Università
degli Studi
di Padova

QUESITI PROVA SCRITTA
SELEZIONE N. 2025S28

Indicazioni svolgimento:

I candidati sono invitati a sviluppare le seguenti domande:

- usando massimo 1 facciata per domanda foglio A4
- carattere Times New Roman dimensione 12
- interlinea 1
- margine superiore 5
- margine inferiore 2,5
- entro un tempo massimo di un’ora e mezza

Elenco n. 1

1. Describe the main steps involved in creating an R package. In your answer, explain the essential components (e.g., DESCRIPTION, NAMESPACE, man, R, tests) and their role in ensuring that the package is functional, maintainable, and shareable with others.

2. Explain the concept of *software containers* (e.g., Docker, Singularity) and how they are used in High-Performance Computing (HPC) environments. Discuss their advantages for reproducibility and portability, with a focus on how they facilitate bioinformatics workflows.

3. Outline the main steps of a standard RNA-seq data analysis workflow (from raw data to biological interpretation). For each step, specify the type of software or tools typically required, and explain the role of the software in the analysis.

Indicazioni svolgimento:

I candidati sono invitati a sviluppare le seguenti domande:

- usando massimo 1 facciata per domanda foglio A4
- carattere Times New Roman dimensione 12
- interlinea 1
- margine superiore 5
- margine inferiore 2,5
- entro un tempo massimo di un'ora e mezza

Elenco n. 2

1. Explain the advantages of developing an R package instead of using a collection of standalone R scripts for data analysis.

2. Define *version control* and explain the key features of version control systems. Discuss the main challenges of managing different software versions in a research environment, and describe strategies or tools that can help address these challenges.

3. Discuss why reproducibility is critical in bioinformatics research. In your answer, highlight common factors that threaten reproducibility and describe best practices or tools that can improve reproducibility.

Indicazioni svolgimento:

I candidati sono invitati a sviluppare le seguenti domane:

- usando massimo 1 facciata per domanda foglio A4
- carattere Times New Roman dimensione 12
- interlinea 1
- margine superiore 5
- margine inferiore 2,5
- entro un tempo massimo di un'ora e mezza

Elenco n. 3

1. Describe effective strategies for analyzing large-scale datasets in R. In your answer, discuss one or more approaches and provide examples where appropriate.

2. Explain how multiple jobs are managed in a High-Performance Computing (HPC) cluster using queue management systems (e.g., Slurm, PBS). Describe the key concepts and provide an example of how a researcher might efficiently schedule and monitor jobs.

3. Describe the main challenges of aligning RNA-seq reads to a reference genome (e.g., splicing, repetitive regions, sequencing errors). Explain how commonly used alignment tools (such as STAR or HISAT2) address these challenges and what outputs they typically produce for downstream analysis.