The `micropan` package vignette

Lars Snipen and Kristian Hovde Liland

1 Using `dplyr` and `stringr`

A major change in the 2.0 version is the use of generic data structures and functions in R instead of creating package specific ones. This makes it possible to use the power of standard data manipulation tools and visualization that R-users are familiar with.

Compared to previous versions some functions have been moved to the `microseq` package.

You will also find no casestudy document or plotting functions. However, if you locate the GitHub site for this package, you will find a tutorial with code making similar plots using `ggplot` or `ggdendro`. This is an example of using generic R tools instead of making functions for each special case.

1.1 Faster reading of BLASt results

A major change in the 2.1 version is faster reading of the BLAST result files, see '?bDist' or the tutorial at GitHub mentioned above for more details.

2 External software

Some functions in this package calls upons external software that must be available on the system. Some of these are 'installed' by simply downloading a binary executable that you put somewhere proper on your computer. To make such programs visible to R, you typically need to update your `PATH` environment variable, to specify where these executables are located. Try it out, and use google for help!

2.1 Software `blast+`

The function *blastpAllAll* uses the free software `blast+` (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/L Source code and installers makes it straightforward to install. In the R console the command

```
> system("blastp -h")
```

should produce some sensible output.

2.2 Software `hmmer`

The functions *hmmerScan()* uses the free software `hmmer` (http://hmmer.org/). This software is developed for UNIX systems (e.g. Mac or Linux), and Windows users may find it a little difficult to install and run from R. In the R console the command

```
> system("hmmscan -h")
```

should produce some sensible output.