
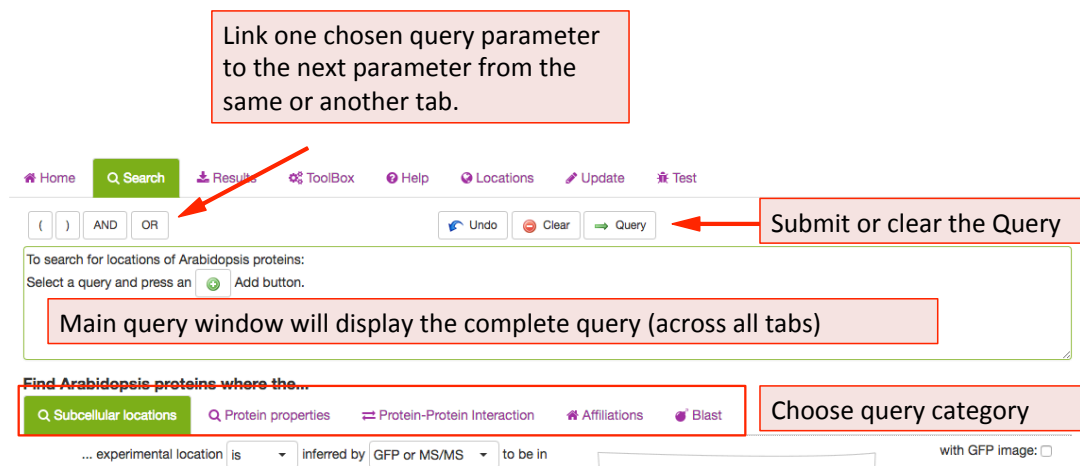


## SUBA4 Search Tab

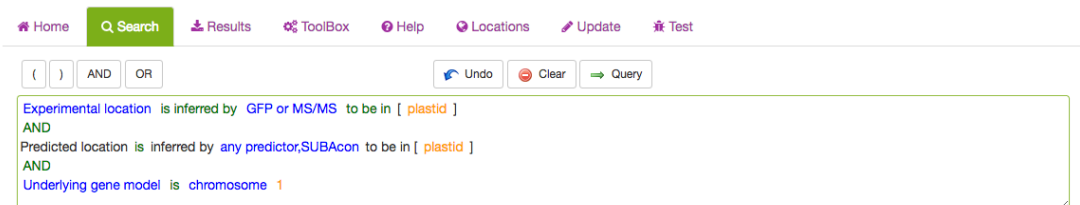
When clicking on the Search function the tab opens the query builder menu. In SUBA4 there are more options and categories of queries you can use to interrogate the SUBA data set. In order to enable the user to find the desired search parameters we have introduced search categories. Each category is stored under a tab. The user can choose a parameter from any tab and add it to the query. The query will appear in the query window at the bottom after clicking the  button. Different search categories can be combined using the AND/OR buttons in between parameters above the query window.

Link one chosen query parameter to the next parameter from the same or another tab.



The screenshot shows the SUBA4 search interface. At the top, there is a navigation bar with links for Home, Search, Results, ToolBox, Help, Locations, Update, and Test. Below this is a query builder menu with buttons for ( ), ), AND, and OR. To the right of these buttons are buttons for Undo, Clear, and Query. A red arrow points from the 'Query' button to a text box that says 'Submit or clear the Query'. Below the menu is a main query window with a text input field and an 'Add button' icon. A red arrow points from the 'Add button' icon to a text box that says 'Main query window will display the complete query (across all tabs)'. Below the main query window is a 'Find Arabidopsis proteins where the...' section with a dropdown menu for 'Subcellular locations' and other categories like 'Protein properties', 'Protein-Protein Interaction', 'Affiliations', and 'Blast'. A red arrow points from the 'Subcellular locations' dropdown to a text box that says 'Choose query category'. Below this section are several dropdown menus for query parameters, such as '... experimental location is inferred by GFP or MS/MS to be in' and 'with GFP image: '.

The full query can be seen in the query window. For details about each category please see search category sections in the tutorial below.



The screenshot shows the SUBA4 search interface with the full query displayed in the query window. The query is: Experimental location is inferred by GFP or MS/MS to be in [ plastid ] AND Predicted location is inferred by any predictor,SUBAcon to be in [ plastid ] AND Underlying gene model is chromosome 1. The query builder menu and navigation bar are also visible at the top.

## Subcellular location Search Tab

This tab contains queries for limiting proteins based on their localisations. SUBA4 has 2 main categories of localisations. You can search for experimental localisations, which is the top query parameter.

The screenshot shows the SUBA4 search interface. At the top, there is a navigation bar with links for Home, Search, Results, ToolBox, Help, Locations, Update, and Test. Below this is a search bar containing the query: "Experimental location is inferred by GFP or MS/MS to be in [ plastid ]". A red box with an arrow points to this query with the text "Check the full query".

Below the search bar, there is a section titled "Find Arabidopsis proteins where the...". Underneath, there are tabs for "Subcellular locations", "Protein properties", "Protein-Protein Interaction", "Affiliations", and "Blast". The "Subcellular locations" tab is selected. The main query area shows "... experimental location is" followed by a dropdown menu set to "is", "inferred by" set to "GFP or MS/MS", and "to be in" set to "plastid". To the right of this is a "with GFP image:" checkbox.

Below the main query area, there are two lists of subcellular locations. The first list includes: nucleus, plastid (checked), Golgi, plasma membrane, cytosol, peroxisome, extracellular, mitochondrion, vacuole, endoplasmic reticulum. A red box with an arrow points to the "plastid" checkbox with the text "2. Choose the parameter". To the right of this list is a cell schematic diagram with a red box around a plus icon and a minus icon, with an arrow pointing to it from the "2. Choose the parameter" box.

The second list is titled "... predicted location is" and includes: nucleus, plastid, Golgi, plasma membrane, cytosol, peroxisome, extracellular, mitochondrion, vacuole, endoplasmic reticulum. A red box with an arrow points to the "plastid" checkbox with the text "3. Add parameter to query". To the right of this list is another cell schematic diagram with a red box around a plus icon and a minus icon, with an arrow pointing to it from the "3. Add parameter to query" box.

The parameters such as in/exclusion of particular compartments and methodology can be chosen from the drop down lists. For choosing a subcellular location, tick any of the box or structures in the cell schematic for the conventional SUBA location categories. For expanded suborganellar categories, click on the >> to expand the list.


The screenshot shows the SUBA4 search interface with expanded subcellular location options. The main query area shows "... experimental location is" followed by a dropdown menu set to "is", "inferred by" set to "GFP or MS/MS", and "to be in" set to "plastid". To the right of this is a "with GFP image:" checkbox.

Below the main query area, there are two lists of subcellular locations. The first list includes: nucleus, plastid (checked), Golgi secretory, plasma membrane secretory, cytosol, matrix, ribosome, cell plate, peroxisome, envelope, matrix, extracellular secretory, apoplast (checked), cell wall, endoplasmic reticulum secretory, membrane, lumen. A red box with an arrow points to the "apoplast" checkbox with the text "Choose from the suborganellar locations". To the right of this list is a cell schematic diagram with a red box around a plus icon and a minus icon, with an arrow pointing to it from the "Choose from the suborganellar locations" box.

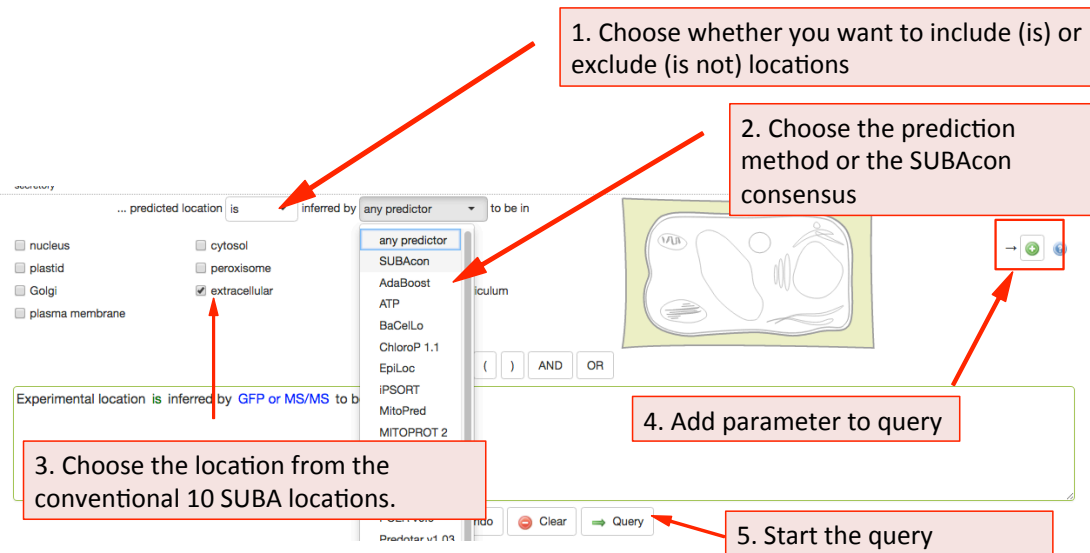
Below the second list, there is a dropdown menu with the following options: GFP or MS/MS, GFP assay, MS/MS assay, endoplasmic reticulum secretory, membrane, lumen. A red box with an arrow points to this dropdown menu with the text "Select >> to access suborganellar locations".

At the bottom, a red box with an arrow points to the "with GFP image:" checkbox with the text "Add the filter to the main query".

This will maximise the localisation view. Click on the desired location. For choosing more than one location, keep ticking more boxes. When choosing the whole compartments (extracellular), this will automatically include the

suborganellar locations (apoplast, cell wall). For only searching for apoplast, untick extracellular and only tick apoplast. Then add your parameter to the query by clicking the  button.

Similarly, to filter for prediction data choose the inclusion and exclusion and the type of predictor from the drop down list. Through this search option, you can also filter by our consensus call output when choosing SUBAcon.



The screenshot shows a search interface with the following elements and callouts:

- 1. Choose whether you want to include (is) or exclude (is not) locations:** Points to the dropdown menu showing "is".
- 2. Choose the prediction method or the SUBAcon consensus:** Points to the "any predictor" dropdown menu, which lists methods like SUBAcon, AdaBoost, ATP, BaCellLo, ChiroP 1.1, EpiLoc, iPSORT, MitoPred, and MITOPROT 2.
- 3. Choose the location from the conventional 10 SUBA locations:** Points to the "extracellular" checkbox, which is checked. Other locations include nucleus, plastid, Golgi, plasma membrane, cytosol, and peroxisome.
- 4. Add parameter to query:** Points to the plus icon button used to add the selected criteria to the query.
- 5. Start the query:** Points to the "Query" button at the bottom of the interface.

Once you have added all desired parameters to the query window you can check your query and submit it using the Query button. Your retrieved results will be automatically displayed in the Results tab when ready.

## Protein Properties Search Parameters

The Protein property tab lets you filter SUBA data for protein annotations, physical properties and chromosomal locations. This tab also contains the option to enter a list of AGIs or text containing AGIs. A new query in SUBA4 lets you also filter for protein aliases, PFAM domains, EC numbers, pathway annotations, structural features and other annotations.

The screenshot shows the SUBA4 Protein Properties Search interface. At the top, there is a navigation bar with links for Home, Search, Results, ToolBox, Help, Locations, Update, and Test. Below this is a search bar with buttons for AND, OR, Undo, Clear, and Query. The main query area contains the following text: "Experimental location is inferred by GFP or MS/MS to be in [plastid] AND Predicted location is inferred by any predictor,SUBAcon to be in [plastid] AND Underlying gene model is chromosome 1". A red box highlights this main query with the text "Main query will appear when adding parameters".

Below the main query is a section titled "Find Arabidopsis proteins where the...". It has tabs for Subcellular locations, Protein properties (selected), Protein-Protein Interaction, Affiliations, and Blast. The Protein properties tab is active, showing a list of search criteria:

- ... protein description contains keyword(s)
- ... physical property of Number of supporting ESTs is equal to
- ... gene model is on chromosome 1
- ... Arabidopsis Gene Initiative (AGI) identifier(s) is in list

A dropdown menu is open for the "Number of supporting ESTs" criterion, showing options: Number of supporting ESTs, Number of supporting fl-cDNAs, Number of amino acids, Molecular weight, Calculated IEP, and GRAVY. A red box highlights the dropdown menu with the text "Choose the query and enter thresholds, parameters or choose from the drop down menus then add to the main query".

There is a text input field labeled "Enter AGI identifier(s) here." with a red box below it containing the text "Enter a list of AGIs".

At the bottom, there is a text input field labeled "Enter AGI xref identifier(s) here." with a red box below it containing the text "Enter PFAM domains, UniProt IDs and other features".

## Protein-Protein Interaction Search Parameters


New in SUBA4: In addition to protein-protein interactions (PPI), there are now experimental localisations from observed protein-protein interactions (PPI) such as Bifocal completion Experimentation. The PPI search tab was included to provide a straightforward access to a number of PPI queries. Besides the conventional search for existing PPI partners by entering AGIs, SUBA4 users can now discover PPI proteins that have been experimentally shown to interact in a particular compartment. At the same time, the drop down menu also allows for the choice of PPI methodology. Other search options for PPI data include the isolation of PPI studies.

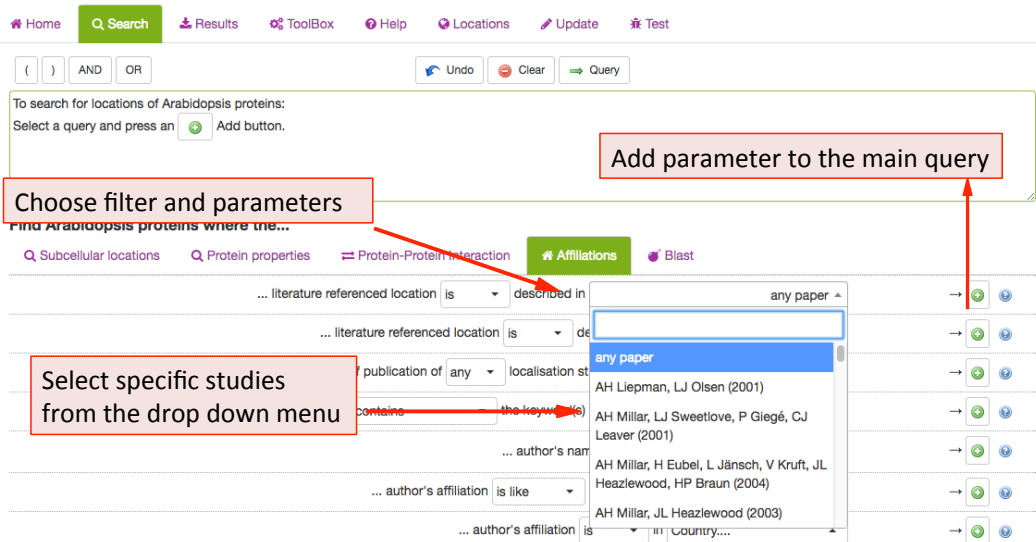
The screenshot shows the SUBA4 search interface with several annotations:


- Search Query:** "protein-protein interactions has been determined by methods [ Bifocal Completion ] AND location was experimentally observed in plastid". A red box highlights this query with the text "Add the filter to the main query".
- Search Tab:** The "Protein-Protein Interaction" tab is selected. A red box highlights the "interact with protein(s) in list" dropdown menu with the text "Enter AGIs and find their interaction partners".
- Subcellular Locations:** The "plastid" checkbox is selected under "experimentally observed in". A red box highlights this section with the text "Search for interactions that have been observed in subcellular locations (experimentally by e.g. BiFC)".
- Methodology:** A dropdown menu for "been determined by method of" is open, showing "Bifocal Completion" as the selected method. A red box highlights this menu with the text "Choose the methodology of interaction as a filter".

The interface also includes a navigation bar with "Home", "Search", "Results", "ToolBox", "Help", "Locations", "Update", and "Test". Below the search bar are buttons for "Undo", "Clear", and "Query". The search results area shows a diagram of a cell with a green plastid highlighted.

## Affiliations Search Parameters


This tab allows SUBA4 users to find experimental localisation data from specific authors, institutions, countries or filter by year of publications. To limit the results a particular study, choose from the publication list in the drop-down menu. The results can also be filtered by any author (not just first author) as well as by year or range of years of publication. For adding any of the parameters to the main query press the  button.



The screenshot shows the SUBA4 search interface with the 'Affiliations' tab selected. The main query area contains the text: 'To search for locations of Arabidopsis proteins: Select a query and press an  Add button.' Below this, there are several search criteria with dropdown menus and 'Add' buttons. A red box highlights the 'Add parameter to the main query' button. Another red box highlights the 'Choose filter and parameters' dropdown menu. A third red box highlights the 'Select specific studies from the drop down menu' dropdown menu. The dropdown menu is open, showing a list of publications: 'any paper', 'AH Liepman, LJ Olsen (2001)', 'AH Millar, LJ Sweetlove, P Giegé, CJ Leaver (2001)', 'AH Millar, H Eubel, L Jänsch, V Kruff, JL Heazlewood, HP Braun (2004)', and 'AH Millar, JL Heazlewood (2003)'. A red arrow points from the 'Select specific studies from the drop down menu' box to the dropdown menu.

Home Search Results ToolBox Help Locations Update Test

( ) AND OR Undo Clear Query

To search for locations of Arabidopsis proteins:  
Select a query and press an  Add button.

**Add parameter to the main query**

**Choose filter and parameters**

Find Arabidopsis proteins where the...

Subcellular locations Protein properties Protein-Protein interaction **Affiliations** Blast

... literature referenced location is described in any paper

... literature referenced location is de

publication of any localisation st

**Select specific studies from the drop down menu**

contains the keyword(s)

... author's nam

... author's affiliation is like

... author's affiliation is Country...

- any paper
- AH Liepman, LJ Olsen (2001)
- AH Millar, LJ Sweetlove, P Giegé, CJ Leaver (2001)
- AH Millar, H Eubel, L Jänsch, V Kruff, JL Heazlewood, HP Braun (2004)
- AH Millar, JL Heazlewood (2003)

SUBA4 allows the search for data by country of origin of the experimental study. Using the drop-down menu shows the countries and number of studies that have contributed to the SUBA4 data set. When using the map for choosing a country, the grey countries indicate a contribution to SUBA4 whereas white countries have not contributed data sets to SUBA4. Green indicates a chosen country.

Find *Arabidopsis* proteins where the...

Subcellular locations  
  Protein properties  
  Protein-Protein Interaction  
  Affiliations  
  Blast

... literature referenced location is described in any paper

... literature referenced location is described in pubmed:

... year of publication of any localisation studies is between and

... publication title or abstract of the localisation study contains the keyword(s)

... author's name is like

author's affiliation is like

... author's affiliation is in Australia

author's affiliation is in:

Choose a specific country from the drop-down menu or from the map to limit the results to data generated in a specific country

| Country        | Total Studies |
|----------------|---------------|
| Argentina      | total 6       |
| Australia      | total 99      |
| Austria        | total 23      |
| Belgium        | total 59      |
| Brazil         | total 18      |
| Canada         | total 74      |
| Chile          | total 7       |
| China          | total 183     |
| Czech Republic | total 21      |
| Denmark        | total 40      |
| Finland        | total 4       |
| France         | total 226     |
| Germany        | total 392     |
| Greece         | total 6       |
| Hong Kong      | total 25      |
| Hungary        | total 5       |
| India          | total 3       |
| Ireland        | total 3       |
| Israel         | total 35      |
| Italy          | total 38      |
| Japan          | total 391     |
| Lithuania      | total 1       |

This work is licensed under a Creative Commons Attribution 4.0 International License. © 2005-2016 Harvev Millar (harvev.millar@lati.uwa.edu.au)

## BLAST Search Parameters

The BLAST tab contains the BLAST tool equal to the one in the BLAST panel labelled 'Find your closest AGI!' on the SUBA4 homepage. The user can enter a sequence and retrieve data from Arabidopsis proteins with sequence similarity. The results can be filtered using the BLAST score as a threshold. The score measures sequence similarity in respect to sequence length. The BLAST hit AGIs are retrieved and the data for the AGI linked protein is retrieved from SUBA.

The screenshot shows the BLAST search interface. At the top, there are navigation links: Home, Search, Results, ToolBox, and Help. Below this is a search bar with a text input field containing "Protein contains any fragment in () with a blast score any". A red box labeled "Choose BLAST score threshold" points to the "any" text in the input field. Below the search bar is a section titled "Find Arabidopsis proteins where the...". There are several tabs: Subcellular locations, Protein properties, Protein-Protein Interaction, Affiliations, and Blast. The "Blast" tab is selected. Below the tabs, there is a "Clear" button and a text input field containing "1 Peptides found". A dropdown menu is open, showing options: "any", "more than 200", "more than 500", and "more than 1000". A red box labeled "Enter your protein fragment" points to the text input field. A red box labeled "Add to the query" points to a green plus icon in the top right corner of the search area.

Score is  $-\log_2(E)$  where  $E = p_{val} \times N_{eff}$  is the p-value times the effective search space size. The larger the score, the more statistically significant the match but since we tried  $N_{eff}$  times to find a match we need to multiply by  $N_{eff}$  (PDF).

The hits are displayed in the results view and each hit shows the BLAST score and aligned protein sequence.

The screenshot shows the BLAST results view. At the top, there are navigation links: Home, Search, Results, ToolBox, Help, Locations, Update, and Test. Below this is a search bar with a text input field containing "1". There are navigation icons (back, forward, search) and a "page size" dropdown set to "20". To the right, it says "Showing page 1 of 1 (5 total hits)". There are also links for "What's this query" and "Download". Below the search bar is a table with columns: AGI, SUBAcon, Predictions, Annotations, GFP, MS/MS, and PPI. The table contains three rows of results. The first row is for AGI AT2G40880.1, with SUBAcon "extracellular", Predictions "cytosol, mitochondrion, peroxisome, plastid, vacuole, golgi, endoplasmic reticulum, extracellular", Annotations "SwissProt: extracellular", GFP, MS/MS, and PPI. The second row is for AGI AT3G12490.1, with SUBAcon "cytosol", Predictions "nucleus, cytosol, vacuole, plastid, mitochondrion", Annotations "SwissProt: extracellular", GFP "cytosol", MS/MS "cytosol, endoplasmic reticulum", and PPI "AT3G56170.1: mitochondrion". The third row is for AGI AT3G12490.2, with SUBAcon "extracellular", Predictions "cytosol, mitochondrion, plastid, vacuole, golgi, endoplasmic reticulum, extracellular", Annotations "SwissProt: extracellular", GFP, MS/MS, and PPI. Below the table, there are two red boxes: "BLAST score" pointing to the "score:64.18" and "BLASTed sequence" pointing to the "MADQQAGTIVGGVVRDIDANANDLQVESLARFAVDEHNKNENLTLEYKRLLGAKTQVAVAGTMHHLTVEVADGETNKVY" sequence.



## SUBA4 results tab

The results tab will automatically be activated when the query is submitted. SUBA4 users will be able to see the query by clicking on the “What’s this query” button in the top left. The results can be downloaded as a table format using the download button.

The results are presented in table format. The columns can be customized towards the preference of the user. The first column shows the AGI of the proteins fitting the submitted query and the description for the protein below. This is followed by the consensus call derived from SUBAcon. Each of the individual localisation data columns show the summary of the data for the category. For a more details view for each category the user can access the factsheet by clicking on the AGI.

The screenshot shows the SUBA4 results interface. At the top, there is a navigation bar with 'Home', 'Search', and 'Results' buttons. A search bar contains the text 'Your Main query' and 'The download of the results table'. Below the search bar, there are navigation controls including a page number '1', a search icon, and a 'page size: 20' dropdown. A 'Showing page 1 of 58 (1142 total hits)' indicator is present. On the right side, there are buttons for 'What's this query' and 'Download'. The main content is a table with columns: AGI, SUBAcon, Predictions, Annotations, GFP, MS/MS, and PPI. The table lists several protein entries with their AGIs and associated data. Annotations include 'golgi mitochondrion', 'endoplasmic reticulum', 'cytosol', 'plasma membrane (2x)', 'nucleus', 'mitochondrion', 'plastid', 'peroxisome', and 'extracellular'. A 'PPI partners' column is also visible, showing a link to 'AT1G01910.1: golgi'. Red arrows point from text boxes to specific elements: 'What's this query' and 'Download' buttons, the 'MS/MS' column header, the first row of the table, the 'Protein description as in TAIR10' text box pointing to the description of AT1G05320.1, and the 'PPI partners' column.

Your Main query  
The download of the results table

What's this query Download

page size: 20 Showing page 1 of 58 (1142 total hits)

| AGI         | SUBAcon                     | Predictions   | Annotations   | GFP                        | MS/MS                             | PPI                |
|-------------|-----------------------------|---|---|----------------------------|-----------------------------------|--------------------|
| AT1G05190.1 | endoplasmic reticulum       | golgi mitochondrion   | golgi mitochondrion   |                            |                                   |                    |
| AT1G05320.1 | endoplasmic reticulum       | cytochrome P450, family 88, subfamily A, polypeptide 3; Encodes an ent-kaurenoic acid hydroxylase, a member of the CYP88 family | cytochrome P450, family 88, subfamily A, polypeptide 3; Encodes an ent-kaurenoic acid hydroxylase, a member of the CYP88 family |                            | (4x) cytosol plasma membrane (2x) |                    |
| AT1G05520.1 | endoplasmic reticulum golgi | endoplasmic reticulum plasma membrane nucleus cytosol mitochondrion   | endoplasmic reticulum plasma membrane nucleus cytosol mitochondrion   |                            | cytosol plasma membrane (2x)      |                    |
| AT1G05575.1 | endoplasmic reticulum       | Sec23/Sec24 protein transport family protein; plastid plasma membrane nucleus peroxisome extracellular endoplasmic reticulum    | plastid plasma membrane nucleus peroxisome extracellular endoplasmic reticulum  | endoplasmic reticulum (2x) |                                   | AT1G01910.1: golgi |

SUBA4 hits fitting your query. The AGI is linked and users can open the detailed flatfile by clicking on the AGI

Columns with the localisation summaries for each data category

Protein description as in TAIR10

PPI partners are linked to their flat file and their SUBAcon location is shown