CS201 DNA Project Analysis[*]

Keping Wang (netid: kw238)

November 2, 2016

1 Benchmarking Method

I use the JMH framework for benchmarking. Please refer to my Markov Project Analysis for detailed explanation.

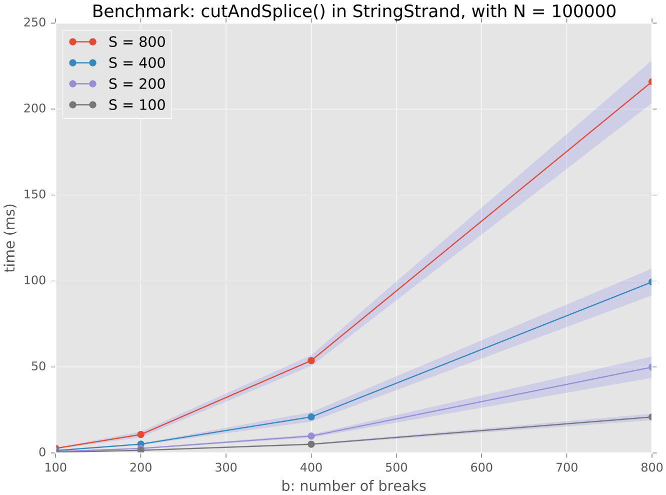
I generate the testing data with the simplest method. The enzyme is always a single character "B". The splicee is a sequence of character "S" with length S. The source is the concatenation of b "B"s and $(N - b)$ "x"s.

For example, with $N = 10$, $b = 2$, $S = 3$, the `cutAndSplice()` method (for `StringStrand`) would be:

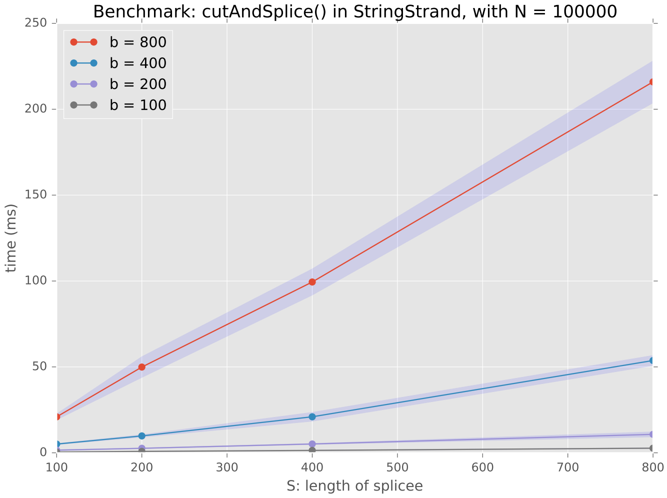
```java
String source = "BBxxxxxxxx";
String enzyme = "B";
String splicee= "SSS";
(new StringStrand(source)).cutAndSplice(enzyme, splicee);
```

2 Non Linked List Hypotheses

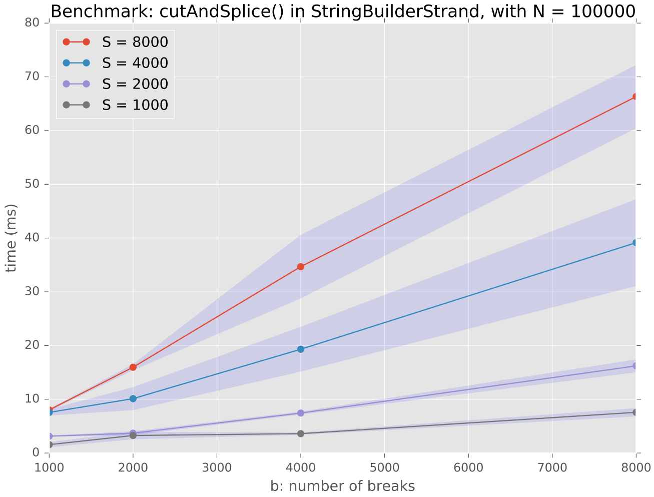
From the following two figures, we know that the `cutAndSplice()` method of `StringStrand` is $O(b^2 S)$.

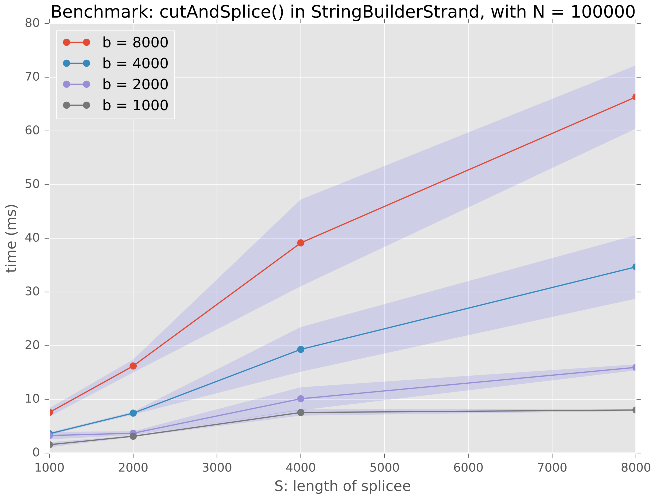


[*] See https://github.com/kepingwang/JavaBenchmark for the full benchmarking project.

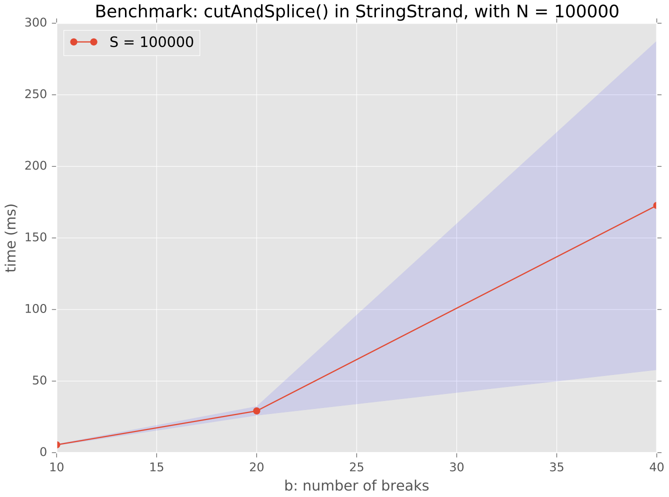


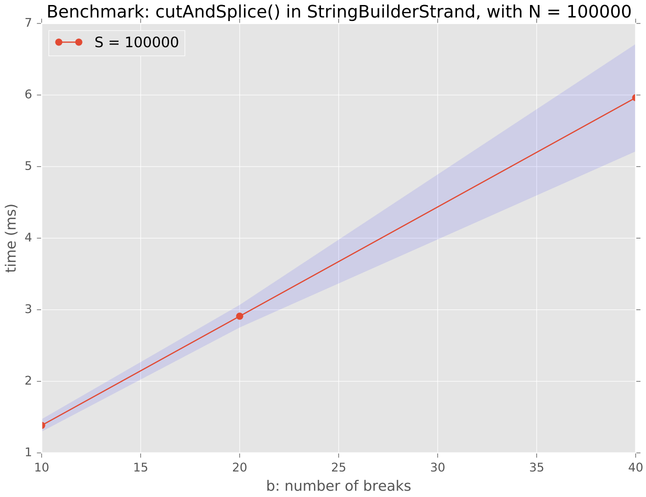
From the following two figures, we know that the `cutAndSplice()` method of `StringBuilderStrand` is $O(bS)$.





From the following two figures, we confirm that b and S are multiplied in the run time of non linked list strands because the run time is quadratic/linear in b even when b is very small relative to S.

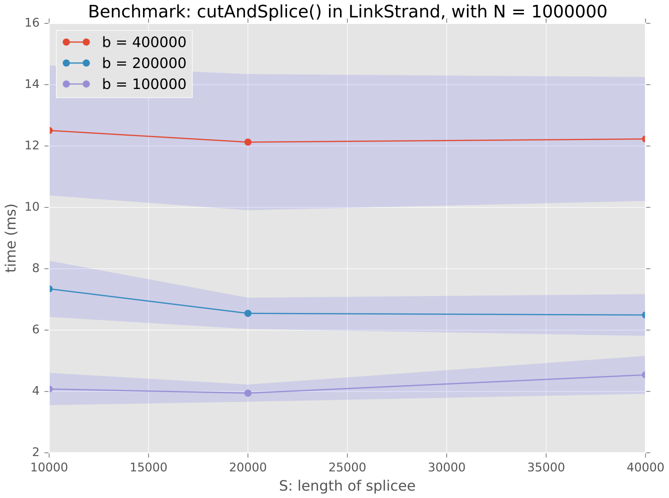




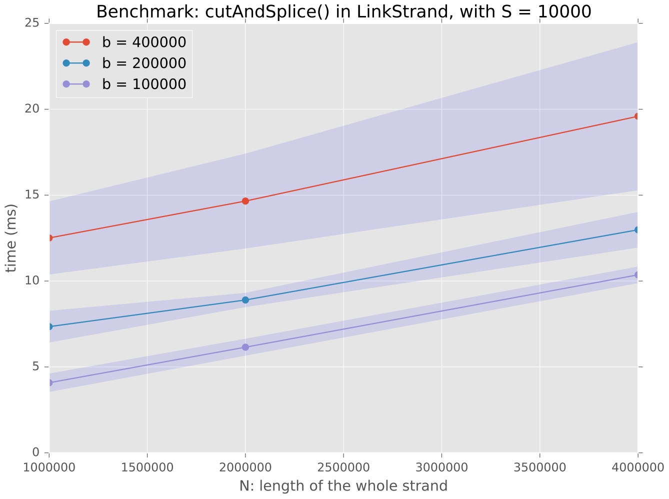
Finally the running time of `cutAndSplice()` for both `StringStrand` and `StringBuilderStrand` have to be at least $O(n)$, since the method has to go through the whole strand to find pieces that matches the enzyme.

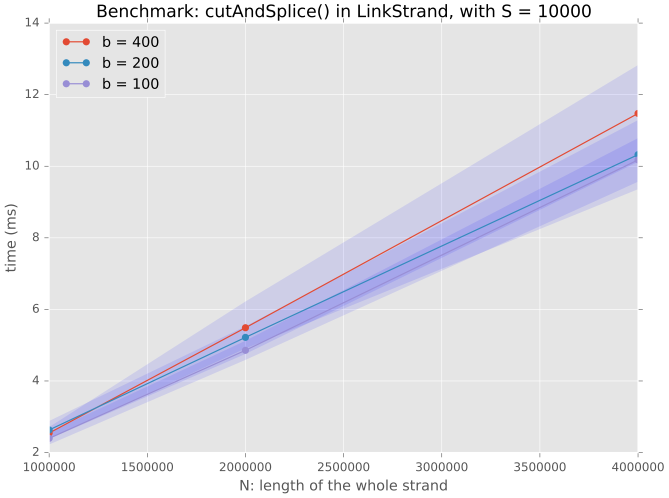
3 LinkStrand Hypothesis

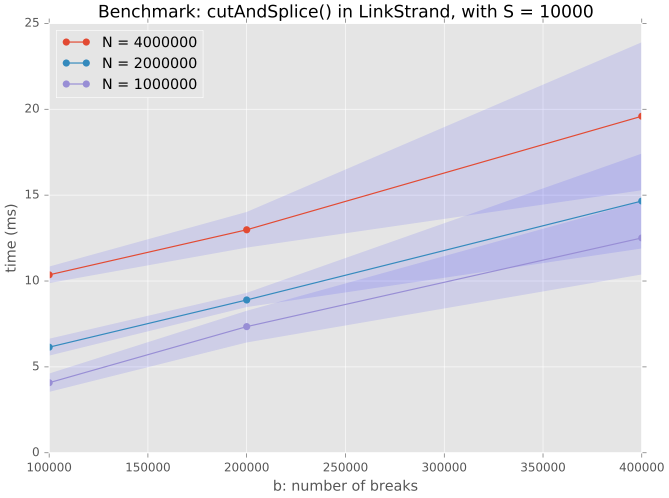
From the following figure, we can see that the run time of `cutAndSplice()` int `LinkStrand` is irrelevant to S (of course when S is much smaller than N).



From the two following figures, we notice that the run time of `cutAndSplice()` int `LinkStrand` is linear in both N and b. Besides, when b becomes very small relative to N, the run time is dominated by N, so we know that the run time is $O(b+N)$ (instead of $O(bN)$).






Benchmark: cutAndSplice() in LinkStrand, with S = 10000
N = 4000000
N = 2000000
N = 1000000
time (ms)
0
5
10
15
20
25
100000
150000
200000
250000
300000
350000
400000
b: number of breaks