Applications of automated bioinformatics procedures for the analysis of biological data

Luciano Milanesi

Institute of Biomedical Technologies, National Research Council, Milan, Italy

Abstract

Due to the increasing use of high throughput technologies, a large amount of information is generated. In order to correlate these information with a range of knowledge domains, it has become necessary to use hundreds of resources and applications available via web forms or graphical user interfaces provided by remote sites.

On the other hand, in large genomics studies, the data produced by high throughput techniques, that have to be analyzed by bioinformatics tools, make it necessary to increase the calculation resources. Therefore, in order to successfully face these new challenges, it will be necessary to develop dedicated supercomputers, parallel computer based on clustering technologies and high performance distributed platforms like GRID.

The next generation of Bioinformatics tools need to integrate several resources ranging from local Supercomputer or Computer Clusters in GRID and to access remote database and tools by using Web Services and Grid Infrastructures. It will be realised by implementing a distributed computing model where easy access to large geographical computing and data management resources will be provided to large multi/inter-disciplinary Virtual Organizations made of both research and user entities.

Indeed, a mix of Computational Workflows, Web Services, Semantic Web and Grids is "de facto" considered as the way to realise the concept of virtual places where scientists and researchers work together to solve complex problems in Bioinformatics, despite their geographic and organizational boundaries. The aim is to operate this widely distributed computing environment as a uniform service, which looks after resource management, exploitation and security independently of individual technology choices.

Bioinformatics studies in biology and in biomedicine need to face complex biological processes, through the analysis of nucleotide and protein sequences, macromolecular structures and interactions data.

A general overview and some examples of high-throughput biomedical and biological data analysis based on these techniques will be presented.