

An introduction to *bamchop*

Zhe Zhang

September 18, 2013

Contents

1	Installation	1
1.1	Install R packages within R	1
2	Inputs	2
3	Run bamchop	2
4	Outputs	2
5	More information	3

1 Installation

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. Bamchop has been tested in R version 2.13 and 2.15. Version 3.01 and above is currently not recommended due to package compliance. 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

Bamchop: Download bamchop from <https://github.com/CBmi-BiG/bamchop>, and copy the whole folder to the location where will be your local home of bamchop. Replace the logo.jpg file with the logo of your own institute/group. There are a couple of examples in /reports. This subfolder will also be the place to put the results of your own runs.

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 Inputs

Open the `bamchop.r` file to edit input variables. The file was split into several sections:

Mandatory variables. These are required variables for each bamchop run. A value must be assigned to each of them.

- **bamchop.path:** the full path to the bamchop home folder.
- **genome.name:** name of the reference genome used to generate the BAM files. Currently only support hg19, mm9, and ce10.
- **bamflies:** assign the full path of each BAM file to a unique ID. The ID will be used to name the output files.

Extra section: bamchop can be potentially expanded to describe application-specific information, such as ChIP-seq and targeted resequencing. This feature is not currently supported by bamchop1.01. So no inputs are required in this section.

Extra parameters: extra variables that further describe the data set.

- **project.name:** name of the data set or project.
- **prepared.by:** the person who made the run.
- **affiliation:** affiliation of the person.
- **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.
- The rest of the variables in this section are for advanced users. Changes are not recommended.

3 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] "/nas/is1/bamchop/bamchop.1.01/source/instruction"
```

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

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

The running will take a while. Wait patiently.

4 Outputs

The outputs of each BAM file will be saved in a subfolder of the `/results` folder, named by the ID of the BAM file. Each folder includes the following files:

Two R objects of processed data imported from the BAM file, one full version and one slim version. These objects can be used later to obtain extra information.

A set of PDF files, each with a plot describing one aspect of the BAM file

A summarized report with all the PDF files and their detailed explanation, named as `ID.pdf`.

5 More information

Publication: <http://www.biomedcentral.com/bmcbioinformatics/supplements/14/S11>

Source code: <https://github.com/CBMi-BiG/bamchop>

Discussion forum: <http://www.biostars.org/p/81285/>