VIP: Vicariance Inference Program
Version 10.02.15
by J. Salvador Arias

User's manual

Index

1. Introduction

The biogeographical analysis within a single lineages (“taxon history”) has been subject of several methods based on the concept of ancestral areas (e.g. Fitch optimization[1] or DIVA[2]). Unfortunately, such methods rest on the idea of predefined areas that are optimized in some way into a phylogenetic trees. There are several problems on the use of predefined areas [3][4], that ultimately end in the doubtful implementation of such methods [5].

Based on insights from Rosen [6], Hovenkamp [3][4] propose a method to deal directly with spatial data in a phylogenetic context. Although his objective is more akin to “cladistic biogeography” (a search for history of earth), ideas from Hovenkamp are useful for use in taxon history. The program present here, Vip uses ideas the ideas of Hovenkamp in a optimality criterion framework allowing the user to set reconstruction costs, use heuristic searches and visualize the results in a geographical defined graphic interface.

1.1 Data requirements

Vip requires two types of primary information. First, it requires a cladogram depicting phylogenetic relationships among the study taxa. The second type of data is georeferences (latitude, longitude) from terminals. Each terminal must have a georeferenced point to be taken into account in the analysis. Additionally, in the case of time calibrated phylogenies (i.e. an analysis including fossils) is possible to assign an age to terminals and nodes.

A detailed account on data formats and data edition can be found in the following sections.

1.2 Program background

Hovenkamp [3][4] argued that the only evidence left from an speciation process in a geographic context are allopatric distributions. Unfortunately, although many clades show an allopatric distribution, many does not. There are some distributions that present some form of overlap, others in that a widespread terminal or clade make uninformative several of its ancestral nodes.

Some of these problems are also found in the “area-based” biogeography. Area based biogeography assumes that distributions can be resumed as few homogeneous areas. Vip does not take that approximation, but it uses some solutions developed under such paradigm [7][8]. If a distribution of a node is ignored, and it is possible to found a solution with more nodes showing allopatric distributions, this solution might be preferred as it increases the explanatory power of allopatry. But also, removing a node is a form of an ad-hoc rejection of conflicting evidence.

Vip uses a system of costs, in which the user can set the cost of a node distribution removal, and the cost of a non-vicariant node (a node that can not be explained geographically). If distribution removal has a cost of 0, the result if the maximum number of possible nodes explained by allopatry, that is “maximum vicariance” [7][8][9]. If removal has an infinite cost, the result is the maximum number of possible nodes explained without removal, as in the original Hovenkamp's method [3][4] and similar to “assumption 0” [3] or “paralogy free” subtrees [10].

Details about the method are described in [5].

1.3 Results

The results of an analysis with vip is displayed in two windows, the map and the tree window. In the map window, allopatric distributions are displayed using red and blue for each descendant, if there are some overlap, it is show in green. It is possible to show a potential barrier, calculated using the Voronoi tessellation. In a tree window, nodes with allopatric distribution are marked with a black square. Nodes with removed distributions are shown as white circles. In some cases, a node is removed, but it also has an allopatric distribution, in such cases, a white square is displayed in the node.

1.4 Running VIP

Vip can be downloaded as binary or in form of source code. Binaries for Linux and Windows are available.

1.4.1 Linux binaries. To run linux version you need to install the gtk-bin package. This package is already installed in Gnome desktop. For KDE, you can install this package using synaptic, get-app or a similar tool from your preferred repository.
1.4.2 Windows binaries. To run the program you must install the gtk windows runtime. The latest version can be downloaded here (http://sourceforge/projects/gtk-win/).

Also, it is possible to compile the code yourself. The source is write on plain C, using gtk. Make sure you have the latest version of gtk-dev library (available here: http://www.gtk.org/download.html, linux users can download and install it with synaptic, app-get or a similar application). In windows you need to link the program including the following libraries gdk-win32.lib, gdk_pixbux.lib and pango.lib.

1.5 License and citation

The source code of vip and its documentation is distributed with a creative-commons license (source: GPL http://creativecommons.org/licenses/GPL/2.0/; documentation: attribution-share alike http://creativecommons.org/licenses/by-sa/3.0/) which means that you can use and modified it as long as you give attribution and keep the same license.

As any scientific work, if you use this program please cite the method description [5], Hovenkamp original proposal [3][4], and the program itself. Also I suggest including algorithms or methods used/modified by the program (such as [8] and [11]) where applicable.

1.6 Contact

I will be happy to hear you, for any questions, suggestions, bug reports, your vip user experience or your code modifications. You can contact me at: jsalarias@csnat.unt.edu.ar. If you like, I also can include you in a list of vip users, so I can send you reminders of new available versions.

For bug reports, please specify how the bug is triggered, and if possible, a data set that produce the error.

Also, I will be happy if you send me a pdf copy of your published results using vip.

2. General layout

Vip is a program that uses a Graphic User Interfaz (GUI) based on different windows. There are three basic windows. The main window has the menus and parameter entries for basic operations, such as data loading, searches, reconstructions and data edition. The tree window displayed the actual tree and node, it has menus for basic tree display. The map window displayed the position of records for the selected node, and a map, if any; it has menus for map configuration, and record display.

In all windows, there are keyboard shortcuts for all menu operations.

In this section, a general overview of the layout is provided, more detailed accounts (based mostly on menu actions) are given in the following chapters.

2.1 Main Window

The main window is used for user input (files and search/reconstruction parameters), and for a basic report of the results.

It has four menus. File menu is similar to most program's file menu, it is used to load input data, export data and exit the program. Grid menu is activated when there are some data, it is used to set the grid. Recons menu is activated when a grid is active, it is used for searches and reconstruction operations. Edit menu is activated when there some data, it is used to some basic edit operations for the records and reconstructions.

The windows has two tabs. The first tab is label as parameters. It is used to set the parameters used during the reconstruction. The box cost of non vicariant nodes is used to set cost of nodes that are not allopatric. The box cost of node removal is used to set the cost of a distribution removal. The cost of the reconstruction is the sum of the cost assigned to each node. The check box use fractional cost is used to give more grain to the reconstruction. If fractional costs are used, then an allopatric node might be with a cost if it has some overlapping, the cost is proportional to the overlap: cost = (non vicariant cost) * (overlap ratio), for example, with a cost of 2.0 and an overlap of the 10% the cost of the node is 2.0 * 0.1 = 0.2.

The max overlap box sets the maximum (as percentage) acceptable overlap for a node to be counted as allopatric. The late overlap box allow the program to search using late overlap, that is, instances of overlapping in a nested node with respect to a node attached more basally (see [5]). The max late overlap box is used to set the maximum late overlap acceptable. If late overlap and fractional costs are used simultaneously, then the removal cost is scaled with the overlap ratio.

If data is time calibrated, then the use time check box is activated. With this option checked, the searches take into account the age difference between the descendants (similar to [11], see [12]).

Apply and cancel buttons are used to set the values, or the return to the previous ones. Execute button is equivalent to the menu Recons\Search... and perform the search using the actual parameters (equivalent to apply and then search).

The reconstruction tab is used to show some basic statistics of the actual reconstruction, and to navigate across the data. The tree navigator shows the number of trees in memory, and the name of the actual tree. With the buttons back and forward is possible to change the tree. The reconstruction navigation box shows a cost box with the actual cost of the reconstruction, the vicariant box show the number of vicariant nodes in the reconstruction, the box nodes with removal indicates the number of nodes with a removed distribution. The reconstructions in memory shows the number of reconstructions found with the searches, and if a searched reconstruction is selected indicates the id (number) of the actual reconstruction, in this case, the back and forward button allows the user to change the reconstruction. The option entries OR reconstruction, Searched reconstruction and Consensus reconstruction are used to select among reconstruction kinds. The node navigator shows the label of the actual node, it can be changed using the back and forward buttons.

2.2 Tree view window

The tree view window is used to visualize the actual tree, and the reconstructions, if any.

To move the tree inside the window it is possible to use arrow keys, or holding the mouse left button (drag), also, if there are a mouse wheel, it can be used to scroll the tree vertically. Tree size can be changed with arrows keys and Ctrl button. To quick size Ctrl + F fit the tree to the actual window, and with Ctrl + T fit the space between terminals to be readable.

With the mouse left button is possible to select nodes by a click. Also using page up and page down, the selected node change. The selected node is marked with a green square. If the node is a vicariant node, then their descendants are marked with a red and blue square. With a click of the right button, information about the node is displayed, and the name of the node can be edited.

If there are a reconstruction, then the vicariant nodes are displayed as black squares. If the node span is removed is displayed with a white circle. If the node span is removed but also is a vicariant node, then it is show with a white square.

2.3 Map window

The map window is used to display the records of the selected node.

The records are displayed as green squares. If the node is a vicariant node, then the records of each descendant are displayed in red and blue, according to each descendant (in coordination with the colors used in the tree view window), in such cases, the overlap is show in green. In the case of consensus reconstructions, not all records are included, so they are showed in white.

With a click on the left button of mouse over a record, information of the particular record is displayed. By holding the right button, it is possible to move the map. At the bottom of the mouse window, there is a label showing the latitude and longitude at the mouse pointer, if it is on the map.

3. File operations (Main Window > File menu)

Basic file operations are packed in the File menu, at Main Window, this operations behave nearly identical to file operations in most GUI applications.

3.1 Open (Ctrl + O)

This menu shows a file chooser dialog, that allows the user to select a file to be open. At this moment, the only valid format is the pKML files (KML like files). Although just only one file can be open each time, several files might be open, allowing the user to use multiple trees.

3.2 Format input

The pKML files are a derivation of the KML files used by Google Earth. Although not any KML file can be easily transformed into a KML file, the pKML files are nearly compatible with the KML standard. It is not expected that the user edit directly this files. A simple TNT macro is provided (topkml.run) to write a phylogenetic tree into a pkml file. Then using vip, the point localities of each terminal can be inputted manually.

All pKML files start with a <Document> token, a name (between <name> and </name> is optional). Each node (internal or terminal) must be identified with the <Folder> token. If it is a terminal, the folder can contain one of several <Placemark>. To close a node the token </Folder> is used.

The <Placemark> token identified a point <Point> and its coordinates <coordinates> using longitude and latitude. Here is an example

```
<Folder>
        <name>Bassaris_itea</name>
        <Placemark>
                <Point>
                        <coordinates>145.5000,-20.0000</coordinates>
                </Point>
                </Placemark>
        <Placemark>
                <Point>
                        <coordinates>168.0000,-45.0000</coordinates>
                </Point>
                </Placemark>
</Folder>
```

3.3 Save (Ctrl + Shift + S)

It is possible to save files using pkml format (for example, after edition), or a kml file (to be open with Google earth).

3.4 Close (Ctrl + W)

This option removes all the loaded data from memory, without closing the program.

3.5 Quit (Ctrl + Q)

Exit the program.

4. Grid operations (Main Window > Grid menu)

Vip uses a grid to represent the spatial distribution of taxa. The grid menu is used to set the basic properties of the grid.

4.1 Set grid (Ctrl + G)

This menu sets the size of each grid, and the filling around. The grid always start at -180 lon, 90 lat. The fill is the number of cells around the record to be counted as presences. For example, a fill of 0, only the cells in which a record fall are counted with a presence, a fill of 1, the eight cells surrounded the cell in which the record fall as presences.

Instead of trying several starting points of the grid, it is better to use an small grid size, and a fill. For example, instead of a 5 degree cells, using a 1.66 degree cells and a fill of 1, make a virtual 5 degree cell for each record.

4.2 Create grid (Ctrl + A)

The grid set does not creates the grid, just set the rules for the grid creation. The create grid menu creates a grid in memory. As the grid is compressed to save memory and computation time, record edition is unavailable after the grid is created. When the grid is created, a default OR reconstruction is performed on the data, activating several of the information boxes in the reconstruction tab.

4.3 Destroy Grid (Ctrl + D)

Removes the grid and the reconstructions from memory. The absense of grid is reflected on reconstruction tab showing “no grid” in several fields. As the grid is removed, it is possible to edit records.

5. Reconstruction operations (Main Window > Recons menu)

This is the basic menu of Vip. It encapsulates the basic operations about reconstructions: searching, consensus and filtering reconstructions.

5.1 Search/Execute (Ctrl + E)

This menu displays the search dialog. The search use an iterative removal [5, 8] of nodes to search for the best reconstruction score, for details see [5]. The dialog display the number of independent iterations (by default 100). Accept equals box is used to stipulate the frequency (in percent units) to keep ties at each iteration, by default it is 50. A low or 0 value of this parameter make the search faster, but for typical data this setting surely fails to even hit the best score. With values near to 100 (i.e. accept nearly all ties) slow down the search and it might fall to found reconstructions as the search get stuck at ties. If there are already reconstructions in memory, checking keep actual reconstructions check box do not clean the reconstruction buffer before the search, and the best score of the actual reconstructions is used as bound. Note that if a better score is found, then the buffer clean all reconstructions (as they are suboptimal).

When a search is performed, a progress dialog shows the status of the seach.

KNOWN BUG: When the search is about to finish, some times (very few) the progress dialog suddenly show an extraneous number (if numbers are shown) and freezes. Actually the search was finished and results are displayed on screen, but the program fail to close the dialog. If you suspect that the search takes longer than expected and the progress bar is just near to the end, try to close the dialog, if it not close, then the search is already running. This error is very rare.

5.2 Consensus (Ctrl + C)

After a search, the consensus of the found reconstructions can be calculated. Note that before each search, the consensus is cleaned. After every search consensus is only calculated the first time that this menu (or the consensus reconstruction option on reconstruction flap) is selected, any other selection of this options just shows the previous consensus.

5.3 Filter (Ctrl + F)

Removed all suboptimal reconstructions and clean the consensus. It can be helpful, for example, how is the best reconstruction found, under other cost parameter setting.

5.4 Clean (Ctrl + L)

Removes all reconstructions in memory.

6. Edit operations (Main Window > Edit menu)

This menu contains some basic edition options. This instructions are only available under certain circumstances.

6.1 Hot mouse (Ctrl + H)

Checking this option allows the user to edit manually the records for each terminal. It is only active when there are no active grid. A terminal must be selected. Records can be added on the map window by pressing the right mouse button holding the shift key. By clicking the right mouse button and holding the Ctrl key over a record, a georeference browser dialog is displayed, showing the selected point, it is possible to change the coordinates or delete records using this dialog box.

6.2 Edit reconstruction (Ctrl + T)

When a searched reconstruction is selected, pressing the right mouse button holding the Ctrl key over a node (in the tree window) change its status (removed or not removed and vice verse). Two points must to be taken into account. First, the selected recostruction is lost, and second partial overlapping is not calculated (so if the changed node has some partial removal, this information is lost).

6.3 Georeference browser

Displays the georeference browser. If the grid is not active and Hot mouse option is selected is possible to edit georeferenced records.

7. Tree formating (Tree window > Format menu)

This menu allows the user to configure the shape and sorting of the tree. If the tree contains time labels, it is possible to draw the tree as a “chrono tree” (Ctrl + C) in which branch forks are displayed according to their associated time.

7. Tree reconstruction operations (Tree window > Reconstruction menu)

This menu is used to show reconstructions when the consensus reconstruction is selected.

After making the consensus, the program check if there are other available nodes that share the same vicariant distribution than the selected node. To be available, a node must be on an independent clade to the selected node (i.e. it can not be a descendant or an ancestor). This is equivalent to search for a “supported vicariance event”.

But different from [3], the search is node based, so instead of finding an event, it find other nodes that share a common allopatric pattern with the selected node. This in that sets might be no reciprocal. For example, a node A share an allopatric pattern with node B and node C, but node C is a descendant of B, so node C will not be valid on the common allopatric set of B.

7.1 All reconstructions (Ctrl + L)

The default option, shows the reconstruction of all nodes.

7.2 Common reconstructions (Ctrl + N)

This option shows only the reconstructions that share the same common allopatric pattern as the selected node.

7.3 Supported reconstructions (Ctrl + S)

This options shows only the reconstruction off all nodes that share a common allopatric pattern with another node. Note that given the nature of the set reconstruction, the showed node might not be simultaneously supported.

8. Map operations (Map Window > Map menu)

This menu is used to basic map operations, such as loading, and scale.

8.1 Open (Ctrl + O)

Load an image to be used as map. Most popular graphic formats as jpg, gif, bmp and png are supported. The map must be in a cilindrical and isometric projection (i.e. equirectangular projection), to coincide with the window coordinate system.

8.2 Map limits (Ctrl + L)

This menu display a map limit dialog. It is used to set the limits of the actual map, so the map will be put on the right coordinates, and in adequate the scale. Note that the limits are just for the loaded image. Records can be scattered in any valid geographic point of the earth.

In absence of loaded image, it can be used to change the scale of the displaying window (it is assumed than an image of 720 x 360 pixels is loaded).

8.3 Centre on (Ctrl + C)

Center the window at the selected coordinates.

9. Drawing operations (Map Window > Drawing menu)

This menu is used to change the display on the map.

9.1 Record radius (Ctrl + R)

Allows the user to change the size used to draw records. The radius is measured in pixels. It also change the mouse radius sensibility above records.

9.2 All records (Ctrl + A)

If checked (the default) shows all records, records that did not pertains to the reconstruction are showed as void boxes. If unchecked, just the records assigned to the actual reconstruction are showed.

9.3 Barrier (Ctrl + B)

If checked, shows the potential barrier associated with the selected node, if the node is reconstructed as vicariant (see Barriers menu).

9.4 Grid (Ctrl + G)

If checked, display the grid based on the actual grid settings.

10. Barrier operations (Map Window > Barriers)

This menu is active only when the barrier option (Map Window > Drawing > Barrier) is checked. It change the display and properties of barrier drawing.

Barriers are calculated using the Voronoi-Delaunay dual graph on the records reconstructed for the selected node. But instead of storing all triangles/cells, just the triangles that contain points from each descendant are kept. Points on overlapping areas are not taken into account for the triangulation.

KNOWN BUG: This is not exactly a bug, just a consequence on how the program works, i.e. assuming a limited earth. For some amphi-pacific distributions, triangles connecting both pacific halves are never searched.

10.1 Voronoi lines (Ctrl + V)

When this option is selected, a line using the Voronoi tessellation is showed. Note that the line did not necessary fall in the “correct” barrier, it is only a devise to easily locate the potential geographic barrier associated with the disjuntion. As the Voronoi segments are mid-distance lines, farther points produce less meaningful barriers.

KNOWN BUG: In some reconstructions, for a problem in the sorting of triangles, in some drawings a “loop” is displayed. These loops are easily identified, and apart of unpleasant aesthetics, they are harmless.

10.2 Delaunay trinangles (Ctrl + D)

When this option is selected, the triangles of the Delaunay triangulation are showed. This is equivalent to show the space between the disjoint distributions.

10.3 Barriers on ancestors (Ctrl + N)

If checked, display the barriers detected on parent nodes of the selected node. The ancestral barriers are displayed as broken and slim lines.

10.4 Common barriers (Ctrl + I)

If checked, and a consensus reconstruction is selected, it display the barrier associate with the common allopatric set associated with the node (see under Tree window > Reconstruction menu) as a thin line. It uses a very conservative procedure, using the disjunctions associated with all nodes in the allopatric set, even if these nodes are not simultaneously supported.

References

[1] Fitch, 1971.
[2] Ronquist, 1997
[3] Hovenkamp, 1997
[4] Hovenkamp, 2001
[5] Arias et al., ms.
[6] Rosen, classic
[7] Page, 1994a. Component
[8] Page, 1994b. Tree map
[9] Ronquist, 1999. libro
[10] Nelson & Ladiges, 1996.
[11] Upchurch & Hunn, chronobiogeography
[12] Arias, in prep.