type="checkbox"/>	hCoV-19/Panama/GMI-PAM17453/2021	Original	EPI_ISL_11681856	2021-05-27	2022-04-01	29,884	Human	North America / Panama	Gorgas Medical
<input checked="" type="checkbox"/>	hCoV-19/Panama/GMI-PAM17180/2021	Original	EPI_ISL_11681855	2021-05-26	2022-04-01	29,884	Human	North America / Panama	Gorgas Medical
<input checked="" type="checkbox"/>	hCoV-19/Panama/GMI-PAM16945/2021	Original	EPI_ISL_11681854	2021-05-26	2022-04-01	29,903	Human	North America / Panama	Gorgas Medical
<input checked="" type="checkbox"/>	hCoV-19/Panama/GMI-PAM16911/2021	Original	EPI_ISL_11681852	2021-05-26	2022-04-01	29,903	Human	North America / Panama	Gorgas Medical
<input type="checkbox"/>	hCoV-19/Panama/GMI-PAM16293/2021	Original	EPI_ISL_11681850	2021-05-25	2022-04-01	29,884	Human	North America / Panama	Gorgas Medical
<input type="checkbox"/>	hCoV-19/Panama/GMI-PAM15608/2021	Original	EPI_ISL_11681849	2021-05-21	2022-04-01	29,884	Human	North America / Panama	Gorgas Medical
<input type="checkbox"/>	hCoV-19/Panama/GMI-PAM15190/2021	Original	EPI_ISL_11681848	2021-05-20	2022-04-01	29,885	Human	North America / Panama	Gorgas Medical
<input type="checkbox"/>	hCoV-19/Panama/GMI-PAM15152/2021	Original	EPI_ISL_11681847	2021-05-20	2022-04-01	29,903	Human	North America / Panama	Gorgas Medical

Total: 5,115,803 viruses

<< < 1 2 3 4 5 > >>

EPI\_SET Select Analysis **Download**

5. Select the first 10 sequences.
6. Select **Download** in the bottom right of the search results.
7. Select **Input for the Augur pipeline** as the download format.

## Download

- Acknowledgement table
- Input for the Augur pipeline



## Note

You may see different download options, but it is fine as long as **Input for the Augur pipeline** is available.

8. Select **Download**.
9. Download/move the .tar file into the `ncov/data/` directory.
10. Extract by opening the downloaded .tar file in your file explorer. It contains a folder prefixed with `gisaid_auspice_input_hcov-19_` containing two files: one ending with `.metadata.tsv` and another with `.sequences.fasta`.
11. Rename the files as `custom.metadata.tsv` and `custom.sequences.fasta`.
12. Move the files up to the `ncov/data/` directory.
13. Delete the empty `gisaid_auspice_input_hcov-19_`-prefixed folder and the .tar file if it is still there.

## Hint

Read [the full data prep guide](#) for other ways to curate custom data.

[Run the workflow](#)

From within the `ncov/` directory, run the `ncov` workflow using a pre-written [config file](#):

```
nextstrain build . --configfile ncov-tutorial/custom-data.yaml
```

[Break down the command](#)

The workflow can take several minutes to run. While it is running, you can investigate the contents of `custom-data.yaml` (comments excluded):

**inputs:**

- **name:** `reference_data`

**metadata:**

`https://data.nextstrain.org/files/ncov/open/reference/metadata.tsv.xz`

**sequences:**

`https://data.nextstrain.org/files/ncov/open/reference/sequences.fasta.xz`

- **name:** `custom_data`

**metadata:** `data/custom.metadata.tsv`

**sequences:** `data/custom.sequences.fasta`

**refine:**

**root:** `"Wuhan-Hu-1/2019"`

**builds:**

**custom-build:**

**title:** `"Build with custom data and example data"`

**subsampling\_scheme:** `all`

**auspice\_config:** `ncov-tutorial/auspice-config-custom-data.json`

This is the same as the previous file, with some additions:

1. A second input for the custom data, referencing the metadata and sequences files downloaded from GISAID.
2. A `builds` section that defines one output `dataset` using:
  1. A custom name `custom-build`, which will be used to create the dataset filename, in this case `auspice/ncov_custom-build.json`.
  2. A custom title `Build with custom data and example data`, which will be shown when you visualize the dataset in Auspice.
  3. A [pre-defined subsampling scheme](#) `all`, that tells the workflow to skip subsampling and use all input data.
  4. An Auspice config file, `ncov-tutorial/auspice-config-custom-data.json`, that defines parameters for how Auspice should display the dataset produced by the workflow. It has the following contents:

```
5. {
6.   "colorings": [
7.     {
8.       "key": "custom_data",
9.       "title": "Custom data",
10.      "type": "categorical"
11.    }
12.  ],
13.  "display_defaults": {
14.    "color_by": "custom_data"
15.  }
16. }
```

This JSON tells Auspice to:

1. Create a new coloring `custom_data` that reflects a special metadata column generated by the `ncov` workflow. When there is more than one input, each data input produces a new final metadata column with categorical values `yes` or `no` representing whether the sequence was from the input.
2. Set the default `Color By` as the new `custom_data` coloring.

Note

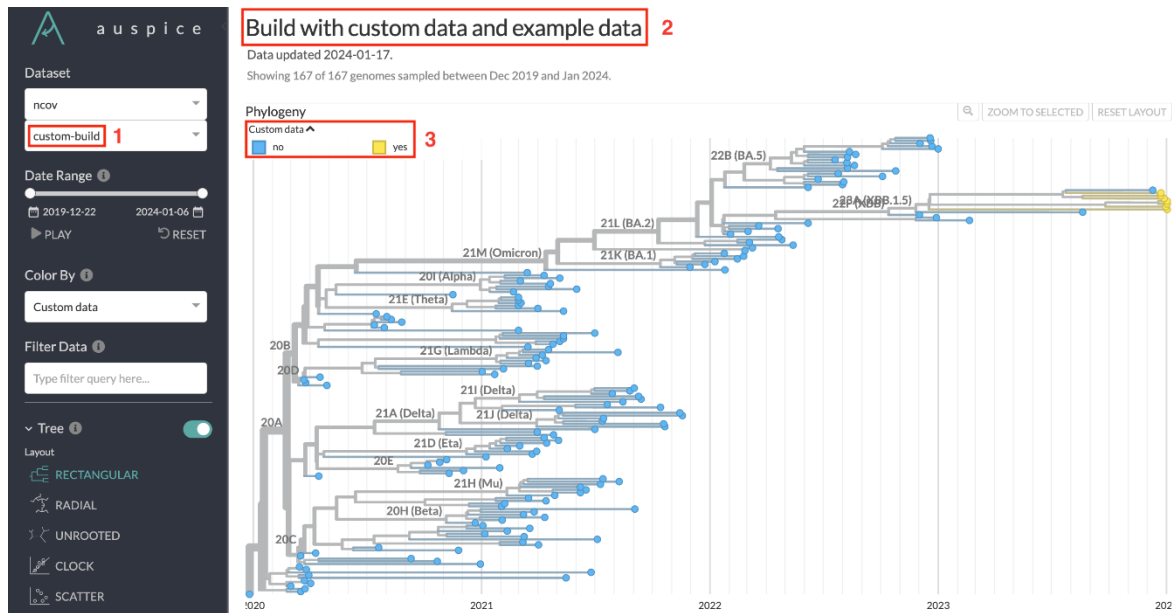
**Build** is a widely used term with various meanings. In the context of the ncov workflow, the builds: section defines output [datasets](#) to be generated by the workflow (i.e. “build” a dataset).

### Visualize the results

Run this command to start the [Auspice](#) server, providing `auspice/` as the directory containing output dataset files:

```
nextstrain view auspice/
```

Navigate to <http://127.0.0.1:4000/ncov/custom-build>. The resulting [dataset](#) should have similar phylogeny to the previous dataset, with additional sequences:



1. The custom dataset name `custom-build` can be seen in the dataset selector, as well as the dataset URL.
2. The custom dataset title can be seen at the top of the page.
3. The custom coloring is used by default. You can see which sequences are from the custom data added in this tutorial.

### Note

You may not see all 10 custom sequences - some can be filtered out due to quality checks built into the ncov workflow.