Which of the following is NOT part of the explanation for how complex functional molecules were assembled, despite the vastness of protein space?

- **Gaia directs protein evolution, through negative feedback loops, to the correct region of protein space.**
- There are multiple unrelated solutions for the same functionality, exemplified by the fact that there are non-homologous enzymes inhabiting completely different regions of protein space with the same function.
- An exact function does not need to be hit upon, because natural selection can take a protein with limited function and make it better.
- Similar structures have similar function, so there are entire regions of protein space occupied by homologs that all function equally well, or nearly so.
- Protein space is made slightly smaller by removing all of the possibilities that cannot be synthesized or they will clog up the ribosome

Bioinformatics does NOT include which of the following activities?

- Analyzing genome data using computers
- **Figuring out a protein structure from X-ray crystallography**
- Detecting homologs using primary sequence similarity
- Detecting homologs using secondary sequence similarity

Which of the following is NOT an example how a new gene can be created?

- Through mutations
- Left over DNA of viruses or other genetic parasite being repurposed
- **Golgi Apparatus packaging of proteins**
- Gene duplication followed by neofunctionalization
- None of the above

The universe and the earth are approximately how old, respectively?

- 20 billion years old and 500 million years old
- 14 billion years old and 4.5 billion years old
- 16 billion years old and 6 billion years old
- 2 billion years old and 450 million years old
- 1 million years old and 4500 thousand years old

Can a protein be 84% homologous to another protein?

- Yes
  - **No, with the exception of cases of domain shuffling.**
- Yes, if they share 84% sequence identity.
- No, without exception.
Darwin considered evolution as a slow, rigorous, and gradual process. Describe 3 processes that might lead to the rapid/increased evolution of an organism. Explain your reasoning.

Gene transfer: Can allow an organism to move into a new environment.
Acquisition of symbionts with new properties: Same as above
Allopolyplloidization: The chromosome sets of the two parents are retained in the offspring, which can lead to instant speciation (both of the parent species have only half of the chromosome numbers)
Acquisition of a mutator gene that creates mutations in local gene neighborhood
Deleting/mutation of the mismatch repair system, which allows homologous recombination between more divergent genes
Mutations in regulatory genes that give rise to hopeful monsters – neoteny may be an example of this.

In the catalytic cycle of the ATPase, all of the catalytic subunits work in different phases of the catalytic cycle at any one point in time.

True/False

It is NOT possible to create a computer program to mimic evolution by means of artificial selection, because computers programs are NOT capable of finding new solutions that a human has never thought of before.

True/False

The finding that the ribosomal protein alone is responsible for the catalysis of translation is an argument against the RNA world hypothesis.

True/False

The place of the root in the tree of life was first determined using which molecule?

Inteins
The signal recognition particle and receptor
ATPsynthase catalytic and non-catalytic subunits
rRNA
None of the above

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

Thousands
Millions
Billions
Limited only by how long life has existed
All of the above
How many peptides (short proteins) of 70 amino acids in length are possible, given that there are 20 possible amino acids? For your answer only consider the principles of combinatorics and ignore possible incompatibilities between amino acids. How does this number compare to the estimated number of protons in the universe?

1.18E91
It is larger by about a factor of 100

How many **types** of subunits (note: this asks for the different types of subunits, encoded by separate genes, not the number of subunits) form the hexamer (the head) of nucleotide binding subunits in the F1 ATPase?

2

Give a one sentence description of the field known as Bioinformatics.

Bioinformatics is the science at the intersection of **Computer Sciences** (informatics) and **Genomics**. Sequences that do not show significant similarity are not homologous. **might never-the-less be homologous** are homologous.

Which of the following statements is correct:

- **All complex sequences that show significant similarity in a pairwise sequence comparison are homologous.**
- **All homologous sequences show significant similarity in a pairwise sequence comparison.**
- Both of the above statements are correct

Having the same or similar function frequently occurs with homologs.

**True/False**

Which elements make up the secondary structure of proteins?

- Hydrophobic domains, nucleotide binding motifs, and protein channels.
- Beta barrels, alpha corkscrews, and delta turns.
- **Beta sheets, alpha helices, and loops.**
- Van der Waals interactions, hydrogen bonds, and disulfide bridges.
- Multiple protein chains interacting to form one macromolecule.

What is homology?

- The creation of artificial life by any means.
- A difference found because of diverging evolutionary paths since the last common ancestor.
- **Similarity due to shared ancestry, i.e. both got it from a common ancestor.**
- When two proteins share a function, such as nucleotide binding, they also share sequence similarities, because of the limited size of protein space.
Shared sequence similarity based on convergent evolution, i.e. the ancestor did not have it.

Which of the following is NOT considered strong evidence for homology?

The arrangement of secondary structure elements in space is the same or very similar.
Significant primary sequence similarity
Significant primary sequence identity
**Identical function**
None of the above

What might be the reason for proteins that show significant similarity in their primary sequence (when no simple repeats or low complexity regions are present) to be homologous.

Sequence space is so big that stumbling onto a significantly similar sequence by chance is very unlikely.
This is a trick question, because NOT all currently known proteins that share significant similarity are homologous.
There is a direct link between sequence similarity and percent homology.
Homologous sequences share a function and there is only a small region of protein space that can accomplish any one function
All of the above

Assuming equal frequency of the different building blocks, two random protein sequences are on average _________ and nucleotide sequences are on average _________?

25% identical and 5% identical.
**5% identical and 25% identical.**
25% identical and 40% identical.
95% identical and 75% identical.
None of the above.

__________ sequences reach saturation before ________ sequences reach saturation, so __________ sequences can be used to look further back in time.

Nucleotide, protein, nucleotide
Protein, nucleotide, nucleotide
**Nucleotide, protein, protein**
Protein, nucleotide, protein
None of the above

Who drew the first phylogenetic trees?

Lamarck
Darwin
Mayr
Henning

Peter Simon Pallas (matter of debate, if the dynamic description of the tree is only due to the translation)

When inteins first begin to decay they lose the DNA-binding domain first, while the protein-binding domain must stay functional or it will destroy the function of the host proteins.

**True/False**

Among Site Rate Variation (ASRV) means that some sites will undergo multiple substitutions while other sites do not undergo any substitutions. Due to ASRV, protein and nucleotide sequences take longer to become saturated with substitutions than without ASRV.

**True/False**

What Boolean operators can be used in NCBI/Entrez searches? **AND OR NOT**

Match the terms on the left with the definitions on the right-

mRNA 7

tRNA 5

rRNA 10

transcription 4

replication 11

translation 1

intein 6

intron 2

exon 8

extein 9

Does not exist 3

1. The process of making a protein from an RNA template
2. A molecular parasite that splices itself out at the RNA level
3. A molecular parasite that splices itself out at the DNA level
4. The process of making RNA from DNA
5. RNA that binds an A.A. & matches it with mRNA triplet
6. A molecular parasite that splices itself out at the protein level
7. An RNA copy of a gene, used in the process of making proteins
8. Part of a host gene's transcript left after RNA parasite is spliced out
9. The host protein, which is spliced back together
10. RNA that makes up the ribosome and catalyzes protein synthesis
11. Process of creating a new DNA molecule, from DNA strand
Inteins are composed of which of the following domains? Choose 2.

- Self-splicing domain
- Walker motif
- Nucleotide binding domain (GRASP)
- Hydrolase domain
- Helix-turn-Helix DNA binding domain
- Homing endonuclease domain

Which of the following are databases available through the NCBI aka Entrez? Circle all that apply:

- BioProject (formerly Genome Project)
- Bookshelf
- Database of Genome Survey Sequences (dbGSS)
- GenBank
- Genome Reference Consortium (GRC)
- NCBI Help Manual
- Nucleotide Database
- Protein Database
- PubMed Central (PMC)
- Taxonomy
- All of the above and many many more.

If the following searches were conducted in PubMed for articles, what would the searches return? Please draw Venn diagrams to illustrate your answers (i.e. depict each of the individual searches as a circle).

A. Gogarten J NOT Gogarten JP
B. Gogarten JP AND Doolittle WF

C. Gogarten J OR ATP synthase

D. Intein AND (Gogarten JP OR Swithers K)
If multiple search terms are connected by Boolean operators without parenthesis, the NCB interface will start evaluation:

- **from the left**
- **from the right**
- Using a priority of AND over OR

What does the abbreviation NCBI stand for?

- **National Center for Biotechnology Information**

What is the Black queen hypothesis?

- A. Life is like an arms race, where all lifeforms have to run faster and faster just to stay in place. **This is the red queen hypothesis**
- B. Gaia favors cooperation, which is the driving force behind multi-species communities being interdependent.
- C. Leaky goods can be taken up by all members in a community, so selection for streamlined genomes will result in all members of a community producing only a subset of the required leaky goods.
- D. DNA-based organisms took over from the RNA world, after DNA was created by a virus in an act of genome warfare.
- E. None of the above.

The Modern Synthesis did not give much weight to the effects of mutations themselves.

- **True/False**

How might mutual aid be selected for?

- A. Trick question: it cannot be selected for, because even if a stingy species is going extinct, it cannot decide to stop being stingy.
- B. When cooperation results in more offspring for both, the entire community thrives and spreads.
- C. When cooperation is between close relatives, like siblings, helping each other survive leads to some of your own genes being passed on.
- D. When one bird helps defend another bird’s nest, that second bird will remember and return the favor when the first bird is under attack.
- E. B-D.

Why did Darwin consider the term "Coral of Life"?

- A) Because an herbarium specimen of a red algae that was wrongly labeled as a coral looked very similar to a phylogenetic tree. Most of the sentences is correct, but the "because" is not.
- B) Because he recognized that fusion of lineages is an important process in evolution of species, and fan corals often have strands that fuse, in contrast, tree branches only branch and never fuse. Again mostly correct, except for the because. This was not the argument Darwin put forward in his notebook.
- C) Because a tree has living cells in the root, stem and leaves, whereas a phylogenetic tree has living representatives only at the tips.
What is the Gaia hypothesis?

A. Earth’s plants control the planet’s temperature by selection for flower color
B. All life on Earth descended from ONE common ancestor
C. The unit of life is the entire Earth and the entire biosphere is alive, because no single species can exist in complete isolation. The ecosystems on the Earth are protected by negative feedback loops that help maintain homeostasis.
D. Mars cooled faster than the Earth and therefore was a more habitable place for life earlier. Life arose first on Mars and traveled to Earth on meteorites
E. Large glaciers and ice ages result from a runaway cold-house, where the Earth gets colder because glaciers reflect more light from the surface
F. Zircon crystals in 3.8 billion year old rocks were produced by ancient life over 4 billion years ago. The zircon crystals are formed by running water and the bias in carbon isotope ratios indicates the presence of life.

Evolution by natural selection requires which three things to occur?

Variation among offspring, a niche, and competition for resources.
Heredity, excess offspring, and a niche.
Excess offspring, competition for resources, and variation among offspring.
A niche, a human to naturally select the best offspring, and oxygen.
None of the above.

We have worked with 3 enzymes in the Friday computer labs, which are shown below.
Please identify them:

Lysozyme with NAG trimer
Intein with HE bound to DNA
F1 ATPase (part of the F-ATPase)

The first genetic material likely was similar to
A. DNA
B. Lipids
C. Proteins
After plotting the blast hits from the two genomes, what does the blue line represent?

The location of all genes in one genome versus the location of ALL the blast hits in the other genome
The location of all genes in one genome versus the location of the top scoring blast hit in the other genome
The location of the gene in the environment
A and B

Map the genome rearrangement shown onto the circular genome below:
The following diagram gives the relation between the number of substitutions that have occurred during evolution (x axis) and the observed fraction of sequence differences. The depicted curve corresponds to the Jukes Cantor correction for a nucleotide sequence. This correction is only correct, if all sites have the same probability to undergo a substitution, and if all nucleotides occur with the same frequency.

Provide a rough sketch of how this relationship would change, if the different sites would have different substitution frequencies, and some sites would only very rarely undergo a substitution.

Indicate how the curve would change, if the sequences have a strong compositional bias (e.g., 50% A, 50% T, 0% G, 0% C), but all have the same probability to undergo a substitution event. (If combinatorics is not your expertise, it might help to think about a sequence that only consists of As and Ts.)