Automated data retrieval of alternative splicing sites and transcript isoformes in eukaryotic organisms

Castrignanò T[1], D'Antonio M[3], D'Onorio De Meo P[1], Anselmo A[2], D'Erchia A[3], Licciulli F[4], Mangiulli M[2], Mignone F[2], Pavesi G[2], Picardi E[3], Rizzi R[5], Bonizzoni P[5], and Pesole G[3-4]

[1] Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca, CASPUR, Rome, Italy,
[2] University of Milan, Dipartimento di Scienze Biomolecolari e Biotecnologie, via Celoria 26, Milan 20133, Italy,
[3] University of Bari, Dipartimento di Biochimica e Biologia Molecolare, via Orabona, 4, Bari 70126, Italy,
[4] Istituto Tecnologie Biomediche del Consiglio Nazionale delle Ricerche, via Amendola 122/D, Bari 70126, Italy,
[5] DISCo, University of Milan Bicocca, via Bicocca degli Arcimboldi, 8, Milan, 20135, Italy.

Motivation

New large scale techniques in biology are producing a rapidly growing amount of public available data. The increasing size and complexity of biological databases has motivated the developers to federate rather than duplicate them. In order to share data among federated databases, a protocol for the exchange mechanism, the Distributed Annotation System (DAS), has been widely adopted in the bioinformatics community. It allows a single machine to gather up biological information from multiple distant web sites, collate the information, and display it to the user in a single view. On the other hand, web services are the distributed computing technology that offers powerful capabilities for scalable interoperation of heterogeneous software across a wide variety of networked platforms. Database information retrieval and analysis services have to be linked, so that search results from one database can be used as starting point of a search in another to perform further analysis. ASPicDB is a database designed to provide access to reliable annotations of the alternative splicing pattern of eukaryotic genes and the functional annotation of predicted isoforms in eukaryotic organism.

Methods

ASPicDB is a relational database developed in MySQL 5.0 designed to manage all data produced by the ASPic program applied to the collection of eucaryotic genes. The data has been obtained performing the multiple alignment, implemented in the ASPic algorithm, of gene-related transcripts (typically a Unigene cluster) to the genomic sequences. A DAS (Distributed Annotation System) server has been added to ASPicDB allowing to share data sets with other database; a SOAP-based web service has been develop to provide programmatic access to ASPicDB. The PHP scripts, running under Apache, allow the user to browse alternative splicing and transcript isoform data or exchange data sets through both DAS server and ASPic web service.

Results

Here we present the following two distributed technologies to exchange data from ASPicDB:
1) a DAS server of ASPicDB that allows integration of the database into the network of biological databases via the Distributed Annotation System (DAS) and
2) a ASPic SOAP-based web service which provides several methods for data retrieval allowing the user to develop more complex applications for data exchange.
Both services provide automated access to up-to-date detected splice sites and modelled full-length transcripts of eukaryotic genomes (Human, Mouse, Rat, Cow, Chicken, Dog, Horse). Users are free to directly use a simple web interface (www.caspur.it/ASPicDB) or export huge gene-centred data sets via the DAS server (http://www.caspur.it /ASPicDB/das) or through the use of ASPic web service (http://www.caspur.it/ASPicDB/ASPicWS.html).

Contact : tiziana.castrignano@caspur.it