MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms

Sudhir Kumar,[†,1,2,3] Glen Stecher,[†,1] Michael Li,[1] Christina Knyaz,[1] and Koichiro Tamura*[,4,5]

[1]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA

[2]Department of Biology, Temple University, Philadelphia, PA

[3]Center for Excellence in Genome Medicine and Research, King Abdulaziz University, Jeddah, Saudi Arabia

[4]Research Center for Genomics and Bioinformatics, Tokyo Metropolitan University, Hachioji, Japan

[5]Department of Biological Sciences, Tokyo Metropolitan University, Hachioji, Japan

[†]Joint first authors.

*Corresponding author: E-mail: ktamura@tmu.ac.jp.

Associate editor: Fabia Ursula Battistuzzi

Abstract

The Molecular Evolutionary Genetics Analysis (MEGA) software implements many analytical methods and tools for phylogenomics and phylomedicine. Here, we report a transformation of MEGA to enable cross-platform use on Microsoft Windows and Linux operating systems. MEGA X does not require virtualization or emulation software and provides a uniform user experience across platforms. MEGA X has additionally been upgraded to use multiple computing cores for many molecular evolutionary analyses. MEGA X is available in two interfaces (graphical and command line) and can be downloaded from www.megasoftware.net free of charge.

Key words: software, evolution, Linux.

Introduction

Over the past decade, genome sequencing has become an efficient and potent means to investigate a broad array of biological systems, from large-scale studies of biological diversity to tracking the evolution and origin of pathogenic microbes. The numerous steps required to glean interpretable and actionable results from raw sequence data invariably require comparative analysis of molecular sequences to discover functional and adaptive genome differences. Molecular Evolutionary Genetics Analysis (MEGA) software provides tools to conduct such analyses. MEGA includes a large repertoire of programs for assembling sequence alignments, inferring evolutionary trees, estimating genetic distances and diversities, inferring ancestral sequences, computing timetrees, and testing selection (Kumar et al. 2016). Over the last 25 years, MEGA's use in evolutionary analysis has been cited in over one hundred thousand studies in diverse biological fields.

MEGA across Computing Platforms (MEGA X)

MEGA was first developed for MS DOS in the early 1990s (Kumar et al. 1994) and then upgraded for use in MS Windows eight times, including MEGA 1 to MEGA 6 and MEGA-CC and MEGA-MD (Kumar et al. 2001, 2016). Some of the MEGA releases have been packaged for Linux systems using the WINE compatibility layer for POSIX-compliant operating systems and the Wineskin tool (built on WINE) for macOS systems. These versions have been downloaded over 200,000 times. But the ad hoc Windows-emulation solution is sluggish and relatively unstable when compared with the performance in MS Windows. Emulators cannot be used effectively for the latest 64-bit version of MEGA that is built to handle memory-intensive analyses of large contemporary data sets (Kumar et al. 2016), so a more comprehensive solution is required for users of alternate platforms.

Therefore, MEGA has been transformed into a cross-platform version that runs natively on Linux and Microsoft Windows. This advancement eliminates the Windows-only limitation of MEGA, which has become particularly acute due to the increasing use of Linux in biological research. This transformation also paves the way for development of a MEGA X version for macOS in the near future.

Cross-Platform Graphical User Interface (GUI) for MEGA X

The GUI has been programmed by refactoring MEGA's source code to be compatible with the Lazarus integrated development environment (Lazarus Team) and the Free Pascal compiler. Extensive reprogramming of visual components was needed to compile cross-platform MEGA, including the Alignment Explorer, Sequence Data Explorer, and Tree Explorer. Additionally, all the internal control code that used the MS Windows system application programming interface (API) was advanced to work across platforms.

The reprogramming was accompanied with visual modernization of many parts of the GUI (fig. 1). To provide a uniform user experience across computing platforms, many user interface elements have been implemented using the Chromium Embedded Framework (CEF), which is an open source framework based on the Google Chromium web
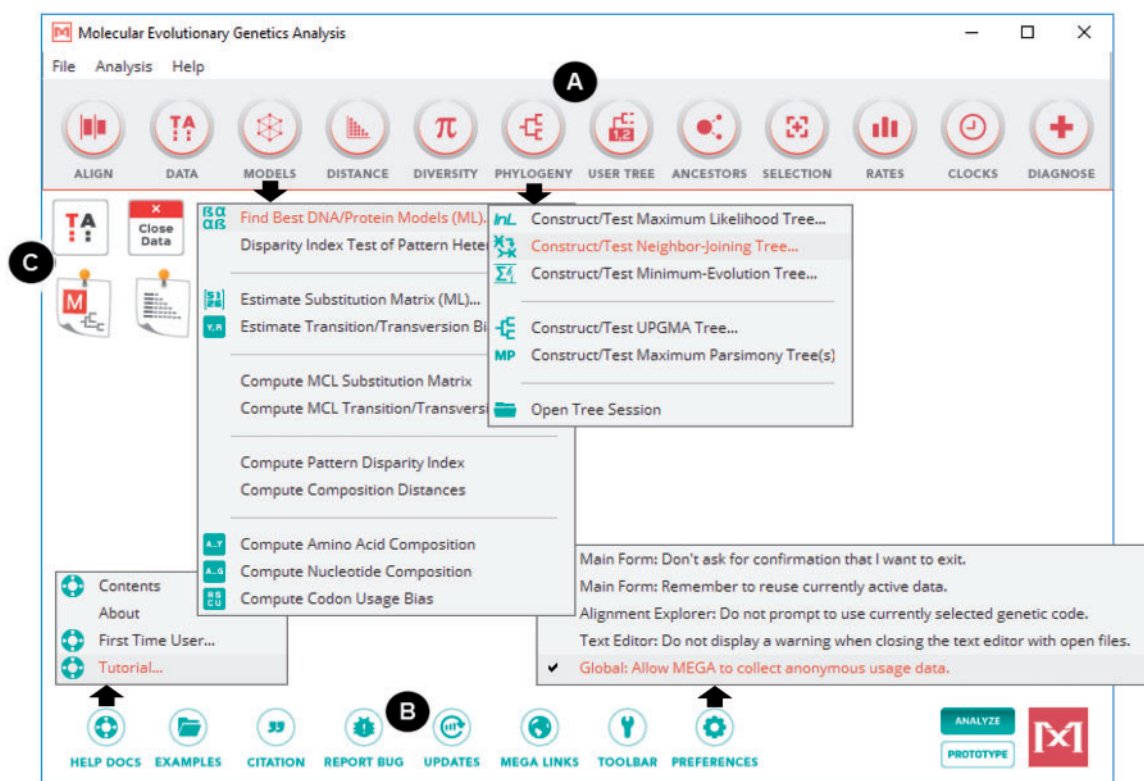
Fig. 1. The MEGA X main form has a modernized look-and-feel, but it maintains the familiar structure of the previous versions of MEGA. (*A*) The top toolbar organizes the large number of analyses available in MEGA X into logical groups accessed via drop down menus. (*B*) The bottom toolbar provides access to utility functions such as the help system, example input data files and application preferences. (*C*) Icons on the main form provide convenient access to input data and results explorer windows. Mega X can now be used in two modes: one for data analysis (ANALYZE mode) and another for prototyping (PROTOTYPE mode).

browser core. With this framework, we have implemented GUI elements using HTML and JavaScript in MEGA X desktop GUI (fig. 2). As a result, MEGA X has the same look-and-feel in Windows and Linux. We are currently developing MEGA X for macOS, which will share this user interface.

Parallelization of Computational Core

We have also improved MEGA's computational core by increasing the number of methods that can utilize multiple processors and cores on the same machine. This will greatly reduce the run time for computationally demanding inferences. Users can specify the maximum number of cores to deploy in the Analysis Preferences dialog box (fig. 2*G*). By default, MEGA X uses all available cores, minus one, such that no >80% of RAM is consumed. Computational methods benefitting from calculation parallelization include those for selecting the best model for nucleotide or amino acid substitutions and bootstrap tests of molecular phylogenies employed with distance-based and parsimony-based methods. Users can already employ multiple cores during the heuristic search for the Maximum Likelihood (ML) tree in MEGA.

Distributions

On Microsoft Windows, MEGA X is intended to be used on 64-bit systems having version 10 and later. It will, however, work smoothly on previous versions of 64-bit Windows. We also

provide a 32-bit version of MEGA X for Windows for those with 32-bit computers. On Linux platforms, MEGA X has been optimized and tested for use on Debian/Ubuntu and RedHat/CentOS based distributions. Because users may customize their Linux environments easily, we recommend use of MEGA X for Linux in tested desktop environments. Specifically, we have tested MEGA X on Ubuntu 16.04.4 (with Unity desktop), Mint 18.1 (with Cinnamon desktop), and CentOS 7 (with GNOME desktop).

MEGA X now generates a configuration file specifying analysis options for user-desired analysis in MEGA CC, which many researchers use for high throughput and iterative analysis in Windows, Linux and macOS environments (Kumar et al. 2012). The "Save Settings" button provides this functionality (fig. 2*H*). In fact, MEGA X can now be used in the prototype mode (see fig. 1, lower right corner), if desired, which obviates the need to distribute and maintain a separate application for generating the configuration file for running MEGA CC, so MEGA-Proto is no longer required or supported.

In conclusion, the cross-platform MEGA X will benefit many existing users and enable many others to consider utilizing MEGA X in their research and teaching endeavors. The consistency of user experience across platforms should be particularly useful in facilitating collaborative analyses within and among research groups as well as in teaching a diverse group of students. In the near future, we will complete the
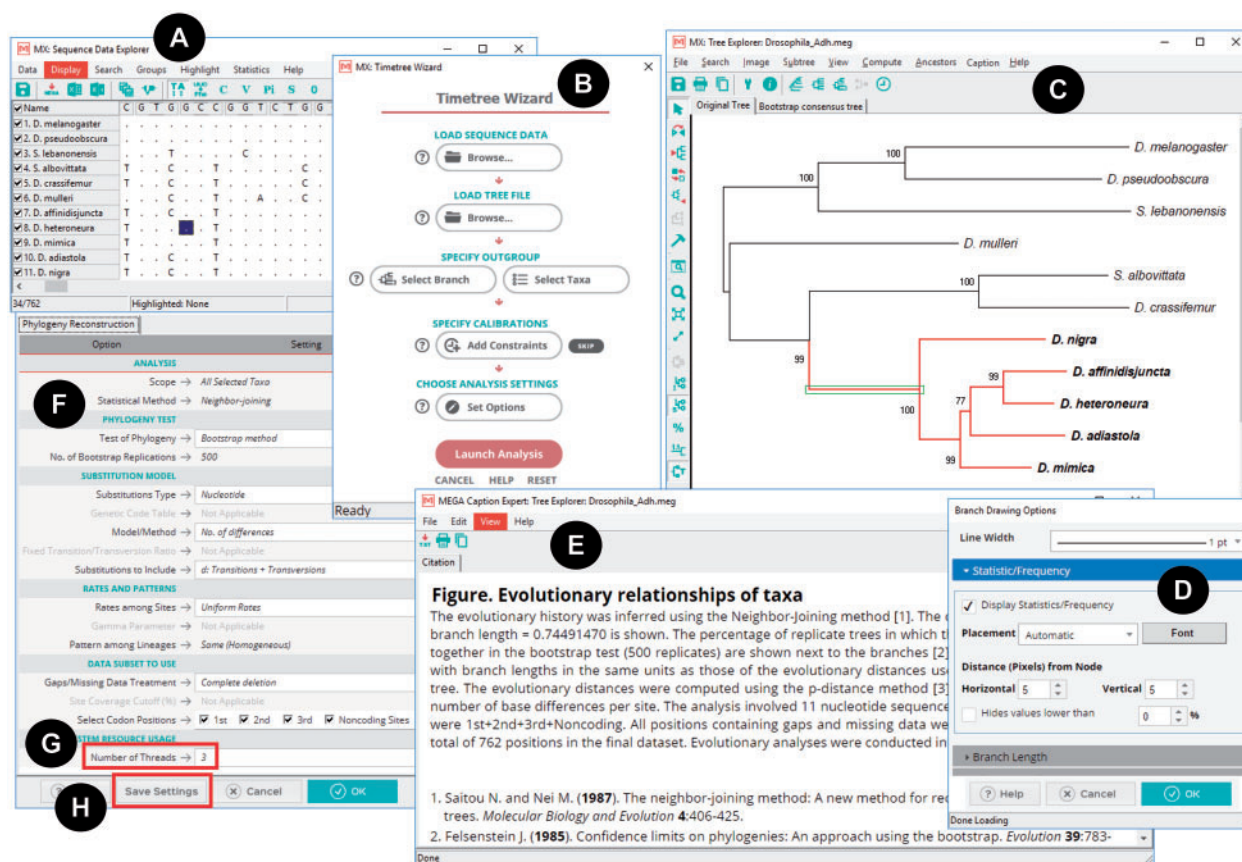
Fig. 2. The redesigned GUI elements in Mega X include explorer windows such as the (*A*) Sequence Data Explorer, (*B*) Timetree Wizard, and (*C*) Tree Explorer. These use the native widget set of the target operating system. (*D*) Options dialogs and (*E*) Caption Expert were developed using HTML and JavaScript and are displayed via Mega's integrated web browser that is built on the Chromium Embedded Framework. (*F*) The redesigned "Analysis Preferences" dialog box for selecting options for data analysis, including (*G*) the number of threads to employ and (*H*) a "Save Settings" button for generating an ".mao" file for use with MEGA-CC for high throughput analysis.

development of a macOS version to make Mega X available for all major operating systems.

References

Kumar S, Stecher G, Peterson D, Tamura K. 2012. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* 28(20):2685–2686.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 33(7):1870–1874.

Kumar S, Tamura K, Jakobsen IB, Nei M. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17(12):1244–1245.

Kumar S, Tamura K, Nei M. 1994. MEGA: molecular Evolutionary Genetics Analysis software for microcomputers. *Comput Appl Biosci* 10(2):189–191.

Lazarus Team. Lazarus Homepage. Available from: https://www.lazarus-ide.org/.