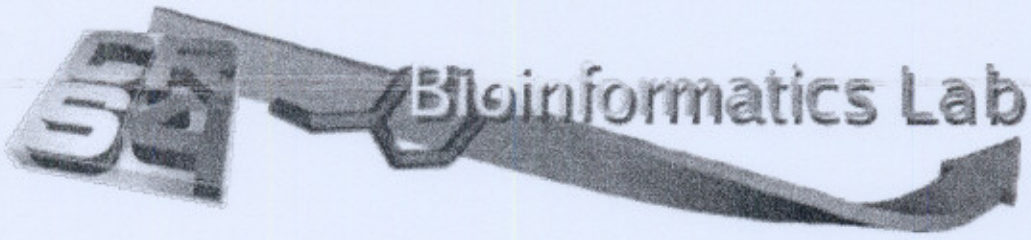

Ottobre 2007- Maggio 2008

Rapporto finale

Dr.ssa Patricia Rodriguez-Tomé

Progetto Cluster Biomedicina e Tecnologie della salute sottoprogetto Bioinformatica

1 - INTRODUZIONE

Il rapporto descrive la realizzazione del Progetto per il periodo che va dal 1 Ottobre 2007 al 15 Maggio 2008, dettagliato per work package.

Il Progetto ha avuto operativamente inizio con incontri informativi tra i diversi soggetti che hanno sottoscritto la dichiarazione di interesse, coordinati dal gruppo Servizi&Sviluppo del Laboratorio di Bioinformatica del CRS4. Durante gli incontri sono state individuate le principali aree di lavoro, utili per stabilire gli argomenti dei corsi di formazione, dei workshop e dei seminari che sono poi stati organizzati secondo gli obiettivi del Progetto stesso.

Il dialogo tra i diversi gruppi di lavoro è rimasto sempre aperto durante lo svolgersi delle varie fasi del Progetto, come descritto nelle sezioni sotto riportate. Nel BioPortale del Laboratorio di Bioinformatica del CRS4 è stata creata una sezione dedicata al Progetto cluster all'indirizzo: http://www.bioinformatica.crs4.org/services/projects/cluster07 (Fig. 1). Per ogni work package è stata dedicata una pagina web, da dove è possibile scaricare il materiale didattico relativo agli eventi di formazione.

Lo stato di avanzamento di ciascuno degli work package è dettagliato nella Gant chart della Fig. 2.

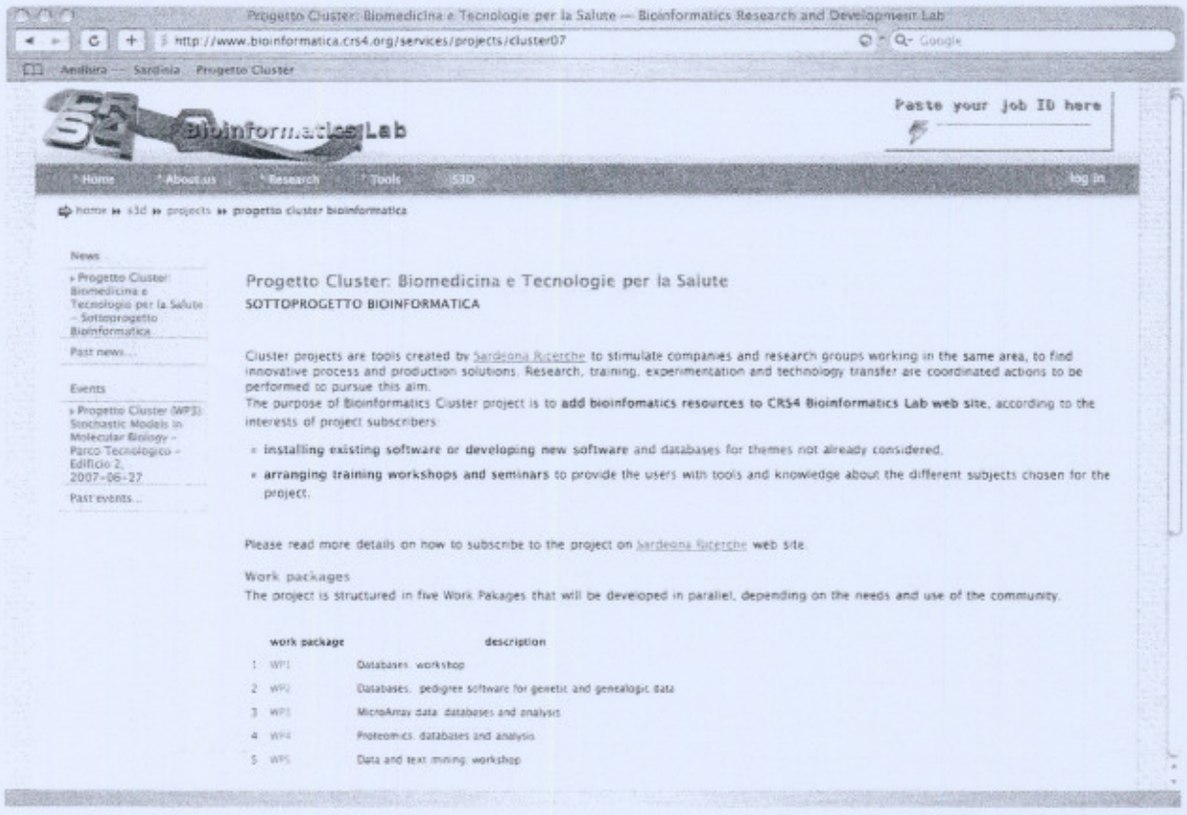

Fig. 1: BioPortale: la homepage dedicata al Progetto cluster.

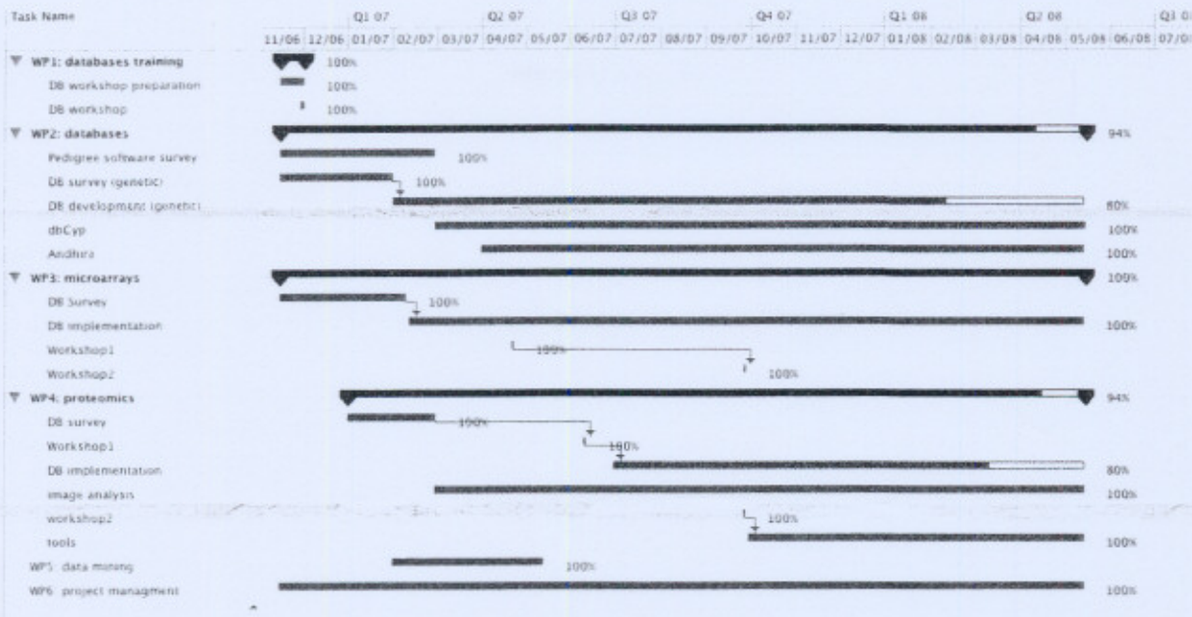

Fig. 2: Gant chart dello stato di avanzamento del Progetto al 15 Maggio 2008.

2 - WP1: Workshop base di dati

I database rappresentano uno strumento indispensabile per i ricercatori in biologia molecolare, per l'archiviazione e l'organizzazione della grande quantità di dati che viene prodotta giornalmente nei laboratori di tutto il mondo. Il primo passo previsto nel progetto per quest'area di interesse era l'organizzazione di un workshop, che si è volto dal 29 Novembre al 1 Dicembre 2006.

Argomenti trattati nel workshop (in allegato al rapporto una descrizione per esteso):

1. introduzione alle diverse teorie e tecnologie dei database;
2. teoria del database relazionale;
3. tecnologia del object oriented database;
4. progettare database: UML e Entity Relationship;
5. introduzione al linguaggio SQL;
6. strumenti di programmazione per l'accesso ai database (Perl, java, Python, PHP);
7. presentazione dei sistemi di database piú utilizzati in biologia molecolare: Oracle, MySQL, PostgreSQL.

Un rapporto sui sistemi di basi di dati esistenti è fornito in allegato.

3 - WP2: Analisi di dati genetici e genealogici

A - WP2.1 & 2.2: Basi di dati

A.1 - Progetto base di dati genetica: PGDS (Portable Genetic Database System)

È ormai noto che molte patologie hanno origini o predisposizioni genetiche. È quindi altamente interessante trovare i tratti genotipici, spesso denominati marker, che sono altamente correlati alle patologie di interesse. Questa attività investigativa è resa più accurata e semplice quando nell'analisi si possono unire ai dati genotipici, i dati genealogici dei soggetti e dei loro familiari.

Tali progetti di analisi genotipica e genealogica necessitano di un supporto informatico per la gestione dell'alto volume di dati coinvolti. Da questa necessità è nato il progetto Portable Genetic Database System (PGDS).

A.1.1 Obiettivi del progetto PGDS

Il progetto PGDS ha come obiettivo quello di adempiere ai particolari requisiti informatici delle analisi di dati genotipici e genealogici con lo scopo di individuare caratteristiche genetiche correlate allo sviluppo di patologie. In specifico si riferisce ai requisiti informatici attinenti allo stoccaggio, l'analisi statistica, e la visualizzazione dei dati. Quindi, il progetto prevede la consegna di:

- un database per lo stoccaggio dei dati genotipici e genealogici;
- i software necessari per accedere ai dati in maniera efficiente e controllata al fine di fornire assistenza in caso di aumento della produttività e di tutelare la riservatezza dei dati secondo la normativa europea;
- il collegamento del database ai software per l'analisi dei dati;
- il collegamento del database all'applicativo per la visualizzazione degli alberi genealogici.

I deliverable possono essere software già esistenti, magari adattati allo scopo di questo progetto. Nel caso che non fosse possibile o vantaggioso adottare applicativi già esistenti si prevede lo sviluppo di nuovi software.

A.1.2. Stato del progetto

a. Collaborazioni

Già prima di aver generato un prodotto facilmente utilizzabile il progetto PGDS ha suscitato interesse nel campo medico. In particolare, è nata una collaborazione tra il gruppo Servizi&Sviluppo del Laboratorio di Bioinformatica del CRS4 e il personale medico del Laboratorio di Ematologia dell'Ospedale "San Francesco" di Nuoro. Il lavoro di ricerca del Laboratorio di Ematologia necessita di un sistema come quello previsto dal progetto PGDS, il personale sta quindi fornendo consigli che assicureranno l'utilità del sistema in campo medico e scientifico. Inoltre, attraverso la collaborazione, si sta sviluppando un componente applicativo da inserire nei processi lavorativi del laboratorio per informatizzare le loro procedure e per raccogliere dati reali da inserire nel database PGDS, con lo scopo di usarli in future ricerche e pubblicazioni scientifiche.

b. Database

È già in funzione una base di dati per lo stoccaggio dei dati importanti per il progetto PGDS e le funzioni del Laboratorio di Ematologia. Dopo uno studio di confronto tra i vari RDBMS disponibili è stato preferito, per motivi tecnici ed economici, PostgreSQL. Lo schema della base di dati, riportato nella Fig. 3, rappresenta le entità identificate nei dati e le loro relazioni, ma potrebbe necessitare di ulteriori evoluzioni per massimizzare le prestazioni del database nelle elaborazioni delle richieste più frequenti. Tali ottimizzazioni dell'efficienza del database verranno considerate in fasi future del progetto.

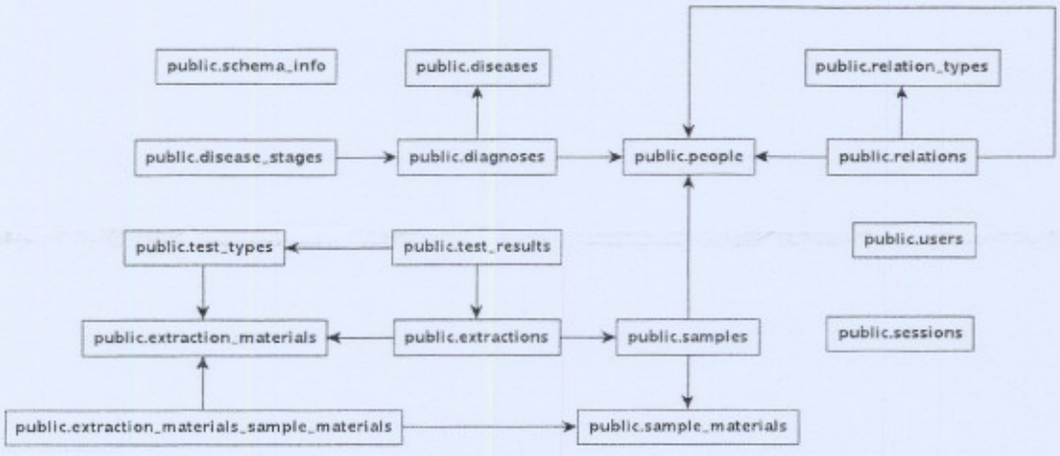

Fig 3: Schema della base di dati PGDS.

A.1.3 Applicativo per l'accesso ai dati

a. Architettura

Una caratteristica fondamentale per il successo del sistema prodotto dal progetto PGDS è la centralizzazione dei dati. Questa caratteristica implica che gli stessi dati, contenuti nel database, siano accessibili da più stazioni di lavoro via rete telematica. In quanto la caratteristiche di tali punti di accesso non sono ben definite si è deciso di adottare un'architettura client-server e implementare l'applicativo per l'accesso ai dati come applicazione web. Quest'architettura consentirà agli utenti di accedere ai loro dati attraverso tutte le piattaforme utilizzate sul desktop (per es. Mac OSX, Linux e Windows), e di accedervi contemporaneamente e da qualsiasi rete autorizzata.

b. Implementazione

L'implementazione dell'applicazione web è basata sul framework Ruby On Rails. Questo framework è stato scelto per la sua buona architettura e funzionalità che consente di arrivare a ottimi livelli di produttività degli sviluppatori. Inoltre le librerie del framework hanno facilitato la creazione di una batteria di test automatici per assicurare l'alta qualità del software prodotto, e minimizzare la durata della fase di test e i disagi agli utenti finali.

c. Funzionalità già presenti

L'applicativo possiede già molte delle funzionalità richieste per la versione finale:

- gestione di un elenco di pazienti e dei loro dati anagrafici (Fig. 4);
- registrazione diagnosi e monitoraggio delle condizioni dei pazienti;

- registrazione prelievi di materiali biologici effettuati (Fig. 5);
- registrazione dei test effettuati sui materiali prelevati, dei loro esiti e dell'operatore che ha effettuato il test (Fig. 6);
- accesso via rete telematica con comunicazione criptata;
- vari riassunti statistici dei dati gestiti dal database;
- gestione e autenticazione del personale che accede all'applicazione, con possibilità di differenziare le funzionalità disponibili in base al ruolo dell'utente.

In quanto alle funzionalità di collegamento ai software per analisi e di collegamento all'applicativo per la visualizzazione dei pedigree, il progetto è ancora in fase di pianificazione.

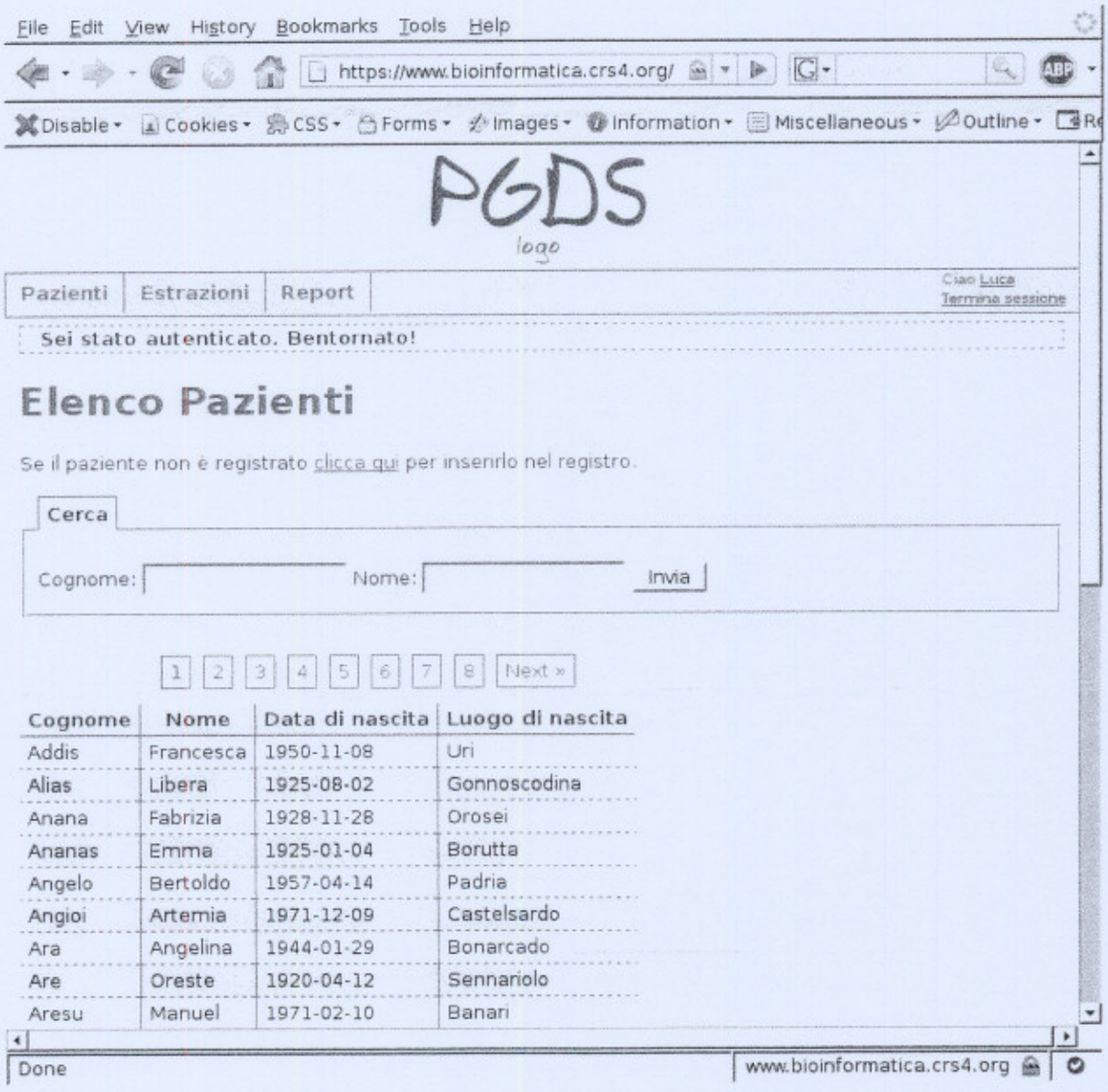

Fig. 4: Elenco Pazienti (i nomi sono esemplificativi).

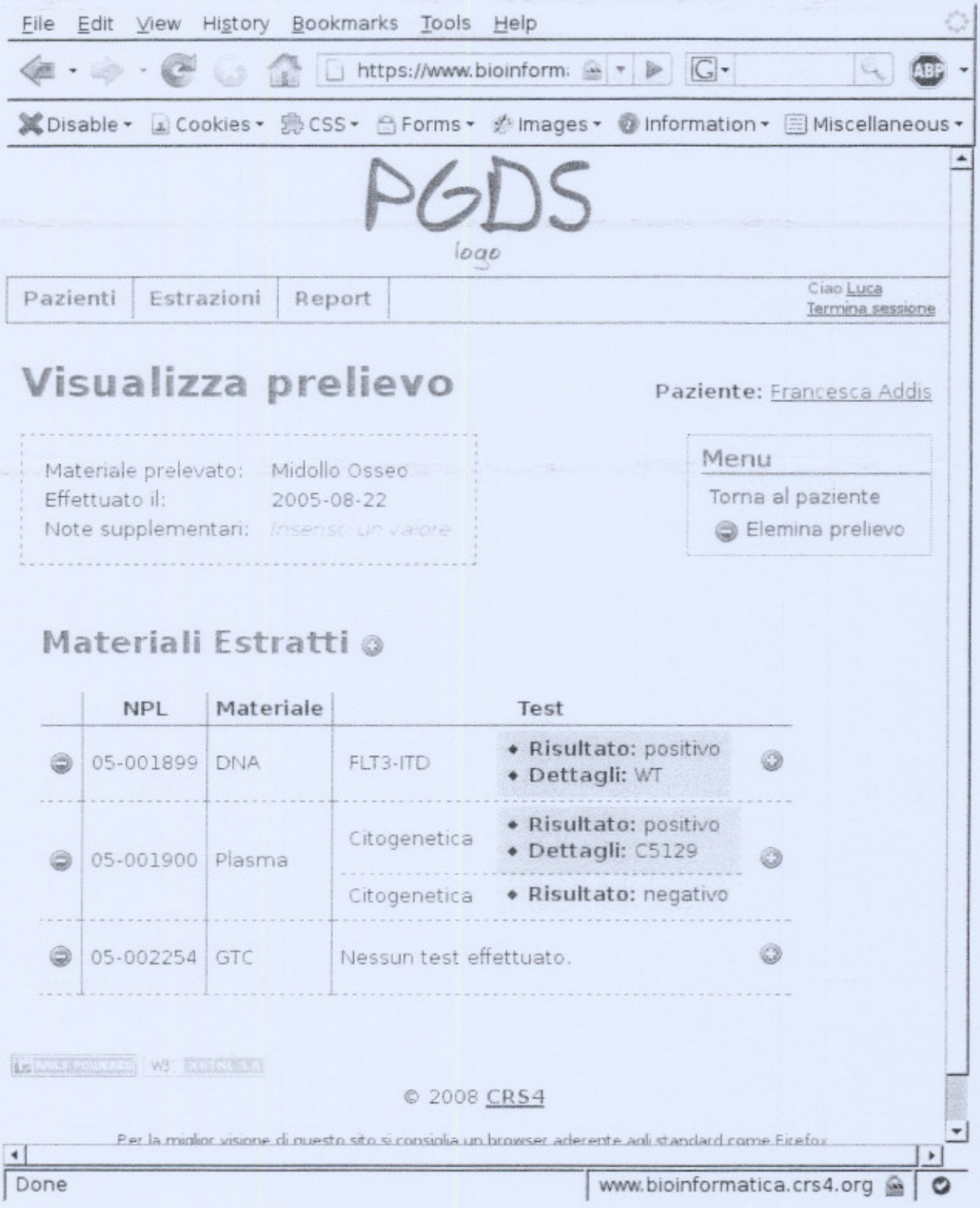

Fig. 5: Visualizzazione di un prelievo.

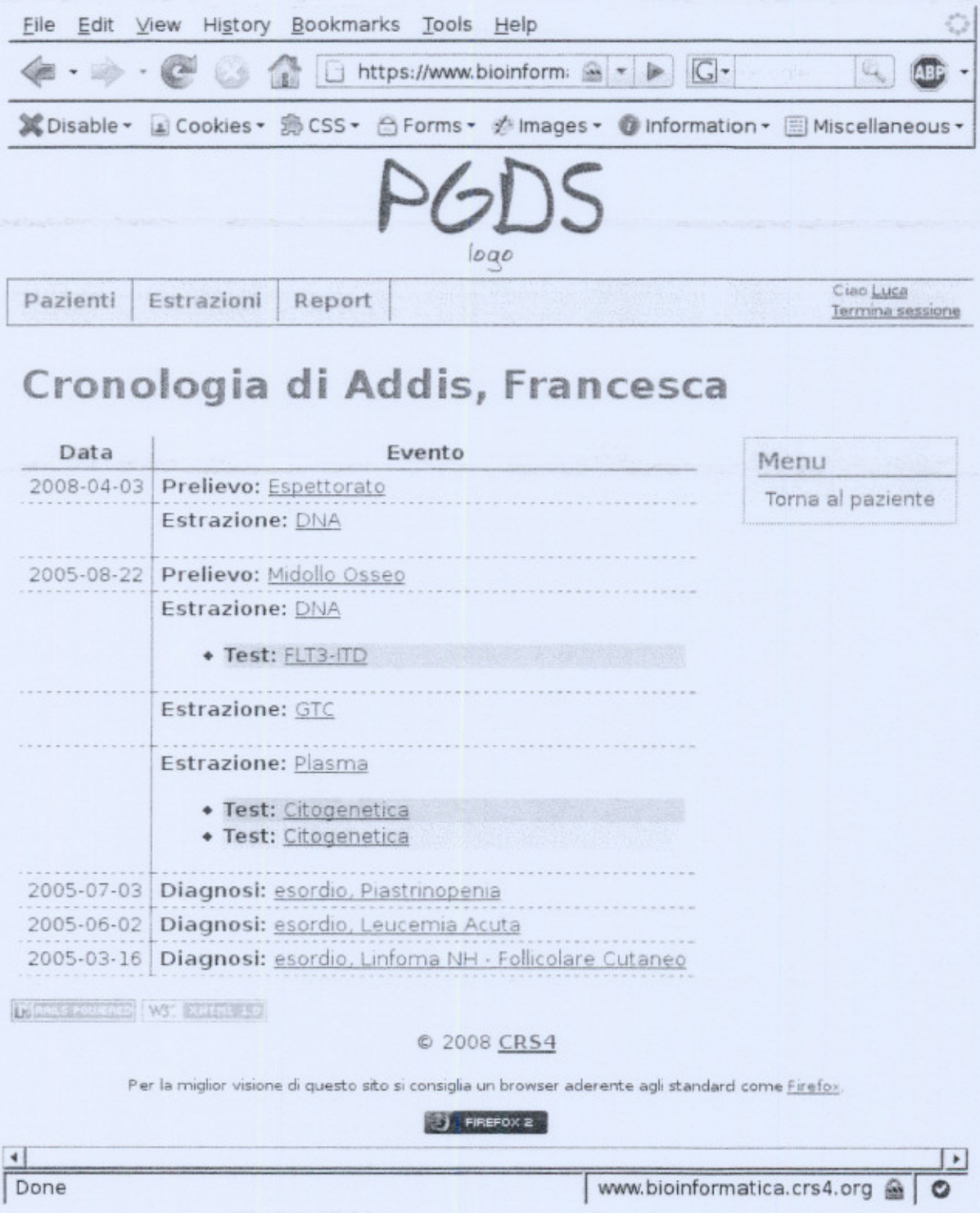

Fig. 6: Cronologia di un paziente (i nomi sono esemplificativi).

A.1.4 Futuro del progetto

a. Collaborazioni

Si prevede di continuare la collaborazione con l'Ospedale "San Francesco" che ha già prodotto buoni risultati e che garantisce che il prodotto del progetto PGDS sia utile agli utenti in campo medico e scientifico.

b. Database

Il database resterà nella sua forma attuale fino alla fase di ottimizzazione dei software.

c. Applicazione

Sono state identificate numerose funzionalità di interesse nell'applicazione PGDS. Di seguito le funzionalità approvate per l'implementazione:

- rappresentazione più flessibile dei risultati dei test registrati, in modo da poter archiviare immagini, sequenze di amminoacidi e nucleotidi, ecc;
- archiviazione della provenienza geografica dei soggetti, per poter studiare correlazioni tra zona di provenienza e probabilità di sviluppare particolari patologie;
- resoconti per fornire statistiche sullo sviluppo di patologie e il funzionamento del laboratorio che usa il sistema;
- archiviazione delle relazioni genealogiche tra pazienti e altri soggetti registrati nel database;
- sistema di registrazione degli accessi per rilevare accessi anomali.

Le sopra elencate funzionalità saranno implementate a breve e sottoposte alla valutazione degli utenti dell'Ospedale "San Francesco".

d. Collegamento ai software per analisi

Il collegamento del database ai software per l'analisi dei dati richiede uno studio più approfondito. Con gli utenti si dovranno definire i tipi di analisi richiesti ed arrivare ad una lista di applicativi con cui è desiderata la connettività. Per le analisi già definite saranno valutate le seguenti opzioni:

- fornire la possibilità di eseguire le analisi server-side;

- fornire la possibilità per l'utente di un'esecuzione indipendente, dopo essersi assicurati che l'interfaccia per l'accesso automatico ai dati fornisca i valori necessari.

Per gli applicativi, invece, bisognerà studiare una strategia ad hoc.

e. Collegamento all'applicativo per la visualizzazione dei pedigree

In quanto alla visualizzazione dei pedigree, precedenti fasi di questo progetto hanno concluso che nessuno degli applicativi già disponibili corrisponde esattamente alle esigenze del progetto PGDS, e che i tempi del progetto non consentono lo sviluppo di un'applicazione realizzata ad hoc. L'unica opzione quindi è di adattare un programma già esistente ai requisiti del progetto. In quanto questa funzionalità è stata richiesta da alcuni utenti, si farà il possibile per implementarla, forse integrando le funzionalità create per accedere agli altri dati.

f. Adeguamento alla direttiva sulla privacy

Sono stati investigati i requisiti imposti dalla direttiva europea sulla privacy in quanto i dati relativi alla salute dei pazienti sono denominati sensibili. La normativa impone:

- la protezione dei dati memorizzati e archiviati con funzioni crittografiche,
- l'autenticazione forte degli utenti,
- un sistema di registrazione degli accessi per rilevare eventuali anomalie.

Per rispettare questi requisiti l'applicazione conserverà i dati utili per identificare i pazienti in forma criptata, utilizzando l'algoritmo AES, ovvero lo standard attuale per la cifratura simmetrica. Questo provvedimento proteggerà i dati nel caso che i dischi venissero smarriti o rubati.

L'accesso ai dati cifrati verrà autorizzato solo agli utenti che ne hanno necessità, ovvero i medici che compilano referti e analizzano i dati per derivarne diagnosi. Il resto degli utenti, come gli operatori tecnici del Laboratorio di Ematologia, lavoreranno esclusivamente con i codici identificativi dei campioni.

Per autenticare gli utenti privilegiati, che hanno diritto ad accedere ai dati sensibili, sarà adoperata una tecnica di autenticazione forte per aderire alla normativa vigente. L'autenticazione forte (Strong Authentication in inglese) richiede dagli utenti due diversi fattori autenticativi. Il piano è di richiedere una parola chiave e un codice generato da un dispositivo identificativo, di cui l'utente sarà munito quando gli verrà approvata la richiesta di autorizzazione all'accesso.

Inoltre, la direttiva sulla privacy richiede che il sistema registri gli accessi al sistema per permettere eventuali verifiche e rilevare accessi anomali. Questa funzionalità è già in via di progettazione per l'applicazione PGDS.

Infine, è necessario proteggere i dati in transito tra sistemi informatici, ovvero mentre passano nelle varie reti telematiche. L'applicazione già rispetta questo requisito utilizzando la tecnologia di cifratura SSL per trasmettere i dati dal server di database al server di applicazione, e l'HTTPS per trasmettere i dati all'utente.

Eventi:

a) 11 Ottobre 2007: PGDS è stato installato all'Ospedale "San Francesco" di Nuoro, Laboratorio di Ematologia, con i dati prodotti dallo stesso laboratorio. Al momento è in fase di test da parte di questi utenti.

b) Fine Ottobre 2007: Sergio Contrino, sviluppatore di PGDS ha lasciato il gruppo di Bioinformatica, Dal 15 gennaio un nuovo sviluppatore ha preso il suo posto nella realizzazione del progetto.

A.2 - Progetto ANDHIRA:

Il gruppo di ricerca del professor Mauro Ballero dell'Università di Cagliari ha aderito al Progetto con la richiesta della progettazione e costruzione di un database della flora endemica della Sardegna. "ANDHIRA" è il nome del DB, sviluppato con PostregSQL, un sistema open source che può essere installato su Linux, PC/Windows e Apple computer. ANDHIRA è stato pensato per organizzare e gestire testi e immagini sulle proprietà botaniche, fitochimiche e farmaceutiche delle piante sarde.

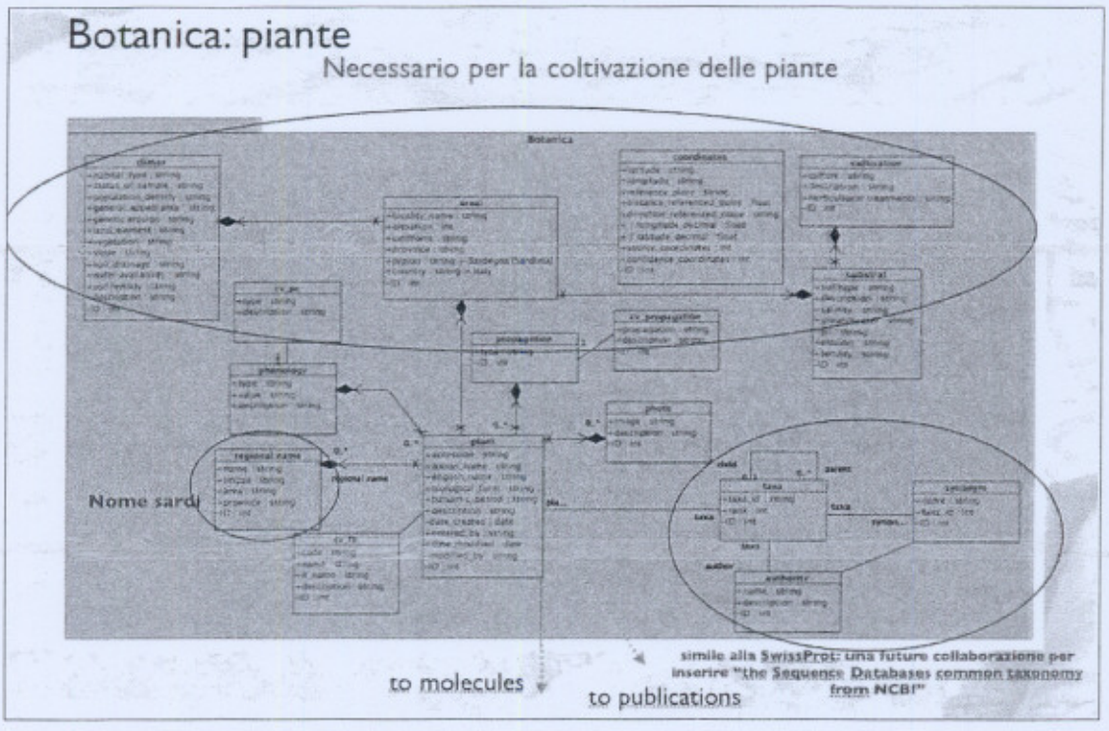

Fig 7: Schema del database ANDHIRA.

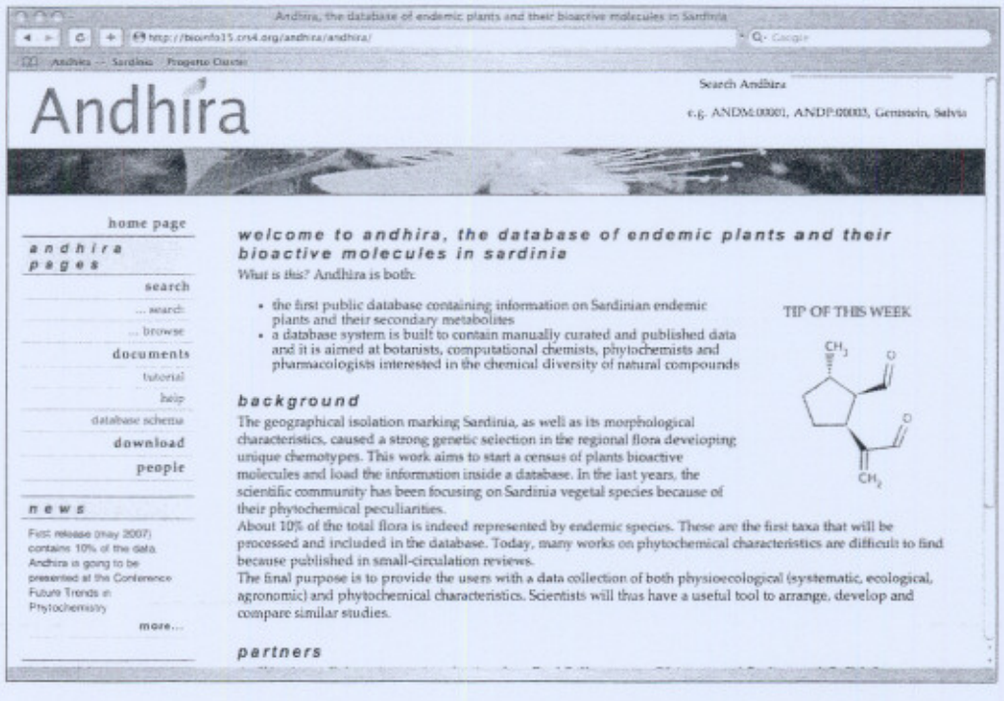

Fig. 8: Pagina Web del progetto ANDHIRA.

Al termine del lavoro sarà disponibile un'interfaccia web user-friendly, attraverso cui effettuare ricerche su ANDHIRA. Sarà possibile, ad esempio, sapere se e quali piante endemiche sarde producono sostanze testate per attività antimicrobica, antitumorale o psicoattiva.

Per l'inserimento manuale dei dati è stato utilizzato il programma SEU-PEDRo, per l'inserimento automatizzato dei dati è stato creato un apposito "script". Sono stati inseriti tutti i dati sulle piante endemiche della Sardegna facilmente reperibili in articoli e libri, ma mancano dati etnobotanici, di più difficile accesso.

Il progetto ha richiesto il coinvolgimento di 3 collaboratori (i cui contratti sono stati finanziati dal progetto cluster) per lo sviluppo del sistema, per il reperimento dei dati botanici e chimici e per il loro inserimento nel database.

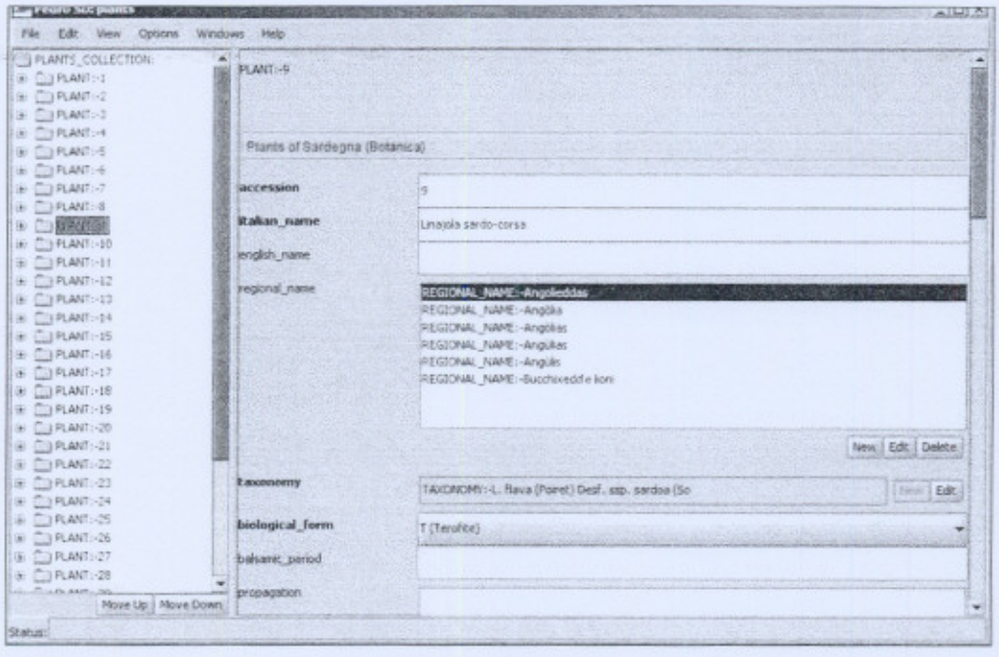

Fig. 9: L'interfaccia seuPedro per i dati botanici.

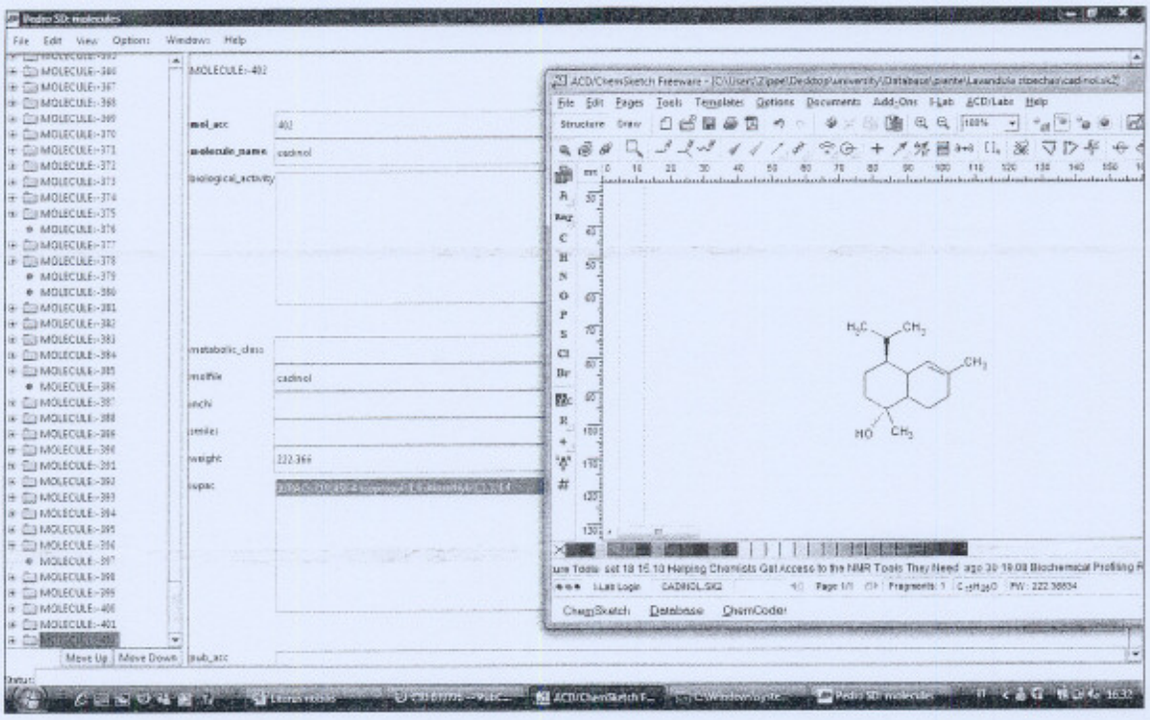

Fig. 10: L'interfaccia seuPedro per i dati chimici.

Eventi:

In collaborazione con l'associazione "Arca verde Otonga" e Sardegna Ricerche, è stato organizzato il convegno "La Biodiversità come opportunità di Sviluppo e Cooperazione: Dal modello sardo alla foresta di Otonga", tenutosi il 26 Ottobre 2007 a Cagliari (in allegato la locandina del convegno). Durante il convegno il progetto ANDHIRA è stato presentato al pubblico e alla stampa (in allegato la presentazione).

LA NUOVA

1 – La Nuova Sardegna
Pagina 2 - Cagliari
BIODIVERSITA'
UN DATA BASE PER LE PIANTE

CAGLIARI. Tutelare la biodiversità significa creare opportunità di sviluppo. Altro che disboscare e costruire. L'ambiente può diventare una grande fonte di reddito non solo per l'agricoltura e il turismo, ma anche per la farmaceutica e la chimica. Perché ciò avvenga, le informazioni devono essere elaborate e sistemate in modo da essere fruibili ai ricercatori. È con questo obbiettivo che Matteo Floris, un ricercatore di bioinformatica del CRS4 ha messo a punto Andhira, un database "intelligente" delle molecole bioattive e delle piante endemiche sarde con proprietà fitochimiche e farmaceutiche. Il database è stato presentato venerdì in un convegno sul tema "La biodiversità come opportunità di sviluppo e cooperazione", organizzato dall'associazione l'Arca verde Otonga, dal CRS4 e da Sardegna Ricerche, con lo scopo di creare un ponte di cooperazione fra la Sardegna e l'Ecuador, dove ha sede la fondazione Otonga, nell'omonima regione andina, una delle aree del pianeta più importanti per la sua biodiversità. La fondazione è impegnata nella salvaguardia della foresta attraverso l'acquisto di aree da sottoporre a tutela, l'educazione delle popolazioni, le adozioni a distanza per finanziare la cultura e la ricerca sulla biodiversità. Il modello di lavoro usato per il database Andhira consentirà di trasferire know how anche a Otonga. A fare da ponte fra i due paesi è l'associazione Arca verde Otonga, nata a Cagliari nel 2006 con lo scopo di tutelare ovunque la biodiversità. «La biodiversità va affrontata a livello globale - ha detto il presidente Giampaolo Ruzzante - e noi cerchiamo di far conoscere situazioni vicine e lontane». Secondo Giovanni Onore, della fondazione Otonga, anche in Sardegna, dove non ci sono tante risorse, l'ambiente va protetto.
Stefania Sidd

Fig.11: Articolo de "La Nuova Sardegna".

Futuro del progetto ANDHIRA:

Il progetto ANDHIRA si è concluso il 15 Maggio 2008.

A.3 - Progetto MMsINC

Come estensione del database ANDHIRA, é stato iniziato il progetto "MMsINC": un database molecolare (circa 5 milioni di molecole), derivato dal vecchio database "Zinc", ripulito di tutte le informazioni ridondanti ed errate. Lo strumento nasce dalla collaborazione con il "Drug Design Laboratory" del Prof. Stefano Moro, dell'Università di Padova, ed è attualmente in via di sviluppo: sono in via di costruzione una piattaforma di virtual screening e un algoritmo per ricerche rapide di similarità.

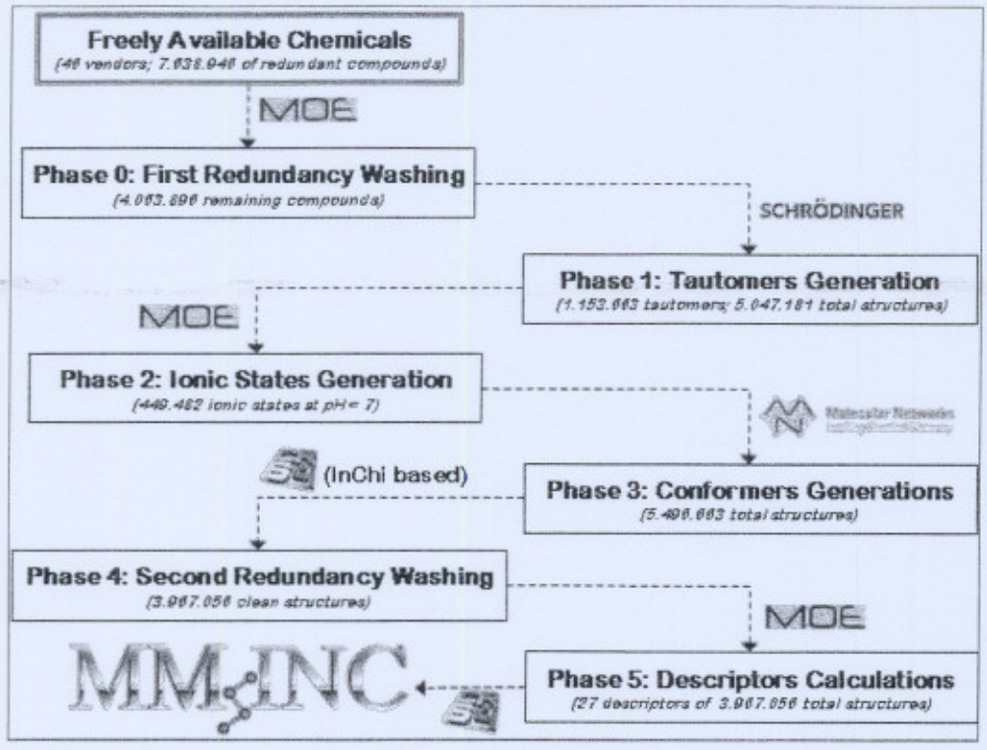

Fig. 12: Fasi di realizzazione di MMsINC.

MMsINC consente di selezionare una molecola contenuta nel DB per le sue proprietà di struttura o chimico fisiche. E' il primo DB al mondo che per ogni molecola può dare la similarità rispetto ai ligandi contenuti nella Protein Data Bank (PDB).

MMsINC è dotato di un'interfaccia web user-friendly, che consente di confrontare una molecola di interesse con quelle contenute nel database. Grazie alla base di dati MMsINC si ha a disposizione una piattaforma per analisi di virtual screening.

Futuro del progetto MMsINC:

Il progetto MMsINC si è concluso il 15 Maggio 2008.

A.4 - Progetto dbCYP

I polimorfismi degli isoenzimi del Citocromo P450 umano (CYPs) sono tra i più studiati in farmacogenetica, perché sono responsabili della maggior parte del metabolismo fase I dipendente di più del 50% dei farmaci usati nella pratica clinica, avendo un ruolo importante nell'attivazione e inattivazione di sostanze cancerogene e tossine, nella biosintesi e inattivazione di diversi ormoni e altri composti endogeni.

Durante l'evoluzione, i geni codificanti i Citocromi P450 hanno accumulato diverse mutazioni e vari tipi di riarrangiamenti genici, causando una variabilità nel fenotipo, che va dalla perdita totale di attività dell'enzima a varianti in cui l'attività è invece accresciuta. La variabilità genetica nei geni codificanti i CYP è perciò determinante nella differente suscettibilità individuale all'azione dei farmaci e altre sostanze chimiche ambientali, così come ha un ruolo nella differente patogenesi di malattie.

Il sempre crescente numero di alleli CYP identificati e caratterizzati ha indotto a creare una risorsa Internet costantemente aggiornata: il sito web del CYP Allele Nomenclature (www.cypalleles.ki.se). Il sito raccoglie un elenco frequentemente aggiornato delle mutazioni conosciute dei geni in esame, etichettando con una sigla ciascun allele secondo la nomenclatura stabilita dal Human Cytochrome P450 Allele Nomenclature Committee.

Si tratta però di informazioni non strutturate, difficilmente utilizzabili da sistemi informatici. E' stato quindi creato un database a cui è stato dato il nome di dbCYP, prendendo spunto, per la scelta dei dati da inserire, dal sito appena menzionato. In particolare ogni allele è stato definito mediante una determinata combinazione di mutazioni.

Per la creazione del dbCYP sono state analizzate le pagine web e, mediante un parser in linguaggio Perl, sono state estratte le seguenti informazioni sui citocromi: il nome dell'allele; la lista delle mutazioni caratteristiche; eventuali link che puntano al database degli SNP; altri nomi con cui è conosciuto l'allele che non rispettano la nomenclatura standard; l'effetto sulla proteina, in termini di variazioni di amminoacidi; l'effetto in vivo e in vitro sull'attività dell'enzima; la lista di pubblicazioni che hanno segnalato l'allele.

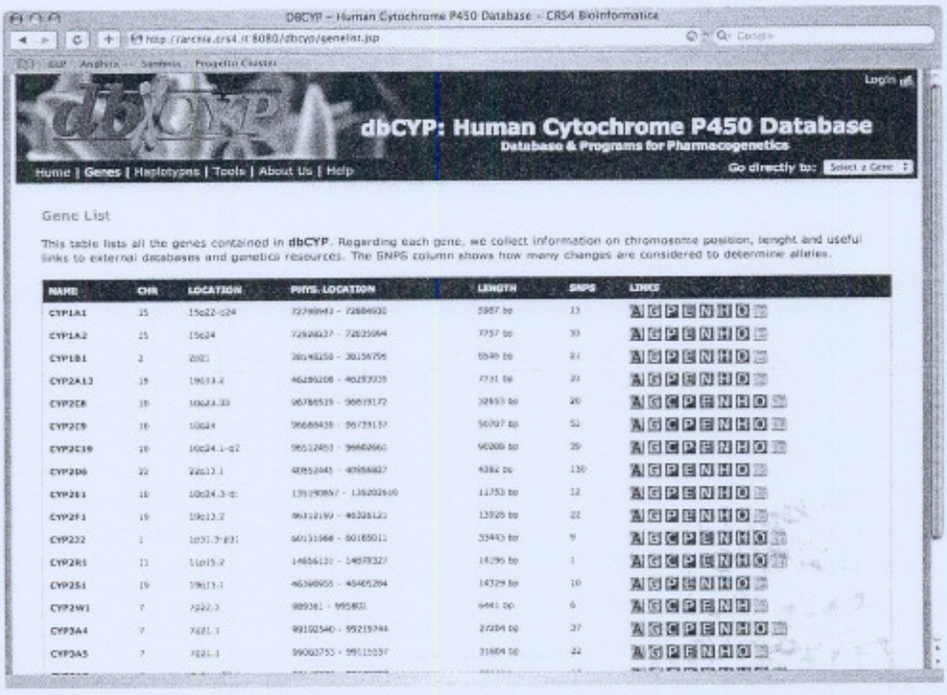

Fig. 13: Pagina Web di dbCyp.

In seguito le informazioni memorizzate nel database sono state utilizzate per creare una serie di strumenti che permettono di (1) stimare in maniera automatica l'aplotipo dei campioni, (2) suggerire il set

minimo di polimorfismi da utilizzare per genotipizzare i campioni, (3) interrogare il database per allele, per mutazione, per variante, per effetto.

Il Laboratorio di Bioinformatica ha sviluppato una base di dati (dbCYP), in cui sono già stati inseriti i dati pubblici disponibili. Lo sviluppo dello strumento è proseguito con la costruzione delle query e dell'interfaccia web per gli utenti.

Futuro del progetto dbCYP:

Il progetto dbCYP si è concluso il 15 Maggio 2008.

B - WP2.3 & 4: Pedigree software

I software per la costruzione di pedigree sono usati per disegnare l'albero genealogico delle famiglie e mostrare graficamente i risultati degli studi di genotyping e le informazioni cliniche. Secondo quanto previsto dal Progetto è stato stilato un rapporto sui programmi esistenti per la costruzione di pedigree, che è stato presentato durante un incontro con i gruppi interessati nel Marzo 2007.

Alcuni programmi sono stati testati ed è stato effettuato un confronto fra loro:

Open-source/GNU licence: programmi che possono essere modificati ed adattati.

1 - COPE (linguaggi di programmazione: Java, piattaforme compatibili: tutte)

E' stato sviluppato per la prima volta all'Istituto Genethon in Francia. Il codice sorgente è disponibile, ma non sono disponibili ulteriori sviluppi. Il programma non permette di disegnare il pedigree in modo interattivo.

2 - MADELINE 2.0 (linguaggi di programmazione: C++, piattaforme compatibili: Linux)

Questo programma richiede un motore grafico molto performante. Non si avevano a disposizione pedigree abbastanza complessi per effettuare un test. L'installazione del programma è complessa e richiede il lavoro di un operatore specializzato su Linux. L'output è fatto in SVG e può essere visto con qualsiasi browser. Il programma è in via di sviluppo.

3 - HAPLOPAINTER (linguaggi di programmazione: Perl, piattaforme compatibili: tutte)

E' un progetto puramente open source, si trova nel sito sourceforge.net ed è scritto in perl/tk. E' stato sviluppato perchè potesse gestire dati sugli aplotipi e per poter visualizzare l'output di progetti di genotipizzazione. E' l'unico programma trovato capace di gestire questo tipo di dati (di tali dimensioni). Sono previste alcune opzioni per gestire i grafici, ma non è prevista la modalità interattiva . Il programma non è adatto per la gestione di dati clinici e di fenotipo.

Prodotti commerciali: non possono essere modificati.

1 - PEDNAVIGATOR (linguaggi di programmazione: Java, piattaforme compatibili: tutte)

E' un programma, sviluppato dalla società SharDna, di cui esistono due versioni: una versione Lite che può essere scaricata e installata gratuitamente, e una versione commerciale. L'applicazione è stata sviluppata con la tecnologia Servlet di Java e si appoggia quindi ad un application server Tomcat-Apache. Il vantaggio di questo tipo di installazione è che tutti gli utenti accedono sempre all'ultima versione del programma, senza doverlo mantenere in locale. Lo svantaggio principale è che il programma non può essere eseguito senza una connessione alla rete.

2 - PED5 (Piattaforme compatibili: Windows)

Secondo quanto riportato da Bennet et al., il programma:

- permette di disegnare interattivamente
- può leggere ed esportare a linkage file
- accetta dati clinici, genealogici e di aplotipo

La versione del programma scaricabile gratuitamente è estremamente limitata e non permette di valutarne tutte le capacità. Non è chiaro, in oltre, se la limitazione del numero di marker con cui si può lavorare (18), vale solo per la versione test.

3 - CYRILLIC2 (Piattaforme compatibili: Windows)

Il programma permette di disegnare interattivamente, di fare analisi di haplotyping e di esportare verso vari programmi. Sono disponibili strumenti per gestire, importare ed esportare i dati sulle famiglie. La versione demo, anche in questo caso, non consente di testare in modo soddisfacente l'applicazione.

4 - CYRILLIC3 (Piattaforme compatibili: Windows)

E' un'altra versione del software CYRILLIC, orientata ai dati clinici, ma che manca della gestione dei dati di genotyping e di aplotipo.

5 - PROGENY (Piattaforme compatibili: Windows; ha una versione web scritta in Java, non accessibile con Firefox/ Linux)

Una delle applicazioni più complete e complesse della serie. E' integrata con un database che include informazioni su marker conosciuti come, ad esempio, la loro localizzazione sul cromosoma. L'applicazione è in grado di gestire vari tipi di informazione anche sul fenotipo e l'aplotipo. Può restituire output in numerosi formati e consente di gestire "microplates studies" generando dati compatibili con i sistemi più diffusi.

L'unico svantaggio potrebbe risiedere nell'impossibilità di estendere l'applicazione a causa dell'uso di formati proprietari. Per esempio non è possibile connettere Progeny ad altri database diversi da quello proprio del software.

6 - PEDDRAW (Piattaforme compatibili: Mac OS)

E' un'applicazione disponibile su MacOSX. La versione precedente è ancora disponibile per vecchi MacOS. Peddraw non consente di disegnare interattivamente i pedigree. Il pedigree viene disegnato nel momento in cui i dati della famiglia sono caricati.

Commenti

Sono state trovate molte pubblicazioni su programmi per il disegno di pedigree sviluppati appositamente per un singolo progetto, ma che non sono stati poi mantenuti o che non sono più disponibili. Esistono poi molti programmi per gli accoppiamenti di animali, utili per il disegno di pedigree complessi, ma non sono stati pensati per i dati genetici.

Si è concluso che:

- nessuno dei programmi testati corrisponde perfettamente alle esigenze degli utenti del Progetto;
- molti dei suddetti programmi sono commercializzati.

Il principale svantaggio dell'uso dei software commerciali è dovuto al loro comportamento rispetto ai formati di input e output dei nuovi programmi di analisi. Un altro problema consiste nella difficoltà di connetterli direttamente ai database, per cui è indispensabile spesso l'uso di file intermediari.

I programmi sviluppati dalle comunità accademiche spesso non sono aggiornati in maniera regolare. Provando a scaricare il codice sorgente di alcuni software per poterlo modificare, appare chiaro che il lavoro richiesto è di tale entità da rendere preferibile lo sviluppo di un nuovo programma.

Nessun utente del Progetto ha espresso la richiesta dello sviluppo di un programma per pedigree. Lo studio effettuato sui programmi esistenti aveva inoltre rivelato che il lavoro richiesto per lo sviluppo di un programma non era compatibile con i tempi del presente Progetto.

Eventi:

Il 4 Febbraio 2008, seminario: "Approcciare lo Studio della Natura con la Chimica Computazionale", Marco Masia -Dipartimento di Chimica, Università degli Studi di Sassari.

4 - WP3 Microarray

La tecnologia MicroArray è sempre più utilizzata per la ricerca sull'espressione genica e il genotyping. Con l'uso di questa tecnologia è possibile produrre un'enorme mole di dati in poco tempo, dati che devono essere archiviati e analizzati.

A - WP3.1 & 3.2: Basi di dati

Il Laboratorio di Bioinformatica ha installato il sistema BASE: un sistema completo per la gestione dei dati di laboratorio, per cui esiste la possibilità di includere strumenti di analisi. In BASE è incluso, come database system, MySQL, che permette l'utilizzo del sistema da parte di più utenti contemporaneamente, ma con un controllo di sicurezza che rende visibili i dati solo da parte del proprietario. Per gli ospiti del Parco tecnologico è possibile testare il sistema BASE, accessibile sulla rete del Parco attraverso l'indirizzo http://www.bioinformatica.crs4.org/tools/dbs/baseDB , o in postazioni dedicate nei locali del Laboratorio di Bioinformatica.

BASE è in fase di test da parte degli operatori della Piattaforma di Genotyping del Parco tecnologico. Sono in fase di sviluppo gli strumenti che rispondono alle esigenze specifiche della Piattaforma di Genotyping.

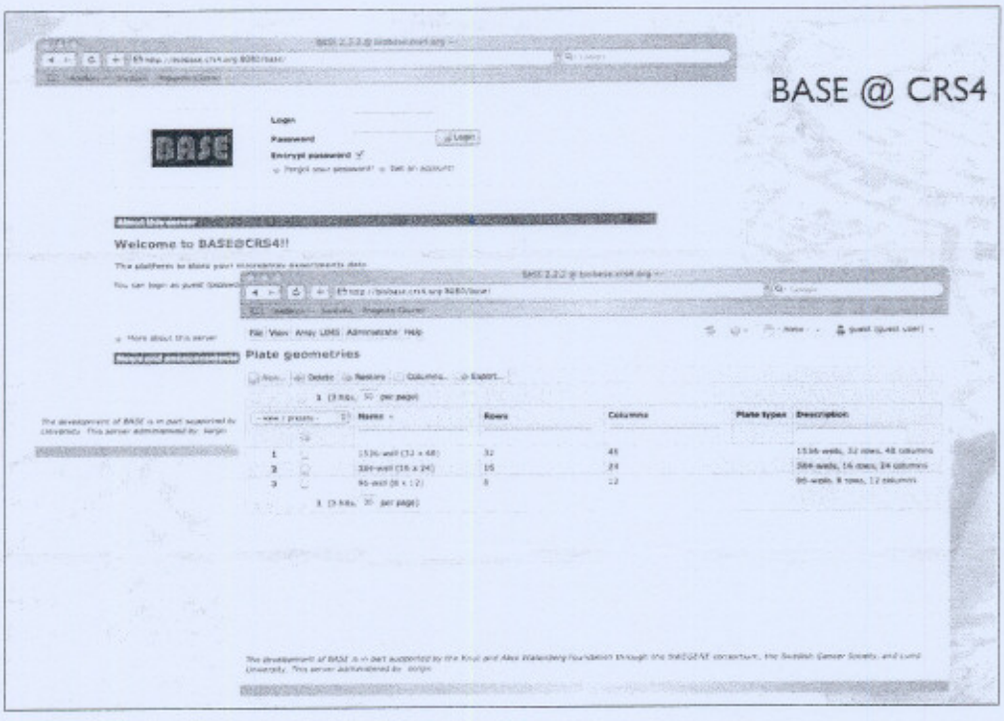

Fig. 14: Istallazione di BASE al CRS4.

Eventi:

a) Workshop: BASE è stato presentato agli utenti del Progetto cluster durante un workshop, il 23 Aprile 2007, dal titolo "Gestione e Analisi di Dati di Microarray".

1. Dr. Andrea Angius (CNR-Istituto di Genetica delle Popolazioni e Laboratorio di Genotyping di Sardegna Ricerche) - "*Utilizzo di GeneChip Affymetrix per l'analisi ad alta processività di SNPs: determinazione dei parametri genetici e applicazioni agli studi di associazione*".

2. Dr. Joel Masciocchi (CRS4 – Bioinformatica) - "*Basic Base*".

3. Dr.ssa Helen Parkinson (European Bioinformatics Institute) - "*ArrayExpress – A Public database for Gene Expression Data at EBI*".

4. Dr. David Kreil (Boku University Vienna) - "*Transcript profiling by microarrays: experiment and analysis*".

b) Seminario del 27 Giugno 2007

Dr. Greg Rempala (University of Louisville, KY, USA) - "*Stochastic Models in Molecular Biology*".

c) Un nuovo collaboratore lavora sul progetto dal 20 Novembre 2007.

Futuro del progetto BASE:

Database e sistema continueranno a essere sviluppati presso il Laboratorio di Bioinformatica del CRS4.

B - WP3.3 & 3.4: Analisi

I programmi di analisi per dati di microarray sono studiati in correlazione con BASE.

Eventi:

Dal 27 al 28 di settembre si è tenuto il workshop sull'argomento "micorarray analysis"

27 Settembre - Lezioni pratiche

1. Dr.ssa Gabriella Rustici (EBI) - "*Array Express*".
2. Dr.ssa Eugenia Migliavacca, (Swiss Institute for Experimental Cancer Research, Lausanne) - "*Normalization, quality controls and differentially expressed genes*".

28 Settembre - Seminari

1. Dr. Korbinian Strimmer (University of Leipzig) - "*Multiple Testing In Genomics And Proteomics: A Novel Unified Approach For Estimating False Discovery Rates*".
2. Dr. Silvio Bicciato (Università di Padova) - "*Integrative analysis of genomic and transcriptional data*".
3. Dr.essa Gabriella Rustici (EBI) - "*ExpressionProfiler*".

5 - WP4: Proteomica

Il sequenziamento completo del genoma umano non ha fornito la spiegazione per tutti i dubbi sul suo funzionamento, anzi, ha contribuito al sorgere di nuove domande. Il genoma umano contiene circa 25 000 geni e probabilmente le informazioni per la sintesi di più di un milione di proteine. La proteomica è lo studio del proteoma, ossia il complesso di proteine espresse da un genoma. Le proteine possono essere analizzate usando strumenti per l'analisi di sequenza e tecnologie per la modellizzazione molecolare, che sono già presenti nel BioPortale. Altri metodi di studio hanno come scopo l'identificazione delle proteine in tessuti specifici, e utilizzano le tecnologie della spettrometria di massa.

Questo work package ha visto la continua interazione tra gli sviluppatori del progetto cluster e gruppi di ricerca interessati all'implementazione di database e di software specifici per gli studi di proteomica.

A - WP4.1 & 4.2: Basi di dati

E' stata installata la base di dati Proteios. In data 12 giugno 2007, per la sezione del database, si è tenuto il workshop "Database and Data management for Proteomics", in cui la base di dati Proteios è stata presentata agli utenti. Gli utenti interessati lavorano per le società: Proteotech e Porto Conte Ricerche.

E' stata installata una versione con maggiore stabilità, a cui possono accedere gli utenti al fine di testare il sistema.

Eventi

Il 12 giugno 2007, per la sezione database, si è tenuto il workshop "Database and Data management for Proteomics":

1. Dr. Henning Hermjakob (European Bioinformatics Institute) - "*Present your data with PRIDE*".

2. Dr. Fredrik Levander (Lund University) - "*Managing and analysing proteomic data using the Proteios database application*".

3. Dr.ssa Maria Filippa Addis (Porto Conte Ricerche) - "*The Proteomics Platform of Porto Conte Ricerche: facilities, activities, and applications*".

4. Dr.ssa Patricia Rodriguez-Tomé (CRS4 – Bioinformatica) - "*Proteomics Databases: managing data for large scale studies*".

Futuro del progetto PROTEIOS:

Il progetto PROTEIOS si è concluso il 15 Maggio 2008.

B - WP4.1 & 4.2: Analisi

L'analisi quantitativa delle immagini generate dall'acquisizione digitale dei gel bidimensionali, è importante per lo studio dell'espressione proteica in un dato sistema biologico.

Nonostante esistano in commercio molti software creati appositamente per l'analisi di immagine, nessuno di questi soddisfa pienamente le richieste dei ricercatori: richiede la spesa di molte risorse in termini di calcolo ed è poco automatizzato (il risultato dipende spesso dal giudizio soggettivo dell'operatore).

In collaborazione con la società Proteotech, che ha aderito al Progetto cluster, è stato sviluppato un programma per l'analisi di immagine dei gel 2D. Il programma scompone le immagini in componenti più piccole, attraverso l'analisi delle quali si cerca di catalogare le immagini in categorie definite.

Il programma serve per automatizzare alcune fasi del processo di lavoro:

- la definizione di famiglie di immagini;
- l'inserimento di una nuova immagine nella famiglia di appartenenza;
- l'inserimento dei parametri corretti di partenza al software di analisi di immagine.

Proteotech ha messo a disposizione i gel 2D e l'esperienza nel campo della proteomica dei propri ricercatori. Il programma di lavoro è stato diviso in due fasi:

a) clusterizzazione di tutte le immagini conosciute, usando l'annotazione manuale attualmente utilizzata da Proteotech, per semplificare l'analisi e la classificazione. Il processo di clusterizzazione può essere applicato alle nuove immagini aggiunte per incrementare l'accuratezza del sistema di classificazione.

b) confronto delle nuove immagini con i dati dei cluster inseriti in un data-base e assegnazione delle stesse a uno dei cluster. Come risultato della classificazione viene prodotto un file informativo, che viene allegato all'immagine. Il file informativo è in un formato compatibile al programma di analisi di immagine.

c) Per questo progetto è stato stipulato un contratto con un collaboratore.

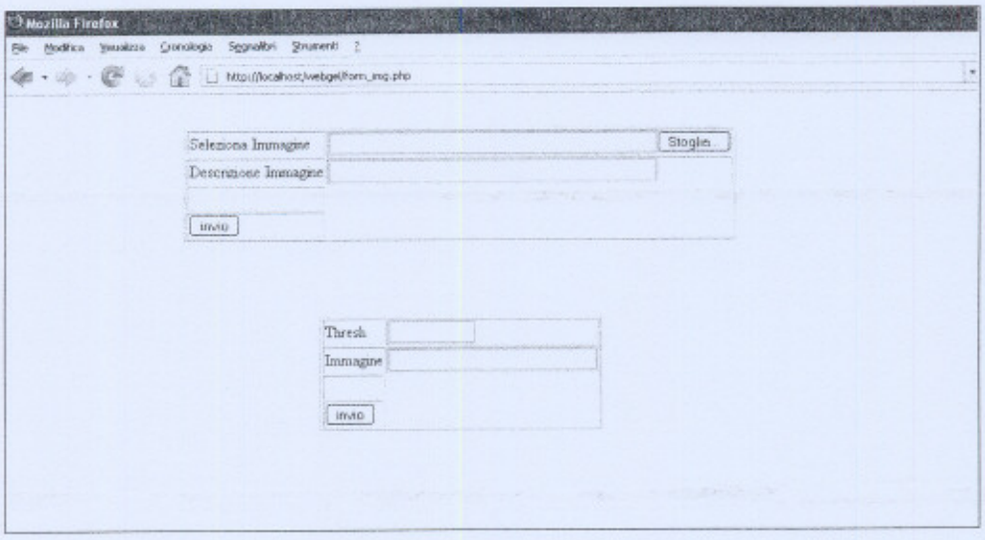

Fig. 15: Interfaccia di 2D gels - l'utente può selezionare l'immagine da analizzare.

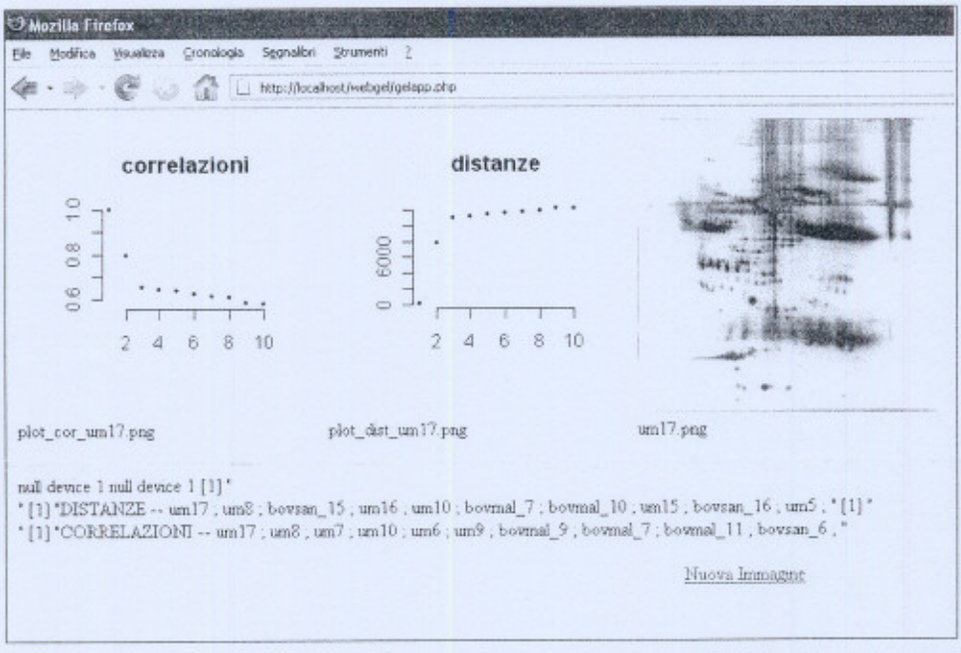

Fig. 16: Il risultato - l'immagine è stata analizzata, sono elencati i migliori risultati.

Eventi:

a) il 24-25 di settembre 2007, per la sezione "Analisi", si è tenuto il workshop "Proteomica" di cui si allegano informazioni sul programma e sui relatori;

24 Settembre - lezioni pratiche

1. Dr. Samuel Kerrien (EBI) - "*Database INTACT*".

2. Dr. Markus Muller (Swiss Federal Institute of technology Zurich) - "*Identificaton 1 : LC-MS data analysis, quantification*".

3. Dr. Jacques Colinge (The Austrian Academy of Sciences) - "*Identification 2* ".

25 Settembre - Seminari tenuti dai docenti del 24/09.

1. Dr. Samuel Kerrien (EBI) - "*The IntAct database*".

2. Dr. Markus Muller (Swiss Federal Institute of technology Zurich) - "*Towards quantitative understanding of proteomes : ongoing projects at the Institute of Molecular Systems Biology*".

3. Dr. Jacques Colinge (The Austrian Academy of Sciences) - "*From mass spectrometry data processing to molecular interactions annotation*".

Futuro del progetto 2DGels:

Il progetto 2DGels si è concluso il 15 Maggio 2008.

6 - WP5: DataMining

L'attività di ricerca in biologia molecolare produce giornalmente una grande quantità di informazioni, per cui diventa sempre più difficile per i ricercatori tenersi aggiornati sugli sviluppi più recenti relativi alla loro area di studio. L'informatica sta sviluppando la tecnologia per cercare velocemente e semplicemente grandi quantità di informazioni sia tra le pubblicazioni scientifiche, sia nelle banche dati o nei siti web (text e data mining).

Eventi:

Si sono svolte due giornate di workshop sull'argomento DataMining.

Workshop del 9 Febbraio 2007: "***Data Mining in Bioinformatics***".

1. Dr.ssa Eloisa Vargiu, Università di Cagliari - *Introduzione*

2. Dr.ssa Manuela Angioni, Dr. Roberto Demontis, Dr. Franco Tuveri, CRS4 - *Categorizzazione di risorse del Web con tecniche di NLP.*

3. Dr. Fabio Maggio, CRS4, Lab. Bioinformatica - *Support Vector Machines for the classification of biological datasets.*

4. Dr. Arek Kasprzyk, European Bioinformatics Institute - *Data mining with BioMart.*

Workshop del 16 Maggio 2007: "***Ontologies and Data Mining: tools for Knowledge in Biology***".

1. Dr. Christopher Brewster (University of Sheffield) - "Abraxas: Bridging the Gap between Text and Knowledge".

2. Dr. Kirill Degtyarenko (European Bioinformatics Institute) - "Open chemical dictionaries and ontologies for biosciences".

7 - BioPortale e TOOLS

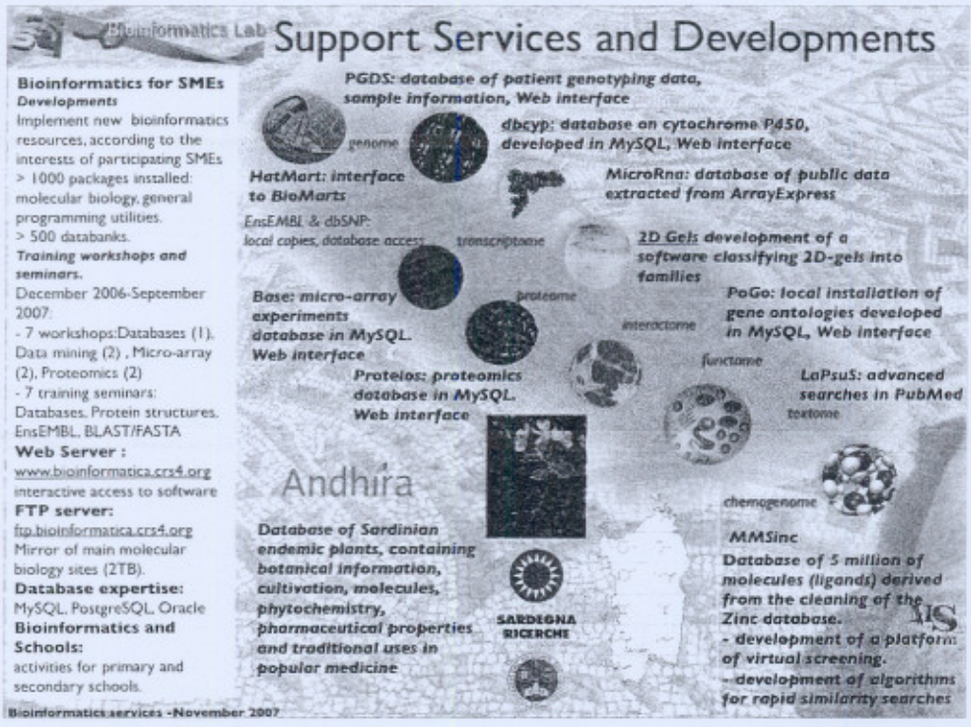

Fig. 17: Poster dei servizi del Laboratorio di Bioinformatica: PGDS, HatMart, dbCyp, 2DGels, LaPsuS, Andhira, MMSinc, Base e Proteios sono basi di dati o strumenti sviluppati/installati per il progetto cluster.

In Fig. 17 sono presentati i servizi che il gruppo Servizi&Sviluppo del Laboratorio di Bioinformatica del CRS4 mette a disposizione della comunità scientifica locale, precedenti e prescindenti dalle attività del progetto cluster, che sono descritti in questo documento. Gli strumenti, le banche e le basi di dati installati in locale, si trovano, fisicamente, nelle macchine situate nella computer room dell'edificio 1 del Parco tecnologico di Pula.

Le applicazioni sviluppate dal gruppo di bioinformatica, sono presentate in relazione a vari settori di ricerca: genomica, transcrittomica, proteomica, interattomica, genomica funzionale, chemogenomica, testomica (dall'inglese textomics, definizione all'inidrizzo: http://omics.org/index.php/What_is_omics).

A - Progetto BioPortale

Il BioPortale corrisponde al sito web del Laboratorio di Bioinformatica del CRS4 ed è stato utilizzato per l'accesso agli strumenti sviluppati per il Progetto cluster, ma anche come canale di comunicazione e informazione del Progetto, con una sezione dedicata e pagine relative ai diversi work package e agli eventi formativi.

Gli strumenti installati in locale e accessibili attraverso il Bio-Portale, permettono di interrogare le basi di dati pubbliche con query diverse rispetto a quelle disponibili nei siti ufficiali. Sono state costruite anche

interfacce web user-friendly per molti degli strumenti installati. Strumenti, banche e basi di dati sono tutti resi disponibili per l'uso di strumenti standard di biologia. Un collaboratore del Progetto Cluster ha effettuato lo sviluppo dei programmi sul BioPortale e ha preso parte all'amministrazione tecnica del sistema utilizzato (Zope e Plone).

Futuro del progetto BioPortale:

Il BioPortale continuerà a essere attivo come sito web del Laboratorio di Bioinformatica del CRS4, ma gli strumenti descritti in seguito non saranno più attivi.

B - Genomica: Progetto HatMart

HatMart è un'applicazione realizzata dal gruppo Servizi&Sviluppo del Laboratorio di Bioinformatica, scritta in java, che permette di interrogare database basati sul sistema BioMart. Il sistema BioMart è stato sviluppato dalla collaborazione del European Bioinformatics Institute (EBI) con il Cold Spring Harbor Laboratory (CSHL), allo scopo di realizzare uno strumento per effettuare ricerche biologiche su database pubblici. Esistono numerosi DB pubblici, costruiti con schemi diversi, quindi per interrogarli è necessario utilizzare un'applicazione specifica per ognuno. BioMart consente di costruire un estratto di ogni DB, che si chiama Data Mart. I Data Mart generati da BioMart hanno un formato standard, per cui sono tutti interrogabili da uno stesso strumento.

Nel Laboratorio del CRS4 è stato costruito un Mart Validator, per verificare che le configurazioni dei sistemi BioMart siano corrette nei Data Mart, e per segnalare eventuali errori. Questo strumento è utile per chi gestisce un sito centrale come quello dell'Istituto Europeo di Bioinformatica (EBI), o per chi vuole costruire un Data Mart; infatti questo strumento è stato sviluppato in collaborazione con l'EBI.

Nel Laboratorio di Bioinformatica è stato sviluppato anche HatMart, che, diversamente dalle interfacce web di BioMart, si può installare in locale come applicazione a sé stante ed ha un'interfaccia di utilizzo più intuitiva, nonché alcune funzionalità aggiuntive, come la possibilità di salvare, recuperare e manipolare manualmente le query.

Fig. 18: HatMart è connesso all'Istituto Europeo di Bioinformatica, l'utente ha selezionato il database di suo interesse.

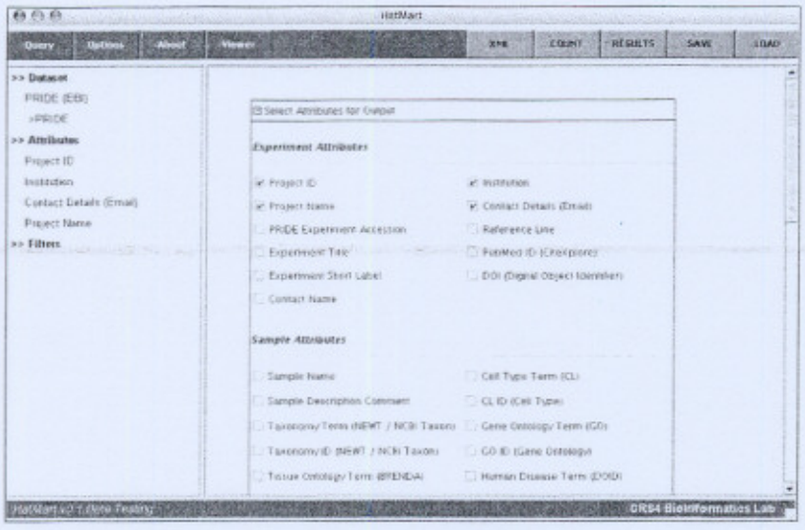

Fig. 19: Selezione degli attributi e dei filtri per la ricerca.

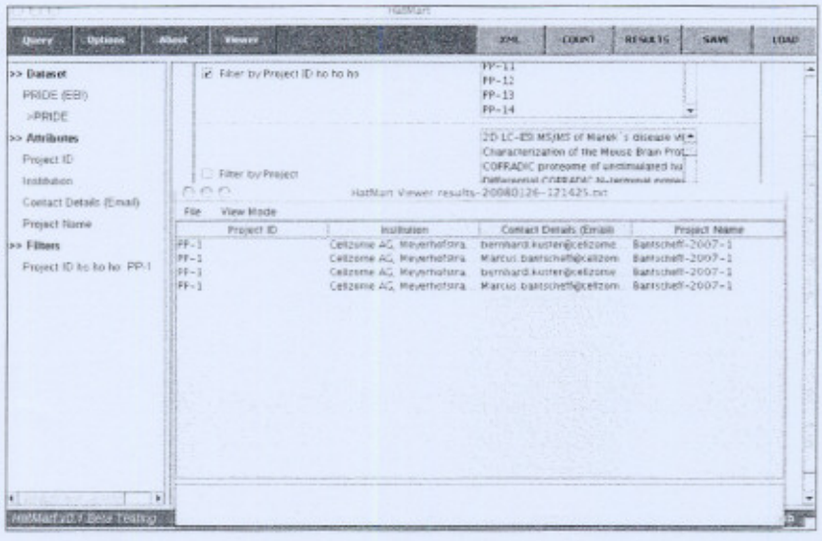

Fig. 20: Il risultato della ricerca.

Futuro del progetto HATMART:

Il progetto HATMART si è concluso il 15 Maggio 2008.

C - Testomica: Progetto LaPsuS

L'applicazione è stata progettata e sviluppata dal Laboratorio di Bioinformatica per effettuare ricerche complesse su PubMed in poco tempo.

Utilizzando LaPsuS l'utente può fornire al programma fino a tre liste di parole da ricercare nella letteratura scientifica indicizzata su PubMed, caricando tre file di testo attraverso un'interfaccia web. Il programma genera tutte le combinazioni possibili utilizzando gli operatori logici scelti dall'utente, che può anche settare i parametri di filtro, identici a quelli presenti sul sito di PubMed.

LaPsuS invia automaticamente la richiesta a PubMed e restituisce come risultato una lista di ID corrispondenti alle pubblicazioni scientifiche.

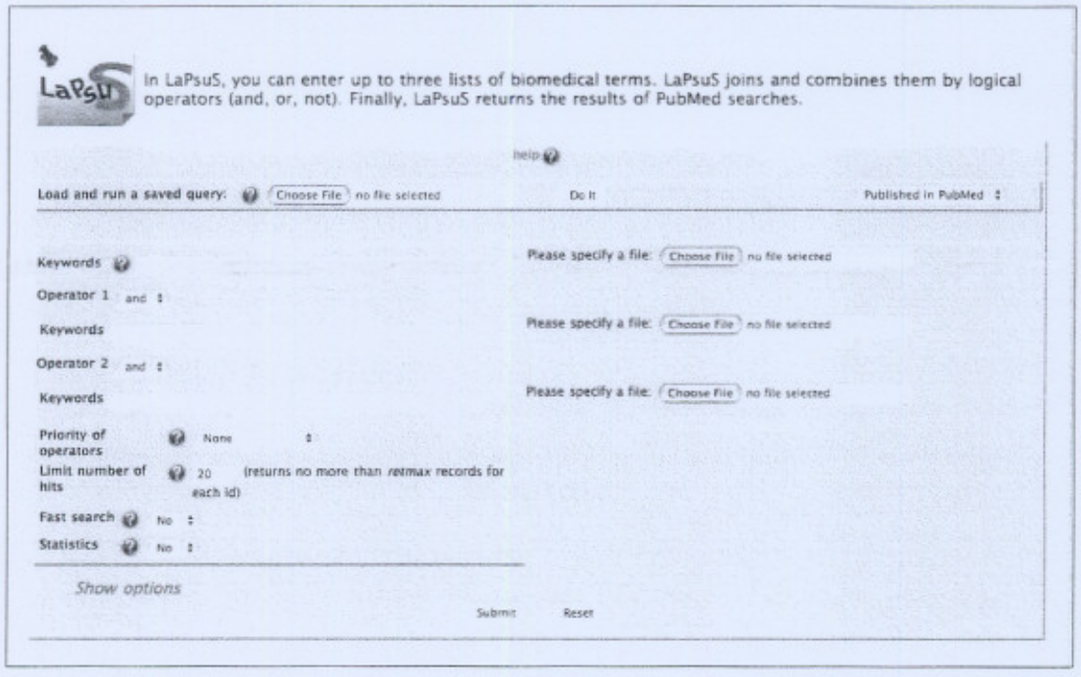

Fig. 21: L'interfaccia di utilizzo di LaPsuS.

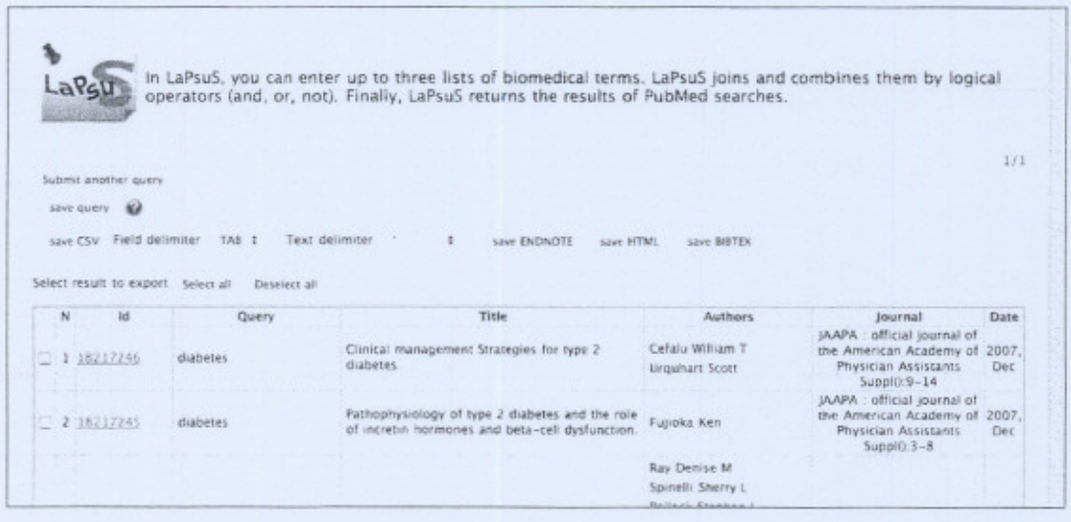

Fig. 22: Risultati della query.

Si possono salvare tutte le impostazioni di una ricerca e ripeterla a distanza di tempo: LaPsuS fornisce come risultato solo le variazioni rispetto alla ricerca precedente.

L'ultimo sviluppo di LaPsuS ha prodotto la possibilità di effettuare ricerche combinate tra una lista di nomi di geni e una lista di nomi di patologie, per misurare il grado di correlazione. L'applicazione permette anche all'utente di effettuare un'analisi statistica sui risultati ottenuti.

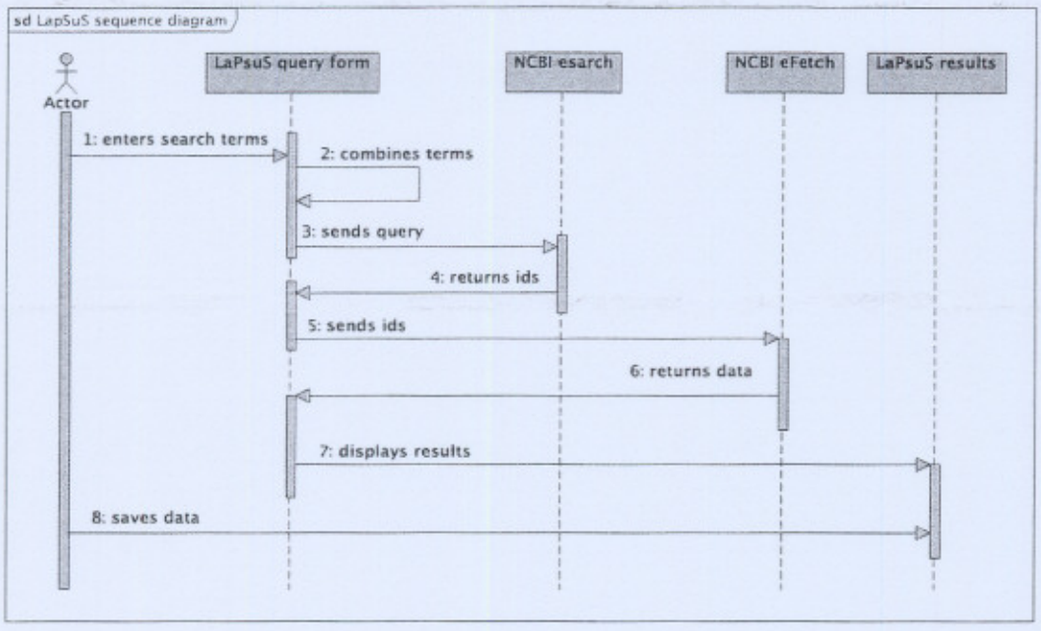

Fig. 23: Schema di funzionamento di LaPsuS.

Futuro del progetto LAPSUS:

Il progetto LAPSUS si è concluso il 15 Maggio 2008.

D - Altri strumenti: Progetto seuPedro

Pedro è un'applicazione sviluppata cinque anni fa in Inghilterra, che crea velocemente un'interfaccia grafica per l'inserimento di dati in un contenitore temporaneo, a partire da uno schema xml.

Per la costruzione di un database complesso si parte dalle richieste di un utente che sarà l'utilizzatore dello strumento. L'utente comunica le richieste al programmatore che costruirà il database. La fase di scambio di informazioni è molto delicata, e spesso di difficile realizzazione, ma fondamentale per lo sviluppo di uno strumento che soddisfi le esigenze dell'utente/utilizzatore.

L'utilizzo dell'applicazione Pedro è di grande aiuto in questa fase di progettazione. Pedro permette di creare molto velocemente un'interfaccia attraverso cui inserire i primi dati, originata da uno schema xml che contiene le prime informazioni fornite dall'utente. Dall'inserimento dei primi dati è possibile capire immediatamente se la comunicazione tra utente e programmatore è stata corretta, se il database richiesto è funzionale e correggere eventuali errori. Dopo questo primo passo realizzato con Pedro è possibile progettare e costruire il database vero e proprio. Pedro permette quindi di mettere in atto velocemente e semplicemente la fase iniziale di progettazione del database.

SeuPedro è un'evoluzione di Pedro, realizzata nel Laboratorio di Bioinformatica, a cui sono stati corretti alcuni errori, e che è stata adattata alle esigenze degli utenti del Progetto cluster.

Futuro del progetto SeuPedro:

Il progetto SeuPedro si è concluso il 15 Maggio 2008.

E - Progetto Chemioteca Sarda: base di dati per la chimica sarda

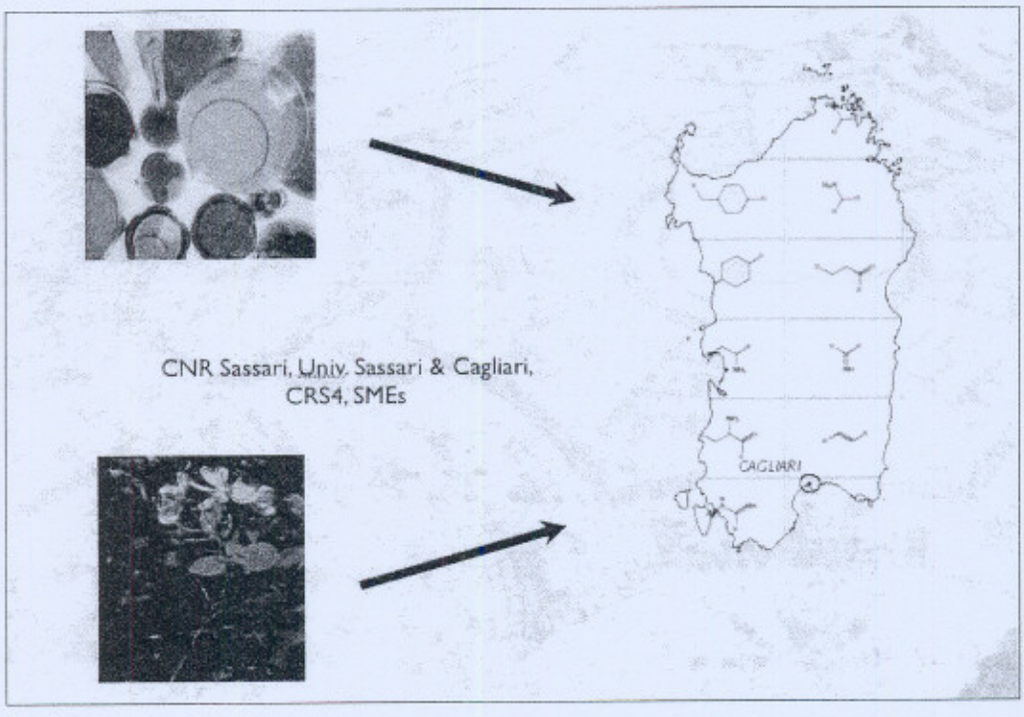

Fig. 23: Progetto per la realizzazione di una chemioteca Sarda.

Questa attività nasce principalmente dall'esigenza di valorizzare il patrimonio produttivo dei ricercatori che operano in Sardegna in termini di sintesi, estrazione, caratterizzazione ed individuazione di nuove molecole a potenziale attività biologica o molecole con proprietà utili in campo ambientale, medicale, farmaceutico, agronomico e della scienza dei materiali.

Questa sarebbe la prima *chemioteca* delle molecole organiche organizzata in Sardegna e raggrupperebbe tutti i centri di ricerca pubblici presenti nella Regione (UNISS, UNICA, CNR).

A differenza di altre *chemioteche*, che iniziano a nascere in varie regioni d'Italia, questa *chemioteca* raggrupperebbe sia molecole di sintesi che molecole o classi di molecole provenienti da estratti naturali.

In un momento di forte concorrenza da parte di Paesi, soprattutto europei, che dedicano un forte budget alla ricerca nazionale ed in mancanza di un corrispondente sostegno per la ricerca italiana, l'istituzione in Sardegna di una *chemioteca* on-line delle molecole organiche è un mezzo per contribuire a superare questa differenza e dare un contributo alla nascita e/o potenziamento di piccole imprese sul territorio interessate a questo tipo di molecole. Considerate le realtà presenti sul territorio, la *chemioteca* on-

line delle molecole organiche può essere un utile strumento per i centri di ricerca e le imprese che lavorano nel campo delle molecole biologicamente attive e nel campo dei materiali o dell'ambiente.

Futuro del progetto Chemioteca Sarda:

Il progetto Chemioteca Sarda si è concluso il 15 Maggio 2008.

8 - Divulgazione risultati

In allegato gli abstract delle conferenze sotto riportate.

A - Conferenza, 29 Maggio 2007

Il lavoro è stato presentato presso Porto Conte Ricerche, ad Alghero, al "42° International Symposium Analytical Technologies: Tools and Implementation Strategies in Animal Science", il 29 maggio 2007. Al poster descrittivo è stato assegnato il secondo premio della competizione per il miglior poster.

B - Conferenza, 6-9 Giugno 2007

Il lavoro su ANDHIRA è stato presentato alla conferenza "Future Trends in Phytochemistry - A Young Scientitsts Symposium", che si è svolto dal 6 al 9 Giugno 2007, a Gargnano, in Italia.

C - Convegno OTONGA, 26 Ottobre 2007

In collaborazione con l'associazione "Arca verde Otonga" e Sardegna Ricerche, è stato organizzato un convegno: "La Biodiversità come opportunità di Sviluppo e Cooperazione: Dal modello sardo alla foresta di Otonga", tenutosi il 26 di ottobre 2007 a Cagliari. Durante il convegno anche il progetto ANDHIRA è stato presentato al pubblico e alla stampa.

ULTIM'ORA

17:34 - mercoledì, 24 ottobre 2007

La biodiversità come opportunità di sviluppo e cooperazione: Dal modello sardo alla foresta di Otonga

Venerdì 26 giugno, alle 17:00, nella Sala Convegni del Banco di Sardegna, in viale Bonaria 33, a Cagliari, si terrà il convegno "La Biodiversità come opportunità di sviluppo e cooperazione. Dal modello sardo alla foresta di Otonga", organizzato dall'associazione L'Arca Verde Otonga, dal CRS4 e da Sardegna Ricerche. La biodiversità può essere vissuta come un'opportunità di sviluppo economico: la consapevolezza delle risorse naturali e l'acquisizione di metodologie di valorizzazione possono offrire nuove opportunità di crescita, anche nelle aree povere del pianeta. L'evento ha l'obiettivo di contribuire a una maggiore conoscenza di tali opportunità. In particolare, il professor Giovanni Onore dell'Università Cattolica dell'Ecuador, ecologo di fama internazionale e presidente della "Fundación Otonga", illustrerà la sua lunga esperienza nella tutela della foresta primaria e nella valorizzazione della biodiversità a favore delle popolazioni. Durante il convegno, inoltre, sarà presentato "Andhira", il database degli endemismi sardi realizzato dal Laboratorio di Bioinformatica del CRS4. In programma, ancora, relazioni e interventi di: Antonello Arru, Fondazione Banco di Sardegna; Giuliano Murgia, Sardegna Ricerche; Mauro Ballero, Dipartimento di Scienze Botaniche dell'Università di

Fig. 24: Il convegno annunciato su un sito di informazione locale.

D - Evento IRC a "MEDICA 2007", 14-16 Novembre 2007

I Centri di Collegamento per l'Innovazione (IRC) della Commissione Europea hanno il compito di aiutare le imprese e le organizzazioni di ricerca a trasferire le nuove tecnologie in Europa, fornendo assistenza specializzata per le iniziative di progettazione, di trasferimento tecnologico e know-how e di partenariato internazionale.

Il gruppo IRC North Rhine Westphalia ha organizzato incontri tra enti di ricerca e piccole industrie durante l'evento MEDICA 2007, nella città di Dusseldorf. In questa occasione la Dr.ssa Patricia Rodriguez-Tomé ha presentato il progetto cluster Bioinformatica (documento allegato) a due enti:

a) East Midlands Development Agency, Gran Bretagna

b) Università di Szeged, Ungaria

L'incontro con due ricercatori della Università di Szeged si è concluso con una proposta di collaborazione per lo sviluppo di strumenti di data mining. I due ricercatori hanno progettato un soggiorno in Sardegna per presentare il loro lavoro. In occasione dell'incontro i ricercatori dovranno anche presentare i risultati di un'analisi di dati di MicroArray.

E - SardiniaChem 2008

Il 30 Maggio 2008 si terrà l'edizione 2008 di SardiniaChem, a Sassari. In tale occasione, durante un talk intitolato "Andhira ed MMSInc, verso una piattaforma web per lo screening virtuale di piccole molecole", saranno presentati i progetti sviluppati per il Progetto Cluster.

E - CheminfoS3

Dal 22 al 25 Giugno 2008 si terrà in Obernai (Francia) la "Strasbourg Summer School on Chemoinformatics" dove verrà presentato il poster "MMsINC: a new public large-scale chemoinformatics database system".

F - Biotechno2008

Dal 29 Giugno al 5 Luglio 2008, si terrà a Bucarest (Romania) la "International Conference on Biocomputation, Bioinformatics, and Biomedical Technologies". Alla conferenza sarà presentato il talk intitolato "MMsINC(R): a new public large-scale chemoinformatics database system".

G - Natural Products

Dal 3 all'8 Agosto 2008, ad Atene (Grecia) si terrà la conferenza "Natural Products" dove sarà presentato il progetto Andhira durante un talk intitolato "Andhira, the database of Sardinian endemic plants and their molecules".

9 - PEOPLE

A - Contratti coperti dal budget del progetto cluster:

Nome	Attività
Giuliana Brunetti	workshop, collaborazioni, editing di testi
Gianfranco Frau	sviluppo Andhira, MMsINC, HatMart, seuPedro
Osvaldo Marullo (part time)	sviluppo 2Dgels
Ricardo Medda (fino a dic 2007)	amministrazione zope (web), sviluppo BioPortale, LaP-suS
Giuseppe Mocci (part time)	inserimento dati per Andhira
Piergiorgio Palla	sviluppo Andhira, MMsINC, HatMart, seuPedro
Vera Uras	inserimento dati per Andhira

B - Contratti coperti dal budget del progetto bioinformatica, hanno partecipato anche al progetto Cluster:

Nome	Attività
Patricia Rodriguez-Tomé	management, Andhira, Proteios, workshop
Sergio Contrino (fino a ottobre 2007)	sviluppo di PGDS, esperto di MicroArray
Matteo Floris (dal gruppo Ricerca)	Andhira, MMSinc, LaPSuS
Joel Masciocchi	dbCyp, Andhira, MMSinc, BASE
Ricardo Medda (da Gennaio 2008)	amministrazione Zope (web), sviluppo BioPortale, LaP-suS
Michele Muggiri	IT support, Proteios
Iliena Zara (dal 19 novembre 2007)	sviluppo BASE e MicroArray tools
Luca Pireddu (da Gennaio 2008)	sviluppo di PGDS

10 - HARDWARE E SOFTWARE

A - Hardware acquistati

Nome	Specificazione technice	utilizzo
bioweb	Processori 2 AMD Opteron Dual Core 2.2 GHz RAM 4GB. 2 HardDisk SATAII 250GB	web server
scarteddu (HP Proliant dl585)	Processori 4 AMD Opteron DualCore 2.2GHz. RAM 8GB. HardDisk 4 x 146GB Controller SmartArray P800. HardDisk per Array 8 x 300GB. OS: RedHat Enterprise	produzione basi di dati
scivedda (HP Proliant dl585)	Processori 4 AMD Opteron DualCore 2.2GHz RAM 16GB. HardDisk 4 x 146GB Controller SmartArray P800 HardDisk per Array 8 x 300GB OS: RedHat Enterprise	sviluppo basi di dati
3 mac mini	Processori Intel Dual Core 2GHz Ram 1GB. HardDisk 120 GB	desktop
3 macbook	Processore Intel Dual Core 2GHz Ram 2GB. HardDisk 80GB	Laptop
laptop	HP DV6580. Processore Intel Dual Core 2GHz. Ram 2GB. HD 200 GB	Laptop per inserimento dati per Andhira

B - Software acquistati

Nome	utilizzo
iWork	equivalente Office per Mac
iTask	programma per project management
Visual Paradigm	programma per il design di database, software
MacOSX10.5	sistema operativo per Mac