

A short instructions for using *Ganoderma lucidum* Database (GaLuDB)

Thanks for using GaLuDB, this document provide a brief introduction for GaLuDB. Please refer to our genome analysis paper and the GaLuDB paper for background information and an overview of GaLuDB. If you have any questions or suggestions, please contact cliu@implad.ac.cn or yjzhu@implad.ac.cn.

As described in the introduction, GaLuDB is the companion website to our genome analysis paper (Chen SL, Xu J, Liu C, Zhu YJ, Nelson DR, Zhou SG, Li CF, Wang LZ, Guo X, Sun YZ, *et al.*: **Genome sequence of the model medicinal mushroom *Ganoderma lucidum***. *Nat commun*, in press.) to support further mining of the genomic, genetic and chemical information of *G. lucidum*. GaLuDB provides a genome browser to navigate around the genome, a data page for the downloading of primary experimental and secondary analytical data for *G. lucidum* genome, a set of tools to mine GaLuDB genome and a wiki page for information exchange and sharing. These sections will be described in details below.

I. Genome browser

The genome browser is implemented with GBrowse, it allows users to access the assembled genome and available biological evidence supporting the predicted gene models. We use a MySQL database serving as the backend of Gbrowse. Through the genome browser, researchers can access tracks with information containing *G. lucidum* gene sequences, RNA-Seq reads, GC compositional domains and predicted protein sequences. In addition, some useful plugins have been incorporated into Gbrowse, including those can be used to dump sequence in fasta format and gene models in GFF format, blast against the displayed sequence, align the sequence to genome with BLAT, and so on.

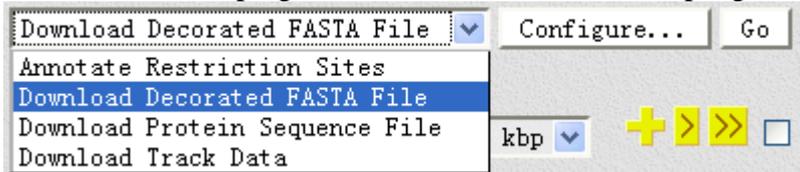
GBrowse consists of five tabbed sections, including Gbrowse, Select Tracks, Snapshots, Custom Tracks and Preferences. Under the Gbrowse section, there are four areas, namely, Search, Overview, Region and Details. Below, we describe some of the hypothetical scenarios for using GBrowse to demonstrate the functions.



- (1). You can log into genome browser by click . After you come into GBrowse page, enter "GaLu96scf" to view all 82 scaffolds. Then, you can click on a particular scaffold you are interested in.
- (2). From the overview graph for the scaffold, you can find the general survey information including GC content on "Overview" track, distribution of genes on "Gene Sequence" track and the regions supported by RNA-Seq data on "RNA-SEQ" track.
- (3). If you are interested in a particular gene, you can move your mouse over to the "Gene Sequence" track. The name of the gene and its coordination will pop up.
- (4). To review the structure of a gene in details, you need to enter the gene name in "Landmark or Region". An overview about gene structure and RNA-Seq evidence will be displayed at default. You can also search the gene annotations by clicking on "Gene

Sequence". You will be redirected to the page on "Tools->Fetch Gene Annotations". Alternatively, you can click "Select Track", and choose the corresponding tracks.

- (5). Additional useful plugins can be accessed from the up-right column by choosing



II. Data

Two categories of data are provided, the primary experimental data and then secondary analysis data. Six types of the data can be downloaded directly from GaLuDB. They are (1) Genome assembly; (2) Gene models. (3) Gene functional annotations; (4) Transcriptome data (5) Gene clusters (6) Compounds and etc. In addition, you can browse to see the gene clusters and chemical compound structures.

III. Tools for GaLuDB

The set of tools fall into four categories and will be described in details below.

1 Search

1.1 Blast+pfamscan

This tool allows a user to search the sequences based on BLAST and Hidden Markov Model (HMM). The query sequences can be mapped to genome sequences including genome, mRNA, CDS and proteins using programs such as Blastn, Blastp, Blastx, tBlastn, tBlastx and pfamscan. When a user searches these databases using Blast, you can the set the option for E-value or output format. However, if he/she chooses pfamscan, he/she will only be allowed to search the Pfam database.

2 Retrieve

2.1 Fetch Gene Annotations

This tool allows a user to obtain annotations for *G. lucidum* predicted genes based on those of the hits from Nr, Nt, KEGG, Swissprot, KOG, COG, Pfam and InterPro. In addition, gene ontology is assigned according to InterPro.

Two methods can be used to retrieve gene annotations. You can retrieve annotation information by gene ID when you choose

search *G. lucidum* gene annotations by gene id (example: GL15091)

In this case, you just can input gene ID like 'GL15091'. Alternatively, you can retrieve the information by keyword

search *G. lucidum* gene annotations by key word (example: P450)

In this case, the program will search the backend MySQL database with the annotation information. Please note that the input is case-insensitive and feel free to try different keywords for the genes of interest.

2.2 Fetch Sequences

This tool allows a user to fetch the *G. lucidum* sequences from a MySQL database including genome, mRNA, CDS and proteins. Optionally, you can extract a particular region partial of a gene when you enter the ranges for the region. It should be pointed out that only one sequence is allowed at a time.

2.3 Fetch IDs of Adjacent Genes

This tool allows a user to fetch the IDs of genes adjacent to a gene of interest. Additional parameters include numbers of genes or sequence ranges on the upstream and downstream of the query gene. The input files need to contain four columns. For “by region” option, the four columns are “gene ID”, the start and end of a region, and “r”. For “by number” option, the four columns are “gene ID”, number of genes from the upstream or downstream, and “n”. The results of genes are separated by new line. The output is compressed file in a .tar.gz format, which can be uncompressed using WinRAR.

3 Gene Models

3.1 Fetch a Gene Model

This tool allows a user to fetch gene models in GFF3 format for manual curation. The user just needs to enter a gene ID like 'GL15091' on the query form. Multiple gene IDs can be entered at once allowed.

3.2. Get a Sequence Alignment

This tool allows a user to retrieve homologous sequences for their query and build a multiple sequence alignment for the curation of the gene structure. At present time, this tool only supports annotating CAZY, FOLy, CYP450s and transporter homologues. The input is a protein sequence that is likely to be homologous sequence of CAZY, FOLy, CYP450s and transporter. The output is a .meg file, which can be opened by [MEGA](#) software.

3.3. Upload a Gene Model

This tool allows a user to upload the curated gene models in GFF3 format. After gene curation using [Apollo](#) or other gene curation tools, the user needs to save them to standard GFF3 format like the GFF created by MAKER. 'ID' and 'Parent' must be included in the 'exon', 'CDS' and 'mRNA' lines. You can upload three GFF files simultaneously. After submission, the gene models will be uploaded into MySQL database. The new models will substitute the old ones. A simple table can be downloaded, which include some changes about gene names (???).

4. Analysis

4.1. Find Gene Clusters

This tool allows a user to search for gene clusters given a list of genes belonging to particular functional groups. Only the list of *G. lucidum* gene IDs is needed, the functional classification of these genes will be retrieved internally based on InterPro. Optionally, you

can set the size of sliding window in number of genes. The P-value is calculated by hypergeometric probability. The output includes identified gene clusters. For genes in the clusters, the “1” in the second column means that the gene have been annotated by InterPro database.

4.2. Map Genes To Chromosomes

This tool allows a user to view the locations of genes on chromosomes. The genes are colored by their functional categories provided by the users. Before you open the output file, you need to download [svgviewer](#) and open .svg file using Google Chrome or Firefox. The chromosomal coordination of a gene can be found by clicking on that gene.

IV. Wiki page

This wiki page has all the regular functions of a wiki page. The page can be used to post topics on ‘Discussion Board’ pages, create or modify annotation information data on gene(s) of interest, communicate news to the research community, and share protocols, images, job postings, etc.