

Multi-species coalescent for polyploids: notes for 6-13th April 2011

Graham Jones

2011-04-09

1 Notation

I use alpha, beta, gamma,... as species names, 1,2,3,... as specimen numbers, eg individual flowers from same species, and A,B,C for different sequence copies of a gene coming from different diploid genomes in same flower.

2 Probabilities for multiply labelled trees

Present time is zero, times increase into past. The root of a tetraploid tree T is at t . Its two genomes arise from a diploid tree D at times $v \geq u > t$. The case $u = v$ occurs when two diploids hybridized at this time, and both diploids leave surviving descendants at present.

The case $u > v$ occurs when one or two of the diploids producing the hybrid have not left descendants at present. Here are two possibilities leading to the same multiply labelled tree.

Suppose λ is speciation rate per branch, μ is extinction rate per branch, and α is the hybridization rate per pair of branches. These are constant, and all events independent. λ, μ apply to the diploid tree and the tetraploid subtree(s). Let $P_n(s)$ be the probability that a tree produced by a birth-death process with parameters λ, μ with origin at time s leaves n descendants at time 0.

I think it may be possible to calculate the probabilities of multiply labelled trees like this, even in closed form. But I haven't worked anything out properly! For the left hand one in the picture, need a speciation of diploid tree at v , then one at u , then at some point h between u and zero, there needs to be a hybridization event, after which both diploid subtrees started at v and u must go extinct, while the hybrid must leave exactly one survivor.

So I think expressions like this arise:

$$\alpha \int_0^u \sum_j j P_j(v-h) P_0(h)^j \sum_i i P_i(u-h) P_0(h)^i P_1(h) dh$$

The P_i s as functions of h are of general form

$$\frac{A + B \exp(\delta h)}{C + D \exp(\delta h)} \left[\frac{E + F \exp(\delta h)}{G + H \exp(\delta h)} \right]^i$$

with $A...H$ and δ constants, and the sums can be evaluated and the substitution $y = \exp(\delta h)$ leads to an integral of a rational function in y . It might look something like

$$\int_1^{\exp(\delta u)} \frac{Q(y)}{R(y)} dy$$

where Q and R are polynomials of degree about 6 or 7, all roots real. Well, maybe...

From a biological point of view, this model is probably not very good because it allows hybridizations between any pair of diploids, whereas you expect this to only happen between fairly similar ones. Also you would expect there to be an increased risk of extinction for the parents of a tetraploid since they face competition from the new hybrid.

References

- [1] Bruce Rannala and Ziheng Yang, *Bayes Estimation of Species Divergence Times and Ancestral Population Sizes Using DNA Sequences From Multiple Loci*, *Genetics* 164: 1645-1656 (August 2003)
- [2] Jeff J. Doyle, Ashley N. Egan, *Dating the origins of polyploidy events*, Article first published online: 16 DEC 2009, DOI: 10.1111/j.1469-8137.2009.03118.x