STSM Scientific Report

Compilation of plant food bioactives' metabolites in the online database PhytoHub

Andreia Leonor Vieira Bento da Silva

***Reference**: COST Action FA1403*

***Host institution**: Centre INRA Auvergne-Rhone-Alpes*

***Period**: 01/03/2016 to 15/04/2016*

Because of interindividual variation in ADME of plant food bioactives, it is essential to develop methods to assess the actual internal exposure of individuals to their different metabolites. Metabolomics based on high resolution mass spectrometry is an approach of choice as it allows the detection of hundreds of plant food bioactive metabolites in human biofluids.

However, currently, about 75% of the detected metabolites cannot be easily identified; because the compound databases are not comprehensive. Therefore, the main objective of this STSM was to enter the known human metabolites of the main plant food bioactives in the database PhytoHub (www.Phytohub.eu), to make the information usable through online query. PhytoHub, initially developed by INRA, is dedicated to dietary phytochemicals and their metabolites and has been designed for metabolomics. Within the WG1, intensive efforts are being made to review the literature and compiling the available data for the different families of plant food bioactives, and provide a list of the human and gut microbiota metabolites of the main representative compounds.

For this STSM, eight families of compounds have been chosen. The metabolites for the main representative compounds (precursors) of each family were collected. In order to add the main precursors and metabolites to the database, the following steps were followed:

1. Interaction with the WG1 compound families' subgroup leaders and presentation of the work in the 3rd WG meeting, in Bucharest, to obtain the list of known metabolites for the representative compounds of each family;
2. Drawing the chemical structures using Marvin Sketch software;
3. Literature search of the relevant information about the metabolites (described in 4.);

4. Entering the relevant information about the metabolites, following the described SOPs (Standard Operating Procedures), namely:
    a. Taxon family, class and subclass of the Precursor;
    b. Metabolite common name;
    c. Metabolite chemical formula;
    d. Synonyms of the metabolite;
    e. Precursor name;
    f. Precursor formula;
    g. 2D structure of each metabolite and respective MolFile;
    h. Species where the metabolite has been found;
    i. Biofluids and/or tissues where the metabolite has been found;
    j. Origins of the metabolite, when known (Microbial, Host Metabolism, or Both);
    k. Biotransformation, metabolism phase and enzymes, when known;
    l. Associated references.
5. Follow-up with the WG1 compound families' subgroup leaders
6. Entering new precursors to the Precursors' Template, following the described SOPs;
7. Interaction with bioinformaticians involved in the updating pipeline of Phytohub.

It was expected that PhytoHub would be enriched by around 160 new metabolites. The objective was overcome [Table 1], since:

1. 316 metabolites' inputs, of which 195 different compounds (sometimes the same metabolite has different precursors), have been added to PhytoHub. Of these 195 compounds, 130 were sufficiently described, and it was possible to draw their molecular structure, and the others have been entered with the information available, without the molecular structure.
2. 190 new precursors have been added, of which 185 have the molecular structure included;
3. Since some precursors are not completely described in the literature, they have been added to a different category: group of precursors (e.g. citrus flavanones), and will later be updated to PhytoHub ("Compounds for later release", Table 1), when this possibility will be developed. In total, 11 precursors and 110 inputs of metabolites (93 different compounds, 37 of which are not yet described in PhytoHub) will be added.
4. 433 different compounds have been collected (190 precursors, 195 metabolites, 11 precursors for later release and 37 metabolites for later release), of which 354 molecules have been drawn.

Table 1. *Inputs added to PhytoHub.*

FAMILY GROUP	NUMBER OF INPUTS (Precursors)		NUMBER OF INPUTS (Metabolites)	
ANTHOCYANINS	Precursors added to PhytoHub	24	Metabolites added to PhytoHub	51
	Precursors For Later Release	1	Metabolites for Later Release	17
CAROTENOIDS	Precursors added to PhytoHub	-	Metabolites added to PhytoHub	14
	Precursors For Later Release	-	Metabolites for Later Release	-
ELLAGITANNINS	Precursors added to PhytoHub	21	Metabolites added to PhytoHub	24
	Precursors For Later Release	3	Metabolites for Later Release	40
FLAVAN-3-OLS	Precursors added to PhytoHub	-	Metabolites added to PhytoHub	42
	Precursors For Later Release	2	Metabolites for Later Release	9
FLAVANONES	Precursors added to PhytoHub	-	Metabolites added to PhytoHub	60
	Precursors For Later Release	1	Metabolites for Later Release	22
FLAVONOLS	Precursors added to PhytoHub	5	Metabolites added to PhytoHub	18
	Precursors For Later Release	1	Metabolites for Later Release	16
PHENOLIC ACIDS	Precursors added to PhytoHub	140	Metabolites added to PhytoHub	86
	Precursors For Later Release	2	Metabolites for Later Release	-
PROCYANIDINS	Precursors added to PhytoHub	-	Metabolites added to PhytoHub	21
	Precursors For Later Release	1	Metabolites for Later Release	6
TOTAL	Precursors added to PhytoHub	190	Metabolites added to PhytoHub	316
	Precursors For Later Release	11	Metabolites for Later Release	110
TOTAL	**Number of Total Inputs**		**627**	
TOTAL	**Number of Different Molecules**		**433**	

In total, the database PhytoHub was enriched in 385 new compounds (190 precursors and 195 different metabolites), and more compounds will be added in the future (37 different metabolites and 11 precursors). This is major contribution as PhytoHub contained only 535 compounds in the previous version.

With the enriched version of PhytoHub, the plant food bioactive metabolites will be more easily identified in metabolomics studies, which will be valuable for both the study of the interindividual variation in their bioavailability and for consideration of their health effects.