Model choice using Approximate Bayesian Computation via random forests: analyzes based on model grouping to make inferences about the genetic history of Pygmy human populations

Arnaud Estoup[*1], Louis Raynal[2], Paul Verdu[3], and Jean-Michel Marin[†2]

[1]CBGP – Institut national de la recherche agronomique (INRA) – France
[2]IMAG – Université de Montpellier, Université de Montpellier – France
[3]MNHN – Muséum National d'Histoire Naturelle (MNHN) : UMR7206, Muséum National d'Histoire Naturelle (MNHN) : UMR7206 – France

Abstract

In evolutionary biology, simulation-based methods such as Approximate Bayesian Computation (ABC) are well adapted to make statistical inferences about complex models of natural population histories, in which serial or independent divergence events, genetic admixture or migration events, and change(s) of population sizes are often suspected. Pudlo et al (2016) recently developed a novel approach based on a machine learning tool named Random Forests (RF): the ABC-Rf scheme. Here we present the results of analyses based on ABC-RF to make inferences about the genetic history of Pygmy human populations from Western Central Africa from a microsatellite genetic dataset. A noticeable novelty of the statistical analyses presented here includes the application of ABC-RF algorithms to make model choice on predefine groups of models. We formalized eight complex evolutionary scenarios which incorporate (or not) three major events: (i) whether there is an independent or non-independent origin of Pygmy populations, (ii) the possibility of introgression/migration events between Pygmy and non-Pygmy populations, and (iii) the possibility of a change in size in the past in the non-Pygmy African population. We show that our scenario grouping approach allows detangling with strong confidence the main events that compose the evolutionary history of interest. The selected scenario corresponds to a common origin of all Western Central African Pygmy groups with the ancestral Pygmy population having diverged from the non-Pygmy population in a remote past. It encompasses asymmetrical introgression events from the non-Pygmy gene-pool into each Pygmy population, and a change of population size in the non-Pygmy African population.

---
[*]Corresponding author: arnaud.estoup@inra.fr
[†]Speaker