

SiteSampler

Simon Y. W. Ho

School of Biological Sciences, University of Sydney

Introduction

SiteSampler is a Java application that can sample the sites of a sequence alignment to produce replicate data sets. Sampling of the sites can be done in two ways:

- 1) **With replacement.** Any particular site in the alignment can be sampled any number of times. This is equivalent to randomly shuffling the sites of the alignment.
- 2) **Without replacement.** Any particular site in the alignment can only be sampled once. This is equivalent to randomly shuffling the sites of the alignment.

Site sampling can be performed within predefined data partitions. This might be useful if you wish to generate bootstrap replicates with the original data partitions preserved.

Sequence labels (names) can also be randomly reassigned. This might be useful for certain randomisation tests, such as date-randomisation in heterochronous data sets.

Using SiteSampler

The application can be executed on any platform provided that Java has been installed.

To execute *SiteSampler*, place the .jar file and the sequence alignment in the same directory. From a command line in the directory, type the following:

```
java -jar SiteSampler.jar
```

The application will prompt the user for the following:

- 1) **Input file.** This is the name of the file containing the sequence alignment, which should be in Fasta format. Individual sequences should not run across multiple lines.
- 2) **Output file root.** This is the prefix to be used for each of the output files.
- 3) **Number of aligned sequences.**
- 4) **Number of randomisations required.**
- 5) **Resample sites?** This determines whether sites are to be resampled or not. If not, then the original alignment is maintained (no bootstrapping or shuffling).

Resample sites with replacement?

- 6) **Randomise sequence labels?**
- 7) **Number of partitions.** Number of partitions in the sequence alignment.

Sites in partition n . The number of sites in the n th partition.

Citation

If you use *SiteSampler* in your research, please cite the following paper.

Ho SYW, and Lanfear R (2010) Improved characterization of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks. *Mitochondrial DNA*, 21: 138-146.