rvd_gen User Manual

Background Information

rvd_gen is a command line tool that can be used to generate a designer TAL effector RVD sequence from a given nucleotide sequence. Because rvd_gen is written in C, it should run on any modern UNIX-like operating system, and has been tested on Elementary OS (Linux kernel 4.4.0-66-generic) and FreeBSD (11.0-RELEASE-p1 GENERIC). rvd_gen is Copyright (c) 2017, Stephen P. Cohen and is distributed under a permissive 3-clause license (see below for license terms).

Installation Instructions

rvd_gen can be cloned directly from github for free using the git command line tool (for installation of git, see the developer website at https://git-scm.com/downloads). To install rvd_gen, use git to clone the distribution, then use make to compile, like so (user input in bold):

```
~$ git clone https://github.com/stephen-cohen/rvd_gen
Cloning into 'rvd_gen'...
remote: Counting objects: 13, done.
remote: Compressing objects: 100% (9/9), done.
remote: Total 13 (delta 5), reused 11 (delta 3), pack-reused 0
Unpacking objects: 100% (13/13), done.
Checking connectivity... done.
~$ cd rvd_gen
~/rvd_gen$ make
cc -O3 -Wall -o rvd_gen rvd_gen.c
```

The binary file is rvd_gen, which can be run directly or copied to a bin directory for easy access (e.g. ~/bin for a single user).

Usage Instructions

To use rvd_gen, you must supply a nucleotide sequence. The sequence may be supplied as a command line argument or as input (user input in bold):

```
~/rvd_gen$ ./rvd_gen ACGTACGTACGTACGT
NI-HD-NK-NG-NI-HD-NK-NG-NI-HD-NK-NG-NI-HD-NK-NG
~/rvd_gen$ ./rvd_gen
Enter nucleotide sequence: ACGTACGTACGTACGT
NI-HD-NK-NG-NI-HD-NK-NG-NI-HD-NK-NG-NI-HD-NK-NG
```

By default, the maximum sequence length allowed is 41. This limit can be raised by altering the source code (rvd_gen.c, line 41), but most available dTALE construction kits do not allow target sequences to be larger than ~30. There is no minimum input for the program.

About the Program

This program generates an optimal TAL effector RVD sequence given a nucleotide sequence. It uses the following RVDs for DNA-binding, which are available in modules from the dTALE construction kit made available by Cermak et al. (2011)[1]:

Nucleotide	RVD	Strength[2]
A	NI	weak
C	HD	strong
G	NK	weak
T	NG	weak

In the case of sequences that would lead to stretches of 5 or more weak RVDs, the program uses NN for binding to A or G to break up the weak stretches, based on recommendations from Streubel et al. (2012)[2]. The rvd_gen logic differs slightly from these recommendations, which recommend that stretches of 6 or more consecutive weak RVDs should be avoided. This limit can be altered in the source code (rvd_gen.c, line 117), and future versions may allow for command line arguments for user options. Input sequences with stretches of 5 or more Ts can not be strengthened and should be avoided as EBEs.

A user may force usage of the RVD NN by using the nucleotide symbol R (user input in bold):

```
~/rvd_gen$ ./rvd_gen AAAAA
NI-NI-NI-NI-NN
~/rvd_gen$ ./rvd_gen ARAAA
NI-NN-NI-NI-NI
```

References

1. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF. Nucleic Acids Res. 2011; 39(12):e82.
2. Streubel J, Blücher C, Landgraf A, Boch J. Nat Biotechnol. 2012;30:593-595.

Software License