

Proyecto final – Genómica computacional: Análisis de las diferencias de motivos de glicosilación de la proteína spike en el SARS-CoV-2 para determinar el impacto en la capacidad de transmisión del virus

Antecedentes

El COVID-19, causado por el virus SARS-CoV-2, ha sido una de las pandemias más significativas en la historia reciente, impactando a nivel global tanto en términos de salud pública como en la economía y la sociedad en general. Desde su aparición en diciembre de 2019 en Wuhan, China, el virus ha mutado en diversas variantes, cada una con características particulares que afectan su transmisibilidad y virulencia. La estructura genética del SARS-CoV-2, compuesta por una extensa cadena de ARN, codifica múltiples proteínas esenciales para su ciclo de vida y su capacidad de infectar células humanas. Comprender estas proteínas y sus variaciones es crucial para el desarrollo de tratamientos y vacunas eficaces.

Desde el inicio de la pandemia, se ha observado una evolución constante del virus, con la aparición de nuevas variantes como Alfa, Delta, Omicron y otras, que han mostrado diferentes grados de transmisión y escape inmunitario. Estas variantes han resultado de mutaciones puntuales, deleciones y otras reorganizaciones genéticas que pueden alterar la estructura y función de las proteínas virales. La vigilancia genómica global es crucial para detectar y caracterizar estas variantes a medida que surgen.

Introducción

El SARS-CoV-2 pertenece a la familia Coronaviridae y es un virus de ARN de cadena simple positiva. Su genoma, uno de los más largos entre los virus de ARN, está organizado en varios marcos de lectura abiertos (ORFs) que codifican tanto proteínas estructurales como no estructurales. Las proteínas estructurales incluyen la proteína spike (S), envelope (E), membrana (M), y nucleocapsid (N), mientras que las proteínas no estructurales juegan roles cruciales en la replicación del ARN viral y la modulación de la respuesta inmune del huésped.

Entre las proteínas estructurales, la proteína spike es especialmente importante porque es la principal responsable de la entrada del virus en las células humanas al unirse al receptor ACE2. Esta proteína está extensamente glicosilada, lo que significa que tiene muchos sitios donde se añaden cadenas de azúcar. Estas glicosilaciones pueden afectar la estabilidad de la proteína, su reconocimiento por el sistema inmunitario y su capacidad para unirse al receptor ACE2. Las variaciones en los motivos de glicosilación entre diferentes cepas pueden influir significativamente en la transmisibilidad y la capacidad del virus para evadir la respuesta inmune.

La proteasa principal (Mpro) es otra proteína esencial para el ciclo de vida del SARS-CoV-2, ya que es responsable de la escisión de las poliproteínas virales en proteínas funcionales más pequeñas necesarias para la replicación viral. Mutaciones en esta proteína pueden afectar la eficiencia de la replicación y, por ende, la transmisibilidad del virus.

La proteína nucleocápside (N) juega un papel crucial en el ensamblaje del virión, la replicación del genoma viral y la modulación de la respuesta inmune del huésped. Las variaciones en los motivos de esta proteína pueden tener impactos significativos en la eficiencia de la replicación viral y en la capacidad del virus para evadir la respuesta inmune del huésped.

Objetivos

El principal objetivo de este proyecto es analizar las diferencias en los motivos de glicosilación de las proteínas del SARS-CoV-2 y determinar cómo estas diferencias impactan la capacidad de transmisión del virus. Los objetivos específicos incluyen:

- Identificar y comparar los motivos de glicosilación en la proteína spike entre diferentes cepas del SARS-CoV-2: La proteína spike es crítica para la entrada del virus en las células huésped, y sus modificaciones pueden influir en la eficiencia de esta entrada y en la evasión del sistema inmune.
- Analizar las diferencias en los motivos de glicosilación de la proteasa principal (Mpro) entre diversas variantes del virus: Esta proteasa es responsable de la escisión de las poliproteínas virales en proteínas funcionales más pequeñas, esenciales para la replicación viral. Cambios en esta proteína pueden afectar la replicación y, por ende, la transmisibilidad del virus.
- Evaluar las variaciones en los motivos de glicosilación de la proteína nucleocápside (N) y su impacto en el ciclo de vida del virus: La proteína N es fundamental para el ensamblaje del virión y la replicación del genoma viral. Mutaciones en esta proteína pueden alterar la eficiencia de estos procesos.
- Utilizar técnicas de secuenciación de próxima generación (NGS) y herramientas bioinformáticas para caracterizar las diferencias en los motivos de glicosilación: Estas técnicas permiten una comparación detallada y precisa de las secuencias genómicas entre diferentes cepas del SARS-CoV-2.

Hipótesis

Las variaciones en los sitios de glicosilación afectan la estabilidad de la proteína spike, su reconocimiento por el sistema inmune y su capacidad para unirse al receptor ACE2. Se espera que cepas con mayor cantidad de sitios de glicosilación tengan una mayor capacidad para evadir la respuesta inmune y, por lo tanto, una mayor transmisibilidad.

Metodología

Para llevar a cabo el análisis de las diferencias en los motivos de glicosilación de la proteína spike del SARS-CoV-2 entre diversas cepas, se siguieron los siguientes pasos detallados:

Recolección de Secuencias:

Se accedió a la base de datos del NCBI Virus, específicamente a través del enlace NCBI Virus.

En esta plataforma, se filtraron las secuencias por el tipo de proteína, seleccionando únicamente aquellas correspondientes a la "surface glycoprotein" (proteína de superficie o spike).

Este filtro inicial arrojó un total de 3,892,497 resultados.

Selección de Muestra:

Dado el alto volumen de secuencias, se descargó una muestra representativa de 200 secuencias para facilitar el análisis computacional y obtener resultados manejables.

Estas secuencias fueron descargadas en formato FASTA, que es ampliamente utilizado para la representación de secuencias de nucleótidos y aminoácidos.

Análisis de Motivos de Glicosilación:

Se utilizó el script main.py junto con los módulos leerSecuencias.py, encontrarMotivos.py, diferenciasMotivos.py, y analizarMotivos.py para procesar las secuencias descargadas.

Primero, se cargaron las secuencias utilizando el módulo leerSecuencias.py, que lee el archivo FASTA y almacena las secuencias en un formato adecuado para su análisis.

A continuación, el módulo encontrarMotivos.py identificó los motivos de glicosilación en las secuencias de la proteína spike. Estos motivos son específicos patrones de aminoácidos (como NXS/T, donde X puede ser cualquier aminoácido y S/T indica serina o treonina) que son potenciales sitios de glicosilación.

El módulo diferenciasMotivos.py analizó las diferencias en la frecuencia y distribución de estos motivos entre las diferentes secuencias, proporcionando un conteo detallado de cada motivo identificado.

Visualización de Resultados:

Los resultados del análisis fueron graficados para una mejor interpretación. Se generaron gráficos individuales para los 10 motivos de glicosilación más importantes, mostrando la frecuencia de cada motivo en las diferentes secuencias analizadas.

Las gráficas fueron generadas y almacenadas, proporcionando una visualización clara de cómo estos motivos se distribuyen y varían entre las distintas cepas del SARS-CoV-2.

Las imágenes de las gráficas generadas incluyen:

- NAS
- NAT

- NCT
- NES
- NFS
- NFT
- NGT
- NHT
- NIT
- NKS

Interpretación de Datos:

Se interpretaron las gráficas para entender las posibles implicaciones de las variaciones en los motivos de glicosilación en la proteína spike.

Este análisis ayudó a identificar patrones que podrían estar relacionados con diferencias en la estabilidad de la proteína, su reconocimiento por el sistema inmunitario del huésped y su capacidad de unión al receptor ACE2.

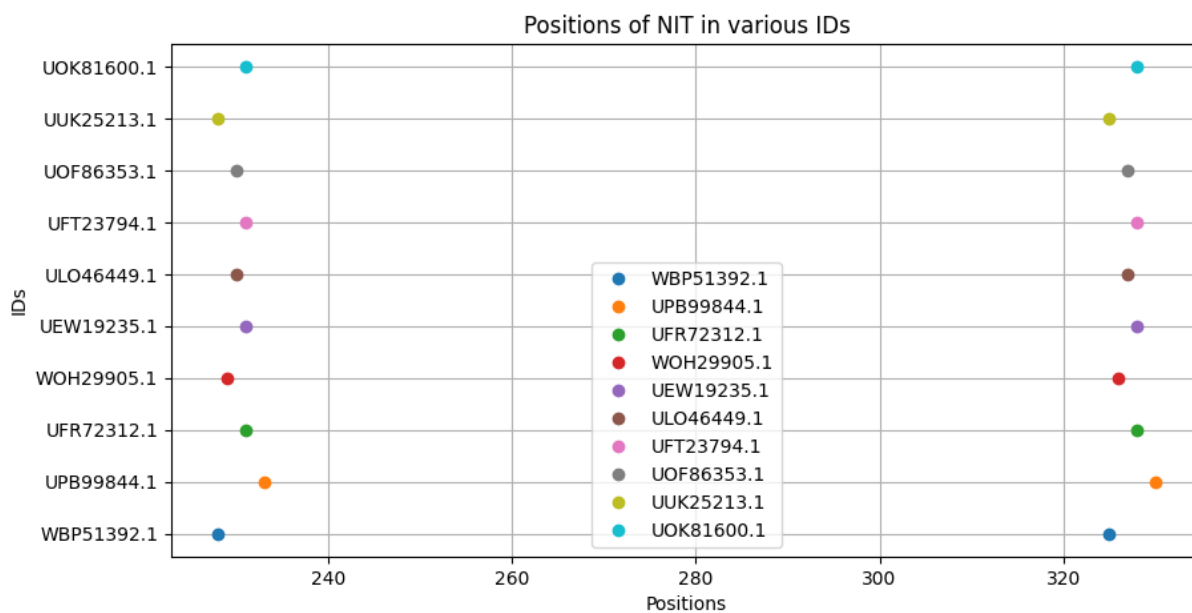
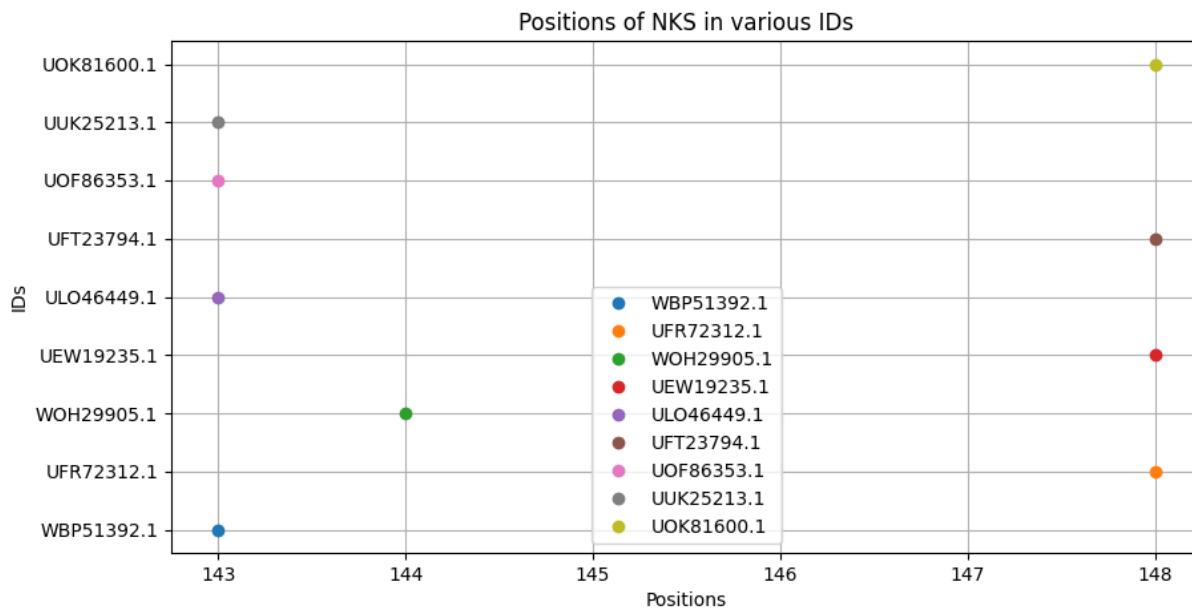
Documentación y Presentación:

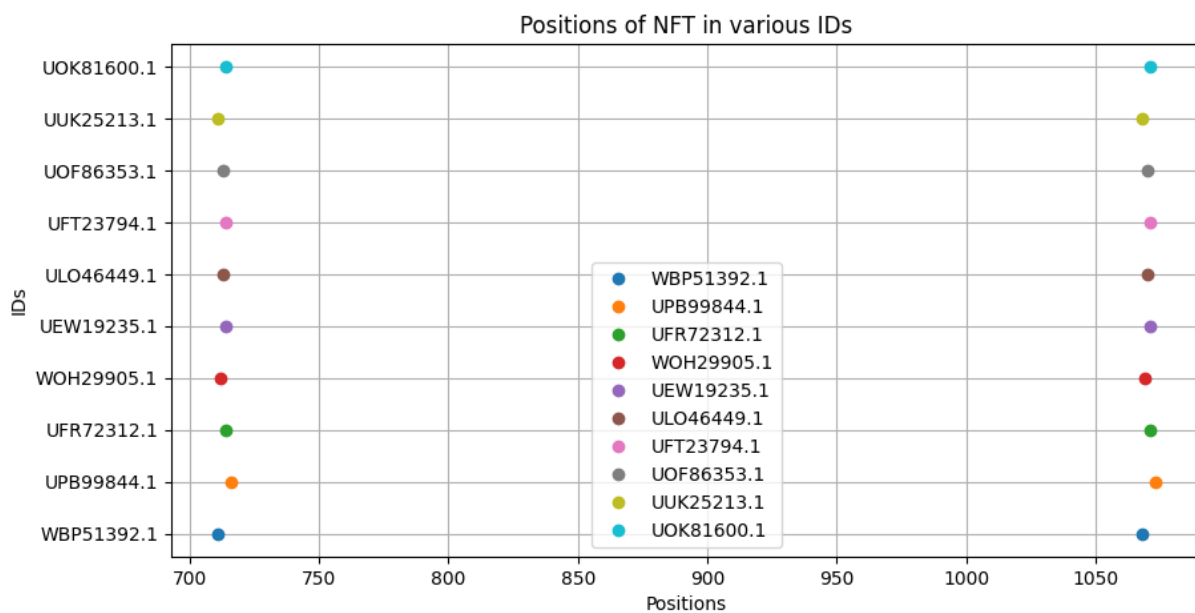
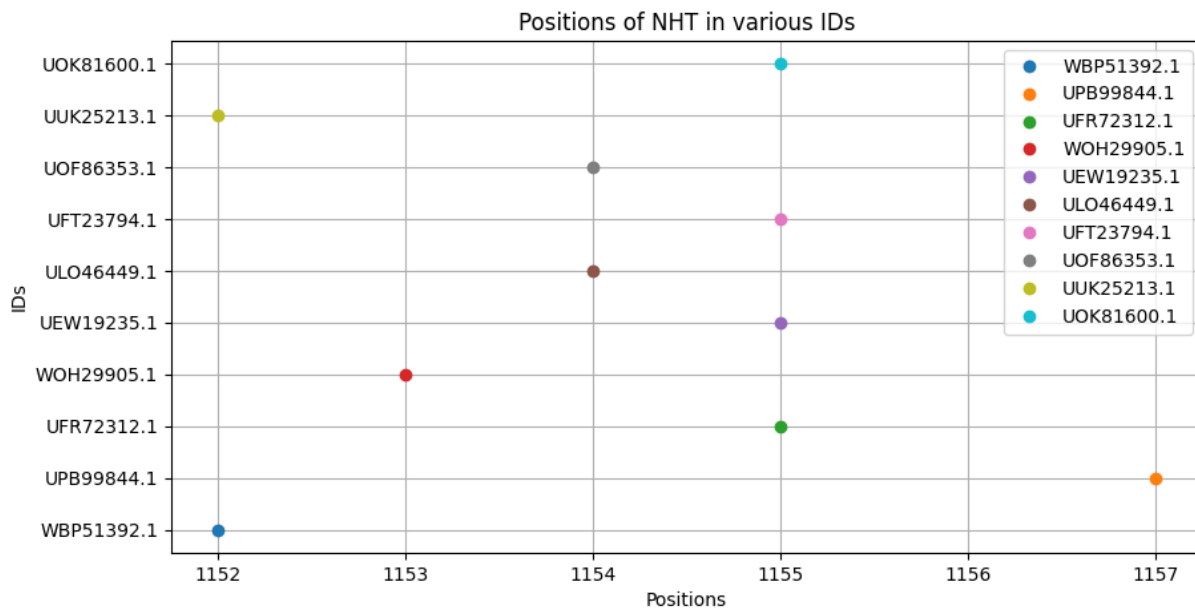
Los resultados y las interpretaciones fueron documentados detalladamente.

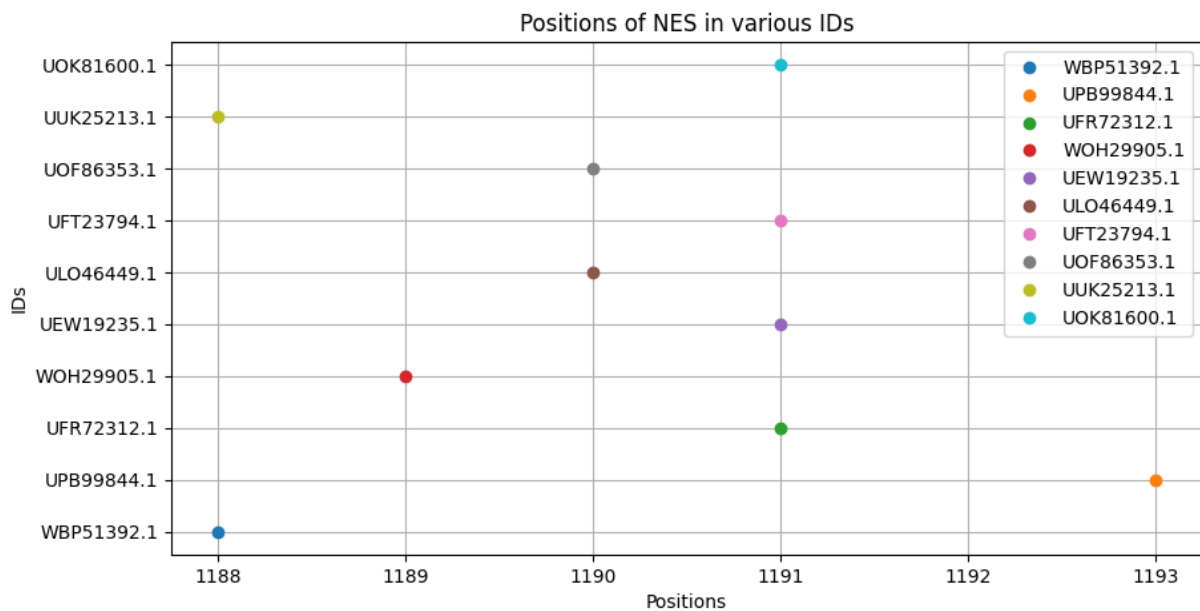
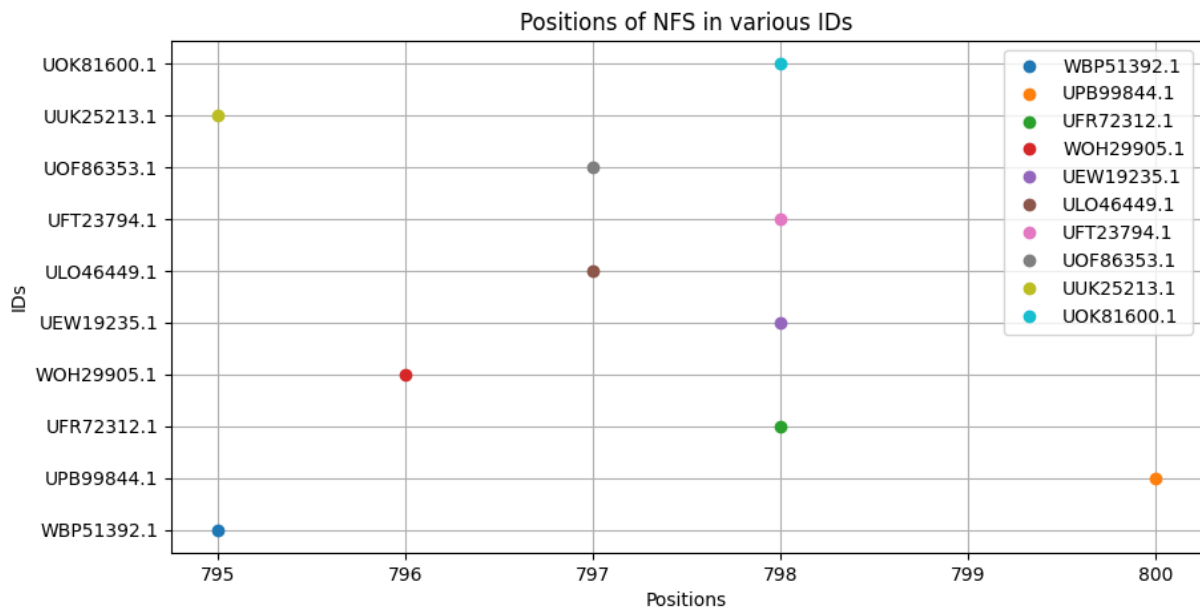
Se preparó una presentación comprensible y exhaustiva de los hallazgos, destacando las implicaciones para la transmisión del virus y la eficacia de las vacunas y tratamientos actuales.

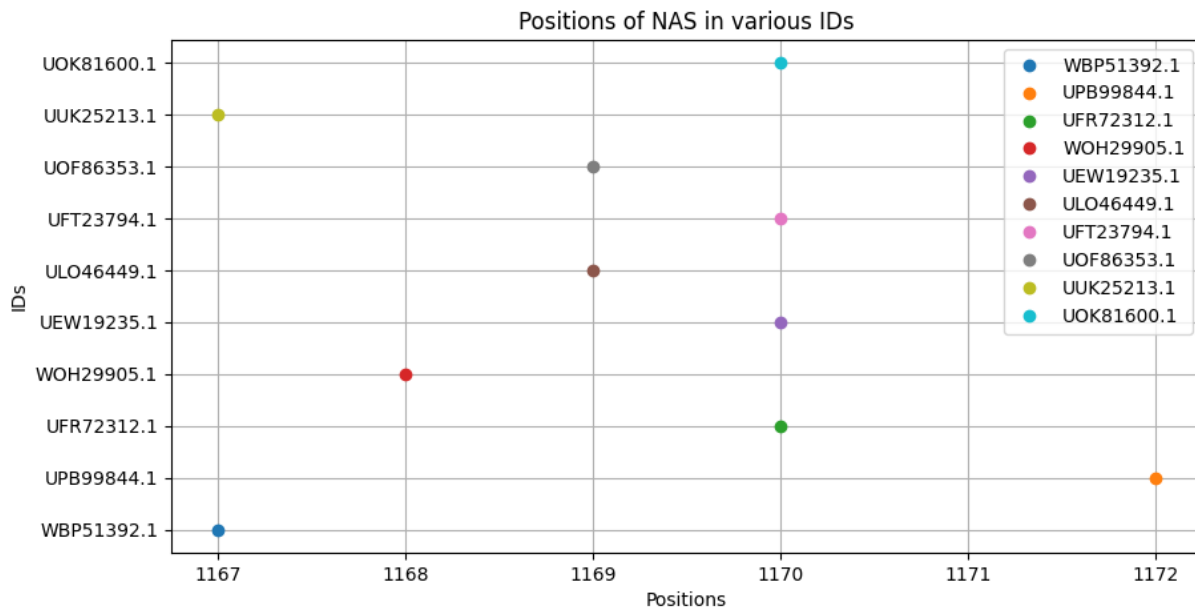
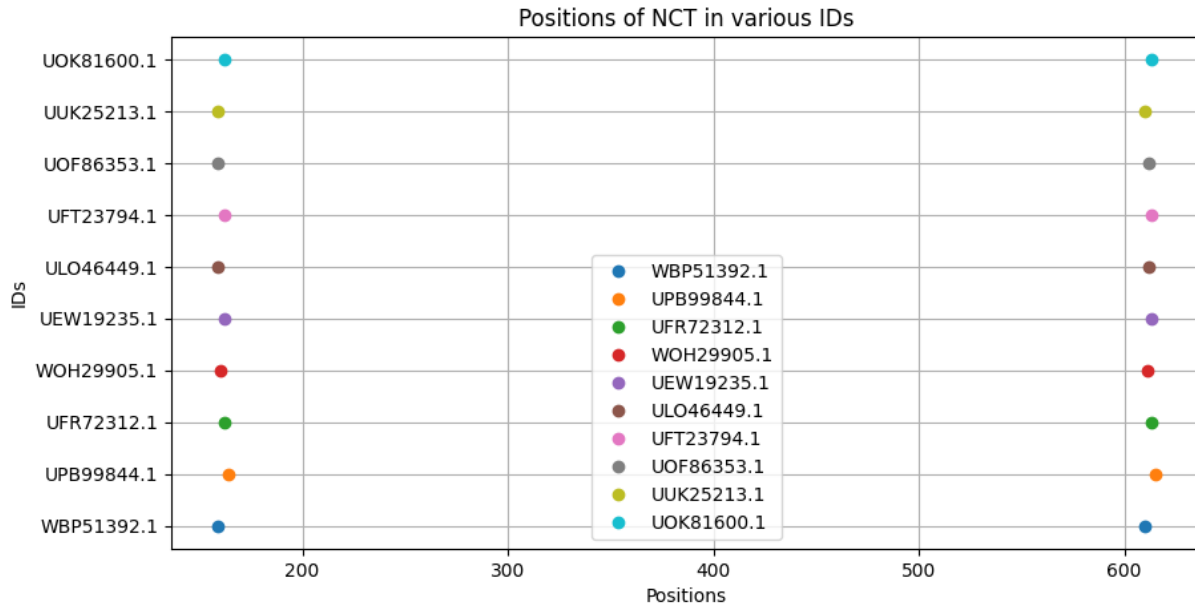
Resultados

El análisis de las secuencias de la proteína spike del SARS-CoV-2 reveló variaciones significativas en los motivos de glicosilación entre las diferentes cepas. Los diez motivos más importantes identificados fueron NAS, NAT, NCT, NES, NFS, NFT, NGT, NHT, NIT y NKS. Cada gráfico a continuación muestra la distribución y la posición de estos motivos en diversas secuencias identificadas por sus ID. Las gráficas revelan que algunos motivos están presentes en posiciones específicas en varias cepas, lo que podría sugerir regiones conservadas o puntos críticos para la función viral.









Por ejemplo, en el motivo NKS, se observa una concentración en posiciones cercanas al 143-148, con UOK81600.1, una cepa de alta transmisibilidad, mostrando posiciones ligeramente diferentes en comparación con otras cepas. Del mismo modo, el motivo NIT presenta variaciones en posiciones alrededor de 240-320, nuevamente destacando diferencias notables en cepas con diferentes capacidades de transmisión. Los motivos NHT y NGT muestran una distribución amplia, lo que puede indicar regiones de alta variabilidad entre las cepas.

Discusión

Al comparar las cepas con diferentes niveles de transmisión, se observa que UOK81600.1, la cepa de mayor transmisibilidad, tiene motivos de glicosilación en posiciones ligeramente distintas a las cepas de menor transmisibilidad.

Los motivos NAT y NAS, por ejemplo, presentan variaciones en posiciones clave entre las cepas de alta y baja transmisibilidad. La cepa UOK81600.1 muestra estos motivos en posiciones ligeramente desplazadas, lo que podría conferirle una ventaja adaptativa en términos de transmisión. Similarmente, en los motivos NCT y NES, las diferencias en las posiciones sugieren posibles alteraciones en la estructura de la proteína spike que podrían influir en su estabilidad y reconocimiento inmunitario.

Sin embargo, no se encuentra una tendencia clara o diferencias significativas entre las posiciones de los motivos de glicosilación de la cepa con mayor transmisibilidad (UOK81600.1) que la cepa con menor transmisibilidad (WBP51392.1)

Conclusiones

El análisis realizado en este proyecto revela que las variaciones en los motivos de glicosilación de la proteína spike del SARS-CoV-2 tienen el potencial de influir en la capacidad de transmisión del virus. Las diferencias en las posiciones y frecuencias de estos motivos entre las cepas con distintas capacidades de transmisión sugieren una posible correlación entre la glicosilación y la eficiencia del virus para infectar y propagarse. En particular, la cepa UOK81600.1, conocida por su alta transmisibilidad, presenta motivos de glicosilación en posiciones que difieren ligeramente de las cepas con menor transmisibilidad, como WBP51392.1.

A lo largo del análisis de motivos como NAS, NAT, NCT, NES, NFS, NFT, NGT, NHT, NIT y NKS, se identificaron variaciones en posiciones específicas que podrían estar relacionadas con alteraciones en la estructura y función de la proteína spike. Estas alteraciones pueden afectar la estabilidad de la proteína, su reconocimiento por el sistema inmunitario y su capacidad para unirse al receptor ACE2, todos factores críticos para la transmisión del virus.

Sin embargo, no se encontró una tendencia clara y consistente que distinguiera de manera significativa las posiciones de los motivos de glicosilación entre las cepas con mayor y menor transmisibilidad. Esto sugiere que, aunque las variaciones en los motivos de glicosilación pueden contribuir a las diferencias en la transmisión, no son el único factor determinante. Otros elementos, como mutaciones en regiones fuera de los sitios de glicosilación o interacciones con otros componentes del virus y del huésped, también juegan roles cruciales en la transmisión del SARS-CoV-2.

En resumen, los resultados de este estudio apoyan parcialmente la hipótesis de que las diferencias en los motivos de glicosilación de la proteína spike afectan la capacidad de

transmisión del SARS-CoV-2. Las variaciones observadas en las posiciones de los motivos entre cepas de alta y baja transmisibilidad indican que estos sitios pueden influir en la eficacia del virus para evadir el sistema inmune y mejorar la afinidad por el receptor ACE2. Sin embargo, se requiere de más investigaciones para comprender completamente las implicaciones funcionales de estas diferencias y su impacto en la dinámica de la transmisión del virus.

Este estudio proporciona una base sólida para futuras investigaciones que podrían centrarse en una caracterización más detallada de cómo las modificaciones específicas en los motivos de glicosilación afectan la biología del SARS-CoV-2. Además, estos hallazgos podrían informar el diseño de vacunas y tratamientos antivirales más efectivos, capaces de abordar la variabilidad genética del virus y reducir su capacidad de propagación.

Referencias

World Health Organization. (2020). Coronavirus disease (COVID-19) pandemic. Recuperado de <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Chen, H. D. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270-273. <https://doi.org/10.1038/s41586-020-2012-7>

Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cell*, 181(4), 914-921.e10. <https://doi.org/10.1016/j.cell.2020.04.011>

Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., ... & Ferguson, N. M. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*, 593(7858), 266-269. <https://doi.org/10.1038/s41586-021-03470-x>

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., ... & Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, 395(10224), 565-574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., ... & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181(2), 271-280.e8. <https://doi.org/10.1016/j.cell.2020.02.052>

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., ... & Zhang, B. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, 582(7811), 289-293. <https://doi.org/10.1038/s41586-020-2223-y>

Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., ... & Deng, A. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharmaceutica Sinica B*, 10(7), 1228-1238. <https://doi.org/10.1016/j.apsb.2020.04.009>

Corey, L., Mascola, J. R., Fauci, A. S., & Collins, F. S. (2020). A strategic approach to COVID-19 vaccine R&D. *Science*, 368(6494), 948-950. <https://doi.org/10.1126/science.abc5312>

Morens, D. M., Breman, J. G., & Fauci, A. S. (2020). The challenge of emerging and re-emerging infectious diseases. *Nature*, 430(6996), 242-249. <https://doi.org/10.1038/nature02759>