MAY 20, 2020

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

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

Marcos Viana¹, Mauricio Mudadu¹, Adhemar Zerlotini¹

¹EMBRAPA

Marcos Viana

ABSTRACT

OPEN ACCESS

DOI:
dx.doi.org/10.17504/protocols.io.bgdgjs3w

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 B. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.bgdgjs3w>

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:
37000

The Plantannot software provides several filters and a text search box that allows searching for molecules by its desired annotation features. These filters are needed to obtain PUFs and to try to relate them to abiotic stressors using RNA-seq expression data and co-expression networks. The Filters 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 remaining filters, 6 protocols were created as examples of different ways to selecting PUFs. Protocol A: using lack of both homology and protein domain signatures. Protocol B: using lack of homology, presence of domain signatures - trying to select Domains of Unknown Function (DUF) from PFAM, and the text search “Unknown function”. Pr

<p>otocol 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 homologous proteins with co-expression data related to abiotic stress.</p>	
<p>Protocol B is intended to find PUFs from organisms whose proteins are not yet in the NCBI's "nr" database and have DUF domains from PFAM found by InterproScan, selected using the text search "Unknown function".</p>	

- 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 whose proteins are not yet in the NCBI's "nr" database and have PFAM's DUF domains found by InterproScan. Proteins will be selected using the text search "Unknown function".

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 thomaeum* 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
apply remove

☐ 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 "Analyses" filter, select both "no diamond matches" and "interproscan matches", and click "apply" to execute the filter:

Analyses
apply remove

☐ diamond matches (0)
☒ interproscan matches (2,545)
☒ no diamond matches (2,545)
☐ no interproscan matches (0)

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

2.4 Leave the "Orthology" and "Coexpression" and "Orthologs_coexpression" filters empty:

Orthology
apply

☐ no orthology (2,347)
☐ orthology (204)

Coexpression
apply

☐ no co-expression groups (2,551)
☐ co-expression groups (0)

Orthologs_coexpression
apply

☐ no co-expression (2,541)
☐ co-expression (10)

2.5 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)

- 2.6 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.

Unknown function

Q

Filters

Selected filters

Results

Organism	Feature Type	Feature ID	Relationship	Display
<i>Oropetium thomaeum</i>	polypeptide	Protein: W54746.1 (AT3G1410.1) - B1A	DE15459	Protein 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=analyses%3Ainterproscan+matches&selected_facets=analyses%3Ano+diamond+matches

Filters

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

Selected filters

search:Unknown function X

organism:Oropetium thomaeum X

so_term:polypeptide X

analyses:interproscan matches X

analyses:no diamond matches X

https://www.machado.cnptia.embrapa.br/plantannot/find/?q=Unknown+function&selected_facets=organism%3AOropetium+thomaeum&selected_facets=so_term%3Apolypeptide&selected_facets=analyses%3Ainterproscan+matches&selected_facets=analyses%3Ano+diamond+matches

Viewing results

- 4 Visualize the "Results" card on the center-right of the screen, we will have the resulting list of *Oropetium*'s PUFs, 8 PUFs were filtered:

Results

Organism	Feature Type	Feature ID	Relationship	Display	Orthologous Group	Coexpression Group
Oropetium thomaeum	polypeptide	Oropetium_20150105_01078Av1.0	miRNA	PF12023 - Domain of unknown function (DUF3511) (DUF3511) (1 of 6)		
Oropetium thomaeum	polypeptide	Oropetium_20150105_03125Av1.0	miRNA	PF06376 - Protein of unknown function (DUF1070) (DUF1070) (1 of 9)		
Oropetium thomaeum	polypeptide	Oropetium_20150105_03513Av1.0	miRNA	PF03478 - Protein of unknown function (DUF295) (DUF295) (1 of 52)		
Oropetium thomaeum	polypeptide	Oropetium_20150105_06370Av1.0	miRNA	PF08213 - Mitochondrial domain of unknown function (DUF1713) (DUF1713) (1 of 1)		
Oropetium thomaeum	polypeptide	Oropetium_20150105_07160Av1.0	miRNA	PF06533 - Protein of unknown function (DUF1110) (DUF1110) (1 of 4)		
Oropetium thomaeum	polypeptide	Oropetium_20150105_19562Av1.0	miRNA	PF03478 - Protein of unknown function (DUF295) (DUF295) (1 of 52)		
Oropetium thomaeum	polypeptide	Oropetium_20150105_23984Av1.0	miRNA	PF03195 - Protein of unknown function DUF260 (DUF260) (1 of 34)		
Oropetium thomaeum	polypeptide	Oropetium_20150105_24873Av1.0	miRNA	PF05097 - Protein of unknown function (DUF688) (DUF688) (1 of 13)	plantannot66633	