May 21, 2020

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

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

Marcos Viana[1], Adhemar Zerlotini[1], Mauricio Mudadu[1]

[1]Embrapa

👤   Mauricio Mudadu

DOI: dx.doi.org/10.17504/protocols.io.bgd6js9e

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

Protocol status: Working
We use this protocol and it's working

Created: May 14, 2020

Last Modified: May 21, 2020

Protocol Integer ID: 37022

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.
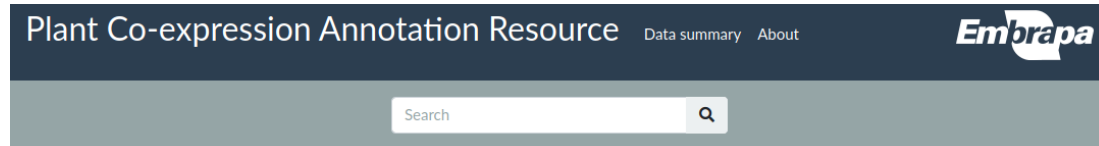
Entering application   1s

1   Enter the Plantannot Result's page, with empty filters and text box search: 1s
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   1s

2   Find PUFs from organisms whose proteins are not yet public in the NCBI's "nr" database   1s
and have no protein domain 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 select all by leaving all boxes empty. We will use *Oropetium tomaeum* as example. Click "apply" to execute the filter:

1s

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%3Apolyp

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%3Apolyp

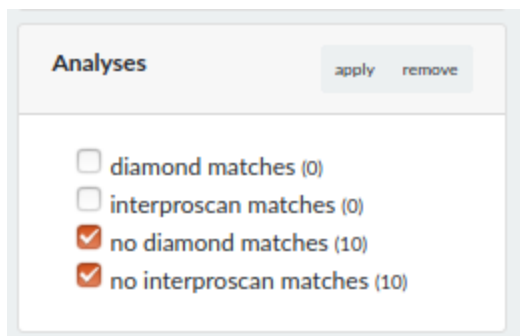**2.4** In the "Orthologs_coexpression" filter, select "co-expression", and click "apply" to execute the filter:

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

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

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

2.6 Leave the "Coexpression" filter empty:

Coexpression apply

☐ no co-expression groups
(18,755)
☐ co-expression groups (0)

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

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

Biomaterial apply

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

Treatment apply

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

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

Filters   1s

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

Selected filters

organism:Oropetium thomaeum ✖
so_term:polypeptide ✖
orthology ✖
coexpression in orthologs ✖
analyses:no diamond matches ✖
analyses:no interproscan matches ✖

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

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`

Results

Organism	Feature Type	Feature ID		Relationship	Display	Orthologous Group	Coexpression Group
Oropetium thomaeum	polypeptide	Oropetium_20150105_03170A.v1.0	mRNA			plantannot17078	
Oropetium thomaeum	polypeptide	Oropetium_20150105_03347A.v1.0	mRNA			plantannot147988	
Oropetium thomaeum	polypeptide	Oropetium_20150105_06293A.v1.0	mRNA			plantannot22668	
Oropetium thomaeum	polypeptide	Oropetium_20150105_07187A.v1.0	mRNA			plantannot9539	
Oropetium thomaeum	polypeptide	Oropetium_20150105_08656A.v1.0	mRNA			plantannot147954	
Oropetium thomaeum	polypeptide	Oropetium_20150105_09096A.v1.0	mRNA		PTHR13020:SF36 - EXPRESSED PROTEIN (1 of 1)	plantannot7088	
Oropetium thomaeum	polypeptide	Oropetium_20150105_10231A.v1.0	mRNA			plantannot148250	
Oropetium thomaeum	polypeptide	Oropetium_20150105_14188A.v1.0	mRNA			plantannot26702	
Oropetium thomaeum	polypeptide	Oropetium_20150105_18142A.v1.0	mRNA			plantannot39098	
Oropetium thomaeum	polypeptide	Oropetium_20150105_24954A.v1.0	mRNA			plantannot21729	

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