

Introducción al Análisis de Datos Genéticos con R

X

Carrera: Licenciatura en Ciencias Biológicas	Código de la carrera: 05
Carrera: Doctorado en Ciencias Biológicas	Código de la carrera: 55
	Código de la materia:

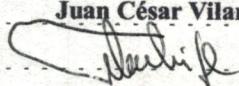
CARÁCTER:	[SI / NO]	PUNTAJE:
Curso obligatorio de licenciatura (plan 19)	NO	--
Curso optativo de licenciatura (plan 1984)	NO	--
Curso de postgrado	SI	3

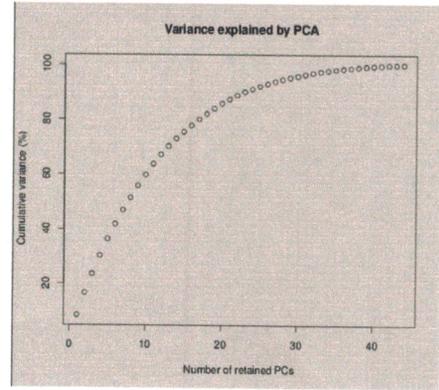
Duración de la materia:	2 Semanas	Cuatrimestre en que dicta:	2do
Frecuencia en que se dicta:			

Horas de clases semanales:	Discriminado por:	Hs.
	Teóricas	20
	Problemas	10
	Laboratorios	0
	Seminarios	0
Carga horaria semanal:		30

Carga horaria total del curso:	
Salidas de Campo (cn días)	

Asignaturas correlativas:	---
Curso PG. Dirigido a:	Lic. En Cs. Biológicas, Ing. Agrónomos y carreras afines.
Forma de Evaluación:	Examen Final y Escrito

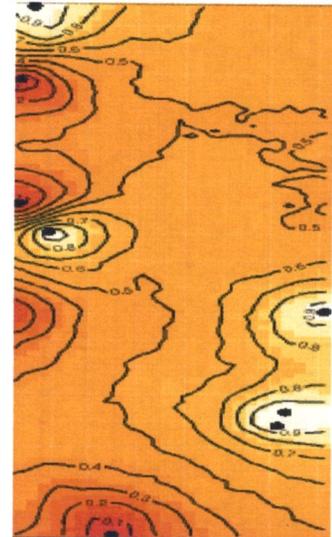
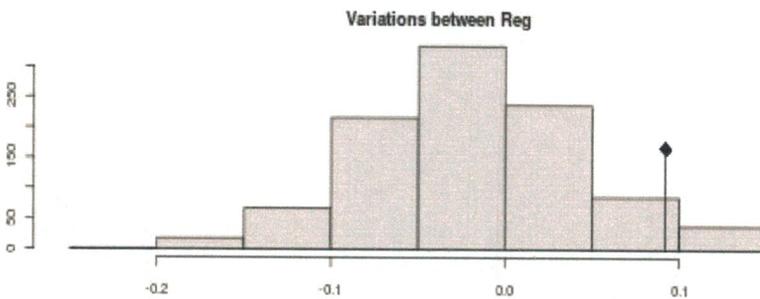
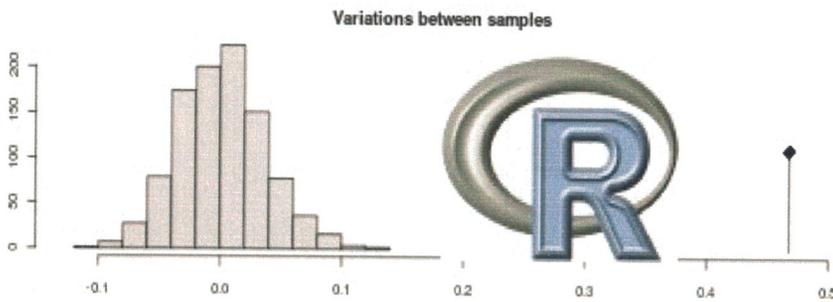
Profesor/a a cargo:	Juan César Vilardi
Firma:	
Aclaración:	Juan César Vilardi Fecha: 30 /03/2011



Introducción al Análisis de



Datos Genéticos con



DR. ESTEBAN R. HASSON
 DIRECTOR
 DEPARTAMENTO DE ECOLOGÍA, GENÉTICA Y EVOLUCIÓN
 IANIGLA

Introducción al Análisis de Datos Genéticos con R

Objetivos del curso:

R es un sistema para análisis estadísticos y gráficos creado por Ross Ihaka y Robert Gentleman (1996) cuyo uso ha ido creciendo fuertemente en el ámbito académico, ampliando de manera exponencial sus aplicaciones a diferentes campos de la ciencia. R tiene una naturaleza doble de programa y lenguaje de programación, por lo que ofrece una amplia gama de posibilidades dependiendo de los intereses particulares y el grado de entrenamiento en el campo de la programación.

El objetivo del curso es proporcionar un punto de partida para investigadores del campo de la genética en el uso de este programa en el manejo y análisis de datos genéticos.

Entre las principales razones para comenzar a utilizar R se podría mencionar:

1. Su incomparable amplitud en relación con los campos de aplicación y la disponibilidad de avanzadas herramientas estadísticas tales como los modelos mixtos y aditivos generalizados.
2. Ante el notorio incremento de contribuciones científicas que presentan sus resultados en el contexto de R, un mayor conocimiento de este programa es importante para una mejor comprensión de la literatura.
3. La contribución de un gran número de investigadores de diferentes áreas proveen continuamente nuevos paquetes que amplían día a día las aplicaciones disponibles.
4. R se distribuye gratuitamente bajo los términos de la GNU General Public Licence, disponible en varias formas, pudiéndose instalar en diferentes sistemas operativos (Linux, Windows, MacIntosh). Esto asegura su permanente actualización y perfeccionamiento.

En lo que se refiere específicamente al campo de la genética, el uso del entorno R tiene la ventaja de facilitar el uso de las tablas de datos para diferentes tipos de análisis, manteniendo una estructura básica generalizada, gracias a la existencia de comandos específicos para importación, exportación y conversión entre formatos. Asimismo existen paquetes para diferentes tipos de análisis, tales como estudios de estructura y variabilidad genética, evolución y filogenia, estudios de asociación entre síndromes clínicos y marcadores genéticos, mapeo genético, genética cuantitativa, etc.


DR. ESTEBAN R. HASSON
DIRECTOR
DEPTO ECOLOGIA, GENETICA Y EVOLUCION
FGEN-UBA

El presente curso propone brindar las herramientas elementales que permitan a los alumnos conocer las principales virtudes de R, aprender a manejar datos genéticos dentro de este ambiente y realizar algunos análisis de uso frecuente en estudios poblacionales. Dada la infinidad de aplicaciones disponibles, más que una visión exhaustiva, se pretende que el curso sea un punto de partida para que los alumnos puedan explorar las posibilidades que ofrece R para estudios más específicos vinculados a las respectivas líneas de trabajo.

Duración del Curso

Clases teóricas: 40 hs

Clases prácticas: 20 hs

Evaluación

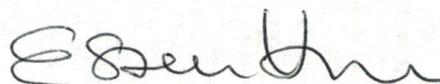
Examen consistente en la preparación individual y análisis de datos experimentales aportados por el alumno

Cupo

15 alumnos.

Requisitos

Graduado universitario con conocimientos e interés en análisis de datos genético/poblacionales.



DR. ESTEBAN R. HASSON
DIRECTOR
DEPTO ECOLOGIA GENETICA Y EVOLUCION
FCEN-UBA

Programa del curso

Introducción.

Qué es R? Por qué usar R? Ventajas de R sobre otros entornos

Tipo de marcadores y Sistemas Genéticos. Marcadores Codominantes y Dominantes (Presencia/Ausencia). Sistemas Haploides, Diploides, Sexuales, Endogámicos, Exogámicos

Objetivos y Métodos de Análisis

Objetivos de Estudios Genéticos Poblacionales. Identificación de grupos Genéticos. Correlación entre Genotipo y Fenotipo. Estimación de Distancias Genéticas. Inferencia de Rutas de Migración

Conceptos básicos del uso de R

Instalación y uso de R en distintos sistemas operativos. La página de inicio. Interfases gráficas disponibles. Espacio de trabajo. Directorio de trabajo. Operadores elementales. Funciones o comandos básicos. Guardado de datos. Guardado de historial de comandos. Instalación y Carga de Paquetes. Paquetes para análisis de datos genéticos. Cierre de sesión

Los objetos en R

Variables. Vectores. Matrices. Tablas de datos ("data frames"). Funciones. Organización y manejo de los datos. Preparación de los datos en hojas de cálculo. Exportación de datos: formatos compatibles. Importación de datos desde R. Recodificación de datos = conversión de objetos. Objetos con datos genéticos. Tablas de datos. Objetos genind. Objetos genpop. Objetos hierfstat. Tipos de marcadores genéticos. Microsatélites. SNP. Secuencias de ADN. Secuencias proteicas.

Análisis elementales de datos genético-poblacionales

Datos Poblacionales Básicos. Número de Alelos por población. Número de Alelos por locus. Frecuencias alélicas. Heterocigosidad observada y esperada. Graficación de resultados. Funciones plot() y text(). Mejora del aspecto de los gráficos. Comparación entre H observada y H esperada. Prueba de ajuste a Hardy Weinberg. Ji cuadrado tradicional y simulación de Monte Carlo. Ajuste para comparaciones múltiples.



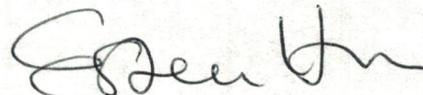
DR. ESTEBAN R. HASSON
DIRECTOR
DEPTO. ECOLOGIA GENETICA Y EVOLUCION
FCEMUN

Análisis de Estructura Poblacional y Diferenciación Genética

Estadísticos F . Significación de los estadísticos F por permutaciones. Análisis Jerarquizado de Estructura Genética. Estimación de Componentes de la Varianza. Intervalos de Confianza para los Componentes de Varianza y Estadísticos F . Estimación de Distancias. Datos genotípicos. Datos de secuencias. Análisis de agrupamiento. Test para Hipótesis de Aislamiento por Distancia. Manejo de datos Presencia/Ausencia. Análisis de Varianza Molecular (AMOVA). Preparación de los datos. Estimación de componentes de varianza y estadísticos Phi (Φ). Significación de los resultados por test de permutaciones.



Prof. Dr. JUAN C. VILARDI
EGE-FCEN-UBA
Investigador CONICET



DR. ESTEBAN R. HASSON
DIRECTOR
DEPTO. ECOLOGIA, GENÉTICA Y EVOLUCION
FCEN-UBA

Bibliografía recomendada

- Benjamini, Y.; Bretz, F.; Sarkar, S. K. 2004. Recent Developments in Multiple Comparison Procedures. IMS lecture notes-monograph series, 47. Beachwood, Ohio : Institute of Mathematical Statistics,.
- Chessel D., Dufour A.B., Thioulouse J. 2004. The ade4 package - I: One-table methods. *R News*, 4:5-10.
- Crawley , M. J. 2007. The R Book. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England.
- Dalgaard, P. 2002. Introductory Statistics with R. Springer-Verlag. New York.
- Excoffier L.; Heckel G . 2006. Computer programs for population genetics data analysis: a survival guide . *Nature Reviews Genetics* 7: 745-758
- Foulkes, A. S. 2009. Applied Statistical Genetics with R For Population-based Association Studies. Springer Dordrecht Heidelberg London New York. ISBN 978-0-387-89553-6
- Goudet J. 2005. Hierfstat, a package for R to compute and test variance components and F-statistics. *Molecular Ecology Notes*. 5:184-186
- Hartl, D. L.; Clark, A. G. 2007. Principles of Population Genetics (4^a Ed.). Sinauer Associates, Inc. Sunderland, Massachusetts.
- Ihaka R.; Gentleman R. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5: 299-314
- Jombart T (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405.
- Maindonald, J. H. 2004 . Using R for Data Analysis and Graphics . Introduction, Code and Commentary. <http://cran.r-project.org/>.
- McCullagh P. and Nelder, J. A. 1989. Generalized Linear Models. London: Chapman and Hall.
- Paradis , E. 2002. R para Principiantes . <http://cran.r-project.org/>.

Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language . Bioinformatics 20: 289-290.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.



DR. ESTEBAN R. HASSON
DIRECTOR
DEPTO ECOLOGIA GENÉTICA Y EVOLUCION
FCEN-UBA

Curso o Seminario de Postgrado y/o Doctorado
Ecología, Genética y Evolución

CARRERA LICENCIATURA EN CIENCIAS BIOLÓGICAS

Nombre del Curso Introducción al Análisis de Datos Genéticos con R

Responsable Juan César Vilardi

En caso de que el responsable del Curso no sea Docente de esta Facultad deberá adjuntarse su CV y nota solicitando la autorización

Docentes que colaboran en el dictado del curso

Adjuntar LISTADO con nombre, apellido y cargo docente. Si no es docente de esta Facultad deberá adjuntarse CV.

Curso es dirigido a Lic. En Cs. Biológicas, Ing. Agrónomos y carreras afines.

Cantidad de días que dura el curso 10 días

Fecha de inicio 05/09/11 **Fecha de finalización** 16/09/11

En ambos casos consignar día y mes aún cuando sea tentativo

Modalidad horaria Lu-Vi 9-12 y 14-17

Informar días y horario aún cuando sea tentativo. Indicar además si el día **sábado** se dicta el curso

Cant. horas totales 60 **Cant. horas semanales** 30

Hs. semanales de teóricas	40 hs.
Hs. semanales de problemas	20 hs
Hs. semanales de laboratorio	00 hs.
Hs. semanales de seminario	00 hs.
Salidas de campo	00 días

En salidas de campo indicar cantidad de días.

Nº mín. de alumnos 10 **Nº max. De alumnos** 15

En caso de nº máximo indicar prioridades de ingreso o método de selección.

Forma de evaluación Monografía

Puntaje para doctorado 3 **Puntos**

Justificar si se difiere de las pautas aconsejadas por la Comisión de Investigación, Publicaciones y Postgrado.

Arancel (Justificar) 200 **Módulos**

En caso de aceptar excepciones al arancel total indicarlos con claridad.

Modalidad de pago El que establece la Facultad

Aprobación programa Resolución CD Nº nuevo

Si aún no fue aprobado poner "nuevo". En todos los casos adjuntar programa !!

Comisión que evaluó el curso Subcom. Doctorado

Vº Bº del Departamento


DR. ESTEBAN R. HASSON
DIRECTOR
DEPTO. ECOLOGIA, GENETICA Y EVOLUCION
FGEN-UBA



Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Referencia Expte. N° 499.496/2011

Buenos Aires,

23 MAY 2011

VISTO

la nota del Dr. Esteban Hasson Director del Departamento de Ecología, Genética y Evolución, mediante la cual eleva la Información y el Programa del Curso de Posgrado **INTRODUCCIÓN AL ANÁLISIS DE DATOS GENÉTICOS CON R**, que será dictado en el Segundo cuatrimestre de 2011 (05 al 16 de setiembre de 2011), por el Dr. Juan César Vilardi

CONSIDERANDO:

lo actuado por la Comisión de Doctorado el 03/05/2011,
lo actuado por la Comisión de Enseñanza, Programas, Planes de Estudio y Posgrado,
lo actuado por la Comisión de Presupuesto y Administración,
lo actuado por este cuerpo en Sesión Ordinaria realizada en el día de la fecha,
en uso de las atribuciones que le confiere el Artículo N° 113° del Estatuto Universitario,

EL CONSEJO DIRECTIVO DE LA FACULTAD DE
CIENCIAS EXACTAS Y NATURALES
RESUELVE:

Artículo 1°: Autorizar el dictado del Curso de Postgrado **INTRODUCCIÓN AL ANÁLISIS DE DATOS GENÉTICOS CON R**, de 60 hs. de duración.

Artículo 2°: Aprobar el Programa del Curso de Posgrado **INTRODUCCIÓN AL ANÁLISIS DE DATOS GENÉTICOS CON R**, obrante a fs 7 a 10 del Expediente de la Referencia.

Artículo 3°: Aprobar un puntaje máximo de tres (3) puntos para la Carrera de Doctorado.

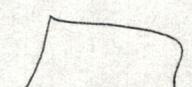
Artículo 4°: Aprobar un Arancel de 200 Módulos. Disponer que los montos recaudados serán utilizados conforme a lo dispuesto por Resolución CD N° 072/03.

Artículo 5°: Comuníquese a la Dirección del Departamento de Ecología, Genética y Evolución, a la Biblioteca de la FCEN y a la Subsecretaría de Postgrado (con fotocopia de Programa incluido: fs 7 a 10); comuníquese a la Dirección de Alumnos (sin fotocopia del Programa). Cumplido archívese.

-- 1178

Resolución CD N° _____
SP/med/03/05/2011


Dra. MARIA ISABEL GASSMANN
SECRETARIA ACADEMICA ADJUNTA


Dr. JORGE ALLAGA
DECANO