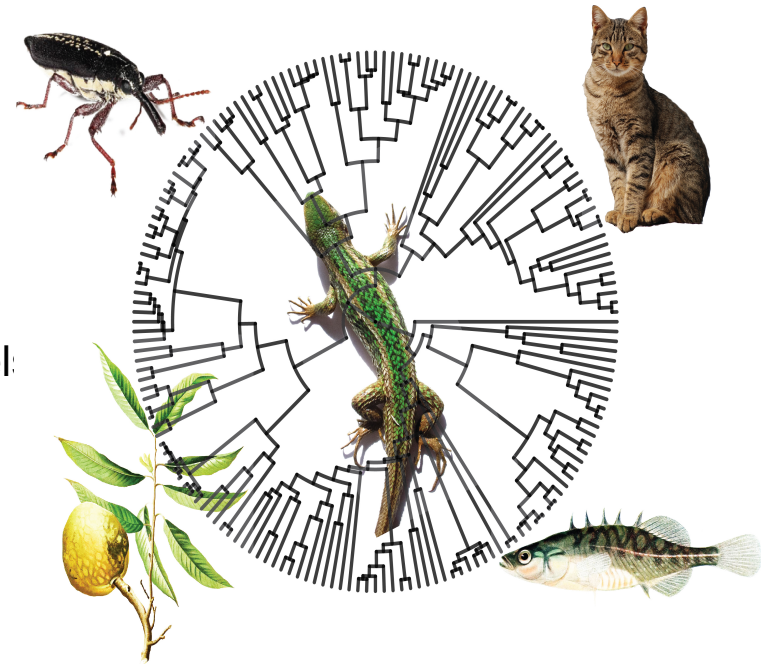


# Biology 559R: Introduction to Phylogenetic Comparative Methods

## Topics for this week (Jan 20 & 22):

- Alignment: Simultaneous alignment and tree estimation
- Alignment visualization and manipulation
- Introduction to statistical estimation of model sequence evolution



# Alignment and Simultaneous Tree Estimation

- We are going to use SATe-2 (SATé - Simultaneous Alignment and Tree Estimation)

URL: <http://phylo.bio.ku.edu/software/sate/sate.html>

*Syst. Biol.* 61(1):90–106, 2011  
© The Author(s) 2011. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.  
For Permissions, please email: journals.permissions@oup.com  
DOI:10.1093/sysbio/syr095  
Advance Access publication on December 1, 2011

## **SATé-II: Very Fast and Accurate Simultaneous Estimation of Multiple Sequence Alignments and Phylogenetic Trees**

KEVIN LIU<sup>1</sup>, TANDY J. WARNOW<sup>1</sup>, MARK T. HOLDER<sup>2</sup>, SERITA M. NELESEN<sup>3</sup>, JIAYE YU<sup>2</sup>,  
ALEXANDROS P. STAMATAKIS<sup>4</sup>, AND C. RANDAL LINDER<sup>5,\*</sup>

<sup>1</sup>*Department of Computer Science, University of Texas at Austin, Austin, TX 78712, USA;* <sup>2</sup>*Department of Ecology and Evolutionary Biology, University of Kansas at Lawrence, Lawrence, KS 66045, USA;* <sup>3</sup>*Department of Computer Science, Calvin College, Grand Rapids, MI 49546, USA; and* <sup>4</sup>*Scientific Computing Group, Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, D-69118 Heidelberg, Germany;* <sup>5</sup>*Section of Integrative Biology, School of Biological Sciences, University of Texas at Austin, Austin, TX, USA;*  
*\*Correspondence to be sent to: The University of Texas at Austin, Austin, TX 78712, USA; E-mail: rlinder@mail.utexas.edu.*

*Received 20 August 2010; reviews returned 9 November 2010; accepted 3 June 2011*

*Associate Editor: Laura Kubatko*

## Alignment and Simultaneous Tree Estimation

- From the Developers' webpage (University of Kansas: Jiaye Yu, Mark Holder, Jeet Sukumaran, Siavash Mirarab, and Jamie Oaks):

SATé is a software package for inferring a sequence alignment and phylogenetic tree. The iterative algorithm involves **repeated alignment and tree searching operations**. The original data set is divided into smaller subproblems by a tree-based decomposition. These sub-problems are aligned and further merged for phylogenetic tree inference.

Currently, the following tools are supported, and are bundled with the SATé distribution:

ClustalW 2.0.12 (sequence alignment program)

MAFFT 6.717 (sequence alignment program)

MUSCLE 3.7 (sequence alignment program)

OPAL 1.0.3 (sequence alignment program)

PRANK 100311 (phylogeny-aware alignment program)

RAxML 7.2.6 (phylogeny estimator program)

FastTree 2.1.4 (phylogeny estimator program)



## SATe-2 needs Python 2.7 (Upgrade Python Instructions)

- MAC OS: Open terminal (go the HD>Applications>Utilities>Terminal)
- MAC OS: Check your version of Python

```
python --version
```

- MAC OS: if necessary upgrade python to 2.7 as required by SATe-II

<http://legacy.python.org/download/>

### Download Python

The current production versions are [Python 3.4.0](#) and [Python 2.7.6](#).

Start with one of these versions for learning Python or if you want the most stability; they're both considered stable production releases.

If you don't know which version to use, try Python 3.4. Some existing third-party software is not yet compatible with Python 3; if you need to use such software, you can download Python 2.7.x instead.

For the MD5 checksums and OpenPGP signatures, look at the [detailed Python 3.4.0](#) page:

- [Python 3.4.0 Windows x86 MSI Installer](#) (Windows binary -- does not include source)
- [Python 3.4.0 Windows X86-64 MSI Installer](#) (Windows AMD64 / Intel 64 / X86-64 binary [1] -- does not include source)
- [Python 3.4.0 Mac OS X 64-bit/32-bit x86-64/i386 Installer](#) (for Mac OS X 10.6 and later [2])
- [Python 3.4.0 Mac OS X 32-bit i386/PPC Installer](#) (for Mac OS X 10.5 and later [2])
- [Python 3.4.0 compressed source tarball](#) (for Linux, Unix or Mac OS X)
- [Python 3.4.0 xzipped source tarball](#) (for Linux, Unix or Mac OS X, better compression)

For the MD5 checksums and OpenPGP signatures, look at the [detailed Python 2.7.6](#) page:

- [Python 2.7.6 Windows Installer](#) (Windows binary -- does not include source)
- [Python 2.7.6 Windows X86-64 Installer](#) (Windows AMD64 / Intel 64 / X86-64 binary [1] -- does not include source)
- [Python 2.7.6 Mac OS X 64-bit/32-bit x86-64/i386 Installer](#) (for Mac OS X 10.6 and later [2])
- [Python 2.7.6 Mac OS X 32-bit i386/PPC Installer](#) (for Mac OS X 10.3 and later [2])
- [Python 2.7.6 compressed source tarball](#) (for Linux, Unix or Mac OS X)
- [Python 2.7.6 xzipped source tarball](#) (for Linux, Unix or Mac OS X, better compression)

## Install SATe-2



- Download SATe-II precompiled from UT-Austin website:

<http://phylo.bio.ku.edu/software/sate/downloads2/mac/satemacs-v2.2.7-2013Feb15.dmg>

Archived bundles of SATe:

File name	Build/Platform	Date	Size (MB)
<a href="#">satesrc-v2.2.7-2013Feb15.tar.gz</a>	src	2013-Feb-15	25.5
<a href="#">satemacs-v2.2.7-2013Feb15.dmg</a>	mac	2013-Feb-15	25.9
<a href="#">satewin-v2.2.7-2013Feb15.zip</a>	win	2013-Feb-15	26.6

\*\*\*For those more adventurous you can download the command based 'SATe-II' from:

<http://phylo.bio.ku.edu/software/sate/downloads2/>

download: [satesrc-v2.2.7-2013Feb15.tar.gz](#)

Follow the instructions in the main webpage

## Preparing FASTA files for SATE-2



- Download the FASTA files from the course website

data 1: [Liolaemus\\_CYTB.fasta](#)

data 2: [Lizard\\_RAG1.fasta](#)

- Create two output folders for the alignment results in your desktop and place the fasta files in the corresponding one

folder: [Liolaemus\\_CYTB](#)

folder: [Lizards\\_RAG1](#)



## Running SATE-2

- Open SATE-II GUI by clicking on the executable on the program folder:
- Explore the console and the options in the SATE-II GUI version:

The screenshot shows the SATE GUI with the following settings:

- External Tools:** Aligner: MAFFT, Merger: MUSCLE, Tree Estimator: FASTTREE, Model: GTR+G20
- Job Settings:** Job Name: satejob, Output Dir.: (empty), CPU(s) Available: 1, Max. Memory (MB): 1024
- Sequences and Tree:** Sequence file: (empty), Multi-Locus Data: , Data Type: DNA, Initial Alignment:  Use for initial tree, Tree file (optional): (empty)
- Workflow Settings:** Algorithm:  Two-Phase (not SATE), Post-Processing:  Extra RAXML Search
- SATe Settings:** Quick Set: SATE-II-fast, Max. Subproblem: Percentage (50) or Size (200), Decomposition: Centroid, Apply Stop Rule: After Last Improvement, Stopping Rule:  Blind Mode Enabled,  Time Limit (hr) (24),  Iteration Limit (1), Return: Best

Buttons: Start

SATe 2.2.7, 2009-2013

Running Log (2015-01-11 15:23:57 MST)

SATe Ready!

## Running SATe-2



- Explore the options in the SATe-II GUI version:

### External Tools:

Aligner: [ClustalW2, MAFFT, PRANK, OPAL]

Merger: [MUSCLE, OPAL]

Tree Estimator: [RAXML, FASTTREE]

Model: [RAXML-options: GTRCAT, GTRGAMMA, GTRGAMMAI;

FASTTREE-options: GTR+G20, GTR+CAT, JC+G20, JC+CAT ]

### Sequences and Tree:

Sequence file ...: [This is the folder where our fasta file resides]

Multi-locus Data [option]

Data Type: [DNA, RNA, Protein]

Initial Alignment [option]

Tree file (optional): [Provide if you have an initial phylogeny associated with the sequences]

### Workflow Settings:

Algorithm [option] Two-Phase (not SATe-II)

Post-Processing [option] Extra RAXML Search

## Running SATe-2



- Explore the options in the SATe-II GUI version:

### Job Settings:

Job Name: [give a name for the job]

Output Dir.: [Select the corresponding directory for the output alignment]

CPU(s) Available: [It will depend on your computer]

Max. Memory (MB): [It will depend on your computer]

### SATe-II Settings

Quick Set: [Presets: SATe-II\_fast, SATe-II\_ML, SATe-II\_simple, custom]

Max. Subproblem:

    Percentage [default 50]

    Size [default 10]

Decomposition:

    Centroid (fast) or Longest (slow)

Apply Stop Rule: [options]

Stopping Rule: Blind Mode Enabled

    Time Limit (hr) [default 24 hours]

    Iteration limit [default 1 iterations]

Return: [Default are Final or Best alignment]

## Running SATe-2: Select the Following Options



### External Tools:

Aligner: [MAFFT]

Merger: [MUSCLE]

Tree Estimator: [RAXML]

Model: [GTRGAMMAI]

### Sequences and Tree:

Sequence file ...: [Liolaemus\_CYTB.fasta]

Data Type: [DNA]

Tree file (optional): [None]

### Workflow Settings:

Algorithm: [None] Two-Phase (not SATe-II)

Post-Processing: [None] Extra RAXML Search

### Job Settings:

Job Name: [Liolaemus\_CYTB\_alignment]

Output Dir.: [Liolaemus\_CYTB] Select the corresponding directory for the output alignment

CPU(s) Available: [2] It will depend on your computer

Max. Memory (MB): [1000] It will depend on your computer

### SATe-II Settings

Quick Set: [SATe-II\_fast]

Iteration Limit: [3]

Leave other options unchanged

# Running SATe-2



The screenshot shows the SATe software interface with the following settings:

- External Tools:** Aligner: MAFFT, Merger: MUSCLE, Tree Estimator: RAXML, Model: GTRGAMMAI
- Job Settings:** Job Name: satejob, Output Dir.: /Users/jcsantos/Desktop/R\_class\_wint, CPU(s) Available: 1, Max. Memory (MB): 1024
- Sequences and Tree:** Sequence file: /Users/jcsanto..., Multi-Locus Data: , Data Type: DNA, Initial Alignment:  Use for initial tree, Tree file (optional):
- Workflow Settings:** Algorithm:  Two-Phase (not SATe), Post-Processing:  Extra RAXML Search
- SATe Settings:** Quick Set: SATe-II-fast, Max. Subproblem:  Percentage (50),  Size (50), Decomposition: Centroid, Apply Stop Rule: After Last Improvement, Stopping Rule:  Blind Mode Enabled,  Time Limit (hr) (24),  Iteration Limit (1), Return: Best

A "Stop" button is located at the bottom center of the interface.

```
SATe INFO: Reading input sequences from '/Users/jcsantos/Desktop/R_class_winter_2015/2_alignnet_using_SATe_II/data/Liolaemus_CYTB/Liolaemus_CYTB.fasta'...
SATe INFO: Directory for temporary files created at /Users/jcsantos/.sate/satejob/temp0AeStA
SATe INFO: Name translation information saved to /Users/jcsantos/Desktop/R_class_winter_2015/2_alignnet_using_SATe_II/data/Liolaemus_CYTB/satejob_temp_name_translation.txt as safe name, original name, blank line format.
SATe INFO: Creating a starting tree for the SATe algorithm...
SATe INFO: Performing initial alignment of the entire data matrix...
SATe INFO: Performing initial tree search to get starting tree...
SATe INFO: Starting SATe algorithm on initial tree...
SATe INFO: Max subproblem set to 50
SATe INFO: Step 0. Realigning with decomposition strategy set to centroid
SATe Running!
```

## Running SATE-2



- Explore the output in a text editor. The alignment is located in these .aln files in fasta format:

satejob.marker001.Liolaemus\_CYTB.aln

```
File Path: ~/Desktop/R_class_winter_2015/2_alignmet_using_SATe_II/data/Liolaemus_CYTB/satejob.marker001.Liolaemus_CYTB.aln
>Liolaemus_wiegmannii_CYTB_KF968961-
-----AACATTTCTGCATGATGAAACTTT
>Liolaemus_tehuelche_CYTB_KF968950-
-----AACATCTCTGCATGATGAAACTTT
>Liolaemus_pseudoanomalus_CYTB_KF968904-
-----AACATCTCCGCATGGTGAACTTT
>Liolaemus_pseudoanomalus_CYTB_KF968903-
-----AACATCTCCGCATGATGAAACTTT
>Liolaemus_pseudoanomalus_CYTB_KF968902-
-----AACATCTCCGCATGATGAAACTTT
>Liolaemus_hermannunezi_CYTB_KF968987-
AAAATGACAATTATACGAAAACACCAATATAAAAAATTATCAATGGCTCATTTATTGACCTACCAACACCATCAAAATCTCTGCATGATGAAACTTT
>Liolaemus_sp_1_MO_2014_CYTB_KF968925-
-----AACATTTCCGCCTGATGAAACTTT
>Liolaemus_salinicola_CYTB_KF968912-
-----AACATCTCTGCATGATGAAACTTC
>Liolaemus_boulengeri_CYTB_KF968990-
AAAATGACAATTATACGAAAACACCAATATAAAAAATTATTAACGGCTCATTTATTGACCTACCAACACCCTCAAACATCTCTGCATGATGAAACTTC
>Liolaemus_sp_4_MO_2014_CYTB_KF968900-
-----AACATCTCTGCATGATGAAACTTT
>Liolaemus_sitesi_CYTB_KF968980-
AAAATGACAATTATACGAAAACCAATATAAAAAATTATCAACGGCTCATTTATTGACCTACCAACACCATCAAAATCTCTGCATGATGAAACTTT
>Liolaemus_sp_4_MO_2014_CYTB_KF968998-
-----TTATACGAAAACCAATATAAAAAATTATTAACGGCTCATTTATTGACCTACCAACACCCTCAAACATCTCTGCATGATGAAACTTT
>Liolaemus_sitesi_CYTB_KF968979-
AAAATGACAATTATACGAAAACCAATATAAAAAATTATCAACGGCTCATTTATTGACCTACCAACACCATCAAAATCTCTGCATGATGAAACTTT
>Liolaemus_sitesi_CYTB_KF968978-
AAAATGACAATTATACGAAAACCAATATAAAAAATTATCAACGGCTCATTTATTGACCTACCAACACCATCAAAATCTCTGCATGATGAAACTTT
>Liolaemus_sp_1_MO_2014_CYTB_KF968938-
-----AACATCTCAGCCTGATGAAACTTC
>Liolaemus_sp_1_MO_2014_CYTB_KF968932-
-----AACATCTCTGCATGATGAAACTTT
>Liolaemus_sp_1_MO_2014_CYTB_KF968933-
-----AACATCTCTGCATGATGAAACTTT
>Liolaemus_sp_1_MO_2014_CYTB_KF968930-
-----AACATCTCAGCCTGATGAAACTTC
>Liolaemus_sp_1_MO_2014_CYTB_KF968931-
-----AATATTTCCGCATGATGAAACTTT
```

Repeat the same process with the Lizard\_RAG1.fasta file

## Mesquite: Visually explore the alignments

- Download mesquite:

<http://mesquiteproject.wikispaces.com/Installation+on+MacOS+X>

<https://github.com/MesquiteProject/MesquiteCore/releases>