

Papel de la bioinformática en la mitigación del impacto de la pandemia del COVID-19

Milton Vindas Rodríguez

mvindasr@ucenfotec.ac.cr

*Universidad Cenfotec
Bachillerato en Ingeniería del Software*

Abstract

El síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2) es responsable de la enfermedad del coronavirus 2019 (COVID-19), que ha ejercido una enorme presión sobre el sistema de salud mundial y ha cobrado millones de vidas a nivel global. Una explosión de investigaciones científicas en el área de la bioinformática ha sido la respuesta para estudiar y afrontar este virus novedoso y que ha aportado de herramientas y técnicas avanzadas con diferentes propósitos que se discutirán brevemente en este artículo: secuenciación de próxima generación del virus, generación de bases de datos biológicas, detección de la enfermedad, tratamiento terapéutico y el rastreo y contención de la propagación de la pandemia.

Palabras Clave

Bioinformática, tecnología, pandemia, COVID-19, vacunación

Introducción

La bioinformática es un campo interdisciplinario que utiliza métodos de simulación computacional para analizar datos biológicos y hacer predicciones de diversa índole basados en estos. Se ha utilizado con éxito en la investigación de patogénesis de enfermedades, en la síntesis de vacunas, en inmunología y en epidemiología (Ishack y Lipner, 2021). La aplicación de la informática en la investigación biológica y en el desarrollo de nuevos fármacos nunca ha sido tan esencial en la lucha contra las

enfermedades infecciosas como en la actualidad. El mayor combate del siglo XXI contra un agente de enfermedad debilitante, el virus SARS-CoV-2 ha provocado un uso sin precedentes de herramientas bioinformáticas para descifrar las caracterizaciones moleculares de un patógeno infeccioso. Con los datos del genoma viral disponibles apenas unas semanas después del informe del brote, las plataformas bioinformáticas se han convertido en una herramienta fundamental para mitigar el impacto de la pandemia.

Desarrollo del tema

En diciembre de 2019, se informó de un brote de neumonía sin precedentes en Wuhan, China. Hubo una rápida propagación global de la enfermedad desde el epicentro al resto del mundo, con miles de personas infectadas y varios casos de muerte reportados, elevando la enfermedad a una pandemia. Para el 31 de diciembre de 2019, el brote de la enfermedad se rastreó hasta una nueva cepa de coronavirus, que más tarde se denominó SARS-CoV-2 y la enfermedad como COVID-19 por la Organización Mundial de la Salud (WHO, 2020). Desde entonces, el SARS-CoV-2 se logró identificar como una nueva cepa del grupo 2B de coronavirus, con

aproximadamente un 70% de homología genética con el SARS-CoV del brote del 2003 en el mismo país (Hui *et al.*, 2020). Debido a que el virus tiene una similitud del 96% con un coronavirus de murciélago, este es ampliamente aceptado en la comunidad científica como su origen. La pandemia resultó en restricciones de viaje sin precedentes a nivel nacional e internacional, junto con graves trastornos económicos. El aporte de la bioinformática durante este proceso de velocidad sin precedentes en la búsqueda de soluciones se discute a continuación.

Detección del virus

El método de detección de rutina para el SARS-CoV-2 es una reacción en cadena de la polimerasa en tiempo real (qRT-PCR). La técnica de PCR, cuya invención le valió a Kary Mullis el premio Nobel de Química en 1993, permite obtener muestras de alta calidad y cantidad de ADN amplificando concentraciones muy bajas de la molécula. La prueba para el COVID-19 se basa en la detección de dos secuencias: el gen de la envoltura del virus (E) y un gen que participa en la replicación del virus (RdRp), y por medio de sistemas informáticos, se realiza una detección en tiempo real (Corman *et al.*, 2020). Además, la especificidad (exclusión de falsos positivos) y la sensibilidad (exclusión de falsos negativos) son dos de los criterios de calidad más importantes para la validez de las pruebas diagnósticas. Para garantizar la identificación única de SARS-CoV-2, se requiere un análisis computacional continuo de los reactivos utilizados para detectar y amplificar estos genes, ya que mutaciones del virus en

estas secciones del genoma podrían afectar los resultados de la prueba.

Secuenciación y bases de datos

Además de la detección por PCR, el análisis del genoma completo juega un papel crucial en las respuestas de salud pública, incluidos los esfuerzos epidemiológicos para rastrear y contener el brote. La secuencia del genoma de SARS-CoV-2 se determinó rápidamente y se compartió en el sitio de bases de datos genómicos GenBank. La metodología de secuenciación de próxima generación (NGS, por sus siglas en inglés) se puede utilizar para secuenciar muestras variadas del virus en paralelo para así evaluar su diversidad genómica. La secuenciación regular de casos clínicos es útil, por ejemplo, para monitorear mutaciones que puedan afectar la prueba de detección. El protocolo de secuenciación (llamado también de nanoporos) para el SARS-CoV-2 ha sido desarrollado y optimizado por la red ARTIC, que tiene una amplia experiencia y conocimientos en el despliegue de esta tecnología en la secuenciación y vigilancia de brotes, incluidos el zika y el ébola (Quick *et al.*, 2017). Esto posibilita el almacenamiento continuo en bases de datos de genomas de alta precisión de SARS-CoV-2 y el rastreo tanto de la transmisión de COVID-19 como la evolución viral a lo largo del tiempo. Estos metadatos están disponibles a través de bases de datos abiertas como, por ejemplo, en el sitio web GenBank, o de acceso a través de un acuerdo como en GISAID (2021), en donde los laboratorios de todo el mundo han compartido datos de secuencias del genoma viral con una velocidad sin precedentes (> 4.5 millones de secuencias

genómicas del SARS-CoV-2 para el 25 de octubre del 2021).

Análisis filogenéticos

Otra grave pandemia que afronta simultáneamente la humanidad, incrementada por el descontrol en las redes sociales, es la de la desinformación. A pesar de todas las pruebas basadas en el método científico, existen grupos alimentados por el miedo y la ignorancia, que han publicado historias de conspiración a través de los medios de comunicación religiosos, políticos y de pseudociencia. Uno de los casos más comentados fue un artículo publicado en Nature (Menachery *et al.*, 2015), donde se sugiere que un coronavirus quimérico de ratón/murciélago, SL-SHC014, fue ensamblado artificialmente y que se utilizó en estudios de mejoramiento de la infecciosidad del coronavirus en ratones; los grupos anteriores especularon que la actual enfermedad podría ser de origen artificial por un escape accidental. Estudios de científicos de renombre, basados en análisis de la filogenia del virus, es decir, el establecimiento del origen y parentesco de un organismo basados en la secuencia del ADN, han brindado pruebas sólidas del origen natural del SARS-CoV-2 y que los coronavirus artificiales conocidos difieren demasiado para ser considerados responsables de la pandemia. Uno de estos es la investigación de Dallavilla y colegas (2020) los cuales realizaron un estudio bioinformático con las secuencias en bases de datos de coronavirus patogénicos, así como de coronavirus artificiales patentados y concluyeron irrefutablemente que el SARS-CoV-2

difiere demasiado de cualquier coronavirus sintetizado en el laboratorio, incluyendo el SL-SHC014 y que los datos apoyan el origen natural del virus, con alta probabilidad de ser derivado de murciélagos, posiblemente transferidos a pangolines, antes de extenderse al ser humano.

Construcción de modelos estructurales

Otro término que ha tenido un resurgimiento gracias a aplicaciones bioinformáticas es el de “reutilización de fármacos”. Como el COVID-19 es reciente y porque los ensayos clínicos suelen durar varios años y existe una alta diversidad, no existen datos de la efectividad de medicamentos ya existentes contra esta enfermedad, los cuales tienen efectos secundarios mínimos y que son efectivos ante otras enfermedades. Como caso específico, Singh y Florez (2020) utilizaron un software de proyección virtual basado en Python, PyRx, que permite hacer predicciones del potencial de unión de la estructura molecular de diversos fármacos con objetivos blanco en el virus para inactivarlos, utilizando bases de datos bioquímicas, con el cual lograron reducir a 13 moléculas aquellas con posibilidad alta de emplearse en ensayos clínicos y en el tratamiento de casos positivos. Otras aplicaciones disponibles en línea de esta clase incluyen CORDITE, VirHostNet y CoVex.

Diseño de vacunas y fármacos

Los métodos de síntesis de vacunas convencionales (VC) son inadecuados para esta pandemia debido a que: (i) la identificación de antígenos – las moléculas que forman parte del virus y

que generan una respuesta del sistema inmune – consume mucho tiempo y debido a que es un microorganismo reciente, la diversidad de antígenos documentados es escasa, (ii) se requiere el cultivo extensivo del patógeno en laboratorios y (iii) los altos costos (Soria *et al.*, 2014). Los métodos *in silico* de la bioinformática se pueden aplicar eficazmente al diseño avanzado de vacunas, con un tiempo de procesamiento más rápido que la metodología convencional. Destacan dos estrategias informáticas principales, la vacunología inversa (RV) y la vacunología estructural (SV). En la sección anterior, se abordó como el análisis estructural de fármacos (SV) ya existentes tiene el potencial de encontrar tratamientos contra el COVID-19. La vacunología RV es un proceso de desarrollo de vacunas que implica la identificación de nuevos antígenos, no analizando *in vitro* estas moléculas como en la VC ni analizando su estructura, si no que se va a la raíz de todo: el ADN del microorganismo. Mediante softwares avanzados como VaxiJen (Doytchinova y Flower, 2007) se identifican aquellas regiones que se expresan en proteínas dentro del genoma del virus y mediante el análisis de datos, se refinan los candidatos al evaluar propiedades antigénicas y fisicoquímicas de interés que puedan utilizarse en la síntesis de vacunas.

Análisis de datos e inteligencia artificial

El análisis y la visualización de datos es una herramienta esencial para explorar y comunicar los resultados de la investigación médica, especialmente en la vigilancia epidemiológica. Puede ayudar a los investigadores y legisladores a identificar tendencias que podrían pasarse

por alto si los datos se revisaran en forma de tabla y esta facilidad ha sido brindada por aplicaciones desarrolladas desde el ámbito bioinformático en la lucha contra la epidemia del SARS-CoV-2. Una de ellas es COVID19-World que produce sistemáticamente la visualización de datos específicos de cada país y análisis de tendencias y proyecciones a corto plazo para casos diagnosticados y muertes, tasa de letalidad, tiempo de infección, y número de reproducción básico (Tebé *et al.*, 2020).

El *machine learning* incorpora la aplicación de inteligencia artificial (IA) que proporciona a la tecnología informática la capacidad automatizada de aprender y mejorar experiencias sin haber sido programada explícitamente. Las herramientas cuyo modo de operación se centra en la inteligencia artificial y el aprendizaje profundo se exploraron por completo durante el auge de la pandemia para predecir y diseñar candidatos a vacunas y fármacos. Algunos softwares de este ámbito que se han utilizado ampliamente para predecir moléculas antigénicas son Sprint, el paquete de Python modlAMP y TepiTool (Chukwudozie *et al.*, 2021).

Conclusión

La bioinformática mundial ha reaccionado rápidamente a la pandemia de COVID-19 al proporcionar herramientas específicas para avanzar en la investigación del SARS-CoV-2 y mejorar la detección, la comprensión y el tratamiento. A pesar de que esta revisión es solamente una pincelada de estos aportes, no cabe duda de que, en esta época de crisis, el desarrollo del área y de las bases de datos biológicas ha sido muy

acelerada, y conforme avancen los meses, se implementarán más herramientas que si siguen la tendencia, serán softwares disponibles de forma gratuita y de código abierto, para invitar a otros desarrolladores a mejorarlo. La bioinformática es un campo poderoso para avanzar en la investigación médica y no debe abandonarse por miedos irracionales. Estos miedos históricamente han impedido el acceso a una mejor atención y contramedidas más rápidas a las amenazas para la salud humana.

Referencias

- Chukwudozie, O. S., Duru, V. C., Ndiribe, C. C., Aborode, A. T., Oyebanji, V. O., & Emikpe, B. O. (2021). The Relevance of Bioinformatics Applications in the Discovery of Vaccine Candidates and Potential Drugs for COVID-19 Treatment. *Bioinformatics and biology insights*, 15, 11779322211002168. <https://doi.org/10.1177/11779322211002168>
- Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M. L., Mulders, D. G., Haagmans, B. L., van der Veer, B., van den Brink, S., Wijsman, L., Goderski, G., Romette, J. L., Ellis, J., Zambon, M., Peiris, M., ... Drosten, C. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro surveillance: bulletin European sur les maladies transmissibles = European communicable disease bulletin*, 25(3), 2000045. <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>
- Dallavilla, T., Bertelli, M., Morresi, A., Bushati, V., Stuppia, L., Beccari, T., Chiurazzi, P., Marceddu, G. (2020). Bioinformatic analysis indicates that SARS-CoV-2 is unrelated to known artificial coronaviruses. *Eur Rev Med Pharmacol Sci* 24(8). 4558-4564. doi: 10.26355/eurrev_202004_21041
- Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics*, 8, 4. <https://doi.org/10.1186/1471-2105-8-4>
- GISAID. (2021). hCoV-19 data sharing via GISAID. Recuperado de: <https://www.gisaid.org/>
- Hui, D. S., I Azhar, E., Madani, T. A., Ntoumi, F., Kock, R., Dar, O., Ippolito, G., Mchugh, T. D., Memish, Z. A., Drosten, C., Zumla, A., & Petersen, E. (2020). The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. *International journal of infectious diseases: IJID: official publication of the International Society for Infectious Diseases*, 91, 264–266. <https://doi.org/10.1016/j.ijid.2020.01.009>
- Ishack, S., & Lipner, S. R. (2021). Bioinformatics and immunoinformatics to support COVID-19 vaccine development. *Journal of medical virology*, 93(9), 5209–5211. <https://doi.org/10.1002/jmv.27017>
- Menachery, V. D., Yount, B. L., Jr, Debbink, K., Agnihothram, S., Gralinski, L. E., Plante, J. A., Graham, R. L., Scobey, T., Ge, X. Y., Donaldson, E. F., Randell, S. H., Lanzavecchia, A., Marasco, W. A., Shi, Z. L., & Baric, R. S. (2015). A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature medicine*, 21(12), 1508-1513. <https://doi.org/10.1038/nm.3985>
- Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T. F., Beutler, N. A., Burton, D. R., Lewis-Ximenez, L. L., de Jesus, J. G., Giovanetti, M., Hill, S. C., Black, A., Bedford, T., Carroll, M. W., Nunes, M., Alcantara, L. C., Jr, ... Loman, N. J. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature protocols*, 12(6), 1261–1276. <https://doi.org/10.1038/nprot.2017.066>
- Singh, S., & Florez, H. (2020). Bioinformatic study to discover natural molecules with activity against COVID-19. *F1000Research*, 9, 1203. <https://doi.org/10.12688/f1000research.26731.1>
- Soria-Guerra, R. E., Nieto-Gomez, R., Govea-Alonso, D.O., Rosales-Mendoza, S. (2014).

An overview of bioinformatics tools for epitope prediction: implications on vaccine development. J Biomed Inform. 53:405-414. doi: 10.1016/j.jbi.2014.11.003.

Tebé, C., Valls, J., Satorra, P., & Tobías, A. (2020). COVID19-world: a shiny application to perform comprehensive country-specific data visualization for SARS-CoV-2 epidemic. BMC medical research methodology, 20(1), 235. <https://doi.org/10.1186/s12874-020-01121-9>

World Health Organization (WHO). (2000). Statement regarding cluster of pneumonia cases in Wuhan, China. Recuperado de: www.who.int.

“Milton Vindas es estudiante de la Universidad Cenfotec y cursa la carrera de Ingeniería del Software. Cuenta con una Ingeniería en Biotecnología del Instituto Tecnológico de Costa Rica y con una maestría en Microbiología de la Universidad de Costa Rica. Ha realizado investigaciones y publicaciones relacionadas con el campo de la Bioinformática.”