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

Justin Payne
University of Nebraska-Lincoln

Erica Lindroth
University of Nebraska-Lincoln, erica.lindroth@huskers.unl.edu

K. M. Kneeland
University of Nebraska-Lincoln, herplvr@aol.com

Steven R. Skoda
USDA-ARS, Steve.Skoda@ars.usda.gov

Fatima Mustafà
University of Agriculture Faisalabad

Follow this and additional works at: http://digitalcommons.unl.edu/entomologyfacpub
Part of the Entomology Commons

Payne, Justin; Lindroth, Erica; Kneeland, K. M.; Skoda, Steven R.; Mustafà, Fatima; Ullah, Muhammad Irfan; and Foster, John E., "BOOTSIE – ESTIMATION OF COEFFICIENT OF VARIATION OF AFLP DATA BY BOOTSTRAP ANALYSIS" (2014). Faculty Publications: Department of Entomology. 601.
http://digitalcommons.unl.edu/entomologyfacpub/601

This Article is brought to you for free and open access by the Entomology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications: Department of Entomology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
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), Poggene (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 Portuguese-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,
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, Pogene, 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).

RESULTS AND DISCUSSION

Output was directly compared between Bootsie and DBOOT, using the same AFLP marker data, at a setting of 1000 bootstraps. 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, Goiânia, GO.


Payne et al. / Pakistan Entomologist 2014, 36(2):79-81


