Application Note

JADE: a distributed Java application for deleterious genomic mutation (DGM) estimation

J.-L. Li 1, M.-X. Li2,3, Y.-F. Guo2, H.-Y. Deng4, and H.-W. Deng2,4,5 *

1. Seattle Biomedical Research Institute, Seattle, WA 98109, USA
2. Laboratory of Molecular and Statistical Genetics, College of Life Sciences, Hunan Normal University, Changsha, P. R. China 410081
3. Shanghai Center for Bioinformation Technology, Shanghai 200235, P.R. China
4. Departments of Orthopedic Surgery and Basic Medical Sciences, School of Medicine, University of Missouri-Kansas City, Kansas City, MO 64108, USA
5. Osteoporosis Research Center, Creighton University Medical Center, Omaha, NE 68131, USA

ABSTRACT

Summary: The characterization of deleterious genomic mutation (DGM) is of central significance for evolutionary biology and genetics studies. Fitness moment method has been developed to efficiently characterize DGM from natural population directly. In order to enable researchers to employ this method for theoretical and empirical research on characterizing DGM, we here present a distributed Java Application for DGM Estimation (JADE).

Availability: http://orclinux.creighton.edu/DGM/index.htm

Contacts: Jinlong.Li@sbri.org

1 INTRODUCTION

Deleterious mutations occur in the genomes of all species and their impact on evolution and human society is profound (Peck and Eyre-Walker 1997). The estimation of the mutation parameters, such as the rate ($\lambda$) of their occurrence and their average deleterious effects, is necessary for studies of evolutionary biology and genetics (For references, see Deng et al. 1999). Several complex methods (e.g., Morton et al. 1956; Bateman 1959; Deng and Lynch 1996; Keightley 1998; Fernandez et al., 2005) have been developed to infer the deleterious genomic mutation (DGM) rate and other parameters.

Among these methods, the fitness moment approach displays the best statistical properties, as reflected by the minimum mean square error that is a composite index of both bias and sampling error for biased estimates (Deng and Fu 1998). Its robustness when applied to natural populations has been investigated extensively through computer simulations under a range of biologically plausible conditions (e.g., Li et. al. 1999).

The results from these simulation studies (e.g., Li et. al. 1999) alleviate concerns in applying the fitness moments approach for efficiently characterizing DGM from natural populations, and also provide a basis for correct inferences about deleterious mutations from natural populations. However, the analytical methods are complex and the implementation of this approach can be hampered by lack of easy access to user-friendly programs. In order to enable interested researchers to adopt this method for their research and to plan optimal/efficient experiments for characterizing DGM, we have developed a distributed Java application for DGM estimation (JADE) to implement the complex analytical methods developed earlier.

2 IMPLEMENTATION

JADE is a three-tier architecture distributed application in which the three-tiers are defined as the client, the database server and the application server. Java Remote Method Invocation (RMI) technology (http://java.sun.com/products/jdk/rmi/) has been employed to implement the three tier architecture in JADE.

Fig. 1. (A) Friendly graphic user interface of JADE; (B) Functionality of JADE.

© The Author (2006). Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
Java RMI client: The RMI client program of JADE has friendly graphical user interfaces (GUI) (Figure 1A). A user can submit multiple simulations to the application server at one time and retrieve the results later. After a simulation is submitted, the simulation runs persistently in the server.

Database server: All simulation results are stored in a central MySQL relational database. The results are accessible to all users. If a user submits a simulation that has been already computed in the RMI server, the result will be retrieved from the database directly without running additional simulations. In this way, it avoids running unnecessary repeat simulations by different users or the same user.

The application server: The application server is the core component of JADE. All simulation programs written in C/C++ are resident in the application server, and this is where all simulations performed. In general, the server is much faster and more powerful than a regular PC. Compared to running DGM simulations on users’ desktop devices of which a majority of them are regular PCs, the simulation time is reduced substantially and the efficiency of the programs is improved.

3 FUNCTIONALITY

Figure 1B shows the functionalities of JADE. JADE can be used to estimate DGM parameters under different biologically plausible situations such as by accounting for overdominance effects, variable mutation effects, etc. The user interfaces and procedures to run the simulations in different situations are similar. However, the input parameters and output results may be different. For a detailed description of all functions in JADE, interested readers may refer to the online user manual (http://orclinux.creighton.edu/DGM/Manual.pdf).

4 DISCUSSION

Most scientific simulation programs have been developed by scientists who typically have a greater interest in algorithm development and implementation than in user interface and program efficiency. These programs are effective when used by developers or used within the developer’s group. However, these programs are less useful to the wider scientific community in which the majority of potential users are not familiar with computing tools and the programming environment. For the wider communities, unfriendly user interface, long computational time, and the requirement for additional computer equipment including servers discourage the adoption of these programs for research.

In this note, we describe a distributed Java application for DGM estimation called JADE. JADE incorporates a database management system to store all simulation results, which avoids unnecessary repeat simulation that can consume considerable amounts of time and computing power. In addition, a distributed system has been developed for JADE that allows all simulations to operate on a Linux server which has been installed, configured and managed by the developers. The benefits are three-fold. First, the running time for DGM estimation is reduced substantially. Second, programming crashes (unexpected terminations) due to insufficient computational power or memory are prevented. Third, the end users, who are primarily scientists with an interest in DGM estimation but not in computing environments, do not require new or specialized computer equipments in their laboratories in order to run DGM estimation programs.

ACKNOWLEDGEMENTS

The authors thank Dr. Patrick Duffy for proofreading and revising the manuscript and Mr. Evan Brown for revising the User Manual of JADE. This study was supported in part by NIH and CNSF.