An introduction to *bamchop*

Zhe Zhang

May 23, 2012

Contents

1 Installation

The easiest way for BiC memebers to run *bamchop* is to use the Rstudio installed on Riboso by Jemery. You can open this link http://10.30.4.241:8787 from a web browser and log in with your Riboso account. You can then skip the following installation steps.

bamchop is based upon **Sweave** (http://en.wikipedia.org/wiki/Sweave) framework, which is an intergration of **R** coding and **LaTeX** documenting. Therefore, *bamchop* requires the installation of:

LaTeX: Download from http://www.latex-project.org/ftp.html and install it. The size of the full Mac distribution is about 1.8 GB. You can also use smaller version such as MacTeXtras from http://mirror.ctan.org/systems/mac/mactex/MacTeXtras.zip.

R: *bamchop* runs within R, so you need to download the newest version of R from http://www.r-project.org and install it. The basic functions of R are extended by thousands of packages and *bamchop* requires the following packages to run:

- **R.oo**
- **gplots**
- **plotrix**
- **vcd**
- **xtable**

Bioconductor: Bioconductor is a project extending basic R with hundreds of bioinformatics-related packages. You can install the Bioconductor core packages within R. *bamchop* for to run, plus the following packages:

- **GenomicRanges**
- **Rsamtools**
- **seqLogo**
- **BSgenome**
- reference genome, such as **BSgenome.Hsapiens.UCSC.hg19** package if hg19 was used for alignment

1.1 Install R packages within R

To install an R package within R:

Install a specific R package, such as **vcd**:

```
> install.packages('vcd', repos="http://lib.stat.cmu.edu/R/CRAN/");
```

Install Bioconductor core packages:

```
> source("http://bioconductor.org/biocLite.R");
> biocLite();
```

Install a specific Bioconductor package, such as **Rsamtools**:

```
> source("http://bioconductor.org/biocLite.R");
> biocLite("Rsamtools");
```

Alternatively, you can download a package to local disk and install it using shell command **R CMD INSTALL**.

2 Prepare input files

Running *bamchop* requires two tab-delimited text files, respectively with information about:

BAM files: this file should contain two columns with no header.

- **Column 1:** unique names of BAM files. The names will be used as the folder names of output files.
- **Column 2:** full path to BAM files.

Requencing targets: this file should contain five columns with no header.

- **Column 1:** unique names of targeted regions.
- **Column 2:** chromosome names of target regions. Try to use names consistent with those used in reference genome. Otherwise, the program will try to map chromosome names with its best guess.
- **Column 3:** start position of target regions.
- **Column 4:** end position of target regions.
- **Column 5:** assign value 1 to all targets of interest (TOI) and 0 to all the other targets. *bamchop* will provide details about individual targets. However, reporting details about many targets will slow down the program and increase the size of ouput substantially. So, we suggest you to specify a subset of them as TOI when there are more than 144 total targets. The TOI can be selected randomly, by size, by importance, or any other criteria.

3 Set parameters

Open and edit the file **paste(program.name, '_param.r', sep="")** to set running parameters. There are three types of parameters:

Required parameters: These parameters should be specified for each data set.

- **bams** The full path and name of the tab-delimited text file with BAM file information (see section **Prepare input files**).
- **targets** The full path and name of the tab-delimited text file with target information (see section **Prepare input files**).
- **path.out** The fule path where all the output files will be stored.

Optional parameters: Change the value of these parameters when necessary

- **project.name**
- **prepared.by**
- **affiliation**
- **genome.name** Name of the reference genome.
- **logo** Name of the file with an image to be put in the footer with affiliation. Copy the file into the program folder if you want to use your own.

Advanced parameters: These are for experienced users only.

4 Run bamchop

We are finally ready to run *bamchop* now. Copy the program folder to a location you prefer and start R within the folder; then use

```
> getwd();

[1] "/home/zhangz/hts/scripts/bamchop/source/instruction"
```

to make sure you in the right working directory; and run *bamchop* with

```
> source('bam2pdf.r');
```

The running will take a while. Wait patiently.

5 Outputs

All output files will be saved to the folder specified by the **path.out** parameter. The minimal outputs are a set of pdf files, each with an individual figure, and a full report named **bamchop.pdf**, within each subfolder.