

Work Log for September

Logan Brown

September 8, 2014

1 Week of September 1st-5th

1.1 Goals for the Week

1. Connect codons with Δ eta values
2. Liz Howell Code
3. Further readings in the R user manual
4. NSE Model

1.2 Progress/Notes

1.2.1 Connect codons with delta eta values

I made a change to `run_roc.r`, line 314. I changed
`mean.b.mat <- cbind(bmat.names, mean.b.mat, sd.b.mat)`
to

```
mean.b.mat <- cbind(names(results$chains[[1]]$b.Mat[[1]]), mean.b.mat, sd.b.mat)
```

`bmat.names` is just the synonym group of the codon. `names(results$chains[[1]]$b.Mat[[1]])` is of the format `aminoacid.codon.value` (for example `A.GCC.log.mu` or `H.CAC.Delta.t` where `log.mu` represents the natural log of the mutation rate and `Delta.t` represents the change in the pausing times (compared to some reference codon). This information gets written to

A note on the reference codon: Cedric claimed the one with the shortest pausing time was the reference codon, I've found it's the last one alphabetically.

1.2.2 Liz Howell Code

The code seems to be done. It can be checked out from github at
<https://github.com/ozway/optimal-pessimal.git>

It takes in a gene (by default, named "ecoli.fasta"), and can write the optimal or pessimal version(s) of the genome using `makeOptimal.r`, `makePessimal.r`, or `makeBoth.r`. The optimal is written to `optimalEcoli.fasta`, and the pessimal is written to

pessimalecoli.fasta. writeOptimal and writePessimalecoli takes about 5 minutes and 49 seconds to run on the ecoli gene. writeBoth takes about 7 minutes and 45 seconds to run on the ecoli gene, likely due to the increased file writing.

Future Considerations:

1. (Add a config.r file for setting global variables like input filenames and output filenames) Done
2. Calculate the total amount of time gained or lost in the changes
3. Do we need to switch from ROC to η ?
4. Try for other genes?

I've also written up readme instructions for how to execute the code.

1.2.3 Further readings in the R user manual

In order to write the Liz Howell code, I had to do more readings into the R user manual. I read deeply into, and played around with

- Matrices
- Lists
- String Manipulation
- the `[]` and `[[[]]` operators
- for and while loops

1.2.4 NSE Model

Following Wei-Chen's instructions, I attempted to run a cubfits NSE model.

1. Change the cubmultichain and cubsinglechain calls on lines 188, 195, 205, 212, changing model="roc" to model="nsef"
2. Before each of those lines, add `.CF.CT$model <- "nsef";`

And then running it as normal.

9/5: I started the model at at 9:02. After generating very little output, it ended at 12:38. It parsed all the inputs and such, but generated no results.

1.3 Goals for next Week

1. NSE Model