

Random Jungle, version 1.2.365

A powerful Random Forests (TM) implementation
Edition 1.2.365, 15 March 2011

by Daniel F. Schwarz et al.
(bug@randomjungle.com)

This manual (15 March 2011) is for Random Jungle (RJ) version 1.2.365, a package containing an implementation of the Random Forests (TM) method.

Copyright © 2008-2011 Daniel F. Schwarz et al..

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License.”

Table of Contents

1	Introduction and preliminaries	3
1.1	Introduction to Random Jungle	3
1.2	Historical references	3
1.3	Problems and bugs	3
1.4	Using this manual	3
2	Compiling and installing	5
2.1	Compiling and installing Random Jungle	5
3	Invoking Random Jungle	7
3.1	Command line options for operation modes	7
4	Input data	13
4.1	Input the whole data	13
4.2	Restricted analysis	13
5	Output data	15
5.1	The standard output files	15
5.2	Log file	15
5.3	Confusion file	15
5.4	Prediction file	15
5.5	Importance file	15
5.6	Verbose file	16
6	Error codes	17
6.1	Runtime error codes	17
7	Some nice examples	19
7.1	Simple example	19
7.2	Simple example with ped file	19
7.3	Estimating variable importance	19
7.4	Using R and rjungle	19
7.5	Using plink and rjungle	20
7.6	Prediction	20
7.7	Imputation	20
7.8	Using plink and rjunglesparse	20
7.9	Using MPI	21
Appendix A How to make copies of the overall Random Jungle package		23
A.1	License for copying the Random Jungle package	23

Appendix B How to make copies of this manual **35**

B.1 License for copying this manual..... 35

Appendix C Indices of concepts and macros .. 43

C.1 Index for all `rrjungle` macros 43

C.2 Index for many concepts..... 44

`Random Jungle` is an implementation of Random Forests (TM). It is mostly compatible to the original fortran code. `Random Jungle` also has additional builtin functions.

`Random Jungle` is written by Daniel F. Schwarz et al.. All names can be found in the files `'randomjungle-1.2.365/AUTHORS'` and `'randomjungle-1.2.365/THANKS'` from the `Random Jungle` distribution.

This is release 1.2.365. It is considered unstable: future releases are meant to add new features, fix bugs, increase speed, or improve documentation. However. . .

1 Introduction and preliminaries

This first chapter explains what **Random Jungle** is, where **Random Jungle** comes from, how to read and use this documentation, how to call the **Random Jungle** program, and how to report bugs about it. It concludes by giving tips for reading the remainder of the manual.

The following chapters then detail all the features of the **Random Jungle** .

1.1 Introduction to Random Jungle

Random Jungle is an implementation of Random Forests (TM). It is supposed to analyse high dimensional data. In genetics, it can be used for analysing big Genome Wide Association (GWA) data. Random Forests (TM) is a powerful machine learning method. Most interesting features are variable selection, missing value imputation, classifier creation, generalization error estimation and sample proximities between pairs of cases.

Random Jungle is mostly compatible with the Linux and Windows (TM). Also, there exists compatibilities with Solaris.

1.2 Historical references

First ideas were seeded in 2006 during GAW15 workshop in Florida. In the beginning of 2008, Daniel F. Schwarz released 0.5.0 and 0.5.1 which was a "way pre-release" and fast, but lagged of features and documentation. However, the williams award was won at the International Genetic Epidemiology Society (IGES) conference 2008 in St. Louis with release 0.5.2. In late 2008 and 2009, intensive work improved the platform compatibility, added documentation and raised the number of features of **Random Jungle**.

1.3 Problems and bugs

If you have problems with **Random Jungle** or think you've found a bug, please report it. Before reporting a bug, make sure you've actually found a real bug. Carefully reread the documentation and see if it really says you can do what you're trying to do. If it's not clear whether you should be able to do something or not, report that too; it's a bug in the documentation!

Before reporting a bug or trying to fix it yourself, try to isolate it to the smallest possible input file that reproduces the problem. Then send us the input file and the exact results **Random Jungle** gave you. Also say what you expected to occur; this will help us decide whether the problem was really in the documentation.

Once you've got a precise problem, send e-mail to bug@randomjungle.com. Please include the version number of **Random Jungle** you are using. You can get this information with the command `rjungle --version` or `rjungle -Z`.

Non-bug suggestions are always welcome as well. If you have questions about things that are unclear in the documentation or are just obscure features, please report them too.

1.4 Using this manual

This manual contains a number of examples of **Random Jungle** input and output, and a simple notation is used to distinguish input, output and error messages from **Random Jungle**. Examples are set out from the normal text, and shown in a fixed width font, like this

This is an example of an example!

To distinguish input from output, all output from `Random Jungle` is prefixed by the string ‘`⇒`’, and all error messages by the string ‘`[error]`’. When showing how command line options affect matters, the command line is shown with a prompt ‘`$ like this`’, otherwise, you can assume that a simple `rjungle` invocation will work. Thus:

```
$ command line to invoke rjungle
```

```
Example of input line
```

```
⇒Output line from rjungle
```

```
[error] and an error message
```

The sequence ‘`^D`’ in an example indicates the end of the input file. The sequence ‘`NL`’ refers to the newline character. The majority of these examples are self-contained, and you can run them with similar results by invoking `rjungle -d`. In fact, the testsuite that is bundled in the `Random Jungle` package consists of the examples in this document! Some of the examples assume that your current directory is located where you unpacked the installation, so if you plan on following along, you may find it helpful to do this now:

```
$ cd randomjungle-1.2.365
```

2 Compiling and installing

2.1 Compiling and installing Random Jungle

There are pre-compiled versions on the internet site <http://www.randomjungle.com/>. Maybe, there exists a pre-compiled version for your platform (computer) and you do not need to compile it.

Random Jungle uses some GNU tools, libraries and other free open source code. The installation routine (`./configure`) prompt you to install packages which are missing on your machine. It is also recommended to compile the sources with an openMP supporting compiler (i.e. GNU gcc > 4.2). Invoke the following commands to compile `randomjungle`:

```
./configure
make
make check
make install
```

For compiling **Random Jungle** with MPI (message passing interface) support, configure RJ using option `--enable-mpi`. The program is optimized for Open MPI (<http://www.open-mpi.org/>). The MPI mode should be used by experienced users only.

```
./configure --enable-mpi
```


3 Invoking Random Jungle

The format of the `Random Jungle` command is:

```
rjungle [option...]
```

or

```
rjunglesparse [option...]
```

All options begin with ‘-’, or if long option names are used, with ‘--’. On some platforms long options might not work. A long option name need not be written completely, any unambiguous prefix is sufficient. POSIX requires `Random Jungle` to recognize arguments intermixed with files, even when `POSIXLY_CORRECT` is set in the environment. Most options take effect at startup regardless of their position, but some are documented below as taking effect after any files that occurred earlier in the command line. The argument ‘--’ is a marker to denote the end of options.

With short options, options that do not take arguments may be combined into a single command line argument with subsequent options, options with mandatory arguments may be provided either as a single command line argument or as two arguments, and options with optional arguments must be provided as a single argument. In other words, `rjungle -QPDfoo -d a -df` is equivalent to `rjungle -Q -P -D foo -d -df -- ./a`, although the latter form is considered canonical.

With long options, options with mandatory arguments may be provided with an equal sign (=) in a single argument, or as two arguments, and options with optional arguments must be provided as a single argument. In other words, `rjungle --def foo --debug a` is equivalent to `rjungle --define=foo --debug= -- ./a`, although the latter form is considered canonical (not to mention more robust, in case a future version of `Random Jungle` introduces an option named ‘--default’).

`Random Jungle` understands the following options, grouped by functionality.

3.1 Command line options for operation modes

Several options control the overall operation of `rjungle`:

`-h`

`--help` Print a help summary on standard output, then immediately exit `rjungle` without reading any input files or performing any other actions.

`-Z`

`--version`

Print the version number of the program on standard output, then immediately exit `rjungle` without reading any input files or performing any other actions.

`-f FILENAME`

`--file=FILENAME`

FILENAME of input file with data. Input data has to be numerical. The default *FILENAME* is `input.dat(.gz)`. In R (<http://www.r-project.org>), save data to file using `write.table` as follows:

```
> make.your.data.in.R
> write.table(yourData, file = "input.dat", row.names = FALSE, quote = FALSE)
> quit()
$ rjungle -f input.dat [...]
```

In plink (pngu.mgh.harvard.edu/~purcell/plink/), save data to raw file using the `recodeA` option, set the ped file option and char memory mode option in `rjungle` as follows:

```
$ plink --file yourDataFile --recodeA
$ rjungle -f yourDataFile.raw -p -M 2 [...]
```

Avoid missing values in data. See See [Chapter 4 \[Input data\]](#), page 13, for more details.

`-o FILEPREFIXNAME`

`--outprefix=FILEPREFIXNAME`

`FILEPREFIXNAME` of output files is the first part of output files (i.e. `rjungle.importance`, `rjungle.prediction`, ...). The default `FILEPREFIXNAME` is `rjungle`. Use for example `-o my_analysis_no123`.

`-e CHAR`

`--delimiter=CHAR`

Set the delimiter in your input file to `CHAR`. Default is a space.

`-w2`

Save model (forest) to a file `FILEPREFIXNAME.jungle.xml`.

`-P FILENAME`

Predict data with model (forest) in file `FILENAME` (data is given with option `-f`). Predictions will be written to `FILEPREFIXNAME.prediction`.

`-y ID`

`--treetype=ID`

Choose base classifier by setting `ID` There are several treetypes but only CART is fully supported. Explanatory or exposure variables will be named: input variables. Explained or response variable will be named: output variable. If you want to use CART like Random Forest (TM) does choose one of three possible values as follows:

= 1 or 5 [ID]

CART, y (output variable): nominal, x (input variable(s)): numeric, `ID = 1` is recommended for less different values in the input variables (i.e. GWA SNP data or integer data). `ID = 5` is recommended for more different values in the input variables (i.e. many floating point numbers). Like original Breiman/Cutler/Friedman algorithm.

= 2 [ID]

CART, y (output variable): nominal, x (input variable(s)): nominal,

= 3 [ID]

CART regression trees, y (output variable): numeric, x (input variable(s)): numeric,

= 4 [ID]
 CART regression trees, CART regression trees, y (output variable): numeric, x (input variable(s)): nominal,

Default is a 1.

-t *SIZE*

--ntree=*SIZE*

SIZE is the number of trees in jungle. If *SIZE*=0 then the size will be set automatically depending on mtry and variable size (experimental feature). Default is 500.

-m *SIZE*

--mtry=*SIZE*

SIZE of randomly chosen variable sets. At each node building step, a variable will be selected out of the set, that serves the biggest information gain. The bigger *SIZE* is set, the higher computing time might be. The bigger *SIZE* is set, the more similar trees in jungle will be. High noised data sets should be processed with a big *SIZE*. Default is square root of number of input variables.

-x *NUM*

--missingcode=*NUM*

Missings should always be coded as NA or *NUM* in your data. The program takes *NUM* as an internal representation of a missing value. Default is -99.

-i *ID*

--impmeasure=*ID*

Variable selection: Choose a method for estimating variable importance as follows:

= 1 [ID]
 Intrinsic Importance (i.e. GINI-Index).

= 2 [ID]
 Permutation Importance by Breiman, Cutler (observed in Fortran code).

= 3 [ID]
 Permutation Importance by Liaw, Wiener (in R-package RandomForest).

= 4 [ID]
 Permutation Importance, raw values, no normalization.

= 5 [ID]
 Permutation Importance by Meng et al.

The results will be written to file *FILEPREFIXNAME.importance*. Default is 1 and you cannot turn off variable importance output.

-B *ID*

Variable selection/ model optimization: Choose a method for estimating variable importance as follows:

= 0 [ID]
 No backward elimination.

- = 1 [ID]
Backward elimination. Discard 50% at each step (slow). Stop if number of variables shrunk to *STOPSIZE*, see option -j.
- = 2 [ID]
Backward elimination. Discard 33% at each step (slow). Stop if number of variables shrunk to *STOPSIZE*, see option -j.
- = 3 [ID]
Backward elimination. Discard only negative values at each step (slow/recommended). (Shown at IGES2007 by Inke R. Konig).

Default is 0.

-j *STOPSIZE*

--nimpvar=*STOPSIZE*

Only necessary if --impmeasure = 2,3,5,6 or 7. How many variable should remain. The lesser *STOPSIZE* is, the reliable the result might be. The smaller *SIZE* is, the higher computing time will be. Default is 100.

-v

--verbose

Print some nice information to screen. Otherwise put information to file *FILEPREFIXNAME.verbose* (default).

-u

--downsampling

Choose randomly samples without replacement. Switched off is default.

-M *ID*

--memmode=*ID*

Usage of the heap memory (RAM) as follows:

- = 0 [ID]
Double precision floating point (BIG).
- = 1 [ID]
Single precision floating point (Normal).
- = 2 [ID]
Char (small). CHAR normally fits in one byte. DATA CELL VALUE HAS TO BE AN INTEGER IN [-127..127].

If you want to use very small data coding, i.e. for SNP analysis, give *rjunglesparse* a try! Default is 0.

-C *FILE*

--colselection=*FILE*

Only use selected columns listed in *FILE*. Example content of *FILE*:

```
var1
var20
var1000
```

`var300`

Default is take all variables.

`-D NAME`

`--depvarname=NAME`

Output variable name in the data SET! If *NAME* is empty then the `rjungle` switches to unsupervised mode.

`-s`

`--sampleproximities`

It computes proximities between pairs of cases that can be used in clustering, locating outliers, or (by scaling) give interesting views of the data. Can be used as the distance matrix for Multidimensional Scaling (MDS). The results will be written to file *FILEPREFIXNAME.samproximity*. Default: switched off

`-z`

`--seed` Seed of random number generators.

`-U`

`--nthreads=NUM`

Maximally use *NUM* threads (CPUs) for parallel processing. Limit for *NUM* is number of CPUs in computer. Default: Number of CPUs in computer.

`-p`

`--pedfile`

Input file has got ped format (i.e. plink output with `recodeA`). Default: switched off.

`-I NUM`

`--impute=NUM`

Impute missings in input data using Random Forest(TM)'s imputation algorithm. The number of iterations is given by *NUM*. For imputing continuous data, use option `-A` (`--impcont`) as well. For more information, have a look at http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.

Do not try to impute untyped SNPs (Schwarz et al. 2009, BMC Proceedings, 3, S65) if case-control-status is missing. Try a different program like: IMPUTE, MACH, PLINK, ... Default: switched off.

`-k NUM`

`--maxtreedepth=NUM`

This is a stop criterium/tunning parameter. Tree growing will stop, when the tree exceeds a depth of *NUM*. Default: switched off.

`-l NUM`

`--targetpartitionsize=NUM`

This is a stop criterium/tunning parameter. Tree growing will stop, when a partion falls below a size of *NUM* samples. Default: switched off.

`-K NUM`

`--condimp=NUM`

Perform conditional importance if option `-i` > 1. *NUM* is the pearson's cor. coef. cutoff. The smaller *NUM*, the bigger a conditional importance permutation

group will be created. (=> More accurate, but slower) Requires: $0 \leq \text{NUM} \leq 1$ $\text{NUM} < 0 \Rightarrow$ switched off DEFAULT: switched off.

-W

Adjust class weights of unbalanced datasets automatically. DEFAULT: switched off.

--oobset Outputs the oobset of forest in file *.oob (each row == one tree) DEFAULT: switched off.

--classweights *STRING*

Sets the weights of classes E.G.: "1.23;22;" ... first class gets weight 1.23 ... second class gets weight 22 DEFAULT: switched off

-d -g -n -r -c -a -b -w -V -S -P -A -G -E -X

Those options are experimental. Be cautious about using one of those.

4 Input data

This chapter describes various input file types of `rjungle`.

4.1 Input the whole data

`Random Jungle` analyses data given in a file and name of response variable, see option `-f` and `-D`. The file format is matrix like and is as follows:

names [Variable]
 The first line of the input file contains the variable names. The variable name must not contain quotes or space characters. The variables are separated by space characters.

cells [Data]
 Each following line represents one sample (observation). Every single sample needs to have one numerical value for each variable. The values are separated by space characters and are ordered corresponding to the variable names (of course).

Here, an example of an input file:

```
responseVar inputVar1 inputVar2 inputVar3 inputVar4
0 1.2 3.4 5.6 7.8
0 1.1 3.3 5.5 7.7
0 2.2 4.4 6.6 8.8
1 1.0 3.0 5.0 7.0
1 1.0 3.0 5.0 7.0
1 2.0 4.0 6.0 8.0
```

If you choose the ped file option `-p` then the input file format is ped like. The file must have variables FID, IID, PAT, MAT, SEX, PHENOTYPE and at least three variables (SNPs). Here, an example of an input file:

```
FID IID PAT MAT SEX PHENOTYPE rs1 rs2 rs3 rs4
1000 NA2001 0 0 2 0 0 1 0 1
1000 NA2002 0 0 2 0 0 2 0 1
1001 NA2003 0 0 1 0 1 2 0 1
1001 NA2004 0 0 1 0 1 1 1 1
1002 NA2005 0 0 2 1 1 1 1 1
1004 NA2006 0 0 2 1 1 2 1 1
1005 NA2007 0 0 2 1 2 2 2 2
1006 NA2008 0 0 1 1 2 2 2 2
1007 NA2008 0 0 1 1 2 2 2 2
```

4.2 Restricted analysis

The `rjungle` can be run also with just a subset of all variables which are given in the input file (option `-f`). See description of option `-C`, for more details.

5 Output data

This chapter describes various output file types of `rjungle`.

5.1 The standard output files

If `rjungle` is executed it will always produce three files. See [Section 5.2 \[Log file\], page 15](#), [Section 5.3 \[Confusion file\], page 15](#) and [Section 5.5 \[Importance file\], page 15](#). You can run an example in your `randomjungle-1.2.365` directory and see it.

```
$ cd tests
$ rjungle -f test.ped -p -v -M2
$ ls
rjungle.importance  rjungle.log  rjungle.prediction  test.ped
```

Or without PED file as follows:

```
$ rjungle -f your_data_file.txt -D response_variable_name -v
```

5.2 Log file

The log file contains all options / parameter of the `rjungle` execution. This is useful to reproduce results.

5.3 Confusion file

The `rjungle` always evaluates the classifier. The accuracy on training and test data. Test data is called oob data which is collected during growing process (similar to cross validation's test sets). The results are shown in so-called confusion matrices in file `FILEPREFIXNAME.confusion`. Columns represent the predicted values and rows represent the real values.

When using option `-y1` a file `FILEPREFIXNAME.confusion2` will be created. The file contains class specific error rates.

5.4 Prediction file

The `rjungle` is able to perform new data to a saved classifier (option `-w2` and `-P`). The predicted results will be saved to file `FILEPREFIXNAME.prediction`.

5.5 Importance file

Random Jungle estimates the importance of variables (option `-i`). The results are saved to file `FILEPREFIXNAME.importance` and/or `FILEPREFIXNAME.importance2`. The first file contains four columns but most interesting columns are `varname` (variable name) and `value` (importance value). The higher the `value`, the more important is the variable with name `varname`. This list is sorted ascending. So, look at the tail of file to see the most important variables. The second file contains various permutation importance informations.

5.6 Verbose file

If Random Jungle is invoked from command line without verbose option it is very schtum and do not output anything to screen. It writes all information occuring during a run to file *FILEPREFIXNAME.verbose*. Nevertheless, if you want a takly *rjungle* which puts all process information to screen then use option (-v)

```
$ cd tests
$ rjungle -f test.ped -p -v
```

```
Start: Thu Nov 27 12:19:26 2008
```

```
+-----+-----+-----+
| RandomJungle | ..... | 2008 |
+-----+-----+-----+
| (C) 2008 Daniel Frederik Schwarz et al., GNU GPL, v3 |
+-----+-----+-----+
```

```
Output to:
```

```
rjungle.*
loading data...
Read 9 row(s) and 10 column(s).
Use 9 row(s) and 6 column(s).
dependent variable name: PHENOTYPE
Growing jungle...
Number of variables: 6 mtry = 2
1 thread(s) growing 500 tree(s)
Growing time estimate: ~0 sec.
Generating and collecting output data...
Compiling trees.
Writing accuracy information...
calculating confusion matrix...
```

```
Elapsed time: 0 sec
```

```
Finished: Thu Nov 27 12:19:26 2008
```

```
$
```

6 Error codes

This chapter describes error codes.

6.1 Runtime error codes

Error code description can be found in `randomjungle-1.2.365/src/library/error_codes.txt` or http://www.randomjungle.com/tags/newest/error_codes.txt.

7 Some nice examples

This chapter describes various examples of working with `rjungle`. You can run the examples in your `randomjungle-1.2.365/tests` directory.

7.1 Simple example

We want to grow a simple forest. So, our input data contains 9 input (V1-V9) and 1 output variable(y). 8 output variables are non informative. Only variable V1 is informative.

```
$ rjungle -f test.dat -U1 -z1 -v -D y -o example1
$ ls example1.*
...
```

7.2 Simple example with ped file

Using Plink PED-files:

```
$ rjungle -f test.ped -p -o example2
...
```

7.3 Estimating variable importance

In the first example we used the data in file `test.dat`. As we know, only variable V1 is informative and therefore it is the most important variable. Firstly, lets estimate the fast GINI-Importance with 1000 trees.

```
$ rjungle -f test.dat -U1 -z1 -t1000 -v -D y -o example3.1
$ cat example3.1.importance
...
```

We see that variable V1 is the most important (highest value).

Secondly, we now want to estimate the permutation importance.

```
$ rjungle -f test.dat -U1 -z1 -t1000 -v -D y -i2 -o example3.2
$ cat example3.2.importance2
...
```

Now, we got even a more reliable result. In most cases, only variable V1 is positive. It means that only V1 contributes to a high accurate classification.

7.4 Using R and rjungle

We want to analyse data from R.

```
$ R
...
> mydata = iris
> mydata$Species = as.integer(mydata$Species) # convert factor to integer
> write.table(mydata, file = "data_from_r.dat",
+   row.names = FALSE, quote = FALSE)
> quit("no")
$ rjungle -f data_from_r.dat -U1 -z1 -t500 -v -D Species -o example4
$ cat example4.confusion
```

```
...
$ cat example4.importance
...
```

7.5 Using plink and rjungle

```
$ plink --file gwadata --recodeA
$ rjungle -f plink.raw -p -M2 -v -o example5
...
$ tail example5.importance
```

or

```
$ plink --file gwadata --recodeA
$ rjungleparse -f plink.raw -p -v -o example5
...
$ tail example5.importance
```

The `rjungleparse` uses less memory than `rjungle`. But your data should only use the values 0,1,2 and 3(missing code).

7.6 Prediction

```
$ rjungle -f train.dat -D DependVar -v -w2 -o example6a
...
$ rjungle -f test.dat -D DependVar -v -P example6a.jungle.xml -o example6b
$ cat example6b.confusion
$ cat example6b.prediction
...
```

7.7 Imputation

SNP data (raw/fast imputation):

```
$ rjungleparse -f mypedfile.raw -p -t1 -I1 -o example7_1_1
```

The imputation result was written to `example7_1_1.imputed.dat.gz` and can be used for analysis. E.g.:

```
$ rjungleparse -f example7_1_1.imputed.dat.gz -p ... -o example8_1_2
```

Imputing continuous data:

```
$ rjungle -f continuous.dat -A -D DependVar -I6 -o example7_2
```

or imputing categorical data:

```
$ rjungle -f cate.dat -D DependVar -I6 -o example7_3
```

or imputing data with no groups to classify on (unsupervised learning):

```
$ rjungle -f continuousAndNoGroups.dat -A -I5 -o example7_4
```

7.8 Using plink and rjungleparse

The `rjungleparse` is the same program like `rjungle`, but you can use just a small set of values: 0,1,2 (and 3 as missing coding). You might want to use `rjungleparse` in conjunction with `plink`. The memory consumption of `rjungleparse` is very small.

```
$ plink --file gwadata --recodeA
$ rjunglesparse -f plink.raw -p -v -o example8
...
$ tail example8.importance
```

7.9 Using MPI

For high speed parallel processing, RJ could be used on computer clusters using MPI mode (RJ has to be compiled for MPI!). The program was performed on huge data successfully using 150 processors (300 processes) in parallel on a high performance cluster. However, the MPI mode should be used by experienced users only. In MPI mode, performing permutation importance (option `-i2`, .., `-i5`) is allow exclusivly. Execute RJ as follows:

```
$ plink --file gwadata --recodeA
$ mpirun -np 200 --host hostname1, hostname2, ... rjunglesparse -f plink.raw -
p -v -i4 -o example9
...
$ tail example9.importance
```

Each process writes temp files to working directory (`example9_mpi_id_*.*`). Final results are written to usual files (`example9.*`).

Appendix A How to make copies of the overall Random Jungle package

This appendix covers the license for copying the source code of the overall Random Jungle package. This manual is under a different set of restrictions, covered later (see [Appendix B \[Copying This Manual\]](#), page 35).

A.1 License for copying the Random Jungle package

Version 3, 29 June 2007

Copyright © 2007 Free Software Foundation, Inc. <http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The GNU General Public License is a free, copyleft license for software and other kinds of works.

The licenses for most software and other practical works are designed to take away your freedom to share and change the works. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change all versions of a program—to make sure it remains free software for all its users. We, the Free Software Foundation, use the GNU General Public License for most of our software; it applies also to any other work released this way by its authors. You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for them if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs, and that you know you can do these things.

To protect your rights, we need to prevent others from denying you these rights or asking you to surrender the rights. Therefore, you have certain responsibilities if you distribute copies of the software, or if you modify it: responsibilities to respect the freedom of others.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must pass on to the recipients the same freedoms that you received. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

Developers that use the GNU GPL protect your rights with two steps: (1) assert copyright on the software, and (2) offer you this License giving you legal permission to copy, distribute and/or modify it.

For the developers' and authors' protection, the GPL clearly explains that there is no warranty for this free software. For both users' and authors' sake, the GPL requires that modified versions be marked as changed, so that their problems will not be attributed erroneously to authors of previous versions.

Some devices are designed to deny users access to install or run modified versions of the software inside them, although the manufacturer can do so. This is fundamentally incompatible with the aim of protecting users' freedom to change the software. The systematic

pattern of such abuse occurs in the area of products for individuals to use, which is precisely where it is most unacceptable. Therefore, we have designed this version of the GPL to prohibit the practice for those products. If such problems arise substantially in other domains, we stand ready to extend this provision to those domains in future versions of the GPL, as needed to protect the freedom of users.

Finally, every program is threatened constantly by software patents. States should not allow patents to restrict development and use of software on general-purpose computers, but in those that do, we wish to avoid the special danger that patents applied to a free program could make it effectively proprietary. To prevent this, the GPL assures that patents cannot be used to render the program non-free.

The precise terms and conditions for copying, distribution and modification follow.

TERMS AND CONDITIONS

0. Definitions.

“This License” refers to version 3 of the GNU General Public License.

“Copyright” also means copyright-like laws that apply to other kinds of works, such as semiconductor masks.

“The Program” refers to any copyrightable work licensed under this License. Each licensee is addressed as “you”. “Licensees” and “recipients” may be individuals or organizations.

To “modify” a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a “modified version” of the earlier work or a work “based on” the earlier work.

A “covered work” means either the unmodified Program or a work based on the Program.

To “propagate” a work means to do anything with it that, without permission, would make you directly or secondarily liable for infringement under applicable copyright law, except executing it on a computer or modifying a private copy. Propagation includes copying, distribution (with or without modification), making available to the public, and in some countries other activities as well.

To “convey” a work means any kind of propagation that enables other parties to make or receive copies. Mere interaction with a user through a computer network, with no transfer of a copy, is not conveying.

An interactive user interface displays “Appropriate Legal Notices” to the extent that it includes a convenient and prominently visible feature that (1) displays an appropriate copyright notice, and (2) tells the user that there is no warranty for the work (except to the extent that warranties are provided), that licensees may convey the work under this License, and how to view a copy of this License. If the interface presents a list of user commands or options, such as a menu, a prominent item in the list meets this criterion.

1. Source Code.

The “source code” for a work means the preferred form of the work for making modifications to it. “Object code” means any non-source form of a work.

A “Standard Interface” means an interface that either is an official standard defined by a recognized standards body, or, in the case of interfaces specified for a particular programming language, one that is widely used among developers working in that language.

The “System Libraries” of an executable work include anything, other than the work as a whole, that (a) is included in the normal form of packaging a Major Component, but which is not part of that Major Component, and (b) serves only to enable use of the work with that Major Component, or to implement a Standard Interface for which an implementation is available to the public in source code form. A “Major Component”, in this context, means a major essential component (kernel, window system, and so on) of the specific operating system (if any) on which the executable work runs, or a compiler used to produce the work, or an object code interpreter used to run it.

The “Corresponding Source” for a work in object code form means all the source code needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to control those activities. However, it does not include the work’s System Libraries, or general-purpose tools or generally available free programs which are used unmodified in performing those activities but which are not part of the work. For example, Corresponding Source includes interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work.

The Corresponding Source need not include anything that users can regenerate automatically from other parts of the Corresponding Source.

The Corresponding Source for a work in source code form is that same work.

2. Basic Permissions.

All rights granted under this License are granted for the term of copyright on the Program, and are irrevocable provided the stated conditions are met. This License explicitly affirms your unlimited permission to run the unmodified Program. The output from running a covered work is covered by this License only if the output, given its content, constitutes a covered work. This License acknowledges your rights of fair use or other equivalent, as provided by copyright law.

You may make, run and propagate covered works that you do not convey, without conditions so long as your license otherwise remains in force. You may convey covered works to others for the sole purpose of having them make modifications exclusively for you, or provide you with facilities for running those works, provided that you comply with the terms of this License in conveying all material for which you do not control copyright. Those thus making or running the covered works for you must do so exclusively on your behalf, under your direction and control, on terms that prohibit them from making any copies of your copyrighted material outside their relationship with you.

Conveying under any other circumstances is permitted solely under the conditions stated below. Sublicensing is not allowed; section 10 makes it unnecessary.

3. Protecting Users’ Legal Rights From Anti-Circumvention Law.

No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures.

When you convey a covered work, you waive any legal power to forbid circumvention of technological measures to the extent such circumvention is effected by exercising rights under this License with respect to the covered work, and you disclaim any intention to limit operation or modification of the work as a means of enforcing, against the work's users, your or third parties' legal rights to forbid circumvention of technological measures.

4. Conveying Verbatim Copies.

You may convey verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice; keep intact all notices stating that this License and any non-permissive terms added in accord with section 7 apply to the code; keep intact all notices of the absence of any warranty; and give all recipients a copy of this License along with the Program.

You may charge any price or no price for each copy that you convey, and you may offer support or warranty protection for a fee.

5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions:

- a. The work must carry prominent notices stating that you modified it, and giving a relevant date.
- b. The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to "keep intact all notices".
- c. You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it.
- d. If the work has interactive user interfaces, each must display Appropriate Legal Notices; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

A compilation of a covered work with other separate and independent works, which are not by their nature extensions of the covered work, and which are not combined with it such as to form a larger program, in or on a volume of a storage or distribution medium, is called an "aggregate" if the compilation and its resulting copyright are not used to limit the access or legal rights of the compilation's users beyond what the individual works permit. Inclusion of a covered work in an aggregate does not cause this License to apply to the other parts of the aggregate.

6. Conveying Non-Source Forms.

You may convey a covered work in object code form under the terms of sections 4 and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License, in one of these ways:

- a. Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by the Corresponding Source fixed on a durable physical medium customarily used for software interchange.
- b. Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by a written offer, valid for at least three years and valid for as long as you offer spare parts or customer support for that product model, to give anyone who possesses the object code either (1) a copy of the Corresponding Source for all the software in the product that is covered by this License, on a durable physical medium customarily used for software interchange, for a price no more than your reasonable cost of physically performing this conveying of source, or (2) access to copy the Corresponding Source from a network server at no charge.
- c. Convey individual copies of the object code with a copy of the written offer to provide the Corresponding Source. This alternative is allowed only occasionally and noncommercially, and only if you received the object code with such an offer, in accord with subsection 6b.
- d. Convey the object code by offering access from a designated place (gratis or for a charge), and offer equivalent access to the Corresponding Source in the same way through the same place at no further charge. You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source may be on a different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements.
- e. Convey the object code using peer-to-peer transmission, provided you inform other peers where the object code and Corresponding Source of the work are being offered to the general public at no charge under subsection 6d.

A separable portion of the object code, whose source code is excluded from the Corresponding Source as a System Library, need not be included in conveying the object code work.

A “User Product” is either (1) a “consumer product”, which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, “normally used” refers to a typical or common use of that class of product, regardless of the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a consumer product regardless of whether

the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.

“Installation Information” for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section must be accompanied by the Installation Information. But this requirement does not apply if neither you nor any third party retains the ability to install modified object code on the User Product (for example, the work has been installed in ROM).

The requirement to provide Installation Information does not include a requirement to continue to provide support service, warranty, or updates for a work that has been modified or installed by the recipient, or for the User Product in which it has been modified or installed. Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network.

Corresponding Source conveyed, and Installation Information provided, in accord with this section must be in a format that is publicly documented (and with an implementation available to the public in source code form), and must require no special password or key for unpacking, reading or copying.

7. Additional Terms.

“Additional permissions” are terms that supplement the terms of this License by making exceptions from one or more of its conditions. Additional permissions that are applicable to the entire Program shall be treated as though they were included in this License, to the extent that they are valid under applicable law. If additional permissions apply only to part of the Program, that part may be used separately under those permissions, but the entire Program remains governed by this License without regard to the additional permissions.

When you convey a copy of a covered work, you may at your option remove any additional permissions from that copy, or from any part of it. (Additional permissions may be written to require their own removal in certain cases when you modify the work.) You may place additional permissions on material, added by you to a covered work, for which you have or can give appropriate copyright permission.

Notwithstanding any other provision of this License, for material you add to a covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms:

- a. Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License; or

- b. Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it; or
- c. Prohibiting misrepresentation of the origin of that material, or requiring that modified versions of such material be marked in reasonable ways as different from the original version; or
- d. Limiting the use for publicity purposes of names of licensors or authors of the material; or
- e. Declining to grant rights under trademark law for use of some trade names, trademarks, or service marks; or
- f. Requiring indemnification of licensors and authors of that material by anyone who conveys the material (or modified versions of it) with contractual assumptions of liability to the recipient, for any liability that these contractual assumptions directly impose on those licensors and authors.

All other non-permissive additional terms are considered “further restrictions” within the meaning of section 10. If the Program as you received it, or any part of it, contains a notice stating that it is governed by this License along with a term that is a further restriction, you may remove that term. If a license document contains a further restriction but permits relicensing or conveying under this License, you may add to a covered work material governed by the terms of that license document, provided that the further restriction does not survive such relicensing or conveying.

If you add terms to a covered work in accord with this section, you must place, in the relevant source files, a statement of the additional terms that apply to those files, or a notice indicating where to find the applicable terms.

Additional terms, permissive or non-permissive, may be stated in the form of a separately written license, or stated as exceptions; the above requirements apply either way.

8. Termination.

You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License (including any patent licenses granted under the third paragraph of section 11).

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have

been terminated and not permanently reinstated, you do not qualify to receive new licenses for the same material under section 10.

9. Acceptance Not Required for Having Copies.

You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work. These actions infringe copyright if you do not accept this License. Therefore, by modifying or propagating a covered work, you indicate your acceptance of this License to do so.

10. Automatic Licensing of Downstream Recipients.

Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

An “entity transaction” is a transaction transferring control of an organization, or substantially all assets of one, or subdividing an organization, or merging organizations. If propagation of a covered work results from an entity transaction, each party to that transaction who receives a copy of the work also receives whatever licenses to the work the party’s predecessor in interest had or could give under the previous paragraph, plus a right to possession of the Corresponding Source of the work from the predecessor in interest, if the predecessor has it or can get it with reasonable efforts.

You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License. For example, you may not impose a license fee, royalty, or other charge for exercise of rights granted under this License, and you may not initiate litigation (including a cross-claim or counterclaim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for sale, or importing the Program or any portion of it.

11. Patents.

A “contributor” is a copyright holder who authorizes use under this License of the Program or a work on which the Program is based. The work thus licensed is called the contributor’s “contributor version”.

A contributor’s “essential patent claims” are all patent claims owned or controlled by the contributor, whether already acquired or hereafter acquired, that would be infringed by some manner, permitted by this License, of making, using, or selling its contributor version, but do not include claims that would be infringed only as a consequence of further modification of the contributor version. For purposes of this definition, “control” includes the right to grant patent sublicenses in a manner consistent with the requirements of this License.

Each contributor grants you a non-exclusive, worldwide, royalty-free patent license under the contributor’s essential patent claims, to make, use, sell, offer for sale, import and otherwise run, modify and propagate the contents of its contributor version.

In the following three paragraphs, a “patent license” is any express agreement or commitment, however denominated, not to enforce a patent (such as an express permission to practice a patent or covenant not to sue for patent infringement). To “grant” such

a patent license to a party means to make such an agreement or commitment not to enforce a patent against the party.

If you convey a covered work, knowingly relying on a patent license, and the Corresponding Source of the work is not available for anyone to copy, free of charge and under the terms of this License, through a publicly available network server or other readily accessible means, then you must either (1) cause the Corresponding Source to be so available, or (2) arrange to deprive yourself of the benefit of the patent license for this particular work, or (3) arrange, in a manner consistent with the requirements of this License, to extend the patent license to downstream recipients. “Knowingly relying” means you have actual knowledge that, but for the patent license, your conveying the covered work in a country, or your recipient’s use of the covered work in a country, would infringe one or more identifiable patents in that country that you have reason to believe are valid.

If, pursuant to or in connection with a single transaction or arrangement, you convey, or propagate by procuring conveyance of, a covered work, and grant a patent license to some of the parties receiving the covered work authorizing them to use, propagate, modify or convey a specific copy of the covered work, then the patent license you grant is automatically extended to all recipients of the covered work and works based on it.

A patent license is “discriminatory” if it does not include within the scope of its coverage, prohibits the exercise of, or is conditioned on the non-exercise of one or more of the rights that are specifically granted under this License. You may not convey a covered work if you are a party to an arrangement with a third party that is in the business of distributing software, under which you make payment to the third party based on the extent of your activity of conveying the work, and under which the third party grants, to any of the parties who would receive the covered work from you, a discriminatory patent license (a) in connection with copies of the covered work conveyed by you (or copies made from those copies), or (b) primarily for and in connection with specific products or compilations that contain the covered work, unless you entered into that arrangement, or that patent license was granted, prior to 28 March 2007.

Nothing in this License shall be construed as excluding or limiting any implied license or other defenses to infringement that may otherwise be available to you under applicable patent law.

12. No Surrender of Others’ Freedom.

If conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot convey a covered work so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not convey it at all. For example, if you agree to terms that obligate you to collect a royalty for further conveying from those to whom you convey the Program, the only way you could satisfy both those terms and this License would be to refrain entirely from conveying the Program.

13. Use with the GNU Affero General Public License.

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU Affero General Public License into a single combined work, and to convey the resulting work.

The terms of this License will continue to apply to the part which is the covered work, but the special requirements of the GNU Affero General Public License, section 13, concerning interaction through a network will apply to the combination as such.

14. Revised Versions of this License.

The Free Software Foundation may publish revised and/or new versions of the GNU General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies that a certain numbered version of the GNU General Public License “or any later version” applies to it, you have the option of following the terms and conditions either of that numbered version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of the GNU General Public License, you may choose any version ever published by the Free Software Foundation.

If the Program specifies that a proxy can decide which future versions of the GNU General Public License can be used, that proxy’s public statement of acceptance of a version permanently authorizes you to choose that version for the Program.

Later license versions may give you additional or different permissions. However, no additional obligations are imposed on any author or copyright holder as a result of your choosing to follow a later version.

15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

16. Limitation of Liability.

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

17. Interpretation of Sections 15 and 16.

If the disclaimer of warranty and limitation of liability provided above cannot be given local legal effect according to their terms, reviewing courts shall apply local law that

most closely approximates an absolute waiver of all civil liability in connection with the Program, unless a warranty or assumption of liability accompanies a copy of the Program in return for a fee.

END OF TERMS AND CONDITIONS

How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively state the exclusion of warranty; and each file should have at least the “copyright” line and a pointer to where the full notice is found.

```
one line to give the program's name and a brief idea of what it does.
Copyright (C) year name of author
```

```
This program is free software: you can redistribute it and/or modify
it under the terms of the GNU General Public License as published by
the Free Software Foundation, either version 3 of the License, or (at
your option) any later version.
```

```
This program is distributed in the hope that it will be useful, but
WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU
General Public License for more details.
```

```
You should have received a copy of the GNU General Public License
along with this program. If not, see http://www.gnu.org/licenses/.
```

Also add information on how to contact you by electronic and paper mail.

If the program does terminal interaction, make it output a short notice like this when it starts in an interactive mode:

```
program Copyright (C) year name of author
This program comes with ABSOLUTELY NO WARRANTY; for details type 'show w'.
This is free software, and you are welcome to redistribute it
under certain conditions; type 'show c' for details.
```

The hypothetical commands ‘show w’ and ‘show c’ should show the appropriate parts of the General Public License. Of course, your program’s commands might be different; for a GUI interface, you would use an “about box”.

You should also get your employer (if you work as a programmer) or school, if any, to sign a “copyright disclaimer” for the program, if necessary. For more information on this, and how to apply and follow the GNU GPL, see <http://www.gnu.org/licenses/>.

The GNU General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Lesser General Public License instead of this License. But first, please read <http://www.gnu.org/philosophy/why-not-lgpl.html>.

Appendix B How to make copies of this manual

This appendix covers the license for copying this manual. Note that some of the longer examples in this manual are also distributed in the directory ‘Random Jungle-1.2.365/examples/’, where a more permissive license is in effect when copying just the examples.

B.1 License for copying this manual

Version 1.2, November 2002

Copyright © 2000,2001,2002 Free Software Foundation, Inc.
51 Franklin St, Fifth Floor, Boston, MA 02110-1301, USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document *free* in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or non-commercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “Document”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as “you”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A “Modified Version” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a

textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none. The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties:

any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing

distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
 - I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version’s license notice. These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements.”

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted

document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of

this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (C)  year  your name.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.2
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover
Texts. A copy of the license is included in the section entitled ‘‘GNU
Free Documentation License’’.
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the “with...Texts.” line with this:

```
with the Invariant Sections being list their titles, with
the Front-Cover Texts being list, and with the Back-Cover Texts
being list.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

Appendix C Indices of concepts and macros

C.1 Index for all `rjungle` macros

This index covers all `rjungle` builtins, as well as several useful composite macros. References are exclusively to the places where a macro is introduced the first time.

=		C	
=	8, 9, 10	cells	13
		N	
		names	13

C.2 Index for many concepts

A	
application	19
B	
bug reports	3
C	
command line	7
command line, options	7
compiling and installing	5
confusion file	15
E	
error codes	17
examples, understanding	3
F	
FDL, GNU Free Documentation License	35
Free Documentation License (FDL), GNU	35
G	
General Public License (GPL), GNU	23
GNU Free Documentation License	35
GNU General Public License	23
GPL, GNU General Public License	23
H	
history of Random Jungle	3
I	
importance file	15
Imputation	20
input data	13
Input the whole data	13
invoking Random Jungle	7
L	
License, code	23
License, manual	35
log file	15
O	
options, command line	7
output data	15
overview of Random Jungle	3
P	
plink rjungle	20
POSIXLY_CORRECT	7
prediction	20
prediction file	15
R	
r rjungle	19
reporting bugs	3
restricted analysis	13
rjungle mpi MPI	21
Runtime error codes	17
S	
simple example	19
simple example with ped file	19
Standard files	15
suggestions, reporting	3
V	
variable importance	19
verbose file	16