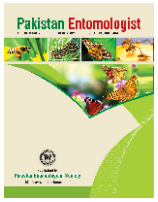
Pakistan Entomologist

Journal homepage: www.pakentomol.com



BOOTSIE – ESTIMATION OF COEFFICIENT OF VARIATION OF AFLP DATA BY BOOTSTRAP ANALYSIS

Justin Payne[1], Erica Lindroth[2], Kate Kneeland[1], Steven R. Skoda[3]*, Fatima Mustafa[4], Muhammad Irfan Ullah[5] and John E. Foster[1]

[1]Department of Entomology, Insect Genetics Laboratory, University of Nebraska, Lincoln, NE 68583-0816
[2]Walter Reed Army Institute of Research, 503 Robert Grant Ave, Silver Spring, MD 20910
[3]USDA-ARS-KBUSLIRL Screwworm Research Unit, 2700 Fredricksburg Road, Kerrville, TX 78028
[4]Department of Entomology, University of Agriculture Faisalabad, Pakistan
[5]University College of Sargodha, University of Sargodha, Sargodha, Pakistan

ARTICLE INFORMATION

Received: September 22, 2014
Received in revised form: November 15, 2014
Accepted: December 12, 2014

***Corresponding Author**:
Steven R. Skoda

E-mail: steve.skoda@ars.usda.gov

ABSTRACT

Bootsie is an English-native replacement for ASG Coelho's "DBOOT" utility for estimating coefficient of variation of a population of AFLP marker data using bootstrapping. Bootsie improves on DBOOT by supporting batch processing, time-to-completion estimation, built-in graphs, and a suite of export tools for creating data files for other population genetics software. Bootsie is released as open-source software under the Apache 2.0 license and is available for any Java SE 6 platform at http://code.google.com/p/bootsie/downloads/list/ .

Keywords: Population genetics; genetic markers; AFLP; utility; software

INTRODUCTION

Amplified Fragment Length Polymorphism has emerged as a well-developed technique for generating data about the genetics of populations (Vos *et al*., 1995) but few standards for the format of these data have emerged. Many applications in use accept marker data only in formats specific to the program, which has necessitated time-consuming editing of data files if a researcher's work chain involves several of these programs. Here we present Bootsie, a tool developed by our laboratory to perform functions poorly represented in other available applications. Primarily developed to replace the DBOOT (Coelho, 2001) program for estimating the coefficient of variation of marker data, the functions were expanded to include export of marker data into formats for Arlequin (Excoffier *et al*., 005), Popgene (Yeh *et al*., 1997), Numerical Taxonomy and Multivariate Analysis System (NTSys) (Rohlf, 2000), and Phylogenetic Analysis Using Parsimony (PAUP) (Swofford, 2003).

The utility "DBOOT" has found use among agronomists studying the genetic variation of crop and pest species, both in the Portugese-speaking community and beyond. But for all its simplicity the software has certain flaws. It appears to perform the entire analysis on the UI event dispatch thread, such that UI function and OS operation are disrupted during analysis. Users learn not to attempt to interact with the computer in any way while DBOOT is running; if Windows attempts to take window focus from DBOOT, the program will usually hang. It offers relatively few options for configuration and operates on only one file at a time; there is no way to set up multiple populations for analysis. In addition, DBOOT offers relatively little indication of progress during the computation and no estimate of total time to completion.

Bootsie is an open-source Java application that attempts to address these and other concerns. It is multi-threaded and takes advantage of multi-core processors. Multiple populations can be loaded into Bootsie for batch analysis,

Cite this article as: Payne, J., E. Lindroth, K. Kneeland, S.R. Skoda, F. Mustafa, M.I. Ullah and J.E. Foster, 2014. Bootsie – estimation of coefficient of variation of AFLP data by bootstrap analysis. Pak. Entomol., 36(2):79-81.

each with their own analysis parameters. Currently by default Bootsie runs as many as two concurrent analysis threads. Unlike DBOOT, Bootsie estimates the total time for the analysis and reports the total time actually taken. A UI progress bar indicates the relative completion of each analysis. Additionally, Bootsie supports the graphical display of coefficient of variation data in PNG format.

Bootsie can serve as the first step in an AFLP analysis tool chain because it can export genetic marker data into a variety of text formats. Currently Bootsie supports the creation of data files for NTsys, Popgene, and Arlequin, plus a generic tab-delimited format and distance matrix export. Additionally, Bootsie has English documentation within the application. The efficiency of Bootsie has been evaluated for number of insect markers obtained from AFLP before running through the various softwares as mentioned above; for example, *Melanoplus bowditchii* (Orthoptera: Acrididae) (Ullah *et al*., 2014) and Spined soldier bugs (Hemiptera: Pentatomidae) (Mustafa *et al*. in press FE).

MATERIALS AND METHODS

Implementation: Bootsie is hosted by Google Code at http://code.google.com/p/bootsie where it can be downloaded both as source files and as a distributable ZIP archive. Test data files can be found at the same location for verification purposes and as examples of appropriate Bootsie input format. Bootsie is written in Java 6 and can run under Windows, Macintosh, and Linux operating systems provided the Java environment is present.

Bootsie estimates the sampling variance of AFLP marker data via a bootstrap procedure, where markers are resampled with replacement and compared (Tiyang *et al*., 1994). Bootsie compares populations using the simple measure of genetic similarity thought to be most appropriate for dominant markers (Sokal and Michener, 1958) but future versions will allow users to select from other measures such as Jaccard's distance (Jaccard, 1908) or Dice-Nei (Dice, 1945; Nei and Li, 1979). Direct code comparison to Coelho's DBOOT program was not possible, but a comparison of output between DBOOT and Bootsie using the same marker data was performed.

Currently Bootsie has five modules that allow for data to be exported for input to Popgene, NTSys, Arlequin, applications that take a tab-delimited file, and a genetic distance matrix for use in phylogenetics software such as PAUP. Programmers familiar with Java can easily add new output modules as they see fit.

Bootsie supports a rudimentary graphing function that produces a chart of mean coefficient of variation values versus number of markers resampled. These charts are exported to the program's default results directory in Portable Network Graphics (PNG) format. Numeric output is in the form of a tab-delimited table of values; these values can be used with software such as Excel (Microsoft, Inc.) or SigmaPlot (Systat Software, Inc.) to produce more elaborate graphical displays of data.

RESULTS AND DISCUSSION

Output was directly compared between Bootsie and DBOOT, using the same AFLP marker data, at a setting of 1000 boostraps. Data were provided by Kate Kneeland from samples of stable flies (*Stomoxys calcitrans* L.) (Kneeland 2011).The test data set was comprised of 122 samples with 191 marker loci per sample.

Output between DBOOT and Bootsie are consonant to a high degree; results varied by 1.6% at one sampled locus, where the effect of random resampling is the strongest, to as low as 0.07% at 191 sampled loci, the number of loci in the test data set. We believe this establishes the validity of Bootsie as a tool for assessing the sampling variance of genetic distance and the number of markers required for a given level of precision.

Future versions of the software will allow the user to determine the number of simultaneous threads, and will include SVG support for scalable, vector-based graphs.

ACKNOWLEDGMENTS

This work was done in cooperation with the Institute of Agriculture and Natural Resources, University of Nebraska, Lincoln, NE. Mention of a proprietary product does not constitute endorsement or recommendation for its use by the USDA. USDA is an equal opportunity provider and employer.

REFERENCES

Coelho, A.S.G., 2001. DBOOT - Avaliação dos erros associados a estimativas de distâncias/similaridades genéticas através do procedimento de bootstrap com número variável de marcadores, v. 1.1. Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiâ-nia, GO.

Dice, LR., 1945. Measures of the amount of ecologic association between species. Eco., 26:297-302.

Excoffier, L., G. Laval and S. Schneider, 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47-50.

Jaccard, P., 1908. Nouvelles recherches sur la distribution florale. Bull. Soc. Vaud. Sci Nat., 44: 223-270.

Kneeland, K., 2011. Genetic variability of the stable fly *Stomoxys calcitrans* (L.) (Diptera: Muscidae) assessed on a global scale using AFLP. PhD dissertation, University of Nebraska, Lincoln.

Mustafa, F., M.I. Ullah K.M. Kneeland, T.A. Coudron, D.W. Stanely, W.W. Hoback, S.R. Skoda, J. Molina-Ocha and J.E. Foster, 2015. Genetic variability of sined soldier bugs (Hemiptera: Pentatomidae) sampled from distinct field sites and laboratory colonies in The United States. Florida Entomol., (In Press).

Nei, M. and W. Li, 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases.

Proc. Natl. Acad. Sci., 76(10): 5269-5273.

Rohlf, F.J., 2000. NTSYSpc. Numerical Taxonomy and Multivariate Analysis System. Version 2.1. Exeter Software, Applied Biostatistics Inc. Port Jefferson, NY.

Sokal R.R. and C.D. Michener, 1958. A Statistical Method for Evaluating Systematic Relationships. The University of Kansas Scientific Bulletin, 38: 1409-1438.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Tivang, J.G., J. Nienhuis and O.S. Smith, 1994. Estimation of sampling variance of molecular marker data using the bootstrap procedure. Theor. Appl. Genet., 89: 259-264.

Ullah, M.I., F. Mustafa, K.M. Kneeland, M.L. Brust, W.W. Hoback, S.T. Kamble and J.E. Foster, 2014. Forms of *Melanoplus bowditchi* (Orthoptera: Acrididae) collected from different host plants are indistinguishable genetically and in aedeagal morphology. Peer J., 2: 418.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van der Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, M. Zabeau, 1995. AFLP: A new technique for DNA finger printing. Nucl. Acids Res., 21: 4407-4441.

Yeh, F.C., R.C. Yang, T.B.J. Boyle, Z.H. Ye and J.X. Mao, 1997. POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada.