

Rich von der Schmidt
3-24-16
DNA Analysis
TA: Efe Aras

Part 1:

- Scientists often forget the power of words. Empirical evidence can fall short in the way of convincing other people, just like the OJ Simpson Trial. DNA evidence didn't change the outcome of the trial now did it? Where is the role of creativity in scientific ingenuity if we just rot in front of Microsoft Excel.
- We can't live an echo chamber, and justify science with science all the time, and I did in fact do this analysis in the fall, so why not have a little fun with this one. You have to read through all these Efe so I might as well try to earn a smile on the other end in a way that graphs and regression equations won't.
- Why is the run time $O(n + B \cdot l)$, and how to show this when the size of the strand and the number of the occurrences of the splicee are linearly related?
- The intuition here is to attempt a flow proof
- Let's say the size of the splicee is the size of the strand itself? That is, what if n and B are equal? And $l = 1$?
- These costs are equal in so far as $n = b$, so clearly, the run time is shared between these two
- Now, what if the string isn't found? The number of occurrences is 0, so the B term drops out, and we just need to traverse the whole strand to find out that the splicee isn't there. So the entire cost is born by the n term
- We can't find a way of having $n = 0$, without b also necessarily being 0, and that is why this is a difficult relationship to demonstrate
- QED

Part 2:

- See attached text file

Part 3:

- Same deal
- Why is LinkStrand faster?
- Ahhhhhh the beauty of a LinkedList which I understand even better the second time around. This is just way easier to debug for starters. Sorry sorry, losing my narrative voice here.
- LinkStrand is faster precisely because the run time does not bear the cost of inserting the entirety of the string- you just need to update two pointers, regardless of the size of the splicee that is coming in.