QUESTION 1: SHORT ANSWER

1	out of 1 points

You want to identify virulence genes in a genome. You have turned all the genes present in the genome into a searchable databank.   You also have a set of 1000 putative virulence genes.   You perform a search of the genome with each of a the 1000 putative virulence genes. The overall aim is a 0.1% probability to identify a false positive. Using the Bonferroni correction, which Evalue should be applied to each of the 1000 individual databank searches?

Given Answer: .001/1000= 1E-6

Correct Answer: ✅ 0.001/1000=1*10^-6

Response Feedback: | Arial ↕ | 3 (12pt) ↕ | | | | | | | | | | | | |

Path: p   Words:0

QUESTION 2:   SHORT ANSWER | 1 | **out of 1 points (Extra Credit)**

Considering the previous question, what features of the collection of 1000 virulence genes might make the Bonferroni correct too stringent. (It might help to consider the extreme case, where all of the 1000 putative virulence genes are homologs of one another).

Given Answer: In the case where all of the 1000 putative virulence genes were homologs of one another, we'd see the Bonferroni correction be a bit too strict because the point of Bonferroni correction is to allow us to reject the null hypothesis. However, if our 1000 genes are homologs or even if just a few of them were to be homologs, we could no longer reject our null hypothesis as none of our individual tests would be different from chance.

The Bonferroni correction seems to assume independence of each gene, but if the 1000 were homologs, the independence of each variable would no longer be a valid aspect of the correction and would therefore be providing extra false positives.

Correct Answer: [None]

Response Feedback: | Arial | 3 (12pt) |

Path: p   Words:0

QUESTION 3:　MULTIPLE CHOICE

| 1 | out of 1 points |

Command line versions of the BLAST programs are available for which platform?

Given Answer:　☑ D.　All of the Above

Correct Answer:　☑ D.　All of the Above

QUESTION 4:　MULTIPLE CHOICE

| 1 | out of 1 points |

Comparing sequence A to sequence B obtains an alignment that matches sequences A and B over their whole length. The P-value for this alignment is $<10^{-13}$. Sequence B also has a significant match to sequence C ($P<10^{-9}$). You consider these P-values as sufficient proof for homology.

Given Answer: ☑ A.

This shows that sequence A is homologous to sequence C

Correct Answer: ☑ A.

This shows that sequence A is homologous to sequence C

QUESTION 5:　MULTIPLE CHOICE

| 1 | out of 1 points |

If BLAST returns a match with an E-value of 5.4 e-11, what is the probability that this match represent a false positive?

Given Answer:　☑ B.　$5.4 \ 10^{-11}$

Correct Answer:　☑ B.　$5.4 \ 10^{-11}$

QUESTION 6:　MULTIPLE CHOICE

| 1 | out of 1 points |

In the above example, what is the frequency of false negatives in the databank?

Given Answer: ✅ D.

The rate of false negatives cannot easily be estimated.

Correct
Answer: ✅ D.

The rate of false negatives cannot easily be estimated.

QUESTION 7:   MULTIPLE CHOICE | 1 | out of 1 points

If you load a multiple sequence FASTA formatted file into an alignment program and the program only recognizes a single sequence, what could have gone wrong?

Given
Answer: ✅ C.

the text file used different end of line conventions than the alignment program.

Correct
Answer: ✅ C.

the text file used different end of line conventions than the alignment program.

QUESTION 8:   MULTIPLE CHOICE | 1 | out of 1 points

If you want to align two sequences that are about 35% identical, which of the following scoring matrices would be most appropriate:

Given Answer: ✅ B. Blosum 35

Correct Answer: ✅ B. Blosum 35

QUESTION 9:   MULTIPLE CHOICE | 1 | out of 1 points

If you want to do a BLAST search of the non-redundant database using a new catalytic RNA sequence as query, which is the BEST search program to use?

Given Answer: ✅ A. blastn,

Correct Answer: ✅ A. blastn,

QUESTION 10: MULTIPLE CHOICE

1	out of 1 points

In a BLAST search, what does the filter for low-complexity do?

Given
Answer: ✅ C.
It replaces regions of low complexity in the query sequence with the symbol for any residue.

Correct
Answer: ✅ C.
It replaces regions of low complexity in the query sequence with the symbol for any residue.

QUESTION 11: MULTIPLE CHOICE

1	out of 1 points

Selection for function can preserve sequence similarity in the pairwise comparison of homologous proteins, across domains separated by how many years of independent evolution?

Given Answer: ✅ E. All of the above

Correct Answer: ✅ E. All of the above

QUESTION 12: MULTIPLE CHOICE

1	out of 1 points

One databank search is done using FASTA with an amino acid sequence as query and the only reported match has an E-value of 52, what does this mean for the homology of the two sequences?

Given
Answer: ✅ A.
An E-value of this magnitude does not prove homology, but the sequences may never-the-less be homologous.

Correct
Answer: ✅ A.

An E-value of this magnitude does not prove homology, but the sequences may never-the-less be homologous.

QUESTION 13: MULTIPLE CHOICE

1   out of 1 points

One databank search is done using FASTA with an amino acid sequence as query and the only reported match has an E-value of 0.000005. What does this mean for the homology of the two sequences?

Given Answer: ✓ A.

This proves (beyond reasonable doubt) that the two sequences are homologs.

Correct Answer: ✓ A.

This proves (beyond reasonable doubt) that the two sequences are homologs.

QUESTION 14: MULTIPLE CHOICE

1   out of 1 points

Some students still have difficulties to discriminate between the term homology (=shared ancestry) and significant similarity. Which of the following statements is correct:

Given Answer: ✓ A.

All complex sequences that show significant similarity in a pairwise sequence comparison are homologous.

Correct Answer: ✓ A.

All complex sequences that show significant similarity in a pairwise sequence comparison are homologous.

QUESTION 15: TRUE/FALSE

1   out of 1 points

True/False A multiple sequence fasta file contains hyperlinks to the actual sequences.

Given Answer: ✓ False

Correct Answer: ✓ False

QUESTION 16:   TRUE/FALSE

1	out of 1 points

True/False RNA alone CAN have catalytic activity, it does NOT need to collaborate with proteins to do so, and it is capable of doing more than providing specificity due to base pairing.

Given Answer: ✅ True
Correct Answer: ✅ True

QUESTION 17:   MULTIPLE CHOICE

1	out of 1 points

Using a random shuffling approach (PRSS) you find that two sequences have an E value (assuming 10000 comparisons) of 950. This

Given Answer: ✅ D.

does not exclude the possibility that the two sequences might be homologous

Correct Answer: ✅ D.

does not exclude the possibility that the two sequences might be homologous

QUESTION 18:   MULTIPLE CHOICE

1	out of 1 points

Usually E values smaller than a certain threshold are considered to demonstrate homology. This threshold is usually about

Given Answer: ✅ C. about 10^{-4},

Correct Answer: ✅ C. about 10^{-4},

QUESTION 19:   MULTIPLE CHOICE

1	out of 1 points

Usually a Z values of which magnitude is considered to demonstrate homology?

Given Answer: ✅ B. larger than 3

Correct Answer: ✅ B. larger than 3

QUESTION 20: MULTIPLE CHOICE

| 1 | out of 1 points |

What are two of the most commonly used scoring matrices for data bank searches and for aligning protein sequences?

Given Answer: ✅ B. PAM and Blosum

Correct Answer: ✅ B. PAM and Blosum

QUESTION 21: SHORT ANSWER

| 1 | out of 1 points |

What does PAM stand for, and what does it mean?

Given Answer:
PAM stands for Point Accepted Mutation or Percent point Accepted Mutation, and can basically be defined as a replacement of a single amino acid in a protein's primary level of structure, however this doesn't include lethal or silent mutations--so essentially only those that would be accepted by natural selection.

Correct Answer: ✅
Point accepted Mutation - Accepted point mutations per 100 residues. The PAM matrix (20x20) gives information on how frequently one amino acid is substituted by another one. The PAM001 matrix was calculated for very closely related sequences, separated by only one substitutions per 100 residues. While the matrix was calculated from closely related sequences, the PAM250 matrix (calculated from applying PAM001 250 times) is appropriate for very divergent sequences.

Response Feedback:

| | | Arial ▲▼ | 3 (12pt) ▲▼ | | | | | | | | |

Path: p　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　Words:0

QUESTION 22: MULTIPLE CHOICE　　　　　　| 1 | **out of 1 points**

What is a Z-value?

Given
Answer:　　✅ C.
　　　　　　Number of standard deviations a match is above mean,
　　　　　　generated by randomizing sequences.

Correct
Answer:　　✅ C.
　　　　　　Number of standard deviations a match is above mean,
　　　　　　generated by randomizing sequences.

QUESTION 23: MULTIPLE CHOICE　　　　　　| 1 | **out of 1 points**

When aligning two sequences that are about 20% identical, which of the
following scoring matrices would be most appropriate?

Given Answer:　✅ D. PAM 210

Correct Answer: ✅ D. PAM 210

QUESTION 24: MULTIPLE CHOICE　　　　　　| 1 | **out of 1 points**

Copy of
One data bank search is done using FASTA with an amino acid sequence
as query and the only reported match has an E-value of 3 E-3, what does

this mean for the homology between the query and the
target sequences?

Given
Answer: ✅ A.

               An E-value of this magnitude is suggestive of homology, but
               further studies need to undertaken to prove homology

---

| | | Exit | Save and Exit | Save and Next |

Answer: An E-value of this magnitude is suggestive of homology, but
               further studies need to undertaken to prove homology
               beyond reasonable doubt.

QUESTION 25: MULTIPLE CHOICE

| 1 | out of 1 points |

In the unix operating system, which command would one use to display
the content of a file on the screen?

Given Answer: ✅ B. cat

Correct Answer: ✅ B. cat

QUESTION 26: MULTIPLE CHOICE

| 1 | out of 1 points |

In the unix operating system, which command would one use to check if
a file is in the current directory?

Given Answer: ✅ A. ls

Correct Answer: ✅ A. ls

QUESTION 27: MULTIPLE CHOICE

| 1 | out of 1 points |

In the unix operating system, which command could one use to copy the
content of two file into a single file?

Given Answer: ✅ B. cat

Correct Answer: ✅ B. cat

QUESTION 28:   MULTIPLE CHOICE

<div style="float:right">

| 1 | out of 1 points |

</div>

In the unix operating system, which command would one use to enter a subdirectory?

Given Answer: ✅ D. cd

Correct Answer: ✅ D. cd

QUESTION 29:   MULTIPLE ANSWER

<div style="float:right">

| 5 | out of 5 points |

</div>

Margaret Dayhoff was one of the founders of the field of bioinformatics. Which of the following were her contributions?

Given
Answers: ✅

She developed log-odds matrices to score substitution events for nucleotide and aminio acid sequences

✅

She reconstructed the evolutionary history of tRNAs that included gene duplication events that had occurred before the ast universal common ancestor.

✅

She compiled an Atlas of protein sequences and structures. This collection became the Protein Information Resource and was a precursor to genbank.

✅

She devised the single letter code for amino acid sequences.

Correct   ✅
Answers:

She developed log-odds matrices to score substitution events for nucleotide and aminio acid sequences

✅

She reconstructed the evolutionary history of tRNAs that included gene duplication events that had occurred before the ast universal common ancestor.

✅

She compiled an Atlas of protein sequences and structures. This collection became the Protein Information Resource and

was a precursor to genbank.

✅

She devised the single letter code for amino acid sequences.