Week 6 – Population Genetics

Summary: Discussion based on papers that examined how coalescent theory (Eldon and Wakely, 2006 *Genetics*) and how continuous vs. discrete spatiatl distributions (Bradburd et al. 2018, *Genetics*) affect genetic structure and differentiation.

16 groups

74 + participants

**1. What are the implications for sweepstakes reproduction on how we model evolutionary processes in the ocean?**

**WSU:**

* If SR is big term in model, wouldn’t be able to model evolutionary processes as we see them in lab/observations
* Need to track individuals through time to see if sweepstakes is truly random
* Looking at input/output distributions for mero/holoplanktonic larvae in the same environment, you can evaluate the relative importance of SR

**Hal-Dames:**

* Could cause population bottlenecks and impact signatures of local adaptation.
* Effective population size is not capturing the number of individuals that pass alleles to the next generation, so won’t scale with TMRCA and pops can diverge quickly.
* Evolutionary history of alleles may be difficult to detect due to repeat bottlenecks.
* How common is SR? Is it plastic?

**Rutgers:**

* SR can play large role.
* Creates large amounts of variability in number of viable offspring/recruits, which increases chances of multiple mergers in coalescent process.
* If variability is large or frequent enough, could cause uncertainty in the effectiveness of simpler models that estimate pop. Size or genetic diversity.

**NoCal:**

* SR can accelerate adaptation if nonrandom in regard to parental genotypes.
* In long term, might reduce genetic variation in populations and decrease adaptive potential.
* How common is SR? assume it is rare or may reduce genetic variation.

**Umass Amherst:**

* How does modeling SR differ between marine and terrestrial systems?
* SR can cause such extreme variance in reproductive success, and ultimately result in substantial temporal changes in allele frequencies in larval systems.
* Suggest sampling across size classes or age classes in addition to unbiased sampling to avoid inaccurate inferences.

**NEU:**

* Literature seems to focus on selection and drift lately, but there are other factors.
* Which dominates larval stage with genetic patchiness? Lately it is selection, earlier was drift.
* Both are at play, but still don’t know if one dominates.
* Should have strong considerations when applying moran model – most analyses assume wright fisher model which may cause erroneous conclusions if the species displays evidence of SR

**Cornell:**

* With SR, expect reduced genetic diversity, increased relatedness, and more rare variants.
* Could have implications for modeling evolutionary processes, especially given the assumptions of coalescent theory and wright-fisher model (assumes small variance in # of offspring, no sweepstakes).
* In Eldon + Wakeley’s model, extends coalescent, offspring of single individual can replace substantial fraction of population.
* Deals with variance in offspring, how much does it improve it compared to traditional methods? Is it closer to reality?
* How would this perform with nuclear markers?

**USC:**

* SR changes null expectations as previous coalescent methods estimate inaccurately low effective population size.
* Selection at larval stage could have implications for recruitment.
* Moran model probably always better as null model than wright-fisher coaslescent models.

**Laval:**

* SR is the chances of a highly fecund marine animal contributing offspring to a future pool of reproductively mature adults, a process envisioned to have few winners and many losers.
* In study, Eldon and Wakely highlight coalescent models for accounting for SRS parameters, which is substantial advance.
* Do we need to implement some intrinsic parameters relative to species of concern?

**MIT:**

* SR acts as bottleneck and can dramatically reduce genetic variation compared to what we would expect if every individual contributed to next generation.
* In coalescent, multiple mergers become important and coalescence events happen more quickly.
* Models that assume wright-fisher processes rather than accounting for rare reproductive eventus with a modified moran process will tend to underestimate the effective population size if SR is important.

**CSU MontBay:**

* In populations with high variance in reproductive success, coalescent process departs from Kinman coalescent so much that the concept of coalescent population size no longer holds. If this is true, can fit Moran model to a population that shows gamma between 1 and 2, then all bets are off with traditional pop gen.
* When family size of reproductive event is large (greater than 25%), then rule of thumb that sample size can be low for coalescent theory no longer applies. Doubt that it would ever be that large.
* In region of parameter space, multiple mergers can lead to star phylogenies and neutrality tests violated in the negative direction. Frequently happens and is usually interpreted as population expansion, selective sweep, or purifying selection. But may be none of the above!

**UT Arlington:**

* SR creates rush for resources, which could select for most advantageous traits (through bottleneck) and promote traits promoting resource attainment and management.
* Longer time scale, genetic diversity could be hindered.
* Depends on scenario and environmental variables on when it promotes or obstructs adaptation and how rapid adaptation comes at the expense of genetic variance in long term.

**MSC:**

* Commonly used models assume equal numbers of offspring and discrete populations, so violations like SR can lead to incorrect inferences.
* Authors say coalescent with multiple mergers is better, but even in this model simultaneous mergers are not allowed, so how important is that?
* Wondered how this would play out with nuclear markers.
* Discussed if SR is drift or selection and how it plays role in evolutionary processes.

**LSU**:

* Most models do not take SR into account.
* Source variation is layered, so genetic drift could potentially be misinterpreted with models not accounting for SR.
* Recency of common ancestor, allele frequency vs. kingman model could falsely conclude that an individual has high fitness among a population. So we need effective population size.
* How large of a sample size is sufficient?

**UQ:**

* SR is probably very important in how we model evolutionary processes in ocean.
* More interrogation into what data would look like with 0<gamma<2 would’ve been nice.
* Should be on lookout for multiple mergers and understand consequences on selection.

**2. Given the complexity of marine dispersal dynamics, do you think we as a field should adopt Bradburd et al. approach for characterizing population structure?**

**WSU:**

* No, need more information.

**Hal-Dames:**

* Might help extract the effect of space, help overcome some issues in detecting local adaptation vs. isolation by distance.

**Rutgers:**

* Can these programs resolve the clusters vs clines problem?
* Compelling way to look at population structure in systems where we think clines may be most responsible for shaping population structure, as ability to run both spatially explicit and non-spatial models gives a large degree of flexibility.

**NoCal:**

* What are assumptions of construct? What are effects of migration/dispersal on construct outcome?
* Does construct assume restricted dispersal?

**Umass Amherst:**

* Spatial components can be case specific. Temporary congregations of individuals can lead to inaccurate inference when identifying population structure. Another source of bias could be invasion populations.
* Adopting approach beneficial to characterizing pop. Structure, but agreed it’s a case-specific process and sampling scheme design is critical.
* Idenfiying populations and management units very important.

**NEU:**

* Bradbury approach is useful when analyzing pop structure in species with isolation by distance.
* STRUCTURE has been applied to many studies studying pop structure of species, but assumptions of model are not always considered carefully.
* Like idea of investigating few different K’s rather than picking single one – different k values can help identify large scale vs. smaller scale processes that might be shaping population structure.
* Limitation is that dispersal distance does not equal geographic distance, and membership in layers does not always reflect this.

**Cornell:**

* construct is expected to perform same as traditional methods when there is no IBD, we wish they had applied it to more no IBD species to confirm.
* Agree that tool is improvement, but may not be flexible enough to apply to marine systems with chaotic patchiness from sweepstakes could generate strongly non-IBD patterns.
* How prevalent is IBD in marine systems, and what spatial scale do we expect to see this pattern?
* Need to make sure construct is appropriate for species/distribution/sampling before we can sample it.

**USC:**

* Long dispersal distances may make it difficult to apply method that leans on spatial autocorrelation.
* But isolation by distance is observed in many marine systems so application of bradburd model is viable.
* We should account for migration and dispersal patterns within their system as strong currents can create areas of close relation across large spaces.
* Bradburd model has more conservative method for estimating number of clusters whereas STRUCTURE and its variants tend to overfit data.

**UChicago/UGA:**

* Bradbury approach useful for pop gen surveys in which interplay of both discrete and continuous components both important.
* construct requires subjective interpretation by the user (multiple K’s, not best K). Subjectivity is one of greatest issues for standardizing field and facilitating cross species comparisons.
* Perhaps instead of single option, validation through multiple agreed upon methods is best (each with different assumptions etc.).
* Approach could generate reproducible datasets along a species range to devise generalities for marine population structure across species range.
* May want to consider user friendly methods, well maintained by authors.
* Also need to find selectively neutral loci, which depends on robust filtering of putatively adaptive loci. Advancements here may be very useful because even small number of non-neutral loci can dramatically alter results.

**Laval:**

* Yes, until a certain level of genetic relatedness.
* Power of approach is that ultra high connectivity often leads to low Fsts.
* What is tradeoff between having many markers or just informative markers? What kind of markers should be used? Maximise number of markers or use only informatives (like Fst outliers?)

**MIT:**

* Continuous genetic variation is probably often important in marine systems so we think the approach of Bradburd would be a good way to generate hypothesise for organisms whose population structure is not known – but shouldn’t be only framework used.   
  Need to evaluate assumptions relative to the study system.
* Bradburd depends on spatial relationships between samples. But dispersal distance doesn’t always correspond to geographic distance. Further, pelagic organisms can have enormous ranges and it isn’t clear. Isolation by distance more important for certain species.

**CSU MontBay**

* Bradbury approach relevant but only relevant to extent.
* Most marine species in “waples zone” of high Ne and high migration, and don’t see evidence that this method will help work in this zone.
* Should be thinking about pop structure based on measured gene flow between populations rather than measured allele frequency differences.
* Coalescent sampler (migration – n) can differentiate populations and identify underlying metapopulation model when Nem is as high as 100 migrants per generation.

**UT Arlington:**

* Spatial model is interesting and will be better approach for some systems, but should not be used to generalize.
* Dispersal distance doesn’t always correlate with geographic distance.
* Model is good step but should be evaluated on system by system basis.

**MSC:**

* Bradbury approach useful in some situations, but not universal panacea.
* Could be modified to deal with other forms of spatial structure in addition to IBD. Best approach comes down to questions and resources.

**LSU**:

* Appropriate for statisticians. However, dispersal distance vs. geographic distance different and important to take barriers into account.
* Both STRUCTURE and ADMIXTURE are great tools, but shouldn’t be used alone. Other models should confirm results. As always, with population data, sample size is extremely important factor, so is question.
* Not 100% on accepting Bradbury approach as field.

**UQ:**

* Bradburd method is super cool but doesn’t help SR problem. Detecting lots of K’s with geographic restricted shifts in allele frequencies if sweepstakes is pervasive.

**3. Given the potential for high variance in reproductive success, overlapping generations, and a combination of continuous and discrete population structure, how should we sample for population studies? (Across space and cohorts)**

**WSU:**

* The more you know about your system, the more appropriate design you can implement.
* Important thing is variance structure in the data, not ecological/evol. Hypotheses.
* Different populations/questions will always differ in their approach, but may be able to address by understanding underlying variance structures in the data.

**Hal-Dames**:

* What about software to help figure out what sampling regime might overcome issues in your system?
* Difficulty in sampling makes overcoming some of these issues nearly impossible.
* Instead of increasing fine-ness of scale, adjust inference about data we have.
* Sampling design contingent on species, environment, and question.
* Simulations needed to determine sampling regime optimal – generate result or model that can be parameterized with system conditions (amount of variation in reproductive success) to determine targeted sampling regimes.

**Rutgerss:**

* Need for larger sampling efforts (more individuals/population) especially in SR systems.
* Systems in highly variable reproductive success will be comprised of individuals that will coalesce back to MRCA in a few generations, most of genetic diversity harbored in a few rare individuals. The more uncommon these individuals are, the larger the sampling effort will need to be to give them a fair chance of being sampled and better represent level of diversity actually present in study population.

**NoCal:**

* If SR has taken place and we use coalescent theory, most of alleles will trace back to same ancestor. If we want to learn about evolutionary potential of species, would have to include several populations and sample deeply to infer most recent common ancestory.
* For construct to pick up isolation by distance, would need to sample across distribution and not just at the extremes.
* How do offspring by SR distribute in space? Will there still be isolation by distance if sweepstaking in a population is strong?

**Umass Amherst:**

* Agree that this is case specific – biology should inform sampling design.
* Reducing geographic scope on a particular site should help reduce noise associated with high variance in reproductive success (in addition to assessing continuous vs. discrete population structure).
* Should consider applying eldon and wakely model to account for organisms that have overlapping generations as opposed to Wright-fisher model that doesn’t.
* To reduce bias in sampling, perhaps vary sampling sites over a temporal scale to cover both spatial and temporal scale – but longer lived organisms may not need temporal consideration.

**NEU:**

* Designing sampling scheme is complex and often limitations in how well you can sample.
* When researcher is trying to understand pop dynamics, important to know system and design study to sample across cohorts when possible.
* Even more critical in species with SR.
* Sampling should be more extensive across multiple years because most individuals could be highly related under SR.

**Cornell:**

* Sampling should be highly dependent on study system. More extensive spatial and/or temporal sampling may be needed in some scenarious.
* IBD is common at some scales, but may be at larger scales than is typical of terrestrial systems, so may need to sample more widely.
* Not clear whether random sampling is preferable to stratified sampling based on hypothetical barriers to dispersal.
* Sequencing pools may be better than sequencing individuals. If variation in pool, can we account for it using construct?

**USC:**

* Knowledge of system (life history) will yield best methods for distinguishing generations and variance in reproductive success.
* When there are overlapping generations, binning according to factors like size class could help disentangle separate generations.
* Genetic markers across multiple loci have power to determine population structure better than previous methods. Especially if this approach is combined with taxon specific knowledge could help with initial parsing of complex systems and could be used as scaffold of experimental design.

**UChicago/UGA:**

* High variance in reproductive success warrants investigation into the stability of inferred structure through time.
* Sampling method appropriate would be re-sampling over multiple (2-3) generations of a focal species and replicating computational methods to see how structure estimates vary.
* Sampling year/generation – need to standardize approximate age/life history stage of individuals
* Revised methods for field sampling may be useful: fine resolution around geographic breaks?
* Unclear how much bias will affect results, or if it would just improve defining fine-scale clusters.
* Within population sampling important where patterns of microgeographic adaptation are at play. Sampling across range of microhabitats within a population likely most robust means to correct for this. Sampling should be consistent across populations.

**Laval:**

* Impact of SRS could be different over time: population size fluctuations (temporal structure, bottlenecks, temporal seasonality) over different time-scales should be considered.
* Sampling over time and over cohorts seems a good approach to catch the SRS effect.
* However, no matter the sampling approach used (follow population through time or sit in one spot and study what is recruiting locally) – each is valid, and when combined, can provide powerful understanding of evolutionary change in marine populations.

**MIT:**

* Sampling regime should reflect biology of study system. Efforts should be made to sample across time and space, and sampling across cohors can become important with species with high variance in reproductive success.
* In marine systems, often critical to understand the physical oceanography of study system, especially for species that have current driven dispersal.

**CSU MontBay**

* Sampling depends on goal, but for coalescent studies, sampling should take place across cohorts present in population.
* Someone should design method based on Eldon and Wakeley to quickly determine whether sampled population sits in Moran space of 1 < gamma < 2 (SRS).
* Curious how frequently this happens in nature, not just in oysters.
* Noted that long generation times should even out generation to generation variance created by SRS.
* For studies based on coalescence, each genome is product of multiple coalescent processes, so large sample size may not be important if you sample widely across genome.

**UT Arlington:**

* Question depends on study system, questions, context.
* Range and dispersal should be weighed, while temporal parameters (gen time) must be accounted for.   
  Good to conduct preliminary study to look at pop structure and life history and discuss what definition of population is being used.

**MSC:**

* This may be productive discussion topic at working group meeting – because optimal sampling method will depend on question being asked and system, may be challenging to come up with guidelines.

**LSU**:

* Dynamic question that could be addressed a few different ways. Could sample across space and cohorts by using preserved historical samples. Could also give insight into historical sweepstakes events.
* Other way is to sample population studies given all of dynamic variables involved to ensure large sample size to account for rare alleles or events.

**UQ:**

* Time is as important as space. Not easy to do and hard to be funded.