Feedback — Course Project **Please Note: No Grace Period**

Thank you. Your submission for this exam was received.

You submitted this exam on **Mon 10 Aug 2015 9:32 PM PDT**. You got a score of **10.00** out of **10.00**.

In this project you will be reading a genomic data science paper and answering some questions to help you learn about how the different fields in genomic data science work together and to evaluate your understanding of some of the concepts we have learned throughout the course.

The paper we are reading is called: "Microbial Genes in the Human Genome: Lateral Transfer or Gene Loss?" You can access the paper from the Science Magazine site here:

http://www.sciencemag.org/content/292/5523/1903.full.

You'll need to register for a free Science account to access the full text. If you aren't able to do that, then you can also access the via this link:

Salzberg et al. 2001

You may also find the annotated version of the manuscript helpful for the project.

Annotated version

Question 1

Why did the authors write this paper?

Your Answer	Score	Explanation
○ To compute the E-values for the BlastP matches to the proteins from the human proteome.		

    ◯ To identify species that share more genes in common
between humans and bacteria.

    ◯ To prove that there are a large number of genes shared
between humans and bacteria.

    ⦿ To propose a plausible alternative to the hypothesis that   ✔   1.00
genes had been "laterally" transferred to humans.

Total	1.00 / 1.00

Question 2

What is "lateral gene transfer"?

Your Answer	Score	Explanation
⦿ When genetic material is passed from the genome of one organism to another through a process other than reproduction.	✔ 1.00	
◯ When genetic material is passed between one human and another human.		
◯ When a gene is transferred out of the DNA and permanently lost.		
◯ When genetic material is transferred from RNA to proteins.		
Total	1.00 / 1.00	

Question 3

Why is lateral gene transfer (LGT) from bacteria to humans unlikely?

Your Answer	Score	Explanation
⦿ Because a bacterium would have to infect a germline cell,	✔ 1.00	

enter the nucleus of that cell, and insert some of its DNA into one of the host's chromosomes, after which the mutation would then have to provide an evolutionary advantage to spread through the population.

○ There hasn't been sufficient time since humans and bacteria diverged for laterally transferred genes to spread through the population.

○ Because bacteria never actually enter human cells during an infection.

○ It is not unlikely; in fact, LGT has occurred and it is an ongoing process in the human population.

Total	1.00 / 1.00

Question 4

What are homologs?

Your Answer	Score	Explanation
○ Two genes in different organisms that have been mutated at the same rate.		
○ Genes that are greater than 99% similar in DNA sequence		
○ Identical mutations that occurred over evolutionary time.		
◉ Two copies of a gene in different organisms that share a common ancestor.	✔ 1.00	
Total	1.00 / 1.00	

Question 5

What was the main method used to rule out lateral gene transfers between humans and bacteria?

Your Answer	Score	Explanation
○ If a homolog of a gene was found in prokaryotic genomes.		
○ If genes were found to have mutated between eukaryotic genomes and human genomes.		
⦿ If a homolog of a gene found in humans was also found in a species of nonvertebrate eukaryotes.	✔ 1.00	
○ If a homolog of a eukaryote was found in another eukaryote but not in humans.		
Total	1.00 / 1.00	

Question 6

Why would this method rule out lateral gene transfers?

Your Answer	Score	Explanation
○ Humans and bacteria are both likely to have shared an evolutionary history with nonvertebrate eukaryotic organisms, so genes are likely to be homologous across all three.		
⦿ Lateral gene transfer is an unusual process compared to standard inheritance and nonvertebrate eukaryotic organisms and humans are evolutionarily "closer" than bacteria and humans. If humans and nonvertebrates share a homologous gene, it was likely not directly passed from bacteria to humans.	✔ 1.00	
○ Lateral gene transfer is a more likely explanation only when organisms that are close together evolutionarily share homologous genetic material.		
○ Nonvertebrate eukaryotic organisms and bacteria are evolutionarily "closer" than invertebrate eukaryotic organisms and humans. If they share a homologous gene, then bacteria are likely to have passed genes directly to humans.		

Total		1.00 / 1.00

Question 7

What are the biological, computational, and statistical parts of Figure 1?

Your Answer	Score	Explanation
○ **Biological**: the argument that lateral transfer should be ruled out if there is a human/nonvertebrate eukaryote homologs. **Computational**: The identification of homologs by performing Blastp searches on known protein sets. **Statistical**: The calculation of the standard error for the sample size curves.		
● **Biological**: the argument that lateral transfer should be ruled out if there is a human/nonvertebrate eukaryote homologs. **Computational**: The identification of homologs by performing Blastp searches on known protein sets. **Statistical**: Observing and quantifying the trend in genes shared versus genome sample size.	✔ 1.00	
○ **Biological**: the argument that gene should be ruled out if there is a human/nonvertebrate eukaryote homologs. **Computational**: Observing and quantifying the trend in genes shared versus genome sample size. **Statistical**: The identification of homologs by performing Blastp searches on known protein sets.		
○ **Biological**: the argument that lateral transfer should be ruled out if there is a human/nonvertebrate eukaryote homologs. **Computational**: the plotting of number of genomes versus number of genes in common **Statistical**: Observing and quantifying the trend in genes shared versus genome sample size.		
Total		1.00 / 1.00

Question 8

What are the biological, computational, and statistical parts of Figure 2?

Your Answer	Score	Explanation
○ **Biological**: the argument that proteins should have more similar sequences if they are evolutionarily closer. **Computational**: The identification of homologs of human HAS genes by iterative BlastP searches and application of the neighbor-joining algorithm to create the phylogenetic tree. **Statistical**: The statistical modeling of protein sequences via a Markov Model.		
○ **Biological**: the argument that proteins should have more similar sequences if they are evolutionarily closer. **Computational**: The calculation of statistical significance of the protein hits in the Blastp search. **Statistical**: The statistical modeling of protein sequences via a Markov Model.		
○ **Biological**: the argument that lateral gene transfer is less common than standard gene flow through reproduction **Computational**: The identification of homologs of human HAS genes by iterative BlastP searches and application of the neighbor-joining algorithm to create the phylogenetic tree. **Statistical**: The inference that humans cluster more closely (have smaller distances to) other eukaryotes than to bacteria.		
⦿ **Biological**: the argument that proteins should have more similar sequences if they are evolutionarily closer. **Computational**: The identification of homologs of human HAS genes by iterative BlastP searches and application of the neighbor-joining algorithm to create the phylogenetic tree. **Statistical**: The inference that humans cluster more closely (have smaller distances to) other eukaryotes than to bacteria.	✔ 1.00	
Total	1.00 / 1.00	

Question 9

The analysis in this paper required multiple data sources. Which of the following data sources was not used in the paper?

Your Answer		Score	Explanation
◯ The set of all known genes (at the time) from the malaria parasite, Plasmodium falciparum.			
⦿ The complete set of noncoding RNA genes from the human genome.	✔	1.00	
◯ The complete set of genes from the fruit fly, nematode worm, yeast, and mustard weed genomes.			
◯ The complete set of annotated genes from two different versions of the human genome.			
Total		1.00 / 1.00	

Question 10

In the end what is the conclusion of the paper?

Your Answer		Score	Explanation
◯ That the argument for lateral gene transfer is statistical because we must average over multiple possible transfer events.			
◯ That genes are more likely to be laterally transferred from certain types of bacteria to humans.			
◯ That clustering organisms by protein similarity is the right way to discover lateral gene transfer events.			
⦿ That a more plausible explanation for the observation of homologous genes found in bacteria and humans but not in non-vertebrate eukaryotes is gene loss and low sample size.	✔	1.00	

Total	1.00 / 1.00