Tutorial – cropPAL search interface

Table of content

The cropPAL home page

The home page view of cropPAL (http://crop-pal.org) provides the user with a introductory summary, links in the menu bar as well as several instant search options.

The *search window* below the summary is the easiest starting point to search for text (gene descriptions, titles or abstracts of studies in cropPAL) or ids (proteins identifiers).

Menu Bar

Quick Search Window

The Quick Search option lets you search text or protein identifiers by cut and paste in the window. Click on the blue (text) or green (ID) search button

The bottom part of the page contains the *query builder* tabs for creating more complex queries regarding locations, geographical data, protein properties or orthology.

The Menu links

Tutorial
You most likely reached this link as it leads to this document.

Localization World Map
This link shows the geographical distribution of the experimental studies that were integrated into cropPAL.

cropPAL stats
The data derived from experimental localization studies was graphed in number against years. This shows the amount of studies published in each crop, each year (accumulative) that are in the cropPAL database.
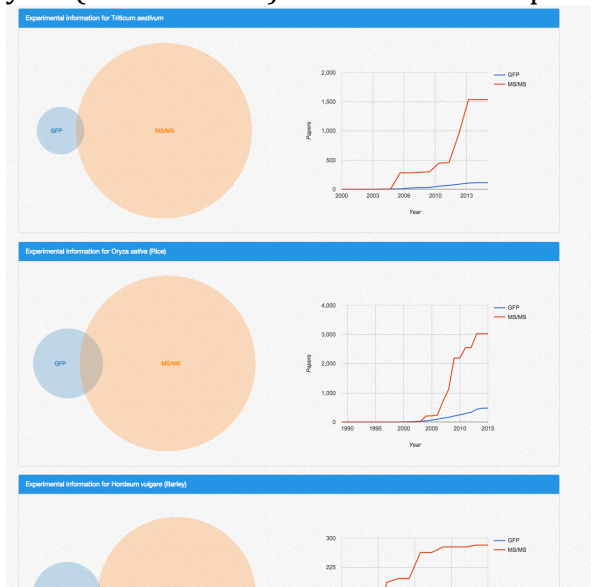
Discover PEB
This link leads to the Research facility home page or the ARC centre of excellence in Plant Energy Biology.

cropPAL search options

The Quick Search option

The first step for any search is choosing the crop species shown on the left. It is possible to choose from one to all crops. In order to do a quick search us the Quick Search window on the right of the front page. This search will allow you to search for keywords in the gene description, title or abstracts of experimental studies. Enter you keyword (e.g. peroxidase) and press the blue text search button. Alternatively, you can search for protein identifiers (e.g. MLOC_19297.1) to retrieve cropPAL data for the matching entries. Using identifiers press the green button to retrieve results.

Crop choice

For starting any type of query, choose the crop species you would like to search. You can click on one single crop or multiples for searching several crop data sets in series

Quick Search Window

The Quick Search option lets you search text or protein identifiers by cut and paste in the window. Click on the blue (text) or green (ID) search button

The Query Builder

The first step for any search is choosing the crop species that you want to query from the left side. It is possible to choose one, several or all crops. Users can access the query builder in the lower part of the portal. There are 4 query categories (Subcellular Location, Protein Properties, Homology, Affiliations) displayed as tabs.

Crop choice

For starting any type of query, choose the crop species you would like to search. You can click on one single crop or multiples for searching several crop data sets in series


Query building tabs

Choose the type of filter you want to apply (or question you want to ask)

Choose the category of queries and then choose the filter options (e.g. BaCeLo, plastid etc.). Then add the filter to the query by pressing the **+ add to query** button on the left.

This will open the query window and display the filter and the query builder buttons **(), AND, OR** to add more filters. Once done with the building. Press the "Query" button and retrieve the results.
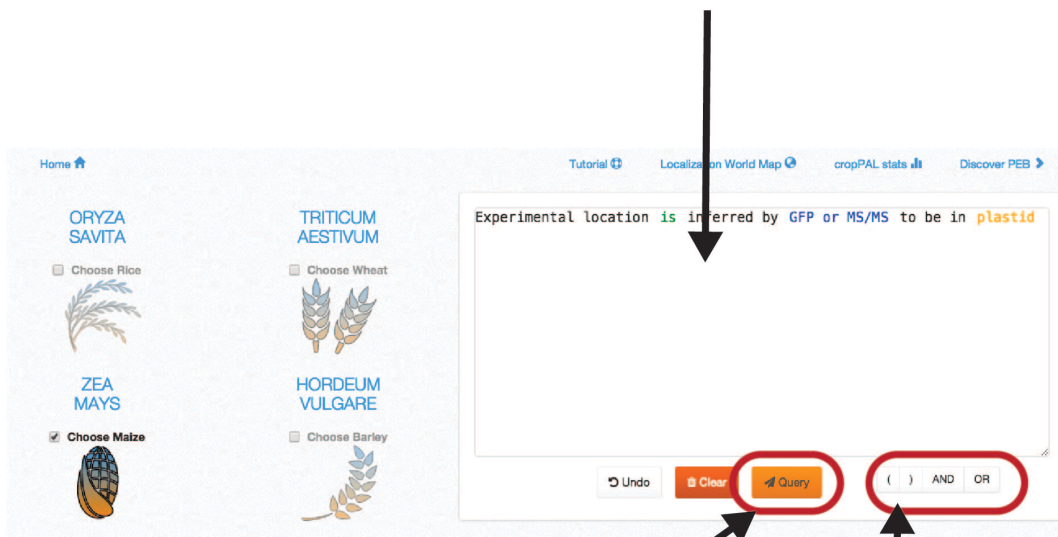

Customize the filter

Each query tab contains several filters to build queries. Customize the filter to create query conditions that will answer your questions you have towards the data
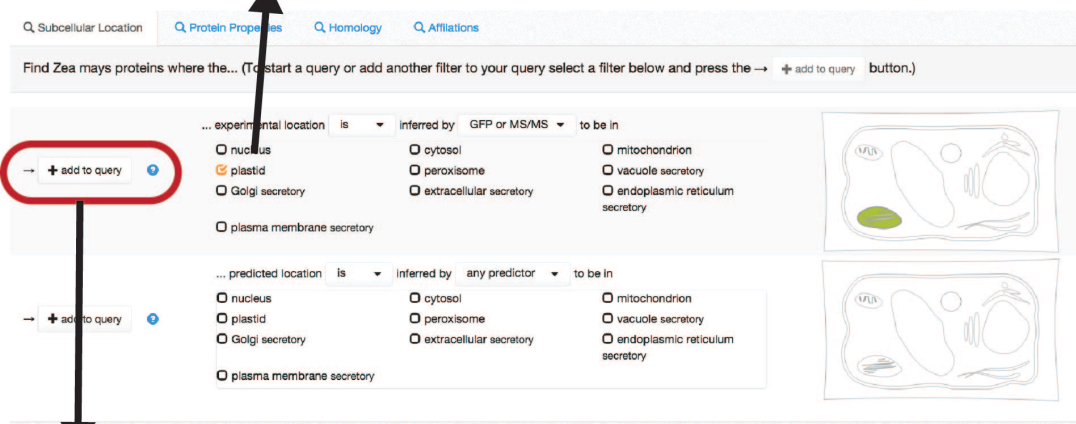
Add filter to the query

Adding the filter will lead to the appearance of the **Query Window**

Submit your query

link more filters from the query tabs into your query

The Query Categories

Subcellular Location Queries

This contains filters for subcellular locations. There are two sections, experimental data (top) and predicted data (bottom). The data can be filtered by type of experimental (GFP or MS or either), predictor (BaCeLo, Predotar, …, any or concensus) and subcellular location (plastid and, or nucleus and, or …). The drop down bars allow inclusion ("is in location xyz") or exclusion ("is not in location xyz") of locations. Subcellular locations can be added using the tick box lists as well as clicking on a location in the cell schematic.

In order to make the quick query window reappear go back to the home page using the **home** page link in the top left corner. Alternatively use the last filter in the Protein Properties tab for the ID query and the first filter for text earch in the protein description.

Property Queries

The cropPAL data sets can be filtered for protein properties. This includes protein descriptions (Ensembl annotations), protein size (amino acid length), molecular weight, isoelectric point, GRAVY, chromosome or assembly location of gene encoding the protein and protein features (GO terms or interpro motifs).

Choose a filter and enter your criteria. Then press the add button on the left to add it to the query.

Query category

Filters

Each row is one filter that can be customized

Start reading here

Add your filter to the Query

Adding the filter will lead to the appearance of the **Query Window**

Submit your query

link more filters from the query tabs into your query

In order to make the quick query window reappear go back to the home page using the **home** page link in the top left corner. Alternatively use the last filter in the Protein Properties tab for the ID query and the first filter for text earch in the protein description.

Homology Queries

The orthology filters included in cropPAL derive from either (the top filter) a reciprocal BLAST of crop proteins against Arabidopsis thaliana proteins (TAIR10) or (the bottom filter) using the homology tree generated by Ensemble using TreeBEST (http://plants.ensembl.org/info/website/ftp/index.html).

Query category

Start reading here

Protein trees
This will find the crop proteins that have orthologues in another crop species you can find paralogs with specification of subcellular location. The homology tree derived from EnsemblPlants

Reciprocal BLAST
This will find the crop proteins that have Arabidopsis orthologues in a subcellular location that you pick. The links are made by choosing the top match between two sequences after doing a BLAST both ways

Add your filter to the Query
Adding the filter will lead to the appearance of the **Query Window**

Submit your query

link more filters from the query tabs into your query

Not all proteins have homology matches due to evolutionary genome differences. In order to make the quick query window reappear go back to the home page using the **home** page link in the top left corner. Alternatively use the last filter in the Protein Properties tab for the ID query and the first filter for text earch in the protein description.

Affiliation Queries

The affiliations filter allow the user to filter cropPAL data for individual experimental studies, the study origin (country, institutions), study author names, year of publication as well as keywords within titles and abstracts of experimental studies.

Query category

Filters

Each row is one filter that can be customized

Start reading here

Add your filter to the Query

Adding the filter will lead to the appearance of the **Query Window**

Submit your query

link more filters from the query tabs into your query

The second last filter lets you choose one country at a time giving you the option of all the countries that contributed studies to cropPAL for the species you have chosen. The last query lets you choose several countries at once from all the countries in the world.

In order to make the quick query window reappear go back to the home page using the **home** page link in the top left corner. Alternatively use the last filter in the Protein Properties tab for the ID query and the first filter for text earch in the protein description.
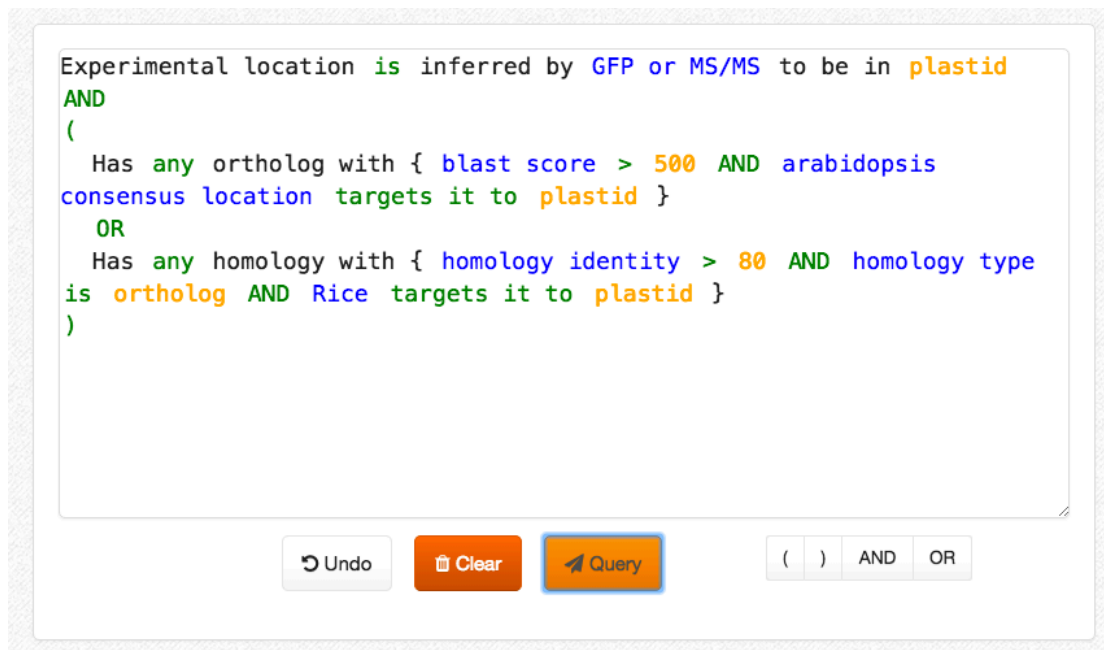
Query stacking

Building multi-filter queries can be achieved by pressing the AND, OR functions underneath the query window inbetween adding each filter from the query tabs. The full query will appear in the query window.

Example:
Question: I want to find all proteins in maize that are on the plastid and have experimental evidence. These proteins have orthologs in Arabidopsis and SUBAcon determined plastidal location or in rice with experimental evidence.

Start:
1. choose maize from the top left species choice menu
2. choose the Subcellular Location tab and the top filter for experimental locations in the plastid
3. add filter and start the query window
4. select **AND** to prepare for the next filter.
5. Since there is an AND as well as an OR we need brackets so that we do not end up with all plastid proteins in rice (query part after the **OR**)
6. go to the Homology tab and add the Arabidopsis comparison for plastid and add to the query
7. select **OR** under the query window
8. Go to the Homology tab and select the EnsemblPlants Homology Tree filter and choose rice, threshold and type of homology
9. add to query and close the brackets in the query window. The query should look similar to the below:
10. Press **Query** and wait for the result tab to appear.

```
Experimental location is inferred by GFP or MS/MS to be in plastid
AND
(
  Has any ortholog with { blast score > 500 AND arabidopsis
consensus location targets it to plastid }
  OR
  Has any homology with { homology identity > 80 AND homology type
is ortholog AND Rice targets it to plastid }
)
```

Undo | Clear | Query | (|) | AND | OR

Result Views

Result Tabs

Each query and each crop selected will result in a new tab that appears after pressing the **Query** button under the query window (or the blue/green Search button next to the Quick Search window. The result tabs are yellow and the query tabs are blue. You can run more queries and create more result tabs. They all stay open. To close them, press the **x** in the top right of the result tabs.

The Result tab displays the hits to your query in tabular format.

Tabular result view

Display columns can be customized

Query hit IDs

Result tab
each species and/or each query opens a separate

Results query information
and **download** option for results table

Best Arabidopsis match
Description: Os01g0191500 protein; Putative mitochondrial processing peptid [Source:UniProtKB/TrEMBL;Acc:Q5SNJ4]
Arabidopsis BestMatch: AT1G51980.1
Arabidopsis Description: Insulinase (Peptidase family M16) protein;
SUBA Consensus: [mitochondrion]

Localization information
including literature references

Homology matches
The closest homolog for each species is given in the homology column as a hyperlink to the respective flat file

In the left column the ID of the hit is displayed under "locus". This columns has an arrow appear when your mouse runs over the right hand column corner. This Arrow opens a drop down menu to customize displaying other available data for the protein hits.

Tabular result view

Display columns can be customized

Query hit IDs

Result
each species and/or each

Best Arabidopsis match

Description: Os01g0191500 protein; Putative mitochondrial processing peptid:
[Source:UniProtKB/TrEMBL;Acc:Q5SNJ4]
Arabidopsis BestMatch: AT1G51980.1
Arabidopsis Description: Insulinase (Peptidase family M16) protein;
SUBA Consensus: [mitochondrion]

Loc

Underneath the locus ID there is the description as given in the EnsemblPlants genome annotation. Below the hit protein description you find the best match in the Arabidopsis proteome (reciprocal BLAST) including the description the the SUBAcon (consensus) location sourced from SUBA.

In the other columns to the right you find the location information. The predictions column displays the different predictions made by all predictors, the experimental columns are devided into GFP and MS. They show any experimental evidence attached to the hit. The PubMed IDs are hyperlinked to the NCBI page.

The right column shows homology matched from the EnsemblPlants tree. If there is more matches than the height of the row allows, use the scroll bar to look at the hits. Each hit displays the EnsemblPlants ID, the type of homology match as well as the identity. All IDs are hypelinked to their individual factsheets.

Result tab
ch species and/or each query opens a separate

Results **query information** and **download** option for results table

Homology matches
The closest homolog for each species is given in the homology column as a hyperlink to the respective flat file

Localization information
including literature references

opsis match

The right top corner shows the query when clicking on it. If you have several results windows open, you can check which query they derived from. Next to the Query info is the Download button. This allows you to download the results table as a csv.

Protein Factsheets

Each protein hit can be selected to open its factsheet. The fact sheet displays an overview of all information regarding localization, protein properties and sequences.

On the left side is the location summary that shows how many times a subcellular location was called (either predicted or experimentally). The cartoon schematic shows the normalized calls (fraction of calls for each compartment).
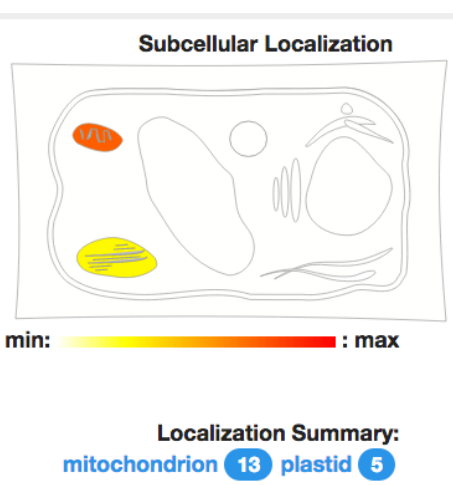
The protein annotations are all hypelinked to their databases. This allows easy access to more detailed information.

For continuing the search go back to the previous tab or click on the **back to CropPAL search** in the top left of the factsheet page.