

Come un informatico si sente a suo agio nelle biotecnologie

Professioni &
Carriere

www.ecostampa.it

Nato professionalmente come esperto di computer, dopo la laurea in Scienze biologiche e un master conseguito online presso una università inglese, Alessandro Guffanti ci racconta la sua esperienza presso l'Ifom di Milano, dove è responsabile dei servizi bioinformatici.

Nato a Milano nel 1998 per iniziativa della Fondazione Italiana per la Ricerca sul Cancro (Firc), l'Istituto Firc di Oncologia Molecolare (Ifom) è un centro di ricerca specializzato nello studio dei meccanismi di formazione e di sviluppo dei tumori. Concepito per creare una potente infrastruttura tecnologica che permette ai gruppi di ricerca di diverse istituzioni accademiche italiane di usufruire di apparecchiature e tecniche evolute, l'Ifom è, al contempo, un canale di trasferimento dell'informazione biotecnologica dal mondo accademico al mondo industriale.

Grazie al facile accesso alle risorse logistiche e alla condivisione del know how, l'Istituto rappresenta un modello altamente funzionale dello sviluppo autonomo della ricerca, attraverso finanziamenti pubblici e privati, la sinergia tra i gruppi di ricerca, l'ottimizzazione delle risorse nonché l'applicabilità dei risultati. Inaugurata nell'aprile del 2003, la nuova sede (vedi foto pag. 26) sorge su un'area ex-industriale di Milano, che occupa 11.200 mq ed è ripartita su 6 edifici, 6.200 mq di laboratori, 2.200 mq di uffici e 2.800 mq di spazi adibiti a biblioteca, auditorium, aule per seminari, mensa, foresteria, con una capacità di accoglienza per oltre 300 ricercatori. Il lavoro di ricerca svolto al suo interno è molto articolato: la maggior parte degli strumenti software utilizzati sono stati sviluppati dagli stessi scienziati Ifom mediante customizzazione di soluzioni open sour-

ce o produzione ex novo. Il profilo degli skill e delle competenze degli operatori, tra cui lavorano ingegneri, matematici e fisici, è verticale e innestato su di un robusto bagaglio cognitivo di tipo informatico. Alessandro Guffanti, analista programmatore e bioinformatico Ifom, ci fa scoprire una realtà dove l'evoluzione dei processi scientifici si coniuga con quelli tecnologici.

Da biologo a bioinformatico: quali sono stati i suoi percorsi professionali?

«Ho cominciato la mia carriera lavorativa nell'86 come computer operator in Digital Equipments su sistemi Vax/Vms, prima di mettermi a studiare Scienze biologiche: fu proprio in uno dei gruppi di interesse aziendali in rete DecNet che scoprii la bioinformatica, ovvero l'applicazione dell'IT alla gestione e all'analisi dei dati di sequenza allo scopo di fornire una risposta a determinati problemi biologici attraverso lo studio dell'espressione genica. Subito dopo la laurea in Scienze biologiche, ho fatto alcune esperienze all'estero, occupandomi della gestione e dell'analisi di database di dati genomici. Dopo un periodo di poco più di un anno di borsa di studio all'Istituto Nazionale Tumori di Mi-

lano come biologo molecolare, sono entrato nel nuovo Istituto Telethon di Genetica e Medicina, formato da un gruppo di genetisti arrivati direttamente dal Genome Center del Baylor College di Houston, che credevano molto nell'applicazione della bioinformati-

ca alla genetica delle malattie umane. Come borsista ho affiancato Gyorgy Simon, oggi attuale direttore di Bioinformatica all'Axxam di Milano, che si occupava di tutta l'infrastruttura tecnologica e di programmazione bioinformatica. Dopo diversi anni al Ifgem, ho lavorato per un anno e mezzo all'Istituto Sanger di Cambridge in Inghilterra, occupandomi della gestione di database e di analisi di dati genomici. Dal novembre 2000 lavoro all'Ifom come analista programmatore nel gruppo di bioinformatica e dal gennaio 2004 sono il responsabile dei servizi bioinformatici».

Quali sono le esigenze principali di un sistema bioinformatico?

«Scalabilità, compatibilità con le piattaforme esistenti, estensibilità in termini di Cpu e spazio disco, potenza computazionale anche per calcolo multithread o parallelo, con sistemi di queueing appropriati, nonché affidabilità. Sul fronte della compatibilità, in particolare è importante l'indirizzamento di memoria, di gestione di grossi file system e via dicendo, con sistemi operativi non proprietari come Linux e FreeBSD, con compilatori come Gcc e, in generale, con software open source. Per capire l'entità della massa critica dei dati che ci troviamo a gestire, basti pensare che se nel 1999 le sequenze di Dna, in termini di singole basi, A, C, G o T, disponibili pubblicamente, erano poco più di tre miliardi, nel febbraio del 2004 sono arrivate a circa 38 miliardi, divise in più di 32 milioni di record (fonte Genetic Sequence

Data Bank, ndr). Per quello che riguarda gli investimenti previsti

per l'estensione del sistema bioinformatico, ci sposteremo sulla tecnologia cluster».

Quali sono i filoni principali di sviluppo dell'attività bioinformatica in Ifom?

«Il cuore del sistema di calcolo Ifom si basa su un server Unix altamente espandibile Silicon Graphics Origin2200, attualmente equipaggiato con 8 Cpu, 8 Gbyte di memoria e 1 Terabyte di spazio disco. Su questo sistema sono installate le principali banche dati pubbliche di acidi nucleici e di

proteine, con un corredo di sistemi di aggiornamento giornaliero interamente automatici e di programmi di ricerca e interrogazione dedicati. Sfruttiamo quest'architettura unitamente a un insieme di programmi bioinformatici che supportano i progetti di ricerca Ifom per procedure automatiche di lettura, filtraggio e analisi di una considerevole quantità di dati di sequenziamento o di espressione genica basati su microarray. Con i risultati sperimentali creiamo banche dati di facile interrogazione, sviluppando strategie di ricerca automatizzate su banche dati per la caratterizzazione e il raggruppamento funzionale dei dati grezzi derivati da sequenze prodotte al-

l'interno dell'Istituto. Inoltre, per facilitare la collaborazione tra Ifom e gli istituti partner, i programmi bioinformatici sono disponibili attraverso il nostro sito Internet».

Grazie a Internet e a nuove ottiche di condivisione e capitalizzazione delle risorse informative, i vecchi approcci alla ricerca sembrano profondamente mutati. Qual è la sua percezione in merito?

«La libera comunicazione degli approcci teorici, delle metodologie sperimentali e dei risultati ha sempre rappresentato il pilastro fondamentale su cui si basa l'attività scientifica. Quello che è cambiato negli ultimi tempi è il mezzo ➔ di comunicazione. Mentre la

Itanium come chiave di sviluppo per le attività di ricerca

Per far fronte a un'attività di elaborazione sempre più complessa e dettagliata, la divisione di bioinformatica agli inizi dell'anno scorso ha iniziato a vagliare la possibilità di aggiornare i propri strumenti It con macchine più evolute. Lo staff dei bioinformatici Ifom, e in particolare il direttore dei sistemi informativi, era curioso di testare la tecnologia Itanium e a fine estate 2003 l'Istituto ha ottenuto in prova una macchina Hp Itanium-based, ovvero il modello Rx2600.

«Su questa macchina abbiamo installato una serie di strumenti complessi - spiega Alessandro Guffanti, responsabile dei servizi bioinformatici di Ifom - e, in particolare, una serie di software specialistici, tra cui Spidey, Anti-Hunter, RepeatMasker e BlastN. Il nostro interesse principale era applicare alla bioinformatica un server più potente, ricco di funzionalità aggiuntive e, soprattutto, capace di

adattarsi a un ambiente multipiattaforma come il nostro. BlastN, per esempio, è molto difficile che riesca a funzionare su diverse architetture con i parametri necessari alla nostra ricerca. In pratica, si tratta di una ricerca in banca dati con cui si va a cercare con una stringa un'altra stringa, e su un database contenente milioni di stringhe, con i parametri utilizzati le criticità sono moltissime».

«Tutti parlavano di Itanium - precisa Michael Kahle, direttore dei sistemi informativi di Ifom -, ma nessuno l'aveva mai visto in azione. Per la nostra tipologia di lavoro, eravamo curiosi di sperimentare il livello di performance possibile grazie a questo nuovo processore. Se era vero quel che si diceva, cioè che a livello di server Itanium dovrà rappresentare il futuro sostituendo le altre tecnologie, allora era necessaria una verifica delle sue effettive potenzialità testando come funzionava "su strada", per esempio, nel porting da una piattaforma all'altra».

La richiesta ad Hp di un server dotato di processore Itanium rispondeva alla necessità di supportare alcune nuove attività di ricerca all'interno del dipartimento di bioinformatica, applicando una procedura Web based per realizzare un'economia di scala più ampia, mediante un'architettura capace di supportare l'analisi dell'intero genoma umano. Un altro progetto su cui sono state testate le performance dell'Rx2600 riguarda la gestione di una consistente mole di dati, generati nel laboratorio di bioinformatica, che consentono di monitorare l'espressione di tantissimi geni in varie condizioni. «Grazie alle performance di Rx2600 - aggiunge James F. Reid, ricercatore bioinformatico Int/Ifom - oggi possiamo analizzare e gestire una campionatura di oltre 10-20mila geni senza problemi di stoccaggio, archiviazione e recupero».

L.Z.



→ pubblicazione su riviste di tipo tradizionale come Nature e Science con revisione scientifica da parte di colleghi anonimi rimane lo sbocco tradizionale, prende sempre di più piede la comunicazione su riviste totalmente elettroniche che utilizzano gli stessi criteri ma con tempi molto più rapidi e, soprattutto, senza far pagare un vero e proprio abbonamento; si tratta di riviste di tipo open access, che richiedono una quota da pagare per articolo, che spesso può venire ridotta o annullata per ricercatori con difficoltà economiche, ma garantisce la stessa qualità scientifica delle riviste tradizionali cartacee e una diffusione dell'informazione molto più capillare anche nei paesi scientificamente

più periferici come quelli del terzo mondo».

In un processo evolutivo scientifico e tecnologico incalzante, come riesce a conciliare le sue necessità di aggiornamento con il suo lavoro?

«A parte i mezzi tradizionali di aggiornamento come corsi serali, articoli, libri o siti Internet, un contributo molto importante sono i corsi a distanza. Ho cominciato con i moduli in computer science della Open University e recentemente ho concluso un master in computer science presso un Poli-

tecnico inglese. Adesso sto cominciando un corso di statistica applicata, sempre presso un'altra Università inglese. Esistono ottimi corsi a distanza anche in Italia, ma non ancora a livello di master. Sicuramente si tratta di un investimento in termini di tempo tolto alla famiglia e anche economico notevole; Ifom mi concede i viaggi per gli esami o per "study

school" che rientrano come "viaggi di lavoro" perché considera ogni spesa in formazione un positivo investimento per la struttura nel suo complesso».

Quali sono i maggiori punti di criticità nel suo lavoro?

«Reperire e trattenere personale tecnicamente qualificato e identificare partner tecnologici affidabili. È relativamente facile trovare personale con qualifiche accademiche apparentemente appropriate ma è piuttosto difficile trovare perso-

nale con un corretta mentalità tecnico-scientifica e pratica. L'università dovrebbe lasciare molto più spazio a enti come il nostro nella formazione post-laurea, perché la sinergia tra ricerca e realtà lavorative nel campo del sapere è fondamentale. Un buon esempio è il master in bioinformatica di Milano Bicocca, dove quest'anno sono presenti tre docenti Ifom per la bioinformatica e l'analisi dei dati».

LAURA ZANOTTI



adattarsi a u



Alessandro Guffanti, analista programmatore e responsabile dei servizi bioinformatici presso l'Ifom (Istituto Firc di Oncologia Molecolare, www.bio.ifom-firc.it), la cui sede di Milano è ritratta nella foto di pag. 26.