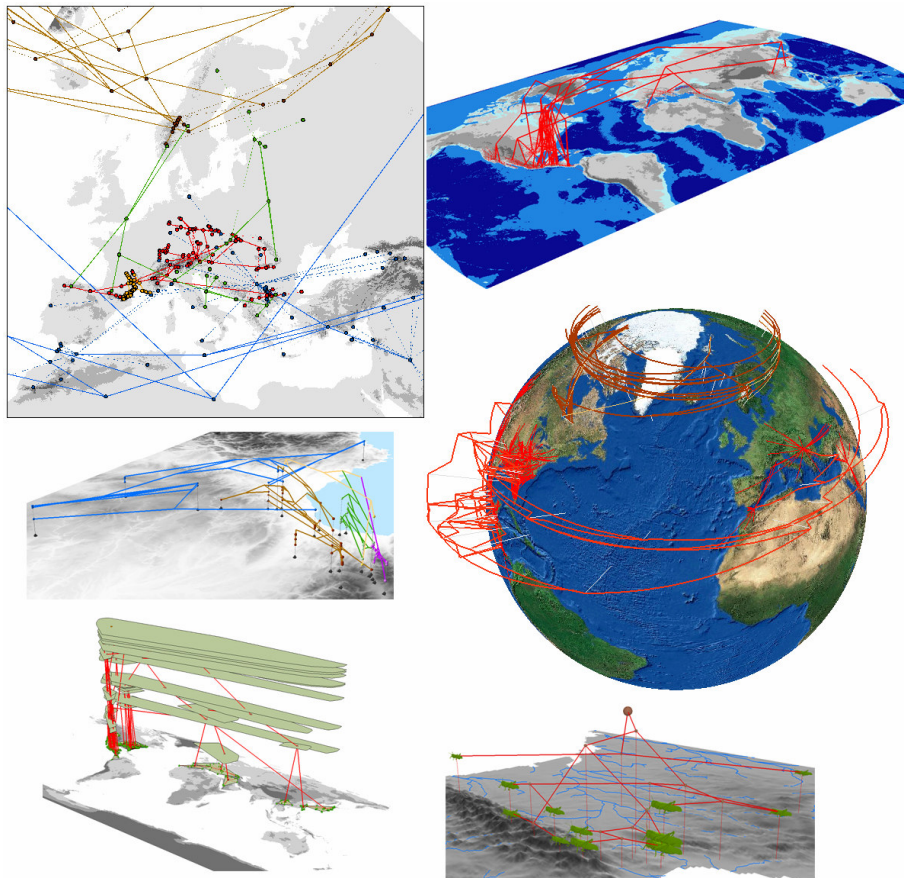


GeoPhyloBuilder v1.0

for ArcGIS



David Kidd & Xianhua Liu

**National Evolutionary Synthesis Center,
Suite A200, 2024 West Main Street, Durham, NC 27705**

2007

GeoPhyloBuilder

Copyright (C) 2007, David Kidd, dk@nescent.org.

Copyright (C) 2007, Xianhua Liu, xl24@duke.edu.

Copyright (C) 2007, National Evolutionary Synthesis Center (NESCent).

This library is free software; you can redistribute it and/or modify it under the terms of the GNU Lesser General Public License as published by the Free Software Foundation; either version 2.1 of the License, or (at your option) any later version.

This library 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 Lesser General Public License for more details.

You should have received a copy of the GNU Lesser General Public License along with this library; if not, write to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

Supported by the National Evolutionary Synthesis Center, (NESCent)

NSF #EF-0423641

Contents

Section	Page
Introduction	1
Installation	1
Help and Feedback	1
Geophylogenies	2-3
Building a geophylogeny with GeoPhyloBuilder	4-7
Example Data	7
Defining Subclades	8
GeoPhyloBuilder Feature Classes	8-10

Introduction

Population genetics, phylogeography and historical biogeography combine information on the geographical distribution of biotic variation between individuals, populations, species or higher taxonomic units with information on their relatedness to identify biogeographical pattern and hence infer spatiotemporal scenarios. GIS integrate disparate data sets through common georeferencing. GeoPhyloBuilder creates a spatial phylogeographical network model from an input phylogenetic model and associated spatial data of observations. Version 1.0 supports only tree models, however, future releases will support phylogenetic networks and pairwise distance matrices.

GeoPhyloBuilder is an extension to [ESRI's ArcGIS geographical information system \(GIS\)](#) that builds a spatiotemporal phylogeographic GIS model, a 'geophylogeny', from a tree and a set of geographical features.

GeoPhyloBuilder is written in .NET using ESRI's ArcObjects library and a custom library of geophylogenetic COM objects.

GeoPhyloBuilder GUI and COM library were developed at the National Evolutionary Synthesis Centre by David Kidd (dk@nescent.org) and Xianhua Liu. Source code is freely available on request. We welcome collaboration with users and developers.

For more on GIS applications in evolutionary science see,

- [Kidd DM, Ritchie MG \(2006\) Phylogeographic information systems; Putting the geography into phylogeography. Journal of Biogeography 33: 1851-1865.](#)
- [Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, et al. \(2006\) Putting the 'landscape' in landscape genetics. Heredity : 1-15. doi: 10.1038/sj.hdy.6800917](#)

This User Manual describes the GeoPhyloBuilder 1.0 data model, functionality and operation. We assume a reasonable working knowledge of ArcGIS, however, it is intended that in time more in depth instructions and case studies will become available on the [EvoViz Wiki](#). Some example data sets are provided.

Installation

To install double-click *GeoPhyloBuilder.msi* and follow instructions.

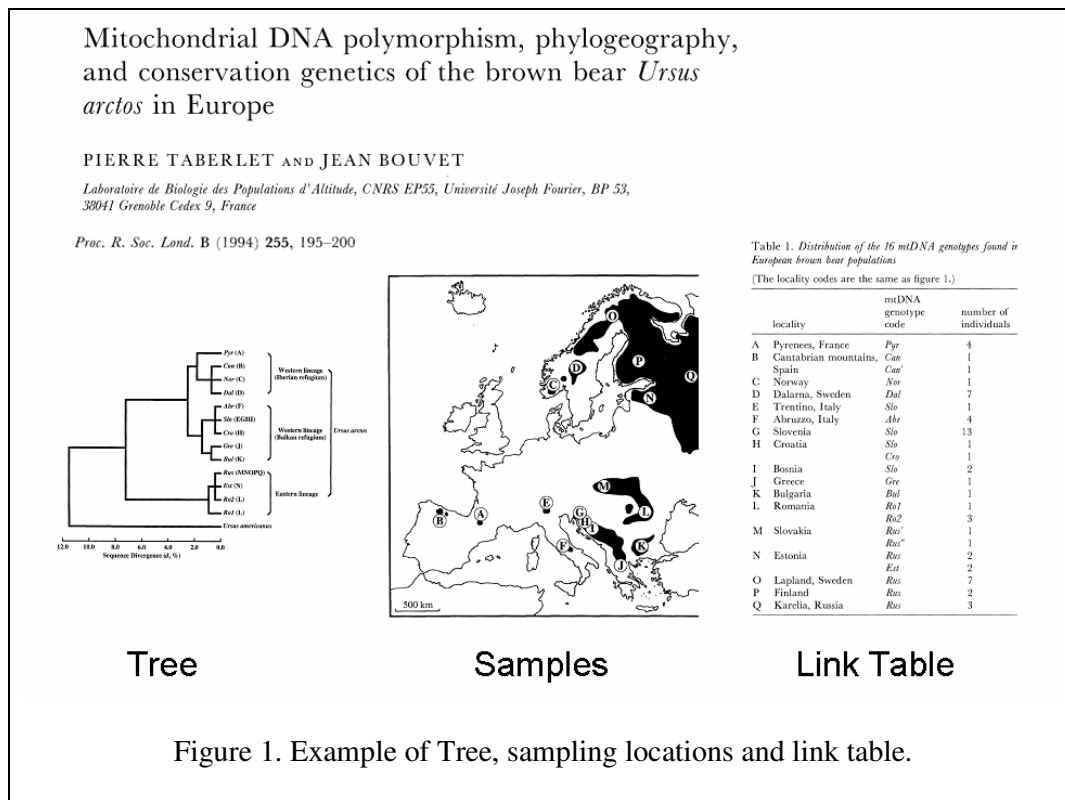
Help and Feedback

Further guidance for using GeoPhyloBuilder as well as viewing and analysing geophylogenies in ArcMap and ArcScene is available through the [GeoPhyloBuilder](#) pages of the [EvoViz Wiki](#). The EvoViz Wiki is a general source of information on visualizing evolution. Bugs and help requests can also be sent to geophylobuilder@nescent.org.

GeoPhylogenies

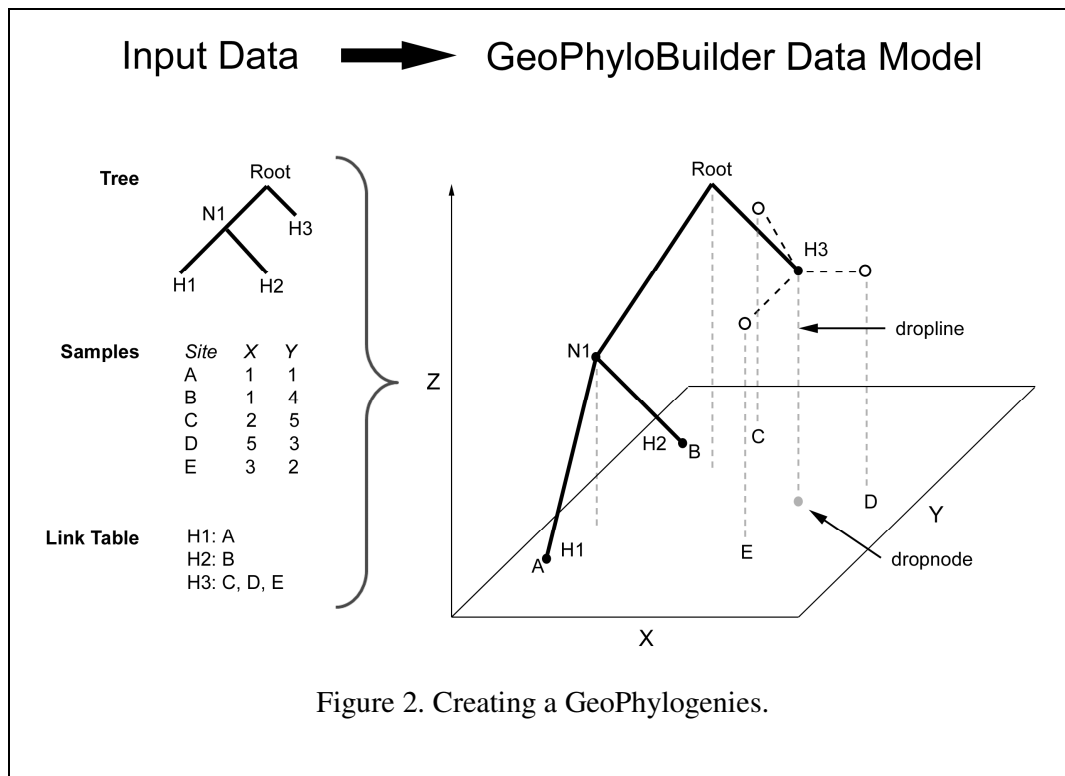
Spatial evolutionary data sets (fig. 1) consist of;

1. *Entities* sampled in space; individuals populations or species
2. *Phylogenetic models* (tree, reticulate network or pair wise graphs) defining ‘relatedness’ between entities.
3. A *link table* defining n:m relationships between entities and phylogenetic model nodes.



GeoPhyloBuilder creates a spatiotemporal GIS network data model from a [NEWICK](#) string defining a tree and an ArcGIS Geodatabase feature class or shapefile defining sample locations. An optional ‘link table’ defines n:m relationships between tree-tips and sample locations (fig. 2). Sample locations may be points, polylines or polygons. The NEWICK string for the tree in fig. 2 is; ‘((H1,H2),H3);’

To create a geophylogeny the location of phylogenetic nodes observed at multiple locations and those inferred from the data but not observed must be estimated. GeoPhyloBuilder v1.0 locates nodes with multiple observations nodes at the spatial centroid of the observations. Similarly, inferred nodes are located at the spatial centroid of the set of nodes one level below in the hierarchy. Other node positioning algorithms, for example, using historical biogeographical reconstruction algorithms, niche models or fossil evidence will be implemented in future releases.



We have a rooted tree with three tip entities (H1-H3) and two inferred nodes (Root & N1) (fig. 2). H3 was observed at three point locations while H1 & H2 were observed at single locations. The geophylogeny tip H1 is positioned at the geographic location of A and H2 at B. H3 was observed at three locations (C-E) so tip H3 is positioned at the centroid of the envelope of C, D and E and ‘observation to tip’ branches between the observations and tip created established. The geographical location of internal inferred node N1 can now be positioned at the centroid of the envelope of H1 and H2 and subsequently the Root is positioned at the centroid of the envelope of N1 and H3.

If the samples are polylines then nodes are derived from the polyline midpoint and if polygons then the polygon centroid.

Phylogenetic distances are assigned as attributes of the geophylogeny branches and nodes. Distances from the root are also assigned to the z-attribute of the ArcGIS network for easy 3D visualization in ArcScene. ‘Droplines’ are design elements that connect geophylogeny nodes to other geographical entities, e.g. a base map, positioned at other z-values. ‘Dropnodes’ are duplicate point objects that can be used to mark the position of nodes on the base map.

Any node of an unrooted tree can be set as the root and the tree ‘suspended’ in 3D from the selected root as desired.

The geographical position of a tree root, a ‘*spatial root*’, can be specified. If set inferred nodes are positioned at the centroid of the envelope of the normal set locations *and* the spatial root.

Building a GeoPhylogeny with GeoPhyloBuilder

1. *Run software.*

Start > All Programs > GeoPhyloBuilder > GeoPhyloBuilder

2. *Set Tree.*

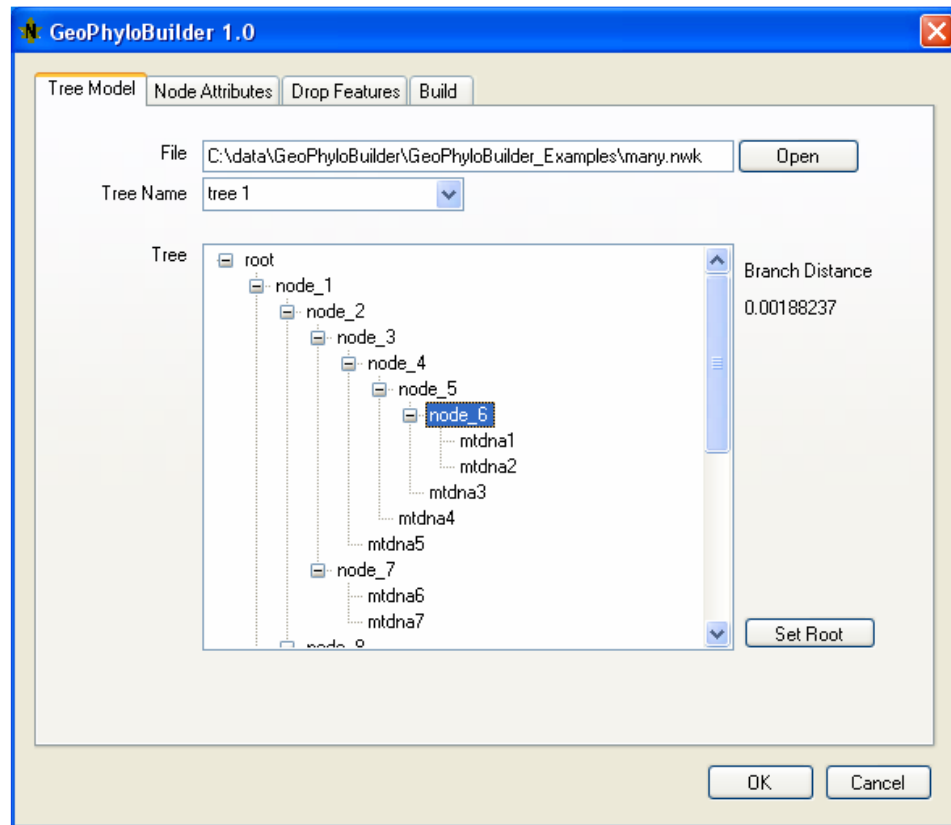


Figure 3. Selecting the Tree Model.

Select ASCII file containing NEWICK string. If more than one tree exists in the file select the desired model from the dropdown 'Tree Name' combobox. The selected tree is displayed in the 'Tree Window'. To re-root an unrooted tree select the node to be designated the root and 'Set Root'. The tree will be redrawn with the selected node set as the new root.

3. Set Tree Attributes.

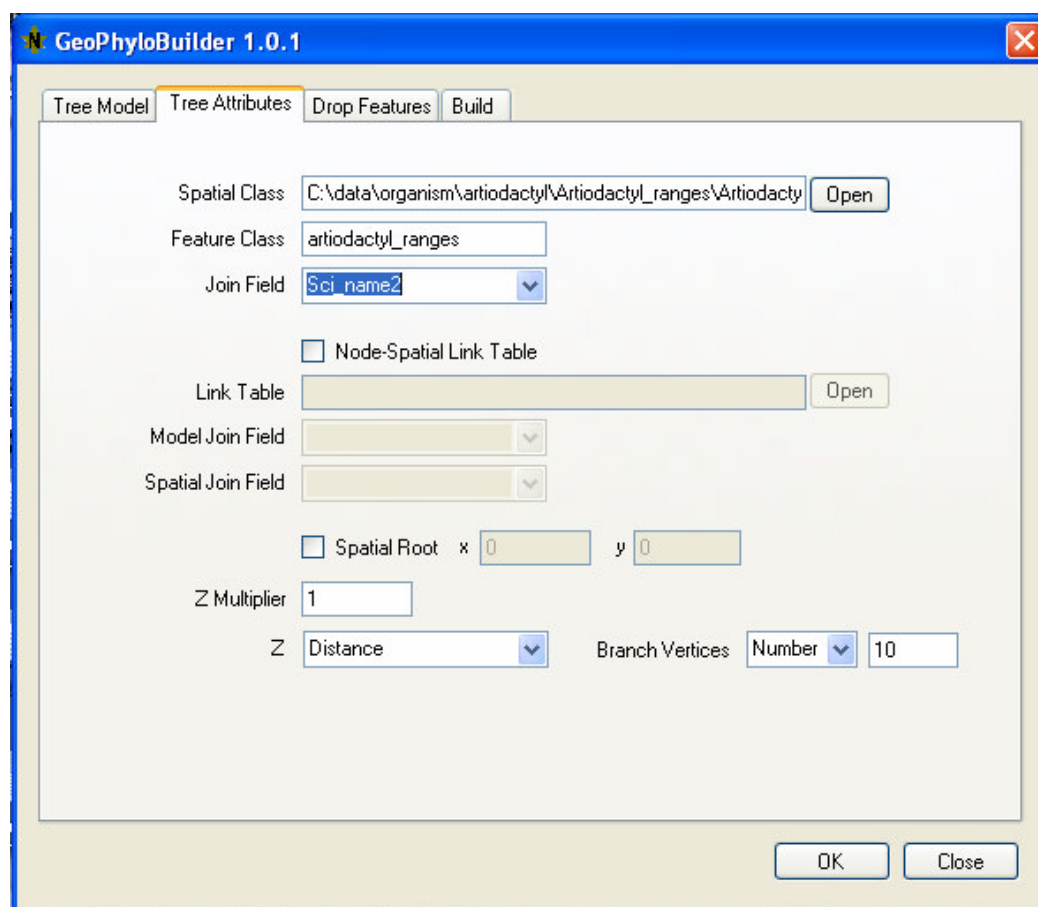


Figure 4. Selecting the Node Attributes.

Select spatial class that contains the sampled features. If there are no n:m relationships defined then the tree can be simply linked to the features by selecting the 'Join Field' of the feature table that stores the names of the terminal tips.

If a link file defining n:m relationships between tips and features exists then check the 'Node –Spatial Link Table' checkbox and select the link table name. Now select the 'Model Join Field' that joins the tree tips to the link table and the 'Spatial Join Field' that joins the link table to the spatial feature class. The link table must be an Access database table.

A 'Spatial Root' can be optionally set as can the 'Z multiplier'. The 'Z multiplier' multiplies tree distances by the value specified to aid 3D visualization. For example, if an entire tree has a depth of 1 unit is displayed on a map with meters as the base unit then the tree will by default only be raised to 1 meter above the ground surface, i.e. not very far. Note that a z-elevation multiplier can be also be set in ArcScene on the 'Base Heights' tab of a features properties.

Branch Vertices sets the number of vertices to break each branch into. Branches that are not broken into vertices remain as straight line when projected into other coordinate systems.

4. *Set Drop Features.*

Select whether to create droplines and/or drop features. Features will be ‘dropped’ to the value of ‘Drop to Z’.

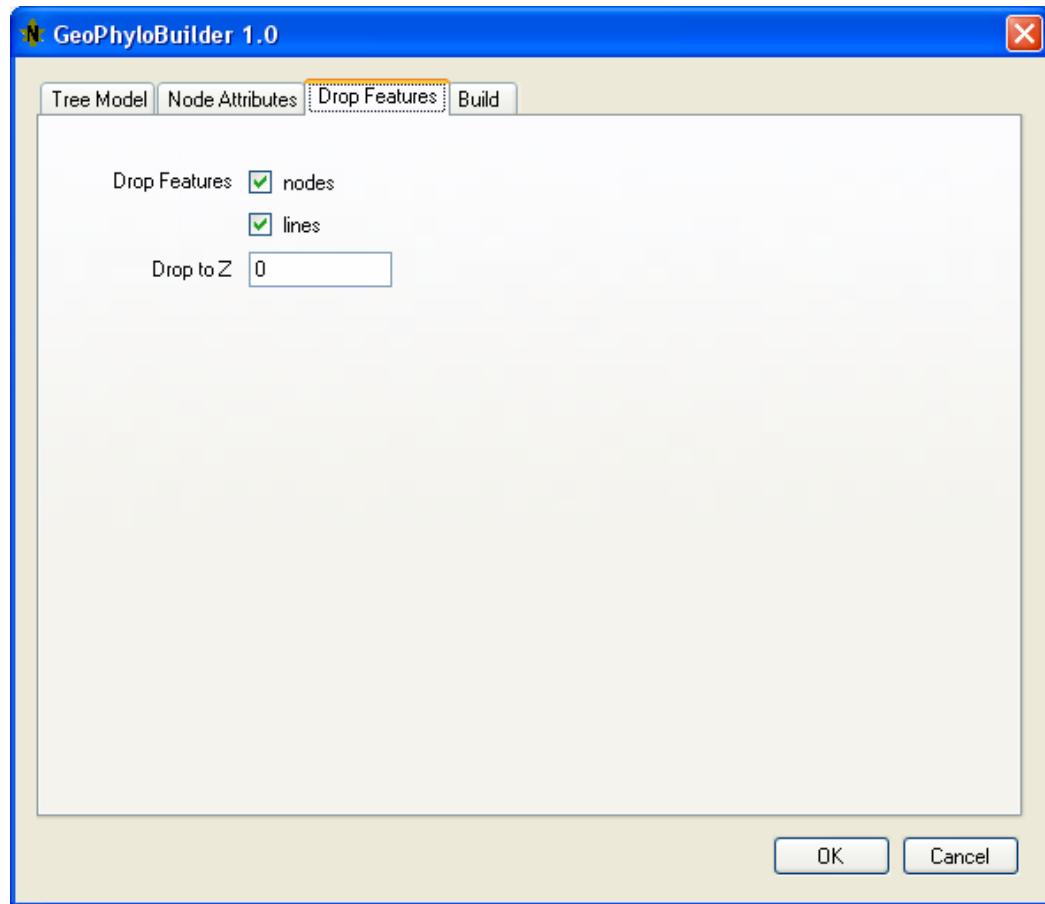


Figure 5. Selecting the Drop Features.

5. Build the Geophylogeny

Select ‘Import Type’ as geodatabase, shapefile or KML (Google Earth, NASA WorldWind, etc. **Note KML output does not presently support branch densification.** Select the location of the geodatabase or shapefile directory. Input the ‘Dataset Name’ which is the name of the geodatabase feature dataset and the prefix of the feature classes created within it. If shape files are output then the data set name is the prefix of the shape file name (table 1).

Object	Geodatabase		Shapefile
	Dataset	Feature Class	
nodes	mytree	mytree_node	mytree_node
branches	mytree	mytree_branch	mytree_branch
dropline	mytree	mytree_dropline	mytree_dropline
dropnode	mytree	mytree_dropnode	mytree_dropnode

Table 1. GeoPhyloBuilder Output with Dataset Name set to “mytree”

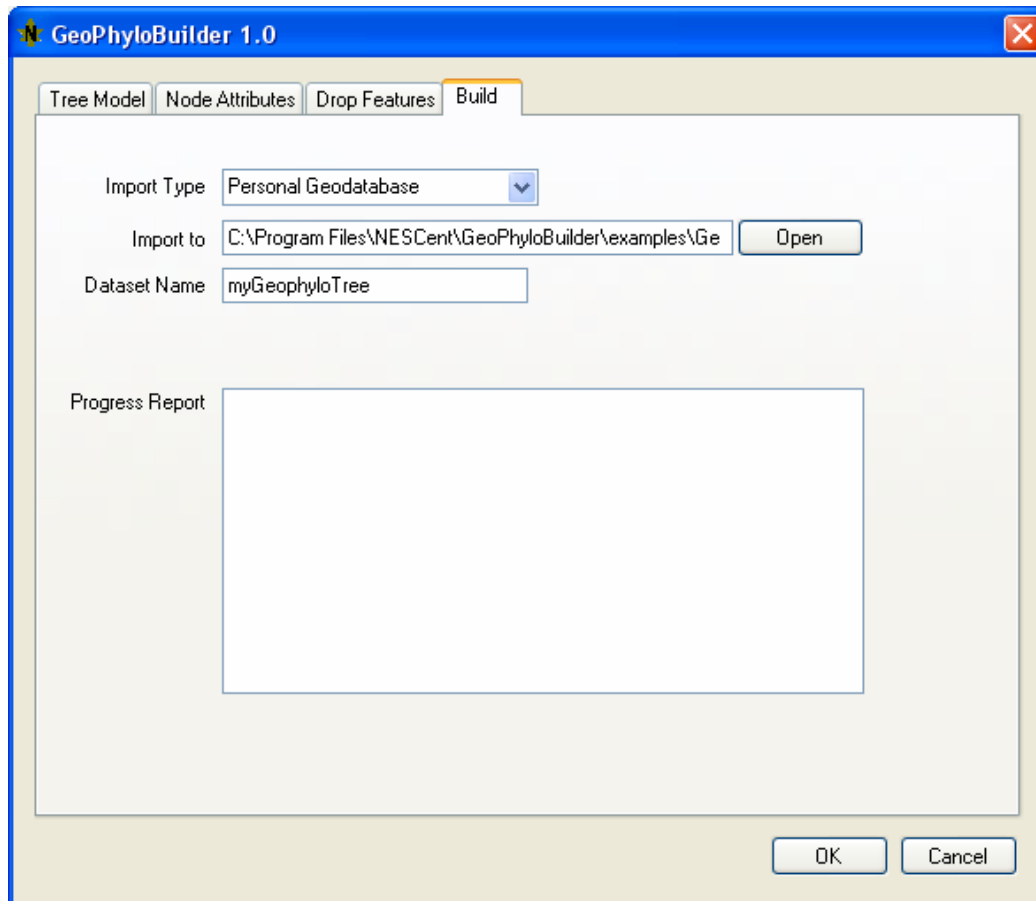


Figure 5. Setting the Build Options.

Example Data

Folder GeoPhyloBuilder_Examples contains two example .nwk files (one.nwk and many.nwk) and a geodatabase containing six feature classes (sites_pts_one, sites_pts_many, sites_lines_one, sites_lines_many, sites_poly_one, sites_poly_many) and a link table 'LinkFile'. For spatial context a digital elevation model for the area covered by the example datasets clipped from <http://www.ngdc.noaa.gov/mgg/topo/globe.html> is included.

Use one.nwk with sites_points_one, sites_lines_one & sites_poly_one to create geophylogenies without a link table joining on the 'indivref' field.

Use many.nwk with sites_points_many, sites_lines_many & sites_poly_many to create geophylogenies without a link table joining on the 'indivref' field.

Defining Subclades

RankId's provide a quick way to select and code subclades through 'path enumeration'. Each node in a tree is coded according to the following schema. The root is designated 0. All nodes at the level below the root are then coded 0.0, 0.1, 0.2, etc. Subsequent divisions are coded through the addition of an extra dot and number hence, 0.0.1, 0.0.2 and 0.0.3 are three nodes below 0.0, and 0.2.0 and 0.2.1 are two nodes below 0.2 (fig 6).

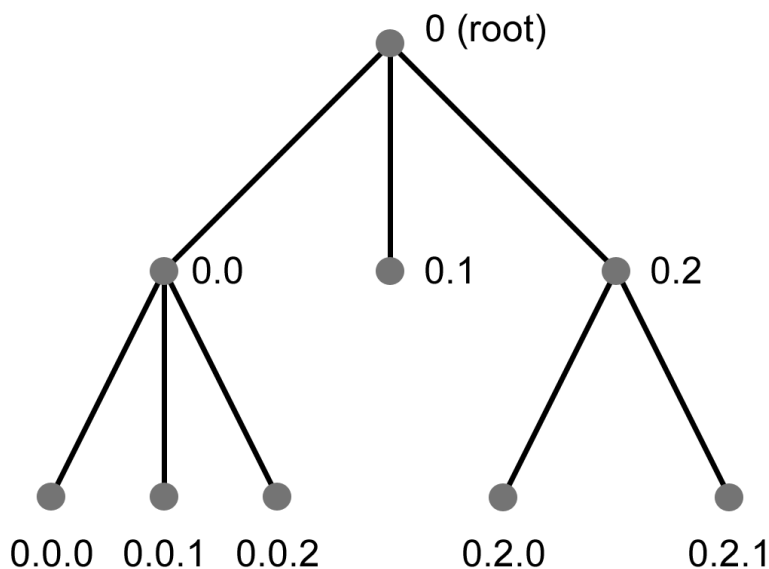


Figure 5. Node RankId.

To code subclades create a new field in the node and branch feature classes, e.g. 'mysubclade'. Select subclade nodes by 'selecting by attributes' where the select string is of the form *SELECT * FROM tree.tree_node WHERE rankid like '0.2*'* that will select the node with rankid = 0.2 and all nodes below it. The 'mysubclade' field

can now be updated to a code describing the subclade group, e.g. ‘mysubclade1’. A similar logic can be applied to select the branches of subclades using the FromRankID or ToRankID fields.

GeoPhyloBuilder Feature Classes

Features are Nodes, Branches, DropNode, DropLine, & NodePolygon (not implemented in v1.0).

1. Nodes

Idx	Field Name	Type	Description
0	OID	Integer	Internal Object Identifier
1	Shape	Point	Geometry
2	NodeID	Integer	User defined unique node identifier
3	NodeName	String(30)	Name of Branch
4	NodeType	Long Integer	The type of node; 0 = (optional) root 1 = tip (single observation or pseudo-observation) 2 = inferred internal (except the optional root) 3 = terminal observation
5	NodeLevel	Integer	Level of node from root where root is level 0
6	NodeDepth	Double	Distance of node from upstream node
7	NodeCumDepth	Double	Cumulative distance of node from root
8	RankID	String (254)	Node rank; root 0; first level nodes 0.0, 0.1, 0.2, etc; second level nodes 0.0.0,0.0.1,0.1.0, etc

2. Branches

Idx	Field Name	Type	Description
0	OID	Integer	Internal Object Identifier
1	Shape	Polyline	Geometry
2	BranchID	Integer	User defined unique node identifier
3	BranchName	String(30)	Name of Branch
4	FromNodeID	Integer	Upstream (toward root) nodeID
5	ToNodeID	Integer	Downstream (toward tip) nodeID
6	BranchType	Integer	The type of branch; 1 = phylogenetic model branch 2 = observation to tip branch
7	PhyloLength	Double	Phylogenetic length of branch
8	GeoLength	Double	Geographic length of

			branch
9	FromRankID	String(254)	RankID of FromNode
10	ToRankId	String(254)	RankID of ToNode

3. DropNode

Idx	Field Name	Type	Description
0	OID	Integer	Internal Object Identifier
1	Shape	Point	Geometry
2	DropNodeID	Integer	User defined unique node identifier
3	DropNodeName	String(30)	Name of Drop Node
4	DropNodeType	Integer	The type of node DropNode is for; 1 = tip (single observation or pseudo-observation) 2 = inferred internal (except the optional root) 3 = (optional) root 4 = terminal observation
5	NodeID	Integer	NodeID from Node FClass

4. DropLine

Idx	Field Name	Type	Description
0	OID	Integer	Internal Object Identifier
1	Shape	Polyline	Geometry
2	DropLineID	Integer	User defined unique line identifier
3	DropLineName	String(30)	Name of DropLine
4	DropLineType	Integer	The type node DropLine is from; 1 = tip (single observation or pseudo-observation) 2 = inferred internal (except the optional root) 3 = (optional) root 4 = terminal observation
5	NodeID	Integer	NodeID from Node FClass

5. NodePolygon (not implemented in v1.0)

Idx	Field Name	Type	Description
0	OID	Integer	Internal Object Identifier
1	Shape	Polygon	Geometry
2	NodePolyID	Integer	User defined unique polygon identifier
3	NodePolyName	String(30)	Name of NodePolygon
4	NodePolyType	Integer	The type node NodePoly is for; 1 = tip (single observation or pseudo-observation) 2 = inferred internal (except the optional root)

			3 = (optional) root 4 = terminal observation
5	NodePolyLevel	Integer	Level from root (0) of NodePoly
6	NodePolyDepth	Double	Depth of NodePoly from upstream node
7	NodePolyCumDepth	Double	Cumulative depth from root
8	NodeID	Integer	NodeID from Node FClass