NAME: ASIF ERFAN KHAN

ROLL NUMBER: 546

COURSE: MSc CS

SUBJECT: BIOINFORMATICS

TOPIC: MULTIPLE SEQUENCE ALIGNMENT

Practical No: 4

Aim: Enter genome of five different organism and write a python/java program to find consensus sequence using Multiple Sequence Alignment (MSA) technique.

Code:

```java
import java.io.*;
import java.util.*;
public class Consensus
{
  public static void main(String str[]) throws IOException
  {
    int n, i,j,k,count;
    String seq[],cons[];
    ArrayList<Integer> a = new ArrayList<Integer>();
    ArrayList s = new ArrayList();
    BufferedReader br=new BufferedReader(new InputStreamReader(System.in));
    System.out.println("Enter the no of Sequences");
    n=Integer.parseInt(br.readLine());
    seq=new String[n];
    System.out.println("Enter sequences");
    for(i=0;i<n;i++)
    seq[i]=br.readLine();
    cons=new String[seq[0].length()];
    for(j=0;j<seq[0].length();j++)
      cons[j]=" ";
    for(j=0;j<seq[0].length();j++)
    {
      a.clear();
      s.clear();
      for(i=0;i<n;i++)
```

```java
{

    count=1;

    for(k=i+1;k<n;k++)

    {


        if(seq[i].charAt(j)==seq[k].charAt(j))

        count++;


    }

    System.out.println("count="+count);

    a.add(count);

    s.add(seq[i].charAt(j));

}
/**Updated Snippet 1**/
Set<String> set = new HashSet<>(s);

ArrayList setlist = new ArrayList(set);

Collections.sort(setlist);

if (setlist.contains('-') &&setlist.size()==2){

cons[j]+="-"+setlist.get(1);

}

else if (setlist.size()==1){

cons[j]+="-"+setlist.get(0);

}

else{

int m = Collections.max(a);

int index=a.indexOf(m);

System.out.println("Max="+m);

cons[j]+=s.get(index);

System.out.println("index="+index);

for(i=index+1;i<a.size();i++)

{
```

```java
    if(a.get(i)==m)

    cons[j]+="/"+s.get(i);

     }

     }

    }

    System.out.println("Consensus=");

    for(j=0;j<seq[0].length();j++){

      /**Updated Snippet 2**/

    if(cons[j].length()==2)

    System.out.print(cons[j].toLowerCase());

    else if(cons[j].length()==3)

    System.out.print(cons[j].replace("-",""));

    else

    System.out.print(cons[j]);

     }

   }
}
```

OUPUT:

```
C:\Windows\system32\cmd.exe

Enter the no of Sequences
5
Enter sequences
ACTG
TCGA
TATG
TGCA
___TA
count=1
count=3
count=2
count=1
count=1
Max=3
index=1
count=2
count=1
count=1
count=1
count=1
Max=2
index=0
count=3
count=1
count=2
count=1
count=1
Max=3
index=0
count=2
count=3
count=1
count=2
count=1
Max=3
index=1
Consensus=
 t c t a
C:\Users\admin\Desktop>Pause
Press any key to continue . . .
```