Ancient Rep Instructions

Vamsee Reddy <vamsee30@gmail.com>

Initial Preparations

1. Create a list of homologs for the family of proteins you are analyzing
2. Generate a multiple alignment using any alignment program (i.e ClustalX). Make sure that there are no large gaps in the middle. If your family has many large gaps in the middle, some proteins will have to be removed. You might have to generate several different groups of alignment files to accurately model the family of interest.
3. Using the AveHAS program (http://www.tcdb.org/progs/?tool=avehas2), generate an average hydropathy plot of your multiple alignment.

Deciding where to make the cut

A single cut needs to be made in your multiple alignments. The cut must be placed so that the two halves contain the repeat.
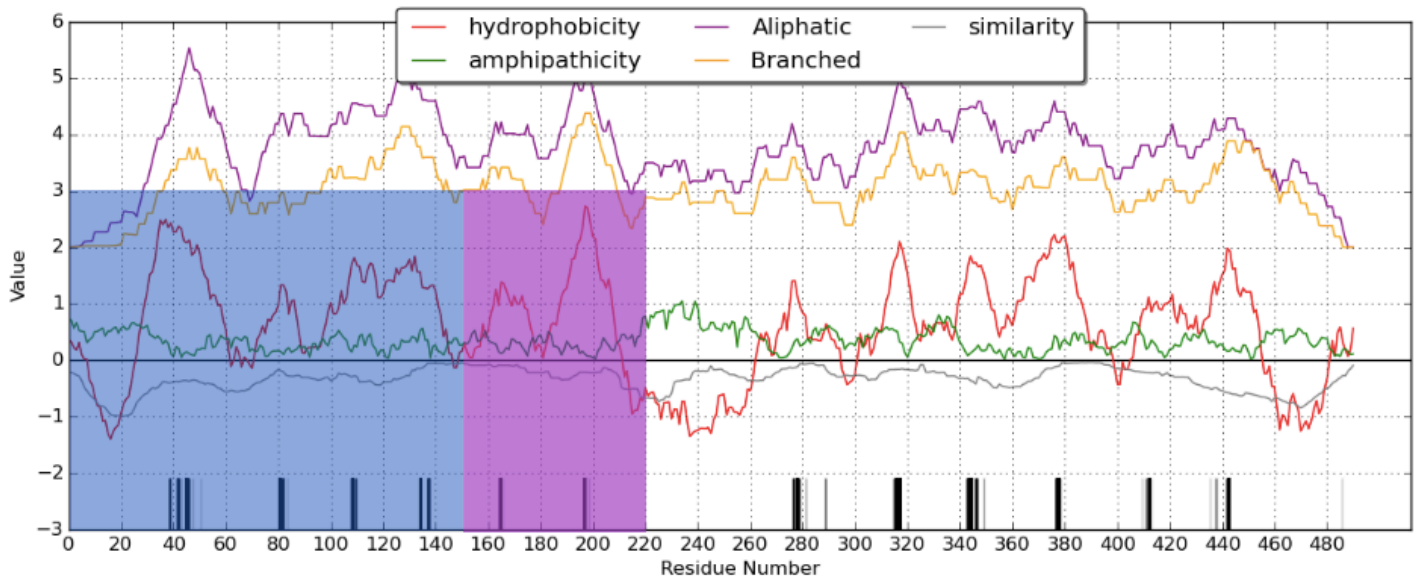
Example 1
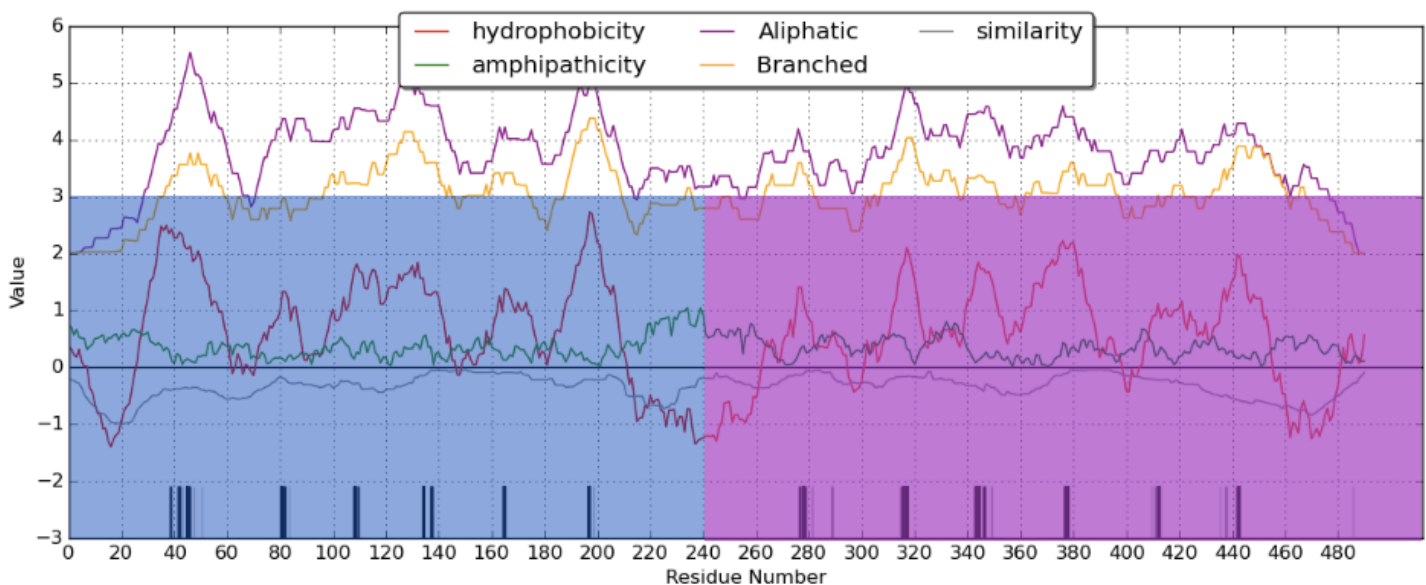
If the repeat is contained somewhere within the blue region and its counterpart somewhere in the purple region, we would compare the two halves illustrated here. For example, if we suspect TMS 1,23 is repeated at 7,8,9 then this delination should be used. The program will automatically find which regions in the blue region are most similar to the regions of the purple region. To communicate this to Ancient, use the following command:

```
ancient.py -i mfs.aln -o mfsout -n 240
```

This command will cut the sequences at position 240 (relative to the multiple alignment) and compare these two halves.
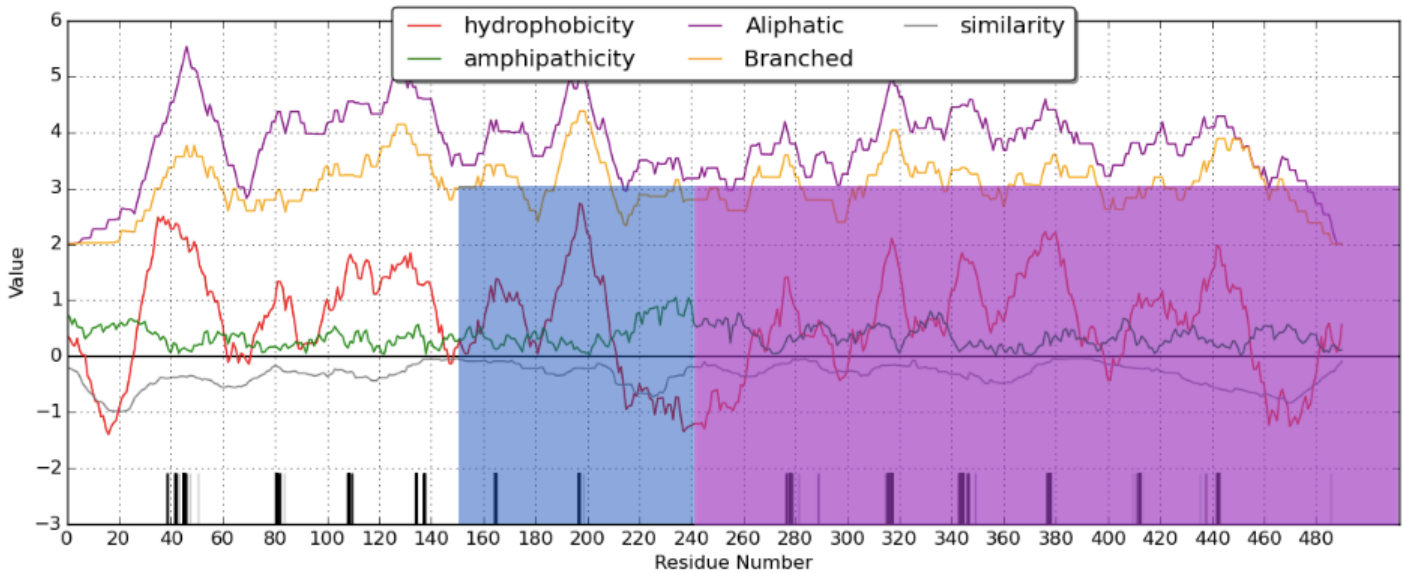
Example 2

You can omit part of the protein if you were interested only in one part of the multiple alignments. In this example, we are interested in the origin of TMSs 5 & 6 with respect to the region marked in blue. To communicate this to the program, we do the same procedure as before. Except, we also tell the program to omit everything after position 220.

```
ancient.py –i mfs.aln –o mfsout –n 150 --omit_after=220
```

Example 3

In this example we are only comparing the region in blue (TMS 5&6) to any region marked in purple. Ancient will identify which regions marked in purple have the most similarity the blue region. As before, we need to place the cut at 240. Except, this time we must omit everything before position 150. The following command conveys this to Ancient:

```
ancient.py -i mfs.aln -o mfsout -n 240 --omit_before=150
```

When to use the "omit" feature

In most cases, simply cutting the alignment in half should work. The program will find the regions of highest similarity and generate a report that contains the repeat. You will only need to restrict the search using the omit feature if you are interested in the origin of older repeats that are more divergent. If these specific repeats are very divergent, then the program might only identify the more recent and obvious repeats when only cut in half. However, by restricting the alignment with the 'omit' tags, you can force the program to consider your hypothesis. You should always run Ancient without the omit feature (example 1) before using the restrict features. It is very likely that it will identify all valid repeats using the default settings. Only follow up with the restrict features if you are certain they exist but they don't appear with the default settings.