

Intelligence of Biological Systems 3 - 19BIO201

Aditya Anil, AM.EN.U4AIE19006

Assignment 3

Question 1

1. Given an arbitrary collection of k-mers Patterns, we form a graph having a node for each k-mer in Patterns and connect k-mers Pattern and Pattern' by a directed edge if $\text{Suffix}(\text{Pattern})$ is equal to $\text{Prefix}(\text{Pattern}')$. The resulting graph is called the overlap graph on these k-mers, denoted $\text{Overlap}(\text{Patterns})$.

We use the terms prefix and suffix to refer to the first $k - 1$ nucleotides and last $k - 1$ nucleotides of a k-mer, respectively.

Overlap Graph Problem

Construct the overlap graph of a collection of k-mers.

Given: A collection Patterns of k-mers.

Return: The overlap graph $\text{Overlap}(\text{Patterns})$, in the form of an adjacency list.

Sample Dataset

ATGCG
GCATG
CATGC
AGGCA
GGCAT

Sample Output

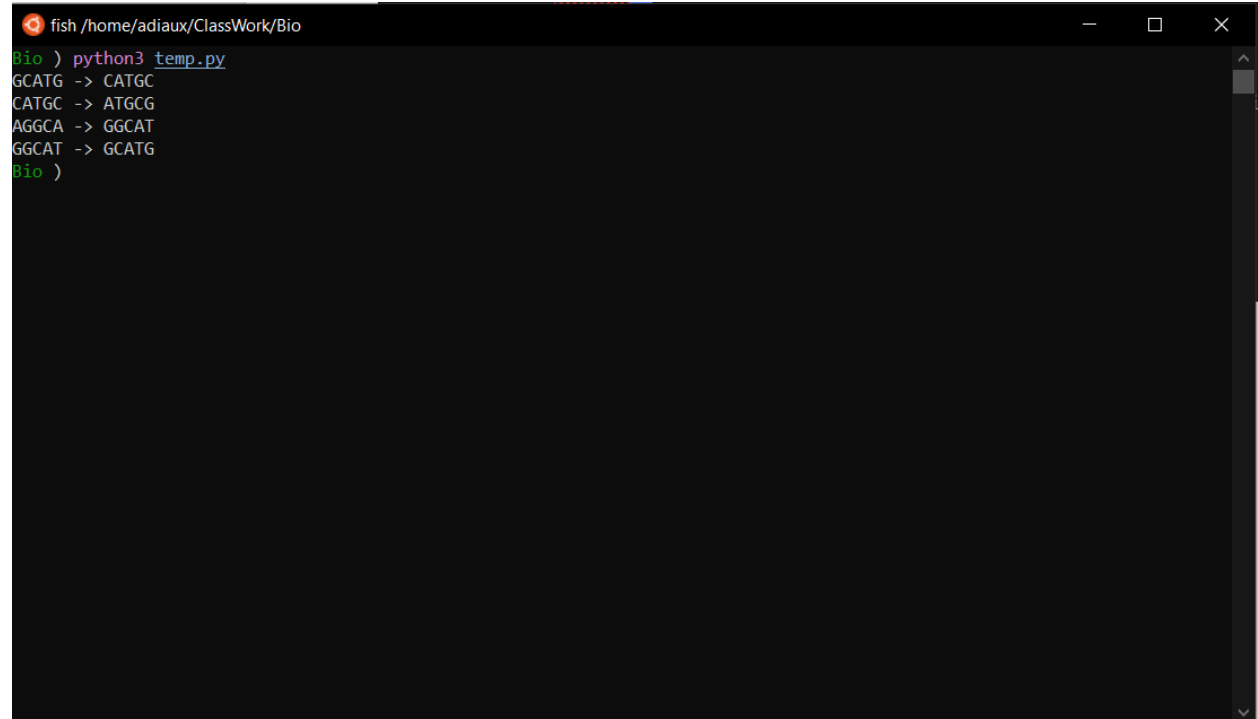
AGGCA -> GGCAT
CATGC -> ATGCG
GCATG -> CATGC
GGCAT -> GCATG

Answer

Code

```
def q3():
    outstr=[]
    linList=open('test.txt', 'r').readlines()
    listFinal=[]
    for line in linList :
        listFinal.append(line.rstrip())
    for line in listFinal:
        for testline in listFinal:
            if (line!=testline and len(line)==len(testline)) and line[1:]==testline[:len(line)-1]:
                str=line+" -> "+testline
                outstr.append(str)
    return outstr
list=q3()
for item in list:
    print(item)
```

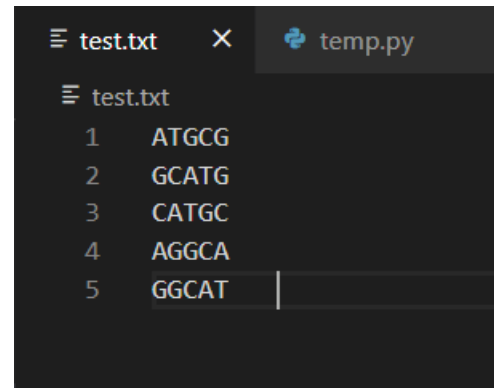
Output:

A terminal window titled 'fish /home/adiaux/ClassWork/Bio' showing the execution of a Python script. The prompt is 'Bio)' and the command 'python3 temp.py' has been entered. The output consists of four lines of DNA sequence pairs connected by arrows: 'GCATG -> CATGC', 'CATGC -> ATGCG', 'AGGCA -> GGCAT', and 'GGCAT -> GCATG'. The prompt 'Bio)' is shown again at the bottom.

```
fish /home/adiaux/ClassWork/Bio
Bio ) python3 temp.py
GCATG -> CATGC
CATGC -> ATGCG
AGGCA -> GGCAT
GGCAT -> GCATG
Bio )
```

Dataset Used:

ATGCG
GCATG
CATGC
AGGCA
GGCAT



```
test.txt  X  temp.py
test.txt
1  ATGCG
2  GCATG
3  CATGC
4  AGGCA
5  GGCAT
```

Question 2

Write short notes on ISOCIAN game. Write about its history, objectives and its relation to isocian calculus.

Answer

The icosian game is a mathematical game invented in 1857 by William Rowan Hamilton. The game's object is finding a Hamiltonian cycle along the edges of a dodecahedron such that every vertex is visited a single time, and the ending point is the same as the starting point. The puzzle was distributed commercially as a pegboard with holes at the nodes of the dodecahedral graph and was subsequently marketed in Europe in many forms.

The motivation for Hamilton was the problem of symmetries of an icosahedron, for which he invented icosian calculus—an algebraic tool to compute the symmetries. The solution of the puzzle is a cycle containing twenty (in ancient Greek icos) edges (i.e. a Hamiltonian circuit on the dodecahedron).

Hamiltonian cycles gained popularity in 1880, when P. G. Tait made the conjecture: “Every cubic polyhedron has a Hamiltonian cycle through all its vertices”. Cubic means that three edges meet at every vertex. Without the cubic requirement, there are smaller polyhedra that are not Hamiltonian. The simplest counterexample is the rhombic dodecahedron. Every edge connects one of six valence-four vertices to one of eight valence-three vertices. The six valence-four vertices would need to occupy every other vertex in the length-14 tour. Six items cannot fill seven slots, so this is impossible.