Theoretical Population Biology 87 (2013) 1–4

Contents lists available at SciVerse ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Editorial

Coalescent theory has many new branches

1. Initial development of coalescent theory

Coalescent theory and the story of its unfolding have become part of the canon of population genetics. This is not a signal that the field is in decline, but a mark of the enduring value of the gene-genealogical way of thinking. Within both the biological and the mathematical literature of coalescent theory, novel extensions and wholly new developments continue to appear. This special issue of the journal presents a snapshot of coalescent theory's leading edge, now roughly thirty years after the birth the field.

Coalescent theory begins by imagining the ancestry of a sample of size n at a single genetic locus without recombination. This ancestry is comprised of exactly $n-1$ coalescent events wherein pairs of genetic lineages join together backward in time. The result is called the gene genealogy of the sample. At the $(n-1)$th coalescent event, the most recent common ancestor of the entire sample has been reached, and the process is stopped because the fundamental aim of coalescent theory is to understand genetic variation within samples. All genetic variation in a sample must be the result of mutations that occurred on these branches of the gene genealogy, between the present and the time of the most recent common ancestor.

Kingman (1982a,c,b) described his n-coalescent using the discrete-time, haploid, exchangeable population model of Cannings (1974), with its now familiar distribution of "offspring numbers". In a population of size N, if ν_i denotes the number of offspring of individual i in some generation, then the overall outcome of reproduction in that generation is the vector $(\nu_1, \ldots, \nu_N)$. These offspring numbers are exchangeable random variables in that they are identically distributed and have only a mild sort of non-independence, specifically that $\sum_{i=1}^{N} \nu_i = N$. Making the further assumptions that N is constant over time and that the genotypes of individuals do not affect the distribution of offspring numbers, we have the basic ingredients of Kingman's model and many of its later extensions.

From a more biological point of view, in addition to this explicit statement of constant population size, we must recognize a number of other assumptions. First, the organisms are haploid. The oft-stated view that the coalescent holds for diploid populations if N is replaced by $2N$ comes from the analysis of the diploid, monecious Wright–Fisher model (Fisher, 1930; Wright, 1931) which is, in essence, a haploid model. Second, there is no selection; all genetic variation is assumed to be neutral. Third, there is no geographic structure. For haploids, this means that the locations of individuals (and resulting population densities, etc.) have no effect on their numbers of offspring. The same restriction applies to diploids, but for diploids it must also be true that geographic distance poses no barrier to mating. All three caveats are aspects of the fundamental assumption behind exchangeability—namely that there is no population structure whatsoever—and one could go on to consider the consequences for other biologically relevant phenomena such as age structure and assortative mating.

The standard neutral coalescent process, or the Kingman coalescent, describes the genetic ancestry of a sample of size n from such a population provided that the second and higher moments of the offspring-number distribution are not too great and that the population size is sufficiently large. The latter assumption is commonly stated as $n \ll N$. Formally, a limiting process is described for n finite and $N \to \infty$, with time rescaled in proportion to N. Under Cannings' model, time is rescaled so that it is measured in units of N/σ^2 generations. When this is done, each pair of ancestral lineages coalesces independently with rate equal to 1. Therefore, we have the following.

(a) The time during which there are i lineages ancestral to the sample follows an exponential distribution,

$$f(t) = \binom{i}{2} e^{-\binom{i}{2}t}. \tag{1}$$

(b) When a coalescent event occurs, it is equally likely to involve any one of the $\binom{i}{2}$ pairs of lineages.

Although the setting of Kingman's proof excludes much of what captivates biologists, by using a different approach, it can be shown that the standard neutral coalescent process holds rather more broadly, in the face of substantial population structure (Möhle, 1998a,b,c). This robustness occurs when N is large because many types of structure, including dioecy, partial selfing, age structure, and some instances of geographic subdivision, exert their influence over time scales much shorter than N generations. Thus, a variety of more complicated population models map onto the Kingman coalescent. In view of this, Sjödin et al. (2005) suggested that the term "coalescent effective population size" be used whenever (a) and (b) hold in the limit $N \to \infty$ with time measured in units of $N_e \equiv N/c$, for some constant c which will depend on the details of the model.

Felsenstein (1989) once reviewed the edited volume by Feldman (1989) using the provocative title "Mathematics vs. Evolution". His point was that we should see the value of mathematical rigor for what is it, and not be swayed by the notion that a work filled with symbols and abstruse logic must be important. It is worthwhile recognizing and reflecting upon the tension between the mathematician's incessant drive toward rigorous generalization and the biologist's comfort with exceedingly concrete models and results. Doing so leads to a greater appreciation of coalescent theory. With its explicit focus on the sampling properties of genetic data, it provides an enduring and fruitful point of contact between mathematics and biology.

0040-5809/$ – see front matter
http://dx.doi.org/10.1016/j.tpb.2013.06.001

The paper by Sjödin et al. (2005) is a good example. In a large and confusing literature following Wright (1931, 1938), the idea of effective population size was to map a given complicated population model onto the simple diploid, monecious Wright–Fisher model. Unfortunately, there is no natural way to do this mathematically. The idea is even of limited heuristic value because the diploid, monecious Wright–Fisher model is not realistic for any organism. The truth is that many different population models map onto the Kingman coalescent in the limit $N \to \infty$. It might seem backwards, but the notion of the (coalescent) effective size is what justifies the heuristic use of the diploid, monecious Wright–Fisher model in obtaining results of broader applicability. Provided that the population size is not too small, Sjödin et al. (2005) emphasize the powerful statement that, when the coalescent effective size exists in the limit, the sampling properties of genetic data in all their complexity are given by the standard neutral coalescent.

Coalescent processes have immediate, tangible application in several areas of science. For example, (a) and (b) arise in the theory of stochastic coalescence of water droplets in clouds (Marcus, 1968). In population genetics, their application is more indirect, and arises by reversing time in the standard modes of population genetics. Here, the initial objects of study are populations of individuals or genetic loci evolving in a type space, forward in time, according to some set of rules. Types are commonly genotypes, and the rules typically describe the processes of mutation and population-level reproduction. The Kingman coalescent and closely related extensions are backward-time dual processes of the standard diffusion models of population genetics (Donnelly and Kurtz, 1999; Möhle, 1999).

It was from this point of view that Kingman originally sought to discover the genealogical process underlying the sampling structure of selectively neutral alleles in one particular model: the step-wise mutation model (Ohta and Kimura, 1973; Moran, 1975; Kingman, 2000). Here, the type space is the integers, on which the individuals perform a random walk. The gene genealogy, with its branches of common ancestry, emerges as a description of the non-independence of these walkers, who do not become ever more dispersed but rather wander around together in a clump. Interestingly, we can see the great-grandchild of Kingman's approach in the recent work of Bolthausen and Sznitman (1998), Brunet et al. (2006, 2007), Berestycki et al. (2012), and others, which incorporates directional selection on these wandering distributions, and holds the promise of novel practical methods for the analysis of sequence data (Neher and Hallatschek, 2013).

Given the intimate relationship between the forward-time models of population genetics and the standard neutral coalescent, there were many precursors to Kingman's insights. One may see the coalescent embedded, for example, in the work of Felsenstein (1971), who used what amounts to Cannings' exchangeable model to study the loss of genetic diversity in a population without mutation; in the recursive equations of Karlin and McGregor (1972), who gave a proof of the Ewens sampling formula (Ewens, 1972); and in the analyses of Watterson (1975), who exploited the conditional independence of mutations given patterns of ancestry to obtain the generating function of the number of polymorphic sites in a sample. However, in none of these works was attention to gene genealogies overt.

The explicit focus on ancestral genetic relationships as key features of population genetic processes is due to Malécot (1946, 1948). It was in this tradition that Hudson (1983b) and Tajima (1983) took gene genealogies as objects worthy of study, and introduced the standard neutral coalescent process to biologists. Both Hudson (1983b) and Tajima (1983) obtained the coalescent process starting from the diploid, monecious Wright–Fisher model which, as noted above, is unnecessarily restrictive but nonetheless leads to many general results. The subsequent flurry of work in the remainder of the 1980s established the fundamental results for standard neutral gene genealogies, for example the relationship between gene trees and species trees (Pamilo and Nei, 1988), and extended coalescent models to include many biologically relevant phenomena, such as recombination (Hudson, 1983a; Kaplan and Hudson, 1985), selection (Hudson and Kaplan, 1988; Kaplan et al., 1988), and population subdivision and migration (Takahata, 1988).

2. New contributions presented in this special issue

The extensions to the standard neutral coalescent mentioned at the close of Section 1 are described in detail, along with many subsequent developments, in Hein et al. (2005) and Wakeley (2008), yet almost nothing of what follows in this special issue can be found in those texts. Several novel lines of research have gained prominence in the intervening years, and the time is ripe for a snapshot of the state of coalescent theory. In sketching backgrounds and emphasizing new results, the contributions speak for themselves. Here, we will consider the major unifying themes.

Following the work of Pitman (1999), Sagitov (1999), Donnelly and Kurtz (1999), and Schweinsberg (2000), research on coalescent processes with multiple mergers of ancestral lines has expanded greatly. The robustness of the Kingman coalescent breaks down under large deviations in the higher moments of the offspring-number distribution. No longer does each pair of ancestral lineages coalesce with rate equal to 1. Instead, mergers involving any numbers of lineages may occur. Models in which at most one such event occurs at any given time are known as Λ-coalescents (Pitman, 1999), and models in which multiple coalescent events may occur simultaneously are called Ξ-coalescents (Schweinsberg, 2000). A number of articles in this volume make novel contributions to this thriving literature.

Huillet and Möhle (2013, pages 5–14) study the $(N \to \infty)$ convergence properties of extended Moran models of reproduction, in which the single individual who reproduces may have an arbitrary number of offspring. In the original model of Moran (1958, 1962), the parent always has two offspring. However, if the number of offspring may be large, in particular of order N, then a wide range of generalized coalescent processes arises. In fact, any Λ-coalescent may arise. These models differ dramatically in their predictions from the Kingman coalescent, and they may be used to interpret data which cannot be explained under the standard neutral model.

Biologists are interested in coalescent theory because it provides a framework for the analysis of DNA sequence data. Steinrücken et al. (2013a, pages 15–24) describe a method of inferring the parameters of a flexible class of coalescent processes with multiple mergers, a subset of Λ-coalescents called Beta-coalescents. Their method uses importance sampling, after the classic approach of Griffiths and Tavaré (1994a,b). They apply their method to data from Pacific oysters and Atlantic cod, both of which have a tremendous capacity for reproduction and thus may violate the assumptions behind the Kingman coalescent.

Natural selection is a defining feature of evolutionary change, yet much of the simplicity of the Kingman coalescent derives from its assumption of neutrality. Two main approaches to selection have been taken in coalescent theory, the approach of Kaplan et al. (1988), which conditions on the trajectories of alleles, and the ancestral selection graph approach of Krone and Neuhauser (1997), which allows ancestral lineages to branch as they are followed backward in time. The next two articles in this volume make progress on the ongoing challenge of extending and applying these approaches.

Analytical results from either approach have been rare. In particular, the ancestral selection graph has appeared intractable

under strong selection due to high rates of branching. Pokalyuk and Pfaffelhuber (2013, pages 25–33) confront this problem directly, using the ancestral selection graph to give a novel proof of the fixation time of a favorable allele when selection is strong. This fixation time has been employed in descriptions of the haplotype structure at genetic loci which have undergone selective sweeps and in the development of methods to detect these loci in genome-wide scans for selection (Kim and Nielsen, 2004; Pennings and Hermisson, 2006).

A major challenge of population genetics is to distinguish among the great variety of forms that natural selection may take. Taylor (2013, pages 34–50) adapts the approach of Hudson and Kaplan (1988) and Barton et al. (2004) to study the effect of fluctuating selection on a linked neutral site. Notably, the form of the diffusion approximation in this case differs from that of the standard models of population genetics (Gillespie, 1991). Taylor's analysis identifies differential signatures of linked genetic variation between subtly different kinds of diversifying selection, suggesting ways in which genomic data might be used to distinguish the varieties of selection.

The robustness of the standard neutral coalescent justifies its use as a prior model for the interpretation and analysis of genetic data in a variety of statistical settings. For example, the methods of Griffiths and Tavaré (1994a,b) are importance-sampling methods in which the proposal distribution is the unconditional coalescent process (Felsenstein et al., 1999). Stephens and Donnelly (2000) used coalescent arguments to devise a more efficient proposal distribution, specifically taking the data into account. The popular program PHASE, which reconstructs haplotypes from diploid SNP genotypes, employs a similar conditional sampling distribution (Stephens et al., 2001). Two articles develop new applications of the standard neutral coalescent, using the model as a prior in a statistical setting.

Steinrücken et al. (2013b, pages 51–61) combine two sets of insights to develop conditional sampling distributions of haplotypes in a model with migration and recombination. On the one hand, they employ an improved version of the methods of Fearnhead and Donnelly (2001) and Li and Stephens (2003) for sampling entire haplotypes in the face of recombination. On the other hand, they use the simplifying approximation of the sequentially Markov coalescent (Wiuf and Hein, 1999; McVean and Cardin, 2005), which models coalescence and recombination as point processes along a sequence. By combining these approaches, Steinrücken, Paul, and Song obtain a tractable (approximate) method of computing likelihoods for the inference of migration rates from DNA sequence data.

Huang et al. (2013, pages 62–74) consider the problem of genotype imputation using a two-population version of the standard neutral coalescent as a prior. Under a model with no migration and no recombination, and focusing on a sample of three sequences, they are able to obtain analytical expressions that assess the accuracy of imputed genotypes. Genotype imputation is proving to be an important tool in genome-wide association studies that use low-coverage sequencing to identify the genes affecting complex traits and diseases (Li et al., 2009).

Recombination is a conceptually straightforward but nonetheless complicating factor which has run through coalescent theory since the beginning (Hudson, 1983a). The articles by Taylor (2013, pages 34–50) and by Steinrücken et al. (2013b, pages 51–61) mentioned above deal directly with recombination. They illustrate, respectively, the two major concerns: (1) the effect of selection on linked sites, and (2) the background, neutral correlation of gene genealogies along the genome. The two articles by Barton et al. (2013a,b, pages 75–89 and pages 105–119), discussed below, likewise address the influence of recombination in these two contexts.

Geographic population structure has been a major theme in the expansion of coalescent theory beyond the simple Kingman coalescent. In contrast to recombination, it is not clear how one should model geographic structure. A main line of division in the theory has been between models with discrete subpopulations (Wright, 1931) and models in which individuals occupy continuous habitats (Wright, 1943; Malécot, 1948). However, relatively little is known about the dynamics of genetically meaningful movement of individuals in nature. Like selection, geographic structure negates the simplifying assumption of exchangeability in the Kingman coalescent, and hence similarly has been the source of many formidable problems. In addition to the articles by Steinrücken et al. (2013b, pages 51–61) and by Huang et al. (2013, pages 62–74), the final three articles in this volume address issues of geographical structure.

Barton et al. (2013a, pages 75–89) study coalescent processes that arise when a selectively favorable allele spreads quickly across a continuously distributed population. In one dimension, the process is similar to selection in the stepwise mutation model mentioned above, which yields a Bolthausen–Sznitman coalescent, but here allelic state is replaced by geographic location. Testing quasi-deterministic approximations to coalescence and recombination during sweeps in two dimensions, Barton, Etheridge, Kelleher, and Véber uncover a non-diffusive jump process for the locations of ancestral lineages inside the leading edge of the sweep.

Heuer and Sturm (2013, pages 90–104) prove a novel robustness result. For a spatial Λ-coalescent with migration between neighboring populations on a two-dimensional lattice, they show that the Kingman coalescent holds in the limit of a large number of populations as long as the samples are taken from sufficiently far apart. In other samples, nearby lineages will coalesce quickly, after which those that remain may be far enough apart for the Kingman coalescent to apply. Examining genetic pseudo-data in simulations, Heuer and Sturm show that this two-phase model may be useful even when the Kingman coalescent is not a good approximation.

Barton et al. (2013b, pages 105–119) capitalize on this separation of times scales to investigate the utility of two different types of data—allele frequencies versus haplotype blocks—in estimating the parameters of populations living in two-dimensional habitats. They consider both discrete-habitat and continuous-habitat models without selection, and make extensive use of an approach embodied in a classic formula of Wright and Malécot, which gives the generating function of pairwise coalescence times. The two types of data contain different sorts of information about the coalescent process in subdivided populations, and correspondingly each allows inferences only about a subset of the parameters.

To summarize, the articles in this special issue display the current broad range of coalescent theory. They present novel analyses motivated by the interpretation of genetic data, and with the attention to inference that is so important to this field. The issues addressed fall under four major headings: (1) coalescent processes with multiple mergers of ancestral lines, (2) the effects of natural selection, (3) coalescents as prior models in statistical inference, and (4) the analysis of geographically structured populations. Yet, despite the attempts of this introduction, they cannot be assigned seamlessly to categories, because they span them in multiple interesting ways. Though it has been said many times, the future still holds great promise that burgeoning genetic data can be combined with quantitative analyses to take our understanding of population genetics and evolution in new directions.

References

Barton, N.H., Etheridge, A.M., Sturm, A.K., 2004. Coalescence in a random background. Ann. Appl. Prob. 14, 754–785.

Barton, N.H., Etheridge, A.M., Kelleher, J., Véber, A., 2013a. Genetic hitchhiking in spatially extended populations. Theoretical Population Biology 87, 75–89.

Barton, N.H., Etheridge, A.M., Kelleher, J., Véber, A., 2013b. Inference in two dimensions: allele frequencies versus lengths of shared sequence blocks. Theoretical Population Biology 87, 105–119.
Berestycki, J., Berestycki, N., Schweinsberg, J., 2012. The genealogy of branching Brownian motion with absorption. Annals Probab. 41, 527–618.
Bolthausen, E., Sznitman, A.S., 1998. On Ruelles probability cascades and an abstract cavity method. Comm. Math. Phys. 197, 247–276.
Brunet, E., Derrida, B., Mueller, A.H., Munier, S., 2006. Noisy traveling waves: effect of selection on genealogies. Europhys. Lett. 76, 1–7.
Brunet, E., Derrida, B., Mueller, A.H., Munier, S., 2007. Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. Phys. Rev. E 76, 041104.
Cannings, C., 1974. The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. Adv. Appl. Prob. 6, 260–290.
Donnelly, P., Kurtz, T.G., 1999. Particle representations for measure-valued population models. Ann. Probab. 27, 166–205.
Ewens, W.J., 1972. The sampling theory of selectively neutral alleles. Theoret. Pop. Biol. 3, 87–112.
Fearnhead, P., Donnelly, P., 2001. Estimating recombination rates from population genetic data. Genetics 159, 1299–1318.
Feldman, M.W. (Ed.), 1989. Mathematical Evolutionary Theory. Princeton University Press, Princeton, NJ.
Felsenstein, J., 1971. The rate of loss of multiple alleles in finite haploid populations. Theoret. Pop. Biol. 2, 391–405.
Felsenstein, J., 1989. Mathematics vs. evolution. Science 246, 941–942.
Felsenstein, J., Kuhner, M.K., Yamato, J., Beerli, P., 1999. Likelihoods on coalescents: a monte carlo sampling approach to inferring parameters from population samples of molecular data. In: Seillier-Moiseiwitsch, F. (Ed.), Statistics in Molecular Biology and Genetics. In: IMS Lecture Notes-Monograph Series, vol. 33. Institute of Mathematical Statistics, Hayward, California, pp. 163–185.
Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Clarendon, Oxford.
Gillespie, J.H., 1991. The Causes of Molecular Evolution. Oxford University Press, New York.
Griffiths, R.C., Tavaré, S., 1994a. Ancestral inference in population genetics. Statistical Science 9, 307–319.
Griffiths, R.C., Tavaré, S., 1994b. Simulating probability distributions in the coalescent. Theoret. Pop. Biol. 46, 131–159.
Hein, J., Schierup, M.H., Wiuf, C., 2005. Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory. Oxford University Press, Oxford.
Heuer, B., Sturm, A., 2013. On spatial coalescents with multiple mergers in two dimensions. Theoretical Population Biology 87, 90–104.
Huang, L., Buzbas, E.O., Rosenberg, N.A., 2013. Genotype imputation in a coalescent model with infinitely-many-sites mutation. Theoretical Population Biology 87, 62–74.
Hudson, R.R., 1983a. Properties of a neutral allele model with intragenic recombination. Theoret. Pop. Biol. 23, 183–201.
Hudson, R.R., 1983b. Testing the constant-rate neutral allele model with protein sequence data. Evolution 37, 203–217.
Hudson, R.R., Kaplan, N.L., 1988. The coalescent process in models with selection and recombination. Genetics 120, 831–840.
Huillet, T., Möhle, M., 2013. On the extended Moran model and its relation to coalescents with multiple collisions. Theoretical Population Biology 87, 5–14.
Kaplan, N.L., Darden, T., Hudson, R.R., 1988. Coalescent process in models with selection. Genetics 120, 819–829.
Kaplan, N.L., Hudson, R.R., 1985. The use of sample genealogies for studying a selectively neutral *m*-loci model with recombination. Theoret. Pop. Biol. 28, 382–396.
Karlin, S., McGregor, J., 1972. Addendum to a paper of W. Ewens. Theoret. Pop. Biol. 3, 113–116.
Kim, Y., Nielsen, R., 2004. Linkage disequilibrium as a signature of selective sweeps. Genetics 167, 1513–1524.
Kingman, J.F.C., 1982a. The coalescent. Stochastic Process. Appl. 13, 235–248.
Kingman, J.F.C., 1982b. Exchangeability and the evolution of large populations. In: Koch, G., Spizzichino, F. (Eds.), Exchangeability in Probability and Statistics. North-Holland, Amsterdam, pp. 97–112.
Kingman, J.F.C., 1982c. On the genealogy of large populations. J. Appl. Prob. 19A, 27–43.
Kingman, J.F.C., 2000. Origins of the coalescent: 1974–1982. Genetics 156, 1461–1463.
Krone, S.M., Neuhauser, C., 1997. Ancestral processes with selection. Theoret. Pop. Biol. 51, 210–237.
Li, N., Stephens, M., 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165, 2213–2233.
Li, Y., Willer, C., Sanna, S., Abecasis, G., 2009. Genotype imputation. Annu. Rev. Genomics Hum. Genet. 10, 387–406.
Malécot, G., 1946. La consaguinité dans une population limitée. C. R. Acad. Sci., Paris 222, 841–843.
Malécot, G., 1948. Les Mathématiques de l'Hérédité. Masson, Paris, Extended translation: The Mathematics of Heredity. W.H. Freeman, San Francisco, 1969.
Marcus, A.H., 1968. Stochastic coalescence. Technometrics 10, 133–143.
McVean, G.A.T., Cardin, N.J., 2005. Approximating the coalescent with recombination. Phil. Trans. R. Soc. B 360, 1387–1393.
Möhle, M., 1998a. Coalescent results for two-sex population models. Adv. Appl. Prob. 30, 513–520.
Möhle, M., 1998b. A convergence theorem for Markov chains arising in population genetics and the coalescent with partial selfing. Adv. Appl. Prob. 30, 493–512.
Möhle, M., 1998c. Robustness results for the coalescent. J. Appl. Prob. 35, 438–447.
Möhle, M., 1999. The concept of duality and applications to Markov processes arising in neutral population genetics models. Bernoulli 5, 761–777.
Moran, P.A.P., 1958. Random processes in genetics. Proc. Camb. Phil. Soc. 54, 60–71.
Moran, P.A.P., 1962. Statistical Processes of Evolutionary Theory. Clarendon Press, Oxford.
Moran, P.A.P., 1975. Wandering distributions and the electrophoretic profile. Theoret. Pop. Biol 8, 318–330.
Neher, R., Hallatschek, O., 2013. Genealogies of rapidly adapting populations. Proc. Natl. Acad. Sci., USA 109, 437–442.
Ohta, T., Kimura, M., 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res., Camb. 22, 201–204.
Pamilo, P., Nei, M., 1988. The relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583.
Pennings, P.S., Hermisson, J., 2006. Soft sweeps II: molecular population genetics of adaptation from recurrent mutation or migration. Mol. Biol. Evol. 23, 1076–1084.
Pitman, J., 1999. Coalescents with multiple collisions. Ann. Probab. 27, 1870–1902.
Pokalyuk, C., Pfaffelhuber, P., 2013. The ancestral selection graph under strong directional selection. Theoretical Population Biology 87, 25–33.
Sagitov, S., 1999. The general coalescent with asynchronous mergers of ancestral lines. J. Appl. Probab. 36, 1116–1125.
Schweinsberg, J., 2000. Coalescents with simultaneous multiple collisions. Electron. J. Probab. 5, 1–50.
Sjödin, P., Kaj, I., Krone, S., Lascoux, M., Nordborg, M., 2005. On the meaning and existence of an effective population size. Genetics 169, 1061–1070.
Stephens, M., Donnelly, P., 2000. Inference in molecular population genetics. J. R. Stat. Soc. Ser. B 62, 605–655.
Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978–989.
Steinrücken, M., Birkner, M., Blath, J., 2013a. Analysis of DNA sequence variation within marine species using beta-coalescents. Theoretical Population Biology 87, 15–24.
Steinrücken, M., Paul, J.S., Song, Y.S., 2013b. A sequentially Markov conditional sampling distribution for structured populations with migration and recombination. Theoretical Population Biology 87, 51–61.
Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105, 437–460.
Takahata, N., 1988. The coalescent in two partially isolated diffusion populations. Genet. Res., Camb. 53, 213–222.
Taylor, J.E., 2013. The effect of fluctuating selection on the genealogy at a linked site. Theoretical Population Biology 87, 34–50.
Wakeley, J., 2008. Coalescent Theory: An Introduction. Roberts & Company Publishers, Greenwood Village, Colorado.
Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. Theoret. Pop. Biol. 7, 256–276.
Wiuf, C., Hein, J., 1999. Recombination as a point process along sequences. Theoret. Pop. Biol. 55, 248–259.
Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.
Wright, S., 1938. Size of population and breeding structure in relation to evolution. Science 76, 430–431.
Wright, S., 1943. Isolation by distance. Genetics 28, 114–138.

John Wakeley
Harvard University, 4096 Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA
E-mail address: wakeley@fas.harvard.edu.

Available online 5 June 2013