Accessing Digital Expression Explorer from R

To facilitate convenient access to these data from within the R environment, a function called “getDEE” has been written that interfaces with the web server and loads the matrix into R. The function can be “sourced” as follows:

```
> source("http://dee.bakeridi.edu.au/getDEE.R")
```

Function syntax is as follows:

```
> getDEE(organism,SRRlist,outfile="NULL"))
```

The function will download the data as a zip archive that contains a count matrix as well as a folder of log files detailing the primary processing of the data including base quality scores, alignment rates, etc. If “outfile” is defined, then files will be downloaded to the current working directory. If it is not defined, then the files are downloaded to the temporary directory of R and deleted immediately after use.

The options for organism currently are

athaliana
celegans
dmelanogaster
drerio
ecoli
hsapiens
mmusculus
rnorvegicus
scerevisiae

The SRR numbers need to exactly match those in SRA.

Here is an example of usage for a small number of *E. coli* datasets that are loaded into a “mytable” and the zip archive is saved in the current working directory as "DEE_count_data.zip".

```
> mytable<-getDEE("Ecoli",c("SRR922260","SRR922261"),outfile="DEE_count_data.zip")

trying URL
'http://dee.bakeridi.edu.au/cgi-bin/extract.sh?org=ecoli&x=SRR922260&x=SRR922261'
Content type 'application/zip' length unknown
opened URL
.......... .......... .......... .........
downloaded 45 Kb

> head(mytable)
             GeneName SRR922260.v1 SRR922261.v1
ECDH10B_0001     thrL          470          452
ECDH10B_0002     thrA         6046         8138
ECDH10B_0003     thrB         1758         2918
ECDH10B_0004     thrC         3082         4188
ECDH10B_0005     yaaX           95          268
ECDH10B_0006     yaaA          169          734
```

Another example this time starting with a list of SRR accession numbers in a plain text file and not saving the zip archive. This approach works well with SRAdb integration.

```
> SRRlist<-read.table("SRRlist.txt")
> SRRlist
          V1
1 SRR1121431
2 SRR1121432
3 SRR1121433
4 SRR1233867
5 SRR1233868
> mytable2<-getDEE("mmusculus",SRRlist$V1)
trying URL
'http://dee.bakeridi.edu.au/cgi-bin/extract.sh?org=mmusculus&x=SRR1121431&x=SRR1121432&
x=SRR1121433&x=SRR1233867&x=SRR1233868'
Content type 'application/zip' length unknown
opened URL
.......... .......... .......... .......... ..........
.......... .......... .......... .......... ..........
.......... .......... .......... .......... ..........
.......... .......... .......... .......... ..........
.......... .......... .......... .......... ..........
.......... .......... .......... .......... ..........
.......... .......... .......... .......... ..........
.......... .......... .......... .......... ..........
.......... .......... .......... .......... ..........
.......... .......... .......... .......... ..........
.......... .......... ...
downloaded 523 Kb

> head(mytable2)
                   GeneName SRR1121431.v1 SRR1121432.v1 SRR1121433.v1
ENSMUSG00000000001    Gnai3          3316          3666          3813
ENSMUSG00000000003     Pbsn             0             1             0
ENSMUSG00000000028    Cdc45           167           191           199
ENSMUSG00000000031      H19            18             0             0
ENSMUSG00000000037    Scml2             2             1             4
ENSMUSG00000000049     Apoh             1             1             4
                   SRR1233867.v1 SRR1233868.v1
ENSMUSG00000000001           794           737
ENSMUSG00000000003             0             2
ENSMUSG00000000028           117           103
ENSMUSG00000000031             2             6
ENSMUSG00000000037            86           113
ENSMUSG00000000049             3            10
```

For convenience, we provide another function called “loadDEE” that extracts the count matrix from the .zip archive and loads into R as follows:

```
> mytable3<-loadDEE("DEE_count_data.zip")
> head(mytable3)
             GeneName SRR922260.v1 SRR922261.v1
ECDH10B_0001     thrL          470          452
ECDH10B_0002     thrA         6046         8138
ECDH10B_0003     thrB         1758         2918
ECDH10B_0004     thrC         3082         4188
ECDH10B_0005     yaaX           95          268
ECDH10B_0006     yaaA          169          734
```

If you need to aggregate data from two of more technical replicates, follow this guide.

```
> head(mytable)
             SRR922260v1 SRR922261v1
ECDH10B_0001         470         452
ECDH10B_0002        6046        8138
ECDH10B_0003        1758        2918
ECDH10B_0004        3082        4188
ECDH10B_0005          95         268
ECDH10B_0006         169         734
> mytable$SRR922260v1_SRR922261v1<-mytable$SRR922260v1 + mytable$SRR922261v1
> head(mytable)
             SRR922260v1 SRR922261v1 SRR922260v1_SRR922261v1
ECDH10B_0001         470         452                     922
ECDH10B_0002        6046        8138                   14184
ECDH10B_0003        1758        2918                    4676
ECDH10B_0004        3082        4188                    7270
ECDH10B_0005          95         268                     363
ECDH10B_0006         169         734                     903
```

getDEE source code

```
loadDEE<-function(zipname){
  CM="CountMatrix.xls"
  TF=tempfile()
  unzip(zipname, files = CM, exdir = tempdir() )
  mxname<-paste0(tempdir(),"/",CM)
  file.rename(mxname,TF)
  dat <- read.table(TF,row.names=1,header=T)
  unlink(TF)
  return(dat)
}


getDEE<-function(species, SRRvec, outfile=NULL ,
            baseURL="http://dee.bakeridi.edu.au/cgi-bin/extract.sh?"){
  llist<-paste0("&x=",paste(SRRvec,collapse = "&x="))
  murl <- paste0(baseURL,"org=",species, llist)
  if(is.null(outfile)){
            zipname=tempfile()
  } else {
            zipname=outfile
            if(!grepl(".zip$",zipname)){zipname=paste0(zipname,".zip")}
  }
  download.file(murl, destfile=zipname)

  dat<-loadDEE(zipname)
  if(is.null(outfile)){unlink(zipname)}
  return(dat)
}
#Simple test
#mytable<-getDEE("Ecoli",c("SRR922260","SRR922261"))
```