

# CoRe-PA - Softwaretool for Cophylogeny Reconstruction: A Short Tutorial

Daniel Merkle<sup>1</sup>, Martin Middendorf<sup>2</sup> and Nicolas Wieseke<sup>2\*</sup>

<sup>1</sup>Department of Mathematics and Computer Science  
University of Southern Denmark, Odense, Denmark  
`daniel@imada.sdu.dk`

<sup>2</sup>Parallel Computing and Complex Systems Group  
Department of Computer Science  
University of Leipzig, Germany  
`{middendorf, wieseke}@informatik.uni-leipzig.de`

November 16, 2009

## Abstract

This is a short tutorial for the cophylogeny reconstruction tool CoRe-PA which can be downloaded at <http://pacosy.informatik.uni-leipzig.de/pv/Software/CoRe-PA/CoRe-PA-index.engl.html>.

CoRe-PA is a softwaretool that can be used to infer the common history of coevolutionary systems, e.g., hosts and their parasites or insect-plant relations. It uses an event-based concept for reconciliation analyses where the possible events are cospeciations, sortings, duplications, and (host) switches. One feature of CoRe-PA is a new parameter-adaptive approach, i.e., no costs have to be assigned to the coevolutionary events in advance. With CoRe-PA it is possible to

- design host parasite scenarios with a graphical editor
- generate random coevolutionary scenarios using the beta-split model with beta 0, -1 or -1.5
- generate random coevolutionary scenarios by simulating coevolution

---

\*This work was supported by the German Research Foundation (DFG) through the project “Deep Metazoan Phylogeny” within SPP 1174.

- generate random coevolutionary scenarios which “keep the characteristics” of given host parasite systems
- handle non-binary host and parasite phylogenies
- choose between different behaviors of how to handle host switches
- use divergence timing information
- compute the best reconstructions for a given event cost vector
- compute the best cost vector for a given host parasite system (where the cost vector fits best to the reconstructed event frequencies)
- do randomization tests for given host parasite systems to analyze the probability of coevolution
- export host parasite scenarios and their reconstructions to svg

More information about CoRe-PA can be found in [1]. The following is a short tutorial on how to use CoRe-PA.

## Start CoRe-PA

To use CoRe-PA just download the platform specific archive and unzip the file to a local folder. Then start the core-pa executable. Please note that a Java Runtime Environment version 1.5 or later has to be installed on your system. You will see CoRe-PA’s main window with the menu and tool bar on the top, the navigator and the property views on the left, and the empty editor part on the right.

As CoRe-PA is based on the Eclipse RCP framework it has the same layout functionalities. Therefore by using drag and drop you can move the views and the editor to any position you like.

## Create a project

CoRe-PA has also derived the Eclipse project structure. So before you can start you have to add a project to your workspace. This is done by clicking File → New → Project in the menu bar. The “New Project” dialog will open (the same can be done by clicking New → Project in the context menu of the “Navigator” view).

In the “Project name”-field specify the name of your new project, e.g. “Samples” and click finish. In the “Navigator” view a folder with the specified project name is shown (Figure 1).

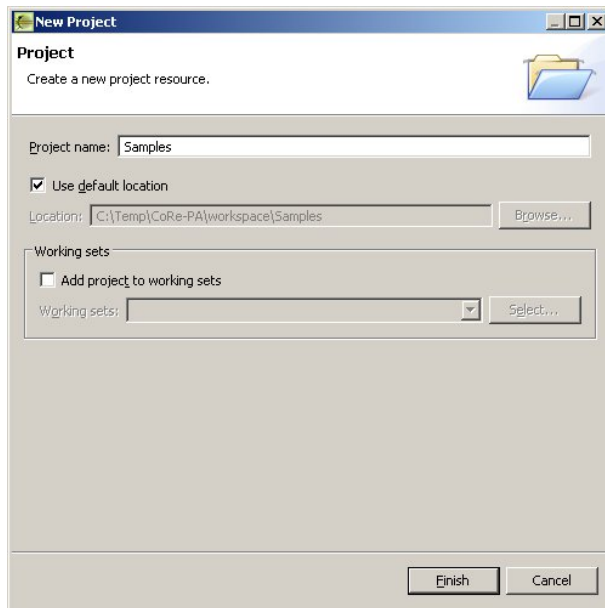


Figure 1: The “New Project” dialog.

## Import files

The easiest way to test CoRe-PA is to use the sample data which can be downloaded from the page <http://pacosy.informatik.uni-leipzig.de/pv/Software/CoRe-PA/CoRe-PA/samples.zip> To import the file click File → Import in the menu bar (or use the “Import” item in the context menu of the “Navigator” view). Select General → Archive File in the “Import” dialog and click “Next” (Figure 2 top). Select the downloaded samples.zip file in the “From archive file”-field, choose the folder of your recently created project as “Input folder” and click finish (Figure 2 bottom). There should be six new .nex files in your sample project.

## Create new files

Another way is to create a new nexus file with File → New → Nexus File. There you have to select the project folder, specify a filename and click finish. The new file is opened immediately and the editor shows only the root nodes of the host and the parasite tree. You can move a node by clicking on that node and moving the mouse to the designated position.

## Using the editor

A .nex file is opened manually by double clicking that file (e.g. seabirds-chewinglice.nex) in the “Navigator” view. In the editor you will see the vi-

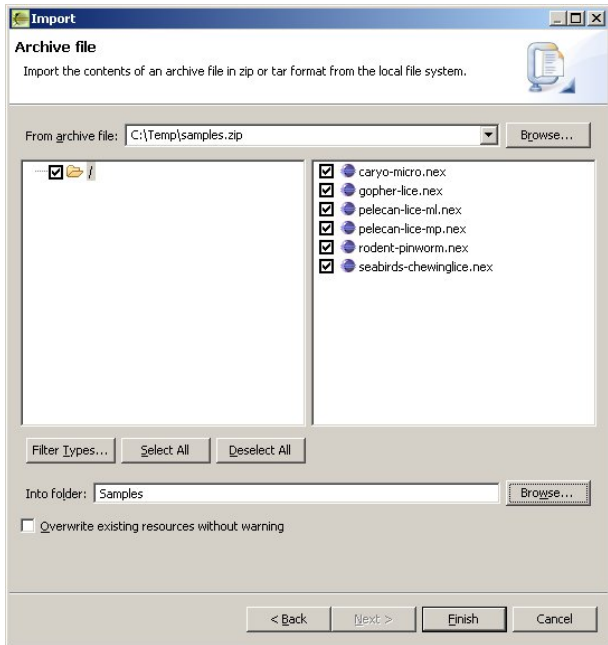
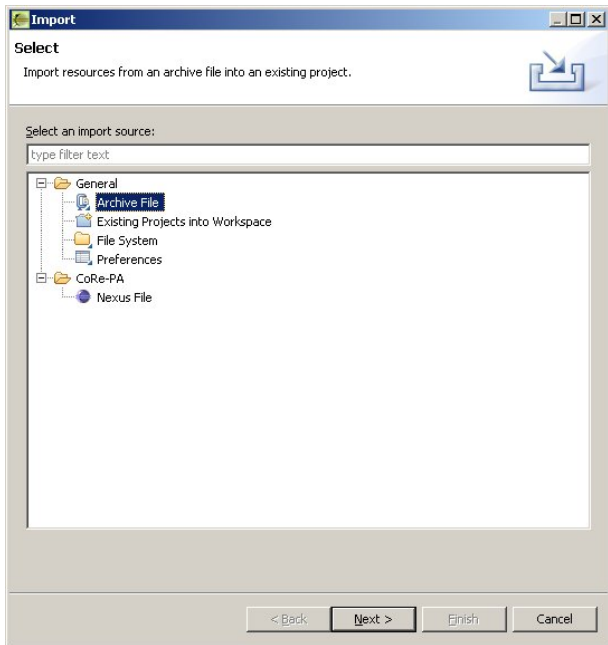


Figure 2: Top: the “Import” dialog; bottom: the “Archive File” dialog.

sualization of the specified host parasite system with the host tree on the left and the parasite tree on the right (Figure 3).

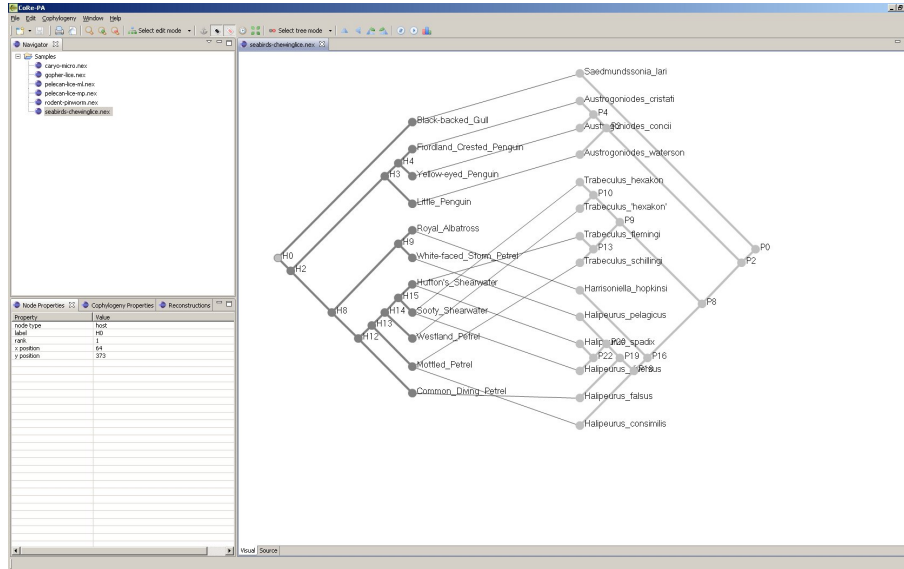



Figure 3: Screenshot of a coevolutionary seabirds-chewinglice system.

Additionally the tool bar was extended. There you find the actions to manipulate the appearance of the visualization and to set-up and start the reconstruction process. See Table 1 for a details.

To edit the copylogenetic data you can right-click at a node and select one of the node actions which are described in Table 2. The same behavior can be achieved by selecting the designated action as edit mode event in the tool bar and double-click on the respective node.

## Reconstruction properties

By clicking the  button in the tool bar you can change the cost values and reconstruction options (Figure 4).

The costs can be set as double values or probability values for each event (cospeciation, sorting duplication or switch) and there are several options which specify the exact behaviour during the reconstruction process.

You can specify which type of switches are permitted (HOSTSWITCH) and which additional sortings (SORTING) and duplications (DUPLICATION) occur during a switch. It is also possible to delay switches (TAKEOFF) after an appropriate speciation by inserting sortings and to allow full switches (FULL-HOSTSWITCH) where no parasite child remain in the subtree of the related host speciation. For host parasite systems it is suggested to use the default options.




















Icon	Description
	Print the editor's content
	Export the editor's content to SVG
	Reset the zoom
	Zoom in
	Zoom out
	Select the double click event for the host and parasite nodes
	Fixing the position of the descendant nodes if a node is moved
	Show/hide host labels
	Show/hide parasite labels
	Show/hide timing information in addition to the labels
	Rearrange the position of all host and parasite nodes
	Select the tree which should be changed by the following actions
	Flip the tree(s) horizontally
	Flip the tree(s) vertically
	Rotate the tree(s) anti-clockwise
	Rotate the tree(s) clockwise
	Set up the preferences for the cophylogenetic reconstruction
	Compute the cost minimal reconstruction for this scenario
	Compute the cost minimal reconstructions for "similar" random scenarios

Table 1: Actions for changing the appearance of the visualization and set-up respectively start the reconstruction process.










Icon	Description
	Fold/Expand a node
	Add a child node
	Delete a child node
	Cut a subtree
	Change the label of a node
	Change the timezone of a node
	Add an association between a host and a parasite leaf
	Delete all associations of this node
	Rearrange the order of the child nodes

Table 2: Actions for manipulating the coevolutionary system.

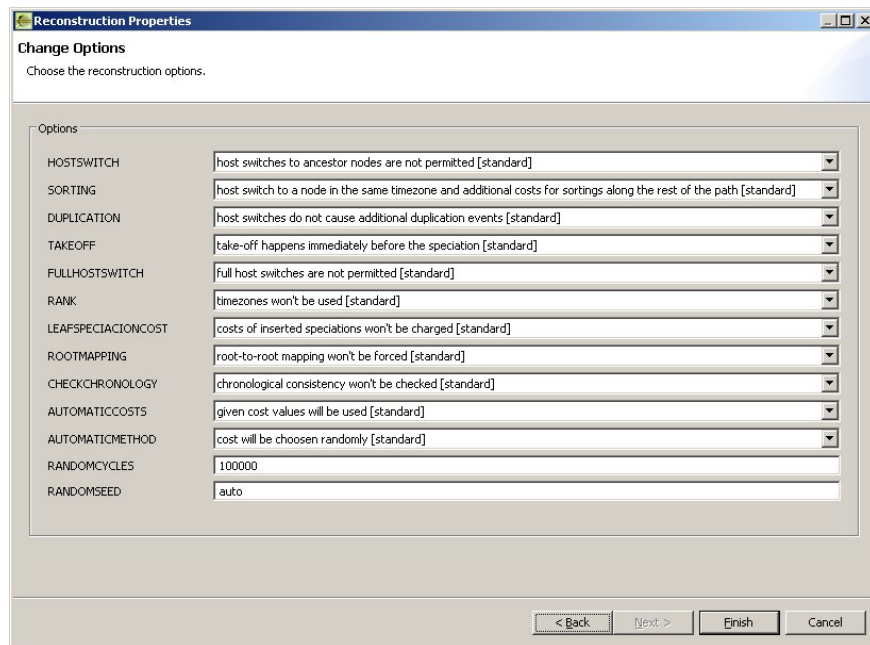
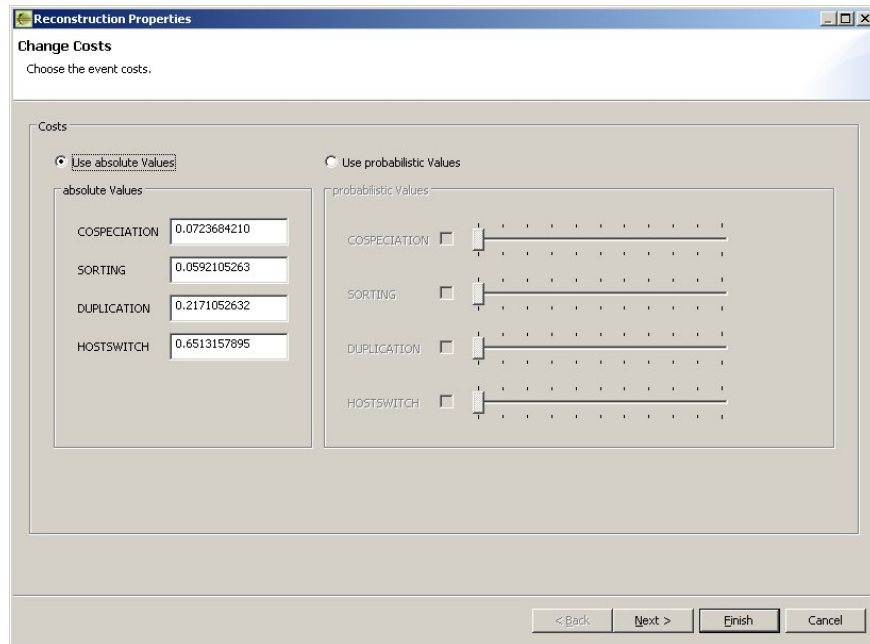



Figure 4: Top: dialog for changing the cost values for a reconstruction; bottom: dialog for changing the reconstruction options.


Additionally you can set up if divergence timing information (RANK) should be used or if a root-to-root mapping (ROOTMAPPING) will be forced. The settings LEAFSPECIACIONCOST and CHECKCHRONOLOGY are for features which are not yet implemented.

Interesting options are AUTOMATICCOSTS and AUTOMATICMETHOD. The first option specifies if the event costs will be computed automatically. If AUTOMATICCOSTS is enabled then with the second option you can choose the method of searching for the best costs. One method is to choose just random values for the cost. An other way is to seek good start values and optimize the solution by a simplex optimisation algorithm. In both cases you can set an integer value for the RANDOMCYCLES property. Either this value gives the number of used random cost values or it specifies how many cost sets should be checked for finding good simplex start values. If you use the random cost method the value should be 100000 or larger. If you use the simplex method values between 1000 and 10000 should be useful. As both methods are only heuristics based on random values you can set a seed value for the random number generator (RANDOMSEED) to reproduce your results.

## Compute a reconstruction

With the  button in the tool bar you can start the reconstruction. In the occurring dialog you can change the previously described options again. Click finish to start the reconstruction. When finished the cost minimal reconstruction will be shown in the editor.

## Statistical tests

For statistical tests you can run several reconstructions based on random co-phylogenetic systems which are in some sense “similar” to the original data. To do this click the  button in the tool bar. There you can set up the reconstruction options as shown in Figure 4. You also have to specify the number of random data sets and the randomisation method (Figure 5 top). Then choose a destination for the log file which will contain the results (Figure 5 bottom).

In the current version of CoRe-PA the results can not be visualized in the tool itself, but you can use frameworks like [R] (<http://www.r-project.org/>) to do this. To import the generated log file into [R] use the command:

```
data <- read.table("filename", sep="\t", head=TRUE)
```

Boxplots for the quality value and the number of cospeciations of the random reconstructions can be created with the commands:

```
boxplot(data$Quality, ylab=expression(q[bar('c')]))  
boxplot(data$CoEvents, ylab='Cospeciations')
```



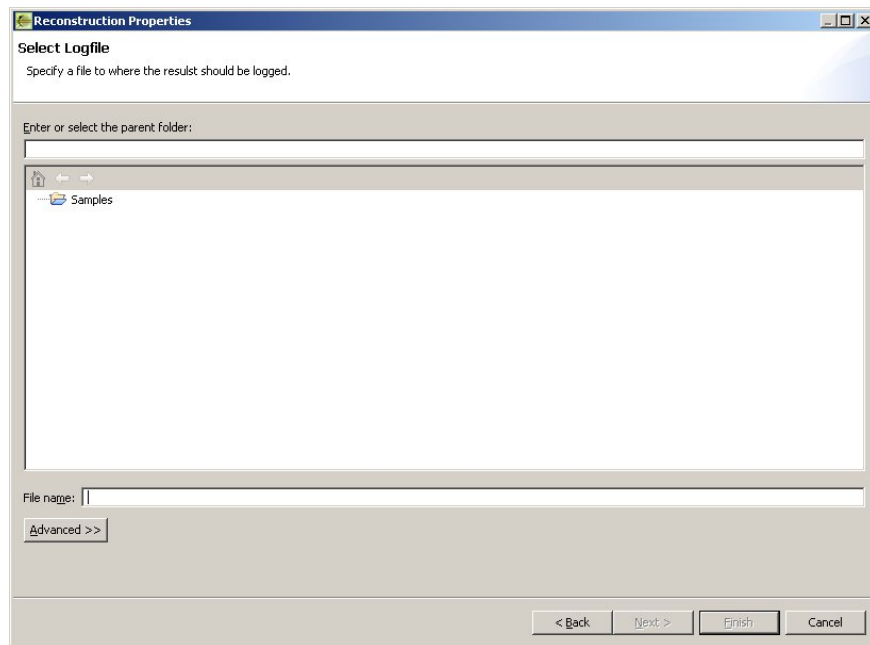
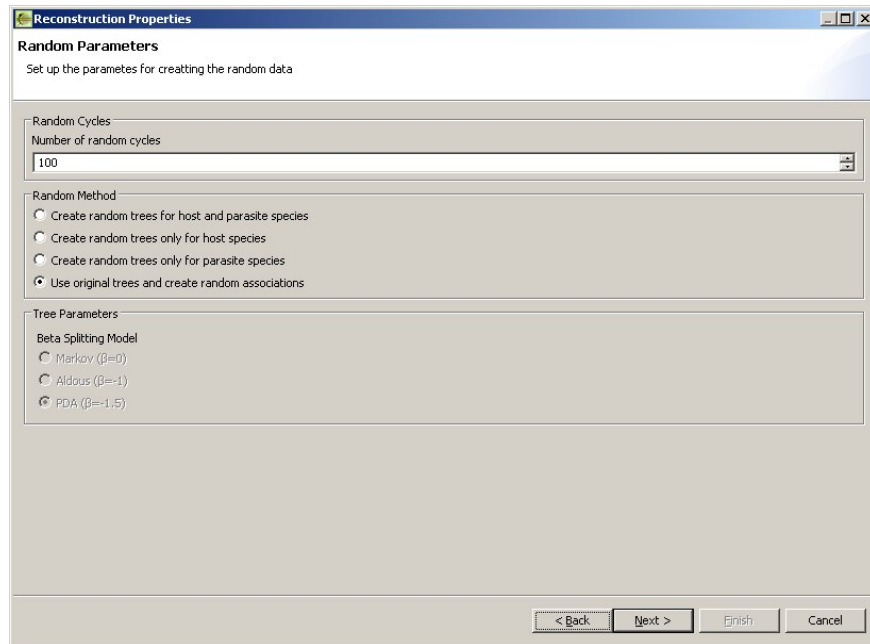


Figure 5: Top: dialog for reconstructing random cophylogenies to do statistical test; bottom: dialog for choosing a log file destination for statistical tests.

A histogram of the number of occurred cospeciations can be generated by:

```
hist(data$CoEvents,
      xlim = c(min(data$CoEvents-1), max(data$CoEvents+1)),
      breaks = c(min(data$CoEvents-0.5):max(data$CoEvents+0.5)),
      main='', ylab='Frequency', xlab='Cospeciations')
```

## Show reconstructions

After computing a cost minimal reconstruction the result will be shown in the editor part (Figure 6). You can get additional information on the respective nodes and on the reconstruction itself by looking at the “Properties” views. While the “Node Properties”-view contain information on the currently selected node (Figure 7), the “Cophylogeny Properties”-view shows several details of the reconstruction, including the quality value, the host association of the parasites root node, the total costs, the occurred event numbers and the used options (Figure 8). There is also a flag indicating if the reconstruction is chronologically valid. If the reconstruction was made with the AUTOMATICCOSTS option turned on, the “Reconstructions” view will contain the distinct cost minimal solutions computed with different cost values (Figure 9). A double-click on such an item will show the respective reconstruction in a new editor.

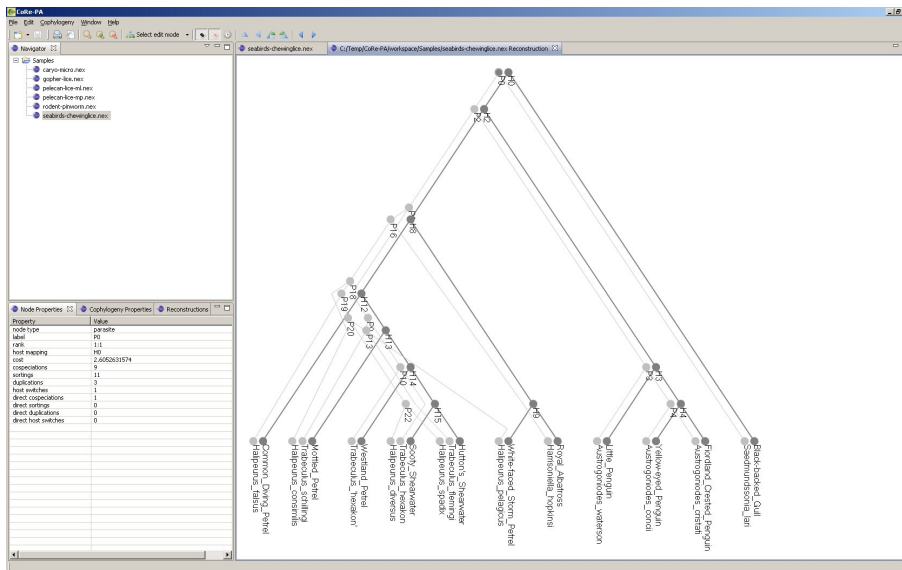


Figure 6: Screenshot of a reconstruction for the seabirds-chewinglice system.

If there were more than one cost minimal reconstruction was found using the same cost values than you can iterate through these reconstructions using the ◀ and ▶ buttons in the tool bar.





Figure 9: The “Reconstructions”-view.

## Additional Notes

Please note that CoRe-PA is in an alpha release stage and there may exist several bugs in the software. So feel free to send bug reports or any other kind of impressions or suggestions on CoRe-PA to [wieseke@informatik.uni-leipzig.de](mailto:wieseke@informatik.uni-leipzig.de).

## References

- [1] D. Merkle, M. Middendorf, N. Wieseke: A Parameter-Adaptive Dynamic Programming Approach for Inferring Cophylogenies. Working Paper PA-COSY 09-01, Department of Computer Science, University of Leipzig, August 2009, submitted.