

## GPTOSU

GPTOSU is a computer program that converts a "classical" genotype data file of the GENEPOP format (Raymond & Rousset 1995) into an infile that can be read by the TEMPOFS program for estimation of genetic drift and effective size from allele frequency data.

GPTOSU is written in TurboPascal 7.0 and compiled for Windows/Dos under the Win32 1.0.8 release of Free Pascal Compiler version 2.0.4 [2006/08/21] for i386 (<http://www.freepascal.org>). The Windows/Dos version (*gptosu.exe*) has been tested under Windows 98 and XP. The Linux version (*gptosu*) was compiled using the Free Pascal Compiler version 2.0.4 [2006/09/20] for i386.

The GPTOSU acronym implies conversion from GENEPOP to Stockholm University (SU) format. The SU format was adopted originally from the Center for Demographic and Population Genetics, Houston, Texas, USA, and has been used at the Division of Population Genetics, Stockholm University, for several programs analyzing allele frequency data. GPTOSU can be downloaded from <http://www.zoologi.su.se/~ryman/>.

### Installing GPTOSU

The downloaded zip file contains the executable program (*gptosu.exe* for Windows/Dos and *gptosu* for Linux), an example data file (*gptest.in*), the corresponding result file (*gptest.out*), and the present brief instruction (*gptosu.pdf*). Copy these files to a separate directory.

### Data files

Input data is read from a file formatted for GENEPOP. Locus designations must be given on separate lines, and cannot contain blank (space) characters. No blank (empty) lines are allowed in the file. Alleles are designated with two digits (01-99), and "00" is used for missing value.

GPTOSU deals with generations rather populations (as in GENEPOP). Thus, in the input file generations are treated as GENEPOP "populations", and the generation number of each individual must be given first on the data line and be followed by a blank (space) character. Any text (up to 200 characters) is allowed after this blank before the comma (","). *Gptest.in* provides an example of a small valid data file (Table 1).

The present version of GPTOSU is compiled for a maximum of 50 generations ("populations"), 100 loci, and 99 alleles per locus.

### Running GPTOSU

The program prompts for the name of the data file. It is convenient, but not necessary, to have the program and the data file located in the same directory. The output (result) is written to a file named *gptosu.out*. Rename this file to save it for future use!

Running the test data set (*gptest.in*; Table 1) should result in an output identical to that of *gptest.out* (Table 2) when the program operates correctly.

*Temporary files*: Intermediate results are stored in temporary files named *locus1*, *locus2*, etc. that are deleted automatically during execution. If the program halts prematurely it may be necessary to erase these files manually. These temporary files may also be visible for a short time when running directly from Windows.

Nils Ryman  
Division of Population Genetics  
Department of Zoology  
Stockholm University,  
S-106 91 Stockholm, Sweden

Email: Nils.Ryman@popgen.su.se

## Reference

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**:248-249.

Table 1. The *gptest.in* example input file comprising data on three loci and two "populations" = generations (generations 0 and 5).

Reduced set of Atlantic herring data from Lena C. Larsson

GPI

MDH4

CPA111

POP

0	MASEB	,	0101	0103	1010
0	MASEB	,	0106	0103	0511
0	MASEB	,	0106	0101	0911
0	MASEB	,	0104	0103	0910
0	MASEB	,	0101	0103	0909
0	MASEB	,	0105	0103	0509
0	MASEB	,	0101	0101	0911
0	MASEB	,	0101	0303	0509
0	MASEB	,	0105	0103	0709
0	MASEB	,	0105	0101	0709
0	MASEB	,	0104	0303	0509
0	MASEB	,	0101	0103	0909
0	MASEB	,	0105	0101	0909
0	MASEB	,	0101	0101	0910
0	MASEB	,	0101	0103	0909
0	MASEB	,	0104	0303	0910
0	MASEB	,	0105	0103	0509
0	MASEB	,	0101	0103	0509
0	MASEB	,	0101	0303	0909
0	MASEB	,	0101	0303	0909
0	MASEB	,	0101	0303	0707
0	MASEB	,	0101	0103	0910
0	MASEB	,	0101	0303	0709
0	MASEB	,	0101	0103	0909

Pop

5	MASEEF	,	0101	0103	0509
5	MASEEF	,	0606	0303	0909
5	MASEEF	,	0105	0101	0910
5	MASEEF	,	0506	0103	0910
5	MASEEF	,	0101	0101	0909
5	MASEEF	,	0101	0103	0910
5	MASEEF	,	0101	0101	0909
5	MASEEF	,	0101	0101	0709
5	MASEEF	,	0103	0101	0910
5	MASEEF	,	0101	0103	0909
5	MASEEF	,	0101	0101	0910
5	MASEEF	,	0101	0303	0509
5	MASEEF	,	0104	0103	0709
5	MASEEF	,	0105	0103	0910
5	MASEEF	,	0101	0103	0909
5	MASEEF	,	0106	0103	0910
5	MASEEF	,	0104	0101	1010
5	MASEEF	,	0101	0103	0910
5	MASEEF	,	0406	0103	0909
5	MASEEF	,	0406	0103	0910
5	MASEEF	,	0104	0103	0909
5	MASEEF	,	0406	0101	0909
5	MASEEF	,	0101	0103	0909

Table 2. The *gptest.out* file representing the output from GPTOSU when processing the *gptest.in* data file (Table 1). The *gptest.out* file can be used as an input to the TEMPOFS program. The figures " 2 3 " in the first line refer to the number of generations ("populations") and loci, respectively, whereas the figures on the second line indicate the number of alleles at each of the three loci.

2 3 Reduced set of Atlantic herring data from Lena C. Larsson

5 2 5

0.7917 0.0000 0.0625 0.1042 0.0417 48 GPI 0

0.6304 0.0217 0.1304 0.0652 0.1522 46 GPI 5

0.4583 0.5417 48 MDH4 0

0.6304 0.3696 46 MDH4 5

0.1250 0.1042 0.5833 0.1250 0.0625 48 CPA111 0

0.0435 0.0435 0.6739 0.2391 0.0000 46 CPA111 5