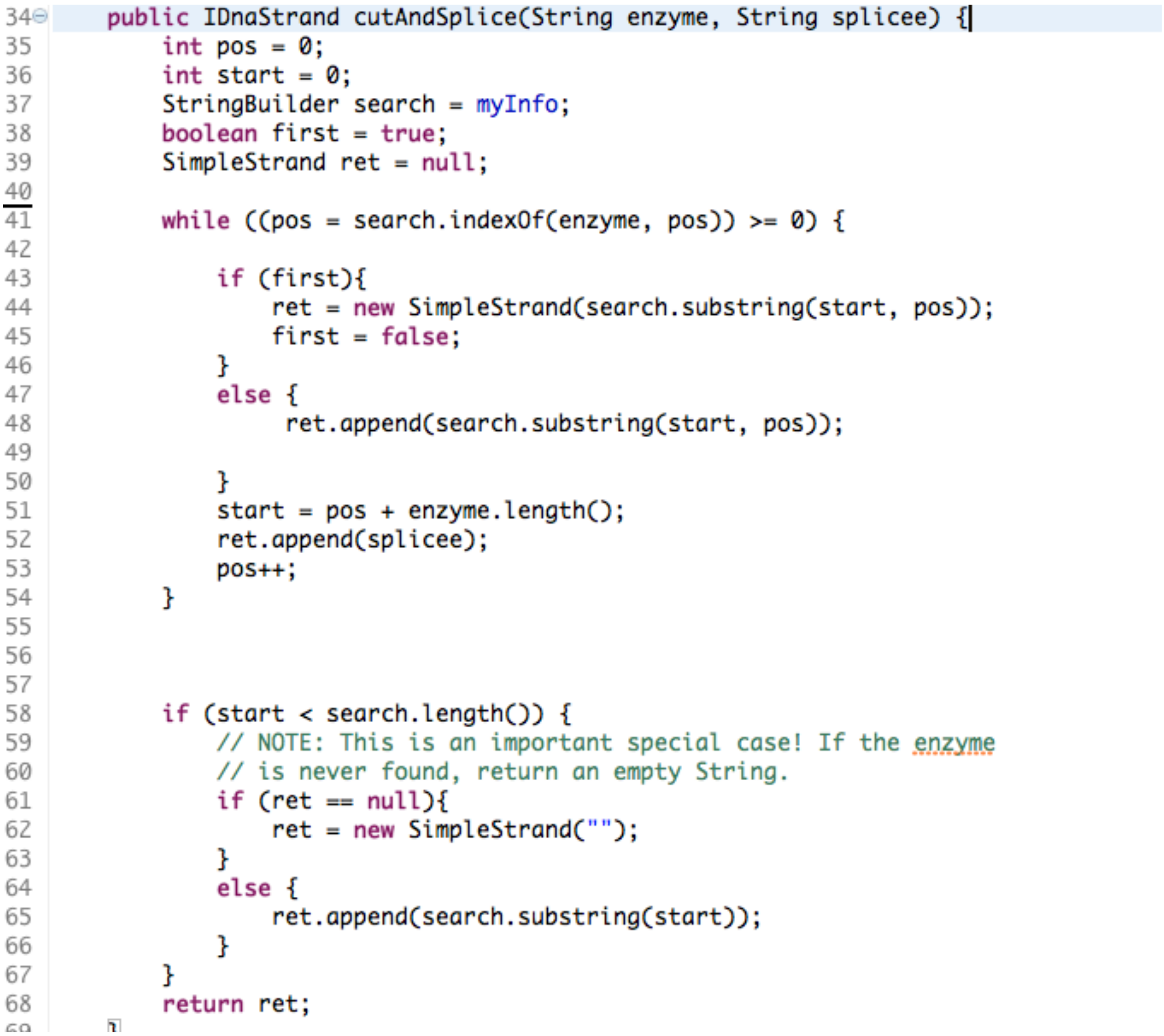
1)
Benchmark

The code in SimpleStrand.cutAndSplice is shown below:



```
    public IDnaStrand cutAndSplice(String enzyme, String splicee) {
        int pos = 0;
        int start = 0;
        StringBuilder search = myInfo;
        boolean first = true;
        SimpleStrand ret = null;

        while ((pos = search.indexOf(enzyme, pos)) >= 0) {

            if (first){
                ret = new SimpleStrand(search.substring(start, pos));
                first = false;
            }
            else {
                 ret.append(search.substring(start, pos));

            }
            start = pos + enzyme.length();
            ret.append(splicee);
            pos++;
        }



        if (start < search.length()) {
            // NOTE: This is an important special case! If the enzyme
            // is never found, return an empty String.
            if (ret == null){
                ret = new SimpleStrand("");
            }
            else {
                ret.append(search.substring(start));
            }
        }
        return ret;
    }
```

O(N) Hypothesis

When calling the cutAndSplice in the SimpleStrand class, a splicee is appended to ret. Because ret is of type StringBuilder, whenever a splicee of size S is appended, each operation will be at least O(S). Since ret is built from null, this implies that at the end of n operations using splices of size S, it will take O(Sn) time, where Sn is the total length of the final recombinant strand (ret).

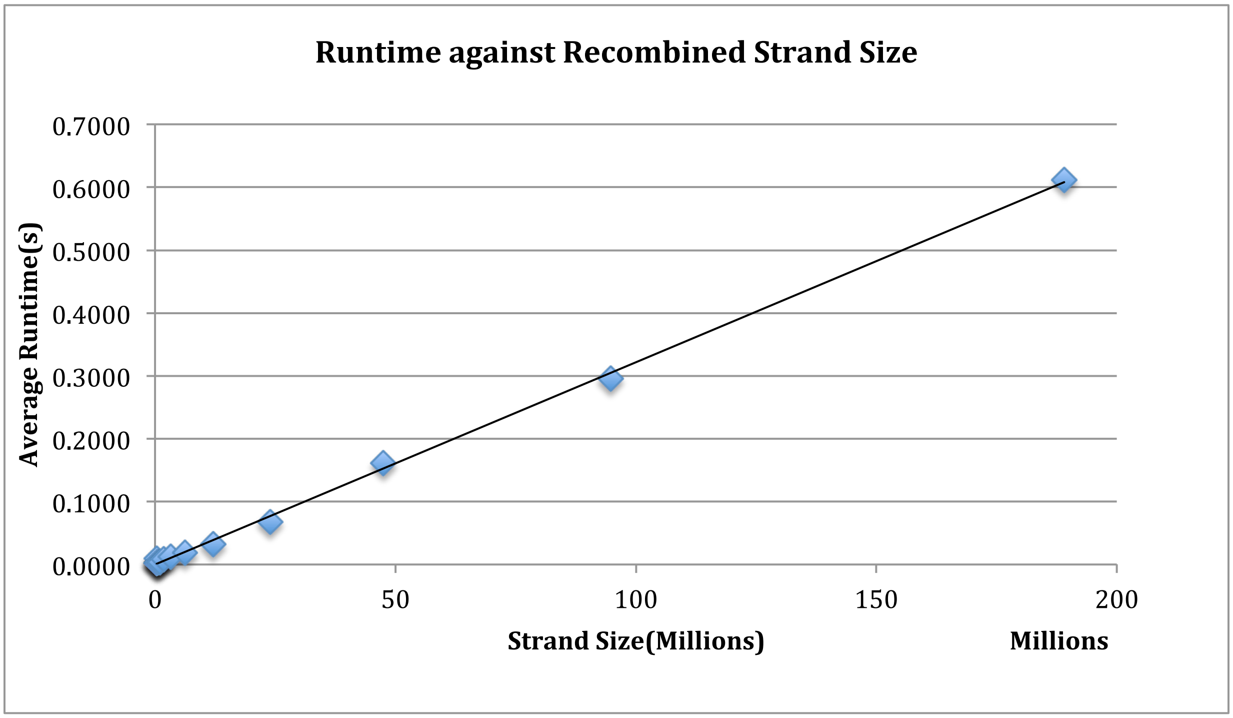
The SimpleStrand.cutAndSplice code was BenchMaked using the class DNABenchMark, and run using 3 different files of different strand lengths. These are :

Ecoli Small : DNA length – 320 126
Ecoli: DNA length - 4,639,221
Ecoli + Ecoli small: DNA length - 4,959,381
Double Ecoli: DNA length - 9,278,443
Tripple Ecoli: DNA length - 13,917,663
Ecoli Quad+Small: DNA length - 18,877,044

For each of the files, DNABenchMark was run 3 times, the running times were then averaged to get a more accurate value.

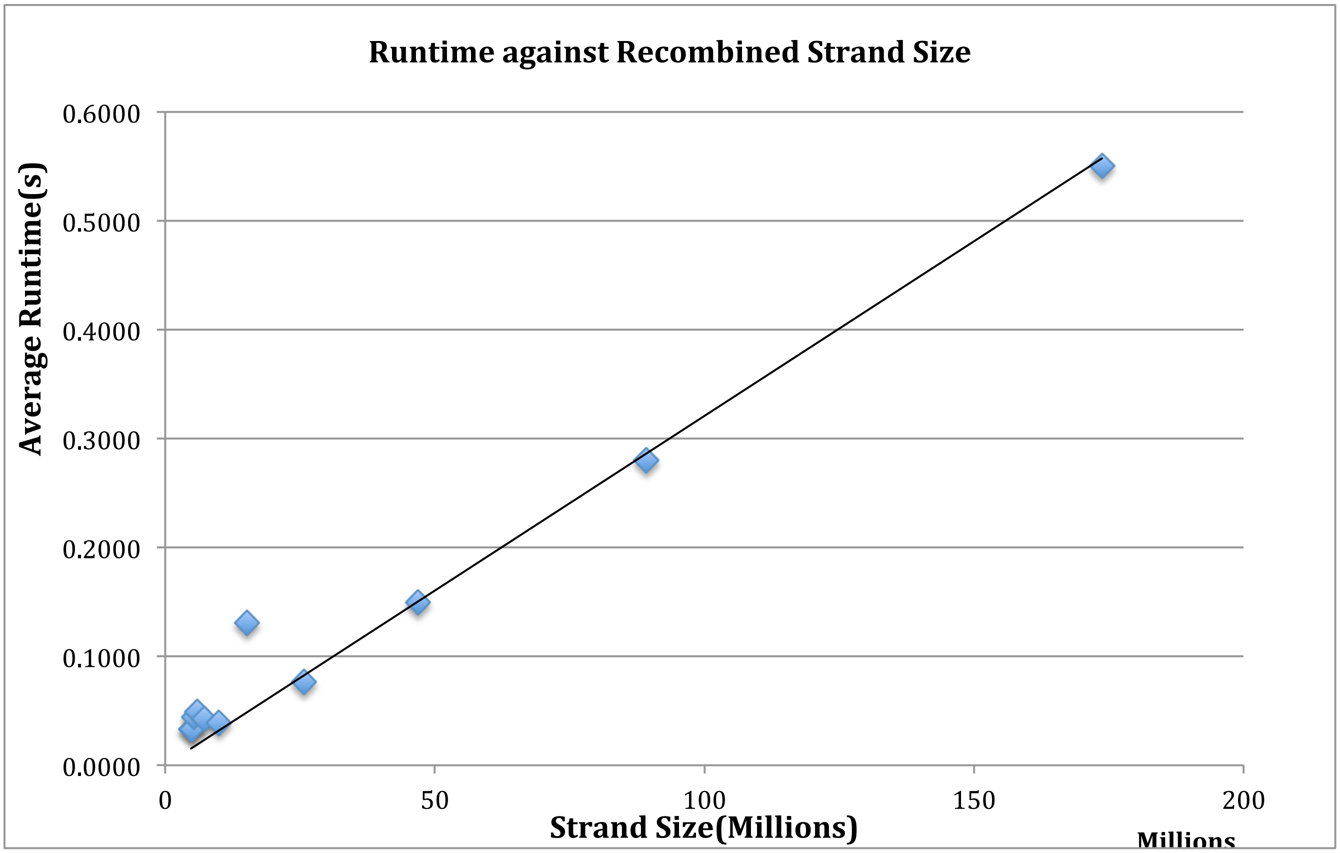
Ecoli Small : DNA length – 320 126

Recombined Strand Size		Runtime 1(s)	Runtime 2(s)	Runtime 3(s)	Average Runtime(s)
331,410		0.003	0.002	0.002	0.0023
342,930		0.003	0.002	0.002	0.0023
365,970		0.003	0.002	0.003	0.0027
412,050		0.009	0.01	0.009	0.0093
504,210		0.003	0.003	0.003	0.0030
688,530		0.003	0.003	0.003	0.0030
1,057,170		0.003	0.003	0.003	0.0030
1,794,450		0.009	0.01	0.008	0.0090
3,269,010		0.012	0.013	0.013	0.0127
6,218,130		0.018	0.018	0.02	0.0187
12,116,370		0.03	0.033	0.032	0.0317
23,912,850		0.067	0.069	0.068	0.0680
47,505,810		0.159	0.162	0.163	0.1613
94,691,730		0.292	0.301	0.295	0.2960
189,063,570		0.636	0.605	0.595	0.6120



Ecoli: DNA length - 4,639,221

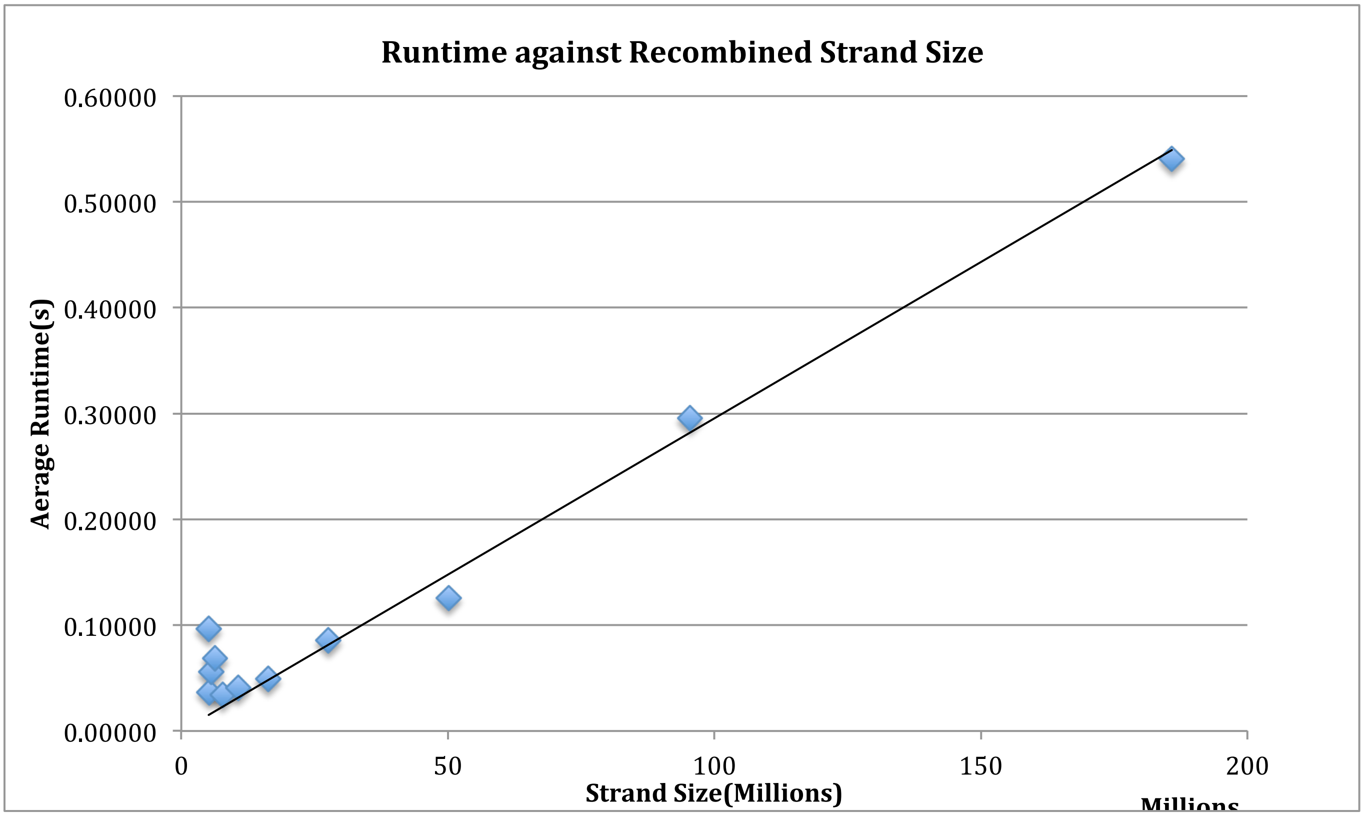
Recombined Strand Size		Runtime 1(s)	Runtime 2(s)	Runtime 3(s)	Average Runtime(s)
4,800,471		0.033	0.035	0.032	0.0333
4,965,591		0.034	0.033	0.033	0.0333
5,295,831		0.046	0.044	0.043	0.0443
5,956,311		0.05	0.047	0.049	0.0487
7,277,271		0.043	0.042	0.042	0.0423
9,919,191		0.04	0.039	0.037	0.0387
15,203,031		0.168	0.109	0.115	0.1307
25,770,711		0.089	0.07	0.07	0.0763
46,906,071		0.15	0.149	0.15	0.1497
89,176,791		0.278	0.288	0.275	0.2803
173,718,231		0.531	0.437	0.683	0.5503



Ecoli + Ecoli small: DNA length - 4,959,381

Recombined Strand Size		Runtime 1(s)	Runtime 2(s)	Runtime 3(s)	AvgRuntime(s)
5,131,881		0.1	0.101	0.088	0.09633
5,308,521		0.036	0.035	0.039	0.03667
5,661,801		0.055	0.059	0.054	0.05600
6,368,361		0.071	0.067	0.067	0.06833
7,781,481		0.034	0.035	0.034	0.03433
10,607,721		0.041	0.041	0.04	0.04067
16,260,201		0.051	0.053	0.043	0.04900
27,565,161		0.0811	0.091	0.0843	0.08547

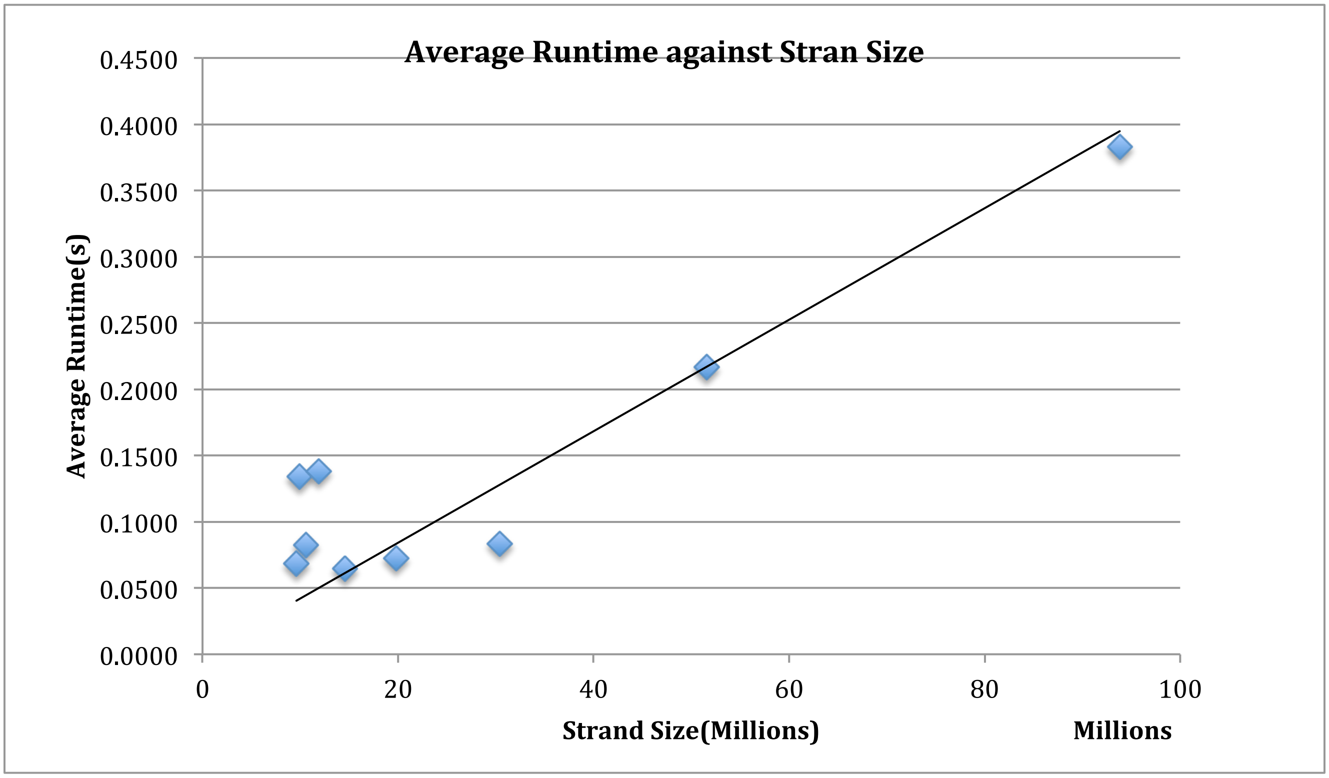
50,175,081		0.138	0.131	0.1066	0.12520
95,394,921		0.307	0.286	0.293	0.29533
185,834,601		0.61	0.582	0.430	0.54067



Double Ecoli: DNA length - 9,278,443

Recombined Strand Size		Runtime 1(s)	Runtime 2(s)	Runtime 3(s)	AvgRuntime(s)
9,600,943		0.065	0.069	0.071	0.0683
9,931,183		0.178	0.122	0.102	0.1340
10,591,663		0.089	0.079	0.079	0.0823
11,912,623		0.152	0.13	0.132	0.1380
14,554,543		0.064	0.066	0.064	0.0647
19,838,383		0.072	0.073	0.0723	0.0724

30,406,063		0.085	0.083	0.083	0.0837
51,541,423		0.233	0.207	0.211	0.2170
93,812,143		0.38	0.392	0.377	0.3830



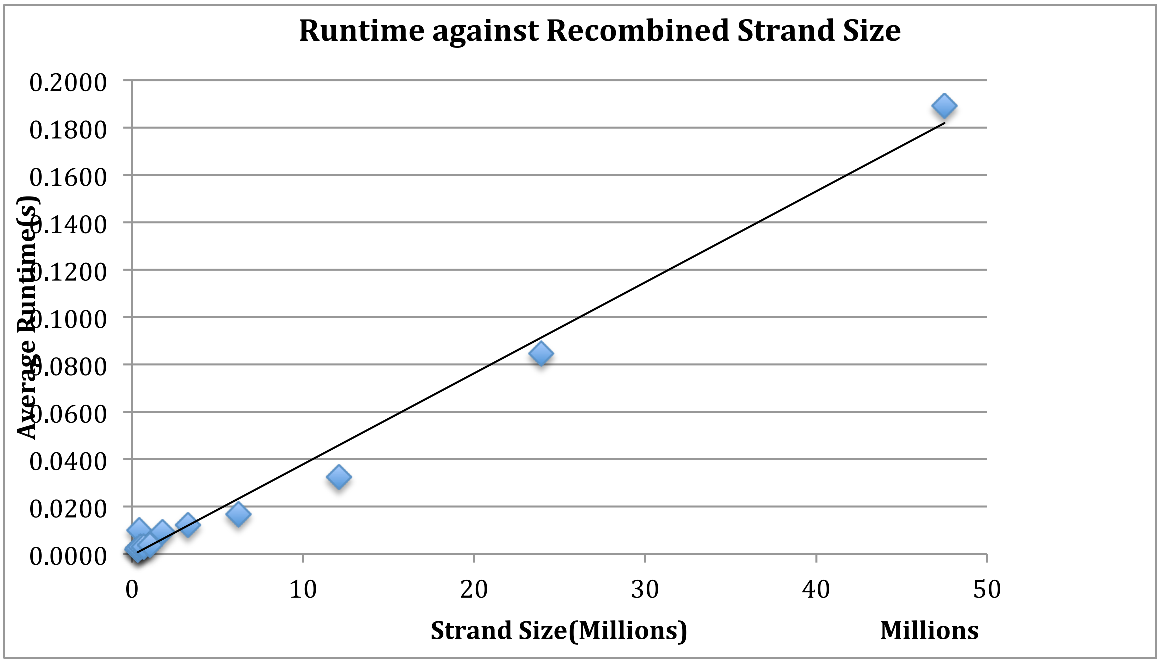
For all 3 files used for benchmarking, a graph of running time against combinant strand return size shows a linear relationship. This supports my Big(O) hypothesis - O(N) time, which is linear time.

2)

Ecoli Small : DNA length – 320 126
Heap Size -Xmx512M

Table Of Results

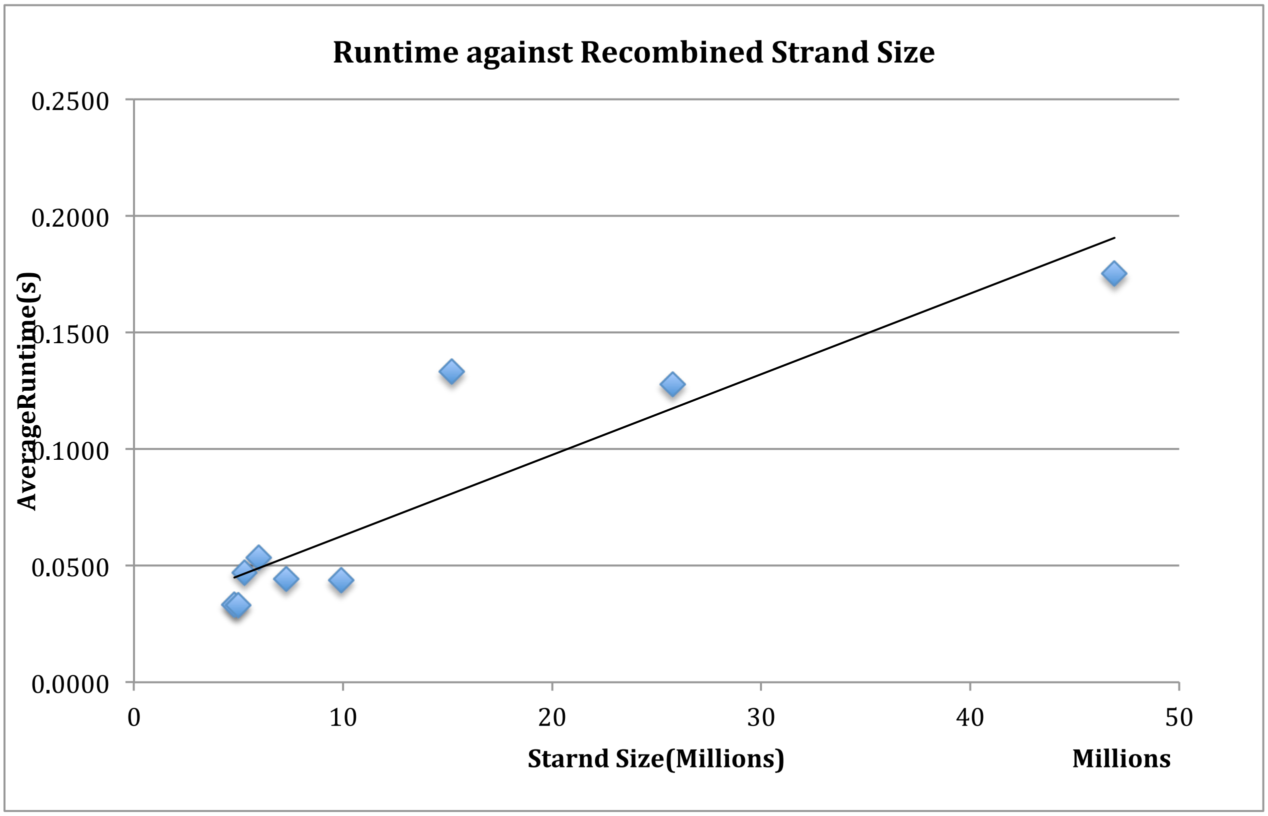
Recombined Strand Size		Runtime 1(s)	Runtime 2(s)	Runtime 3(s)	Average Runtime(s)
331,410		0.002	0.002	0.002	0.0020
342,930		0.002	0.002	0.002	0.0020
365,970		0.003	0.003	0.002	0.0027
412,050		0.01	0.011	0.009	0.0100
504,210		0.003	0.003	0.003	0.0030
688,530		0.003	0.003	0.003	0.0030
1,057,170		0.004	0.004	0.004	0.0040
1,794,450		0.01	0.009	0.009	0.0093
3,269,010		0.012	0.013	0.012	0.0123
6,218,130		0.017	0.017	0.017	0.0170
12,116,370		0.033	0.032	0.033	0.0327
23,912,850		0.097	0.079	0.078	0.0847
47,505,810		0.197	0.189	0.182	0.1893



For Ecoli small, the largest splicee that worked without exhausting memory for a heap-size of Xmx512M is of size **1,048,576**.

Ecoli: DNA length - 4,639,221
Heap Size -Xmx512M

Recombined Strand Size		Runtime 1(s)	Runtime 2(s)	Runtime 3(s)	AvgRuntime(s)
4,800,471		0.034	0.033	0.033	0.0333
4,965,591		0.033	0.032	0.034	0.0330
5,295,831		0.049	0.044	0.048	0.0470
5,956,311		0.055	0.058	0.047	0.0533
7,277,271		0.051	0.038	0.044	0.0443
9,919,191		0.047	0.045	0.039	0.0437
15,203,031		0.12	0.17	0.11	0.1333
25,770,711		0.139	0.126	0.118	0.1277
46,906,071		0.178	0.178	0.17	0.1753



For Ecoli, the largest splicee that worked without exhausting memory for a heap-size of Xmx512M is of size **65,536**

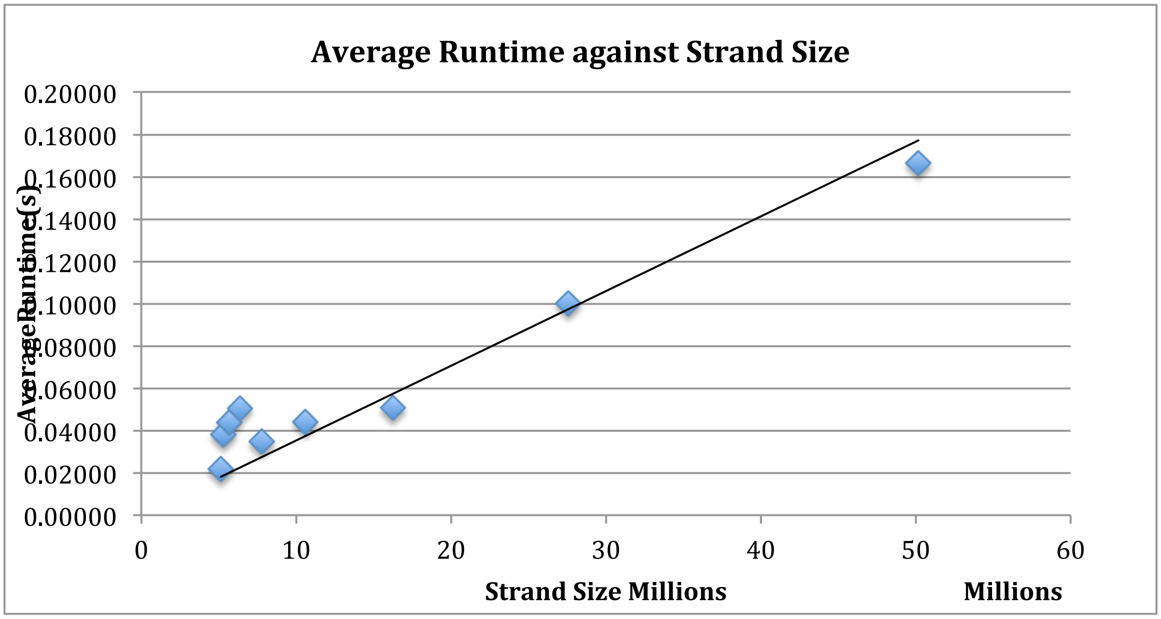
.

Ecoli + Ecoli small: DNA length - 4,959,381

Heap Size -Xmx512M

Recombined Strand Size		Runtime 1(s)	Runtime 2(s)	Runtime 3(s)	Average Runtime(s)
5,131,881		0.032	0.021	0.012	0.02167
5,308,521		0.042	0.035	0.037	0.03800
5,661,801		0.052	0.0478	0.0321	0.04397

6,368,361		0.0521	0.0556	0.0438	0.05050
7,781,481		0.035	0.034	0.035	0.03467
10,607,721		0.047	0.042	0.043	0.04400
16,260,201		0.05	0.05	0.052	0.05067
27,565,161		0.106	0.098	0.097	0.10033
50,175,081		0.16	0.171	0.169	0.16667

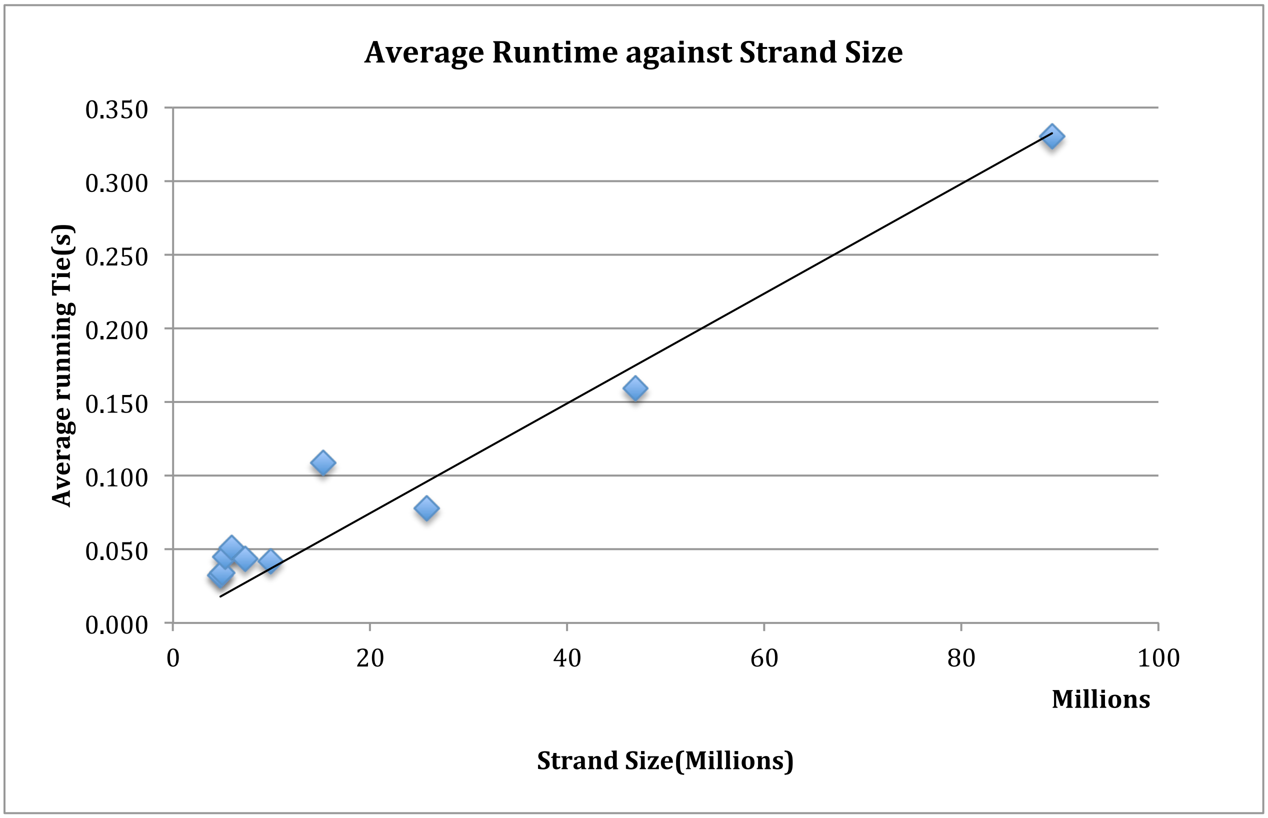


For Ecoli + EcoliSmall, the largest splicee that worked without exhausting memory for a heap-size of Xmx512M is of size **65,536.**

.

Ecoli: DNA length - 4,639,221
Heap Size - -Xmx1024M

Recombined Strand Size		Runtime 1(s)	Runtime 2(s)	Runtime 3(s)	Average Runtime(s)
4,800,471		0.032	0.033	0.032	0.032
4,965,591		0.033	0.036	0.033	0.034
5,295,831		0.044	0.043	0.048	0.045
5,956,311		0.052	0.054	0.047	0.051
7,277,271		0.048	0.041	0.042	0.044
9,919,191		0.045	0.04	0.041	0.042
15,203,031		0.113	0.101	0.112	0.109
25,770,711		0.084	0.074	0.076	0.078
46,906,071		0.174	0.148	0.155	0.159
89,176,791		0.423	0.293	0.275	0.330

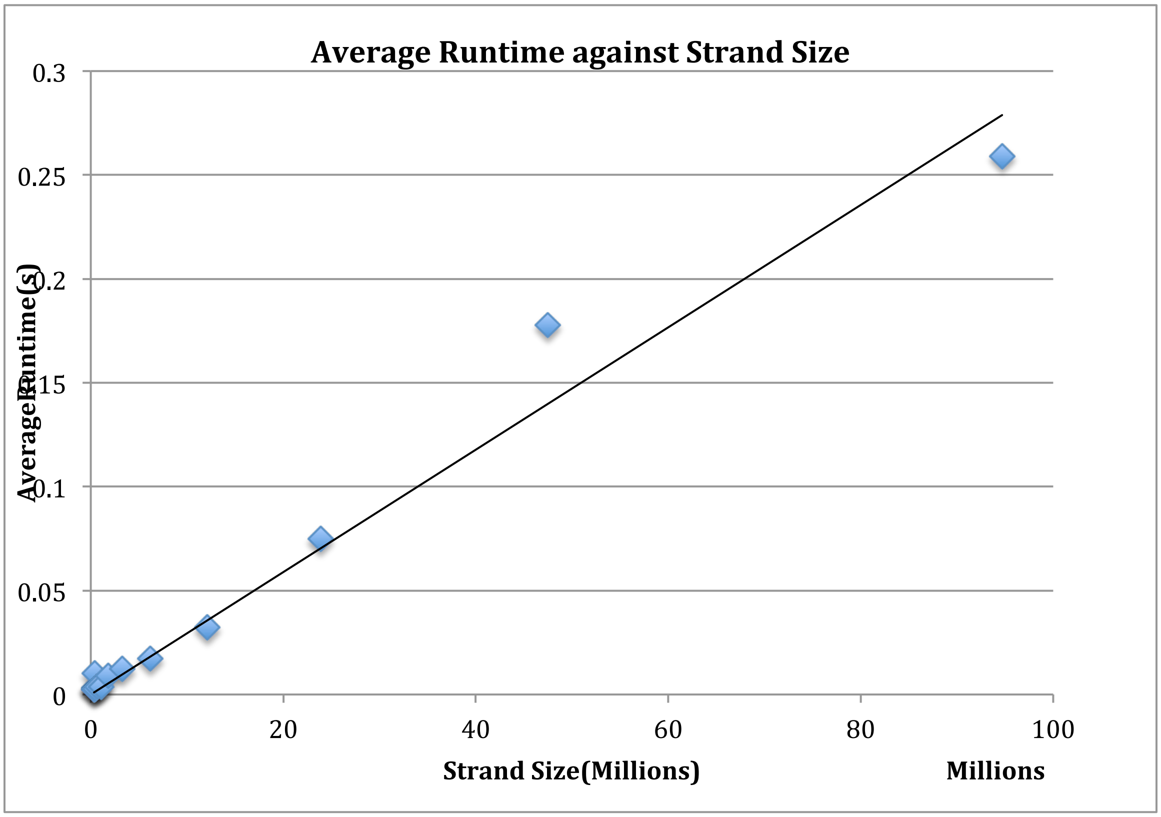


For Ecoli, the largest splicee that worked without exhausting memory for a heap-size of Xmx1024M is of size **131,072.**

Ecoli Small : DNA length – 320 126

Heap Size –Xmx1024M

Recombined Strand Size		Runtime 1(s)	Runtime 2(s)	Runtime 3(s)	AvgRuntime(s)
331,410		0.004	0.003	0.002	0.0030
342,930		0.003	0.002	0.002	0.0023
365,970		0.002	0.003	0.002	0.0023
412,050		0.011	0.01	0.01	0.0103
504,210		0.002	0.003	0.003	0.0027
688,530		0.003	0.004	0.004	0.0037
1,057,170		0.004	0.003	0.004	0.0037
1,794,450		0.009	0.009	0.009	0.0090
3,269,010		0.012	0.012	0.013	0.0123
6,218,130		0.018	0.017	0.017	0.0173
12,116,370		0.033	0.032	0.032	0.0323
23,912,850		0.077	0.079	0.069	0.0750
47,505,810		0.186	0.186	0.161	0.1777
94,691,730		0.257	0.212	0.308	0.2590



For Ecoli_Small, the largest splicee that worked without exhausting memory for a heap-size of Xmx1024M is of size **2,097,152**

Longest Splicee size for Ecoli and Ecoli small relative to memory

	ecoli	**ecoli_small**
Heap Size -Xmx512M	**65,536**	**1,048,576**
Heap Size -Xmx1024M	**131,072**	**2,097,152**

From the table above, the longest splice that worked before memory was exhausted, doubled in both cases(ecoli and ecoli_small) when heap size memory allocation was doubled from Xmx512M to Xmx1024M.
This shows that in SimpleStrand, the length of the Splicee influences overall performance and efficiency of the program. This supports the hypothesis

that due to the cutAndSplice method which appends splicee, the size of the splicee results in n operations where n is the size of the splicee, hence as the splicee gets bigger, the operation becomes more inefficient and requires more memory.

4)
.LinkStrand Hypothesis

When implementing LinkStrand, the complexity of splicing is independent of the size of the strand being spliced in. In this implementation, the complexity of creating the recombinant strand is dependent on the number of breaks caused by the restriction enzyme in the original DNA strand.

This is because when calling cutAndSplice in LinkStrand, each splice will cost O(1) time due to fact that at each break, appending is done by simply setting the pointer to the splicee, which is simply O(1) and more efficient than the SimpleStrand append method. Therefore since each break costs O(1), the runtime for cutAndSplice in LinkStrand should be O(1*B) = O(B) where B is the number of breaks caused by the restriction enzyme.

It is for this reason that LinkStrand is more efficient than SimpleStrand in terms of both memory, and time. This can be seen in the results below.

The LinkStrand.cutAndSplice code was BenchMarked using the class DNABenchMark, and run using six different files of different strand lengths. These are :

Ecoli Small : DNA length – 320 126
Ecoli: DNA length - 4,639,221
Ecoli + Ecoli small: DNA length - 4,959,381
Double Ecoli: DNA length - 9,278,443
Tripple Ecoli: DNA length - 13,917,663
Ecoli Quad+Small: DNA length - 18,877,044

The results are shown below

```
ecoli_small    1024M

dna length = 320,160
cutting at enzyme gaattc
-----

Class          splicee           recomb       time
-----
LinkStrand:    256           331,410          0.004        # append calls =90
LinkStrand:    512           342,930          0.003        # append calls =90
LinkStrand:    1,024         365,970          0.002        # append calls =90
LinkStrand:    2,048         412,050          0.003        # append calls =90
LinkStrand:    4,096         504,210          0.002        # append calls =90
LinkStrand:    8,192         688,530          0.003        # append calls =90
LinkStrand:    16,384        1,057,170        0.002        # append calls =90
LinkStrand:    32,768        1,794,450        0.003        # append calls =90
LinkStrand:    65,536        3,269,010        0.002        # append calls =90
LinkStrand:    131,072       6,218,130        0.002        # append calls =90
LinkStrand:    262,144       12,116,370       0.002        # append calls =90
LinkStrand:    524,288       23,912,850       0.001        # append calls =90
LinkStrand:    1,048,576     47,505,810       0.002        # append calls =90
LinkStrand:    2,097,152     94,691,730       0.002        # append calls =90
LinkStrand:    4,194,304     189,063,570      0.002        # append calls =90
LinkStrand:    8,388,608     377,807,250      0.002        # append calls =90
LinkStrand:    16,777,216    755,294,610      0.002        # append calls =90
LinkStrand:    33,554,432    1,510,269,330    0.002        # append calls =90
LinkStrand:    67,108,864    3,020,218,770    0.001        # append calls =90




Ecoli



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

```
-----

Class          splicee            recomb         time
-----
                                                            # append calls =
LinkStrand:    256             4,800,471         0.025      1290
                                                            # append calls =
LinkStrand:    512             4,965,591         0.026      1290
                                                            # append calls =
LinkStrand:    1,024           5,295,831         0.026      1290
                                                            # append calls =
LinkStrand:    2,048           5,956,311         0.025      1290
                                                            # append calls =
LinkStrand:    4,096           7,277,271         0.028      1290
                                                            # append calls =
LinkStrand:    8,192           9,919,191         0.032      1290
                                                            # append calls =
LinkStrand:    16,384          15,203,031        0.036      1290
                                                            # append calls =
LinkStrand:    32,768          25,770,711        0.029      1290
                                                            # append calls =
LinkStrand:    65,536          46,906,071        0.025      1290
                                                            # append calls =
LinkStrand:    131,072         89,176,791        0.025      1290
                                                            # append calls =
LinkStrand:    262,144         173,718,231       0.026      1290
                                                            # append calls =
LinkStrand:    524,288         342,801,111       0.028      1290
                                                            # append calls =
LinkStrand:    1,048,576       680,966,871       0.026      1290
                                                            # append calls =
LinkStrand:    2,097,152       1,357,298,391     0.028      1290
                                                            # append calls =
LinkStrand:    4,194,304       2,709,961,431     0.03       1290
                                                            # append calls =
LinkStrand:    8,388,608       5,415,287,511     0.025      1290
```

```
                                                      # append calls =
LinkStrand:  16,777,216    10,825,939,671   0.032     1290
                                                      # append calls =
LinkStrand:  33,554,432    21,647,243,991   0.028     1290
                                                      # append calls =
LinkStrand:  67,108,864    43,289,852,631   0.153     1290
                                                      # append calls =
LinkStrand:  134,217,728   86,575,069,911   0.025     1290



Ecoli+small

dna length = 4,959,381
cutting at enzyme gaattc
-----

Class        splicee           recomb       time
-----
                                                      # append calls =
LinkStrand:  256           5,131,881        0.028     1380
                                                      # append calls =
LinkStrand:  512           5,308,521        0.028     1380
                                                      # append calls =
LinkStrand:  1,024         5,661,801        0.097     1380
                                                      # append calls =
LinkStrand:  2,048         6,368,361        0.027     1380
                                                      # append calls =
LinkStrand:  4,096         7,781,481        0.03      1380
                                                      # append calls =
LinkStrand:  8,192         10,607,721       0.029     1380
                                                      # append calls =
LinkStrand:  16,384        16,260,201       0.027     1380
                                                      # append calls =
LinkStrand:  32,768        27,565,161       0.029     1380
LinkStrand:  65,536        50,175,081       0.027     # append calls =
                                                         1380
                                                         # append calls =
 LinkStrand:  131,072      95,394,921      0.028         1380
                                                         # append calls =
 LinkStrand:  262,144      185,834,601     0.029         1380
                                                         # append calls =
 LinkStrand:  524,288      366,713,961     0.031         1380
                                                         # append calls =
 LinkStrand:  1,048,576    728,472,681     0.031         1380
                                                         # append calls =
 LinkStrand:  2,097,152    1,451,990,121   0.035         1380
                                                         # append calls =
 LinkStrand:  4,194,304    2,899,025,001   0.027         1380
                                                         # append calls =
 LinkStrand:  8,388,608    5,793,094,761   0.027         1380
                                                         # append calls =
 LinkStrand:  16,777,216   11,581,234,281  0.029         1380
                                                         # append calls =
 LinkStrand:  33,554,432   23,157,513,321  0.027         1380
                                                         # append calls =
 LinkStrand:  67,108,864   46,310,071,401  0.031         1380
                                                         # append calls =
 LinkStrand:  134,217,728  92,615,187,561  0.434         1380
```

double ecoli

```
dna length = 9,278,443
cutting at enzyme gaattc
-----
Class              splicee            recomb time
-----
LinkStrand:            256            9,600,943     0.052 # append calls = 2580
LinkStrand:            512            9,931,183     0.056 # append calls = 2580
LinkStrand:          1,024           10,591,663     0.055 # append calls = 2580
LinkStrand:          2,048           11,912,623     0.052 # append calls = 2580
```

```
LinkStrand:        4,096           14,554,543    0.052 # append calls = 2580
LinkStrand:        8,192           19,838,383    0.053 # append calls = 2580
LinkStrand:       16,384           30,406,063    0.122 # append calls = 2580
LinkStrand:       32,768           51,541,423    0.056 # append calls = 2580
LinkStrand:       65,536           93,812,143    0.053 # append calls = 2580
LinkStrand:      131,072          178,353,583    0.051 # append calls = 2580
LinkStrand:      262,144          347,436,463    0.053 # append calls = 2580
LinkStrand:      524,288          685,602,223    0.051 # append calls = 2580
LinkStrand:    1,048,576        1,361,933,743    0.051 # append calls = 2580
LinkStrand:    2,097,152        2,714,596,783    0.054 # append calls = 2580
LinkStrand:    4,194,304        5,419,922,863    0.062 # append calls = 2580
LinkStrand:    8,388,608       10,830,575,023    0.051 # append calls = 2580
LinkStrand:   16,777,216       21,651,879,343    0.050 # append calls = 2580
LinkStrand:   33,554,432       43,294,487,983    0.051 # append calls = 2580
LinkStrand:   67,108,864       86,579,705,263    0.051 # append calls = 2580
LinkStrand:  134,217,728      173,150,139,823    0.051 # append calls = 2580
```

Tripple Ecoli

```
ecoli_tripple

dna length = 13,917,663
cutting at enzyme gaattc
-----

Class         splicee          recomb        time
-----
LinkStrand                                              # append calls =
:             256         14,401,413         0.086      3870
LinkStrand                                              # append calls =
:             512         14,896,773         0.092      3870
LinkStrand                                              # append calls =
:             1,024       15,887,493         0.096      3870
LinkStrand    2,048       17,868,933         0.137      # append calls =
```

:				3870
LinkStrand :	4,096	21,831,813	0.077	# append calls = 3870
LinkStrand :	8,192	29,757,573	0.08	# append calls = 3870
LinkStrand :	16,384	45,609,093	0.078	# append calls = 3870
LinkStrand :	32,768	77,312,133	0.079	# append calls = 3870
LinkStrand :	65,536	140,718,213	0.079	# append calls = 3870
LinkStrand :	131,072	267,530,373	0.077	# append calls = 3870
LinkStrand :	262,144	521,154,693	0.077	# append calls = 3870
LinkStrand :	524,288	1,028,403,333	0.082	# append calls = 3870
LinkStrand :	1,048,576	2,042,900,613	0.089	# append calls = 3870
LinkStrand :	2,097,152	4,071,895,173	0.076	# append calls = 3870
LinkStrand :	4,194,304	8,129,884,293	0.079	# append calls = 3870
LinkStrand :	8,388,608	16,245,862,533	0.076	# append calls = 3870
LinkStrand :	16,777,216	32,477,819,013	0.091	# append calls = 3870
LinkStrand :	33,554,432	64,941,731,973	0.147	# append calls = 3870
LinkStrand :	67,108,864	129,869,557,893	0.076	# append calls = 3870
LinkStrand :	134,217,728	259,725,209,733	4.916	# append calls = 3870

Quad Ecoli

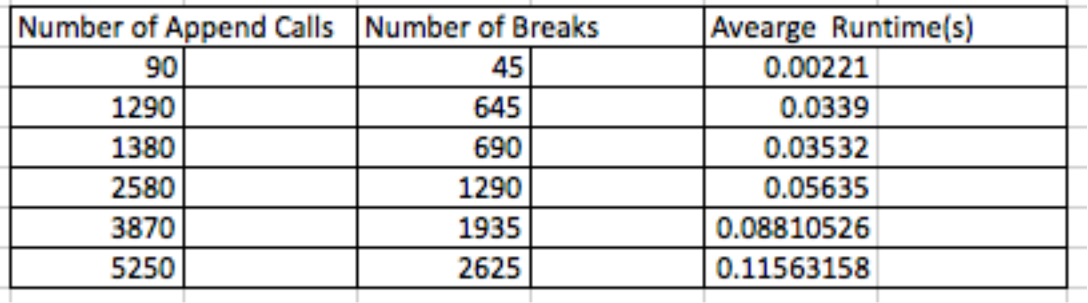
```
ecoli_quad+small

dna length =
18,877,044
cutting at enzyme gaattc
-----

Class         splicee          recomb          time
-----
                                                        # append calls =
LinkStrand:   256              19,533,294      0.213    5250
                                                        # append calls =
LinkStrand:   512              20,205,294      0.106    5250
                                                        # append calls =
LinkStrand:   1,024            21,549,294      0.111    5250
                                                        # append calls =
LinkStrand:   2,048            24,237,294      0.11     5250
                                                        # append calls =
LinkStrand:   4,096            29,613,294      0.105    5250
                                                        # append calls =
LinkStrand:   8,192            40,365,294      0.107    5250
                                                        # append calls =
LinkStrand:   16,384           61,869,294      0.11     5250
                                                        # append calls =
LinkStrand:   32,768           104,877,294     0.105    5250
                                                        # append calls =
LinkStrand:   65,536           190,893,294     0.103    5250
                                                        # append calls =
LinkStrand:   131,072          362,925,294     0.104    5250
                                                        # append calls =
LinkStrand:   262,144          706,989,294     0.105    5250
                                                        # append calls =
LinkStrand:   524,288          1,395,117,294   0.122    5250
```

LinkStrand:	1,048,576	2,771,373,294	0.106	# append calls = 5250
LinkStrand:	2,097,152	5,523,885,294	0.103	# append calls = 5250
LinkStrand:	4,194,304	11,028,909,294	0.107	# append calls = 5250
LinkStrand:	8,388,608	22,038,957,294	0.103	# append calls = 5250
LinkStrand:	16,777,216	44,059,053,294	0.103	# append calls = 5250
LinkStrand:	33,554,432	88,099,245,294	0.167	# append calls = 5250
LinkStrand:	67,108,864	176,179,629,294	0.107	# append calls = 5250

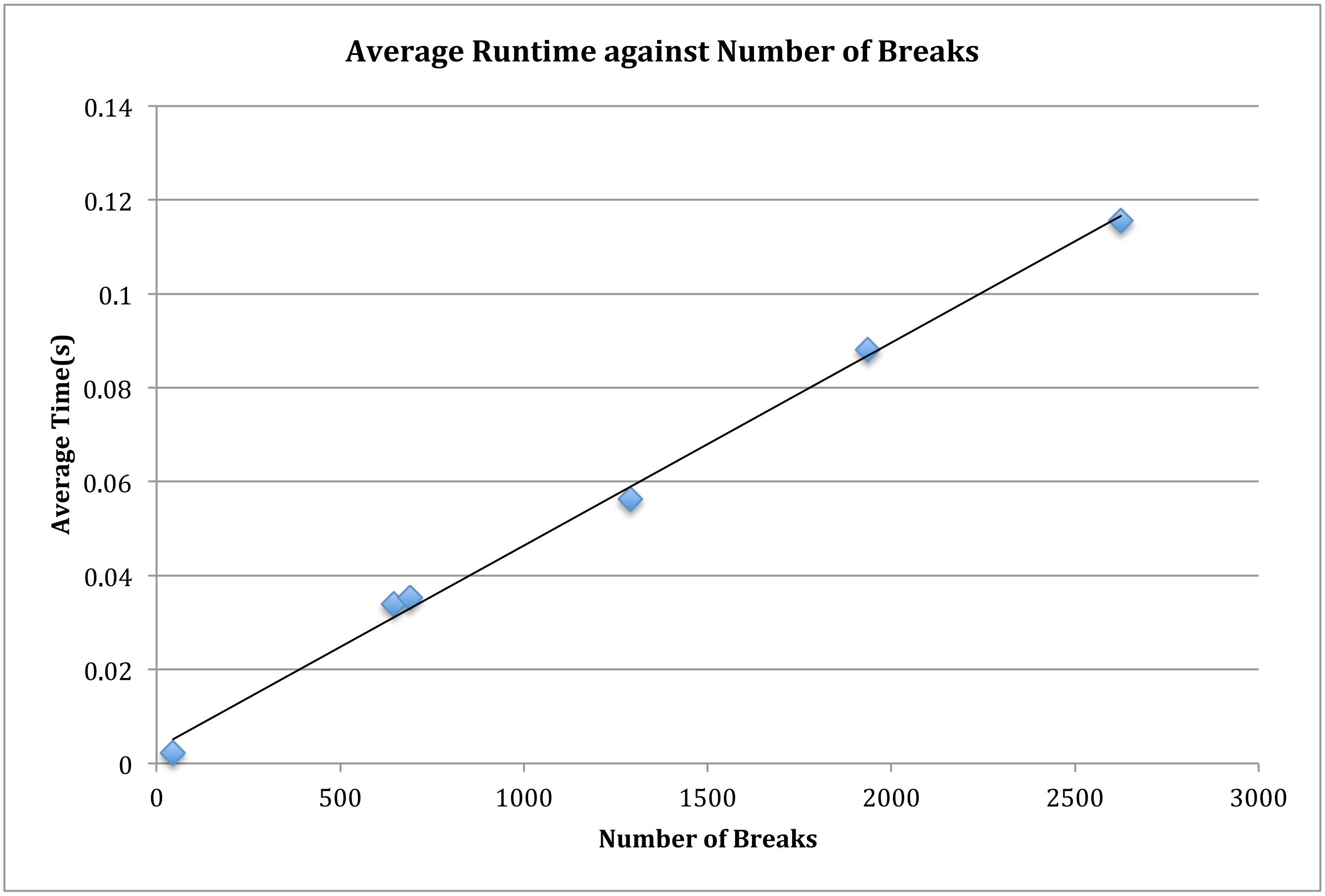
The running time for each text file was averaged for the different splice sizes. As can be seen from the tables above, the runtime is independent of the splice size.

The number of breaks in the DNA strand were found by dividing the number of append calls by 2, since at each break, the append method was called twice. (For joining the front and back of the splicee node to DNA strand).

The results are in the table below:



Number of Append Calls	Number of Breaks	Avearge Runtime(s)
90	45	0.00221
1290	645	0.0339
1380	690	0.03532
2580	1290	0.05635
3870	1935	0.08810526
5250	2625	0.11563158



The graph of Average runtime against the number of breaks in a DNA strand shows a linear time relationship. This supports the hypothesis from Big(O) that the runtime is O(B) where B is the number of breaks in the DNA strand.

The runtime for LinkStrand was more efficient than that of SimpleStrand, and it was independent of the length of the splicee itself, but only depended on the number of breaks. Which is shown in the reults above.

LinkStrand also performed better than SimleStrand in terms of the lengths of the reconmbinant strand it could return before running out of memory. This was due to the efficiency of the cutAndSplice method in LinkStrand which was discussed earlier.