Report thesis LATRILLE Thibault

31st of October 2020

Fabienne Macro
Université Claude Bernard - Lyon 1

UPPSALA UNIVERSITET

Carina Farah Mugal

PhD

Department of Ecology and Genetics,
Uppsala University

Address:
Norbyvägen 18D
75236 Uppsala
Sweden

Phone:
018-471 64 66
+46 18 471 64 66

www.ieg.uu.se

carina.mugal@ebc.uu.se

Report on the PhD thesis work by Thibault Latrille

The PhD thesis work by Thibault Latrille is focused on phylogenetic codon models. More specifically, it discusses the necessity of simplifying model assumptions for statistical inference, and then raises the problematics that result from those simplifying model assumptions in current approaches. Three specific topics are addressed, which result in interdisciplinary model extensions of codon evolution covering population genetics theory and protein biophysics.

The first study addresses the robustness of classical codon models to mutation bias, and proposes a model extension that improves the accuracy in statistical inference of mutation bias. The second study develops a statistical framework that enables the inference of effective population size (N_e) based on mechanistic codon models. Application of the new software to a range of datasets results in an unexpectedly narrow range of observed variation in inferred N_e, which is suggested to result from the underlying model assumptions. The problematic is taken up in the third study, and relates phylogenetic codon models to the field of protein biophysics.

At the same time as the focus of the PhD thesis is very specific, phylogenetic codon models, it covers a broad range of aspects within computational molecular evolution, ranging from mathematical theory, methodological developments, software implementation to data analysis. The breadth of the accomplished work is impressive. The accomplished work is also very creative and of rigorous scientific quality.

The thesis manuscript opens with a very comprehensive introduction (Part I), that provides relevant background information in an easily accessible manner. I really enjoyed reading this section. The historical perspective on molecular evolution summarizes early developments and different approaches to molecular evolution, and thereby nicely outlines

the problematic of integrating theories across disciplines, which is also the main topic of the PhD thesis work. The following chapters of part I provide an overview of relevant mathematical and biological theory and statistical tools, which prepare and motivate the three thesis objectives in a clean and sound fashion.

Part II of the thesis manuscript constitutes of the three scientific studies briefly summarized above. A summary and discussion of the three studies is provided in Part III, which also puts the three studies in a broader scientific context and brings up possible model extensions of interest.

The presentation of the PhD thesis work is comprehensive and of shows excellent scientific maturity of the PhD candidate Thibault Latrille.

New insights often lead to new questions, which is nicely illustrated by the transition from the second to the third study. In relation to this, some more questions came into my mind, which I hope to discuss during the PhD defense.

1) The second study emphasizes the role of epistasis to explain the observed narrow range of variation in N_e. I was wondering if the assumption of a fixed selection coefficient across the phylogenetic tree also could contribute to the observation of a narrow range of variation in N_e? If the fitness effect of new mutations decreases with increasing N_e, wouldn't also this result in a narrow range of inferred variation in N_e?

2) Several definitions of N_e exist in the theoretical population genetics literature. This results in the non-trivial question, what N_e is it that is inferred in the second study? Besides, note that it is strictly speaking not correct to simply replace census population size (N) by N_e, in the probability of fixation derived by Kimura (equation 2.19 in the thesis manuscript). This approximation is based on simplifying assumptions, which could be interesting to discuss in more detail.

3) The third study investigates the behavior of ω after changes in N_e over time in 100 million years for different models of fitness. It could be interesting to elaborate on the choice of the time-scale, and also relate it to scenarios observed in nature.

Based on my evaluation of the PhD thesis work by Thibault Latrille, I am clearly favorable that the defense can take place. I am looking forward to discussing his work with Thibault Latrille.