

String Strand Timing Data

dna length = 320,160

cutting at enzyme gaattc

Class	dna,N	splicee,S	recomb	time(ms)	breaks,b
StringStra:	320,160	10,000	769,890	14	44
StringStra:	320,160	20,000	1,219,890	12	44
StringStra:	320,160	40,000	2,119,890	73	44
StringStra:	320,160	80,000	3,919,890	36	44
StringStra:	320,160	160,000	7,519,890	69	44
StringStra:	320,160	320,000	14,719,890	69	44
StringStra:	320,160	640,000	29,119,890	157	44
StringStra:	320,160	1,280,000	57,919,890	185	44
StringStra:	320,160	10,000	769,890	9	44
StringStra:	640,320	10,000	1,539,780	16	89
StringStra:	1,280,640	10,000	3,079,560	34	179
StringStra:	2,561,280	10,000	6,159,120	120	359
StringStra:	5,122,560	10,000	12,318,240	496	719
----- Memory Exhaustion Benchmark -----					

DNABenchmark

dna length = 320,160

cutting at enzyme gaattc

Class	dna,N	splicee,S	recomb	time(ms)	breaks,b
StringStra:	320,160	10,000	769,890	6	44
StringStra:	320,160	20,000	1,219,890	17	44
StringStra:	320,160	40,000	2,119,890	10	44
StringStra:	320,160	80,000	3,919,890	13	44
StringStra:	320,160	160,000	7,519,890	19	44
StringStra:	320,160	320,000	14,719,890	36	44
StringStra:	320,160	640,000	29,119,890	67	44
StringStra:	320,160	1,280,000	57,919,890	139	44
StringStra:	320,160	10,000	769,890	6	44
StringStra:	640,320	10,000	1,539,780	11	89
StringStra:	1,280,640	10,000	3,079,560	33	179
StringStra:	2,561,280	10,000	6,159,120	115	359
StringStra:	5,122,560	10,000	12,318,240	442	719
----- Memory Exhaustion Benchmark -----					

Class	splicee	recomb	time	appends
StringStra:	256	4,800,471	0.042	1290
StringStra:	512	4,965,591	0.044	1290
StringStra:	1,024	5,295,831	0.037	1290
StringStra:	2,048	5,956,311	0.041	1290
StringStra:	4,096	7,277,271	0.051	1290
StringStra:	8,192	9,919,191	0.069	1290
StringStra:	16,384	15,203,031	0.100	1290
StringStra:	32,768	25,770,711	0.178	1290
StringStra:	65,536	46,906,071	0.313	1290
StringStra:	131,072	89,176,791	0.691	1290
StringStra:	262,144	173,718,231	1.289	1290
StringStra:	524,288	342,801,111	2.418	1290

StringBuilder timing Data

dna length = 320,160

cutting at enzyme gaattc

Class	dna,N	splicee,S	recomb	time(ms)	breaks,b
StringBuil:	320,160	10,000	769,890	3	44
StringBuil:	320,160	20,000	1,219,890	3	44
StringBuil:	320,160	40,000	2,119,890	3	44
StringBuil:	320,160	80,000	3,919,890	3	44
StringBuil:	320,160	160,000	7,519,890	3	44
StringBuil:	320,160	320,000	14,719,890	7	44
StringBuil:	320,160	640,000	29,119,890	9	44
StringBuil:	320,160	1,280,000	57,919,890	21	44
StringBuil:	320,160	10,000	769,890	1	44
StringBuil:	640,320	10,000	1,539,780	3	89
StringBuil:	1,280,640	10,000	3,079,560	6	179
StringBuil:	2,561,280	10,000	6,159,120	10	359
StringBuil:	5,122,560	10,000	12,318,240	18	719

----- Memory Exhaustion Benchmark -----

dna length = 320,160

cutting at enzyme gaattc

Class	dna,N	splicee,S	recomb	time(ms)	breaks,b
-------	-------	-----------	--------	----------	----------

```

-----
StringBuil: 320,160 10,000 769,890 2 44
StringBuil: 320,160 20,000 1,219,890 2 44
StringBuil: 320,160 40,000 2,119,890 3 44
StringBuil: 320,160 80,000 3,919,890 3 44
StringBuil: 320,160 160,000 7,519,890 5 44
StringBuil: 320,160 320,000 14,719,890 8 44
StringBuil: 320,160 640,000 29,119,890 10 44
StringBuil: 320,160 1,280,000 57,919,890 11 44
StringBuil: 320,160 10,000 769,890 2 44
StringBuil: 640,320 10,000 1,539,780 3 89
StringBuil: 1,280,640 10,000 3,079,560 7 179
StringBuil: 2,561,280 10,000 6,159,120 9 359
StringBuil: 5,122,560 10,000 12,318,240 18 719
----- Memory Exhaustion Benchmark -----

```

LinkStrand timing Data
dna length = 320,160
cutting at enzyme gaattc

```

-----
Class      dna,N splicee,S   recomb time(ms) breaks,b
-----
LinkStrand: 320,160 10,000 769,890 2 44
LinkStrand: 320,160 20,000 1,219,890 3 44
LinkStrand: 320,160 40,000 2,119,890 2 44
LinkStrand: 320,160 80,000 3,919,890 2 44
LinkStrand: 320,160 160,000 7,519,890 2 44
LinkStrand: 320,160 320,000 14,719,890 2 44
LinkStrand: 320,160 640,000 29,119,890 1 44
LinkStrand: 320,160 1,280,000 57,919,890 1 44
LinkStrand: 320,160 10,000 769,890 1 44
LinkStrand: 640,320 10,000 1,539,780 3 89
LinkStrand: 1,280,640 10,000 3,079,560 6 179
LinkStrand: 2,561,280 10,000 6,159,120 9 359
LinkStrand: 5,122,560 10,000 12,318,240 17 719
----- Memory Exhaustion Benchmark -----

```

dna length = 320,160
cutting at enzyme gaattc

```

-----
Class      dna,N splicee,S   recomb time(ms) breaks,b

```

```

-----
LinkStrand: 320,160 10,000 769,890 2 44
LinkStrand: 320,160 20,000 1,219,890 2 44
LinkStrand: 320,160 40,000 2,119,890 2 44
LinkStrand: 320,160 80,000 3,919,890 2 44
LinkStrand: 320,160 160,000 7,519,890 2 44
LinkStrand: 320,160 320,000 14,719,890 2 44
LinkStrand: 320,160 640,000 29,119,890 2 44
LinkStrand: 320,160 1,280,000 57,919,890 1 44
LinkStrand: 320,160 10,000 769,890 1 44
LinkStrand: 640,320 10,000 1,539,780 2 89
LinkStrand: 1,280,640 10,000 3,079,560 5 179
LinkStrand: 2,561,280 10,000 6,159,120 10 359
LinkStrand: 5,122,560 10,000 12,318,240 17 719
----- Memory Exhaustion Benchmark -----

```

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

```

-----
Class      dna,N splicee,S   recomb time(ms) breaks,b
-----
LinkStrand: 4,639,221 10,000 11,085,351 20 644
LinkStrand: 4,639,221 20,000 17,535,351 16 644
LinkStrand: 4,639,221 40,000 30,435,351 15 644
LinkStrand: 4,639,221 80,000 56,235,351 15 644
LinkStrand: 4,639,221 160,000 107,835,351 14 644
LinkStrand: 4,639,221 320,000 211,035,351 14 644
LinkStrand: 4,639,221 640,000 417,435,351 15 644
LinkStrand: 4,639,221 1,280,000 830,235,351 15 644
LinkStrand: 4,639,221 10,000 11,085,351 14 644
LinkStrand: 9,278,442 10,000 22,170,702 32 1,289
LinkStrand: 18,556,884 10,000 44,341,404 59 2,579
LinkStrand: 37,113,768 10,000 88,682,808 127 5,159
LinkStrand: 74,227,536 10,000 177,365,616 256 10,319
----- Memory Exhaustion Benchmark -----

```

Question 1 .

The time complexity of the cutAndSplice function in StringStrand should be expressed as $O(Nb + b^2S)$. This is due to the quadratic runtime (b^2S) involved in appending the new string intended for replacement, where b represents the number of strings to be created. Additionally, the time required for adding the non-replaced string is proportional to N/b , and since this operation is performed b times, it contributes Nb to the overall complexity which is added to b^2S . The observed empirical results align with the anticipated theoretical expectations, demonstrating a quadratic increase in time with each execution.

Question 2.

The time complexity of the cutAndSplice function in StringStrand is $O(N + bS)$. This is derived from the bS component, representing the runtime for appending a string of size S to a StringBuilder, repeated b times. The N component accounts for appending N/b -sized strings b times, resulting in a total of N . These two components are added together since they are not nested but occur in the same run. Empirical results confirm the theoretical expectations, showcasing a runtime that appears to follow a linear time complexity.

Question 3

A StringStrand object would consume $N + bS$ bytes of memory. This is determined by the total length of the resulting String in the StringStrand, with N representing the length of the appended strings and bS accounting for the size of each appended string, multiplied by the number of times the operation is performed. This memory usage is more efficient compared to StringBuilder, as the latter essentially operates as an array and needs to double in size when its limit is exceeded. This doubling process may result in unused memory spaces in the array, contributing to a higher overall memory footprint.

Question 4

The asymptotic runtime complexity of using the LinkStrand in the cutAndSplice method is $O(N)$. This is because the method iterates through the DNA Strand b times, creating a LinkStrand object and appending a substring of length S to each of these strands. Although the operation involves appending substrings of length S b times, the time complexity simplifies down to being dependent only on the size of the original LinkStrand, which is represented as N . Consequently, the overall time complexity of the method is $O(N)$.

Empirical results from DNABenchmark support this analysis, demonstrating a linear increase in runtime with the size of the original DNA Strand (N) and no discernible pattern beyond that. This empirical evidence reinforces the conclusion that the time complexity of the cutAndSplice method, when using LinkStrand, is empirically observed to be $O(N)$.

Question 5

In summary, the total memory required to store the result of a cutAndSplice operation on a LinkStrand object is approximately $(8 \times 2B) + N + S$ bytes. The initial DNA strand is composed

of a series of linked lists, with each node (totaling $2B$ nodes) containing strings of length S , resulting in a memory requirement of N bytes. The spliced DNA strand is stored as a string of length S . During the cutAndSplice operation, as B nodes are removed and replaced with the splice, each node will then occupy 8 bytes of memory storage. The overall memory needed for the new DNA strand after the cutAndSplice operation is given by $(2B) * 8$ bytes.

Additionally, considering the total number of characters needed to store the strands from the cutAndSplice operation within LinkStrand, which includes N (number of strands) and S (final strand), the overall memory requirement for the LinkStrand object with the operation is calculated as $(8*2B) + N + S$ bytes.