



Universidad de Buenos Aires  
Facultad de Ciencias Exactas y Naturales

Planilla a completar para presentación de Cursos de Posgrado

1.- DEPARTAMENTO de COMPUTACIÓN

2.- NOMBRE DEL CURSO: Algoritmos para el análisis de secuencias biológicas

3.- DOCENTES:

RESPONSABLE/S: Pablo Guillermo Turjanski

COLABORADORES: .....

AUXILIARES: .....

4.- CARRERA de DOCTORADO

5.- AÑO: 2018

CUATRIMESTRE/S: 2do. Cuatrimestre

6.- PUNTAJE PROPUESTO PARA CARRERA DE DOCTORADO: 3

7.- DURACIÓN (anual, cuatrimestral, bimestral u otra): Cuatrimestral

8.- CARGA HORARIA SEMANAL:

Teóricas: 2hs

Problemas: 0hs

Laboratorio: 2hs

Seminarios: 0hs.

Teórico – Práctico: 0hs

Salida a Campo: 0hs

9.- CARGA HORARIA TOTAL: 4hs x 16 semanas = 64hs

10.- FORMA DE EVALUACIÓN: Trabajos de laboratorio y examen final escrito

11.- PROGRAMA ANALÍTICO:

El curso cubrirá algunos de los algoritmos más utilizados en la bioinformática para el análisis de secuencias biológicas. Se hará hincapié en la aplicación de los algoritmos en términos de programas computacionales. Durante el curso serán presentados y analizados

problemas biológicos vinculados al análisis de secuencias biológicas, con el propósito de poner de relieve los puntos fuertes y débiles de los diferentes algoritmos que atacan dichos problemas. Se cubrirán los siguientes temas:

- Alineamiento utilizando programación dinámica: Needleman-Wunsch, Smith-Waterman
- Heurísticas de alineamiento múltiple
- Algoritmo BLAST y sus variantes (BLASTN, BLASTP, PSI-BLAST): Búsquedas por similitud de secuencia en bases de datos.
- Búsqueda de patrones repetitivos en secuencias: Maximal Repeats, Super Maximal Repeats
- Perfiles estadísticos para representar familias de proteínas basados en Modelos Ocultos de Markov
- Clustering de secuencias: algoritmo de Hobohm
- Algoritmos de ensamblado parte I: k-mers
- Algoritmos de ensamblado parte II: Grafos de De Bruijn
- Algoritmos de identificación de genes: Glimmer, GeneMark, Prodigal.

## 12.- BIBLIOGRAFÍA:

### Libros

- Phillip Compeau, Pavel Pevzner. *Bioinformatics Algorithms. An Active Learning Approach*. Active Learning Publishers, 2014
- David W. Moun. *Bioinformatics: Sequence and Genome Analysis*. 2nd Edition. Cold Spring Harbor, 2004.
- Dan Gusfield. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press, 1997.

### 6 Papers

- Krogh A, Brown M, Mian IS, Sjölander K, Haussler D. *Hidden Markov models in computational biology. Applications to protein modeling*. **J Mol Biol**. 235(5):1501- 1531. (1994)
- Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, David J. Lipman, *Basic local alignment search tool*, **Journal of Molecular Biology**, J Mol Biol 215(3): 403-410. (1990)
- Doug Hyatt, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer and Loren J Hauser. *Prodigal: prokaryotic gene recognition and translation initiation site identification* **BMC Bioinformatics** 11:119, (2010)
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J.. *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. **Nucleic Acids Research**, 25(17), 3389–3402. (1997)
- Hobohm, U., Scharf, M., Schneider, R., Sander, C. *Selection of representative protein data sets*. **Protein Science : A Publication of the Protein Society**, 1(3), 409–417. (1992)
- Zhu, W., Lomsadze, A., Borodovsky, M. *Ab initio gene identification in metagenomic sequences*. **Nucleic Acids Research**, 38(12), e132. (2010)

- Delcher, A. L., Harmon, D., Kasif, S., White, O., Salzberg, S. L. *Improved microbial gene identification with GLIMMER*. **Nucleic Acids Research**, 27(23), 4636–4641. (1999)
- Delcher, A. L., Bratke, K. A., Powers, E. C., Salzberg, S. L. *Identifying bacterial genes and endosymbiont DNA with Glimmer*. **Bioinformatics** (Oxford, England), 23(6), 673–679. (2007)
- Thompson, J. D., Higgins, D. G., Gibson, T. J. *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. **Nucleic Acids Research**, 22(22), 4673–4680. (1994)
- Eddy, S. R. *Accelerated Profile HMM Searches*. **PLoS Computational Biology**, 7(10), e1002195. (2011)
- Zhang, W., Chen, J., Yang, Y., Tang, Y., Shang, J., Shen, B. *A Practical Comparison of De Novo Genome Assembly Software Tools for Next-Generation Sequencing Technologies*. **PLoS ONE**, 6(3), e17915. (2011)