Connor Ghazaleh
DNA Assignment
Due 11/6/17

Non-LinkedList Hypothesis

For this part of the data analysis, a new class was made (DNABenchmark2) to process the IDnaStrand objects. It retained some of the code from the original DNABenchmark class, but included some important changes. The general structure of the code is as follows.

Instance variables are set at the top for ease of access later, then a method dnaData was created to generate strings that represent DNA strands. It creates these strings using the StringBuilder class and first, creates one large StringBuiler object by appending a "c" in a for loop how ever many times needed to achieve the desired length, but then in a second for loop uses the replace method to replace some of those characters with a different character "t", which is used to identify splice locations. The final output of the method is then converted to a string and returned.

The main method first calls dnaData to create the DNA strand, and then uses that string to initialize either a StringStrand or StringBuilderStrand object. Following that, the time is recorded using System.nanoTime(), the object is cut and splices using the cutAndSplice() method, and time is recorded one final time. The first and second recordings of time are then used to calculate the time it took to execute the cutAndSplice() method. The final time is then printed. To test the hypotheses in each of the first two parts of the analysis, some of the initial variables were either changed or maintained constant.

Runtime data for NLLH was collected maintaining the length of the overall string, number of trials, and size of the splicee constant at 10,000 characters, 10 trials, and 3 characters respectively.

	Runtime (seconds)		
Number of Splices	StringStrand	StringBuilderStrand	LinkStrand
1000	$3.73(10)^{-3}$	$2.86(10)^{-4}$	$4.99(10)^{-5}$
2000	$9.76(10)^{-3}$	$4.51(10)^{-4}$	$5.47(10)^{-5}$
3000	$2.08(10)^{-2}$	$6.02(10)^{-4}$	$6.37(10)^{-5}$
4000	$4.04(10)^{-2}$	$7.21(10)^{-4}$	$6.97(10)^{-5}$
5000	$6.50(10)^{-2}$	$8.74(10)^{-4}$	$7.12(10)^{-5}$

The data illustrates a pattern showing that StringBuilderStrand executes in O(bS) time and String Strand executes in $O(b^2S)$ time where b represents the number of splices and S represents the length of the splicee.

LinkStrand Hypothesis:

Runtime data for LSH was collected maintaining the length of the overall string and number of trials constant at 10,000 characters, 10 trials respectively.

Number of Splices	Size of splice (in characters)	LinkStrand Runtime (seconds)
1000	1	$5.05(10)^{-5}$
	2	$5.11(10)^{-5}$
	3	$5.07(10)^{-5}$
	4	$5.08(10)^{-5}$
	5	$5.05(10)^{-5}$
2000	1	$5.30(10)^{-5}$
	2	$5.32(10)^{-5}$
	3	$5.28(10)^{-5}$
	4	$5.32(10)^{-5}$
	5	$5.35(10)^{-5}$

The data illustrates a pattern showing that LinkStrand executes in O(b) time where b represents the number of splices.

Timing Results:

From StringStrand:

dna length = 4,639,221
cutting at enzyme gaattc

Class	splicee	recomb	time	appends
StringStrand:	256	4,800,471	3.468	1290
StringStrand:	512	4,965,591	3.716	1290
StringStrand:	1,024	5,295,831	3.974	1290
StringStrand:	2,048	5,956,311	4.379	1290
StringStrand:	4,096	7,277,271	5.386	1290
StringStrand:	8,192	9,919,191	7.346	1290
StringStrand:	16,384	15,203,031	11.506	1290
StringStrand:	32,768	25,770,711	21.054	1290
StringStrand:	65,536	46,906,071	47.727	1290
StringStrand:	131,072	89,176,791	115.106	1290
StringStrand:	262,144	173,718,231	240.757	1290
StringStrand:	524,288	342,801,111	581.569	1290

From StringBuilderStrand:

dna length = 4,639,221
cutting at enzyme gaattc

Class	splicee	recomb	time	appends
StringBuilderStrand:	256	4,800,471	0.041	1290
StringBuilderStrand:	512	4,965,591	0.035	1290
StringBuilderStrand:	1,024	5,295,831	0.050	1290
StringBuilderStrand:	2,048	5,956,311	0.024	1290
StringBuilderStrand:	4,096	7,277,271	0.034	1290
StringBuilderStrand:	8,192	9,919,191	0.028	1290
StringBuilderStrand:	16,384	15,203,031	0.062	1290
StringBuilderStrand:	32,768	25,770,711	0.092	1290
StringBuilderStrand:	65,536	46,906,071	0.165	1290
StringBuilderStrand:	131,072	89,176,791	0.249	1290
StringBuilderStrand:	262,144	173,718,231	0.536	1290
StringBuilderStrand:	524,288	342,801,111	0.934	1290

From LinkStrand:

dna length = 4,639,221
cutting at enzyme gaattc

Class	splicee	recomb	time	appends
LinkStrand:	256	4,800,471	0.028	1290
LinkStrand:	512	4,965,591	0.025	1290
LinkStrand:	1,024	5,295,831	0.031	1290
LinkStrand:	2,048	5,956,311	0.026	1290
LinkStrand:	4,096	7,277,271	0.035	1290
LinkStrand:	8,192	9,919,191	0.040	1290
LinkStrand:	16,384	15,203,031	0.026	1290
LinkStrand:	32,768	25,770,711	0.027	1290
LinkStrand:	65,536	46,906,071	0.031	1290
LinkStrand:	131,072	89,176,791	0.024	1290
LinkStrand:	262,144	173,718,231	0.024	1290
LinkStrand:	524,288	342,801,111	0.026	1290
LinkStrand:	1,048,576	680,966,871	0.038	1290
LinkStrand:	2,097,152	1,357,298,391	0.035	1290
LinkStrand:	4,194,304	2,709,961,431	0.033	1290
LinkStrand:	8,388,608	5,415,287,511	0.027	1290
LinkStrand:	16,777,216	10,825,939,671	0.026	1290
LinkStrand:	33,554,432	21,647,243,991	0.027	1290
LinkStrand:	67,108,864	43,289,852,631	0.041	1290
LinkStrand:	134,217,728	86,575,069,911	0.026	1290
LinkStrand:	268,435,456	173,145,504,471	0.036	1290
LinkStrand:	536,870,912	346,286,373,591	0.036	1290