

SimiTri

This is a java applet to view similarity relationships between sets of genes and three datasets - Similarity Triangle (geddit). The example supplied here shows the blast score relationships of all *C. elegans* proteins to three protein databases (Yeast, Fly and Human). To run you will need to download the files :

```
example.html
elegans.dat
SimiTri.class
```

To run use the command

```
appletviewer example.html
```

The applet displays a similarity triangle in which each *elegans* protein is represented as a coloured tile. Each vertex of the triangle represents a single dataset (Yeast, Fly and Human). The position of the tile represents its relative relationship to each of these datasets, the colour of the tile represents the highest Blast score of that protein to all three datasets. A protein which is found to have a blast score to only two datasets will be found on the line joining the vertices representing the datasets. Proteins which have a Blast score to only one dataset would occupy the vertex representing that dataset and are not shown on the graphic. In calculating the *elegans* dataset only Blast bit scores in excess of 50 were included.

The applet is fully interactive. Left click on a tile to display the gene name in the top left window. Use the slide bar at the bottom to zoom in and out. Use the sliders on the colour bars to isolate datasets within a range of similarity scores. Click on the main view and move the mouse to move the view.

To use SimiTri on your own data you will need to create a flat file similar to *elegans.dat* which has the format :

```
GeneID    X-coordinate    Y-coordinate    Colour (e.g. highest blast
                                score on a scale of 50-500)
```

For the *elegans* data this was calculated using the perl script *SimiTri.pl* using a flatfile of Blast scores in the format :

```
GeneID    Blastscore_db1    Blastscore_db2    Blastscore_db3
```

In essence the core of the script is :

```
$x=250;          ### Coordinates of the center of the triangle
$y=273;
if($a==1) { $a=0; }    ### Similarity score vrs DB A
if($b==1) { $b=0; }    ### Similarity score vrs DB B
if($c==1) { $c=0; }    ### Similarity score vrs DB C
if($a+$b!=0 && $a+$c!=0 && $b+$c!=0)
{
  $A=174;            ### A scaling term specific for the applet
  $A/=( $a+$b+$c );  ### The rest calculates the 2D position
  $y-=( $A*$a );     ### for each point
  $y+=(cos($PI/3))*($A*$b);
  $x-=(sin($PI/3))*($A*$b);
  $y+=(cos($PI/3))*($A*$c);
  $x+=(sin($PI/3))*($A*$c);
}
```

Need help ? - try emailing me

john.parkinson@ed.ac.uk

To reference Simitri us :

Parkinson, J & Blaxter, M. Simitri - Visualising similarity relationships for groups of sequences. Bioinformatics in press.