Phylogenetics

PADRE: a package for analyzing and displaying reticulate evolution

Martin Lott*, Andreas Spillner, Katharina T. Huber and Vincent Moulton
School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Received on January 7, 2009; revised and accepted on March 3, 2009
Advance Access publication March 5, 2009

Associate Editor: Martin Bishop

ABSTRACT

Summary: Recent advances in gene sequencing for polyploid species, coupled with standard phylogenetic tree reconstruction, leads to gene trees in which the same species can label several leaves. Such multi-labeled trees are then used to reconstruct the evolutionary history of the polyploid species in question. However, this reconstruction process requires new techniques that are not available in current phylogenetic software packages. Here, we describe the software package PADRE (Package for Analyzing and Displaying Reticulate Evolution) that implements such techniques, allowing the reconstruction of complex evolutionary histories for polyploids in the form of phylogenetic networks.

Availability: PADRE is an open-source Java program freely available from http://www.uea.ac.uk/cmp/research/cmpbio/PADRE.

Contact: martin.lott@uea.ac.uk

1 INTRODUCTION

Recent advances in techniques for identifying and characterizing homoeologous DNA regions for polyploid species, coupled with standard phylogenetic tree reconstruction (using e.g. parsimony; Swofford, 2000), leads to gene trees in which the same species can label several leaves (see e.g. Popp and Oxelman, 2001). A consensus of such multi-labeled trees (or MUL-trees for short) can then be used to reconstruct the evolutionary history of the polyploid species in question in the form of a phylogenetic network (see e.g. Popp et al., 2005; Huber and Moulton, 2006; Brysting et al., 2007). However, to date, this whole reconstruction process has been done mainly by hand since the necessary techniques are not available in current phylogenetic software packages [see e.g. Gambette (2008), for a comprehensive list of currently available approaches and software tools]. This is because such packages are not designed to cope with MUL-trees. Here, we present the Package for Analyzing and Displaying Reticulate Evolution (PADRE) software package that has been developed to perform this reconstruction process automatically.

2 DESCRIPTION OF THE SOFTWARE PACKAGE

2.1 A consensus of MUL-trees

Given a collection of MUL-trees, PADRE computes a consensus MUL-tree which summarizes these trees using an approach developed in (Huber et al., 2008; Lott et al. submitted for publication). The input must be provided as a file in the widely accepted NEWICK format for phylogenetic trees (Archie et al., 1986), which can easily accommodate MUL-trees.

Essentially, to compute a consensus MUL-tree the input trees are broken down into a collection of clusters (from a multi-set) of the species in question. In Figure 1, we present a screen-shot from PADRE that shows an input MUL-tree which, for example, contains the cluster \{O, S\} twice. These clusters are then used by PADRE to build a new MUL-tree, which forms the consensus of the input MUL-trees.

This whole process is similar to the greedy consensus method for phylogenetic trees (see e.g. Bryant, 2003) that processes clusters greedily according to the number of input trees that exhibit a cluster. However, in contrast to this method, various technical difficulties might occur due to ambiguities that can arise from inserting a cluster into a MUL-tree [leading in fact to an NP-hard problem (Huber et al., 2008)]. PADRE, therefore, implements a heuristic and the user can provide a threshold \(t\) such that clusters that appear in less than \(t\) input trees are not taken into account for computing a consensus MUL-tree.

Note that PADRE will also accept collections of MUL-trees as input in which the number of copies of each species differs (which may arise e.g. in case a gene is absent from a species or cannot be sequenced). In this case, the software first determines the number of copies of each taxon to be placed in the consensus MUL-tree based on a majority rule.

*To whom correspondence should be addressed.
Virtual Machine 1.5 or above is installed. The software provides the PADRE package (current version 2.0) is an open-source Java precise details]. Two ancestral lineages hybridize [see Huber and Moulton (2006) for

phylogenetic network is guaranteed to have a minimum number of the merging of isomorphic subtrees identified at every iteration as part of the network construction process.

leave separate the identified isomorphic subtrees. This option can be removed. At each iteration, the user has the option to combine or and to merge these until all but one of these subtrees has been removed. At each iteration, the user has the option to combine or leave separate the identified isomorphic subtrees. This option can be used to incorporate, for example, additional biological information as part of the network construction process.

An important feature of the algorithm is that if the user accepts the merging of isomorphic subtrees identified at every iteration (or just runs the method in non-interactive mode), then the resulting phylogenetic network is guaranteed to have a minimum number of reticulation nodes, that is, nodes representing events in which two ancestral lineages hybridize [see Huber and Moulton (2006) for precise details].

2.3 Implementation and testing

The PADRE package (current version 2.0) is an open-source Java software package which will work on any platform where the Java Virtual Machine 1.5 or above is installed. The software provides a graphical user interface (GUI) for ease of use. A drag-and-drop facility allows the user to adapt the layout of the trees or networks that is generated automatically by the program for further analysis or printing. The fonts of labels and line width for drawing edges can also be adjusted, and all graphics can be saved directly to EPS format for inclusion in publications. A manual and example files are provided on the web site alongside the software itself.

In collaboration with biologists the software has been tested on various collections of Silene species. For example, see Figure 2 for a network generated using PADRE for a collection of such species [which appeared as part of a larger study in Popp et al. (2005)]. The tests showed that our program was able to detect known polyploidization events and also proposed additional ones, which are largely consistent with what is currently known about these species. In all these tests, the run time of our program was a few seconds on a modern desktop PC. Based on this, we expect that the algorithm for constructing a consensus tree can handle up to 30 taxa and the algorithm for constructing a phylogenetic network from a MUL-tree up to 1000 taxa.

3 DISCUSSION

We have presented a software package that allows biologists to manipulate multi-labeled gene trees in order to reconstruct the evolutionary history of polyploid species. The algorithm for computing a consensus MUL-tree will probably need to be improved to deal with the much larger datasets that we expect to arise from the new pyro-sequencing technologies. However, as it stands, PADRE provides a convenient tool that allows the exploration of the complex evolutionary scenarios that commonly arise when considering polyploid species.

ACKNOWLEDGEMENTS

The authors would like to thank Prof B. Oxlølman and A. Petri, whose input and testing greatly helped with the development of PADRE.

Funding: Engineering and Physical Sciences Research Council [grant number EP/D068900/1] to A.S and V.M.

Conflict of Interest: none declared.

REFERENCES


