Universidad Nacional de Río Cuarto

Facultad de Ciencias Exactas, Físico-Químicas y Naturales

Departamento de Ciencias Naturales

Carrera: Microbiología

Bioinformática Aplicada (2195)

Año académico: 2018

Régimen: Bimestral (Cuarto Bimestre)

Horas totales: 64

Modalidad: Teóricos-Prácticos

Responsable: Dr. Edgardo Jofré

Colaboradores: Dra. Sonia Fischer, Dra. Daniela Medeot, Dra. Analía Príncipe, Dr. Fernando Ibañez, Dra. María Flores-Cáceres, Mic. Juan Liaudat.

PROGRAMA:

UNIDAD 1: Bioinformática: Consideraciones generales. Definiciones. Principales desafíos de la Bioinformática. **Principales bases de datos de bioinformática**: NCBI, EMBL, DDBJ, PDB, SWISS-PROT. Base de datos GenBank: descripción y alcances. Bases de datos derivadas de NCBI: RefSeqs. Alineamientos globales y locales. Algoritmos heurísticos FASTA y BLAST. Búsqueda de similitudes en bases de datos con BLAST del NCBI. Comparación de secuencias en diferentes bases de datos (blastn**,** blastp, blastx, tblastn, tblastx). Selección de algoritmo (megablast, megablast discontinuo, blastN, blastp, PSI-Blast, PHI-Blast).
Parámetros del algoritmo: valor E, matrices de puntuación (BLOSUM, PAM), penalización de gaps, filtros y máscaras. Interpretación de salida de Blast. Herramientas del NCBI: Búsqueda de marcos de lectura abiertos (ORFinder), sistema para encontrar contaminación con vectores en secuencias de ácidos nucleícos (VecScreen).

Unidad 2: Búsqueda de similitud entre dos o más secuencias nucleotídicas o aminoacídicas. Métodos basados en matrices de puntos (Dot plot). Identificación de: regiones de similitud entre

dos secuencias, inserciones-deleciones (Eventos INDEL), regiones repetitivas, repetidos invertidos.

Unidad 3: Métodos de secuenciamiento. Métodos y Tecnologías de secuenciación de ácidos nucleicos. Aplicaciones de la secuenciación de ADN. Métodos de secuenciación de primera generación: Sanger (manual y automatizado). Métodos de secuenciación de segunda generación: Pirosecuenciamiento, Ion Torrent, Illumina. Estrategias para la preparación de los ADN molde: PCR en emulsión, amplificación en fase sólida, molde de molécula única. Métodos de secuenciación de tercera generación: Secuenciación de ADN de molécula única en tiempo real. Ventajas y desventajas de cada método de secuenciación. Análisis de secuencias crudas de ADN. Interpretación de cromatogramas. Corrección de errores de secuenciamiento. Empalme de secuencias. Búsqueda de marcos de lectura abiertos ORF.

Análisis de genomas. Bases de datos de genomas y análisis. Cronología de genomas secuenciados. Análisis de genomas procariotas. Transferencia horizontal de genes. Islas genómicas. Características de genomas eucariotas. Secuencias repetitivas, elementos transponibles, pseudogenes. Ensamblaje de las secuencias e identificación de genes. Genómica comparativa. Sintenía. Identificación de distintos tipos de rearreglos: inversión, transposición, transposición reversa, translocaciones, duplicaciones. Bases de datos COGs, KEGGs, KOGs, Pfam y TIGRfam. Búsqueda de genes en distintos genomas. Búsqueda de funciones en distintos genomas. Búsqueda de ortólogos en distintos genomas. Análisis de ejemplos en servidores internacionales, IMG (Integrated Microbial Genome) para el análisis comparativo de genomas microbianos.

Unidad 4: Bioinformática aplicada al análisis filogenético de secuencias. Las moléculas como documentos de la historia evolutiva. Árboles filogenéticos enraizados y no-enraizados. Modelos de sustitución. Métodos para la inferencia filogenética. Bootstrap como prueba de confiabilidad de las topologías inferidas.

Unidad 5: Búsqueda de dominios y motivos en secuencias nucleotídicas y aminoacídicas. Reglas de sintaxis para la descripción de motivos. Principales algoritmos. Principales bases de datos para la búsqueda de motivos o patrones: Basados en métodos probabilísticos (Pfam) y no porbabilísticos

(Prosite). Análisis de ejemplos en servidores internacionales. Búsqueda de secuencias promotoras en genes procariotas. Logo de secuencias.

Unidad 6: El análisis de secuencias y la predicción de estructuras secundarias en proteínas. Parámetros físico-químicos de una secuencia de proteínas: composición atómica y de aminoácidos, punto isoeléctrico, coeficiente de extinción, peso molecular (ProtParam) (Compute pI/Mw) (ScanSite pI/Mw). Predicción de lipoproteínas y Péptido señal en bacterias Gram Negativas (LipoP). Predicción de sitios de clivaje de péptido señal (SignalP). Predicción de Localización PSORT - Predicción de localización subcelular de proteínas (TargetP). Predicción de regiones de transmembrana (DAS). Predicción de regiones de transmembrana y localización en proteínas procariotas (HMMTOP) (PredictProtein) (SOSUI) (TMHMM) (TMpred). Detección de regiones de transmembrana basado en el alineamiento de secuencias múltiples (TMAP). Predicción de topologías en proteínas de membrane (TopPred). Predicción de sitios de fosforilación y glicosilación en proteínas. Predicción de regions coiled-coil en proteínas (Coils) (Paircoil) (Paircoil2) (Multicoil). Predicción de Leucine Zippers (2ZIP). Predicción de estructuras secundarias en proteínas (Jpred2). Principales criterios y algoritmos (hidropatía, anfipaticidad). Análisis de ejemplos en servidores internacionales. Introducción al modelado de proteínas.

Unidad 7: Expresión de proteínas Heterólogas. Características de la Inteínas: Mecanismo químico del *Protein Splicing*. Utilización de Inteínas para la expresión de proteínas. Desarrollo de plásmidos de expresión conteniendo secuencias que codifican para Inteínas. Purificación de proteínas recombinantes en *E. coli* por sistema de Inteínas. Detección y análisis de proteínas expresadas a partir de genes clonados. Optimización de la expresión de proteínas recombinantes. Genotipos bacterianos más utilizados. Otros sistemas comerciales de uso más frecuente His-Tag, GST, Flag; características, ventajas y desventajas de los diferentes métodos de purificación. Sistemas de expresión comúnmente usados. Producción de proteínas recombinantes en *E. coli*. Expresión en Baculovirus, *Saccharomyces cerevisiae* y Células de Mamíferos. Elección del método de expresión más eficiente. Estrategias de clonado. Métodos de purificación. Interacción Proteína-Proteína: Ensayos de dos híbridos Construcción de vectores y detección de interacción de proteínas. Cross-Linking. Detección de interacción proteína-proteína *in vitro*.

UNIDAD 8: Análisis global de la expresión génica: Transcriptómica, Proteómica. Tecnologías genómicas. Caracterización del transcriptoma y el proteoma celular. Estudios de expresión diferencial. Estudios masivos de expresión génica. Arreglos génicos (*macro/microarrays, chip-technology*). **Tecnología de expresión "in vivo" (IVET):** Fundamentos y componentes del sistema. Estrategias de selección IVET: basadas en auxotrofía, en resistencia a antibióticos, utilizando un único gen como reportero dual, y selección basada en recombinación

(RIVET). Ventajas y desventajas de las distintas estrategias de selección IVET. Ejemplos del uso de IVET para estudios de genes de patogenicidad y en la interacción planta-bacteria. RNAseq. Métodos para la ingeniería y edición de genomas. Nucleasas Zinc-finger (ZFNs), nucleasas efectoras tipo activadores transcripcionales (TALENs) y endonucleasas de ADN guiadas por RNA basadas en repetidos palindrómicos cortos regularmente interespaciados (CRISPR)/Cas.

TRABAJOS PRÁCTICOS: Se realizarán 6 trabajos Prácticos durante el cursado de la asignatura.

Requisitos y materias correlativas:
Para el cursado de Bioinformática Aplicada se requiere: Tener aprobadas las siguientes asignaturas: 2163 Genética Microbiana.

Evaluación: Examen parcial integrador. La asignatura contará con un régimen de promoción. Para la promoción el alumno deberá haber aprobado el parcial con nota igual o superior a 7, haber aprobado el trabajo práctico y contar con el 80% de asistencia. Para la regularidad el alumno deberá haber aprobado el parcial y el trabajo práctico y contar con el 80% de asistencia. Los alumnos regulares deberán rendir un examen final oral para aprobar la materia.

Bibliografía:

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25:** 3389-3402.

- Birren B, Green E, Hieter P, Klapholz S, Myers R, Riethman H. (1999). Genome analysis: A laboratory manual. Vol.3: Cloning Systems. Cold Spring Harbor Laboratory Press.

- Brosch, M., Yu, L., Hubbard, T., Choudhary, J. (2009). Accurate and sensitive peptide identification with Mascot. *J. Proteome Res*. **8:** 3176-3181.

- Casali N., Preston A. (2003). *Escherichia coli* plasmid vectors. Methods and Applications. Humana Press. v. 235.

- Coligan J.E, Dunn B.M, Speicher D.W, Wingfield P.T. (2010). Current Protocols in Protein Science. ISSN 1934-3663. Wiley. England.

- Fidelma Boy, E., Almagro-Moreno, S., Parent, M. (2008). Genomic islands are dynamic, ancient, integrative elements in bacterial evolution. *Trends in Microbiology* **17:** 47-53.

- Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A. **(**2010). The Pfam protein families database. *Nucleic Acids Research* 38(Database issue):D211-22.

- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Research* **32:**277-280.

- Karimova G, Pidoux J, Ullman A, Ladant D. (1998). A bacterial two-hybrid system based on reconstituted signal transduction pathweay. *PNAS*. **95:**5752-5756.

- Kaufmann, M. (2006).The Role of the COG Database in Comparative and Functional Genomics. *Current Bioinformatic*. **1:** 291-300.

- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:**376-80.

- McDonnell A.V., Jiang T., Keating A.E., Berger B. (2006) Paircoil2: Improved prediction of coiled coils from sequence. *Bioinformatics* **22:** 356-358.

- Mount, D. (2004). Bioinformatics: Sequence and Genome Analysis.. Second Edition, Ed: Cold Spring Harbor Laboratory Press, New York.

- Pevsner J. (2009). Bioinformatic and functional genome**.** Ed. Wiley-Blackwell.

- Reece R. (2004). Analysis of genes and genomes. University of Manchester, UK. Ed. Wiley.

- Selle K and Barrangou R. (2015). Harnessing CRISPR–Cas systems for bacterial genome editing. *Trends in Microbiology*. http://dx.doi.org/10.1016/j.tim.2015.01.008

- Selengut, J., Haft, D., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson W., Richter, A., White, O. (2007). TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acid Research* **35:**260-264.

- Smith, Donald and Johnson, Kevin. (1988). Single-step purification of polypeptides in *E. coli* as fusion with glutathione *S*-transferase. *Gene* **67:** 31-40.

- Steen, H., Mann, M. (2004). The ABC'S (and XYZ'S) of peptide sequencing. *Nature Reviews. Molecular Cell Biology* **5:** 699-711.
- Thiede, B., Hohenwarter, W., Krah, A., Mattow, J., Schmid, M., Schmidt, F., Jungblut, P.R. (2005). Peptide mass fingerprinting. *Methods* **35:** 23-247

- Vaillancourt, P. E. (2003). *Escherichia coli* Gene Expression Protocols. Humana Press Inc. USA.

- Westermeier, R., Naven, T. (2002). Proteomics in Practice: A laboratory manual of proteome analysis. Wiley-VCH Verlag GmbH. Germany.