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 with little to no modification. It has been tested and verified as functional on GNU/Linux, macOS, FreeBSD, and Windows (via Cygwin). For the rvd_gen software license and terms, see the end of this document.

Installation Instructions

GNU/Linux, macOS, or similar command line operating systems

rvd_gen can be installed by cloning from github 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, as below:

```
~$ git clone https://github.com/stephen-cohen/rvd_gen
Cloning into 'rvd_gen'...
Checking connectivity... done.
~$ cd rvd_gen
~/rvd_gen$ make
cc -Os -Wall -o rvd_gen rvd_gen.c
```

The output binary file is **rvd_gen**, which can be run directly or copied to a directory in $PATH (e.g. ~/bin for a single user, or /usr/local/bin in Linux for multiple users) for easy access.

Windows instructions

rvd_gen can be installed in Windows using GNU tools in Cygwin (downloadable at https://www.cygwin.com/). When installing Cygwin, install the development tools **make**, **gcc-core**, and **git**. Using the **git** and **make** commands as described above will produce **rvd_gen.exe**.

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, as below:

```
~/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
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, because most available dTALE construction kits do not allow target sequences to be larger than ~30. Source code (rvd_gen.c, line 41) may be altered to raise or lower this limit. There is no minimum input for the program. Input may be upper- and/or lowercase.

About the Program

This program generates an optimal TAL effector RVD sequence given a nucleotide sequence. Table 1 shows the nucleotide-RVD pairs used by rvd_gen, which are available in modules from the dTALE construction kit made available by Cermak et al. (2011)[1].

Table 1. Nucleotide-RVD pairs used by rvd_gen and the relative strength of each corresponding interaction.

Nucleotide(s)	RVD	Strength[2]
A	NI	weak
C	HD	strong
G	NH / NK	weak
T	NG	weak
A / G	NN	strong

In the case of sequences that would lead to stretches of 6 or more weak RVDs, rvd_gen substitutes NN for binding to A or G to break up the weak stretches, based on recommendations from Streubel et al. (2012)[2]. Input sequences with stretches of 6 or more Ts cannot be strengthened and should be avoided as EBEs. rvd_gen outputs NK for G specificity, but NH may be substituted.

A user may force usage of the RVD NN in a specific location by using the nucleotide symbol R, as below:

```
~/rvd_gen$ ./rvd_gen AAAAAA
NI NI NI NI NI NN
~/rvd_gen$ ./rvd_gen ARAAAA
NI NN NI NI NI NI
```

Uninterpretable input returns “??” in place of valid RVDs.

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.

Citing rvd_gen

To cite rvd_gen in publications, use:

> Cohen SP. 2018. rvd_gen: a command line tool for generating designer TAL effector RVD sequences. Available from: https://github.com/stephen-cohen/rvd_gen.

Software License