

August 5, 2023

Dear Editor in Chief,

Please find attached our manuscript titled “How to validate a Bayesian evolutionary model” by Fábio K. Mendes, Remco Bouckaert, Luiz M. Carvalho and Alexei J. Drummond for consideration as a research article in *PLoS Computational Biology*.

Large amounts of biological data of different kinds (molecular, morphological, geographical) are increasingly available in public databases. As a result, computational/statistical methods for analyzing this wealth of information are also increasingly available (many of them published in this very journal). In evolutionary biology, **methods implementing probabilistic models are immensely popular** – specifically those within the Bayesian paradigm. Bayesian software platforms we have contributed to ourselves, e.g., BEAST, BEAST 2, RevBayes, have accrued tens of thousands of citations.

As researchers developing and reviewing Bayesian methods on a daily basis, we noticed that different computational tools can be widely variable with respect to how they are tested and validated. This variation partly reflects the idiosyncrasies of each methodological contribution, e.g., the empirical analyses therein. But perhaps more critically, it also reflects a lack of consensus or common ground in the community as to what validation means and how to achieve it. We believe this is because computational biology method development is relatively young and largely multidisciplinary. As such, **information on good validation practices is yet to be synthesized and made available in accessible language for researchers in disparate fields**.

Here, we define validation as the verification of method *correctness*, and summarize best practices in probabilistic model validation for method developers, with an emphasis on Bayesian methods. We execute two different validation protocols on variations of a simple hierarchical model, discuss the results, and expand on how to interpret other potential outcomes. We further introduce a suite of methods for automating these protocols within the BEAST 2 platform. Lastly, we propose method development guidelines that will assist method users, developers and referees in quickly **finding common ground when evaluating new modeling work**. We expect that in defining standards that are desirable and reasonable, these guidelines will **enhance both the release rate and standards of statistical software** for biology.

Sincerely,

Fábio K. Mendes, Remco Bouckaert, Luiz M. Carvalho and Alexei J. Drummond