I. Choose ONE most appropriate answer (51%)

1. Which pair of the following genes is "orthologue"?
   a. human hemoglobin α and human hemoglobin β
   b. human hemoglobin α and horse hemoglobin β
   c. human hemoglobin α and horse hemoglobin α
   d. human hemoglobin α and human trypsin
   e. human hemoglobin β and human trypsin

2. tRNAs
   a. commonly contain modified bases
   b. at lease 61 different tRNAs, one for each codon, are required for an organism.
   c. are approximately 250-300 nucleotide in length
   d. exhibit a random coil structure
   e. all of the above statements are correct

3. A bacterial coding sequence of 300 nucleotides is capable of coding for a protein of how many amino acids?
   a. 3   b. 50   c. 100   d. 300   e. > 300

4. _____ is the transfer of normal genes into body cells to correct a genetic defect.
   a. reverse transcription   b. nucleic acid hybridization   c. gene mutation   d. gene therapy
   e. nucleotide sequencing

5. About protein structure folding and modeling
   a. Ramachandran plot is used to describe allowed mainchain conformation
   b. Amino acid side chains offer the physicochemical properties to generate protein folding pattern
   c. The SCOP is a database that organizes protein structures according to evolutionary origin and structural similarity
   d. The ROSETTA is a computer program that predicts protein structure from primary sequence by assimilating information from known structure
   e. All of the above statements are correct

6. DNA replication results in
   a. two double-stranded DNA, one contains two new DNA strands and the other is the parental molecule
   b. two double-stranded molecules, both contain one newly made strand and one old strand
   c. three double-stranded molecules, one with new strands and two that are discarded
   d. three single-stranded molecules, two are newly made and one is discarded
   e. None of the above

7. The "BLAST" is a commonly used computer tool for
   a. comparing DNA or protein sequence similarity
   b. searching for published medical article database
   c. simulating protein tertiary structure
   d. predicting the isoelectric point of a gene product
   e. predicting restriction enzyme cutting site and draw the map.

8. Which of the following descriptions about the genome size and gene number is NOT correct.
   a. E. coli K12 genome is about 4 million base pairs.
   b. The total gene number of E. coli K12 is about 4000-5000 genes.
   c. Human haploid genome is about 3×10^9 base pairs
   d. Human has 4×10^6 different genes
   e. The smallest cell such as Mycoplasma genitalium contains ~500 genes

9. Which one of the following descriptions about polymerase chain reaction (PCR) is NOT correct?
   a. it relies on a thermoresistant DNA polymerase
   b. it can amplify a DNA region specifically
   c. the specificity of PCR amplification depends on a pair of oligonucleotide primers
   d. it requires 3 reaction steps: annealing, polymerization, and denature
   e. after 6 PCR cycles, the target region can theoretically be amplified 12 folds

10. Which of the following statements about single nucleotide polymorphisms (SNPs) is NOT correct?
    a. are useful markers for genome mapping
    b. occur once every 1-2 kb in human genome
    c. are useful for establishing parent-progeny relationship
    d. defined by their SNPs, every human being is unique
    e. most SNPs cause severe diseases in humans.

11. Aminoacyl-tRNA synthetases
    a. are responsible for translocating aminoacyl-tRNA from nucleus to cytoplasm
    b. in an organism, isoacceptor tRNAs such as tRNA-Arg(AGA) and tRNA-Arg(AGG), are normally recognized by distinct aminoacyl-tRNA synthetases
    c. in general charging one kind of amino acid requires one aminoacyl-tRNA synthetase
    d. there are at least 61 different aminoacyl-tRNA synthetases in the *E. coli* genome
    e. the enzymes are essential for eucaryotes but not procaryotes

12. Which statements about programmed frameshifting translation is NOT correct?
    a. It can be caused by some mutant tRNA suppressors recognize a "codon" for 4 bases
    b. It can occur at the slippery sequences that allows a tRNA to shift by 1 base after it has paired with its anticodon
    c. It always leads to apoptosis
    d. It is common in RNA viruses that synthesize polyproteins
    e. Frameshifting at a stop codon causes readthrough of the protein

13. Which statements about transcription termination process in *E. coli* is NOT correct?
    a. Intrinsic terminators consist of a AT rich hairpin in the RNA product followed by a U-rich region in which termination occurs.
    b. Approximately half of the genes in *E. coli* have intrinsic terminators.
    c. The rho factor is a terminator protein that binds to nascent RNA and translocates to a sequence that rich in C and poor in G residues preceding the actual termination sit in the DNA
    d. Rho factor contains a helicase activity
    e. Anti-termination proteins such as lambda phage pN and pQ can act on RNA polymerase to cause it readthrough a specific terminator

14. Restriction enzyme *EcoR*I recognizes a sequence of 6 base pair in length. How many cleaved fragments do you anticipate when you use the enzyme to completely digest a 100-kb DNA fragment of random sequence?
    a. 4   b. 10   c. 40   d. 100   e. 400.

15. An open reading frame
    a. starts with UGA codon most frequently
    b. continues in nucleotide triplet to transcription terminator
    c. normally refers to the region that can code for protein
    d. is commonly found in many pseudogenes
    e. all of above are correct

16. Which of the following statements about "sequence divergence" is NOT correct?
    a. It is the basis for the evolutionary clock
    b. In general mutation accumulates at silent sites is ~10 times slower than at replacement sites
    c. Human glutathione reductase gene should be more related with that of monkey than Drosophila
    d. Neutral substitutions mean those mutations without any effect on the functions of the protein
    e. There are 10 differences in 146 residues of human β and δ globin chains, so the divergence is 6.9%.

17. The Hamming distance, defined between two strings of equal length, is the number of positions with mismatched characters. Therefore, the Hamming distance between DNA sequences 5'-ATGTGTGTGCC-3' and 5'-ATGGGTGTACC-3' is
    a. 1   b. 2   c. 9   d. 10   e. none of the above.

II. Answer following questions in details (49%)
1. The number of gene products has been estimated to be much bigger than the number of genes. What are the likely causes for this phenomenon?
2. What is RNA "secondary structure"? What are the possible roles of secondary structure on the function of RNA?
3. What is "codon usage frequency"? If you are given a complete sequence of a bacterial genome, how are you going to determine the codon usage frequency for this microorganism?
4. You are given a mouse genomic contig sequence of 5,000 kb in length. Please explain how you can identify exons, introns, promoters, and intergenic sequences within this sequence either by computation strategy or testing experimentally.
5. Briefly describe what kind of information does the National Center of Biotechnology Information (NCBI) provide?
6. Please describe the theory behind "cDNA microarray" technology? What are the major applications of this technology?
7. Please distinguish the following related terms: expression sequence tag (EST); cDNA; transcriptome; proteome.