

November 21, 2005

## Composite likelihood analysis of incomplete selective sweeps

Yuseob Kim  
Arizona State University  
yuseob.kim@asu.edu

This manual gives instruction to compile and run CLics.c. This program has numerous similarities with clsw, the program analyzing complete selective sweeps. I therefore urge you to read CLmanual.pdf (“Programs for simulation and composite likelihood analysis of selective sweep”, included in KimCLA.tar) first.

CLics.c and other necessary files are packaged in ICS.tar. After extracting files in a new directory (`>> tar -xf ICS.tar`), compile the program by typing ‘make’. Then, the program is compiled into ‘clics’.

clics runs like clsw, taking inputs in ‘ms’ format. If the input has just one segment without coding regions, it can be run without a parameter file. If not, run with parameter file (clparm.inp in this case):

```
>> ./clics -p clparm.inp < ssw_first.out
```

clparm.inp and ssw\_first.out were given as examples in the package. Then you will get the following result in the output file, clics.out:

Composite Likelihood Analysis of Incomplete Sweeps 11/22/05

parameters from clparm.inp

nSam= 10

seqL=10000, LB=1, RB=10000, Rn=1.000000e-01

theta=-1.000000e+00

Ntry=5

epsilon = 0.50/alpha

1	47	theta = 0.003938	LR1=	4.3775	alf=	400.91	x=	3752	LR1B=	0.2222	alf=
435.76	x=	3785	fB=	0.909637							
2	36	theta = 0.003016	LR1=	4.0743	alf=	871.03	x=	4497	LR1B=	0.0000	alf=
871.03	x=	4497	fB=	1.000000							
3	37	theta = 0.003100	LR1=	2.6614	alf=	184.67	x=	4276	LR1B=	0.0163	alf=
173.83	x=	4292	fB=	1.000000							
4	52	theta = 0.004357	LR1=	2.6073	alf=	189.15	x=	7137	LR1B=	0.8396	alf=
371.98	x=	6754	fB=	0.940361							
5	42	theta = 0.003519	LR1=	1.3162	alf=	234.58	x=	3017	LR1B=	0.0007	alf=

232.08 x= 3029 fB= 0.999712

Alpha= 4.169366e+02 (2.457034e+02), X= 4.472054e+03 (1.248612e+03), fB= 0.969942  
(0.037960)

In the example above, five data sets (with numbers of segregating sites 47, 36, 37, 52 and 42) were analyzed. LR1 and LR1B correspond to LR\_1 and LR\_2 of Meiklejohn et al. (2004). LR1B is thus the likelihood ratio of incomplete sweep vs. complete sweep. Note that this program assumes that you distinguish ancestral (0) versus derived (1) alleles in the input. fB is the estimated frequency of the putative beneficial mutation at the time of sampling. On the bottom line, means and s.d.s of Alpha, X and fB under incomplete sweep model over replicates were given. See Meiklejohn et al. (2004) for more information.

Reference: Meiklejohn, C. D., Y. Kim, D. L. Hartl, and J. Parsch 2004 *Genetics* 168: 265-279.