Name:

Bioinformatics Take Home Test #4

Due Date Monday 10/26/2015 before class

(This is an open book exam based on the honors system -- you can use notes, lecture notes, online manuals, and text books.

Teamwork is not allowed *on the exams*, write down your own answers, do not cut and paste from webpages. If your answer uses a citation, give the source of the quoted text.)

Make sure each answer is only on one page, by using page breaks. Splitting an answer onto two pages leads to grading errors.

Do not write or type in font smaller than 12 point or write in cursive. Doing so will lose you 2 pts.

If you have an emergency and cannot submit a quiz in person, email it in by the start of class on the due date. If you do so, first remove the instructions and extras (blank lines, alternative answers for multiple choice questions) from your document, so that only your answers, a minimal amount white space, and optionally the questions, are left.

Note on Late Quizzes: Late quizzes are an inconvenience and cannot be accepted at all after the answers have been released. If your quiz is submitted within the first 12 hours after the deadline, you will receive 5% off. Each additional 12 hours is an additional 5% off, up until the graded quizzes are returned or the answers released.

All questions worth 1pt

1. Write a question for this quiz and email it to your TA- Due Wednesday at 4pm

2. According to Hennig a natural taxonomy should be based on which of the following?

A. shared primitive characters.
B. shared derived characters.
C. homoplasies.
D. non-shared derived characters.
E. None of the above.

3. Which of the following is true regarding genome rearrangements?

A. They occur frequently in evolution.
B. They are often responsible for erecting species barriers.
C. They can trigger speciation events.
D. Closely related organisms show fewer genome arrangements than more distantly related organisms
E. All of the above.

4. True/False In BLAST searches using only a single genome as query, proteins have more than one match because of paralogs.

5. What is GC strand bias (based on location with respect to the origin and terminus of replication)?
A.There are more GC dinucleotide simple repeats near the origin.
B.The G versus C content of the leading strand versus lagging strand changes.
C.The CG versus AT content of a genome changes.
D.The GC versus AT content of the leading strand versus lagging strand changes.
E.None of the above.

6. Which of the following is NOT an advantage of performing BLAST on the command line?
A. It is easy to BLAST an entire genome against another entire genome
B. The same script to perform command line BLAST searches can be reused anytime one wants to add a BLAST step to a pipeline.
C. It is simple enough for anyone, even people with no computer skills, to point and click their way to results.
D. It is possible to write a script to run 100,000 BLAST searches in one go
E. Scripted BLAST searches can be put into a pipeline with other computer scripts, to perform a complex task for you, leaving one free to do other things

7. In the unix operating system, which command would one use to check if a file is in the current directory?
A. ls B. cat C. pwd D. cd E. qlogin

8. In the unix operating system, which command would one use to enter a subdirectory?
A. ls B. chmod C. pwd D. cd W. qlogin

9. When running a perl script, no error is returned to the screen, but also no output file is created. What went wrong? (Hint, play around with purposely creating these errors and see what happens)
A. The perl script is not present in the directory you are in.
b. You have a space in the name of the perl script. You cannot put spaces in the script name.
c. Typo. You spelled the name of the file wrong.
d. The perl script didn't fully transfer over to the cluster. There is only an empty file with that script name.
e. Incorrect case. Unix is case sensitive.

10. In the principle component analysis, JALVIEW uses which of the following to define protein space?
A. The presence or absence of a conserved sequence motif to define protein space.
B. A tree based on percent identity to define groups that are close to each other in sequence space.

C. Each alignment column as a dimension to define protein space.
D. GC bias on the leading versus lagging strand.
E. None of the above.

11. What is a Principle Component Analysis?
A. A way to visualize n-dimensional protein space by breaking projecting it onto a 2-dimensional screen.
B. BLAST is used to identify common motifs that together using domain shuffling make up the components of a large number of proteins.
C. A measure of how many rearrangements a genome has undergone.
D. A method for detecting duplicated genes.
E. None of the above.

12. An example of strand bias is:
A. One strand of DNA has more Guanine than the other strand
B. One strand of DNA is more frequently found in the cell than the other strand
C. One strand of DNA is more prone to mutations than the other strand
D. One strand of DNA can be degraded more quickly than the other strand
E. All of the above

13. True/False Group 3 Introns are present in the eukaryotic nucleus and DO NOT need spliceosomes to get spliced out, because they have their own self splicing machinery.

14. Which of the following is NOT one of the possible fates of a pseudogene?
A. Decay, lost, and deletion.
B. Gain a homing endonuclease domain and turn into inteins.
C. Subfuctionalization (Both copies retain only part of the original function).
D. Neofunctionalization (Acquires a new function).
E. Sit around semi-permanently as junk DNA.

15. Which statement is NOT in support of the Duplication-Degeneration-Complementation model for the creation of new genes:
A. The fraction of genes preserved following polyploidization events is higher than predicted by the classic model
B. Most loci observed in preserved gene lineages appear to have nonfunctional members in some related tetraploid species
C. In *Xenopus laevis*, nucleotide substitution patterns are consistent with the action of purifying selection on both copies of the duplicated genes
D. For loci that have avoided nonfunctionalization in both duplicate copies, there is only a small amount of null alleles segregating in extant populations
E. The majority of duplicate gene copies have acquired new functions that did not already exist in the ancestral genome

16. Which of the following is NOT an example how a new gene can be created?
a) Through mutations
b) Left over DNA of viruses or other genetic parasite being repurposed
c) Golgi Apparatus packaging of proteins
d) Gene duplication followed by neofunctionalization
e) None of the above

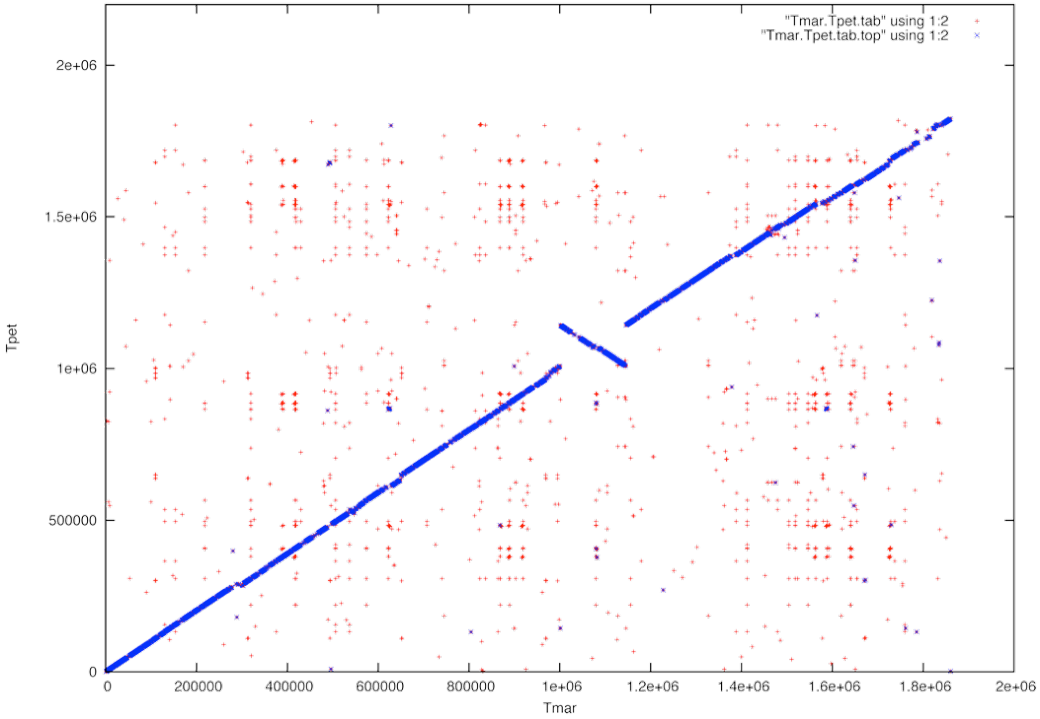
17. True/False Most duplicated genes go on to preform a new functions in an organism.

18. Which one of the following is NOT an outcome in the evolution of duplicate genes?
A. Neofunctionalization
B. Retrofunctionalization
C. Subfunctionalization
D. Nonfunctionalization
E. All are examples of outcomes

19. True/False Plastids are descended from free living Cyanobacteria (also sometimes called blue green algae).

20. True/False There are many unrelated Eukaryotes that acquired the ability to photosynthesize by capturing as an endosymbiont an Eukaryote with a plastid endosymbiont. This is called secondary endosymbiosis.

For Questions 21 through 24, refer to the following graph

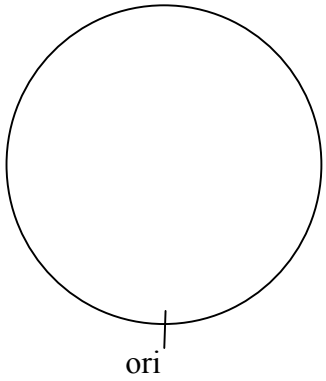


21. What mechanism is this graph depicting when blue dots appear on the downward sloping diagonal?
A. Translocation
B. Deletion or Insertion
C. Neofunctionalization
D. Inversion
E. All of the above

22. What mechanism is this graph depicting when there is a gap in the blue line, with the blue line picking up a notch higher or lower on the y-axis after the gap?
A. Translocation
B. Deletion or Insertion
C. Neofunctionalization
D. Inversion
E. All of the above

23. After plotting the blast hits from the two genomes, what does the blue line represent?
A. The location of all genes in one genome versus the location of ALL the blast hits in the other genome
B. The location of all genes in one genome versus the location of the top scoring blast hit in the other genome
C. The location of the gene in the environment
D. A and B
E. None of the above

24. Map the genome rearrangement shown onto the circular genome below:



25. What happened in the plot on the right?

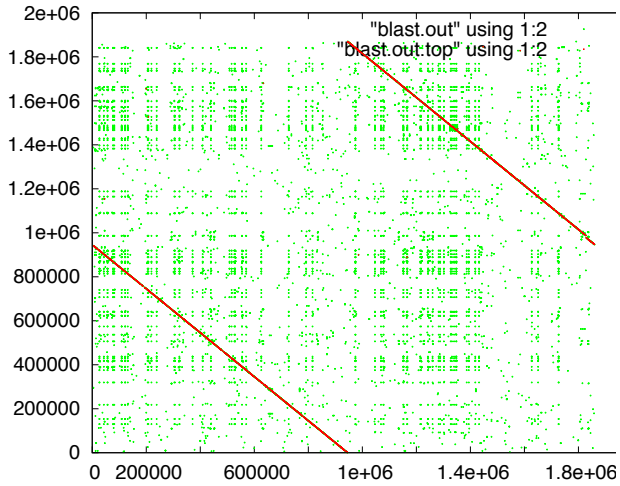
A. Two nearly identical genomes were used, but the origin of replication was miscalled in one.

B. One massive genome inversion, involving half of the genome

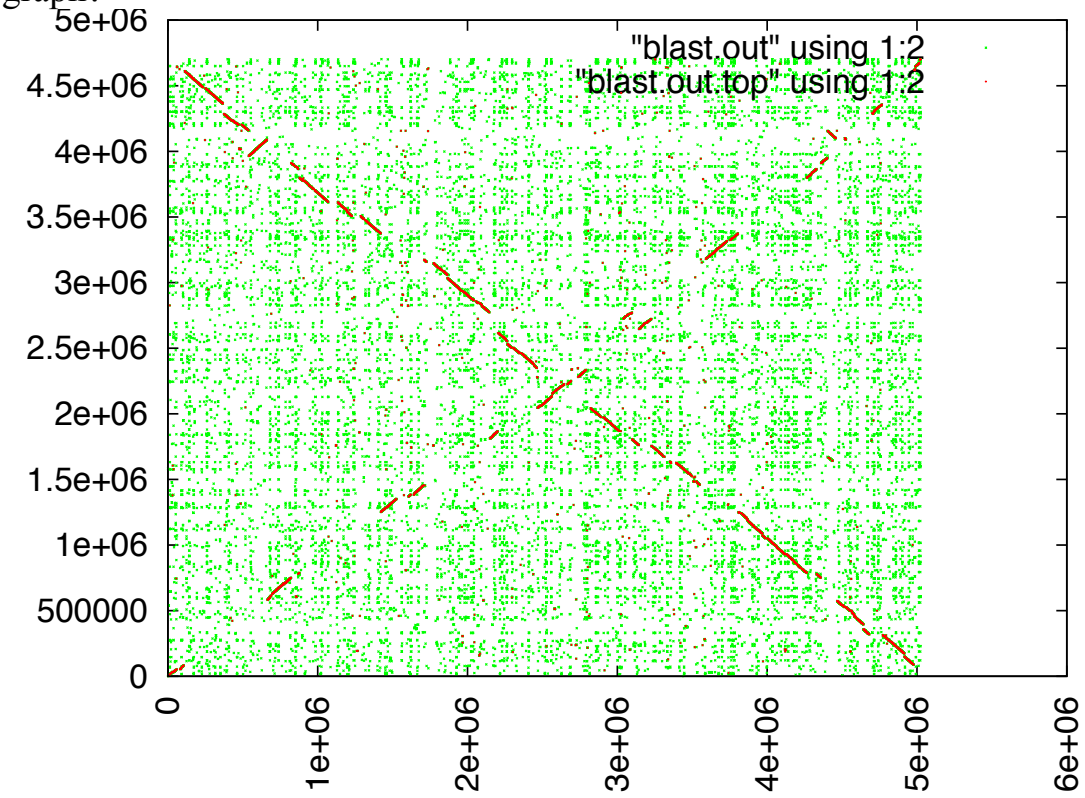
C. One round of whole genome duplication, so that every gene is present in one of the two genomes twice and only once in the other

D. These genomes are so divergent that synteny can no longer be observed.

E. There is a strong strand bias.



26. How many genome rearrangement events are needed to produce the following graph?



A. 1 B. 2 C. 3 D. 4 E. 5 F. 6 or more

27. A monophyletic group of organisms that is defined by which of the following?
A) synapomorpy
B) autapomorphy
C) paraphyly
D) polyphyly
E) symplesiomorpy

28. A paraphyletic group of organisms that is defined by which of the following?
A) synapomorpy
B) autapomorphy
C) paraphyly
D) polyphyly
E) symplesiomorpy

Extra credit:

1. Draw the rearrangements shown in question 26