GLBIO 2015 POSTER ABSTRACT

Authors:

Edward A. Salinas, MS, Independent Researcher, Dallas, TX, USA, eddie@eddiesalinas.com
John Horne, PhD, Dept. of Biology, Dalhousie Univ., Halifax, Nova Scotia, Canada, jhorne3@jhu.edu
Rita Castilho, PhD, CCMAR, Univ. of Algarve, Faro, Portugal, rcastil@ualg.pt

---

ABSTRACT TITLE

COMET ("COalescence with Multiple Mergers Employing Thermodynamic Integration")

ABSTRACT: Alternatives to the Kingman coalescent have recently been developed proposing hypothetical genealogies with simultaneous multiple mergers [1, 2]. These multiple merger models differ in several assumptions from the classical Kingman coalescent, viz., overlapping generations, reproductive variance between individuals and no limitations on the number of offspring an individual can contribute to the effective population size. We want to model such demographic parameters with multiple merger coalescent gene genealogies that are ignored by the Kingman coalescent (e.g. Y the "degree of simultaneity of multiple mergers" and φ a "generational overlap factor" from "Algorithm #1" in [2]). Further, to aid in model selection, we desire to compute Bayes Factors [3]. To accomplish such tasks so we have written an MCMC computer program in the Java programming language called COMET ("COalescence with Multiple Mergers Employing Thermodynamic Integration") that samples from the posterior distribution of parameters of the models [2,3,4] by sampling genealogy trees. COMET, an independence-style sampler incorporating multiple algorithms for sampling (e.g. Kingman Coalescent [4] and "Algorithm #1" from [2]), implements a heated chain MCMC approach that allows for the calculation of marginal probabilities using thermodynamic integration. As genealogies are considered, parameters from the simulation (e.g. Tree Height for any tree, and Y and φ, from "Algorithm #1" in [2]) are logged and thus sampled from the posterior distribution. The program computes likelihoods using the HKY substitution model with PhyML (v20150115) [5] or BEAGLE [6]. By allowing for overlapping generations and reproductive variance in genealogies, COMET has potential to improve coalescence analyses for organisms that depart from the assumptions of the Kingman coalescent. Improved coalescence analyses could have an immediate benefit for scientific research aiming to conserve biodiversity and manage natural resources, such as wild stocks of fishery species and tourist-attracting ecosystems like coral reefs. Here, we present initial use and output of COMET and results suggesting validity. We also demonstrate computation of marginal likelihoods and Bayes factors using an empirical mtDNA population genetic dataset from the blue-spine unicornfish (Naso unicornis). COMET formats its output for easy use with the Tracer [7] program; inspection of traces indicates good mixing is necessary. We finally briefly describe current and future steps to improve both mixing and efficiency and new datasets to explore and characterize. Our code is accessible via the bitbucket repository here: https://bitbucket.org/eddieasalinas/comet

References:

1. Aurélien Tellier, Christophe Lemaire, Coalescence 2.0: a multiple branching of recent theoretical developments and their applications , Molecular Ecology (2014) Volume 23, Issue 11, pp 2637–2652
2. Ori Sargsyan, John Wakeley, A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms, Theoretical Population Biology 74 (2008) 104–11
3. Peter Beerli, Michal Palczewski, Unified Framework to Evaluate Panmixia and Migration Direction among Multiple Sampling Locations, Genetics 185: 313–326 (May 2010)
4. Wakely, John, 2009, Coalescent Theory: An Introduction. Roberts and Company Publishing, 2009
5. Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O., New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0, Systematic Biology, 59(3):307-21, 2010
6. Daniel L.A., et al., BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics, Systematic Biology, 2012 Jan; 61(1): 170–173
7. Rambaut A, Suchard MA, Xie D & Drummond AJ (2014) Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer