

# MrBayes Best Practices for Intel® Cluster Ready



BEST PRACTICES

## 1. Introduction

The following best practices document is provided as courtesy of the HPC Advisory Council.

## 2. Application Description:

MrBayes is a program for the Bayesian estimation of phylogeny. Bayesian inference of phylogeny is based upon the posterior probability distribution of trees. Trees are labelled T1, T2, ..., Tn, where n is the number of possible trees. The posterior probability of the i-th tree is calculated using Bayes's formula as:

$$\Pr[T_i | X] = \Pr[X | T_i] \times \Pr[T_i] / \Pr[X]$$

where X is a character matrix. Here, "Pr[Ti | X]" is the posterior probability of the i-th tree, "Pr[X | Ti]" is the likelihood of the i-th tree, and "Pr[Ti]" is the prior probability of the i-th tree. The denominator of Bayes's formula ("Pr[X]") is a normalizing constant that involves a summation over all possible trees. The likelihood, as described above, cannot be calculated with knowledge of only the tree's topology. You also need to have information on the lengths of the branches and on the mechanism of character change. Hence, the likelihood ("Pr[X | Ti]") involves a multidimensional integral over all possible combinations of branch lengths and substitution model parameters.

MrBayes including its sources and pointers to the authors can be found at:

<http://mrbayes.csit.fsu.edu>

## 3. Version Information:

Version of this date is used: MrBayes v3.1.2h

The Hybrid version of MrBayes can be found from this URL:

<http://wwwkramer.in.tum.de/exelixis/software.html>

## 4. Prerequisites:

The instructions from this best practice have been tested with the following configuration:

### 4.1 Hardware:

- Dell PowerEdge M610 38-node cluster
- Intel Xeon X5670 CPUs @ 2.93 MHz
- Memory: 24GB per node @ 1333MHz

- Mellanox ConnectX-2 QDR InfiniBand Adapters
- Mellanox QDR InfiniBand Switch

### 4.2 Software:

- Intel® Cluster Ready running RHEL 5.5
- Mellanox OFED 1.5.2 InfiniBand Software Stack
- Application: MRBAYES
- Compilers: Intel compilers
- MPI: Intel MPI 4, Open MPI 1.5.3, Platform MPI 8.0.1
- Benchmark workload: primates.nex

## 5. Obtaining the source code

The original version of MrBayes can be found from this URL:

<http://mrbayes.csit.fsu.edu/download.php>

The Hybrid MPI/OpenMP version of MrBayes can be found from this URL:

<http://wwwkramer.in.tum.de/exelixis/software.html>

## 6. Building MrBayes

After extracted the MrBayes-3.1.2h.tar.gz or mrbayes-3.1.2.tar.gz files, you will need to copy one of the Makefile available and modify for your environment.

For making hybrid OpenMP/MPI to work, you need to use the Makefile.icc.hybrid. The default settings should work for Intel Compilers and MKL.

For GCC with OpenMP, you will need to add "-lgomp" for OPTFLAGS to make it compiled.

```
% cp Makefile.icc.hybrid Makefile
```

```
% make
```

At the end of the compilation, the executable named "mb" will be generated.

If you get this error, you will need to install the readline-devel RPM.

```
gcc -DUNIX_VERSION -DUSE_READLINE -O3 -Wall -c -o bayes.o bayes.c
```

```
bayes.c:45:31: error: readline/readline.h: No such file or directory
```

```
bayes.c:46:30: error: readline/history.h: No such file or directory
```

## 7. Running MrBayes

The input dataset is defined in the \*.nex (Nexus) input files. In the nexus file, it should have information on the top of the files that defines the dimension (number of taxa and characters) and the format of the dataset. In the nexus file, there may or may not be a block named "mrbayes" included. If the block is not included, you will need to define some of the information from a separate input file.

A note about the "nrns" and "nchains" parameters in the "mrbayes" block.

The reason for doing two "runs" is to see whether they obtain similar consensus trees and parameters. Making at least two runs is always recommended. The interface permits nrns=2 and nrns=4. Running odd numbers is inefficient AND slows your run down.

By default, MrBayes will set nchains to 4, and nrns to 2.

It means that MrBayes will use 3 heated chains and one "cold" chain. The MrBayes manual shows that heating is essential for problems with more than about 50 taxa. For smaller problems MrBayes can analyze successfully without adding heat. However, adding more than three heated chains may be helpful in analyzing large and difficult data sets.

Typically user can use nrns=2 and nchains=4 without problems. The number of nrns \* nchains needs to be a multiple of 2. One thing to keep in mind is that in the batch file that you create:

- The number of cores for the mpirun is equal to the product of number of chains and the number of runs in the analysis.

### Running with Intel MPI

```
% mpdboot -parallel-startup -r ssh -f <PATH_TO_
HOSTFILE > -n 16
```

```
% mpiexec -ppn 1 -np 16 -IB <MRBAYES_HOME>/
mb < <MRBAYES_HOME>/input
```

```
%mpdallexit
```

### Running with Platform MPI

```
% mpirun -np 16 -IBV -prot -hostfile <PATH_TO_
HOSTFILE> <MRBAYES_HOME>/mb < <MRBAYES_
HOME>/input
```

### Running with Open MPI

```
% mpirun -np 16 -mca btl self,sm,openib -hostfile
<PATH_TO_HOSTFILE> < <MRBAYES_HOME>/mb
<MRBAYES_HOME>/input
```

