

## **Bases de Datos en Bioinformática**

**CICLO I- 2020**

**SIGLA PF-5023**

**CRÉDITOS: 2**

**HORARIO:** 2 horas por semana (miércoles 17:00-19:00)

**PÁGINA WEB:** [https://marceloarayasalas.weebly.com/datos\\_bioinformatica\\_r.html](https://marceloarayasalas.weebly.com/datos_bioinformatica_r.html)

**PROFESOR:**

- Marcelo Araya Salas; correo: [marcelo.araya@ucr.ac.cr](mailto:marcelo.araya@ucr.ac.cr)
  - Página web: [marceloarayasalas.weebly.com](https://marceloarayasalas.weebly.com)
- Oficina: Biología 4A

### **DESCRIPCIÓN DEL CURSO**

La investigación en ciencias biomédicas suele generar una gran cantidad de datos cuantitativos que deben ser analizados de forma meticulosa para obtener resultados confiables. Para dicho fin, se han desarrollado una gran variedad de programas estadísticos que permiten procesar extensas bases de datos de distintos tipos a gran velocidad. Sin embargo, muchas veces estos programas o paquetes estadísticos tienen la limitante de que son difíciles de adquirir por su alto costo, deben ser actualizados periódicamente, no proveen herramientas para la manipulación de datos, o están restringidos por un número definido de pruebas estadísticas o funciones matemáticas. Además, debido a que existe un gran número de paquetes estadísticos, a veces es necesario aprender a manipular diferentes programas para realizar análisis específicos.

En la última década el lenguaje de programación R ha surgido como una alternativa de acceso libre que le permite al usuario programar sus propias funciones, realizar pruebas estadísticas, graficar, manipular bases de datos extensas y de naturaleza distinta, así como compartir nuevas herramientas desarrolladas independientemente por los usuarios (Crawley 2007, Fuchs & Barrantes 2015). Esto ha hecho que R se haya convertido en una de las plataformas más populares y utilizadas en una amplia variedad de disciplinas (Crawley 2007, R Development Core Team 2014). Este curso pretende profundizar en los elementos de programación computacional, manipulación de bases de datos, y graficación personalizada en R como plataforma (Touchon & McCoy 2016, Warton et al. 2016). Las clases serán impartidas de forma que los conceptos teóricos de programación sean demostrados y aplicados en prácticas durante las mismas sesiones. Además, durante el curso cada estudiante deberá presentar un paquete o extensión (conjunto de herramientas aplicables a análisis específicos) de R, en donde profundizará sobre sus aplicaciones en el campo biomédico demostrando en clase en que consiste el análisis.

### **OBJETIVO GENERAL**

- Enseñar a estudiantes elementos básicos y avanzados de programación que les permitan manipular, explorar y graficar datos bioinformáticos utilizando R como plataforma de trabajo.

### **OBJETIVOS ESPECÍFICOS**

- Familiarizar al estudiante con la programación en R

### **Bases de Datos en Bioinformática**

- Brindar herramientas para la manipulación de bases de datos bioinformáticos usando la plataforma R
- Emplear métodos de visualización de datos usando la plataforma R
- Proveer a los estudiantes con experiencia en la aplicación de las herramientas brindadas por medio de prácticas y proyectos de investigación

#### **CRONOGRAMA**

<b>SEMANA</b>	<b>Temas</b>
1	Introducción a R
2	Clases básicas de objetos y funciones en R
3	Construcción de bucles "for"
4	Construcción de funciones 1
5	Construcción de funciones 2
6	Gráficos con R básico
7	Formatos de datos en bioinformática
8	Acceso programático a bases de datos biológicas y herramientas de análisis
9	Manipulación de bases de datos 1
10	Manipulación de bases de datos 2
11	Construcción de funciones 2
12	Diseño de bases de datos biológicas
13	Funciones Xapply
14	Informes dinámicos con Rmarkdown
15	Prueba Final
16	Presentación de paquetes

#### **BIBLIOGRAFÍA**

- Borcard D, Gillet F, Legendre P (2011) Numerical Ecology with R. Springer, London
- Crawley MJ (2007) The R Book, 2nd Editio. Wiley, Southern Gate
- Fuchs EJ, Barrantes G (2015) El lenguaje estadístico R aplicado a las ciencias biológicas. Editorial de la Universidad de Costa Rica, San José, Costa Rica
- Gentleman, R. (2008). R programming for bioinformatics. Chapman and Hall.
- Gentleman, R., Carey, V., Huber, W., Irizarry, R., & Dudoit, S. (Eds.). (2006). Bioinformatics and computational biology solutions using R and Bioconductor. Springer Science & Business Media.
- James G, Witten D, Hastie T, Tibshirani R (2013) An Introduction to Statistical Learning with Applications in R. Springer
- Maindonald J, Braun WJ (210AD) Data Analysis and Graphics Using R - an Example-Based Approach (J Maindonald and W. Braum, Eds.), Third edit. Cambridge University Press, London
- R Development Core Team (2014) R: a language and environment for statistical computing.