PROJECT SUMMARY

Overview:

The availability of large multilocus datasets is becoming increasingly common in systematic biology, leading to improvements in phylogenetic inference. However, such datasets can also make species tree inference more difficult because there may be conflicting signals across the genome. Though evolutionary models provide expectations for these discordant signals, a number of processes can lead to deviations from these expectations. In particular, ancestral structure, which has been uncovered in many diverse species, can skew gene tree frequencies, thereby hindering the performance of many methods for estimating species trees. Here, the PIs propose to develop novel likelihood-based methods that can infer phylogenies in the presence of ancestral structure, and to apply these methods to genomic data from several model and non-model organisms. The model organisms considered will be mouse, yeast, and mosquito, for which previous studies have observed skewed gene tree frequencies that may be the result of ancestral structure. The PIs will also apply their methods to the threatened non-model coral group, Orbicella, which is of particular interest because morphological data provide evidence for hybridization, suggesting that the taxa exhibit skewed gene tree distributions. Thus, the objectives of this proposal are to design model-based methods for inferring species relationship in the presence of ancestral structure, and to apply these methods to genomic data from model and non-model organisms to shed light on the evolutionary dynamics of these systems.

Intellectual Merit :

The proposed work involves development of novel likelihood-based methods for inferring species trees in the presence of ancestral structure, and application of these methods to several model and non-model organisms. As ancestral structure has been uncovered in a number of diverse species, our methods will enable systematic biologists to infer more accurate species trees across taxa. In particular, our methods will model how ancestral structure skews gene tree frequencies, thereby aiding researchers in distinguishing between ancestral structure and hybridization. Because skewed gene frequencies in mouse, yeast, and mosquito have been used as evidence for hybridization, we will apply our methods to revisit this hypothesis and elucidate the evolutionary histories of these systems. We will also apply our methods to the threatened non-model coral group Orbicella, for which evidence of hybridization was recently uncovered. The Orbicella group is of particular interest because they compose 25% of marine biodiversity and are rapidly declining. Thus, understanding their evolutionary dynamics may provide insight into current and future conservation approaches.
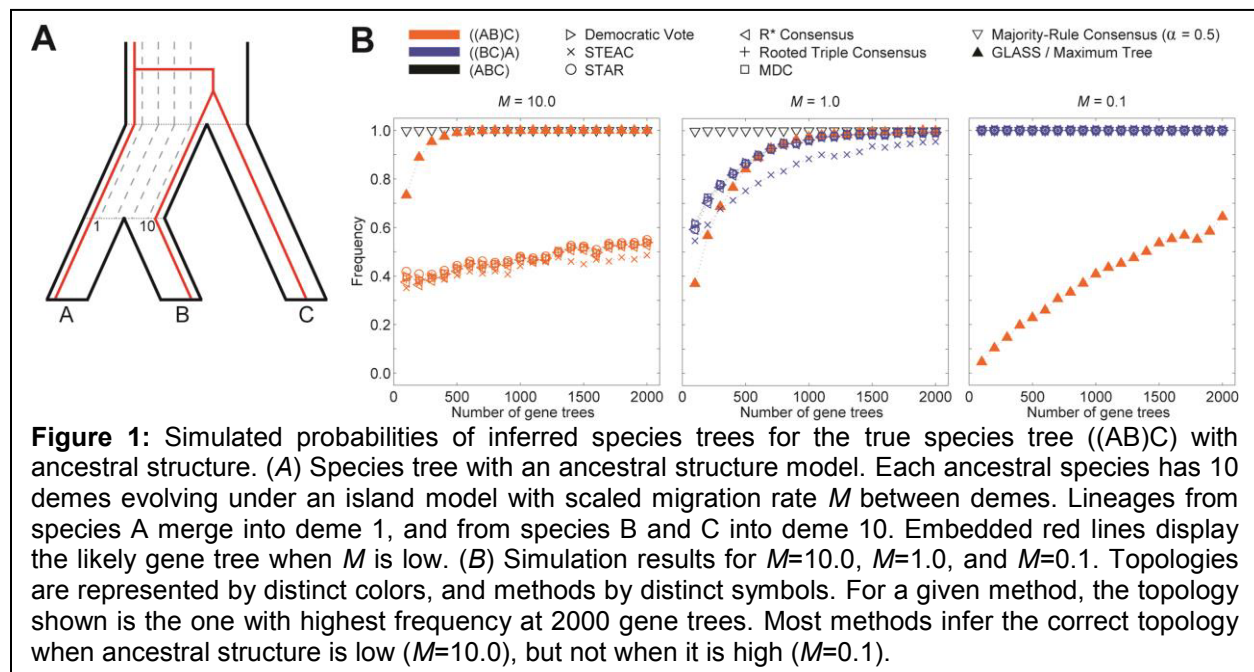
Broader Impacts :

All methods developed will be available as open source, findings will be published in open-access format, and short general-audience videos describing results will be created and hosted on the PIs websites and YouTube. The PIs will also continue to provide outreach to women and underrepresented groups in the sciences. In particular, Dr. DeGiorgio will continue serving as an organizer, advisory committee member, and instructor of the Summer Internship for Native Americans in Genomics workshop, which introduces Native Americans from different educational backgrounds to genomics and its associated legal, ethical, and social issues. To advance this program, Dr. DeGiorgio will continue his efforts to keep the curriculum up-to-date based on the latest research in evolutionary genomics, and adapt the curriculum to the needs and interests of the students. Further, Dr. DeGiorgio hopes to host future workshops, exposing students to the diverse biological research at Penn State. He will also develop hands-on phylogenetic activities for K-12 students by participating in the Penn State Science-U program. Dr. Medina will continue to develop courses that provide hands-on experiences in marine science and genomics. She is also active in both Penn State's Eberly College of Science outreach programs and in K-5 educational coral reef science activities in the State College School District. Further, Dr. Medina will continue organizing a cnidarian genomics workshop for scientists. Finally, the PIs will attend and present findings from the proposed work at annual SACNAS meetings, and will use this conference as a venue to recruit underrepresented students to work in their labs.

1. Background

The use of large multilocus datasets is becoming increasingly common in systematic biology (Cranston *et al*. 2009, Song *et al*. 2012, McCormack *et al*. 2013, Salichos and Rokas 2013, DeGiorgio *et al*. 2014). However, while such data generally improves phylogenetic inference, it can also result in gene tree discordance, in which tree topologies inferred for different loci do not match. A major source of gene tree discordance is incomplete lineage sorting, which occurs when sampled lineages fail to coalesce, or find a common ancestor, in the population in which they are first able. For this reason, the performance of a species tree estimator is often tested for robustness to incomplete lineage sorting under the multispecies coalescent model (e.g., Degnan *et al*. 2009, Mirarab *et al*. 2014).

The multispecies coalescent assumes that each modern and ancestral species is unstructured and has a constant population size, and that each pair of lineages within a given ancestral species has an equal chance of coalescing (Nakhleh 2013). These assumptions lead to symmetric gene tree distributions. For example, if the true species tree relating taxa A, B, and C has a branching pattern ((AB)C), for which A and B are sisters and C is an outgroup, then the frequencies of gene tree topologies ((AC)B) and ((BC)A) are equal. Common processes that lead to asymmetric gene tree distributions are continuous gene flow among species, horizontal gene transfer, and hybridization (Yu *et al*. 2011, Yu *et al*. 2012, Leache *et al*. 2014). A key distinction between these processes and incomplete lineage sorting is that species boundaries are blurred, imposing a reticulate rather than a tree-like relationship.

However, the true species relationship can be tree-like, despite asymmetric gene tree distributions, when ancestral species are structured (Slatkin and Pollack 2008). There are several examples of ancestral species that are structured (e.g., Garrigan *et al*. 2005, Thalmann *et al*. 2007, White *et al*. 2009), and genomic signatures of ancestral structure have been uncovered in a number of diverse lineages, including mouse (White *et al.* 2009) and yeast (Yu *et al*. 2012). Our preliminary results (Figure 1) suggest that many methods for inferring species trees from multilocus data are not robust to ancestral structure. An exception is the maximum likelihood estimate of the species tree, GLASS (Mossel and Roch 2010), though it can be misled by inferred gene trees (Leaché and Rannala 2011, DeGiorgio and Degnan 2014).

Figure 1: Simulated probabilities of inferred species trees for the true species tree ((AB)C) with ancestral structure. (*A*) Species tree with an ancestral structure model. Each ancestral species has 10 demes evolving under an island model with scaled migration rate *M* between demes. Lineages from species A merge into deme 1, and from species B and C into deme 10. Embedded red lines display the likely gene tree when *M* is low. (*B*) Simulation results for *M*=10.0, *M*=1.0, and *M*=0.1. Topologies are represented by distinct colors, and methods by distinct symbols. For a given method, the topology shown is the one with highest frequency at 2000 gene trees. Most methods infer the correct topology when ancestral structure is low (*M*=10.0), but not when it is high (*M*=0.1).

We propose to develop novel methods that can infer phylogenies in the presence of ancestral structure, and to apply these methods to several model and non-model organisms. The model organisms we will focus on are mouse, yeast, and mosquito, for which previous studies have found asymmetric gene tree distributions that may be a result of ancestral structure. We will also apply our methods to the non-model group *Orbicella*, a coral species complex that is currently listed as threatened under the Endangered

Species Act. *Orbicella* is of particular interest because morphological extant and fossil data provide evidence for hybridization (Budd and Pandolfi 2004, 2010), suggesting that gene tree frequencies are skewed. Moreover, we have generated ideal data for the application of our methods, with at least 50x coverage genomes from three strains of *O. faveolata*, one *O. annularis*, and one *O. fanksi*. Further, understanding more about the evolutionary history of *Orbicella* may be helpful in conservation efforts.

2. Specific Aims

The objectives of this proposal are to develop novel likelihood-based methods for inferring species trees when there is ancestral structure, and to apply these methods to data from model and non-model organisms. In particular, we will revisit hybridization hypotheses in mouse, yeast, and mosquito lineages by using ancestral structure as an alternative model, and will also refine species boundaries and study evolutionary dynamics of the reef building Caribbean coral *Orbicella*. Future studies can use our techniques to obtain more accurate phylogenies in a number of diverse taxa.

Aim 1. To develop novel approaches for inferring species trees in the presence of ancestral population structure, and to examine their performance relative to current state-of-the-art methods.

Aim 2. To apply our methods to investigate the evolutionary processes shaping genetic variation in mouse, yeast, mosquito, and *Orbicella*.

3. Development of model-based methods to infer species trees under ancestral structure (Aim 1)

We will develop a likelihood method for inferring species trees that is based on the Slatkin and Pollack (2006) ancestral structure model depicted in Figure 2. Though the extended population structure in the model is an oversimplification, it may be plausible for many groups of marine or terrestrial organisms periodically separated by barriers to gene flow through environmental changes such as sea level fluctuations or glaciations (e.g., Toms *et al*. 2014). This model leads to asymmetries in gene tree distributions and thus provides an alternative scenario to other models (e.g., hybridization) that can similarly impact gene tree distributions.

By applying the model in Figure 2 to a single lineage sampled per species, we will construct Markov chains describing dynamics under an isolation with migration model within each of the two ancestral species (e.g., Hobolth *et al*. 2011). We will compute gene tree probabilities from the Markov chains using the matrix exponentiation technique provided by Hobolth *et al*. (2011). A similar approach was taken by Tian and Kubatko (2015) to obtain gene tree probabilities under a model with continuous gene flow between species. Using this framework, we will compute the probability density of a gene tree (topology and branch lengths) given the model parameters (species tree topology, divergence times, migrations rates, and population sizes). As is common with likelihood methods for inferring species trees from multilocus data, the input data will be a set of estimated gene trees. Using estimated gene trees makes the calculation of the likelihood computationally tractable, as the likelihood is simply the product across loci of the probabilities of observing the estimated gene trees under the model. We will choose the species tree and associated model parameters (Figure 2) that maximize the likelihood.

Figure 2: Model of the relationship among species A, B, and C in a species tree with divergence times t_1 and t_2. Ancestral species each have two demes, with size θ_k in deme $k=1,\ldots,4$. Symmetric migration occurs between demes at rate m_1 and m_2.

The above formulation calculates the likelihood of a three-taxon species tree with ancestral structure. However, data for many systems include more than three species. Thus, we plan to use our framework to also construct *n*-taxon species trees under ancestral structure. Because the full *n*-taxon model is computationally prohibitive, we will use approximations to infer the full *n*-taxon species tree. As a first approach, we will estimate all sets of rooted three-taxon tree topologies under our three-taxon model using multilocus gene tree estimates. Because a rooted *n*-taxon tree is defined by its set of rooted triples (Steel 1992), we will use a supertree approach (e.g., DeGiorgio and Degnan 2010) to estimate the *n*-taxon species tree topology from the set of rooted three-taxon topologies. The performance of this
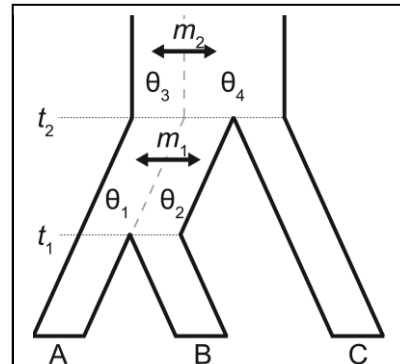
approach will depend heavily on the chosen supertree algorithm. However, the estimation of each rooted triple can be parallelized, thereby enabling a computationally efficient solution. We will also apply a second approach involving maximum pseudo-likelihoods (e.g., Liu *et al*. 2010, Yu and Nakhleh 2015). For this approach, we will assume a model *n*-taxon species tree with branch lengths, and consider a simplification of the ancestral structure model in which the migration rates (*m*) and population sizes (θ) are identical for all branches. This assumption substantially reduces the parameter space of the model, while also permitting a structured model in a larger tree. The likelihood of this extended model is then the product of the likelihoods of the model, conditional on input gene trees restricted to three taxa.

To evaluate our methods, we will use *ms* (Hudson 2002) to simulate gene trees under species trees with ancestral structure, and will examine a range of model parameters, including population sizes, divergence times, migration rates, and numbers of demes. We will employ *Seq-Gen* (Rambaut and Grassly 1997) to simulate sequence alignments that are conditional on these trees for a range of mutation rates and alignment sizes. Gene trees will be inferred using maximum likelihood (e.g., Felsenstein 1989) and Bayesian approaches (e.g., Ronqiust and Huelsenbeck 2003), and will be used as input for our methods. We will investigate the accuracy of species tree estimates in the presence of ancestral structure, assess the ability to infer species tree topologies and underlying parameters, and examine robustness to sources of error, such as gene tree uncertainty (Liu *et al*. 2010, Yu *et al*. 2012) and recombination (Lanier and Knowles 2012). We will also consider how input type (*i.e.*, full gene trees or topologies) affects inferences. Finally, as complete genomes are becoming available for a variety of species, accounting for correlations among loci may aid in correct inference of species relationships. We will therefore also consider coalescent hidden Markov model approaches (e.g., Hobolth *et al*. 2007, Dutheil *et al*. 2009, Mailund *et al*. 2012, Liu *et al*. 2014) as potential alternative solutions.

4. Elucidating past evolutionary dynamics of model and non-model organisms (Aim 2)
Coral reefs are the most diverse marine ecosystems on Earth, harboring at least 25% of marine biodiversity (Kleypass 1997, Reaka-Kudla 1997, Spalding *et al.* 2001) in less than 1% of the marine environment. Despite their diversity and billion dollar industry (Contanza *et al.* 1997), reefs are disappearing rapidly, particularly Caribbean reefs, which have experienced over 50% loss in coral cover in the past few decades (Jackson *et al.* 2014). As a result, dominant species, such as those in the genus *Orbicella* (*O. annularis*, *O. franksi*, and *O. faveolata*), are now listed as threatened (Aronson *et al*. 2008, Williams *et al*. 2015). Because *Orbicella* compose an integral part of the marine ecosystem, further study of this group is important for current and future conservation efforts.

Despite the importance of studying corals, delineating the most fundamental units for coral conservation—species—has been difficult, making phylogenetic inference challenging. These difficulties are partially due to the plastic morphologies, lack of synapomorphies, and relatively slow rate of mitochondrial evolution (Hellberg 2006) in corals. Here, we propose using the methods developed in Aim 1 to make phylogenetic inferences that will help refine species boundaries and elucidate the evolutionary dynamics of the reef-building Caribbean coral *Orbicella*. This system is ideal for the application of our methods for several reasons. First, *Orbicella* has few species, making this system a good starting point to study coral evolutionary relationships. Second, *Orbicella* species exhibit morphological differences (Pandolfi and Budd 2008), have different depth distributions (Levitan *et al*. 2010), and display postzygotic and spawning time isolation (Levitan *et al*. 2004). Third, studies have uncovered structure within *Orbicella* species (Severance and Karl 2006, Foster et al. 2012). Fourth, morphological extant and fossil evidence suggest that hybridization exists within *Orbicella* (Budd and Pandolfi 2004, 2010), implying a potential skew in gene tree frequencies. Finally, we have generated an ideal dataset for the application of our methods, with at least 50x coverage genomes for three strains of *O. faveolata*, one *O. annularis*, and one *O. fanksi*. In particular, our data consist of strains from Panama (all three species), Florida (*O. faveolata*), and Mexico (*O. faveolata*). However, because hybridization levels are thought to differ in the Bahamas (Levitan *et al.* 2004), and Puerto Rican populations are genetically differentiated from those in Panama and the Bahamas (Foster *et al*. 2012), we plan to also sample strains from the Bahamas and Puerto Rico. To refine our understanding of the evolutionary relationships among the *Orbicella* species, we will infer gene trees (Felsenstein 1989, Ronqiust and Huelsenbeck 2003) from multilocus alignments. We will apply our maximum likelihood procedure to gene tree estimates to infer the species tree topology and associated model parameters in our ancestral structure model (Figure 2). Further, from the set of inferred

gene trees, we will estimate the species tree topology and associated model parameters for three other maximum likelihood procedures: the standard multispecies coalescent model (Kubatko *et al*. 2009, Wu 2012), a model of continuous gene flow between species (Tian and Kubatko 2015), and a model with hybridization (Yu *et al*. 2014). Depending on the amount of variation in the alignments, confidence in gene tree topologies may be higher than in full gene trees. The likelihood under each of these models can also be computed for data based only on topologies, rather than on full gene trees, and we will choose the best approach based on initial exploration of the data. Because not all of the models we consider are nested, we will perform model selection by comparing their likelihoods using the Akaike Information Criterion (Akaike 1973, Sullivan and Joyce 2005). Using this procedure, we will simultaneously infer the species topology (or network, in case of hybridization) along with the model (*i.e.*, hybridization, gene flow, ancestral structure, or standard model) that best fit the data relating the set of three *Orbicella* species.

A variety of studies in model organisms have used asymmetric gene tree frequencies to infer hybridization. Examples include studies in mouse (White *et al*. 2009), yeast (Yu *et al*. 2012 from Rokas *et al*. 2003 data), and mosquito (Fontaine *et al*. 2015) lineages. Thus, we will revisit this hypothesis by applying similar strategies as for corals to these publicly available datasets and evaluating whether an alternative model of ancestral structure better explains the data.

5. Intellectual Merit
The proposed methods will enable systematics researchers to infer more accurate phylogenies and test diverse hypotheses about the evolutionary dynamics of their systems. In particular, our methods will model how ancestral structure skews gene tree frequencies, and therefore provide an alternative explanatory model to hybridization. Because hybridization has been proposed in a number of model organisms based on evidence from skewed gene tree frequencies, we will reevaluate these hypotheses to refine our understanding of the evolution of these systems. Moreover, the proposed research involves application of our methods to the threatened non-model coral group *Orbicella*. Because they provide the ecosystem for 25% of marine biodiversity and are rapidly declining, it is particularly important to understand their evolutionary history, which may aid in future conservation efforts.

6. Broader Impacts
Accessibility of results. We will make all methods into open source programs and publish our findings in open-access format. In addition, we will produce short general-audience videos describing the major findings of our publications. These videos will be available both on our labs' websites and on YouTube.

Outreach. Drs. DeGiorgio and Medina both have established track records of increasing the participation of underrepresented groups in science. Dr. DeGiorgio's past activities include mentoring four female undergraduate students leading to three publications, reviewing fellowships and designing bioinformatics lab exercises for the Berkeley Biology Scholars Program, and reviewing fellowships for the Graduate Women in Science program. He has also been an organizer, advisory committee member, and instructor of the Summer Internship for Native Americans in Genomics Workshop since its inception in 2011. The purpose of this workshop is to increase the participation of Native Americans in science, particularly in genomics. Dr. DeGiorgio will continue to participate in and enhance this program, and will request funds to host future workshops at Penn State. He will also develop hands-on phylogenetic activities for local K-12 students through participation in Penn State's Science-U program. Dr. Medina has extensive experience working with underrepresented groups. Her efforts include courses with hands-on experience in marine science and genomics. Her genome science course has led to several publications with undergraduates, some of whom are from underrepresented groups. Her laboratory currently hosts about a dozen undergraduates assisting with different aspects of coral genomics research, and actively participates both in Penn State's Eberly College of Science outreach programs and in K-5 educational coral reef science activities in the State College School District. Dr. Medina also organizes a cnidarian genomics workshop for scientists. Moreover, both PIs plan to attend and present at annual SACNAS meetings, and will use this venue to recruit underrepresented students to work in their labs.

References Cited

H Akaike (1973) Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, ed PN Petrov, F Csaki. pp 267-281. Budapest: Akad Kiado.

R Aronson, A Bruckner, J Moore, B Precht, E Weil (2008) "Montastrea faveolata". IUCN Red List of Threatened Species. Version 2013.2. International Union for Conservation of Nature. Retrieved 01-30-2015.

AF Budd, JM Pandolfi (2004) Overlapping species boundaries and hybridization within the *Montastraea "annularis"* reef coral complex in the Pleistocene of the Bahama Islands. *Paleobiology* 30:396-425.

AF Budd, JM Pandolfi (2010) Evolutionary novelty is concentrated at the edge of coral species distributions. *Science* 328:1558-1561.

R Costanza, R D'Arge, S de Groot, M Farber, B Grasso, S Hannon, S Naeem, K Limburg, J Paruelo, RV O'Neill, R Raskin, P Sutton, M van den Belt (1997) The value of the world's ecosystem services and natural capital. *Nature* 387:253-260.

KA Cranston, B Hurwitz, D Ware, L Stein, RA Wing (2009) Species trees from highly incongruent gene trees in rice. *Syst Biol* 58:489-600.

M DeGiorgio, JH Degnan (2014) Robustness to divergence time underestimation when inferring species trees from estimated gene trees. *Syst Biol* 64:66-82.

M DeGiorgio, J Syring, AJ Eckert, A Liston, R Cronn, DB Neale, NA Rosenberg (2014) An empirical evaluation of two-stage species tree inference strategies using a muultilocus dataset from North American pines. *BMC Evol Biol* 14:67.

JH Degnan, M DeGiorgio, D Bryant, NA Rosenberg (2009) Properties of consensus methods for estimating species trees from gene trees. *Syst Biol* 58:35-54.

JY Dutheil, G Ganapathy, A Hobolth, T Mailund, MK Uyenoyama, MH Schierup (2009) Ancestral population genomics: the coalescent hidden Markov model approach. *Genetics* 183:259-274.

J Felsenstein (1989) PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* 5:164-166.

MC Fontaine, JB Pease, A Steele, RM Waterhouse, DE Neafsey, IV Sharakhov, X Jiang, AB Hall, F Catteruccia, E Kakahni, SN Mitchell, Y-C Wu, HA Smith, RR Love, MK Lawniczak, MA Slotman, SJ Emrich, MW Hahn, NJ Besansky. (2015) Extensive introgression in malaria vector species revealed by phylogenomics. *Science* 347:1248524.

NL Foster, CB Paris, JT Kool, IB Baums, JR Stevens, JA Sanchez, C Bastidas, C Agudelo, P Bush, O Day, R Ferrari, P Gonzalez, S Gore, R Guppy, MA McCartney, C McCoy, J Mendes, A Srinivasan, S Steiner, MJA Vermeij, E Weil, PJ Mumby (2012) Connectivity of Caribbean coral populations: complementary insights from empirical and modeled gene flow. *Mol Ecol* 21:1143-1157.

D Garrigan, Z Mobasher, SB Kingan, JA Wilder, MF Hammer (2005) Deep haplotype divergence and long-range linkage disequilibrium at Xp21.1 provide evidence that humans descend from a structured ancestral population. *Genetics* 170:1849.186.

ME Hellberg (2006) No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evol Biol* 6:24.

A Hobolth, L Nørvang Andersen, T Mailund (2011) On computing the coalescence time density in an isolation-with-migration model with few samples. *Genetics* 187:1241-1243.

A Hobolth, OF Christensen, T Maillund, MH Schierup (2007) Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet* 3:e7.

RR Hudson (2002) Generating samples under a Wright-Fisher neutral model. *Bioinformatics* 18:337-338.

JBC Jackson, MK Donovan, KL Cramer (2014) Status and trends of Caribbean coral reefs. ed VV Lam. pp 1970-2012. Gland, Switzerland: Global Coral Reef Monitoring Network, IUCN.

JA Kleypas (1997) Modeled estimates of global reef habitat and carbonate production since the Last Glacial Maximum. *Paleoceanography* 12:533-545.

LS Kubatko, BC Carstens, LL Knowles (2009) STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25:971-973.

HC Lanier, LL Knowles (2012) Is recombination a problem for species-tree analyses? *Syst Biol* 61:691-701.

AD Leaché, RB Harris, B Rannala, Z Yang (2014) The influence of gene flow on species tree estimation: a simulation study. *Syst Biol* 63:17-30.

AD Leaché, B Rannala (2011) The accuracy of species tree estimation under simulation: a comparison of methods. *Syst Biol* 60:126-137.

DR Levitan, ND Fogarty, J Jara, KE Lotterhos, N Knowlton (2010) Genetic, spatial, and temporal components of precise spawning synchrony in reef building corals of the *Montastraea annularis* species complex. *Evolution* 65:1254-1270.

DR Levitan, H Fukami, J Jara, D Kline, TM McGovern, KE McGhee, CA Swanson, N Knowlton (2004) Mechanisms of reproductive isolation among sympatric broadcast-spawning corals of the *Montastraea annularis* species complex. *Evolution* 58:308-323.

K Liu, J Dai, K Truong, Y Song, MH Kohn, L Nakhleh (2014) An HMM-based comparative genomic framework for detecting introgression in eukaryotes. *PLoS Comput Biol* 10:e1003649.

L Liu, L Yu, SV Edwards (2010) A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol Biol* 10:302.

T Mailund, AE Halager, M Westergaard, JY Dutheil, K Munch, LN Andersen, G Lunder, K Prüfer, A Scally, A Hobolth, MH Schierup (2012) A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species. *PLoS Genet* 8:e1003125.

JE McCormack, MG Harvey, BC Faircloth, NG Crawford, TC Glenn, RT Brumfield (2013) A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS One* 8:e54848.

S Mirarab, R Reaz, MdS Bayzid, T Zimmerman, MS Swenson, T Warnow (2014) ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30:541-548.

E Mossel, S Roch (2010) Incomplete lineage sorting: consistent phylogeny estimation from multiple loci. *IEEE/ACM Trans Comp Biol Bioinf* 7:166-171.

L Nakhleh (2013) Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends Ecol Evol* 28:719-728.

JM Pandolfi, AF Budd (2008) Morphology and ecological zonation of Caribbean reef corals: the *Montastraea 'annularis'* species complex. *Mar Ecol Prog Ser* 369:89-102.

A Rambaut, NC Grassly (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution among evolutionary trees. *Comput Appl Biosci* 13:235-238.

M Reaka-Kudla (1997) The global biodiversity of coral reefs: a comparison with rain forests. In *Biodiversity II: understanding and protecting our biological resources.*, ed M Reaka-Kudla, DE Wilson, EO Wilson. pp 83-108. Washington, DC: Joseph Henry Press.

A Rokas, BL Williams, N King, SB Carroll (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798-804.

F Ronquist, JP Huelsenbeck (200) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.

L Salichos, A Rokas (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497:327-331.

EG Severance, SA Karl (2006) Contrasing population genetic structures of sympatric, mass-spawning Caribbean corals. *Marine Biology* 150:57-68.

M Slatkin, JL Pollack (2008) Subdivision in an ancestral species creates asymmetry in gene trees. *Mol Biol Evol* 25:2241-2246.

S Song, L Liu, SV Edwards, S Wu (2012) Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc Natl Acad Sci USA* 109-14942-14947.

MD Spalding, C Ravilious, EP Green (2001) United Nations Environment Programme, World Conservation Monitoring Centre. In *World Atlas of Coral Reefs*. pp 416. Berkeley, CA: University of California Press.

M Steel (1992) The complexity of reconstructing trees from qualitative characters and subtrees. *J Classif* 9:91-116.

J Sullivan, P Joyce (2005) Model selection in phylogenetics. *Annu Rev Ecol Evol Syst* 36:445-466.

O Thalmann, A Fischer, F Lankester, S Pääbo, L Vigilant (2007) The complex evolutionary histories of gorillas: insights from genomic data. *Mol Biol Evol* 24:146-158.

Y Tian, L Kubatko (2015) Distribution of gene tree histories under the coalescent model with gene flow. *BioRxiv* doi:http://dx/doi.org/10.1101/023937.

JA Toms, JS Compton, M Smale, S von der Heyden (2014) Variation in palaeo-shorelines explains contemporary population genetic patterns of rocky shore species. *Biol Lett* 10:20140330.

MA White, C Ané, CN Dewey, BR Larget, BA Payseur (2009) Fine-scale phylogenetic discordance across the house mouse genome. *PLoS Genet* 5:e1000729.

SM Williams, PJ Mumby, I Chollett, J Cortés (2015) Importance of differentiating *Orbicella* reefs from gorgonian plains for ecological assessments of Caribbean reefs. *Mar Ecol Prog Ser* 530:93-101.

Y Wu (2012) Coalescent-based species tree inference from gene tree topologies under incomplete lineage sorting by maximum likelihood. *Evolution* 66:763-775.

Y Yu, JH Degnan, L Nakhleh (2012) The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genet* 8:e1002660.

Y Yu, J Dong, K Liu, L Nakhleh (2014) Maximum likelihood inference of reticulate evolutionary histories. *Proc Natl Acad Sci USA* 111:16448-16453.

Y Yu, L Nakhleh (2015) A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics* 16:S10.

Y Yu, C Than, JH Degnan, L Nakhleh (2011) Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. *Syst Biol* 60:138-149.