

SUBA Tutorial

This interface is written for and optimized for Mozilla FireFox, a free browser (download ~ 7 Mb at many download sites world wide), it will also work on Microsoft Explorer (6 and above), but is a bit slower in the database searching.

Then you open the SUBA URL at : <http://www.suba.bcs.uwa.edu.au>

You will see the contents of the Search Tab (below)

Search	SQL Search	Results	Format	Columns	Rows
SUBA Search Interface					
Search for Arabidopsis genes where...					
AND OR () CLEAR LAST CLEAR ALL					
found by	GFP analysis in	any location	Add to query [?]		
found in reference	Collections	Add to query [?]			
predicted by	TargetP to be	Mitochondrial	Add to query [?]		
gene information contains	Add to query [?]				
TIGR paralog family ID	>	Add to query [?]			
is on chromosome	1	Add to query [?]			
is in functional category	Cellular Communication / Signal Transduction	Add to query [?]			
loci in list	Add to query [?]				
Query Database					

In the **Top Box** the “Search for Arabidopsis genes where...” is the beginning of the Boolean query sentence that will form your query of the database. To add to and refine this query you use one of the lines of pull down menus below the box.

For example:

If you look at the first line and pull down the first box to select “mass spectrometry” and the location menu to select “plastid”, then add this to the query using the “add to query” button on the far right of this row. The words will appear in the **Top Box** to show you are now asking for “Search for Arabidopsis genes where found by mass spectrometry in plastids”. Further additions can now be made, but first you must select the Boolean linker, the “and” or “or” command buttons (or brackets for more complex queries) are below the query sentence box. If you make a mistake, CLEAR LAST will remove the last part of your query, if you hit CLEAR ALL, you will delete the whole query. If you select “and” and then use the pull down boxes for found in “GFP analysis in plastids” and add to query, then you will be selecting the set confirmed to be in plastids by both GFP and mass spectrometry. If you used “or” instead between these two commands, you would be

selecting the much larger set where GFP, mass spectrometry or both GFP and mass spectrometry had located gene products in plastids.

Much more complicated queries than this can be made using this interface. Note that you can look at the claims of particular research papers using the “found in reference” line, this allows combinations of sets from specific papers (by using “or” linkers) or looking for the set of proteins claimed by multiple papers (using “and” linkers).

The targeting predictor sets can also added to queries, and they can be further refined by using word searches of the descriptions of genes (eg. you can limit your search to genes with the word “dehydrogenase” or “kinase” in the description). Alternatively , or in combination with the above, you can list specific loci (Atxgxxxx.1) in your search criteria using the Bottom Box to rapidly view a specific subset of gene products and the evidence for their location.

Once a query has been built and the database queried, then the **Results Tab** will automatically be displayed.

Search SQL Search Results Format Columns Rows						
Query Results						
		<input type="button" value="Total Hits: 1017"/>		<input type="button" value="Next >>"/>		<input type="button" value="Help"/>
Check	Locus << X >>	Gene Name << X >>	Lab Description << X >>	GFP Sources << X >>	MS Sources << X >>	SP Sources << X >>
<input checked="" type="radio"/>	At1g01080.1	ND	chloroplast RNA-binding protein cp33, putative		Nucleus (14617066) Plastid (12938931,15028209)	
<input checked="" type="radio"/>	At1g01090.1	PDC E1 alpha-1	pyruvate dehydrogenase E1 alpha subunit		Plastid (12938931)	
<input checked="" type="radio"/>	At1g01620.1	PIP1C	aquaporin (plasma membrane intrinsic protein)		Plasma Membrane (15574830) Plastid (15028209)	Plasma Membrane (Q08733)
<input checked="" type="radio"/>	At1g01790.1	ND	K Efflux antiporter KEA1		Plastid (12766230,12938931,15028209)	
<input checked="" type="radio"/>	At1g02140.1	ND	Mago Nashi-like protein	Nucleus (15496452)	Nucleus (15496452) Plastid (15028209)	Nucleus (O23676)
<input checked="" type="radio"/>	At1g02280.1	TOC33	subunit of TOC complex		Plastid (12766230,12938931)	
<input checked="" type="radio"/>	At1g02560.1	ClpP5	ClpP5 subunit of chloroplast Clp protease complex		Plastid (12766230,12938931,11278690)	
<input checked="" type="radio"/>	At1g02980.1	ND	hypothetical protein		Plastid (12938931)	
<input checked="" type="radio"/>	At1g03130.1	ND	putative photosystem I reaction center subunit II precursor		Plastid (11826309,12938931,14729914,15028209,15322131)	Plastid (O9SA56)
<input checked="" type="radio"/>	At1g03475.1		coproporphyrinogen III oxidase	Plastid (11733507)	Plastid (15028209)	
<input checked="" type="radio"/>	At1g03800.1		photosystem II protein family		Plastid (11826309,14729914,15028209,15322131)	

The default is to show six columns: the chromosome locus number (chrlocus), the short gene name (if known) the description, (ie. long gene name), and the reduced information for GFP, MS and SP source data. One hundred results are shown at a time, further pages can be displayed by using the forward and back arrows at the top of the page. The total number of results is noted at the top of the page.

If you click on the **Columns Tab** you can now select other columns to be added to this search.

Search SQL Search Results Format Columns Rows

Column Selection

Add Selected Data Help

(Toggle) Commonly Used

<input checked="" type="checkbox"/>	Locus	Chromosomal locus of the gene
<input checked="" type="checkbox"/>	Gene Name	Gene name
<input checked="" type="checkbox"/>	Lab Description	Lab description
<input checked="" type="checkbox"/>	GFP Sources	Shows all GFP localisations with associated references
<input checked="" type="checkbox"/>	MS Sources	Shows all mass spec. localisations with associated references
<input checked="" type="checkbox"/>	SP Sources	Shows all SwissProt localisations with associated reference
<input type="checkbox"/>	GFP Localisation	Show all localisations determined via GFP
<input type="checkbox"/>	MS Localisation	Show all localisations determined by mass spectrometry
<input type="checkbox"/>	SP Localisation	Show all localisations mentioned in SwissProt
<input type="checkbox"/>	SwissProt ID	The SwissProt identification number

(Toggle) Characteristics

(Toggle) Predicted Location

(Toggle) Predictor Details

(Toggle) Data Sources

These are broken into sections:

“Commonly used” being the default columns

“**Characteristics**” being protein sequence attributes such as location in the genome, number of matching ESTs, descriptions of gene function, MW, pI, GRAVY etc.

“**Prediction**” and “**Prediction details**” provides much more detail on the prediction programme results, most of these programmes provide numeric scores that relate to the likelihood of prediction, all these can be displayed. Refer to the details of each programme in the original articles for interpretation of these numbers. The columns with “x_x_x_location” are the final claim of location by each programme.

“**Data sources**” provides the individual columns for FP (Fluorescent Protein Localisation), MS (Mass spectrometry localization) and SP (Swiss-Prot database information) in different locations that can be useful if downloading larger datasets to excel and wanting to sort by location details.

Individual columns can be added, or alternatively, add them all by ‘mouse over’ of the blue dots to the left of the column descriptions, and “Add Selected Data” on the top tag. If you return to the Results Tab, these columns will be added. If any are now not wanted, these can be hidden by simply clicking on the top of the column, or a return to the Columns Tag can change the columns required.

All the data selected and present in the **Results Tab** can be downloaded using the button on the bottom of the Results Tag to Excel for further analysis by the user.

Each column of the “Data Sources” data provides a link to the data. For GFP and MS this is a PubMed ID number or ISI ID number, these can be easily used to find the articles, these are hyperlinked to the articles online. For the Swiss-Prot columns these are Swiss-Prot IDs, again these are hyperlinked to the Swiss-Prot database.

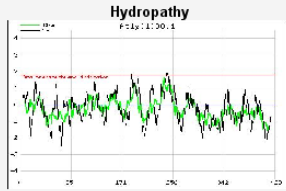
Each loci number (Atxgxxxx) operates as a link to a ‘flat-file’ presentation that provides a variety of information and links that we hope are helpful to the user. These include a display of the different names of the gene, a variety of database lookups to well known

resources online, and a picture presentation of the hydropathy plot so the user can get some idea of the likely orientation of the protein in membranes etc. In addition a series of pull down menus at the base of the page provide information on the physical characteristics of the protein, InterPro domains and targeting prediction programme outputs.

At1g01090.1

Name: PDC E1 alpha-1
BAC: 68408.m00009
Lab Description: pyruvate dehydrogenase E1 alpha subunit
TIGR Description: pyruvate dehydrogenase E1 component alpha subunit, chloroplast identical to pyruvate dehydrogenase E1 alpha subunit GB:AAB86803 GI:2454182 from [Arabidopsis thaliana]; identical to cDNA pyruvate dehydrogenase E1 alpha subunit mRNA, nuclear gene encoding plastid protein GI:2454181
Misc Comment: Plastidial Pyruvate Dehydrogenase E1alpha subunit
Functional Category: Energy

DB Lookups:



Protein Sequence:

```

MATAFAPTKL TATVPLHGSH ENRLLPIRL APPSSFLGST RSLSLRRLNH SNATRRSPVV SVQEVVKEKQ
STNNTSLLIT KEEGLELYED MILGRSFEDM CAQMYRGMK FGFVHLYNGQ EAVSTGFIKL LTKSDSVVST
YRDVHALSK GVSARVMSE LFGKVTGCCR GGGSMHMF8 KEHNMLGGFA FIGEGIPVAT GAAFSSKYRR
EVLKQDCDDV TVAFFGDGTC NNGOFFECLN MAALYKLP11 FVVENNLWAI GMSHLRATSD PEIWKGPAP
GHPGVHVDGM DVLKRVKAVK EAVTRARRGE GPTLVECEY RFRGHSLADP DELRDAAEKA KYAARDPIAA
LKKYLIENKL AKEAELKSIE KKIDELVEEA VEFADASPQP GRSQLELVNF ADPKGFGIGP DGRYRCEDPK
FTEGTAQV

```