INTERNATIONAL AVICENNA OLYMPIAD IN BIOLOGY

UZBEKISTAN 2024

EXAM PAPER

PRACTICAL BIOINFORMATICS

TOTAL POINT: 50

General Instructions:

You have 90 minutes to complete **TWO** tasks in this practical exam. You can do the tasks in **ANY ORDER**.

- **Mitochondrial genome – 35 points;**
- **Medical genetics – 15 points.**

Important notes:

- No answers on the exam paper will be graded. For your answers use Answer sheet **ONLY**.
- It is strictly forbidden to look at the computer screen and/or work of another participant and talk to other participant. Otherwise, it leads to disqualification from the Olympiad.
- Using other tools or going to sites other than those given in the exam paper leads to disqualification from the Olympiad.
- After finishing, close all the web pages, **ERASE** the history of the browser, delete any temporary files.

Good luck!

Useful information:

Ctrl+A – select the entire sequence;

Ctrl+C – copy the specified sequence;

Ctrl+V – paste the copied sequence.

Copy the screenshot - press the PrtScr key and select the desired location by clicking the left mouse button.

Mouse select + Shift + Mouse select – selects the region.

1ST TASK: MITOCHONDRIAL GENOME

Mitochondria are organelles found in eukaryotic cells. Mitochondria is a semi-autonomous cell organelle surrounded by a membrane that produces a large part of the chemical energy necessary for the biochemical reactions of the cell. This practical assignment is dedicated to the analysis of fish mitochondrial DNA using bioinformatics methods and tools.

1. In the IBO folder, you are given the complete mitochondrial genome of an unknown fish, named **Fish_mtDNA**. Open this file using **SnapGene Viewer**. Study the mtDNA map carefully. It shows open reading frames, tRNA, and rRNA genes. Observe the nucleotide/amino acid sequences of cytochrome oxidase (COXI) and NADH dehydrogenase subunit 1 (Nd1), and ATPase 8 open reading frames (ORFs). To see the nucleotide sequence of the gene, as well as the amino acid sequence of the encoded protein, click on the **'sequence'** tab in the bottom left.

Check the statements below and mark 'X' if they are true or false, respectively. Maximum 5 points for correct answers. Write your answers into **Q1** section.

Q1.	Statement	True	False
1.1.	Unlike human mitochondrial DNA, fish mitochondrial DNA is circular.		
1.2.	The protein-coding ORFs of the mitochondrial DNA of this fish contain introns.		
1.3.	There are overlapping ORFs in fish mtDNA.		
1.4.	Methionyl-tRNA initiates the translation of all proteins encoded in a given mtDNA.		
1.5.	All protein-coding ORFs are encoded in only one of the mtDNA strands.		
1.6.	Each mitochondrial ribosomes contain fewer RNA molecules than cytoplasmic ones (assume no cytoplasmic RNAs enter the given mitochondrion).		
1.7.	The copy number of protein-coding genes encoded in mitochondrial DNA is greater than that of genes encoded in the nuclear genome.		
1.8.	Mitochondrial DNA encodes all mitochondrial proteins.		
1.9.	The genes in the mitochondria are polycistronic.		
1.10.	AT and GC percentage is equal in given mtDNA.		

2. Find the ORF coordinates (start and stop nucleotide positions) for the following genes given below. Maximum 3 points for correct answers. Write your answers into **Q2** section.

Q2.	**Genes**	**Start position**	**Stop position**
2.1.	*COX I*		
2.2.	*COX II*		
2.3.	*COX III*		
2.4.	*12 S rRNA*		
2.5.	*16 S rRNA*		
2.6.	*Cytb*		

3. Denatured circular double-stranded DNAs, including plasmids and mitochondrial genomes, can be separated into **Heavy** (H) chain and **Light** (L) chain using CsCl gradient. The two strands have different masses due to different distribution of nucleotides between the strands. As the result, the template strands for the genes can be distinguished into Heavy (H) chain and Light (L) chain. For example, **Cytochrome c oxidase** subunits are synthesized from the heavy chain.

Check the statements below and mark 'X' if they are true or false, respectively. Write your answers into **Q3** section. Maximum 4 points for correct answers.

Q3.	**Statements**	**True**	**False**
3.1.	The percentage of pyrimidines is higher in the heavy strand than the light one.		
3.2.	Cyt b mRNA is transcribed from light chain.		
3.3.	Light chain is the coding strand for Nd 6.		
3.4.	Heavy chain is template for the most of the tRNAs		

4. Restriction mapping maybe used to identify plasmids. Researchers have created a restriction map of the given and several other mitochondrial genomes. For this, in the **first set**, the mixture of *Pme*I + *Xho*I + *Sac*I enzymes; in the **second set**, the mixture of NdeI and *Dra*III enzymes were used and the resulting restriction products were identified by gel electrophoresis.

In this task, based on gel electrophoresis picture (Figure **1**), determine which lane(s) belong to the given mitochondrial genome. To determine restriction reaction products, press Ctrl+E in **SnapGene viewer** and remove all enzyme and then select the given enzymes, and analyze the expected restriction products for both sets accordingly. Finally, mark the lane(s) with 'X' that belongs to your target genome in the **Q4**. Maximum 4 points for the correct answer.

Q4.	**Lanes**							
Mark the lane(s) with 'X' that belongs to your target genome	1	2	3	4	5	6	7	8

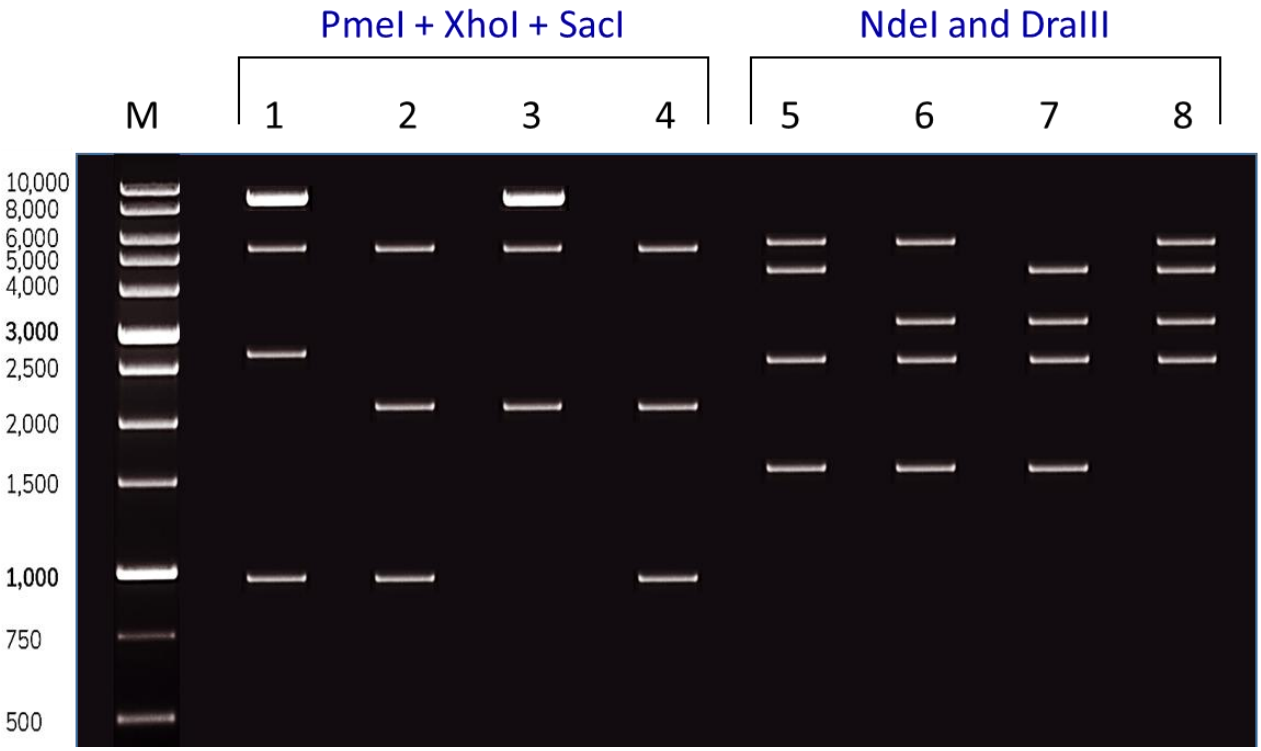

Figure **1.** Restriction analyis of different types of DNAs with 1-4: *Pme*I+*Xho*I+*Sac*I and 5-8: NdeI and *Dra*III.

5. Cytochrome c oxidase subunit I (*COXI*) gene is widely used to study phylogeny. For this purpose, mitochondrial DNA is first isolated and amplified using specially designed primers for the *COXI* gene, and then the gene is sequenced by Sanger sequencing and aligned with other available corresponding sequences.

In this assignment, you need to design **forward** and **reverse** primers of 20 bp each for Polymerase Chain Reaction (PCR) amplification of the *COXI* gene. When creating a primer, cover the entire protein coding region (except for the stop codon). Enter your answers in **Q5**. Maximum 2 points for the correct answer.

Important: primer sequences must be entered in 5' -- > 3' direction.

Q5.	**Questions**	**Sequences**
5.1.	Forward primer sequence 5' -- > 3'	
5.2.	Reverse primer sequence 5' -- > 3'	

6. It is important to check that the primers only bind to the target region and do not bind to another part of the fish mitochondrial genome, and to check the probability of non-target products being formed by showing complementarity with the genome of the PCR test conductor (human), because assume that the samples prepared for PCR have a possibility of contamination with the human genome.

To check this, access the NCBI Primer blast program via the link provided: https://www.ncbi.nlm.nih.gov/tools/primer-blast/ and place your designed forward and reverse primers in the appropriate locations. In the PCR template section, load the **Fish_mtDNA** file or enter the sequence from it. Adjust the primer pair specificity checking parameters as shown in the image below (Figure **2**). Select 'RefSeq representative genomes' for Database type. Then click the 'Get primers' button.

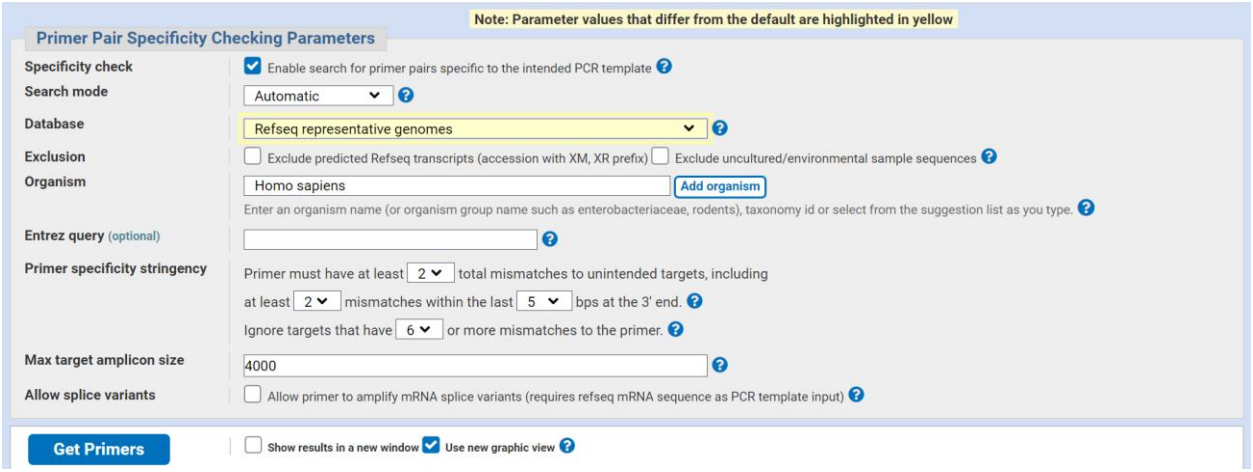

Figure **2.** Parameters for Primer Specificity Checking.

Enter the following parameters in **Q6** after the results are displayed. Maximum 4 points for the correct answer.

Q6.	**Questions**	**Answers**
6.1.	Screen copy and paste the graphical representation of primer pairs.	
6.2.	Length of target PCR product to be generated (bp)	
6.3.	1- Non-target product length (bp) that can be generated	
6.4.	2- Non-target product length (bp) that can be generated	
6.5.	3- Non-target product length (bp) that can be generated	
6.6.	4- Non-target product length (bp) that can be generated	

7. In this task you need to fill the missing data (A-D) for Taq polymerase (elongation speed 1 kb/min) based endpoint PCR conditions (given below in Figure **3**) according to the melting temperature (Tm) value of your primers, PCR product length and your prior knowledge. Write your answers into **Q7** section. Maximum 3 points for correct answers.

There are several formulas available to calculate Tm value. However, we ask you to use the given two following formulas.

For sequences less than 14 nucleotides the formula is:

$$\mathbf{Tm= (A+T) * 2 + (G+C) * 4}$$

For sequences longer than 13 nucleotides, the equation used is

$$\mathbf{Tm= 64.9 +41*(G+C-16.4)/(A+T+G+C)}$$

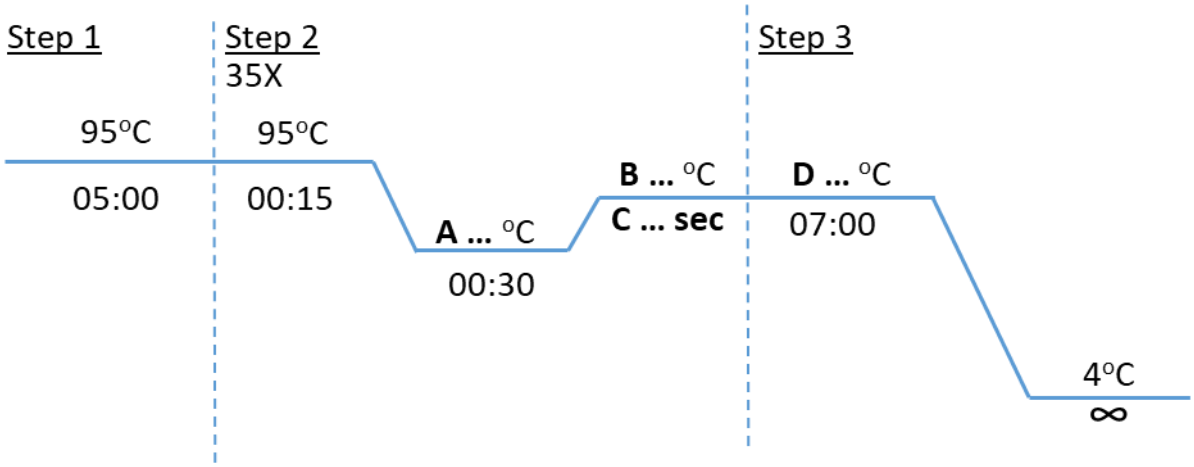

Figure **3.** Incomplete Endpoint PCR reaction conditions.

Q7.	**Missing details**	**Answers**
7.1.	A … °C	
7.2.	B … °C	
7.3.	C … sec	
7.4.	D … °C	

8. To identify the species of fish based on this gene sequence enter to the BOLD platform via the link https://www.boldsystems.org/index.php and from the **Identification section, select Animal identification --> Species level barcode records.** Copy the PCR product from the SnapGene viewer, paste it in the appropriate space on the site and press the submit button. The result of barcoding comes out within 30-60 seconds. Write the species the given mtDNA belongs to in the **Q8** section. Maximum 2 point for the corrects answer.

Q8.	Question	Species name
8.1.	The given mtDNA belongs to	

9. Based on the protein sequence in the **Nd1 protein** coding region of the mitochondria of this fish, the scientists wanted to amplify this gene of other fish by PCR. In such cases, a set of **degenerate** primers is created taking into account possible differences in the gene sequence. A degenerate primer is constructed when the protein sequence is known but the DNA sequence is unknown, covering all possible codon bases. As a result, during the PCR reaction, one of these primers binds to the template DNA and a PCR product is formed.

In this task, you are requested to create forward degenerate primer covering the first 5 amino acid codons and reverse degenerative primer covering last 4 amino acid codons + stop codon of the **Nd1 protein** and calculate how many primers types exist in the primer pool for each primer separately. Write your answers into **Q9** section. Maximum 8 points for correct answers.

Important 1: A degenerate primer designed using standard codons is provided as an example to give you direction. Also, for your convenience, you are provided with conventional symbols of nucleotides (Figure **4**) and a table of genetic code of vertebrate mitochondria (Figure **5**).

Important 2: Primer sequences should be presented in 5’ -- > 3’ direction.

Example:

Amino acid sequence	Met Lys Gly Asn Phe
Forward degenerate primer 5’ -- > 3’	ATG AAR GGN AAY TTY
Number of primer types	2*4*2*2=32

IUPAC nt code	Base	IUPAC nt code	Base
A	Adenine	**K**	G or T
C	Cytosine	**M**	A or C
G	Guanine	**B**	C or G or T
T (or U)	Thymine (or Uracil)	**D**	A or G or T
R	A or G	**H**	A or C or T
Y	C or T	**V**	A or C or G
S	G or C	**N**	any base
W	A or T	**. or -**	Gap

Figure **4.** Conventional symbols of nucleotides.

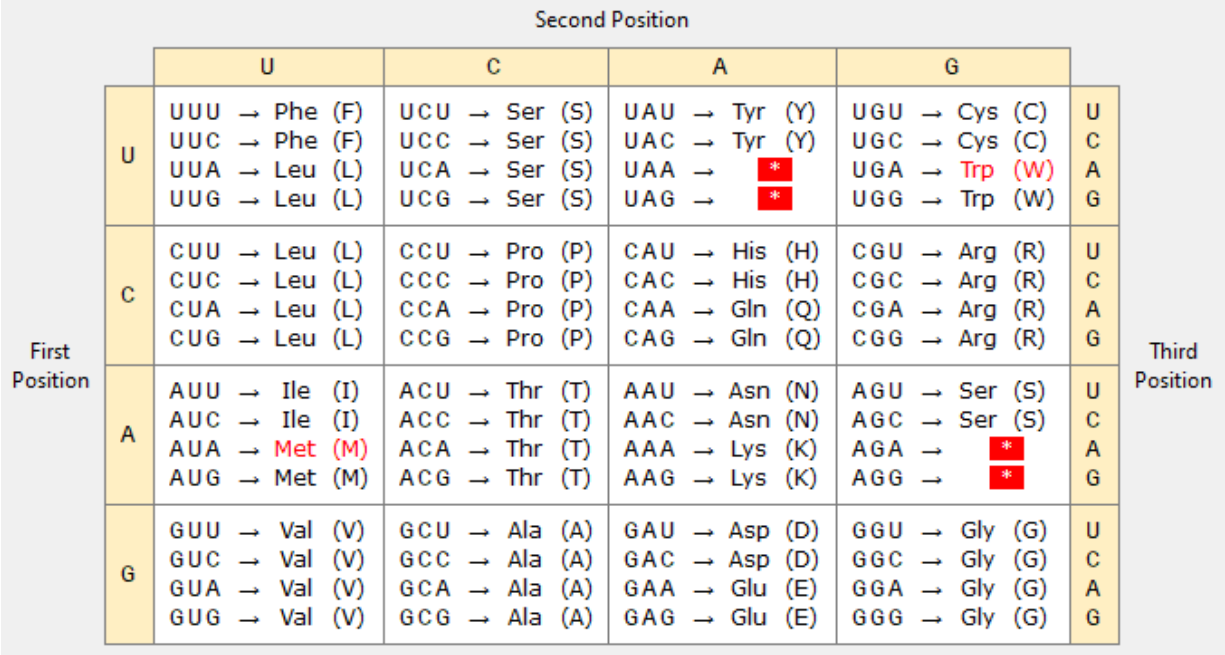

Figure **5.** Vertebrate Mitochondria Genetic Code Chart.

Q9.	Questions	Answers
9.1.	Sequence of the Forward degenerate primer in 5’ -- > 3’ direction	
9.2.	How many Forward primer types exist in the pool?	
9.3.	Sequence of the Reverse degenerate primer in 5’ -- > 3’ direction	
9.4.	How many Reverse primer types exist in the pool?	

2ND TASK: MEDICAL GENETICS

Angiotensin-converting enzyme 2 (ACE2) is an integral part of the renin-angiotensin-aldosterone system (RAAS), which is important for controlling the body's blood pressure. Membrane ACE2 also serves as a cellular entry point for some coronaviruses, such as HCoV-NL63, SARS-CoV, and SARS-CoV-2. For this reason, mutations in this gene may affect the severity of the COVID-19 infection.

In this work, you need to identify changes in the ACE-2 protein/gene, create allele-specific PCR primers to detect the mutation in the ACE-2 gene and perform some other relevant analysis.

2.1. To identify mutations in ACE-2 protein, In the IBO folder, you are given the ACE-2 protein sequences of 6 individuals named Samples 1-6 and the ACE-2 reference protein sequence in the **ACE-2.text** file. First, open Clustal Multisequence Alignment: https://www.ebi.ac.uk/jdispatcher/msa/clustalo and upload the file or copy paste the given sequences at once into corresponding space and click the **'run'** button. View the results. Finally, carefully check the aligned sample sequences by comparing to the reference protein. Divide the samples into Wild Type (WT) and Mutant (M) genotypes by marking with 'X'. Write your answers into **Q2.1.** section. Maximum 3 points for the correct answer.

Important: In results, sequence consensus will be visible as '*'.

Q2.1.	Samples	WT	M
2.1.1.	Sample1		
2.1.2.	Sample2		
2.1.3.	Sample3		
2.1.4.	Sample4		
2.1.5.	Sample5		
2.1.6.	Sample6		

2.2. Scientists decided to develop an allele-specific PCR (ARMS PCR) test system to determine the frequency of one of such mutations in the population. Typically, for the development of this type of PCR test system, in addition to the usual 2 types: forward and reverse primers, additional 2 types of primers are used to distinguish WT and M variants (see the Figure **6**). As a result, depending on the type of allele, in addition to external forward and reverse primer products, additional PCR products specific to WT and/or M are formed. As a result, alleles are easily identified.

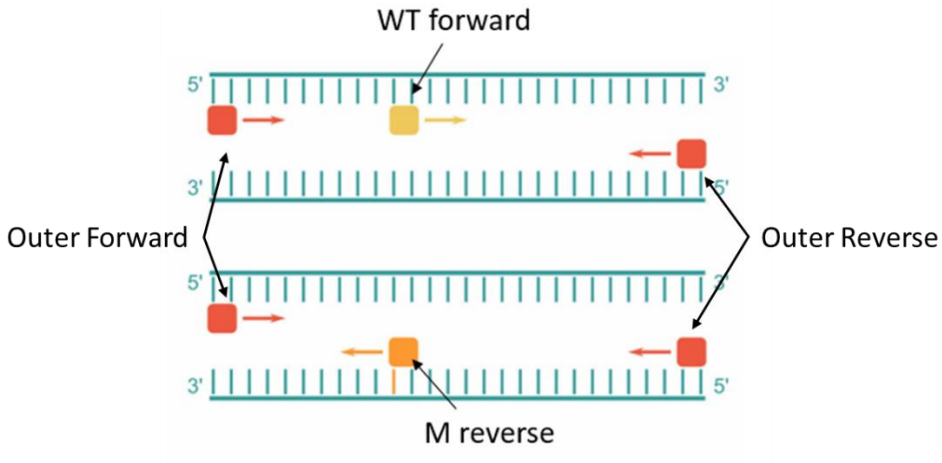

Figure **6.** ARMS-PCR primers.

2.2.1. First, using the Blast program https://blast.ncbi.nlm.nih.gov/Blast.cgi#alnHdr_Query_35679 compare the WT and M variants given in the **ARMS.txt** file and record the mutation position and observable allele genotypes corresponding to **Q2.2.1.** Maximum 4 points for the correct answer.

Q2.2.1.	**Questions**	**Answers**
2.2.1.1.	Mutation position	
2.2.1.2.	WT Homozygote allele	
2.2.1.3.	Heterozygote allele	
2.2.1.4.	M Homozygote allele	

2.2.2. Now, to identify alleles using ARMS PCR, you need to create **WT forward and M reverse** primers and determine the length of the PCR products generated in different alleles.

To do this, open the provided **ARMS_WT** and **ARMS_M** files in the SnapGene viewer. Ensure that outer primers are already included to cover outer the regions where the Single nucleotide polymorphism (SNP) occurs, and design a 22 bp WT forward and a 26 bp M reverse primer with the SNP at the 3' end of the primers and finally determine the length of PCR products for all possible alleles. Enter your answers to **Q2.2.2**. Maximum 5 points for the correct answer.

Q2.2.2.	**Questions**	**Answers**
2.2.2.1.	WT forward primer sequence 5’ -- > 3’ .	
2.2.2.2.	M reverse primer sequence 5’ -- > 3’ .	
2.2.2.3.	Expected PCR product sizes in bp for WT Homozygote allele	
2.2.2.4.	Expected PCR product sizes in bp for Heterozygote allele	
2.2.2.5.	Expected PCR product sizes in bp for M. Homozygote	

2.3. Mutations in ACE-2 may affect the entry of SARS-CoV-2 into the cells and have an effect on the severity of COVID-19 infection. Based on this, check the statements below and mark ‘X’ if they are true or false, respectively. Write your answers into **Q2.3.** section. Maximum 3 points for the correct answer.

Q2.3.	**Statements**	**True**	**False**
2.3.1.	If the mutation increases the stability of ACE-2 protein then the chance of SARS-CoV-2 entry into the cells will be lower.		
2.3.2	If the mutation is not in the SARS-CoV-2 virus interaction site than the mutation does not affect severity of COVID-19.		
2.3.3.	If the mutation increases the affinity for receptor binding domain of SARS-CoV-2, the chance of the virus entry into the cell will increase		

THE END