**How to run HMMER with adjusted rates program:**

**Items Needed:**

1. Starting HMM: obtained from dfam.org generally. AllHMMs folder contains the complete database locally.
2. PSM File: I have these for 22 human chromosomes. Due to some troubles with the importing of classes in python you will most likely need to create new PSM files. PSM files were created either in simulation or by calling the create PSM files function in the FileMaker.py class. Note changes will need to be made to go to path of new PRM files.
3. Fasta File: This is the file with the DNA a sequence in it to be searched.
4. Installated version of existing HMMER software. <http://hmmer.janelia.org/>

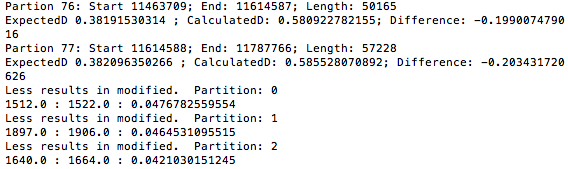
**How to:**

1. The main program is Hmmeraroc.py The following run time parameters
   1. –f the location of the fasta file to be searched. The default is “./FastaFiles/Simulation.fa
   2. –m the location of the HMM file. The default is ./AllHMMs/AluSx.hmm
   3. – b location of the HMMER executables. The default is “/usr/local/bin/”
   4. –p location of PSM file that containts other known repeats. The default is“./PSMs/simulation.psm”
   5. –r location of the results folder. The default is HMMResults followed by the current time and date
   6. –s is a flag indicating that the run is a simulation and additional summary files should be generated
   7. –n is a flag indicating that only changes that reduce the hmm’s probability for original bases
   8. –d is a flag that indicates not to delete the directory containing the temporary files to run hmmer on for each partition
   9. –e is a flag that indicates not to delete the original results file and only keeps the cumulative summary of all the matches

Also just run python hmmeraroc.py –h to get a summary of this.

**What will happen:**

Currently a list of the information of the partitions will be made. In other words, it will indicate the age of the partition vs. expected and how many repeat bases will be made. A quick summary of the results for each partition will then print out.



A random folder will be made to contain all of the utility files to be used by HMMER for these new runs. The name of the folder is a large string of hexnumbers for example f483e144-84ec-43ed-83dd-8879c763f048. Hmmer will run on each set of newly created files and place the results in the folder specified above. In this folder, you fill find 4 files for every partition.

The files will be named: #.res, #.rescleaned, #.resorig, #.resorigcleaned

1. The .res file is the HMMER original results file for the modified HMM calibrated to partition relative rate of change.
2. The .rescleaned file is just the list of repeats found, their statistical values such as e value, and their indices relative to the original fasta file input. Note the .res file will have all indices relative to the temporary files. If the start index comes after the end index, it means the repeat is on the reverse strand.
3. The .resorig and .resorigcleaned files are the parallel files for the unmodified Hmm.

Also in the folder will be the cumulative.rescleaned and the cumulative.resorigcleaned files which are just a concatenation of all the .rescleaned and .resorigcleaned files. If you are running a simulation other files called summary.txt, summaryOfMod.txt, and summaryOfOrig.txt. Since we know where the repeats are in these cases. More detail is given about bases matched, total number of bases that could be matched, bases missed, bases identified as a match but actually weren’t.

