

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

Forked from Find proteins of unknown function (PUF) using Plantannot - Protocol C

Marcos Viana¹, Adhemar Zerlotini¹, Mauricio Mudadu¹

¹Embrapa

Works for me dx.doi.org/10.17504/protocols.io.bgd6js9e

Mauricio Mudadu

ABSTRACT

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 stresses 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". 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 D is intended to find PUFs from organisms whose proteins are not yet public in the NCBI's "nr" database and have no protein domain signatures found by InterproScan. We will use ortholog groups and co-expression networks to relate proteins to abiotic stresses.

EXTERNAL LINK

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

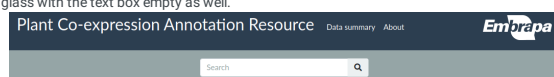
Entering application

1s

- 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>

1s

Filtering

1s

- 2 Find PUFs from organisms whose proteins are not yet public in the NCBI's "nr" database and have no protein domain^{1s} signatures found by InterproScan. We will use ortholog groups and co-expression networks 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^{1s} 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:

1s

Feature type apply

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

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:

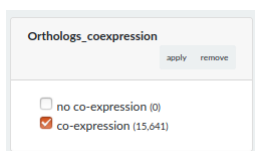
1s

Orthology apply remove

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

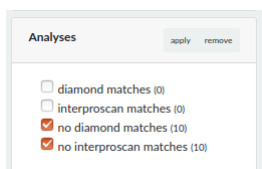
https://www.machado.cnptia.embrapa.br/plantannot/find/?q=&selected_facets=organism%3AOropetium+thomaeum&selected_facets=so_term%3Apolypeptide&selected_facets=orthology%3A1

- 2.4 In the "Orthologs_coexpression" filter, select "co-expression", and click "apply" to execute the filter: 1s



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

- 2.5 In the "Analyses" filter, select both "no diamond matches" and "no interproscan matches", and click "apply" to execute the filter: 1s



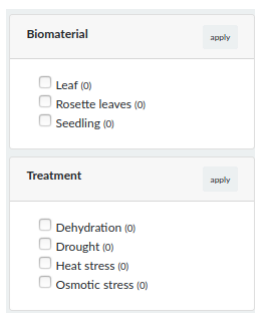
https://www.machado.cnptia.embrapa.br/plantannot/find/?q=&selected_facets=organism%3AOropetium+thomaeum&selected_facets=so_term%3Apolypeptide&selected_facets=orthology%3A1&selected_facets=orthologs_coexpression%3Atrue&selected_facets=analyses%3A%5Bno%20diamond%20matches%2Cno%20interproscan%20matches%5D

- 2.6 Leave the "Coexpression" filter empty: 1s



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

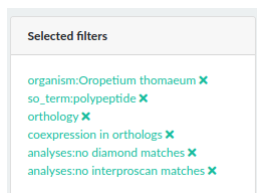
- 2.7 Leave the "Biomaterial" and "Treatment" filters empty: 1s



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%3A%5Bno%20diamond%20matches%2Cno%20interproscan%20matches%5D

Filters 1s

- 3 After execution of all filters we will have the following list of filters: 1s



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%3A%5Bno%20diamond%20matches%2Cno%20interproscan%20matches%5D

Viewing results 1s

- 4 Visualize the "Results" card on the center-right of the screen, we will have the resulting list of *Oropetium*'s PUFs supposedly related to abiotic stress by using orthology and co-expression networks. 10 PUFs were selected: 1s

https://www.machado.cnptia.embrapa.br/plantannot/find/?q=&selected_facets=organism%3AOpuntia+humifolia&selected_facets=so_term%3APolypeptide&selected_facets=orthology%3A1&selected_facets=orthologs_coexpression%3Atrue&selected_facets=analyses%3AAno+diamond+matches&selected_fac