5.

a. Cycle 5 (zero based) of file A is probably the most anomalous because it has a large section of 2's all in a row, and the rest of the file is pretty uniformly in the 20s and 30s. However, cycle 37 of file C also has a large quantity of low scores, with the surrounding reads being much higher. The only reason I think cycle 5 is more anomalous is that the later cycles of file C all have very low scores, so looking beyond the immediate surroundings cycle 37 looks less anomalous.

b. I think that file B is barcoded because for most of the reads there is a collection of all the bases in a pattern at the beginning, all next to each other. Looking through the rest of the sequences, it seems like it's unusual in to encounter them like this, which leads me to believe it may be the barcode.

c. I think that file A is Plasmodium falciparum because in the Wikipedia it mentions that the structure is about 80% AT, and this file is most rich in A and T bases.

6.

a. A length-20 substring that is only present in the Delta variant is “GCTACTCCTTTAGATTTTGT”.

b. A length-20 substring present in both the Wuhan and Alpha variants but not the Delta variant is “TCACCGGTGGAATTGCTATC”.

7.

a. The difference between the two versions is that the extended bad character rule allows for bigger skips by matching the template to the pattern, rather than the pattern to the template. When a mismatch occurs, we look for the alignment where the mismatched character in the template matches the next alignment that puts that character in the pattern in the right place, rather than looking for the place where the mismatched character in the pattern matches the template next.

b. In class, we learned the extended bad character rule.

c.
```
 T: TCATGTCTGACCTAA
 P: TCTTATAT
BC:     TCTTATAT
EBC:        TCTTATAT
```

8. This would not be correct.
T: AAATA
P: AATA
1st skip: AATA

Leftmost search misses the match.

9.

a. The difference between the weak good suffix rule and the strong good suffix rule is that in the weak good suffix rule, we don't check that the character preceding the suffix is different in the next occurrence. In the strong suffix rule, we make sure that the preceding character is not the same as the one that just caused a mismatch.

b. This example in the book shows how the two rules differ.

```
    T: PRSTABSTUBABVQXRST
    P: QCABDABDAB
WGS:     QCABDABDAB
SGS:           QCABDABDAB
```

c. Galil is the person who suggested the approach that gives linear time for all cases.