MassBank
Past, Present and Future

Masanori Arita (NIG and RIKEN CSRS)

On behalf of MassBank Developer Team in Japan

MassBank History

2006: JST-BIRD Project was initiated by Prof. Takaaki Nishioka at Keio University, Tsuruoka.

PI: Takaaki Nishioka (Keio U)

Co-PI: Shigehiko Kanaya (NAIST), Masanori Arita (U Tokyo)

With Dr. Hisayuki Horai, Mr. Yoshito Nihei, and Mr. Tasuku Ikeda.

2008: IPB Halle (S Neumann) and others joined the consortium.

<u>At that time:</u>

1. There were no free, public database for ESI mass spectra.
2. Massbank was designed as ‘data’-distributed style; each researcher is expected to publish data locally.

MassBank

History continued

2008: Record editor and download service

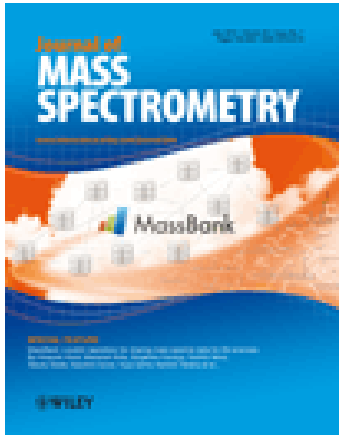
2009: JST-BIRD Project was extended for 3 more years.

BIRD project also established KNApSAcK DB and Metabolomics.JP.
Mr. Yuya Ojima started fragment annotation.

SourceForge site created.

2010: Web API and English manuals ready.
19 groups, >30,000 spectra, 8 data servers

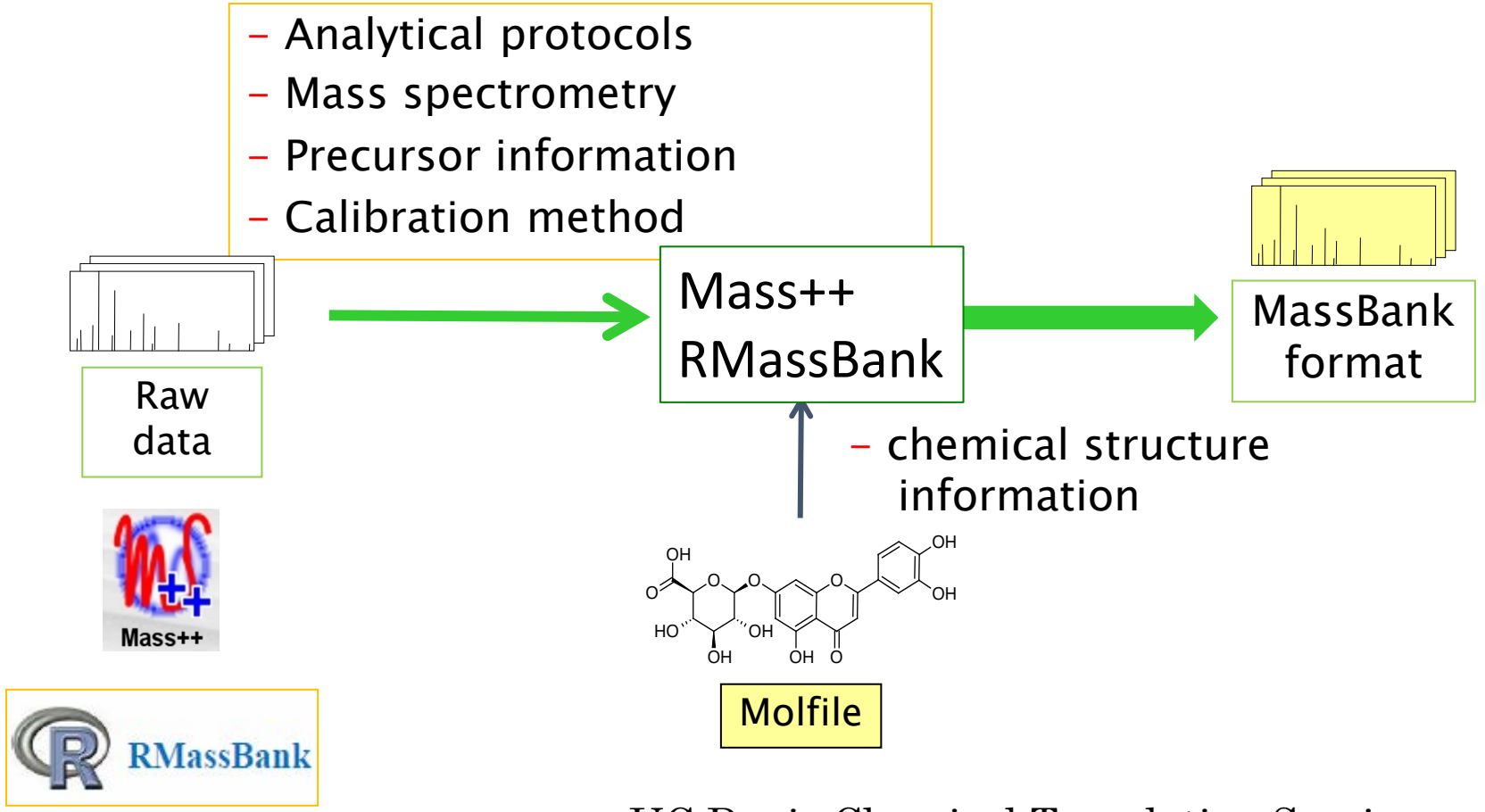
2011: Fragmentation library on wiki



Cover Article:
J. Mass Spectrom.
45(7), 2010

MassBank

Automatic Preparation of MassBank Records



UC Davis Chemical Translation Service
NCI Chemical Identifier Resolver

Format conversion software

Mass++ by Shimadzu project

(AB Sciex, Agilent, Bruker, Shimadzu, Thermo, Waters)

http://www.first-ms3d.jp/english/achievement/software

Reifycs ABF converter

(AB Sciex, Agilent, Bruker, Shimadzu, Thermo)

http://www.reifycs.com/english/AbfConverter/

MassBank

Achievements

- MassBank record … MSP version for MS/MS is available. We will upload the library on the MassBank site.
- API & Search tools … Java-applet was the initial choice
- Wiki-based annotations … details later

MassBank

MassBank Format

Dr. Horai is software engineer.

COMPOUND Section
- License and metabolite info.
- Listing of DB links

Compound identifiers and database links are supplied manually by the Tsuruoka staff.

METHOD Section
- Analytical methods

Methods are often unique to each laboratory.

Details are available.

SPECTRUM Section
- Precursor information
- Actual spectrum

Similar to MSP format at NIST, but precursor info is valuable.

MassBank

Record sample

<Good>

- Easy to understand and to edit.
- Detailed metadata.
- Formal specification.

<Bad>

- No systematic way to describe MS^n information.
- Redundant and hard to comply with.

MassBank

```
ACCESSION: KNA00365
RECORD_TITLE: Phosphoenolpyruvic acid; LC-ESI-ITFT; MS2; m/z:132.10; POS
DATE: 2011.08.03 (Created 2009.11.18)
AUTHORS: Takahashi H, Kanaya S, Ogasawara N, Graduate School of Information
LICENSE: CC BY-SA

CH$NAME: Phosphoenolpyruvate
CH$NAME: Phosphoenolpyruvic acid
CH$NAME: PEP
CH$COMPOUND_CLASS: Natural Product
CH$FORMULA: C3H5O6P
CH$EXACT_MASS: 167.98237
CH$SMILES: C=C(C(O)=O)OP(O)(O)=O
CH$IUPAC: InChI=1S/C3H5O6P/c1-2(3(4)5)9-10(6,7)8/h1H2,(H,4,5)(H2,6,7,8)
CH$LINK: CAS 138-08-9
CH$LINK: CHEBI 18021
CH$LINK: KEGG C00074
CH$LINK: KNAPSACK C00000798
CH$LINK: NIKKAJI J9.706C
CH$LINK: PUBCHEM 3374

AC$INSTRUMENT: LTQ Orbitrap XL, Thermo Scientfic
AC$INSTRUMENT_TYPE: LC-ESI-ITFT
AC$MASS_SPECTROMETRY: MS_TYPE MS2
AC$MASS_SPECTROMETRY: ION_MODE POSITIVE
AC$MASS_SPECTROMETRY: COLLISION_ENERGY 35eV
AC$MASS_SPECTROMETRY: IONIZATION ESI
AC$CHROMATOGRAPHY: COLUMN_NAME TOSOH TSKgel ODS-100V  5um Part no. 21456
AC$CHROMATOGRAPHY: FLOW_GRADIENT 0min:3%, 45min:97%, 50min:97%, 50.1:3%, 57
AC$CHROMATOGRAPHY: FLOW_RATE 0.5 ml/min
AC$CHROMATOGRAPHY: RETENTION_TIME 6.996967 min
AC$CHROMATOGRAPHY: SOLVENT 0.1%formate-water / 0.1%formate-acetonitrile

MS$FOCUSED_ION: BASE_PEAK 86.010551
MS$FOCUSED_ION: PRECURSOR_M/Z 132.10

PK$NUM_PEAK: 2
PK$PEAK: m/z int. rel.int.
```

Search Tools

Matching method was implemented after the GC/MS dot-product.

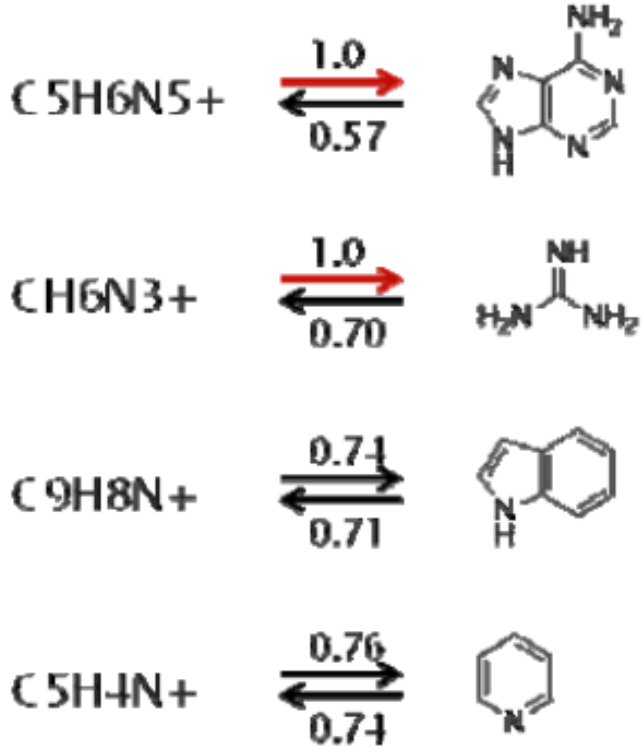
Problems:

- Fewer peaks and the dependence on collision-energy.

→Spectral merging and the use of characteristic peaks
(This is why we emphasize peak annotations.)

- Mass accuracy.

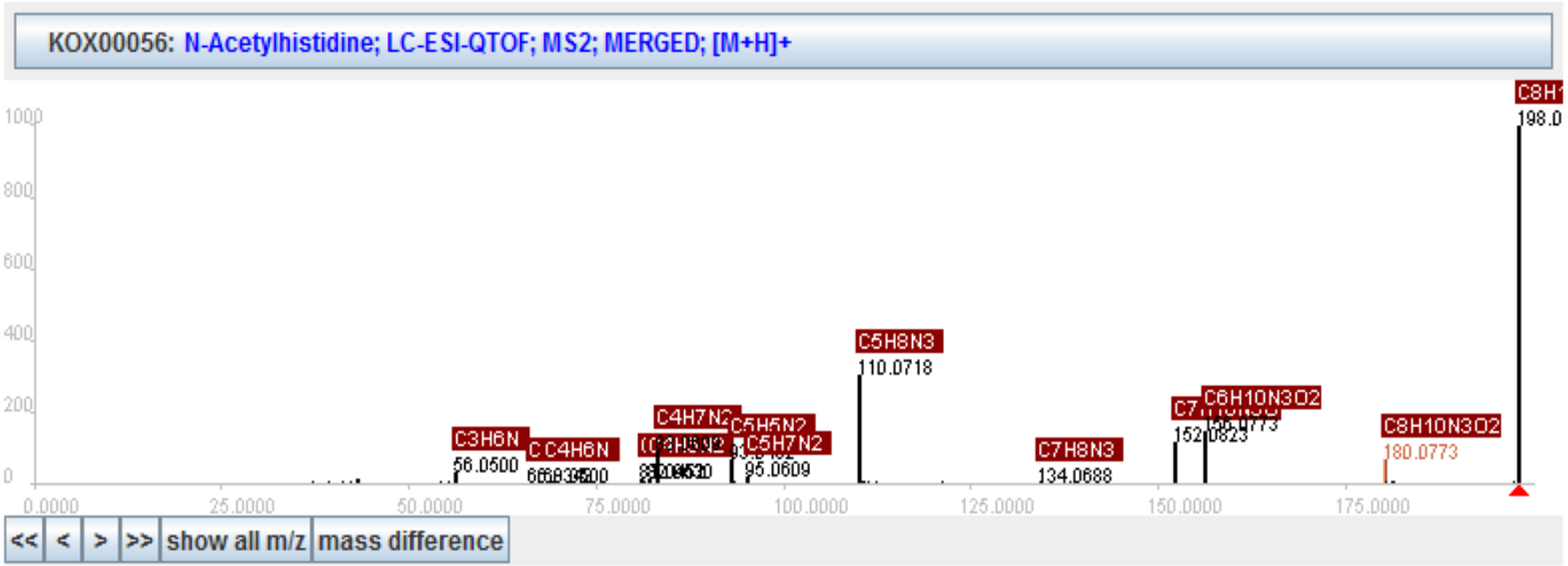
→ Annotation by chemical formulas



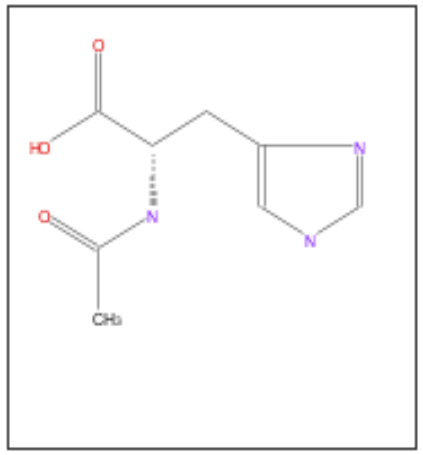
MassBank

“Chemically Accurate” ESI-MS^2 Data

Home | Spectrum | Quick | Peak | Substructure | Identification | Browser | Batch | Browse | Index | MassBank ID: Go



Formula: C8H11N3O3
Exact Mass: 197.08004



Annotated Peak	C3H6N (56.05)	C4H4N (66.034)	C4H6N (68.05)	C4H5N2 (81.045)	C4H6N2 (82.053)	C4H7N2 (83.061)	C5H5N2 (93.045)	C5H7N2 (95.061)	C5H8N3 (110.072)	C7H8N3 (134.072)	C7H10N3O (152.082)	C6H10N3O2 (156.077)	C8H10N3O2 (180.077)	C8H12N3O3 (198.088)
KOX00056	*	*	*	*	*	*	*	*	*	*	*	*	*	*

MassBank

Web API and interface

- SOAP vs REST: We need to offer spectral search queries.
- BATCH search service: If possible, why not?

The SOAP API is usually hard to use (SOAP + XML), but a Perl wrapper is available.

We insist on chemical annotation!

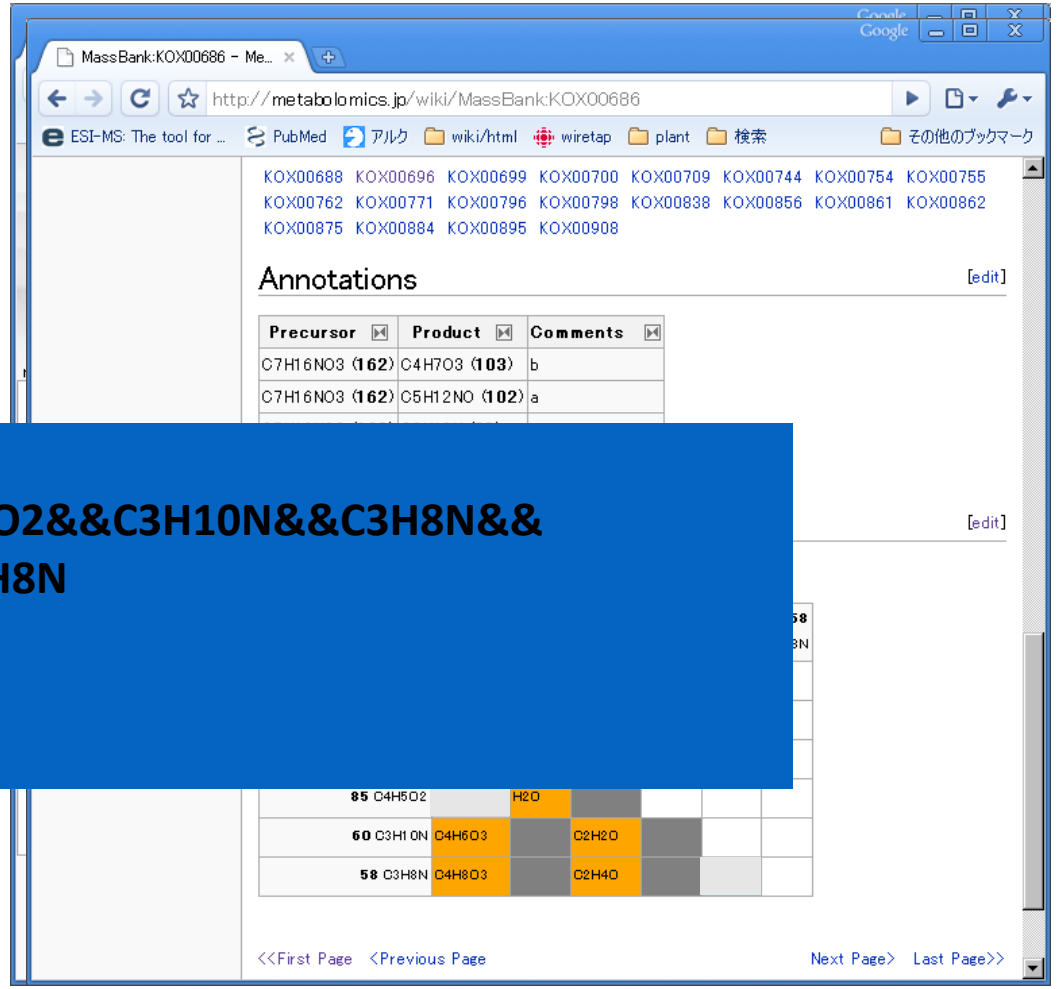
```
$msbk = &connectMassBank();
@inst = &getInstrumentTypes($msbk);

%rec = &getRecordInof($msbk, $id1, $id2, ...);

%specs = &getPeak($msbk, $id1, $id2, ...);

@res = &searchSpectrum($msbk, {%spec},
          $max, $ion, [@inst], $unit, $tol, $cut);
@res = &searchPeak($msbk, [@mz],
          $max, $ion, [@inst], $cut, $tol);
@res = &searchPeakDiff($msbk, [@diff],
          $max, $ion, [@inst], $cut, $tol);
$jobId = &execBatchJob($msbk,
          {%specset}, $ion, [@inst], $mail);
($stat, $code, $date) = &getJobStatus($msbk,
          $jobId);
@bres = &getJobResults($msbk, $jobId);
```

MassBank

Wikifying annotation

From spectra, we want to obtain fragmentation rules for ESI-MS.

→ Use wiki to collect annotations.




```
{{MassBank/Matrix|data=
&&C7H16NO3&&C4H7O3&&C5H12NO&&C4H5O2&&C3H10N&&C3H8N&&
&&C7H16NO3: C4H7O3  C5H12NO  C3H10N  C3H8N
&&C4H7O3: C4H5O2
&&C5H12NO: C3H10N  C3H8N
}}
```

MassBank

Data Description in Wiki

http://metabolomics.jp/wiki/Index:MassBank

```
{{MassBank/Matrix|data=
&&C7H16NO3&&C4H7O3&&C5H12NO&&C4H5O2&&C3H10N&&C3H8N&&
&&C7H16NO3: C4H7O3 C5H12NO
            C3H10N C3H8N
&&C4H7O3: C4H5O2
&&C5H12NO: C3H10N C3H8N
}}
```

Identified ions in mass spectra (written in molecular formula)

Fragmentation pattern for each ion (detected by considering structural information)

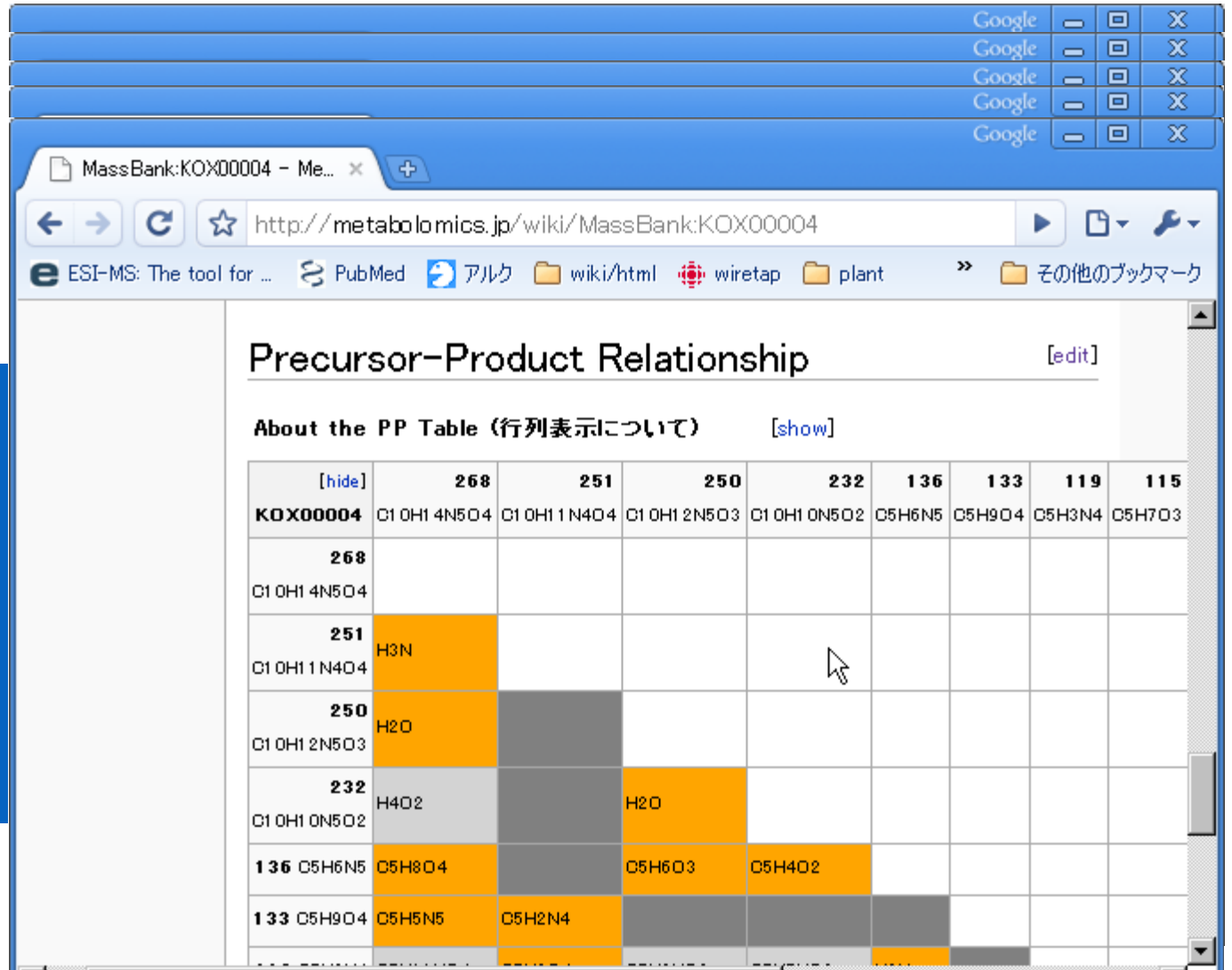
Open development. Open discussion.

MassBank

Example. Annotation of MS

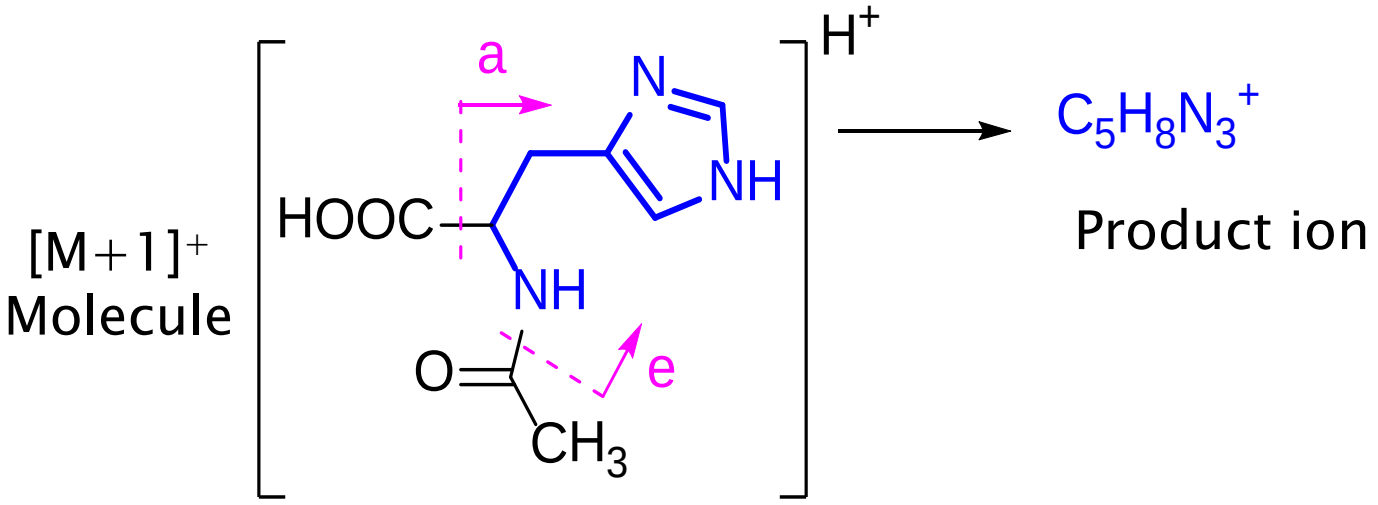
Collaborative edit changes all the related pages automatically.

```
{{MassBank/Matrix|data=
&&C7H16NO3&&C4H7O3&&C5H12NO&&C4H5O2
&&C3H10N&&C3H8N&&
&&C7H16NO3: C4H7O3 C5H12NO      C3H10N
C3H8N
                comments
&&C4H7O3: C4H5O2
                comments
&&C5H12NO: C3H10N C3H8N
                comments
}}
```



MassBank

Relationships between product ions and substructures



Cleavage sites **a** & **e**
are predicted.

http://metabolomics.jp/wiki/Index:MassBank

History continued

2012: JST Togo(Integration)-DB Project succeeded the MassBank.

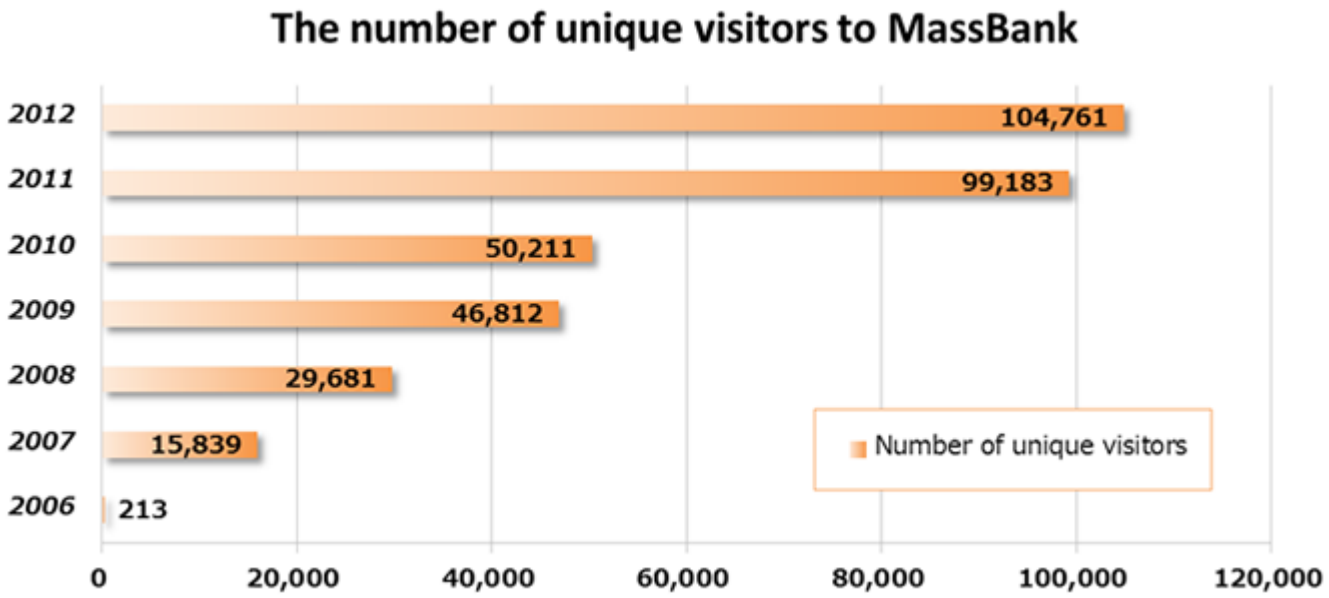
PI: Shigehiko Kanaya (NAIST) co-PI: Arita and Nozomu Sakurai (Kazusa DNA)

Launch of Bio-massbank

Eawag & EU-mirror

2013: Project scale-down

Closing data servers and Tsuruoka office.



MassBank

Bio.massbank.jp

Repository for LC- and GC-MS results from biological samples in collaboration with Kazusa DNA Institute.

Most peaks are “unknowns”.

Metadata are stored in the Metabolonote database.

http://metabolonote.kazusa.or.jp/Main_Page

But needs an overhaul.

MassBank

Future

2014: JST Togo (Integration) DB Project continues for additional 2 years.

PI: Masanori Arita (NIG), Co-PI: Shigehiko Kanaya, Nozomu Sakurai

Tasks

- Replace all applets with JavaScript components.
- Overhaul and refactor the Bio-massbank.

2015: Construction of “model” metabolome site as the alternative to Bio-massbank.

Research Organization of Information and Systems
National Institute of Genetics

MassBank

Policy

- MassBank allows users to freely download and to contribute.
  - CC license and free use.
  - No institutional logos on the website, no representative face.
- International alliance is most important.
  - We share data but not necessarily the search interfaces.
  - MassBank in Japan, Norman MassBank in EU, and MassBank of North America
- Code sharing (GitHub) and international developer team
  - Establish a mailing list and exchange information live.

MassBank

Ongoing

- Retrieving all the source codes and developer items from the Tsuruoka (now NAIST) server.

  We will put everything first on the GitHub (early Oct)

- Create a cloud server that will replace the current NAIST one.
  - Batch API service will be temporarily closed.
  - Replace all Java applets with JavaScript codes
  - 3D spectral comparison will be discontinued.
  - Substructure-search will be discontinued.
  - Query interfaces will be simplified.

MassBank

Year 2015

- We keep the data sharing with EU-MassBank and MONA, and MassBank format needs to change.
  - MS^n consideration is necessary.
  - Method section will move to the Metabolonote DB at Kazusa DNA Institute.
  - Compound section may be discontinued, if we have a ChemSpider link.
  - Easier bulk download option.
  - Curation and merging of spectra, i.e., reduction in the spectral number.
  - Original spectra will be kept on the wiki site.

MassBank

Future

- Collaboration with DDBJ (DNA Data Bank of Japan), which allies with GenBank and ENA (EBI).
- Establishment of “consensus” metabolome information for model species such as Arabidopsis and Yeast.
- Creation of “verified” theoretical spectra for various classes of metabolites.

MassBank

Acknowledgments (Japan side)

MassBank Team:

Takaaki Nishioka, Hisayuki Horai, Yoshito Nihei, Yuya Ojima, Tasuku Ikeda, Rie Matsuzawa

KNApSAcK Team:

Shigehiko Kanaya, Aki Hirai, Hiroki Takahashi, Kenichi Tanaka

MetabolomicsJP Team:

Kazuhiro Suwa, Miwa Yoshimoto

Kazusa DNA Team:

Nozomu Sakurai, Takeshi Ara, Mitsuo Enomoto

MassBank