Exam starts at 12:05 and ends at 12:55

There are nine pages including this cover page
Please write your name on each page.

Please...

- Look over the entire exam so you don’t spend too much time on hard questions leaving easy questions unanswered.

- Check your answers to make sure that they make sense.

- To help us give partial credit, show your work and state any assumptions that you make.

<table>
<thead>
<tr>
<th>Question 1</th>
<th>36 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 2</td>
<td>30 points</td>
</tr>
<tr>
<td>Question 3</td>
<td>34 points</td>
</tr>
</tbody>
</table>
1. You are studying the relationship between the genome of the roundworm *C. elegans* to that of a related species *C. briggsae*.

(a 12 points) On the next page the major organization of the chromosomes in the two species are compared by dot-plots. Each species has six chromosomes, annotated as 1-6 for *C. elegans* and A-F for *C. briggsae*. In the space below the plots, complete the diagram showing the organization of the *C. briggsae* chromosomes relative to the *C. elegans* chromosomes. The *C. elegans* chromosomes are already drawn, and the relationship between Chromosome 5 and chromosome D is already diagramed to give an example of what your answer should look like. Put arrows and nucleotide numbers (coordinates) as needed. Note: the dashed lines in the plots are for your convenience to trace the coordinates, and the chromosome sizes may not be identical (e.g. Chromosome D is 24Mb and chromosome 5 is 18Mb).

(b 6 points) What is the likely basis for the relationship between Chromosome 2 in *C. elegans* and chromosome F in *C. briggsae*? Be as explicit as possible.

Chr. F Section 0-3 in *C. briggsae* is duplicated twice in *C. elegans* at chr. 2 6-9 & 9-12.

(c 6 points) Are there sequences that are that are present in one species, but completely absent from the other? If so, identify the chromosomal position of the unique sequences. Your answer should be in a format such as this: “Chromosome 12 between 6Mb and 14Mb”

Yes, D 12-18 in *C. briggsae*

(d 6 points) Knowing that *C. briggsae* is a more pathogenic parasite than *C. elegans*, what is a likely candidate region for a pathogenicity island? Your answer should be in the same format as for part c). Explain your reasoning.

3pts D 12-18

This is the only part of the *C. briggsae* genome that differentiates it from *C. elegans*. Everything else is contained in both.
**Name:** KEY

(e 6 points) You now examine the known annotation of the *C. elegans* genome, comparing one annotated ORF in *C. elegans* to the corresponding sequences in three other worm genomes. In the figure below, we use the following code:

- Annotated start
- Annotated stop
- in/del
- conserved
- predicted stop (in other species)

Is this ORF valid (represents a real protein coding gene) or would you remove it as a mistake? Explain your reasoning.

**Note:** There is no need to examine the actual nucleotide sequence!

3 pts NOT an ORF

1 pt - Low conservation

1 pt - many in/dels that are not multiples of 3
2. You study a wild population of the *Anopheles* mosquito in Senegal, and are interested in the loci for two genes Pla1 and Pla2, encoded on Chromosome II, which are important for the susceptibility of the fly to infection by the malaria parasite. For your convenience, the critical values of the chi-squared distribution are:

<table>
<thead>
<tr>
<th>df</th>
<th>0.1</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.706</td>
<td>3.841</td>
<td>5.024</td>
<td>6.635</td>
<td>10.828</td>
</tr>
</tbody>
</table>

(a 10 points) You capture 2500 mosquitoes and test each of the chromosomes. We denote by D,E the common SNP for each of the genes, and by d,e the less common SNP (i.e. two alleles for each gene). You obtain the following numbers for haplotypes:

| P_{DE} | 3610 |
| P_{De} | 41   |
| P_{dE} | 1333 |
| P_{de} | 15   |

Are the two loci in LD or not? Estimate your confidence.

\[ P_D = \frac{3651}{5000} = .730 \quad P_d = .270 \]

\[ P_E = \frac{3610 + 1333}{5000} = .989 \quad P_e = .011 \]

\[ \chi^2 = \frac{(O - E)^2}{E} \]

\[
\begin{array}{ccc}
\text{DE} & 3610 & \text{E} \\
\text{De} & 41 & 40 \\
\text{dE} & 1333 & 1335 \\
\text{de} & 15 & 15
\end{array}
\]

\[ E_{DE} = P_D \cdot P_E \cdot 5000 = 3610 \]
\[ E_{De} = P_D \cdot P_e \cdot 5000 = 40 \]
\[ E_{dE} = P_d \cdot P_E \cdot 5000 = 1335 \]
\[ E_{de} = P_d \cdot P_e \cdot 5000 = 15 \]

\[ df = 1 \]

\[ P > 0.05 \]

⇒ Not in linkage disequilibrium
(b 10 points) When you visit your study site in the next season, you learn from your colleague that the local Senegalese population has just been invaded by a second population of mosquitoes from Malawi. The size of the invading population is estimated to be a third of the native (existing) one. Your colleague in Malawi sends you the following gamete frequencies for the mixed population, based on 10,000 chromosomes. She claims that the loci for the Malawi population were in Linkage Equilibrium (LE) prior to mixing.

<table>
<thead>
<tr>
<th>P_{DE}</th>
<th>0.0277</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_{Dd}</td>
<td>0.0628</td>
</tr>
<tr>
<td>P_{dE}</td>
<td>0.2786</td>
</tr>
<tr>
<td>P_{de}</td>
<td>0.6309</td>
</tr>
</tbody>
</table>

Is your colleague correct? Estimate your confidence.

\[
P_D = \frac{277 + 628}{10,000} = 0.0905 \quad \quad P_d = 0.9095
\]

\[
P_E = \frac{277 + 2786}{10,000} = 0.3063 \quad \quad P_e = 0.6937
\]

\[
P_{DE} = 277 \quad \quad P_{Dd} = 628 \quad \quad P_{dE} = 0.001595
\]

\[
P_{de} = 6309 \quad \quad \chi^2 = 0.001595
\]

\[
\text{df} = 1, \quad P < 0.05
\]

Yes, colleague is correct.

(d 10 points) What would be the degree of Linkage Disequilibrium (D') in the first generation in the mixed fly population? How correlated would the two loci be?

\[
P_{DE\text{mixed}} = (0.25 \times 0.0277) + (0.75 \times 0.722) = 0.55
\]

\[
P_{de} = (0.25 \times 0.628) + (0.75 \times 0.0082) = 0.2185
\]

\[
P_{dE} = (0.25 \times 0.2786) + (0.75 \times 0.2666) = 0.269
\]

\[
P_{de} = (0.25 \times 0.6309) + (0.75 \times 0.003) = 0.158
\]

\[
P_D = (0.25 \times 0.0905) + (0.75 \times 0.73) = 0.57 \quad \quad D = P_{DE} \times P_{de} - P_{Dd} \times P_{dE}
\]

\[
D = 0.081
\]

\[
D' = \frac{0.081}{0.1026} = 0.789
\]

\[
\frac{\chi^2}{P_{DD}P_{EE}P} = 0.18136 \quad \leftarrow \text{How correlated two loci are}
\]
3. (a 3 points) A rare genetic disorder occurs at a frequency of $3 \times 10^{-4}$ in males and a frequency of about $10^{-7}$ in females. What is the likely mode of inheritance of this trait? Explain.

X-linked recessive.
In a population, females must have two copies of the disease allele to have the disease whereas males only need one copy of the allele.

(b 3 points) What is the allele frequency ($q$) for the trait?

$$q = 3 \times 10^{-4}$$ (all diseased males have the allele)

(c 5 points) Does this population appear to be undergoing random mating with respect to the genotypes for the trait (i.e. does the population appear to be in Hardy-Weinberg equilibrium)? Explain your reasoning.

$$q^2 = 9 \times 10^{-8} \times 10^{-7}$$
The disease frequency in females should be $q^2$ if in Hardy-Weinberg equilibrium. So, yes.

(d 3 points) Estimate the frequency of females in this population who are heterozygous for the trait.

$$2pq \times 2q = 6 \times 10^{-4}$$

(e 4 points) A dominant inherited disease occurs at a frequency of $4 \times 10^{-4}$ in a large population. Family histories reveal that for about half of the individuals with the disease, neither parent has the disease. Assuming complete penetrance (and reliable information about paternity) estimate the mutation rate ($\mu$) for the disease.

Half of the people with the disease are new mutations. This new mutation could arise in the maternal or paternal gamete. Thus $(4 \times 10^{-4})(\frac{1}{2})(\frac{1}{2}) = 1 \times 10^{-4}$

(f 4 points) Assuming that the allele frequency for the disease in part e) was set by a balance between new mutations and selection against individuals with the disease, estimate the selective disadvantage ($S$) for the disease that existed when the allele frequency for the disease was established.

Steady state: $\Delta q_{sel} + \Delta q_{mut} = 0$

$-Sq + \mu = 0$

$S = \frac{\mu}{q}$

$$2pq \approx 2q = 4 \times 10^{-4}$$

$$q = 2 \times 10^{-4}$$

$$S = \frac{1 \times 10^{-4}}{2 \times 10^{-4}} = 0.5$$
(g 6 points) Calculate the inbreeding coefficient (F) for the following pedigree. (Be careful to take into account all of the different ways that the individual marked by “?” could become homozygous for an allele by descent.)

4 ways to inherit 2 copies of disease allele:
- Carrier grandparents are siblings (2 ways):
  \[ F = \left(\frac{1}{2}\right)^2 \cdot 4 \text{ alleles} = 0.03125 \]
- Carrier grandparents are cousins (2 ways):
  \[ F = \left(\frac{1}{2}\right)^8 \cdot 4 \text{ alleles} = 0.015625 \]

\[ F_{\text{tot}} = 2 \cdot 0.03125 + 2 \cdot 0.015625 = 0.09375 \]

(h 3 points) What is the probability that the individual marked by “?” will be affected by an **autosomal recessive** trait with allele frequency \(2 \times 10^{-4}\) ?

\[ F_q = (2 \times 10^{-4})(0.09375) = 1.9 \times 10^{-5} \]

(i 3 points) What is the probability that the individual marked by “?” will be affected by an **autosomal dominant** trait with allele frequency \(2 \times 10^{-4}\) ?

\[ 2q = 4 \times 10^{-4} \]

**Inbreeding doesn’t affect autosomal dominant traits**
Grading section

Question 1 36 points:_________________

Question 2 30 points:_________________

Question 3 34 points:_________________

Total :_________________