

## How to use PAINT, the Protein Annotation and Inference Tool

Requirements: PAINT will run on both PC's and Macs using Java 1.6 (PC name) or Java 6 (Mac name) or higher.

Note: PAINT is currently available as a beta release. Consequently, some bugs persist. Please be flexible and patient.

Obtaining PAINT: PAINT is available by searching SourceForge for "PANTHER" or connecting to the following url: <http://sourceforge.net/projects/pantherdb/files/> . Download and unzip PAINT on your desktop.

Note: Depending on your location, it may be necessary to modify or replace the hibernate.cfg.xml file to use different GO mirrors to speed up PAINT. Choose the closest mirror from Berkeley, Princeton, or EBI.

Launching PAINT:

1) In a new terminal window, run the following command:

```
Desktop/paint1.0_beta32/launchPAINT.sh
```

Note that this assumes you are using PAINT beta 32 and that the PAINT folder is on the Desktop. Modify your command if this is not the case, or cd to the PAINT directory and use this command instead:

```
./launchPAINT.sh
```

(Don't forget the initial "." when typing this command.)

2) Choose File → Login. A dialog box will open with the username and password filled in. Click "OK."

3) Open your protein family of interest. Choose File → Open\_from\_database. A dialog box will appear. If you know the PANTHER family ID of the protein(s) you are interested in, you can enter it into the lower dialog box. Otherwise, you can search for the gene symbol, gene identifier, protein identifier, or description of your gene of interest in the upper dialog box. Note that you will have to select which type of search you would like using the radio buttons. Click the "Search" button to begin the search. One or more family ID's will appear in the "Search Results" box; double-click on the family id to open that family in PAINT.

Opening the family may take several minutes.

## Appearance and Basic Operation

### Windows

PAINT is organized into tabbed panes that may be resized or split out into windows. Initially, the top pane shows a phylogenetic tree on the left and table on the right; you may switch back and forth between the table and a multiple sequence alignment (MSA) using the button above the table.

The bottom pane contains three tabs: Associations, Evidence, and Annotation Matrix. You may also see a minimized tab for Status. Click on a tab to bring it to the front. Click the icons in the tabs or the upper right corner to Undock/Dock, Minimize, Maximize, or close individual tabs or groups of tabs. Tabs and panes may also be rearranged within a window by dragging. Windows may be closed, arranged, or resized in the standard ways.

We recommend that you use PAINT in a 4-window configuration with the Association, Evidence, and Annotation Matrix tabs undocked into separate windows and arranged comfortably on your screen.

### Phylogenetic tree and identifier table

Proteins are arranged in a phylogenetic tree showing their relationships; the tree is aligned with a table showing additional information and linkouts to various databases. You can adjust the relative sizes of each within the window by dragging the dot in the column separating them. Click on the triangles at the top of the central column to expand or retract either of the panes. Note that the identifier table contains a lot of information that can be observed by scrolling to the right.

Proteins with experimental annotations (IDA, EXP, IMP, IGI, IPI, or IEP evidence codes) for a particular ontology are colored and shown in boldface. You may select one ontology at a time to examine using the radio buttons at the top of the window. You may change the color used to indicate annotations using Edit → Curation status colors... . Descriptions here refer to the default colors.

The root and internal nodes of the tree are shown as circles (speciation events) and squares (gene duplication events). The nodes are numbered in a defined order starting with the root, AN0. ("AN" = "Ancestor.") Mouse over a node to see its identifier. If you right-click on a node, a menu will appear with the options to "Collapse or expand node" to hide/show its descendants or "Reroot to node" to focus only on its descendants. As long as there is room, you may reroot downwards; however, the only option to go back up the tree is Tree → Reset Root to Main.

Click on a protein name in the tree to highlight the protein in the tree and the table. Clicking anywhere within a row in the table highlights the protein in the tree and the table; clicking on one of the blue linkouts will also open a link in your web browser. Left-click on a node in the tree to highlight the entire clade descended from it.

### Associations window

Click on a protein in the tree or table. Annotations associated with that protein appear in the Associations window. Click the term name to link out to AmiGO. Click the reference to link out to the reference at PubMed or the appropriate model organism database.

The ECO/QUAL (Evidence code and qualifier) column shows the evidence code supporting the annotation and icons indicating any qualifiers, such as NOT (a red circle with a white X), colocalizes\_with (???), or contributes\_to (???). The icon resembling a ball-and-stick molecular figure indicates an experimental annotation, as opposed to an inferred annotation (see below).

### Annotation matrix

The annotation matrix displays annotations associated with each protein in table format. Click on a protein in the tree and the corresponding row will be highlighted in the matrix. Each column indicates whether an annotation exists for a given term. Red squares indicate experimental annotations, blue squares inferred annotations. Black dots within the squares indicate that the annotation is directly to the term, white dots that the annotation is to a child term. NOT annotations are indicated with the same icon as in the Associations window, a red circle with a white X. Mouse over an annotation square to popup the protein and the term.

In the upper pane, click on a term to highlight it; every protein annotated to that term will also be highlighted. If one protein is annotated to the selected term, its name and annotations will appear in the Associations window; if more than one protein is annotated to the term, the Associations window will indicate, in the upper left corner, which ancestral node is the most recent ancestor to all the annotated proteins.

Technical note: The terms in the annotation matrix are arranged in a non-redundant first in-first out (FIFO) order, with all parent terms shown for each term used. A few very broad terms such as “protein binding” are not shown, even though they are listed in the associations table.

## Evidence window

The evidence window is a text editor used to record notes on the curation process. It is pre-seeded with some boilerplate text.

## “Find” function

The Find function (Edit → Find...) allows you to search for either a gene or a GO term. Select a gene or term search using the radio buttons. Searches are case-insensitive.

A gene search matches against any text in the identifier table. Scroll through the list of matches and click on a specific match to highlight it in the tree, table, and annotation matrix, and to display its annotations in the Associations window.

You may search GO terms using text, or you may numbers to search for GO IDs.

## **Making an inference: Transferring annotations**

To transfer an annotation from a protein to one of its ancestral nodes, drag the GO term from the top pane of the Annotation Matrix window to the ancestral node of choice in the phylogenetic tree. The ancestral node will now be annotated with that term using the evidence code “IDS” (“Inferred from Descendant Sequence”). All descendants of that ancestral node will now be annotated with that term using the evidence code “IAS” (“Inferred from Ancestral Sequence”).

To remove an inferred annotation from the ancestral node to which it was made, click on the node in the phylogenetic tree, and then click on the trash can icon in the Undo column of the Associations window.

You may propagate a NOT annotation if you feel the evidence warrants it for one of the following reasons, listed here in order of decreasing strength:

- 1) There are experimental annotations indicating that a function has been lost from one or more proteins. (Evidence code IDS = Inferred from Descendant Sequences)
- 2) Specific residues have been mutated at, for example, an enzyme’s catalytic site, and a specified function is no longer possible. (Evidence code IMR = Inferred from Mutant Residues)
- 3) A protein or clade has evolved rapidly, losing the original function and gaining a new one. This may be visible as a long branch in the tree, but the meaning of “long” varies by context, and a visibly long branch is not strictly required. (Evidence code IRD = Inferred from Rapid Divergence)

To add a NOT modifier to an inferred annotation, select a node or protein from the tree. In the EC/QUAL column of the Associations window, click on the “IAS” and a popup menu will appear. Under “NOT,” select which evidence code justifies the NOT qualifier. All annotations to proteins and nodes descended from that node will have the NOT qualifier added.

If you do not wish to allow a positive annotation to propagate into a particular clade and you do not wish to make a statement as strong as a NOT annotation, you may block an annotation from propagating into the clade. Instead of selecting a choice from the NOT menu as above, instead choose “STOP.” In addition to the annotation no longer propagating downward, a small hash mark will appear near the node in the tree to indicate that the block exists. Note that a hash mark only indicates the existence of at least one block, *not* that every annotation through that node is blocked.

You may remove a NOT or STOP qualifier by clicking the trash can in the Undo column. Note that you can only remove these qualifiers from the specific node to which they were made.

### **Saving annotations**

File → Save annotations

### **Reloading annotations**

File → Open from files...

### **Troubleshooting**

If the annotation matrix does not function as expected, try closing and reopening its window.

## Useful links

To see existing annotations and notes:

<http://pantree.org/>

PAINT CVS:

<http://cvsweb.geneontology.org/cgi-bin/cvsweb.cgi/go/gene-associations/submission/paint/>

Wnt signaling pathway page, with links to wiki pages with curator questions:

[http://gocwiki.geneontology.org/index.php/Wnt\\_signaling\\_Pathway](http://gocwiki.geneontology.org/index.php/Wnt_signaling_Pathway)

Wnt family questions:

<http://gocwiki.geneontology.org/index.php/PTHR12027>

PAINT wiki manual:

[http://wiki.geneontology.org/index.php?title=PAINT\\_User\\_Guide&action=history](http://wiki.geneontology.org/index.php?title=PAINT_User_Guide&action=history)

Annotation tracker:

<http://amigo.berkeleybop.org/cgi-bin/amigo/phyloree>

or, if you're desperate and AmiGO Labs is down:

<http://amigo-sven.princeton.edu/cgi-bin/amigo/phyloree?mode=cluster&key=PTHR22573&dbname=PantherDB>

PAINT CVS instructions:

[http://apollo.berkeleybop.org/go\\_cvs\\_submission/](http://apollo.berkeleybop.org/go_cvs_submission/)