

- E coli dataset from Bloom
 - * Has tree.
 - * Laboratory population from single ancestor evolved over time.
- Bloom created tree using one gene using public datasets on E coli.
- Deep sequencing information on one gene.
- Lenski student at UNH's data?

2 ROC SEMPPR

Cedric

Project State:

- Drop in s_ϕ correlates with drop in posterior pr.
- Did not post question on discussion board.
- If s_ϕ changes while ϕ are fixed, the problem disappears.

Current Tasks:

- Look at Gelman paper that Russ quoted. Maybe due to improper prior due to s_ϕ prior on log scale.
- Assuming flat prior on log scale. Is this leading to the problem.
- Is proposal ratio correct? Check with Russ.
- Why doesn't this problem appear in the one category runs?
- Two s_ϕ values accepted and rejected together.
- TODO
 - Meet Russ
 - Read Gelman paper
 - Look at acceptance ratio and compare to log posterior trace.
 - Change position where Pr trace is calculated.
- After more discussion with Cedric, it has become clear that the model fit calculation ignores the gene specific probabilities of being in a category. That is, in calculating the plotted posterior pr., he is using p_i where i is the category as opposed to $p_{i,j} = \sum_i p_i \frac{\text{Lik}(\vec{c}_j|i)}{\sum_k \text{Lik}(\vec{c}_j|k)}$.
TODO: Compare model fit calculations using p_i and $p_{i,j}$, does the model posterior pr actually drop when using $p_{i,j}$?

Future Tasks:

3 31 Mar 2016

Cedric

Project State:

- Drops seen in previous runs disappeared!
- However, in a fixed, observed phi run with the Kluyveri genome cedric saw a drop of a couple 100 LLik units and is re-running that analysis.
- This fixed ϕ runs gave good categorization but started with good categorization.

Current Tasks:

- TODO: flag for doing this or set $s_\epsilon = 0$. Flag seems easier.
- Started run where initial categorization is random.
- If we see drop. Focus on understanding it better.
 - Does fixing s_ϕ to a constant value fix this? (we expect so)
 - Does using a non-flat prior on s_ϕ , such as `IGamma()`, fix this?
 - How does the drop model fit relate to log of acceptance ratios?
- Fixed or known ϕ with unknown assignment.
- With ϕ runs starting with known & unknown assignment.
- Make proposal using covariance rather than just variance (1 day).
- Altered prior for s_ϕ .

Future Tasks:

04 April 2016

1 SelAC

Cedric

- Cedric looking at E coli paper (Fernberg ... ostermeyer, MBE 2014) as test for SelAC.
- Jessie Bloom used same data for simulating of sequence evolution.
- Synonymous CUB needs to be added by Me.
- Getting deep sequencing data ready to run.

04 April 2016

Jeremy Beaulieu

Project State: Jobs still running

Current Tasks: UCEs (SelON)

- Problems with Ne
- Smoothing counts? Yes, try using counts of 1/1000.
- Russ had good ideas of using the observed variance at each site to estimate the strength of selection. Make the most sense for coming up with initial values for model fitting.

Future Tasks: Add smoothing, think about if optimal nt changes over tree branches (simpler than opt aa changing).

GRANT IDEA (from Russ): Combine UCE and protein sequence data for phylogenetic inference.

2 SEMPPR

Cedric

Project State:

- Model working well if you know ϕ exactly. Doesn't matter if your initial categorization is poor or good.
(Is this paper? Russ thinks so. Cedric likes the idea as well.). Categorization (left vs. non-left) seems to be working well.
- Are the updates right in the with ϕ model? Russ is concerned about this.
- Regarding the drop in model fit
 - He cannot recreate the problem and only saw it once so for now we set this issue aside.
 - Wonky acceptance ratio for low acceptance Pr are similar.
- With ϕ run: Started with known categorization, known assignment, SCUO for initial ϕ . Ends up looking like a w/o phi run.
 - Running fit longer.
 - Need to look at s_e

Current Tasks:

- TODO: Fix s_e to value estimate for B. yeast and then do order of magnitude smaller.