

# SUBA3 tutorial

The SUBA interface works best in the browsers Mozilla Firefox, Chrome, and Safari. It will also work on Microsoft Explorer (6 and above), but is a bit slower in the database searching.

Open the SUBA URL at: <http://suba.plantenergy.uwa.edu.au/>

The four tabs on top of the page (HOME, SEARCH, RESULTS, HELP) will allow you to navigate through SUBA.

## 1 HOME page

'SUBA3 News' informs the user on any updates of the database including when localisation data from bibliographic references were last updated.

'Your Data in SUBA3' gives the user the possibility to submit data to SUBA. Click on the 'Upload' button and follow the instructions in the new window. We will assess your data and subsequently upload them into the SUBA database.

'Quick Search' allows the user to enter an AGI number or a keyword and search SUBA3 without entering the SEARCH page.

Buttons on the left give information about the SUBA database, its content and how to cite SUBA. In addition, a 'Search Example' can be clicked to view the RESULTS page and a link to SUBA2 is available.

The screenshot shows the SUBA3 homepage with several annotations in green text and red arrows:

- Additional Info**: Points to the left sidebar containing links for 'About SUBA3', 'Stats in SUBA3', 'Cite SUBA3', 'Search Example: AT2G33210.1', and 'SUBA2'.
- 'News'**: Points to the 'SUBA3 News' section, which lists updates such as protein-protein interaction data, localisation predictions from TAIR10, and the SUBAcon approach.
- 'Your Data'**: Points to the 'Your data in SUBA3' section, which includes an 'Upload' button and instructions for submitting subcellular location data.
- 'Quick Search'**: Points to the 'Quick Search' section, which features a text input field for AGI identifiers or keywords and a 'Query' button.

The page header includes navigation tabs: HOME, SEARCH, RESULTS, and HELP. The main heading is 'Welcome to SUBA3' with the subtitle 'The SUBcellular localisation database for Arabidopsis proteins.' The footer contains copyright information: © 2005-2012 Harvey Millar harvey.millar(at)uwa.edu.au and a revision date of 2012-09-05.

## 2 SEARCH page

To query the database on localisation information of Arabidopsis proteins select the SEARCH tab.

The phrase '**Find Arabidopsis proteins where the...**', displayed at the top, represents the beginning of the Boolean sentence that will form your query. To refine and add to your query use the rows of pull down menus. Each row of pull down menus works in combination to construct a query and each combination can be added to the main query pane by clicking the '**Add**' button at the end of the row. The '**Help**' buttons next to the 'Add' buttons open a window with instructions for each menu row.

The screenshot shows the SUBA search interface. At the top, there are navigation tabs: HOME, SEARCH (highlighted), RESULTS, and HELP. Below the tabs, the main heading is "Find Arabidopsis proteins where the...". The interface consists of several rows of search criteria, each with a "Clear" button on the left and an "Add" button (green circle with a plus) and a "Help" button (blue circle with a question mark) on the right. The criteria rows are:

- ... experimentally observed location is [dropdown] inferred by any source [dropdown] to be in any location [dropdown] and should have GFP image [checkbox]
- ... predicted location is [dropdown] predicted by any predictor [dropdown] to be in any location [dropdown]
- ... literature referenced location is [dropdown] described in select a paper [dropdown]
- New! ... protein does [dropdown] interact with protein(s) in list [dropdown]
- New! ... interacting protein is [dropdown] described in select a paper [dropdown]
- New! ... literature title or abstract contains [dropdown] keyword(s) [text input]
- ... protein description contains [dropdown] keyword(s) [text input]
- ... physical property of Number of supporting ESTs [dropdown] is equal to [dropdown]
- ... gene model is [dropdown] on chromosome 1 [dropdown]
- ... Arabidopsis Gene Initiative (AGI) identifier(s) is [dropdown] in list [dropdown]

Below the criteria rows, there are two text input fields for "Enter AGI identifier(s) here." and a "Clear" button. At the bottom, there are buttons for parentheses, AND, and OR, and a section titled "To search for locations of Arabidopsis proteins:" with a "Select a query and press an Add button." instruction. At the very bottom, there are "Undo", "Clear", and "Query" buttons.

### Search SUBA by pull down menus

The first row of pull down menus provides access to the majority of **experimental localisation data**. Using the three pull down menus, a query can be built to investigate, for example, *plastid localised proteins by mass spectrometry*. To access such a set of proteins:

- Keep the first pull down menu as 'is'.
- Select 'MS/MS assay' from the second pull down menu.
- The third pull down menu contains the subcellular location tags, select 'plastid'.
- Activate your query combination using the 'Add' button at the end of this row.
- The built query sentence will appear in the bottom window with the statement "Protein location is inferred by MS/MS assay to be in plastid".

Find Arabidopsis proteins where the...

... experimentally observed location is inferred by MS/MS assay to be in any location and should have GFP image

... predicted location is predicted to be in any location

... literature referenced location is

Clear

Enter AGI identifier(s) here.

New! ... interacting protein is

New! ... literature title or abstract

... protein description contains keyword(s)

... physical property of Number of supporting ESTs is equal to

... gene model is on chromosome 1

Clear

... Arabidopsis Gene Initiative (AGI) identifier(s) is in list

Enter AGI identifier(s) here.

( ) AND OR

Protein location is inferred by MS/MS assay to be in plastid

Undo Clear Query

Query sentence

'Undo', 'Clear' and 'Query' buttons

'AND' and 'OR' buttons and brackets

A gold standard (GS) set of subcellular location of 5688 proteins in Arabidopsis was compiled by manual analysis of TAIR10 annotation (Lamesch et al., 2012) and MapMan evaluation of biochemical pathway and functional groups (Usadel et al., 2009). This GS set can be investigated through the SUBA3 search interface using the first row of pull down menus. The GS dataset will be maintained and extended in the future.

Further additions can be made by using the Boolean linkers 'AND' or 'OR' command buttons (or brackets for more complex queries). If you select 'AND' and then use the pull down boxes for *plastid localised proteins by GFP analysis* and add to the query, you will select the protein set confirmed to be in plastids by both GFP and mass spectrometry. If you select 'OR' between these two commands, you will select the much larger protein set of GFP, mass spectrometry or both GFP and mass spectrometry located gene products in plastids.

- Query the database by clicking the 'Query' button at the bottom of the page.
- Use the 'Undo' button to remove any mistakes from the forming query. This removes the last entered step of the query.
- Use the 'Clear' button to remove the entire constructed query.

## Search SUBA by AGI

Alternatively, or in combination with the above, you can search for specific loci (Atxgxxxx) using the box below the 'Arabidopsis Gene Identifier(s) (is/is not) in list' menu row to rapidly view a specific subset of gene products and the evidence for their location. Alternative splice information may be included if desired. A list of AGIs can be **copy-paste** from Excel, Word, etc. or you may also enter a text containing AGIs. Any AGIs present in a text will be extracted and the **numbers of identified AGIs** will be reported above the box on the left. Press the '**Clear**' button (above the box far left) to delete any content from this box.

**'Clear'**      **Numbers of identified AGIs**      **Copy-paste list of AGIs into this box**

The screenshot shows the SUBA3 search interface. At the top are tabs for HOME, SEARCH, RESULTS, and HELP. Below is a section titled 'Find Arabidopsis proteins where the...' with various search criteria like 'experimentally observed location', 'predicted location', and 'literature referenced location'. A 'Clear' button is located to the left of the 'Enter AGI identifier(s) here.' text box. Below this text box, it says '22 AGIs found' followed by a list of AGI identifiers. At the bottom, there is a 'Query sentence' section showing a constructed query: 'Protein location is inferred by MS/MS assay to be in plastid AND Arabidopsis Gene Identifier is (AT4G04020.1 AT4G22240.1 AT2G35490.1 AT3G26070.1 AT3G26080.1 AT3G23400.1 AT5G09820.1 AT5G19940.1 AT3G58010.1 AT2G42130.4 AT2G46910.1 AT4G00030.1 AT1G51110.1 AT1G18060.1 AT5G05200.1 AT4G31390.1 AT1G79600.1 AT4G19170.1 AT4G32770.1 AT1G54570.1 AT5G08740.1 AT1G32220.1)'. At the very bottom are 'Undo', 'Clear', and 'Query' buttons.

22 AGIs found

AT4G04020.1 AT4G22240.1 AT2G35490.1 AT3G26070.1 AT3G26080.1 AT3G23400.1 AT5G09820.1 AT5G19940.1 AT3G58010.1 AT2G42130.4 AT2G46910.1 AT4G00030.1 AT1G51110.1 AT1G18060.1 AT5G05200.1 AT4G31390.1 AT1G79600.1 AT4G19170.1 AT4G32770.1 AT1G54570.1 AT5G08740.1 AT1G32220.1

Query sentence

## Search SUBA by protein-protein interaction

A new feature in SUBA3 is the search for protein-protein interacting pairs (PPIs). A list of AGIs can be entered into the box below the 'Protein (does/does not) interact with proteins(s) in list' menu row to rapidly view a specific subset of interacting proteins and the evidence for their location. Alternative splice information may be included if desired. A list of AGIs can be **copy-paste** from Excel, Word, etc. or you may also enter a text containing AGIs. Any AGIs present in a text will be extracted and the number of identified AGIs will be reported above the box on the left. Press the 'Clear' button (above the box far left) to delete any content from the box.



'Clear'      Number of identified AGIs      Copy-paste list of AGIs into this box

Find **Arabidopsis proteins** where the...

... experimentally observed location is inferred by any source to be in any location and should have GFP image

... predicted location is predicted by any predictor to be in any location

... literature referenced location is described in select a paper

Clear 18 AGIs found

New! ... protein does interact with protein(s) in list

AT3G26080.1 AT3G23400.1 AT5G09820.1 AT5G19940.1 AT3G58010.1 AT2G42130.4 AT2G46910.1 AT4G00030.1 AT1G51110.1 AT1G18060.1 AT5G05200.1 AT4G31390.1 AT1G79600.1 AT4G19170.1 AT4G32770.1 AT1G54570.1 AT5G08740.1 AT1G32220.1

New! ... interacting protein is described in select a paper

New! ... literature title or abstract contains keyword(s)

... protein description contains keyword(s)

... physical property of Number of supporting ESTs is equal to

... gene model is on chromosome 1

Clear

... Arabidopsis Gene Initiative (AGI) identifier(s) is in list

Enter AGI identifier(s) here.

( ) AND OR

Protein does interact with protein(s) AT3G26080.1 AT3G23400.1 AT5G09820.1 AT5G19940.1 AT3G58010.1 AT2G42130.4 AT2G46910.1 AT4G00030.1 AT1G51110.1 AT1G18060.1 AT5G05200.1 AT4G31390.1 AT1G79600.1 AT4G19170.1 AT4G32770.1 AT1G54570.1 AT5G08740.1 AT1G32220.1

Undo Clear Query

Query sentence

Other search parameters include:

- Pre-computed localisation **predictions** of 22 prediction programs for all ~35,400 Arabidopsis proteins. Access predicted protein sets through the 'Predicted location (is/is not) predicted by (predictor) to be in (location)' menu row.
- Subcellular localisation claims of particular **research papers** using the 'Literature referenced location (is/is not) described in (paper)' menu row. This allows combinations of protein sets from specific papers (by using 'OR' linker) or looking for the set of proteins claimed by multiple papers (using 'AND' linker). The 'look ahead typing' feature allows the user to quickly find a paper in question.
- Interacting protein claims of different **research papers** using the 'Interacting protein (is/is not) described in (paper)' menu row. The 'look ahead typing' feature allows the user to quickly find a paper in question.
- Text based searches of **titles or abstracts** through the 'Literature title or abstract (contains/does not contain/matches) keyword(s)' menu row. A search will be conducted against the literature titles and abstracts in the SUBA database. Choosing 'matches' will give you access to the [match syntax of MySQL](#), e.g. entering +leaf – seed\* in the keyword(s) box matches a title/abstract that contains leaf but that *does not* contain seed, seeds, or seedling etc.

- Text based searches of **descriptor fields** through the 'Protein description (contains/does not contain) keyword(s)' menu row. For example, you can limit your search to genes with the word “dehydrogenase” or “kinase” in the description. A search will be conducted against the TAIR10 descriptions of proteins in the SUBA database. Choosing ‘matches’ will give you access to the [match syntax of MySQL](#), e.g. entering +leaf –seed\* in the keyword(s) box matches a TAIR description that contains leaf but that *does not* contain seed, seeds, or seedling etc.
- **Physio-chemical parameters** such as number of ESTs, number of full-length cDNAs, number of amino acids, molecular weight, isoelectric point, and gravity. Use the 'Physical property of' menu row to filter for proteins based on these numeric data derived from TAIR10 or calculated locally.
- Use the 'Gene model' menu row to filter for proteins that are translated from genes on (or not on) a specific **chromosome**.

### **3 RESULTS page**

Once a query has been submitted the contents of the 'RESULTS' tab will automatically be displayed. By default seven columns will be displayed: AGI (loci), SUBA consensus location, locations of all predictions, locations of all annotations, locations by GFP, locations by MS/MS, and PPI (interacting proteins).

Results can be sorted (ascending/descending) by field using the **function menu**. The function menu is activated by tracking the mouse over the column header and then selecting the emerging arrow. New columns can be added to the Result tab window by selecting 'Columns' in the function menu. Columns can be organized using mouse drag and drop functionality.

The TAIR description is shown below each result in grey and can be hidden by clicking the **'no descriptions'** button (top left of the page).

Download results

Hide descriptions

Function menu

AGI

SUBAcon

Predictions

Annotations

GFP

MS/MS

PPI

AT1G49970.1

plastid

Sort Ascending

Sort Descending

Columns

SwissProt: plastid

TAIR: plastid 16207701

TAIR: plastid 18431481

TAIR: plastid 14593120

TAIR: plastid 11278690

AGI

TAIR Description

SUBAcon

Predictions

Annotations

AmiGO

SwissProt

TAIR

GFP

MS/MS

PPI

All papers

Papers GFP

Papers MS/MS

Papers PPI

AT1G50170.1

plastid

mitochondrion

plastid

nucleus

plasma membrane

AT1G50320.1

plastid

cytosol

mitochondrion

plastid

AT1G50900.1

plastid

cytosol

mitochondrion

nucleus

plastid

AT1G54500.1

plastid

mitochondrion

plastid

nucleus

extracellular

AT1G56190.1

plastid

cytosol

mitochondrion

plastid

Amigo: plastid 16207701

Amigo: plastid 18431481

Amigo: plastid 18633119

Amigo: plastid 15028209

SwissProt: plastid

plastid: 19036033

mitochondrion: 14671022

plasma membrane: 19334764

plastid: 21539947

plastid: 21531424

plastid: 20423899

Page 1 of 5

Displaying 1 - 50 of 248

AGI, link to SUBA flatfile

'Next Page' and 'Last Page' buttons

PubMed ID or doi hyperlink to publication

Total number of results

Each loci number (**AGI**) provides a link to a summary page, the **SUBA flatfile** (see below).

The codes or unique identifiers (uid) beside a location are **PubMed IDs or doi** and provide links to the publication for each entry at PubMed or ISI Web of Science.

The **total number of results** is noted in the bottom right corner of the page. Fifty results are shown at a time, further pages can be displayed by using the '**Next Page**' or '**Last Page**' buttons at the bottom left of the page.

### Download SUBA results

For further analysis by the user, all results can be downloaded as a tab delimited file (csv, tsv, xls) by using the '**Download All Results**' button in the top left corner of the page. By default it downloads as an Excel file.

The columns of the exported file include:

- AGI number
- TAIR10 description
- Number of the chromosome the gene is located on
- Expressed sequence tags (ESTs), fragments of mRNA sequences

- Full-length cDNAs
- Number of amino acid residues
- Molecular weight of a protein
- Isoelectric point (IEP), pH at which a protein carries no net charge
- Grand average of hydropathy (GRAVY), calculated as the sum of hydropathy values of all amino acids and divided by the number of residues in the sequence
- Location determined by SUBAcon (the Bayes approach to give consensus location)
- List of locations by all predictors
- Locations of each of the 22 predictors: AdaBoost (Niu et al., 2008), ATP (Mitschke et al., 2009), BaCelLo (Pierleoni et al., 2006), ChloroP 1.1 (Emanuelsson et al., 1999), EpiLoc (Brady and Shatkay, 2008), iPSORT (Bannai et al., 2002), MitoPred (Guda et al., 2004), MitoProt (Claros and Vincens, 1996), MultiLoc2 (Blum et al., 2009), Nucleo (Hawkins et al., 2007), PCLR 0.9 (Schein et al., 2001), Plant-mPLOC (Chou and Shen, 2010), PProwler 1.2 (Hawkins and Boden, 2006), Predotar v1.03 (Small et al., 2004), PredSL (Petsalaki et al., 2006), PTS1 (Neuberger et al., 2003), SLPFA (Tamura and Akutsu, 2007), SLP-Local (Matsuda et al., 2005), SubLoc (Hua and Sun, 2001), TargetP 1.1 (Emanuelsson et al., 2000), WoLF PSORT (Horton et al., 2007), and YLoc (Briesemeister et al., 2010)
- List of locations of all annotators
- Locations of each of the 3 annotators: AmiGO (Carbon et al., 2009), Swiss-Prot (Schneider et al., 2009) and Swiss-Prot identification number, and TAIR10 (Lamesch et al., 2012).
- Location of GFP experimental evidence
- Location of MS/MS experimental evidence
- Protein-protein interaction (PPI) data (AGI numbers of interacting proteins)
- List of all papers relating to experimental evidence
- List of papers for MS/MS evidence
- List of papers for GFP evidence
- List of papers for PPI data

#### **4 SUBA flatfile**

The SUBA flatfile provides a variety of information and links that we hope are helpful to the user. These include a detailed breakdown of **subcellular localisation information**: predictions, annotations, GFP, MS/MS, PPI data and in house GFP localisation images for some of the entries. Each of the GFP images is clickable to enlarge the image.

To determine as objectively as possible where a particular protein is located, we have developed a naive Bayes classifier (SUBAcon) that incorporates experimental localisation and all the targeting prediction data to best estimate a protein's location in the cell. The probabilities of subcellular location for each protein are provided and displayed as a **pictographic heat map** of a plant cell (top left). The probabilities of each subcellular compartment can be viewed by tracking the mouse over the pictographic heat map.

The flatfile further contains **descriptions**, **curator summaries**, **coordinates** and **sequence information** from TAIR10 and calculated **physio-chemical characteristics**. A picture presentation of a **hydropathy plot** gives the user an idea of the likely orientation of the protein in membranes. Links to the same entry at **other Arabidopsis databases** can be found below the hydropathy plot.



## Pictographic heat map (SUBAcon)

## Subcellular localisation information

## GFP images

Plant Energy Biology ARC Centre of Excellence  
Computational Systems Biology Centre of Excellence

AT2G33210.1

Subcellular localization

Predictors

Annotations

GFP

MSMS

PPI

GFP Images

Arabidopsis cell culture (mitochondrial marker)

AT2g33210-GFP (full-length) mitochondrial marker-RFP overlay

Arabidopsis cell culture (plastidial marker)

AT2g33210-GFP (full-length) plastidial marker-RFP overlay

Images by Sandra Tana

Description (TAIR10) protein\_coding heat shock protein 60-2

Curator Summary (TAIR10)

Computational Description (TAIR10) heat shock protein 60-2 (HSP60-2); FUNCTIONS IN: copper ion binding, ATP binding; INVOLVED IN: inflammatory response, response to salt stress; LOCATED IN: mitochondrion, plasma membrane, chloroplast; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAINS: Chaperonin Cpn60/TCP-1 (InterPro:IPR002423), Chaperonin Cpn60, conserved site (InterPro:IPR018370), Chaperonin Cpn60 (InterPro:IPR001844); BEST Arabidopsis orthologous protein in match is: heat shock protein 60 (TAIR:AT3G33990.1); Has 33915 BLAST hits to 33873 proteins in 8692 species: Archae - 707; Bacteria - 21782; Metazoa - 1660; Fungi - 1548; Plants - 607; Viruses - 2; Other Eukaryotes - 7369 (source: NCBI BLINK).

Coordinates (TAIR10) chr2:-14075093..14078568

Molecular Weight (calculated) 61982.2 Da

IEP (calculated) 6.59

GRAVY (calculated) -0.04

Length 584 amino acids

Sequence (TAIR10) MYSLVNV ASKARAR KCTSDGS RLSNTRY AADIRFG VEARALM ROVEDLAD AVKYVMP KGRNVLE QNGAPVY KCOGVTA KSEIFKR IKNVGSL YKVQVANF NOVAGDT TCATVLR ALTFEGK SVAGRNA MLAGRIK LAVDTVT NLQSRAM ISSEETIA QVOTISAN GDSIGEL IAKAMTV GKEGVITI QGKTLFN ELEYVGM KIDRGVY PFYITNPK TQCELE PLILIEK KISINAM VYVLEAL KQKPLLI VAEIVSD ALATLIN KLRANKY CAVKAPG GENRKANL HDLAALTG AQVITEEL QNGLNDID LNFQNRK KYVYKDD PVYLDGAG DQALGER CEQIRHW EATSDYD KKKLQKL AKLGGVA VLKDGAS STEVERK DRVTALN ATKANVE GIVPGGV ALLYASKE LKRLSTAN FQKIQVQ IIQNALTK PVYTIAS AGGVNAV VOKLLIQD NPLQDIA AGGVYTM EKAGTID LUTRALY VQNAVSE LITTEVAY VEEIPPE VAEVQSG GQMGSGE KEGACAG P

Descriptions, curator summaries, coordinates, protein sequence, etc.

Hydropathy Plot (raw data)

Hydropathy: AT2G33210.1

See Also Aramemnon AIProteome DBGET Inparanoid MIPS MPSS Plus PPDB PlantSpecDB ProMEX Proteins.Wiki SALK(inserts) SALK(signal) TAIR SwissProt

Citation

If you find this resource useful please cite one of the following publications:

Heazlewood JL, Verboom RE, Tonis-Filipini J, Small I and Millar AH. (2007) SUBA: the Arabidopsis Subcellular Database. Nucleic Acids Res. 35(D):213-218. (PubMed)

Heazlewood JL, Tonis-Filipini J, Verboom RE, Millar AH. (2008) Combining experimental and predicted datasets for determination of the subcellular location of proteins in Arabidopsis. Plant Physiol. 139(2):598-608. (PubMed)

Hydropathy plot

Links to other Arabidopsis databases

## 5 HELP page

'How to search SUBA3?' provides a step-by-step tutorial that can be downloaded as a PDF file.

'How to submit your data to SUBA3?' gives the user the possibility to submit data to SUBA. Click on the 'Upload' button and follow the instructions in the new window. We will assess your data and subsequently upload them into the SUBA database.

'Any problems with SUBA3?' please contact the people listed.

## References

- Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., and Miyano, S. (2002). Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* 18, 298-305.
- Blum, T., Briesemeister, S., and Kohlbacher, O. (2009). MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *BMC bioinformatics* 10, 274.
- Brady, S., and Shatkay, H. (2008). EpiLoc: a (working) text-based system for predicting protein subcellular location. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 604-615.
- Briesemeister, S., Rahnenfuhrer, J., and Kohlbacher, O. (2010). YLoc--an interpretable web server for predicting subcellular localization. *Nucleic acids research* 38, W497-502.
- Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., and Lewis, S. (2009). AmiGO: online access to ontology and annotation data. *Bioinformatics* 25, 288-289.
- Chou, K.C., and Shen, H.B. (2010). Plant-mPLOC: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PloS one* 5, e11335.
- Claros, M.G., and Vincens, P. (1996). Computational method to predict mitochondrially imported proteins and their targeting sequences. *European journal of biochemistry / FEBS* 241, 779-786.
- Emanuelsson, O., Nielsen, H., Brunak, S., and Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of molecular biology* 300, 1005-1016.
- Emanuelsson, O., Nielsen, H., and Von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein science : a publication of the Protein Society* 8, 978-984.
- Guda, C., Guda, P., Fahy, E., and Subramaniam, S. (2004). MITOPRED: a web server for the prediction of mitochondrial proteins. *Nucleic acids research* 32, W372-374.
- Hawkins, J., and Boden, M. (2006). Detecting and sorting targeting peptides with neural networks and support vector machines. *Journal of bioinformatics and computational biology* 4, 1-18.
- Hawkins, J., Davis, L., and Boden, M. (2007). Predicting nuclear localization. *Journal of proteome research* 6, 1402-1409.
- Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic acids research* 35, W585-587.
- Hua, S., and Sun, Z. (2001). Support vector machine approach for protein subcellular localization prediction. *Bioinformatics* 17, 721-728.
- Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A., and Huala, E. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic acids research* 40, D1202-1210.
- Matsuda, S., Vert, J.P., Saigo, H., Ueda, N., Toh, H., and Akutsu, T. (2005). A novel representation of protein sequences for prediction of subcellular location using support vector machines. *Protein science : a publication of the Protein Society* 14, 2804-2813.
- Mitschke, J., Fuss, J., Blum, T., Hoglund, A., Reski, R., Kohlbacher, O., and Rensing, S.A. (2009). Prediction of dual protein targeting to plant organelles. *The New phytologist* 183, 224-235.

- Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A., and Eisenhaber, F. (2003). Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. *Journal of molecular biology* 328, 581-592.
- Niu, B., Jin, Y.H., Feng, K.Y., Lu, W.C., Cai, Y.D., and Li, G.Z. (2008). Using AdaBoost for the prediction of subcellular location of prokaryotic and eukaryotic proteins. *Molecular diversity* 12, 41-45.
- Petsalaki, E.I., Bagos, P.G., Litou, Z.I., and Hamodrakas, S.J. (2006). PredSL: a tool for the N-terminal sequence-based prediction of protein subcellular localization. *Genomics, proteomics & bioinformatics / Beijing Genomics Institute* 4, 48-55.
- Pierleoni, A., Martelli, P.L., Fariselli, P., and Casadio, R. (2006). BaCellLo: a balanced subcellular localization predictor. *Bioinformatics* 22, e408-416.
- Schein, A.I., Kissinger, J.C., and Ungar, L.H. (2001). Chloroplast transit peptide prediction: a peek inside the black box. *Nucleic acids research* 29, E82.
- Schneider, M., Lane, L., Boutet, E., Lieberherr, D., Tognolli, M., Bougueleret, L., and Bairoch, A. (2009). The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program. *Journal of proteomics* 72, 567-573.
- Small, I., Peeters, N., Legeai, F., and Lurin, C. (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 4, 1581-1590.
- Tamura, T., and Akutsu, T. (2007). Subcellular location prediction of proteins using support vector machines with alignment of block sequences utilizing amino acid composition. *BMC bioinformatics* 8, 466.
- Usadel, B., Poree, F., Nagel, A., Lohse, M., Czedik-Eysenberg, A., and Stitt, M. (2009). A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. *Plant, cell & environment* 32, 1211-1229.