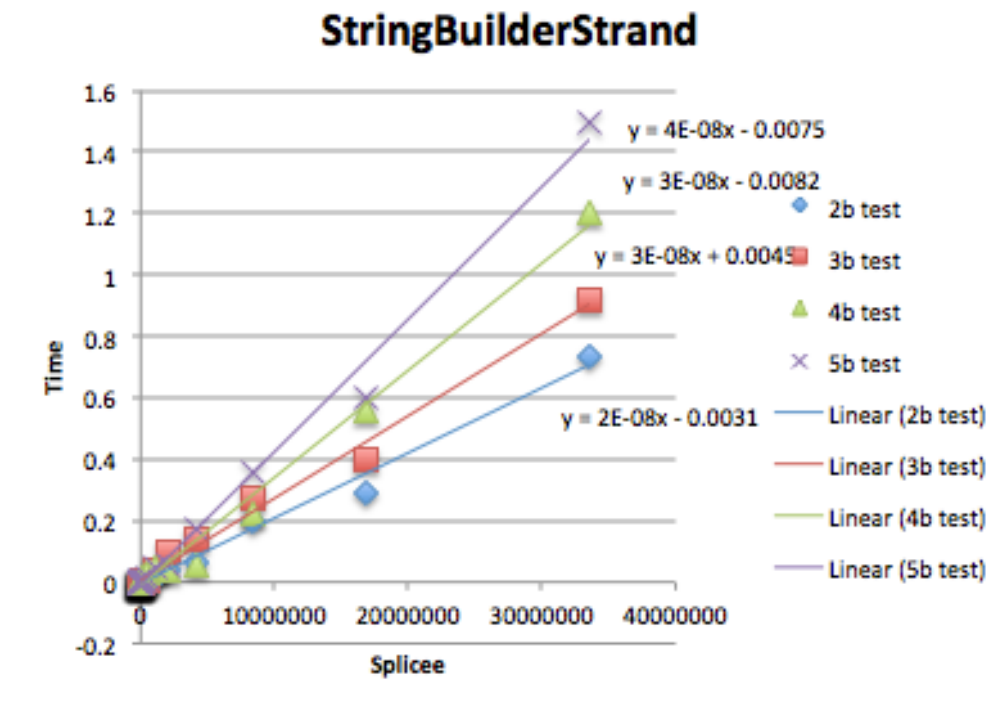
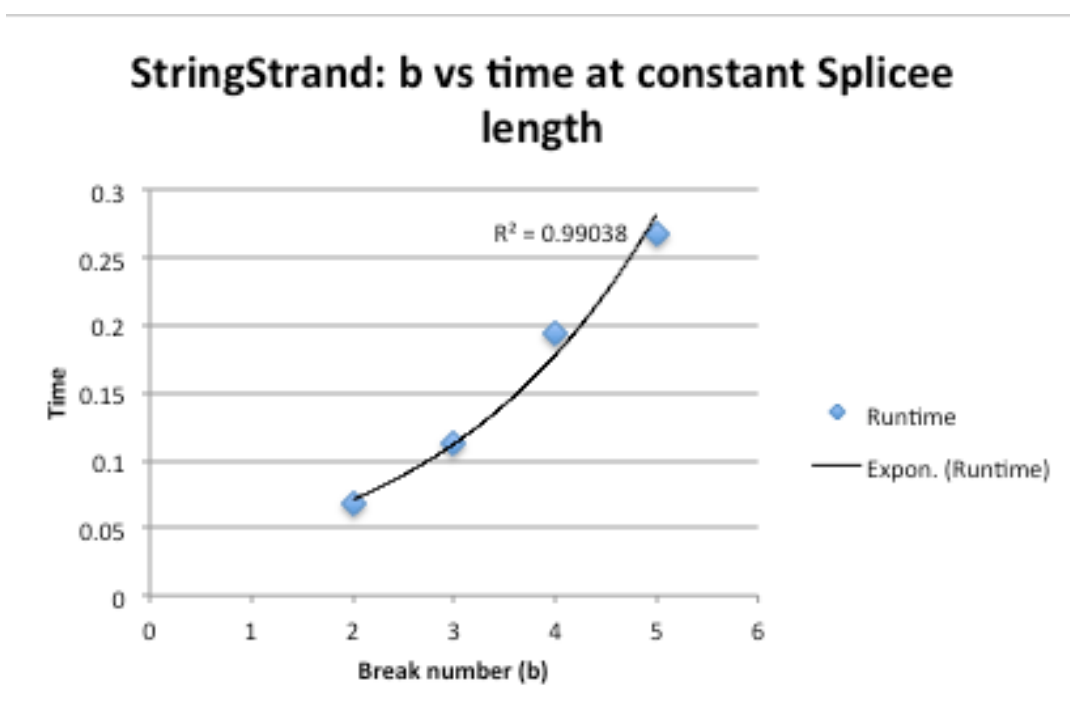
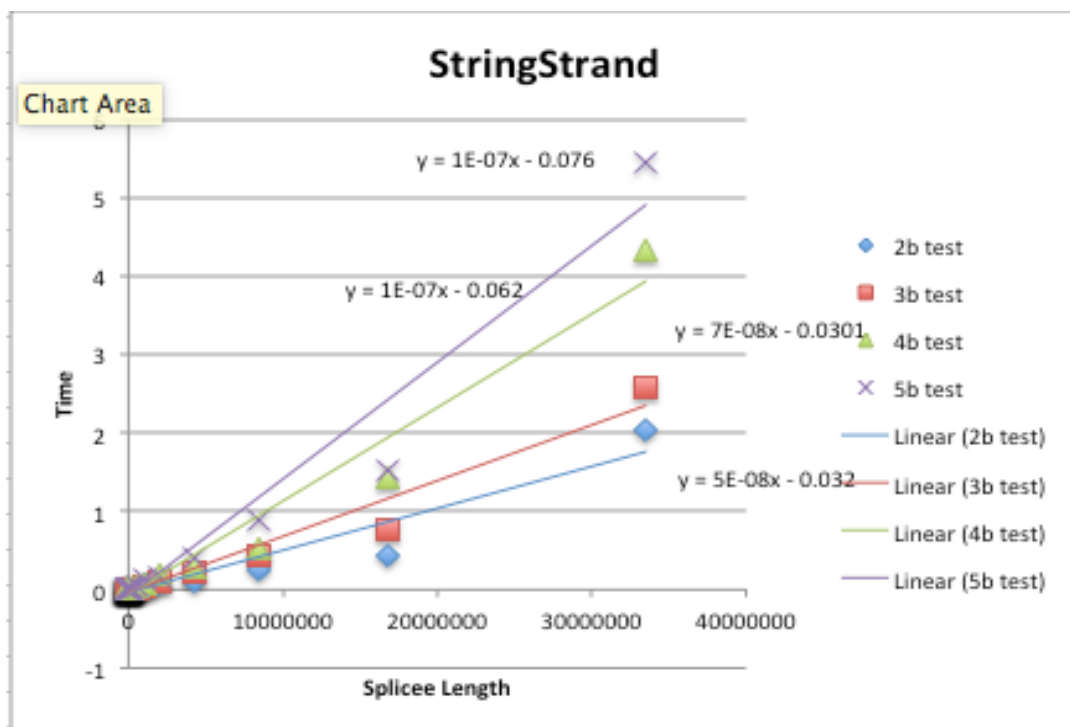


## DNA Analysis

### Non-Linked Strand Hypothesis

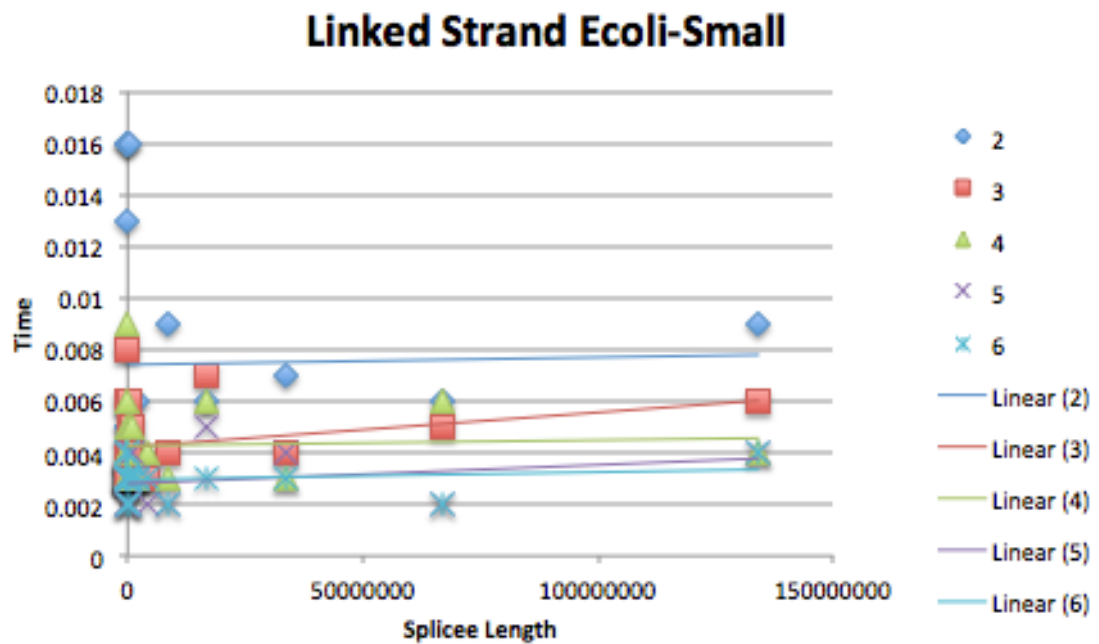
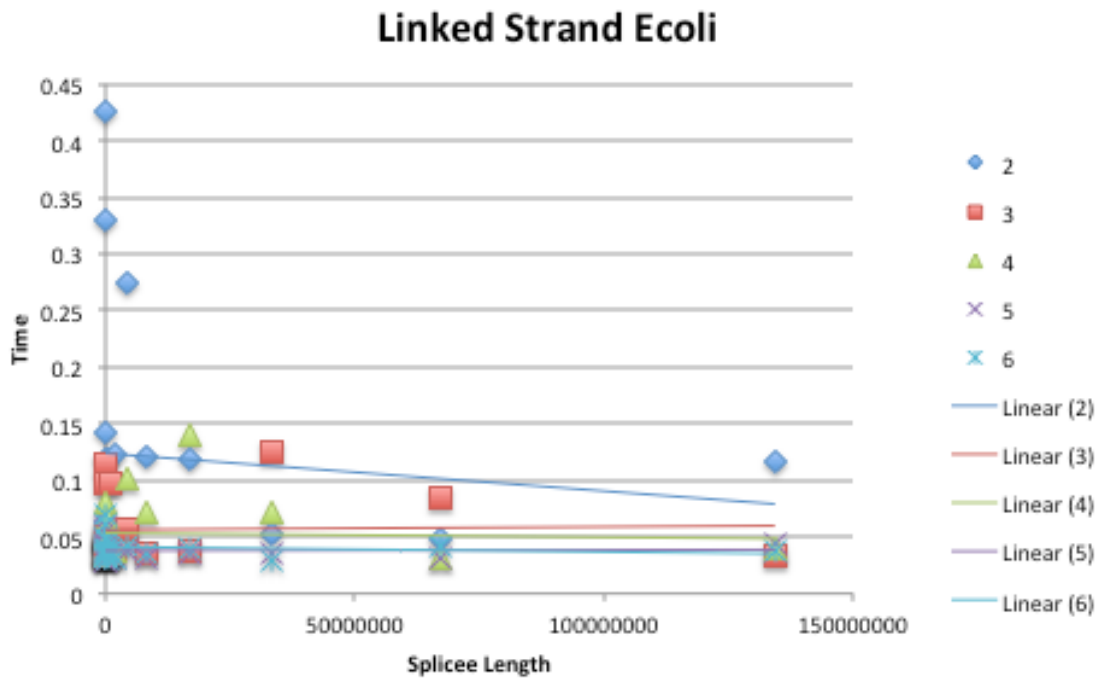


The above graph represents the correlation between the splicee length in the DNA strand and the run time for StringBuilderStrand. The test for this graph used a new text file with a DNA strand of length 25 (N) and a varying number of cut points (b). The cut points ranged from 2 to 5 and the linear trend line for each test is shown. This experimental data supports the non Linked List Hypothesis that StringBuilderStrand would have a runtime of  $O(bS)$  where S is length of splicee. The positive linear trends of the 2b, 3b, 4b, and 5b tests prove the linear correlation of the length of splicee on runtime. These trend lines also increase in slope as b increases across the different test. The similar slope between 3b and 4b trend lines can be attributed to excel being unable to show a more specific number with such a low value, however the graphs clearly show a steeper slope at 4b that is concurrent with the increases between the other trend lines. This constant increase in slope between tests as b increases proves the linear effect the number of breaks has on the runtime. These two linear trends combine to prove the hypothesis that the run time for StringBuilderStrand is  $O(bS)$ .



The first graph represents the correlation between splicee length and runtime in String Strand. The four separate data sets in this graph represent four test of this correlation, each with a constant N value of 25, and with a varying b value (break points for enzyme) as seen in the key. The positive slopes of these linear trend lines show the positive linear correlation between splice length (S) and runtime in each of the tests, proving the splicee length aspect of the non Linked Strand hypothesis of a runtime being  $O(b^2S)$ . The second graph represents the correlation between the number of breaks (b) and the runtime when S is held constant at 2097152. An exponential trend line was added to this test with an  $R^2$  value of 0.99. This strong exponential correlation helps to prove the second part of this hypothesis, that runtime for StrandString has a  $b^2$  aspect. These two graphs together prove the hypothesis of a runtime  $O(b^2S)$  for StrandString.

**LinkStrand Hypothesis:** LinkStrand runs in  $O(b)$  time when  $N$  is fixed, regardless of the size of the splicee.



The above graphs prove the LinkStrand Hypothesis, that with a fixed length of DNA (N), LinkStrand will run in time  $O(b)$  based solely on the number of breaks (b) in the DNA regardless of the size of the splicee. These two graphs show the relationship between splicee length and time for five tests using two different sized DNA texts, the ecoli text and the ecoli\_small text. Each test used a different length enzyme (from 2 – 6 as seen by the key in the graph), causing the number of breaks to decrease as the enzyme size goes up. The horizontal slope of the trend lines for these tests proves that the runtime for LinkStrand is independent of splicee length, since as splicee length increases, there is no increase in runtime overall. The slight slope in some of these lines can be attributed to the runtime being so small and the program only being able to calculate runtime to the nanosecond causing slight discrepancies, however the overall horizontal trend can still be seen. As b length increases between tests, there is a decrease in the number of appends in

| Enzyme Length (b) | Number of Appends Ecoli_small | Number of Appends Ecoli |
|-------------------|-------------------------------|-------------------------|
| 2                 | 40801                         | 619584                  |
| 3                 | 9752                          | 153948                  |
| 4                 | 2698                          | 43780                   |
| 5                 | 740                           | 11544                   |
| 6                 | 210                           | 3202                    |

the DNA sequence, as seen by the table to the left. The decrease in appends correlates to decreasing runtime between tests as seen by the order of the trend lines in each graph. Therefore, as the number of breaks in the DNA strand decreases, the runtime decreases, proving the Linked Strand hypothesis of a runtime of  $O(b)$ .