

Genome Database Development

In NGS, a large amount of data is being generated assembled, annotated & comparative analysis is being carried out. Therefore it is most essential to consolidate all the relevant information obtained by NGS analysis in database format. Xcelris Genomics aim to construct a powerful database with excellent interface for users. Xcelris Genomics database development team has developed a series of useful genome database, integration, RNA database. So far, over 15 important genome databases have been successfully constructed and well maintained. These databases and services, such as rice, silkworm, chicken, cattle, human etc., have been widely used by researchers around the world.



Accordingly, Xcelris Genomics has accumulated rich experience in building genomics databases, constructing related applications, and developing genome browser platforms for providing customized solution to each customers need.

Requirements

Genome Database

- Genome information
- Platform/technology
- Tools used for analysis.
- Raw data genome, assembled data and statistics
- Functional annotated data and associated files

Metagenome Database

- Metadata file which includes information about metagenome.
- Platform/technology
- Sequencing type (16S/whole metagenome sequencing)
- Tools used for analysis
- Raw data, assembled data and statistics
- Taxonomic abundance
- Functional abundance, associated data and statistics

RNA-Seq Database

- Transcriptome information
- Platform/technology
- Raw data, assembled data and statistics
- Tools used for analysis
- Functional annotated data and associated files
- Pathways

Small RNA Database

- Small RNA information
 - Platform/technology
 - Raw data, assembled data and statistics
 - Tools used for analysis
 - Assembled data and statistics
 - pre-cursor miRNA, hairpin loops structures, miRNA target information and associated files
-

Experimental Pipeline

Preparing the data

The preparatory work is divided into several parts: conducting data compilation and classification, designing the database, schema, importing relevant data, optimizing the process, etc.. MySQL, an opensource RDBMS with high performance, is applied to our database development.

Designing the core background services

We provide the services of mapping features to relevant genomes and exhibiting the result as a graph. Therefore a data engine and graphical design are pivotal elements.

Designing the user interface

A user friendly interface is very important for a good database. Firstly, we transfer the information selected by users to the core background services, then return an easy to use graphical page, which provides the basic operation of dragging, zooming and selecting.

Data retrieval and global searchS

Data retrieval is another core service. Users query the database to get the data they are interested in, and the system returns a complete report forms including annotations, sequence information and protein information etc.

Provide online services

Taking alignment tool as an example, we provide BLAST online for users to carry out alignment analysis.

Front-End of the database

We develop web and command line interface using HTML, CSS and CGI script as per the client's requirement.

Case Studies

Rare Disease Database

RareDDB is an integrated catalog of rare disease for Indian sub-population and global population by assembling various information including chromosome, genetic variations, disease, associated drug at one browse where users can customize easily. The database may be beneficial to clinicians, researchers or anyone working on rare disease or associated genes at molecular and clinical level.

Analysis Contents

- Construction of a genome database includes:
 - Database design, data collation, data uploading and database optimization: the database design must meet the requirement of massive data processing.
 - Integrating genome browser (Mapview): GBrowse.
 - Providing online services, such as alignment analysis based on the database: BLAST is available by default at present. Other online services can be integrated into genome database by need.
 - Data retrieval: We developed data retrieval service based on the database to generate reports that meet user's search criteria.
-

6. Frequently Asked Questions

Q: Is there anyone to maintain the database after the construction?

A: The efficient computing resources, sufficient storage resources, and professional database administrators and system administrators guarantee the genome database effective and normal application programs when users ask for new features.

Q: How to update the database after the construction of database?

A: Based on the database structure, we will make some modification of the database or some application programs when users ask for new features.

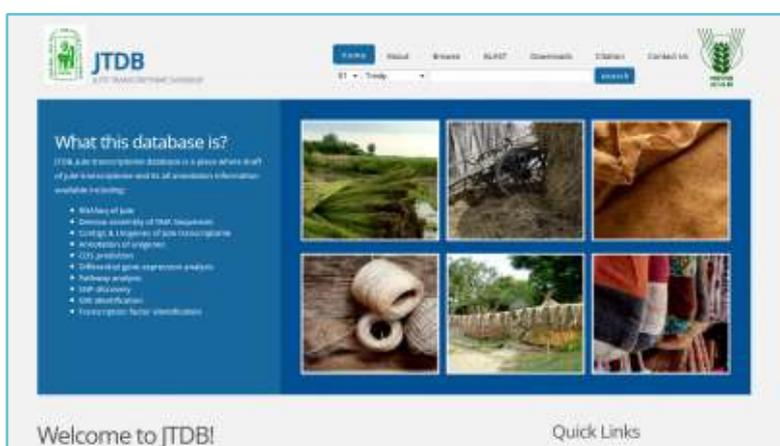
Q: Time period required to construct a database?

A: Four weeks is needed if the required genome database is based on GBrowse framework, and only two weeks if it is based on our self-developed genome database. However, at least six weeks is required when the users need to re-order the database.



The screenshot shows the RareDDB website interface. At the top, there is a navigation menu with links for HOME, ABOUT US, LITERATURE, TECHNOLOGY/STATISTICS, and CONTACT US. The main heading is "Cystic fibrosis". Below this, there is a "DESCRIPTION" section with text about the disease. To the right, there is a "Welcome, admin" message and a "Search database" form with a search button. The URL "www.rareddb.xcelrislabs.com" is visible at the bottom right of the page.

Figure 1: RareDDB Database Home Page (<http://www.rareddb.xcelrislabs.com/>)



The screenshot shows the Jute Transcriptome Database (JTDB) website. The header includes the JTDB logo and navigation links for Home, About, News, About, Downloads, Contact Us, and a search bar. The main content area features a section titled "What this database is?" with a list of features: "Millions of jute", "Diverse varieties of jute", "Cultivation & history of jute transcriptome", "Adaptation of jute", "JTE sequences", "Differential gene expression analysis", "Pathway analysis", "Gene ontology", "GO enrichment", and "Transcriptome factor identification". Below this, there are six small images showing jute plants and products. At the bottom, there is a "Welcome to JTDB!" message and a "Quick Links" section.

Figure 2: Jute Transcriptome Database Home Page