HAL
archives-ouvertes.fr

Aevol-4b: Toward a new simulation platform to benchmark phylogenetic tools

Nicolas Comte, Vincent Liard, Carole Knibbe, Guillaume Beslon

▶ **To cite this version:**

Nicolas Comte, Vincent Liard, Carole Knibbe, Guillaume Beslon. Aevol-4b: Toward a new simulation platform to benchmark phylogenetic tools. ALPHY (ALignments and PHYlogeny), Feb 2017, Villeurbanne, France. hal-01569078

HAL Id: hal-01569078
https://hal.archives-ouvertes.fr/hal-01569078

Submitted on 26 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Aevol-4b: Toward a new simulation platform to benchmark phylogenetic tools.

Nicolas Comte, Vincent Liard, Carole Knibbe, Guillaume Beslon

Aevol (www.aevol.fr) is a computational platform that allows for the study and manipulation of populations of digital organisms evolving under different conditions. Using Aevol simulations, one can better understand evolutionary forces and mechanisms leading to specific genome and transcriptome structures, as well as indirect selection pressures involved in the evolution of cooperation and genetic information transfer.

Recently, we used aevol as a benchmarking tool. Indeed, Molecular evolutionary methods and tools are difficult to validate, as we have almost no direct access to ancient molecules. Inference methods may be tested with simulated data but this requires that the inference methods and the simulation be design independently (Biller et al., Computation in Europe 2016; Biller et al., Jobim 2016). Using aevol we can simulate perfectly characterized phylogenies and obtain a final population that evolved accordingly. Then we can use this final population to try to recover the initial phylogeny using various tools and assess their efficiency in doing so.

This approach has recently been applied to test various estimators of inversion distance, revealing their limits and suggesting important improvement directions (Biller et al., Genome Biology and Evolution 2016). However, current aevol structure – more specifically the use of a binary representation for the genomic sequence – strongly limits its usability as a benchmarking tool. That is why we recently started the development of a new version of the software in which the genome sequence will use a four-nucleotides code and the translation from genetic sequence to polypeptide sequences will use the extant genetic code to map the 4-bases alphabet to the 20-amino-acids one.

Although the development of this new version is in its infancy a first prototype has been developed and we would like to discuss the main modelling choices with the Alphy community that will be the potential users of the generated benchmarks. In particular, in this prototype the genotype-to-phenotype map is be based a mathematical description of traits under selection and on A.D. Solis (Proteins, 2015) classification of amino-acids, two crucial modelling choices that deserve discussion before we start final software implementation.

Priscila Biller, Carole Knibbe, Guillaume Beslon, Eric Tannier. Comparative Genomics on Artificial Life. *Computability in Europe*, 2016, Paris, France. Lecture Notes in Computer Science 9709, pp. 35-44

Priscila Biller, Eric Tannier, Guillaume Beslon, Carole Knibbe. In silico experimental evolution provides independent and challenging benchmarks for comparative genomics. *Journées ouvertes Biologie Informatique Mathématiques*, Jun 2016, Lyon, France. pp. 79-82, 2016

Priscila Biller, Laurent Guéguen, Carole Knibbe, Eric Tannier. Breaking Good: Accounting for Fragility of Genomic Regions in Rearrangement Distance Estimation. *Genome Biology and Evolution*, 2016, 8(5), pp.1427-1439.

Armando D. Solis. Amino acid alphabet reduction preserves fold information contained in contact interactions in proteins. *Proteins*, 2015, 83(12), pp. 2198–2216