

Alla scoperta del Centro di ricerca biotecnologica: come funziona la banca dati sulla genomica

# Ceinge, portale al servizio della scienza

www.ecostampa.it

Quando al Ceinge si dice Capri non è dell'isola che si parla, ma di un programma realizzato dal laboratorio di Bioinformatica. A dispetto della dolce suggestione estiva, Capri (Common Application Program Resource Interface) è in realtà un software che consente di lavorare con numerosi dati biologici conservati nelle banche dati del Ceinge, come le sequenze di amminoacidi o di nucleotidi. Essendo accessibile a tutti i ricercatori, il browser Capri ha un'usabilità praticamente infinita, e gli utenti registrati possono acquisire dati via web con grande facilità: li memorizzano sul server e li richiamano al momento opportuno.

di GIUSEPPE NUTINI

Lettere a disposizione degli studiosi programmi che accelerano i tempi di accesso ai dati "è il primo dovere scientifico del laboratorio di Bioinformatica - come spiega il direttore Giovanni Paoletta - Siamo un mirror locale del mondo come i principali centri di ricerca". Ogni notte il software scarica dati dalle memorie dell'Ncbi o dell'Ebi (European Bioinformatics Institute) per aggiornarsi. Conserviamo sequenze di acidi nucleici, di proteine, patologie e mutazioni: migliaia di informazioni biologiche. Dati utilizzabili mediante programmi installati in house. Capri e i programmi che permettono di operare sui dati sono utilizzati in diversi istituti di ricerca campani. Co-

si, dal 2001, Biogem, Igb (Istituto di Genetica e Biofisica del Cnr), e Federico II attingono, con un proprio username, tutti dallo stesso pozzo. Ma è l'elevato uso interno e la conservazione dei dati sperimentali che giustificano il lavoro dei bioinformatici: gli studiosi del Ceinge, infatti, consultano le collezioni di dati sistematicamente e sulla base di dati sperimentali. "Noi la chiamiamo gestione in house dei dati - aggiunge Paoletta - Siamo la memoria informatica del Ceinge. Conservare migliaia di informazioni sul Dna consente di effettuare studi come la ricerca nel campo della Genomica Comparata".

Una banca dati molto estesa è stata realizzata in collaborazione con Tigem, Biogem, Università di Milano e Ifom (Istituto Firc di Oncologia Molecolare), una banca che è arrivata a contenere 70 mila regioni genomiche annotate cui si accede grazie al programma abbinato Dg Cst (Disease Genes Conserved Sequence Tags Database), sviluppato al Ceinge. In questo tipo di ricerca, quando si identificano sequenze di Dna simili in specie diverse ma affini come uomo e topo, si ha un punto di partenza per spiegare la funzione secondo le osservazioni genetiche.

Anche se finora è un atollo nell'oceano, rispetto cioè ai milioni di sequenze del genoma umano, si è già data una motivazione scientifica di alcune delle regioni schedate: il caso di proteine o malattie comuni alle due specie. E dalle condivisioni di sequenze geniche si passa alle condivisioni di calcolo. Anche qui l'équipe dei sei

ricercatori diretti dal professor Paoletta collabora

con altri centri di ricerca, italiani o europei: come nei progetti per l'uso di Grid per la condivisione di risorse di calcolo, o come in precedenti connessioni con la University of Dundee in Scozia o con lo Embl di Heidelberg sulla Analisi delle Immagini'. Grazie ai fondi del Centro di Competenza

Gear, presto il Laboratorio potenzierà il suo hardware con l'acquisto di un mega-processore.

Il cervellone sarà un cluster, avrà una capacità di processing pari a 100 processori, ma configurati in un'unica macchina e potrà spostare il calcolo da uno all'altro molte volte al secondo. Il cluster segna un tappa significativa nell'elaborazione dei dati: potendo interfacciarsi con macchine esterne, renderà disponibili ai ricercatori elevate capacità di calcolo.

Molto vivace è l'attività di formazione del Laboratorio. Due dei sei ricercatori sono dottorandi, altri sono inseriti in altri piani formativi. Nel 2003 un Master di Bioinformatica, iniziato all'Università del Sannio, si è concluso in uno stage in loco di quattro giovani specializzandi che hanno trovato occupazione in istituti di ricerca campani. C'è spazio anche per corsi molto compatti, finalizzati all'apprendimento rapido di conoscenze informatiche. Sono gli short courses: quattro giorni full immersion di lezioni di bioinformatica per dottorandi e studenti di scuole di specializzazione. La divulgazione è presente sul sito del Laboratorio (<http://bioinfo.ceinge.unina.it>) con il programma Mini Genome Browser che descrive la posizione di un gene nel Dna, la composizione di un genoma e così via.

BIOINFORMATICA



Il gruppo di ricercatori del laboratorio di Bioinformatica

