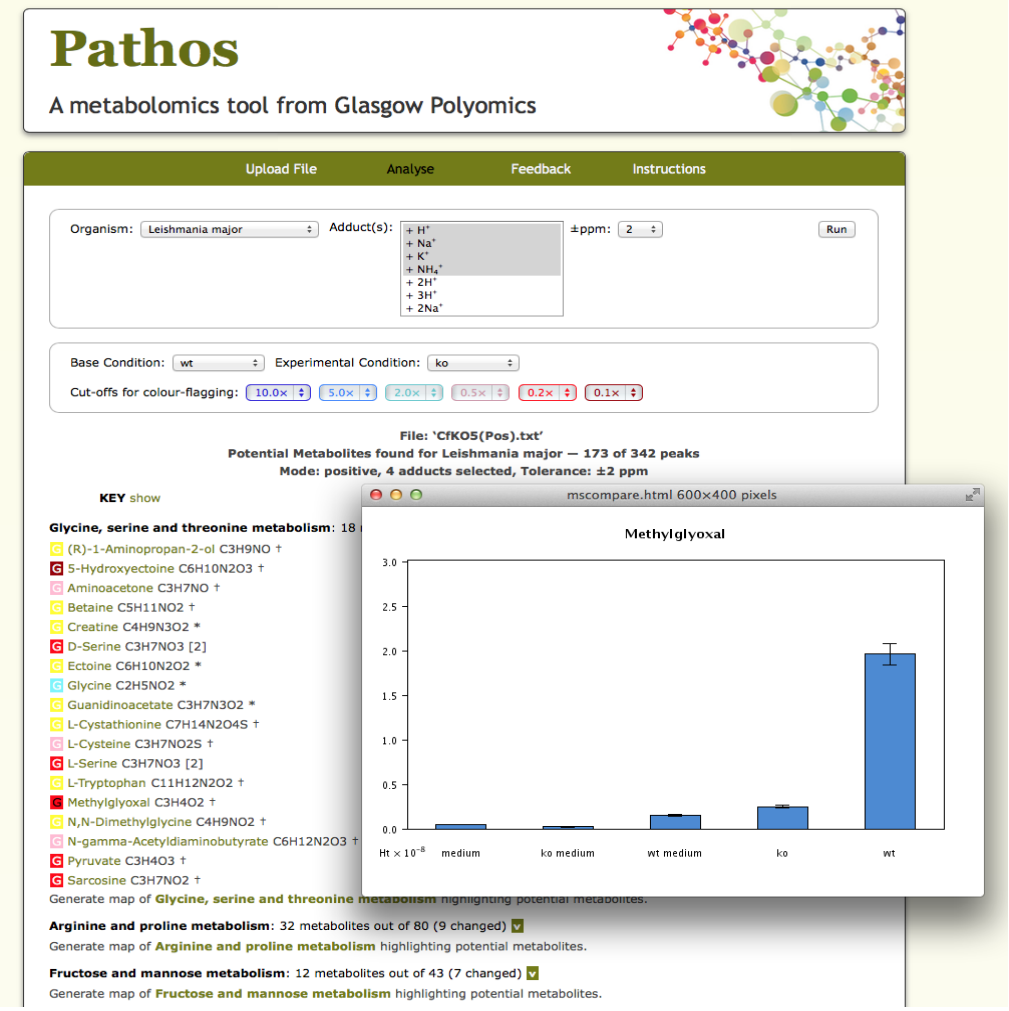
Pathos

User's Manual



revised May 2014

CONTENTS

This is a minor revision to the October 2011 version of the manual to document changes in the sort order of experimental results. Most of the screenshots do not reflect the new design that was implemented at the same time, but are functionally correct.

1. Introduction

Pathos is a web facility that allows one to display metabolites identified by mass spectrometry in the context of the metabolic pathways or broader areas of metabolism in which they occur.

Input files for analysis can be of the following types:

- Lists of M/z values of peaks from mass spectrometric analysis,
- Lists of *Kegg* compound IDs for metabolites already identified,
- Lists of *MetaCyc* compound IDs for metabolites already identified.

If quantitative data from comparison of the abundance of metabolites in different experimental conditions are available, they can also be analysed and displayed.

Output options are as follows:

- Text listings of pathways (according to *Kegg* maps) with identified metabolites, colour coded by degree of experimental change, where relevant,
- *Kegg* pathway maps with identified metabolites highlighted, and colour coded by degree of experimental change, where relevant,
- Bar charts of experimental changes for particular metabolites, where relevant.

URL

The url for *Pathos* is:

http://motif.gla.ac.uk/Pathos/index.html

Requirements

Pathos can be used with a modern standards-compliant web browser such as *Safari, Chrome, Firefox* or *Opera*, provided that JavaScript is *not* disabled. Unfortunately the manner in which *Microsoft Internet Explorer* handles JavaScript makes it impossible to recommend that web browser for use with *Pathos*.

Acknowledgements

The construction of this web application was guided by the requirements of Michael Barrett and colleagues in the Wellcome Trust Centre for Molecular Parasitology, and Mark Burgess and Darren Creek in the Scottish Metabolomics Facility, at the University of Glasgow. It grew out of work performed by Hani Ajahdali, porting part of David Wildridge's desktop *MS Access* application to the web.

Most of the data used in the database underlying *Pathos* were downloaded from the *Kegg* public ftp site, and the metabolic pathway maps are generated using *Kegg* web services. (The *Kegg* website is http://www.genome.jp/kegg/.)

Citation in Publications

In any publication in which you mention *Pathos*, please cite the following paper:
Leader, D.P., Burgess, K., Creek, D. and Barrett, M.P. (2011). Pathos: A web facility that uses metabolic maps to display experimental changes in metabolites identified by mass spectrometry. Rapid Communications in Mass Spectrometry, 25, 3422–3426.

Contacting the Author

Bug reports, feedback and feature requests are welcomed. These may be sent to the author directly (david.leader@glasgow.ac.uk) or through the feedback form on the website.

2. Input File Formats

Input files for *Pathos* require a different format depending on the type of data. The first line informs the application of the format of the data that follows on subsequent lines, and, where experimental data are involved, should be inserted into the text file rather than in an *Excel* spreadsheet of data.

In all cases files should consist of plain tab-separated text. If saving from a *Microsoft Excel* spreadsheet, for example, select 'Tab Delimited Text' as the output format (*not* CSV or UTF-16).

I. Simple M/z Input

(i) The first line should consist of a zero ('0') preceded by a character indicating the mode: 'P' for +ve mode, 'N' for –ve mode, or 'U' for 'neutral' mode (i.e. +ve or –ve mode masses corrected for one proton).

(ii) Each data line should contain a single M/z value.

(iii) The first few lines of an example file are shown opposite. (The file — 'simplemz.txt' — can be found in a folder of examples that can be downloaded from the website.)

```
P0
199.1692421
199.1691189
384.1444594
314.2690337
241.0707449
259.0924694
```

II. Simple Kegg ID Input

(i) The first line should consist solely of an upper-case C followed by a zero ('C0').

(ii) Each data line should contain a single Kegg ID.

(iii) The first few lines of an example file are shown opposite. (The file — 'keggid.txt' — is also in the folder of examples.)

```
C0
C00019
C00022
C00025
C00062
C00064
C00077
```

III. Simple MetaCyc UID Input

(i) The first line should consist only of an upper-case M followed by a zero ('M0').

(ii) Each data line should contain a single MetaCyc UID.

(iii) The first few lines of an example file are shown below. (The file — 'metacycuid.txt' — is also in the folder of examples.)

```
M0
S-ADENOSYLMETHIONINE
PYRUVATE
GLT
ARG
GLN
L-ORNITHINE
```

IV. M/z Input with Experimental values (no standard errors)

(i) The first line should consist of the following tab-separated items:

modeNo. base:expt name1 name2 name3 (etc.)

where:

mode = 'P', 'N', or 'U' for positive, negative or 'neutral' mode.

No. = number of conditions in the experiment

base = position number of condition which is the base (i.e. 100%) for comparison

expt = position number of the experimental condition of primary interest, and for which the percentage change will be flagged in colour

name1 etc. = names of conditions (for labelling bar charts)

N.B. It is unwise to try to insert this line in *Excel* as it will mangle the term with the colon.

(ii) Each data line should contain a single M/z value followed by the tab-separated experimental values. i.e.
M/z. condition1 condition 2 condition3 (etc.)

(iii) The first few lines of an example file are shown below.
The identifier line indicates positive mode mass spectrometry for two conditions, the first (wt) being the base and the second (mutant) being the experimental one of interest.

P2	1:2	wt	mutant
116.0705562		1450	175000
229.1180895		1720	7770
308.0907387		25000	92600
189.1231812		400000	1430000
148.0603455		508000	1510000
130.0497478		17000	28200

V. M/z Input with Experimental values and Standard Errors

(i) The format of the first line is exactly as in IV, above.

(ii) Each data line (tab-separated) should contain a single M/z value followed by all the experimental values and then all the standard errors, in the same order. i.e.
M/z. condition1 condition 2 condition3 (etc.) SE1 SE2 SE3 (etc.)

(iii) The first few lines of an example file, which contains the same experimental values as in IV, are shown below. In this example on the first data line the normal value is 1450 ± 1560, and the value for the drug condition is 175000 ± 26000. (The file — 'exptmz.txt' — can be found in a folder of examples that can be downloaded from the website.)

P2	1:2	wt	mutant		
116.0705562		1450	175000	1560	26000
229.1180895		1720	7770	1550	2190
308.0907387		25000	92600	13600	45200
189.1231812		400000	1430000	133000	348000
148.0603455		508000	1510000	165000	330000
130.0497478		17000	28200	8130	13600

VI. Kegg ID Input with Experimental values

(i) The first line should consist of the following tab-separated items:
CNo. base:expt name1 name2 name3 (etc.)
where:
CNo. = Upper-case C followed by number of conditions (e.g. C3)
and everything else is exactly as for M/z experimental input..

(ii) Each data line (tab-separated) should contain a single Kegg ID followed by all the experimental values and then all the standard errors (if present) in the same order. i.e.
KeggID condition1 condition 2 condition3 (etc.) SE1 SE2 SE3 (etc.)

(iii) The first few lines of the previous example, but with Kegg rather than M/z data, might be:

C2	1:2	wt	mutant		
C00022		1450	175000	1560	26000
C00025		1720	7770	1550	2190

VI. MetaCyc UID Input with Experimental values

(i) The first line should consist of the following tab-separated items:
MNo. base:expt name1 name2 name3 (etc.)
where:
MNo. = Upper-case M followed by number of conditions (e.g. M3)
and everything else is exactly as for M/z experimental input.

(ii) Each data line (tab-separated) should contain a single MetaCyc UID followed by all the experimental values and then all the standard errors (if present) in the same order. i.e.
MetaCycUID condition1 condition 2 condition3 SE1 SE2 SE3

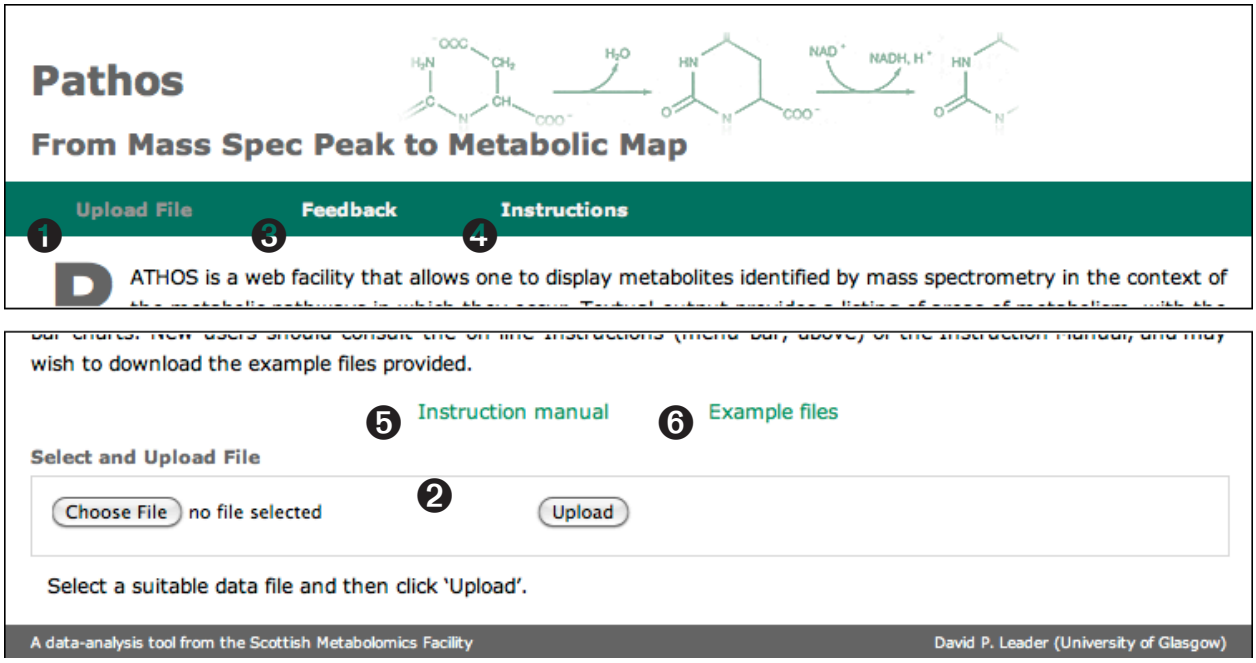
(iii) The first few lines of the previous example, but with MetaCyc rather than Kegg data, might be:

M2	**1:2**	**wt**	**mutant**		
PYRUVATE		**1450**	**175000**	**1560**	**26000**
GLT		**1720**	**7770**	**1550**	**2190**

3. Home Page & File Upload

On connecting to the *Pathos* home page (below), one has access to the single functionality of file upload. This is indicated by the dimmed 'Upload File' item in the menu bar (1) and the 'Select and Upload File' control buttons (2).

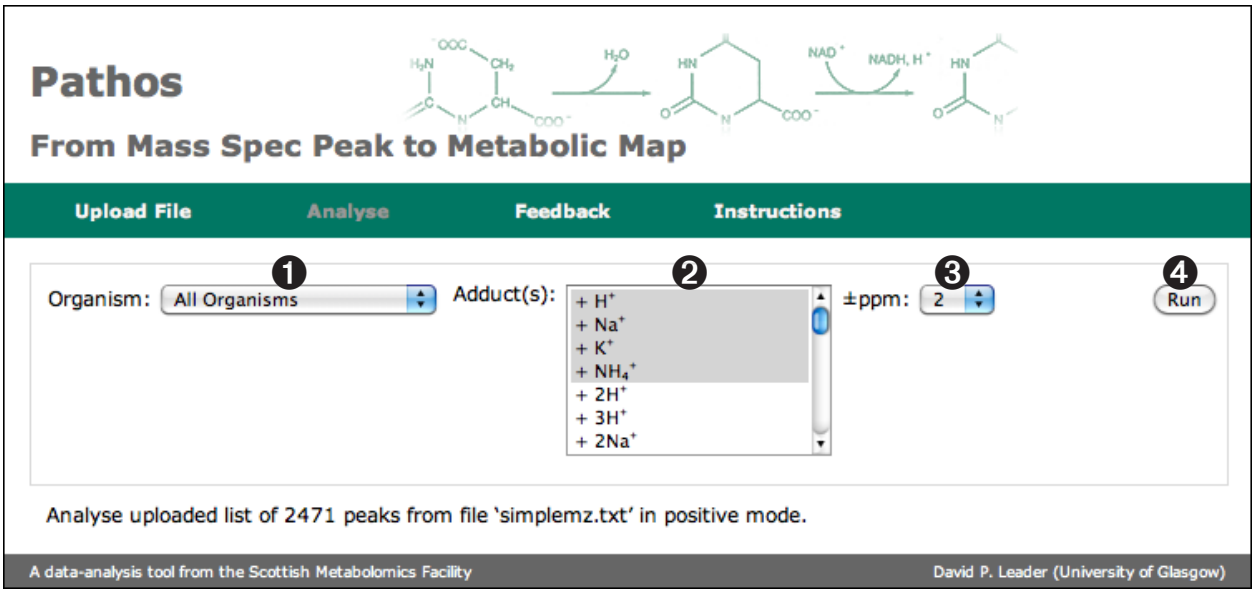
The menu bar also contains links to a 'Feedback' page (3) and on-line instructions that will appear in a small pop-up window (4). Below the introductory rubric are links that allow one to download a copy of this 'Instruction Manual' (5) and 'Example Files' (6).



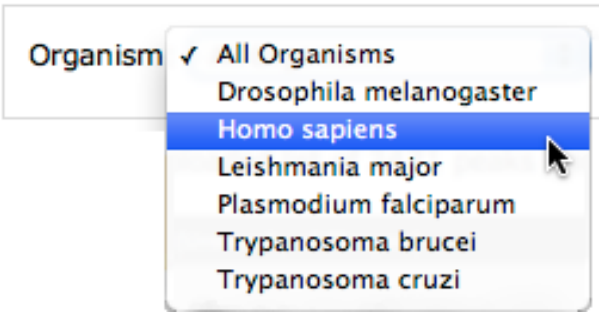
File upload (2) follows standard web-form procedure. The input file is located and selected in one's file system after pressing the 'Choose File' button (for *Safari* and *Chrome* — the button is labelled 'Browse' in *Firefox* and 'Choose' in *Opera*), after which one presses the 'Upload' button. One should then be taken to a new page with confirmation that the file has been uploaded.

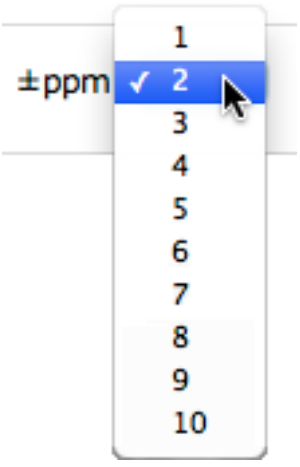
4. Initiating Analysis of Simple M/z Data

On successfully uploading an M/z data file (such as the example provided) one is taken to the analysis page shown below. The menu bar is similar to that on the home page, with the single addition of an 'Upload File' item, for use if one wishes to analyse another file. Below this is a panel with up to three options (1–3 below) which should be set before clicking the associated 'Run' button (4).



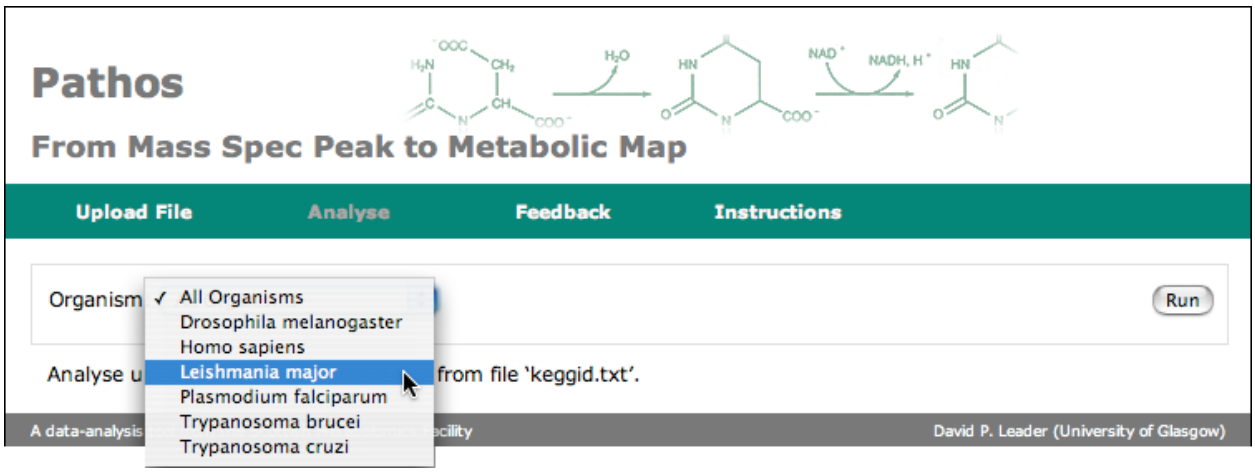
1. The 'Organism' drop-down menu allows one, if one wishes, to restrict the results of the analysis to metabolic pathways known to be present in a given organism, provided that it is listed. The current list is of organisms relevant to local users. Requests for the addition of other organisms may be made using the Feedback form.
2. The 'Adduct(s)' selection window relates to the adducts associated with metabolites when mass spectrometry is performed in positive or negative mode. (Hence it is absent for 'neutral' mode.) There are 32 adducts available in positive mode and 15 in negative mode, with a suggested 'core' set preselected (*see* Appendix for listings). Adducts may be selected and deselected by control-clicking (Windows) or command-clicking (Mac) in the normal manner.
3. The 'ppm' drop-down menu allows one to specify the stringency with which the masses calculated from the M/z values of the peaks are to be correlated to the exact masses of the metabolites.





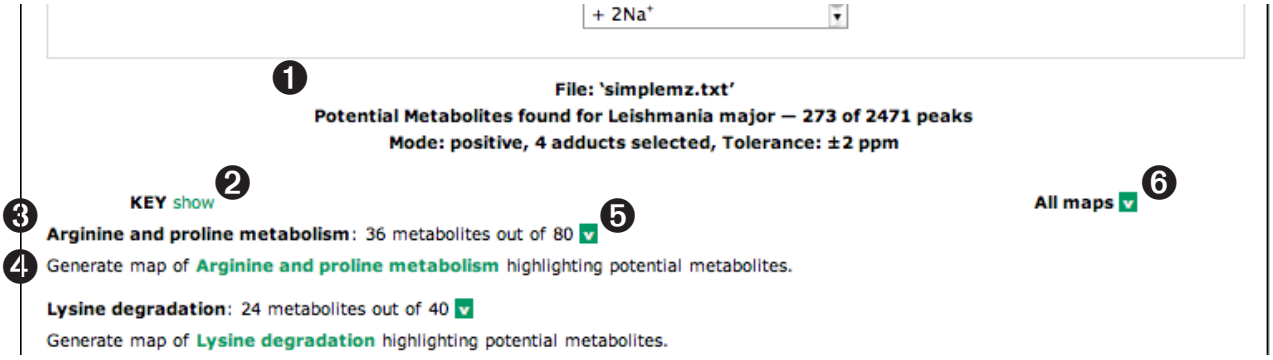
5. Initiating Analysis of Simple Kegg or MetaCyc Data

On successfully uploading a file containing a list of metabolites with Kegg or MetaCyc identifiers one is taken to the analysis page shown below. This is similar to that for M/z data (p. 7), but only has the single drop-down menu, 'Organism', which, as already described, allows one to restrict the results of the analysis to metabolic pathways known to be present in a given organism (provided that it is listed).

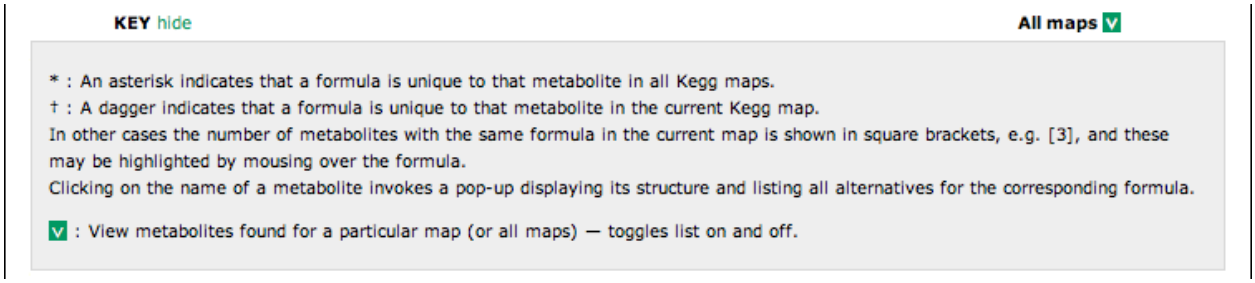


6. Output from Analysis of Simple M/z Data

The initial format of the output page is shown below, and, after repeating the analysis option panel, shows a header summarizing the selections and the number of peaks identified (1), a hide–show link to a key to the symbols used (2), titles of the pathways in which potential metabolites have been identified, sorted by number of metabolites (3), together with links to annotated pathway maps (4). Hide-show toggles allow one to view the metabolites from individual pathways (5) or all pathways (6). A view of the 'opened' key is presented on the opposite page, after which the metabolite listings and the pathway maps are described in more detail.

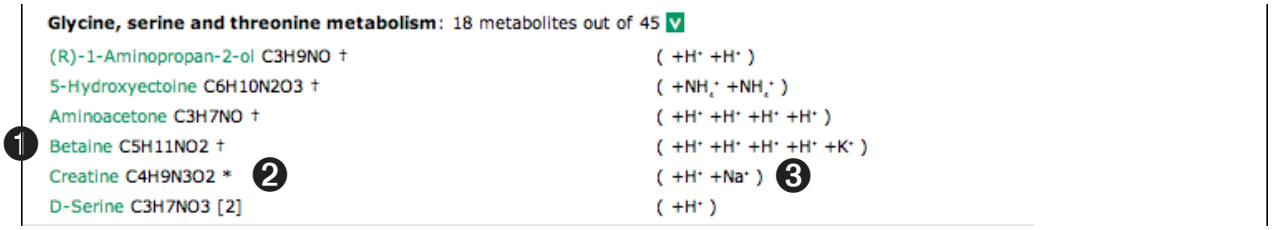


At the foot of the analysis page (one normally has to scroll to reach this) are hide–show links to listings of maps that contain no potential metabolites from the experiment, and peaks that do not correspond to any metabolite in the Kegg maps for the organism. There is also a listing of peaks that have been identified, with the corresponding molecular formula and adduct (if any).

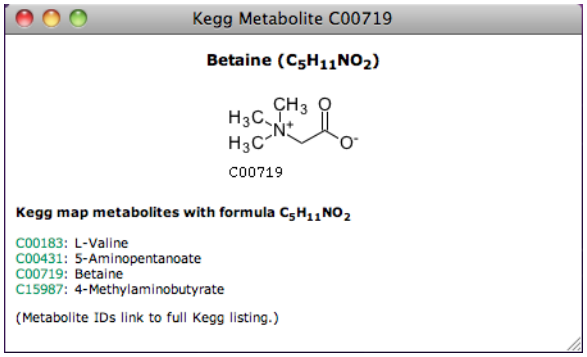


Metabolite Listings

As can be seen from the key above, the list of metabolites for a pathway of interest is viewed by clicking the adjacent V-symbol, highlighted in green. An example, showing part of the output for glycine, serine and threonine metabolism, is presented below.

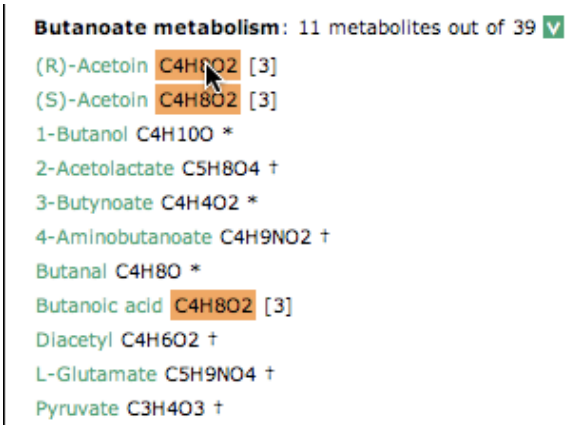


1. The name of the metabolite also provides a link to a small pop-up window showing its structure (opposite) and a listing of any other metabolites with the same formula.
2. The formula of the metabolite is followed by an asterisk or a dagger in cases where it is unique among the metabolites present in all maps or the current map, respectively. Otherwise the number of isomers present in the current map is indicated in square brackets.
3. The adduct(s) of the metabolite detected are listed (when the data are from analysis in positive or negative mode).
   Clicking on the 'V' again closes the listing.



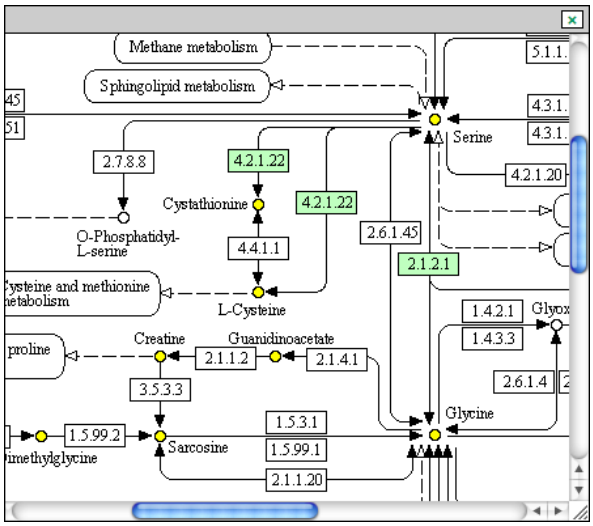
A Note about Isomers

When the input to *Pathos* is M/z data one is faced with the problem that a particular formula often corresponds to more than one metabolite. Where *Pathos* flags with a dagger metabolites that are unique in the current map, one should remember that isomers will be present in other maps. Where the listing indicates a number of isomers in square brackets, these can be highlighted by holding the cursor over a formula (opposite). Again, further isomers may exist in other pathways — the full list is given in the structure pop-up (above).



Pathway Maps

Clicking on the name of the pathway in the 'Generate map…' line (4 on p. 8) sends a request to Kegg (in Kyoto) for a map of the pathway, customized for the potential metabolites that have been identified. This takes on average about 15 seconds to generate and return, which can seem like an age on the modern internet. However the rotating 'busy' icon in the grey 'window' that appears is a genuine indication that the request has not failed, so please be patient.



A section of a typical map is shown, where one can see that potential metabolites are coloured yellow. (The maps are generally large, and it may be necessary to scroll or to resize the window. It can be moved by dragging the top bar.) If one has selected a particular organism one will see that some of the enzymes on the map (indicated by EC numbers) have a green background, indicating that their genes are present in the genome of the organism in question. Those with a white background are thought to be absent.

7. Output from Analysis of Simple Kegg or MetaCyc Data

The output page from the analysis of simple list of Kegg of MetaCyc IDs is in the same format as that for simple M/z data (section 7), but without peak-specific information. An extract from an example is shown below, with the text listing revealed.

Organism: Leishmania major Run

File: 'keggid.txt'
Kegg Maps for Leishmania major with compounds from input list

KEY show **All maps**

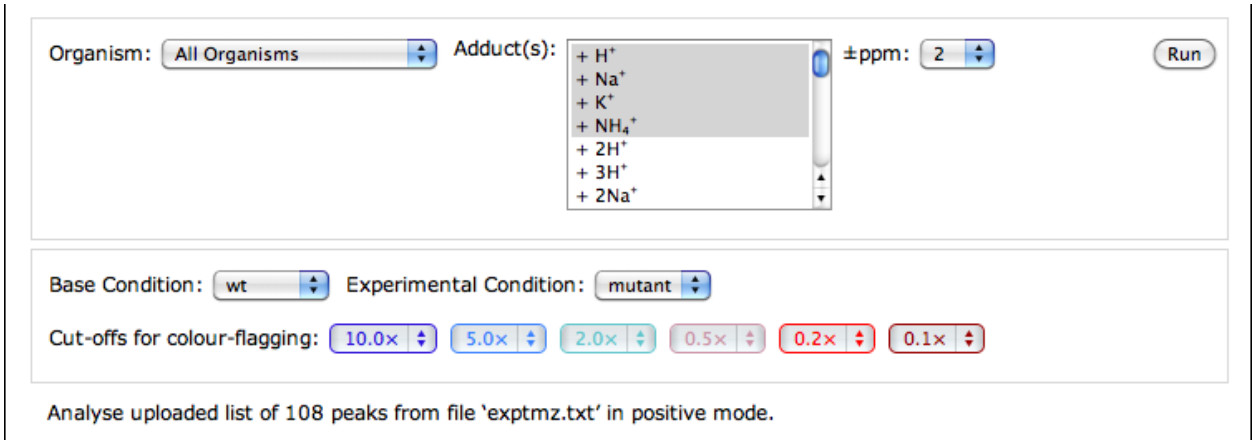
Arginine and proline metabolism: 36 metabolites out of 80
Generate map of **Arginine and proline metabolism** highlighting potential metabolites.

Alanine, aspartate and glutamate metabolism: 5 metabolites out of 24
4-Aminobutanoate C4H9NO2
L-Glutamate C5H9NO4
L-Glutamine C5H10N2O3
N-(L-Arginino)succinate C10H18N4O6
Pyruvate C3H4O3
Generate map of **Alanine, aspartate and glutamate metabolism** highlighting potential metabolites.

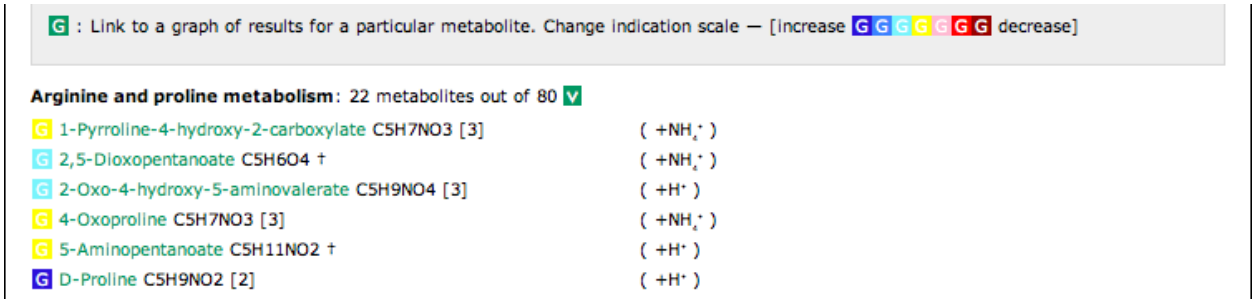
Exactly as with M/z data, clicking on the name of the pathway in the 'Generate map…' line sends a request to Kegg for a map of the pathway on which the metabolites that have been identified are marked in yellow. There is also a link from the name of each metabolite in the text listing, but the pop-up window this invokes contains the Kegg page for that metabolite, rather than the customized structure and ambiguity listing shown on p. 9.

8. Analysis of Experimental Data

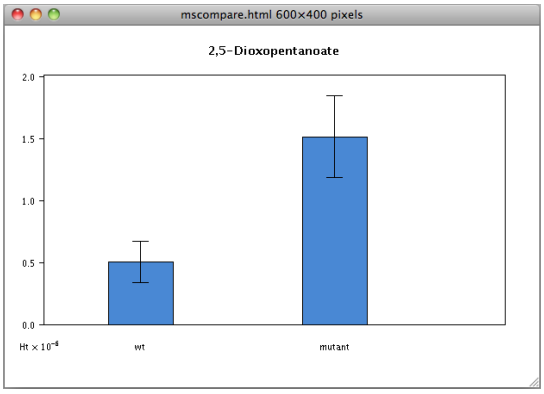
Where the input files contain experimental values for different conditions, additional features are present on the initial analysis page, as shown below. The conditions specified as 'base' and 'experimental' (p. 4) are presented as selections on pull-down menus. In the case of experiments with several different conditions, this allows one to change the reference conditions during a session. These are used to determine the extent of change in the experiment, which is flagged by colour coding (p. 11), the default cut-off values for which are also presented on pull-down menus. These defaults are the ones we use in our own metabolomics facility, but the user is free to change them, if desired.



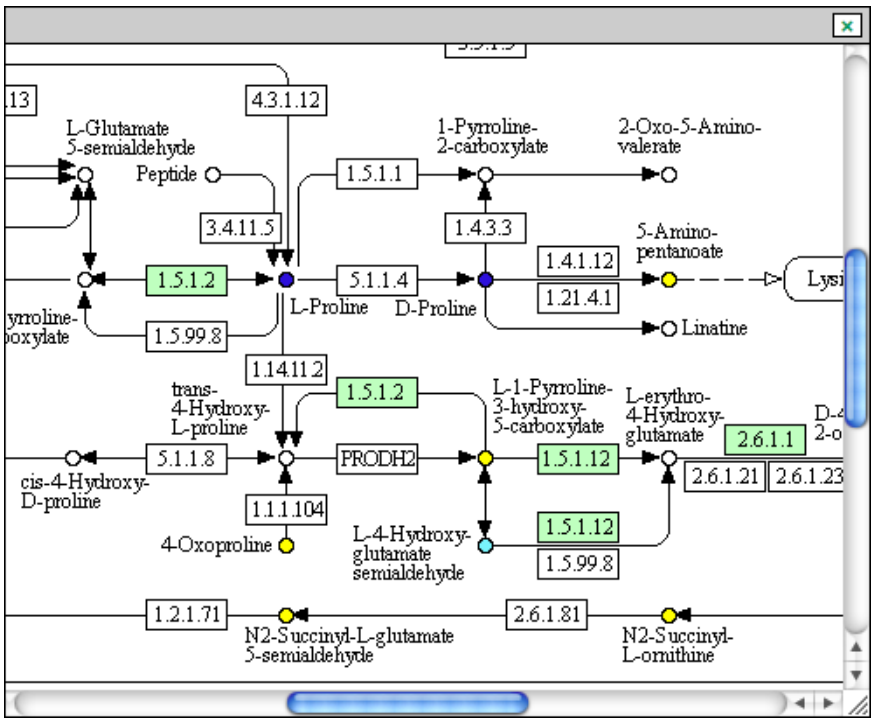
A portion of the initial output page is shown below, with a part of the 'Key' visible in its 'show' state. The names of the metabolites are preceded by a white letter 'G' on a coloured background. This colouring gives an indication of the extent of change in the experimental condition compared to the control. Increases are shown in shades of blue, and decreases in shades of red — the greater the change, the deeper the colour.



Clicking on the 'G' invokes a pop-up graphic window, with the values displayed on a bar chart. One difference of the corresponding display from the analysis of a simple list of M/z values is that only the single adduct from the most intense peak is shown, as this is the one used for the comparison and bar chart. (However, all adducts are listed in the summary of peaks identified.)



The customized Kegg metabolite maps differ from those from simple analyses in that the



colouring of the identified metabolites (or potential metabolites) conforms to the red–blue coding scheme already mentioned, reflecting the extent of experimental change in the metabolite. Furthermore, clicking on the metabolite on the map should also invoke the pop-up bar chart directly.

The output from an analysis of experimental data is sorted in the following manner. The maps in which there are most metabolites showing a change in amount are listed first. Those that contain no metabolites showing changes are then sorted by number of metabolites detected.

9. Printing and Saving

Printing the textual results of an analysis produces output which includes all the data (i.e. including any hidden from view on screen) and with irrelevant web decoration removed. (Use the web browser's preview to see how it will appear.) There is no dedicated facility in *Pathos* to save the textual version of your results to an editable file, but this can be achieved simply by selecting the web page, copying it, and then pasting it into an new word-processor or text-editor page. Technical users who wish to process the results with their own software should save the source of the page, which as strict XHTML is in a suitable format for processing as XML.

The graphic results (pathway maps and bar charts) can be saved by dragging to the desktop in most modern operating systems. Failing this, right-click (control-click on a one-button mouse) on the graphic and select the option in the pop-up menu to save the image to disc. Files may have generic names that users will probably wish to change.

Appendix: Adduct details

The adducts available for positive and negative mode are show below, with M representing the metabolite. Those designated 'core' are preselected as defaults.

Positive

'Core'

$M + H^+$
$M + NH_4^+$
$M + Na^+$
$M + K^+$

Others available

$M + 3H^+$
$M + 2H^+ + Na^+$
$M + H^+ + 2Na^+$
$M + 3Na^+$
$M + 2H^+$
$M + H^+ + NH_4^+$
$M + H^+ + Na^+$
$M + H^+ + K^+$
M + Acetonitrile + $2H^+$
$M + 2Na^+$
M + 2Acetonitrile + $2H^+$
M + 3Acetonitrile + $2H^+$
M + Methanol + H^+
M + Acetonitrile + H^+
$M + 2Na^+ – H^+$
M + Isopropropanol + H^+
M + Acetonitrile + Na^+
$M + 2K^+ – H^+$
M + Dimethylsulphoxide + H^+
M + 2Acetonitrile + H^+
M + Isopropropanol + $Na^+ + H^+$
$2M + H^+$
$2M + NH_4^+$
$2M + Na^+$
$2M + 3H_2O + 2H^+$
$2M + K^+$
2M + Acetonitrile + H^+
2M + Acetonitrile + Na^+

Negative

'Core'

$M – H^+$
$M – H_2O – H^+$

Others available

$M – 3H^+$
$M – 2H^+$
$M + Na^+ – 2H^+$
$M + Cl^–$
$M + K^+ – 2H^+$
M + Formic acid – H^+
M + Acetic acid – H^+
$M + Br^–$
M + Trifluoroacetic acid – H^+
$2M – H^+$
2M + Formic acid – H^+
2M + Acetic acid – H^+
$3M – H^+$