

Simulation of WGS

The simulation of a read library based on an existing sequence is often used for quality control and benchmarking of next generation sequencing methods. This is possible because the simulation provides a read library under controlled conditions. Thus, libraries can be simulated under perfect, natural but also under the most adverse conditions. Simulating read sets as they are obtained from large scale sequencing projects is meanwhile common, and many different tools have been developed for this purpose. An overview is given in [Escalona et al. \(2016\)](#). Figure 1 from this publication gives an overview of the available tools. Simulated data have the advantage that we can basically control each step in the data generation, and thus we are in possession of a gold standard for each step in a biosequence analysis. This will help to very precisely assess the performance of each algorithm used during data analysis ([Greshake, et al. 2016](#)).



Figure 1: Decision tree from [Escalona et al. \(2016\)](#) of how to use which simulator for high-throughput sequencing data sets.

[ART \(Huang, et al. 2012\)](#) is a simulation tool to generate synthetic next-generation sequencing reads. ART simulates sequencing reads by mimicking the real sequencing process with empirical error models or quality profiles summarized from large re-calibrated sequencing data.

From:

<https://fsbioinf.biologie.uni-frankfurt.de/teaching/wiki/> - Teaching

Permanent link:

<https://fsbioinf.biologie.uni-frankfurt.de/teaching/wiki/doku.php?id=general:bioseqanalysis:readsimulation>

Last update: **2021/10/19 20:16**

