GPT: a web-server to map phylogenetic trees on a virtual globe

Pere Puigbò[1,*] and Jacqueline M. Major[2]

[1] National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD 20894, USA

[2] Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, FDA, Silver Spring, MD 20993, USA

* To whom correspondence should be addressed. Tel: 301-496-9196; Fax: 301-480-2288; Email: puigboap@ncbi.nlm.nih.gov

ABSTRACT

GPT (Global Positioning Trees) is a web-server that maps phylogenetic trees on a virtual globe. The minimum requirements are a phylogenetic tree and geographical coordinates of leaves to generate a Keyhole Markup Language (KML) file that can be viewed on Google Earth. An advantage of GPT is the results may be pre-visualized directly on the web. This web-server also implements several tools to display geolocation and geotrack data. GPT has been designed to be an easy-to-use tool to track evolutionary processes and will be useful for phylogeographical and spatial epidemiological studies. It covers a wide-range of visualizations divided in three components increasingly complex: geolocation, geotrack and GPT. This web-server is freely available at http://ppuigbo.me/programs/GPT and only requires Internet access, a web browser, and an earth browser able to read KML files. Several examples and a tutorial are accessible from the web-server's home page.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.840v1 | CC-BY 4.0 Open Access | rec: 17 Feb 2015, publ: 17 Feb 2015

INTRODUCTION

The term phylogeography was first coined in 1987 (Avise, et al., 1987) as the hybrid between population genetics and biogeography. During the last decade, the number of publications using phylogeographical analysis has increase substantially. Several articles have proposed new theoretical models/methods (Carstens, et al., 2009; Johnson and Crandall, 2009; Lemey, et al., 2009; Lemey, et al., 2010; Shapiro, et al., 2006) and tools (Hill, et al., 2009; Janies, et al., 2007; Shapiro, et al., 2006). Since Google launched in 2004 the program Google Earth, applications have been developed to map phylogenetic trees on a virtual globe (Bielejec, et al., 2011; Hill and Guralnick, 2010; Janies, et al., 2007; Krishnamurthy, et al., 2006; Maddison and Maddison, 2008; Parks, et al., 2009), which incorporate Global Positioning System (GPS) data. In this article, we introduce GPT (Global Positioning Trees), a user-friendly web-server designed to map phylogenetic trees on Google Earth. This unique web-server also includes several features designed for displaying GPS data in an earth browser allows inclusion of time information, an advantage over existing applications. In addition, GPT includes a Google Earth plug-in, which automatically loads generated data into the web-browser thereby facilitating fine-tuning of display settings. GPT is designed to be used by both novice and experienced users and can be utilized to track several evolutionary processes on a virtual globe, including migrations and speciation. GPT provides a valuable resource for spatial epidemiologists interested in tracking evolutionary dynamics (Zehender, et al., 2013) and geographic distribution of disease (Saito, et al., 2013), including outbreaks (Hill, et al., 2009; Janies, et al., 2007). This web-server is designed to cover a wide-range of visualizations for Graphical Information Systems (GIS). The first two visualizations are based on GPS data (geolocation and geotrack) and the third one (GPT) combines geographical locations and tree topologies.

PROGRAM OVERVIEW

Implementation and minimal requirements

GPT is an easy-to-use web-server that executes a Perl/CGI script. This is a cross-platform web-server that generates a Keyhole Markup Language (KML) file to be displayed in an earth browser such as Google Earth (http://earth.google.com). KML is a standard markup language focused on expressing notations and geographic visualizations on virtual maps and earth

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.840v1 | CC-BY 4.0 Open Access | rec: 17 Feb 2015, publ: 17 Feb 2015

browsers. The KML file generated by the GPT web-server can be pre-visualized directly on the web-browser or easily downloaded and displayed on Google Earth. The server has been tested in Windows, Macintosh, and several portable devices (iPhone and iPad). In the rare event the file generated cannot be displayed automatically, the server additionally provides all the information in a textbox that can be copied and modified. Currently the plug-in necessary to run Google Earth directly on the web is not available for iPad and iPhone.

Geolocation

The identification of the geographic location of species (geolocation) is the first layer of complexity (see green pins in Figure 1). Notice that in this article we use the word species as a unit or object of study, but may be substituted by the words genes, populations, events, objects, etc. The GPT web-server includes the possibility to easily visualize geolocation data. An option is to simply geographically allocate species in the virtual globe. In other words, users may map a set of species on the earth browser to visualize their distribution (Figure 2a). This option is available by filling the textbox for geographical coordinates as in (i) and leaving empty the textbox for the tree. In this initial version, the GPT web-server does not provide geographical coordinates, but they can be obtained elsewhere, e.g. GPS Visualizer (http://www.gpsvisualizer.com/geocoder).

i. A(41.46826580,2.22612210)
B(41.16695420,1.37522490)
C(42.5462450,1.6015540)
D(43.391160,5.5323519)

Users may also display geolocation data in a time-series of events by including time information as in (ii). In addition, more experienced users may plot area maps and 3D histograms in the final output (Figure 2b).

ii. A(2012-01-01T21:05:02Z)
B(2012-01-02T21:05:02Z)
C(2012-01-03T21:05:02Z)

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.840v1 | CC-BY 4.0 Open Access | rec: 17 Feb 2015, publ: 17 Feb 2015

D(2012-01-04T21:05:02Z)

Geotrack

The GPT web-server includes the possibility to visualize geotrack data (see blue line in Figure 1), adding a second layer of information to a GIS analysis. This option not only allows allocation of species in the virtual globe, but also adds information on how species spread. The GPT web-server makes the visualization of geotrack data a relatively easy task. The data must be inserted into the geotrack textbox in the format shown in (iii). A second example of geotrack visualization (Figure 2c) is available from the on-line tutorial (example number six)

iii. PATH(2012-01-01T21:05:02Z,2.22612210,41.46826580,100)
PATH(2012-01-02T21:05:02Z,1.37522490,41.16695420,100)
PATH(2012-01-03T21:05:02Z,1.6015540,42.5462450,100)
PATH(2012-01-04T21:05:02Z,5.5323519,43.391160,100)

Global Positioning Trees

Global Positioning Trees (GPT) is a method to position tree topologies (phylograms or dendrograms) in a virtual globe (Figure 1). This option adds even more information to GIS data analyses. Whereas the first two components of the GIS analysis (geolocation and geotrack) are useful to know where, when and how species distribute, the GPT can be used to visualize similarities among species. The GPT web-server minimally requires two inputs, a phylogenetic tree and geographical coordinates of the species to display a tree topology on Google Earth (Figure 2d-f). Once again, species may also refer to individuals, populations and objects. The input phylogenetic tree must be binary and in standard newick format (single line), including branch lengths (iv) or only the tree topology (v). Species names in the tree may contain only alphanumerical characters. Geographical coordinates of species (terminal branches of the tree) must be inserted in separate lines as shown in (i).

iv. ((A:1.0,B:1.2):1.1,(C:0.6,D:0.4):0.3);

v. ((A,B),(C,D));

Additional features

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.840v1 | CC-BY 4.0 Open Access | rec: 17 Feb 2015, publ: 17 Feb 2015

GPT contains several options and parameters that may be modified by more experienced users. There is a textbox to include a brief description of the species in the format shown in (vi) and an option to modify the title for the KML file. Descriptions are incorporated to the final display and they may be helpful to find species within a large and complex tree.

vi. A(Description_Species_A)

B(Description_Species_B)

C(Description_Species_C)

D(Description_Species_D)

In order to illustrate wide range of phylogenetic trees, the parameters scale and ground have been adjusted to 100,000 and 5, respectively, but users may modify these parameters. The parameter scale converts branch lengths to altitude; whereas ground is a parameter to prevent lines going underground when crossing the virtual globe (Figure 2f). The pre-visualization of the tree directly on the web allows users to easily modify colors and thickness of branches.

Conclusion

The GPT web-server is a simple, yet powerful, tool for a wide-range of phylogeographical and spatial epidemiological studies. The web-server allows mapping of phylogenetic trees, geolocation and geotracking of events, and depicting magnitude through use of bar graphs on the earth browser; all of which can be visualized directly on the web-browser. Whereas GPS data provides an objective measure for detecting spatial and temporal variations, GPT provides a powerful new resource for visualization and understanding this complex multi-dimensional data.

ACKNOWLEDGEMENTS

PP is supported by intramural funds of the US Department of Health and Human Services (NLM, NIH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views expressed are the authors and not necessarily those of FDA, NIH or DHHS.

FIGURES

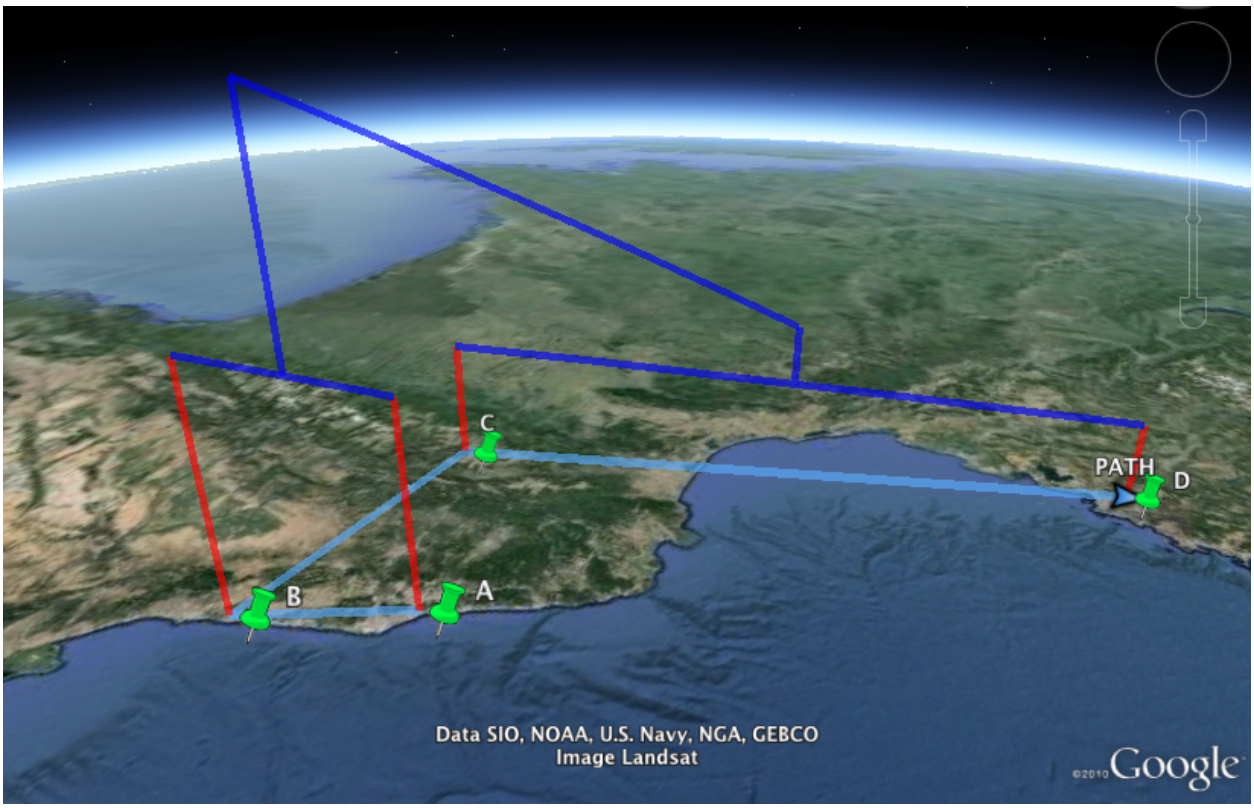

Figure 1 – Graphical Information Systems (geolocation, geotrack and GPT)

The GPT web-server includes a third layer of display to increasing the versatility of Graphic Information Systems (GIS) analysis. The first layer, geolocation (green pins with labels) shows the presence of species in geographical locations. The second layer, geotrack, contains information of the real path from one location to another. The third layer, GPT, depicts similarities and differences among species. In this example, species are present (geolocation) in 4 geographical locations (Andorra, Barcelona, Marseille and Tarragona). The geotrack layer shows that the spread occurred as follows: Barcelona → Tarragona → Andorra → Marseille. The GPT shows two main clusters: 1) Barcelona and Tarragona; 2) Andorra and Marseille.

Map data: Google, Landsat.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.840v1 | CC-BY 4.0 Open Access | rec: 17 Feb 2015, publ: 17 Feb 2015

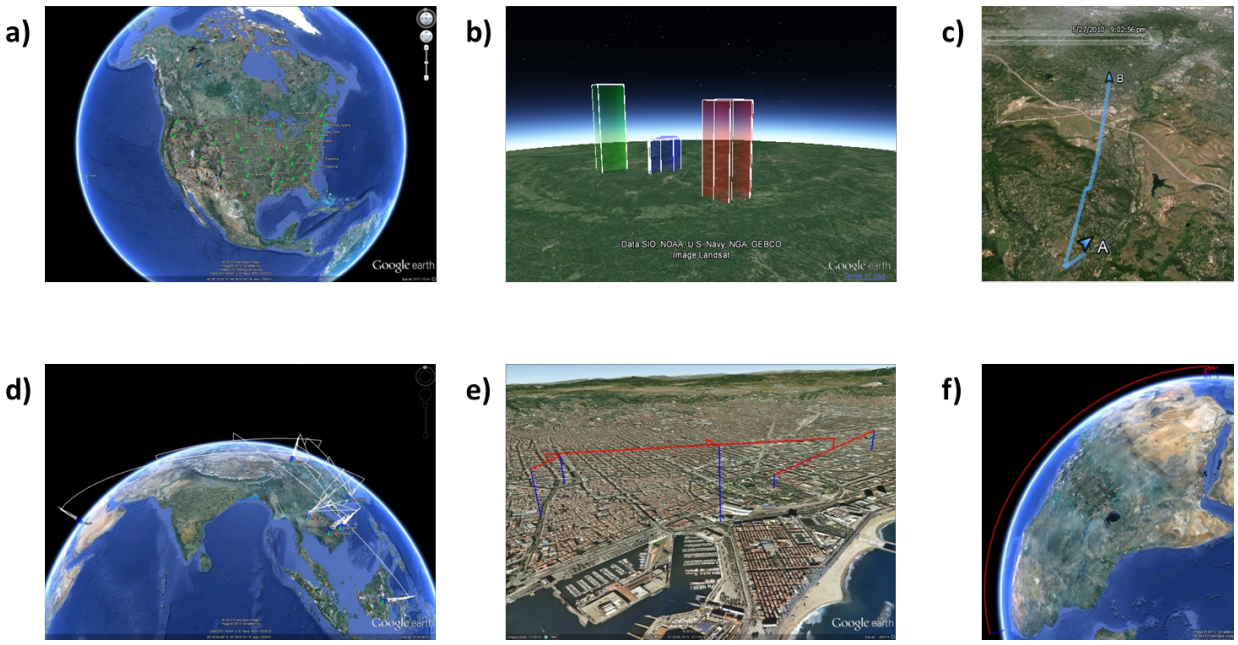

Figure 2 – Examples of six displays available in GPT

a) Geolocalization

b) 3D bar graphs on the virtual globe

c) Geotrack

d) An earth-wide phylogenetic tree displayed on the virtual globe

e) A five-leaf tree displayed on a small region of the virtual globe

f) A Phylogenetic tree with horizontal branches follows the curvature of the earth.

Map data: Google, Landsat

REFERENCES

Avise, J.C., *et al.* (1987) Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. Annual Review of Ecology and Systematics, pp. 489-522.

Bielejec, F., *et al.* (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics, *Bioinformatics*, **27**, 2910-2912.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.840v1 | CC-BY 4.0 Open Access | rec: 17 Feb 2015, publ: 17 Feb 2015

Carstens, B.C., Stoute, H.N. and Reid, N.M. (2009) An information-theoretical approach to phylogeography, *Mol Ecol*, **18**, 4270-4282.
Hill, A.W. and Guralnick, R.P. (2010) GeoPhylo: an online tool for developing visualizations of phylogenetic trees in geographic space. Ecography.
Hill, A.W.*, et al.* (2009) Evolution of drug resistance in multiple distinct lineages of H5N1 avian influenza, *Infect Genet Evol*, **9**, 169-178.
Janies, D.*, et al.* (2007) Genomic analysis and geographic visualization of the spread of avian influenza (H5N1), *Syst Biol*, **56**, 321-329.
Johnson, J.B. and Crandall, K.A. (2009) Expanding the toolbox for phylogeographic analysis, *Mol Ecol*, **18**, 4137-4139.
Krishnamurthy, N.*, et al.* (2006) PhyloFacts: an online structural phylogenomic encyclopedia for protein functional and structural classification, *Genome Biol*, **7**, R83.
Lemey, P.*, et al.* (2009) Bayesian phylogeography finds its roots, *PLoS Comput Biol*, **5**, e1000520.
Lemey, P.*, et al.* (2010) Phylogeography takes a relaxed random walk in continuous space and time, *Mol Biol Evol*, **27**, 1877-1885.
Maddison, D.R. and Maddison, W.P. (2008) Cartographer, a Mesquite package for plotting geographic data Version 1.3.
Parks, D.H.*, et al.* (2009) GenGIS: A geospatial information system for genomic data, *Genome Res*, **19**, 1896-1904.
Saito, M.*, et al.* (2013) Genetic diversity and geographic distribution of genetically distinct rabies viruses in the Philippines, *PLoS Negl Trop Dis*, **7**, e2144.
Shapiro, L.H., Strazanac, J.S. and Roderick, G.K. (2006) Molecular phylogeny of Banza (Orthoptera: Tettigoniidae), the endemic katydids of the Hawaiian Archipelago, *Mol Phylogenet Evol*, **41**, 53-63.
Zehender, G.*, et al.* (2013) Reconstruction of the evolutionary dynamics of hepatitis C virus subtypes in Montenegro and the Balkan region, *Infect Genet Evol*, **17C**, 223-230.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.840v1 | CC-BY 4.0 Open Access | rec: 17 Feb 2015, publ: 17 Feb 2015