MSAProbs-MPI

Reference Manual
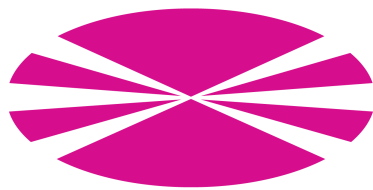
Authors:
Jorge González-Domínguez
Yongchao Liu
Juan Touriño
Bertil Schmidt

Institutions:
Department of Electronics and Systems, University of a Coruña, Spain
Institute of Computer Science, Johannes Gutenberg University Mainz, Germany

Date:
May 2, 2016

Contents

1 Introduction

MSAProbs-MPI is a parallel multiple protein sequence aligner designed by combining hidden Markov models and partition functions to calculate posterior probabilities. The tool is implemented with MPI in order to exploit the parallel capabilities of multicore clusters.

The *MSAProbs-MPI* tool is developed as a research collaboration of the Department of Electronics and Systems at the University of A Coruña (Spain) and the Institute of Computer Science at the Johannes Gutenberg University in Mainz (Germany). The tool is distributed as free software and publicly available under the GPLv3 license at:

<p align="center"><code>http://msaprobs.sourceforge.net</code></p>

The corresponding license file is shipped with the software but can also be accessed via:

<p align="center"><code>http://www.gnu.org/licenses/gpl-3.0.en.html</code></p>

If you want to reuse the code, please ensure compliance to the aforementioned license and a proper attribution/citation of the original work/authors.

2 Installation

To complete the installation of *MSAProbs-MPI* follow these steps:

1. Untar the archive and move into the *MSAProbs-MPI* directory.

2. Update the file *Makefile* of the root directory in order to indicate the correct path and libraries for the MPI compiler installed in your system.

3. Type `make` to build *MSAProbs-MPI*.

3   Execution

MSAProbs-MPI can be executed with any MPI running command (e.g., `mpirun`, `mpiexec`). The only compulsory argument is the input FASTA dataset. The optional arguments for the program are:

- *-o, –outfile.* String to specify the output file name (STDOUT by default).

- *-num_threads.* Integer with the number of threads per MPI process (1 by default).

- *-clustalw.* If specified, use CLUSTALW output format instead of FASTA format.

- *-c, –consistency.* Integer with the number of passes of consistency transformation (2 by default, 0 minimum, and 5 maximum).

- *-ir, –iterative-refinement.* Integer with the number of passes of iterative refinement (10 by default, 0 minimum, and 1000 maximum).

- *-v, –verbose.* If specified, report progress while alignment.

- *-annot.* String to specify an annotation file for the multiple alignment.

- *-a, –alignment-order.* If specified, print output sequences in alignment other rather than input order.

- *-version.* If specified print out version of *MSAProbs-MPI*.

For instance, the following command prints in the file *outAlign.fasta* the multiple sequence alignment of *dataset1.fasta* on eight cores using two MPI processes with four OpenMP threads each, three consistency transformation passes and 100 iterations for refinement.

```
mpirun -n 2 ./msaprobs-mpi dataset1.fasta -num_threads 4 -c 3
-it 100 -o outAlign.fasta
```