Todo List for *Houston We Have A Problem*

Bob Week

3/9/2022

Results/Figures

The Main Result: Error rates of the t-test approach as functions of genetic architecture (monogenic/oligogenic), spatial autcorrelation (continuous space vs island model), symmetry of coevolution (host locally adapted, parasite locally adapted, no local adaptation). Since the t-test approach was published under the assumption of discrete space, comparing our results under the island model to the original paper provides a baseline comparison which provides a stepping-stone for comparing with our results in the case of spatial autocorrelation. The symmetry of coevolution can be adjusted by tweaking relative dispersal abilities, strengths of biotic selection, or local carrying capacities for each species. There's a lot of empirical work suggesting parasites tend to have a simple genetic architecture relative to their host, so it will be interesting to see the effect of differing genetic architectures between the two species on error rates.

Additional Result:

- Identify scenarios where the t-test approach fails due to spurious interspecific correlations generated by spatially autocorrelated allele frequencies. Identify scenarios where the method works.
- Iscaf as a function of distance in bp's from selected site.

Methods

Simulate n replicated data sets (with $n \leq 1000$) for each of the following (where items of same depth are parallel while nested items are permutations of the parental item):

- neutral scenario (no biotic selection)
    - discrete space (island model)
    - continuous space
        - four combinations of dispersal distances (short/short, short/long, long/short, long/long)
- four combinations of selection strengths (weak/weak, weak/strong, strong/weak, strong/strong)
    - four combinations of causal architecture (monogenic/monogenic, monogenic/oligogenic, oligogenic/monogenic, oligogenic/oligogenic)
        - discrete space (island model)
        - continuous space
            - four combinations of dispersal distances (short/short, short/long, long/short, long/long)

That's $4n$ neutral sims + $64n$ non-neutral sims. Assuming each sim takes ten minutes implies a total of $2.2n$ hours if run in serial.

With these data apply the nuismerian t-test approach and:

- compute distribution of type-1 error rates for each neutral scenario
- compute distributions of type-1 and type-2 error rates for each non-neutral scenario

Development

- next steps for sims
  - two phases: debugging & automating
    - debugging: make sure structure of code "correct" somehow
    - automating: for a given set of parameters
      - make sure spatial coefficient of variation of abundance is low for both species
        - what is the ideal abundance density? as large as possible without being computationally prohibitive?
      - make sure spatial structure occurs in both species (which stat? morans i?)
      - make sure genome long enough wrt to recomb rate (what are we looking for? one recomb event per _?)
      - record associated stats in metadata with each run
  - next steps for downstream analysis:
    - make sure ld thinning is not too strong and not too weak (check ld threshold)
    - get scale of absolute divergence to see if the species is reprod isolated at large enough dists