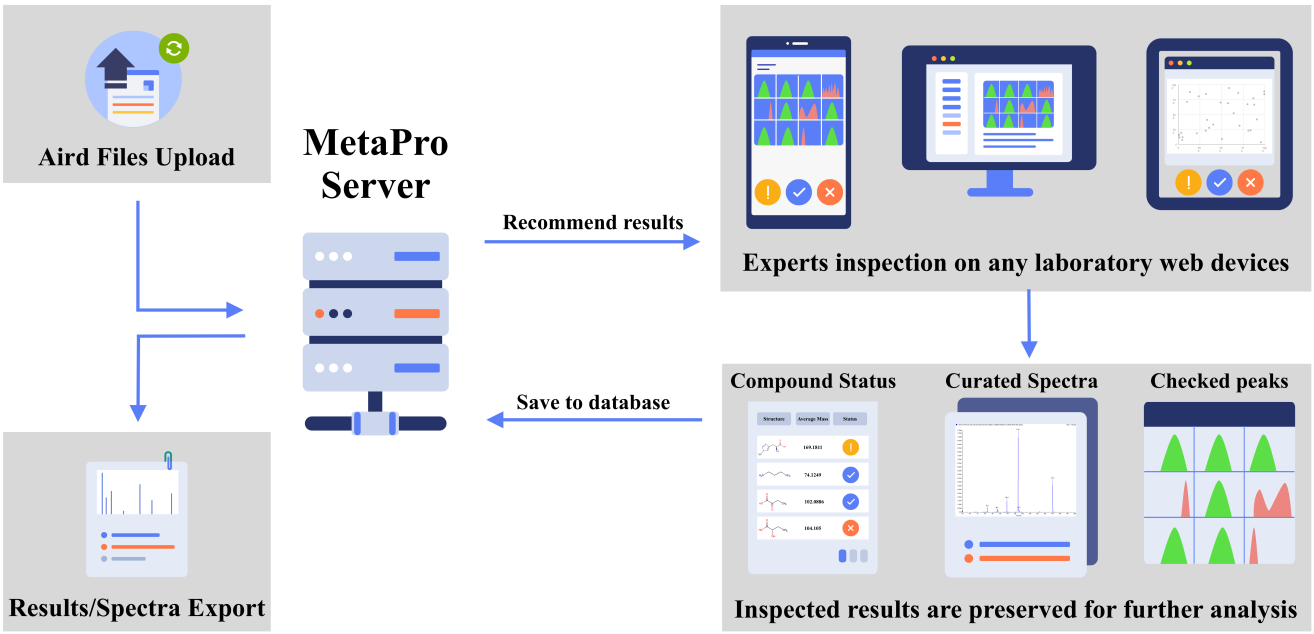
User Manual

Main Interface of MetaPro

Framework of MetaPro



Demo website

http://47.88.77.176:8080/

1 General Information

1.1 Scope of MetaPro Software

Currently high-resolution mass spectrometry instruments are widely used in metabolomics studies, but software in this filed usually have a lack of easily executed quality control and spectra management abilities. Besides, collaboration inconvenience on data processing also restrict analysis on efficiency.

1.2 Computer System Requirements

MetaPro is available for Windows 7(or above), Linux and MacOS. The minimum and recommended system specifications for **single-node deployment** are described in Table 1.

Specifications	Minumum	Recommended
CPU	Intel Core i5 10-gen or AMD R5	Intel Core i9 12-gen or AMD R9 5900
Hard drive	50GB free space	2x data set size
RAM	16GB	64GB(Core i9-10gen), 128GB(Core i9-12gen)

Table 1. **single-node deployment requirements**

the **multi-nodes deployment** requirements are described in Table 2.

Server Requirements	Deploy software
Core i7-10gen, 32GB RAM, 5TB SSD	MongoDB(Main Node), Redis
Core i7-10gen, 32GB RAM, 5TB SSD	MongoDB(Child Node, optional)
NAS(such as Synology RS2821), 100TB HDD	Data Repository
Core i9-12gen, 128GB RAM, 500GB SSD	MetaPro(Core computing node)
Core i7-10gen, 16GB RAM, 100GB HDD	AirdPro(Data Conversion)

Table 2. **multi-nodes deployment requirements recommended**

1.3 Acquisition Method and Vendor File Conversion

MetaPro supports for **DDA** data acquisition mode. Due to the need to support the whole platform of mass spectrometry file processing, we use a self-developed Computing-Oriented high compression data format-**Aird**. That means users should convert the vendor file to the aird format files using **AirdPro.** AirdPro supports the vendor files including **.raw, .wiff and .d** formats.

Aird Paper link: https://www.biorxiv.org/content/10.1101/2020.10.14.338921v1

1.4 Code Available

AirdPro: https://github.com/CSi-Studio/AirdPro

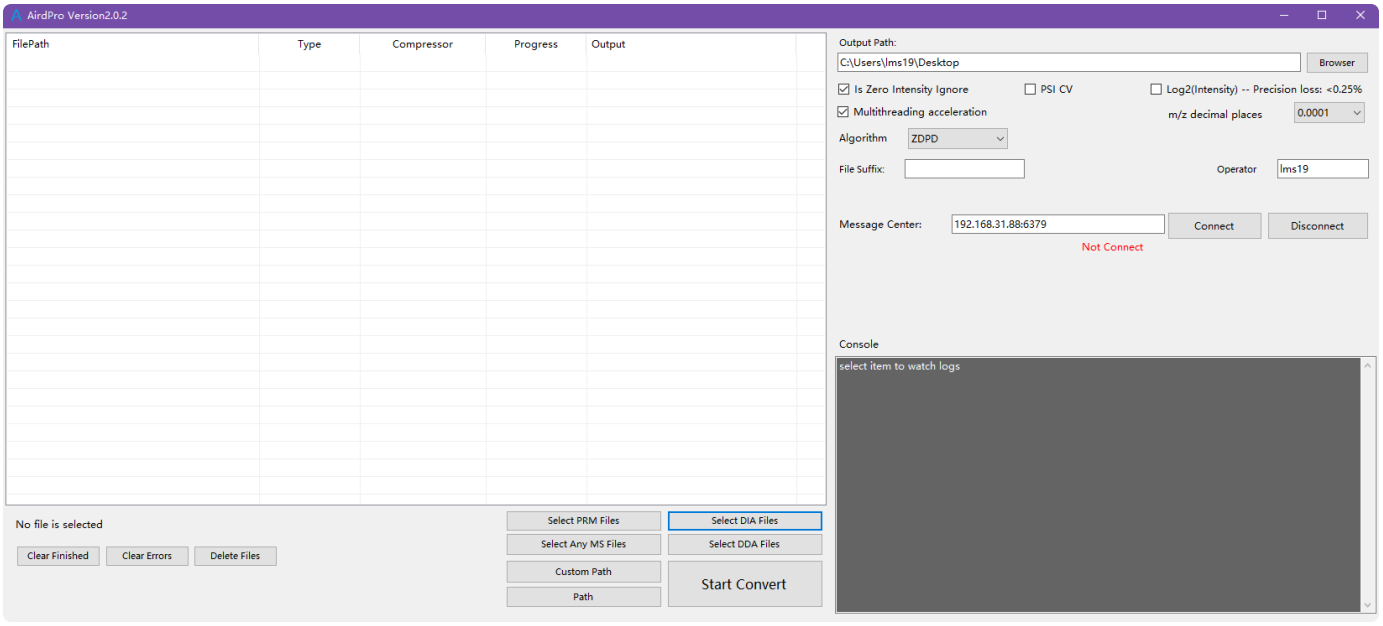
MetaPro: https://github.com/CSi-Studio/MetaPro

2 Getting Started

2.1 Associated software

MetaPro depend on the database and basic software below. All dependencies are free to use.

Software	Version	Download Link
Java(or OpenJDK)	15.0.1	OracleJDK: https://www.oracle.com/java/technologies/downloads/ OpenJDK: http://openjdk.java.net/projects/jdk/15/
Redis	6.2.6	Windows: https://github.com/MicrosoftArchive/redis/releases Others: https://redis.io/download
MongoDB	5.0.5	https://www.mongodb.com/try/download/community
AirdPro	2.0.0	https://github.com/CSi-Studio/AirdPro/releases



AirdPro Screenshot

2.2 Deployment

For windows/mac/linux users

Download the Zip setup file and unzip the file to the recommended disk which has an enough disk space.

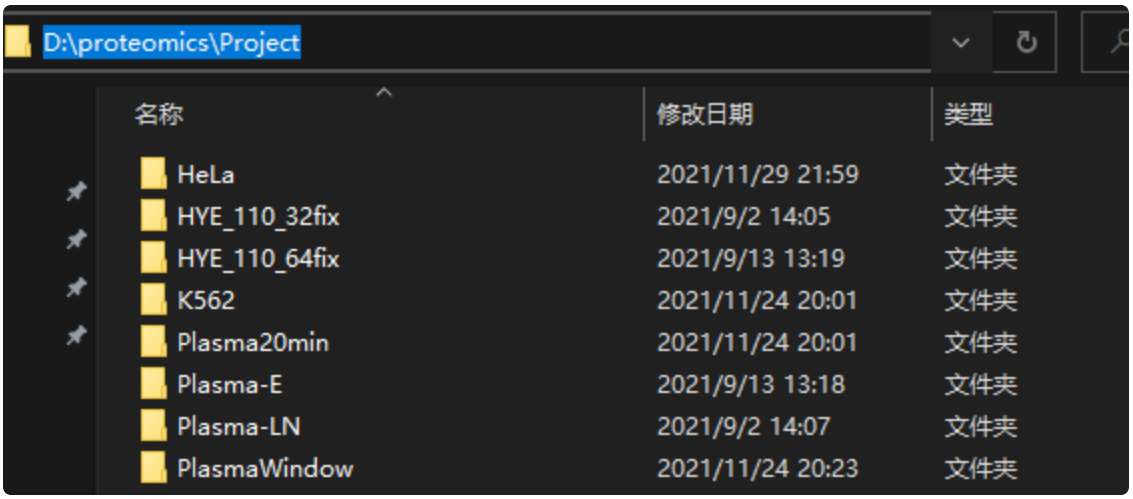
2.3 Configuration

2.3.1 MetaPro

User can edit the configuration file "application.properties" under the "libs" directory

Properties	Default	Comments
spring.data.mongodb.port	27017	default port for mongodb
spring.data.mongodb.database	metapro	default collection name for mongodb
spring.data.mongodb.host	127.0.0.1	default IP for mongodb
repository	**D:/proteomics**	**set your aird data repository location.Like D:/proteomics**
spring.redis.host	127.0.0.1	default IP for redis
spring.redis.port	6379	default Port for redis
spring.redis.database	0	default db for redis
spring.redis.password		no password as default

repository: User must change the repository before using MetaPro. Select a proper disk which contains enough disk space to store the data file(here means Aird format data). Suppose you choosed D:/Proteomics as the repository. Then you need to add a empty directory "Project" under the repository. Every Project Folders need to be placed under the "Project" directory. MetaPro will scan the "Project" folder automatically.



The Project Folder structure

3 How to use MetaPro

3.1 System conception description

In MetaPro, we provide these main database modules: experiment, compound, method, library, overview, project, spectra, task. The following table illustrates the definition for each module.

Experiment: An experiment is a basic analysis unit in MetaPro. It represents the mass spectrometry data derived from one sample. The actual storage mode of an experiment is an Aird file under the user-defined document folder. A few experiments can be divided into one batch.

Project: A project can have a few batches. Each batch has a few experiments. This is the largest unit where experiments could be organized.

Compound: A compound refers to a specific metabolite with name, id, m/z, retention time and other information stored in the database. Chosen compounds could be used to conduct targeted analysis.

Method: All parameters used in the analysis steps are extracted into a method. Different parameters composition could be saved into different methods.

Library: A library contains a few compounds.

Overview: An overview is the storage of the analyzed results. By checking a overview, multiple quality control operations can be achieved.

Spectra: A spectrum is correlated to a specific compound. Instrument type, collision energy, ionization mode and other description information about the mass spectrometry characteristic is stored in a spectrum.

Task: Task refers to a metabolomics analysis process to be done or undertaken. It's timely status and records could be seen in the task page. Currently three task templates are offered in MetaPro.

3.2 Project management page

This is the first page which users will see when they get into the system. Projects are created by users by clicking on the button in the top column. Users can input the name, owner, description information to generate a new project. Delete, update, search options are also available on this page. The project management page shows all the users' projects and their addition information, including batch numbers, experiment numbers,

overview numbers, owner and create date. To add experimental files into a project, users need to put all the aird files into a folder having the same name with the created project. After doing that, clicking on “scan files” will start reading these files into the system. Their brief information will be stored. Batch number and experiment number are automatically modified according to the actual number. Users can also check EIC for each experiment after the index is created in the system. Clicking on one project will get into its detailed information page, where its experiments information are fully displayed. Updating, deleting, modifying experiment information are operated here.

3.3 Library management page

This page has the similar visualization hierarchical with the project management page. Creating, updating, modifying, deleting options are similar but towards compound libraries. When creating a new library, MetaPro allow users to upload a formated CSV file to fill the compounds information in this library. Clicking a library means checking all the compounds that are concluded in this library.

3.4 Method management page

What is shown here is the extracted parameters which we call methods. Default parameters are offered for the users to generate a default method. Each parameter could be modified. A method could be used for multiple projects, batches and experiments, thus minimizing the time cost every time when users try to find the best parameters.

3.5 Task management page

All the task generated in the system could be found here. Currently three task types are available. They are “File Scan” task, “Targeted Analysis” task and “Untargeted Analysis” task. Each task has its create date, modified date, status and time cost for users to review.

3.6 How to start a targeted analysis task

To start a targeted analysis task, several actions should be done. First, the project should be properly set and all the files under this project should be scanned and indexed. Second, a prepared compounds list of interest needs to be imported as a library in MetaPro. Compounds information would be used during the results generation process. Third, a method used for these experiments should be created with proper parameters. After all the mentioned are prepared, users can go to the project management page and find the project that they want to analyze. Then clicking on “Prepare to analyze” will lead to the analysis preparation page. Here users need to choose the batches they want to analyze, the library used for each batch and the method for each batch. Then clicking on “targeted analyze” will start a task that could be seen in the task management page. Detailed progress inside this task is printed as system log and showed in the task management page. Once the status shows the task is done, users could go to the QC page to check the results. The QC inspection entrance is at the project management page.

3.7 How to use QC inspection page for targeted analysis results

In this page, users will learn how to check the system recommended results. The primary design conception of MetaPro QC interface is to cover as more as needed information to satisfy inspection for MSI level–1 confidence and provide fast inspection and modification operations. Based on experience from other metabolomics software and skilled engineer, visualization modules and inspection operations are optimized on peak picking, spectrum viewer, manual integration and so on.
On left of the page lies the compounds switch column. Compounds information like id, m/z, rt are showed and every dimension could be used as the ordering rule for these compounds. Compounds are organized into pages for users to swiftly search and check results compound by compound. The right part of the page is the peak picking results for this compound in all the samples in one batch. Users could check all results in a batch at one sight. To be more specific, different modules are displayed on the right side. The manual integration module allows users to modify the shape of the detected peaks. The spectra viewer module mainly illustrates the quality of the MS1 and MS2 spectra from the peak summit point. The batch peak inspection module let users to mark the results as success, fail and unknown. By doing this, only the approved results will be exported to the report for quantification results. Furthermore, it happens many times when users find

the retention time excurse from the given value. Thus, we provide a function named recalculate by new rt, meaning the results could be recalculated if users find the peak is far away from the expected situation. Besides, users could also change the parameters to recalculate the pick detection results towards one compound. This would be of help for some specific compounds.

3.8 How to start an untargeted analysis task

To start an untargeted analysis task, several actions should be done. First, the project should be properly set and all the files under this project should be scanned and indexed. Second, a method used for these experiments should be created with proper parameters. After all the mentioned are prepared, users can go to the project management page and find the project that they want to analyze. Then clicking on “Prepare to analyze” will lead to the analysis preparation page. Here users need to choose the batches they want to analyze and the method for each batch. Then clicking on “untargeted analyze” will start a task that could be seen in the task management page. Detailed progress inside this task is printed as system log and showed in the task management page. Once the status shows the task is done, users could go to the QC page to check the results. The QC inspection entrance is at the project management page.

3.9 How to use QC inspection page for untargeted analysis results

At this page, the main function is to check the common features among all experiments in a batch. Common features are called a bin here. On top of the page is the bin column, here all the bins are correlated and grouped based on their correlation index. To change the bin groups, input new minimum correlation index value and click on “Refresh”. All the bins are automatically generated based on the default bin parameters. If users want to change the bin numbers, they should click on “reset bin parameters”. This means generating bins from a batch of experiments with new parameters which could affect this process. On the bottom of this page lie two modules. One is ion peak inspection module while the other is bin correlation inspection module. The ion peak inspection module shows all the ion peaks that fall into this bin. The bin correlation module shows the correlation figure between two bins. Clicking on “Generate into a new library” will regard all the chosen bins as unknown compounds and save them into a new library.