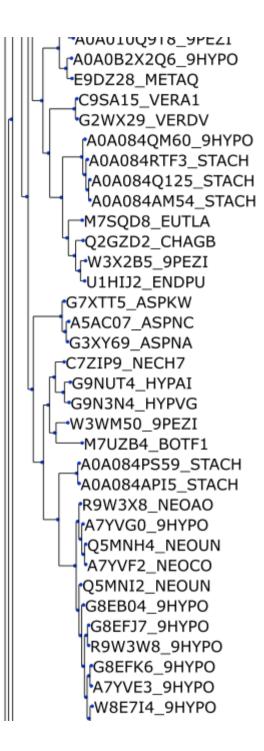
Why is it useful to know a programming language when doing phylogenomics?



Sequences obtained from UniProt

```
>tr|C8VAU1|C8VAU1 EMENI MAT2 proteinMating type HMG-box protein [Source:UniProtKB/TrEMBLAcc:Q7Z8M2] OS=Emericella nidulans (strain FGSC A4 /
ATCC 38163 / CBS 112.46 / NRRL 194 / M139) GN=MAT2 PE=4 SV=1
MAAVSIAMKSPTOSPDSITELLWKDALRHLGSTNDEVLLPTNVVDIIGODNVEKIKSRLS
ALLGAPVVSFVDESINALRVLRTPTFSGSSISVASPSRALDSWPSEPPNKPRPASMKPAK
IPRPPNAFILYRQHHYPKVKEARPDLSNNEISVIIGKKWRAEPEEGKLHFKNLAEEFKKK
HAEEYPDYOYTPRKPSEKKRRAASRISPKNSKRTVALENPGSMTAPSSNVFTPOMYPGIO
NGOLAGAGYIGYLDGLNSMVNTGGLTDEPTNFGTNAFNSLFOOPOSDYGRTALFPOLEFA
GPSLGDSLEFPEFAADYF
>tr|Q7Z8M2|Q7Z8M2 EMEND MAT2 protein OS=Emericella nidulans GN=matA-2 PE=2 SV=1
MAAVSIAMKSPTOSPDSTTELLWKDALRHLGSTNDEVLLPTNVVDIIGODNVEKIKSRLS
ALLGAPVVSFVDESINALRVLRTPTFSGSSISVASPSRALDSWPSEPPNKPRPASMKPAK
IPRPPNAFILYROHHYPKVKEARPDLSNNEISVIIGKKWRAEPEEGKLHFKNLAEEFKKK
HAEEYPDYQYTPRKPSEKKRRAASRISPKNSKRTVALENPGSMTAPSSNVFTPQMYPGIQ
NGOLAGAGYIGYLDGLNSMVNTGGLTDEPTNFGTNAFNSLF00P0SDYGRTALFP0LEFA
GPSLGDSLEFPEFAADYF
>tr|Q5B3Z6|Q5B3Z6 EMENI Uncharacterized protein OS=Emericella nidulans (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) GN=AN4734.2
PE=4 SV=1
MAAVSIAMKSPTOSPDSITELLWKDALRHLGSTNDEVLLPTNVVDIIGODNVEKIKSRLS
ALLGAPVVSFVDESINALRVLRTPTFSGSSISVASPSRALDSWPSEPPNKPRPASMKPAK
IPRPPNAFILYROHHYPKVKEARPDLSNNEISEEFKKKHAEEYPDY0YTPRKPSEKKRRA
ASRISPKNSKRTVALENPGSMTAPSSNVFTPQMYPGIQNGQLAGAGYIGYLDGLNSMVNT
GGLTDEPTNFGTNAFNSLFQQPQSDYGRTALFPQLEFAGPSLGDSLEFPEFAADYF
>tr|A0A0U5G7S6|A0A0U5G7S6 9EURO Putative Mating type protein MAT1-2 OS=Aspergillus calidoustus GN=ASPCAL10920 PE=4 SV=1
MATVSIAMKSTAOSPENVMERLWODALRHLGSTNNEVLLPTNVVDVIGONNVEEIKSRLC
ALLGAPVVAFIDESINALRVMRTPAFSGTTISVASHQGMAGVKSVEAPGKPRAAIAKPPK
IPRPPNAFILYRQHHHPKVKEAYPDLSNNEISVILGKQWKSEPDEARLRFKSMAEEFKKK
HAEEYPDYQYTPRKPSEKKRRATSRQSPKSKRSIPLESPPSIAAPSPSAFTPSMYSELQG
NDTMIEGYPNPLDNFDNLNFMIDSGALADEHTGYDMNSFDALLHOAPNDYGRGTFFOHLG
IADQSIGDSFEFPDYAANCF
>tr|A0A0F8XMY0|A0A0F8XMY0 9EURO Uncharacterized protein OS=Aspergillus rambellii GN=ARAM 000112 PE=4 SV=1
MTAVSIAMKSVTTSTDNLTELLWODALRHLGATNNEVLLPTNVMDLIGODNVDKIKSRLC
VLLGAPVVAFVDESINVLRIMRTPVFSGSSISIASHDRMFGNKSNEASSKVRAASTKSTK
IPRPPNAFILYROHHHPRVKEAYPDLSNNOISIILGKOWKAESEEVKVHFKDLAEEFKKK
HAEDHPDYQYTPRKPSEKKRRATSRRSPKHVKSSQDLKYPTPTSASPNAFTPTTYPDVQN
GNIAITGYTDDLTDLNLMFDSSSVSERHSSFDSNAFDTLFQQPQDEYARATLYPHMNLAE
QSFGDSFEFPNFPGDCF
```

How can I get my sequences to have a shorter, more informative, header?



You're interested in a group of genes that perform a certain metabolic function.

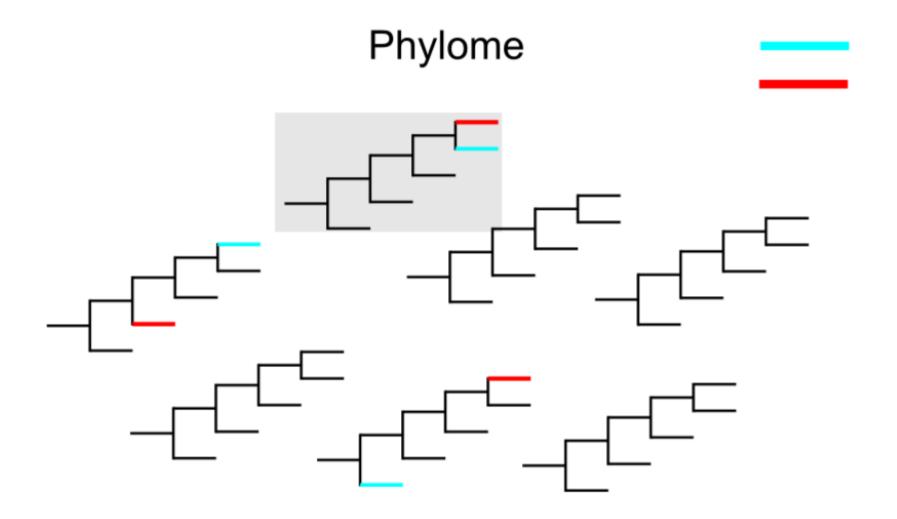
To study their evolution you have reconstructed a gene tree for every gene in the pathway.

You have found something interesting in the trees and now you want to put them nicely for your paper

AUAUIUQ918_9PEZI A0A0B2X2Q6 9HYPO E9DZ28_METAQ rC9SA15 VERA1 G2WX29 VERDV rA0A084QM60_9HYPO A0A084RTF3 STACH rA0A084Q125 STACH A0A084AM54 STACH →M7SQD8_EUTLA O2GZD2 CHAGB rW3X2B5 9PEZI U1HIJ2 ENDPU G7XTT5 ASPKW A5AC07_ASPNC G3XY69 ASPNA C7ZIP9 NECH7 r•G9NUT4 HYPAI G9N3N4 HYPVG ⊸W3WM50_9PEZI M7UZB4_BOTF1 rA0A084PS59 STACH A0A084API5 STACH rR9W3X8 NEOAO A7YVG0_9HYPO Q5MNH4 NEOUN A7YVF2 NEOCO Q5MNI2_NEOUN G8EB04 9HYPO G8EFJ7_9HYPO R9W3W8 9HYPO G8EFK6_9HYPO A7YVE3_9HYPO W8E7I4_9HYPO

Python can help us create images that would otherwise take a lot of time

```
U1HIJ2_ENDPU
                                     Endocarpon pusillum
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Eurotiomycetes; Chaetothyriomycetidae; Verrucariales; Verrucariaceae; Endocarpon
G7XTT5 ASPKW
                                     Aspergillus kawachii
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiomycetidae; Eurotiales; Aspergillaceae; Aspergillus
rA5AC07 ASPNC
                                     Aspergillus niger
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiomycetidae; Eurotiales; Aspergillaceae; Aspergillus
 G3XY69 ASPNA
                                     Aspergillus niger
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiomycetidae; Eurotiales; Aspergillaceae; Aspergillus
•C7ZIP9 NECH7
                                      Nectria haematococca
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Fusarium; Fusarium solani species complex
rG9NUT4 HYPAI
                                     Hypocrea atroviridis
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Trichoderma
 G9N3N4 HYPVG
                                     Hypocrea virens
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Trichoderma
→W3WM50_9PEZI
                                     Pezizomycotina
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota
  M7UZB4 BOTF1
                                      Botryotinia fuckeliana
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Leotiomycetes; Helotiales; Sclerotiniaceae; Botrytis
  rA0A084PS59_STACH
                                     Stachybotrys chartarum
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Stachybotrys
  A0A084API5 STACH
                                     Stachybotrys chartarum
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Stachybotrys
  reswax8 NEOAO
                                     Neotyphodium aotearoae
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Neotyphodium
  hA7YVG0 9HYPO
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae
                                     Hypocreales
    Q5MNH4_NEOUN
                                     Neotyphodium uncinatum
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Neotyphodium
   A7YVF2 NEOCO
                                     Neotyphodium coenophialum
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Neotyphodium
   O5MNI2 NEOUN
                                     Neotyphodium uncinatum
                                                                   Eukaryota; Funqi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Neotyphodium
  G8EB04_9HYPO
                                     Hypocreales
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae
    rG8EFJ7 9HYPO
                                     Hypocreales
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae
    R9W3W8_9HYPO
                                                                   Eukaryota; Funqi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae
                                     Hypocreales
    G8EFK6 9HYPO
                                     Hypocreales
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae
    A7YVE3 9HYPO
                                     Hypocreales
                                                                   Eukaryota; Funqi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae
    rW8E7I4_9HYPO
                                     Hypocreales
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae
     G8EB14 9HYPO
                                     Hypocreales
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae
O5AV82 EMENI
                                     Emericella nidulans
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiomycetidae; Eurotiales; Aspergillaceae; Aspergillus
 A1CSG6_ASPCL
                                     Aspergillus clavatus
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiomycetidae; Eurotiales; Aspergillaceae; Aspergillus
  G7XA53 ASPKW
                                      Aspergillus kawachii
                                                                   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiomycetidae; Eurotiales; Aspergillaceae; Aspergillus
```



A phylome is the complete collection of phylogenetic trees for each gene encoded in a genome. We can obtain a lot of data from it, but it implies working with a huge amount of trees.

You will probably need a tree manipulating software to get any kind of information from it.

For those of you that would like to learn python, here are two tutorials:

http://www.cmbi.ru.nl/pythoncourse/

http://cscircles.cemc.uwaterloo.ca/

For those of you already know how to program in python: ETE

http://etetoolkit.org/



Home

Gallery

Documentation -

TreeView

Support

About

Download atest v3.0.0b35

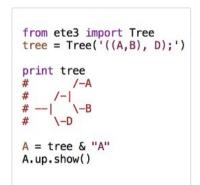
A Python framework for the analysis and visualization of trees.





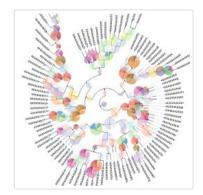
Phylogenomic tools

Contribute



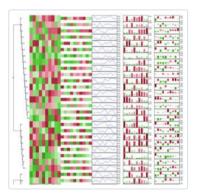
Trees as Python objects

Load, create, traverse, search, prune, or modify hierarchical tree structures with ease using the ETE Python API.



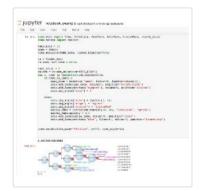
Programmatic tree visualization

Get full control of your tree images. Browse them interatively or render SVG, PNG of PDF images.



Tree annotation

Custom node attributes can be rendered as graphical elements. Choose among external images, charts, symbols, text labels, and more!



Jupyter notebook support

Prototype your methods using the Jupyter notebook framework including inline visualization of trees.

What is ETE?

- ETE (Environment for Tree Exploration) is a package of tools to work with trees in a programmatic way.
- We need some simple concepts of scripting programming
- There are other programming toolkits with more specific scopes: (DendroPy, BioPython, BioPerl).

ETE will allow you to:

- Read and write newick format
- Allows for node annotation
- Manipulate large tree structures
- Calculate distances among nodes
- Re-root trees
- Search nodes and partitions
- Iterate along the tree topology
- Create nice tree images
- And a lot more...

Tree Basics

A phylogenetic tree can be uploaded in ETE and this will create a tree object.

In ETE a tree object is nothing more than a group of tree nodes connected in a hierarchical way

To create a new tree object in ETE you only need to call the Tree() command.

```
import ete3
t = ete3.Tree()
```

This will just create an empty tree node.

You can upload trees from either a file or a string:

```
t = ete3.Tree("file_name")

t = ete3.Tree("((A,B),C);")
```

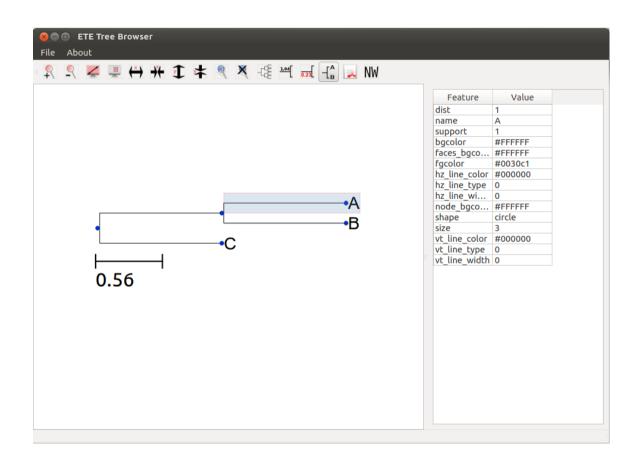
You can also write the tree either on the screen or into a file

```
t = ete3.write()
t = ete3.write(outfile=file_name)
```

Tree Basics

ETE also includes a visualizer that allows you to look at your tree:

t.show()



The window is interactive and will allow you to manually check out information and modify the tree if you want.

Tree Basics

What is the difference between a Tree and a PhyloTree in ETE?

A PhyloTree includes all the options of a Tree but includes many more that are strictly designed to work with phylogenetic trees.

For a PhyloTree we will need to provide the species information. By default ETE will assume that the first three letters of each leaf indicate the species.

```
t= ete3.PhyloTree("((Prot1_Human,Prot1_Chimp),Prot1_Rat);")
```

Species will be: Pro for all leaves.

To define a new way to get the species names:

```
def species_name(node):
    return node.split("_")[1]
```

This small function says that each name should be split by the "_" and the second element should be returned.

```
t=ete3.PhyloTree("((Prot1_Human,Prot1_Chimp),Prot1_Rat);",sp_naming_function=species_name)
```