synDss Tutorial
Peter Chi and Vladimir Minin

Objective: This is a tutorial to illustrate the use of the synDss package in `R`.

- The first step is to download and install PAML, from: `http://abacus.gene.ucl.ac.uk/software/paml.html`.
  - The two programs we need from this software package are `codeml` and `evolver`.
    - For `codeml`, the included MS Windows executable `codeml.exe` will work as is on Windows, and on Linux/Mac, one can compile `codeml.c` following the PAML installation instructions.
    - For `evolver`, we need to compile it in a slightly different way in order to allow for variable dN/dS rates at each site (i.e. the M3 model). Essentially, we need to go to the PAML subdirectory that contains `evolver.c`, and type:

      `gcc -O4 -DCodonNSsites -o evolverNSsites evolver.c tools.c -lm`

      This is assuming that a gcc compiler is installed on your machine, e.g. through RTools if you are using Windows. See

      `petrov.stanford.edu/software/src/paml3.15/Technical/Simulation/Codon/CodonSimulation.txt`

      for more details.
  - The `codeml` and `evolverNSsites` executables should then be placed in a folder of local programs pointed to by your PATH variable, so that they can be run from any directory. See the PAML installation instructions for more details. Alternatively, if this cannot be successfully performed, one will need to specify the directory where the PAML programs are located when running certain functions (see below).

- Next, we need to open `R`, and install and load the `synDss` package:

```
> #install.packages("synDss", repos="http:/R-Forge.R-project.org")
> library(synDss)
```

- Now, let us try to calculate the original and synonymous Dss statistics from a DNA sequence alignment. First, we will read in the fasta file for a sequence alignment of 18 salmonella sequences included in the package:

```
> salmonella<-fasta.to.alignment(system.file("seqdata/Salmonella_fimH_aligned-18alleles.txt",package="
>
```

- Now, let us calculate the Dss statistic for this dataset, with a window size of 105 nucleotides, and step size of 9 nucleotides:

```
> dss.stat<-calc.Dss(salmonella, l=270, m=9)
>
```

- We can also calculate the synonymous Dss statistic (for a dataset of this size, this will take about 5 minutes):

```
> syn.stat<-calc.Dss.syn(salmonella, l=270, m=9,syn.matrix=regist.synonym())
>
```

- To look at our statistic landscapes, we can plot them:

```
> m<-9
> l<-270
> length<-dim(as.matrix.alignment(salmonella))[2]
> n.windows<-((length-(2*l))/m+1)
> index<-c(1:n.windows)
> title<-"sal18.pdf"
> pdf(title,height=4,width=8)
> max<-max(c(dss.stat,syn.stat),na.rm=T)
> plot(dss.stat~index,type="l",xlab="window number",ylab="",ylim=c(0,max),lty=2)
> lines(syn.stat~index,type="l")
> legend("topright",c("Original Dss","Synonymous Dss"),lty=c(2,1))
> dev.off()
```