Hw1:

1. Read in fasta file
2. Part 1: buildsa
   1. Make suffix array and write string, suffix array (and maybe lookup table—prefex table) to file in binary
   2. Compute statistics and write to text
   3. Report
      1. What did you find to be the most challenging part of implementing the buildsa program?
      2. For references of various size:
         1. How long does building the suffix array take?
         2. How large is the resulting serialized file?
         3. For the times and sizes above, how do they change if you use the --preftab option with some different values of k?
         4. Given the scaling above, how large of a genome do you think you could construct the suffix array for on a machine with 32GB of RAM, why?
3. Part 2: querysa
   1. Read in suffix array and FASTA –query with naïve or simple accelerant algorithm
   2. Output query results (original query name and hits)
   3. Do not time query reading and writing
   4. Report
      1. What did you find to be the most challenging part of implementing the buildsa program?
      2. For references of various size:
         1. How long does building the suffix array take?
         2. How large is the resulting serialized file?
         3. For the times and sizes above, how do they change if you use the --preftab option with some different values of k?
         4. Given the scaling above, how large of a genome do you think you could construct the suffix array for on a machine with 32GB of RAM, why?