

May 20, 2020

Find Proteins of Unknown Function (PUFs) using Plantannot - Protocol F



Forked from [Find Proteins of Unknown Function \(PUFs\) using Plantannot - Protocol E](#)

DOI

dx.doi.org/10.17504/protocols.io.bgdkjs4w

Marcos Viana¹, Mauricio Mudadu¹, Adhemar Zerlotini¹

¹EMBRAPA



Marcos Viana

OPEN  ACCESS



DOI: dx.doi.org/10.17504/protocols.io.bgdkjs4w

External link: <https://www.machado.cnptia.embrapa.br/plantannot>

Protocol Citation: Marcos Viana, Mauricio Mudadu, Adhemar Zerlotini 2020. Find Proteins of Unknown Function (PUFs) using Plantannot - Protocol F. [protocols.io](https://dx.doi.org/10.17504/protocols.io.bgdkjs4w) <https://dx.doi.org/10.17504/protocols.io.bgdkjs4w>

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

It is working

Created: May 13, 2020

Last Modified: May 20, 2020

Protocol Integer ID: 37004



Abstract

The Plant annot softw are provi des sever al filters and a text searc h box that allow s searc hing for mole cules by its desir ed annot ation featur es. Thes e filters are need ed to obtai n PUFs and to try to relate them to abioti c stress es using RNA-seq expre ssion data and co-expre ssion netw orks.



The Filter menu is separated in 8 fields, of those we are going to use only five: "Organism", "Feature type", "Orthology", "Orthologs _coexpression" and "Analyses". The "Feature Type" filter has three molecule types, from those the polypeptide box is the only that is going to be always checked and the others



blank.
By
using
the
other
4
remain
ing
filters
, 6
proto
cols
were
creat
ed as
exam
ples
of
differ
ent
ways
to
select
ing
PUFs.
Proto
col A:
using
lack
of
both
homo
logy
and
protei
n
doma
in
signa
tures.
Proto
col B:
using
lack
of
homo
logy,
prese
nce
of
doma
in
signa
tures
-
trying
to
select
Doma
ins of
Unkn
own
Funct
ion



(DUF) from PFAM, and the text search "Unknown function". Protocol C: using homology, lack of protein domain signatures and the text search "Unknown function". Protocol D-F: same protocols of A-C but using ortholog groups to find homolog proteins with co-expression data related to abiotic stress.



Protocol F is intended to find PUFs from organisms that proteins are already public in the NCBI's "nr" database and have no protein domain found by Interproscan. Proteins will be selected using the text search "Unknown function". Also, ortholog groups and co-expression networks will be

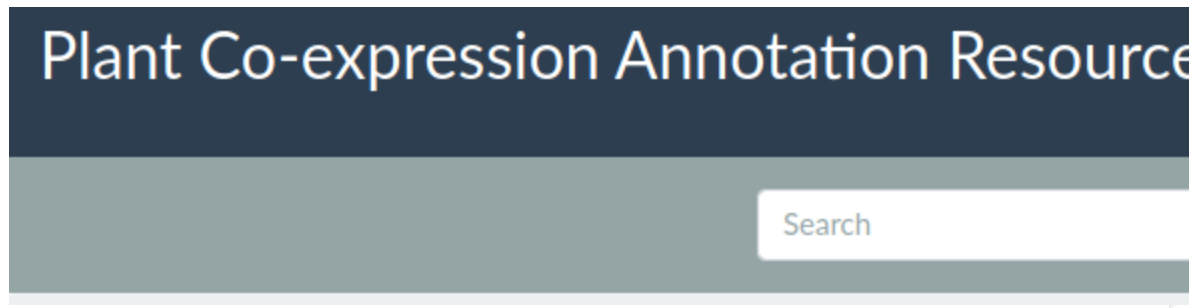


used
to
relate
proteins
to
abiotic
stresses.

Entering application

- 1 Enter the Plantannot Result's page, with empty filters and text box search:
<https://www.machado.cnptia.embrapa.br/plantannot/find/?q=>

Or you can enter the <https://www.machado.cnptia.embrapa.br/plantannot> initial page and click on the magnifying glass with the text box empty as well.



<https://www.machado.cnptia.embrapa.br/plantannot>

Filtering

- 2 Find PUFs from organisms that proteins are already public in the NCBI's "nr" database and have no protein domain found by Interproscan. Proteins will be selected using the text search "Unknown function". Also, ortholog groups and co-expression networks will be used to relate proteins to abiotic stresses.

Visualize the "Filters" card on the left of the page from step1:



Filters

Organism (53)

[apply](#)

- ☐ *Amaranthus hypochondriacus* (69,156)
- ☐ *Amborella trichopoda* (80,538)
- ☐ *Ananas comosus* (81,072)
- ☐ *Aquilegia coerulea* (117,123)



Feature type

[apply](#)

- ☐ gene (1,862,010)
- ☐ mRNA (2,332,974)
- ☐ polypeptide (2,332,974)

Orthology

[apply](#)

- ☐ no orthology (4,636,180)
- ☐ orthology (1,891,778)

Coexpression

[apply](#)

- ☐ no co-expression groups (6,381,557)
- ☐ co-expression groups (146,401)

Orthologs_coexpression

[apply](#)

- ☐ no co-expression (5,097,464)
- ☐ co-expression (1,430,494)

Analyses

[apply](#)

- ☐ diamond matches (2,209,087)
- ☐ interproscan matches (1,903,332)
- ☐ no diamond matches (4,318,871)
- ☐ no interproscan matches



(4,624,626)

Biomaterial apply

☐ Leaf (144,826)
☐ Rosette leaves (21,968)
☐ Seedling (26,971)

Treatment apply

☐ Dehydration (66,121)
☐ Drought (134,012)
☐ Heat stress (50,409)
☐ Osmotic stress (130,599)

<https://www.machado.cnptia.embrapa.br/plantannot/find/?q=>

- 2.1 In the "Organisms" filter, select any organisms (expand the organism's list using the green arrow) or select all by leaving all boxes empty. We will use *Oropetium tomaeum* as example. Click "apply" to execute the filter:

**Organism (53)**

apply

- ☐ *Amaranthus hypochondriacus*
(69,156)
- ☐ *Amborella trichopoda* (80,538)
- ☐ *Ananas comosus* (81,072)
- ☐ *Aquilegia coerulea* (117,123)
- ☐ *Arabidopsis halleri* (78,830)
- ☐ *Arabidopsis lyrata* (97,337)
- ☐ *Arabidopsis thaliana* (98,188)
- ☐ *Boea hygrometrica* (143,334)
- ☐ *Boechera stricta* (87,040)
- ☐ *Brachypodium distachyon*
(140,254)
- ☐ *Brachypodium stacei* (102,612)
- ☐ *Brassica oleracea* (106,200)
- ☐ *Brassica rapa* (127,232)
- ☐ *Capsella grandiflora* (77,927)
- ☐ *Capsella rubella* (83,415)
- ☐ *Carica papaya* (83,355)
- ☐ *Citrus clementina* (92,391)
- ☐ *Citrus sinensis* (117,673)
- ☐ *Cucumis sativus* (82,231)
- ☐ *Daucus carota* (96,349)
- ☐ *Eucalyptus grandis* (128,909)
- ☐ *Eutrema salsugineum* (84,919)
- ☐ *Fragaria vesca* (98,493)
- ☐ *Glycine max* (233,338)
- ☐ *Gossypium raimondii* (192,039)
- ☐ *Kalanchoe fedtschenkoi* (121,344)
- ☐ *Kalanchoe laxiflora* (188,815)
- ☐ *Linum usitatissimum* (130,439)
- ☐ *Malus domestica* (190,548)
- ☐ *Manihot esculenta* (115,795)
- ☐ *Medicago truncatula* (175,532)
- ☐ *Mimulus guttatus* (95,286)
- ☐ *Musa acuminata* (109,584)
- ☒ *Oropetium thomaeum* (85,338)
- ☐ *Oryza sativa* (147,037)
- ☐ *Panicum hallii* (136,936)
- ☐ *Panicum virgatum* (348,885)
- ☐ *Phaseolus vulgaris* (101,423)
- ☐ *Populus trichocarpa* (187,361)
- ☐ *Prunus persica* (121,051)
- ☐ *Ricinus communis* (93,663)
- ☐ *Salix purpurea* (160,905)



- ☐ *Setaria italica* (120,586)
- ☐ *Setaria viridis* (132,402)
- ☐ *Solanum lycopersicum* (104,175)
- ☐ *Solanum tuberosum* (151,458)
- ☐ *Sorghum bicolor* (128,371)
- ☐ *Spirodela polyrhiza* (58,869)
- ☐ *Theobroma cacao* (118,260)
- ☐ *Trifolium pratense* (122,552)
- ☐ *Vitis vinifera* (79,038)
- ☐ *Zea mays* (241,000)
- ☐ *Zostera marina* (61,350)

https://www.machado.cnptia.embrapa.br/plantannot/find/?q=&selected_facets=organism%3AOropetium+thomaeum

2.2 In the "Feature type" filter, select "polypeptide", and click "apply" to execute the filter:

Feature type

applyremove

☐ gene (0)
☐ mRNA (0)
☒ polypeptide (2,551)

https://www.machado.cnptia.embrapa.br/plantannot/find/?q=&selected_facets=organism%3AOropetium+thomaeum&selected_facets=so_term%3Apolypeptide

2.3 In the Orthology filter, select "orthology", and click "apply" to execute the filter:

**Orthology**

apply remove

- ☐ no orthology (0)
- ☒ orthology (18,755)

https://www.machado.cnptia.embrapa.br/plantannot/find/?q=&selected_facets=organism%3AOropetium+thomaeum&selected_facets=so_term%3Apolypeptide&selected_facets=orthology%3A1

2.4 Leave the "Coexpression" filter empty:

Coexpression

apply

- ☐ no co-expression groups (711)
- ☐ co-expression groups (0)

2.5 In the "Orthologs_coexpression" filter, select "co-expression", and click "apply" to execute the filter:

Orthologs_coexpression

apply remove

- ☐ no co-expression (0)
- ☒ co-expression (15,641)

https://www.machado.cnptia.embrapa.br/plantannot/find/?q=&selected_facets=organism%3AOropetium+thomaeum&selected_facets=so_term%3Apolypeptide&selected_facets=orthology%3A1&selected_facets=orthologs_coexpression%3Atrue

- 2.6 In addition to the applied filters, text search for "Unknown function" in the search box. After entering the text click on the magnifying glass icon to apply the search.

Analysesapplyremove

☒ diamond matches (894)
☐ interproscan matches (0)
☐ no diamond matches (0)
☒ no interproscan matches (894)

https://www.machado.cnptia.embrapa.br/plantannot/find/?q=&selected_facets=organism%3AOropetium+thomaeum&selected_facets=so_term%3Apolypeptide&selected_facets=orthology%3A1&selected_facets=orthologs_coexpression%3Atrue&selected_facets=analyses%3Adiamond+matches&selected_facets=analyses%3Ano+interproscan+matches

- 2.7 Leave the "Biomaterial" and "Treatment" filters empty:


Biomaterial
apply

☐ Leaf (0)
☐ Rosette leaves (0)
☐ Seedling (0)

Treatment
apply

☐ Dehydration (0)
☐ Drought (0)
☐ Heat stress (0)
☐ Osmotic stress (0)

- 2.8 In addition to the applied filters, text search for "Unknown function" using the search box. After entering the text click on the magnifying glass icon.

Unknown function 

Filters

Selected filters

Results

Organism	Feature Type	Feature ID	Relationship	Display
Oropetium	polypeptide	Oropetium_20150405_01079A.v1.0_mRNA	DE12022	Domain of unknown function

https://www.machado.cnptia.embrapa.br/plantannot/find/?q=Unknown+function&selected_facets=organism%3AOropetium+thomaeum&selected_facets=so_term%3Apolypeptide&selected_facets=orthology%3A1&selected_facets=orthologs_coexpression%3Atrue&selected_facets=analyses%3Adiamond+matches&selected_facets=analyses%3Ano+interproscan+matches

Filters

3 After execution of all filters we will have the following list of filters:

Selected filters

search:Unknown function ✕

organism:Oropetium thomaeum ✕

so_term:polypeptide ✕

orthology ✕

coexpression in orthologs ✕

analyses:diamond matches ✕

analyses:no interproscan matches ✕

https://www.machado.cnptia.embrapa.br/plantannot/find/?q=Unknown+function&selected_facets=organism%3AOropetium+thomaeum&selected_facets=so_term%3Apolypeptide&selected_facets=orthology%3A1&selected_facets=orthologs_coexpression%3Atrue&selected_facets=analyses%3Adiamond+matches&selected_facets=analyses%3Ano+interproscan+matches

Viewing results

4 Visualize the "Results" card on the center-right of the screen. There will be the resulting list of *Oropetium*'s PUFs. 4 PUFs were filtered:

Organism	Feature Type	Feature ID	Relationship	Display	Orthologous Group	Coexpression Group
Oropetium thomaeum	polypeptide	Oropetium_20150105_04788Av1.0	mRNA	PF14365 - Domain of unknown function (DUF4409) (DUF4409) (1 of 10)	plantannot2042	
Oropetium thomaeum	polypeptide	Oropetium_20150105_05984Av1.0	mRNA	PF03478 - Protein of unknown function (DUF295) (DUF295) (1 of 52)	plantannot22362	
Oropetium thomaeum	polypeptide	Oropetium_20150105_05988Av1.0	mRNA	PF03478 - Protein of unknown function (DUF295) (DUF295) (1 of 52)	plantannot22362	
Oropetium thomaeum	polypeptide	Oropetium_20150105_06185Av1.0	mRNA	PF04578//PF13968 - Protein of unknown function, DUF594 (DUF594) // Domain of unknown function (DUF4220) (DUF4220) (1 of 12)	plantannot17852	

