Overview

The following sections contain sample inputs and outputs to help check the correctness of your algorithms. We will be grading mostly on your output, and if you are getting the same answers that are presented here for the same inputs, you will most likely be awarded most of the points for the assignment.

In order to retrieve the files referred to in the examples below, do the following:

1. cd into a place where you wouldn’t mind having all of these test files, a good place might be outside of your actual pr1 folder, so that you don’t accidentally overwrite any of your files

2. Run cs182_setup pr1Examples when ready, and it will copy over the pr1Examples folder that contains all the files referenced in the examples below, aside from the .sh files of course.

If you aren’t getting the same outputs as indicated, take a look at both your output and the example output. A lot of the test cases are simplified so you can pinpoint what exactly is going wrong (for example, maybe right extensions are failing, but left extensions are working fine). If you continually notice your output is wrong, come to TA hours or post on Piazza; if you think there is a mistake in the examples, also feel free to post on Piazza.
1 Implementing Local Alignment

These examples make use of DNA sequences. If your code does depend on the input being protein sequences, don’t worry; just mention this in your README, and there are protein examples on the next page. Furthermore, note that multiple local alignments are often possible, but as long as your score matches, your alignment should be fine.

```bash
> sh local.sh localSeq1.txt unitary.m -1
C
C
1
> sh local.sh localSeq1.txt unitary.m negInf
C
C
1
> sh local.sh localSeq2.txt unitary.m -1
ACGT-GTCAACGT
ACGTCGT-AGC-T
5
> sh local.sh localSeq2.txt unitary.m negInf
ACGT
ACGT
4
> sh local.sh localSeq3.txt unitary.m -1
AA-ATTTT
AAGATTTT
6
> sh local.sh localSeq3.txt unitary.m negInf
AAAAATT
AAGATTTT
6
> sh local.sh localSeq4.txt unitary.m -1

0
> sh local.sh localSeq4.txt unitary.m negInf

0
> sh local.sh localSeq5.txt unitary.m -1
<See Below>
> sh local.sh localSeq5.txt unitary.m negInf
CCCCGGGGGCTCGGGC
CCCCGGGGGCCCCCGGG
12
```

From above; this example is just formatted differently because it won’t appear nicely in the Latex environment above.

```
GTTGGCTGTCCCCGGTCTCTGCTCTGTCTCCGCCGGGGGCCTC-GGGC-CGGCTCAGCTTTAAGGAAAGAATCCGAGG
GGGGCCGTGAGG-GCCACAGGCAAGGAGGCTGGAGT--GGGTGGCTCCTCCGGGGCTCTCTCCCCGGGGGGCTC-G
GGG-GGGCGGGGGCGGCGGGGGG
```

21
What follows are a few protein examples of local alignment:

```
> sh local.sh localSeq6.txt pam250.m -1
K-HAWQEIGKEF--S-HEIS-H-H--GR-HSVRD----HCMNS--LEYIAIGDEH
KTHA-MSV---FWMTCEAAAQLRQAGKV--TVRETTLKR-LGATHLRY-GVAD-GH
71.0
> sh local.sh localSeq6.txt pam250.m negInf
EFSHEISHGHRHSVRDHCMNSL
EAAAQLRQAGKVTVRETTLKR
32
> sh local.sh localSeq7.txt pam250.m -1
DEP
DEP
14
> sh local.sh localSeq8.txt pam250.m -1
DEPP
DEPA
15
> sh local.sh localSeq9.txt pam250.m -1
0
> sh local.sh localSeq10.txt pam250.m -1
AW-C
AWTC
30.0
> sh local.sh localSeq10.txt pam250.m negInf
AW
AW
19
```
2 Seeding

What follows are a few examples of us running the seeding script from Problem 2. Rather than showing the terminal output directly here, it is stored in the indicated files. Also note that the hit positions can be in a different order in your output, but this is not a problem. Just be sure that you are indeed finding all of the seeds (a good way to initially check for this might be to see if the total number of seeds found is the same).

<table>
<thead>
<tr>
<th>Terminal Input</th>
<th>Output File</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; sh seed.sh db1.txt query1.txt pam250.m 1 11</td>
<td>seedOutput1.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db1.txt query1.txt pam250.m 1 -3</td>
<td>seedOutput2.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db1.txt query2.txt pam250.m 1 11</td>
<td>seedOutput3.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db1.txt query2.txt pam250.m 1 -5</td>
<td>seedOutput4.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db1.txt query2.txt pam250.m 2 10</td>
<td>seedOutput5.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db1.txt query2.txt pam250.m 2 7</td>
<td>seedOutput6.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db1.txt query2.txt pam250.m 2 6</td>
<td>seedOutput7.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db1.txt query2.txt pam250.m 2 -6</td>
<td>seedOutput8.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db2.txt query2.txt pam250.m 2 -11</td>
<td>seedOutput9.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db2.txt query3.txt pam250.m 4 5</td>
<td>seedOutput10.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db2.txt query3.txt pam250.m 2 2</td>
<td>seedOutput11.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db2.txt query4.txt pam250.m 2 0</td>
<td>seedOutput12.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db3.txt query5.txt pam250.m 3 10</td>
<td>seedOutput13.txt</td>
</tr>
<tr>
<td>&gt; sh seed.sh db3.txt query5.txt pam250.m 4 18</td>
<td>seedOutput14.txt</td>
</tr>
</tbody>
</table>

3 Extension

What follows are examples of what your extension phase should output. Again, the order of the extensions doesn’t matter, but you should be yielding the appropriate alignments (with correct lengths), the correct scores, and also all of the HSPs that are in the examples (no more, no less).

<table>
<thead>
<tr>
<th>Terminal Input</th>
<th>Output File</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; sh extension.sh db3.txt pam250.m extInput1.txt 1 0</td>
<td>extOutput1.txt</td>
</tr>
<tr>
<td>&gt; sh extension.sh db3.txt pam250.m extInput2.txt 1 0</td>
<td>extOutput2.txt</td>
</tr>
<tr>
<td>&gt; sh extension.sh db4.txt pam250.m extInput3.txt 1 15</td>
<td>extOutput3.txt</td>
</tr>
<tr>
<td>&gt; sh extension.sh db4.txt pam250.m extInput3.txt 1 19</td>
<td>extOutput4.txt</td>
</tr>
<tr>
<td>&gt; sh extension.sh db5.txt pam250.m extInput4.txt 1 15</td>
<td>extOutput5.txt</td>
</tr>
<tr>
<td>&gt; sh extension.sh db5.txt pam250.m extInput4.txt 1 0</td>
<td>extOutput6.txt</td>
</tr>
<tr>
<td>&gt; sh extension.sh db5.txt pam250.m extInput4.txt 5 0</td>
<td>extOutput7.txt</td>
</tr>
<tr>
<td>&gt; sh extension.sh db5.txt pam250.m extInput4.txt 10 0</td>
<td>extOutput8.txt</td>
</tr>
<tr>
<td>&gt; sh extension.sh db5.txt pam250.m extInput4.txt 10 20</td>
<td>extOutput9.txt</td>
</tr>
</tbody>
</table>

4 Statistics

There are no example outputs for this section.