IMGui—A Desktop GUI Application for Isolation with Migration Analyses

Jared Knoblauch,[1] Arun Sethuraman,*[1,2] and Jody Hey[1]

[1]Department of Biology, Center for Computational Genetics and Genomics, Temple University, Philadelphia, PA

[2]Department of Biological Sciences, California State University San Marcos, San Marcos, CA

*Corresponding author: E-mail: asethuraman@csusm.edu.

Associate editor: Keith Crandall

Abstract

The Isolation with Migration (IM) programs (e.g., IMa2) have been utilized extensively by evolutionary biologists for model-based inference of demographic parameters including effective population sizes, migration rates, and divergence times. Here, we describe a graphical user interface for the latest IM program. IMGui provides a comprehensive set of tools for performing demographic analyses, tracking progress of runs, and visualizing results. Developed using *node. js* and the *Electron* framework, IMGui is an application that runs on any desktop operating system, and is available for download at https://github.com/jaredgk/IMgui-electron-packages.

Key words: graphical user interface, isolation with migration, desktop application.

Introduction

IMa2 (Hey 2010b) and its parallel implementation IMa2p (Sethuraman and Hey 2016) are the most recent versions of a series of programs for the population genetics inference of divergence times, ancestral and current effective population sizes, and migration rates under the Isolation with Migration (IM) model (Nielsen and Wakeley 2001; Hey and Nielsen 2004, 2007). These tools belong to a class of methods called genealogy samplers (Kuhner 2009), which use a Bayesian Markov Chain Monte Carlo (MCMC) framework for the estimation of posterior density distributions of evolutionary parameters, given haploytpic, or genotypic data. A typical IM run is divided into an MCMC phase (also called "M" mode) and an analysis phase (called "L" mode) wherein quantities of interest are generated using the samples obtained in the MCMC phase. The analyses typically comprise parameter estimation, but can also include likelihood ratio tests of nested demographic models (Hey and Nielsen 2004). To help ensure better mixing of the Markov chain simulation, the programs implement Metropolis coupled MCMC (MC3); and to improve runtimes, a parallelized version of the program has recently been made available (Sethuraman and Hey 2016).

The IM programs implement a command-line interface (CLI), which needs user-specification of numerous required and optional values that determine the details of screen and file output, the MCMC sampling process, the model being analyzed, and the analyses to be carried out. For a given analysis, the programs often require long runs, or multiple independent runs, as well as periodic checks of run performance to ensure sufficient MCMC mixing. For larger problems that run slowly, users generally mine periodic output files to track trajectories of sampled divergence times, likelihoods, and marginal posterior densities. Unfortunately, all these steps come with a learning curve for new users that are not entirely familiar with command lines, shell scripting, or methods for assessing the mixing properties of Markov chain simulations.

In an effort to improve usability, ease of mining and visualizing output, and tracking runs, we developed IMGui, a user-friendly desktop-based front-end implementation of IMa2p that provides users with several interactive functions. Developed with the *Electron* v. 1.3.1 framework for node.js, IMGui can be used across operating systems and platforms, allows interactive validation and monitoring of runs, produces trace-plots and histograms of run parameters at regular intervals, and assists with mining run outputs to create publication quality graphs and visualizations. The IMGui software bundle is freely accessible in packages for a variety of operating systems through https://github.com/jaredgk/IMgui-electron-packages

IMGui

IMGui is a *Node.js* v0.10.25 (https://nodejs.org/en/foundation/) based application developed using the *Electron* v. 1.3.1 (http://electron.atom.io/) framework which allows the development of a cross-platform desktop application using Javascript/HTML/CSS. IMGui bundles a fully functional front-end GUI for IMa2/IMa2p, includes basic output analyses and visualization tools, and a GUI for IMfig, a plotting script which produces IMa2/IMa2p output summaries. The included IMa2 executable generates XML-formatted output files that are parsed to generate tables and histograms of output data in the visualization component of the app. The IMGui app is composed of three pages:

FIG. 1. Screen shot of the IMa2p command line module of IMGui, showing user input fields for prior values on estimable parameters, and run-time parameters.
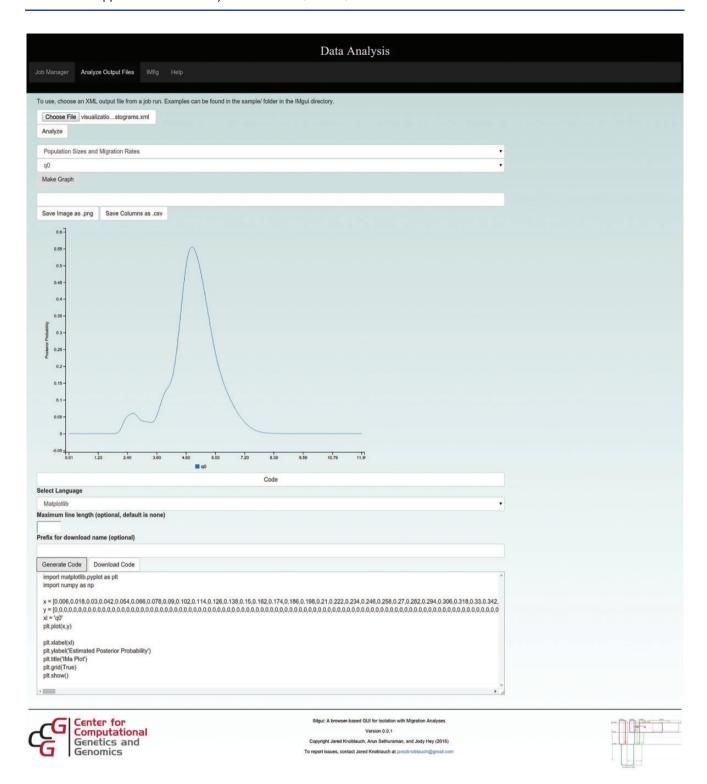
Fig. 2. Screen shot of the Data Analysis module of IMGui displaying a histogram of estimated parameters. Also shown is the *MATLAB* and *matplotlib* extension for generation of code for user-defined plotting.

(1) Job Manager—This page has two components: a command line generator and a job status control panel. The command line generator allows users to select a run mode ("M" or "L") and to input job arguments and input files. The job manager will verify that all required options are present and no incompatible options have been selected. An option can be selected to verify that all provided input file paths are present on the user's computer (but will not verify their contents). If errors are present, they will be displayed and highlighted for the user to fix (see fig. 1). If a job passes all checks, it is started and logged in the app's job list. The control panel allows the user to view output from the job and restart/delete active jobs. The user can also control when to end burn-in, and begin sampling, or to end sampling using options on the control panel.
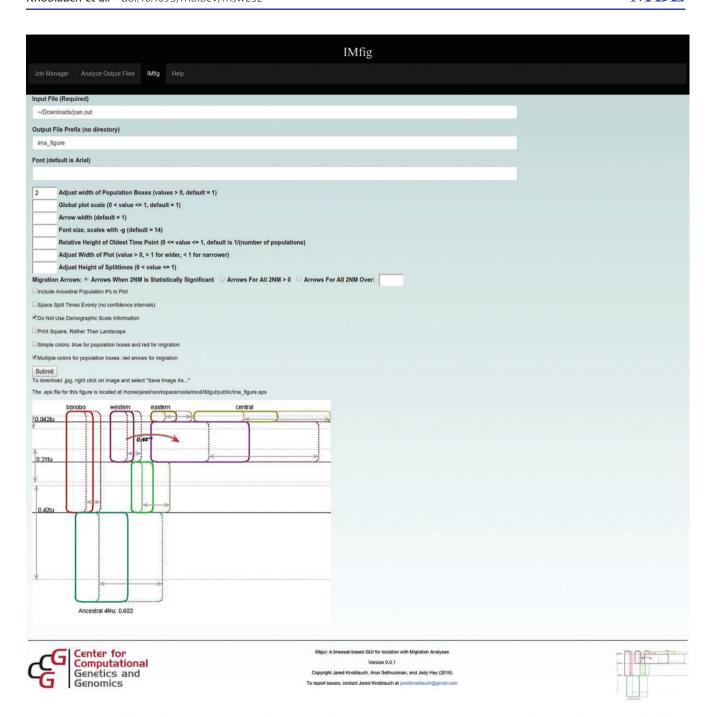
3

FIG. 3. Screen shot of the IMfig generator module of IMGui, visualizing IM parameters estimated using the four-population IM model of Hey (2010a), comprising the Western common chimpanzee (*Pan troglodytes verus*), Central common chimpanzee (*P. troglodytes troglodytes*), Eastern common chimpanzee (*P. troglodytes schweinfurthii*), and bonobos (*Pan paniscus*). Species are coded by color, with the width of solid boxes indicative of their highest posterior density estimate of mutation scaled population sizes. 95% confidence intervals are shown by dotted boxes and solid arrows indicate significant migration events in evolutionary time.

All output generated during the run are saved in the XML format for further analyses.

(2) Output Analyzer—Parses XML output and displays it in the app window. There are three XML output files: main output, intervals, and histograms. The interval file can be loaded to view run-time tables for use with determining when to end user-controlled burn-in or run modes. After a run ends, the main XML file can be loaded to view final output tables (which summarize the estimated parameter distributions). Tables can also be exported to .png or .csv formats. In addition to tables, the histogram file contains visualizations of marginal posterior density distributions for population demographic parameters (e.g., population sizes, divergence times, migration rates). Histograms in IMGui are generated with *C3.js* (http://c3js.org/), an extension of the visualization library *D3.js* (https://d3js.org/). Additionally, output histograms can be

exported as a .png file, or the tables can be exported to a .csv file. In addition to in-app visualization, a code snippet generator is available to create short scripts in R, Python (requires *matplotlib*), and *MATLAB* for user customized plotting options (fig. 2).

(3)   IMfig generator—Provides a GUI for the IMfig plotting script that produces a visualization of divergence times, population sizes, and migration rates. IMfig is a Python program that generates a graphic representing estimates of demographic history from the IMa2/IMa2p run. Population sizes are represented by boxes with height corresponding to the time duration of the population (before/after a split) and width representing population size. Confidence intervals for these population sizes are represented by faint boxes. Migration is represented by curved arrows between population boxes. Various output options can be selected to toggle output features and scale/color the image. The IMfig generator parses user input to validate that all options are in a legal range, then generates an EPS file for use in publications and a JPG file for viewing in the app—see figure 3.

IMGui has been developed to provide users access to either run IMa2p in serial (i.e., on a single CPU machine), or in parallel. Additionally, the IMGui app has been tested on numerous commonly utilized operating system platforms, including Windows, Linux (CentOS, SUSE, Red Hat, Ubuntu, and Fedora), and OS-X. A demo job that includes an input file (comprising five loci, and testing a four-population IM model), and suggested parameters has been bundled with all distributions of IMGui. A detailed description on how to set up parameters, run, and visualize output from this job is available on the IMGui GitHub page: https://github.com/jaredgk/IMgui-electron-packages.

Acknowledgments

References

Hey J. 2010a. The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol Biol Evol.* 27(4):921–933.

Hey J. 2010b. Isolation with migration models for more than two populations. *Mol Biol Evol.* 27(4):905–920.

Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis. Genetics* 167(2):747–760.

Hey J, Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci.* 104(8):2785–2790.

Kuhner MK. 2009. Coalescent genealogy samplers: windows into population history. *Trends Ecol Evol.* 24(2):86–93.

Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158(2):885–896.

Sethuraman A, Hey J. 2016. IMa2p—parallel MCMC and inference of ancient demography under the Isolation with migration (IM) model. *Mol Ecol Resour.* 16(1):206–215.