Virology, Genome Sequencing, and Bioinformatics

Resource Type: Curriculum: Classroom

Publication Date: 7/2/2001

Authors

Erica Suchman
Colorado State University
Fort Collins, Colorado 80523
USA
Email: Erica.Suchman@colostate.edu

Mark Gallo
Niagara University
Lewiston, New York
USA
Email: mgallo@niagara.edu

Carol Blair
Colorado State University
Fort Collins, Colorado
USA
Email: Carol.Blair@colostate.edu

Abstract

Students are given instructor-designed computer-generated "sequencing gels" that they are told were generated after sequencing PCR from samples isolated during a hemorrhagic fever outbreak. They must read the sequencing gel and then using this sequence go to the NCBI home page and perform a BLAST search to determine the identity of the virus. The class compares the results of these searches, and students answer a series of questions about the virus for which they have sequence information. The class then attempts to determine the cause of the outbreak as a group.

Activity

Invitation for User Feedback. If you have used the activity and would like to provide feedback, please send an e-mail to MicrobeLibrary@asmusa.org. Feedback can include ideas which complement the activity and new approaches for implementing the activity. Your comments will be added to the activity under a separate section labeled "Feedback." Comments may be edited.

INTRODUCTION

Time Required.
One 50-minute class period; however, students will be required to work outside of class as well for at least a few hours.

Learning Objectives.
This activity was designed to introduce students to how scientists read dideoxy DNA sequencing gels and to the World Wide Web-based search engines (BLAST in particular) that can be used to analyze these sequences. Furthermore, students compare data with the group to determine the identity of the virus causing a hemorrhagic fever outbreak.

Background.
To prepare for this activity students must be familiar with PCR and DNA sequencing. We include a section in our laboratory manual describing how sequencing reactions and gels are prepared and how a sequencing gel is read. We take students to our computer lab and lead them through the process of performing a BLAST search and analyzing the results. This process is described in their lab manuals.

PROCEDURE

Materials.
Included are the computer-based sequencing gels, the assignment the students are given, the questions they must answer, and a description of how sequencing reactions and gels are prepared and how sequencing gels are read.
1. Explanation of activity (PDF)
2. Sequence gels (PDF)
3. Sequence gel keys (PDF)
4. Questions students must answer about search (PDF)
5. Description of preparing sequencing reactions and gels and reading gels (PDF)

Instructor Version.

1. Each student is given an instructor-generated sequencing gel and asked to use the explanation in their lab manual to

learn to read the gel. They have a few days at home to read the gel and are expected to know the sequence 5' to 3' by the next lab period. However, this could be done during the period if there is enough time. The instructor has prepared about 25 different sequences and each student's sequence is labeled uniquely with a number; however, the virus responsible for the outbreak has been given to many students, as you would expect this to be the most commonly amplified virus in an outbreak.

2. During the next period students are taken to the computer lab and led through a BLAST search of their sequence.
3. After all students have completed their BLAST searches they must use their textbooks to answer questions about their virus, such as does it cause a hemorrhagic fever, is it commonly found where the outbreak is occurring, etc.
4. Each student writes the identity of his/her virus on the board, and the class compares their results in order to determine the identity of the virus causing the outbreak.

Safety Issues. None.

ASSESSMENT and OUTCOMES

Suggestions for Assessment.
Students are graded on the answers they provide to the list of questions they are given. Note: each student is given a different sequence and therefore will be providing different answers; thus no rubric has been included.

Problems and Caveats.

- The most difficult part was preparing all of the sequencing gels. We found that about 35 base pairs was the shortest sequence that could be given and still return a usable BLAST search. If an instructor wants to design new sequences, they should test each sequence by performing a BLAST search to insure it will be specific enough.
- We have a student computer lab near our teaching lab, therefore we were easily able to go to the computer lab to perform our searches as a group. However, many students already had their searches done when they came into class, so if you do not have a computer lab, students can easily perform these searches on their own.

SUPPLEMENTARY MATERIALS

Possible Modifications.
We perform this exercise in our virology and cell culture laboratory, but it could just as easily be performed in a lecture course. We take our students to the computer lab and go through performing a BLAST search and analyzing the results, but many students had already performed the BLAST search before class and had easily followed our directions. Every year we plan to change the location of the outbreak so that a different virus will be responsible for the disease.

References.

NCBI BLAST search - http://www.ncbi.nlm.nih.gov

User Feedback

We performed this activity for the first time in the fall of 2000; the students really enjoyed it and were amazed at how easy a BLAST search is to do. One student commented about how happy she was that we had done this because she imagined a BLAST search to be a very difficult process and was pleased that she could so easily perform one.

Identification of Viruses by Genome Sequences

Complete nucleotide sequences of many virus genomes have been determined and placed in readily accessible databases. These data can be used, as in the following exercise, to identify unknown viruses and to determine possible relationships of new viruses to those already characterized.

Nucleotide sequences of DNA from viral genomes or PCR or RT-PCR products derived from viral genomes are usually obtained by the di-deoxynucleotide method developed by Sanger. The four sequencing reactions include DNA or cDNA template, DNA polymerase, an oligonucleotide primer, and 4 deoxynucleoside triphosphates (dNTPs), one of which carries a radioactive label. In each reaction, there is also a low concentration of one of the nucleotides in di-deoxy form (lacking both the 2' and the 3' hydroxyl groups required to form phosphodiester bonds during DNA synthesis). In each synthesis reaction, incorporation of the di-deoxynucleotide terminates the nascent DNA chain; therefore, every product in the ddATP-containing reaction, for example, will have the base A at its 3' end. After incubation at $37^{o}C$ for 30-60 min, the four reaction mixtures are run side-by-side on a polyacrylamide gel that separates the labeled product DNA by size, with the smallest DNA molecules migrating through the gel most rapidly. The incorporation of the radioactive deoxynucleotide allows visualization of the migration distance of each DNA product after exposure of the gel against x-ray film. The beta-particle emissions result in black bands on the film (see figure).

The National Center for Biotechnology Information (NCBI) web page, located at http://www.ncbi.nlm.nih.gov/ is maintained by the National Library of Medicine and the National Institutes of Health. It allows you to compare sequences you have obtained to all known sequences in the BLAST search database.

Problem: The Centers for Disease Control and Prevention (CDC) has been contacted for help by public health authorities responsible for the region surrounding a small town called Aldehuela Grande in South America. The Minister of Health reports that several people in Aldehuela Grande suddenly developed high fever with hemorrhagic syndrome and about 40% of affected individuals died. Patient specimens were cultured for the presence of bacteria, protoza, and fungi, and none of these infectious agents was isolated. Therefore, it is believed that a virus caused the disease outbreak. Immunohistochemistry was attempted using antibodies against known viruses; however, due to limitations of laboratory facilities and reagents, no positive results have been obtained. Health officials in Aldehuela Grande sent samples from local rodents and arthropods that could have served as viral reservoirs or vectors as well as patient specimens to the CDC for analysis by reverse transcription-polymerase chain reaction (RT-PCR). Technical staff at the CDC performed RT-PCR on all samples, with positive controls consisting of genomes from viruses known to cause hemorrhagic fevers, using primers designed to amplify conserved regions in the viral genomes. Nucleotide sequencing reactions were then performed on small segments of the RT-PCR products. Unfortunately, the notebook with the key to origin of each of the samples was destroyed in a freak laboratory accident. Knowing the skill and reliability of MB425 students, the CDC has asked us to determine the identities of the viral genome amplicons. They will then try to put together the pieces of this puzzle.

Procedure:

1. Use the diagram of the sequencing gel you are given to determine the nucleotide sequence of the viral genome fragment.

2. Perform a BLAST search to determine if your sequence is from the genome of any virus in the database. Type in the NCBI webpage address, click on BLAST, then click on Basic BLAST. You will perform a blastn (for nucleotide) using the nr database. You will have to type in the sequence you determined and compare it to the entire database. (You could restrict or tailor the type of search if you knew the sequence was from the Drosophila or human genome or if you had a peptide sequence.) Your results will be given from top to bottom with the best matches at the top. You may get more than one equally good match (with the same probability score), so be sure to look at the scores. Indicate which virus genome(s) your sequence matches best.

3. Is it probable that your sequence could have come from one of the patient samples, a rodent reservoir, or a mosquito? (This information won't come from your BLAST search, you will have to rely on your knowledge of virology, your MB420 lecture notes, or a virology text or reference book.)

4.   With the limited clinical data you have, do you think it is probable that the virus you identified could be responsible for this outbreak? That is, is this virus known to cause a hemorrhagic fever? Does it usually have a high mortality? Does it occur in South America?

A

A   T   G   C

B

C

A   T   G   C
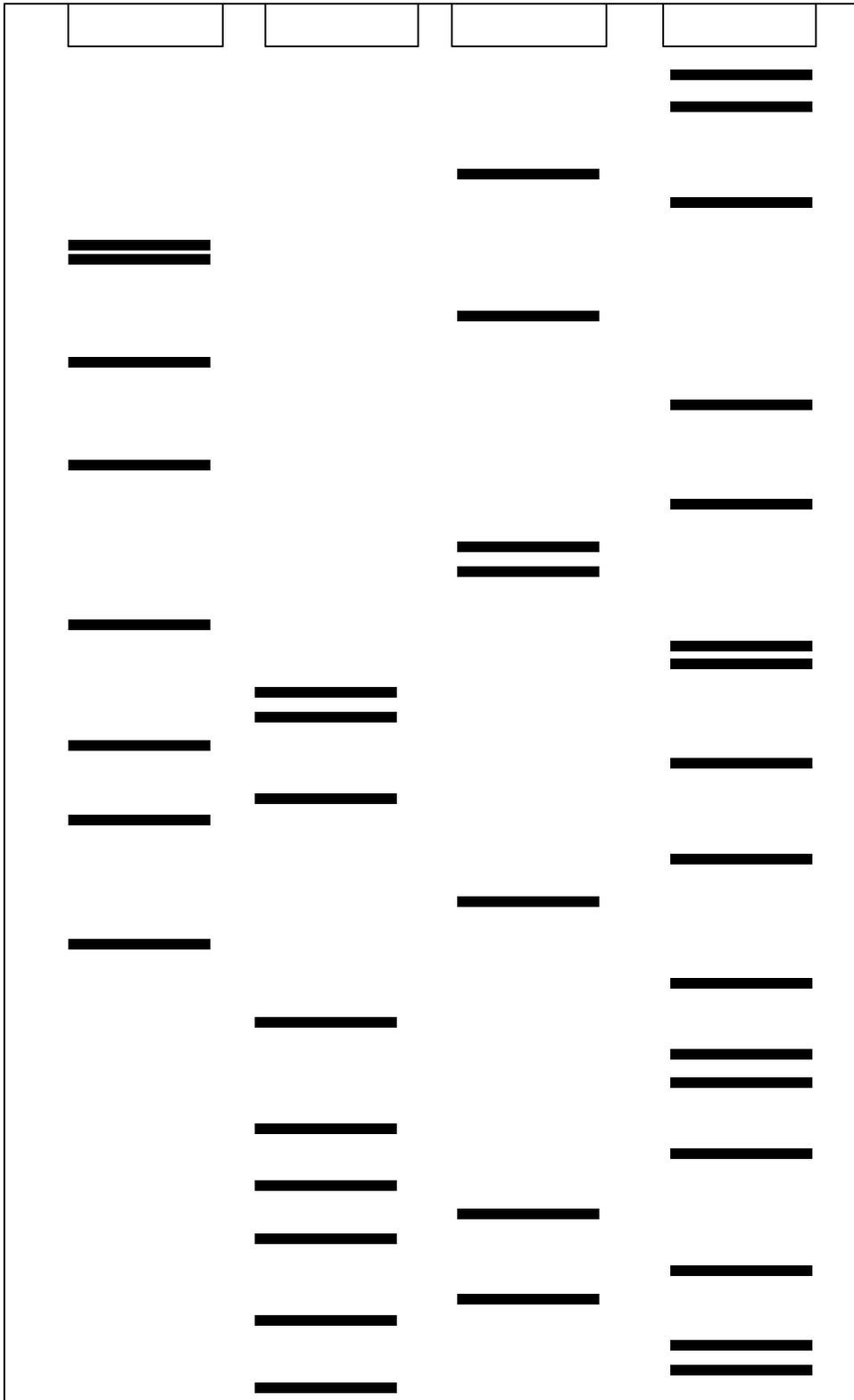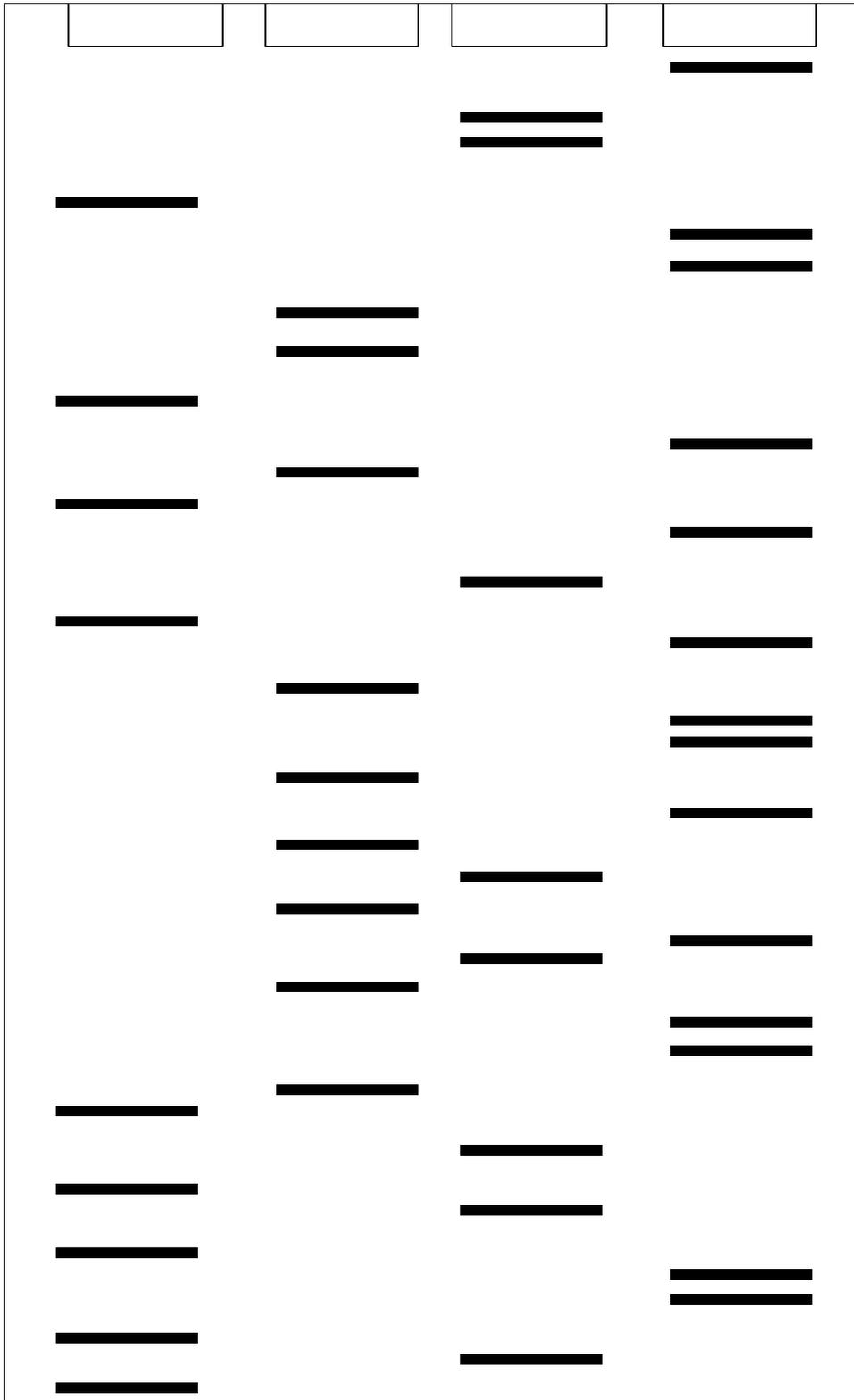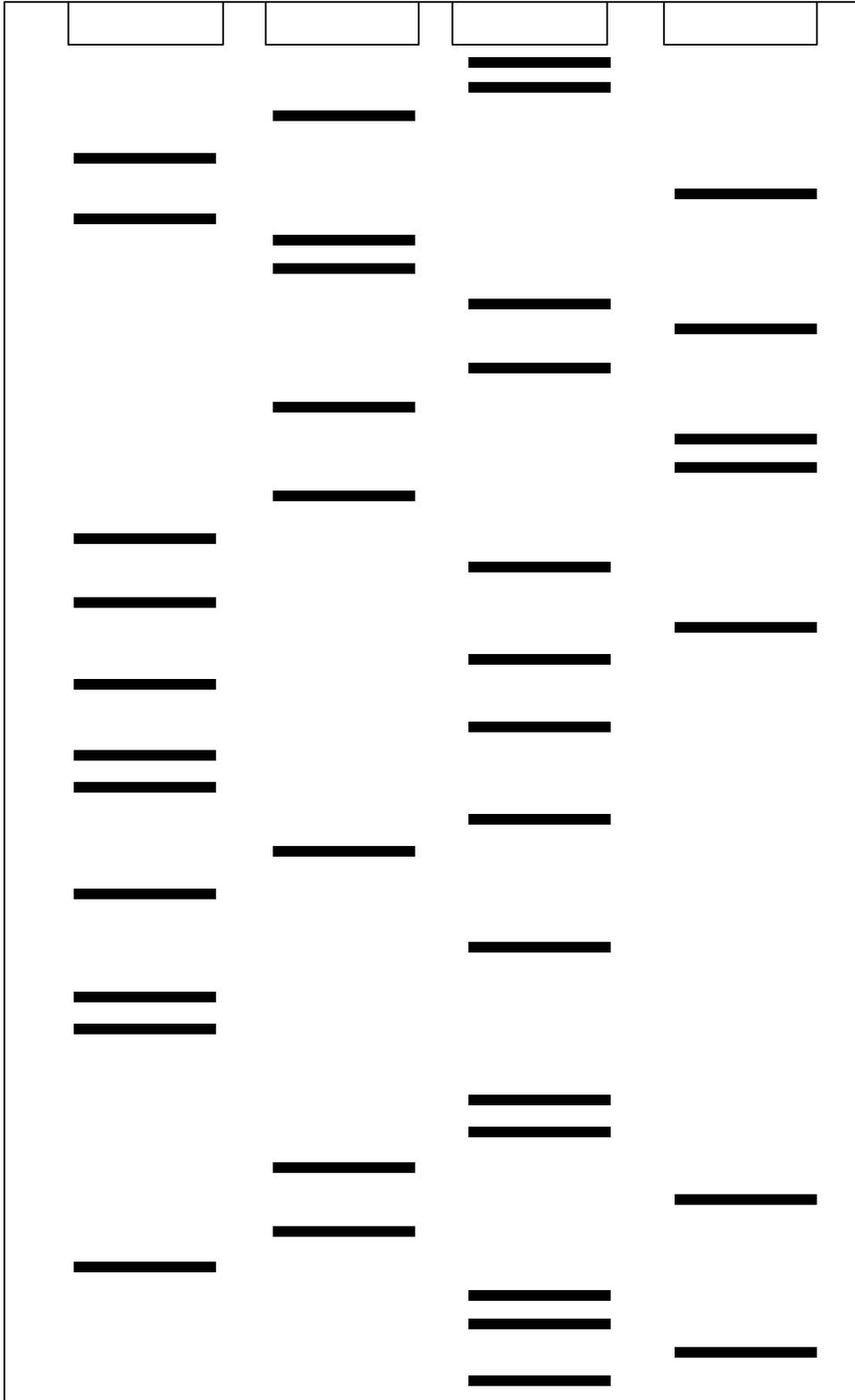
E

A   T   G   C

F

A   T   G   C

G
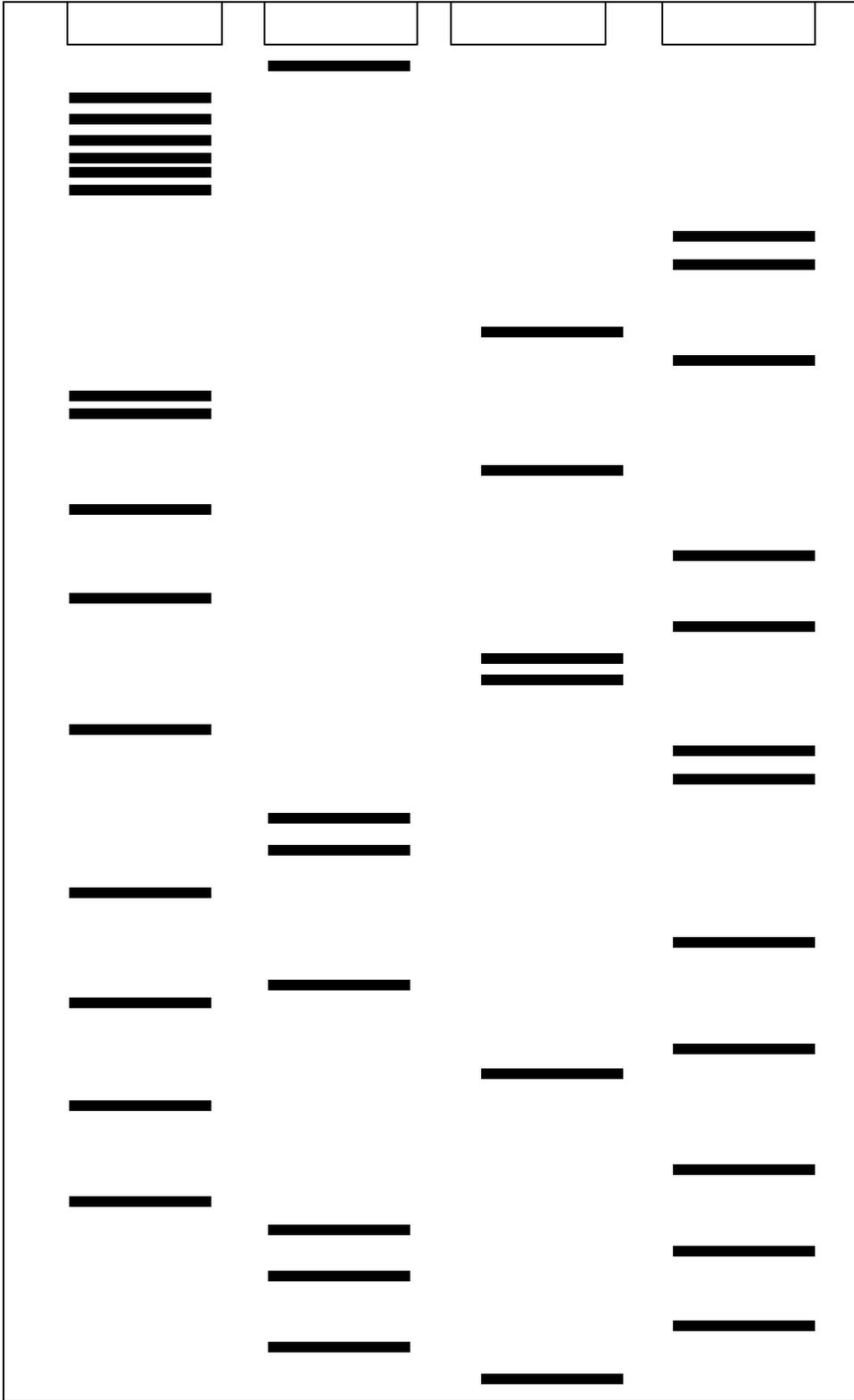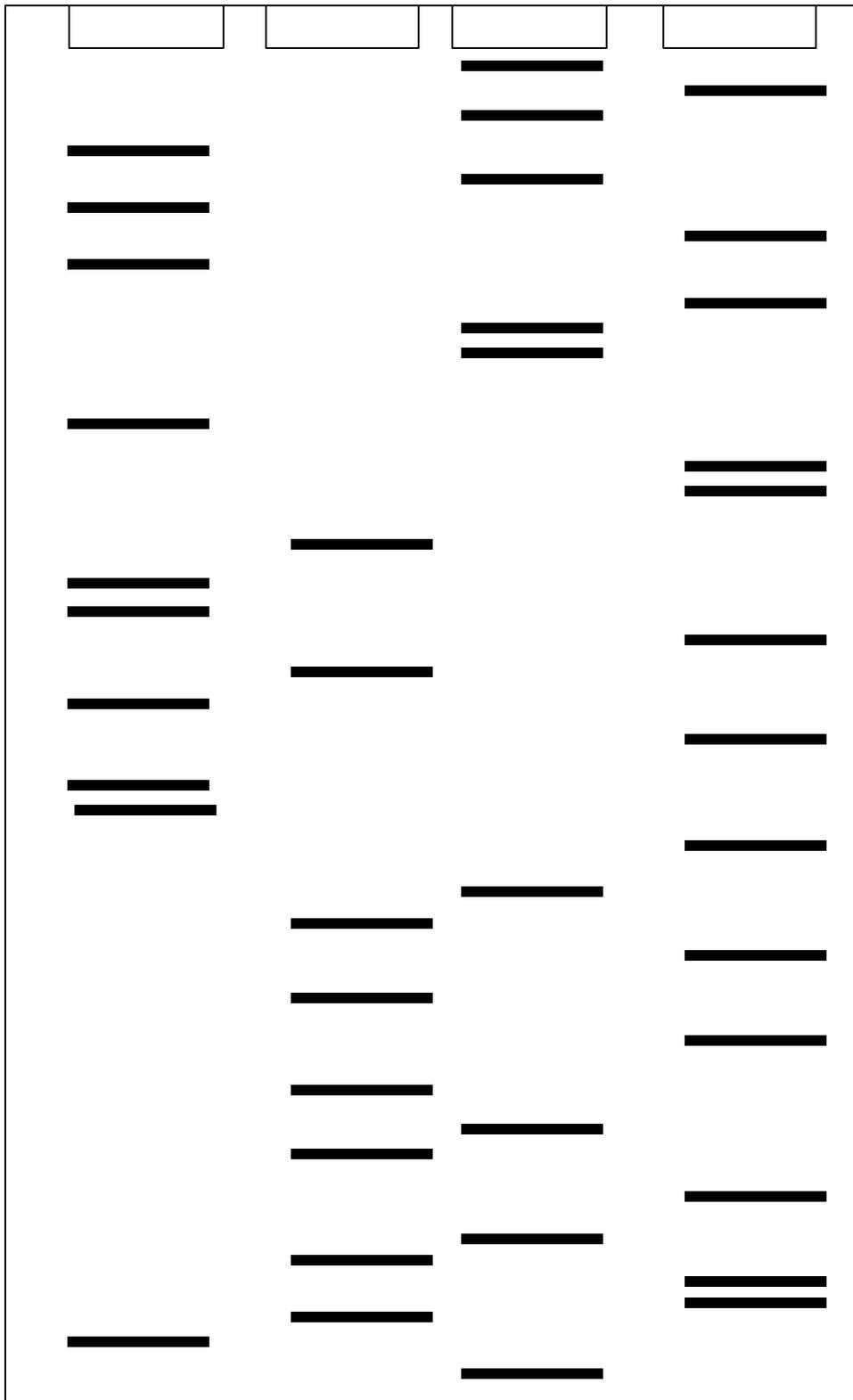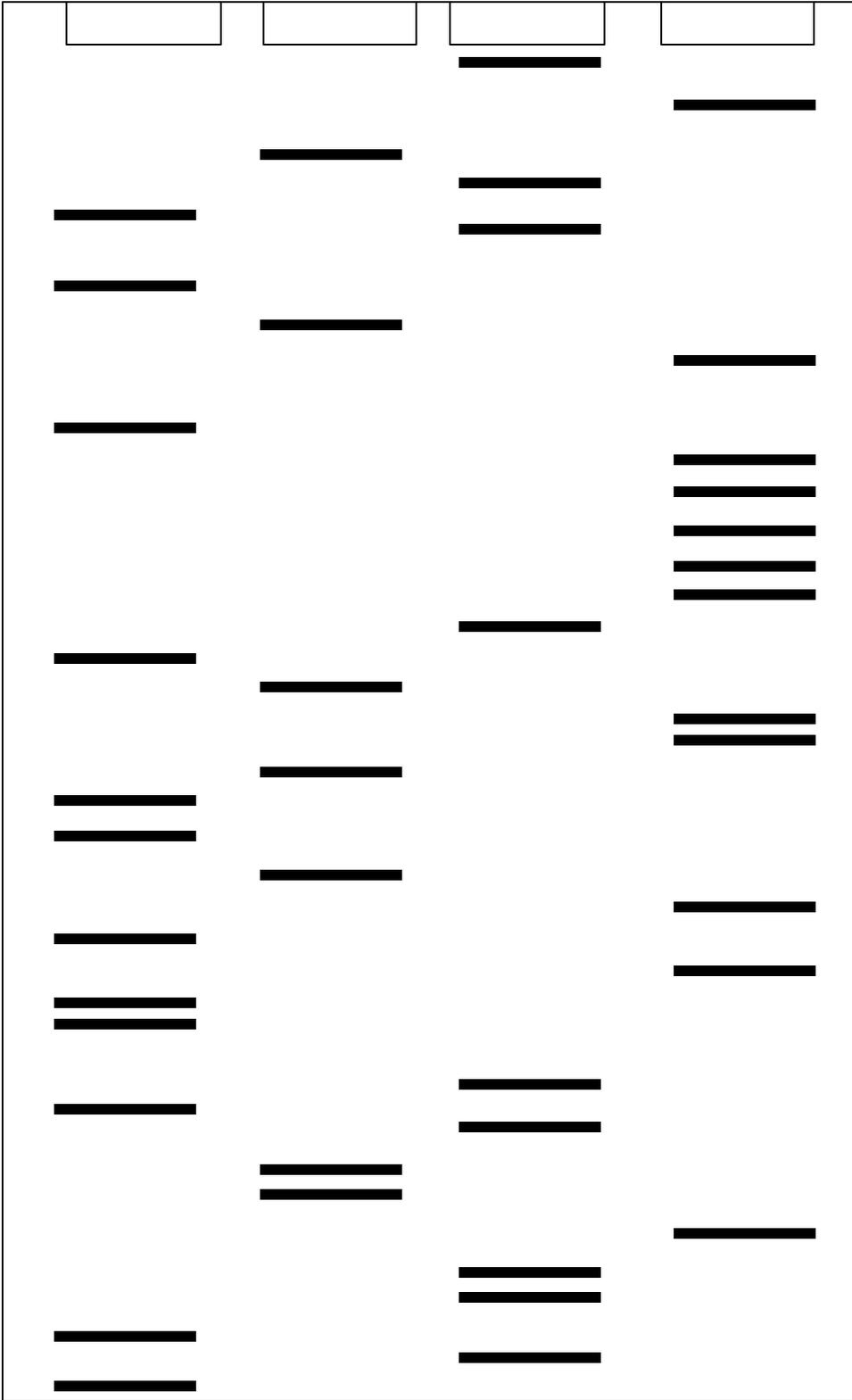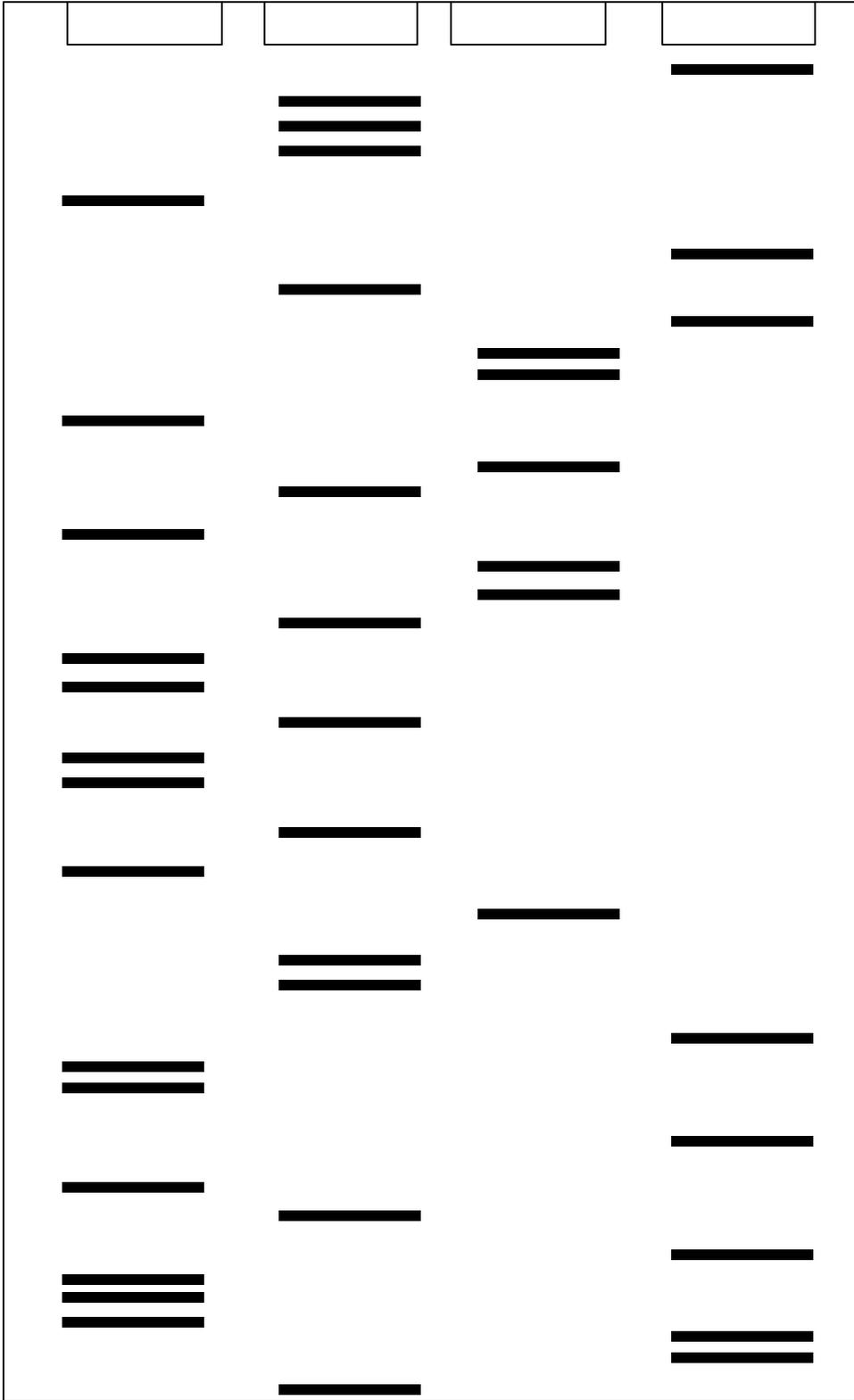
A T G C
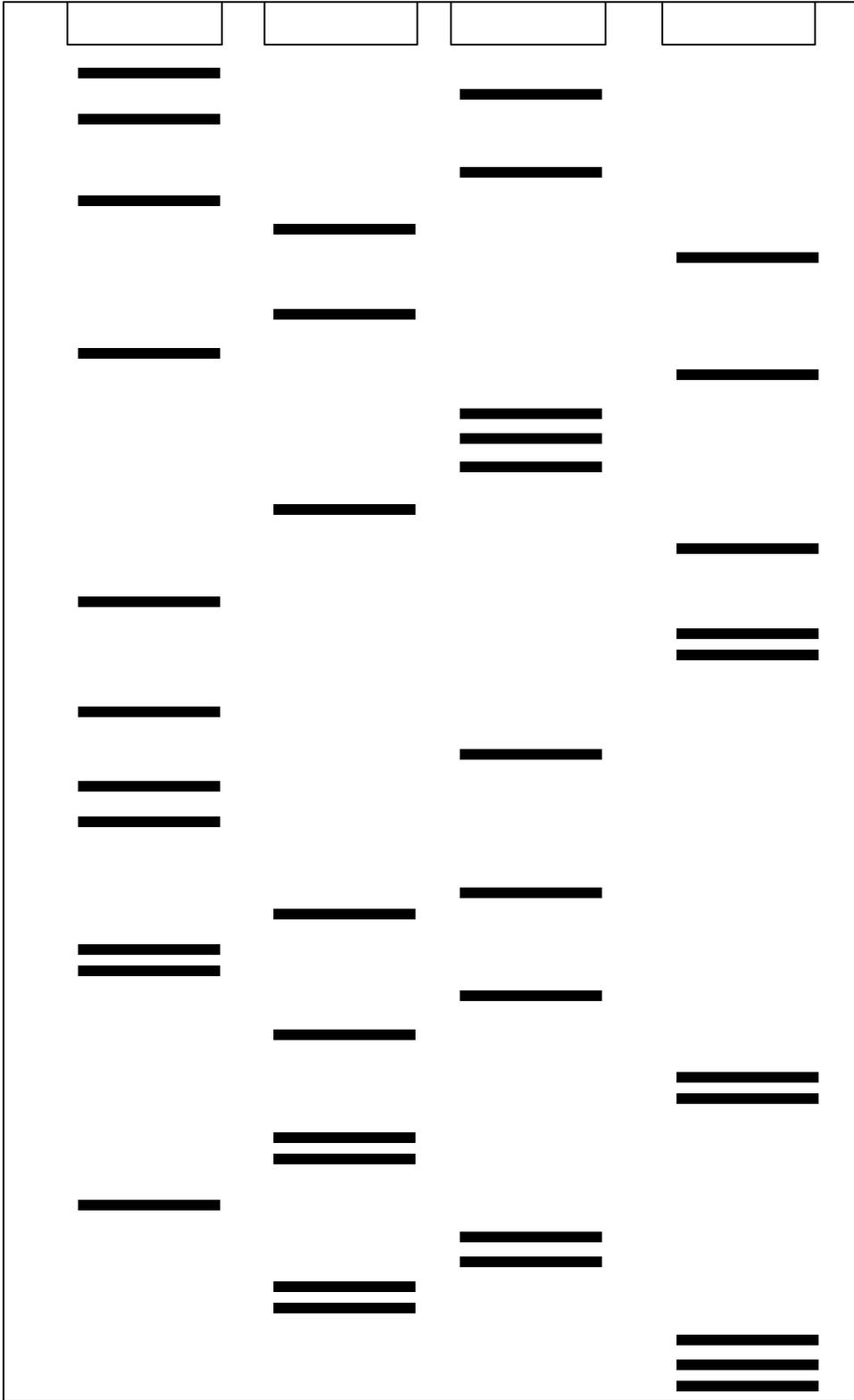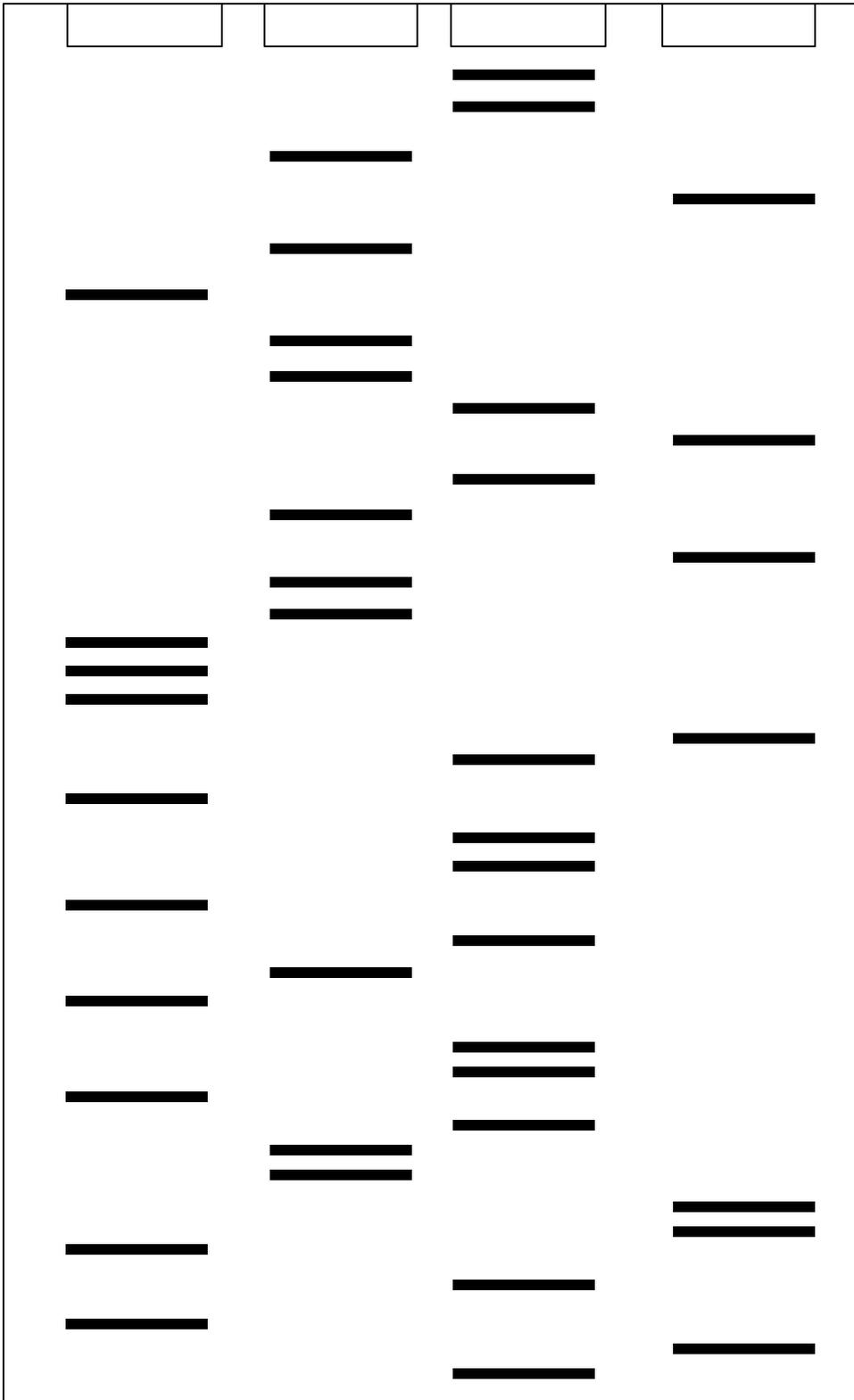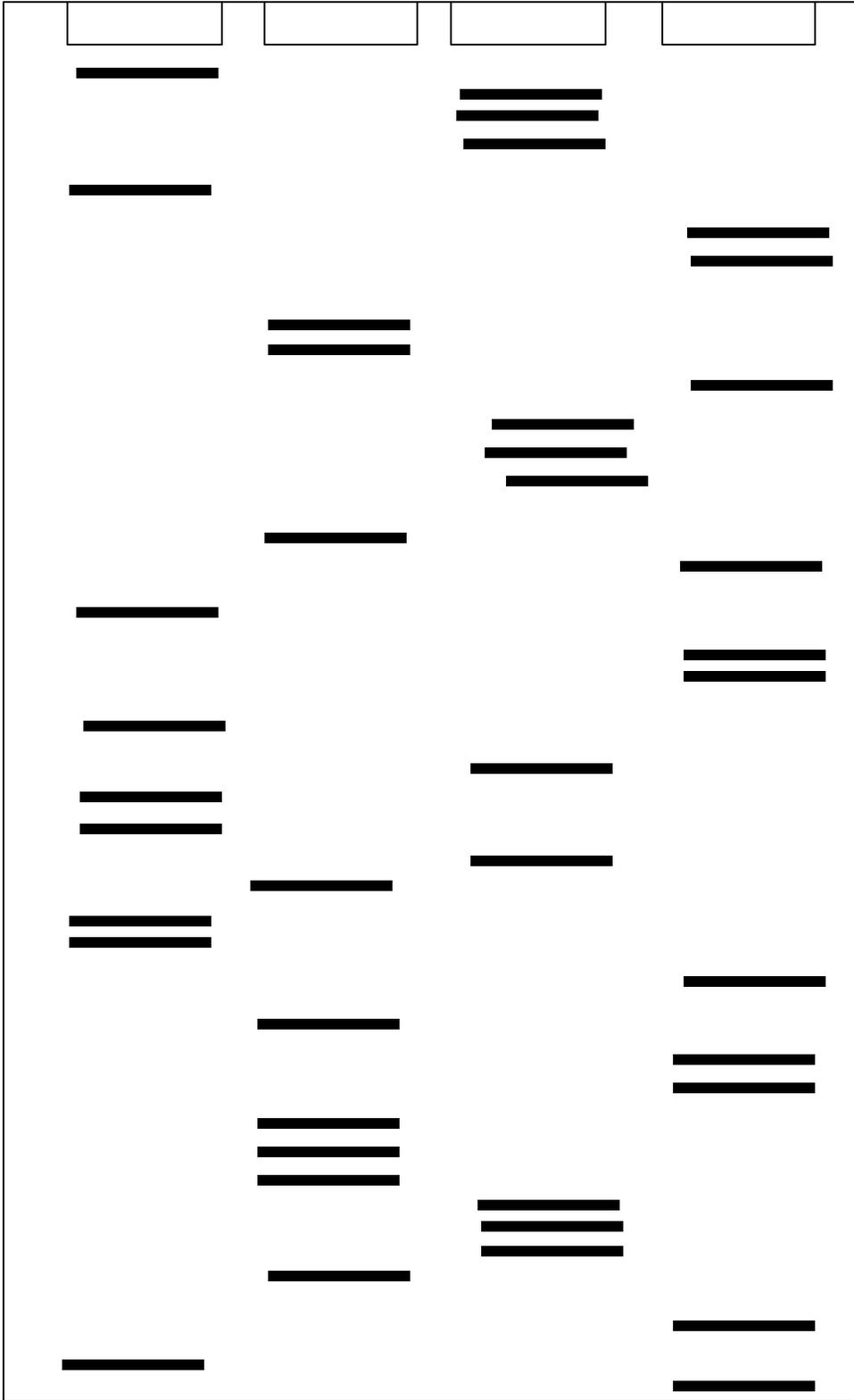
I

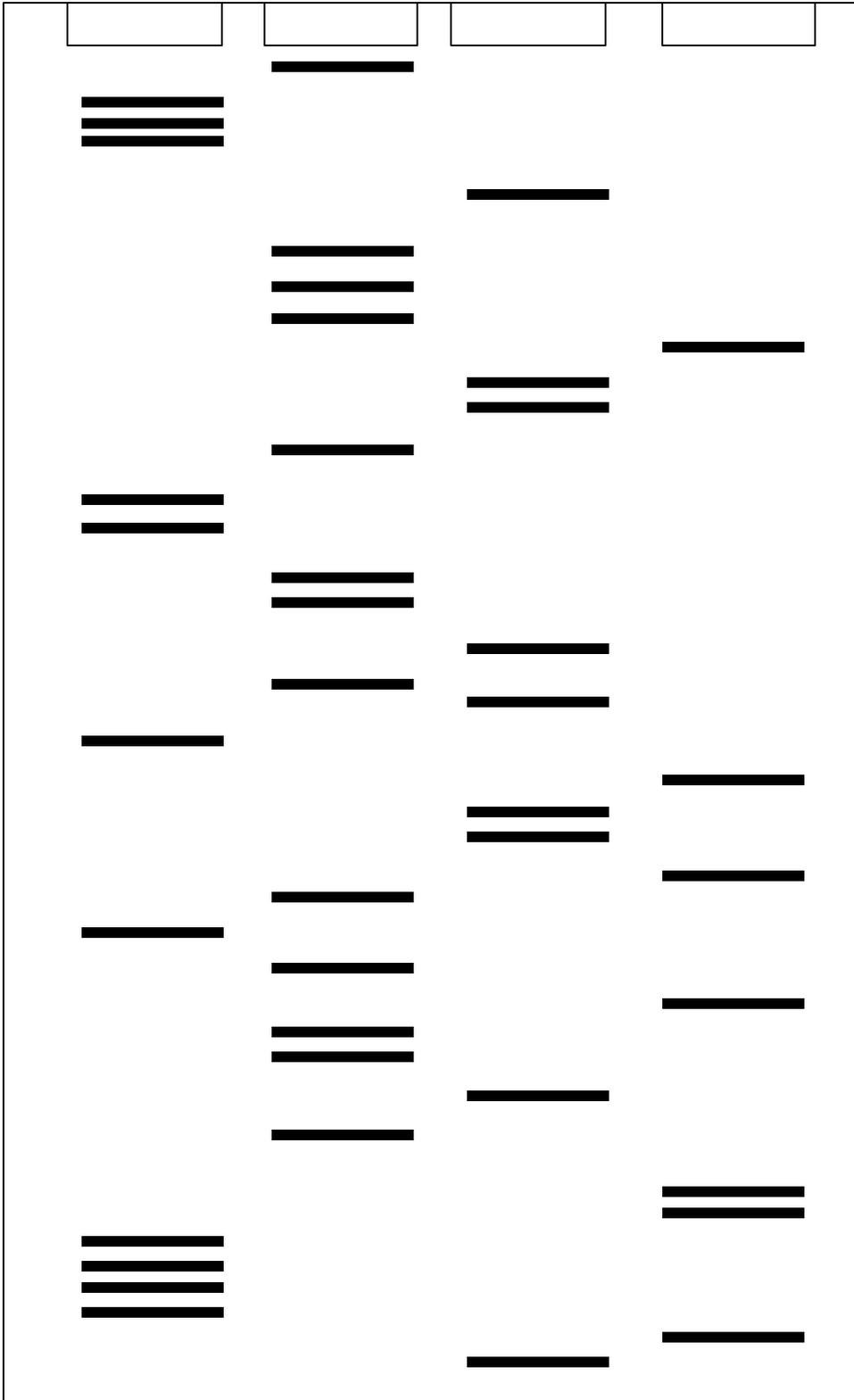A   T   G   C

J

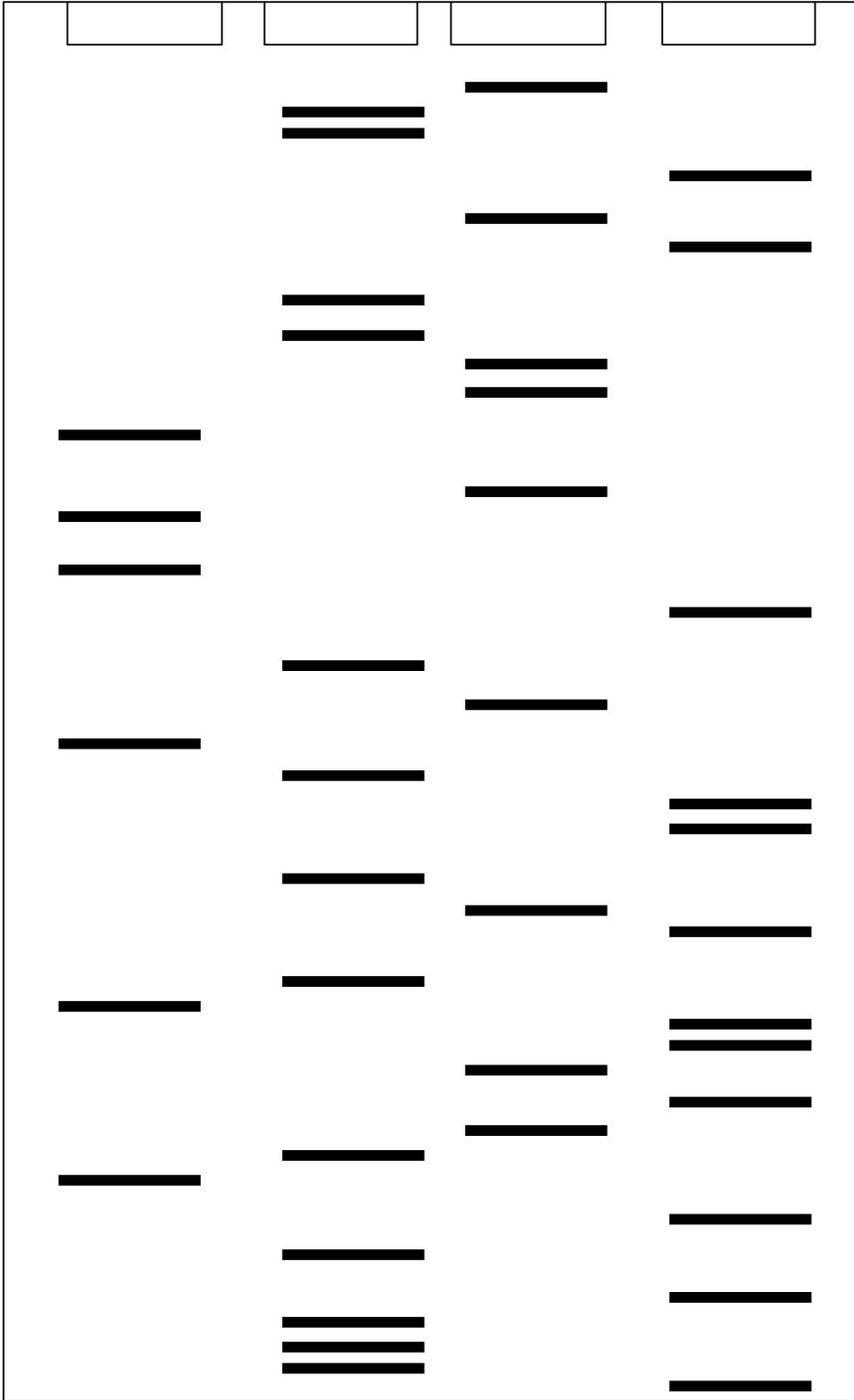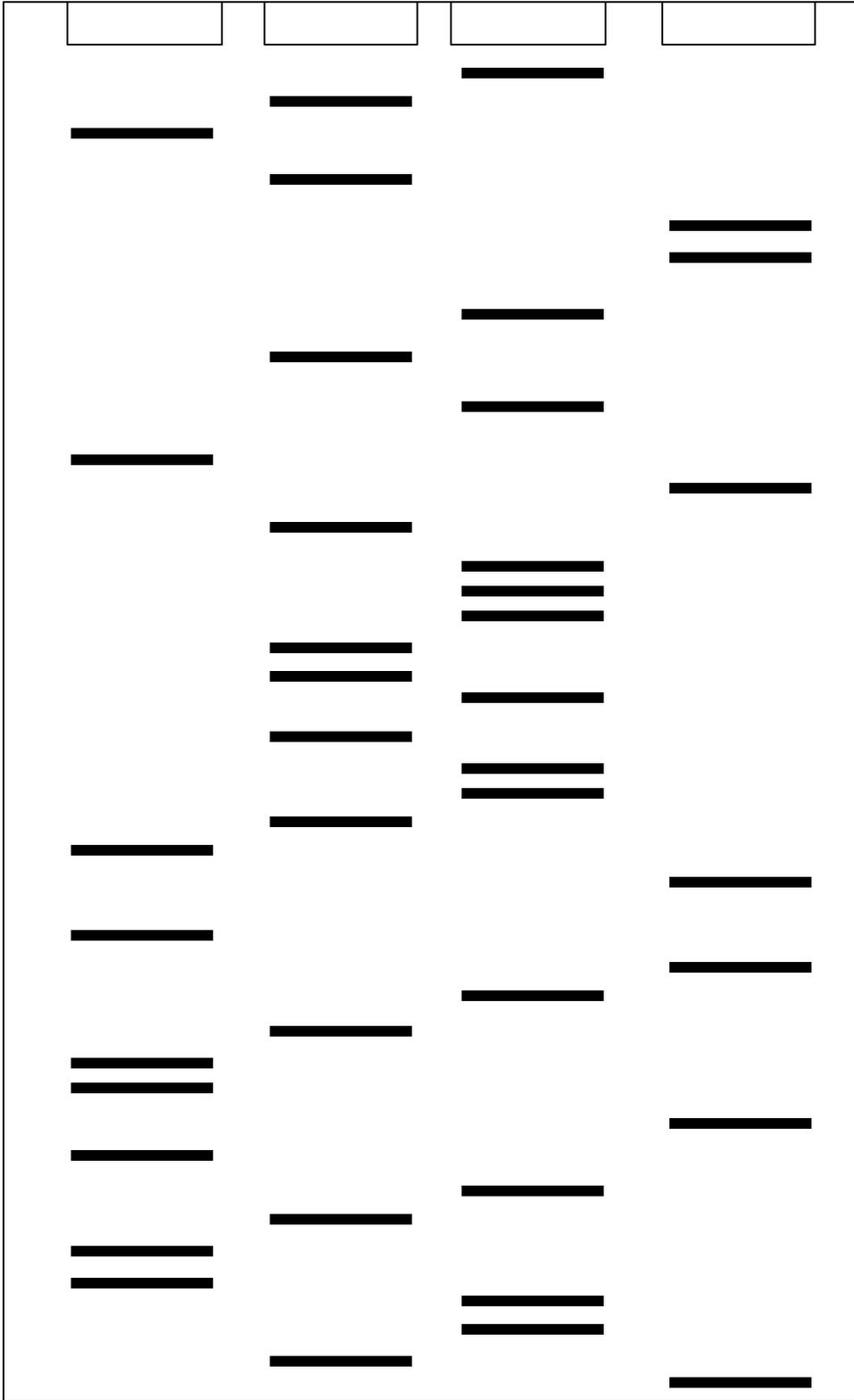A T G C

K

A T G C

L

A T G C

M

A T G C

N

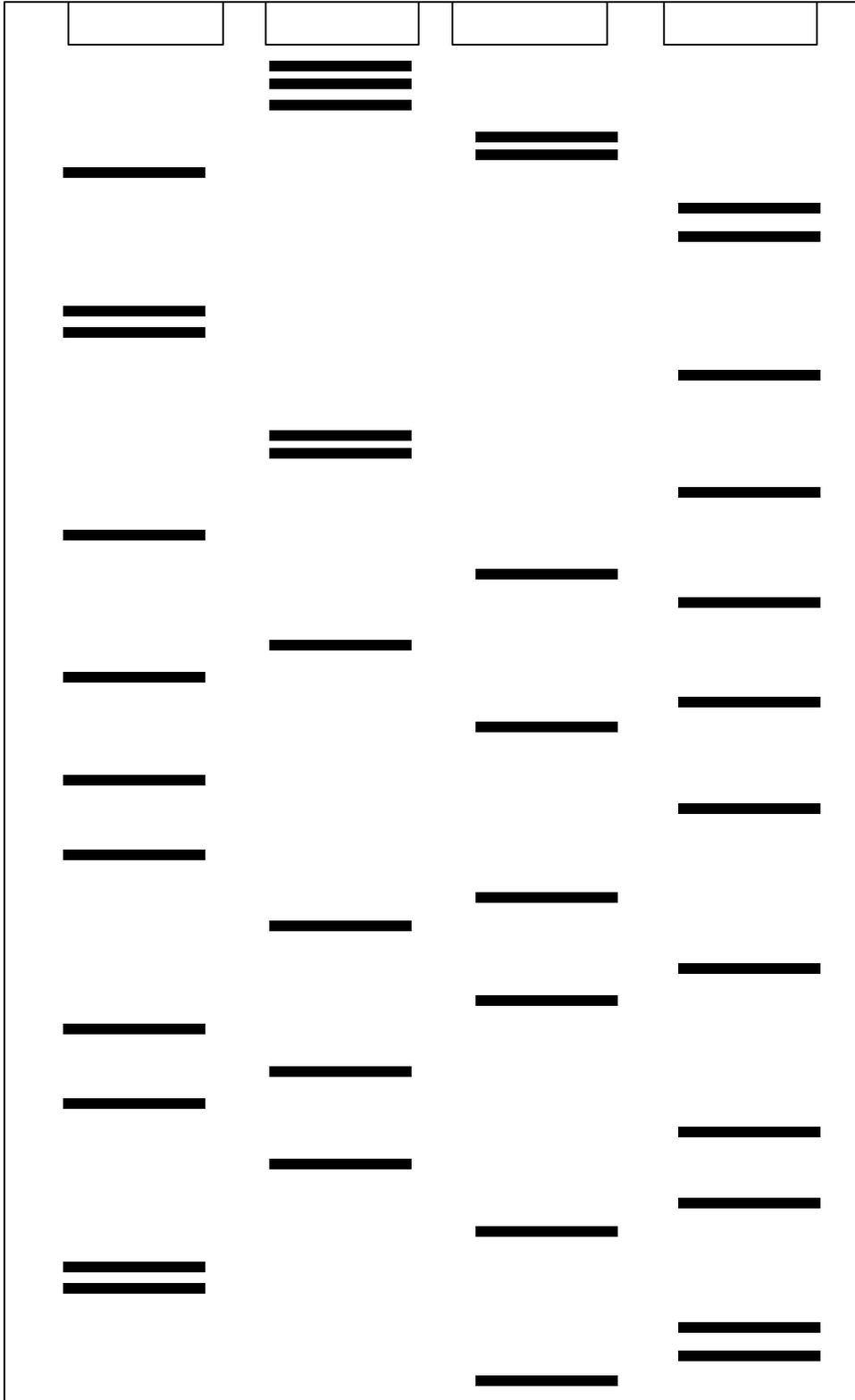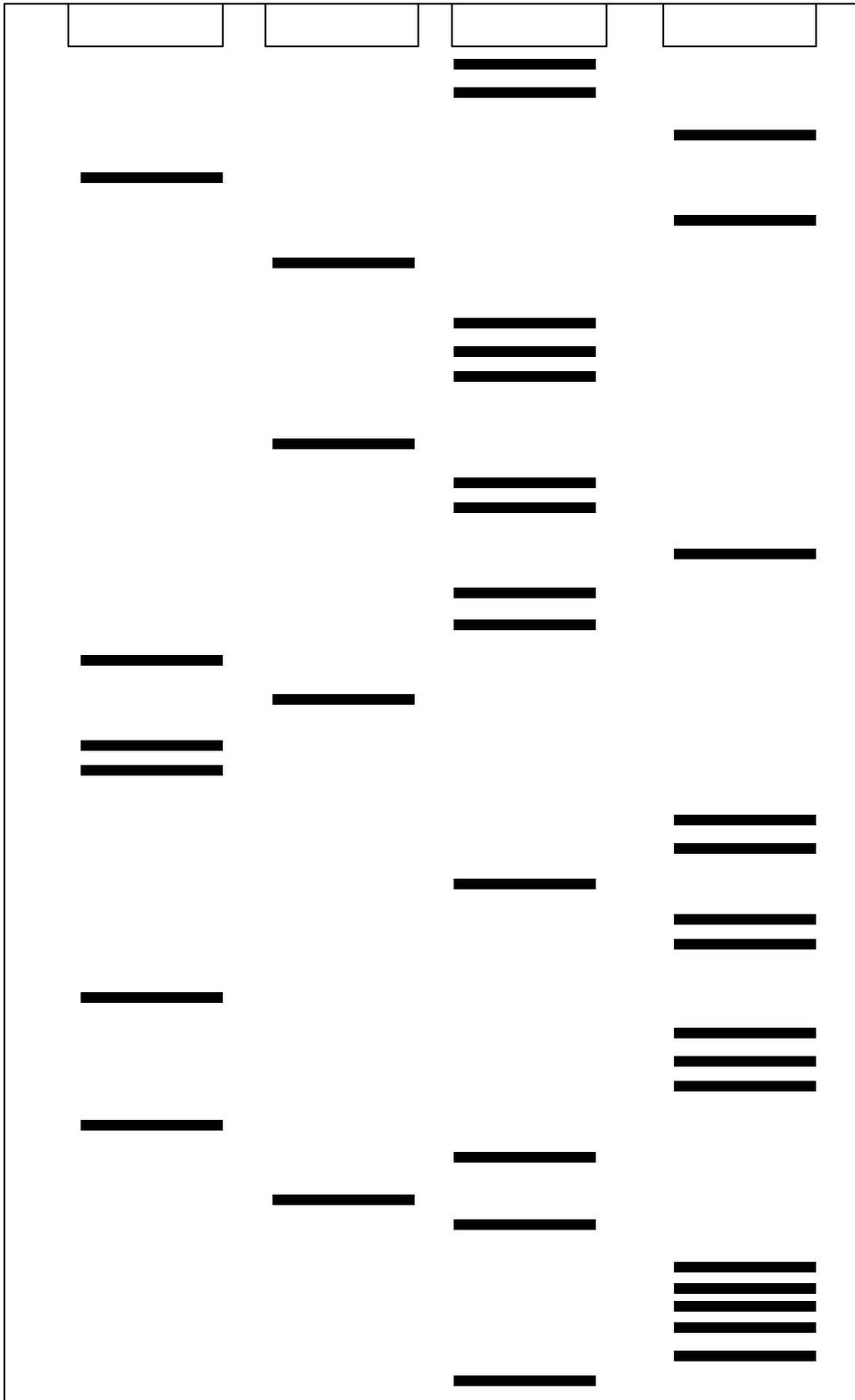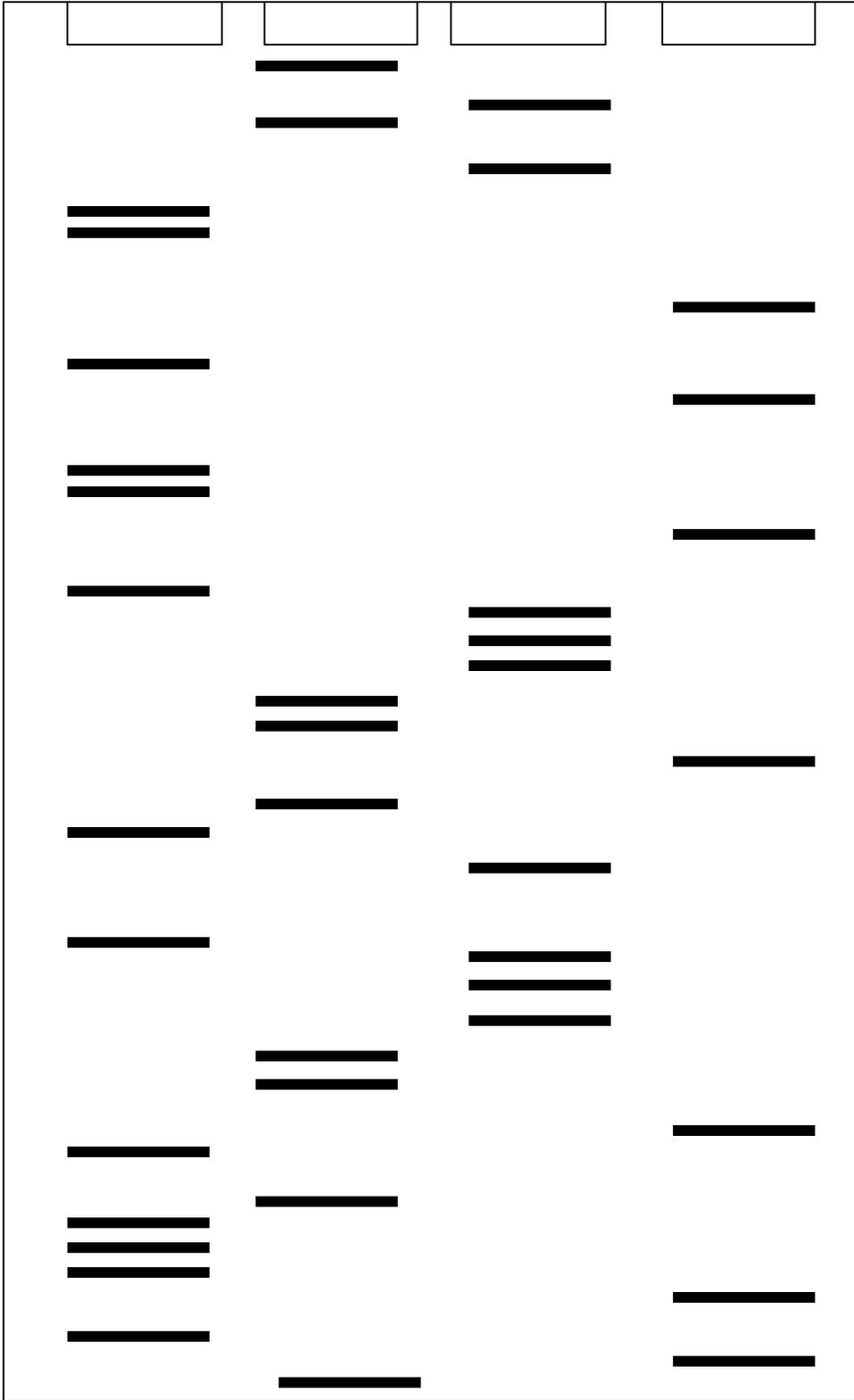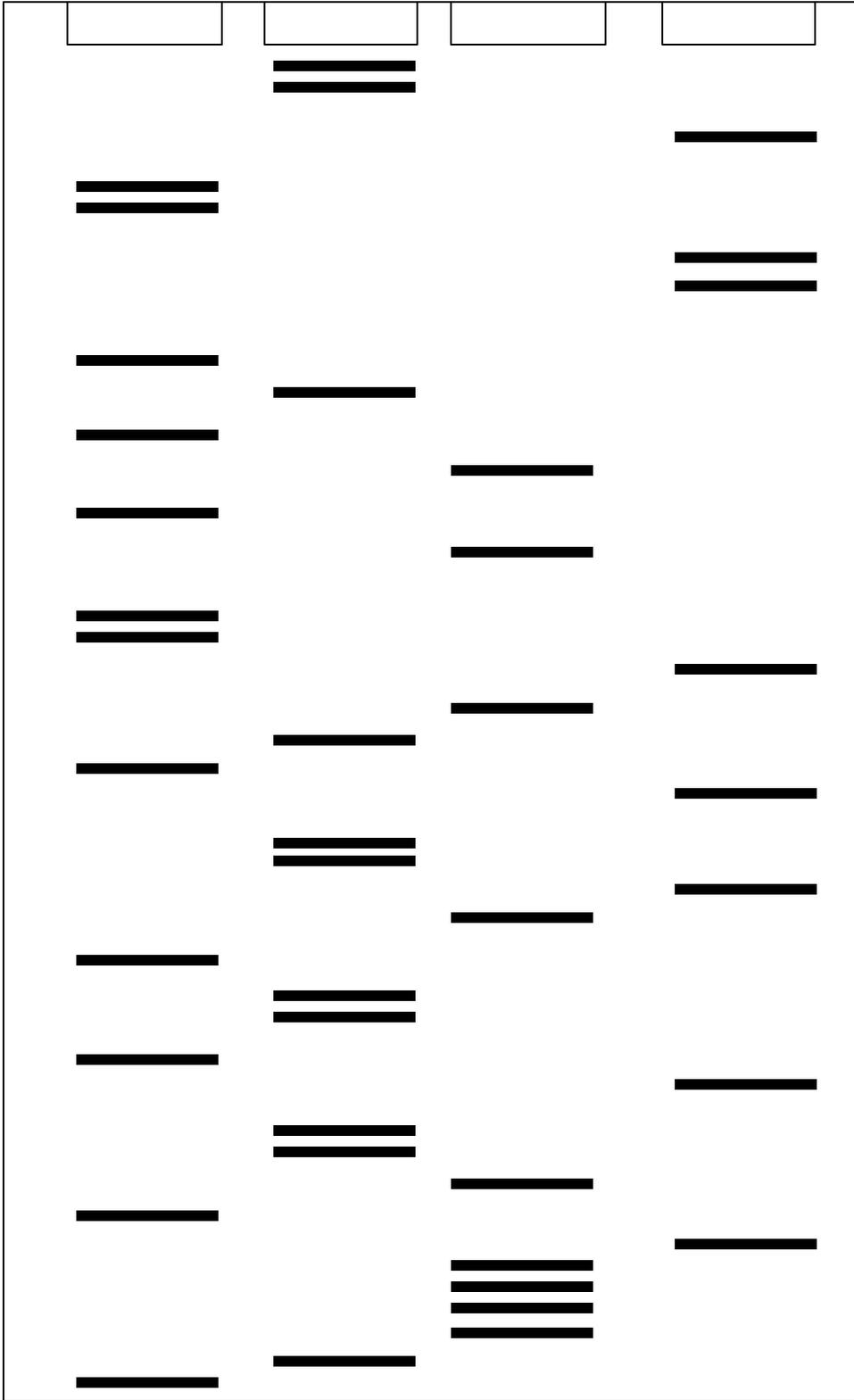A   T   G   C

O

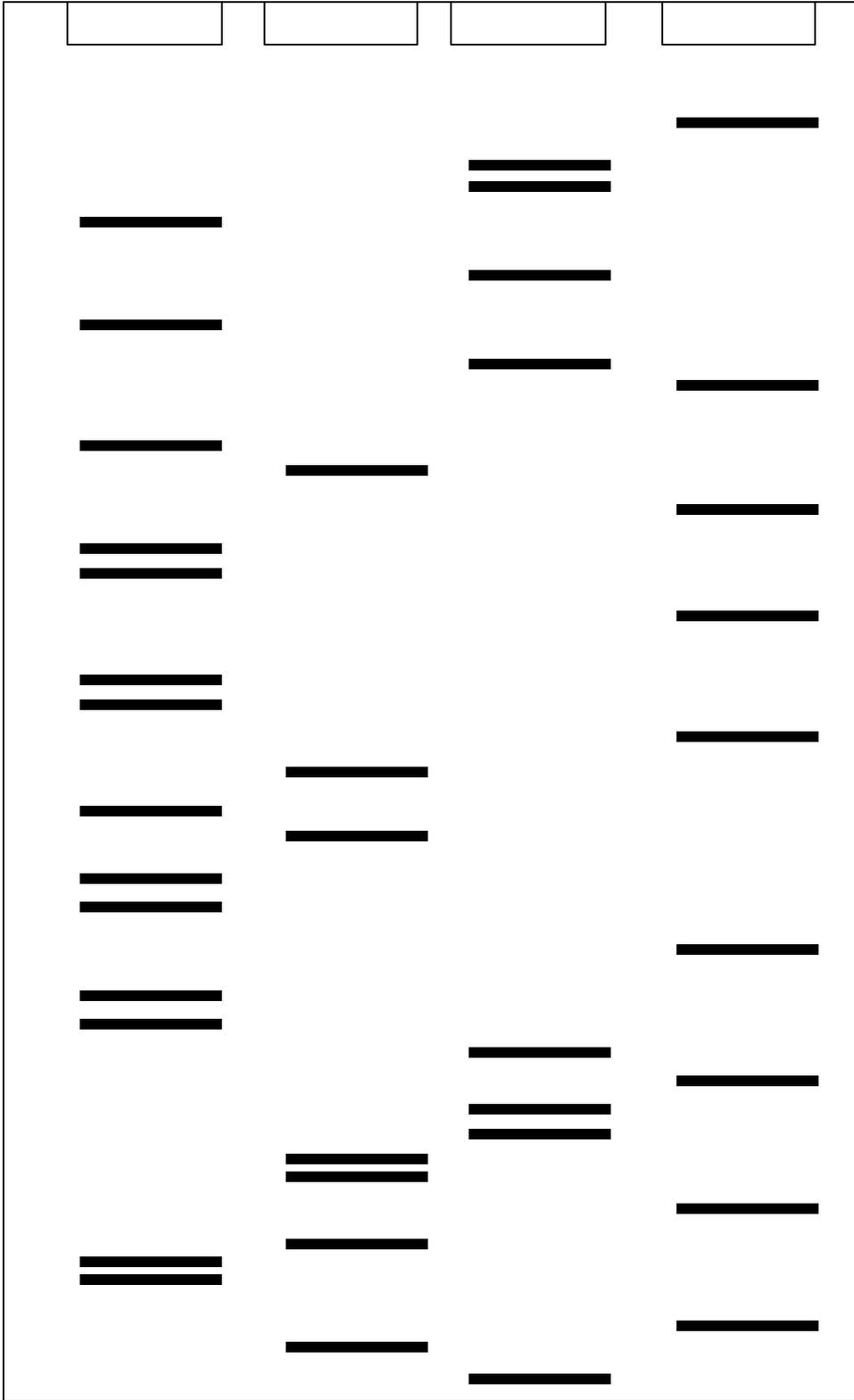A T G C

P

A   T   G   C
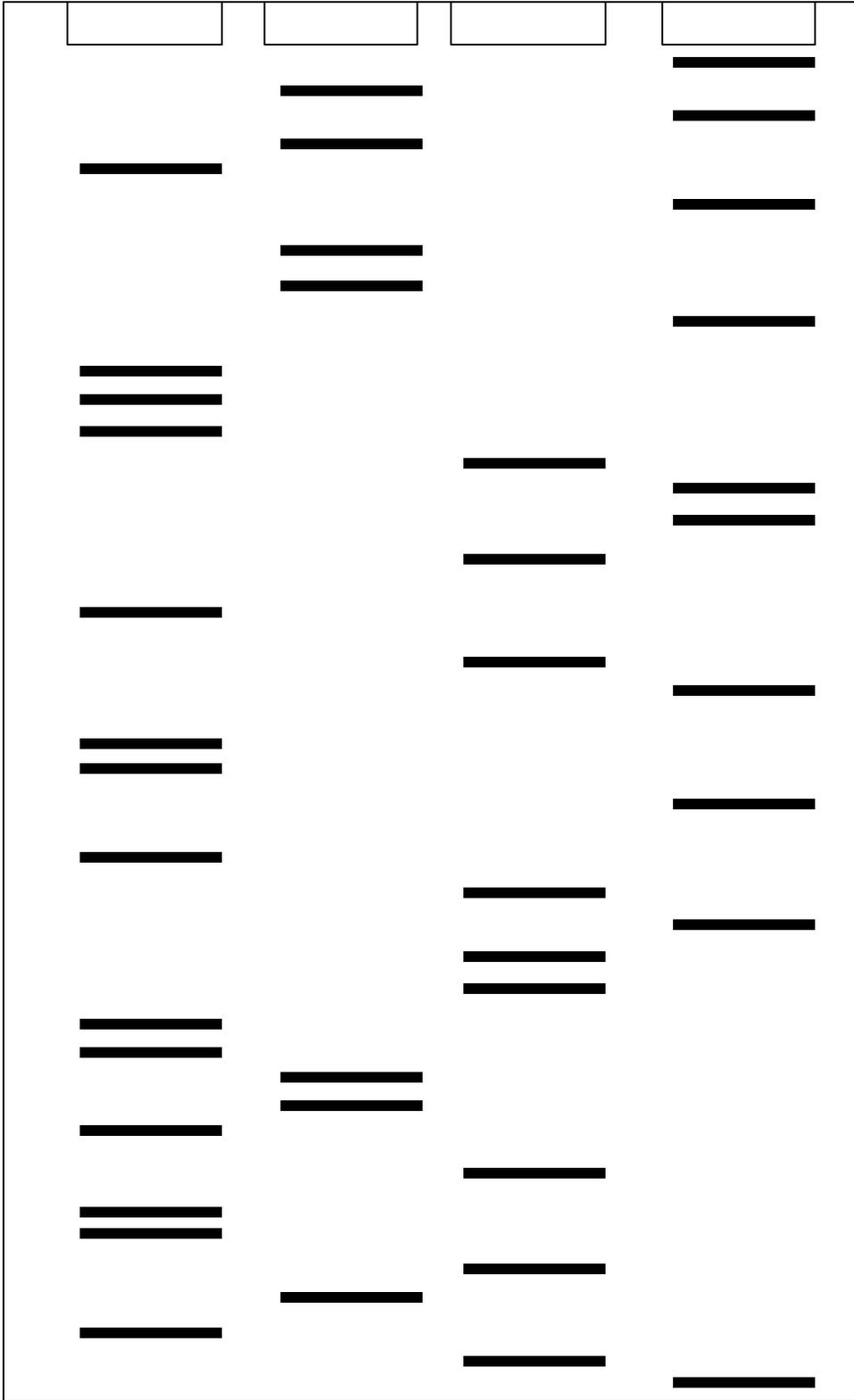
T

A T G C

V

A   T   G   C

W

A   T   G   C
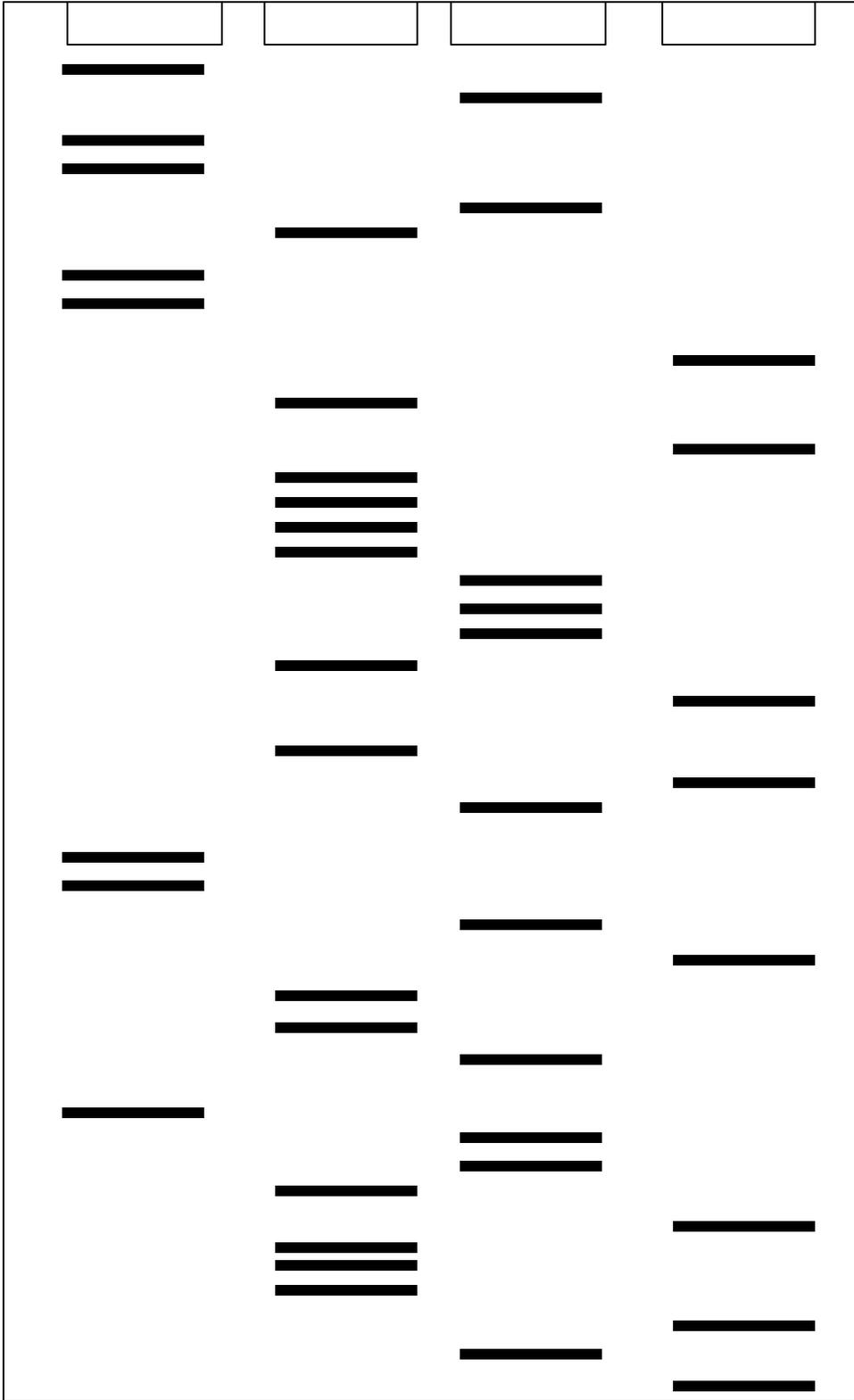
X

A   T   G   C
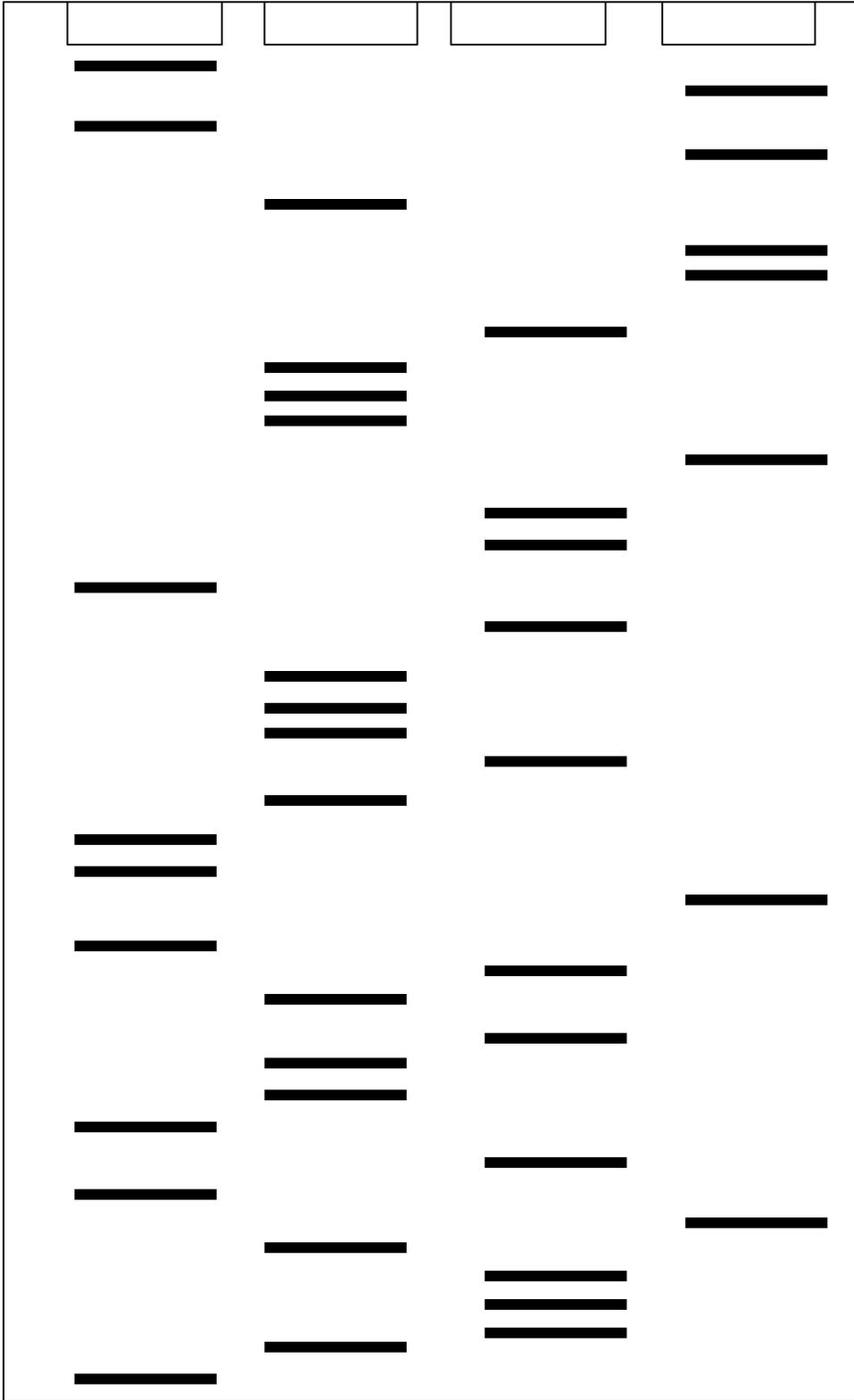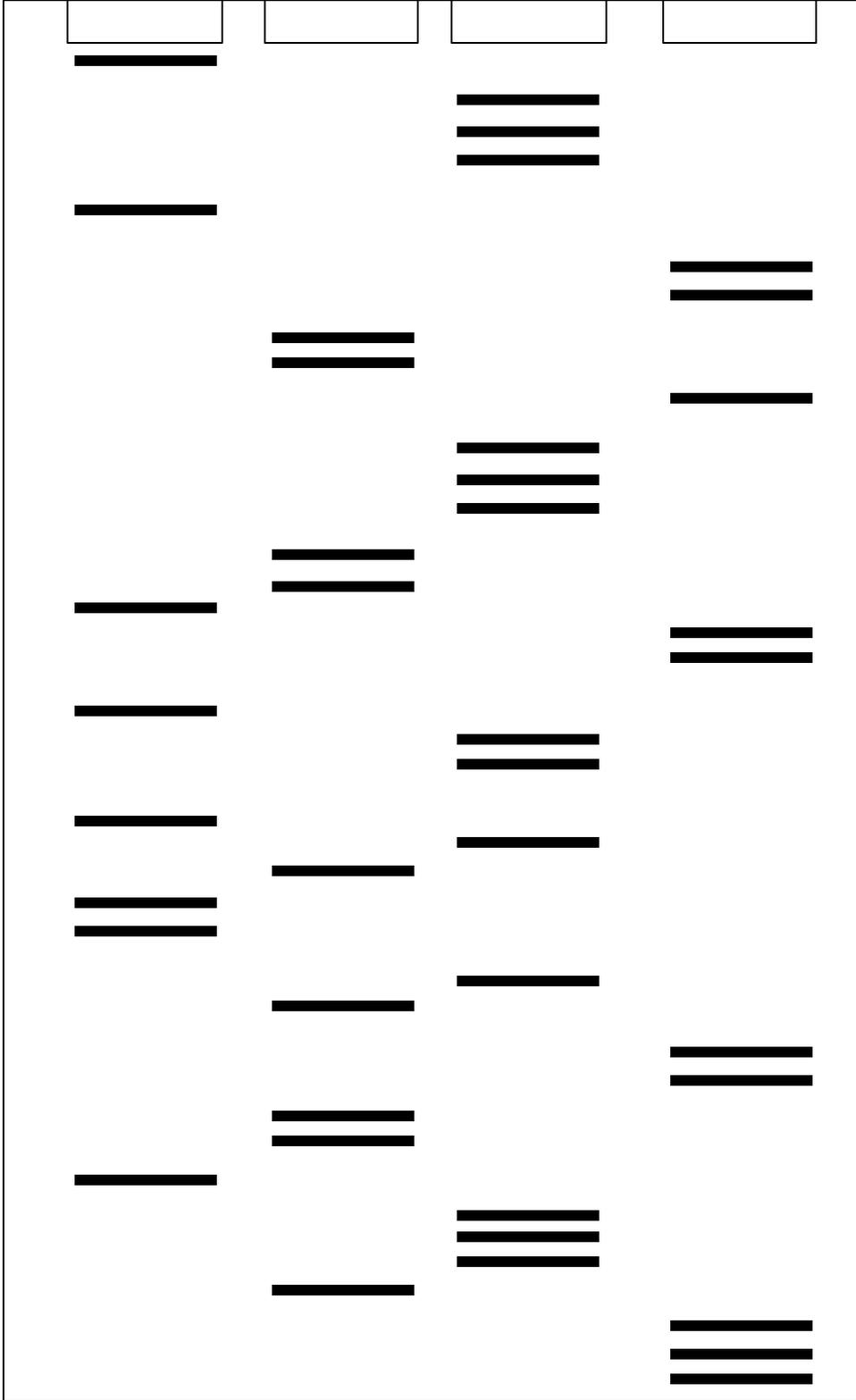
Y

A T G C

Z

A T G C

Sequencing Gels Key

<u>Flaviviridae</u>

A. Dengue type 1   gtctctacagcatcattccaggcacagaacgccaaaaaat

B. Dengue 4   gatcctgctgtctctgcaacatcaatccaggcacagagcg

C. Dengue 2   agaggcttgagaacactaatcctagcccccactagagtcg

D. Dengue 3   tccaaactac aacttgataa taatggatga ggctcatttc

E. Dengue 2 & 3   tcctgctgtctcctcagcatcattccaggcacagaacgcc
F. Dengue 2, 3,   agaccagagatcctgctgtctcctcagcatcattccaggc
G. Dengue 1, 3   gtagcagactagtggttagaggagacccctccca

H. Omsk   gcggatctggaagatgaagagcagatcctgcgttacatgg
I.   cccttggattcctgaatgaagaccactgggcatctagaga

J. Kyasanur forest gcagaccttgaggatgaggagcaaattctgcgttatctg
K.   cactgggtttcctcaatgaagaccactgggcttccaggga

<u>Bunyiviridae</u>

L. Congo-Crimean gcaaaacctgttctatcggcagtgttaatggctttgaaat
M. Rift valley   ctttctcatgcaccatcgtcctagtcacgaggttcgcttgc
N.   Hanta   ctggaatgacaatgcacatggtgttgggtcagtgcctatg

<u>Arenaviridae</u>

O. Machupo   gccaagctcatagctgacagcatcgacttcaaccaggttt
P. Lassa   gcccccgtgacccaccgccaataggcggtgggtcacgg
Q. Guanarito   tcacaaatacttgggagatcttgggacaacacaagtgt
R. Sabia   gcgcaccggg gatcctaggc gtttttagt cacgcttaaa
S. Junin   atggggcagttcattagcttcatgcaagagataccaactt

<u>Filoviridae</u>

T. Marburg   gtcaatgttggcgaacaatatcaacaactacgagaggc
  Ebola

U. Reston   cgatgaagattaaggcgacaacgagccgaaacttcatctc
V. Ziare   cgatgaagattaagccgacagtgagcgtaatcttcatctc

<u>Distracters</u>

W. West Nile   cgctttctggagttcgaagctctgggttttctcaatgaaga

X. Banzi virus   cgctttctggagtttgaggcactgggattcctcaatgaaga

Y. Lymphocytic   atgggtcaga ttgtgacaat gtttgaggct ttgcctcaca
choriomeningitis

Z. Yellow fever   ccctgggattcctgaatgaggaccattgggcttccaggga

Name _____ Sequence ID_____

MB425 Questions Set for Genomics and Sequencing:

1. To what four families do hemorrhagic fever viruses belong?

2. What two biochemical characteristics do all of the hemorrhagic viruses have in common?

3. What is the sequence of your virus?

4. What is the identity of your virus?

5. Can you determine what gene of the virus you have amplified and sequenced? If so please indicate what gene it is (you may not be able to determine this info).

6. Does this virus generally cause a hemorrhagic fever, rarely cause a hemorrhagic fever, or never cause a hemorrhagic fever?

7. Is this virus generally found in South America?

8. From what type of organisms could your virus be amplified (human, insect, rodent)?

9. Using the information from questions 6-8 from what sample type(s) (human, insect, rodent, or control) was your sample most likely amplified? Briefly explain how you made your decision.

10. If your virus causes hemorrhagic fevers is there generally a 20% or higher mortality rate (this data may not be available for a few viruses)?

11. Using the information from questions 6, 7, and 10, could your virus be the cause of the outbreak? Briefly explain how you made your decision.

11. By comparing the classes' isolates what is the most likely cause of this disease? Briefly explain how you made your decision.

Questions:

1. To what four families do hemorrhagic fever viruses belong?

Filioviridae, flaviviridae, bunyviridae, and arenaviridae

2. What two biochemical characteristics other than the symptoms they cause do all of the hemorrhagic viruses have in common?

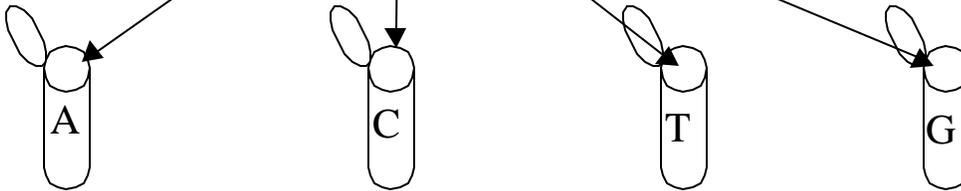They all have an RNA genome and are enveloped.

Sequencing and Reading Gels

5' ACGGTTCCAAATTTGGGGACTTTCCCCAA   3' DNA to be sequenced

1. Add primer, anneal
3' GGGGTT 5'

5' ACGGTTCCAAATTTGGGGACTTTCCCCAA   3'
                                GGGGTT

2. Put DNA annealed to primer in four tubes.

A   C   T   G

3. Add dNTPs,   Add dNTPs,   Add dNTPs,   Add dNTPs,
ddATP,   ddCTP,   ddTTP,   ddGTP,
and DNA pol.   and DNA pol.   and DNA pol.   and DNA pol.

Also add radiolabeled dCTP^{32} to each reaction. Note: ddNTP is a dideoxy NTP which has no OH group on the 3' end available for elongation; therefore, when one of these nucleotides are incorporated, elongation is terminated.

4. Run each of these four reactions on its own lane on a polyacrylamide gel.

A   T   G   C

T
G
C
C
A
A
A
G
G
T
T
T
A
A
A
C
C
C
C
T
G
A
A
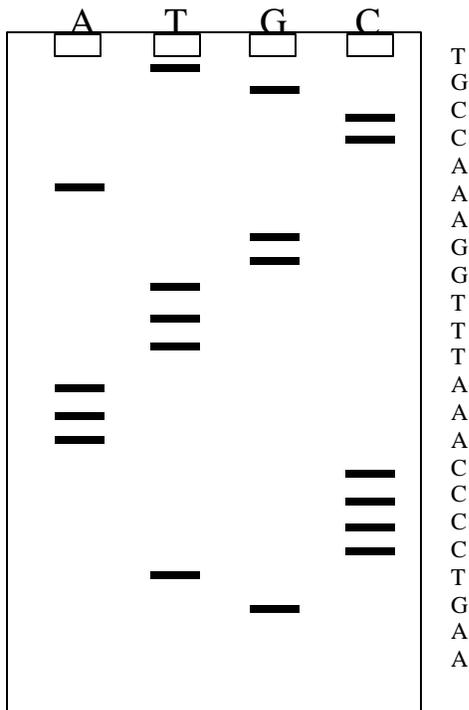
5. Put gel under film, read autoradiograph. Read from bottom to top is to the primer. (5' to 3'). The smaller the fragment the closer it is to the primer.