

1. Título del Curso Fundamental	
Introducción a la Bioinformática	
http://www.ccg.unam.mx/~vinuesa/intro2bioinfo/	
2. Tutores responsables	
Nombre completo	Pablo Vinuesa
Adscripción	CCG- UNAM
Teléfono	(777) 317 5867
Correo electrónico	vinuesa@ccg.unam.mx
Nombre completo	Merino Pérez Enrique
Adscripción	Instituto de Biotecnología-UNAM
Teléfono	(777) 329-1634
Correo electrónico	merino@ibt.unam.mx
<i>Nombre completo</i>	Rosa María Gutiérrez Ríos,
Adscripción	IBT- UNAM
Teléfono	(777) 329-1634
Correo electrónico	rmaria@ibt.unam.mx
3 . Estudiantes Coordinadores del Curso o Tópico	
<i>Nombre completo</i>	
Adscripción	
Teléfono	
Correo electrónico	
<i>Nombre completo</i>	
Adscripción	
Teléfono	
Correo electrónico	
<i>Nombre completo</i>	
Adscripción	
Teléfono	
4. Profesores invitados	
Lic. Karel Estrada Guerra. Unidad de apoyo bioinformático. UNAM.	
5. Introducción/justificación del Curso/Objetivos	
Introducción.	
<p>Una tendencia generalizada dentro de las ciencias biológicas del siglo pasado ha sido el estudiar a los componentes celulares y sus funciones de manera independiente, dentro de un esquema que pudiera llamarse "<i>reduccionista</i>". Basado en dicho enfoque se ha podido generar conocimiento significativo dentro de cada una de las ramas de la Biología. Más recientemente, gracias a las tecnologías de nueva generación se ha cambiado el paradigma, que ahora implica el análisis de genomas completos. Claros ejemplos de ello son la secuenciación genómica completa de organismos, o bien en la cuantificación masiva de los transcritos del organismo o tejidos particulares del mismo bajo condiciones específicas de crecimiento. Existe un gran reto en el</p>	

procesamiento de estos datos masivos para la generación de conocimiento con el fin de progresar hacia una nueva etapa que integre las relaciones existentes entre los diferentes componentes del genoma y su estructura. Empieza a ser generalmente reconocido que las células deben de ser vistas como sistemas que representan las interacciones entre sus productos génicos, mismos que son expresados en el momento y cantidad necesaria de acuerdo a una compleja red de regulación de la expresión genética y estado del cromosoma. Esta tendencia de análisis global o masivo de secuencias, proteomas, transcritos o metabolitos se hace patente en muchas ramas de la biología.

Es por ello que se hace imprescindible el uso de herramientas computacionales que nos permitan la obtención, caracterización, análisis e interpretación de dicha información en las Ciencias Genómicas.

OBJETIVO

Lograr que el alumno conozca los fundamentos básicos de las metodologías más comunes utilizadas en el área de bioinformática y tenga la capacidad de manipular las herramientas computacionales disponibles públicamente en la red.

6. Características para la impartición del Curso

Indique el lugar en donde se realizará el Curso	Centro de Ciencias Genómicas UNAM, Auditorio Guillermo Soberón,
Número de sesiones y duración en horas por sesión (mínimo 36 horas)	12 sesiones, cada viernes de 10-13 hrs.
Disponibilidad de impartirlo por videoconferencia	NO
Número total de alumnos que puede aceptar	15
Número de alumnos EXTERNOS AL PM y DCB que puede aceptar 15	

7. Método de evaluación

Por favor incluya en este apartado el % de la contribución relativa de:	
Exámenes (número)	
Participación en clase	15%
Asistencia	0
Presentación de un proyecto	
Otros: Tareas de cada bloque o sesión: 85% Resolución de problemas que utilicen en cada tema las herramientas bioinformáticas presentadas en el curso	

8. Temario del Curso o Tópico

<http://www.ccg.unam.mx/~vinuesa/intro2bioinfo/>

Alineamientos y Algoritmos

1.- Introducción a las bases de datos

Tema 1.1. Secuencias de nucleótidos: EMBL y Gene Bank

Tema 1.2. Secuencias de aminoácidos: SWISSPROT

Tema 1.3. Estructuras de proteínas y su clasificación: PDB, CATH, SCOP

2. Alineamientos y Algoritmos (Enrique Merino Pérez IBT-UNAM)

Tema 2.1. Las matrices de calificación: BLOSUM, PAM, y matrices basadas en clasificaciones de aminoácidos.

Tema 2.2 Alineamiento en pares: Smith-waterman, Needleman-Wunsch.

Tema 2.3. Programación dinámica y criterios de penalización.

Tema 2.4. Búsqueda por similitud usando FASTA y BLAST.

Tema 2.5. Alineamiento multiples: CLUSTAL, Tcoffee, muscle

3.- Búsqueda de Motivos de proteínas y motivos de regulación

Tema 3.1. Motivos: descripción y construcción: Matrices de Peso

Tema 3.2. Búsqueda de motivos mediante de Modelos de Markov Escondidos

Tema 3.3. Bases de datos y Herramientas: Programas de predicción: Motivos, Prosite, MEME, Wconsensus,

DyadDetector, PFAM, BLOCKS, PROSITE,

Tema 3.4. Usos de los motivos. Predicción de función y búsqueda de homólogos remotos

4.- Filogenética Molecular (Pablo Vinuesa CCG-UNAM)

Tema 4.1. Métodos de Reconstrucción de Filogenia: Métodos basados en distancia, Métodos basados en parsimonia, Máxima Verosimilitud

Tema 4.2. Métodos basados en parsimonia

Tema 4.3. Herramientas BioEdit, DAMBE, MEGA5, jModeltest, ModelGenerator, PhyML 3, MrBayes

5. Análisis de transcripción diferencial de genes

Tema 5.1. Microarreglos, Bases de datos de microarreglos (MIAMI).

Tema 5.2. Plataformas de tercera generación.

Tema 5.3. Diferentes tipos de formatos de secuencia: ssf. Fastq, gff

Tema 5.4. Limpieza de datos y análisis de calidad

Tema 5.5. Filtrado, Ensamblado, ensamblado de novo

Tema 5.6. Herramientas de análisis: Maq, Bowtie, SOAP2, Cuffdiff, Limma Voom

Tema 5.7. Normalización y agrupamiento y análisis de la expresión diferencial.

6. Estructura de Proteínas y Función

Tema 6.1. Identificación de dominios en proteínas: Superfamily, PFAM

Tema 6.2. Predicción de estructura secundaria: nnPredict, Predator, PHD, Psipred.

Tema 6.3. Predicción de estructura tridimensional

Tema 6.4. Docking molecular

Tema 6.5. Programas para la visualización de estructuras de proteínas

9. Bibliografía

Mount, D.W. Bioinformatics: Sequence and Genome Analysis. CSHL Press.

Alberts, Bray, Lewis, Raff, Roberts, Watson: Molecular biology of the Cell. Garland Publ. Inc., 1992.

Henikoff and Henikoff: Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA. 89 : 10915-10919, Nov., 1992.

Gribskov, McLachlan and Eisenberg: Profile analysis: detection of distantly related proteins. Proc. Natl. Acad. Sci. USA 84 : 4355-4358, July 1987.

Pearson: Rapid and sensitive sequence comparison with FASTP and FASTA. Meth. in Enz. 183 : 63-98, 1990.

Altschul, Gish, Miller, Myers and Lipman: Basic local alignment search tool. J. Mol. Biol. 215 : 403-410, 1990.

Thompson, Higgins and Gibson: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acid. Res. 22(22), : 4673-4680, 1994.

Neuwald and Green: Detecting Patterns in Protein Sequences. J. Mol. Biol, 239 : 689-712, 1994.

Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, Makeev VJ, Mironov AA, Noble WS, Pavesi G, Pesole G, Regnier M, Simonis N, Sinha S, Thijs G, van Helden J, Vandenberg M, Weng Z, Workman C, Ye C, Zhu Z: Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol. Jan;23(1):137-44. 2005

Rost and Sander: Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Bio. 232 : 584-599, 1993.

Kohane IS, Kho AT and Butte AJ : Microarrays for integrative genomics. MIT Press. Chapter 3-4. 2003.

Draghici S, Khatri P, Eklund AC and Szallasi Z: Reliability and reproducibility issues in DNA microarrays measurements. Trends in Genetics 22(2) 101-109, 2006.

Yang Z, Rannala B. Molecular phylogenetics: principles and practice. Nat Rev Genet. 2012 Mar 28;13(5):303-14. doi: 10.1038/nrg3186. Review. PubMed PMID: 22456349.