2021-22

BIO F242
Introduction to Bioinformatics

Experiment - 2

Name : Suchismita Tripathy
ID : 2019A7PS0554P

Gene Name : HUS1 Checkpoint Clamp Component (HUS1)
Organism : Theropithecus gelada
Accession Number : XM_025380797.1

7 February, 2022

1 Aim

Write a Python script to find:

1. Number of base pairs (i.e A, T, G, C)
2. Number start, stop codons
3. The GC Content
4. mRNA Transcript
5. Reverse Complement
6. Total EcoRI, BamHI, HindIII sites

in Theropithecus gelada HUS1 Checkpoint Clamp Component (HUS1) transcript variant X1.

2 Materials Required

Gene sequence (HUS1 Checkpoint Clamp Component (HUS1) transcript variant X1), Jupyter Notebook

3 Commands

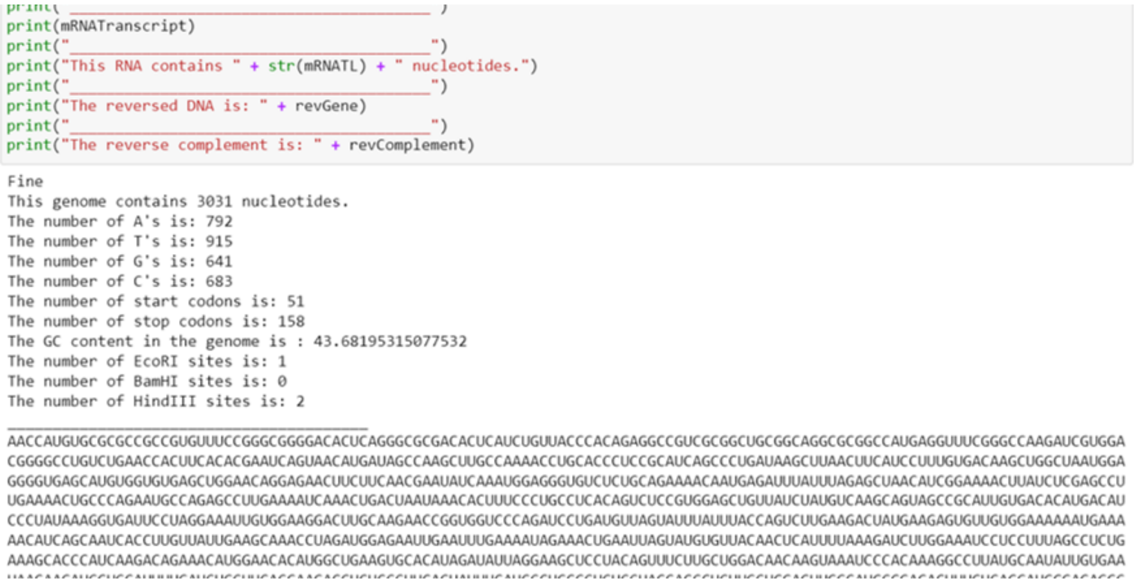
The code was written and executed on Jupyter Notebook for which print commands are not generally required, but the code has been adapted and print statements have been added here :

```
fileread = open("hus1.fasta")
#FASTA format file for HUS1
#Checkpoint Clamp Component (HUS1) transcript variant X1
#as downloaded from NCBI

rem = fileread.readline()

gene = fileread.read()

gene = gene.replace("\n", "")

totalL = len(gene)
```

```
numA = gene.count('A')

numT = gene.count('T')

numG = gene.count('G')

numC = gene.count('C')

basesSum = gene.count('A') + gene.count('T') +
    gene.count('G') + gene.count('C')
if (basesSum == totalL):
    print('Fine')
#Sanity check

#Start Codon: AUG
#Stop Codons: UAG, UGA, UAA

numStart = gene.count("ATG") #Number of start codons

numStop = gene.count("TAG") + gene.count("TGA") +
    gene.count("TAA")
#Number of stop codons

GCCont = ((gene.count('G') + gene.count('C')) * 100) /
    totalL

#EcoRI Site: GAATTC
#BamHI Site: GGATCC
#HindIII Site: AAGCTT

numE = gene.count('GAATTC') #Number of EcoRI sites

numB = gene.count('GGATCC') #Number of BamHI sites

numH = gene.count('AAGCTT') #Number of HindIII sites

mRNATranscript = gene.replace('T', 'U')
mRNATL = len(mRNATranscript)

revGene = gene[::-1]
```

```
revComplement = revGene.replace('A', 't')
revComplement = revComplement.replace('T', 'a')
revComplement = revComplement.replace('G', 'c')
revComplement = revComplement.replace('C', 'g')
revComplement = revComplement.upper()

#Print statements
# print(gene)
# print("----------------------------------------")
print ("This genome contains " + str(totalL) + " nucleotides.")
print("The number of A's is: " + str(numA))
print("The number of T's is: " + str(numT))
print("The number of G's is: " + str(numG))
print("The number of C's is: " + str(numC))
print("The number of start codons is: " + str(numStart))
print("The number of stop codons is: " + str(numStop))
print("The GC content in the genome is : " + str(GCCont))
print("The number of EcoRI sites is: " + str(numE))
print("The number of BamHI sites is: " + str(numB))
print("The number of HindIII sites is: " + str(numH))
print("----------------------------------------")
print(mRNATranscript)
print("----------------------------------------")
print("This RNA contains " + str(mRNATL) + " nucleotides.")
print("----------------------------------------")
print("The reversed DNA is: " + revGene)
print("----------------------------------------")
print("The reverse complement is: " + revComplement)
```

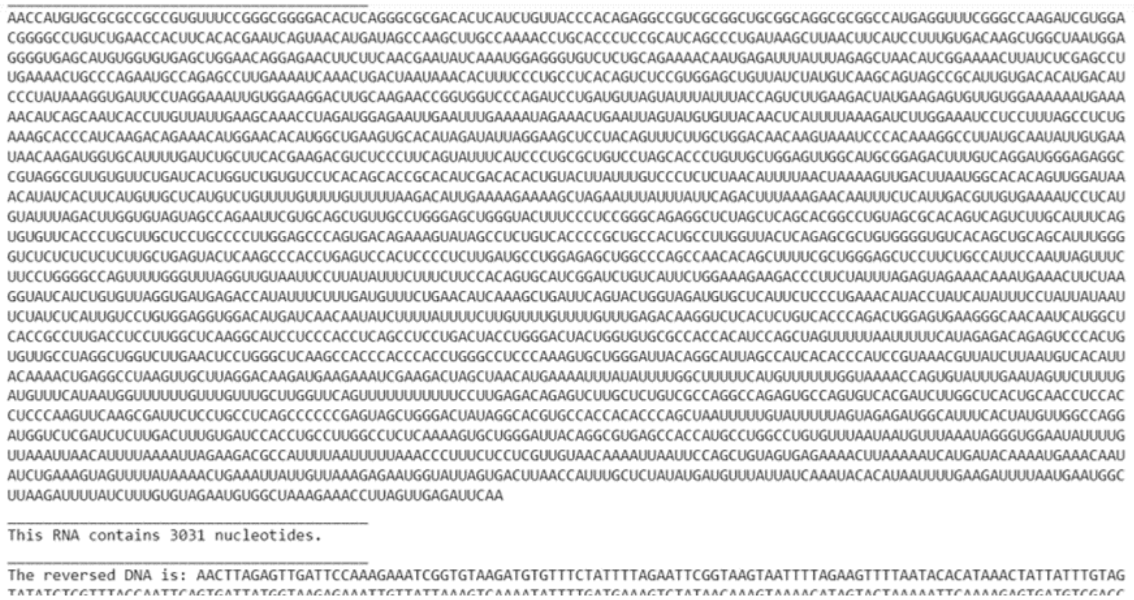
4 Observations



```
print("_____________________________________________")
print(mRNATranscript)
print("_____________________________________________")
print("This RNA contains " + str(mRNATL) + " nucleotides.")
print("_____________________________________________")
print("The reversed DNA is: " + revGene)
print("_____________________________________________")
print("The reverse complement is: " + revComplement)
```

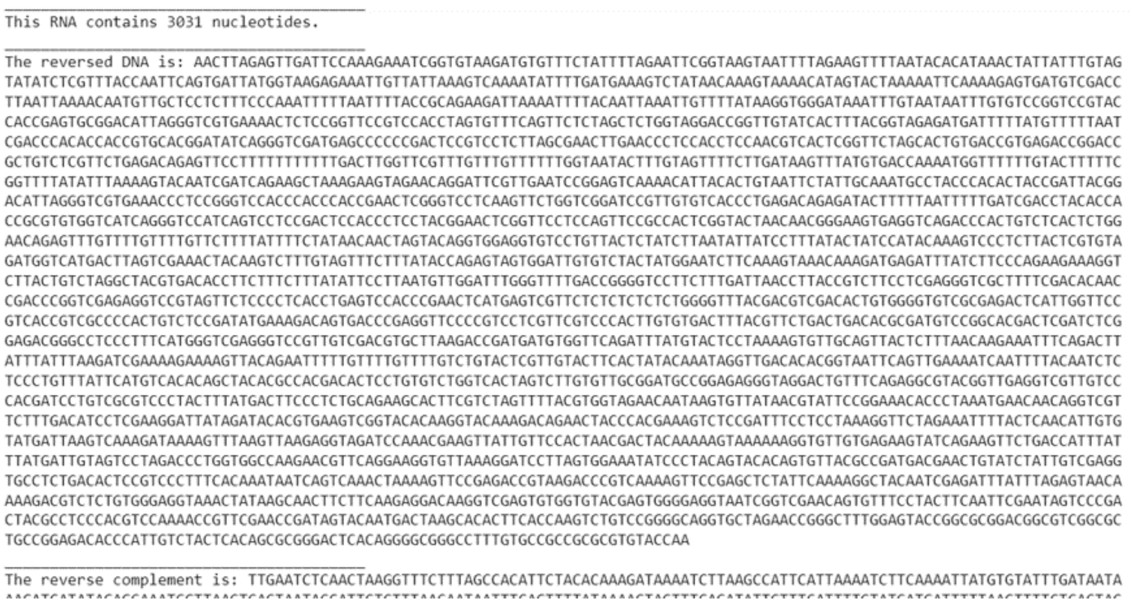
```
Fine
This genome contains 3031 nucleotides.
The number of A's is: 792
The number of T's is: 915
The number of G's is: 641
The number of C's is: 683
The number of start codons is: 51
The number of stop codons is: 158
The GC content in the genome is : 43.68195315077532
The number of EcoRI sites is: 1
The number of BamHI sites is: 0
The number of HindIII sites is: 2

_____________________________________________
AACCAUGUGCGCGCCGCCGUGUUUCCGGGCGGGGACACUCAGGGCGCGACACUCAUCUGUUACCCACAGAGGCCGUCGCGGCUGCGGCAGGCGCGGCCAUGAGGUUUCGGGCCAAGAUCGUGGA
CGGGGCCUGUCUGAACCACUUCACACGAAUCAGUAACAUGAUAGCCAAGCUUGCCAAAACCUGCACCCUCCGCAUCAGCCCUGAUAAGCUUAACUUCAUCCUUUGUGACAAGCUGGCUAAUGGA
GGGGUGAGCAUGUGGUGUGAGCUGGAACAGGAGAACUUCUUCAACGAAUAUCAAAUGGAGGGUGUCUCUGCAGAAAACAAUGAGAUUUAUUUAGAGCUAACAUCGGAAAACUUAUCUCGAGCCU
UGAAAACUGCCCAGAAUGCCAGAGCCUUGAAAAUCAAACUGACUAAUAAACACUUUCCCUGCCUCACAGUCUCCGUGGAGCUGUUAUCUAUGUCAAGCAGUAGCCGCAUUGUGACACAUGACAU
CCCUAUAAAGGUGAUUCCUAGGAAAUUGUGGAAGGACUUGCAAGAACCGGUGGUCCCAGAUCCUGAUGUUAGUAUUUAUUUACCAGUCUUGAAGACUAUGAAGAGUGUUGUGGAAAAAAUGAAA
AACAUCAGCAAUCACCUUGUUAUUGAAGCAAACCUAGAUGGAGAAUUGAAUUGAAAAUAGAAACUGAAUUAGUAUGUGUUACAACUCAUUUUAAAGAUCUUGGAAAUCCUCCUUUAGCCUCUG
AAAGCACCCAUCAAGACAGAAACAUGGAACACAUGGCUGAAGUGCACAUAGAUAUUAGGAAGCUCCUACAGUUUCUUGCUGGACAACAAGUAAAUCCCACAAAGGCCUUAUGCAAUAUUGUGAA
```

Figure 1: Output Screenshot 1



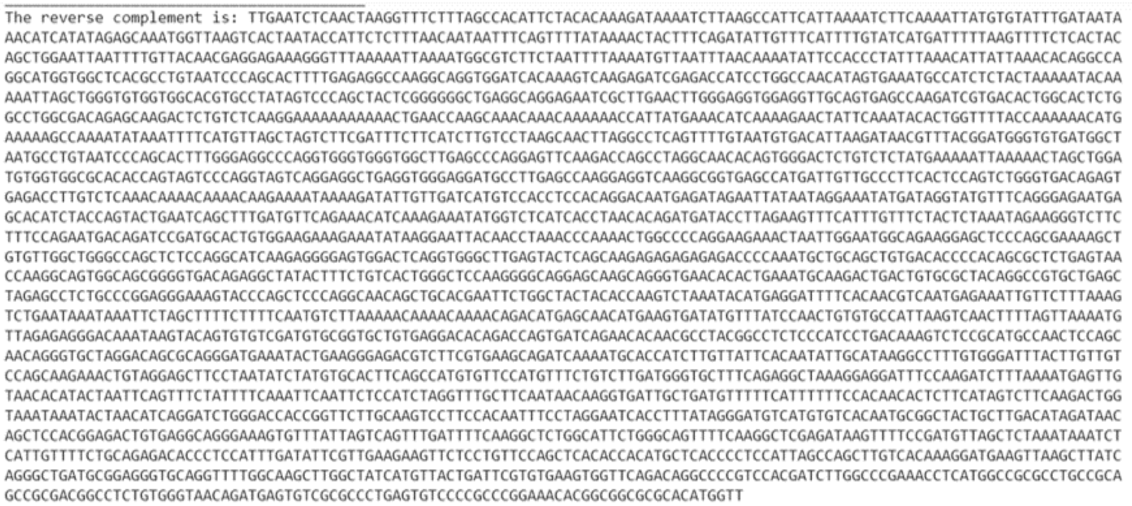
```
_____________________________________________
AACCAUGUGCGCGCCGCCGUGUUUCCGGGCGGGGACACUCAGGGCGCGACACUCAUCUGUUACCCACAGAGGCCGUCGCGGCUGCGGCAGGCGCGGCCAUGAGGUUUCGGGCCAAGAUCGUGGA
CGGGGCCUGUCUGAACCACUUCACACGAAUCAGUAACAUGAUAGCCAAGCUUGCCAAAACCUGCACCCUCCGCAUCAGCCCUGAUAAGCUUAACUUCAUCCUUUGUGACAAGCUGGCUAAUGGA
GGGGUGAGCAUGUGGUGUGAGCUGGAACAGGAGAACUUCUUCAACGAAUAUCAAAUGGAGGGUGUCUCUGCAGAAAACAAUGAGAUUUAUUUAGAGCUAACAUCGGAAAACUUAUCUCGAGCCU
UGAAAACUGCCCAGAAUGCCAGAGCCUUGAAAAUCAAACUGACUAAUAAACACUUUCCCUGCCUCACAGUCUCCGUGGAGCUGUUAUCUAUGUCAAGCAGUAGCCGCAUUGUGACACAUGACAU
CCCUAUAAAGGUGAUUCCUAGGAAAUUGUGGAAGGACUUGCAAGAACCGGUGGUCCCAGAUCCUGAUGUUAGUAUUUAUUUACCAGUCUUGAAGACUAUGAAGAGUGUUGUGGAAAAAAUGAAA
AACAUCAGCAAUCACCUUGUUAUUGAAGCAAACCUAGAUGGAGAAUUGAAUUGAAAAUAGAAACUGAAUUAGUAUGUGUUACAACUCAUUUUAAAGAUCUUGGAAAUCCUCCUUUAGCCUCUG
AAAGCACCCAUCAAGACAGAAACAUGGAACACAUGGCUGAAGUGCACAUAGAUAUUAGGAAGCUCCUACAGUUUCUUGCUGGACAACAAGUAAAUCCCACAAAGGCCUUAUGCAAUAUUGUGAA
UAACAAGAUGGUGCAUUUUGAUCUGCUUCACGAAGACGUCUCCCUUCAGUAUUUCAUCCCUGCGCUGUCCUAGCACCCUGUUGCUGGAGUUGGCAUGCGGAGACUUUGUCAGGAUGGGAGAGGC
CGUAGGCGUUGUGUUCUGAUCACUGGUCUGUGUCCUCACAGCACCGCACAUCGACACACUGUACUUAUUUGUCCCUCUCUAACAUUUUAACUAAAAGUUGACUUAAUGGCACACAGUUGGAUAA
ACAUAUCACUUCAUGUUGCUCAUGUCUGUUUUGUUUUGUUUUUAAGACAUUGAAAAGAAAAGCUAGAAUUUAUUUAUUCAGACUUUAAAGAACAAUUUCUCAUUGACGUUGUGAAAAUCCUCAU
GUAUUUAGACUUGGUGUAGUAGCCAGAAUUCGUGCAGCUGUUGCCUGGGAGCUGGGUACUUUCCCUCCGGGCAGAGGCUCUAGCUCAGCACGGCCUGUAGCGCACAGUCAGUCUUGCAUUUCAG
UGUGUUCACCCUGCUUGCUCCUGCCCCUUGGAGCCCAGUGACAGAAAGUAUAGCCUCUGUCACCCCGCUGCCACUGCCUUGGUUACUCAGAGCGCUGUGGGGUGUCACAGCUGCAGCAUUUGGG
GUCUCUCUCUCUCUUGCUGAGUACUCAAGCCCACCUGAGUCCACUCCCCUCUUGAUGCCUGGAGAGCUGGCCCAGCCAACACAGCUUUUCGCUGGGAGCUCCUUCUGCCAUUCCAAUUAGUUUC
UUCCUGGGGCCAGUUUUGGGUUUAGGUUGUAAUUCCUUAUAUUUCUUUCUUCCACAGUGCAUCGGAUCUGUCAUUCUGGAAAGAAGACCCUUCUAUUUAGAGUAGAAACAAAUGAAACUUCUAA
GGUAUCAUCUGUGUUAGGUGAUGAGACCAUAUUUCUUUGAUGUUUCUGAACAUCAAAGCUGAUUCAGUACUGGUAGAUGUGCUCAUUCUCCCUGAAACAUACCUAUCAUAUUUCCUAUUAUAAU
UCUAUCUCAUUGUCCUGUGGAGGUGGACAUGAUCAACAAUAUCUUUUAUUUUCUUGUUUUGUUUUGUUUGAGACAAGGUCUCACUCUGUCACCCAGACUGGAGUGAAGGGCAACAAUCAUGGCU
CACCGCCUUGACCUCCUUGGCUCAAGGCAUCCUCCCACCUCAGCCUCCUGACUACCUGGGACUACUGGUGUGCGCCACCACAUCCAGCUAGUUUUUAAUUUUUCAUAGAGACAGAGUCCCACUG
UGUUGCCUAGGCUGGUCUUGAACUCCUGGGCUCAAGCCACCCACCCACCUGGGCCUCCCAAAGUGCUGGGAUUACAGGCAUUAGCCAUCACACCCAUCCGUAAACGUUAUCUUAAUGUCACAUU
ACAAAACUGAGGCCUAAGUUGCUUAGGACAAGAUGAAGAAAUCGAAGACUAGCUAACAUGAAAAUUUAUAUUUUGGCUUUUUCAUGUUUUUUGGUAAAACCAGUGUAUUUGAAUAGUUCUUUUG
AUGUUUCAUAAUGGUUUUUUGUUUGUUUGCUUGGUUCAGUUUUUUUUUUUCCUUGAGACAGAGUCUUGCUCUGUCGCCAGGCCAGAGUGCCAGUGUCACGAUCUUGGCUCACUGCAACCUCCAC
CUCCCAAGUUCAAGCGAUUCUCCUGCCUCAGCCCCCCGAGUAGCUGGGACUAUAGGCACGUGCCACCACACCCAGCUAAUUUUUGUAUUUUUAGUAGAGAUGGCAUUUCACUAUGUUGGCCAGG
AUGGUCUCGAUCUCUUGACUUUGUGAUCCACCUGCCUUGGCCUCUCAAAAGUGCUGGGAUUACAGGCGUGAGCCACCAUGCCUGGCCUGUGUUUAAUAAUGUUUAAAUAGGGUGGAAUAUUUUG
UUAAAUUAACAUUUUAAAAUUAGAAGACGCCAUUUUAAUUUUUAAACCCUUUCUCCUCGUUGUAACAAAAUUAAUUCCAGCUGUAGUGAGAAAACUUAAAAAUCAUGAUACAAAAUGAAACAAU
AUCUGAAAGUAGUUUUAUAAAACUGAAAUUAUUGUUAAAGAGAAUGGUAUUAGUGACUUAACCAUUUGCUCUAUAUGAUGUUUAUUAUCAAAUACACAUAAUUUUGAAGAUUUUAAUGAAUGGC
UUAAGAUUUUAUCUUUGUGUAGAAUGUGGCUAAAGAAACCUUAGUUGAGAUUCAA

_____________________________________________
This RNA contains 3031 nucleotides.

_____________________________________________
The reversed DNA is: AACTTAGAGTTGATTCCAAAGAAATCGGTGTAAGATGTGTTTCTATTTTAGAATTCGGTAAGTAATTTTAGAAGTTTTAATACACATAAACTATTATTTGTAG
TATATCTCGTTTACCAATTCAGTGATTATGGTAAGAGAAATTGTTATTAAAGTCAAAATATTTTGATGAAAGTCTATAACAAAGTAAAACATAGTACTAAAAATTCAAAAGAGTGATGTCGACC
```

Figure 2: Output Screenshot 2



```
________________________________________
This RNA contains 3031 nucleotides.

________________________________________
The reversed DNA is: AACTTAGAGTTGATTCCAAAGAAATCGGTGTAAGATGTGTTTCTATTTTAGAATTCGGTAAGTAATTTTAGAAGTTTTAATACACATAAACTATTATTTGTAG
TATATCTCGTTTACCAATTCAGTGATTATGGTAAGAGAAATTGTTATTAAAGTCAAAATATTTTGATGAAAGTCTATAACAAAGTAAAACATAGTACTAAAAATTCAAAAGAGTGATGTCGACC
TTAATTAAAACAATGTTGCTCCTCTTTCCCAAATTTTTAATTTTACCGCAGAAGATTAAAATTTTACAATTAAATTGTTTTATAAGGTGGGATAAATTTGTAATAATTTGTGTCCGGTCCGTAC
CACCGAGTGCGGACATTAGGGTCGTGAAAACTCTCCGGTTCCGTCCACCTAGTGTTTCAGTTCTCTAGCTCTGGTAGGACCGGTTGTATCACTTTACGGTAGAGATGATTTTTATGTTTTTAAT
CGACCCACACCACCGTGCACGGATATCAGGGTCGATGAGCCCCCGACTCCGTCCTCTTAGCGAACTTGAACCCTCCACCTCCAACGTCACTCGGTTCTAGCACTGTGACCGTGAGACCGGACC
GCTGTCTCGTTCTGAGACAGAGTTCCTTTTTTTTTTTGACTTGGTTCGTTTGTTTGTTTTTTGGTAATACTTTGTAGTTTTCTTGATAAGTTTATGTGACCAAAATGGTTTTTTGTACTTTTTC
GGTTTTATATTTAAAAGTACAATCGATCAGAAGCTAAAGAAGTAGAACAGGATTCGTTGAATCCGGAGTCAAAACATTACACTGTAATTCTATTGCAAATGCCTACCCACACTACCGATTACGG
ACATTAGGGTCGTGAAACCCTCCGGGTCCACCCACCCACCGAACTCGGGTCCTCAAGTTCTGGTCGGATCCGTTGTGTCACCCTGAGACAGAGATACTTTTTAATTTTTGATCGACCTACACCA
CCGCGTGTGGTCATCAGGGTCCATCAGTCCTCCGACTCCACCCTCCTACGGAACTCGGTTCCTCCAGTTCCGCCACTCGGTACTAACAACGGGAAGTGAGGTCAGACCCACTGTCTCACTCTGG
AACAGAGTTTGTTTTGTTTTGTTCTTTTATTTTCTATAACAACTAGTACAGGTGGAGGTGTCCTGTTACTCTATCTTAATATTATCCTTTATACTATCCATACAAAGTCCCTCTTACTCGTGTA
GATGGTCATGACTTAGTCGAAACTACAAGTCTTTGTAGTTTCTTTATACCAGAGTAGTGGATTGTGTCTACTATGGAATCTTCAAAGTAAACAAAGATGAGATTTATCTTCCCAGAAGAAAGGT
CTTACTGTCTAGGCTACGTGACACCTTCTTTCTTTATATTCCTTAATGTTGGATTTGGGTTTTGACCGGGGTCCTTCTTTGATTAACCTTACCGTCTTCCTCGAGGGTCGCTTTTCGACACAAC
CGACCCGGTCGAGAGGTCCGTAGTTCTCCCCTCACCTGAGTCCACCCGAACTCATGAGTCGTTCTCTCTCTCTCTGGGGTTTACGACGTCGACACTGTGGGGTGTCGCGAGACTCATTGGTTCC
GTCACCGTCGCCCCACTGTCTCCGATATGAAAGACAGTGACCCGAGGTTCCCCGTCCTCGTTCGTCCCACTTGTGTGACTTTACGTTCTGACTGACACGCGATGTCCGGCACGACTCGATCTCG
GAGACGGGCCTCCCTTTCATGGGTCGAGGGTCCGTTGTCGACGTGCTTAAGACCGATGATGTGGTTCAGATTTATGTACTCCTAAAAGTGTTGCAGTTACTCTTTAACAAGAAATTTCAGACTT
ATTTATTTAAGATCGAAAAGAAAAGTTACAGAATTTTTGTTTTGTTTTGTCTGTACTCGTTGTACTTCACTATACAAATAGGTTGACACACGGTAATTCAGTTGAAAATCAATTTTACAATCTC
TCCCTGTTTATTCATGTCACACAGCTACACGCCACGACACTCCTGTGTCTGGTCACTAGTCTTGTGTTGCGGATGCCGGAGAGGGTAGGACTGTTTCAGAGGCGTACGGTTGAGGTCGTTGTCC
CACGATCCTGTCGCGTCCCTACTTTATGACTTCCCTCTGCAGAAGCACTTCGTCTAGTTTTACGTGGTAGAACAATAAGTGTTATAACGTATTCCGGAAACACCCTAAATGAACAACAGGTCGT
TCTTTGACATCCTCGAAGGATTATAGATACACGTGAAGTCGGTACACAAGGTACAAAGACAGAACTACCCACGAAAGTCTCCGATTTCCTCCTAAAGGTTCTAGAAATTTTACTCAACATTGTG
TATGATTAAGTCAAAGATAAAAGTTTAAGTTAAGAGGTAGATCCAAACGAAGTTATTGTTCCACTAACGACTACAAAAAGTAAAAAAGGTGTTGTGAGAAGTATCAGAAGTTCTGACCATTTAT
TTATGATTGTAGTCCTAGACCCTGGTGGCCAAGAACGTTCAGGAAGGTGTTAAAGGATCCTTAGTGGAAATATCCCTACAGTACACAGTGTTACGCCGATGACGAACTGTATCTATTGTCGAGG
TGCCTCTGACACTCCGTCCCTTTCACAAATAATCAGTCAAACTAAAAGTTCCGAGACCGTAAGACCCGTCAAAAGTTCCGAGCTCTATTCAAAAGGCTACAATCGAGATTTATTTAGAGTAACA
AAAGACGTCTCTGTGGGAGGTAAACTATAAGCAACTTCTTCAAGAGGACAAGGTCGAGTGTGGTGTACGAGTGGGGAGGTAATCGGTCGAACAGTGTTTCCTACTTCAATTCGAATAGTCCCGA
CTACGCCTCCCACGTCCAAAACCGTTCGAACCGATAGTACAATGACTAAGCACACTTCACCAAGTCTGTCCGGGGCAGGTGCTAGAACCGGGCTTTGGAGTACCGGCGCGGACGGCGTCGGCGC
TGCCGGAGACACCCATTGTCTACTCACAGCGCGGGACTCACAGGGGCGGGCCTTTGTGCCGCCGCGCGTGTACCAA

________________________________________
The reverse complement is: TTGAATCTCAACTAAGGTTTCTTTAGCCACATTCTACACAAAGATAAAATCTTAAGCCATTCATTAAAATCTTCAAAATTATGTGTATTTGATAATA
```

Figure 3: Output Screenshot 3



```
________________________________________
The reverse complement is: TTGAATCTCAACTAAGGTTTCTTTAGCCACATTCTACACAAAGATAAAATCTTAAGCCATTCATTAAAATCTTCAAAATTATGTGTATTTGATAATA
AACATCATATAGAGCAAATGGTTAAGTCACTAATACCATTCTCTTTAACAATAATTTCAGTTTTATAAAACTACTTTCAGATATTGTTTCATTTTGTATCATGATTTTTAAGTTTTCTCACTAC
AGCTGGAATTAATTTTGTTACAACGAGGAGAAAGGGTTTAAAAATTAAAATGGCGTCTTCTAATTTTAAAATGTTAATTTAACAAAATATTCCACCCTATTTAAACATTATTAAACACAGGCCA
GGCATGGTGGCTCACGCCTGTAATCCCAGCACTTTTGAGAGGCCAAGGCAGGTGGATCACAAAGTCAAGAGATCGAGACCATCCTGGCCAACATAGTGAAATGCCATCTCTACTAAAAATACAA
AAATTAGCTGGGTGTGGTGGCACGTGCCTATAGTCCCAGCTACTCGGGGGGCTGAGGCAGGAGAATCGCTTGAACTTGGGAGGTGGAGGTTGCAGTGAGCCAAGATCGTGACACTGGCACTCTG
GCCTGGCGACAGAGCAAGACTCTGTCTCAAGGAAAAAAAAAAACTGAACCAAGCAAACAAACAAAAAACCATTATGAAACATCAAAAGAACTATTCAAATACACTGGTTTTACCAAAAAACATG
AAAAAGCCAAAATATAAATTTTCATGTTAGCTAGTCTTCGATTTCTTCATCTTGTCCTAAGCAACTTAGGCCTCAGTTTTGTAATGTGACATTAAGATAACGTTTACGGATGGGTGTGATGGCT
AATGCCTGTAATCCCAGCACTTTGGGAGGCCCAGGTGGGTGGGTGGCTTGAGCCCAGGAGTTCAAGACCAGCCTAGGCAACACAGTGGGACTCTGTCTCTATGAAAAATTAAAAACTAGCTGGA
TGTGGTGGCGCACACCAGTAGTCCCAGGTAGTCAGGAGGCTGAGGTGGGAGGATGCCTTGAGCCAAGGAGGTCAAGGCGGTGAGCCATGATTGTTGCCCTTCACTCCAGTCTGGGTGACAGAGT
GAGACCTTGTCTCAAACAAAACAAAACAAGAAAATAAAAGATATTGTTGATCATGTCCACCTCCACAGGACAATGAGATAGAATTATAATAGGAAATATGATAGGTATGTTTCAGGGAGAATGA
GCACATCTACCAGTACTGAATCAGCTTTGATGTTCAGAAACATCAAAGAAATATGGTCTCATCACCTAACACAGATGATACCTTAGAAGTTTCATTTGTTTCTACTCTAAATAGAAGGGTCTTC
TTTCCAGAATGACAGATCCGATGCACTGTGGAAGAAAGAAATATAAGGAATTACAACCTAAACCCAAAACTGGCCCCAGGAAGAAACTAATTGGAATGGCAGAAGGAGCTCCCAGCGAAAAGCT
GTGTTGGCTGGGCCAGCTCTCCAGGCATCAAGAGGGGAGTGGACTCAGGTGGGCTTGAGTACTCAGCAAGAGAGAGAGAGAGACCCCAAATGCTGCAGCTGTGACACCCCACAGCGCTCTGAGTAA
CCAAGGCAGTGGCAGCGGGGTGACAGAGGCTATACTTTCTGTCACTGGGCTCCAAGGGGCAGGAGCAAGCAGGGTGAACACACTGAAATGCAAGACTGACTGTGCGCTACAGGCCGTGCTGAGC
TAGAGCCTCTGCCCGGAGGGAAAGTACCCAGCTCCCAGGCAACAGCTGCACGAATTCTGGCTACTACACCAAGTCTAAATACATGAGGATTTTCACAACGTCAATGAGAAATTGTTCTTTAAAG
TCTGAATAAATAAATTCTAGCTTTTCTTTTCAATGTCTTAAAAACAAAACAAAACAGACATGAGCAACATGAAGTGATATGTTTATCCAACTGTGTGCCATTAAGTCAACTTTTAGTTAAAATG
TTAGAGAGGGACAAATAAGTACAGTGTGTCGATGTGCGGTGCTGTGAGGACACAGACCAGTGATCAGAACACAACGCCTACGGCCTCTCCCATCCTGACAAAGTCTCCGCATGCCAACTCCAGC
AACAGGGTGCTAGGACAGCGCAGGGATGAAATACTGAAGGGAGACGTCTTCGTGAAGCAGATCAAAATGCACCATCTTGTTATTCACAATATTGCATAAGGCCTTTGTGGGATTTACTTGTTGT
CCAGCAAGAAACTGTAGGAGCTTCCTAATATCTATGTGCACTTCAGCCATGTGTTCCATGTTTCTGTCTTGATGGGTGCTTTCAGAGGCTAAAGGAGGATTTCCAAGATCTTTAAAATGAGTTG
TAACACATACTAATTCAGTTTCTATTTTCAAATTCAATTCTCCATCTAGGTTTGCTTCAATAACAAGGTGATTGCTGATGTTTTTCATTTTTTCCACAACACTCTTCATAGTCTTCAAGACTGG
TAAATAAATACTAACATCAGGATCTGGGACCACCGGTTCTTGCAAGTCCTTCCACAATTTCCTAGGAATCACCTTTATAGGGATGTCATGTGTCACAATGCGGCTACTGCTTGACATAGATAAC
AGCTCCACGGAGACTGTGAGGCAGGGAAAGTGTTTATTAGTCAGTTTGATTTTCAAGGCTCTGGCATTCTGGGCAGTTTTCAAGGCTCGAGATAAGTTTTCCGATGTTAGCTCTAAATAAATCT
CATTGTTTTCTGCAGAGACACCCTCCATTTGATATTCGTTGAAGAAGTTCTCCTGTTCCAGCTCACACCACATGCTCACCCCTCCATTAGCCAGCTTGTCACAAAGGATGAAGTTAAGCTTATC
AGGGCTGATGCGGAGGGTGCAGGTTTTGGCAAGCTTGGCTATCATGTTACTGATTCGTGTGAAGTGGTTCAGACAGGCCCCGTCCACGATCTTGGCCCGAAACCTCATGGCCGCGCCTGCCGCA
GCCGCGACGGCCTCTGTGGGTAACAGATGAGTGTCGCGCCCTGAGTGTCCCCGCCCGGAAACACGGCGGCGCGCACATGGTT
```

Figure 4: Output Screenshot 4

5 Inferences

1. The given sequence of Theropithecus gelada HUS1 Checkpoint Clamp Component (HUS1) transcript variant X1 consists of:
   A: 792
   T: 915
   G: 641
   C: 683

2. ATG (Start Codons): 51

3. TAG + TGA + TAA (Stop Codons): 158

4. GC %: 43.68195315077532

5. Length of mRNA transcript: 3031 nt

6. Number of EcoRI sites: 1

7. Number of BamHI sites: 0

8. Number of HindIII sites: 2