Ryan Stratychuk 43762103 Oct 31 2017

I Introduction

-WHAT IS CLONALITY?

-Clonality is the condition of being clonal; the offspring is genetically identical to the parent.

-Selfing is related but it involves sexual recombination with only 1 parent.

-BENEFITS OF CLONAL REPRODUCTION

-Multiple benefits of clonal reproduction that offset the negatives of lower genotype diversity and lack of sexual recombination.

-Gives reproductive assurance > always will be able to spread in conditions that do not favour sexual reproduction (isolation, competition, disturbance, etc.)

-Range expansion and reclamation of disturbed areas by founders

-For plants, increases floral display size (more pollinators), increases competitive ability for light and water resources (quick and effective expansion of a genotype).

-Can see evidence of this in our very own forest. Himalayan Blackberry. This forest understory bush spreads through clonal growth to very quickly outcompete other plants for sun and space. I would bet that you would find large swaths of the same genotype.

-WHY SHOULD WE CARE ABOUT CLONALITY?

-Many important organisms of human value like fungi, bacteria, plants use asexual reproduction, producing clonal organisms and populations. Eg. Yeast, pathogens, etc.

-About 80% of plants have some way of reproducing asexually or through selfing, meaning that clonality plays a major part in studying the population genetics of plants.

-So if you are trying to do genetic or ecological work on plants microscopic organisms, you are most likely going to encounter an organism that have clonal reproduction and this must be accounted for or else you will have inaccurate analyses.

-You can find asexual reproduction in plants in 2 ways. The technical term for this in plants is Apomixis which is the replacement of sexual reproduction with asexual reproduction. There is clonal growth (think strawberry stolons) or agamospermy, clonal reproduction through seeds.
-Most importantly I think though is THAT SELECTION ACTS DIFFERENTLY ON ORGANISMS THAT REPRODUCE WITH A LOT OF CLONALITY: THE UNIT OF SELECTION BECOME THE CLONAL GENOTYPE WHICH IS SPREAD AMONG MANY MEMBERS

II. Population Structure

-For a long time, there was no standardized methodology for working with clonal organisms as it was hard to distinguish clonal organisms both visually and genetically before the advent of more powerful genotyping methods.

-I would like to present some of the work done by Arnaud-Haond et al 2007 which puts forward some standardized methods in working with clonal organisms and population genetics.

-First we must make some definitions:
  - Multilocus Genotype (MLG): combination of alleles found at 2 or more loci in a single individual. Eg. If at one locus, a diploid individual has an A allele and an a allele on the two chromosomes, and at the other locus you have Z and z, your MLG is \{A/a, Z/z\}
  - Multilocus Lineage (MLL): asexual descendants of a genotype differing only by mutation and mitotic recombination
  - Genet: distinct MLLs, essentially the MLG for a group of ramets
  - Ramet: sampling units of a genet

-Assigning Samples to MLGs

-After our definitions, we want to assign membership of our ramets to genets (ie. find all unique MLGs and MLLs in our sample)

-Why do we want to find all separate lineages? So that we can do accurate population genetics!
- By doing this, we get an estimate of the rate of clonal reproduction ie. how many clones for every sexual reproduction event do we get. This is c in the upcoming measurements. Need to determine the belongingness of samples to look at inbreeding, heterozygote selection, dispersal, migration via sexual versus clonal spread.

-Two questions we must ask to avoid overestimation of the number of MLGs in the sample.
  1. Were these two MLGs produced through sexual recombination and therefore not of the same genet?
  2. Has somatic mutations occurred in these ramets making it appear that they belong to different genets when they are actually in the same lineage?
-Genotype each sample for X loci, using this to assign MLGs to each sample ramet. Microsatellites are the best markers to use for their power.

**Genotype vs Clonal Membership:**

- We have the chance that a single genotype belongs to a clone, \( p_{gen} \)
- Under HWE, this is \( p_{gen} = \sum_{i=1}^{L} (f_i)^2 \)
  - \( L \) is the number of loci
  - \( F_i \) is the frequency of each allele at the \( i^{th} \) locus
  - \( H \) is the number of heterozygous loci in the sample

-This one has an issue of overestimation due to situations where the population is not in HWE, better to be more conservative in this instance
- Departures from HWE:
  \[
p_{gen}(F_{IS}) = \prod_{i=1}^{L} [(f_i g_i) \times (1 + (z_i \times (F_{IS(i)}))]^{2^h}
\]
  - \( I \) is the number of loci
  - \( h \) is the number of heterozygous loci
  - \( f \) and \( g \) are the allelic frequencies at the \( i^{th} \) locus, \( f \) and \( g \) are identical in homozygotes,
  - \( F_{IS} \) is the \( F_{IS} \) of the \( i^{th} \) locus
  - \( z_i \) is 1 if the \( i^{th} \) locus is homozygous and -1 if heterozygous

- \( p_{gen} \) gives the probability of finding an MLG when analysing only one sampling unit and does not give the probability of finding that MLG in the N sampling units collected/analysed.

- \( p_{gen} \) leads us to \( P_{sex} \) which will actually answer our question: what is the chance that this genotype arose from two different zygotes and therefore are not the same genet.

\[
P_{sex} = \sum_{i=n}^{N} \frac{n!}{i!(N-i)!} (p_{gen})^i (1-p_{gen})^{N-i}
\]
  - \( N \) samples
  - \( n \) times that this genotype was detected in your sample population

- This is used to determine if all replicates of the same MLG are part of the same clone

- The higher \( p_{sex} \) is, the more likely this genotype was created via sexual recombination and not clonality
-To decrease \( p_{\text{sex}} \) we should increase the number of loci being evaluated, thus making it less likely that 2 MLGs came from different zygotes

**-MLL Membership**

-Next to answer the second question posited before, we resolve all distinct clonal lineages in our sample (ie. Make sure each MLG belongs to an MLL as some MLGs will belong to the same MLL due to scoring errors and somatic mutations)

-Inspecting the frequency distribution of genetic distance among pairs of MLGs, and there is a lower threshold below which we can say that these MLGs are the same MLLs (ie. they are the same asexual lineage).

-This threshold is \( p_{\text{sex}} \). An example value is 0.01

![Genetic Distances](image.png)

-In a frequency distribution of genetic differences between MLGs there is a point below which the distribution of genetic differences is too low to deny the null hypothesis that distinct MLGs belong to the same genets. In essence, they are the same multi-locus linneage MLL

-Why would this occur? The accumulation of minor somatic mutations and mitotic recombination causes subtle variations. They are slightly different but actually belong together as an MLL

- After knowing the MLL membership of the samples, we can look at the population structure
**Why** do we need to know about clonal population structure? So that once the population genetics have been evaluated, we have information with which to make biological inferences. Also, pop structure informs genetic structure.

**Clonal Richness:**

-What is richness? The number of, in this case, MLGs, detected in the sample. It is dependent on our sample size.

-A useful ratio is \( R = \frac{(G-1)}{G} \) Monoclonal stand is 0 and 1 where all lineages are independent or no clonality

**Clonal Evenness:**

-Describes clonal equitability in your population

-How close in numbers each MLL is. Low evenness means the population is dominated by a few and high evenness is...even.

-Evenness allows us to compare populations with different numbers of clones in it. Compare the evenness of two populations.

**Clonal Heterogeneity:**

-Heterogeneity is the combination of species richness and diversity. A more heterogenous population has both richness and evenness.

-Importantly, Heterogeneity determines ecology and evolution in a population

-Quickly: an example of heterogeneity is the Simpson’s Index

-Simpson’s Index becomes a measure of estimating whether two randomly chosen individuals in a population will belong to the same clonal lineage

\[ \lambda = \sum_{i=1}^{G_{pop}} p_i^2 \] is the simpson's index

-The complement of Simpson's index, \( 1-\lambda \), is the chance of encountering 2 distinct MLLs when randomly sampling 2 ramets. 0 is no chance of them being different, 1 is that they are always different.

**Diagram of Richness, Heterogeneity, and Evenness**
III. Effects of Clonality in Population Genetics

Some General Effects of Clonality in Population Genetics

- Clonality creates strong correlation with alleles at different loci.
- Meselson effect occurs (accumulation of mutations without sex to recombine)
- Therefore, Heterozygosity will increase over time in clonal populations
- In completely clonal organisms, there is a lack of genetic drift due to no gamete formation, keeping heterozygosity
- High rates of clonality will increase heterozygosity and maintain up to TWICE the number of alleles at a locus (in a polyallelic situation) in a population
- Effective population size will increase and then go crazy as you get closer to complete clonality
- Higher genetic diversity, lower genotypic diversity (duh).
- Initial loss of diversity while the genotype is having selection act on it and clonal propagation occurs, but then there is an increase in genetic diversity as mutations accrue.

- So now we have a rate of clonality, so what?

- Now let’s consider how the F statistics change when clonality and migration are introduced.

F Statistics
- Quick refresh as this is important and also alternative definitions
- F is the inbreeding coefficient and the probability that two alleles drawn at random from a single individual share identity by descent
- Θ is F but for 2 alleles from different individuals in the same population, the coancestry of individuals
- α is the coancestry of alleles from 2 different random populations
- \( F_{IS} = \frac{\bar{\alpha} - \alpha}{1 - \alpha} \) identity within individuals compared to the identity between alleles drawn from the subpopulation. It is a measure of INBREEDING biologically, an amount of deviation from random mating.
- \( F_{ST} = \frac{\bar{\alpha} - \alpha}{1 - \alpha} \) is the identity of alleles randomly drawn from a subpopulation compared to the total population, and is a measure of heterozygote deficiency (remember Wahlund’s Effect) which is caused by population subdivision. A low Fst will show that a population is significantly different in its heterozygosity compared to the entire population.
- \( F_{IT} = \frac{\bar{\alpha} - \alpha}{1 - \alpha} \) The measure of inbreeding considering the individual from the entire population, caused by inbreeding and population subdivision.

The Within Population Subdivision \( F_{IS} \)

- Let’s start with getting the recurrence equations used in the model presented by Balloux et al. 2003

\[
F' = \gamma(cF + (1-c)(s(\frac{1+F}{2}) + (1-s)\Theta))
\]

- Let’s Break This Down

So the probability that two alleles in an individual share identity by descent is the chance that the two are still identical after possible mutation \( \gamma = 1-u^2 \) AND If the individual was reproduced clonally (which means that the alleles are all identical by descent) \( cF \) or if the organism was reproduced not through clonality \( (1-c) \) where \( c \) is the rate of clonal propagation, and if it was produced via self-fertilization, rate \( s \), then it has \( (1+F)/2 \) chance of being identical again, or if it was produced via sexual recombination with another individual at rate \( (1-s) \) then there is a \( \Theta \) chance that the alleles share identity by descent in the population. WOO RECURSION AND BIOLOGICAL EXPLANATION.

- This is only for an individual however, and we must include interesting migration probabilities when considering the \( \Theta \) and \( \alpha \) of a clonal population. The migration probabilities are captured in \( q_s \) and \( q_d \).

- \( q_s \) is the probability that two individuals taken at random from the same population after migration were born in the same population. This helps correct for spatial autocorrelation.

- \( q_d \) is the probability that two organism from different populations taken after migration were originally from the same population.
- Don't focus on this too much as it turns out that migration isn't the most important thing with clonal populations.

- Alright, so we know the recurrence equations, and we know the official definitions of $F_{is}$ and $F_{ST}$, we can plug in and make inferences!

- What is $F_{is}$? Measure of identity of alleles within individuals relative to the identity between alleles drawn randomly from the population. It represents the deviation from random mating cause by the reproductive system

- When completed, $F_{is}$ becomes:
  $$F_{IS} = \frac{\gamma(q_1 - c)^2(q_1 - q_2)^2 - 1}{2N(1 - c)(q_1 - q_2)^2(q_1 - q_2)(1 - q_1)(1 - q_2)(1 - q_3)(1 - q_4)}$$

- So we can see that as clonality increases, $F_{is}$ will decrease, hitting negative values even for non-selfing organisms. A negative $F_{is}$ value means there is a large heterozygote excess. Essentially, everyone is heterozygous for different things. This also shows that the system is very far departed from what is expected under random mating. This can allow for extra alleles to be stably sustained in the population. But if mutation still occurs, it is impossible to get -1 $F_{is}$. In a graph of $F_{is}$ vs. c, there is a long period until around 95% clonality where $F_{is}$ resembles the expected values under panmixia, if a little lower. When completely clonally reproducing, everyone will be heterozygous even with mutation therefore it $F_{is}$ will be negative.

- $F_{ST}$, how much are populations different from one another? This changes under clonal reproduction as well. Heterozygote deficiency due to subpopulation

- We skip to neglecting mutation, as the base equation is very long like before.
  $$F_{ST} = \frac{(1 - c)(q_1 - q_2)}{N(1 - c)(1 - q_1)(1 - q_2)}$$

- $F_{ST}$ becomes strongly reduced by high amounts of clonality

- When there is complete clonality, Fst is 0 so populations are very similar to one another (which makes sense as every unit is a clone). Otherwise if there is even a little bit of sexual reproduction, Fst won't change that much in a system, much like Fis.

- In both $F_{is}$ and $F_{ST}$ the changes due to clonality aren’t very apparent until around a clonal reproduction rate of 0.9

- Draw graph of Fst to clonality
Effective population size

-SHOULD NOTE THAT THESE MODELS DO NOT INCLUDE NATURAL SELECTION
-As long as the selection differential s is not much greater than 1/N, while this is true, genetic drift is more important.

IV- A Case Study in the Effects of Clonality in Population Genetics

-Meloni et al 2013 Effects of Clonality on the genetic variability of rare, insular species: the case of *Ruta microcarpa* from the Canary Islands

-The title is the goal of the study
-*R. Microcarpa* is a Rutaceae shrub found on La Gomera of the Canary Islands. So it is isolated and in a tough environment
-They took 73 individuals from the island and extracted DNA, finding 9 loci to use to genotype them. They had to sample greater than 10m from each plant to avoid re-sampling the same ramet, another issue of clonality, you have to account for it in sampling design.
-Then they did work very similar to what I had described.
-Assigning MLGs manually, using an analogue of $P_{sex}$ to estimate the probability that two individuals randomly sampled from the same population don’t have the same MLG due to random chance or sexual reproduction (ie. came from 2 different zygotes).
-Then they used a program to get the pairwise genetic distance frequencies for all of the samples and set a threshold below which samples were considered to be within an MLG. This is to prevent overestimation of the number of MLGs caused by somatic mutation.
-They calculated clonal richness of using G/N (the less useful measure)
-Calculated MLG diversity index $D_o$ where 0 is one dominant clone and 1 is everyone has a different genotype, much like the Simpson’s Diversity Index
-And finally calculated Evenness, which allows for comparison of populations having different numbers of clones in terms of diversity. 0 is where there is one genotype for the population and 1 is every genotype is represented at the same frequency.
-Based on their results, they found 17 MLGs between all populations with no MLG being shared. They used this as $N$.
-Results: the researchers found that there were an excess of heterozygotes in departure from Hardy Weinberg! An average of about 1 MLG per 5 ramets sampled, with 3 populations being dominated by a single genotype and others being fairly even. Interestingly, one population that was dominated by a single MLG was just across the road from a fairly even in its distribution of MLGs and it didn’t share any! So as you can see, clonality helps deal with dispersion issues. The overall $F_{st}$ was also really high at 0.446, showing that each population was quite different.
-They concluded that clonality helps counteract the negative effects of "small population size and isolation by increasing heterozygosity, polymorphism, and allele richness". The amount of genetic variation is higher than expected given the situation and it is attributed to clonality.

Literature Cited:


