

Generating phylogenetic trees with Phylomatic and dendrograms of functional traits in R



Zhang Jinlong

IBCAS

2010-8-1

Contents

1. Introduction
2. Phylomatic: a step by step guide
3. Import phylogenetic trees to R
4. Dendrograms of functional traits

Contents

1. Introduction
2. Phylomatic: a step by step guide
3. Import phylogenetic trees to R
4. Dendrograms of functional traits

Molecular phylogenetic trees

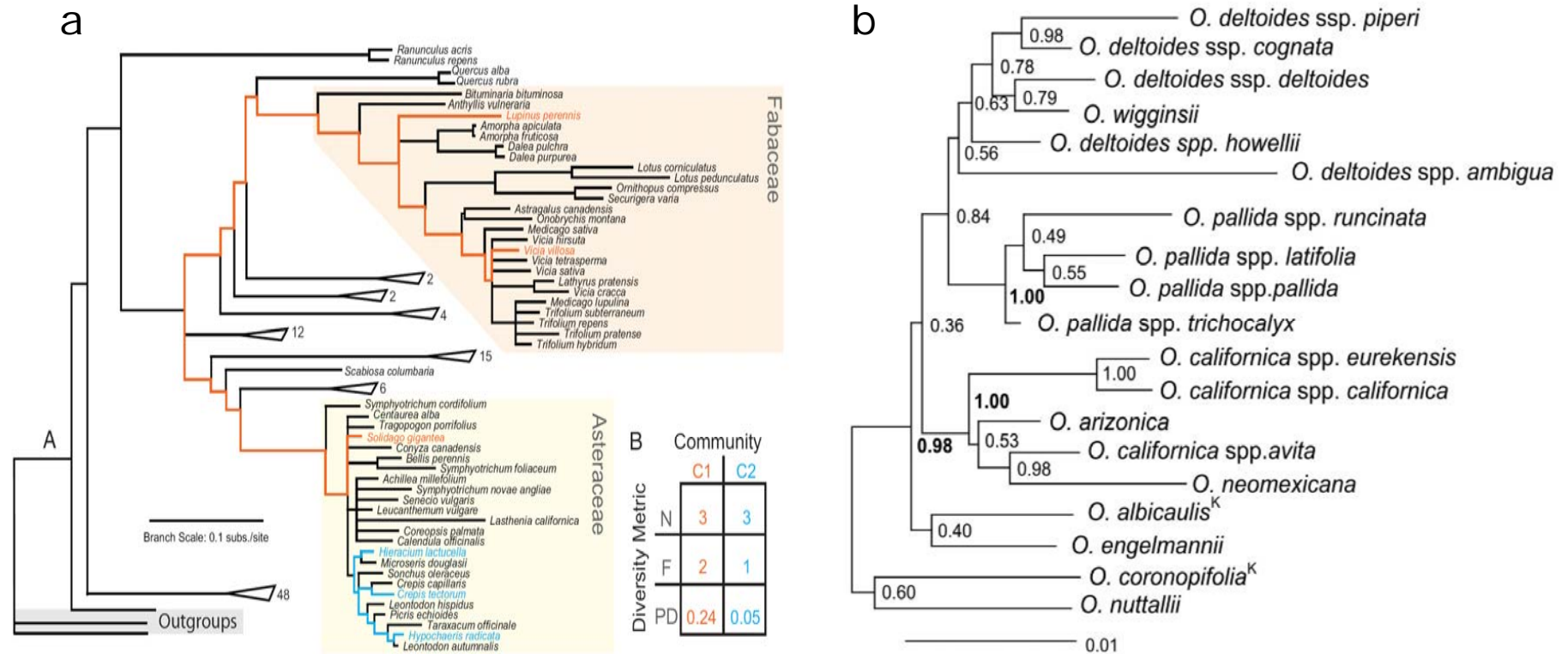
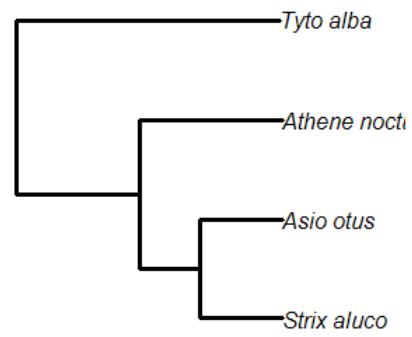
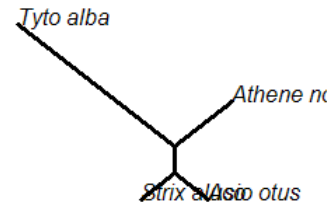


Fig. 1a Cadotte M. *et al.* PNAS, 2008

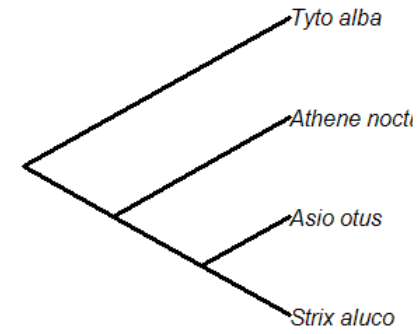
Fig. 1b Evans M. *et al.* The American Naturalist, 2009



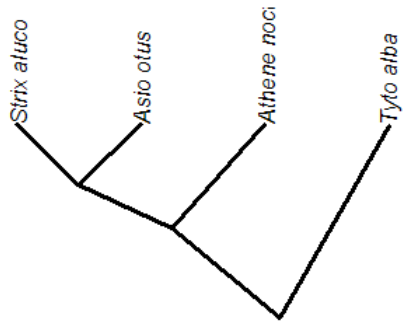
Phylogram



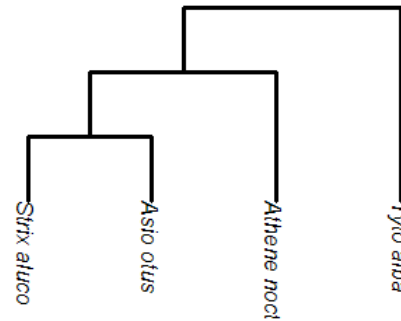
Unrooted



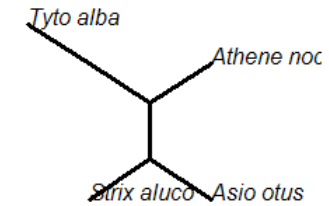
Chronogram



Chronogram



Phylogram



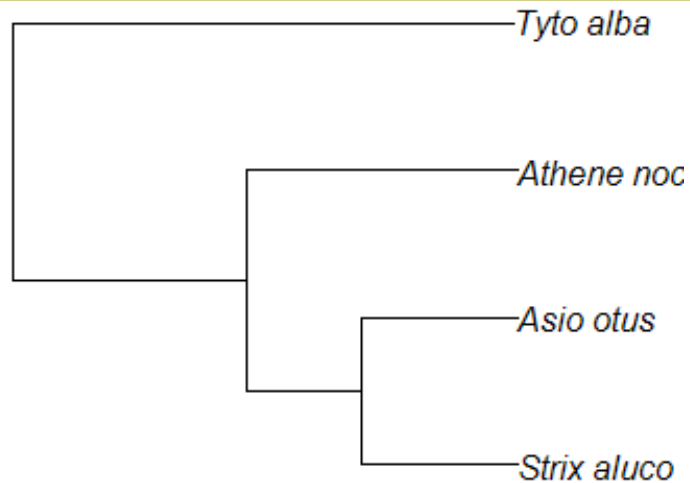
Unrooted

Fig.2 Types of phylogenetic trees

Glossary

- **Phylogram:** length of branches represents likelihood.
- **Chronogram:** length of branches represents time.
- **Outgroup:** A taxon (or taxa) known to lie outside of the group being examined.
- **Bootstrap value:** The confidence level test by bootstrap method. Use to test the reliability of an inferred tree. Nodes with bootstrap value less than 50% will be collapsed and generating polytomies.

The Newick format



```
owls(((Strix_aluco:4.2,Asio_otus:4.2):3.1,Athene_noctua:7.3):6.3,  
Tyto_alba:13.5);
```

Fig. 3 Newick format example

Newick format was developed by Archie J *et al.* in 1986. at Newick's restaurant in Dover, New Hampshire, USA.

The NEXUS format

```
Begin trees;  
  
tree winteraceae =  
owls(((Strix_aluco:4.2,Asio_otus:4.2):3.1,Athene_noc  
tua:7.3):6.3,Tyto_alba:13.5);  
  
End;
```

- Nexus format was defined by Madison *et al*, 1997. Trees are stored as one of the blocks in nexus file. Used in PAUP*, MrBayes, r8s.

Methods for phylogenetic inferences

Tab.1 Methods and software for inferring phylogenetic trees from DNA sequences

Methods	Software
Genetic distance	Phylip
Neighbor joining	PAUP*
Maximum likelihood	PAUP*, Phylip, PHYML, RAxML
Most parsimony	PAUP*
Bayesian inferences	MrBayes, BEAST

Molecular phylogenetics Wiki

Bodega Phylogenetics Wiki

Topics

Front Page

2010 Workshop

Topics

Links

Recent Changes

Contacts

Topics Sorted by Type of Analysis:

[Reconstructing Phylogenetic Trees](#) w/PAUP*, RAxML, GARLI, MrBayes, and BEST Tutorials

[Divergence Time Estimation](#) w/BEAST Tutorial

[Delimiting Populations & Species](#) w/Structure and Structurama Tutorials

[Phylogenetic Comparative Methods](#) w/Mesquite, R, and BayesTraits Tutorials

[Morphological Diversification and Rates of Evolution](#) w/Brownie Tutorial

[Community Phylogenetics](#) w/Phylocom Tutorial

[Biogeography](#) w/Lagrange Tutorial

[Species Diversification Rates](#)

[Phylogenomics](#)

- Fig.4 Bodega wiki <http://bodegaphylo.wikispot.org/Topics>

R-CRAN Task View-Phylogenetics

- Most of the computational molecular phylogeny methods require exhaustive computational power. Therefore, R is not sufficient in building phylogenetic trees, but can call other software such as PHYML, MrBayes and generating required results.
- R- CRAN Task View Phylogenetics summarized the comparative methods that can be implemented in R.
- Package **picante** developed by Steve Kembel *et al.* can be used to analyze community phylogenetics.
- Package **ape**, **phylobase**, **geiger**, **ouch** and **ade4** are also used very often.

Inferring molecular phylogenetic trees

- **Step1** DNA extraction and sequencing in different species.
- **Step2** Appropriate method for alignment, inferences and computational power. Additionally, for molecular phylogeny from distantly related taxon: the super matrix will have to be build.
(alignment, inference method, reliability test)
- **Step3** Non ultrametric to ultrametric tree: methods in molecular dating.
- **A bit complicated and expensive.**

The APGIII phylogenetic tree

The Angiosperm Phylogeny Group

To represents consensus opinions of systematic botanists.

- **APG:** 1998 An ordinal classification for the families of flowering plants
- **APGII:** 2003
- **APGIII:** 2009

Phylomatic and Phylocom

Phylomatic

News

Phylocom

Software for the Analysis of Phylogenetic Community Structure and Character Evolution, with Phylomatic

Current version: 4.1 (released 2009-08-25)

© 2009 **Campbell Webb, David Ackerly, Steven Kembel**

Fig. 5 Page of Phylomatic and Phylocom

Contents

1. Introduction
2. **Phylomatic: a step by step guide**
3. Import phylogenetic trees to R
4. Trees of functional traits

From plant checklist to phylogeny

- Assume we have a checklist, what to do next in order to build phylogeny according to APGIII ?
- *Castanopsis eyrei*, *Rhododendron latoucheae*, *Vaccinium carlessi*, *Schima superba*, *Adinandra millettii*, *Eurya muricata*, *Camellia fraterna*, *Machilus thunbergii*...
- First, we have to reshape the input checklist to the format that can be used in Phylomatic.

The input Data format for Phylomatic

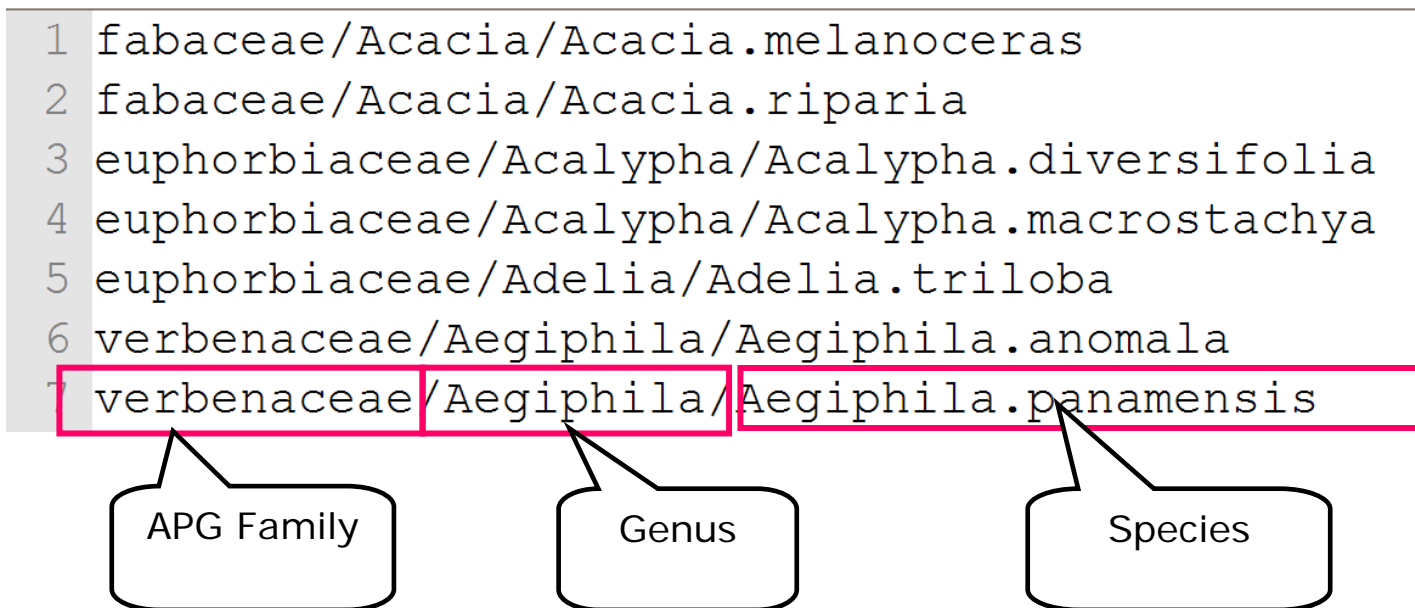


Fig. 6 Format of input data used by Phylomatic

Step 1 Look up the genera for APG families

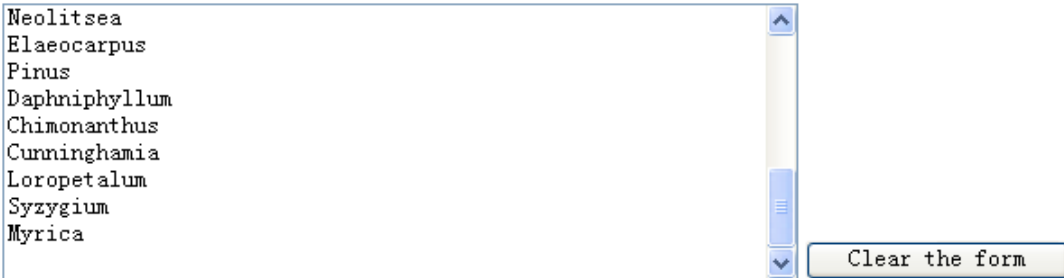
Kew Genus + APG Family Lookup

Consults the genera-in-families list at [Kew](#), and then the families-in-order list of Peter Stevens at [APweb](#). **Many thanks to both groups!**

[NB: the APG may have moved a genus out of the family to which it is assigned by Kew. I.e. this tool does not necessarily assign genera to the 'correct' clade on APweb.]

1. Paste your genus list

Format: Genus RET



Neolitsea
Elaeocarpus
Pinus
Daphniphyllum
Chimonanthus
Curninghamia
Loropetalum
Syzygium
Myrica

Clear the form

2. Check genus synonym matches:

Submit

Fig.7 Check the APG family for each genus

Step 2 Data preparation

- Copy and paste the family, genera, species separated by “/” in plain text. You may use notepad and MS Excel. Notepad++ is an alternative to MS notepad.

Paste and submit the input data to Phylomatic when you finished data preparation .

- **Notice:**

Only the first letter of the APG III families in the **lower case** can be recognized by Phylomatic. Remember to substitute the first letter to lower case.

Output tree

```
((((((((((Castanopsis_eyrei,Castanopsis_carlesii,Castanopsis_fargesii,Ca
stanopsis_tibetana)castanopsis,(Cyclobalanopsis_gracilis,Cyclobalanopsi
s_glauca)cyclobalanopsis,Lithocarpus_glaber)fagaceae,Myrica_rubra),Elae
ocarpus_decipiens),Syzygium_buxifolium),(Daphniphyllum_oldhamii,Loropet
alum_chinense)),((Rhododendron_latoucheae,Rhododendron_ovatum)rhododen
dron,Vaccinium_carlesii)ericaceae,(Schima_superba,Adinandra_millettii,E
urya_muricata,Camellia_fraterna,Ternstroemia_gymnanthera)theaceae)),((M
achilus_thunbergii,Neolitsea.aurata.var.chekiangensis)lauraceae,Chimona
nthus_salicifolius)laurales)magnoliales_to_asterales)euphyllophyte;
```

Fig. 8 The phylomatic output: tree without branch length

- The output is a newick tree that retains the relationship of APGIII families and without branch length. Therefore, we have to assign the branch length using BLADJ block in Phylocom.

Step 3 Preparations before using BLADJ

Download Phylocom

Software for the Analysis of Phylogenetic Community Structure and Character Evolution, with Phylomatic

Current version: 4.0 beta

© 2008 Campbell Webb, David Ackerly, Steven Kembel

User information

Please fill in the required information. We will then be able to keep you informed about

Your name:

Your email:

Please let us know what data you might use this software on:

(no problem if you're just `browsing!`):

Fig. 9 Submit information and download Phylocom

Step 3 Preparations before using BLADJ

- Download and unzip the file ***phylocom4.1.zip***

If you use windows:

- Copy the executable file ***phylocom.exe*** in "`\phylocom-4.1\w32\`" to the directory you wish. For example: "`C:\phylocom\`"
- Copy the file named ***wikstrom.ages*** in "`phylocom-4.1\example_data\blad_j_example\`" to "`C:\phylocom`" and change its name to "***ages***".
- Save the phylogenetic tree generated by Phylomatic in name "***phylo***", and put it in "`C:\phylocom\`"

Step 4 Run phylocom in console

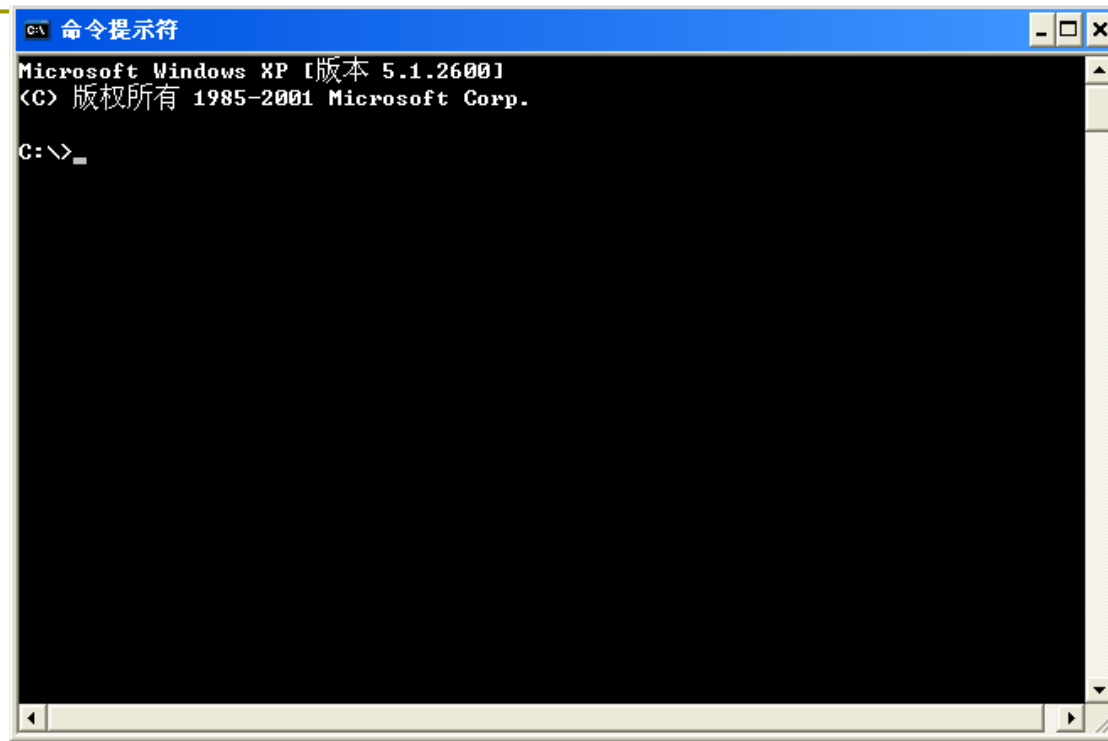


Fig. 10 Screenshot of the console window

- **Start>run>type "cmd"** # start console
- **Type cd C:\phylocom** # change the working directory
- **phylocom bladj>output** # run bladj in phylocom and save to file "output"

Step 5 Plot phylogenetic tree

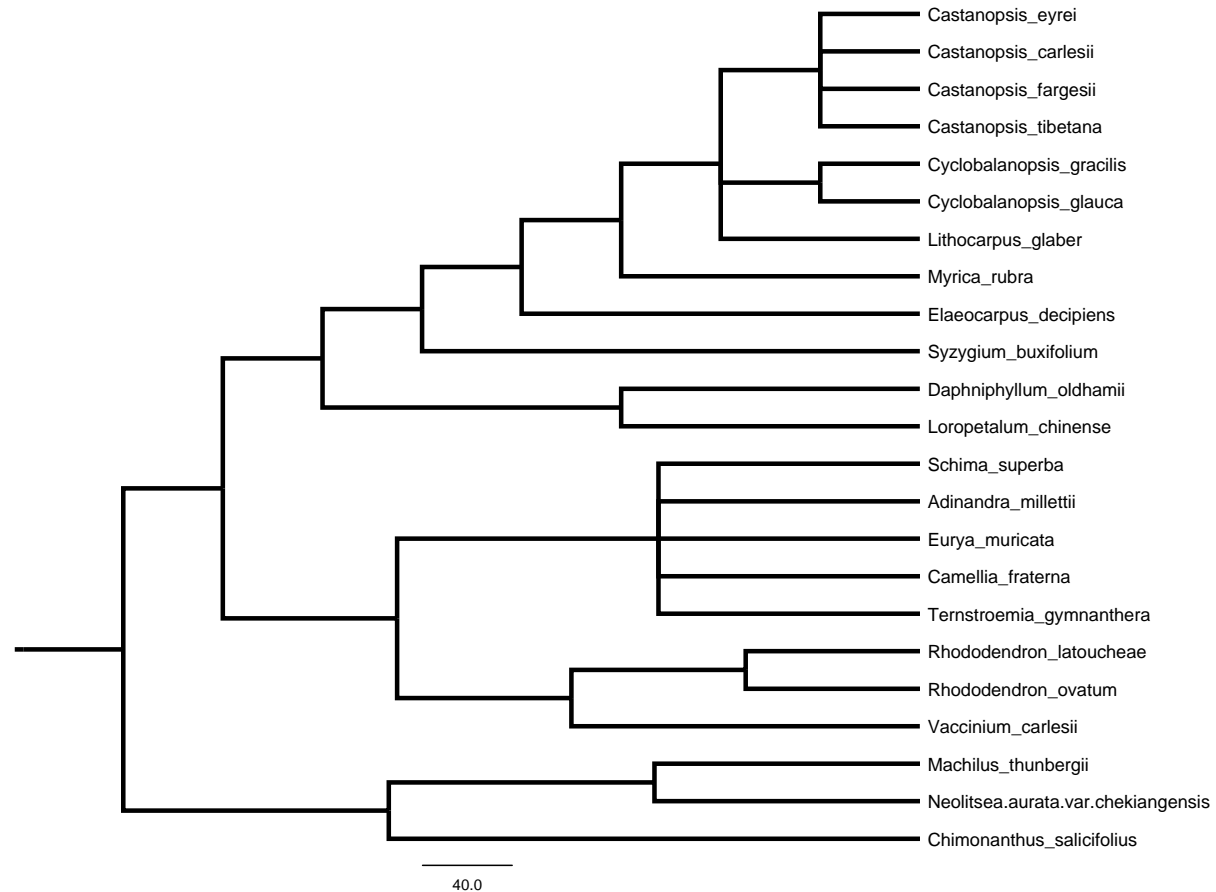


Fig. 11 The phylogenetic tree with branch length

Contents

1. Introduction
2. Phylomatic: a step by step guide
3. **Import phylogenetic trees to R**
4. Dendrograms of functional traits

Import phylogenetic trees to R

Users of ape package often encounter the following error when read the trees generate by Phylomatic.

```
library(ape)
```

```
tre <- read.tree("C:/phylocom/phylo")
```

```
> Error in read.tree("C:/phylocom/phylo"):
```

```
"There is apparently two root edges in your file: cannot read  
tree file.Reading Newick file aborted at tree no.1"
```

In order to read the tree in R by ape, users have to delete the "(" and the ")euphyllphyte:1.000000" at beginning or at the end of the text.

But the ";" at the end of the line should be retained.

Contents

1. Introduction
2. Phylomatic: a step by step guide
3. Import phylogenetic trees to R
4. **Dendrograms of functional traits**

Steps for generating dendrograms of functional traits

Although the shape is similar, result is not a tree but a **dendrogram**.

Building the dendrogram of functional traits in R usually consists of the following steps:

1. Read the matrix of functional traits matrix to R
2. Calculate the distance of functional traits between species
3. Clustering analysis
4. Further analysis: Calculating Functional diversity

Matrix of functional traits

Users have to first build a matrix of functional traits, in which each row for one species and each column for one trait. We will use package FD in which different types of traits and NAs are allowed.

Tab. 2 Matrix of functional Traits

	num1	Num2	fac1	fac2	ord1	ord2	bin1	bin2
species1	9	4.5	A	X	3	2	0	1
species2	8.1	6	A	Z	NA	1	0	1
species3	NA	2.3	C	Y	5	3	1	1
species4	3.2	5.4	B	Z	1	7	0	0
species5	5.8	1.2	C	X	2	6	NA	0
species6	3.4	8.5	C	Y	2	1	1	1
species7	7.5	2.1	B	X	3	2	1	0

Step1 Data input in R

Save the matrix of functional traits as txt file named “*traitmatrix.txt*” to

`"C:/phylocom/"`

MS Excel may facilitate your work.

The R code:

```
setwd("C:/phylocom/")
```

```
traitmatrix <- read.table("traitmatrix.txt", header = TRUE)
```

Step 2. Distances between species in traits's space

Euclidean distance

$$ED_{ab} = \sqrt{\sum_{i=1}^N (x_{ia} - x_{ib})^2}$$

Gower's distance (cope with data with mixed types)

$$GD_{ab} = \frac{\sum_{i=1}^N w_{iab} d_{iab}}{\sum_{i=1}^N w_{iab}}$$

Distance matrix between species

R code:

```
library(FD)
```

```
gowdis(dummy$trait)
```

Tab. 3 Example of distance matrix

	sp1	sp2	Sp3	sp4	sp5	sp6	sp7
sp2	0.218188						
sp3	0.524005	0.667808					
sp4	0.673744	0.561003	0.82257				
sp5	0.529111	0.81457	0.486225	0.484326			
sp6	0.610016	0.593259	0.278474	0.707393	0.606732		
sp7	0.448424	0.686337	0.484866	0.557513	0.302342	0.618784	
sp8	0.407283	0.203944	0.59589	0.239096	0.558553	0.447021	0.703019

Methods of hierarchical clustering

Tab. 4 Methods of hierarchical clustering

Clustering method	Acronym
Single linkage	SL
Complete linkage	CL
Unweighted pair group method using arithmetic averages	UPGMA
Weighted pair group method using arithmetic averages	WPGMA
Unweighted pair group centroid method	UPGMC
Weighted pair group centroid method	WPGMC
Ward's method	Ward

(Mouchet et al., 2008)

Clustering algorithm

`?hclust` for help

method =

"ward", "single", "complete", "average ",

"mcquitty", "median", "centroid"

Test for all the possible combinations of distances and clustering method

(Mouchet *et al.*, 2008).

Example

R code for generating dendrogram of functional traits:

```
library(FD)
dd <- gowdis(dummy$trait)
hresult <- hclust(dd)
plot(hresult, hang = -1)
dendro <- as.dendrogram(hresult)
par(lwd = 2.5, cex = 1.5)
plot(dendro, horiz = TRUE)
```

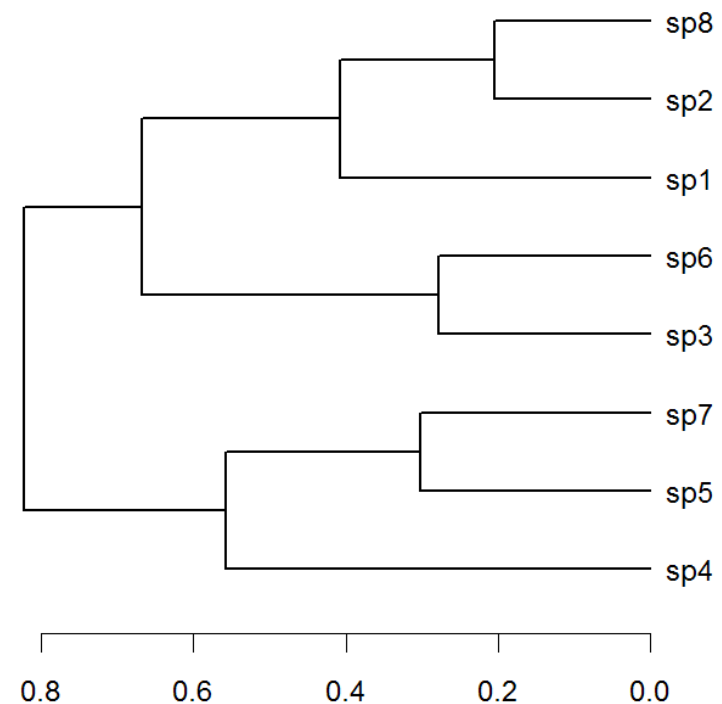


Fig. 12 The relationship between species in functional traits

Summary

1. Phylomatic could generate phylogenetic trees using checklists according to APGIII families.
2. The branch length of the tree generated by Phylomatic could be assigned by BLADJ in PhyloCom.
3. There are various combinations of distances and clustering methods to generate dendrogram of functional traits.

A large, leafy tree stands in the center of a field of tall green grass. In the background, there are hazy mountains under a soft, yellowish sky. The text "Many thanks !" is overlaid in pink on the tree.

Many thanks !