2

Data Integration in Bioinformatics: Current Efforts and Challenges

Zhang Zhang[1], Vladimir B. Bajic[1], Jun Yu[2],
Kei-Hoi Cheung[3,4,5,6] and Jeffrey P. Townsend[6,7]
[1]*Computational Bioscience Research Center (CBRC),*
King Abdullah University of Science and Technology (KAUST), Thuwal
[2]*CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics,*
Chinese Academy of Sciences, Beijing
[3]*Center for Medical Informatics,* [4]*Department of Computer Science,*
[5]*Department of Genetics,* [6]*Program in Computational Biology and Bioinformatics,*
[7]*Department of Ecology and Evolutionary Biology, Yale University,*
New Haven, Connecticut
[1]*Kingdom of Saudi Arabia*
[2]*China*
[3,4,5,6,7]*United States of America*

1. Introduction

With the rapid advancements in next-generation sequencing (NGS) technologies and the consequently fast-growing volume of biological data, a diversity of data sources (databases and web servers) have been created to facilitate data management, accessibility, and analysis. A prerequisite of bioinformatics research has been the ability to find, maneuver and access data deposited in various data sources. For a given bioinformatic task, researchers often need to be skillful in interrogating these data sources, and in the use of extracted information for further data analysis/information search. For example, one must obtain data from one data source, reformat the data and submit to another data source for analysis, parse the analyzed result, and then combine the result with data obtained from the third data source, etc. Undisputedly, data integration becomes tedious and time-consuming, especially regarding the import and export of enormous files of modern NGS and other data. Thus, integration of data from distributed, heterogeneous and voluminous data sources turns out to be a significant obstacle to fully exploit the wealth of big biological data (Davidson, et al., 1995; Stein, 2002). The importance of the integration component of research stemming from studies based on high-throughput technologies (such as NGS), is twofold: (1) due to the great level of automation of the actual experimental procedures, the effort of obtaining the experimental data takes only about 20% or less of the overall research effort in an NGS project; approximately four fifths of the effort goes to the integration and analysis of a collection of the experimental data (Mardis, 2010); (2) the answers to the most important, complex biological questions today are rarely provided directly through the experimental

results; to bring potential answers to the surface, downstream bioinformatics analysis often involves the integration of diverse data from multiple data sources.

The objective of data integration in bioinformatics is to establish automated and efficient ways to integrate large, heterogeneous biological datasets from multiple sources. However, this objective is challenged by data sources that are geographically distributed and heterogeneous in terms of their functions, structures, data access methods and dissemination formats. According to the 2010 update on the Bioinformatics Links Directory (Brazas, et al., 2010), there are almost 1500 unique publicly-available data sources. Based on their functions, data sources can be classified into diverse categories: (1) sequence databases, e.g., GenBank (Benson, et al., 2006), RefSeq (Pruitt, et al., 2009), CMR (Comprehensive Microbial Resource) (Davidsen, et al., 2010); (2) functional genomics databases, e.g., ArrayExpress (Parkinson, et al., 2011), FFGED (Filamentous Fungal Gene Expression Database) (Zhang and Townsend, 2010), GEO (Gene Expression Omnibus) (Barrett, et al., 2011); (3) protein-protein interaction databases, e.g., BIND (Biomolecular Interaction Network Database) (Bader, et al., 2003), DIP (Database of Interacting Proteins) (Salwinski, et al., 2004), IntAct (Aranda, et al., 2010), MINT (Molecular Interactions Database) (Ceol, et al., 2010); (4) pathway databases, e.g., KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa, et al., 2010); (5) structure databases, e.g., CATH (Greene, et al., 2007), PDB (Protein Data Bank) (Rose, et al., 2011); (6) annotation databases, e.g., GO (Gene Ontology) (Ashburner, et al., 2000), NCBI Taxonomy (Sayers, et al., 2011). Moreover, data sources differ in data accessibility and dissemination. That is, different levels of provision are made by the data source managers for human-reading, computer-reading, or both. Certainly, data sources can also be classified by species of interest, such as, filamentous fungi (Zhang and Townsend, 2010), fly (Gilbert, 2007), mouse (Blake, et al., 2011), and yeast (Engel, et al., 2010).

Despite the challenges, the promise of data integration is high: heterogeneous data sources provide biological data encompassing a wide range of research fields. Therefore, data integration has the potential to facilitate a better and more comprehensive scope of inference for biological studies. Although efforts have been devoted to biological data integration over the past two decades, it remains challenging and laborious. Here we review current efforts and illustrate several approaches used for data integration. With a specific consideration of the exponentially-growing NGS data, we also describe challenges in this context and discuss potential trends.

2. Current efforts of data integration in bioinformatics

Several major approaches have been proposed for data integration, which can be roughly classified into five groups (Goble and Stevens, 2008; Zhang, et al., 2009): data warehousing, federated databasing, service-oriented integration, semantic integration and wiki-based integration. Across all of these groups, to a significant extent, an increasingly important component of data integration is the community effort in developing a variety of biomedical ontologies (see Section 3.2), to deal in a more specific manner with the technicality and globality of descriptors and identifiers of information that has to be shared and integrated across various resources (Antezana, et al., 2009; Maojo, et al., 2011; Rubin, et al., 2008).

2.1 Data warehousing

The data warehouse approach offers a "one-stop shop" solution to ease access and management of a large variety of biological data from different data sources. Data warehouses focus on data translation, fetching all accessible data from many disparate data

sources, transforming the data and importing it into the data warehouse. Representative examples of data warehousing include:

- Atlas (Shah, et al., 2005) is a biological data warehouse that locally stores and integrates biological sequences, molecular interactions, homology information, functional annotations of genes, and biological ontologies. It includes data from BIND, DIP, Entrez Gene (Maglott, et al., 2011), GO, GenBank, HomoloGene, HPRD (Human Protein Reference Database) (Keshava Prasad, et al., 2009), IntAct, LocusLink (Pruitt and Maglott, 2001), MINT, RefSeq, OMIM (Online Mendelian Inheritance in Man) (Amberger, et al., 2009), Taxonomy, and UniProt (The UniProt Consortium, 2011).
- BioWarehouse (Lee, et al., 2006) is an open source toolkit for constructing data warehouses. It incorporates data from BioCyc (Karp, et al., 2005), CMR, ENZYME (Bairoch, 2000), GenBank, GO, KEGG, Taxonomy, and UniProt and integrates its component databases into a common representational framework within a single database management system.
- BIOZON (Birkland and Yona, 2006) is a unified biological resource on DNA sequences, proteins, complexes and cellular pathways. It relies on an extensive database schema that integrates information at the macro-molecular level as well as at the cellular level from a variety of data sources, including BIND, DIP, Genbank, InterPro (Hunter, et al., 2009), KEGG, PDB, RefSeq, Swiss-Prot (Bairoch, et al., 2004), UniGene (Sayers, et al., 2011), and UniProt.
- COLUMBA (Trissl, et al., 2005) is an integrated database of information on proteins, structures and annotations. It integrates twelve different databases, including CATH, ENZYME, GO, KEGG, PDB, SCOP (Andreeva, et al., 2008), and Swiss-Prot.
- VINEdb (Hariharaputran, et al., 2007) is a data warehouse for integration and interactive exploration of life science data. It manages diverse data from GO, IntAct, KEGG, OMIM, and UniProt and emphasizes the visualization of the integrated data in a comprehensible manner.

The data warehouse approach has several advantages. (1) The user does not need to access many web sites for multiple data sources. Data warehouses provide one single access point to conveniently manipulate a large variety of data. (2) All queries requested by users are executed within the data warehouse (rather than on distributed data sources) and therefore, data warehousing eliminates network bottlenecks and obtains high performance with fast response. (3) Due to data storage at a single managed point, data warehousing obtains benefits in data control, yielding easy customization to meet users' needs.

Despite its advantages, the data warehouse approach has a major problem; it requires continuous and often human-guided updates to keep the data comprehensive of the evolution of data sources, resulting in high costs for maintenance. In general, there are two kinds of changes. (1) Changes in data volume or revisions of data. Whenever extant data is revised or the volume of data in any data source is changed, the data warehouse must monitor for such remote changes and update the warehouse to store the new data. (2) Changes in data structure, including adding new data types and tables, changing database tables and their relationships, and changing output formats. Many biological data sources change their data structures roughly twice a year (Stein, 2003). Whenever the data sources change their data structures, consequent data translation into the data warehouse must be updated in response. Usually, modification of data translation is labor-intensive and expensive.

2.2 Federated databasing

Unlike data warehousing (with its focus on data translation), federated databasing focuses on query translation. The federated databasing approach executes all queries on the distributed sources by translating a query against the federated database into a query against many data sources. The federated database fetches the data from disparate data sources and then displays the fetched data for its user base. Representative examples for federated databasing include:

- BioMart (Haider, et al., 2009) is a query-oriented data integration system developed jointly by the Ontario Institute for Cancer Research (OICR) and the European Bioinformatics Institute (EBI). It provides a user-friendly and unified way to retrieve data from one or multiple data sources located at diverse geographical locations, including Ensembl (Flicek, et al., 2011), HGNC, Uniprot, Reactome (Croft, et al., 2011), Wormbase, and PRIDE (Jones, et al., 2008).
- DiscoveryLink (Haas, et al., 2001) developed by IBM is a system for integrated access to life sciences data from heterogeneous data sources, including GenBank, MedLine and Swiss-Prot. It features query optimization and cross-source queries that access relational databases and retrieve the data from diverse data sources.
- K2/Kleisli (Chung and Wong, 1999; Davidson, et al., 2001) is a federated database system, integrating data from EcoCyc (Keseler, et al., 2011), GenBank, GSDB (Harger, et al., 1998), dbEST (Boguski, et al., 1993), GDB (Letovsky, et al., 1998), KEGG and SRS-indexed databases. Kleisli uses a high-level query language called Collection Programming Language (CPL) as its query language, which was developed specifically for parsing, optimizing and executing queries. K2 is the newer version of Kleisli and replaces CPL by a powerful and easy-to-use SQL-like query language, Object Query Language (OQL).
- MRS (Hekkelman and Vriend, 2005) allows for very rapid queries in a large number of flat-file data banks, including EMBL, UniProt, OMIM, dbEST, PDB, KEGG. It combines a fast and reliable backend with a very user-friendly implementation of all the commonly used information retrieval facilities.
- QIS (Query Integrator System) is based on a set of distributed network-based servers, data source servers, integration servers, and ontology servers and relies on a combination of SQL-like syntax and XML (eXtensible Markup Language; a widely used standard for data description and exchange), to formulate a query (Marenco, et al., 2004). It stores diverse queries for data integration from continuously changing heterogeneous data sources in the biosciences, including CellPropDB (Crasto and Shepherd, 2007), Brain Architecture Management System (Bota and Swanson, 2010), Yale Microarray Database (Cheung, et al., 2002), a local Gene Annotation Database and GO.
- SRS (Sequence Retrieval System) is an index-based integration system and combines some features of data warehousing and federated databasing (Zdobnov, et al., 2002). SRS uses a keyword-based indexing language ICARUS to describe each integrated data source and locally creates a full-text index over all data sources. Meanwhile, it allows a single query to execute on multiple data sources based on local indexed entries. SRS contains a number of biological databases (see details in http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+databanks+-noSession).
- TAMBIS (Transparent Access to Multiple Bioinformatics Information Sources) is an integration application to perform bioinformatics tasks over multiple data sources by

using an ontology of biological concepts (Stevens, et al., 2000). The prototype version of TAMBIS contains five data sources, viz., BLAST, CATH, ENZYME, PROSITE (Sigrist, et al., 2010) , and Swiss-Prot.

Queries in federated databases are executed within remote data sources and results displayed in federated databases are extracted remotely from the data sources. Due to this capability, federated databasing has two major advantages. (1) Federated databases can be regarded as an on-demand approach to provide immediate access to up-to-date data deposited in multiple data sources. (2) Compared with data warehousing, federated databasing does not replicate data in data sources; therefore, it presents relatively inexpensive costs for storage and curation. However, federated databasing still has to update its query translation to keep pace with data access methods at diverse remote data sources. In addition, since data is retrieved from remote data sources, federated databasing depends heavily on network connectivity and query complexity, which may lead to low efficiency and speed in data retrieval.

2.3 Service-oriented integration

Data warehousing and federated databasing both focus on centralizing data access, through data translation and query translation, respectively. They confront some similar problems stemming from data storage and curation, frequent updates, and high costs for data exchange and/or maintenance. In part to evade these issues, a decentralized approach has also been advanced, in which individual data sources agree to open their data via Web Services (WS). WS are designed for communication between computers over the Web and described by the Web Services Description Language (WSDL). There are several different protocols for WS, e.g., SOAP (Simple Object Access Protocol; a protocol for exchanging XML-based messages over computer networks), REST (REpresentational State Transfer; a simple protocol implemented using HTTP methods). WS support computer-to-computer interaction through Web Application Programming Interface (Web API) (Shi, 2007) and can perform a database query or computation. In the context of data integration, data can be programmatically accessed via WS and data sources serve as service providers. Therefore, this approach can be seen as a service-oriented approach. The service-oriented approach enables data integration from multiple heterogeneous data sources through computer interoperability. Several representative examples for service-oriented integration include:

- BioMOBY (Kawas, et al., 2006; Wilkinson and Links, 2002; Wilkinson, et al., 2008) is an open source ontology-based integration system for accessing distributed and heterogeneous data sources via WS. It implements a WS registry and uses standard ontology terms to annotate WS. BioMOBY adopts SOAP for data exchange and allows interoperability among different data sources to achieve automated data integration and sharing (Neerincx and Leunissen, 2005).
- DAS (Distributed Annotation System) is a client-server system to provide access to complete distributed genome annotations using SOAP-based WS (Dowell, et al., 2001; Katayama, et al., 2010; Olason, 2005). It allows a single machine to collect all annotations from multiple distributed data sources and display them to the user in a single view. DAS is widely used in the genome annotation community (http://en.wikipedia.org/wiki/Distributed_Annotation_System) and adopted by several systems, including Ensembl, WormBase, and the Berkeley Drosophila Genome Project (Jenkinson, et al., 2008; Messina and Sonnhammer, 2009; Olason, 2005).

- Taverna (Oinn, et al., 2004), a part of MyGrid (Stevens, et al., 2003), is a graphical workflow workbench application, aiming to integrate the growing number of molecular biology tools and databases (Hull, et al., 2006). Workflows in Taverna, written by a custom XML-based language called Simple Conceptual Unified Flow Language (SCUFL), can automatically record all data involved, provenance metadata, and results, facilitating complex data processing in a dynamic distributed environment.

The service-oriented approach features data integration through computer-to-computer communication via Web API and up-to-date data retrieval from diverse data sources. Thus, it befits well with the dynamic nature of bioinformatics. However, it remains challenging, primarily because its success in heterogeneous data integration requires that many data sources should become service providers by opening their data via WS and by standardizing data identities and nomenclature to ease data exchange and analysis. In addition, a unified WS registry is also necessitated, not only to establish standards for WS registration, but also to formulate standards for service-oriented workflows or pipelines (Zhang, et al., 2009).

2.4 Semantic integration

Most web pages in biological data sources are designed for human reading (e.g., HTML). The Semantic Web (Dibernardo, et al., 2008; Good and Wilkinson, 2006; Hendler, 2003; Lord, et al., 2004) aims to describe data in a way that computers can understand and to build an interconnected network that computers can easily and unambiguously process. According to the statement of definition from the World Wide Web Consortium (W3C), the purpose of the Semantic Web is to create a universal medium for the exchange of data using several standards, including Resource Description Framework (RDF; http://www.w3.org/RDF), RDF schema (RDFS—RDF Vocabulary Description Language; http://www.w3.org/TR/rdf-schema), Web Ontology Language (OWL; http://www.w3.org/owl), and standard Web query language SPARQL (http://www.w3.org/TR/rdf-sparql-query) for RDF. RDF provides standard formats (e.g, XML format) for data interchange and describes data as a simple statement, containing a set of triples: a *subject*, a *predicate* and an *object*. Any two statements can be linked by an identical *subject* or *object*. OWL builds on RDF and Uniform Resource Identifier (URI) and describes data structure and meaning based on ontology, which enables automated data reasoning and inferences by computers. The Semantic Web provides an machine-readable way for data representation and interoperability (Antezana, et al., 2009). Several studies have applied the Semantic Web technologies in data integration and representative examples of semantic integration are described below.

- Bio2RDF (Belleau, et al., 2008) is a mashup system that creates an integrated space of RDF documents linked together with normalized URIs. Bio2RDF applies the Semantic Web technologies to multiple data sources, such as Entrez Gene, HGNC, KEGG, MGI, OMIM PDB, PubMed and UniProt, and converts data into RDF format based on RDFizer (a set of tools for converting various data formats into RDF; http://simile.mit.edu/wiki/RDFizers), Sesame (an open source framework for storage, inference and querying of RDF data; http://www.openrdf.org) and OWL ontology. In Bio2RDF, each RDF document is expressed as a URI. When a query is requested to Bio2RDF for a given URI, for example, http://bio2rdf.org/go:0004396, the URI identifies RDF triples containing the GO term of Hexokinase (GO:0004396). Bio2RDF supports query via SPARQL.

- HCLS (The Health Care and Life Sciences Interest Group; http://www.w3.org/2001/sw/hcls/), established by W3C, aims to explore the potential benefits of the Semantic Web in the health care and life sciences domains (Cheung, et al., 2008) and advocates the application of the Semantic Web for advancing translational research (Ruttenberg, et al., 2007). The HCLS Knowledge Base (HCLS-KB; http://www.w3.org/TR/hcls-kb) is a Semantic Web system that imports data from many data sources in multiple domains of life sciences, including not only general sources, e.g., Entrez Gene, GO, HomoloGene, but also domain-specific sources, e.g., Allen Brain Atlas (an interactive, genome-wide image database of gene expression in the mouse brain; http://www.brain-map.org) (Lein, et al., 2007), SenseLab (a collection of neuroscience data; http://neuroweb.med.yale.edu/senselab) (Crasto, et al., 2007) and SWAN (Semantic Web Applications in Neuromedicine; aiming to organize and annotate scientific knowledge about Alzheimer disease and other neurodegenerative disorders) (Ciccarese, et al., 2008; Clark and Kinoshita, 2007; Kinoshita and Clark, 2007).
- YeastHub (Cheung, et al., 2005) is an integrated database in RDF format for the yeast community. It creates a RDF repository for RDF storage and provides a utility to convert tabular format into RDF format. YeastHub integrates different types of yeast data provided by different data sources (SGD, YGDP, MIPS, BIND, GO and TRIPLES) and supports RDF-based queries to retrieve and query the data.

Application of the Semantic Web technologies to biological data integration is a significant advancement for bioinformatics, enabling automated data processing and reasoning. The semantic integration uses ontologies for data description and thus represents ontology-based integration (Noy, 2004). However, the Semantic Web continues to evolve and its application in biological data integration has several limitations. The semantic integration locally stores a large collection of RDF documents, by copying data from multiple data sources and converting data into RDF format. From this view, the semantic integration can be regarded as a special data warehouse with data in RDF format. As a consequence, it inherits the pros and cons of data warehousing and is vulnerable to updates in data sources. To keep the RDF documents up-to-date, it requires tedious and periodical data retrieval and RDF conversion. In addition, once any data source changes data structure, the RDF conversion scripts must be updated consequently.

Currently, there is an ongoing project, the World Wide Web Consortium's SWEO (Semantic Web Education and Outreach) Linking Open Data Project (Bizer, 2009; Zhao, et al., 2009) that uses the Semantic Web technologies to connect related distributed data across the Web. Technically, linked data rely on RDF to create typed links between data from different data sources. Linked data is machine-readable, explicitly defined, and inter-linked to other data, promising to facilitate data integration, exposure, sharing, and connecting.

2.5 Wiki-based integration

A weakness common to all the above approaches is that the quantity of users' participations in the process is inadequate. With the increasing volume of biological data, data integration inevitably will require a large number of users' participations. A successful example that harnesses collective intelligence for data aggregation and knowledge collection is Wikipedia, an online encyclopedia (http://www.wikipedia.org) that allows any user to create and edit content. Wikipedia features collaborative integration, continuous and frequent update, up-to-date content, huge content coverage and low cost for maintenance

(McLean, et al., 2007). Although there are fears of inconsistency and inaccuracy since users can freely and anonymously change any content and/or add new content in the wiki (Arita, 2009; Bidartondo, 2008), it is testified that Wikipedia outperforms the traditional Encyclopedia in accuracy (Giles, 2005).

In consideration of the success of Wikipedia, a wiki-based approach has been on the horizon to store, manage and organize biological data (Giles, 2007; Salzberg, 2007; Waldrop, 2008; Yager, 2006). The wiki-based integration makes full use of collective intelligence and efforts for biological data integration. Representative examples include: WikiGenes (a wiki system that combines gene annotation with explicit authorship; Hoffmann, 2008), WikiProteins (a wiki-based system for protein annotation; Mons, et al., 2008), BOWiki (a ontology-based wiki for data annotation and knowledge integration; Hoehndorf, et al., 2009), Gene Wiki (a wiki for human gene annotation; Huss, et al., 2010; Huss, et al., 2008) and PDBWiki (a scientific wiki for the community annotation of protein structures; Stehr, et al., 2010). However, the wiki-based integration has its own shortcomings, including the unstructured data generated, the lack of a standard format for data exchange, the lack of credit for authorship and vulnerability to malicious editing (Lee, 2008; Potthast, et al., 2008).

3. Challenges ahead

Although a number of current efforts have been devoted to data integration, none of them have achieved a pre-eminent impact on their field yet. Since NGS data are growing at an exponential rate, the need for data integration is continually demanding and challenges for data integration are greatly increasing.

3.1 Data as a service

The low-cost and high-throughput NGS technologies can generate huge amounts of data at a relatively short period. To keep pace with the revolution of sequencing technologies, genome sequencing projects have transitioned from classical model organisms (e.g., fly, mouse, yeast), to other organisms (e.g., camel, dog, panda) and eventually, to sequencing individuals within populations, exemplified by the 1000 Genomes Project—a collection of the genomes of 1,000 humans (http://www.1000genomes.org) and the Genome 10K Project—a genomic zoo of genome sequences of 10,000 vertebrate species (http://www.genome10k.org). The era of $1000 personal genome sequencing is approaching within the following years and would produce unparalleled large-scale data, presenting considerable challenges for data integration.

It is infeasible to integrate such large amounts of data into a single point (such as a data warehouse). Data sources are developed for different purposes and fulfill different functions. Therefore, it is promising to establish an efficient way for data exchange among these distributed and heterogeneous data sources. However, a dozen of data sources are designed merely for data storage, but not for data exchange. The growing volume of biological data also requires "computer-readable" approaches for data integration. To ease data integration, data sources need to turn into service providers. In other words, data sources should not only serve as data providers that provide data for human reading with web interfaces (e.g., HTML), but also function as service providers that provide data for computer interoperability via WS. Service providers supply data as a WS, facilitating computer-to-computer interactions and thus enabling automated data integration from multiple data sources (Hansen, et al., 2003). As mentioned, there are several different

protocols that can be used for creating WS. Among them, SOAP and REST have been widely adopted (Figure 1). SOAP is a well-defined standard with XML-structured messaging for request and response, whereas REST is relatively lightweight, relying on HTTP methods (viz., POST, GET, PUT or DELETE). Most commercial applications expose their services as RESTful Web APIs (Figure 1), largely due to its simplicity and easy implementation.

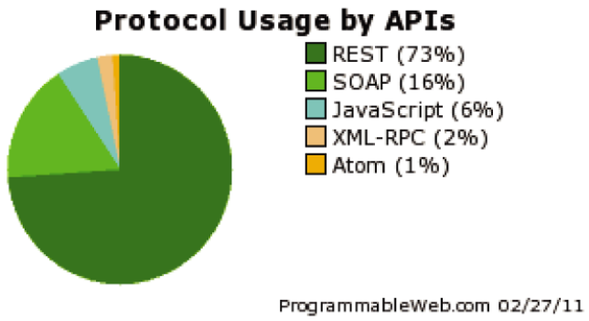


Fig. 1. Statistics of Web API protocols (obtained from http://www.programmableweb.com/apis, which collects more than 3,000 Web APIs; last access: February 27, 2011).

3.2 Standards for biological data

Due to the complex nature of biology, there are a wide variety of biological data types, e.g., sequence data, gene expression data, protein-protein interaction data, pathway data (Karasavvas, et al., 2004). Data sources store different data types as different formats (Li, 2006): flat file (e.g., tab-delimited file), sequence file (e.g., FASTA), structure file (e.g., PSF—Protein Structure File), and XML file (e.g., KGML—KEGG Markup Language for describing graph objects). Data sources often adopt their preferable data formats; even for a same data type, data formats in different sources are often incompatible. It is also noted that new data formats are often invented along with the development of related technologies. Examples of newly invented file formats include SAM (Sequence Alignment/MAP; a generic nucleotide alignment format that describes the alignment of query sequences or sequencing reads to a reference sequence or assembly; Li, et al., 2009), and GVF (Genome Variation Format; a simple tab-delimited format for describing genome variation data; Reese, et al., 2010). In addition, data sources output their data in diverse formats, such as HTML, raw file formats, and XML-based file formats. Taken together, diverse and heterogeneous data formats complicate data exchange, posing challenges for data integration.

Standards for biological data formats can ease data exchange and integration. There has been a successful attempt for standardizing biological pathway data. Pathway-related data sources differed in their data representation, making data integration difficult and inefficient. For this reason, BioPAX (Demir, et al., 2010) has been developed to deliver a compatible standard, facilitating integration, exchange, visualization and analysis of biological pathway data. Another effort related to cope with data incompatibilities of bioinformatics repositories has been devoted to the standardization issues of data exchange formats and WS (Katayama, et al., 2010). In short, establishing standard formats for biological data can realize efficient data exchange and integration. In return, standard data formats facilitate subsequent data analysis and visualization as well as downstream software development.

Equally important, data integration also requires standardizing nomenclature and ontologies for biological data (Rubin, et al., 2008). Suppose two data sources need to exchange gene annotations. They must share a standard regarding gene name. Otherwise, any ambiguity or inconsistency in nomenclature would bring a burden to data integration. Attention has been paid to standardizing nomenclature and ontologies for biological data, e.g., BioPortal (Noy, et al., 2009; Rubin, et al., 2006) for integrating and sharing biomedical ontologies in National Center for Biomedical Ontology, GO (Ashburner, et al., 2000) for standardizing the representation of gene and gene product attributes, HGNC (Seal, et al., 2011) for standardizing human gene symbols and names, OBO (Open Biomedical Ontologies) (Smith, et al., 2007) for creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. However, a centralized system for nomenclature and ontologies standardization may not keep good pace with the rapid accumulation of biological data and any gap in standardization would provoke difficulties for data integration. A wiki-based system might be promising to harness all communities' efforts in standardizing nomenclature and ontologies collaboratively and efficiently.

3.3 WS-based pipelines

The goal of data integration is to enable combining information from different resources in an automated fashion without human intervention, so as to handle the increasing accumulation of biological data (Sarkar, et al., 2008). Towards this goal, data to be integrated should be re-defined in a broader manner, which include not merely sequences and other raw data, but also methods, tools, algorithms, analyzed results, discovered knowledge (see a paper for knowledge integration; Clark, 2007) and even connections among people (Zhang, et al., 2009). All kinds of data can be provided as a service. That is, raw data should be accessible via WS, methods, tools, and algorithms that are used to analyze data should be offered as WS (that is SaaS, Software as a Service), and analyzed results and discovered knowledge should be also delivered as WS (Zhang, et al., 2009). As a result, WS perform a variety of data manipulation, including data retrieval, integration, analysis, visualization, and sharing.

A pipeline with a combination of multiple WS can achieve data integration (Zhang, et al., 2009). Such WS-based pipelines lower technological entrance barriers and provide users with a lightweight programming environment. WS-based pipelines feature computer-to-computer data exchange, simplify data integration and analysis, maximize the scope of sharing and reuse, and function as a medium to link users located anywhere with similar research interests, and finally to form a scientific social community (SSC). SSC reflects several key elements of Web 2.0 and enables data integration, analysis and sharing with greater convenience, speed and efficiency (Zhang, et al., 2009). Any user may easily create WS-based pipelines (adding value), publish them online, and subscribe to pipelines created by other users. Consequently, pipelines may be widely shared, re-used and even integrated into other pipelines. As a result, communications and collaborations among users in SSC can be greatly increased, making knowledge discovery through collective intelligence possible. In addition, SSC can also serve as a registry for collecting WS (Bhagat, et al., 2010; Pettifer, et al., 2010).

3.4 Semantic Web Services

The ever-evolving next-generation Web (NGW), characterized as the Semantic Web, aims to provide information not only for human, but also for computers to semantically process

large-scale data and automatically discover knowledge. From this view, the Semantic Web befits well with the exponential growth of biological data and promises in providing solutions for data integration and advancing translational research (Ruttenberg, et al., 2007). Semantic Web technologies have been applied for data integration as mentioned above. Nevertheless, these applications in essence belong to semantic warehouses and still have pains for integrating dynamic data. One potential solution is to combine WS with Semantic Web technologies and to provide Semantic WS (Matos, et al., 2010; Vandervalk, et al., 2009), namely, RDF-based WS for automated data processing and reasoning. As mentioned, WS are designed not only to perform a query, but also to conduct a computation. Considering that NGS technologies can swiftly generate hundreds of gigabases of sequencing data, WS would become increasingly data-intensive and computation-intensive (e.g., alignment of multiple large-scale sequences). Therefore, to deal with such large-scale data management and analysis, Semantic WS necessitate to adopt advances in high performance computing (Schadt, et al., 2010), such as, cloud/grid computing (Bateman and Wood, 2009; Stein, 2010) and Service-Oriented Computing (Papazoglou, et al., 2008). In addition, a Semantic WS framework (Wilkinson, et al., 2010) is also needed, in order to set up Semantic WS workflows or pipelines.

4. Conclusions

As a critical topic in bioinformatics, data integration bears fundamental significance for biological studies. Efforts have been devoted to this topic and the corresponding approaches for data integration have moved from traditional ones, e.g., data warehousing and federated databasing, to modern ones based on several advanced technologies, e.g., Web Service, Semantic Web and Wiki. The rapid development of sequencing technologies poses tremendous challenges for data integration. Integration of large-scale data not only requires adoption of informatics advances, but also needs communications and collaborations among people in related biological communities to maximize data openness via WS, set up standards for biological data, create Semantic WS-based pipelines and form a scientific social community. Such community harnesses collective intelligence and collaborative efforts for data integration, analysis and sharing, having the potential to be an ideal community of the people, by the people, and for the people.

5. References

Amberger, J., *et al.* (2009) McKusick's Online Mendelian Inheritance in Man (OMIM), *Nucleic Acids Res*, 37, D793-796.

Andreeva, A., *et al.* (2008) Data growth and its impact on the SCOP database: new developments, *Nucleic Acids Res*, 36, D419-425.

Antezana, E., *et al.* (2009) Biological knowledge management: the emerging role of the Semantic Web technologies, *Brief Bioinform*, 10, 392-407.

Aranda, B., *et al.* (2010) The IntAct molecular interaction database in 2010, *Nucleic Acids Res*, 38, D525-531.

Arita, M. (2009) A pitfall of wiki solution for biological databases, *Brief Bioinform*, 10, 295-296.

Ashburner, M., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat Genet*, 25, 25-29.

Bader, G.D., *et al.* (2003) BIND: the Biomolecular Interaction Network Database, *Nucleic Acids Res*, 31, 248-250.

Bairoch, A. (2000) The ENZYME database in 2000, *Nucleic Acids Res*, 28, 304-305.

Bairoch, A., *et al.* (2004) Swiss-Prot: juggling between evolution and stability, *Brief Bioinform*, 5, 39-55.

Barrett, T., *et al.* (2011) NCBI GEO: archive for functional genomics data sets--10 years on, *Nucleic Acids Res*, 39, D1005-1010.

Bateman, A. and Wood, M. (2009) Cloud computing, *Bioinformatics*, 25, 1475.

Belleau, F., *et al.* (2008) Bio2RDF: towards a mashup to build bioinformatics knowledge systems, *J Biomed Inform*, 41, 706-716.

Benson, D.A., *et al.* (2006) GenBank, *Nucleic Acids Res*, 34, D16-20.

Bhagat, J., *et al.* (2010) BioCatalogue: a universal catalogue of web services for the life sciences, *Nucleic Acids Res*, 38, W689-694.

Bidartondo, M.I. (2008) Preserving accuracy in GenBank, *Science*, 319, 1616.

Birkland, A. and Yona, G. (2006) BIOZON: a hub of heterogeneous biological data, *Nucleic acids research*, 34, D235-242.

Bizer, C. (2009) The Emerging Web of Linked Data, *Ieee Intell Syst*, 24, 87-92.

Blake, J.A., *et al.* (2011) The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics, *Nucleic Acids Res*, 39, D842-848.

Boguski, M.S., *et al.* (1993) dbEST--database for "expressed sequence tags", *Nat Genet*, 4, 332-333.

Bota, M. and Swanson, L.W. (2010) Collating and Curating Neuroanatomical Nomenclatures: Principles and Use of the Brain Architecture Knowledge Management System (BAMS), *Front Neuroinformatics*, 4, 3.

Brazas, M.D., *et al.* (2010) Providing web servers and training in Bioinformatics: 2010 update on the Bioinformatics Links Directory, *Nucleic Acids Res*, 38, W3-6.

Ceol, A., *et al.* (2010) MINT, the molecular interaction database: 2009 update, *Nucleic Acids Res*, 38, D532-539.

Cheung, K.H., *et al.* (2002) YMD: a microarray database for large-scale gene expression analysis, *Proc AMIA Symp*, 140-144.

Cheung, K.H., *et al.* (2005) YeastHub: a semantic web use case for integrating data in the life sciences domain, *Bioinformatics*, 21 Suppl 1, i85-96.

Cheung, K.H., *et al.* (2008) HCLS 2.0/3.0: health care and life sciences data mashup using Web 2.0/3.0, *J Biomed Inform*, 41, 694-705.

Chung, S.Y. and Wong, L. (1999) Kleisli: a new tool for data integration in biology, *Trends Biotechnol*, 17, 351-355.

Ciccarese, P., *et al.* (2008) The SWAN biomedical discourse ontology, *J Biomed Inform*, 41, 739-751.

Clark, T. (2007) Knowledge Integration in Biomedicine: Technology and Community, *Briefings in bioinformatics*, 8, E1-E3.

Clark, T. and Kinoshita, J. (2007) Alzforum and SWAN: the present and future of scientific web communities, *Briefings in bioinformatics*, 8, 163-171.

Crasto, C.J., *et al.* (2007) SenseLab: new developments in disseminating neuroscience information, *Brief Bioinform*, 8, 150-162.

Crasto, C.J. and Shepherd, G.M. (2007) Managing knowledge in neuroscience, *Methods Mol Biol*, 401, 3-21.

Croft, D., *et al.* (2011) Reactome: a database of reactions, pathways and biological processes, *Nucleic Acids Res*, 39, D691-697.

Davidsen, T., *et al.* (2010) The comprehensive microbial resource, *Nucleic Acids Res*, 38, D340-345.

Davidson, S.B., *et al.* (2001) K2/Kleisli and GUS: Experiments in integrated access to genomic data sources, *Ibm Syst J*, 40, 512-531.

Davidson, S.B., *et al.* (1995) Challenges in integrating biological data sources, *J Comput Biol*, 2, 557-572.

Demir, E., *et al.* (2010) The BioPAX community standard for pathway data sharing, *Nat Biotechnol*, 28, 935-942.

Dibernardo, M., *et al.* (2008) Semi-automatic web service composition for the life sciences using the BioMoby semantic web framework, *Journal of biomedical informatics*.

Dowell, R.D., *et al.* (2001) The distributed annotation system, *BMC bioinformatics*, 2, 7.

Engel, S.R., *et al.* (2010) Saccharomyces Genome Database provides mutant phenotype data, *Nucleic Acids Res*, 38, D433-436.

Flicek, P., *et al.* (2011) Ensembl 2011, *Nucleic Acids Res*, 39, D800-806.

Gilbert, D.G. (2007) DroSpeGe: rapid access database for new Drosophila species genomes, *Nucleic Acids Res*, 35, D480-485.

Giles, J. (2005) Internet encyclopaedias go head to head, *Nature*, 438, 900-901.

Giles, J. (2007) Key biology databases go wiki, *Nature*, 445, 691.

Goble, C. and Stevens, R. (2008) State of the nation in data integration for bioinformatics, *J Biomed Inform*, 41, 687-693.

Good, B.M. and Wilkinson, M.D. (2006) The Life Sciences Semantic Web is full of creeps!, *Briefings in bioinformatics*, 7, 275-286.

Greene, L.H., *et al.* (2007) The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution, *Nucleic Acids Res*, 35, D291-297.

Haas, L.M., *et al.* (2001) DiscoveryLink: A system for integrated access to life sciences data sources, *Ibm Syst J*, 40, 489-511.

Haider, S., *et al.* (2009) BioMart Central Portal--unified access to biological data, *Nucleic Acids Res*, 37, W23-27.

Hansen, M., *et al.* (2003) Data integration using Web Services, *Lect Notes Comput Sc*, 2590, 165-182.

Harger, C., *et al.* (1998) The Genome Sequence DataBase (GSDB): improving data quality and data access, *Nucleic Acids Res*, 26, 21-26.

Hariharaputran, S., *et al.* (2007) VINEdb: a data warehouse for integration and interactive exploration of life science data, *Journal of Integrative Bioinformatics*, 4, 63.

Hekkelman, M.L. and Vriend, G. (2005) MRS: a fast and compact retrieval system for biological data, *Nucleic Acids Res*, 33, W766-769.

Hendler, J. (2003) Science and the semantic web, *Science (New York, N.Y*, 299, 520-521.

Hoehndorf, R., *et al.* (2009) BOWiki: an ontology-based wiki for annotation of data and integration of knowledge in biology, *BMC Bioinformatics*, 10 Suppl 5, S5.

Hoffmann, R. (2008) A wiki for the life sciences where authorship matters, *Nat Genet*, 40, 1047-1051.

Hull, D., *et al.* (2006) Taverna: a tool for building and running workflows of services, *Nucleic acids research*, 34, W729-732.

Hunter, S., *et al.* (2009) InterPro: the integrative protein signature database, *Nucleic Acids Res*, 37, D211-215.

Huss, J.W., 3rd, *et al.* (2010) The Gene Wiki: community intelligence applied to human gene annotation, *Nucleic Acids Res*, 38, D633-639.

Huss, J.W., 3rd, *et al.* (2008) A gene wiki for community annotation of gene function, *PLoS biology*, 6, e175.

Jenkinson, A.M., *et al.* (2008) Integrating biological data--the Distributed Annotation System, *BMC Bioinformatics*, 9 Suppl 8, S3.

Jones, P., *et al.* (2008) PRIDE: new developments and new datasets, *Nucleic Acids Res*, 36, D878-883.

Kanehisa, M., *et al.* (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs, *Nucleic Acids Res*, 38, D355-360.

Karasavvas, K.A., *et al.* (2004) Bioinformatics integration and agent technology, *J Biomed Inform*, 37, 205-219.

Karp, P.D., *et al.* (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes, *Nucleic Acids Res*, 33, 6083-6089.

Katayama, T., *et al.* (2010) The DBCLS BioHackathon: standardization and interoperability for bioinformatics web services and workflows. The DBCLS BioHackathon Consortium*, *J Biomed Semantics*, 1, 8.

Kawas, E., *et al.* (2006) BioMoby extensions to the Taverna workflow management and enactment software, *BMC bioinformatics*, 7, 523.

Keseler, I.M., *et al.* (2011) EcoCyc: a comprehensive database of Escherichia coli biology, *Nucleic Acids Res*, 39, D583-590.

Keshava Prasad, T.S., *et al.* (2009) Human Protein Reference Database--2009 update, *Nucleic Acids Res*, 37, D767-772.

Kinoshita, J. and Clark, T. (2007) Alzforum, *Methods in molecular biology (Clifton, N.J*, 401, 365-381.

Lee, T.J., *et al.* (2006) BioWarehouse: a bioinformatics database warehouse toolkit, *BMC bioinformatics*, 7, 170.

Lee, T.L. (2008) Big data: open-source format needed to aid wiki collaboration, *Nature*, 455, 461.

Lein, E.S., *et al.* (2007) Genome-wide atlas of gene expression in the adult mouse brain, *Nature*, 445, 168-176.

Letovsky, S.I., *et al.* (1998) GDB: the Human Genome Database, *Nucleic Acids Res*, 26, 94-99.

Li, A. (2006) Facing the challenges of data integreation in biosciences, *Engineering Letters*, 13, EL_13_13_13.

Li, H., *et al.* (2009) The Sequence Alignment/Map format and SAMtools, *Bioinformatics*, 25, 2078-2079.

Lord, P., *et al.* (2004) Applying Semantic Web services to bioinformatics: Experiences gained, lessons learnt, *Semantic Web - Iswc 2004, Proceedings*, 3298, 350-364.

Maglott, D., *et al.* (2011) Entrez Gene: gene-centered information at NCBI, *Nucleic Acids Res*, 39, D52-57.

Maojo, V., *et al.* (2011) Biomedical Ontologies: Toward Scientific Debate, *Methods Inf Med*, 50, [Epub ahead of print].

Mardis, E.R. (2010) The $1,000 genome, the $100,000 analysis?, *Genome Med*, 2, 84.

Marenco, L., *et al.* (2004) QIS: A framework for biomedical database federation, *J Am Med Inform Assoc*, 11, 523-534.

Matos, E.E., *et al.* (2010) CelOWS: an ontology based framework for the provision of semantic web services related to biological models, *J Biomed Inform*, 43, 125-136.

McLean, R., *et al.* (2007) The effect of Web 2.0 on the future of medical practice and education: Darwikinian evolution or folksonomic revolution?, *Medical Journal of Australia*, 187, 174-177.

Messina, D.N. and Sonnhammer, E.L. (2009) DASher: a stand-alone protein sequence client for DAS, the Distributed Annotation System, *Bioinformatics*, 25, 1333-1334.

Mons, B., *et al.* (2008) Calling on a million minds for community annotation in WikiProteins, *Genome Biol*, 9, R89.

Neerincx, P.B. and Leunissen, J.A. (2005) Evolution of web services in bioinformatics, *Brief Bioinform*, 6, 178-188.

Noy, N.F. (2004) Semantic integration: A survey of ontology-based approaches, *Sigmod Record*, 33, 65-70.

Noy, N.F., *et al.* (2009) BioPortal: ontologies and integrated data resources at the click of a mouse, *Nucleic Acids Res*, 37, W170-173.

Oinn, T., *et al.* (2004) Taverna: a tool for the composition and enactment of bioinformatics workflows, *Bioinformatics*, 20, 3045-3054.

Olason, P.I. (2005) Integrating protein annotation resources through the Distributed Annotation System, *Nucleic acids research*, 33, W468-470.

Papazoglou, M.P., *et al.* (2008) Service-oriented computing: a research roadmap, *International Journal of Cooperative Information Systems*, 17, 223-255.

Parkinson, H., *et al.* (2011) ArrayExpress update--an archive of microarray and high-throughput sequencing-based functional genomics experiments, *Nucleic Acids Res*, 39, D1002-1004.

Pettifer, S., *et al.* (2010) The EMBRACE web service collection, *Nucleic Acids Res*, 38, W683-688.

Potthast, M., *et al.* (2008) Automatic vandalism detection in Wikipedia, *Advances in Information Retrieval*, 4956, 663-668.

Pruitt, K.D. and Maglott, D.R. (2001) RefSeq and LocusLink: NCBI gene-centered resources, *Nucleic Acids Res*, 29, 137-140.

Pruitt, K.D., *et al.* (2009) NCBI Reference Sequences: current status, policy and new initiatives, *Nucleic Acids Res*, 37, D32-36.

Reese, M.G., *et al.* (2010) A standard variation file format for human genome sequences, *Genome Biol*, 11, R88.

Rose, P.W., *et al.* (2011) The RCSB Protein Data Bank: redesigned web site and web services, *Nucleic Acids Res*, 39, D392-401.

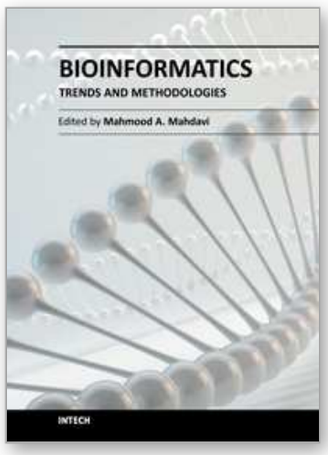
Rubin, D.L., *et al.* (2006) National Center for Biomedical Ontology: advancing biomedicine through structured organization of scientific knowledge, *OMICS*, 10, 185-198.

Rubin, D.L., *et al.* (2008) Biomedical ontologies: a functional perspective, *Brief Bioinform*, 9, 75-90.

Ruttenberg, A., *et al.* (2007) Advancing translational research with the Semantic Web, *BMC Bioinformatics*, 8 Suppl 3, S2.

Salwinski, L., *et al.* (2004) The Database of Interacting Proteins: 2004 update, *Nucleic Acids Res*, 32, D449-451.

Salzberg, S.L. (2007) Genome re-annotation: a wiki solution?, *Genome biology*, 8, 102.
Sarkar, I.N., *et al.* (2008) Automated simultaneous analysis phylogenetics (ASAP): an enabling tool for phlyogenomics, *BMC bioinformatics*, 9, 103.
Sayers, E.W., *et al.* (2011) Database resources of the National Center for Biotechnology Information, *Nucleic Acids Res*, 39, D38-51.
Schadt, E.E., *et al.* (2010) Computational solutions to large-scale data management and analysis, *Nat Rev Genet*, 11, 647-657.
Seal, R.L., *et al.* (2011) genenames.org: the HGNC resources in 2011, *Nucleic Acids Res*, 39, D514-519.
Shah, S.P., *et al.* (2005) Atlas - a data warehouse for integrative bioinformatics, *BMC Bioinformatics*, 6, 34.
Shi, X. (2007) Semantic Web Services: An Unfulfilled Promise, *IEEE IT Professional*, 9, 42-45.
Sigrist, C.J., *et al.* (2010) PROSITE, a protein domain database for functional characterization and annotation, *Nucleic Acids Res*, 38, D161-166.
Smith, B., *et al.* (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nat Biotechnol*, 25, 1251-1255.
Stehr, H., *et al.* (2010) PDBWiki: added value through community annotation of the Protein Data Bank, *Database (Oxford)*, 2010, baq009.
Stein, L. (2002) Creating a bioinformatics nation, *Nature*, 417, 119-120.
Stein, L.D. (2003) Integrating biological databases, *Nat Rev Genet*, 4, 337-345.
Stein, L.D. (2010) The case for cloud computing in genome informatics, *Genome Biol*, 11, 207.
Stevens, R., *et al.* (2000) TAMBIS: transparent access to multiple bioinformatics information sources, *Bioinformatics*, 16, 184-185.
Stevens, R.D., *et al.* (2003) myGrid: personalised bioinformatics on the information grid, *Bioinformatics (Oxford, England)*, 19 Suppl 1, i302-304.
The UniProt Consortium (2011) Ongoing and future developments at the Universal Protein Resource, *Nucleic Acids Res*, 39, D214-219.
Trissl, S., *et al.* (2005) Columba: an integrated database of proteins, structures, and annotations, *BMC Bioinformatics*, 6, 81.
Vandervalk, B.P., *et al.* (2009) Moby and Moby 2: creatures of the deep (web), *Brief Bioinform*, 10, 114-128.
Waldrop, M. (2008) Big data: Wikiomics, *Nature*, 455, 22-25.
Wilkinson, M.D. and Links, M. (2002) BioMOBY: an open source biological web services proposal, *Briefings in bioinformatics*, 3, 331-341.
Wilkinson, M.D., *et al.* (2010) SADI, SHARE, and the in silico scientific method, *BMC Bioinformatics*, 11 Suppl 12, S7.
Wilkinson, M.D., *et al.* (2008) Interoperability with Moby 1.0--it's better than sharing your toothbrush!, *Briefings in bioinformatics*, 9, 220-231.
Yager, K. (2006) Wiki ware could harness the Internet for science, *Nature*, 440, 278.
Zdobnov, E.M., *et al.* (2002) The EBI SRS server--recent developments, *Bioinformatics*, 18, 368-373.
Zhang, Z., *et al.* (2009) Bringing Web 2.0 to bioinformatics, *Brief Bioinform*, 10, 1-10.
Zhang, Z. and Townsend, J.P. (2010) The filamentous fungal gene expression database (FFGED), *Fungal Genet Biol*, 47, 199-204.
Zhao, J., *et al.* (2009) Linked data and provenance in biological data webs, *Brief Bioinform*, 10, 139-152.



Bioinformatics - Trends and Methodologies

Edited by Dr. Mahmood A. Mahdavi

ISBN 978-953-307-282-1
Hard cover, 722 pages
Publisher InTech
Published online 02, November, 2011
Published in print edition November, 2011

Bioinformatics - Trends and Methodologies is a collection of different views on most recent topics and basic concepts in bioinformatics. This book suits young researchers who seek basic fundamentals of bioinformatic skills such as data mining, data integration, sequence analysis and gene expression analysis as well as scientists who are interested in current research in computational biology and bioinformatics including next generation sequencing, transcriptional analysis and drug design. Because of the rapid development of new technologies in molecular biology, new bioinformatic techniques emerge accordingly to keep the pace of in silico development of life science. This book focuses partly on such new techniques and their applications in biomedical science. These techniques maybe useful in identification of some diseases and cellular disorders and narrow down the number of experiments required for medical diagnostic.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen