

What is the answer to question 6?

Question 6 asked about the probability of “reciprocal monophyly” when two samples were drawn from each of two species, diverged for t generations, with effective population sizes N in each species and N in their common ancestor (actually the size of the common ancestor does not matter).

I admit it: it was too hard a question for this exam. Here is the correct answer, closely approached by only two people, obtained exactly by no one.

Suppose that the probability that two lineages will coalesce in t generations is Q . The probability that both pairs will coalesce is then Q^2 . That is one way the coalescent tree can show reciprocal monophyly. It is also possible that one will coalesce and the other will not, which has probability $Q(1 - Q)$. There are two ways this can happen (the coalescence could be in either species). If it does happen, there are then three lineages in the common ancestor (say a , $b1$, and $b2$). The probability that the next (next going back in time) coalescence will be between $b1$ and $b2$ is $1/3$ as there are three possible pairs that could coalesce. That is the probability that we will see reciprocal monophyly when three lineages get back into the common ancestor.

If both pairs of lineages fail to coalesce and thus four lineages get back to the common ancestor, that event has probability $(1 - Q)^2$. If they do, then there will still be reciprocal monophyly if the first (first going back in time) coalescence after that is either $a1$ with $a2$ or $b1$ with $b2$ and the next coalescence is the other one of those two coalescences. As there are 4 lineages, the probability of this sequence of events is 2 out of 6 (there are 6 combinations of 4 things taken two at a time), times 1 out of 3, or

$$\frac{2}{6} \times \frac{1}{3} = \frac{1}{9}$$

This leads to the expression

$$Q^2 + \frac{2}{3}Q(1 - Q) + \frac{1}{9}(1 - Q)^2$$

Now all we need is Q . The probability that a pair of gene lineages will not coalesce in time t , i.e. $(1 - Q)$, is the probability that an event whose probability is $\frac{1}{2N}$ per unit time does not happen in t units of time. This is like waiting for a radioactive decay. The fraction of pairs of lineages still undecayed is calculated using an exponential:

$$e^{-t/(2N)}$$

So we get

$$Q = 1 - e^{-t/(2N)}$$

and you substitute that into the expression we got. I wouldn't have required that you simplify the expression further.

Thank you all for showing as much work as you did and I'm sorry if the question was a worry to you at the exam.