



**Identificación, clonación y análisis bioinformático de la proteína de la
Nucleocápside N del SARS-CoV-2 de aislados ecuatorianos provenientes de pacientes**

Andrade Cedeño, Hugo Ernesto

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Trabajo de titulación, previo a la obtención del título de Ingeniero en Biotecnología

Ph.D. Reyna Bello, Armando

7 de Febrero del 2022



Escrito de TESIS_Hugo Andrade.docx

Scanned on: 1:49 February 7, 2022 UTC



Overall Similarity Score



Results Found



Total Words in Text

Identical Words	392
Words with Minor Changes	0
Paraphrased Words	0
Omitted Words	2351



Website | Education | Businesses

Firma:

Firmado digitalmente por
 ARMANDO REYNA BELLO
 Nombre de reconocimiento (DN):
 c=EC, o=SECURITY DATA S.A. 1,
 ou=ENTIDAD DE CERTIFICACION DE
 INFORMACION,
 serialNumber=290620182334,
 cn=ARMANDO REYNA BELLO
 Fecha: 2022.02.07 08:13:57 -05'00'

.....
Armando Reyna Bello, Ph.D.

C. C.: 1758891830

Director de proyecto de Investigación



DEPARTAMENTO DE CIENCIAS DE LA VIDA Y DE LA AGRICULTURA
CARRERA DE INGENIERÍA EN BIOTECNOLOGÍA

CERTIFICACIÓN

Certifico que el trabajo de integración curricular, **“Identificación, clonación y análisis bioinformático de la proteína de la Nucleocápside N del SARS-CoV-2 de aislados ecuatorianos provenientes de pacientes”** fue realizado por el señor **Andrade Cedeño Hugo Ernesto** el cual ha sido revisado y analizado en su totalidad por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Santo Domingo de los Tsáchilas, 7 de febrero de 2022

Firma:

Firmado digitalmente por ARMANDO REYNA BELLO
Nombre de reconocimiento (DN): c=EC,
o=SECURITY DATA S.A. 1, ou=ENTIDAD DE
CERTIFICACION DE INFORMACION,
serialNumber=290620182334, cn=ARMANDO
REYNA BELLO
Fecha: 2022.02.07 08:14:24 -05'00'

.....
Armando Reyna Bello, Ph.D.

C. C.: 1758891830

Director de proyecto de Investigación



DEPARTAMENTO DE CIENCIAS DE LA VIDA Y AGRICULTURA
CARRERA DE INGENIERIA EN BIOTECNOLOGÍA

RESPONSABILIDAD DE AUTORÍA

Yo, **Andrade Cedeño Hugo Ernesto**, con cédula de ciudadanía N° 172321300-3, declaro que el contenido, ideas y criterios del trabajo de integración curricular: **“Identificación, clonación y análisis bioinformático la proteína de la Nucleocápside N del SARS-CoV-2 de aislados ecuatorianos provenientes de pacientes”** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Santo Domingo de los Tsáchilas, 7 de febrero de 2022

Firma

Andrade Cedeño Hugo Ernesto

C.C.: 172321300-3



**DEPARTAMENTO DE CIENCIAS DE LA VIDA Y AGRICULTURA
CARRERA DE INGENIERIA EN BIOTECNOLOGÍA**

AUTORIZACIÓN DE PUBLICACIÓN

Yo **Andrade Cedeño, Hugo Ernesto**, con cédula de ciudadanía N°172321300-3, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de integración curricular: **“Identificación, clonación y análisis bioinformático la proteína de la nucleocápside n del sars-cov-2 de aislados ecuatorianos provenientes de pacientes”** en el repositorio institucional, cuyo contenido, ideas y criterios son de mi/nuestra responsabilidad.

Santo Domingo de los Tsáchilas, 7 de febrero de 2022

Firma

Andrade Cedeño Hugo Ernesto

C.C.: 172321300-3

Dedicatoria

Con todo mi amor este trabajo se lo dedico a mis padres Luis Heraldo Ortiz Lezano y Synthia Caterine Andrade Moncayo, quienes han sido el pilar fundamental en mi desarrollo académico, y siempre han estado cuidando de mi aconsejándome e impulsándome a ser una persona mejor en este mundo. La herencia más grande que los padres pueden dar a sus hijos es el amor, la cultura y la educación, por todo esto y más estaré eternamente agradecidos con ellos y espero a futuro retribuirles mucho más de lo que me han dado.

“El científico no tiene por objeto un resultado inmediato. El no espera que sus ideas avanzadas sean fácilmente aceptadas. Su deber es sentar las bases para aquellos que están por venir y señalar el camino”

-Nikola Tesla

Agradecimiento

A Armando Reyna Ph.D., quien me ha instruido a lo largo de mi educación universitaria y despertado en mí una pasión por la ciencia y la investigación, gracias a su confianza en mí para participar en varios proyectos y generar ciencia de calidad e importancia.

A mi gran amigo Mateo Botero Pérez, con quien he compartido varias experiencias y pensamientos intelectuales para el desarrollo de este proyecto y con el cual espero poder seguir colaborando.

A Amed Botero que colaboró financiándonos los vectores de clonación, una parte esencial de este proyecto y sin lo cual no podríamos seguir desarrollándolo.

A Dr. Camilo Zurita dueño de Zurita & Zurita Laboratorios, quien nos colaboró con el material genético del RNA de pacientes positivos a SARS-CoV-2 para la amplificación de los genes.

A Jacobus H. de Waard Ph.D., Dayana Aguilar y Salome Guerrero, pertenecientes a la Universidad de las Américas UDLA, quienes nos colaboraron con la recepción y el procesamiento de muestras para la secuenciación.

A mis amigas Ing.Cristina Cholota y Michelle Yugcha quienes me impartieron sus conocimientos de biología sintética sin los cuales no hubiera podido desarrollar mi tesis de la mejor manera. A Jorge Ron Ph.D. por ser de gran apoyo moral durante el desarrollo de este proyecto. Y a Marbel Torres Ph.D. quién nos colaboró con reactivos para este proyecto.

A Ing. Vanessa Armijos e Ing.Katty Medina que nos colaboraron con la accesibilidad al laboratorio y reactivos necesarios a lo largo del proyecto.

A José Ruben Ph.D. quien nos colaboró con sus conocimientos en docking y bioinformática de análisis de modelos proteicos utilizados en el proyecto.

A mi amiga Belén Jácome quien fue de importancia para el contacto de la institución que nos provea el material genético, y me ha llenado de apoyo moral a lo largo del proyecto.

A mi amigo Jefferson Jumbo quién colaboro estandarizando ciertas técnicas utilizadas en este proyecto

A Paula Pérez, Amed Botero, Luna Botero, Clara Botero, Jerónimo Botero, Lucas Botero, Tomas Botero y Manuela Botero, quienes me recibieron en su hogar y me dieron apoyo moral en el desarrollo del proyecto.

A mis profesores Armando Reyna Bello, Natasha Baer, Pedro Puente, Paola Proaño, quienes me impartieron sus conocimientos de la mejor manera y ahora los considero mis amigos.

A mis amigos Mateo Botero, Jefferson Jumbo, Jorge Cueva, Jordan Sarango, Harrison Delgado, Brandon Cusme, Jhonny Maldonado, Jorge Alvear, Stefy Mena, Carlos Estrella y Stalyn Cunalata, Andrés Almachi con los que compartí varias experiencias a lo largo de mi carrera universitaria y me ayudaron en las buenas y en las malas.

A mis abuelos Hugo Andrade Valdez y Ana Moncayo quienes me acogieron en pandemia y me impulsan a ser una persona productiva y de bien en la sociedad. Y a mi hermano perro Nino que me da cariño constantemente.

Índice de contenidos

Caratula.....	1
Reporte de verificación de contenido.....	2
Certificación	3
Responsabilidad de Autoría	4
Autorización de Publicación.....	5
Dedicatoria.....	6
Agradecimiento.....	7
Índice de contenidos.....	9
Listado de Tablas.....	14
Listado de figuras	15
Listado de abreviaturas.....	17
Resumen	19
Abstract.....	20
Capítulo I: Introducción	21
Formulación del Problema.....	21
Justificación del problema	23
Objetivos de la Investigación	26
Objetivo General.....	26

	10
Objetivos Específicos	26
Capítulo II: Marco teórico	27
SARS-CoV-2 Historia.....	27
SARS-CoV-2 en América Latina	28
Características del SARS-CoV-2.....	29
Composición y estructura del SARS-CoV-2	29
Taxonomía del SARS-CoV-2.....	30
Ciclo del SARS-CoV-2, transmisión y diagnóstico.....	31
Proteína de la nucleocápside (N).....	34
Características de la proteína N.....	34
Importancia de la proteína N.....	36
Capítulo III: Materiales y métodos.....	37
Análisis <i>in silico</i> de la proteína N.....	37
Análisis de secuencias de Plataformas bioinformáticas (NCBI Virus y GISAID) en Mega X ...	37
Análisis de Regiones Antigénicas en Antheprot.....	41
Diseño de Primers (Primer Blast, Sequence Extractor, Oligo Analyzer)	41
Síntesis de material Genético	45
Producción de cDNA por retro transcripción.....	45
RT-PCR one Step Kit	45
Reacción en cadena de la polimerasa (PCR) convencional.....	46

Electroforesis en gel de agarosa	47
Clonación	48
Clonación al vector pGEM®-T Easy Vector	48
Clonación al vector de expresión pTrcHis2-TOPO	50
Cultivo en Medio Líquido LB y LB Broth Miller	51
Cultivo en Medio LB suplementado con Ampicilina, X-Gal, IPT	52
Células Químicamente Competentes E.coli JM109	52
Transformación	53
Colony PCR	53
Extracción de Plásmido	54
Corte con enzimas de Restricción	55
Generación de Modelos Bioinformáticos	56
Análisis Estadístico	56
Diseño no experimental	56
Variable respuesta	57
Error aleatorio	57
Secuenciación	57
Análisis Filogenético	57
Capítulo IV: Resultados	58
Identificación de la secuencia de la proteína N	58

	12
Amplificación de la proteína N.....	61
Estandarización de PCR.....	61
PCR convencional.....	63
Secuenciación	64
Clonación del gen de la Proteína N.....	65
Screening de colonias a través de Colony PCR	66
DNA plasmídico de cepas de clonación	67
Amplificación del gen N a partir de DNA plasmídico	68
Análisis por corte con enzimas de restricción al vector de clonación	69
Secuenciación de Insertos.....	71
Análisis de las colonias de Expresión	73
Análisis bioinformático de la secuencia de expresión	76
Filogenia.....	84
Capítulo V: Discusión	88
Antigenicidad de la Proteína N	88
Mutaciones de la proteína N	89
Identidad y Estructura de la Proteína N.....	90
Análisis Filogenético.....	91
Capítulo VI: Conclusiones.....	92
Capítulo VII: Recomendaciones	94

Capitulo VIII: Bibliografía	95
-----------------------------------	----

Listado de Tablas

Tabla 1. <i>Secuencias pertenecientes a la proteína N de la plataforma NCBI Virus</i>	38
Tabla 2. <i>Números de accesoión del GISAID</i>	39
Tabla 3. <i>Propiedades de los pares de Primers</i>	44
Tabla 4. <i>Componentes para RT-PCR One Step Kit</i>	46
Tabla 5. <i>Componentes para PCR convencional</i>	47
Tabla 6. <i>Componentes de Reacción para Ligación en vector de Clonación</i>	48
Tabla 7. <i>Componentes de Reacción para Ligación en vector de Expresión</i>	50
Tabla 8. <i>Componentes de Digestión del vector de clonación</i>	55
Tabla 9. <i>Componentes de digestión del vector de expresión</i>	56
Tabla 10. <i>Mejores candidatos para Epitopes</i>	58
Tabla 11. <i>Cambios aminoacídicos de la proteína N presentes en la región latinoamericana</i>	60
Tabla 12. <i>Regiones antigénicas de proteína expresada</i>	77

Listado de figuras

Figura 1. <i>Composición del genoma del SARS-CoV-2</i>	29
Figura 2. <i>Representación de la estructura del SARS-CoV-2</i>	30
Figura 3. <i>Ciclo de del SARS-CoV-2</i>	32
Figura 4. <i>Organización Estructural de la proteína N del SARS-CoV-2</i>	34
Figura 5. <i>Potencial de Inmunogenicidad de la proteína N</i>	36
Figura 6. <i>Alineamiento de secuencias en MEGA X</i>	40
Figura 7. <i>Configuración de PrimerBlast</i>	42
Figura 8. <i>Primers Forward en Fase</i>	43
Figura 9. <i>Primers Reverse en Fase</i>	43
Figura 10. <i>Mapa del vector pGEM®-T Easy</i>	49
Figura 11. <i>Mapa del vector pTrcHis2-TOPO</i>	51
Figura 12. <i>Perfiles de Antheprot para la selección de epitopes</i>	59
Figura 13. <i>Representación de cambios aminoacídicos de la Proteína N</i>	61
Figura 14. <i>Gradiente de temperatura del gen de la proteína N</i>	62
Figura 15. <i>Múltiples clones del gen N- E. coli Jm109</i>	65
Figura 16. <i>Amplificación del gen N de colonias transformadas con pGEM®-T Easy</i>	66
Figura 17. <i>DNA plasmídico de cepas positivas</i>	67
Figura 18. <i>PCR de la proteína N a distintas concentraciones del DNA plasmídico</i>	69
Figura 19. <i>Digestión de NP4 con EcoRI</i>	70
Figura 20. <i>Identificación de Gen N en Fase</i>	75
Figura 21. <i>Perfil de proteína expresada por cepa N3</i>	78
Figura 22. <i>Dominio de Unión al RNA</i>	79
Figura 23. <i>Enlazador Central (LINKER)</i>	80

Figura 24. <i>Dominio de Dimerización</i>	81
Figura 25. <i>Estructura de la superficie de la proteína N</i>	82
Figura 26. <i>Interacción de Residuos del RBD con el RNA</i>	83
Figura 27. <i>Árbol filogenético Aislados Latinoamericanos</i>	84
Figura 28. <i>Árbol Filogenético de Aislados Ecuatorianos y variantes del SARS-CoV-2</i>	86

Listado de abreviaturas

SARS-CoV-2:	Síndrome Respiratorio Agudo Severo por Coronavirus 2
SARS-CoV :	Síndrome Respiratorio Agudo Severo por Coronavirus
MERS-CoV:	Síndrome respiratorio de Oriente Medio por Coronavirus
DNA:	Ácido Desoxirribonucleico
cDNA:	Ácido Desoxirribonucleico Complementario
RNA:	Ácido Ribonucleico
PCR:	Reacción en Cadena de la Polimerasa
RT-PCR:	Reacción en Cadena de la Polimerasa con Retro transcripción
RBD:	Dominio de Unión al RNA
NCBI:	Centro Nacional para la Información Biotecnológica
bp:	Pares de Bases
aa:	Aminoácidos
L:	Litro
mL:	Militro
μL:	Microlitro
M:	Molar
mM:	Milimolar

μM:	Micromolar
g:	Gramos
mg:	Miligramos
ng:	Nanogramos
U:	Unidad de TaqPolimerasa

Resumen

El SARS-CoV-2 es el virus causante de la COVID-19, enfermedad de la cual se han reportado más de 4 millones de muertes alrededor del mundo. Es por ello por lo que es necesario estudiar sus distintos genes con la finalidad de desarrollar métodos de diagnóstico y ensayos que nos permitan determinar el estado de la enfermedad en la población. El presente estudio tubo la finalidad de identificar, clonar y analizar por bioinformática la proteína de la Nucleocápside N del SARS-CoV-2 de aislados ecuatorianos, con vías al desarrollo de una prueba de ELISA. A través de herramientas de biología molecular se logró aislar el gen de la proteína N, nuestro gen consto de 1218bp y alrededor de 406 residuos aminoacídicos y con una identidad del 100% en relación con otras secuencias del GenBank. Una vez obtenida la secuencia que se clonó se realizó un análisis de regiones antigénicas, en donde se estableció que la región del Enlazador Central (LINKER) posee la mayor cantidad de epítopes, también las mutaciones más comunes en donde se aprecia sustituciones en los residuos 203 y204, además los modelos del Dominio de Unión al RNA (RBD), la región LINKER, y el Dominio de Dimerización; y por último los análisis filogenéticos en donde se pudo observar distintas relaciones entre aislados ecuatorianos en relación a la región latinoamericana y las Variantes del SARS-CoV-2 considerando al gen N.

Palabras Clave

- SARS-COV-2
- PROTEÍNA N
- ANÁLISIS BIOINFORMÁTICO
- MODELAMIENTO PROTEICO
- CLONACIÓN

Abstract

SARS-CoV-2 is the virus that causes COVID-19, a disease from which more than 4 million deaths have been reported around the world. That is why it is necessary to study its different genes to develop diagnostic methods and tests that allow us to determine the state of the disease in the population. The purpose of this study was to identify, clone and analyse by bioinformatics the Nucleocapsid N protein of SARS-CoV-2 from Ecuadorian isolates, with pathways to the development of an ELISA test. Through molecular biology tools, it was possible to isolate the N protein gene, our gene consisted of 1218bp and around 406 amino acid residues and with 100% identity in relation to other GenBank sequences. Once the sequence that was cloned was obtained, an analysis of antigenic regions was carried out, where it was established that the Central Linker region (LINKER) has the largest number of epitopes, as well as the most common mutations where substitutions can be seen in residues 203 y204, in addition to the models of the RNA Binding Domain (RBD), the LINKER region, and the Dimerization Domain; and finally the phylogenetic analyses where it was possible to observe different relationships between Ecuadorian isolates in relation to the Latin American region and the Variants of SARS-CoV-2 considering the N gene.

Key Words

- **SARS-COV-2**
- **N PROTEIN**
- **BIOINFORMATIC ANALYSIS**
- **PROTEIN MODELLING**
- **CLONING**

Capítulo I: Introducción

Formulación del Problema

El virus del síndrome respiratorio agudo severo (SARS-CoV-2), es un tipo de coronavirus humano causante de la COVID-19, hasta el momento se han reportado más de 190 millones de contagios y aproximadamente más de 4 millones de muertes alrededor de todo el mundo, dicha enfermedad fue declarada por la Organización Mundial de la Salud (OMS) el 11 de Marzo del 2020 como una emergencia de salud pública de importancia mundial, tras la identificación de la enfermedad como una pandemia(World Health Organization, 2020a). Lo que generó en el mundo una emergencia sanitaria debido a su modo de transmisión, genética, mutaciones, entre otros(Zeng et al., 2020). Desde entonces reportes del SARS-CoV-2 han mostrado que este tiene características similares a otros coronavirus de origen zoonótico como el SARS-CoV y el MERS-CoV, en donde sus genomas son compartidos en un 79% y 50% respectivamente. El virus ha sido clasificado el grupo de los beta coronavirus, característicos por generar infecciones graves en las vías respiratorias superiores e inferiores(Hu et al., 2020), al ser compuesto por una simple hebra de ARN , de aproximadamente 30 000 nucleótidos, el cual consta de seis fragmentos abiertos de lectura (ORF's) funcionales: replicasa (ORF1a/ORF1b), Spike (S), Envoltura (E), Membrana (M) y nucleocápside (N), la clínica que presentan los pacientes entre los 2 y 14 días tras la infección incluye fiebre, fatiga, tos, disnea, producción de esputo, dolor de garganta y cabeza, en un cuadro neumónico grave que puede conllevar a la muerte del paciente o darse de forma asintomática (Habas et al., 2020).

En Latinoamérica el SARS-CoV-2 afecto a la población tras el primer reporte en Brasil el 29 de Febrero del 2020, posteriormente Ecuador sería el segundo país en reportar su primer caso el 10 de Marzo del 2020, a partir de aquí Ecuador se convertiría en uno de los países con mayor índice de mortalidad en la región, en su mayoría debido a poca conciencia del riesgo por parte de la población y baja rigurosidad de las políticas de salud pública, en sus primeras etapas de pandemia(Márquez et al., 2020). Hasta el momento Ecuador se mantiene en pandemia con alrededor de 486 544 casos confirmados y 31 591 fallecidos(Ministerio de Salud Pública, 2021), estos datos han sido generados gracias a que hasta el momento se han desarrollado distintas técnicas moleculares y serológicas para el diagnóstico del SARS-CoV-2, como la Reacción Cuantitativa en Cadena de la Polimerasa con Transcriptasa Inversa (RT-qPCR) o el Ensayo Inmunoabsorbente Ligado a Enzimas (ELISA) respectivamente, sin embargo, aunque en RT-qPCR se utilice como un método de diagnóstico directo con la detección de uno o más genes, los inmunoensayos basados en proteínas estructurales usan regularmente la proteína Spike (S), considerando que esta es inmunodominante y es la mayor proteína transmembrana(Alpdagtas et al., 2020).

Por otra parte, en el Ecuador se está incrementando el total de dosis administradas en su plan de vacunación, por el momento con un total de 10 584 021 dosis administradas, las cuales se basan en generar una respuesta inmune específica contra la proteína S (Pan American Health Organization, 2021), considerando que se utilizan como dianas inmunoglobulinas totales IgM e IgG para los inmunoensayos basados en anticuerpos (Ab), y a su vez es necesario ensayos que permitan detectar anticuerpos específicos del virus para comprender la prevalencia de la infección (Burbelo et al., 2020), es por ello que se deben considerar otros tipos de proteínas estructurales, proteína E o proteína N, como marcadores para la identificación de infecciones, incluso cuando los pacientes se encuentran vacunados (Alpdagtas et al., 2020).

Justificación del problema

Aunque para el diagnóstico directo del SARS-CoV-2 se utiliza RT-qPCR, los inmunoensayos ayudan a determinar la previa exposición al virus a través de la respuesta específica de anticuerpos (Alpdagtas et al., 2020). Debido a la presente inmunización que genera respuesta de anticuerpos específicos contra la proteína S, es necesario generar ensayos serológicos que nos permitan estimar la proporción de la población expuesta al virus, es por ello que se deben considerar otro tipo de proteínas estructurales para su uso en inmunoensayos (Rump et al., 2021). En el presente trabajo nos centramos en la utilización de la proteína de nucleocápside (NP) como marcador para dicha estimación de la exposición, considerando que se ha reportado como un marcador esencial para la discriminación entre la inmunidad adquirida por una infección o una vacunación (Vos et

al., 2021). Aunque la presencia de Ab se presentan entre los 10 y 21 días después de la infección(Kellam & Barclay, 2020), reportes indican que la presencia de los mismos en otros tipo de coronavirus se mantienen entre los 3 y 13 meses tras no presentar síntomas(Rump et al., 2021) por lo que se podría estimar la seroprevalencia del SARS-CoV-2 aun la población se encuentre inmunizada.

Las proteínas N del SARS-CoV-2 es un grupo de proteínas de unión multidominantes al RNA que cumplen una función critica en el ciclo de vida del virus, debido a que participa en la replicación y el empaquetamiento del ARN genómico, formando un complejo helicoidal de ribonucleoproteina con el RNA viral(Pei et al., 2005). Esta proteína consta de una magnitud de 419 aminoácidos (aa) en donde se pueden diferenciar cinco dominios: Dominio N-Terminal (NTD), Dominio de Unión al RNA (RBD), región de enlace estructurada (LINKER), Dominio de dimerización y un Dominio C-Terminal (CTD)(Cubuk et al., 2021). Esta proteína tiene una gran importancia en los diagnósticos serológicos considerando que existen reportes de alta sensibilidad en sueros de pacientes al encontrar anticuerpos específicos contra la proteína N, además de presentar una mayor persistencia que otras proteínas estructurales en el caso del SARS-CoV(Zeng et al., 2020), con el cual comparte aproximadamente un 90% de similitud(Satarker & Nampoothiri, 2020). Otra característica que la hace clave para su uso es su bajo nivel de mutación (Vos et al., 2021),y que es la mayor proteína estructural expresada durante la infección generando anticuerpos específicos con una tasa de vida más larga y en mayor abundancia en los pacientes(Rump et al., 2021).

En el Ecuador no existen reportes que evaluaran la seroprevalencia de población infectada considerando pacientes ya vacunados y la proteína N como marcador, sin embargo, para realizar este tipo de análisis es necesario generar las herramientas biotecnológicas adecuadas. Es por lo que el presente trabajo se centra en la identificación, clonación y análisis bioinformático de la proteína de la nucleocápside N del SARS-CoV-2 de aislados ecuatorianos provenientes de pacientes, con miras al desarrollo a una prueba de ELISA en la cual se utilice como antígeno recombinante la proteína N para el diagnóstico serológico y a su vez impulsando el desarrollo de la investigación en el Ecuador que permita a posterior, realizar análisis del comportamiento del SARS-CoV-2 en la población, lo que pueden ayudar en la toma de decisiones con respecto medidas sanitarias con la finalidad de evitar nuevamente el aumento masivo de contagios ante posibles variantes que puedan escaparse de los anticuerpos inducidos por las vacunas de la población.

Objetivos de la Investigación

Objetivo General

Identificar, clonar y analizar por bioinformática la proteína de la Nucleocápside N del SARS-CoV-2 de aislados ecuatorianos provenientes de pacientes

Objetivos Específicos

- Realizar un análisis *In-silico* de la proteína de la Nucleocápside N del SARS-CoV-2, a través de herramientas de bioinformática.
- Realizar la clonación de la proteína de la Nucleocápside N del SARS-CoV-2, por medio de herramientas de Biología Molecular
- Analizar por bioinformática la secuencia clonada en el vector de expresión

Capítulo II: Marco teórico

SARS-CoV-2 Historia

El primer caso registrado de neumonía fue el 8 de diciembre del 2019 en Wuhan, China, posteriormente a finales de diciembre del mismo año se dio el primer reporte de 27 casos de neumonía por causas desconocidas en la misma ciudad. El 9 de enero del 2020, China anunciaría que el causante de los nuevos brotes de neumonía fue identificado como un nuevo coronavirus. La OMS el 10 de enero de 2020, generaría un paquete de orientación técnica para la identificación y manejo de casos probables considerando experiencia previa con el SARS y el MERS(World Health Organization, 2020a). El 12 de enero del 2020, China compartiría públicamente la secuencia genética del virus y el 13 de enero del 2020 las autoridades de Tailandia confirmarían el primer caso de neumonía causante del SARS.CoV.2 fuera de China. El 22 de enero del 2020 la OMS confirma que existe evidencia de transmisión persona a persona en Wuhan, y por siguiente el 23 de enero del 2020 la ciudad de Wuhan fue cerrada. Considerando que el 29 de enero del 2020 se informa que el coronavirus se ha dispersado a través de 34 provincias en China, la OMS el 30 de enero del 2020 declara una emergencia de salud pública de importancia internacional, informando a su vez un total de 7818 casos alrededor de todo el mundo. El 11 de febrero del 2020 el Comité Internacional de Taxonomía de Virus (ICTV), nombró al virus SARS-CoV-2 y la OMS nombro a la enfermedad causante COVID-19. Es por ello, en el 11 de marzo del 2020 que la OMS debido a la gravedad de la enfermedad y el grado de propagación, evaluara a la COVID-19 y la caracterizará como una pandemia, llegando a generar en el transcurso de los 10 meses posteriores, específicamente el 2 de octubre del 2020, más de 30 millones de casos y 1 millón de muertes(Hu et al., 2020).

SARS-CoV-2 en América Latina

El primer reporte de la presencia del SARS-CoV-2 en América Latina fue el 25 de febrero del 2020 en Brasil, según reportes de su ministerio de salud. Posteriormente el primer reporte de fallecimiento se generó en marzo 7 en Argentina, producto de un contagio por viajeros provenientes principalmente de Italia.

Aunque los gobiernos intentaron tomar medidas de contingencia para evitar la propagación masiva del COVID-19 en noviembre del 2020, Latinoamérica fue considerada la región con mayores casos confirmados, representando la cuarta parte de los casos notificados en el mundo. Además, se tuvieron casos característicos como el de Ecuador en donde las ciudades de Quito y Guayaquil tuvieron una saturación extremadamente rápida del sistema sanitario, generando posteriormente su colapso. Por Otra parte, en el Perú, aunque fue uno de los primeros países en poner restricciones su mortalidad posterior de los cinco primeros meses ascendió a los 87.53 muertes por cada 100 000 habitantes, poniendo su índice sobre países como México, Brasil y Estados Unidos.

Se considera que el impacto que genero el SARS.CoV-2 en América Latina fue mayor con respecto a países de Asia y Europa, debido a las medidas de bioseguridad que no se practicaron por la sociedad en los inicios de la pandemia como son el distanciamiento social, la restricción de movilidad, el cierre de negocios y la generación de cuarentena por parte de los enfermos de manera voluntario o por mandato del gobierno(Callejas et al., 2020).

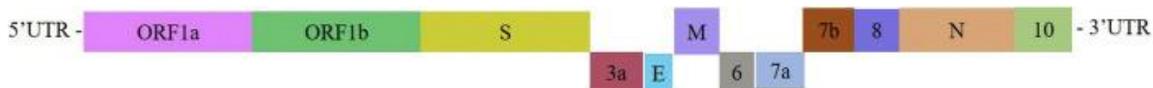
Características del SARS-CoV-2

Composición y estructura del SARS-CoV-2

El SARS-CoV-2 es un tipo de β coronavirus el cual consta de un genoma de RNA de cadena positiva con una longitud de 30 kilo bases (Yang et al., 2020). Reportes indican que el genoma está envuelto dentro de una serie de proteínas de nucleocápside y consiste de diez marcos de lectura abiertos (ORF): el ORF1a/1b se encarga de codificar las poliproteínas 1a/1b respectivamente así como de 1-16 proteínas no estructurales, los demás ORF's restantes se encargan de codificar las proteínas estructurales Spike (S), Envoltura (E), Membrana (M), nucleocápside (N) y siete proteínas accesorias que se intercalan entre los genes estructurales (Satarker & Nampoothiri, 2020), en la Figura 1 se puede apreciar un esquema de la configuración de los genes del SARS-CoV-2 y en la Figura 2 su estructura .

Figura 1.

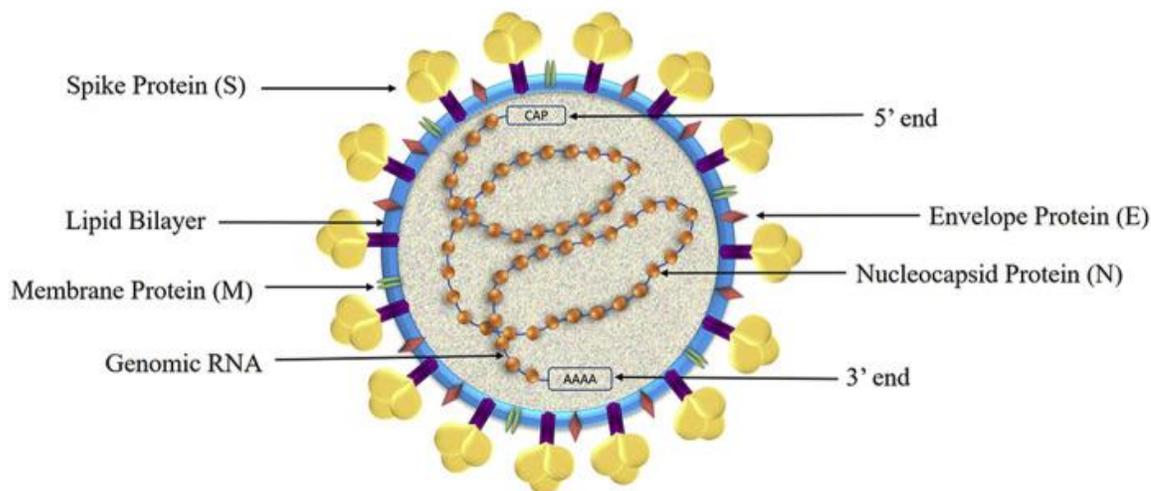
Composición del genoma del SARS-CoV-2



Nota 1. La figura muestra los distintos ORF del genoma del SARS-CoV-2, entre ellos las proteínas estructurales S, M, E y N, en conjunto a las no estructurales. Tomado de Satarker, S., & Nampoothiri, M. (2020). Structural Proteins in Severe Acute Respiratory Syndrome Coronavirus-2. *Archives of Medical Research*, 51(6), 482

Figura 2.

Representación de la estructura del SARS-CoV-2



Nota 2. La figura muestra la estructura del SARS-CoV-2, se puede apreciar desde el exterior se encuentra rodeado por las proteínas Spike, las cuales se asientan en una bicapa lipídica, en donde se encuentran proteínas de membrana (M), en conjunto a proteínas de envoltura (E), y dentro de dicha bicapa se encuentra el RNA vírico protegido por la proteína de la nucleocápside (N). Tomado de Satarker, S., & Nampoothiri, M. (2020). Structural Proteins in Severe Acute Respiratory Syndrome Coronavirus-2. *Archives of Medical Research*, 51(6), 482

Taxonomía del SARS-CoV-2

Taxonomía tomada de: (Helmy et al., 2020)

SuperReino: Viruses

Reino: Orthornavirae

Phylum: Pisuviricota

Clase: Pisoniviricetes

Orden: Nidovirales

Familia: Coronavirinae

Género: Betacoronavirus

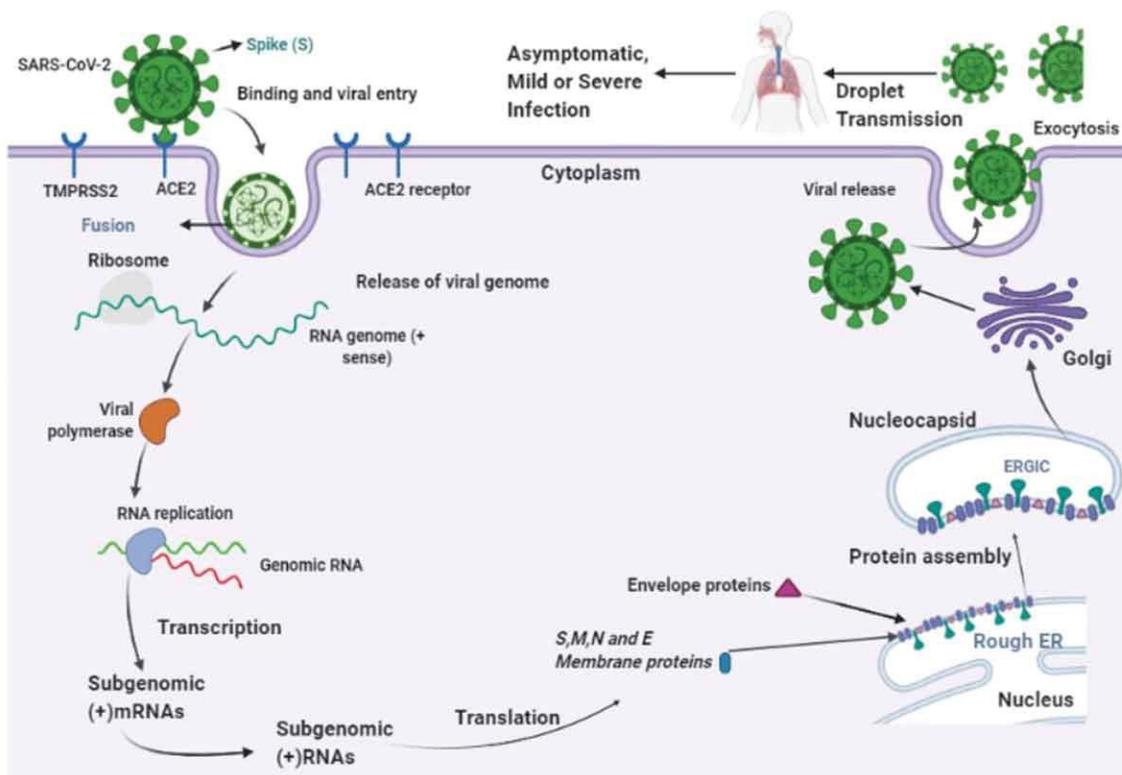
Especie: *Severe acute respiratory syndrome-related coronavirus (SARS-CoV-2)*

Ciclo del SARS-CoV-2, transmisión y diagnóstico

El ciclo del SARS-CoV-2, Figura 3, inicia cuando al ingresar a un huésped, las proteínas Spike (S) se unen al receptor de la enzima convertidora de angiotensina (ACE2), receptor que se encuentra presente en el epitelio celular de los pulmones, vasos sanguíneos y riñones. Tras su unión existe una fusión entre la membrana del virus y la célula huésped por endocitosis. Tras esto el ARN genómico ingresa al citoplasma celular y a través del casquete metilado 5' y la cola 3' poliadenilada, se genera una unión entre el ribosoma y el RNA que permite la traducción de la polimerasa dependiente de RNA, dicha polimerasa ayuda en la síntesis del RNAm genómico. Por medio de los ribosomas se convertirán en proteínas de membrana y proteínas accesorias, en conjunto a las proteínas estructurales del virus, Spike (S), Envoltura (E), Membrana (M) y Nucleocápside (N), el proceso de traducción de proteínas y su ensamblaje se realiza en el retículo endoplasmático rugoso, posteriormente la progenie del virus se libera por exocitosis a través de las vesículas de Golgi a la membrana celular y ulteriormente al espacio extracelular con la finalidad de continuar la transmisión (Habas et al., 2020).

Figura 3.

Ciclo de del SARS-CoV-2



Nota 3. La figura muestra el ciclo de vida del SARS-CoV-2, Este ingresa a la célula a través de los receptores ACE2, por medio de los ribosomas sintetizan una polimerasa viral que permitirá generar RNAm que contendrán la información para la traducción de las proteínas estructurales y no estructurales del virus, dichas proteínas se ensamblan en el retículo endoplasmático rugoso, generando una nuevo progenie que por exocitosis saldrá de la célula y permitirá la transmisión o infección de otras células. Tomado de Habas, K., Nganwuchu, C., Shahzad, F., Gopalan, R., Haque, M., Rahman, S., Majumder, A. A., & Nasim, T. (2020). Resolution of coronavirus disease 2019 (COVID-19). In *Expert Review of Anti-Infective Therapy* (Vol. 18, Issue 12, pp. 1201–1211). Taylor and Francis Ltd.

El modo de transmisión entre persona a persona del SARS-CoV-2, es a través de aerosoles respiratorios expulsados al hablar, cantar, estornudar, respirar o toser (World Health Organization, 2020b). Otra forma de transmisión es a través del contacto de manos de las personas en superficies contaminadas y posterior contacto hacia sus vías mucosas, manos, ojos, oídos y nariz; una persona infectada asintomática que no respete las medidas de seguridad podría ser un fuente de infección para 2 a 4 personas que se encuentren sanas (Callejas et al., 2020).

En cuanto al diagnóstico, existen distintos tipos de detección del SARS-CoV-2, los más utilizados son RT-qPCR y los test serológicos de identificación de anticuerpos del SARS-CoV-2. En el caso del RT-qPCR se utiliza para la rápida detección de RNA viral de manera rápida y altamente sensitivo, sobretodo cuando existe extremadamente bajas concentraciones de RNA viral, permitiendo identificar incluso a pacientes asintomáticos. Otro ensayo molecular es la amplificación isotérmica mediada por bucle de transcriptasa inversa rápida (RT-LAMP), su ventaja es que permite a los laboratorios analizar 2,5 más muestras en comparación a la RT-qPCR estándar. En el caso de los ensayos serológicos que permiten detectar anticuerpos como el ensayo inmunofluorescente (IFA), la prueba inmunocromatográfica (ICT) y el ensayo de inmunoadsorción ligada a enzimas (ELISA), se utilizan principalmente para confirmar las infecciones, al identificar la presencia de IgM e IgG en los sueros de los pacientes (Habas et al., 2020), aunque no es el principal ensayo directo para el diagnóstico concluyente, debido a que reportes muestran que los pacientes seroconvierten entre los 7 a 11 días tras la exposición al virus (Patel et al., 2020).

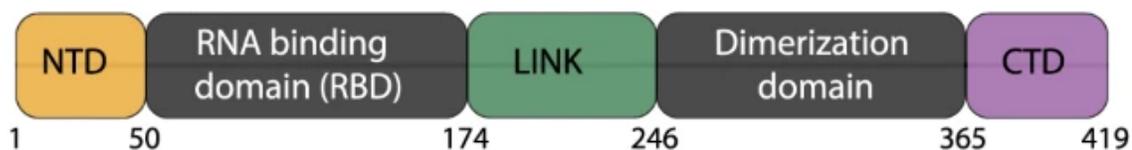
Proteína de la nucleocápside (N)

Características de la proteína N

La proteína N juega un importante rol en el empaquetamiento del genoma viral, formando un complejo helicoidal ribonucleoproteico con el RNA viral (Rump et al., 2021), dicho complejo ayuda al material genético a interactuar con el proceso celular tras la fusión celular con el virus (Satarker & Nampoothiri, 2020). Esta proteína consta de 419 aminoácidos, figura 1, los cuales pueden diferenciarse en cinco dominios: Dominio N-Terminal (NTD) del 1 aa al 50 aa, Dominio de Unión al RNA (RBD) del 51 aa al 174 aa, región de enlace estructurada (LINKER) del 176 aa al 246 aa, Dominio de dimerización del 247 aa al 365 aa, y un Dominio C-Terminal (CTD) del 366aa al 419aa; estos cinco dominios están unidos al ARN sin embargo el dominio de dimerización permite que se formen dímeros estequiométricos de manera bien definida (Cubuk et al., 2021).

Figura 4.

Organización Estructural de la proteína N del SARS-CoV-2



Nota 4. Estructura de la proteína N, las distintas regiones proteicas codificadas por los aminoácidos se pueden observar en diferentes colores, mientras que sus límites de los dominios se presentan en la parte inferior. Tomado de Cubuk, J., Alston, J. J., Incicco, J. J., Singh, S., Stuchell-Breton, M. D., Ward, M. D., Zimmerman, M. I., Vithani, N., Griffith, D., Wagoner, J. A., Bowman, G. R., Hall, K. B., Soranno, A., & Holehouse, A. S. (2021). The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nature Communications*, 12(1).

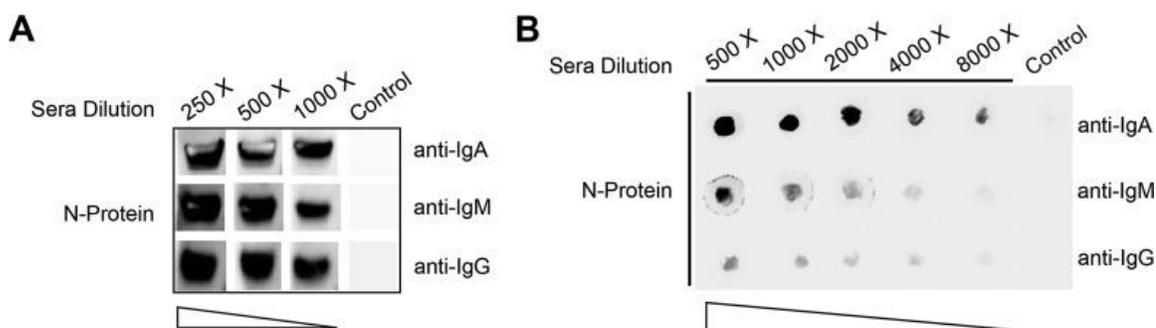
La proteína N es considerada como la proteína que se expresa de forma más abundante sobre todo en las primeras etapas de infección (Ye et al., 2020), además que con respecto al SARS-CoV tiene una identidad similar de aproximadamente el 90%. Se sabe además que la síntesis de RNA viral aumenta cuando tiene interacción con la ribonucleoproteína nuclear heterogénea celular humana A1, relacionada con el empalme pre-RNA en el núcleo y la regulación traduccional en el citosol, de modo que la ribonucleoproteína nuclear heterogénea celular humana A1, puede tener una comunicación cruzada con la proteína N que permite formar un complejo de regulación de transcripción y replicación del SARS-CoV (Luo et al., 2005). Del mismo modo dentro del organismo dichas proteínas promueven la activación de la ciclooxigenasa-2 que en consecuencia produce inflamación de los pulmones (Satarker & Nampoothiri, 2020), adicionalmente interactúa con la subunidad del proteosoma p42, inhibiendo la capacidad proteolítica de los proteosomas (Q. Wang et al., 2010), encargada de la degradación de las proteínas virales, inhibe igualmente el interferón tipo I, restringiendo las respuestas inmunitarias ante infecciones de virus (Lu et al., 2011).

Importancia de la proteína N

Esta es una proteína considerada altamente antigénica, además de iniciar una buena respuesta de células T tras ser bien expresada en células presentadoras de antígenos(Rump et al., 2021). Por otra parte reportes indican que en otros tipos de coronavirus como el coronavirus de Turquía, el coronavirus murino y el coronavirus porcino, este es un antígeno bastante potente(Timani et al., 2004). Otros análisis como los de la Figura 5, muestran gran antigenicidad ante anticuerpos IgG, IgA e IgM, permitiendo confirmar que este es un buen candidato para el diagnóstico de la enfermedad en los pacientes(Zeng et al., 2020), además que la proteína N se mantiene en total expresión durante la infección(Timani et al., 2004), generando que los anticuerpos contra dicha proteína tengan una duración de hasta 52 semanas tras no presentar síntomas de la enfermedad(Rump et al., 2021).

Figura 5.

Potencial de Inmunogenicidad de la proteína N



Nota 5. Se puede apreciar la Antigenicidad de la proteína N, en la Figura 5^a, se muestra un WesternBlot con distintas soluciones de Suero de anticuerpos IgA, IgM e IgG, contra el antígeno de la proteína N. En la Figura5B, se muestra un DotBlot en distintas soluciones de Suero contra anticuerpos iGA, IgM e IgG, contra el antígeno de la proteína N. Tomado de Zeng, W., Liu, G., Ma, H., Zhao, D., Yang, Y., Liu, M., Mohammed, A., Zhao, C., Yang, Y., Xie, J., Ding, C., Ma, X., Weng, J., Gao, Y., He, H., & Jin, T. (2020a). Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochemical and Biophysical Research Communications*, 527(3), 618.

Capítulo III: Materiales y métodos

Análisis *in silico* de la proteína N

Es necesario identificar las características de la secuencia a estudiar, así como el fragmento de la misma que será amplificada, es por ello que inicialmente se realizó un análisis *in silico* con la finalidad de identificar las diferencias aminoacídicas dentro de la región Latinoamericana y Ecuatoriana, además después de identificar dichas diferencias es necesario observar si afectan en la antigenicidad de la proteína, tras delimitar aminoacidicamente la secuencia de interés se procedió a un diseño de primers con la finalidad de amplificar por PCR nuestra secuencia seleccionada

Análisis de secuencias de Plataformas bioinformáticas (NCBI Virus y GISAID) en Mega X

Se utilizaron dos plataformas para el análisis aminoacídico de las secuencias. En el primer caso se usó NCBI-Virus, un portal que recopila datos de secuencias virales de tipo RefSeq de repositorios del GenBank y afines a NCBI que es el Centro Nacional de Información Biotecnológica de los Estados Unidos. Con la ayuda del buscador de la plataforma se delimitó la especie “Severe acute respiratory síndrome coronavirus 2, taxid:2697049” y la región geográfica América del Sur, en la tabla 1, se puede apreciar el número de acceso aminoacídico y nucleotídico de las secuencias seleccionadas.

Tabla 1.*Secuencias pertenecientes a la proteína N de la plataforma NCBI Virus*

Secuencia Aminoacídica	Secuencia Nucleotídica	País
QNH88891	MT907515.1	Venezuela
QNH88903	MT907516.1	Venezuela
QNH88915	MT907517.1	Venezuela
QNH88927	MT907518.1	Venezuela
QNH88939	MT907519.1	Venezuela
QNH88950	MT907520.2	Venezuela
QNH88961	MT907521.1	Venezuela
QMT98148	MT835383.1	Brasil
QMU26003	MT844030.1	Brasil
QMT97881	MT835027.1	Brasil
QMT27585	MT827074.1	Brasil
QMT27598	MT827075.1	Brasil
QMT27612	MT827190.1	Brasil
QMT27624	MT827202.1	Brasil
QMT29017	MT827872.1	Brasil
QMT29093	MT827940.1	Brasil
QMB22619	MT807936.1	Brasil
QLF80224	MT807936.1	Brasil
QLF80264	MT738173.1	Brasil
QLD32036	MT710714.1	Brasil
QKY74794	MT670013.1	Chile
QKY74806	MT670014.1	Chile
QKY74818	MT670015.1	Chile
QKY74830	MT670016.1	Chile
QKY74842	MT670017.1	Chile
QKY74854	MT670018.1	Chile
QKY74866	MT670019.1	Chile
QKY74878	MT670020.1	Chile
QKY74890	MT670021.1	Chile
QKY74902	MT670022.1	Chile
QKY74914	MT670023.1	Chile
QJT43400	MT466071.1	Uruguay
QJT73054	MT470219.1	Colombia
QJA41649	MT350282.1	Brasil
QIS60296	MT263074.1	Perú
QIS30062	MT256924.2	Colombia
QIG56001	MT126808.1	Brasil

Otra plataforma que se tomó a consideración para el análisis fue GISAID, esta base de datos se encarga del intercambio de datos de virus de diferentes especies, con la finalidad de ayudar a los investigadores a la comprensión de su propagación y evolución. En la tabla 2, se puede apreciar el número de accesión de las secuencias nucleotídicas por país, cabe destacar que en esta plataforma se encuentran las secuencias del genoma, por lo cual se delimito el gen de la proteína N y a través del programa Mega X, se transformó de secuencia nucleotídica a aminoacídica.

Tabla 2.

Números de accesión del GISAID

Argentina	Brasil	Chile	Colombia
EPI_ISL_420598	EPI_ISL_541373	EPI_ISL_445343	EPI_ISL_456127
EPI_ISL_420599	EPI_ISL_541393	EPI_ISL_445362	EPI_ISL_456154
EPI_ISL_430799	EPI_ISL_541395	EPI_ISL_445372	EPI_ISL_498154
EPI_ISL_430802	EPI_ISL_547570	EPI_ISL_445377	EPI_ISL_498168
EPI_ISL_430804	EPI_ISL_547573	EPI_ISL_459861	EPI_ISL_526934
EPI_ISL_430808	EPI_ISL_547578	EPI_ISL_468751	EPI_ISL_526964
EPI_ISL_430815	EPI_ISL_572334	EPI_ISL_468758	EPI_ISL_526965
EPI_ISL_430818	EPI_ISL_574579	EPI_ISL_591531	EPI_ISL_526967
EPI_ISL_476563	EPI_ISL_574594	EPI_ISL_591533	EPI_ISL_536399
EPI_ISL_476571	EPI_ISL_583503	EPI_ISL_591534	
Ecuador	Perú	Venezuela	
EPI_ISL_477016	EPI_ISL_540977	EPI_ISL_476702	
EPI_ISL_491935	EPI_ISL_547910	EPI_ISL_476703	
EPI_ISL_491941	EPI_ISL_547931	EPI_ISL_476704	
EPI_ISL_491945	EPI_ISL_568513		
EPI_ISL_516649	EPI_ISL_568535		
EPI_ISL_525434	EPI_ISL_568545		
EPI_ISL_527811	EPI_ISL_568548		
EPI_ISL_527814	EPI_ISL_568554		
EPI_ISL_539786	EPI_ISL_593773		
EPI_ISL_539792	EPI_ISL_593775		
EPI_ISL_574431			

Para el análisis de las secuencias se utilizó el programa MEGA X, versión 10.1.8, este es un software bioinformático que permite un control manual del alineamiento de secuencias y la generación de árboles filogenéticos(Kumar et al., 2018). En la Figura 6, se puede apreciar el alineamiento de secuencias dentro del programa, como ya se lo mencionó anteriormente el objetivo de esto es observar que cambios aminoacídicos existen entre las secuencias de la región y posteriormente a través de una secuencia consenso establecer si dichos cambios afectan en las regiones inmunogénicas.

Figura 6.

Alineamiento de secuencias en MEGA X

Species/Abbrv	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36																																																				
1. QNH88961.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R	
2. QNH88950.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
3. QNH88939.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
4. QNH88915.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
5. QNH88903.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
6. QNH88911.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
7. QNH88909.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
8. QNH88903.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
9. QMT978148.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
10. QMT97881.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
11. QMT92093.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
12. QMT92017.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
13. QMT92024.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
14. QMT92012.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
15. QMT92058.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
16. QMT92055.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
17. QMB22019.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
18. QLF80204.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
19. QLF80224.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
20. QLD32036.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
21. QKY74914.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
22. QKY74902.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
23. QKY74900.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L	L	L	L	D	R	N	L	Q	L	E	S	K	M	S	G	K	G	Q	Q	Q	G	G	T	V	T	K	K	S	A	A	E	A	S	K	K	P	R	K	R	A	T	K	A	Y	N	V	T	A	F	A	A	R	R
24. QKY74866.1_nucleocapsid_phosphoprotein_Severe_acute_respiratory_syndrome_coronavirus_2	S	S	R	R	N	S	S	R	N	S	T	P	Q	S	S	R	G	T	S	P	A	R	M	A	H	G	G	D	A	A	L	A	L	L	L																																																					

Análisis de Regiones Antigénicas en Antheprot

Para el análisis de las regiones antigénicas de las distintas secuencias aminoacídicas de la proteína N se utilizó el software Antheprot, versión 6.9.3, este es un programa diseñado para el análisis de secuencias proteicas por medio de una interfaz gráfica (G et al., 2001). En nuestro caso se guardó la secuencia a analizar en el formato “.SEQ”, y en la interfaz de antheprot se analizaron todos los perfiles con respecto a Hidrofobicidad, Hidrofobicidad, y accesión al solvente, con la finalidad de identificar las regiones más antigénicas de nuestras secuencias.

Diseño de Primers (Primer Blast, Sequence Extractor, Oligo Analyzer)

Se diseñaron 2 pares de primers, para ello, se utilizó la herramienta Primer-BLAST, utilizando como base la secuencia MN908947.3, variando las opciones del rango de los primer siendo para el Foward del nucleótido 1 al 132, y para Reverse del nucleótido 1092-1260; el tamaño de producto PCR min: 70 y max: 2000; y la temperatura de melting de los primers min: 54°C, Opt:55°C, max: 56°C, con una Tm de diferencia máxima de 2°C. La configuración utilizada en la herramienta se presenta en la figura 7.

Figura 7.

Configuración de PrimerBlast

The screenshot shows the Primer-BLAST web interface. At the top, there are logos for NIH (U.S. National Library of Medicine) and NCBI (National Center for Biotechnology Information), along with a 'Sign in to NCBI' link. The main heading is 'Primer-BLAST' with the subtitle 'A tool for finding specific primers'. Below this, it states 'Finding primers specific to your PCR template (using Primer3 and BLAST)'. The interface is divided into several sections:

- PCR Template:** Contains a text input field for the sequence, a 'Clear' button, and a 'Range' selector with 'From' and 'To' fields. The 'From' field is set to 1 and 'To' to 132. Below this, there are fields for 'Forward primer' (1092) and 'Reverse primer' (1280). There is also an option to 'Or, upload FASTA file' with a 'Seleccionar archivo' button and a 'No se eligió archivo' message.
- Primer Parameters:** Includes fields for 'Use my own forward primer (5'->3' on plus strand)', 'Use my own reverse primer (5'->3' on minus strand)', 'PCR product size' (Min: 70, Max: 2000), '# of primers to return' (4), and 'Primer melting temperatures (Tm)' (Min: 54, Opt: 55, Max: 56, Max Tm difference: 2).
- Exon/intron selection:** Includes a dropdown for 'Exon junction span' (No preference), 'Exon junction match' (Min 5' match: 7, Min 3' match: 4, Max 3' match: 8), and an 'Intron inclusion' checkbox (unchecked) with a note: 'Primer pair must be separated by at least one intron on the corresponding genomic DNA'.

Tras la obtención de primers iniciales, se procedió a ponerlos en fase, con respecto a los aminoácidos de la secuencia MN908947.3, en la página Sequence Extractor (bioinformatics.org/seqext/), aquellos primers que no se encontraban en fase, se modificaron y compararon sus propiedades individuales y con las respectivas parejas, para evitar dímeros, horquillas, y que los pares tengan una Tm similar entre ellos, este análisis se realizó en el Software OligoAnalyzer versión 1.0.2. En la Figura 8, se puede apreciar los primers Forward en fase, y en la Figura 9, los primers Reverse en fase.

Figura 8.

Primers Forward en Fase

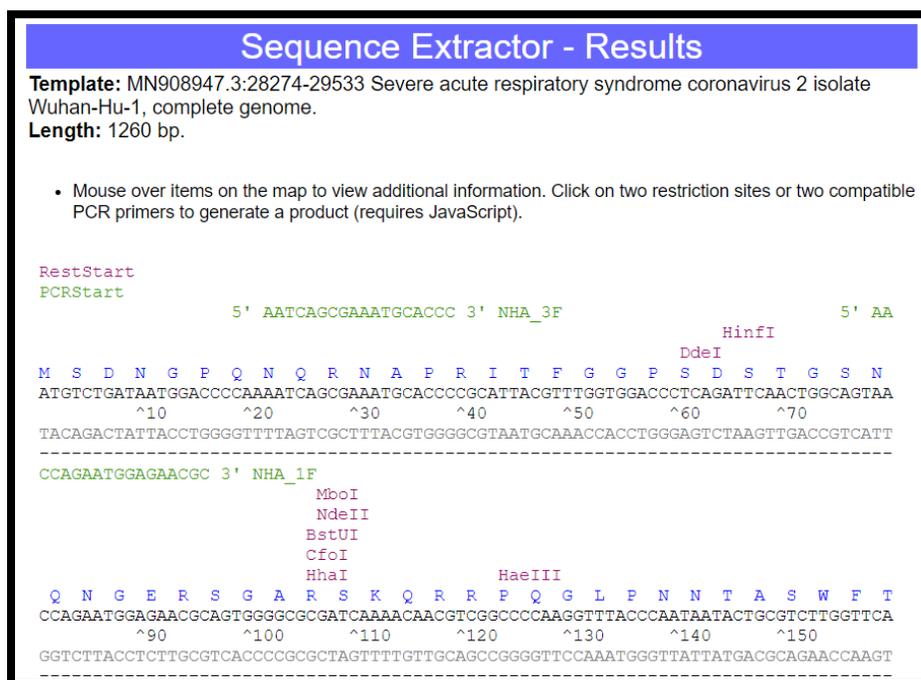
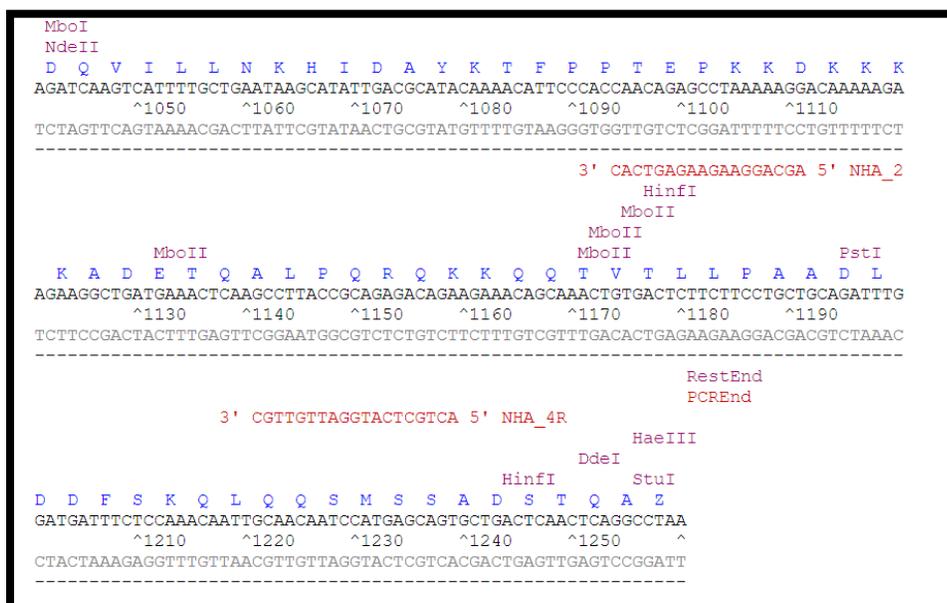


Figura 9.

Primers Reverse en Fase



En la Tabla 3, se presenta un resumen de las propiedades de cada primer. Cabe mencionar que el análisis por pares realizado en OligoAnalyzer, no presento complementariedad de bases y horquillas con un $\Delta G < -2$ kcal/mol. De manera óptima, generalmente se tolera un auto dímero del extremo 3 'con un ΔG de -5 kcal / mol y De manera óptima, se tolera generalmente una horquilla del extremo 3 'con un ΔG de -2 kcal / mol(PREMIER Biosoft, 2021).

Tabla 3.

Propiedades de los pares de Primers

Primer	Tm	% GC	Longitud	Secuencia
NHA_1F	54,93 °C	50 %	18 pb	5'-AACCAGAAATGGAGAACGC-3'
NHA_2 R	53,44 °C	50 %	18pb	5'-AGCAGGAAGAAGAGTCAC-3'
NHA_3F	56.05 °C	50 %	18pb	5'-AATCAGCGAAATGCACCC-3'
NHA_4R	56.19 °C	47,4 %	19pb	5'-ACTGCTCATGGATTGTTGC-3'

Síntesis de material Genético

Producción de cDNA por retro transcripción

Para la obtención de cDNA por retrotranscripción se utilizó RNA SARS-CoV-2 (+), provisto por (Zurita & Zurita Laboratorios, 2021). Para la retrotranscripción (RT) se realizó en tubos eppendorf 200uL y consta de 1 uL Oligo (dT) (18418012, Invitrogen), 1 uL de dNTP Mix 10mM (18427088, Invitrogen), 5uL de agua ultrapura, 5uL de RNA, dicha mezcla se homogenizaría por Spin y se encabaría en bloque de calentamiento por 10 minutos a 65°C, tras la incubación se colocó en hielo por 2 min. A la mezcla se adiciona 4uL de 5X First-Strand Buffer, en conjunto con 2uL 0.1M DTT, nuevamente se homogeneizo por Spin e incubó a 37°C por 2 minutos. Finalmente se agregó 0.8 uL de enzima M-MLV RT (28025021, Thermo Fisher Scientific) y 1.2 uL de agua ultrapura, el proceso finaliza tras una incubación a 37°C por 50 min y una inactivación a 75°C por 15 minutos en bloque de calentamiento. Tras la síntesis es recomendable almacenar a -20°C(Invitrogen, 2015).

RT-PCR one Step Kit

Otro procedimiento que se utilizó para la estandarización de PCR fue el kit “MegaFi™ One-Step RT-PCR” (Cat. No. G597), debido a su alta sensibilidad y especificidad en la retrotranscripción a partir de una simple hebra de RNA(abm, 2021). El mix de reacción y el programa de termociclador tuvieron variaciones con respecto a lo recomendado por el proveedor, esta reacción se utilizó en reacción por gradiente para establecer la temperatura de Annealing óptima para los primers NHA_3F y NHA_4R. En la tabla 4 se puede apreciar el mix de reacción utilizado.

Tabla 4.*Componentes para RT-PCR One Step Kit*

Componentes	Concentración Final	Volumen Final (uL)
H ₂ O	-	3,90
2X One-Step RT-PCR Buffer	1X	7,50
RT-PCR Enzyme Mix	-	1,2
Forward Primer NHA_3F (10 μM)	0.3 μM	0,45
Reverse Primer NHA_4R (10 μM)	0.3 μM	0,45
RNA	1 ng/ μL	2
Volumen Total		15

Para el programa de termociclador que se utilizó: Síntesis de cDNA por 15 min a 60°C, Desnaturalización inicial por 30 seg a 98°C, seguido por 35 ciclos de Desnaturalización a 98°C por 30 seg, Hibridación por 30 seg en las temperaturas (54.3°C-56.4°C-58.7°C), y una Extensión por 35 seg a 72°C; por último, una Extensión final por 2 min a 72°C.

Reacción en cadena de la polimerasa (PCR) convencional

Tras la estandarización de la PCR, se utilizó una PCR convencional con la finalidad de optimizar el material genético utilizando el cDNA, los componentes de reacción son una variación a partir de los recomendados por el fabricante (Invitrogen, 2014) . En la tabla 5, se pueden apreciar los componentes para la PCR.

Tabla 5.*Componentes para PCR convencional*

Componentes	Concentración Final	Volumen Final (µL)
H ₂ O	-	17,13
10X Buffer	1X	2,50
Forward Primer NHA_3F (10 µM)	0.4 µM	1,00
Reverse Primer NHA_4R (10 µM)	0.4 µM	1,00
MgCl ₂	1.5 mM	0,75
dNTPs Mix	0.8 µM	0,50
Taq polimerasa	0.63 U/ µL	0,13
cDNA	1 ng/ µL	2
Volumen Total		25

Para el programa de termociclador que se utilizó: Desnaturalización inicial por 30 seg a 98°C, seguido por 35 ciclos de Desnaturalización a 98°C por 30 seg, Hibridación por 30 seg a 58°C, y una Extensión por 35 seg a 72°C; por último, una Extensión final por 2 min a 72°C.

Electroforesis en gel de agarosa

Los productos PCR obtenidos en los procedimientos fueron analizados en geles de agarosa al 1,5% p/v, dicho gel fue impregnado con Bromuro de etidio en una relación 1/10000 con la finalidad de teñir los amplicones, la cantidad de producto PCR o material genético agregado a cada pocillo para su observación es entre 5 µL a 10 µL y de marcador molecular 3 µL. La electroforesis de los geles de agarosa se realizó a un voltaje de 100V por 45min, y posteriormente se observó en un transluminador de luz UV (SPECTROLINE, Bi-O-Vision™ Series)

Clonación

Clonación al vector pGEM[®]-T Easy Vector

El fragmento de PCR de la proteína N, fue ligado a pGEM[®]-T Easy Vector (Promega), Figura 10, considerando el mix de reacción de la tabla 6, dicha reacción se incubó por 1 hora a temperatura ambiente y posteriormente a 4°C durante toda la noche para incrementar el número de transformantes (Promega, 2018). Dicho vector posteriormente fue transformado en cepas Jm109

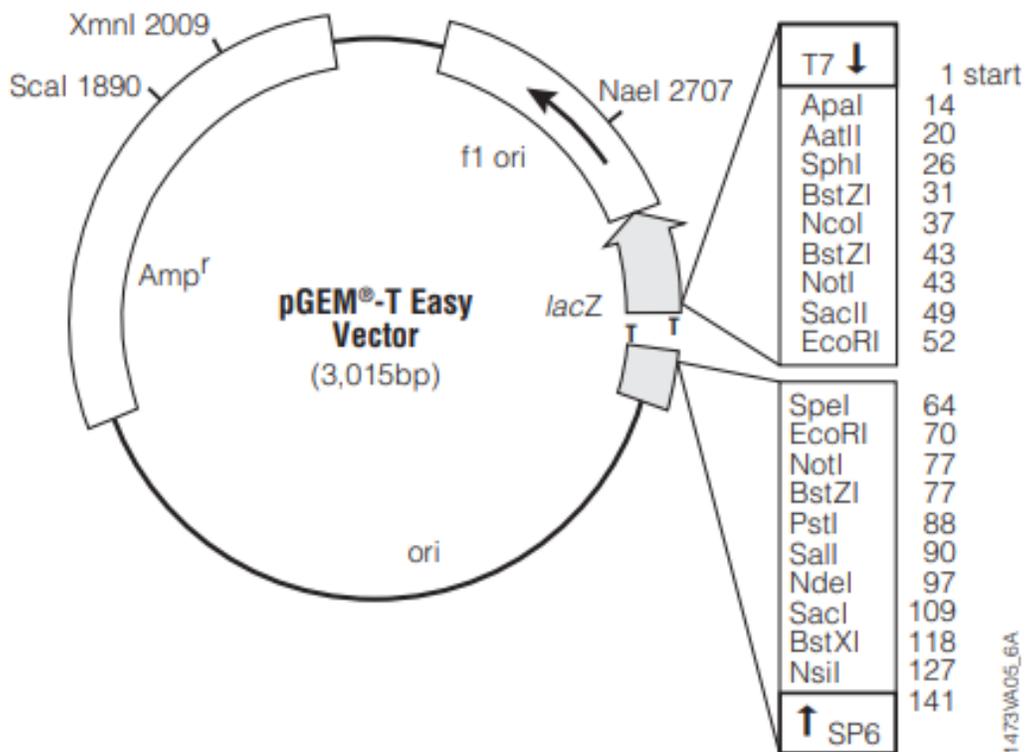
Tabla 6.

Componentes de Reacción para Ligación en vector de Clonación

Componentes	Volumen Final (uL)
H ₂ O	-
2X Rapid Ligation Buffer, T4 DNA Ligase	5,00
pGEM [®] -T Easy Vector (50ng)	1,00
Producto PCR	3,00
Ligasa T4 DNA (3 unidades Weiss/ μL)	1,00
Volumen Total	10,00

Figura 10.

Mapa del vector pGEM[®]-T Easy



Nota. Se puede apreciar un mapa referencial de los genes que posee nuestro vector de clonación, consta específicamente de un gen reportero *lacZ*, para identificar colonias que no poseen nuestro inserto, y un gen *Amp* para proporcionar resistencia a ampicilina a las bacterias que tengan el vector. Tomado de Promega, C. (2018). *pGEM[®]-T and pGEM[®]-T Easy Vector Systems Technical Manual #TM042*

Clonación al vector de expresión pTrcHis2-TOPO

Una vez se puedan generar copias de manera masiva a través de nuestro fragmento de la proteína N clonado en el vector pGEM[®]-T Easy, estos fragmentos se ligaron al vector de expresión pTrcHis2-TOPO (Invitrogen), Figura 11, considerando el mix de reacción de la tabla 7, dicha reacción se mezcló gentilmente e incubó por cinco minutos a temperatura ambiente, es importante recalcar que dicha incubación no debe ser mayor a 5 minutos debido a que la eficiencia de la transformación puede disminuir(Invitrogen, 2006), posterior a la incubación se transformó a cepas JM109.

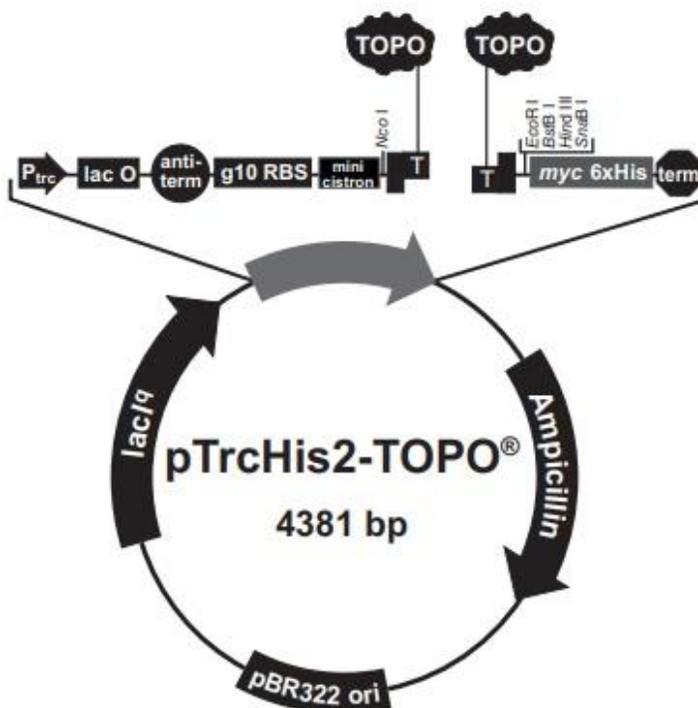
Tabla 7.

Componentes de Reacción para Ligación en vector de Expresión

Componentes	Volumen Final (uL)
H₂O	2,00
Producto PCR	2,00
pTrcHis2-TOPO Vector	1,00
Volumen Total	5,00

Figura 11.

Mapa del vector pTrcHis2-TOPO



Nota. Se puede apreciar un mapa referencial de los genes que posee nuestro vector de expresión, entre las regiones más representativas que podemos apreciar son su gen de resistencia a ampicilina y la región de clonaje TOPO. Tomado de Invitrogen. (2006). pTrcHis and pTrcHis2 TOPO® TA Expression Kits Five-minute cloning of Taq polymerase-amplified PCR products for expression in *E. coli* Catalog no. K4410-01 (pTrcHis TOPO®) Catalog nos. K4400-01 and K4400-40 (pTrcHis2 TOPO®). www.invitrogen.com

Cultivo en Medio Líquido LB y LB Broth Miller

Para la preparación de medio LB (Luria-Bertani), por cada litro se utiliza 10g de Bacto Triptona, 5 g de Bacto- Extracto de Levadura y 5 g de NaCl, dicha solución se ajusta a pH 7.0 con NaOH, y posterior se autoclavó por un ciclo de 121°C por 20 min, de este modo estará listo para cualquier procedimiento estéril (Promega, 2018). Adicionalmente para el cultivo de *E.coli* transformadas con el vector de expresión se utilizó la variante LB Broth Miller, en donde su composición se basa en Digerido enzimático de caseína 10g/L, Extracto de levadura 5g/L y NaCl 10g/L.

Cultivo en Medio LB suplementado con Ampicilina, X-Gal, IPT

Para la clonación se realizaron 3 placas de LB suplementadas con ampicilina y un control de LB. Dichas placas se preparan añadiendo 15g de agar por cada litro de medio LB, tras autoclavar por un ciclo liquido de 121°C por 20 minutos, se procede a dejar enfriar hasta que se encuentre debajo de los 50°C para añadir ampicilina a una concentración final de 100 ug/mL, posteriormente se pone entre 30 a 35ml de medio en placas Petri de 85mm, hay que considerar que previamente se debe separar nuestra placa control con LB, tanto las placas suplementadas como la control se deja almacenada a 37°C toda una noche para descartar posibles contaminaciones. Adicionalmente para la Transformación las placas son suplementadas con 100ul a 100mM de IPTG y 20ul a 50 mg/ml de X-Gal, distribuidas uniformemente por la caja LB-Ampicilina y fue almacenada por 30 min a 37°C previo uso(Promega, 2018).

Células Químicamente Competentes *E.coli JM109*

Se utilizaron células químicamente competentes *E. coli JM109*, alícuotas de 200uL de un solo uso tanto para las transformaciones con los vectores de clonaje y expresión, fueron utilizadas según las recomendaciones del fabricante con ciertas modificaciones(Promega, 2018).

Transformación

Se mantuvo a las células químicamente competentes en hielo, mientras tanto previamente enfriado se agregó 3 uL de ligación a 50uL de células competentes, se mezcló gentilmente y se incubó la mezcla en hielo 30 minutos. A continuación, se incubó en bloque de calentamiento durante 45 segundos a 42°C sin agitación para producir el choque térmico, inmediatamente después se trasladó a hielo y se agregaron 250uL de medio SOC a temperatura ambiente, nuestro eppendorf de 1.5uL con nuestras células transformadas se las incubaron horizontalmente durante 1.5 hora a 300 rpm a 37°C, posteriormente se sembró en placas LBP-Amp e incubaron a 37°C toda una noche.

Colony PCR

Este procedimiento se realizó con la finalidad de determinar si nuestro plásmido de clonación u expresión ligado con el gen e la proteína N, se encuentra dentro de nuestras colonias, para lo cual se utiliza el mismo master mix descrito en la tabla 5, a diferencia que en vez de poner 1 uL de ADN se integra una porción de colonia directamente a la mezcla y el programa de PCR consta de: Desnaturalización inicial por 5min a 95°C, seguido por 35 ciclos de Desnaturalización a 98°C por 30 seg, Hibridación por 30 seg a 58°C, y una Extensión por 35 seg a 72°C; por ultimo una Extensión final por 2 min a 72°C.

Extracción de Plásmido

Para extraer el plásmido, previamente se cultivó las colonias en 10ml de medio LB para cepas de clonaje, y LB Broth Miller para cepas de expresión, durante toda una noche. Posterior se tomó 1.5ml de cultivo y se extrajo el plásmido según el protocolo de “Preparación de DNA Plasmídico por Lisis Alcalina con SDS: Mini preparación” (Sambrook & Russell, 2001) , con ciertas modificaciones descritas a continuación. Centrifugar las colonias durante 1 minuto a 10.000 xg y descartar el medio, posteriormente, se le realiza un tratamiento con STE, en donde se adiciona 400uL de la solución, se resuspende el pellet de la bacteria y se vuelve a centrifugar por 1 min a velocidad máxima, posterior se removió la solución y todas las posibles trazas, debido a que esto puede afectar en procedimientos posteriores como corte con enzimas o PCR, por siguiente se adicionó 100uL de la solución I, se resuspende la bacteria e inmediatamente se adiciono 200uL de la solución II, se mezcló gentilmente por inversión 5 veces, se mantuvo el tubo en hielo, y a continuación se agregó 150 uL de la solución III, se mezcló gentilmente por inversión 5 veces y se dejó reposar durante cinco minutos en hielo.

A continuación, se centrifugo por 5 minutos a 10.000 xg nuevamente y el sobrenadante se transfirió a un tubo limpio y se precipito los ácidos nucleicos con dos volúmenes de etanol absoluto, en este paso se observaron surcos en la mezcla lo cual advierte que existe presencia de ácidos nucleicos, dicha mezcla se dejó reposar por 5 min a -20°C y posterior se volvió a centrifugar a máxima velocidad por dos minutos, de manera sucesiva se removió el sobrenadante, teniendo cuidado de no remover el pellet y se dejó secar a 65°C en una incubadora para eliminar cualquier rastro de etanol, a continuación se añadió 1 ml de etanol 70% e invirtió el tubo varias veces, y se centrifugó durante 2

minutos, nuevamente se remueve el etanol y se lo dejó secar a 65°C por 15 min. El plásmido fue reconstituido en agua con RNAsa a 20 µg/ml, la presencia del plásmido fue verificado en un gel de agarosa al 0.8%.

Corte con enzimas de Restricción

Para la digestión del plásmido de clonación se utilizó la enzima EcoRI (Thermo Fisher Scientific Inc, 2012), debido a que permite liberar del plásmido a nuestro gen y adicionalmente realiza un corte en el nucleótido 278 de la secuencia de la proteína N, la reacción de la digestión se presenta en la tabla 8.

Tabla 8.

Componentes de Digestión del vector de clonación

Componentes	Volumen Final (uL)
H₂O	15,00
10X Buffer EcoRI	2,00
AND plasmídico	2,00
Enzima EcoRI	1,00
Volumen Total	20,00

Para la digestión del plásmido de expresión se utilizó la enzima EcoRI (Promega, 2011), debido a que el tiempo de incubación es de 15 min, y genera un corte en el nucleótido 278 y en el vector, lo que nos permite observar si nuestro gen se encuentra en la dirección correcta para su posterior traducción, los componentes de la digestión se presentan en la tabla 9.

Tabla 9.*Componentes de digestión del vector de expresión*

Componentes	Volumen Final (uL)
H ₂ O	14,80
10X Buffer EcoRI	2,00
BSA	0,20
AND plasmídico	2,00
Enzima EcoRI	1,00
Volumen Total	20,00

Generación de Modelos Bioinformáticos

Para el modelamiento de la proteína expresada se utilizó la plataforma SwissModel (Waterhouse et al., 2018) en conjunto con la herramienta PyMol versión 4.60. Los PDB utilizados para la generación del modelo se pueden encontrar en el Protein Data Bank con los números de identificación “6zco.1.A”, “7act.1”, “7pku.1.B”.

Análisis Estadístico**Diseño no experimental**

La presente investigación tiene la finalidad de la amplificación del gen de la proteína N para su posterior clonación a través de herramientas de biología molecular y análisis bioinformático. Adicionalmente se realizó una secuenciación para generar un árbol filogenético y comparar nuestro gen con respecto a la nación, la región latinoamericana y las distintas variantes del SARS-CoV-2.

Variable respuesta

Las variables respuesta de la investigación son las amplificaciones del gen de la proteína N, con verificación por medio de secuenciación y para el caso de la clonación la presencia de amplificación del gen a través de una Colony PCR y corte por enzimas de restricción.

Error aleatorio

El error procedente en esta investigación puede ser generado por calidad y estado de reactivos utilizados, así como errores de operación.

Secuenciación

Para la secuenciación de nuestra amplificación de la proteína N se utilizó el servicio de secuenciación tipo Sanger de la Universidad de las Américas (UDLA), en donde utilizan la preparación química BigDye V3.1, para la purificación de productos PCR.

Análisis Filogenético

La generación del árbol filogenético se generó en el programa Mega-X, versión 10.1.8, por el método de Maximum Likelihood con una prueba filogenética de tipo “Bootstrap method” con 1000 replicaciones y un modelo de sustitución “Tamura-Nei model”

Capítulo IV: Resultados

Identificación de la secuencia de la proteína N

El alineamiento de las secuencias aminoacídicas tanto las de NCBI-Virus como de la base de datos de GSAID, permitió generar una secuencia consenso la cual es similar a la secuencia de la traducción del aislado de la proteína N, del SARS CoV-2 aislado Wuhan-Hu-1, con número de acceso de la proteína QHD43423 y nucleotídica MN908947.3, y consta del 28274-29533 pb con respecto al genoma completo.

De este modo se determinaron que las regiones más antigénicas o mejores candidatos para epítomos, los cuales se pueden observar en la tabla 10, por otra parte, el perfil de predicción de Antheprot se muestra en la figura 12.

Tabla 10.

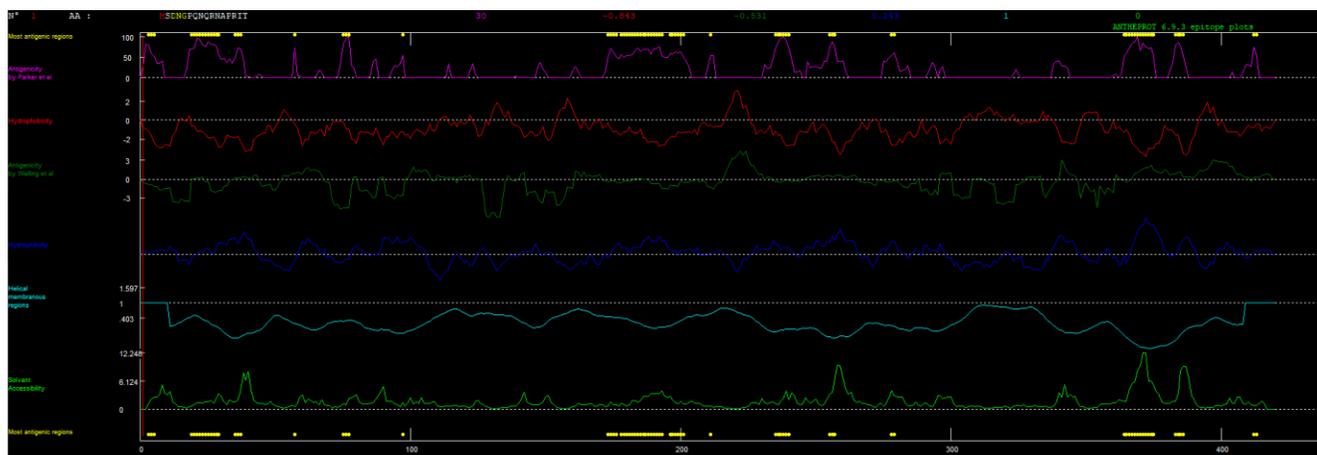
Mejores candidatos para Epítomos

Numero de aminoácido inicial	Aminoácidos pertenecientes al Epítome	Número del aminoácido final
3	DNG	5
19	GPSDSTGSNQN	29
35	ARS	37
57	T	57
75	NTN	77
97	G	97
173	AEGS	176
178	GGSQASSRSSSRNS	193
196	NSTPGS	201
211	A	211
235	SGKGQQ	240
255	SKK	257
278	GP	279
364	PPTEPKDKKKK	375
383	PQRQ	386
412	SS	413

Nota. El programa Antheprot muestra los 16 mejores candidatos para epítomos de la proteína N considerando los perfiles de Hidrofobicidad, Hidrofilicidad y acceso al solvente.

Figura 12.

Perfiles de Antheprot para la selección de epitopes



Nota. El programa Antheprot muestra los 16 mejores candidatos para epitopes de la proteína N considerando los perfiles de Hidrofobicidad (perfil rojo), Hidrofilicidad (perfil azul) y accesión al solvente (perfil verde).

Los cambios aminoacídicos que se encontraron en la secuencia proteica de la proteína N fueron: R191C, S197L, R203K, G204R, R209I, G238C, I292T, A313S, S413I. Sin embargo dichos cambios en su mayoría no presentaron una variación de las regiones antigénicas con respecto a la consensu, a excepción de la variación S197L, que generó una pérdida de las regiones antigénicas de los aminoácidos 196-197; por otra parte una variación mayor se presenta con el cambio G238C en donde los aminoácidos 235-236 ya no son antigénicos y se generan nuevas porciones antigénicas en los aminoácidos 194-195, 211 y 277-280, en la figura 13, se puede apreciar una representación gráfica de lo descrito, adicionalmente se presenta en la tabla 11, los cambios presentados dentro de la región latinoamericana según el análisis de las secuencias GISAID.

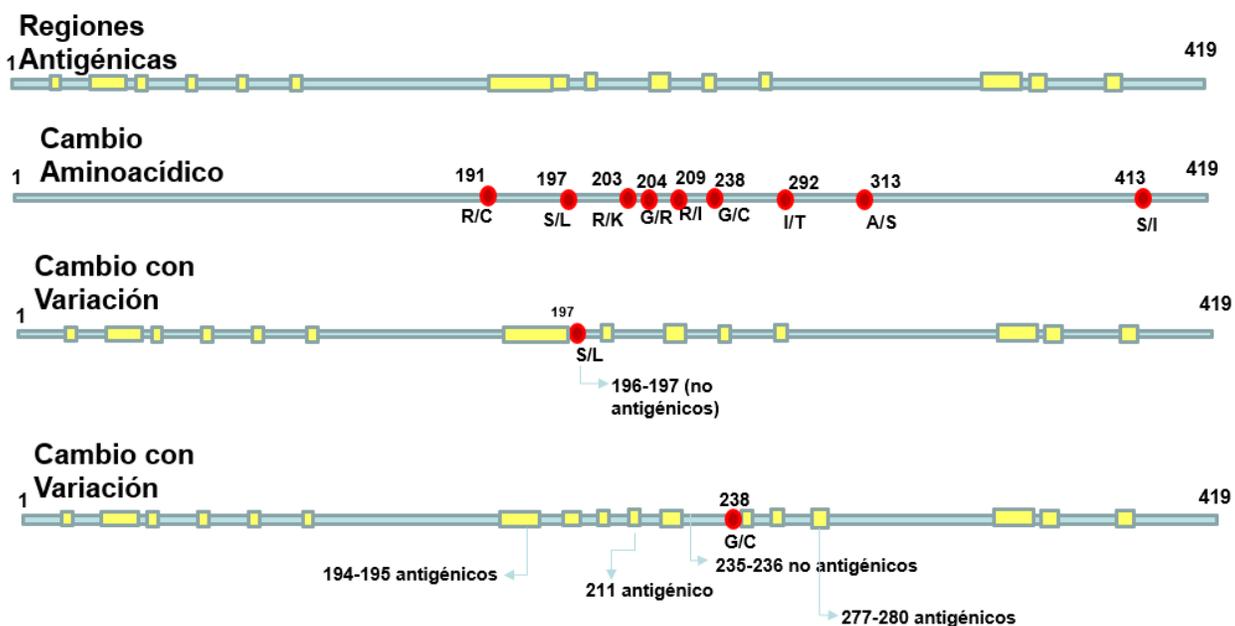
Tabla 11.*Cambios aminoacídicos de la proteína N presentes en la región latinoamericana*

Numero de secuencias analizadas	País	Posición	Aminoácido Base / Aminoácido de Cambio	Presencia del cambio en secuencias analizadas
n=10	Ecuador	203	R / K	8
		204	G / R	8
n=3	Venezuela	203	R / K	1
		204	G / R	1
n=10	Brasil	203	R / K	9
		204	G / R	9
		292	I / T	2
		413	S/I	1
n=10	Colombia	203	R / K	2
		204	G / R	2
		209	R/I	1
n=10	Chile	203	R / K	6
		204	G / R	6
		292	I / T	2
		292	I/V	2
n=9	Perú	203	R / K	8
		204	G / R	8
n=10	Argentina	191	R / C	1
		197	S/ L	3
		203	R/ K	4
		204	G/ R	4
		292	I / T	1

Nota. En la presente tabla se muestra el #n secuencias analizadas, el sitio de secuenciación según GISAID. El número de aminoácido donde existe variación. El aminoácido perteneciente a la traducción de MN908947.3:28274-29533, y el aminoácido al cual cambio. Por último, la cantidad de veces que se dio esta variación de la #n cantidad de secuencias.

Figura 13.

Representación de cambios aminoacídicos de la Proteína N



Nota. Representación de cambios aminoacídicos en la proteína N. La línea celeste representa la proteína N, las regiones amarillas representan las porciones antigénicas y los puntos rojos representan los cambios existentes dentro de la proteína.

Amplificación de la proteína N

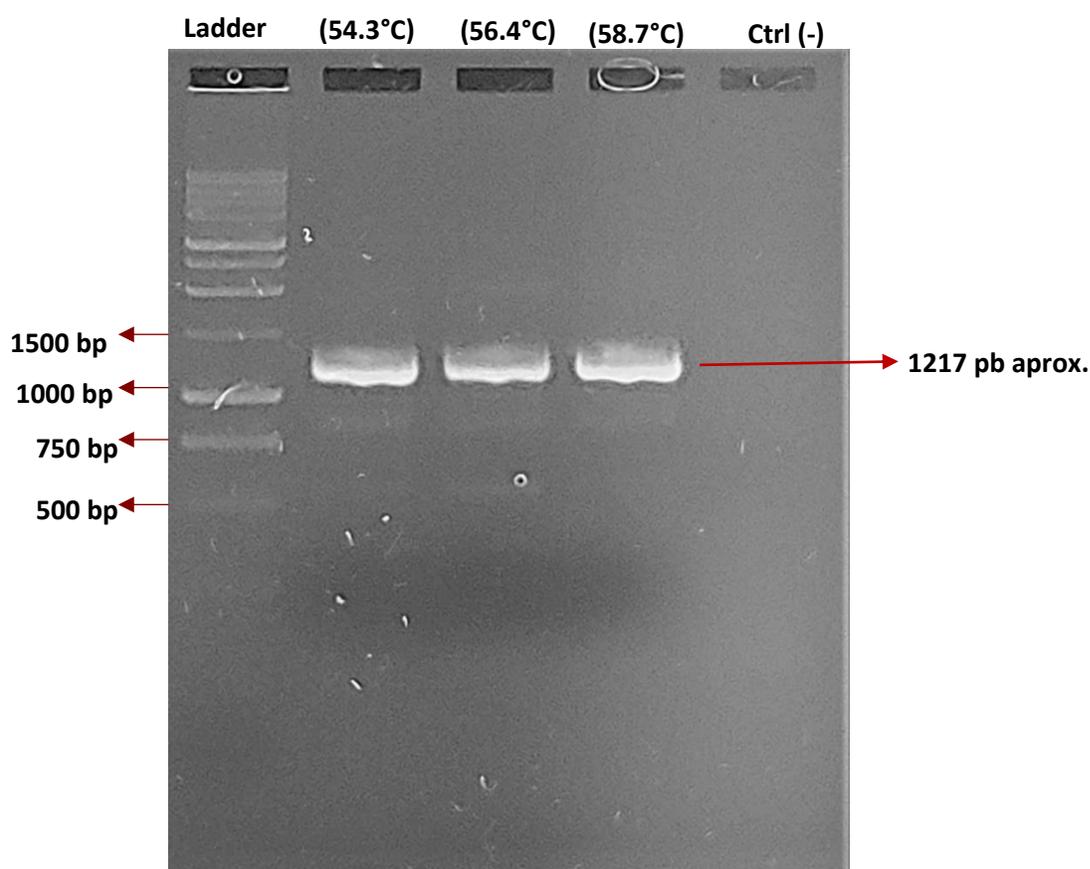
Estandarización de PCR

A través del kit OneStep RT-PCR se realizó un gradiente de temperatura para establecer la mejor condición para un amplicón del gen N bien definida y sin bandas inespecíficas, esto resultó en determinar que con los primers NHA_3F-NHA_4R, la mejor temperatura de melting con el programa de termociclador determinado se encuentra a 58°C, en la Figura 14 se puede apreciar que en dicha temperatura no existen bandas

inespecíficas y se encuentra bien definida la amplificación, además se aprecia que el tamaño de nuestro gen se encuentra entre los 1000pb a 1500pb, considerando la determinación teórica a partir del gen, debería constar aproximadamente de 1217pb, por lo que entraría dentro de los parámetros deseados.

Figura 14.

Gradiente de temperatura del gen de la proteína N



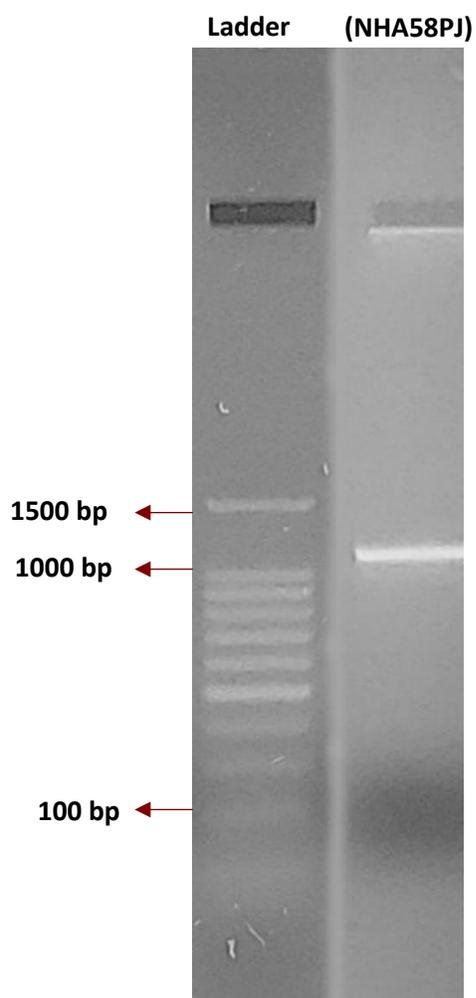
Nota Gradiente de proteína N.- En este experimento se amplifico un fragmento de aproximadamente 1200 pb a diferentes temperaturas de gradiente (NHA3F-NHA4R primers utilizados). Carril 1 pertenece al leader de peso molecular, Carril 2 amplificación del gen N a una temperatura de 54.3°C, Carril 3 amplificación del gen N a una temperatura de 56.3°C, Carril 4 amplificación del gen N a una temperatura de 58.7°C, Carril 5 pertenece al control negativo. Nótese que el carril 5 presenta la mejor banda con pocos productos inespecíficos. Dicho gel de agarosa se encuentra al 1.5%

PCR convencional

Tras la determinación de las condiciones ideales se realizó una amplificación del gen N a partir de cDNA, del mismo modo en la figura 14 se puede apreciar una banda bien definida sin inespecificidad que consta entre los 1000pb a 1500pb, dicha amplificación fue denominada NHA58PJ y fue utilizada para la secuenciación.

Figura 1.

Amplificación del Gen N por PCR convencional



Nota Amplicón de proteína N.- En este ensayo se amplificó un fragmento de aproximadamente 1200 pb en el primer carril, (NHA3F-NHA4R primers utilizados). Nótese que el carril 1 presenta una fuerte banda sin productos inespecíficos. Dicho gel de agarosa se encuentra al 1.5%.

Secuenciación

A través del programa Mega-X se realizó un ensamblaje para determinar la secuencia de NHA58PJ, resultando en una secuencia de 1128pb, para comprobar que dicha secuencia pertenecía al organismo SARS-CoV-2, se procedió a utilizar la herramienta blastn, la cual nos presentó una identidad de 100% con 100 secuencias pertenecientes a la proteína N indexadas al GenBank. A continuación, se presenta la secuenciación de NHA58PJ:

>NHA58PJ

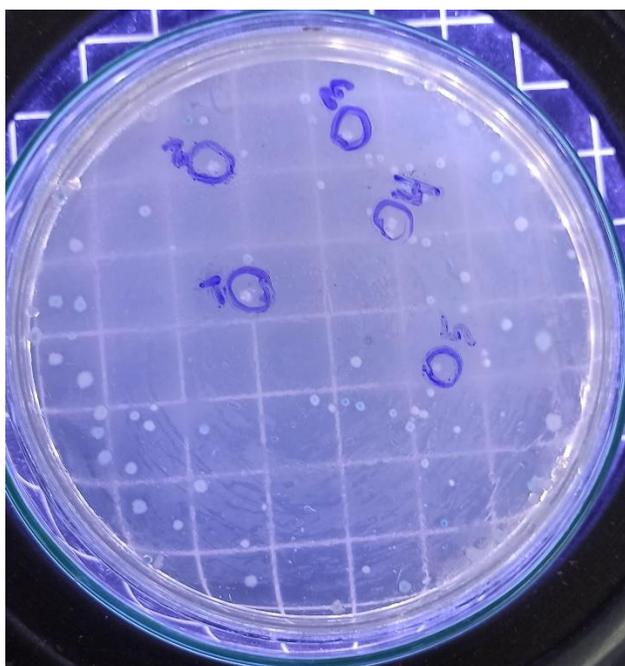
```
CTGGCAGTAACCAGAATGGAGAACGCAGTGGGGCGCGATCAAAAACAACGTCGGCCCC
AAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGGAA
GACCTTAAATTCCCTCGAGGACAAGGCGTTCCAATTAACACCAATAGCAGTCCAGATG
ACCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGTAAAAT
GAAAGATCTCAGTCCAAGATGGTATTTCTACTACCTAGGAACTGGGCCAGAAGCTGGA
CTTCCCTATGGTGCTAACAAAGACGGCATCATATGGGTTGCAACTGAGGGAGCCTTGA
ATACACCAAAAAGATCACATTGGCACCCGCAATCCTGCTAACAAATGCTGCAATCGTGCTA
CAACTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAGGC
GGCAGTCAAGCCTTCTCGTTCCTCATCACGTAGTCGCAACAGTTCAAGAAATCAAC
TCCAGGCAGCAGTAAACGAACTTCTCCTGCTAGAATGGCTGGCAATGGCGGTGATGCT
GCTCTTGCTTTGCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTCTGGTA
AAGGCCAACAAACAAGGCCAAACTGTCACTAAGAAATCTGCTGCTGAGGCTTCTAA
GAAGCCTCGGCAAAAACGTAAGTCCACTAAAGCATACAATGTAACACAAGCTTTTCGGC
AGACGTGGTCCAGAACAACCCAAGGAAATTTTGGGGACCAGGAACTAATCAGACAA
GGAAGTATTACAAACATTGGCCGCAAATTGCACAATTTGCCCCAGCGCTTCAGCGTT
CTTCGGAATGTCGCGCATTGGCATGGAAGTCACACCTTCGGGAACGTGGTTGACCTAC
ACAGGTGCCATCAAATTGGATGACAAAGATCCAAATTTCAAAGATCAAGTCATTTTGCT
GAATAAGCATATTGACGCATACAAAACATTCCCACCAACAGAGCCTAAAAGGACAAA
AAGAAGAAGGCTGATGAAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAAC
GTGACTCTTCTCCTGCTGCAGATT
```

Clonación del gen de la Proteína N

Tras verificar que el gen que se está amplificando pertenece a la proteína N, se procedió a ligar el fragmento al vector pGEM®-T Easy a través de la metodología antes descrita, hay que destacar que el fragmento ligado no pertenece al aislado de la amplificación de la secuencia NHA58PJ, sino a un aislado de la muestra 190, las secuencias pertenecientes se observaran posteriormente al secuenciar el gen insertado en los clones, figura 15, los clones fueron verificados a partir de la técnica colony PCR y corte por enzimas de restricción

Figura 15.

Múltiples clones del gen N- E. coli Jm109



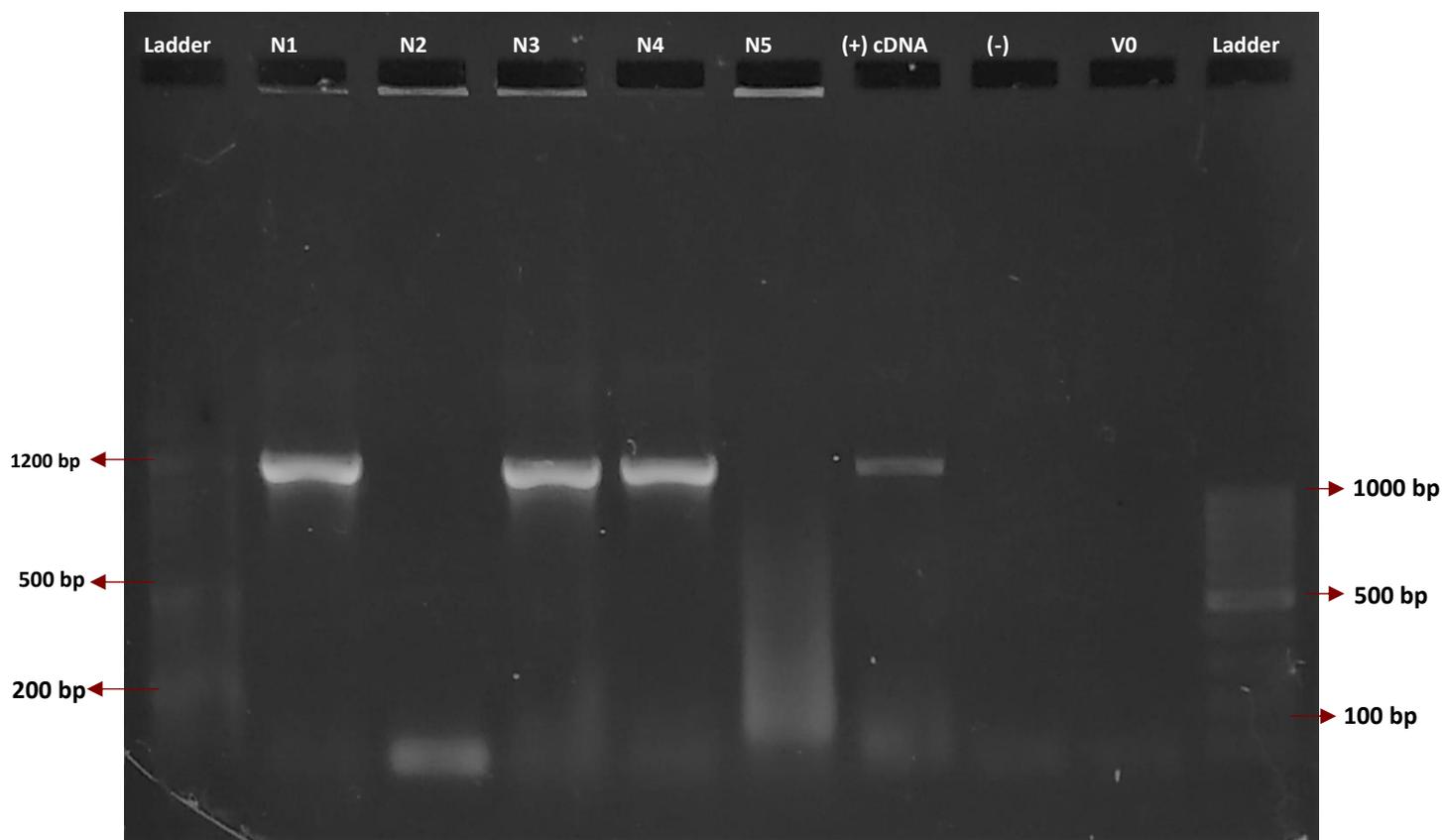
Nota Diferenciación de colonias.- Nótese la presencia de colonias azules y blancas alrededor de toda la superficie de la placa, hay que tener a consideración que las colonias azules representan el vector cerrado sin ningún inserto, la tonalidad azul es debido a la expresión del gen LacZ inducido por IPTG presente en la composición del medio, adicionalmente las colonias blancas representan vectores con un inserto, debido a que no se realizó una purificación al producto PCR, no necesariamente todas las colonias blancas pueden considerarse clones con nuestro inserto.

Screening de colonias a través de Colony PCR

A través de una PCR convencional se realizó una colony PCR de manera descrita anteriormente, el resultado fue la obtención de 4 colonias positivas, figura 16, es decir poseían el inserto deseado, las ampliaciones observadas pueden contrastarse con el control positivo, de la muestra 109, este se generó a partir del cDNA que se ligó al vector, es de destacar que dichos amplicones muestran una mayor intensidad con respecto al control.

Figura 16.

Amplificación del gen N de colonias transformadas con pGEM®-T Easy



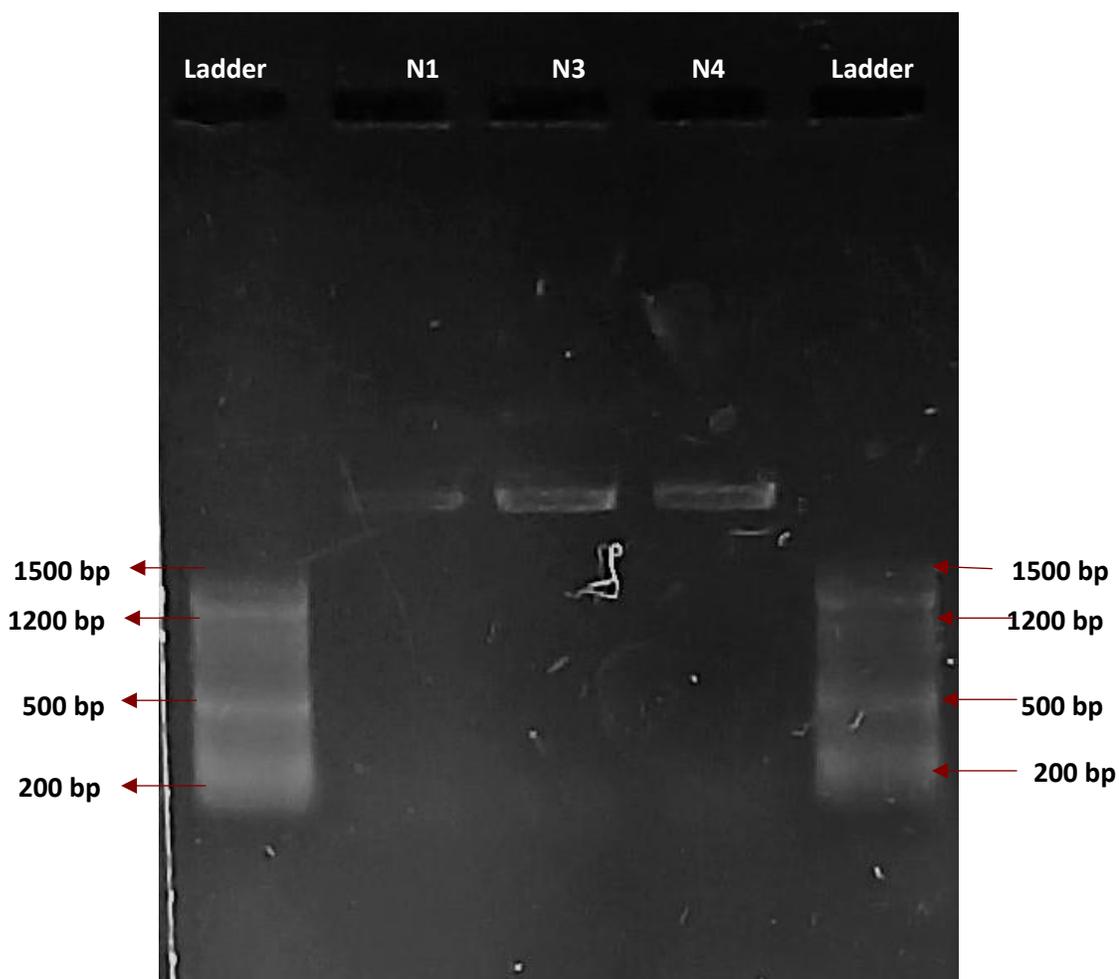
Nota Amplificación del gen N de diferentes Colonias Transformadas. - Se obtuvo presencia del gen N en las colonias 1- 3 -4, estos amplicones corresponden a los tamaños correspondientes y mantienen un peso similar al del control positivo generado a partir de cDNA. Dicho gel de agarosa se encuentra al 1.5%

DNA plasmídico de cepas de clonación

Se obtuvo DNA plasmídico de las cepas N1, N3 y N4, debido a la calidad del material genético, como se puede apreciar en la figura 17, tras su amplificación por PCR se realizó un corte a dichos fragmentos, lo que permitieron dilucidar de manera más precisa que se trataba de nuestro gen.

Figura 17.

DNA plasmídico de cepas positivas



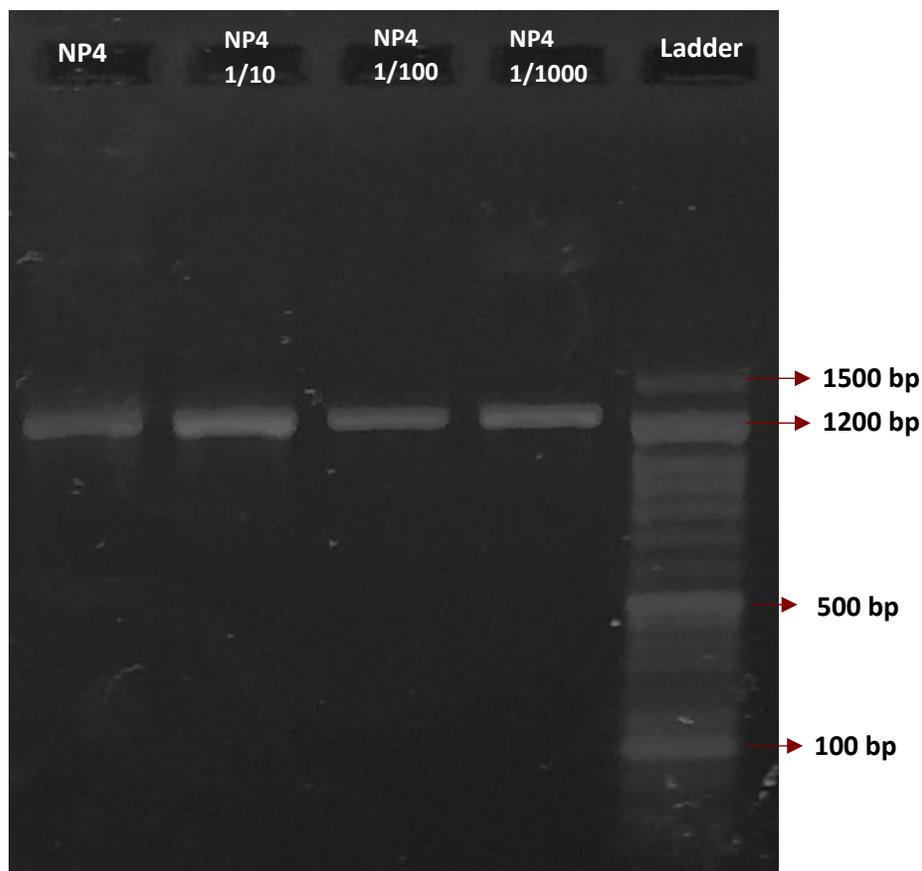
Nota Plásmidos obtenidos por Miniprep. – Se obtuvieron los plásmidos de las cepas N1, N2, N3; estas presentan un tamaño mayor a 1500bp . Dicho gel de agarosa se encuentra al 0.8%

Amplificación del gen N a partir de DNA plasmídico

Tras aislar los plásmidos se procedió a realizar diluciones de estos a concentraciones de 1/10, 1/100 y 1/1000, esto con la finalidad de tener gran cantidad de material genético amplificable por PCR a partir de concentraciones pequeñas. Adicionalmente esto permite generar amplicones mucho más puros eliminando fragmentos inespecíficos, que pueden interferir en futuros procedimientos de clonación e incluso de secuenciación. En la figura 18, se puede apreciar que a medida que disminuye la concentración del plásmido NP4, extraído de la cepa N4, se incrementa la definición del fragmento amplificado siendo la mejor dilución 1/1000, en donde no se presentan fragmentos inespecíficos. Dicho procedimiento también se realizó con los plásmidos NP1 y NP3, extraídos de las cepas N1 y N3 respectivamente, del mismo modo las diluciones 1/1000 fueron las más definidas y sin impurezas, dichos productos fueron secuenciados para comprobar nuevamente que los genes insertados en nuestro vector de clonación pertenecen a la proteína N.

Figura 18.

PCR de la proteína N a distintas concentraciones del DNA plasmídico



Nota PCR de plásmido NP4 a distintas concentraciones. – Se puede apreciar que la PCR realizada al plásmido NP4 a distintas concentraciones permite disminuir fragmentos inespecíficos existentes en el producto PCR. Dicho gel de agarosa se encuentra al 1.5%

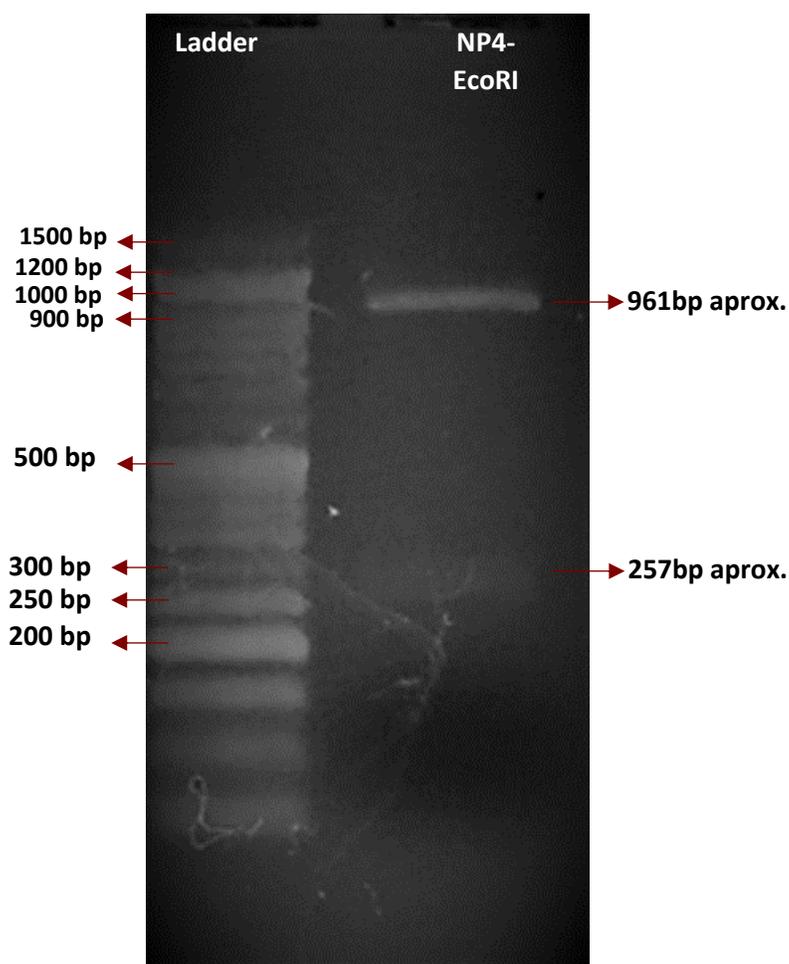
Análisis por corte con enzimas de restricción al vector de clonación

Utilizando amplificaciones a partir del plásmido NP4 se realizó un corte con la enzima EcoRI, figura 19, dicha reacción libera nuestro fragmento del plásmido y nos proporcionó en el gen un corte produciendo dos fragmentos, uno poco perceptible con respecto al estándar de entre 300-250bp, este fragmento debe contener teóricamente alrededor de 257bp, por lo cual entra dentro del rango observable, y el fragmento

restante de manera teórica debe ser de 961bp, que con respecto al estándar de peso también entra dentro del rango apreciable. Hay que recalcar de dicho corte, nos dio una noción de que el gen insertado en nuestro vector y el amplificado, se tratan del gen de la proteína N sin embargo, para corroborar de manera más eficaz la identidad de nuestra secuencia se procedió a secuenciar los amplicones producidos de los plásmidos NP1, NP3 y NP4, dichas secuencias se presentan posteriormente.

Figura 19.

Digestión de NP4 con EcoRI



Nota Digestión Enzimática del gen N. – Considerando las posiciones tóricas de corte de la enzima EcoRI, se puede apreciar que los fragmentos observables se encuentran dentro del peso molecular, siendo los mismos un fragmento de 257bp aprox., y 961bp. El presente gel de agarosa se encuentra al 1.5%

Secuenciación de Insertos

Aunque la digestión enzimática nos permite tener una noción con respecto al gen con el cual estamos trabajando, fue necesario realizar una secuenciación para verificar que el gen insertado perteneciera al de la proteína N. Efectivamente las secuencias insertadas en los plásmidos NP1, NP3 y NP4 poseen una identidad del 100% entre ellas, y a su vez a través de la herramienta blastn, presentó una identidad de 100% con 100 secuencias pertenecientes a la proteína N indexadas al GenBank. A Continuación, se presentan las secuencias generadas tras el ensamblaje realizado en Mega-X.

>NP1

```
ATACTGCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGGAAGACCTTAAATCCCT
CGAGGACAAGGCGTTCCAATTAACACCAATAGCAGTCCAGATGACCAAATTGGCTACT
ACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGTAAAATGAAAGATCTCAGTC
CAAGATGGTATTTCTACTACCTAGGAAGTGGGCCAGAAGCTGGACTTCCCTATGGTGC
TAACAAAGACGGCATCATATGGGTTGCAACTGAGGGAGCCTTGAATACACAAAAGA
TCACATTGGCACCCGCAATCCTGCTAACAAATGCTGCAATCGTGCTACAATTCCTCAAG
GAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAGGCGGCAGTCAAGCCT
CTTCTCGTTCCTCATCACGTAGTCGCAACAGTTCAAGAAATTCAACTCCAGGCAGCAGT
AAACGAACTTCTCCTGCTAGAATGGCTGGCAATGGCGGTGATGCTGCTTTGCTTTGCT
GCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTCTGGTAAAGGCCAACAACA
ACAAGGCCAACTGTCACTAAGAAATCTGCTGCTGAGGCTTCTAAGAAGCCTCGGCAA
AAACGTAAGTCCACTAAAGCATACAATGTAACACAAGCTTTCGGCAGACGTGGTCCAG
AACAAACCCAAGGAAATTTGGGGACCAGGAACTAATCAGACAAGGAACTGATTACA
AACATTGGCCGAAATTGCACAATTTGCCCCAGCGCTTCAGCGTTCTTCGGAATGTCG
CGCATTGGCATGGAAGTCACACCTTCGGGAACGTGGTTGACCTACACAGGTGCCATCA
AATTGGATGACAAAGATCCAAATTTCAAAGATCAAGTCATTTTGCTGAATAAGCATATT
GACGCATACAAAACATTCCACCAACAGAGCCTAAAAAGGACAAAAAGAAGAAGGCT
GATGAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAAGTGTGACTCT
```

>NP3

ATACTGCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGGAAGACCTTAAATCCCT
CGAGGACAAGGCGTTCCAATTAACACCAATAGCAGTCCAGATGACCAAATTGGCTACT
ACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGTAAAATGAAAGATCTCAGTC
CAAGATGGTATTTCTACTACCTAGGAACTGGGCCAGAAGCTGGACTTCCCTATGGTGC
TAACAAAGACGGCATCATATGGGTTGCAACTGAGGGAGCCTTGAATACACAAAAGA
TCACATTGGCACCCGCAATCCTGCTAACAATGCTGCAATCGTGCTACAACTTCTCAAG
GAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAGGCGGCAGTCAAGCCT
CTTCTCGTTCCTCATCACGTAGTCGCAACAGTTCAAGAAATCAACTCCAGGCAGCAGT
AAACGAACTTCTCCTGCTAGAATGGCTGGCAATGGCGGTGATGCTGCTTTGCTTTGCT
GCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTCTGGTAAAGGCCAACACA
ACAAGGCCAACTGTACTAAGAAATCTGCTGCTGAGGCTTCTAAGAAGCCTCGGCAA
AAACGTAAGTCCAACTAAAGCATAACAATGTAACACAAGCTTTCGGCAGACGTGGTCCAG
AACAAACCCAAGGAAATTTGGGGACCAGGAACTAATCAGACAAGGAACTGATTACA
AACATTGGCCGCAAATTGCACAATTTGCCCCAGCGCTTCAGCGTTCTTCGGAATGTGCG
CGCATTGGCATGGAAGTCACACCTTCGGGAACGTGGTTGACCTACACAGGTGCCATCA
AATTGGATGACAAAGATCCAAATTTCAAAGATCAAGTCATTTTGCTGAATAAGCATATT
GACGCATACAAAACATTCCCACCAACAGAGCCTAAAAAGGACAAAAAGAAGAAGGCT
GATGAAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAACCTGTGACTCT

>NP4

GGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGGAAGA
CCTTAAATCCCTCGAGGACAAGGCGTTCCAATTAACACCAATAGCAGTCCAGATGAC
CAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGTAAAATG
AAAGATCTCAGTCCAAGATGGTATTTCTACTACCTAGGAACTGGGCCAGAAGCTGGAC
TCCCTATGGTGCTAACAAGACGGCATCATATGGGTTGCAACTGAGGGAGCCTTCAA
TACACAAAAGATCACATTGGCACCCGCAATCCTGCTAACAATGCTGCAATCGTGCTAC
AACTTCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAGGCG
GCAGTCAAGCCTTTCTCGTTCCTCATCACGTAGTCGCAACAGTTCAAGAAATCAACT
CCAGGCAGCAGTAAACGAACTTCTCCTGCTAGAATGGCTGGCAATGGCGGTGATGCT
GCTCTTGCTTTGCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTCTGGTA
AAGGCCAACACAACAAGGCCAACTGTACTAAGAAATCTGCTGCTGAGGCTTCTAA
GAAGCCTCGGCAAAAACGTACTGCCACTAAAGCATAACAATGTAACACAAGCTTTCGGC
AGACGTGGTCCAGAACAACCCAAGGAAATTTGGGGACCAGGAACTAATCAGACAA
GGAAGTATTACAAACATTGGCCGCAAATTGCACAATTTGCCCCAGCGCTTCAGCGTT
CTTCGGAATGTGCGCATTGGCATGGAAGTCACACCTTCGGGAACGTGGTTGACCTAC
ACAGGTGCCATCAAATTGGATGACAAAGATCCAAATTTCAAAGATCAAGTCATTTTGCT
GAATAAGCATATTGACGCATACAAAACATTCCCACCAACAGAGCCTAAAAAGGACAAA
AAGAAGAAGGCTGATGAAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAACCT
GTGACTCT

El alineamiento de las secuencias NHA58PJ, NP1, NP3 y NP4 con respecto al aislado Wuhan-Hu-1-MN908947.3, presentan ciertas diferencias y similitudes. Entre ellas podemos destacar que el nucleótido 704 posee una “C” en el aislado MN908947.3, el cual comparte con nuestras insertadas en el vector pertenecientes a NP1, NP3 y NP4; sin embargo, esta similitud no la comparten con nuestro primer aislado NHA58PJ el cual posee una “T” en la posición de dicho nucleótido, esto genera un cambio aminoacídico S235F, en donde la serina se traduciría en NP1, NP3, NP4 y MN908947.3, mientras que la fenilalanina pertenecería a NHA58PJ. Sin embargo, la diferencia más notoria son los aminoácidos del 608 al 610 en donde “GGG” pertenecientes a la secuencia MN908947.3, son reemplazados en los aislados ecuatorianos por los aminoácidos “AAC”, esto genera un cambio aminoacídico en dos posiciones con respecto a la secuencia de Wuhan los cuales son R203K y G204R, estas diferencias se han reportado previamente en aislados ecuatorianos y aislados de otros países de Latinoamérica como se observó previamente en el análisis *in-silico*. Considerando estos resultados se utilizó el vector NP4 1/1000, para futuros procedimientos de clonación al vector de expresión.

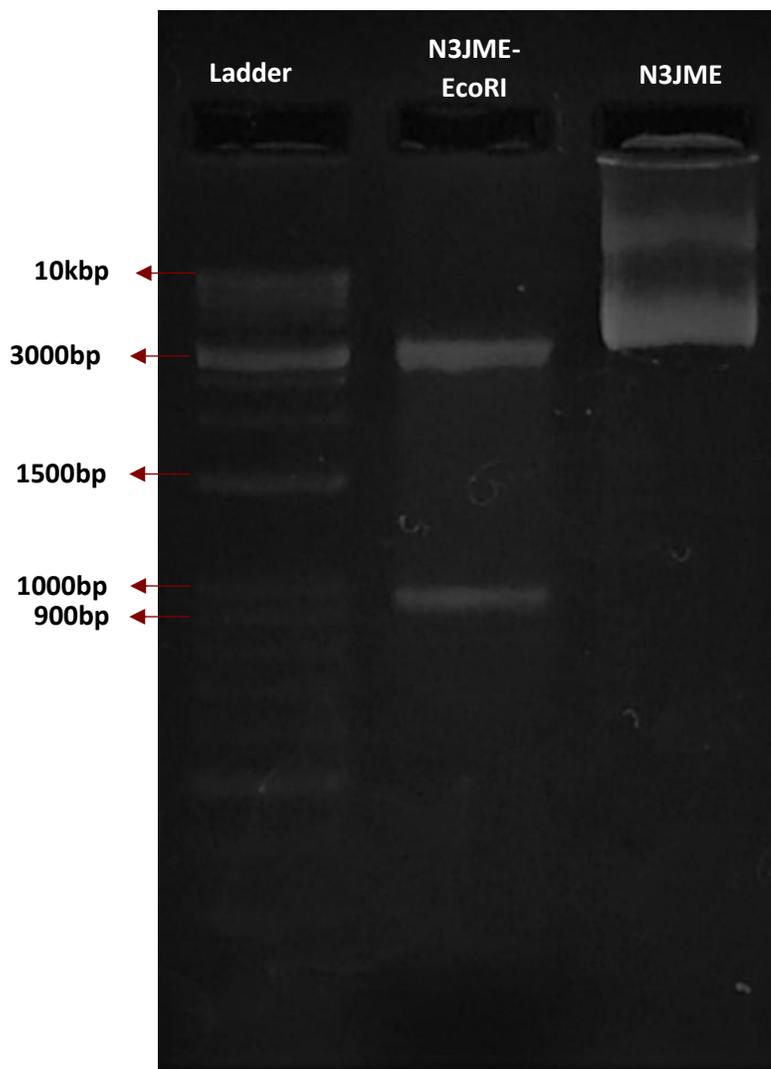
Análisis de las colonias de Expresión

El procedimiento para la transformación y screening de las colonias por colony PCR fue muy similar al realizado anteriormente, con diferencia de utilizar el vector pTrcHis2-TOPO y ligarlo con amplificaciones provenientes del vector NP4 de clonación. El principal análisis que realizamos para dilucidar si nuestro vector se encuentra en Fase es a través de digestión enzimática por EcoRI a los plásmidos extraídos de las cepas con nuestro vector de Expresión. Según el mapa de restricción del gen la proteína N, este debe cortar en la posición 257, considerando el mapa de restricción proporcionado en el

manual del vector (Invitrogen, 2006), la misma enzima corta 6 nucleótidos posteriores al sitio de clonaje, lo que nos permitió diferenciar que los fragmentos observables para que nuestro gen se encuentre en fase son de 4638bp, plásmido más porción del gen hasta el nucleótido 257; y 967bp, porción desde el nucleótido 257 hasta el corte en el sitio de restricción perteneciente al plásmido. Dichos fragmentos establecidos de manera teórica se mantienen como se aprecia en la figura 20, de este modo se identificó que la cepa N3JME posee nuestro inserto en Fase.

Figura 20.

Identificación de Gen N en Fase



Nota Digestión de Plásmido N3JME. – Se puede apreciar el plásmido extraído de la cepa de expresión JM109 en el carril 3 (N3JME) y su respectiva digestión en el carril 2 (N3JME-EcoRI), los fragmentos son correspondientes a un vector con su inserto en fase. Dicho gel de agarosa se encuentra al 1.5%

Análisis bioinformático de la secuencia de expresión

Considerando que la secuencia del vector NP4 clonada al vector de expresión y se ligara en fase desde el inicio de los primers, en la cepa N3JME expresaría una proteína de 406 aminoácidos, alrededor de 44.4kDa, a continuación, se presenta la secuencia ligada y por siguiente su respectiva secuencia aminoacídica.

>N3 (Fragmento insertado en el vector de expresión)

```
AATCAGCGAAATGCACCCCGCATTACGTTTGGTGGACCCTCAGATTCAACTGGCAGTA
ACCAGAATGGAGAACGCAGTGGGGCGCGATCAAAACAACGTCGGCCCAAGGTTTAC
CCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGGAAGACCTAAA
TTCCCTCGAGGACAAGGCGTCCAATTAACACCAATAGCAGTCCAGATGACCAAATTG
GCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGTAAAATGAAAGATC
TCAGTCCAAGATGGTATTTCTACTACCTAGGAACTGGGCCAGAAGCTGGACTTCCCTAT
GGTGCTAACAAAGACGGCATCATATGGGTTGCAACTGAGGGAGCCTTGAATACACCA
AAAGATCACATTGGCACCCGCAATCCTGCTAACAAATGCTGCAATCGTGCTACAACCTCC
TCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAGGCGGCAGTCA
AGCCTCTTCTCGTTCCTCATCACGTAGTCGCAACAGTTCAAGAAATTCAACTCCAGGCA
GCAGTAAACGAACTTCTCCTGCTAGAATGGCTGGCAATGGCGGTGATGCTGCTCTTGC
TTTGCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAATGTCTGGTAAAGGCCAA
CAACAACAAGGCCAAACTGTACTAAGAAATCTGCTGCTGAGGCTTCTAAGAAGCCTC
GGCAAAAACGTACTGCCACTAAAGCATAACAATGTAACACAAGCTTTCGGCAGACGTGG
TCCAGAACAACCCAAGGAAATTTGGGGACCAGGAACTAATCAGACAAGGAACTGA
TTACAAACATTGGCCGCAAATTGCACAATTTGCCCCAGCGTTCAGCGTCTCTCGGAA
TGTCGCGCATTGGCATGGAAGTACACCTTCGGGAACGTGGTTGACCTACACAGGTGC
CATCAAATTGGATGACAAAGATCCAAATTTCAAAGATCAAGTCATTTTGCTGAATAAGC
ATATTGACGCATACAAAACATTCCCAACACAGAGCCTAAAAAGGACAAAAAGAAGA
AGGCTGATGAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAACCTGTGACTC
TTCTCCTGCTGCAGATTTGGATGATTTCTCCAAACAATTGCAACAATCCATGAGCAGT
```

>N3 (Secuencia aminoacídica)

```
NQRNAPRITFGGSPDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHGKEDLKF
PRGQGVPIINTNSSPDDQIGYRRATRRIRGGDGKMKDLSRWYFYLLGTGPEAGLPYGA
NKDGIWVATEGALNTPKDHIGTRNPANNAIIVLQLPQGTTLPKGFYAEGSRGGSQASSR
SSRSRNSSRNSTPGSSKRTSPARMAGNGGDAALALLLDRLNQLSKMSGKQQQQGQ
TVTKKSAEASKKPRQKRTATKAYNVTQAFRRGPEQTQGNFGDQELIRQGTDYKHWP
QIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTF
PPTPEPKDKKKKADETQALPQRQKKQQTVTLPAADLDDFSKQLQQSMSS
```

A partir de la secuencia aminoacídica por medio de antheprot se obtuvo el siguiente perfil de las regiones más antigénicas. Entre los cambios más notables, con respecto al perfil mostrado en la tabla 10, se aprecia una pérdida de dos regiones pertenecientes a los residuos “3DNG5” y “412SS413” y se genera una región grande desde el residuo 171 al 193, tabla 12. El perfil de predicción se muestra en la figura 21.

Tabla 12.

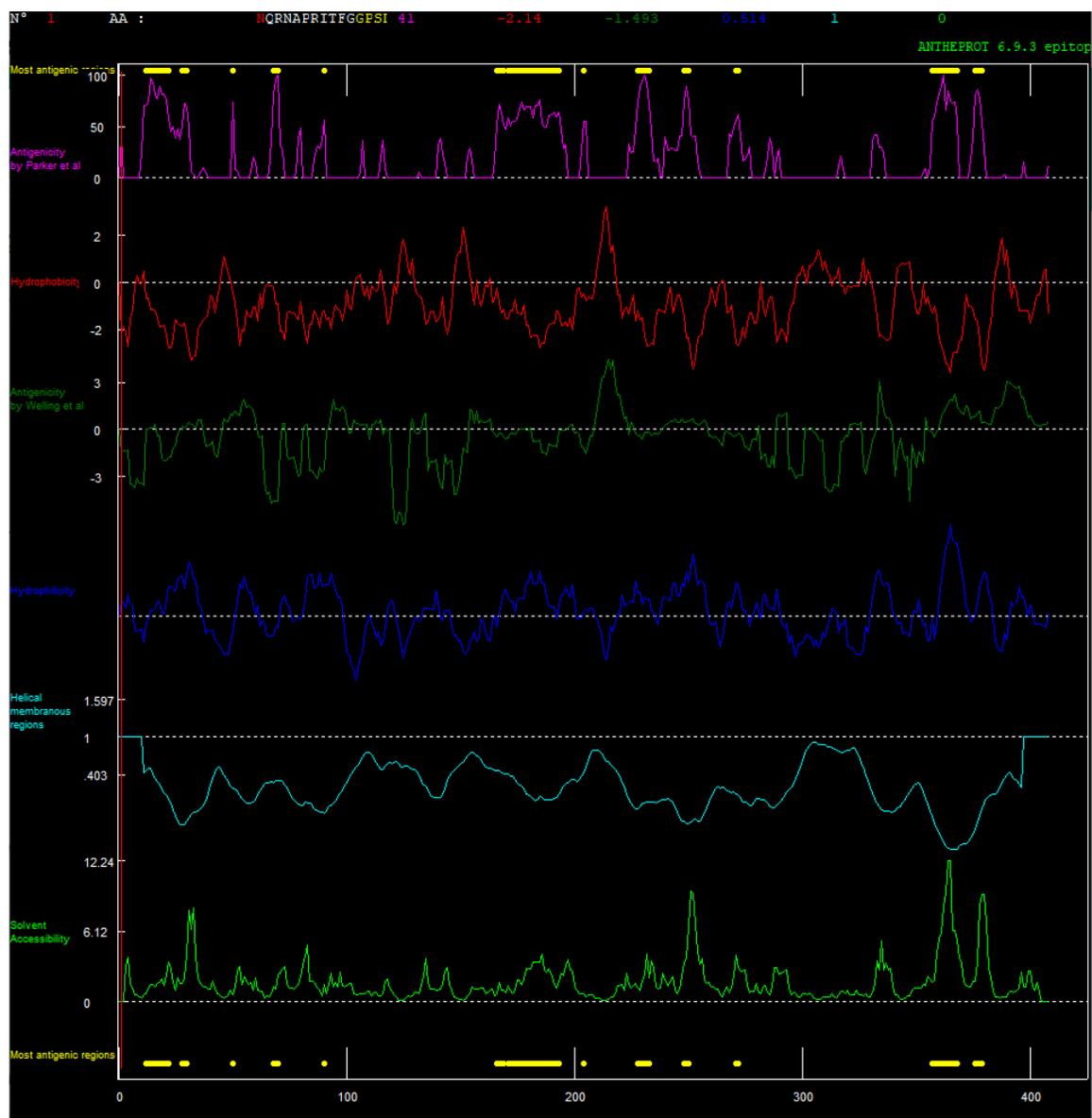
Regiones antigénicas de proteína expresada

Numero de aminoácido inicial	Aminoácidos pertenecientes al Epítotope	Número del aminoácido final
12	GPSDSTGSNQN	22
28	ARS	30
50	T	50
68	NTN	70
90	G	90
166	AEGS	169
171	GGSQASSRSSSRNSTRNTPG	193
204	A	204
228	SGKGQQ	233
248	SKK	250
271	GP	272

Nota. El programa Antheprot muestra los 11 mejores candidatos para epítotope de la proteína N, expresada en la cepa N3, considerando los perfiles de Hidrofobicidad, Hidrofilicidad y accesión al solvente.

Figura 21.

Perfil de proteína expresada por cepa N3

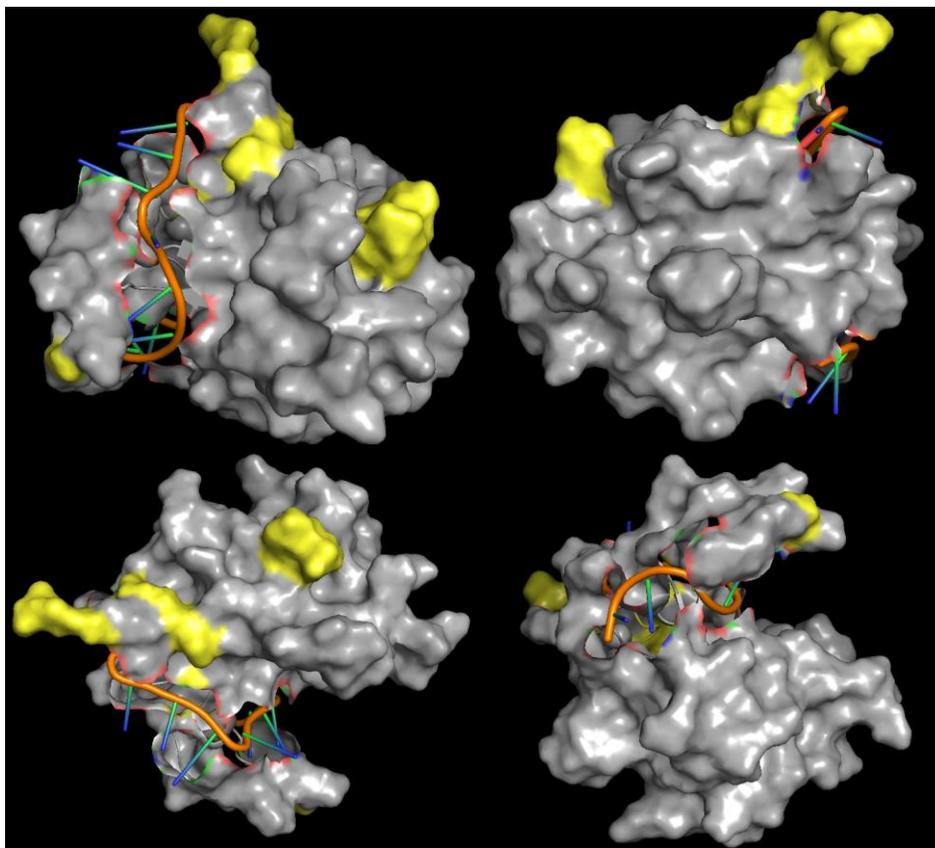


Nota. El programa Antheprot muestra los 11 mejores candidatos para epitopes de la proteína N, expresada en la cepa N3, considerando los perfiles de Hidrofobicidad (perfil rojo), Hidrofilicidad (perfil azul) y acceso al solvente (perfil verde).

Por medio de la herramienta Swissmodel, tras utilizar como templado nuestra secuencia aminoacídica, se pudo observar por medio de PyMOL las regiones antigénicas de los dominios de unión al ARN, de dimerización y el enlazador central. En la figura 22, se aprecia una representación de superficie del Dominio de Unión al RNA (RBD) en color plateado, en amarillo se puede apreciar las regiones antigénicas previamente descritas en la tabla 12, la sección naranja representa una hebra del RNA, cabe destacar que el modelamiento se generó desde el residuo 37 al 173, considerando la secuencia aminoacídica de la cepa N3.

Figura 22.

Dominio de Unión al RNA

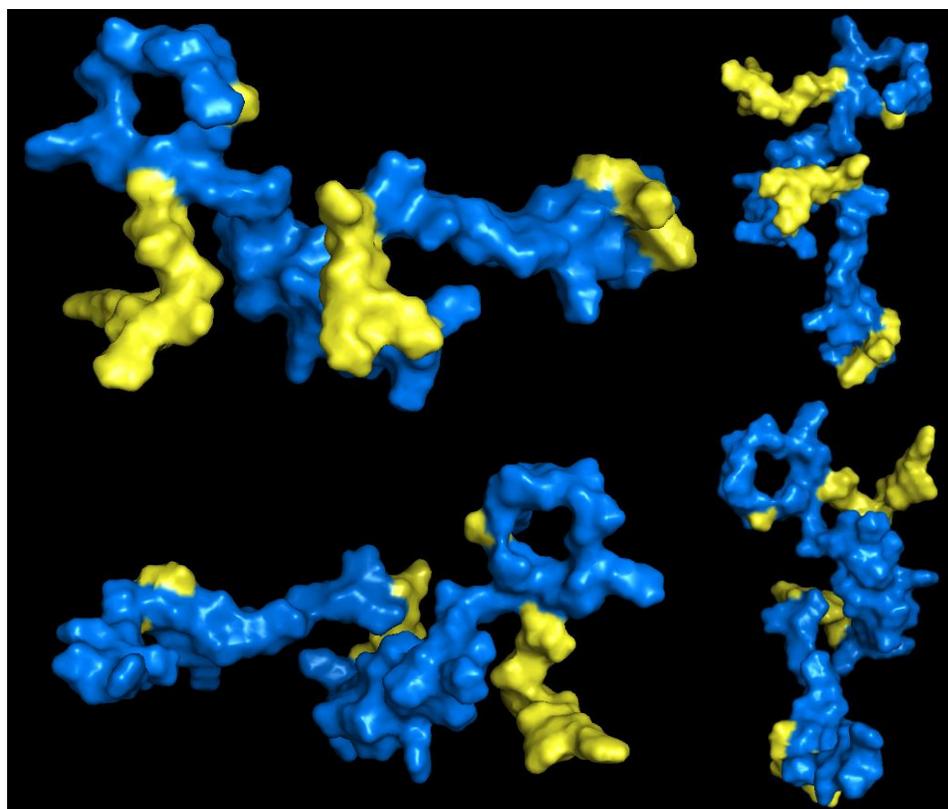


Nota. La superficie del dominio de Unión al ARN (RBD) de la proteína N se representa en cuatro orientaciones diferentes y la manera que interacciona con el RNA, las secciones de color amarillo representan sus regiones más antigénicas.

En la figura 23, se aprecia una representación del Enlazador Central (LINKER) en azul, en amarillo se puede apreciar las regiones antigénicas previamente descritas en la tabla 12, cabe destacar que el modelamiento se generó desde el residuo 184 al 256, considerando la secuencia aminoacídica de la cepa N3.

Figura 23.

Enlazador Central (LINKER)

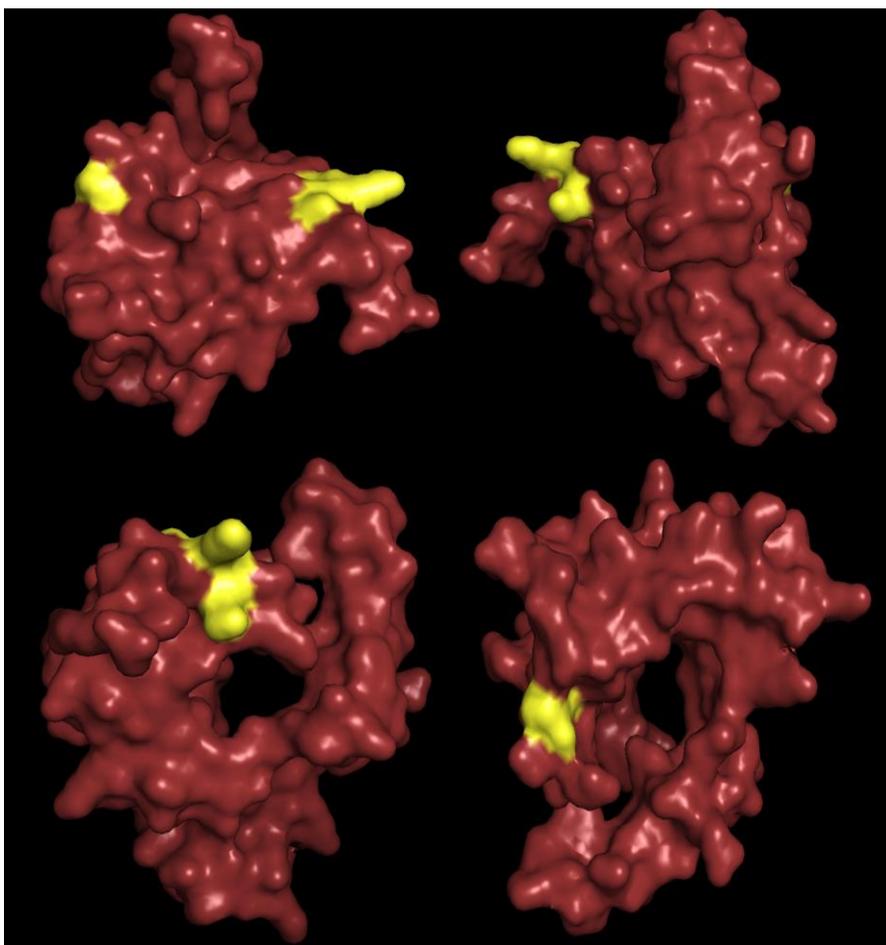


Nota. La superficie del Enlazador Central (Linker) de la proteína N se representa en cuatro orientaciones diferentes, las secciones de color amarillo representan sus regiones más antigénicas.

En la figura 24, se aprecia una representación del Dominio de Dimerización en Rojo, en amarillo se puede apreciar las regiones antigénicas previamente descritas en la tabla 12, cabe destacar que el modelamiento se generó desde el residuo 240 al 357, considerando la secuencia aminoacídica de la cepa N3.

Figura 24.

Dominio de Dimerización

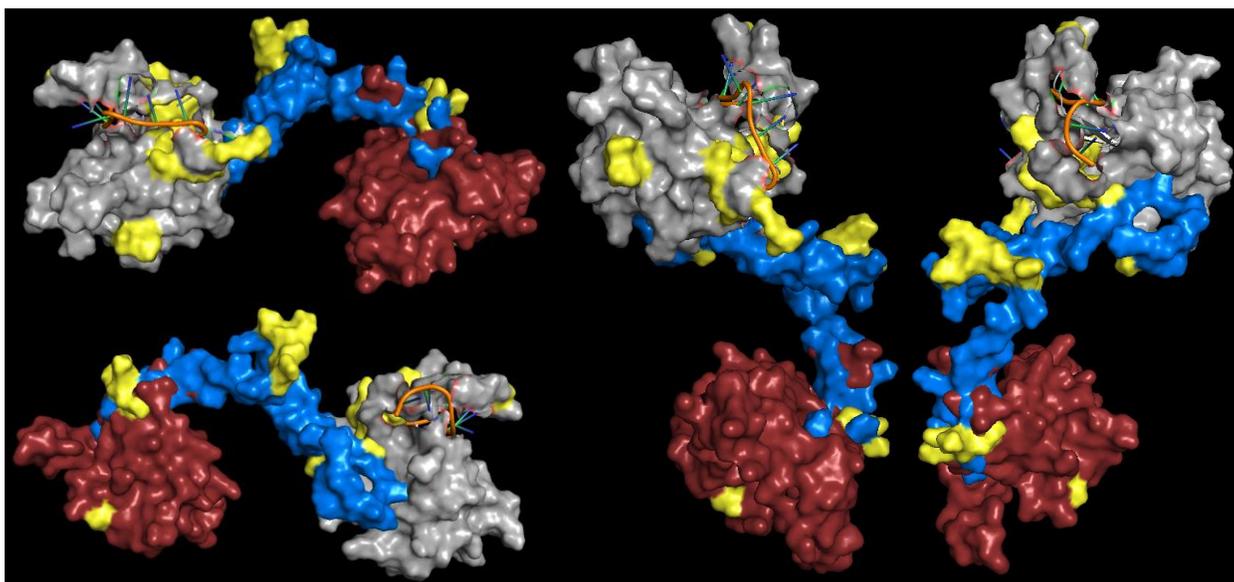


Nota. La superficie del Dominio de Dimerización de la proteína N se representa en cuatro orientaciones diferentes, las secciones de color amarillo representan sus regiones más antigénicas.

Los modelos previos permitieron generar una representación de la proteína N, a partir del alineamiento de sus aminoácidos, en la figura 25, se puede apreciar los tres dominios desde distintas perspectivas, en amarillo se puede apreciar las regiones antigénicas de la proteína N.

Figura 25.

Estructura de la superficie de la proteína N



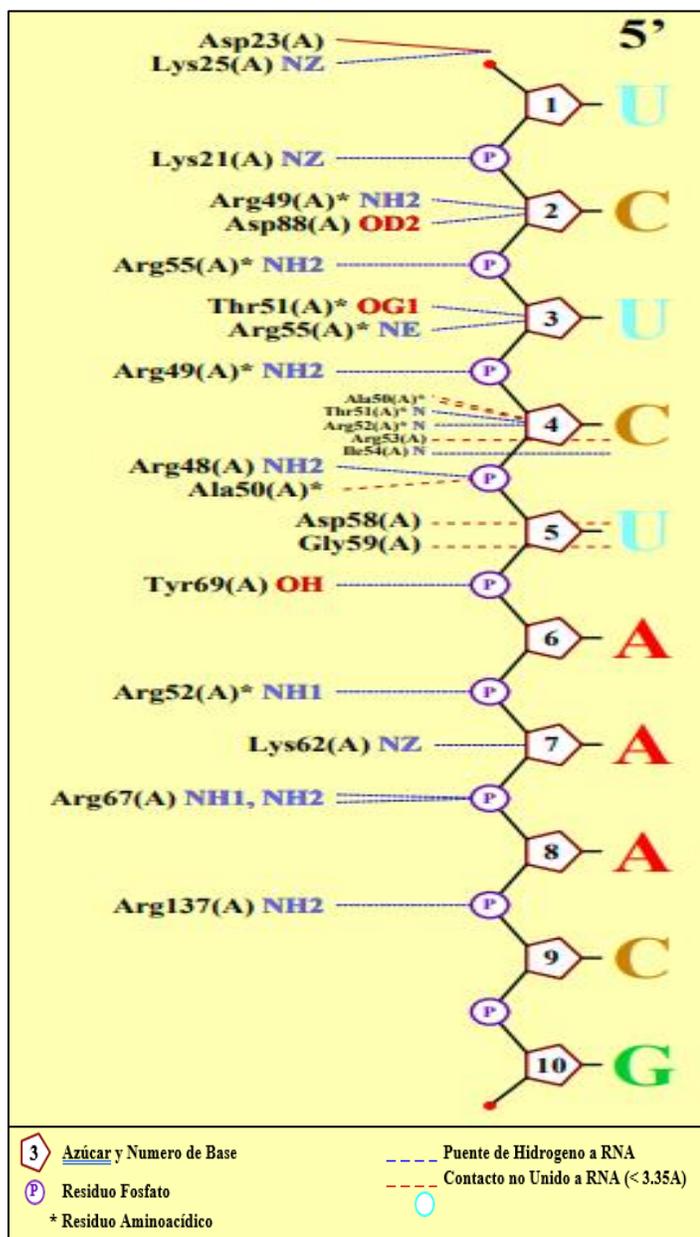
Nota. La superficie la proteína N se representa en cuatro orientaciones diferentes, las secciones de color amarillo representan sus regiones más antigénicas. En Plateado el Dominio de Unión al RBD, en azul el Enlazador Central (LINKER) y en rojo el Dominio de Dimerización.

La utilización de PDBsum nos permitió observar que residuos aminoácidos en el Dominio de Unión al RNA interaccionan con el RNA, en la Figura 26, se puede apreciar que las distintas porciones de los aminoácidos interaccionando a través de puentes de hidrogeno o por medio de uniones sin contacto. Entre los residuos que más interaccionan tenemos a los de Arginina y Lisina. Nótese que las interacciones con fosfatos y azucres

generalmente se generan por medio de puentes de hidrogeno, mientras que las interacciones con las bases nitrogenadas por medio de uniones sin contacto como las fuerzas de Van der Valls.

Figura 26.

Interacción de Residuos del RBD con el RNA



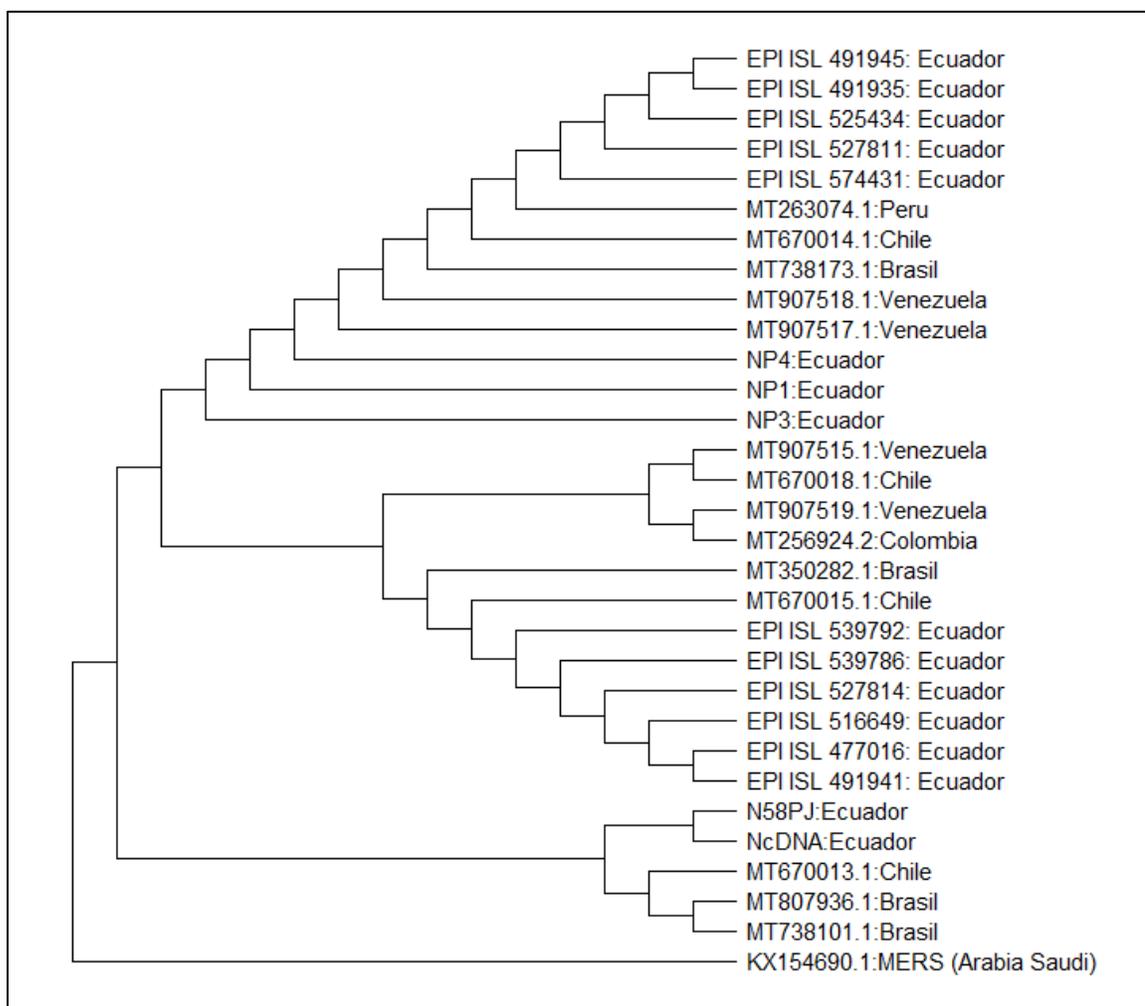
Nota. Interacción del Dominio de unión al RNA con 10 nucleótidos de RNA, el análisis de interacciones fue realizado a por medio de la herramienta PDBsum.

Filogenia

Se generaron dos análisis filogenéticos, el primero compara las secuencias previamente descritas en relación con Latinoamérica, figura 26. El segundo análisis tiene la finalidad de englobar a las distintas variantes del SARS-CoV-2 con respecto a secuencias de aislados ecuatorianos, figura 27.

Figura 27.

Árbol filogenético Asilados Latinoamericanos

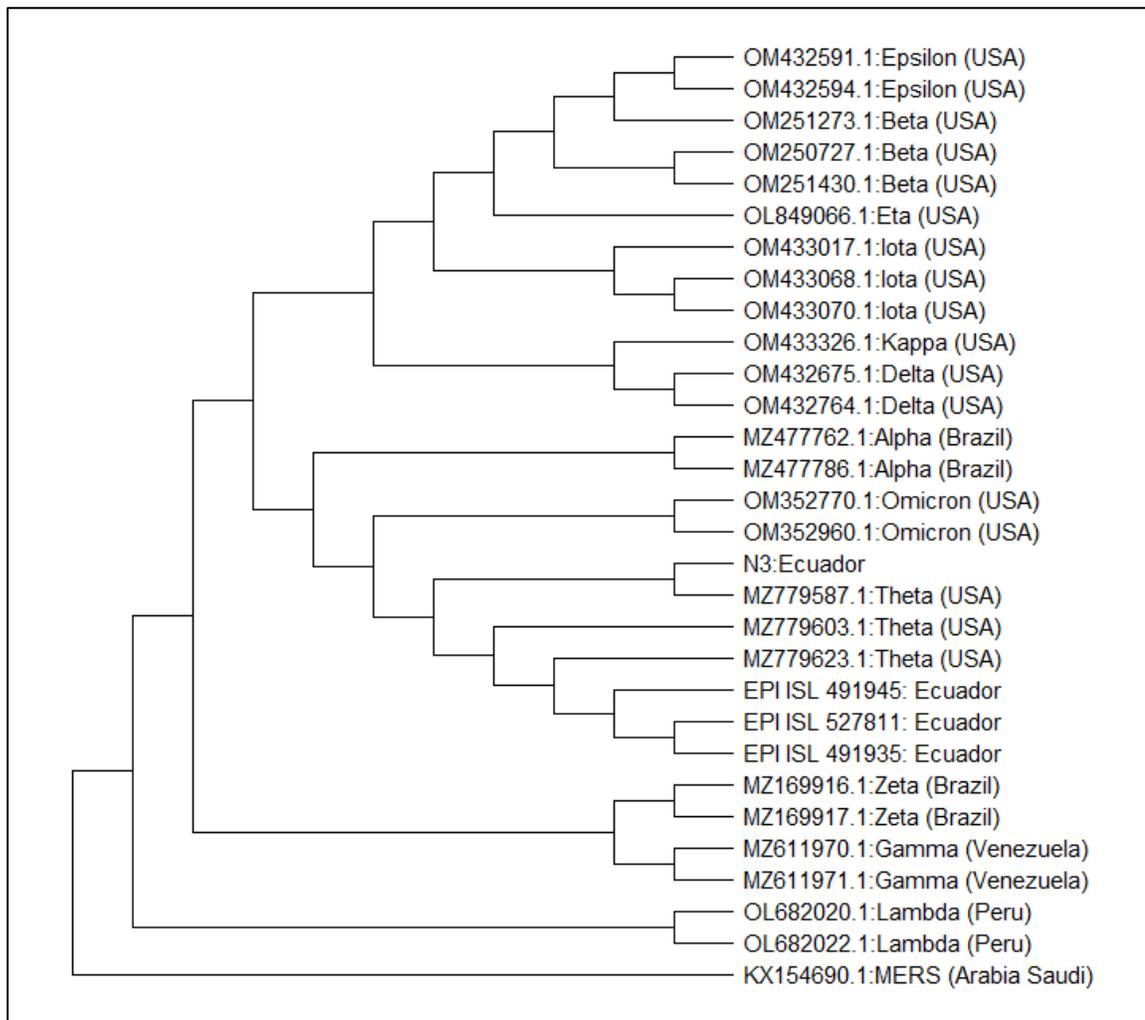


Nota. Árbol filogenético de secuencias nucleotídicas de distintos aislados latinoamericanos, generados por el método de Maximum Likelihood con una prueba filogenética de tipo “Bootstrap method” con 1000 replicaciones y un modelo de sustitución “Tamura-Nei model”.

Se aprecia en el árbol filogenético de los aislados latinoamericanos, figura 26, que entre las secuencias pertenecientes a la proteína N y su grupo externo la proteína N perteneciente al MERS, considerando el análisis de diversidad entre las distintas secuencias, existe una divergencia del 2.1% del MERS con respecto a las secuencias de la proteína N. Entre las secuencias de la proteína N del SARS-CoV-2 existe una divergencia lo suficiente para generar una topología, pero bastante pequeña, menor a cero considerando el análisis de diversidad entre secuencias, para señalar una diferencia en la especie. Se aprecia que las secuencias N58PJ y NcDNA, pertenecen a otro clado en relación con NP1, NP3 y NP4, en donde se encuentran relacionadas con aislados ecuatorianos indexados en la plataforma GISAID, tabla 2, esto no representa que dichas secuencias no pertenezcan a aislados ecuatorianos, simplemente que comparte otro ancestro en común con aislados de Chile y Brasil, esto sucede de manera similar con aislados ecuatorianos pertenecientes a otros grupos.

Figura 28.

Árbol Filogenético de Aislados Ecuatorianos y variantes del SARS-CoV-2



Nota. Árbol filogenético de secuencias nucleotídicas de distintos aislados ecuatorianos y variantes del SARS-CoV-2, generados por el método de Maximum Likelihood con una prueba filogenética de tipo “Bootstrap method” con 1000 replicaciones y un modelo de sustitución “Tamura-Nei model”.

En el caso de las diferentes variantes reportadas del SARS-CoV-2, se consideró de interés analizar cuanta variabilidad podría generar los cambios en el material genético con respecto a la proteína N, se aprecia nuevamente que la divergencia existente entre las secuencias de la proteína N del SARS-CoV-2 ante la de la proteína N del MERS es de 1.2%, considerando el análisis de diversidad entre secuencias. De manera similar se aprecia que la divergencia entre secuencia y secuencia de las distintas variantes es sumamente pequeña dentro de la proteína N para marcar una diferencia dentro de la especie, pero es la suficiente para permitirnos observar una topología, dentro de este árbol es destacable que nuestra secuencia clonada N3 y distintos aislados ecuatorianos en conjunto con aislados la variante Theta comparten un ancestro común, por lo que se puede considerar que la secuencia clonada pertenece a la misma variante al compartir un nodo particular con MZ779587.1; por otra parte la variante Lambda perteneciente a aislados peruanos, presentan una separación del resto de variantes, teniendo en cuenta que dichas secuencias fueron alineadas antes de generar el árbol filogenético, se observó que muchas de estas variaciones pertenecían a sustituciones de varios nucleótidos dentro de su secuencia en relación a las demás lo que puede haber generado la bifurcación en relación al clado de las demás variantes.

Capítulo V: Discusión

Antigenicidad de la Proteína N

En este trabajo confirmamos que la proteína N posee múltiples epítopes que la convierten en un buen candidato como antígeno de pruebas de detección del SARS-CoV-2 (Tian et al., 2022), considerando que la mayoría de los sitios antigénicos se encuentran presentes en el LINKER y el Dominio de Dimerización, podemos considerar estas secciones de esencial expresión. Lutomski et al., 2021 reportaron que, aunque el Dominio de Unión al RNA se mantiene interaccionando constantemente con el RNA, este no es inmunogénico y por lo tanto es débilmente reconocido por IgG, en relación a la respuesta que presentan los residuos desde los aminoácidos 156 al 419, los cuales pertenecen a la región LNKER y al Dominio de Dimerización (Lutomski et al., 2021). Otro estudio utilizando varios servidores bioinformáticos reafirma nuestra predicción de epítopes del gen clonado, resaltando que los residuos predichos del 229 al 269 son epítopes inmunodominantes, dichas secciones pertenecen de igual modo al LINKER y el Dominio de Dimerización (Serra et al., 2012). Sin embargo, tanto en nuestra predicción bioinformática como en nuestros modelos estructurales se puede apreciar que la región que más antigenicidad presenta es el LINKER, incluso ensayos demuestran que los epítopes de células B más dominantes se encuentran localizados entre los residuos 176 y 206 (Rakib et al., 2020), esta es la región de epítopes pertenecientes al LINKER más extensa que se pudo generar en nuestro análisis y que igualmente, se ha reportado en otros estudios.

Mutaciones de la proteína N

Dentro de nuestras secuencias se han encontrado las mutaciones R203K y G204R que contrastan a nuestro previo análisis en la plataforma GISAID de aislados ecuatorianos. Estas son las sustituciones más comunes e igual que en este estudio suelen ocurrir comúnmente con las sustituciones de los nucleótidos “GGG” a “AAC”(Ye et al., 2020). Se reporta que este tipo de sustituciones son originarios de cepas de Norte América (Lee et al., 2021) y considerando que Ecuador es un país turístico, es probable que esta cepa se introdujo al país en las etapas iniciales de la Pandemia. Por otra parte, estas mutaciones en específico no son detectadas en otras poblaciones como en aislados coreanos y corresponde a una mutación de adaptabilidad a una nueva región geográfica en poblaciones específicas(Mercatelli & Giorgi, 2020). De acuerdo a Rahman et al., 2021 incluso estas inserciones favorecen la torción, flexibilidad y aumenta la estabilidad de la proteína(Rahman et al., 2021).

Otro estudio bioinformáticos reporta que la proteína N es altamente conservada, aunque han establecido que el 36% de las cepas posee las sustituciones en los aminoácidos 203 y 204, a su vez sugieren que dichas mutaciones son coevolutivas, es decir aparecieron juntas(Del Veliz et al., 2021), esto tiene concordancia debido a que la variación de los nucleótidos se presentan entre dos codones, los mismos que generan los residuos variantes. Estos residuos mutantes se puede apreciar que se encuentra en la región LINKER, la cual en nuestro análisis ha presentado gran porción de sitios antigénicos, sin embargo no encontramos que dicha sustitución afecte en una posible pérdida de regiones antigénicas, en Brasil que también ha reportado las mutaciones 203-

204 en sus aislados sugieren que puede tener un impacto en las RT-PCR de tiempo real, generando falsos negativos, por lo que recomiendan utilizar más de un gen objetivo en el diagnóstico además de la proteína N.

Identidad y Estructura de la Proteína N

Estudios reportan que la proteína N que está compuesta por 419 residuos aminoacídicos, esto representa una porción de su genoma de 1260bp(Zeng et al., 2020). A lo largo de los diferentes ensayos se ha podido corroborar la identidad del gen de la proteína N. En nuestro trabajo logramos amplificar 1218bp, lo que representa 406 residuos aminoacídicos, y que en nuestro sistema de expresión se generarían los dominios principales de la proteína, los cuales son: el Dominio de Unión al RNA, la Región enlazadora rica en Serina y Arginina y el Dominio de Dimerización. Estos dominios son esenciales para la oligomerización y empaquetamiento del RNA(Wu et al., 2021). Cabe destacar que varios aislados identificados e indexados de secuencias nucleotídicas y aminoacídicas del GenBank tuvieron una similitud del 100% con todas nuestras secuencias obtenidas de nuestros aislados ecuatorianos (GenBank, 2022).

Con respecto a los modelos estructurales generados, el Dominio de Unión al RNA coincide con la estructura reportada por Dinesh et al. , 2020 en donde consta de dos secciones representativas la sección del dedo extendido y la palma, entre ellas interacciona el RNA para su empaquetamiento progresivo (Dinesh et al., 2020). Aunque Dinesh et al. , 2020 presenten los residuos que interaccionan con el RNA, sus resultados difieren con nuestro modelo generado por PDBsum, esto puede ser debido a la diferencia de capacidad de cómputo que pueden generar los servidores informáticos utilizados para el procesamiento y la predicción de interacciones más o menos acertadas, sin embargo

en ambas predicciones de residuos interaccionando con el RNA la mayoría presentan regiones de Arginina y Lisina (Dinesh et al., 2020). La estructura del Dominio de Dimerización es reportada como una superficie altamente básica y generalmente dimerizada (Cubuk et al., 2021), su representación estructural es muy similar a la nuestra con la diferencia que en nuestro caso la presentamos en su forma no dimerizada. En el caso del Enlazador Central LINKER, la estructura similar a la generada en nuestro modelo se lo presenta en estudios donde lo reportan como un cofactor importante para la replicación y transcripción del virus considerando que es esencial su interacción con un dominio de la nsp3 (Bessa et al., 2022).

Análisis Filogenético

Se ha observado que la proteína N es un gen altamente conservado y aunque las distintas variantes del SARS-CoV-2, tengan inserciones, deleciones y sustituciones en varias porciones del gen, no proporcionan la suficiente divergencia para observar diferencias altamente evolutivas, al menos en la proteína N. Sin embargo, como es apreciable en nuestros árboles la diversidad existente entre secuencias es la suficiente para generar una topología y permitirnos una idea de que cepas han evolucionado más, es de recalcar que considerando que la obtención de secuencias fue a distintas etapas de la pandemia por el SARS-CoV-2, se pudo observar grupos que pertenecen a asilados del mismo país pero no se encuentran dentro del mismo clúster, esto puede generarse al contagió proveniente de turistas de distintas partes del mundo o a su vez es debido a una alta tasa de replicación del virus en pocos periodos de tiempo que generan una topología de mayor cantidad de ancestros en relación a otros aun perteneciendo a la misma población (Z. G. Wang et al., 2007).

Por otra parte, la poca divergencia presentada incluso en los genes de las variantes es debida a que si bien el gen N es ortólogo y es una característica ideal para generar este tipo de análisis filogenéticos, este ha sido descartado para ser utilizado como un diferenciador de variantes o especies, en su lugar es preferente la utilización del ORF1ab o varios genes ortólogos para la generación de análisis filogenéticos(Li et al., 2020).

Capítulo VI: Conclusiones

En el presente estudio se ha logrado clonar y trasladar el gen de la proteína N en un vector de expresión, con una longitud de 1218pb y de 406 aminoácidos. Este fragmento clonado mantiene las regiones del Dominio de Unión al RNA, el Enlazador Central (LINKER) y su Dominio de Dimerización. Dicha proteína se encuentra lista para posterior expresión dentro de la cepa N3 a través de un sistema de inducción al vector de expresión con el gen de la proteína N en *E.coli JM109*.

Su análisis bioinformático nos generó un modelo aproximado de los dominios de la proteína N, la interacción del RNA con el Dominio de Unión al RNA, en donde la mayoría de los residuos que interaccionan con el RNA son de Arginina y Lisina, y el enlazador central LINKER y el Dominio de Dimerización.

De estos dominios, la porción con mayor antigenicidad se encuentra en el LINKER y cabe destacar que en esta región se presentaron las mutaciones de los residuos 203 y 204, sin embargo, éstas no originaron cambio alguno en las regiones antigénicas predichas.

Rahman et al., 2021 reportaron que esta es una mutación de adaptabilidad a regiones específicas que mejora propiedades de la proteína N. Para finalizar, el análisis filogenético nos ha permitido afianzar la identidad de nuestra secuencia amplificada con respecto a aislados de la región, del mismo modo el árbol de la proteína N de las distintas variantes del SARS-CoV-2, nos ha permitido determinar que los aislados Ecuatorianos tienen estrecha relación con la variante Theta por lo que se pueden determinar como parte de la misma.

Por lo cual se puede establecer que la proteína N es altamente conservada debido a la funcionalidad que posee de empaquetar el RNA y ayudar en la transcripción de este (Rump et al., 2021). Las mutaciones observadas, no representan un cambio que le afecte tanto en las funciones como en su antigenicidad; Aspectos ambos primordiales para la detección de la proteína a través de anticuerpos, sin embargo, por otra parte, dichas mutaciones no representan una divergencia lo suficientemente alta para que la proteína N sea un gen prioritario en los análisis filogenéticos y modelos de evolución, por otra parte permite observar a través de la topología generada en los árboles filogenéticos a que tipo de variante del SARS-CoV-2 puede pertenecer un aislado de la proteína N.

Capítulo VII: Recomendaciones

Luego de los diferentes estudios bioinformáticos donde se observa que la proteína N puede ser un antígeno conservado para usarse en el diagnóstico de la COVID-19, el cual, además, está excluido de las mayorías de las vacunas, puede permitir diferenciar pacientes enfermos de COVID-19, de otras neumonías, al tener una gran mayoría de la población vacunada con inmunógenos que utilizan como única diana la proteína de RBD. Es por ello, que una conclusión de este trabajo es elaborar pruebas de diagnóstico donde se utilicen ahora la RBD para detectar vacunados y con la proteína N para detectar pacientes con el virus del SARS-Cov-2.

Durante este trabajo el gen de la proteína N, fue clonado y pasado a un vector de expresión para su utilización en pruebas de diagnóstico, es por ello, recomendamos continuar este trabajo con la purificación de la proteína recombinante y su uso posterior en pruebas de diagnóstico.

Considerando las distintas variantes que pueden en cierto momento generar un cambio drástico como deleciones, inserciones o sustituciones que no se observaron en este estudio, es recomendable realizar de manera constante un análisis de identidad de las secuencias utilizadas en relación a las secuencias de aislados recientes, puesto que al disminuir la identidad de la proteína N puede afectar a la identificación del SARS-CoV-2 al momento de estandarizar la prueba de ELISA y generar falsos negativos.

Se recomienda que los análisis bioinformáticos como de predicción de epítopes o docking entre proteínas o RNA se desarrollen a futuro a partir de servidores informáticos con mayor poder de procesamiento para obtener modelos más acertados.

Para finalizar es recomendable realizar análisis bioinformáticos de docking molecular de la proteína N con distintas proteínas que interaccionan para la replicación del virus con la finalidad de comprender que relaciones pueden ser primordiales para el desarrollo de fármacos futuros

Capítulo VIII: Bibliografía

abm. (2021). *MegaFi™ One-Step RT-PCR*.

https://www.abmgood.com/pub/media/productdocument/document//g/5/g597_datasheet_v2-2.pdf

Alpdagtas, S., Ilhan, E., Uysal, E., Sengor, M., Ustundag, C. B., & Gunduz, O. (2020). Evaluation of current diagnostic methods for COVID-19. *APL Bioengineering*, 4(4), 041506.

<https://doi.org/10.1063/5.0021554>

Bessa, L. M., Guseva, S., Camacho-Zarco, A. R., Salvi, N., Maurin, D., Perez, L. M., Botova, M., Malki, A., Nanao, M., Jensen, M. R., Ruigrok, R. W. H., & Blackledge, M. (2022). The intrinsically disordered SARS-CoV-2 nucleoprotein in dynamic complex with its viral partner nsp3a. *Science Advances*, 8(3), 4034.

https://doi.org/10.1126/SCIADV.ABM4034/SUPPL_FILE/SCIADV.ABM4034_SM.PDF

Burbelo, P. D., Riedo, F. X., Morishima, C., Rawlings, S., Smith, D., Das, S., Strich, J. R., Chertow, D. S., Davey, R. T., Jr, & Cohen, J. I. (2020). Sensitivity in Detection of Antibodies to Nucleocapsid and Spike Proteins of Severe Acute Respiratory Syndrome Coronavirus 2 in Patients With Coronavirus Disease 2019. *The Journal of Infectious Diseases*, 222, 206–213.

<https://doi.org/10.1093/INFDIS/JIAA273>

Callejas, D., Echevarría, J. M., Carrero, Y., Rodríguez-Morales, A. J., & Moreira, R. (2020). The

SARS-CoV-2 Pandemic in Latin America: the Need for Multidisciplinary Approaches. *Current Tropical Medicine Reports*, 7(4), 1. <https://doi.org/10.1007/S40475-020-00219-W>

Cubuk, J., Alston, J. J., Incicco, J. J., Singh, S., Stuchell-Brereton, M. D., Ward, M. D., Zimmerman, M. I., Vithani, N., Griffith, D., Wagoner, J. A., Bowman, G. R., Hall, K. B., Soranno, A., & Holehouse, A. S. (2021). The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nature Communications*, 12(1).
<https://doi.org/10.1038/S41467-021-21953-3>

Del Veliz, S., Rivera, L., Bustos, D. M., & Uhart, M. (2021). Analysis of SARS-CoV-2 nucleocapsid phosphoprotein N variations in the binding site to human 14-3-3 proteins. *Biochemical and Biophysical Research Communications*, 569, 154.
<https://doi.org/10.1016/J.BBRC.2021.06.100>

Dinesh, D. C., Chalupska, D., Silhan, J., Koutna, E., Nencka, R., Veverka, V., & Boura, E. (2020). Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLOS Pathogens*, 16(12), e1009100. <https://doi.org/10.1371/JOURNAL.PPAT.1009100>

G, D., C, C., C, B., & C, G. (2001). ANTHEPROT: an integrated protein sequence analysis software with client/server capabilities. *Computers in Biology and Medicine*, 31(4), 259–267.
[https://doi.org/10.1016/S0010-4825\(01\)00008-7](https://doi.org/10.1016/S0010-4825(01)00008-7)

GenBank. (2022). *GenBank Overview*. <https://www.ncbi.nlm.nih.gov/genbank/>

Habas, K., Nganwuchu, C., Shahzad, F., Gopalan, R., Haque, M., Rahman, S., Majumder, A. A., & Nasim, T. (2020). Resolution of coronavirus disease 2019 (COVID-19). In *Expert Review of Anti-Infective Therapy* (Vol. 18, Issue 12, pp. 1201–1211). Taylor and Francis Ltd.
<https://doi.org/10.1080/14787210.2020.1797487>

- Helmy, Y. A., Fawzy, M., Elswad, A., Sobieh, A., Kenney, S. P., & Shehata, A. A. (2020). The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *Journal of Clinical Medicine*, 9(4), 1225. <https://doi.org/10.3390/JCM9041225>
- Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2020). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology* 2020 19:3, 19(3), 141–154. <https://doi.org/10.1038/s41579-020-00459-7>
- Invitrogen. (2006). *pTrcHis and pTrcHis2 TOPO[®] TA Expression Kits Five-minute cloning of Taq polymerase-amplified PCR products for expression in E. coli Catalog no. K4410-01 (pTrcHis TOPO[®]) Catalog nos. K4400-01 and K4400-40 (pTrcHis2 TOPO[®])*. www.invitrogen.com
- Invitrogen. (2014, February 28). *Taq DNA Polymerase, recombinant MANUAL*. https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2F0814_Taq_DNA_Polymerase_recombinant.pdf
- Invitrogen, T. S. (2015). *M-MLV Reverse Transcriptase*. https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2Fmmlv_rt_man.pdf
- Kellam, P., & Barclay, W. (2020). The dynamics of humoral immune responses following SARS-CoV-2 infection and the potential for reinfection. *The Journal of General Virology*, 101(8), 791. <https://doi.org/10.1099/JGV.0.001439>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary

Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/MOLBEV/MSY096>

Lee, S., Won, D. J., Kim, C. K., Ahn, J., Lee, Y., Na, H., Kim, Y. T., Lee, M. K., Choi, J. R., Lim, H. S., & Lee, K. R. (2021). Novel indel mutation in the N gene of SARS-CoV-2 clinical samples that were diagnosed positive in a commercial RT-PCR assay. *Virus Research*, 297, 198398. <https://doi.org/10.1016/J.VIRUSRES.2021.198398>

Li, T., Liu, D., Yang, Y., Guo, J., Feng, Y., Zhang, X., Cheng, S., & Feng, J. (2020). Phylogenetic supertree reveals detailed evolution of SARS-CoV-2. *Scientific Reports*, 10(1). <https://doi.org/10.1038/S41598-020-79484-8>

Lu, X., Pan, J., Tao, J., & Guo, D. (2011). SARS-CoV nucleocapsid protein antagonizes IFN- β response by targeting initial step of IFN- β induction pathway, and its C-terminal region is critical for the antagonism. *Virus Genes*, 42(1), 37. <https://doi.org/10.1007/S11262-010-0544-X>

Luo, H., Chen, Q., Chen, J., Chen, K., Shen, X., & Jiang, H. (2005). The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1. *Febs Letters*, 579(12), 2623. <https://doi.org/10.1016/J.FEBSLET.2005.03.080>

Lutowski, C. A., El-Baba, T. J., Bolla, J. R., & Robinson, C. V. (2021). Multiple Roles of SARS-CoV-2 N Protein Facilitated by Proteoform-Specific Interactions with RNA, Host Proteins, and Convalescent Antibodies. *JACS Au*, 1(8), 1147–1157. <https://doi.org/10.1021/JACSAU.1C00139>

Márquez, S., Prado-Vivar, B., Guadalupe, J. J., Gutierrez, B., Jibaja, M., Tobar, M., Mora, F.,

Gaviria, J., García, M., Espinosa, F., Ligña, E., Reyes, J., Barragán, V., Rojas-Silva, P., Trueba, G., Grunauer, M., & Cárdenas, P. (2020). Genome sequencing of the first SARS-CoV-2 reported from patients with COVID-19 in Ecuador. *MedRxiv*.

<https://doi.org/10.1101/2020.06.11.20128330>

Mercatelli, D., & Giorgi, F. M. (2020). Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Frontiers in Microbiology, 11*, 1800.

<https://doi.org/10.3389/FMICB.2020.01800/FULL>

Ministerio de Salud Pública. (2021, July 30). *Actualización de casos de coronavirus en Ecuador – Ministerio de Salud Pública*. <https://www.salud.gob.ec/actualizacion-de-casos-de-coronavirus-en-ecuador/>

Pan American Health Organization. (2021, July 31). *COVID-19 vaccine doses administered in the Americas*. https://ais.paho.org/imm/IM_DosisAdmin-Vacunacion.asp

Patel, R., Babady, E., Theel, E. S., Storch, G. A., Pinsky, B. A., George, K. S., Smith, T. C., & Bertuzzi, S. (2020). Report from the american society for microbiology covid-19 international summit, 23 march 2020: Value of diagnostic testing for sars–cov-2/covid-19. *MBio, 11*(2). <https://doi.org/10.1128/MBIO.00722-20>

Pei, H., Liu, J., Cheng, Y., Sun, C., Wang, C., Lu, Y., Ding, J., Zhou, J., & Xiang, H. (2005). Expression of SARS-coronavirus nucleocapsid protein in *Escherichia coli* and *Lactococcus lactis* for serodiagnosis and mucosal vaccination. *Applied Microbiology and Biotechnology, 68*(2), 220. <https://doi.org/10.1007/S00253-004-1869-Y>

PREMIER Biosoft. (2021). *Primer Design Guide for PCR :: Learn Designing Primers for PCR*.

http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html

Promega, C. (2011). *Assembly of Restriction Enzyme Reactions Technical Manual, TM367*.

www.promega.com/guides/,

Promega, C. (2018). *pGEM[®]-T and pGEM[®]-T Easy Vector Systems Technical Manual #TM042*.

www.promega.com

Rahman, M. S., Islam, M. R., Alam, A. S. M. R. U., Islam, I., Hoque, M. N., Akter, S., Rahaman, M.

M., Sultana, M., & Hossain, M. A. (2021). Evolutionary dynamics of SARS-CoV-2

nucleocapsid protein and its consequences. *Journal of Medical Virology*, *93*(4), 2177–2195.

<https://doi.org/10.1002/JMV.26626>

Rakib, A., Sami, S. A., Islam, M. A., Ahmed, S., Faiz, F. B., Khanam, B. H., Marma, K. K. S., Rahman,

M., Uddin, M. M. N., Nainu, F., Emran, T. Bin, & Simal-Gandara, J. (2020). Epitope-Based

Immunoinformatics Approach on Nucleocapsid Protein of Severe Acute Respiratory

Syndrome-Coronavirus-2. *Molecules*, *25*(21).

<https://doi.org/10.3390/MOLECULES25215088>

Rump, A., Risti, R., Kristal, M.-L., Reut, J., Syritski, V., Lookene, A., & Boudinot, S. R. (2021). Dual

ELISA using SARS-CoV-2 nucleocapsid protein produced in E. coli and CHO cells reveals

epitope masking by N-glycosylation. *Biochemical and Biophysical Research*

Communications, *534*, 457. <https://doi.org/10.1016/J.BBRC.2020.11.060>

Sambrook, J., & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd ed., Vols 1,2

and 3. *Molecular Cloning: A Laboratory Manual*, 16.33-16.36, 16.54.

Satarker, S., & Nampoothiri, M. (2020). Structural Proteins in Severe Acute Respiratory

Syndrome Coronavirus-2. *Archives of Medical Research*, *51*(6), 482.

<https://doi.org/10.1016/J.ARCMED.2020.05.012>

Serra, H. A., Roganovich, J. M., & Rizzo, L. F. L. (2012). *GLUCOCORTICOIDES: PARADIGMA DE MEDICINA TRASLACIONAL DE LO MOLECULAR AL USO CLINICO*.

Thermo Fisher Scientific Inc. (2012). *PRODUCT INFORMATION EcoRI*.

https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FMAN0012089_EcoRI_10_UuL_5000U_UG.pdf

Tian, Y., Zhang, G., Liu, H., Ding, P., Jia, R., Zhou, J., Chen, Y., Qi, Y., Du, J., Liang, C., Zhu, X., & Wang, A. (2022). Screening and identification of B cell epitope of the nucleocapsid protein in SARS-CoV-2 using the monoclonal antibodies. *Applied Microbiology and Biotechnology*, 1. <https://doi.org/10.1007/S00253-022-11769-6>

Timani, K. A., Ye, L., Ye, L., Zhu, Y., Wu, Z., & Gong, Z. (2004). Cloning, sequencing, expression, and purification of SARS-associated coronavirus nucleocapsid protein for serodiagnosis of SARS. *Journal of Clinical Virology*, 30(4), 309. <https://doi.org/10.1016/J.JCV.2004.01.001>

Vos, J. De, Aguilar, P. P., Köppl, C., Fischer, A., Grünwald-Gruber, C., Dürkop, M., Klausberger, M., Mairhofer, J., Striedner, G., Cserjan-Puschmann, M., Jungbauer, A., & Lingg, N. (2021). Production of full-length SARS-CoV-2 nucleocapsid protein from *Escherichia coli* optimized by native hydrophobic interaction chromatography hyphenated to multi-angle light scattering detection. *Talanta*, 235, 122691. <https://doi.org/10.1016/J.TALANTA.2021.122691>

Wang, Q., Li, C., Zhang, Q., Wang, T., Li, J., Guan, W., Yu, J., Liang, M., & Li, D. (2010).

Interactions of SARS Coronavirus Nucleocapsid Protein with the host cell proteasome subunit p42. *Virology Journal*, 7, 99. <https://doi.org/10.1186/1743-422X-7-99>

- Wang, Z. G., Xu, S. P., Zhang, Y. J., & Bao, Q. Y. (2007). Genetic distance of SARS coronavirus from the recent natural case. *Veterinary Microbiology*, *120*(1), 167.
<https://doi.org/10.1016/J.VETMIC.2006.10.022>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. <https://doi.org/10.1093/NAR/GKY427>
- World Health Organization. (2020a, April 27). *Archived: WHO Timeline - COVID-19*.
<https://www.who.int/news/item/27-04-2020-who-timeline---covid-19>
- World Health Organization. (2020b, December 13). *Preguntas y respuestas sobre la transmisión de la COVID-19*. <https://www.who.int/es/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted>
- Wu, C., Qavi, A. J., Hachim, A., Kavian, N., Cole, A. R., Moyle, A. B., Wagner, N. D., Sweeney-Gibbons, J., Rohrs, H. W., Gross, M. L., Peiris, J. S. M., Basler, C. F., Farnsworth, C. W., Valkenburg, S. A., Amarasinghe, G. K., & Leung, D. W. (2021). Characterization of SARS-CoV-2 N protein reveals multiple functional consequences of the C-terminal domain. *BioRxiv*, *24*(6). <https://doi.org/10.1101/2020.11.30.404905>
- Yang, Y., Xiao, Z., Ye, K., He, X., Sun, B., Qin, Z., Yu, J., Yao, J., Wu, Q., Bao, Z., & Zhao, W. (2020). SARS-CoV-2: characteristics and current advances in research. *Virology Journal* *2020* *17*:1, *17*(1), 1–17. <https://doi.org/10.1186/S12985-020-01369-Z>
- Ye, Q., West, A. M. V., Silletti, S., & Corbett, K. D. (2020). Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Science : A Publication of the Protein Society*,

29(9), 1890. <https://doi.org/10.1002/PRO.3909>

Zeng, W., Liu, G., Ma, H., Zhao, D., Yang, Y., Liu, M., Mohammed, A., Zhao, C., Yang, Y., Xie, J., Ding, C., Ma, X., Weng, J., Gao, Y., He, H., & Jin, T. (2020). Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochemical and Biophysical Research Communications*, 527(3), 618. <https://doi.org/10.1016/J.BBRC.2020.04.136>

Zurita & Zurita Laboratorios. (2021). *Unidad de Investigaciones en Biomedicina , Zurita & Zurita Laboratorios*. <http://www.zuritalaboratorios.com/>