

## PROJECT INTRODUCTION

### Objectives

Evolutionary Tree Reconstruction using Optimization Alignment and Bayesian Likelihood Inference.

### Project Investigator / Manager

A/Prof. Rudolf Meier  
National University of Singapore  
dbsmr@nus.edu.sg

### Period of Project

01 Feb 2005 – 31 Dec 2008

### Website

<http://evolution.science.nus.edu.sg/>  
[www.inhs.uiuc.edu/cee/FLYTREE](http://www.inhs.uiuc.edu/cee/FLYTREE)

### Abstract

Evolutionary tree reconstruction is NP-complete and is computational intensive. The resources on the National Grid Pilot Platform (NGPP) have been used for the reconstruction of evolutionary trees of several taxa.

## PROJECT DETAILS

### Description

The Evolutionary Biology Laboratory (EBL) at the Department of Biological Sciences of National University of Singapore is generating a large amount of DNA sequence data for a wide variety of organisms ranging from mussels, insects to fish and primates (see <http://evolution.science.nus.edu.sg/>). Sequences for the same gene from different species need to be aligned. Multiple sequence alignment is NP-hard thus requiring parallel computing for finding optimal or near-optimal solutions. Once an alignment is found, it can be used for tree reconstruction.

Tree reconstruction is again NP-hard, thus requiring parallel computing. The problem is the large number of different trees that exist for even a small number of taxa. For example, there are 316 billion different rooted trees for a data set with only 13 taxa and our data sets now routinely have between 30-220 taxa and sequence information for 3-55 genes.

EBL recently started a very large project (Tree-of-Life for Diptera, partially funded by US National Science Foundation; see [www.inhs.uiuc.edu/cee/FLYTREE](http://www.inhs.uiuc.edu/cee/FLYTREE)) that will require the analysis of a data set with 1500 species. Without parallel computing, all these datasets can only be insufficiently evaluated.

We are using the computational power of the NGPP for implementing:

- (1) POY, a program that combines alignment and tree search and
- (2) MrBayes, a program that evaluates trees based on likelihood scores under a variety of complex evolutionary models.

We have so far analyzed three datasets. We used POY for analyzing the relationships among 79 species of Scathophagidae (Diptera: Insecta) based on seven genes. We were able to reconstruct the evolution of phytophagy within this group. We have also reconstructed an evolutionary tree for Coelopidae (Diptera: Insecta) for 31 species. Both projects are part of the Tree-of-Life Project on Diptera. The third dataset was subjected to an analysis in MrBayes and comprised 33 species of salticid spiders. We were able to demonstrate that eyes with acute vision have evolved twice in the Salticidae. Several additional datasets await analysis. This applies most notably to a very large data matrix for primates (216 species, 55 genes).

### Collaborating Organizations:

- FLYTREE: Consortium of 15 international collaborators from 7 countries (see list on website)
- University of Zurich
- Copenhagen University