

## Using tacg restriction analysis program with BioEdit.

Author: Stan Vitha (vitha@mic.tamu.edu).

### Background:

While the free BioEdit software has a built-in restriction analysis feature, some people prefer the output format of the [tacg program](#) (see [Appendix](#)). tacg is a unix program for pattern search in DNA sequences. Its most popular application is for restriction analysis. It is a very fast command line program, thus is very suitable for execution from other programs. In order to run this software on Windows machines, it is necessary to install the basic CygWin release ([www.cygwin.com](http://www.cygwin.com)) that will create a unix-like environment on your computer.

Once the CygWin is installed, download and install the tacg executable that was compiled for CygWin (<http://members.tripod.com/mhoenicka/tacg.html>). Please note that you must use the unix commands in cygwin shell to unzip and untar the tacg distribution package (file named “**cygwin1.1.4-tacg-3.04.tar.gz**”):

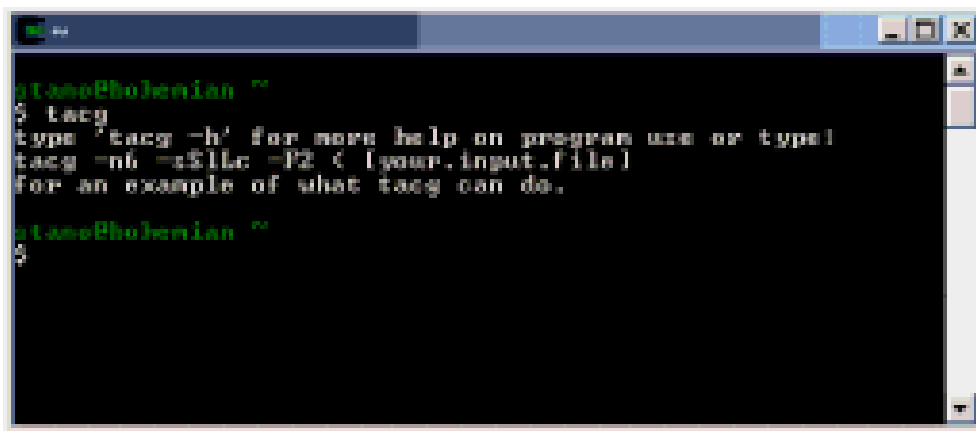
```
$ gunzip cygwin*.gz
$ tar -xvf cygwin*.tar
```

Follow the installation instructions in the “**readme-cygwin.txt**”:

- copy tacg.exe into /usr/local/bin.
- Copy the tacg.1 man file in the /Docs subdirectory to your man1 directory
- Copy all files in the /Docs and /Data subdirectories into one directory with read access for all users, e.g. /usr/local/lib/tacg,
- Create a system-wide environment variable TACGLIB which points to this directory- edit the file /etc/profile and add the following lines:

```
# set the TACG Library environmet variable
export TACGLIB=/usr/local/lib/tacg
```

Also make sure the PATH environment variable includes the location where tacg is installed (/usr/local/bin). After the installation is complete, you should be able to start cygwin shell and run tacg without any parameters. It should return a short message (see below).



```
stano@hulsterian ~
$ tacg
type 'tacg -h' for more help on program use or type!
tacg -h -s$1lc -f2 < [your.input.file]
for an example of what tacg can do.
stano@hulsterian ~
$
```

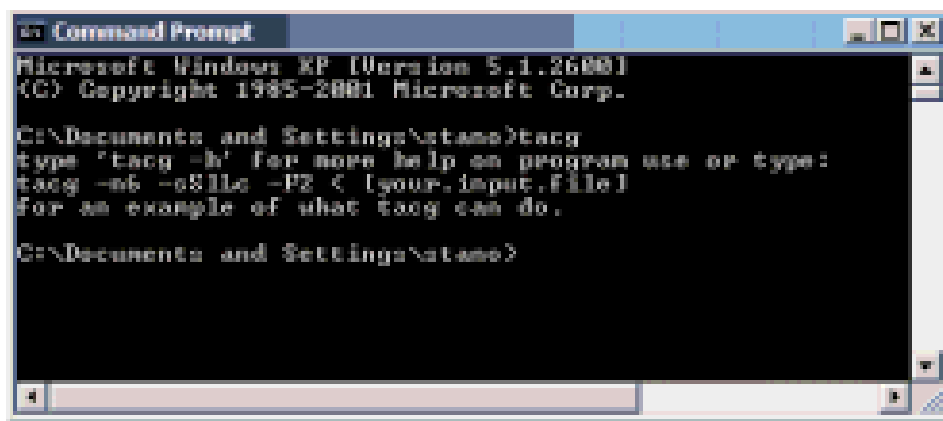
Now that tacg is installed and functional, the next step is to invoke the tacg program as if it was a DOS command, i.e. without starting the cygwin shell window. For this to work, the Windows PATH environment must include the path to the tacg executable and to its library files. In the Windows Control Panel – System, choose the “Advanced” , “Environment Variables”.

- Edit the system variable PATH and add the following to the end of the existing path (beware that the individual paths must be separated by a semicolon):

```
;C:\cygwin\bin;C:\cygwin\usr\local\bin
```

- Create a new system variable TACGLIB, its value is `c:\cygwin\usr\local\lib\tacg`

Now open a Command Prompt window and run tacg without any parameters. You should get the same response as when it was run from the cygwin shell:



```
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\stame>tacg
type 'tacg -h' for more help on program use or type:
tacg -n6 -c811c -P2 < [your.input.file]
for an example of what tacg can do.

C:\Documents and Settings\stame>
```

Finally, In order to create the accessory application menu item for tacg in Bioedit, add the following lines to the end of “accApp.ini” file located in the “Bioedit\apps” folder:

```
[tacg]
Program=C:\cygwin\usr\local\bin\tacg.exe
Don't use interface=0
Don't create menu item=0
Auto-Feed=1
Degap Sequences=1
Separate Files=0
Auto-Feed File Format=1
Auto-Feed File Name Required=0
Remove redundancies=0
Truncate titles=0
Change spaces in titles=0
Specific File Name=
Prompt for Input File=0
Prompt for Output File=0
Open Output as Alignment=0
Open Output as Text=1
Open Output with External Program=0
External Program Name=C:\WINDOWS\system32\write.exe
Open Output with Accessory Application=0
```

```

Accessory Application to open with=
Input File Prefix=
Output File Prefix=
Specify Input File Name=0
Specify Output File Name=0
Input File Name=input.fas
Output File Name=out.txt
Input File Prefix Required=0
Space after input prefix=0
Output File Prefix Required=0
Space after output prefix=0
Input File Name Arbitrary=0
Output File Name Arbitrary=0
Redirect input from stdin=0
Redirect output from stdout=0
Default Command Line=-s < input.fas
View Documentation Option=0
Documentation File=
Description=
Include Additional Options Box=0
Redirect General stdout=1
Stdout Redirected Filename=out.txt
Redirect General stdin=0
Stdin Redirected Filename=
Add input file to command line=0
Add output file to command line=0
output command at beginning=0
input command at beginning=0
output command at end=0
input command at end=0
c0 Title=6-cutters only
c0YES="-n6"
c0NO="-n4"
c0 Default=1
c1 Title=Circular DNA
c1YES="-f0"
c1NO="-f1"
c1 Default=0
c2 Title=Table of Fragment Sizes
c2YES="-F 1"
c2NO=""
c2 Default=0
c3 Title=Ladder Map
c3YES="-l"
c3NO=""
c3 Default=0
c4 Title=Table of Cut Sites
c4YES="-S "
c4NO=""
c4 Default=1
c5 Title=dam methylated?
c5YES="--dam"
c5NO=""
c5 Default=0
c6 Title=dcm methylated?
c6YES="--dcm"
c6NO=""

```

```

c6 Default=0
c7 Title=Double Stranded Output?
c7YES="--strands 2"
c7NO="--strands 1"
c7 Default=1
c8 Title=Tickmarks
c8YES=""
c8NO="--notics"
c8 Default=1
i0 Title=ORF Analysis (frames,minimum size)
i0 Prefix="-O "
i0 Default=123,25
i0 CheckBox=1
i0 CheckBox Title=ORF Analysis
i0 CheckBox Default=0
i1 Title=5=5', 3=3', 0=blunt, 1=all
i1 Prefix="-o "
i1 Default=1
i1 CheckBox=1
i1 CheckBox Title=Overhang of Ends
i1 CheckBox Default=1
i2 Title=Output Width (n x 15 characters)
i2 Prefix="-w "
i2 Default=75
i2 CheckBox=0
i2 CheckBox Title=0
i2 CheckBox Default=0
i3 Title=Frames 1,3, or 6, Code 1 or 3-letter
i3 Prefix="-T "
i3 Default=3,1
i3 CheckBox=1
i3 CheckBox Title=Linear Cotranslation
i3 CheckBox Default=1
i4 Title=1 = 1st base of sequence
i4 Prefix="-b "
i4 Default=1
i4 CheckBox=1
i4 CheckBox Title=Beginning of Subsequence
i4 CheckBox Default=1
i5 Title=0 = last base of sequence
i5 Prefix="-e "
i5 Default=0
i5 CheckBox=1
i5 CheckBox Title=End of Subsequence
i5 CheckBox Default=1
  
```

Note that more options can be added, using the "Add/Edit Accessory Application" feature in Bioedit. Now start BioEdit, load some sequences, highlight one and run the "tag" accessory application. Enjoy!

**Appendix: An example of the tacg program output.**

## Sequence: #1; Format: FASTA; ID: (null); Description: At5g42480\_Wt-  
Ws\_cDNA\_pSV399 2438 bases.

== Sequence info:

140 bases; 33 A(23.57 %) 50 C(35.71 %) 28 G(20.00 %) 29 T(20.71 %)

== Total Number of Hits per Enzyme:

BbsI	1	BtrI	1	PflMI	1
BsaI	1	EciI	1	SgrAI	1

== Match Sites by Enzyme

BbsI	GAAGACnn'nnnn_ (0 Err) - 1 Match(s) found (0.06 sites exp)	129
BsaI	GGTCTCn'nnnn_ (0 Err) - 1 Match(s) found (0.07 sites exp)	36
BtrI	CAC'GTC (0 Err) - 1 Match(s) found (0.08 sites exp)	20
EciI	GCGGANnnnnnnnn'_nn' (0 Err) - 1 Match(s) found (0.11 sites exp)	104
PflMI	CCAn_nnn'nTGG (0 Err) - 1 Match(s) found (0.07 sites exp)	124
SgrAI	Cr'CCGG_yG (0 Err) - 1 Match(s) found (0.03 sites exp)	62

== Open Reading Frame Analysis:

== ORF Analysis for Frame 2: 1 ORF(s) > 25 AAs

F#	ORF#	Begin(bp/AAs)	End(bp/AAs)	#AAs	MWt(KDa)	pI
> 2	1	0 / 0	139 / 46	46	5053.15	10.089

MEALSHVIGIGLSPFQLCRLPPATTKLRRSHNTSTTICSASKWADRL

== Summary of Enzymes that hit \*\* 1 \*\* times or less:

										PflMI@124			
BtrI@20				BsaI@36		SgrAI@62				EciI@104		BbsI@129	
-----													
:	:	:	:	:	:	:	:	:	:	:	:	:	:
10	20	30	40	50	60	70	80	90	100	110	120	130	140

== Linear Map of Sequence:

														PflMI@124	
BtrI				BsaI		SgrAI				EciI<		BbsI<			
-----															
:	:	:	:	:	:	:	:	:	:	:	:	:	:		
10	20	30	40	50	60	70	80	90	100	110	120	130	140		
1	catggaagctctgagtcacgctcggcattggtctctccccattccaattatgccgattaccaccggcgcgacaaa	75													
	gtaccttcgagactcagtgccagccgtaaccagagaggggtaagggttaatacggctaatggtggccgctgctgttt														
	^ * ^ * ^ * ^ * ^ * ^ * ^ * ^ *														
1	H G S S E S R R H W S L P I P I M P I T T G D D K														
2	M E A L S H V G I G L S P F Q L C R L P P A T T K														
3	W K L * V T S A L V S P H S N Y A D Y H R R R Q S														
76	gctccgacgtagccacaacacctctacaactatctgctccgccagcaaatgggcccgaccgtcttc	150													
	cgaggctgcacgctggtttgtggagatggtgatagacgagggcggtcgtttaccggctggcagaag														
	^ * ^ * ^ * ^ * ^ * ^ * ^ * ^ *														
1	A P T * P Q H L Y N Y L L R Q Q M G R P S														
2	L R R S H N T S T T I C S A S K W A D R L														
3	S D V A T T P L Q L S A P P A N G P T V F														