SimpleStrand

Time vs Recomb Size

A scatter plot titled "Time vs Recomb Size" with y-axis labeled "Time (s)" ranging from 0 to 1.4, and x-axis labeled "Recomb Size (nucleotides)" ranging from 0 to 400000000. The data points show an approximately linear relationship, with points near (25000000, 0.1), (50000000, 0.15), (95000000, 0.3), (185000000, 0.58), and (380000000, 1.2).
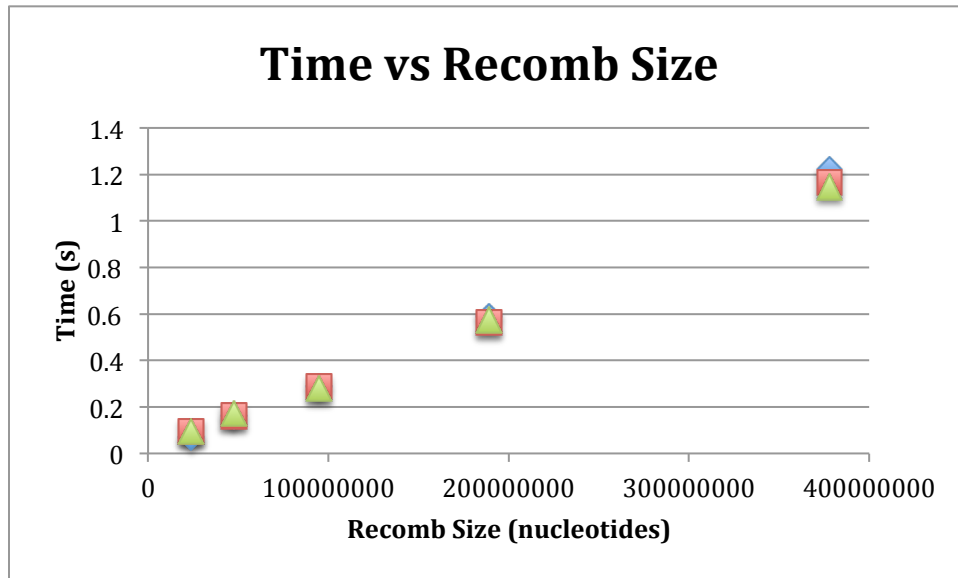
Running DNABenchmark.java using a SimpleStrand implementation with ecoli_small and an -Xmx4096M heap-size produces the graph above, which shows that an linear relationship between Recomb Size and Time. The graph begins with data points with a Recomb Size above 20 million nucleotides and includes all data points until DNABenchmark threw a memory error. DNABenchmark.java was run three times with these specifications, and each color represents a different trial. All data points fit with the conclusion that SimpleStrand is O(N).

The largest splicee string that works without a memory error using the JVM with an -Xmx512M heap-size with the input file ecoli.dat is 65,536 base pairs. This result was determined using DNABenchmark.java, which started by testing cutAndSplice() with a splicee strand that was 2^8 base pairs long and increased by a power of two until it reached 2^16, which is 65,536 base pairs, at which point the memory error was thrown. On my machine it took .136 seconds to determine this. When the heap size was doubled to -Xmx1024M, my machine was able to support the next power of two strand, which took .257 seconds to do with ecoli.dat.

LinkedStrand

Time vs Number of Appends
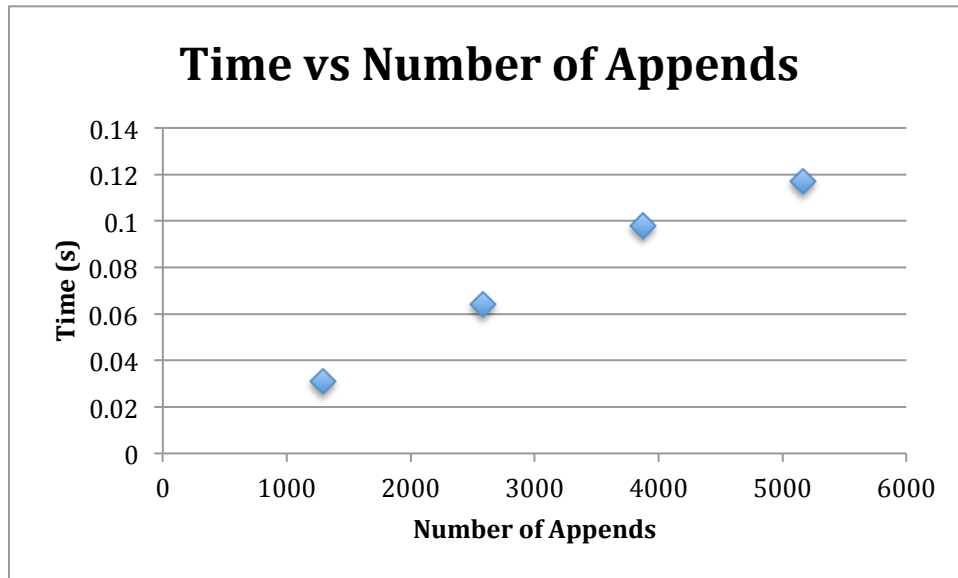
DNABenchmark was run with LinkedStrand implementation on ecoli.dat, ecoli x2.dat, ecoli 3.dat, and ecoli x4.dat, which represent various times that the DNA information from ecoli.dat was copied. The average of the time from 9 different recombination operations on the same file was used for time because each recombination had the same time, regardless of the size of the splicee, since the number of appends was the same. Number of appends was used because it was proportional to the number of breaks in the DNA sequence. The data points support the conclusion that LinkStrand is O(B).