

Coalescent Theory; An Introduction by John Wakeley

Reviewed by Bernhard Haubold
Max-Planck-Institute for Evolutionary Biology
Plön, Germany

March 3, 2009

Revolutions in science tend to be rather quiet affairs. Mendel's attempt in 1866 to augment Darwin's *Origin* by an atomic theory of inheritance was fully absorbed into evolutionary biology only half a century later. The resulting field of population genetics was in turn revolutionized by the introduction of the coalescent in the early 1980's. Yet many biologists are still unsure what the coalescent is and how it might be applied to their own data. John Wakeley's beautiful book *Coalescent Theory; An Introduction* is for them.

Genetic variation is ubiquitous and its best known molecular equivalent, single nucleotide polymorphisms, are uncovered at an ever growing rate in contemporary comparative sequencing projects. Understanding the distribution of this genetic polymorphism within populations and along chromosomes is one of the central challenges of biology. One reason for this is that the distribution of polymorphisms helps to identify regions under selection in complex genomes such as ours.

Since the inception of population genetics in the wake of Mendel's rediscovery, gene evolution was modeled forward in time. This perspective made intuitive sense and provided the basis for important insights into the evolutionary process. However, one of its major disadvantages was that in computer simulations entire populations needed to be represented over many generations. In the early 1980's the English mathematician John Kingman proposed an alternative model of evolution: starting from a sample of homologous genes, he traced back their ancestry. At certain points in time, two lines of descent would coalesce into one, until the last common ancestor of the sample was reached, at which point the process stopped. The resulting gene genealogy is called *the coalescent* (Figure 1).

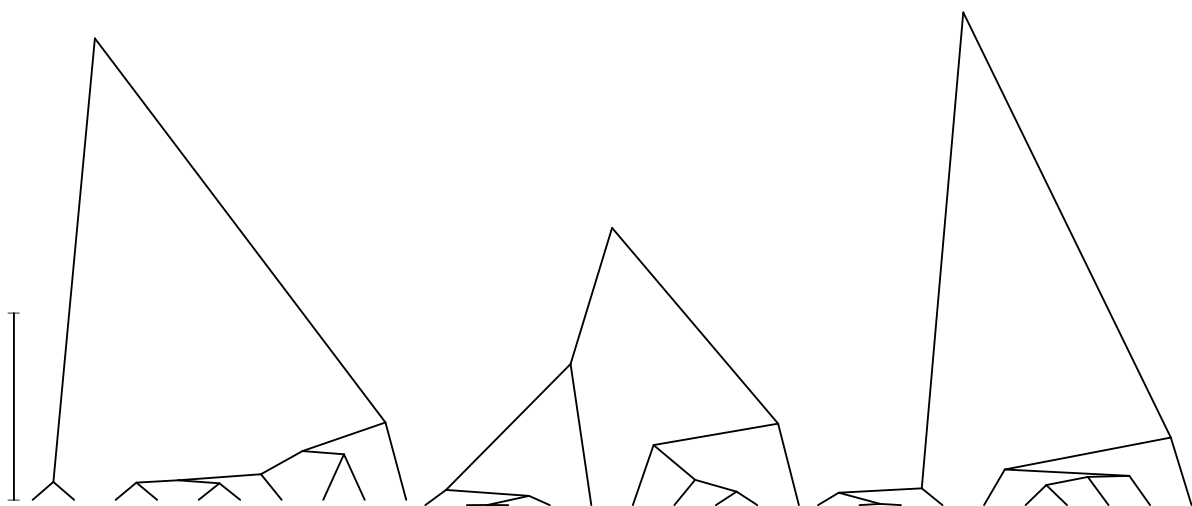


Figure 1: Three random genealogies (coalescents) of ten genes. The rule indicates $2N$ generations, where N is the population size.

Wakeley's book is a definitive exposition of the theory that has developed over the past quarter century around Kingman's model of evolution. He starts with some motivating background including the application of coalescent

theory in one of the two original publications of the human genome in 2001. Since the coalescent is a stochastic model, chapter 2 explains the probability theory necessary to understand the rest of the book. Derivations in subsequent chapters are carefully cross-referenced to the material presented here. Chapter 3 then formally introduces Kingman's coalescent and shows in what sense it approximates the two most important forward in time models of population genetics: the Wright Fisher model and the Moran model. In the next chapter mutations are added to the coalescent yielding a description of selectively neutral genetic variation. The illuminating connection between the frequency spectrum of mutations and the structure of the underlying coalescent is explained, leading to a discussion of widely used statistics for testing deviations from the standard neutral model of evolution, including Tajima's D . This sets the scene for the subsequent chapters where the standard neutral model is extended by the addition of migration, selection, and recombination.

The final chapter introduces the important field of coalescent simulation. Most contemporary analyses of population genetic data are based on such simulations. These were the focus of an earlier textbook on the coalescent, *Gene Genealogies, Variation, and Evolution* published in 2005 by Jotun Hein, Mikkel Schierup and Carsten Wiuf. In contrast to their book, Wakeley emphasizes the conceptual and mathematical structure of coalescent theory. He does so in a delightfully lucid way. His intended target audience are biologists, though it ought to be clear from what has been said so far that these should not be scared by mathematical argument. But given such a mind set, reading *Coalescent Theory* will be a highly rewarding experience, for it is the first treatment of the subject that gently takes the reader from the fundamentals right up to contemporary research problems. Wakeley emphasizes that the purpose of coalescent theory is the interpretation of real sequence data and he presents numerous detailed examples of such analyses taken from the research literature. Mastery of the material presented in *Coalescent Theory* will therefore help biologists to think fruitfully about their own comparative sequence data. My only wish for subsequent editions would be the addition of a set of solved problems to each chapter. This would make the book even more useful for teachers and students of evolutionary biology.

A powerful idea in science evolves from revolutionary to main stream by the arrival of one or more canonical text books. Wakeley's text may well turn out to be the book that delivered coalescent theory to mainstream biology.

Details

Coalescent Theory; An Introduction

Publisher: Roberts and Company Publishers

Publication year: 2008

ISBN: 0-9747077-5-9