

Università degli Studi di Milano
Laurea Specialistica in Genomica Funzionale e Bioinformatica
Corso di Linguaggi di Programmazione per la Bioinformatica

Introduzione a R

Giorgio Valentini
e-mail: valentini@dsi.unimi.it

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

1

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 e proteomici
- E' uno dei linguaggi maggiormente utilizzati dalla comunità internazionale dei bioinformatici

2

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.

3

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/

4

Dove reperire R

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

Sono disponibili distribuzioni binarie per :

- *Windows* 95, 98, NT e 2000
- *Macintosh* (System 8.6 - 9.1, MacOS X)
- *Linux*

L'installazione domestica sul proprio PC è semplice.

5

Bibliografia per R

Libri e manuali introduttivi disponibili on-line:

- W. Venables and D.M. Smith, *An Introduction to R*: <http://cran.r-project.org/doc/manuals/R-intro.pdf>, 2004
- J. Maindonald, *Using R for Data Analysis and Graphics*: <http://www.maths.anu.edu.au/~johnm>, 2000
- A. Mineo, *Una guida all' utilizzo dell' ambiente Statistico R*:

Libri e manuali specifici sulla definizione del linguaggio e per lo sviluppo di Package R

- R Development Core Team, *R Language Definition*: <http://cran.r-project.org/doc/manuals/R-def.pdf>, 2004
- R Development Core Team, *Writing R extensions*: <http://cran.r-project.org/doc/manuals/R-exts.pdf>, 2004

Testo di riferimento per il linguaggio S

- R. Becker, J. Chambers and A. Wilks, *The new S language*. Chapman & Hall, New York, 1988

6