

Ola Hössjer  
Dept. of Mathematics  
Stockholm University  
ola@math.su.se  
Fall 2014

## Coalescence Theory and Population Genetics

### General Information

During November 2014, I will give a 2.5 credits PhD course, “Coalescence Theory and Population Genetics”. The purpose of the course is to develop tools for understanding variation of human DNA between individuals and populations. This variation is caused by a number of different phenomena, including:

1. Random inheritance (Mendelian laws).
2. Changes of DNA due to mutations in sperm or ova cells.
3. Demographics, such as population growth, bottlenecks (when populations get very small), and migration between different geographic regions.
4. Recombination of maternal and paternal DNA.
5. Selection due to varying fitness.

Population genetics is the discipline where mathematical models (in particular probability theory and stochastic processes) are used to assess how 1-5 affect DNA variation.

The course is intended to introduce the student to some of the mathematical models of population genetics, with particular focus on coalescence theory. This is a powerful technique for describing how the genealogy or ancestral tree of a set of individuals (or chromosomes) is distributed. This distribution reflects our incomplete knowledge of the past when we go many generations back in time. In the course we will mainly focus on 1-3, although 4-5 will be mentioned briefly as well.

## Lecturer and examiner

Ola Hössjer, Department of Mathematics, Division of Mathematical Statistics, Kräftriket, House 6, Room 322. Email: ola@math.su.se.

## Literature

Durrett, R. (2008). *Probability models for DNA sequence evolution*, 2nd edition. Springer, New York. ISBN: 978-0-387-78168-6. (Parts of Chapters 1,2 and 4.)

Some articles and a primer, to make the reader familiar with some genetics.

## Prerequisites

It is assumed that the student has taken one basic and at least one advanced course in probability theory and stochastic processes, as well as a theoretical course in statistics (inference) theory. No prior knowledge of genetics is required.

The prerequisites correspond to the contents of the following courses at Stockholm University: Probability theory 1 and 2 (MT3001, MT5002), Stochastic Processes 1 and 2 (MT4002, MT5004), Inference theory (MT5003) and one of the two courses Probability theory 3 or Stochastic Processes 3 (MT7001, MT7023).

## Examination

Two home assignments (HA 1-2) will be handed out, with three exercises each. These are to be solved and turned in individually, with latest days of return as indicated in the schedule below. In order to pass the 2.5 credits PhD course, about half of the maximum number of credits is required.

The course may also be taken as the first third of the 7.5 credits master course MT7030 “Selected topics in biostatistics” (“Valda ämnen i biostatistik”). Its last two thirds will be taught by Federica Giardina in December. The final grade (A-F) on MT7030 is then a weighted average of the two grades from each part.

## Schedule

We meet five times at Kräftriket, House 5, Room 32, 10.15-12.00, according to the following schedule:

Day	Topic	Chapter in RD	HAs
Nov 10	Basic genetics, various population genetic models (Wright-Fisher, Moran, infinite alleles).	1.1-1.2	1 out
Nov 14	Coalescence theory, Ewen's sampling formula, Hoppe's urn scheme, random permutations.	1.2-1.3, 4.1	
Nov 17	Infinite sites model, statistics for assessing variation of DNA between individuals.	1.4.1-1.4.3, 2.1-2.4	2 out
Nov 21	Population complications: Varying size or migration/geographical subdivision.	4.2-4.6	
Nov 24	Geographic substructure, contd. Inference and testing of different genealogies.	4.6, 2.1-2.4	1 in
Dec 1			2 in

Welcome!

Ola