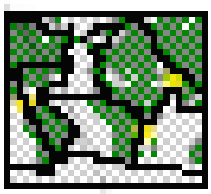
Croizat 2.0

A Software Package for Quantitative Analysis in Panbiogeography

User's Guide



Ecoinformatics Studio
P.O. Box 46521, CEP 20551-970
Rio de Janeiro, RJ, BRAZIL
Email: maurobio@gmail.com

Contents

Overview

Croizat is a freely available cross-platform software package for mapping analyzing biodiversity data using a graph-theoretic, quantitative panbiogeographic approach.

Panbiogeography (Croizat, 1958, 1964; Craw *et al.*, 1999 Heads, 2012) provides a method for analyzing the geographic structure of species distributions in order to generate predictions about the evolution of biological organisms in space and time.

The software is written in Python[1], an interpreted, interactive, object-oriented programming language, very suitable for scientific computing (Beazley, 2000; Bassi, 2007; Oliphant, 2007), and includes external libraries that are also written in Python and C/C++. Therefore, the program is platform-independent, running without modifications on any PC compatible with the x86 architecture, under GNU/Linux, MS-Windows, and Mac OS X.

It allows fast computing of individual tracks (created by the software itself using optimized routines) and generalized tracks (created using the *MartiTracks*[2] program which is automatically run by ***Croizat***).

Both ESRI shapefiles[3] and Google Earth KML[4] files are supported as output; it is therefore possible to save the generalized tracks generated by the *MartiTracks* program as shapefiles and not only in KML format.

1 https://www.python.org
2 http://code.google.com/p/martitracks
3 http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf
4 http://earth.google.com/kml

Download and Installation

Source code and binary installaton packages are available from SourceForge. There is no need to download external files and perform complicated system configurations; everything needed is already included in the installation package. **Notice: the binary installation package is available only for MS-Windows at this time.**

Running the Program

Running the program will bring up the main window, which consists of a single text area for output listing. Across the top of the main window is the main menu bar, which provides access to functions to read data and run analyzes. The various choices of the main menu are described on another section.

Below the menu bar, there is a toolbar, which allows quick access to some commonly used functions.

A status bar at the bottom of the main window displays, from left to right, a message describing the function of the current menu choice.

The main window is taken up by a digitized map of the world (in Equirectangular Cylindrical projection) to its right and, to its left, the associated table of contents (TOC). The user can resize those parts with the mouse pointer. The map may include country borders, as well as major lakes and rivers. By default, country borders and rivers are toggled off. The display of borders and rivers can be toggled on/off by choosing the "Map| Borders" and "Map|Rivers" options from the main menu.

The map window comes with a handy navigation toolbar, which includes the following buttons:

- The Forward and Backward buttons are used to navigate back and forth between previously defined map views. They have no function unless the user have already navigated using the Pan and Zoom buttons.
- The Home button always returns to the first (usually the default) map view.
- The Pan button allows to define new map views. A map may be panned by the clicking the left mouse button and holding it, while moving the mouse to dragging the map view to a new position. Pressing the 'x' or 'y' keys while panning, the motion will be constrained to the x or y axis, respectively.
- The Zoom button allows to zoom the map view limits to a rectangle a user can defined using the mouse pointer. To enable zoom, the mouse must be put somewhere over the map and the left mouse button have to be pressed. Dragging the mouse while holding the button to a new location and releasing it will display a zoomed view of the selected area. There is also an experimental "zoom out to rectangle" with the right button, which will place the entire axes in the region defined by the zoom out rectangle.
- The Save button launches a File Save dialog, to save the currently displayed map as an image file (in PNG, PS, EPS, SVG, or PDF formats).

The table of contents lists all the "layers" (also known as "themes") which comprise a given project. Layers consist of data files (in several formats) and shapefiles. Layers can be made (in)visible on the map using the checkbox next to their name on the TOC. The program automatically takes care of the ordering in which the layers are drawn on the map, so that a layer higher in the TOC will never cover a lower one. Because of this, the order of the layers cannot be changed by the user.

Layers in the Project panel can be added or removed by choosing the "Add File" or "Remove File" options from the File menu (or by clicking on the "+" or "-" buttons in the toolbar). Right-clicking on a layer in the TOC will bring up a dialog box that allows changes to be made to some of the layer's properties (such as symbol colors, sizes, and types for points and lines) on the map.

A project comprises all of the data related to a particular study. It may consist of many data files and shapefiles. A project file is a simple plain text file with the .prj extension and should not be edited. Project files contain pointers to the physical locations of the associated data files. Therefore, editing may destroy the functionality of a project file. When a project is saved, the data themselves are not saved but just the path names to the data and the changes the user may have done to the project (as adding or removing files from it). This means that one will not be able to take the project to another computer, unless the data are already available on that computer. Otherwise, it will be necessary to copy the data as well before running the project on the other computer. If the data are available on the other computer but is in a different folder, the user will need to help the software to find the data before it can reconstruct the project.

Program Menu Choices

Project

New

Creates a new project. If a project is already open and has been modified, its contents are automatically saved.

Open

Displays a File Open dialog box to open a project file and load its contents. If a project is already open, its contents are automatically saved.

Save As

Saves the contents of the current project to a file. If the project was loaded from a file (using the Open Project menu option) or was saved to a file previously, the current state of the project will be written to that file. If a file already exists with the specified name, the user will be asked to overwrite the old file. If the project is new and has not been saved before, a File Save dialog will automatically appear to prompt the user to save the file.

Add File

Adds a file to the current project.

Remove File

Removes a file from the current project. Note that this action does **not** delete the file from disk.

Import

Displays a dialog box allowing to import data from external sources (GBIF[5], VertNet[6], iDigBio[7], OBIS[8]) to the current project. If the project is new and has not been saved before, a File Save dialog will automatically appear to prompt the user to save the file.

Export

Displays a dialog box allowing to save the contents of the current project to a ESRI shapefile or to a KML file.

Quit

Terminates the program. This has the same effect as clicking on the close box on the application frame. A dialog box will appear asking for confirmation to exit. If a project is open, it will be automatically saved to disk before terminating the program.

5 http://www.gbif.org
6 http://www.vertnet.org
7 http://www.idigbio.org
8 http://www.iobis.org

Map

Borders

Toggles the display of country borders.

Rivers

Toggles the display of major rivers.

Limits

Displays a dialog box allowing to set latitude and longitude boundaries, in coordinate degrees, of the geographic region of interest, in case a custom view is desired (defaults to the whole earth). For example, coordinates 10, 162 W; -48 S, 182 E will result in a map of New Zealand and coordinates 35 N, -85 W; 10 S, -115 E will create a map of Mexico.

Center

Displays a dialog box allowing to center the map at the selected coordinates.

Resolution

Displays a dialog box allowing to select the map resolution, with a choice of crude, low, intermediate, and high. Resolution drops off by roughly 80% between options. Higher resolution maps are much slower to draw and take a vast amount of memory.

Blue Marble

Toggles the display of NASA *Blue Marble* image as a map background. Default image size is 5400x2700 pixels, which can be quite slow to drawing and use a large amount of memory. The scale option can be used to downsample the image (for example, setting a scale=0.50 downsamples the image to 2700x1350 pixels).

Colors

Displays a dialog box allowing to set the colors of land surface, water bodies (oceans, seas, lakes), and country borders.

Title

Presents a dialog box that allows to set or change the title of the current project. The title is centered on the top of the map.

Copy

Copies the currently displayed map to the clipboard for pasting into another program (*e.g.*, a graphics program or word processor).

Reload

Reloads data and redraws the map.

Analysis

Individual Tracks

Computes minimum-length spanning trees (MST) as undirected graphs connecting the distribution records of selected species. The output is written to disk as both a shapefile and a KML file.

Generalized Tracks

Finds the largest groups of mutually compatible tracks (generalized tracks), using the *MartiTracks* program which is automatically run. The program will ask for the values of the parameters to be used in the analysis. See the *MartiTracks* manual for details. The output is written to disk as both a shapefile and a KML file.

Help

About

Brings up a dialog box which gives the author's name, a copyright notice, and information about version number of the program, the version of the Python interpreter and required libraries, as well as the operating system.

Data File Format

Croizat accepts input from data files in ASCII format (plain text files, not binary files), that can be created and modified with any text editor (in non-document or unformatted mode) or electronic spreadsheet, as well as files in MS Excel format (.xls, .xlsx), Open Document Format (.ods), dBase format (.dbf), and Keyhole Markup Language format (.kml). The easiest way to create and maintain data files for input to the program is to use a spreadsheet like MS Excel or OO/LO Calc.

The first line of the input file **must** be a header line, which gives field names for the dataset. The input file may contain any number of fields (columns), but the **first four** fields **must** be in this order: species, locality identifier, longitude, latitude. If the input file is in csv format, the fields can be separated by commas, semicolons, or tabs; the program will try to guess the delimiter used to separate the data values and the text qualifier (a character delimiting ambiguous data), which must a single or double-quote (or none). All coordinates must be in decimal format. There can be only one species in the same data file. Note that missing values (blank cells) are not allowed for by the program.

Here is an example of a valid data file in .csv format:

```
Species,Locality,Longitude,Latitude
Zygodontmys,Bush Bush Forest,-61.05,10.4
Zygodontmys,Cerro Azul,-79.4333333333,9.15
Zygodontmys,Dividive,-70.6666666667,9.53333333333
Zygodontmys,Hato El Frio,-63.1166666667,7.91666666667
Zygodontmys,Finca Vuelta Larga,-63.1166666667,10.55
Zygodontmys,Isla Cebaco,-81.1833333333,7.51666666667
Zygodontmys,Kayserberg Airstrip,-56.4833333333,3.1
Zygodontmys,Limao,-60.5,3.93333333333
Zygodontmys,Montijo Bay,-81.0166666667,7.66666666667
Zygodontmys,Parcela 200,-67.4333333333,8.93333333333
Zygodontmys,Rio Chico,-65.9666666667,10.3166666667
Zygodontmys,San Miguel Island,-78.9333333333,8.38333333333
Zygodontmys,Tukuko,-72.8666666667,9.83333333333
Zygodontmys,Urama,-68.4,10.6166666667
Zygodontmys,Valledup,-72.9833333333,10.6166666667
```

Croizat also accepts input data directly from KML (*Keyhole Markup Language*) files, an XML-based language schema for expressing geographic data developed for use with Google Earth.

Croizat can also read and optionally draw ESRI *shapefiles* (ESRI, 1998). A shapefile consists of three separate files with the same name but with different extensions (.shp, .shx, .dbf), which are treated as one single file.

Distribution

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see http://www.gnu.org/licenses.

The latest versions of ***Croizat*** will be made available directly over the Internet from the program website at https://sourceforge.net/projects/croizat/files.

Announcements of updates, bug fixes and other significant changes in the program will be distributed via the Panbiog-L discussion list at Google Groups (http://groups.google.com.br/group/panbiog).

Credits

I thank Brad McFall, Claudio Quezada, Daniel Previatelli, Douglas Riff, Edinaldo Nelson dos Santos Silva, Gervásio Carvalho, Ian Henderson, John Grehan, Jonathan Liria, José Maria Cardoso da Silva, Juan Morrone, Maria Lucia Lorini, Nicholas Matzke, Michael Heads, Renato Bernils, Robin Craw, Ronaldo Alperin, and Tania Escalante for contributing knowledge, moral support, and bug reports.

I also thank the members of the Matplotlib-users mailing list for useful comments for the improvement of this software. Special thanks to Jeffrey Whitaker, Scott Sinclair, John Hunter (*in memoriam*), Ryan May, Stef Mientki, and Benjamin Root for providing assistance and technical expertise.

The support of Conservation International do Brasil and Projeto Biotupé, Instituto Nacional de Pesquisas da Amazônia is much appreciated.

This program has been developed under Microsoft Windows 7 Professional, using Python 2.7.

The user interface was designed with the portable, multi-platform Qt/PyQt[9] interface management library.

All numerical computations are done in double precision arithmetic using routines from the NumPy[10] library written by Travis Oliphant (Oliphant, 2007).

Minimum spanning trees are computed after Page (1987) using Kruskal (1956) algorithm. The *MartiTracks* program (Echeverria-Londoño & Miranda-Esquivel, 2011) is used to find groups of compatible tracks.

The plotting routines are from the Matplotlib[11] library written by John Hunter (Hunter, 2007) with the Basemap[12] module written by Jeffrey Whitaker.

Coastline data is from the GSHHS[13]. Country and river datasets are from the Generic Mapping Tools[14]. The *Blue Marble* image is from NASA Visible Earth[15].

9 http://www.riverbankcomputing.com/software/pyqt/intro
10 http://www.numpy.org
11 http://matplotlib.org
12 https://matplotlib.org/basemap
13 http://www.soest.hawaii.edu/wessel/gshhs/gshhs.html
14 http://gmt.soest.hawaii.edu
15 http://visibleearth.nasa.gov

Citation

Croizat can be cited in a publication as follows:

CAVALCANTI, M.J. 2009. Croizat: A Software Package for Quantitative Analysis in Panbiogeography. *Biogeografía* **4**: 4-6.

References

BASSI, S. 2007. A primer on Python for life science researchers. *PLoS Computational Biology* **3**: 2052-2057.

BEAZLEY, D.M. 2000. Scientific computing with Python. In: MANSET, N., VEILLET, C. & CRABTREE, D., eds. *Astronomical Data Analysis Software and Systems IX*. ASP Conference Series, Vol. 216, pp. 49-58.

CROIZAT, L. 1958. *Panbiogeography*. Published by the author, Caracas. 1731 p.

CROIZAT, L. 1964. *Space, Time, Form: The Biological Synthesis*. Published by the author, Caracas. 881 p.

CRAW, R.C. 1989. Quantitative panbiogeography: introduction to methods. *New Zealand Journal of Zoology* **16**: 485-494.

CRAW, R.C., GREHAN, J.R. & HEADS, M.J. 1999. *Panbiogeography: Tracking the History of Life*. Oxford University Press, New York. 229 p.

ECHEVERRÍA-LONDOÑO, S. & MIRANDA-ESQUIVEL, D.R. 2011. MartiTracks: a geometrical approach for identifying geographical patterns of distribution. *PLoS ONE* **6**(4): 0018460.

ESRI. 1998. *ESRI Shapefile Technical Description*. Environmental Systems Research Institute, Inc. New York. 28 p.

HEADS, M. 2012. *Molecular Panbiogeography of the Tropics*. University of California Press, Berkeley. 576 p.

HENDERSON, I. 1989. Quantitative panbiogeography: An investigation into concepts and methods. *New Zealand Journal of Zoology* **16**: 495-510.

HENDERSON, I. 1991. Biogeography without area? *Australian Systematic Botany* **4**: 59-71.

HUNTER, J.D. 2007. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9**: 90-95.

KRUSKAL, J.B. 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society* **7**: 48-50.

OLIPHANT, T.E. 2007. Python in scientific computing. *Computing in Science & Engineering* **9**: 10-20.

PAGE, R.D.M. 1987. Graphs and generalized tracks: Quantifying Croizat's panbiogeography. *Systematic Zoology* **36**: 1-17.