



Licenciatura en **CIENCIAS GENÓMICAS**

Formato para proponer cursos Semestre 2024-1

El curso ya ha sido impartido: Sí x No

1. Indicar modalidad: (Optativo, Seminario, curso regular (con profesor invitado)
Curso regular

2. Título: Se sugiere que sea conciso y refleje el contenido general
Ensamble, anotación y análisis comparativo de genomas microbianos

3. Tutor responsable: Dr. Luis Lozano y Dr. Pablo Vinuesa

Nombre completo	Luis Lozano
-----------------	--------------------

Entidad académica	CCG - UNAM
-------------------	------------

Nombre completo	Pablo Vinuesa Fleischmann
-----------------	----------------------------------

Entidad académica	Centro de Ciencias Genómicas, UNAM
-------------------	------------------------------------

6. Descripción del curso

Este curso está dividido en dos bloques, cada uno de 11 / 8 temas o lecciones, respectivamente. El primero estará a cargo del Dr. Luis Lozano Aguirre (CCG-UNAM) y cubrirá las bases teóricas y el manejo práctico de las herramientas computacionales para el ensamblado de genomas *de novo* y basados en referencia. El segundo bloque estará a cargo del Dr. Pablo Vinuesa (CCG-UNAM) y se enfocará al análisis comparativo de genomas y filogenómica. El curso tiene una estructura de taller, con lecciones teóricas seguidas siempre de sesiones prácticas en un entorno de Linux.

Objetivos

El alumno conocerá los aspectos fundamentales del análisis de datos genómicos desde la obtención de archivos de secuenciación, análisis de calidad, herramientas para el ensamblado y anotación funcional, y aplicará correctamente las bases prácticas para resolver las preguntas que se planteen con estas herramientas. Una vez revisados los principios del ensamblado y anotación de genomas, el curso progresa hacia el análisis comparativo de los mismos, con un enfoque en pangenómica y filogenómica. Se enseñarán los principios teóricos que subyacen a estas disciplinas, incluyendo métodos filogenéticos, así como el manejo práctico de una amplia gama de programas desde la línea de comandos de Linux.

Evaluación

Ambos bloques tendrán un trabajo de grupo y una serie de ejercicios puntuales para cada tema. La calificación final será el promedio de todos los ejercicios de ambos bloques.

Temario:

Lección 1.

Bases de datos genómicos en el NCBI, EMBL y DDBJ.
Formatos de archivos de secuenciación. Calidad Phred.

Lección 2.

Visualización de calidad de lecturas (FastQC).
Control de calidad y filtrado de lecturas (Trim Galore, TRIMMOMATIC)

Lección 3.

Aspectos teóricos y prácticos de los ensambladores de lecturas cortas (single-end, paired-end). Parte 1.
Velvet, Spades.

Lección 4.

Aspectos teóricos y prácticos de los ensambladores de lecturas cortas (single-end, paired-end). Parte 2.
Velvet, Spades.

Lección 5.

Aspectos teóricos y prácticos de los ensambladores de lecturas largas (PacBio, Nanopore). Parte 1.
Canu.

Lección 6.

Aspectos teóricos y prácticos de los ensambladores de lecturas largas (PacBio, Nanopore). Parte 2.
Canu.

Lección 7.

Aspectos teóricos y prácticos de los ensambladores híbridos.
Spades.

Lección 8.

Refinamiento de Ensamblados.
CheckM, Gapfiller, Quast.

Lección 9.

Alineamiento de lecturas de secuenciación a ensamblados de genomas (Samtools, Bowtie2).

Lección 10.

Predicción de genes y anotación funcional (Glimmer3, PROKKA). Parte 1.
Visualización de lecturas alineadas y genes anotados (ARTEMIS).

Lección 11.

Predicción de genes y anotación funcional (Glimmer3, PROKKA). Parte 2.
Visualización de lecturas alineadas y genes anotados (ARTEMIS).

Lección 12.	Conceptos básicos de genómica comparativa y pan-genómica (1)
Lección 13.	Principios de genómica comparativa: búsquedas de homólogos mediante BLAST (2)
Lección 14.	Alineamientos múltiples de secuencias homólogas (1)
Lección 15.	Introducción a los métodos filogenéticos, árboles de genes y de árboles de especies (2)
Lección 16.	Modelos y matrices de sustitución y máxima verosimilitud (1)
Lección 17.	Selección de modelos y matrices e inferencia de filogenias de máxima verosimilitud (2)
Lección 18.	Análisis pangenómico y filogenómico usando GET_HOMOLOGUES y GET_PHYLOMARKERS (2)
Lección 19.	Pangenómica y evolución de rasgos complejos en patógenos oportunistas multidrogo-resistentes (seminario integrativo de investigación; 1 sesión)

7. Características para la impartición del curso :	
Lugar donde se realizará	LCG (salón #2)
Duración en horas por sesión y número de sesiones	Son un total de 16 semanas [30 sesiones/clases (excluyendo festivos) de 2 hrs max; con un receso en medio de ~10 min]
Disponibilidad de impartirlo por videoconferencia	Sí <u>X</u> No _____
8. Método de evaluación:	
Por favor incluya en este apartado el % de la contribución relativa de:	
Participación en clase	
Presentación en clase	
Proyecto de investigación	50 % de cada módulo tiene un proyecto de investigación
Trabajos	50 % trabajos que deberán resolver para cada bloque del curso, de manera individual
Otros	

10. Bibliografía

Referencias:

- Koonin EV. Orthologs, paralogs, and evolutionary genomics. [Annu Rev Genet. 2005;39:309-38](#)
- Vernikos G, Medini D, Riley DR, Tettelin H. Ten years of pan-genome analyses. [Curr Opin Microbiol. 2015 Feb;23:148-54](#)
- McInerney JO, McNally A, O'Connell MJ. Why prokaryotes have pangenomes. [Nat Microbiol. 2017 Mar 28;2:17040](#)
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. [BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421](#)
- Kristensen DM, Wolf YI, Mushegian AR, Koonin EV. [Computational methods for Gene Orthology inference. Brief Bioinform. 2011 Sep;12\(5\):379-91. doi: 10.1093/bib/bbr030](#)
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. [Mol Syst Biol. 2011 Oct 11;7:539](#)
- Yang Z, Rannala B. Molecular phylogenetics: principles and practice. [Nat Rev Genet. 2012 Mar 28;13\(5\):303-14](#)
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. [Syst Biol. 2010 May;59\(3\):307-21](#)
- Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. [Nat Methods. 2012 Jul 30;9\(8\):772. doi: 10.1038/nmeth.2109.](#)
- Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. [Appl Environ Microbiol. 2013 Dec;79\(24\):7696-701](#)
- Golicz AA, Bayer PE, Bhalla PL, Batley J, Edwards D. Pangenomics Comes of Age: From Bacteria to Plant and Animal Applications. [Trends Genet. 2020 Feb;36\(2\):132-145. doi: 10.1016/j.tig.2019.11.006](#)
- Vinuesa P, Ochoa-Sánchez LE, Contreras-Moreira B. GET_PHYLOMARKERS, a Software Package to Select Optimal Orthologous Clusters for Phylogenomics and Inferring Pan-Genome Phylogenies, Used for a Critical Geno-Taxonomic Revision of the Genus *Stenotrophomonas*. [Front Microbiol. 2018 May 1;9:771](#)