
phylogenetics Documentation

Release 0.4.2

Zach Sailer

Jul 11, 2018

Contents

1	Basic Example	3
2	Installation	5
3	Dependencies	7
4	Table of Contents	9

Phylogenetics is a minimal Python API for doing phylogenetics. It manages the annoying aspects of phylogenetics (i.e. file conversion) for you and lets you focus on exploring and interpreting the data.

CHAPTER 1

Basic Example

The main object in phylogenetics is the `PhylogeneticsProject`. Start by initializing the `PhylogeneticsProject` object, pointing to a directory where you'd like to store all the phylogenetic data/output.

```
# Imports
from phylogenetics import PhylogeneticsProject

# Initialize a project class
project = PhylogeneticsProject(project_dir='project')
```

Then, add an alignment (using any file schema you'd like). These sequences will be the tips of your tree. You can now begin building trees and reconstructing ancestral sequences.

```
project.read_data(dtype='tips', path='alignment.fasta', schema='fasta')

# Run PhyML to construct a phylogenetic
# tree by maximum likelihood.
project.run_tree()

# Reconstruct ancestral sequences using default settings.
project.run_reconstruction()
```

File Edit View Run Kernel Tabs Settings Help

tutorial.ipynb

Python 3 (phylogenetics)

Files

Running

Commands

Cell Tools

Tabs

Intro to Phylogenetics

```

In [1]: from phylogenetics import PhylogeneticsProject

In [2]: # Initialize
project = PhylogeneticsProject('example', overwrite=True)

# Add an alignment
project.read_data(
    dtype='tips',
    path='PF08793_seed.fasta',
    schema='fasta'
)

# Run a tree
project.run_tree()

# Run a reconstruction
project.run_reconstruction()

```

branchScale 150
heightScale 100
[Export as PNG](#) [View Source](#) [Open in Vega Editor](#)

PF08793_seed

	alt_posterior	alt_sequence	id	ml_posterior	ml_sequence	support	unique_id
0	0.906800	YCPEFAR----DP-TRNPRRTGRTIKRGPGTYRLEAECSD	27	0.939629	YCDEFAR----DP-TRNPRRTGRTIKRGPGTYRLEAECAD	1.680162	anc00000000
1	0.679886	KCQEFK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	23	0.714800	KCQEFK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	0.718698	anc00000001
2	0.620800	RCKQFEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	34	0.667229	KCNEFEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	2.806276	anc00000002
3	0.935914	YCDEFER----NP-TRNPRRTGRTIKRGPGTYRLEAECAD	26	0.950114	YCDEFER----NP-TRNPRRTGRTIKRGPGTYRLEAECAD	8.545520	anc00000003
4	0.738706	RCEWEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	30	0.794147	KCDEWEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	2.531435	anc00000004
5	0.617066	LCETFKK----DP-TVNPYTGRIKPGGPIYKLEKMCSD	20	0.694571	LCAEFKR----DP-TVNPYTGRIKPGGPIYKLEKMCSD	6.040602	anc00000005
6	0.756706	KCDEWEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	31	0.789118	KCDEWEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	0.218717	anc00000006
7	0.780588	KCEFEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	29	0.798000	RCDEFER----NP-TVNPYTGRIKPGGPIYKLEKMCSD	0.265099	anc00000007
8	0.790647	RCSKWEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	33	0.829824	KCSKWEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	1.909391	anc00000008
9	0.686857	RCTEFER----NP-TVNPYTGRIKPGGPIYKLEKMCSD	22	0.716829	RCTEFER----NP-TVNPYTGRIKPGGPIYKLEKMCSD	2.503850	anc00000009
10	0.700371	LCADYKK----SP-TVNPYTGRIKPGGPIYKLEKMCSD	19	0.764629	LCAEFKR----SP-TVNPYTGRIKPGGPIYKLEKMCSD	None	anc00000010
11	0.690590	KCDEWEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	32	0.728974	KCDEWEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	7.484104	anc00000011
12	0.663714	LCAEFHK----DP-TVNPYTGRIKPGGPIYKLEKMCSD	21	0.699171	LCDEFER----DP-TVNPYTGRIKPGGPIYKLEKMCSD	1.162623	anc00000012
13	0.659714	TCEFAFK----DP-TVNPYTGRIKPGGPIYKLEKMCSD	28	0.682314	TCEFCR----DP-TVNPYTGRIKPGGPIYKLEKMCSD	2.745285	anc00000013
14	0.758057	KCEFEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	25	0.794343	RCDEFER----NP-TVNPYTGRIKPGGPIYKLEKMCSD	6.209727	anc00000014
15	0.685543	KCNEFEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	24	0.712286	KCNEFEK----NP-TVNPYTGRIKPGGPIYKLEKMCSD	0.677978	anc00000015

CHAPTER 2

Installation

Install from PyPi:

```
pip install phylogenetics
```

To install a development version:

```
git clone https://github.com/Zsailer/phylogenetics
cd phylogenetics
pip install -e .
```


CHAPTER 3

Dependencies

phylogenetics manages phylogenetics data. Currently, it doesn't do any of the phylogenetic calculations itself. For this, use external tools like:

1. **PhyML**: building maximum likelihood trees.
2. **PAML**: reconstructing ancestors.

phylogenetics is built on top of following python stack:

1. **DendroPy**: A Python library for phylogenetic scripting, simulation, data processing and manipulation.
2. **ToyTree**: A minimalist tree plotting library using toyplot graphs
3. **PhyloPandas**: Pandas for phylogenetics
4. **PyASR**: Ancestral Sequence Reconstruction in Python

4.1 Installing PhyML and PAML

It's important that the following executables are callable from any location for `phylogenetics` to work properly with external tools:

1. **PhyML_**: building maximum likelihood trees.
2. **PAML**: reconstructing ancestors.

4.1.1 PhyML

Get the latest, stable release of '**PhyML_**'

Download, unpack, and install PhymL. Export path to the `src` file in `bashrc` file.

4.2 TreeProject API

The following pages is the TreeProject API documentation. The TreeProject is the main class you'll use in the `phylogenetics` package.

4.2.1 TreeProject Class