

Docente: **Giorgio Valentini**

Istruttore: **Matteo Re**

Università degli
studi di milano



Insegnamento: **Bioinformatica**

A.A. : 2011/2012 (sem II)

C.d.I. BIOTECNOLOGIE DEL FARMACO

Introduzione al linguaggio R

Giorgio Valentini

e –mail: valentini@dsi.unimi.it

<http://homes.dsi.unimi.it/~valenti>

Matteo Re

e –mail: re@dsi.unimi.it

<http://homes.dsi.unimi.it/~re>

DSI – Dipartimento di Scienze dell' Informazione
Università degli Studi di Milano

Caratteristiche di R

- Linguaggio ad alto livello *interpretato*
- Dotato di insiemi di operatori ad alto livello per *calcoli su array e matrici*
- Supporta paradigmi di programmazione *imperativa, object-oriented e funzionale*.
- Fornisce un ambiente per la *elaborazione interattiva* dei dati
- *Ambiente integrato di risorse software* per la gestione ed elaborazione di dati e la visualizzazione di grafici
- Dispone di *interfacce* verso programmi e moduli sw scritti con altri linguaggi
- Ambiente di sviluppo e package open source disponibili liberamente in internet.

R come linguaggio per la bioinformatica

- Linguaggio ad alto livello orientato alla analisi dei dati
- Permette di strutturare dati complessi ed eterogenei
- Dispone di un ambiente di lavoro e di sviluppo per lavorare interattivamente con i dati
- Dispone di package (librerie) specifiche per la bioinformatica
- R è il linguaggio utilizzato dal progetto internazionale open source *Bioconductor* per la gestione ed elaborazione di dati genomici, proteomici, farmacologici ...
- E' uno dei linguaggi maggiormente utilizzati dalla comunità internazionale dei bioinformatici

Breve storia di R

- Deriva da *S*, un linguaggio ed un sistema sviluppati da *John Chambers* e collaboratori negli anni '80 presso i Laboratori Bell.
- *S* è valso l' *ACM Software Systems Award* al suo principale progettista *J. Chamber* nel 1999.
- *R* è un progetto *Open Source* conforme per la maggior parte ad *S*:
 - Sviluppato inizialmente da *Ross Ihaka and Robert Gentleman* all'Università di Auckland (Nuova Zelanda)
 - Attualmente sviluppato da una comunità internazionale di ricercatori e sviluppatori in ambito sia accademico sia industriale
 - Opera attraverso il web: www.r-project.org
 - Archivi software e documentazione: cran.r-project.org/

Da dove “scaricare” R

CRAN - the Comprehensive R Archive Network:
<http://cran.r-project.org/> (ci sono anche mirror locali)

Sono disponibili distribuzioni binarie per :

- *Windows*
- *Macintosh*
- *Linux*

L’ installazione sul proprio PC è (in genere) semplice.

Documentazione e bibliografia su R

Materiale didattico scaricabile dalle pagine web del corso: “*Bioinformatica*”:

<http://homes.dsi.unimi.it/~re/corsobioinfo12.html>

- Un corso introduttivo scaricabile dal web:

W. Venables and D.M. Smith, *An Introduction to R*: <http://cran.r-project.org/doc/manuals/R-intro.pdf>, 2010

- Libri sulla programmazione “avanzata” in R:

1) Robert Gentleman *R Programming for Bioinformatics*, CRC/Computer Science & Data Analysis Volume 12 , Chapman & Hall, 2008

2) J. Chambers *Software for Data Analysis: Programming with R*, Springer, 2008

- Un libro specifico sull’ utilizzo di R per la bioinformatica e

Bioconductor:

Gentleman, R.; Carey, V.; Huber, W.; Irizarry, R.; Dudoit, S.

Bioinformatics and Computational Biology Solutions using R and Bioconductor, Springer, 2005