LZ5: a Flexible Data Simulation Program for Animal Breeding Area

R. da Fonseca[*], V. Verniz[*], A.T.H. Utsunomiya[§], A. Queiróz[¥], R.K. Ono[§],
M.P. Pires[§], M.M. Farah[£] and A.V. Pires[£]

Introduction

Simulation is an important tool for animal breeders. It has been used to compare metodologies, test some assumptions and create scenarios to ilustrate the use of new methodologies. Another appealing aspect of simulation is its capacity of diminishing costs by yielding results that otherwise would be attained only by expensive experimental procedures. For instance, in genomics area, SNPs have been discovered and its use as a molecular marker in animal breeding procedures has being heavily exploited, however, genotyping individuals is still expensive and simulation can be useful in this case.

The scenarios needed for research vary largely, thus, simulators are build as small pieces of program which intends to accomplish one or few specific tasks. As a result, data simulator softwares have a short life and it is not suitable to be shared by others groups but the one who develop it. Additionaly, because of these characteristics, they are not documented and do not have an user interface.

General Information

LZ5 intends to be a data simulation software which focus on colaborative efforts to develop procedures that will allow build a flexible and longer-life program to research and teaching.

The functions are developed in order that they can be used together with other functions to build the desired scenario (similar to R software). In situations that the tools to accomplish a task is not yet implemented, the interested user can code the algorithms and couple them into the simulation software. This is possible because LZ5 is a Free-Software which is licensed under Gnu General Public License (GPL v.3 or later).

The development of the program is organized in three modules:

1. Simulation
2. Selection
3. Genetic Evaluation

---

[*] UNESP/Dracena, Rodovia Comandante João Ribeiro de Barros, km 651, 17900-000 Dracena, Brazil
[*] UNESP/Dracena, Rodovia Comandante João Ribeiro de Barros, km 651, 17900-000 Dracena, Brazil
[§] UNESP/Jaboticabal, via de Acesso Prof.Paulo Donato Castellane s/n 14884-900, Jaboticabal, Brazil
[¥] FUNDEC, rua Bahia, 17900-000, Dracena, Brazil
[§] UNESP/Jaboticabal, via de Acesso Prof.Paulo Donato Castellane s/n 14884-900, Jaboticabal, Brazil
[£] UFVJM, Rodovia MGT 367 – Km 583, nº 5000, 39100-000, Diamantina, Brazil
[£] UFVJM, Rodovia MGT 367 – Km 583, nº 5000, 39100-000, Diamantina, Brazil

Within each module are developed the procedures needed respectively to simulate populations, to perform selection and to perform genetic evaluation.

The modules work integrated in the data simulation software allowing the user to design several scenarios. In the future, it is intended that Selection and Genetic Evaluation modules also work independently, being used not only to manipulate simulated data but also manipulate real data.

Some tasks that LZ5 is able to perform is:

1. Simulate molecular data (markers)
2. Yield outputs suitable for the most populars genetic analyses softwares
3. Read real data and simulate additional data considering the covariance structure provided.

In order to simulate data for a population which evolves for g generations by means of random mating and random selection, the basic steps need to be performed:

1. Create a base population
2. Select m males and f females which will be the parents of the initial population
3. Mating the selected individuals
4. Create o new individuals from each mate forming the initial population
5. Repeat steps 2-4 g times to create the new generations

To accomplish this in LZ5, the following commands should be activated in program's prompt:

1. createBasePop(n)
   Create a base population with n individuals, $n/2$ males and $n/2$ females.
2. defineCandidates(´n´)
   Define that candidates will be chosen from non-overlapping generations.
3. selection(m,f,sc)
   Select m males, f females according to the selection criterion sc.
4. createGenRM( )
   Create a new generation by random mating the selected individuals and generating o offspring per mate.

LZ5 is a gene level data simulator program being developed for Linux plataform using C++ language. Additionaly, GSL library is used in its development.

The binary and the corresponding sources can be freely downlonded from (www.unesp.dracena.br/luccaz/lz5).

The one interested in collaboration should make contact by the following e-mail: ricardo@dracena.unesp.br

References

R project: www.r-project.org/

Free Software Fundation: http://www.fsf.org/licensing/licenses/quick-guide-gplv3.html

GNU Scientific Library: http://www.gnu.org/software/gsl/