**Peer assessment practice exam**

My name is: \_Huizhi Lin\_\_\_\_

I am grading the exam of: \_Yan Wang\_\_\_\_\_\_\_\_\_

|  |  |  |
| --- | --- | --- |
| Parsing fasta | 1 | **0.5** |
| Parsing needle | 2 | **2** |
| Run needle | 2 | **2** |
| Calculations | 2 | **2** |
| Output | 1 | **0.5** |
| Main code | 1 | **1** |
| Top code! | 1 | **0.5** |
| Style | 1 | **1** |
|  | 11 | **9.5** |

|  |
| --- |
| **Motivation/feedback functionality** |
| Most function runs successfully!  However, when parsing the fasta file, there is a sequences with the different label structure, which you extract the label as “sp|P26981|GPA1\_SOLLC”. But in the needle output file, the label of this sequence is “GPA1\_SOLLC”. This create the error “KeyError: ('GPA1\_ARATH', 'sp|P26981|GPA1\_SOLLC')” when you later try to calculate the hamm and ident. This also resulted in a not fully corrected output result file. |
| **Motivation/feedback style** |
| Appropriate docstrings, input/output specified!  Clear separation of logic.  Code mostly readable. Variable names largely informative . |
| **Tops: what did you like about the implementation?** |
| It used dictionary in a lot of cases.  It is very readable and clear.  No extract useless code. |
| **Tips: how can this person improve his/her coding skills?** |
| Learn more about print format:  Line108: if you use .format method, then use {} instead of %s in the strings. You can see that in the final result output file, it create a wrong header:  ‘%10s %6d %10s %6d %4d %-5d’.  In the function ‘parse\_needle(needle)’, line85:  You reassigned the tuple (sequence1,sequence2) at each end of sequence2’s line to the dictionary aligned\_pair. If the file is much bigger, this repeat process may slow down the whole script. |

When finished, submit your feedback form on BlackBoard under “P5 assessment” and email it to the author of the script.