**KHMN Scripts User Manual**

Part 1: Introduction and Installation

Part 2: Feature Table Input and Alignment

Part 3: MS2 Input and Filtering

3.1: Simple Conversion

3.2: Batch Conversion

Part 4: New MS2 Loading

Part 5: Seed Metabolites MS2 Input

Part 6: Seed Metabolites Identification and Marking

Part 7: MS2 Similarity and Molecular Network Generation

Part 8: Candidates Features Screening

Part 9: Nodes and Edges Annotation

9.1: Nodes and Edges Annotation of HCAAs

9.2 Nodes and Edges Annotation of CGFs

9.3 Edges Annotation of BOAs

Part 10: Redundancy Removal

**Part 1: Introduction and Installation**

KHMN is used to assist in screening and annotating functional metabolites.

To use KHMN, it is recommended to install python3.6.0 or above.

**Part 2: Feature Table Input and Alignment**

Multiple .xlsx files from different tissues can be dealt with at the same time. Tissue\_ID is needed to provide in feature tables. Considering the emergence of redundancy, features in different tissues would be aligned to generate a uniform Align\_ID for later redundancy removal.

**Part 3: MS2 Input and Filtering**

Every MS/MS spectrum in newly generated .mgf MS2 files corresponds to an independent feature. The MSCluster algorithm doesn’t need to be activated during molecular network construction. There are two situations in the MS/MS acquisition that the same acquisition condition contains multiple MS2 files or only one MS2 file. We separate the two situations into simple conversion and batch conversion.

**3.1: Simple Conversion**

A single acquisition condition contains multiple MS2 files.

**3.2: Batch Conversion**

Every acquisition condition includes a single MS2 file.

**Part 4: New MS2 Loading**

Each MS/MS spectrum in newly generated .mgf files contains a specific Tissue\_ID and Align\_ID.

**Part 5: Seed Metabolites MS2 Input**

Once the class of functional metabolites is selected, the .mgf MS2 file of seed metabolites will be imported, which is from the seed metabolites LC-MS/MS database, containing the predicted retention time.

**Part 6: Seed Metabolites Identification and Marking**

Seed metabolites from non-targeted data are identified and marked based on the seed metabolites LC-MS/MS database.

**Part 7: MS2 Similarity and Molecular Network Generation**

A heterogeneous molecular network of seed metabolites and non-targeted data is constructed. The MS/MS similarity algorithm used here refers to gnpsalignment (Wang M, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828-837 (2016). https://www.nature.com/articles/nbt.3597). There are many MS2 similarity algorithms currently available (https://github.com/mwang87/GNPS\_SpectralSimilarityHub). Users could select them according to requirements. However, the computing power of personal computers is limited. It is recommended to complete the molecular network construction on the powerful GNPS (https://gnps.ucsd.edu/) platform.

**Part 8: Candidates Features Screening**

The features are retained existing in the same molecular cluster with the seed metabolites, regardless of the seed metabolites from databases or from non-targeted data.

**Part 9: Nodes and Edges Annotation**

For HCAAs (hydroxycinnamic acid amides), nodes are searched for characteristic ions and neutral losses from given list. For CGFs (C-glycosylflavones), nodes are searched for motifs and glycosyls from given lists. The positive searching results for each node will be returned to a node table. The edges of the three classes of antiherbivore metabolites are searched for a given list of enzyme-catalyzed reactions. The positive searching results were recorded to an edge table for each edge. If you search for compounds other than the three types of antiherbivore metabolites, you can modify the corresponding knowledge list.

**9.1: Nodes and Edges Annotation of HCAAs**

**9.2 Nodes and Edges Annotation of CGFs**

**9.3 Edges Annotation of BOAs (benzoxazinoids)**

**Part 10: Redundancy Removal**

The nodes with the same Align\_ID will be merged, which refers to redundant nodes in the same acquisition condition but different tissues. There are some small changes for the connection of network clusters. But it does not affect the final result.