

Un GRID server per la bio-informatica presso ENEA

S.Migliori, R. Balducchi, A. Maggio, V.Rosato

ENEA - Centro Ricerche Casaccia

L'ENEA ha intrapreso, da qualche anno, una serie di attività di sequenziamento di geni da colture di interesse agrario. Nell'ambito delle attività previste dal Progetto PROBIO (PROduzioni vegetali mediante BIOTecnologie) del Programma Fondi Strutturali 1994-99 e degli obiettivi del neocostituito Centro di Innovazione Integrato (<http://agrobiopolis.trisaia.enea.it>) è stato realizzato, presso il Centro Ricerche Trisaia, un laboratorio di sequenziamento in grado di generare ed analizzare un elevato numero di sequenze grazie alla disponibilità di strumentazione d'avanguardia (ABI Prism 3700) e di un supporto bioinformatico disegnato ad hoc. Le attività del laboratorio sono incentrate su due linee fondamentali, una di ricerca ed una di servizio. Nell'ambito dell'attività di ricerca sono in corso programmi di sequenziamento di ESTs (Expressed Sequence Tag) di specie di interesse agrario associati ad una più complessa analisi funzionale dei patterns di espressione genica attraverso microarray technology. L'attività di servizio è in via di definizione e sarà organizzata in modo da offrire ad organismi pubblici e privati diversi livelli di analisi genomica. La parallela costituzione di una stazione di bioinformatica ad alta potenzialità e flessibilità è uno dei punti di forza del laboratorio ed è il risultato di una stretta collaborazione tra il Centro Ricerche Trisaia-ProgettoProbio ed il Centro Ricerche Casaccia- Funzione Centrale Informatica e Servizio di Calcolo e Modellistica. Al fine di mettere a punto una banca dati genica sono stati messi in opera una serie di tools standard per l'analisi delle sequenze (GCG package), installate le principali banche dati pubbliche (GeneBank, PDB, SwissProt etc.) e, inoltre, tramite l'accordo con il Biocomputing Group del Centro Interdipartimentale di Ricerche Biotecnologiche dell'Università di Bologna, è in corso di installazione presso i server di calcolo ENEA il predittore di struttura secondaria SECPRED. L'ENEA dispone di un sistema di calcolo scientifico basato sui concetti di "griglia computazionale". Il sistema è costituito da un insieme di risorse, hardware e software, dislocate nei principali centri ENEA e visibili come un'unica entità logica. L'interconnessione fisica fra le risorse è assicurata dalle reti locali connesse a livello nazionale ed internazionale tramite la rete nazionale della ricerca GARR, l'architettura logica è ottenuta tramite l'integrazione di software specialistici (AFS,LSF, Tecnologia Java, Citrix, Patron) atti a realizzare un'immagine unica del sistema dal lato dell'utente finale. L'accesso al sistema è garantito attraverso dei portali che consentono la connessione in modo semplice ed efficace sia dall'interno dell'ENEA che da qualunque postazione di lavoro connessa ad Internet. La sicurezza informatica è assicurata da adeguati meccanismi. In questo ambiente è stato inserito il sistema di sequenziamento automatico che fornisce i dati del sequenziamento all'intero sistema di calcolo e quindi pronti ad essere automaticamente elaborati sulla griglia geografica. Sono stati inoltre implementati i softwares commerciali SeqWeb (un'interfaccia HTML per il GCG package) e GCG-link, che consentono di utilizzare molte delle funzioni del GCG package da un qualsiasi PC. Tutte le funzionalità realizzate saranno a disposizione attraverso un portale web (<http://agrobiopolis.trisaia.enea.it>) in fase di implementazione tramite il quale sarà possibile effettuare le interrogazioni al DB commerciale e per gli utenti autorizzati sottomettere particolari elaborazioni. Queste verranno prese in carico da una serie di server di calcolo ENEA (SUN, IBM, SGI, Linux Intel, Linux Alpha) in funzione del loro carico e della disponibilità degli eseguibili per quella specifica piattaforma, consentendo, a regime, di minimizzare i tempi tra l'interrogazione e la risposta alle richieste di elaborazione. ENEA ha, inoltre, acquisito una serie di piattaforme di calcolo ad alte prestazioni (SGI Onyx2/3) corredate da ambienti per la grafica avanzata e la realtà virtuale immersiva ed interattiva. Tali strumenti possono essere utilizzati attraverso una serie di packages (InsightII, VRML, AVS, IDL, CaveLib,*) che consentono la visualizzazione e il modeling molecolare. Gli strumenti di grafica possono essere messi a disposizione di gruppi che ne facciano richiesta attraverso proposte di collaborazione tra i gruppi medesimi e ENEA. Alcune delle funzionalità grafiche sono fruibili anche da remoto utilizzando le tecnologie implementate.