Remaking the journal article

Toby Dylan Hocking

March 13, 2013

To redo the comparison of breakpoint detection models in the article, first put this code in your .Rprofile

```
> options(repos=c(
+           "http://www.bioconductor.org/packages/release/bioc",
+           "http://r-forge.r-project.org",
+           "http://cran.ism.ac.jp"))
> works_with_R <- function(Rvers,...){
+   pkg_ok_have <- function(pkg,ok,have){
+     stopifnot(is.character(ok))
+     if(!as.character(have) %in% ok){
+       warning("works with ",pkg," version ",
+               paste(ok,collapse=" or "),
+               ", have ",have)
+     }
+   }
+   pkg_ok_have("R",Rvers,getRversion())
+   pkg.vers <- list(...)
+   for(pkg in names(pkg.vers)){
+     if(!suppressWarnings(require(pkg, character.only=TRUE))){
+       install.packages(pkg)
+     }
+     pkg_ok_have(pkg, pkg.vers[[pkg]], packageVersion(pkg))
+     library(pkg, character.only=TRUE)
+   }
+ }
```

This is for installing required packages and checking if version numbers match. More info here
http://sugiyama-www.cs.titech.ac.jp/~toby/org/HOCKING-reproducible-research-with-R.html
Then install bams and its dependencies.

```
> install.packages("bams",dep=TRUE)
```

All of the article source files can then be found in the `bams/article` directory, which you can find by executing the following code in R:

```
> system.file("article",package="bams")
```

To redo the calculations in the article, simply type `make` in the `bams/article` directory, which should create `HOCKING-model-selection-breakpoint-annotations.pdf`.

One of the first steps of the analysis is to download the `segmentation.list.RData` file, which saves the breakpoint locations detected for all the profiles, algorithms, and parameters. The reason why these calculations are not re-done by default is that they take a LONG TIME. If you want to re-do these calculations,

`seg-commands.R` can help, optionally using a cluster that has qsub. The results are saved to the `~/seg` directory, which is parsed and converted to `segmentation.list.RData` by `segmentation.list.R`.

To speed up this process, you can redo the smoothing models for each profile in parallel, if you have access to a cluster with the command line program `qsub`. Try installing the `bams` package on the cluster, then executing the code in `bams/article/smoothing-commands.R`, which should use the cluster to make a `smooth` directory with result files. Then from the cluster, execute the code in `bams/article/make.all.stats.R` to remake zzz.stats.RData, copy this file to the `bams/article` directory, and type `make`.