

SOFTWARE

Open Access



# Hybrid-Lambda: simulation of multiple merger and Kingman gene genealogies in species networks and species trees

Sha Zhu<sup>1\*</sup>, James H. Degnan<sup>2</sup>, Sharyn J. Goldstien<sup>3</sup> and Bjarki Eldon<sup>4</sup>

## Abstract

**Background:** There has been increasing interest in coalescent models which admit multiple mergers of ancestral lineages; and to model hybridization and coalescence simultaneously.

**Results:** *Hybrid-Lambda* is a software package that simulates gene genealogies under multiple merger and Kingman's coalescent processes within species networks or species trees. *Hybrid-Lambda* allows different coalescent processes to be specified for different populations, and allows for time to be converted between generations and coalescent units, by specifying a population size for each population. In addition, *Hybrid-Lambda* can generate simulated datasets, assuming the infinitely many sites mutation model, and compute the  $F_{ST}$  statistic. As an illustration, we apply *Hybrid-Lambda* to infer the time of subdivision of certain marine invertebrates under different coalescent processes.

**Conclusions:** *Hybrid-Lambda* makes it possible to investigate biogeographic concordance among high fecundity species exhibiting skewed offspring distribution.

**Keywords:** Hybridization, Multiple merger, Gene tree, Coalescent,  $F_{ST}$ , Infinite sites model, Hybrid-lambda, Skewed offspring distribution

## Background

Species trees describe ancestral relations among species. Gene genealogies describe the random ancestral relations of alleles sampled within species. Species trees are often assumed to be bifurcating [6], and gene genealogies to follow the Kingman coalescent [23, 27] in allowing at most two lineages to coalesce at a time.

Recently, there has been increasing interest in coalescent models which admit multiple mergers of ancestral lineages [1, 2, 9, 12, 36, 38, 39] and to model hybridization and coalescence simultaneously [3, 25, 26, 28, 46]. For high fecundity species exhibiting sweepstake-like reproduction, such as oysters and other marine organisms [1, 4, 9, 11, 17, 18, 38], the Kingman coalescent may not be appropriate, as it is based on low offspring number population models (see recent reviews by [19] and [42]). Thus, we consider  $\Lambda$  coalescents [8, 35, 36] derived from

sweepstake-like reproduction models, and allow *more* than two lineages to coalesce at a time.

We introduce the software *Hybrid-Lambda* for simulating gene trees under two models of  $\Lambda$ -coalescents within rooted species trees and rooted species networks. Our program differs from existing software which also allows multiple mergers, such as SIMCOAL 2.0 [29] — which allows multiple mergers in gene trees due to small population sizes under the Wright-Fisher model — in that we apply coalescent processes that are obtained from population models explicitly modelling skewed offspring distributions, as opposed to bottlenecks.

Species trees may also fail to be bifurcating due to either polytomies or hybridization events. The simulation of gene genealogies within a species network which admits hybridization is another application of *Hybrid-Lambda*. The package *ms* [24] can also simulate gene genealogies within species networks under Kingman's coalescent. However the input of *ms* is difficult to automate when the network is sophisticated or generated from other

\*Correspondence: joe.zhu@well.ox.ac.uk

<sup>1</sup> Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK  
Full list of author information is available at the end of the article

software. Other simulation studies using species networks have either used a small number of network topologies coded individually (for example, in phylonet [43, 45, 46]) or have assumed that gene trees have evolved on species trees embedded within the species network [22, 28, 31]. Hybrid-Lambda will help to automate simulation studies of hybridization by allowing for a large number of species network topologies and allowing gene trees to evolve directly within the network. Hybrid-Lambda can simulate both Kingman and  $\Lambda$ -coalescent processes within species networks. A comparison of features of several software packages that output gene genealogies under coalescent models is given in Table 1.

### Implementation

The program input file for Hybrid-Lambda is a character string that describes relationships between species. Standard Newick format [33] is used for the input of species trees and the output of gene trees, whose interior nodes are not labelled. An extended Newick formatted string [5, 25] labels all internal nodes, and is used for the input of species networks (see Fig. 1).

### Parameters

Hybrid-Lambda can use multiple lineages sampled from each species and simulate Kingman or multiple merger ( $\Lambda$ )-coalescent processes within a given species network. In addition, separate coalescent processes can be specified on different branches of the species network. The coalescent is a continuous-time Markov process, in which times between coalescent events are independent exponential random variables with different rates. The rates are determined by a so-called coalescent parameter that can be input via command line, or a(n) (extended) Newick formatted string with specific

coalescent parameters as branch lengths. By default, the Kingman coalescent is used, for which two of  $b$  active lineages coalesce at rate  $\lambda_{b,2} = \binom{b}{2}$ . One can choose between two different examples of a  $\Lambda$ -coalescent, whose parameters have clear biological interpretation. While we cannot hope to cover the huge class of Lambda-coalescents, our two examples are the ones that have been most studied in the literature [2, 7, 13]. If the coalescent parameter is between 0 and 1, then we use  $\psi$  for the coalescent parameter, and the rate  $\lambda_{b,k}$  at which  $k$  out of  $b$  ( $2 \leq k \leq b$ ) active ancestral lineages merge is

$$\lambda_{b,k} = \binom{b}{k} \psi^{k-2} (1 - \psi)^{b-k}, \quad \psi \in [0, 1], \quad (1)$$

Eldon and Wakeley [9]. If the coalescent parameter is between 1 and 2, then we use  $\alpha$  for the coalescent parameter, and the rate of  $k$ -mergers ( $2 \leq k \leq b$ ) is

$$\lambda_{b,k} = \binom{b}{k} \frac{B(k - \alpha, b - k + \alpha)}{B(2 - \alpha, \alpha)}, \quad \alpha \in (1, 2), \quad (2)$$

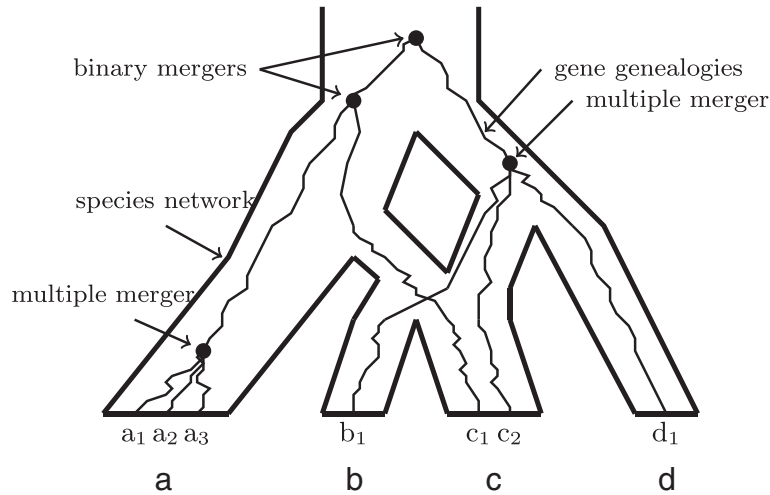
where  $B(\cdot, \cdot)$  is the beta function [39].

Hybrid-Lambda assumes by default that the input network (tree) branch lengths are in coalescent units. However, this is not essential. Coalescent units can be converted through an alternative input file with numbers of generations as branch lengths, which are then divided by their corresponding effective population sizes. By default, effective population sizes on all this parameter using the command line, or using a(n) (extended) Newickbranches are assumed to be equal and unchanged. Users can change formatted string to specify population sizes on all branches through another input file.

The simulation requires ultrametric species networks, i.e. equal lengths of all paths from tip to root.

**Table 1** Comparison of software programs simulating gene trees in species trees and networks. *Migration* refers to modeling post-speciation gene flow

	Gene trees in networks	$\Lambda$ -coalescent	Small pop. mult. merger	Infinite sites model	Recombination	Migration
COAL [6]	no	no	no	no	no	no
CoaSim [32]	no	no	no	yes	yes	yes
fastsimcoal2 [15]	yes	no	yes	no	yes	yes
Genome [30]	yes	no	no	yes	yes	yes
GUMS [20]	no	no	no	no	no	yes
ms [24]	yes	no	no	yes	yes	yes
msHOT [21]	yes	no	no	yes	yes	yes
msms [14]	yes	no	no	yes	yes	yes
scrm [41]	yes	no	no	yes	yes	yes
simcoal2 [27]	yes	no	yes	no	yes	yes
Hybrid-Lambda	yes	yes	no	yes	no	no



**Fig. 1** Demonstration of a multiple merger genealogy within a species network. A multiple merger gene genealogy with topology  $(( (a_1, a_2, a_3), c_1), (b_1, c_2, d_1))$ , of which the coalescence events pointed to by arrows labeled "multiple merger" indicate coalescence of 3 lines, simulated in a species network with topology  $(( (B, C) s_1) h_1 \# H_1, A) s_2, (h_1 \# H_1, D) s_3) r$ , where  $H_1$  is the probability that a lineage has its ancestry from its left parental population

Hybrid-Lambda checks the distances in coalescent units between the root and all tip nodes and prints out warning messages if the ultrametric assumption is violated.

### Results and discussion

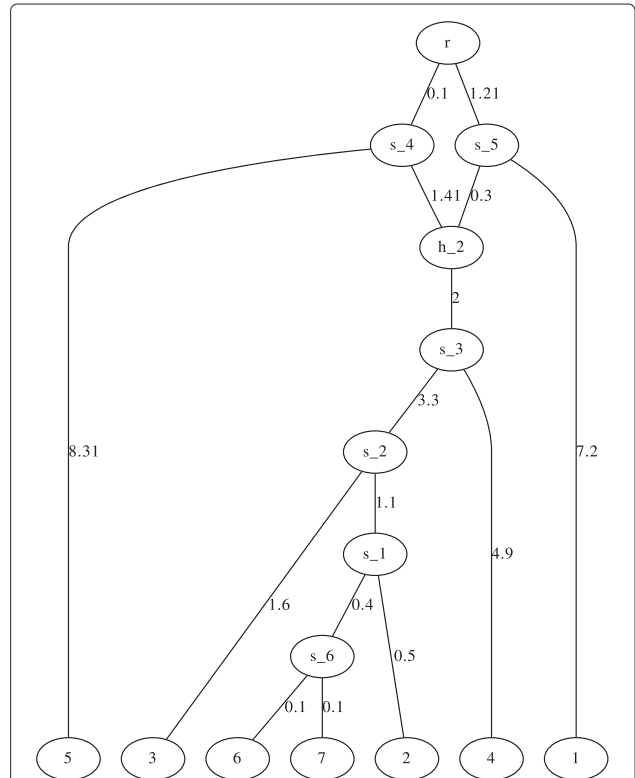
Hybrid-Lambda outputs simulated gene trees in three different files: one contains gene trees with branch lengths in coalescent units, another uses the number of generations as branch lengths, and the third uses the number of expected mutations as branch lengths.

Besides outputting gene tree files, Hybrid-Lambda also provides several functions for analysis purposes:

- user-defined random seed for simulation,
- output simulated data in 0/1 format assuming the infinitely many sites mutation model,
- a frequency table of gene tree topologies,
- a figure of the species network or tree (this function only works when `!TEX` or `dot` is installed) (Fig. 2),
- the expected  $F_{ST}$  value for a split model between two populations,
- when gene trees are simulated from two populations, the software Hybrid-Lambda can generate a table of relative frequencies of reciprocal monophyly, paraphyly, and polyphyly.

### Simulation example

We give a simulation example showing the impact of the particular coalescent model on estimating the divergence time for two populations. Results can be confirmed using analytic approximations to  $F_{ST}$ . This is shown

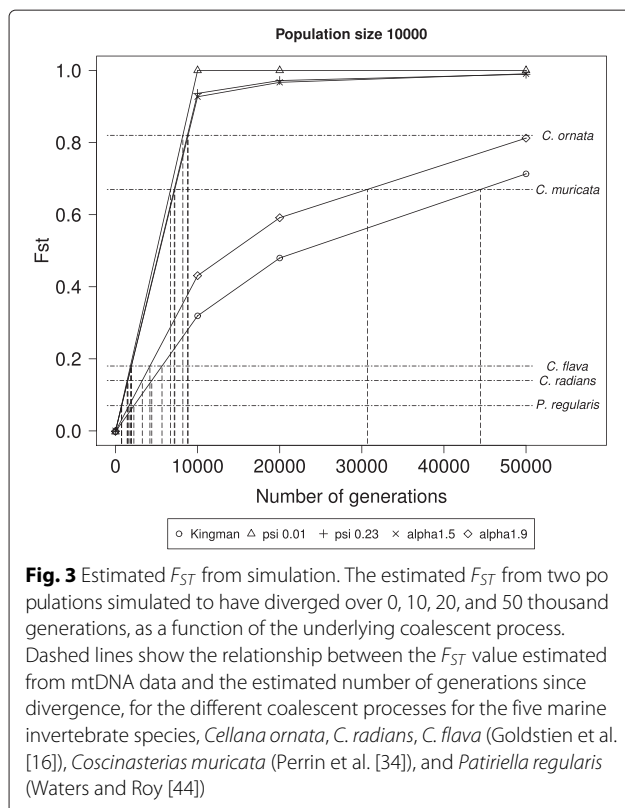


**Fig. 2** Demonstration of a network figure generated by Hybrid-Lambda. The network is automatically generated by Hybrid-Lambda as dot and .pdf files from the extended newick string `"((((((6:.1,7:.1) s_6:.4,2:.5) s_1:1.1,3:1.6) s_2:3.3, 4:4.9) s_3:2) h_2#.5:1.41,5:8.31) s_4:0.1, (1:7.2,h_2#.5:.3) s_5:1.21)r;"`

in the Appendix along with example code for using Hybrid-Lambda for this example.

Eldon B and Wakeley J [10] showed that population subdivision can be observed in genetic data despite high migration between populations. One of the most widely used measures of population differentiation is the  $F_{ST}$  statistic. The relationship between  $F_{ST}$  and biogeography depends on the underlying coalescent process, which might be especially important for the interpretation of divergence and demographic history of many marine species. Here we used Hybrid-Lambda to simulate divergence between two populations based on different  $\Lambda$ -coalescents, as well as the standard Kingman coalescent. Mutations were simulated in Hybrid-Lambda under the infinite-sites model. The summary statistic  $F_{ST}$  was estimated for these data and was used to compare  $F_{ST}$  estimated from mtDNA from five species of marine invertebrates. These species were used in previous studies to test the hypothesis that contemporary oceanic conditions are creating subdivisions between the North Island and South Island reef populations of New Zealand [16, 34, 44]. These studies represent some of the earliest mitochondrial studies on the marine disjunction between the North and South Islands of New Zealand.

The  $F_{ST}$  statistic between North Island and South Island populations reported for these species ranges from approximately 0.07 to 0.8 (Fig. 3). *Cellana ornata* displays



a very strong split, which was estimated to have occurred around 0.2–0.3 million years ago based on published estimates of divergence rates and reciprocal monophyly displayed in the data set. This result may be supported by our simulations using the Kingman coalescent. However, when multiple mergers and a higher fraction of replacement by a single parent is allowed to occur then our simulations support much younger splits between the populations  $\sim 9,000$  generations or  $\sim 48,000$  generations ago (Fig. 3). Similarly, the strong split observed for *Coscinasterias muricata* could be placed anywhere from  $\sim 9,000$  to 45,000 generations ago depending on the degree to which multiple mergers are allowed to occur. While the range for *Patiriella regularis*, *Cellana radians* and *C. flava* is much smaller, it is still not clear cut as to whether divergence would be observed under different coalescent models. Here we used  $\psi = 0.01$  and  $\psi = 0.23$ , and  $\alpha = 1.5$  and  $\alpha = 1.9$ , with larger values of  $\psi$  and smaller values of  $\alpha$  corresponding to higher probabilities of multiple mergers. Our choice of parameter values corresponds to the estimated values obtained for mtDNA of oysters and Atlantic cod. An estimate for Pacific oysters based on mitochondrial DNA for  $\psi$  was 0.075 [9]. The results for our choice of parameter values suggest that our conclusions about a much earlier split of the populations than previously estimated are robust with regard to parameter choice. A recent study of Atlantic cod [2] estimated  $\psi$  between 0.07 and 0.23 for nuclear genes and near 0.01 for mitochondrial genes. The same study estimated  $\alpha$  to be 1.0 and 1.28 for nuclear genes and between 1.53 and 2.0 for mitochondrial genes.

## Conclusions

The implications for using alternative coalescent models are far reaching. Many marine organisms reproduce through broadcast spawning of thousands to millions of gametes, and while the expected survival of these offspring is low, there is the potential for a small subset of the adults to have a greater contribution to the next generation than assumed by the Kingman coalescent. Hybrid-Lambda makes it possible to investigate the effect of high fecundity on biogeographic concordance among species that exhibit high fecundity and high offspring mortality, including in complex demographic scenarios that allow hybridization.

## Availability and requirements

Hybrid-Lambda can be downloaded from <http://hybridlambda.github.io/>. The program is written in C++ (requires compilers that support C++11 standard to build), and released under the GNU General Public License (GPL) version 3 or later. Users can modify and make new distributions under the terms of this license. For full details of this license, visit

<http://www.gnu.org/licenses/>. Hybrid-Lambda works on Unix-like operating systems. We have used travis continuous integration to test compiling the program on Linux and Mac OS. An API in R [37] is currently under development.

**Appendix:  $F_{ST}$  calculations**

Here we show analytic calculations that can be used to obtain expressions for  $F_{ST}$  when mutation rates are low. The effect of  $\alpha$  on  $F_{ST}$  for fixed generation times is shown in Fig. 4.

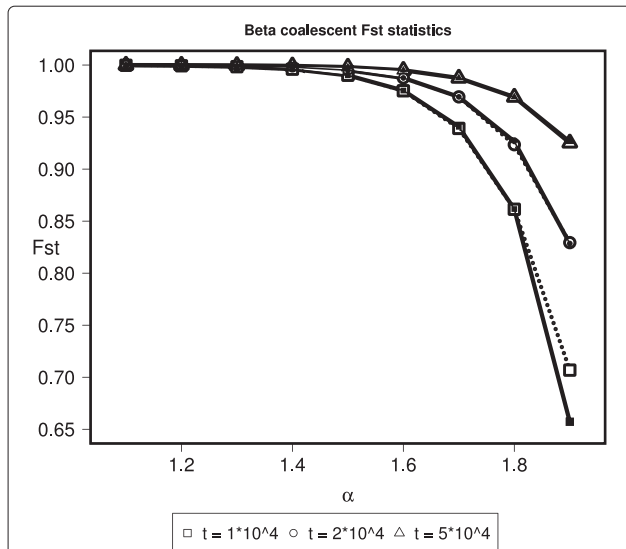
Assume two populations  $A$  and  $B$  have been isolated until time  $\tau$  in the past as measured from the present. Assume also that the same coalescent process is operating in populations  $A$  and  $B$ . Let  $T_w$  denote the time until coalescence for two lines when drawn from the same population, and  $T_b$  when drawn from different populations. Let  $\lambda_A$  denote the coalescence rate for two lines in population  $A$ , and  $\lambda_{AB}$  for the common ancestral population  $AB$ . For the Beta( $2 - \alpha, \alpha$ )-coalescent,  $\lambda_A = 1$ , for the point-mass process  $\lambda_A = \psi^2$ . One now obtains

$$E[T_w] = (1 - e^{-\lambda_A \tau})\lambda_A^{-1} + e^{-\lambda_A \tau} (\tau + \lambda_{AB}^{-1}), \tag{3}$$

$$E[T_b] = \tau + \lambda_{AB}^{-1}.$$

Slatkin [40] obtained the approximation, where  $\mu$  is the per generation mutation rate,

$$F_{ST}^{(0)} := \lim_{\mu \rightarrow 0} F_{ST} = 1 - \frac{E[T_w]}{E[T_b]} \tag{4}$$



**Fig. 4** Comparison of estimated  $F_{ST}$  values from simulation and analytical predictions. Values of  $F_{ST}$  as a function of the parameter  $\alpha$  for  $1 < \alpha < 2$  for different numbers of generations of separation for two populations. Simulations (dotted lines) are based on 1 individual from each of two populations separated by  $t$  generations with  $10^3$  replicates and  $\alpha \in \{1.1, 1.2, \dots, 1.9\}$ . Analytical predictions (solid lines) of  $F_{ST}$  were calculated using (6)

Thus, using (3) gives

$$F_{ST}^{(0)} = (1 - e^{-\lambda_A \tau}) \left( 1 - \frac{1}{(\tau + \lambda_{AB}^{-1})\lambda_A} \right) \tag{5}$$

The result (5) seems to make sense, since  $\lim_{\tau \rightarrow 0} F_{ST}^{(0)} = 0$  and  $\lim_{\tau \rightarrow \infty} F_{ST}^{(0)} = 1$ . By way of example, if all populations exhibit a Beta( $2 - \alpha, \alpha$ )-coalescent,  $\lambda_A = \lambda_{AB} = 1$ , and

$$F_{ST}^{(0)} = (1 - e^{-\tau}) \frac{\tau}{1 + \tau}. \tag{6}$$

However, deciding the timeunit of  $\tau$  now becomes important, since the timescale of a Beta( $2 - \alpha, \alpha$ )-coalescent is proportional to  $N^{\alpha-1}$ ,  $1 < \alpha < 2$  [39], where  $N$  is the population size. One can obtain a more accurate expression of the timescale given knowledge about the mean of the potential offspring distribution (see [39]). However, since the mean is unknown in most cases, we apply the approximation  $N^{\alpha-1}$ . Assuming  $n \geq 2$  sequences from each population, the ‘observed’ FST ( $\hat{F}_{ST}$ ) was computed as  $\hat{F}_{ST} = 1 - \frac{n}{n-1} \frac{H_w}{H_b}$  where  $H_w$  is the average pairwise differences within populations,  $H_w = \frac{1}{2}(H_{w,1} + H_{w,2})$ , and  $H_b$  is the average of  $n^2$  pairwise differences between populations.

The following command-line argument for Hybrid-Lambda simulates 1,000 genealogies with 10 lineages sampled from each of two populations separated by one coalescent unit with mutation rate  $\mu = 0.00001$  using a  $\beta$ -coalescent with parameter  $\alpha = 1.5$ :

```
hybrid-Lambda -spng '(A:10000,B:10000)';
-num 1000 -seed 45 -mu 0.00001 -S 10 10 \
-mm 1.5 -sim_num_mut -seg -fst
```

where

- -spng '(A:10000,B:10000)'; denotes the population structure of a split model of one population splits to two at 10,000 generations in the past.
- -num 1000 simulates 1,000 genealogies from this model.
- -seed 45 initializes the random seed for the simulation.
- -mu 0.00001 specifies the mutation rate of 0.00001 per generation.
- -S 10 10 samples 10 individuals from each population.
- -mm 1.5 specifies the  $\Lambda$ -coalescent parameter.
- -sim\_num\_mut outputs simulated genealogies in Newick string, of which the number of mutations on internal branches are labelled.
- -seg generates haplotype data set.
- -fst computes  $F_{ST}$  of the generated haplotype data set.

One can use this example to generate the data for Fig. 4 by setting the `-S` flag to `-S 1 1`.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

SZ was responsible for the software development. JD and BE supervised the project. BE derived all the  $F_{ST}$  calculations in the Appendix. SG provided the simulation results and time estimates in Fig. 3. All the authors have contributed to the manuscript writing. All authors read and approved the final manuscript.

#### Acknowledgements

This work was supported by New Zealand Marsden Fund (SZ and JD), EPSRC grant EP/G052026/1 and DFG grant BL 1105/3-1 through the SPP Priority Programme 1590 "Probabilistic Structures in Evolution" (BE). This work was partly conducted while JD was a Sabbatical Fellow at the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation, the U.S. Department of Homeland Security, and the U.S. Department of Agriculture through NSF Award #EF-0832858, with additional support from The University of Tennessee, Knoxville.

#### Author details

<sup>1</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>2</sup>Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico, USA. <sup>3</sup>Department of Biology, University of Canterbury, Christchurch, New Zealand. <sup>4</sup>Institut für Mathematik, Technische Universität Berlin, Berlin, Germany.

Received: 18 May 2015 Accepted: 25 August 2015

Published online: 15 September 2015

#### References

- Árnason E. Mitochondrial cytochrome b variation in the high-fecundity Atlantic cod: trans-Atlantic clines and shallow gene genealogy. *Genetics*. 2004;166:1871–85.
- Árnason E, Halldórsson K. Nucleotide variation and balancing selection at the *Ckma* gene in Atlantic cod: analysis with multiple merger coalescent models. *PeerJ*. 2015;e786:3. doi:10.7717/peerj.786.
- Bartoszek K, Jones G, Oxelman B, Sagitov S. Time to a single hybridization event in a group of species with unknown ancestral history. *J Theor Biol*. 2004;322:1–6.
- Beckenbach AT. In: Golding B, editor. Mitochondrial haplotype frequencies in oysters: neutral alternatives to selection models, Non-neutral Evolution. New York: Chapman & Hall; 1994, pp. 188–98.
- Cardona G, Rossell F, Valiente G. Extended Newick: it is time for a standard representation of phylogenetic networks. *BMC Bioinform*. 2008;9:532.
- Degnan JH, Salter LA. Gene tree distributions under the coalescent process. *Evolution*. 2005;59:24–37.
- Delmas JF, Dhersin JS, Siri-Jégousse A. Asymptotic results on the length of coalescent trees. *Ann Appl Probab*. 2008;18:997–1025.
- Donnelly P, Kurtz TG. Particle representations for measure-valued population models. *Ann Probab*. 1999;27:166–205.
- Eldon B, Wakeley J. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*. 2006;172:2621–33.
- Eldon B, Wakeley J. Coalescence times and  $F_{ST}$  under a skewed offspring distribution among individuals in a population. *Genetics*. 2009;181:615–29.
- Eldon B. Estimation of parameters in large offspring number models and ratios of coalescence times. *Theor Popul Biol*. 2011;80:16–28.
- Eldon B, Degnan JH. Multiple merger gene genealogies in two species: monophyly, paraphyly, and polyphyly for two examples of Lambda coalescents. *Theor Popul Biol*. 2012;82:117–30.
- Eldon B, Birkner M, Blath J, Freund F. Can the Site-Frequency Spectrum Distinguish Exponential Population Growth from Multiple-Merger Coalescents? *Genetics*. 2015;199:841–56.
- Ewing G, Hermisson J. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics*. 2010;26:2064–65.
- Excoffier L, Foll M. Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*. 2011;27:9.
- Goldstien SJ, Schiel DR, Gemmill NJ. Comparative phylogeography of coastal limpets across a marine disjunction in New Zealand. *Mol Ecol*. 2009;15:3259–68.
- Hedgewood D. In: Beaumont A, editor. Does variance in reproductive success limit effective population sizes of marine organisms? *Genetics and Evolution of Aquatic Organisms*. London: Chapman and Hall; 1994, pp. 1222–344.
- Hedgewood D, Tracey M, Nelson K. In: Abele LG, editor. *Genetics, The Biology of Crustacea* vol. 2. New York: Academic Press; 1982, pp. 297–403.
- Hedgewood D, Pudovkin AI. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bull Mar Sci*. 2011;87:971–1002.
- Heled J, Bryant D, Drummond AJ. *BMC Evol Biol*. 2013;13:44.
- Hellenthal G, Stephens M. msHOT: modifying Hudson's ms simulator to incorporate crossover and gene conversion hotspots. *Bioinformatics*. 2007;23:520–21.
- Holland BR, Benthin S, Lockhart PJ, Moulton V, Huber KT. *BMC Evol Biol*. 2008;8:202.
- Hudson RR. Gene genealogies and the coalescent process. *Oxford Surv Evol Biol*. 1990;7:1–44.
- Hudson RR. Generating samples under a Wright-Fisher neutral model. *Bioinformatics*. 2002;18:337–38.
- Huson D, Rupp R, Scornavacca C. *Phylogenetic Networks: Concepts, Algorithms and Applications*. Cambridge University Press; 2010.
- Jones G, Sagitov S, Oxelman B. Statistical inference of allopolyploid species networks in the presence of incomplete lineage sorting. *Syst Biol*. 2013;62:467–78.
- Kingman JFC. On the genealogy of large populations. *J App Probab*. 1982;19A:27–43.
- Kubatko LS. Identifying hybridization events in the presence of coalescence via model selection. *Syst Biol*. 2009;58:478–88.
- Laval G, Excoffier L. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*. 2004;20:2485–87.
- Liang L, Zöllner S, Abecasis GR. GENOME: a rapid coalescent-based whole genome simulator. *Bioinformatics*. 2007;23:1565–67.
- Meng C, Kubatko LS. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. *Theor Popul Biol*. 2009;75:35–45.
- Mailund T, Schierup H, Pedersen CNS, Mechlenborg PJM, Madsen JN, Schausser L, et al. CoaSim a flexible environment for simulating genetic data under coalescent models. *BMC Bioinforma*. 2005;6:252.
- Olsen G. Gary Olsen's interpretation of the "Newick's 8:45" tree format standard. 1990. [http://evolution.genetics.washington.edu/phylip/newick\\_doc.html](http://evolution.genetics.washington.edu/phylip/newick_doc.html). Access date 2/Sep/2015.
- Perrin C, Wing SR, Roy MS. Effects of hydrographic barriers on population genetic structure of the sea star *Coscinasterias muricata* (Echinodermata, Asteroidea) in the New Zealand fiords. *Mol Ecol*. 2004;13:2183–95.
- Pitman J. Coalescents with multiple collisions. *Ann Probab*. 1999;27:1870–902.
- Sagitov S. The general coalescent with asynchronous mergers of ancestral lines. *J Appl Probab*. 1999;36:1116–125.
- R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015. <http://www.R-project.org/>.
- Sargsyan O, Wakeley J. A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. *Theor Popul Biol*. 2008;74:104–114.
- Schweinsberg J. Coalescent processes obtained from supercritical Galton-Watson processes. *Stoch Proc Appl*. 2003;106:107–39.
- Slatkin M. Inbreeding coefficients and coalescence times. *Genet. Res*. 1991;58:167–175.
- Staab PR, Zhu S, Metzler D, Lunter G. Scrm: efficiently simulating long sequences using the approximated coalescent with recombination. *Bioinformatics*. 2015;31(10):1680–82.
- Tellier A, Lemaire C. Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. *Mol Ecol*. 2014;23:2637–52.

43. Than C, Ruths D, Nakhleh L. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinforma.* 2008;9:322. doi:10.1186/1471-2105-9-322.
44. Waters JM, Roy MS. Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow *Mol Ecol.* 2004;13:2797–806.
45. Yu Y, Than C, Degnan JH, Nakhleh L. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. *Syst Biol.* 2011;60:138–49.
46. Yu Y, Degnan JH, Nakhleh L. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genet.* 2012;e1002660:8. doi:10.1371/journal.pgen.1002660.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

