NAME: ASIF ERFAN KHAN

ROLL NUMBER: 546

COURSE: MSc CS

SUBJECT: BIOINFORMATICS

TOPIC: REGULAR EXPRESSION

Practical No: 8

Aim: Generate a regular expression enter three protein sequence of three different organism. Write Python/Java code to find regular expression for these sequences.

Code:

```
def gen_reg_exp(seq_list, no_of_col):
    final_list=[]
    for colnum in range(no_of_col):
        collist=[]
        for colseq in seq_list:
            collist.append(colseq[colnum])
        if len(set(collist))==len(collist):
            #print(final_list)
            final_list.append('x')
        else:
            if len(set(collist))==1:
                final_list.append(collist[0])
            else:
                final_list.append(''.join(set(collist)))
                display_output(final_list)
def display_output(final_list):
    print(*final_list, sep='-')
no_of_seq=int(input("Enter the number of sequence: "))
print("Enter all the sequences")
seq_list=[]
for _ in range(no_of_seq):
    seq_list.append(list(map(str, input("").split())))
    gen_reg_exp(seq_list, len(seq_list[0]))
```

OUPUT:


IDLE Shell 3.11.0
File Edit Shell Debug Options Window Help
Python 3.11.0 (main, Oct 24 2022, 18:26:48) [MSC v.1933 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
========== RESTART: C:/Users/asif0/Desktop/Test.py ==========
Enter the number of sequence: 4
Enter all the sequences
A D L G A V F A L C D R Y F Q
... S D V G P R S C F C E R F Y Q
... A D L G R T Q L R C D R Y Y Q
... A D I G Q P H S L C E R Y F Q
... SA-D-IVL-G-x-x-x-x-FRL-C-ED-R-YF-YF-Q
AS
AS-D-LV
AS-D-LV-G-x-x-x-x-x-C-ED
AS-D-LV-G-x-x-x-x-x-C-ED-R-YF
AS-D-LV-G-x-x-x-x-x-C-ED-R-YF-YF
AS
AS-D-LIV
AS-D-LIV-G-x-x-x-x-RLF
AS-D-LIV-G-x-x-x-x-RLF-C-ED
AS-D-LIV-G-x-x-x-x-RLF-C-ED-R-YF
AS-D-LIV-G-x-x-x-x-RLF-C-ED-R-YF-YF
>>>
Ln: 23 Col: 0