

Visualización de reportes genómicos automatizados de SARS-Cov2

Autores Colombia unit

- Johan Fabian Bernal
- Pilar Donado-Godoy
- Gabriel Beltrán
- Ángela García

Autores India unit

- Varun Shamanna
- Ravi Kumar

Autores central hub

- Anthony Underwood
- David M Aanensen

*Colombian Integrated Program for Antimicrobial Resistance Surveillance – Coipars, CI Tibaitatá, Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA), Tibaitatá – Mosquera, Cundinamarca, Colombia

^a Central Research Laboratory, Kempegowda Institute of Medical Sciences, Bengaluru, India

^o Center for Genomic Pathogen Surveillance – NIHR Global Health Research Unit (GHRU). Wellcome Sanger Institute. United Kingdom

Introducción

La reciente pandemia por COVID-19 ha acelerado la implementación de metodologías de laboratorio molecular y análisis bioinformático, como la secuenciación de genomas completos de microorganismos (mWGS) en la red de laboratorios en salud pública. Existe evidencia que las mutaciones y variaciones pueden influir en factores como la transmisibilidad, el desenlace de la enfermedad y la eficacia de la vacuna, a medida que evoluciona la población viral. En Colombia, la consolidación de un consorcio de instituciones con capacidades en mWGS de virus, permite una respuesta más apropiada a los requerimientos globales y locales de la actual emergencia sanitaria. Frente al crecimiento acelerado de la aplicación de mWGS en Colombia, surgió una demanda inusitada de personal capacitado tanto en el laboratorio como en bioinformática que, hiciera frente a la escasez en el país de dicho personal, lo que representa un cuello de botella en la implementación de mWGS y su uso como herramienta de decisión. El objetivo de este trabajo es proveer un reporte automático del análisis genómico de SARS-CoV2, resultado de un proceso computacional de referencia, bajo en requerimientos, fácilmente comunicable y de acceso público.

Métodos

La “Global Health Research Unit (GHRU)” las instituciones AGROSAVIA, KIMSBI y SANGER, se propusieron replicar la estructura de análisis automatizado desarrollado por el consorcio en resistencia antimicrobiana, principalmente el Centro de vigilancia genómica de patógenos (cGPS) de la Universidad de Oxford y el Instituto Sanger Wellcome Trust, usando plataformas y recursos de referencia de libre acceso, para la generación de un proceso para SARS-Cov2 con bajos requerimientos computacionales que incluyera: Filtrado de control de calidad de secuencias, alineación de secuencias a un genoma de referencia, construcción de árbol filogenético, anotación de mutaciones SNPs, cambios en marco de lectura y delecciones de aminoácidos, asignación de

linaje mediante la herramienta pangolin y scorio, estructuración e integración de datos y visualización inteligente a través de Microreact.org.

Resultados

Un proceso descargable desde un repositorio público con una conexión de internet regular, requiere la instalación de Docker (recomendado) o Conda, puede usarse ajustando dos líneas en el comando de corrida, usa menos de 1GB de memoria física en los 11 sub-procesos administrados por Nextflow con dependencias Conda, requiere menos de 2.5 GB RAM/300 genomas y 20 minutos, en una maquina virtual Ubuntu16.04 de 4 GB RAM. El crecimiento de los requerimientos computacionales frente al número de genomas tiene una tendencia lineal, pero puede ser administrado por proceso y ajustado a capacidades específicas.

Discusión

Con esta herramienta resultado del proyecto GHRU-AMR financiado por el “National Institute for Health Research” del Reino Unido, se espera colaborar con científicos, investigadores y personal de la salud en países con recursos reducidos o en fases tempranas de implementación de mWGS, a comunicar los riesgos de una manera eficiente usando los datos genómicos emergentes. En Colombia e India, este proceso ha sido incorporado como alternativa para complementar los informes locales o particulares dentro del consorcio genómico nacional SARS-CoV2 y apoyar los informes en otros países y la región <https://www.ins.gov.co/Noticias/Paginas/coronavirus-genoma.aspx>. El código e instrucciones del proceso pueden encontrarse en: <https://gitlab.com/johan.bernal.morales/sarscov2.git>