The Chinese University of Hong Kong
School of Life Sciences
Biochemistry Program
CUGEN Ltd.

DNA Forensic and Agarose Gel Electrophoresis [1]

Prof. Stephen K.W. Tsui, Dr. Patrick Law and Miss Fion Fong
Version Sept 2012

OBJECTIVES

After performing this experiment, students should be able to:

1. know that both variable number of tandem repeats (VNTR) and short tandem repeats (STR) are two kinds of genetic polymorphism;
2. explain how VNTR and STR can be employed for forensic DNA testing;
3. realize the power of polymerase chain reaction (PCR) in DNA analysis;
4. generate and analyze fingerprint from DNA samples; and
5. perform agarose gel electrophoresis which separates DNA molecules based on molecular weight differences.

A) Introduction

The Human Genome and Genetic Polymorphism

Genetic polymorphism refers to the differences in genome DNA sequences among individuals. Human genome is estimated to have 3 billion base pairs, with less than 5% codes for functional proteins (the functional genes). Most of the remaining DNA has no clear function and is called 'junk DNA'. 'Junk DNA' contains many repeat sequences. There are two kinds of repeat sequences, VNTR and STR, depending on the size of the repeat sequence.

One kind of repeat sequence, VNTR (variable number of tandem repeats), has its repetitive sequence elements arranged in tandem only at certain sites. They are a rich source of genetic polymorphism: the number of repeats in a given VNTR site varies widely in the human population. We can consider the VNTR site as a locus with multiple 'alleles' (various number of repeats) and use it in many applications such as pedigrees and forensic analysis.

STR (Short tandem repeat), another kind of repeat sequence, surrounds the chromosomal centromere (the structural centre of the chromosomes) with similar characteristics as VNTR. Different from VNTR, which has repeating unit of 9-100 base pairs long, STR has only 2-9 base pairs long in its repeating unit. Every person gets (inherits) one copy of a STR (allele) from each parent, which may be of same or different repeat size. Because the number of repeats of STR varies widely from person to person, STR has been proven a very effective tool in human identification, especially in forensic DNA testing.

In 19th century, a person was identified by blood group typing. With the advance of scientific technology, RFLP (Restriction Fragment Length Polymorphism) analysis was invented as the first DNA forensic tool, followed by PCR analysis. However, it is quite a challenging task in obtaining PCR amplification product from forensic samples because the problem of degraded or mixed DNA is common in criminal and civil cases such as in sexual assault case. STR has smaller size and is easier to be amplified with low mutation rates. Therefore, STR can serve as a good candidate of DNA markers in identifying victim, perpetrator, missing persons and others.

Currently, 13 internationally recognized STR loci are routinely used in STR typing test in the FBI Laboratory. These loci also include the core genetic loci for CODIS (Combined DNA Index System). They

[1] This manual is available for download at http://www.bch.cuhk.edu.hk/workshop-DNA/

are CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S818, D8S1179, D13S539, D18S51, D21S11 and AMEL.

An example of STR is a tandem repeat family called D8S1179, located on chromosomal 8 with a repeating sequence of 4 base pairs (TCTA) (Fig, 1). Every normal person should inherit one copy of STR (allele) from each of the paternal and maternal chromosome. An example of a D8S1179 locus is shown as below. There are 21 and 15 repeating units on the paternal and maternal chromosomes, respectively. Those repeating sequences are usually bracketed by regions of conserved sequences, called flanking sequences.

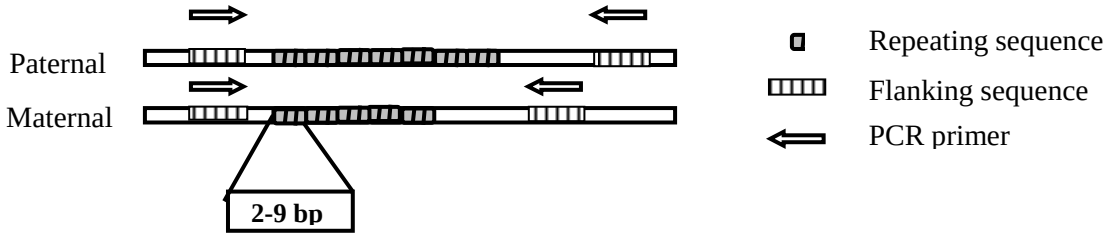
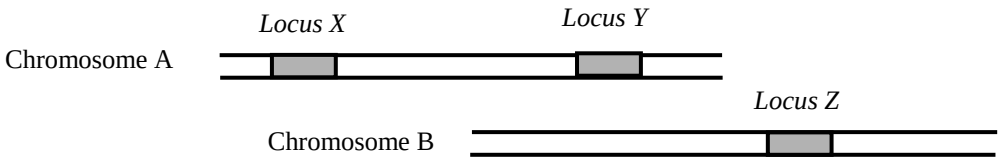


Fig. 1 Schematic structure of D8S1179

If we examine a particular STR locus for a criminal case, say the crime stain (crime scene evidence) has repeat numbers 9, 12, 13 (9, 12,13) and the victim has repeat pattern (12, 13), the offender should have the pattern: heterozygous (9, 12) or heterozygous (9, 13) or homozygous (9, 9) (Fig. 2). However, there is a chance that an unrelated individual carries the same STR repeat pattern in the population. To reduce the occurrence of false positive, we always examine several STR loci in the forensic DNA analysis. The example below illustrating the repeat numbers of three STR loci (X, Y, Z) determined from the victim, the crime stain and the suspects in a criminal case. Do you know who the real offender is?



Locus	X	Y	Z
Crime Stain	5, 11, 16, 17	9, 15, 14	8, 10, 13
Victim	11, 16	15, 14	10,13
Suspect 1	5, 17	9, 14	8, 8
Suspect 2	27, 15	22, 22	8,19

Fig. 2 Genotypes of locus X, Y and Z

Polymerase chain reaction

To detect the difference of the tandem repeat length at a given locus between individuals, we need to go through a series of procedures:

1) Collect DNA sample ;
2) Undergo polymerase chain reaction (PCR) to amplify the specific tandem repeat sequences;
3) Perform agarose gel electrophoresis to separate the amplified DNA sequences based on their sizes; and
4) Visualize the separated DNA pattern by staining dye.

Cheek cells, hair root follicle cells and blood cells are common sources for DNA extraction. DNA found at a crime scene might also include semen, tissue and body fluid from either victims or offenders. The DNA obtained from these tissue samples is in very minute quantity. We need to amplify the DNA regions of interests by PCR, a revolutionary technique developed by Dr. Kary B. Mullis in 1985 (Fig. 3). Dr. Mullis was awarded the Nobel Prize in Chemistry for this invention. The idea of PCR is simple. It starts with a reaction tube containing (1) the template DNA, (2) primers (short synthetic oligonucleotides that are complementary to the two ends of the DNA sequence to be amplified), (3) thermostable DNA polymerase and (4) free nucleotides.

The reaction only comprises three steps:

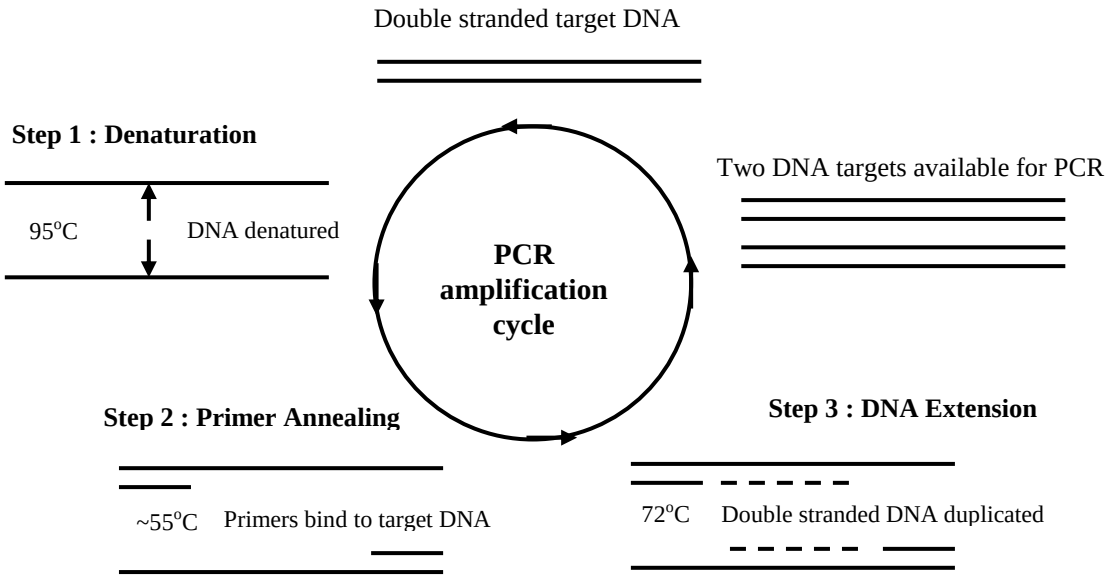


Fig. 3 Schematic diagram of the steps in PCR

These three steps together are called a cycle, which is repeated 20 to 40 times. DNA amount is doubled in each cycle. The amount of DNA obtained after a PCR reaction can be described by the formula

$$2^n \times C$$

where n = number of PCR cycles; and
C = the initial number of copies of DNA template present in the tube.
So, you will get 1,048,576 copies of DNA after 20 cycles of PCR reaction even you start with only one copy of DNA template initially. What a fantastic and powerful reaction!

Agarose Gel electrophoresis

DNA is negative in charge due to its sugar-phosphate backbone. Therefore, DNA will migrate towards the positive anode in the presence of an electric field (Fig. 4). Agarose is a chain of sugar molecules extracted from seaweed. If we dissolve agarose in boiling water and let it cool down, the sugar chains will cross-link with each other (a process called polymerization), creating 'pores' in between, and finally forms a semi-solid gel-like matrix. Those 'pores' present in agarose gel are very tiny and comparable in size to DNA molecules. Agarose gel can confer a sieving effect when DNA is passing through. Agarose gel electrophoresis is a process that separates DNA fragments by applying electricity to cause DNA passing through an agarose gel. DNA fragments shorter in size will move faster and longer DNA fragments will lag

behind. The distances moved by linear DNA molecules are inversely proportional to the $\log_{10}$ of their molecular size.

DNA molecule is colourless. A staining step is needed so that it can be visualized on an agarose gel. Ethidium bromide is a routine chemical used to stain DNA. Once it intercalated into the DNA molecule, it emits orange fluorescent when irradiated by ultraviolet light. However, ethidium bromide is carcinogenic and we will use SYBR safe, a much safer staining dye, in this experiment. SYBR safe staining solution is non-mutagenic and is safe to the environment. With this new SYBR safe staining solution, post overnight staining and destaining steps are no longer necessary. The SYBR stain will bind to the DNA in proportion with the DNA quantity. The DNA patterns can be visualized under the blue light box immediately after the electrophoresis and the time of whole experiment can be shortened to 3 hours. Under the blue light wavelength excitation, the SYBR bounded DNA will emit another wavelength at 530nm which is visible with the help of an amber filter. SYBR safe staining solution and blue light box with amber filter are available separately from CUGEN Ltd at a reasonable price.

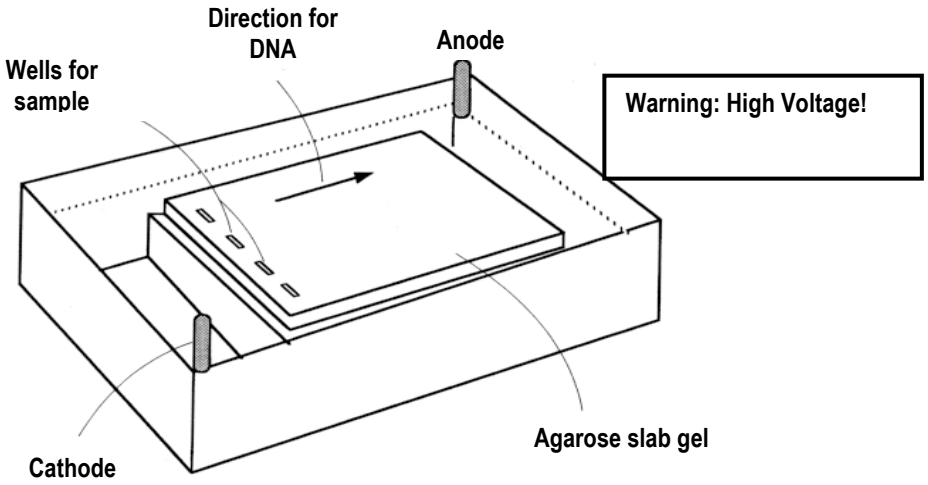


Fig 4. Agarose gel electrophoresis for separation of DNA fragments. [Adapted from D.W.S. Wong (1997) The ABCs of Gene Cloning. Chapman and Hall, pp 80]

Experiment

In this experiment, six STR DNA samples and one STR markers will be provided to perform the agarose gel electrophoresis. Each sample contains artificially generated DNA fragments that simulate tandem repeat containing DNA fragments amplified from three STR loci by PCR method. STR A is a STR locus for gender determination. STR B and C demonstrate the different repeat number of STR in the population of Hong Kong. DNA fragments will be resolved in the gel can be viewed as a fingerprint of a particular individual. Assuming in this experiment all samples come from a crime scene. After the DNA extraction and purification, we did the PCR amplification on the three loci A, B and C on those DNA samples prepared from the crime scene. You need to prepare an agarose gel for the electrophoresis of the STR samples and determine the possible offender in this crime case. You are reminded that this experiment is a simulated scenario of a crime case. In reality, the STR patterns would be analyzed using a capillary based genetic analyzer and the simulated DNA fragments in this experiment are resolved using an agarose gel instead. The corresponding repeat numbers of the DNA fragments used in this experiment are for demonstration only and do not reflect the actual size of the DNA fragments.

A) Materials

Apparatus

i. Electrophoresis units
ii. Power pack (80 -120V) (available separately from CUGEN Ltd)
iii. 1 ml syringe fitted with plastic tip
iv. Microwave oven
v. 250 ml conical flask
vi. Mini centrifuge (optional, available separately from CUGEN Ltd)
vii. 2-20 µl pipetman (optional, available separately from CUGEN Ltd)
viii. Blue light box with amber filter (available separately from CUGEN Ltd)

Reagents

DNA Forensic Kit – BOX A and B

i. 6X gel-loading buffer
(Note: You do not need to prepare the 6X gel-loading buffer. The DNA samples in Box A were already mixed with 6X gel loading. The buffer contains 0.2% bromophenol blue, 0.2% xylene cyanol, 60% glycerol in H_2O and 60mM EDTA. Bromophenol blue and xylene cyanol are tracking dyes. Glycerol can increase the density of the loading samples.)
ii. STR DNA Samples (DNA Forensic Kit – BOX A)
   a. STR-R
   b. V
   c. B
   d. S1
   e. S2
   f. S3
   g. S4
   h. S5

B) Procedures

i . Preparation of 1X TAE buffer (DNA Forensic Kit – BOX B)

To prepare 600ml of 1X TAE buffer, mix 60ml 10X TAE buffer with 540ml distilled water and mix well.
(Note: Working solution of 1X TAE buffer contains 0.04M Tris-acetate and 0.001M EDTA)

ii. Preparation of Agarose Gel for Electrophoresis of STR DNA

1. Prepare gel-casting unit by sealing both ends with tape. Make sure the tape is properly sealed to avoid leakage
2. Dissolve 0.9g agarose powder from Box B in 60ml 1X TAE buffer in a 250ml conical flask with a loose-fitting cap. The buffer should not occupy more than 50% of the volume of the flask. Heat the slurry in a microwave oven until the agarose dissolves (normally 800W for 1 to 3 mins is enough).
   (Caution: Make sure the agarose is dissolved completely before proceeding to the next step. The agarose solution can become superheated and may boil violently under prolonged heating.)
3. Cool the solution to 70^{o}C. Add 0.5ml SYBR safe staining solution to the melted agarose gel and mix gently. Pour the melted agarose gel immediately into the gel-casting unit placed horizontally on the bench. Insert the comb in one end of the casting unit and check for the presence of air bubbles underneath the comb. Remove and insert the comb again to remove any bubbles. Allow the gel to set at room temperature for at least 20 minutes.

iii. Gel Electrophoresis of STR DNA

1. While waiting the gel to set, check the gel tank for leakage by adding 300ml water and observe the tank for leakage for a few minutes. Discard the water if no leakage is found. (Note: If there is a leakage, discard the water and use chloroform to repair)
2. After the gel is completely set, remove the comb and the sealing tape carefully, and mount the gel in the electrophoresis tank with the wells close to the cathode side.
3. Mix 200ml of 1 X TAE buffer with 0.5ml SYBR safe staining solution.
4. Pour the 1X TAE buffer with SYBR staining safe staining solution to the gel tank to cover the gel with about 1 mm depth. Add additional 1X TAE buffer if necessary. *(Note: It is important to use the same batch of electrophoresis buffer in both the electrophoresis tank and the gel.)*
5. Spin down the STR DNA samples using a mini centrifuge or swing the DNA samples to the bottom using your hands. Assembly the 1 ml syringe with the plastic tip. Slowly take and load all the DNA (~20ul) samples into the wells of the gel using the syringe pipette (avoid puncture the wells) according to the sequence in Fig. 4. Wash the syringe pipette with 1X TAE buffer between successive loading to avoid cross contamination. STR-R is the STR references for STR A, B and C. Alternatively, you may load the samples using a 20µl pipette (available from CUGEN Ltd).
6. Close the lid of the gel tank and attach the electrical leads.
7. Apply a voltage of about 80-120V. If leads have been attached correctly, bubbles should be generated at the electrodes and the DNA will migrate toward the anode (red lead) side. (**Remarks: remember to check for the generation of bubbles at both cathode and anode ends to make sure the electrophoresis is undergoing)**.
8. During running you will observe the movement of two dye fronts. Turn off the power supply and remove the gel from the gel tank until the lower dye front bromophenol blue (purple in color) migrated two-third of the gel (depending on the voltage, around 45 min to 1.5 hr is required).
9. After electrophoresis, secure the blue light box inside the Perspex stage and place the gel on the blue light box. Put the amber filter (amber in color) on top of the stage. Turn off the light and turn on the blue light box. Observe the DNA patterns and take pictures simply using an iPhone without the flash. (**Remarks: The strong blue light is irritating to eyes and please don't look at the blue light directly without the amber filter. The Blue light box will generate a lot of heat and don't leave it unattended for a long period of time. Turn it off immediately after use**.)

C) Expected Results

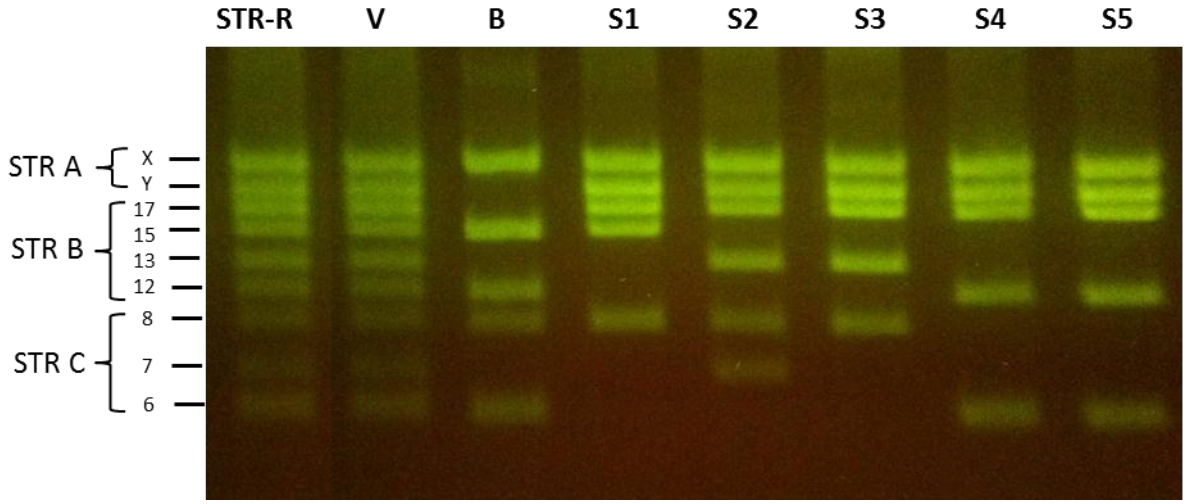
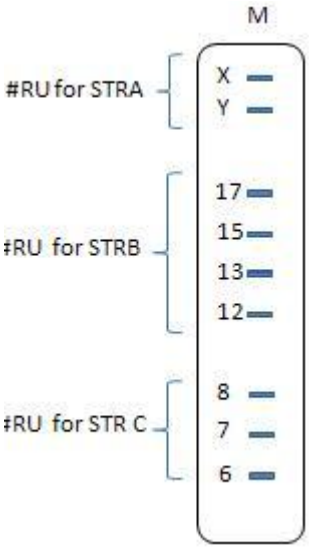


Fig 4

STR R: STR references of STRA, B and C
V: Vaginal swab from the victim
B: Buccal swab from the victim
S1: Buccal swab from suspect 1
S2: Buccal swab from suspect 2
S3: Buccal swab from suspect 3
S4: Buccal swab from suspect 4
S5: Buccal swab from suspect 5

Question 1 : Table of STR typing analysis

Each band represents certain number of repeating units (allele) for each STR locus. Assume that the alleles for the three different loci do not overlap and the sizes of the repeat units are the same.



Analyze the fingerprint patterns on your gel and complete the table. Indicate the gender (F for female, M for male), genotype (#RU, the Number of Repeating Unit at a STR locus). Please specify whether the genotype of each sample is either heterozygous or homozygous as well. The first row is completed for you.

	STR –A Gender	STR- B		STR-C	
Sample	F/M	Genotype	Heterozygous/homozygous	Genotype	Heterozygous/homozygous
V	F and M	12, 13, 15, 17	NA	6, 7, 8	NA
B					
S1					
S2					
S3					
S4					
S5					

Question 2 : Who is the most possible offender in this criminal case? Please give explanations.

S1: _____ (innocent / guilty)

S2: _____ (innocent / guilty)

S3: _____ (innocent / guilty)

S4: _____ (innocent / guilty)

S5: _____ (innocent / guilty)

Question 3 : How could we interpret the gender using the fingerprints from STR A?

Question 4 : Why do some STR typing results at a locus show a single band while some show two bands?

Question 5 : We usually come across the expected frequency of a certain genotype at a locus in reality. It is very useful to calculate the probability of finding the same STR pattern in the population of interest. Taking STR B as an example, if the expected frequency of the genotype (12,12) is 1%, it means that 1 in 100 persons in the population would carry the same genotype (12,12). Given that the male offender's expected frequency of genotypes at STR-B and STR-C be 1% and 0.5% respectively. What is the probability that the population of interest (say Chinese) carry the same genotypes as the offender's?

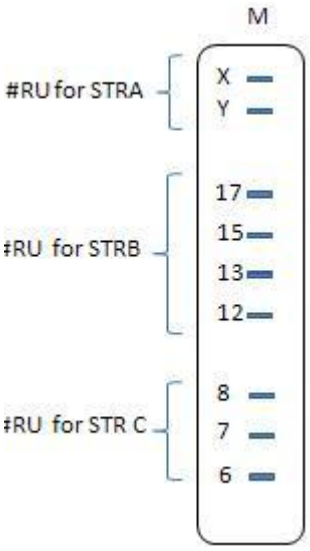
Question 6 : What is the limitation of STR analysis?

Question 7: What you can conclude if suspect 4 and 5 are come from the same family.

Suggested answers:

Question 1 : Table of STR typing analysis

Each band represents certain number of repeating units (allele) for each STR locus. Assume that the alleles for the three different loci do not overlap and the sizes of the repeat units are the same.



Analyze the fingerprint patterns on your gel and complete the table. Indicate the gender (F for female, M for male), genotype (#RU, the Number of Repeating Unit at a STR locus). Please specify whether the genotype of each sample is either heterozygous or homozygous as well. The first row is completed for you.

	STR –A Gender	STR- B		STR-C	
Sample	F/M	Genotype	Heterozygous/homozygous	Genotype	Heterozygous/homozygous
V	F and M	12, 13, 15, 17	***	6, 7, 8	***
B	F	12, 15	Heterozygous	6, 8	Heterozygous
S1	M	15, 17	Heterozygous	8	Homozygous
S2	M	13, 17	Heterozygous	7, 8	Heterozygous
S3	M	13, 17	Heterozygous	8	Homozygous
S4	M	12, 17	Heterozygous	6	Homozygous
S5	M	12, 17	Heterozygous	6	Homozygous

Question 2 : Who is the most possible offender in this criminal case? Please explains the answer:

S1: _____ (innocent / guilty)

S2: _____ (innocent / guilty)

S3: _____ (innocent / guilty)

S4: _____ (innocent / guilty)

S5: _____ (innocent / guilty)

Explanation:

At STR A, sample from the vaginal swab showed the presence of Y-chromosome. Since the victim did not have Y-chromosome, the offender must be a paternally inherited male carrying Y-DNA.

At STR B, sample from vaginal swab (criminal scene evidence) presents the allele of 12, 13, 15, 17 while buccal swab from the victim has the genotype (12, 15) (carrying the allele 12 and 15 only). Hence, the offender's genotype of STR B must be (13, 17). Hence, samples from Suspect 2 and Suspect 3 matched the offender's genotype of STR B and Suspect 1,4,5 can be excluded. Next, when you look at STR C, only Suspect 2 carries the additional allele (7) identified in the vaginal swab. Thus, Suspect 1, Suspect 3, Suspect4 and Suspect5 are excluded and only Suspect 2 is the possible offender in this crime case.

Question 3 : How could we interpret the gender using the fingerprints from STR A?

Explanation:

The gender can be determined by STR A, which test the presence of fragment existing in Y-chromosome. If the genotype of STR A of a sample is (X, Y), the individual must be a paternally inherited male carrying Y-DNA. STR A is used for distinguishing males (X, Y) and females (X, X).

Question 4 : Why do some STR typing results at a locus show a single band while some show two bands?

Explanation:

Some STR typing results show a single band because the genotype of a particular STR is homozygous. It means that the person gets a copy of STR with the same number of repeating units (same repeat size) from each parent. When the genotype of STR is heterozygous, the person gets a copy of STR with different repeat sizes. Hence, the STR result shows a double band.

Question 5 :

We usually come across the expected frequency of a certain genotype at a locus in reality. It is very useful to calculate the probability of finding the same STR pattern in the population of interest. Taking STR B as an example, if the expected frequency of the genotype (12,12) is 1%, it means that 1 in 100 persons in the population would carry the same genotype (12,12). Given that the male offender's expected frequency of genotypes at STR-B and STR-C be 1% and 0.5% respectively. What is the probability that the population of interest (say Chinese) carry the same genotypes as the offender's?

Explanation:

The Probability of carrying the same genotype at STR-B and STR-C is 0.0005 (0.01 * 0.005). That is, the frequency of offender's DNA profile would be 1 in 20000 Chinese.

Question 6 : What is the limitation of STR analysis?

Answer: Cannot identify identical twins.

Question 7: What you can conclude if suspect 4 and 5 are come from the same family.

Answer: They are identical twins