Liane Yanglian,
March 18, 2016

DNA Analysis Questions

1.

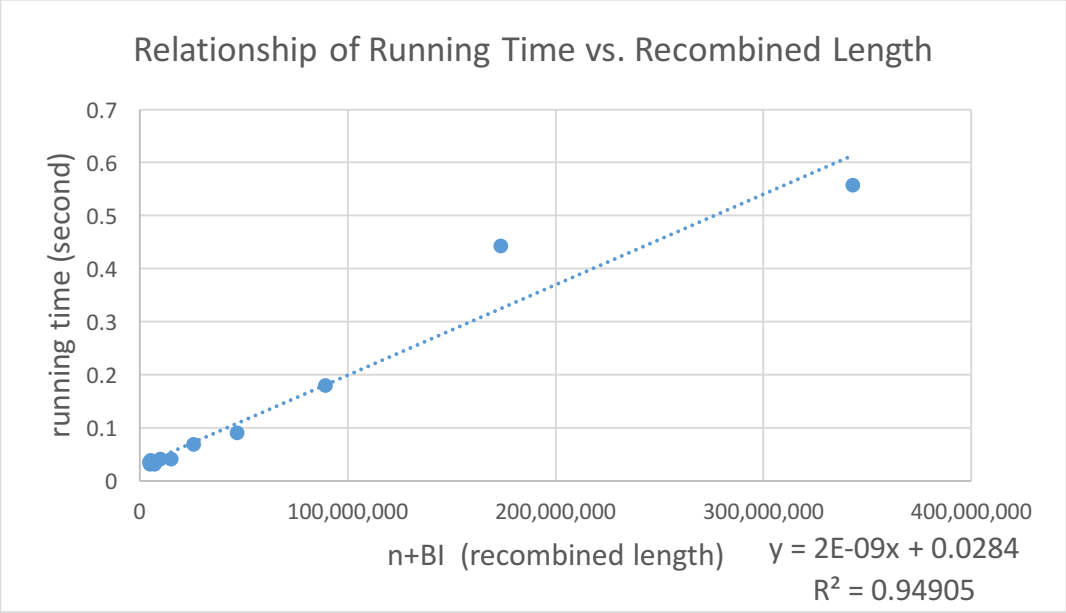
When l changes

Data:

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

Class	splicee	recomb	time	
SimpleStrand:	256	4,800,471	0.035	# append calls = 1290
SimpleStrand:	512	4,965,591	0.031	# append calls = 1290
SimpleStrand:	1,024	5,295,831	0.038	# append calls = 1290
SimpleStrand:	2,048	5,956,311	0.033	# append calls = 1290
SimpleStrand:	4,096	7,277,271	0.031	# append calls = 1290
SimpleStrand:	8,192	9,919,191	0.041	# append calls = 1290
SimpleStrand:	16,384	15,203,031	0.041	# append calls = 1290
SimpleStrand:	32,768	25,770,711	0.069	# append calls = 1290
SimpleStrand:	65,536	46,906,071	0.090	# append calls = 1290
SimpleStrand:	131,072	89,176,791	0.180	# append calls = 1290
SimpleStrand:	262,144	173,718,231	0.443	# append calls = 1290
SimpleStrand:	524,288	342,801,111	0.557	# append calls = 1290

Graph:



Analysis:

I ran benchmark on ecoli.txt. The graph of running time vs. recombined length, as shown above, shows the linear relationship between the two variables is linear. This is supported by the fact that R^2 is 0.94905 for the graph, which is close to 1. The closer R^2 is to 1, the more linear the graph is. This is when l (number of splicee) changes.

When only n changes

Data:

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

Class	splicee	recomb	time	
SimpleStrand:	256	4,800,471	0.032	# append calls = 1290
SimpleStrand:	512	4,965,591	0.032	# append calls = 1290
SimpleStrand:	1,024	5,295,831	0.029	# append calls = 1290
SimpleStrand:	2,048	5,956,311	0.031	# append calls = 1290
SimpleStrand:	4,096	7,277,271	0.032	# append calls = 1290
SimpleStrand:	8,192	9,919,191	0.035	# append calls = 1290
SimpleStrand:	16,384	15,203,031	0.042	# append calls = 1290
SimpleStrand:	32,768	25,770,711	0.061	# append calls = 1290

dna length = 9,278,442
cutting at enzyme gaattc

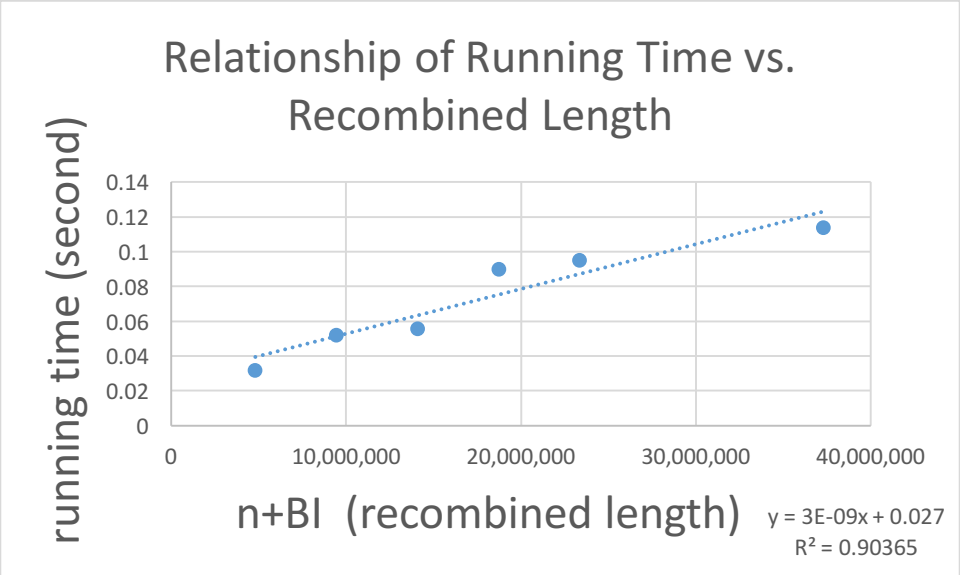
Class	splicee	recomb	time	
SimpleStrand:	256	9,439,692	0.052	# append calls = 1290
SimpleStrand:	512	9,604,812	0.039	# append calls = 1290
SimpleStrand:	1,024	9,935,052	0.041	# append calls = 1290
SimpleStrand:	2,048	10,595,532	0.039	# append calls = 1290
SimpleStrand:	4,096	11,916,492	0.037	# append calls = 1290
SimpleStrand:	8,192	14,558,412	0.040	# append calls = 1290
SimpleStrand:	16,384	19,842,252	0.053	# append calls = 1290
SimpleStrand:	32,768	30,409,932	0.061	# append calls = 1290

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

Class	splicee	recomb	time	
SimpleStrand:	256	14,078,913	0.056	# append calls = 1290

```
SimpleStrand:          512        14,244,033 0.048   # append calls = 1290
SimpleStrand:        1,024        14,574,273 0.046   # append calls = 1290
SimpleStrand:        2,048        15,234,753 0.043   # append calls = 1290
SimpleStrand:        4,096        16,555,713 0.043   # append calls = 1290
SimpleStrand:        8,192        19,197,633 0.054   # append calls = 1290
SimpleStrand:       16,384        24,481,473 0.054   # append calls = 1290
SimpleStrand:       32,768        35,049,153 0.061   # append calls = 1290
dna length = 18,556,884
cutting at enzyme gaattc
-----
Class               splicee          recomb     time
-----
SimpleStrand:          256        18,718,134 0.090   # append calls = 1290
SimpleStrand:          512        18,883,254 0.077   # append calls = 1290
SimpleStrand:        1,024        19,213,494 0.073   # append calls = 1290
SimpleStrand:        2,048        19,873,974 0.071   # append calls = 1290
SimpleStrand:        4,096        21,194,934 0.074   # append calls = 1290
SimpleStrand:        8,192        23,836,854 0.078   # append calls = 1290
SimpleStrand:       16,384        29,120,694 0.078   # append calls = 1290

dna length = 23,196,105
cutting at enzyme gaattc
-----
Class               splicee          recomb     time
-----
SimpleStrand:          256        23,357,355 0.095   # append calls = 1290
SimpleStrand:          512        23,522,475 0.107   # append calls = 1290
SimpleStrand:        1,024        23,852,715 0.097   # append calls = 1290
SimpleStrand:        2,048        24,513,195 0.099   # append calls = 1290
SimpleStrand:        4,096        25,834,155 0.102   # append calls = 1290
SimpleStrand:        8,192        28,476,075 0.104   # append calls = 1290
SimpleStrand:       16,384        33,759,915 0.095   # append calls = 1290

dna length = 37,113,768
cutting at enzyme gaattc
-----
Class               splicee          recomb     time
-----
SimpleStrand:          256        37,275,018 0.114   # append calls = 1290
SimpleStrand:          512        37,440,138 0.113   # append calls = 1290
SimpleStrand:        1,024        37,770,378 0.083   # append calls = 1290
SimpleStrand:        2,048        38,430,858 0.082   # append calls = 1290
SimpleStrand:        4,096        39,751,818 0.084   # append calls = 1290
SimpleStrand:        8,192        42,393,738 0.086   # append calls = 1290
```

Graph:



Analysis:

The graph of running time vs. recombined length, as shown above, shows the linear relationship between the two variables is linear. This is supported by the fact that R^2 is 0.90365 for the graph, which is close to 1. This is when n (the length of the original strand) increases while B and l stays constant.

To solve the n/B relationship, I first copy pasted the ecoli.txt file onto a new text file, ecoli3, after which I replaced all "g"s with "a"s so there will be no enzymes matching "gaattc" in this new file of the same length as ecoli.txt. Afterwards, I created ecoli2, and I copy pasted the content of ecoli.txt on to ecoli2 as well as the content of ecoli3 on to ecoli2. Now ecoli2 is twice as long as ecoli.txt but contains the same number of "gaattc"s as ecoli.txt. Similarly, when I want to increase the length of the ecoli2 without changing the number of "gaattc"s, I just copy paste the content of ecoli3 onto ecoli2 more times. To keep the number of splicee constant, I chose the recombined length and running time that correspond to 256 splicee in each of the data tables to form the graph above.

When only B changes

Data:

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

Class	splicee	recomb	time	
SimpleStrand:	256	4,874,721	0.037	# append calls = 1884
SimpleStrand:	512	5,115,873	0.031	# append calls = 1884
SimpleStrand:	1,024	5,598,177	0.032	# append calls = 1884

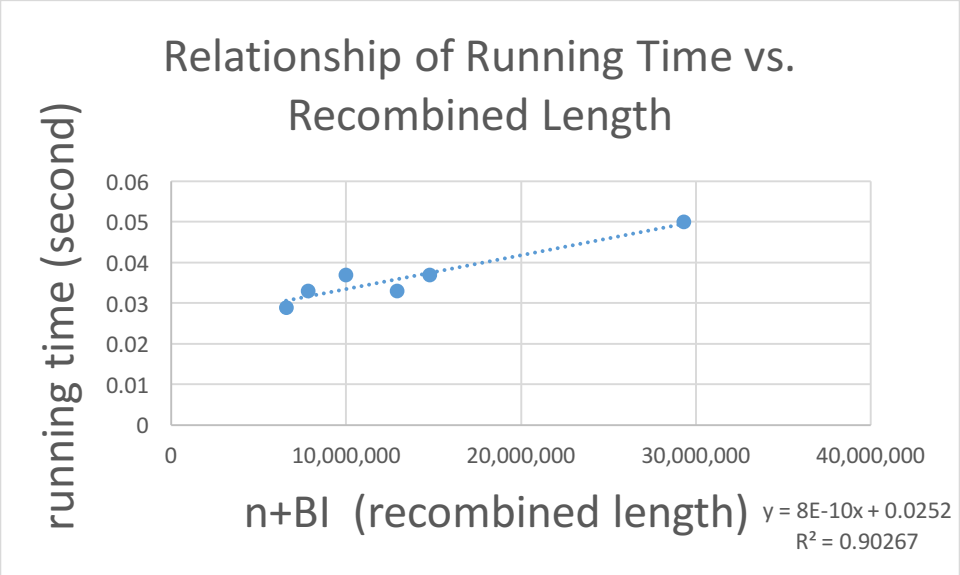
```
SimpleStrand:         2,048        6,562,785  0.029   # append calls = 1884
SimpleStrand:         4,096        8,492,001  0.038   # append calls = 1884
SimpleStrand:         8,192       12,350,433 0.060   # append calls = 1884
SimpleStrand:        16,384       20,067,297 0.048   # append calls = 1884
SimpleStrand:        32,768       35,501,025 0.058   # append calls = 1884

dna length = 4,639,221
cutting at enzyme gaattc
-----
Class               splicee          recomb     time
-----
SimpleStrand:           256        5,029,721  0.029   # append calls = 3124
SimpleStrand:           512        5,429,593  0.026   # append calls = 3124
SimpleStrand:         1,024        6,229,337  0.026   # append calls = 3124
SimpleStrand:         2,048        7,828,825  0.033   # append calls = 3124
SimpleStrand:         4,096       11,027,801 0.036   # append calls = 3124
SimpleStrand:         8,192       17,425,753 0.039   # append calls = 3124
SimpleStrand:        16,384       30,221,657 0.060   # append calls = 3124
dna length = 4,639,221
cutting at enzyme gaattc
-----
Class               splicee          recomb     time
-----
SimpleStrand:           256        5,292,971  0.030   # append calls = 5230
SimpleStrand:           512        5,962,411  0.031   # append calls = 5230
SimpleStrand:         1,024        7,301,291  0.037   # append calls = 5230
SimpleStrand:         2,048        9,979,051  0.037   # append calls = 5230
SimpleStrand:         4,096       15,334,571 0.040   # append calls = 5230
SimpleStrand:         8,192       26,045,611 0.059   # append calls = 5230
SimpleStrand:        16,384       47,467,691 0.092   # append calls = 5230

dna length = 4,639,221
cutting at enzyme gaattc
-----
Class               splicee          recomb     time
-----
SimpleStrand:           256        5,649,721  0.030   # append calls = 8084
SimpleStrand:           512        6,684,473  0.029   # append calls = 8084
SimpleStrand:         1,024        8,753,977  0.027   # append calls = 8084
SimpleStrand:         2,048       12,892,985 0.033   # append calls = 8084
SimpleStrand:         4,096       21,171,001 0.066   # append calls = 8084
SimpleStrand:         8,192       37,727,033 0.079   # append calls = 8084
dna length = 4,639,221
cutting at enzyme gaattc
-----
Class               splicee          recomb     time
```

SimpleStrand:	256	5,882,471	0.032	# append calls = 9946
SimpleStrand:	512	7,155,559	0.030	# append calls = 9946
SimpleStrand:	1,024	9,701,735	0.029	# append calls = 9946
SimpleStrand:	2,048	14,794,087	0.037	# append calls = 9946
SimpleStrand:	4,096	24,978,791	0.055	# append calls = 9946

dna length = 4,642,839
cutting at enzyme gaattc

Class	splicee	recomb	time	
SimpleStrand:	256	7,661,589	0.036	# append calls = 24150
SimpleStrand:	512	10,752,789	0.041	# append calls = 24150
SimpleStrand:	1,024	16,935,189	0.038	# append calls = 24150
SimpleStrand:	2,048	29,299,989	0.050	# append calls = 24150

Graph:



Analysis:

The graph of running time vs. recombined length, as shown above, shows the linear relationship between the two variables is linear. This is supported by the fact that R^2 is 0.90267 for the graph, which is close to 1. I increased B (the number of occurrences of the enzyme each time without changing the length of the original text (ecoli.txt).

To solve the n/B relationship, I created a new text, ecoli4, and copy pasted ecoli.txt onto it. I replaced other six-letter enzymes, such as “tttgt”s in the file, with “gaattc”s. Each time I run benchmark I replaced more enzymes with “gaattc”. To keep the number of splicee constant, I

chose the recombined length and running time that correspond to 2048 splicee in each of the data tables to form the graph above.

Conclusion

In conclusion, the relationship of running time with recombined length (n+BI) is linear.

2.

Data:

Class	largest splicee	recomb	time	
512M SimpleStrand:	32,768	25,770,711	0.055	# append calls = 1290
1024M SimpleStrand:	131,072	89,176,791	0.147	# append calls = 1290
2048 M SimpleStrand	262,144	173,718,231	0.265	# append calls = 1290
4096 M SimpleStrand	524,288	342,801,111	0.506	# append calls = 1290
8192 M SimpleStrand:	1,048,576	680,966,871	0.980	# append calls = 1290
16384M SimpleStrand	1,048,576	680,966,871	2.607	# append calls = 1290

Analysis:

This is the data I obtained from varying the memory sizes. I obtained this data by running benchmark on ecoli.txt while changing the memory sizes. After running benchmark, I determined that the largest splicee size for 512 M memory size is 32,768, with running time 0.055 second. The largest splicee size for 1024M momeory size is 131,072, with running time 0.147 second. The part of code in benchmark that changes the string length is as follows:

```
for (int j = 8; j <= 32; j++) {
                    StringBuilder b = new StringBuilder("");
                    int spSize = (int) Math.pow(2, j);
                    for (int k = 0; k < spSize; k++) {
                            b.append("c");
                    }
                    String splicee = b.toString();
                    String results = strandSpliceBenchmark(enzyme, splicee,
strandType);
```

As heap size increases, the recombined length (n+Bl) increases. However, as the memory size reaches around 8000MB, the largest splicee size and recombined length stop changing. This corresponds to the memory of my computer, which is 8 GB.

3.

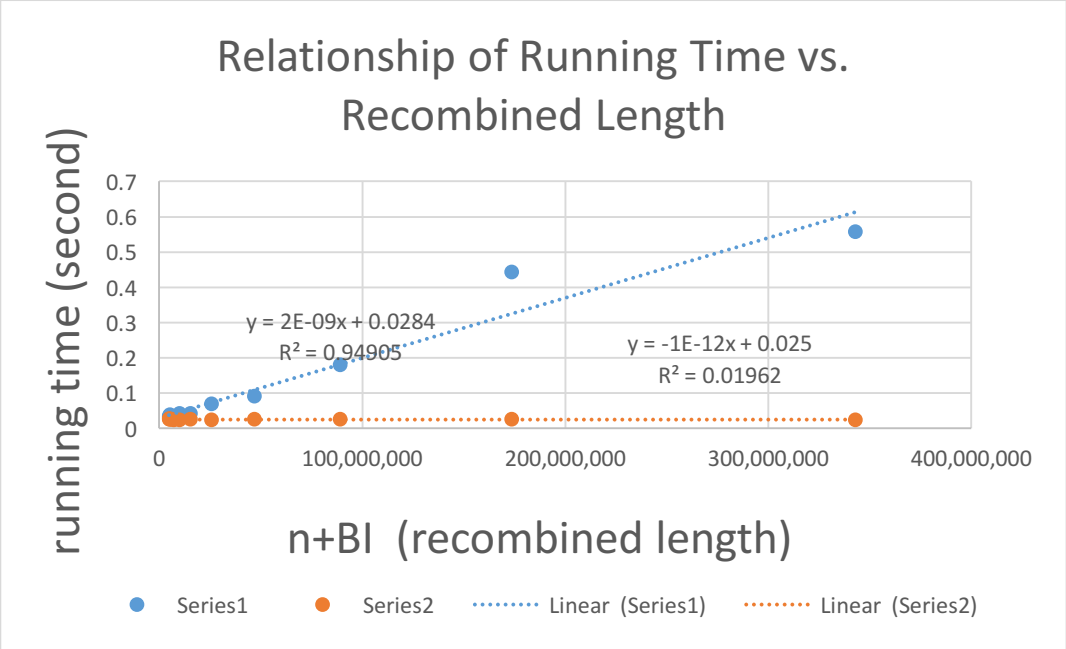
When l changes (data for SimpleStrand is already included in the answer to Question 1)

Data:

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

Class	splicee	recomb	time	
LinkStrand:	256	4,800,471	0.027	# append calls = 1290
LinkStrand:	512	4,965,591	0.025	# append calls = 1290
LinkStrand:	1,024	5,295,831	0.025	# append calls = 1290
LinkStrand:	2,048	5,956,311	0.025	# append calls = 1290
LinkStrand:	4,096	7,277,271	0.024	# append calls = 1290
LinkStrand:	8,192	9,919,191	0.024	# append calls = 1290
LinkStrand:	16,384	15,203,031	0.025	# append calls = 1290
LinkStrand:	32,768	25,770,711	0.024	# append calls = 1290
LinkStrand:	65,536	46,906,071	0.025	# append calls = 1290
LinkStrand:	131,072	89,176,791	0.025	# append calls = 1290
LinkStrand:	262,144	173,718,231	0.026	# append calls = 1290
LinkStrand:	524,288	342,801,111	0.024	# append calls = 1290
LinkStrand:	1,048,576	680,966,871	0.024	# append calls = 1290
LinkStrand:	2,097,152	1,357,298,391	0.028	# append calls = 1290
LinkStrand:	4,194,304	2,709,961,431	0.025	# append calls = 1290
LinkStrand:	8,388,608	5,415,287,511	0.027	# append calls = 1290
LinkStrand:	16,777,216	10,825,939,671	0.027	# append calls = 1290
LinkStrand:	33,554,432	21,647,243,991	0.025	# append calls = 1290
LinkStrand:	67,108,864	43,289,852,631	0.024	# append calls = 1290

Graph:



Analysis:

As the graph shows, the orange line represents LinkStrand and the blue line represents SimpleStrand. It is evident that Linkstrand improves upon running time, as for the same recombined length its running time is less than that of SimpleStrand.

When only n changes (data for SimpleStrand is already included in the answer to Question 1)

Data:

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

Class	splicee	recomb	time	
LinkStrand:	256	4,800,471	0.027	# append calls = 1290
LinkStrand:	512	4,965,591	0.026	# append calls = 1290
LinkStrand:	1,024	5,295,831	0.025	# append calls = 1290
LinkStrand:	2,048	5,956,311	0.024	# append calls = 1290
LinkStrand:	4,096	7,277,271	0.025	# append calls = 1290
LinkStrand:	8,192	9,919,191	0.028	# append calls = 1290
LinkStrand:	16,384	15,203,031	0.026	# append calls = 1290
LinkStrand:	32,768	25,770,711	0.025	# append calls = 1290
LinkStrand:	65,536	46,906,071	0.027	# append calls = 1290
LinkStrand:	131,072	89,176,791	0.027	# append calls = 1290
LinkStrand:	262,144	173,718,231	0.025	# append calls = 1290
LinkStrand:	524,288	342,801,111	0.026	# append calls = 1290
LinkStrand:	1,048,576	680,966,871	0.028	# append calls = 1290
LinkStrand:	2,097,152	1,357,298,391	0.026	# append calls = 1290
LinkStrand:	4,194,304	2,709,961,431	0.026	# append calls = 1290
LinkStrand:	8,388,608	5,415,287,511	0.026	# append calls = 1290
LinkStrand:	16,777,216	10,825,939,671	0.025	# append calls = 1290
LinkStrand:	33,554,432	21,647,243,991	0.031	# append calls = 1290
LinkStrand:	67,108,864	43,289,852,631	0.029	# append calls = 1290

dna length = 9,278,442
cutting at enzyme gaattc

Class	splicee	recomb	time	
LinkStrand:	256	9,439,692	0.033	# append calls = 1290
LinkStrand:	512	9,604,812	0.032	# append calls = 1290
LinkStrand:	1,024	9,935,052	0.029	# append calls = 1290
LinkStrand:	2,048	10,595,532	0.029	# append calls = 1290
LinkStrand:	4,096	11,916,492	0.031	# append calls = 1290
LinkStrand:	8,192	14,558,412	0.034	# append calls = 1290
LinkStrand:	16,384	19,842,252	0.032	# append calls = 1290

```
LinkStrand:          32,768       30,409,932 0.033  # append calls = 1290
LinkStrand:          65,536       51,545,292 0.031  # append calls = 1290
LinkStrand:         131,072       93,816,012 0.035  # append calls = 1290
LinkStrand:         262,144      178,357,452        0.033  # append calls = 1290
LinkStrand:         524,288      347,440,332        0.034  # append calls = 1290
LinkStrand:       1,048,576      685,606,092        0.030  # append calls = 1290
LinkStrand:       2,097,152    1,361,937,612        0.035  # append calls = 1290
LinkStrand:       4,194,304    2,714,600,652        0.031  # append calls = 1290
LinkStrand:       8,388,608    5,419,926,732        0.035  # append calls = 1290
LinkStrand:      16,777,216   10,830,578,892        0.033  # append calls = 1290
LinkStrand:      33,554,432   21,651,883,212        0.034  # append calls = 1290
LinkStrand:      67,108,864   43,294,491,852        0.036  # append calls = 1290

dna length = 13,917,663
cutting at enzyme gaattc
-----
Class              splicee          recomb      time
-----
LinkStrand:          256          14,078,913 0.043  # append calls = 1290
LinkStrand:          512          14,244,033 0.040  # append calls = 1290
LinkStrand:        1,024          14,574,273 0.036  # append calls = 1290
LinkStrand:        2,048          15,234,753 0.033  # append calls = 1290
LinkStrand:        4,096          16,555,713 0.039  # append calls = 1290
LinkStrand:        8,192          19,197,633 0.042  # append calls = 1290
LinkStrand:       16,384          24,481,473 0.041  # append calls = 1290
LinkStrand:       32,768          35,049,153 0.046  # append calls = 1290
LinkStrand:       65,536          56,184,513 0.038  # append calls = 1290
LinkStrand:      131,072          98,455,233 0.039  # append calls = 1290
LinkStrand:      262,144         182,996,673        0.040  # append calls = 1290
LinkStrand:      524,288         352,079,553        0.039  # append calls = 1290
LinkStrand:    1,048,576         690,245,313        0.039  # append calls = 1290
LinkStrand:    2,097,152       1,366,576,833        0.044  # append calls = 1290
LinkStrand:    4,194,304       2,719,239,873        0.038  # append calls = 1290
LinkStrand:    8,388,608       5,424,565,953        0.040  # append calls = 1290
LinkStrand:   16,777,216      10,835,218,113        0.040  # append calls = 1290
LinkStrand:   33,554,432      21,656,522,433        0.037  # append calls = 1290
LinkStrand:   67,108,864      43,299,131,073        0.042  # append calls = 1290

dna length = 18,556,884
cutting at enzyme gaattc
-----
Class              splicee          recomb      time
-----
LinkStrand:          256          18,718,134 0.053  # append calls = 1290
LinkStrand:          512          18,883,254 0.049  # append calls = 1290
LinkStrand:        1,024          19,213,494 0.042  # append calls = 1290
```

LinkStrand:	2,048	19,873,974	0.044	# append calls = 1290
LinkStrand:	4,096	21,194,934	0.046	# append calls = 1290
LinkStrand:	8,192	23,836,854	0.058	# append calls = 1290
LinkStrand:	16,384	29,120,694	0.053	# append calls = 1290
LinkStrand:	32,768	39,688,374	0.052	# append calls = 1290
LinkStrand:	65,536	60,823,734	0.050	# append calls = 1290
LinkStrand:	131,072	103,094,454	0.055	# append calls = 1290
LinkStrand:	262,144	187,635,894	0.046	# append calls = 1290
LinkStrand:	524,288	356,718,774	0.052	# append calls = 1290
LinkStrand:	1,048,576	694,884,534	0.048	# append calls = 1290
LinkStrand:	2,097,152	1,371,216,054	0.050	# append calls = 1290
LinkStrand:	4,194,304	2,723,879,094	0.051	# append calls = 1290
LinkStrand:	8,388,608	5,429,205,174	0.052	# append calls = 1290
LinkStrand:	16,777,216	10,839,857,334	0.057	# append calls = 1290
LinkStrand:	33,554,432	21,661,161,654	0.058	# append calls = 1290
LinkStrand:	67,108,864	43,303,770,294	0.056	# append calls = 1290

dna length = 23,196,105
cutting at enzyme gaattc

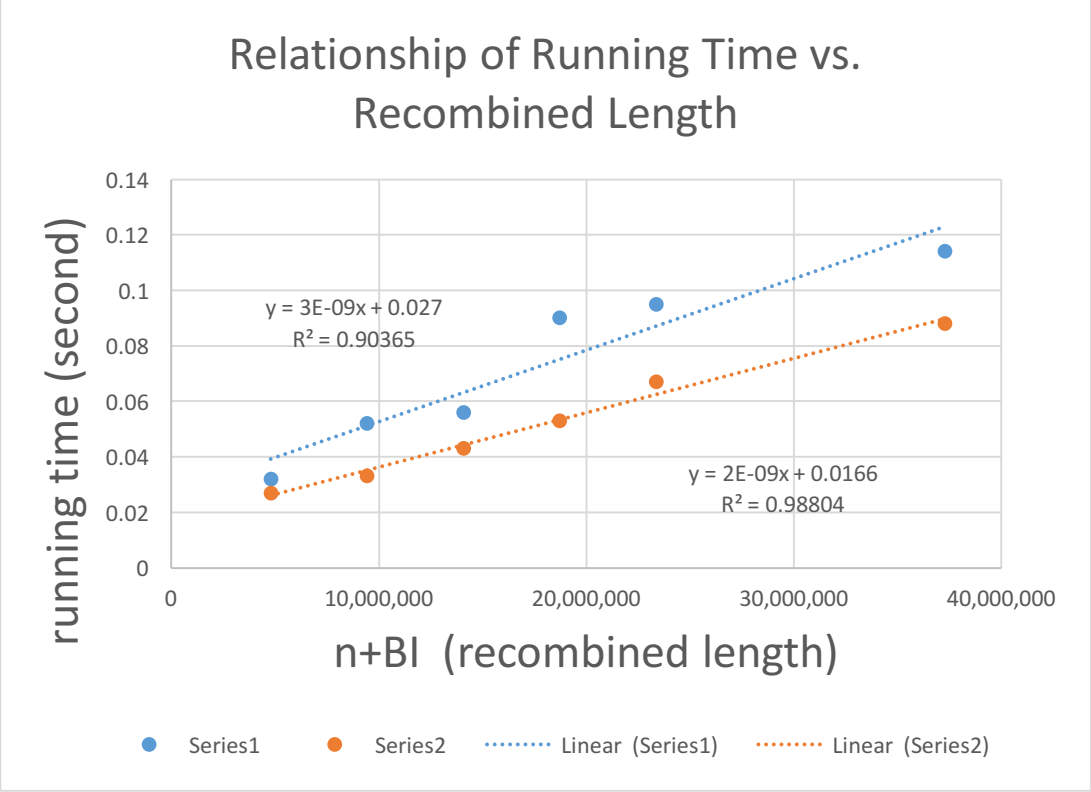
Class	splicee	recomb	time	
LinkStrand:	256	23,357,355	0.067	# append calls = 1290
LinkStrand:	512	23,522,475	0.058	# append calls = 1290
LinkStrand:	1,024	23,852,715	0.052	# append calls = 1290
LinkStrand:	2,048	24,513,195	0.058	# append calls = 1290
LinkStrand:	4,096	25,834,155	0.056	# append calls = 1290
LinkStrand:	8,192	28,476,075	0.056	# append calls = 1290
LinkStrand:	16,384	33,759,915	0.055	# append calls = 1290
LinkStrand:	32,768	44,327,595	0.055	# append calls = 1290
LinkStrand:	65,536	65,462,955	0.056	# append calls = 1290
LinkStrand:	131,072	107,733,675	0.056	# append calls = 1290
LinkStrand:	262,144	192,275,115	0.056	# append calls = 1290
LinkStrand:	524,288	361,357,995	0.056	# append calls = 1290
LinkStrand:	1,048,576	699,523,755	0.055	# append calls = 1290
LinkStrand:	2,097,152	1,375,855,275	0.056	# append calls = 1290
LinkStrand:	4,194,304	2,728,518,315	0.055	# append calls = 1290
LinkStrand:	8,388,608	5,433,844,395	0.057	# append calls = 1290
LinkStrand:	16,777,216	10,844,496,555	0.054	# append calls = 1290
LinkStrand:	33,554,432	21,665,800,875	0.053	# append calls = 1290
LinkStrand:	67,108,864	43,308,409,515	0.058	# append calls = 1290

dna length = 37,113,768
cutting at enzyme gaattc

Class	splicee	recomb	time	

LinkStrand:	256	37,275,018	0.088	# append calls = 1290
LinkStrand:	512	37,440,138	0.078	# append calls = 1290
LinkStrand:	1,024	37,770,378	0.081	# append calls = 1290
LinkStrand:	2,048	38,430,858	0.080	# append calls = 1290
LinkStrand:	4,096	39,751,818	0.082	# append calls = 1290
LinkStrand:	8,192	42,393,738	0.081	# append calls = 1290
LinkStrand:	16,384	47,677,578	0.081	# append calls = 1290
LinkStrand:	32,768	58,245,258	0.079	# append calls = 1290
LinkStrand:	65,536	79,380,618	0.083	# append calls = 1290
LinkStrand:	131,072	121,651,338	0.077	# append calls = 1290
LinkStrand:	262,144	206,192,778	0.082	# append calls = 1290
LinkStrand:	524,288	375,275,658	0.084	# append calls = 1290
LinkStrand:	1,048,576	713,441,418	0.087	# append calls = 1290
LinkStrand:	2,097,152	1,389,772,938	0.086	# append calls = 1290
LinkStrand:	4,194,304	2,742,435,978	0.087	# append calls = 1290
LinkStrand:	8,388,608	5,447,762,058	0.095	# append calls = 1290
LinkStrand:	16,777,216	10,858,414,218	0.095	# append calls = 1290
LinkStrand:	33,554,432	21,679,718,538	0.107	# append calls = 1290

Graph:



Analysis:

As the graph shows, the orange line represents LinkStrand and the blue line represents SimpleStrand. It is evident that Linkstrand improves upon running time, as for the same recombined length its running time is less than that of SimpleStrand.

When only B changes (data for SimpleStrand is already included in the answer to Question 1)

Data:

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

Class	splicee	recomb	time	
LinkStrand:	256	4,874,721	0.028	# append calls = 1884
LinkStrand:	512	5,115,873	0.028	# append calls = 1884
LinkStrand:	1,024	5,598,177	0.028	# append calls = 1884
LinkStrand:	2,048	6,562,785	0.027	# append calls = 1884
LinkStrand:	4,096	8,492,001	0.027	# append calls = 1884
LinkStrand:	8,192	12,350,433	0.026	# append calls = 1884
LinkStrand:	16,384	20,067,297	0.028	# append calls = 1884
LinkStrand:	32,768	35,501,025	0.024	# append calls = 1884
LinkStrand:	65,536	66,368,481	0.024	# append calls = 1884
LinkStrand:	131,072	128,103,393	0.024	# append calls = 1884
LinkStrand:	262,144	251,573,217	0.027	# append calls = 1884
LinkStrand:	524,288	498,512,865	0.028	# append calls = 1884
LinkStrand:	1,048,576	992,392,161	0.024	# append calls = 1884
LinkStrand:	2,097,152	1,980,150,753	0.025	# append calls = 1884
LinkStrand:	4,194,304	3,955,667,937	0.026	# append calls = 1884
LinkStrand:	8,388,608	7,906,702,305	0.026	# append calls = 1884
LinkStrand:	16,777,216	15,808,771,041	0.027	# append calls = 1884
LinkStrand:	33,554,432	31,612,908,513	0.027	# append calls = 1884
LinkStrand:	67,108,864	63,221,183,457	0.027	# append calls = 1884

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

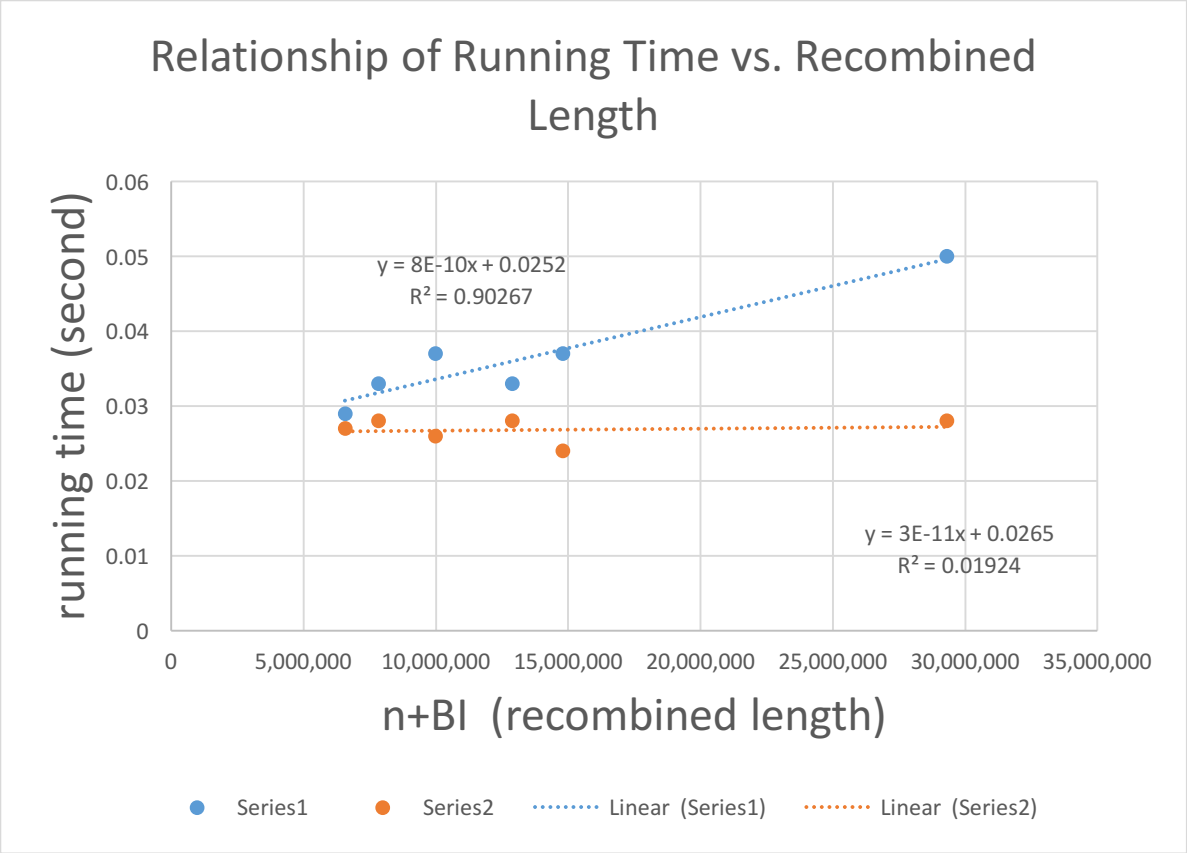
Class	splicee	recomb	time	
LinkStrand:	256	5,029,721	0.029	# append calls = 3124
LinkStrand:	512	5,429,593	0.029	# append calls = 3124
LinkStrand:	1,024	6,229,337	0.027	# append calls = 3124
LinkStrand:	2,048	7,828,825	0.028	# append calls = 3124
LinkStrand:	4,096	11,027,801	0.028	# append calls = 3124
LinkStrand:	8,192	17,425,753	0.027	# append calls = 3124
LinkStrand:	16,384	30,221,657	0.026	# append calls = 3124

```
LinkStrand:          32,768       55,813,465 0.026  # append calls = 3124
LinkStrand:          65,536      106,997,081        0.025  # append calls = 3124
LinkStrand:         131,072      209,364,313        0.026  # append calls = 3124
LinkStrand:         262,144      414,098,777        0.027  # append calls = 3124
LinkStrand:         524,288      823,567,705        0.026  # append calls = 3124
LinkStrand:       1,048,576    1,642,505,561        0.025  # append calls = 3124
LinkStrand:       2,097,152    3,280,381,273        0.024  # append calls = 3124
LinkStrand:       4,194,304    6,556,132,697        0.025  # append calls = 3124
LinkStrand:       8,388,608   13,107,635,545        0.029  # append calls = 3124
LinkStrand:      16,777,216   26,210,641,241        0.034  # append calls = 3124
LinkStrand:      33,554,432   52,416,652,633        0.028  # append calls = 3124
LinkStrand:      67,108,864  104,828,675,417        0.028  # append calls = 3124

dna length = 4,639,221
cutting at enzyme gaattc
-----
Class              splicee          recomb     time
-----
LinkStrand:          256          5,292,971  0.027  # append calls = 5230
LinkStrand:          512          5,962,411  0.026  # append calls = 5230
LinkStrand:        1,024          7,301,291  0.025  # append calls = 5230
LinkStrand:        2,048          9,979,051  0.026  # append calls = 5230
LinkStrand:        4,096         15,334,571  0.025  # append calls = 5230
LinkStrand:        8,192         26,045,611  0.024  # append calls = 5230
LinkStrand:       16,384         47,467,691  0.024  # append calls = 5230
LinkStrand:       32,768         90,311,851  0.024  # append calls = 5230
LinkStrand:       65,536        176,000,171        0.024  # append calls = 5230
LinkStrand:      131,072        347,376,811        0.023  # append calls = 5230
LinkStrand:      262,144        690,130,091        0.024  # append calls = 5230
LinkStrand:      524,288      1,375,636,651        0.027  # append calls = 5230
LinkStrand:    1,048,576      2,746,649,771        0.029  # append calls = 5230
LinkStrand:    2,097,152      5,488,676,011        0.025  # append calls = 5230
LinkStrand:    4,194,304     10,972,728,491        0.025  # append calls = 5230
LinkStrand:    8,388,608     21,940,833,451        0.027  # append calls = 5230
LinkStrand:   16,777,216     43,877,043,371        0.026  # append calls = 5230
LinkStrand:   33,554,432     87,749,463,211        0.026  # append calls = 5230
LinkStrand:   67,108,864    175,494,302,891        0.028  # append calls = 5230

dna length = 4,639,221
cutting at enzyme gaattc
-----
Class              splicee          recomb     time
-----
LinkStrand:          256          5,649,721  0.028  # append calls = 8084
LinkStrand:          512          6,684,473  0.026  # append calls = 8084
LinkStrand:        1,024          8,753,977  0.026  # append calls = 8084
```

```
LinkStrand:          2,048        12,892,985 0.028  # append calls = 8084
LinkStrand:          4,096        21,171,001 0.026  # append calls = 8084
LinkStrand:          8,192        37,727,033 0.027  # append calls = 8084
LinkStrand:         16,384        70,839,097 0.028  # append calls = 8084
LinkStrand:         32,768       137,063,225        0.028  # append calls = 8084
LinkStrand:         65,536       269,511,481        0.028  # append calls = 8084
LinkStrand:        131,072       534,407,993        0.028  # append calls = 8084
LinkStrand:        262,144     1,064,201,017        0.026  # append calls = 8084
LinkStrand:        524,288     2,123,787,065        0.029  # append calls = 8084
LinkStrand:      1,048,576     4,242,959,161        0.027  # append calls = 8084
LinkStrand:      2,097,152     8,481,303,353        0.026  # append calls = 8084
LinkStrand:      4,194,304    16,957,991,737        0.026  # append calls = 8084
LinkStrand:      8,388,608    33,911,368,505        0.030  # append calls = 8084
LinkStrand:     16,777,216    67,818,122,041        0.029  # append calls = 8084
LinkStrand:     33,554,432   135,631,629,113        0.026  # append calls = 8084
LinkStrand:     67,108,864   271,258,643,257        0.029  # append calls = 8084
dna length = 4,639,221
cutting at enzyme gaattc
-----
Class              splicee           recomb      time
-----
LinkStrand:            256         5,882,471  0.027  # append calls = 9946
LinkStrand:            512         7,155,559  0.026  # append calls = 9946
LinkStrand:          1,024         9,701,735  0.025  # append calls = 9946
LinkStrand:          2,048        14,794,087 0.024  # append calls = 9946
LinkStrand:          4,096        24,978,791 0.024  # append calls = 9946
LinkStrand:          8,192        45,348,199 0.025  # append calls = 9946
LinkStrand:         16,384        86,087,015 0.025  # append calls = 9946
LinkStrand:         32,768       167,564,647        0.024  # append calls = 9946
LinkStrand:         65,536       330,519,911        0.025  # append calls = 9946
LinkStrand:        131,072       656,430,439        0.026  # append calls = 9946
LinkStrand:        262,144     1,308,251,495        0.028  # append calls = 9946
LinkStrand:        524,288     2,611,893,607        0.027  # append calls = 9946
LinkStrand:      1,048,576     5,219,177,831        0.027  # append calls = 9946
LinkStrand:      2,097,152    10,433,746,279        0.027  # append calls = 9946
LinkStrand:      4,194,304    20,862,883,175        0.029  # append calls = 9946
LinkStrand:      8,388,608    41,721,156,967        0.028  # append calls = 9946
LinkStrand:     16,777,216    83,437,704,551        0.033  # append calls = 9946
LinkStrand:     33,554,432   166,870,799,719        0.029  # append calls = 9946
LinkStrand:     67,108,864   333,736,990,055        0.030  # append calls = 9946


dna length = 4,642,839
cutting at enzyme gaattc
-----
Class              splicee           recomb      time
```

LinkStrand:	256	7,661,589	0.029	# append calls = 24150
LinkStrand:	512	10,752,789	0.029	# append calls = 24150
LinkStrand:	1,024	16,935,189	0.028	# append calls = 24150
LinkStrand:	2,048	29,299,989	0.028	# append calls = 24150
LinkStrand:	4,096	54,029,589	0.029	# append calls = 24150
LinkStrand:	8,192	103,488,789	0.027	# append calls = 24150
LinkStrand:	16,384	202,407,189	0.027	# append calls = 24150
LinkStrand:	32,768	400,243,989	0.028	# append calls = 24150
LinkStrand:	65,536	795,917,589	0.030	# append calls = 24150
LinkStrand:	131,072	1,587,264,789	0.027	# append calls = 24150
LinkStrand:	262,144	3,169,959,189	0.026	# append calls = 24150
LinkStrand:	524,288	6,335,347,989	0.028	# append calls = 24150
LinkStrand:	1,048,576	12,666,125,589	0.029	# append calls = 24150
LinkStrand:	2,097,152	25,327,680,789	0.029	# append calls = 24150
LinkStrand:	4,194,304	50,650,791,189	0.028	# append calls = 24150
LinkStrand:	8,388,608	101,297,011,989	0.028	# append calls = 24150
LinkStrand:	16,777,216	202,589,453,589	0.032	# append calls = 24150
LinkStrand:	33,554,432	405,174,336,789	0.029	# append calls = 24150
LinkStrand:	67,108,864	810,344,103,189	0.029	# append calls = 24150

Graph:



Analysis:

As the graph shows, the orange line represents LinkStrand and the blue line represents SimpleStrand. It is evident that Linkstrand improves upon running time, as for the same recombined length its running time is less than that of SimpleStrand.

Conclusion:

Linkstrand improves upon running time in regards to SimpleStrand. This is because SimpleStrand uses StringBuilder to append Strings while LinkStrand does not. When Strings are appended to StringBuilder in SimpleStrand, the characters of the String argument are appended individually, in order. However, for LinkStrand, when a node is appended to another node, all of the content of this node, without going through the content of that node one by one like for StringBuilder, gets appended to the other node. Therefore, LinkStrand is faster than SimpleStrand which uses StringBuilder.