

Editorial for the Special Issue on Models and Inference in Population Genetics

Jean-Michel Marin¹ and François Rousset²

Population genetics is a scientific field that is concerned with the study of genetic variation within and between populations. That is a core sub-field of evolutionary biology, which studies phenomena such as adaptation, speciation, or population structure. In that scientific area, the concept of genetic polymorphism is central. It refers to the coexistence of several alleles for a given gene (or locus) in different populations. Genetic polymorphism is a major reason for individuals from the same population to have different phenotypic characters. It is an essential component of the adaptation of populations to changing environments.

Long before Rosalind Franklin, James Watson and Francis Crick introduce the double-helix structure of DNA, mathematicians and statisticians of the last century laid the conceptual foundations and mathematical formalization of the evolution of genetic variation. The pioneering works of Sewall Wright, John Haldane and Ronald Fisher, generally considered as the founders of population genetics, have been extended to describe and understand the causes of polymorphisms in data generated by modern techniques, such as genomic sequences.

With the recent development of molecular biology technologies, the size of genomic datasets has increased considerably. At the same time, mathematicians and theoretical population geneticists have provided efficient stochastic modelling for genomic sequences. In the field of population genetics, one of the most popular is the coalescent process introduced by John Kingman. Contrary to the Wright-Fisher process, Kingman's coalescent describes the distribution of genetic variation for a sample of individuals coming from one population by considering only the ancestors of the sample, not the whole population. It involves a small number of genetic parameters and, by extending that process to several populations connected by present or past migrations, it is possible to consider complex demographic scenarios with clearly interpretable demographic parameters.

From statistical and inferential point of views, Kingman's coalescent creates specific difficulties. This model do not generally lead to explicit likelihoods. Indeed, the presence of complex latent structures, such as genealogies of genes, greatly complicates the calculation of the likelihood. This issue has led to very fertile collaborations between statisticians and population geneticists and innovative inferential methodologies have been introduced. The impact of these methodologies has now largely surpassed the framework of evolutionary biology. Likelihood-

¹ IMAG, Univ Montpellier, CNRS, Montpellier, France

E-mail: jean-michel.marin@umontpellier.fr

² Institut des Sciences de l'Évolution, Univ Montpellier, CNRS, IRD, EPHE, Montpellier, France

E-mail: francois.rousset@umontpellier.fr

free inference methods to estimate the genetic and demographic parameters or to discriminate between evolutionary scenarios have becoming generic statistical inferential tools.

The goal of this Special Issue is to show some advances in how to use genetic data from related populations to infer about their recent evolutionary history and to introduce some likelihood-free inferential techniques.

In a first paper, Pierre Pudlo and Mohammed Sedki describe some population genetic models under neutrality, involving genetic drift and mutations. Starting with Kingman's coalescent, they show how structured populations can be modeled.

François Rousset and coauthors survey an approach using sequential importance sampling techniques derived from coalescent and diffusion theory. This approach requires the re-implementation of methods often considered in the context of computer experiments methods, in particular of Kriging which is used as a smoothing technique to infer a likelihood surface from likelihoods estimated in various parameter points.

In a last paper, Arnaud Estoup and coauthors show how simulation-based methods such as Approximate Bayesian Computation (ABC) are well adapted to make statistical inferences about complex models of natural population histories. A noticeable novelty of the statistical analyses they present includes the application of the ABC Random Forests methodology to make model choice on predefined groups of models to make inferences about the history of Pygmy human populations from Western Central Africa from a microsatellite genetic dataset.