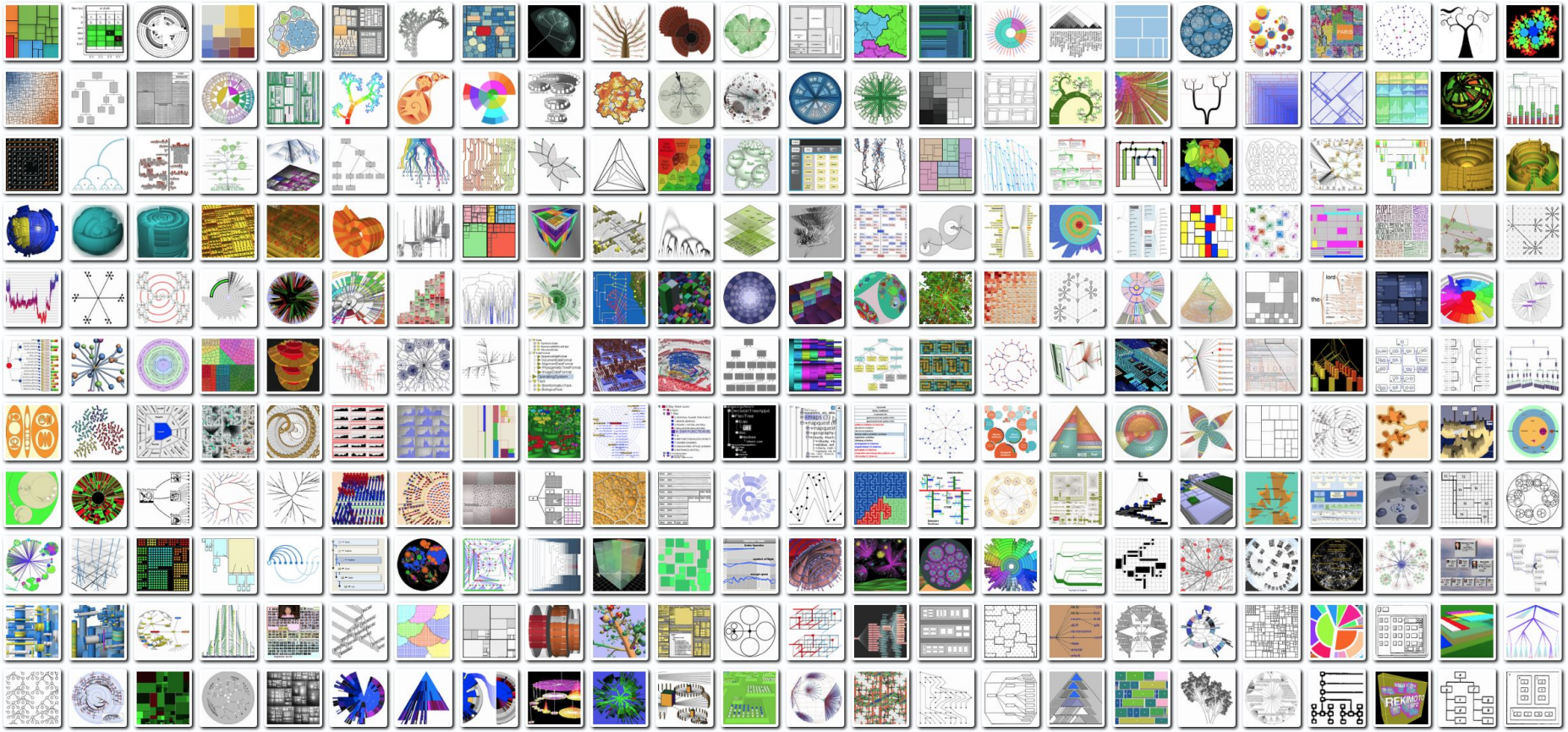
Visualisation des ((très) grands) arbres phylogénétiques* et des données associées

* et autres hiérarchies

@damdevienne
@lifemap_tol

Damien de Vienne
Laboratoire de Biométrie et Biologie Évolutive
CNRS – Université Lyon 1
damien.de-vienne@univ-lyon1.fr



https://treevis.net/

Visualisation d’arbres

Qu’est ce qu’un arbre ?

Quelques outils utiles pour visualiser des arbres (pas trop grands)

Le problème de la visualisation de l’arbre de la vie

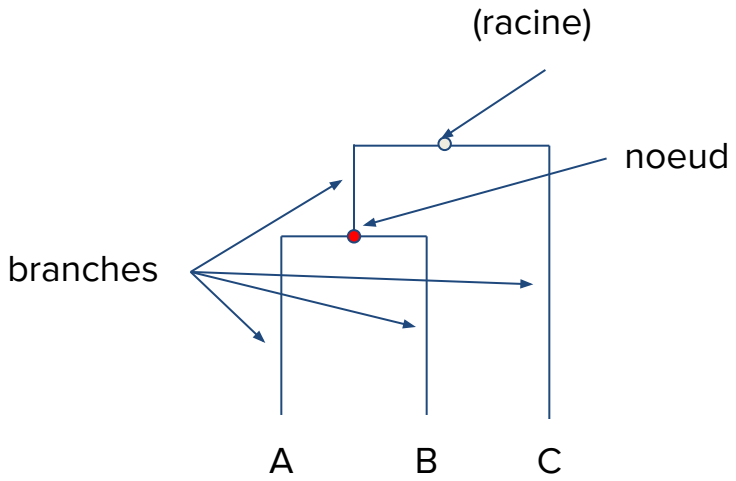
La solution trouvée avec Lifemap

Limites et perspectives de développement

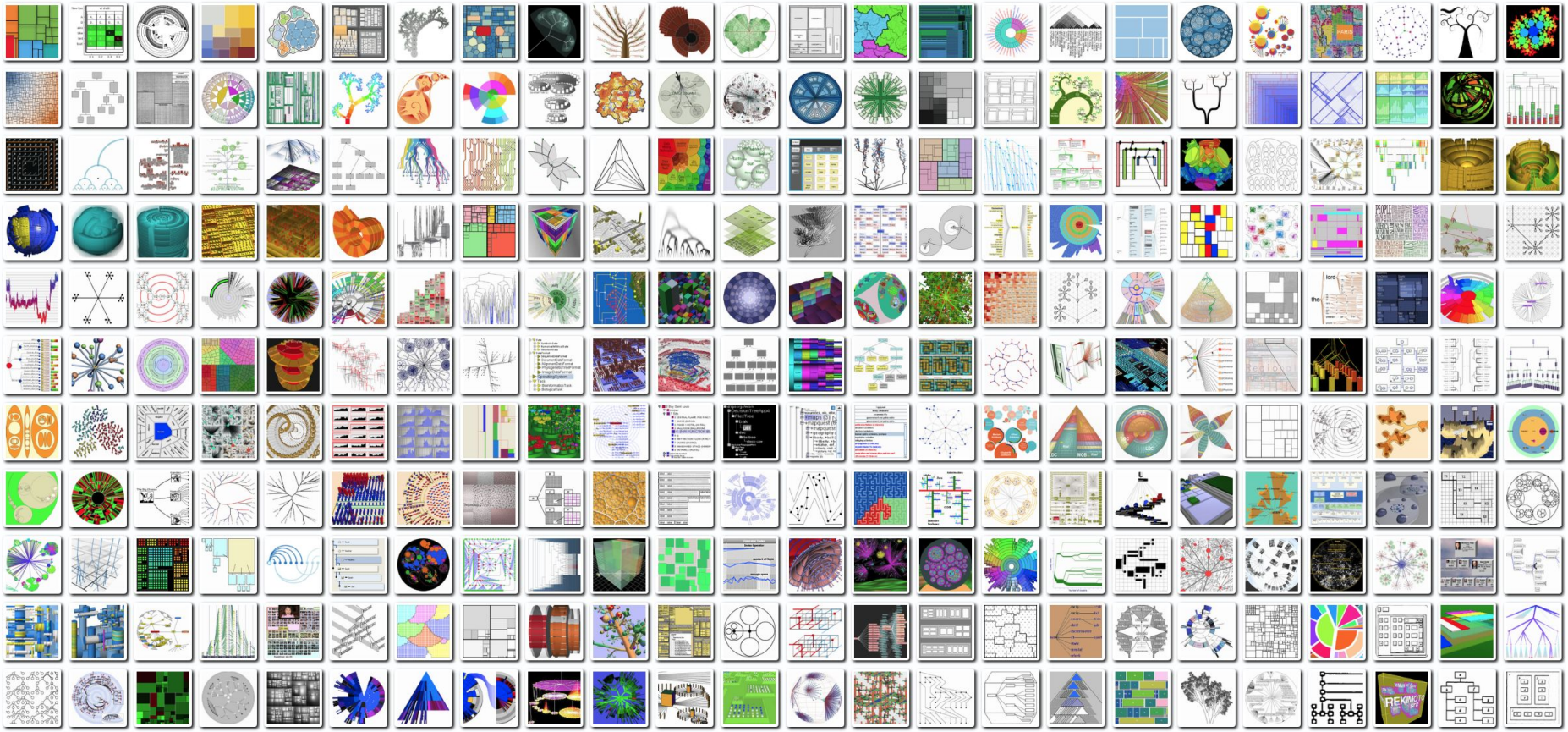
TP

Qu’est-ce qu’un arbre ?

phylogénie
généalogie
système de fichiers
clustering
classification
arbres de décision
mindmap
…

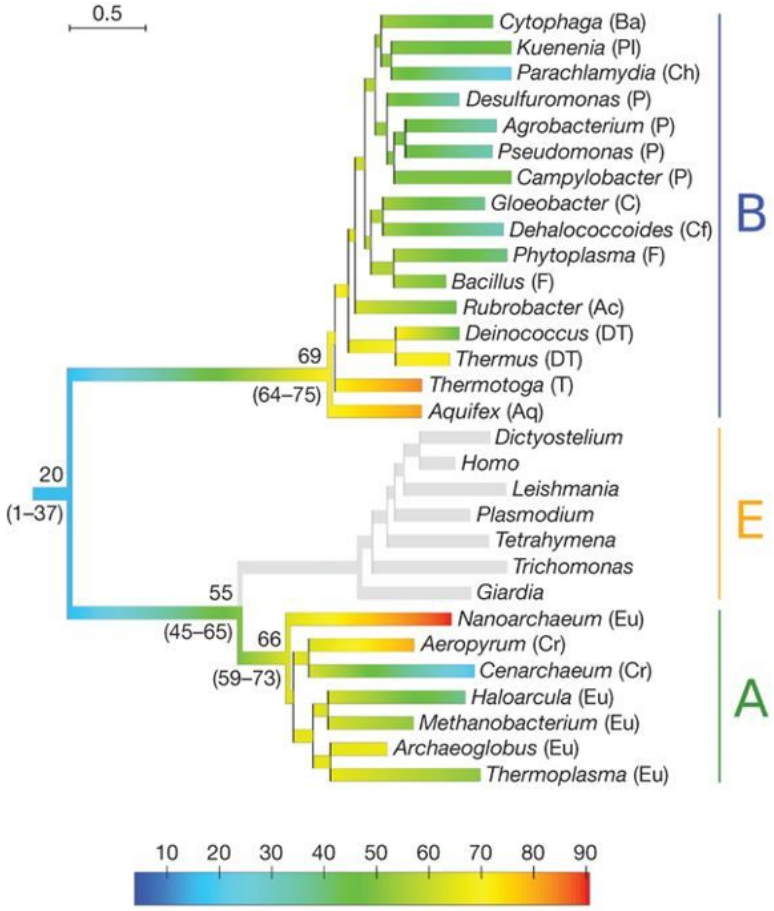


graphe acyclique et connexe



https://treevis.net/

Qu’est-ce qu’un arbre ?



Boussau et al. 2008. *Nature*

Qu’est-ce qu’un arbre ?

L’arbre lui-même qui est une donnée biologique

- racine
- noeuds
- feuilles
- branches

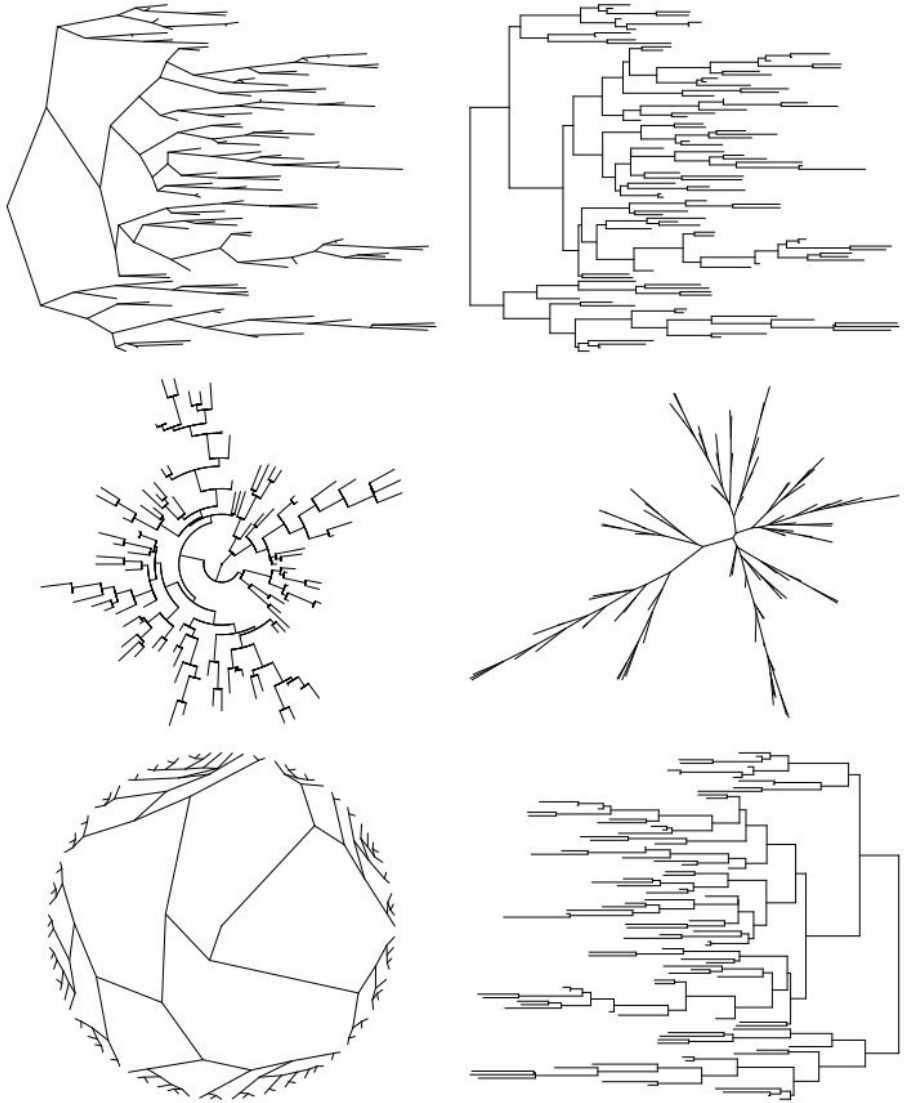
Les informations associées

- à la racine
- aux noeuds
- aux feuilles
- aux branches

Qu’est-ce qu’un arbre ?

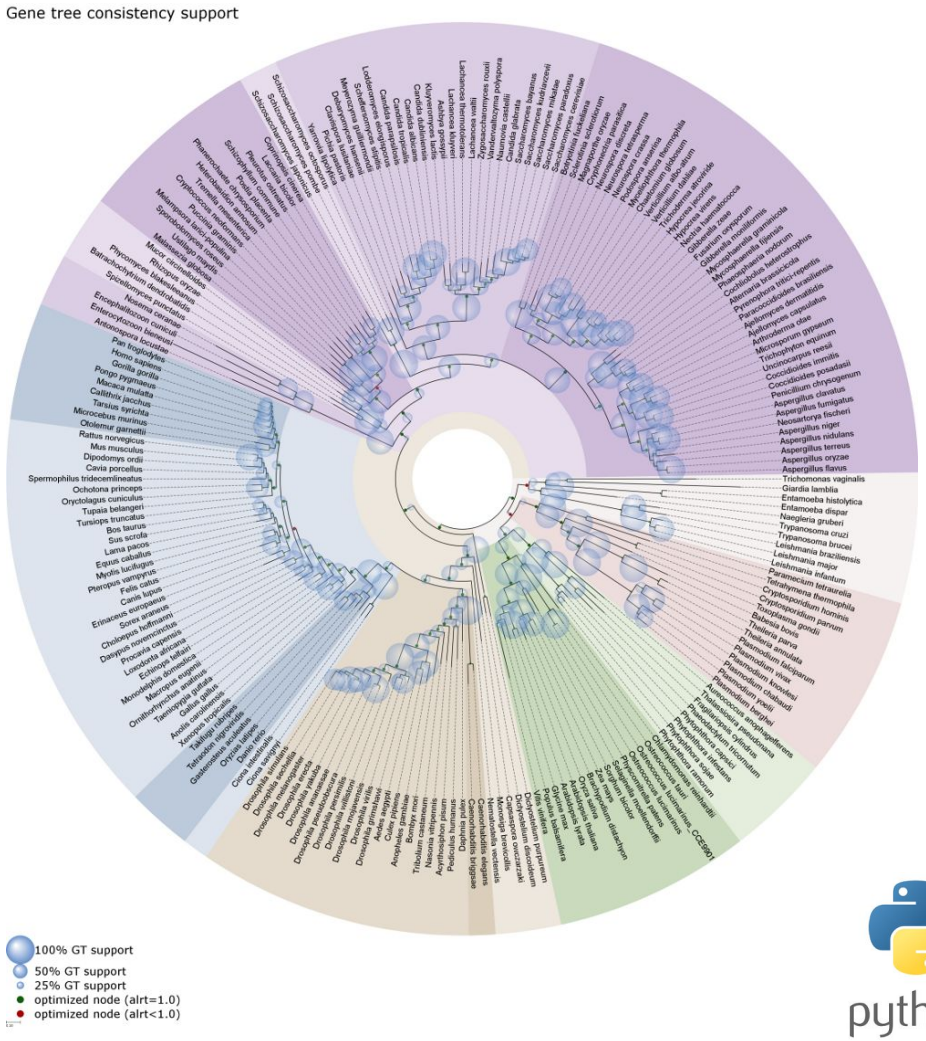
L’arbre lui-même qui est une donnée biologique

- racine
- noeuds
- feuilles
- branches



```
require(ape)
a<-rtree(100)
plot(a, type="cladogram", show.tip.label=F)
plot(a, type="phylogram", show.tip.label=F)
plot(a, type="fan", show.tip.label=F)
plot(a, type="unrooted", show.tip.label=F)
plot(a, type="radial", show.tip.label=F)
```

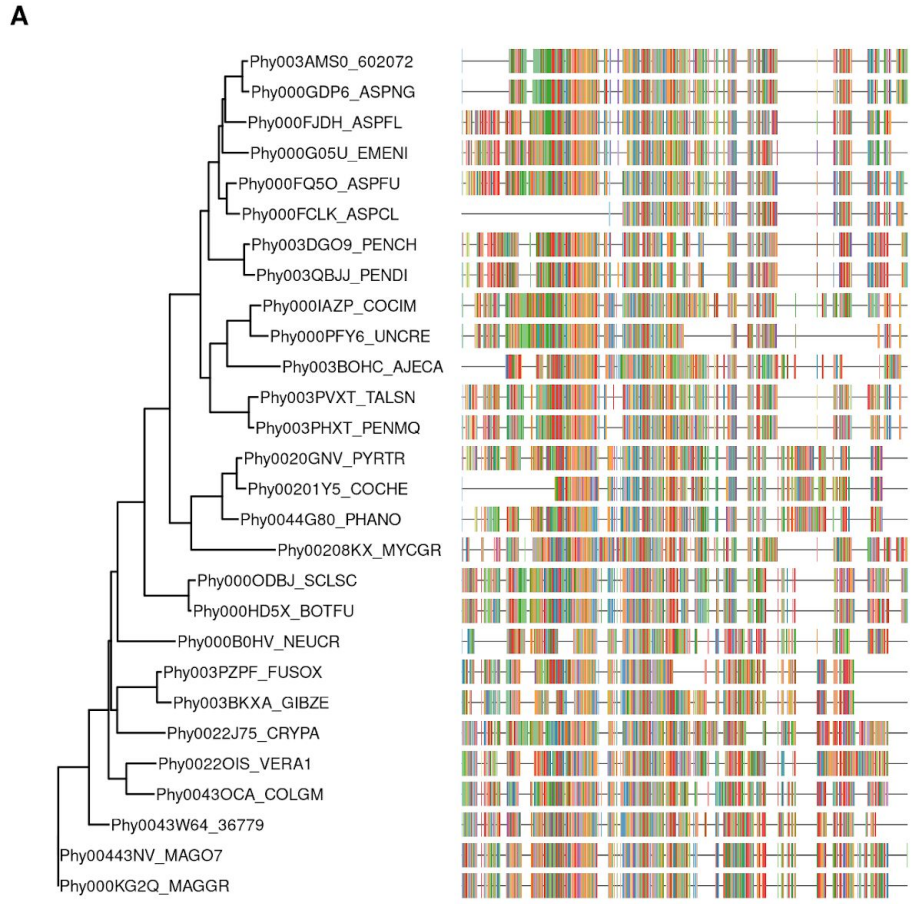
Qu’est-ce qu’un arbre ?



Les informations associées

- à la racine
- aux noeuds
- aux feuilles
- aux branches

Qu’est-ce qu’un arbre ?

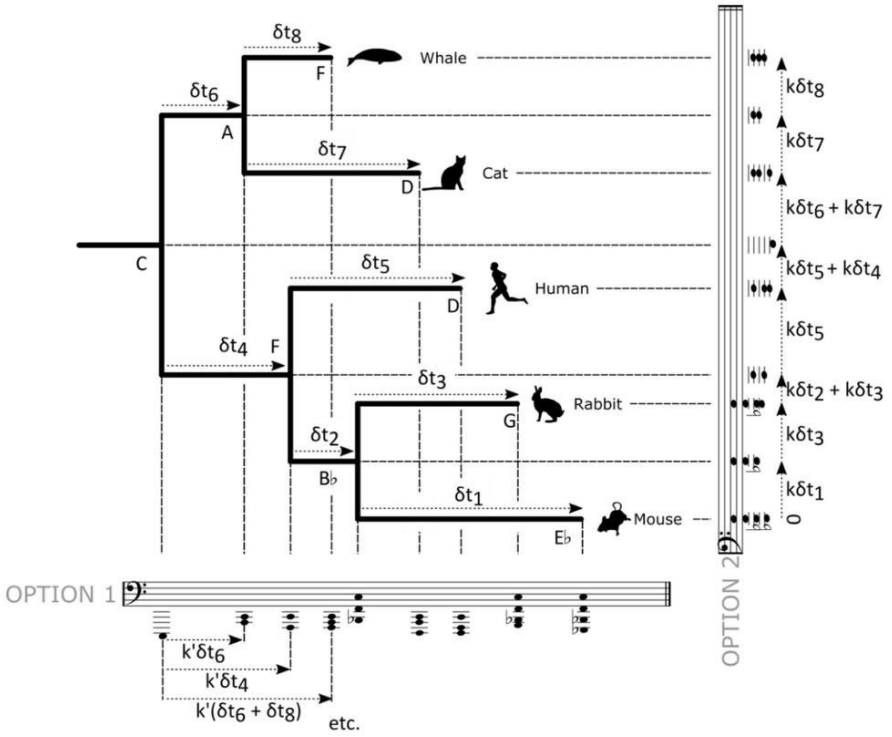


Les informations associées

- à la racine
- aux noeuds
- aux feuilles
- aux branches

```
require(ggtree)
tree <- read.tree("data/tree.nwk")
p <- ggtree(tree) + geom_tiplab(size=3)
msaplot(p, "data/sequence.fasta", offset=3, width=2)
```

Qu’est-ce qu’un arbre ?



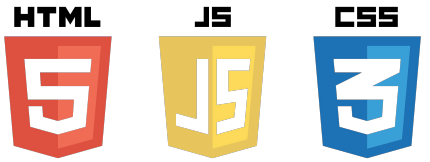
Les informations associées

- à la racine
- aux noeuds
- aux feuilles
- aux branches

Boutin, de Vienne, 2017

Quelques outils utiles pour visualiser les arbres et les infos associées

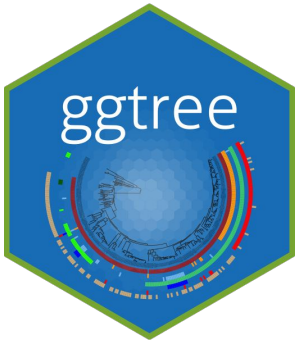
Quelques outils utiles pour visualiser les arbres et les infos associées



ETE Toolkit

http://etetoolkit.org/

https://www.evolgenius.info/evolview



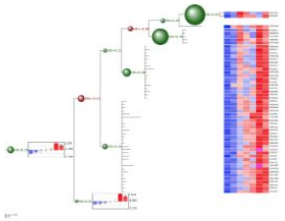
https://guangchuangyu.github.io/software/ggtree/documentation/

https://itol.embl.de/

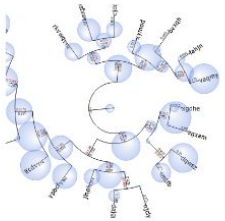
Quelques outils utiles pour visualiser les arbres et les infos associées

ETE Toolkit http://etetoolkit.org/

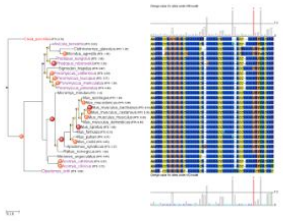
A Python framework to work with trees



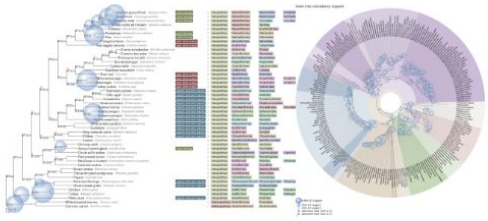
clustering with node validation and heatmap



A bubble tree map

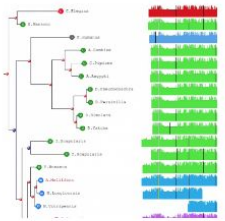


Tree showing the result of an adaptation test with CodeML

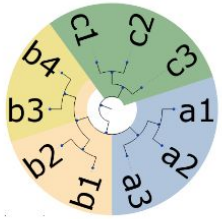


A birds tree with node labeling and NCBI taxonomy lineage

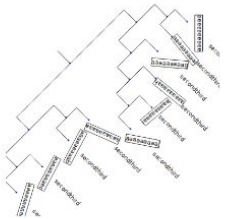
A tree of life representation



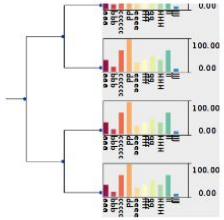
Custom tree showing intron information using aligned faces



standard circular tree representation

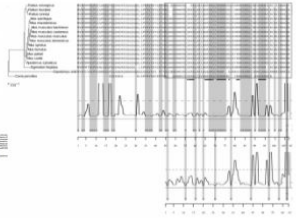


Tree using rotated faces

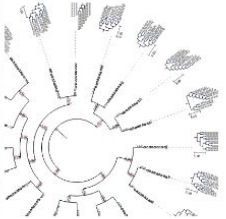


Tree using BarChart faces

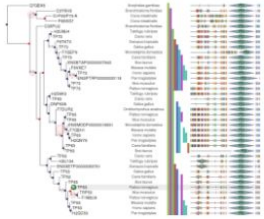
Per node analysis of a fungal tree of life



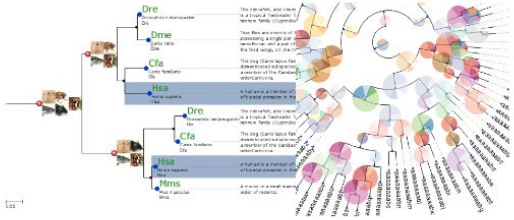
tree and site selection histogram



A tree of trees using TreeFace

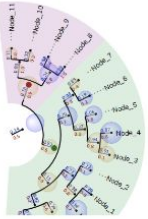


a tree showing domain structure and taxonomic classificaiton

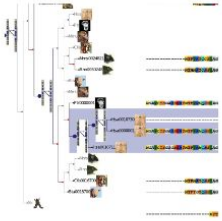


tree with custom annotations

PieChart tree example



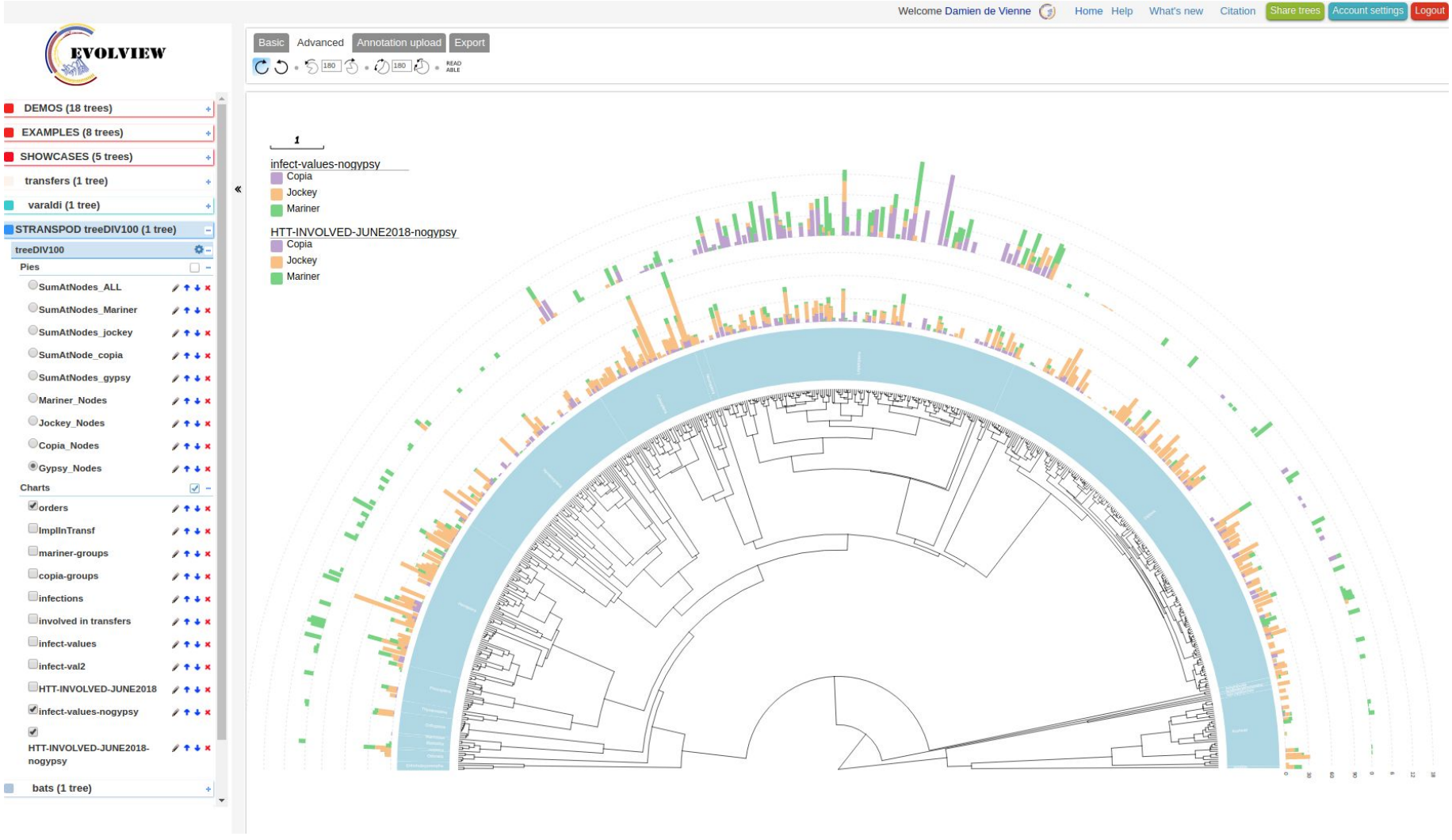
A semicircular tree with dynamic node styles



Reconciled tree visualization with node images

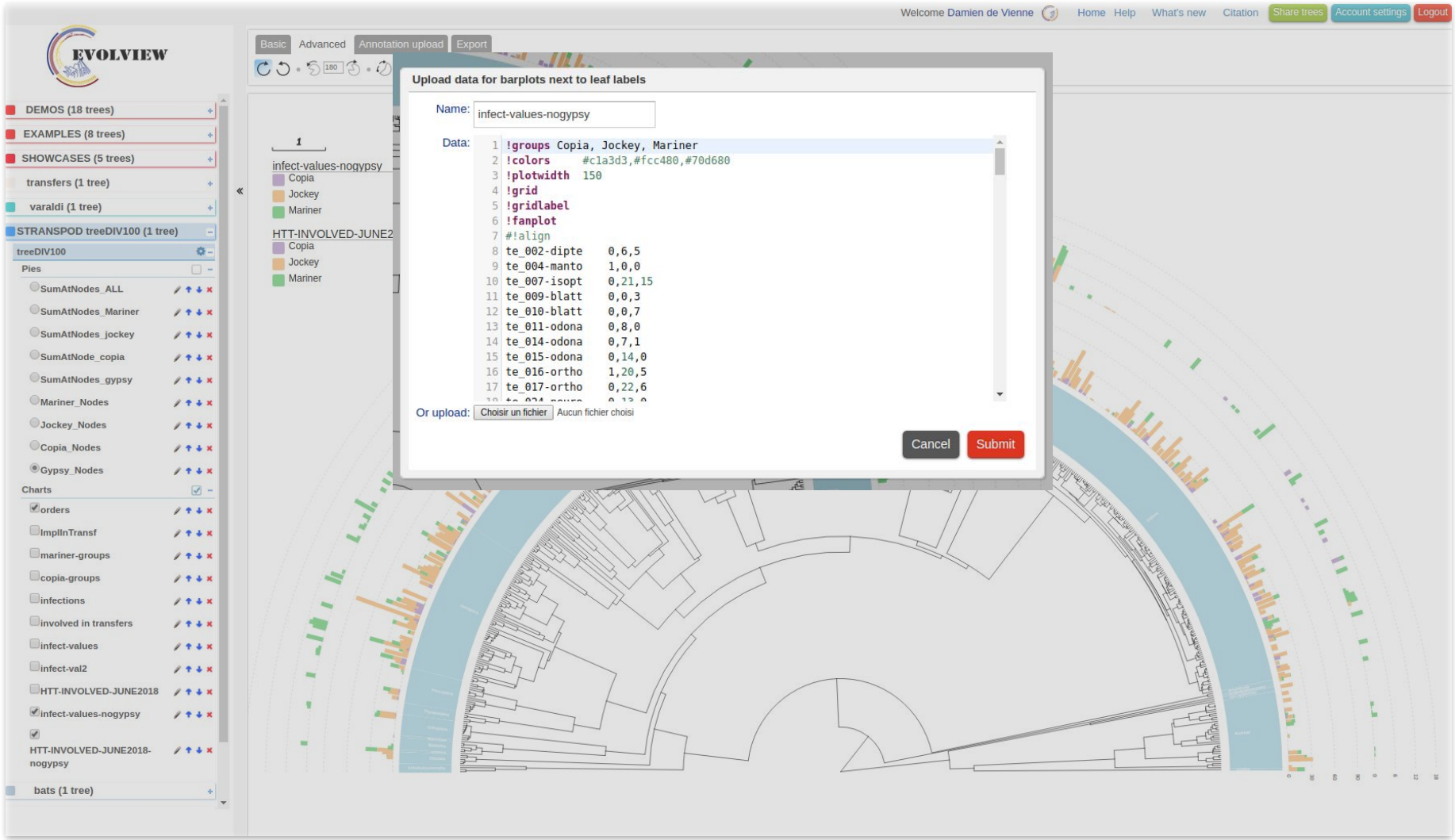
Quelques outils utiles pour visualiser les arbres et les infos associées

https://www.evolgenius.info/evolview/

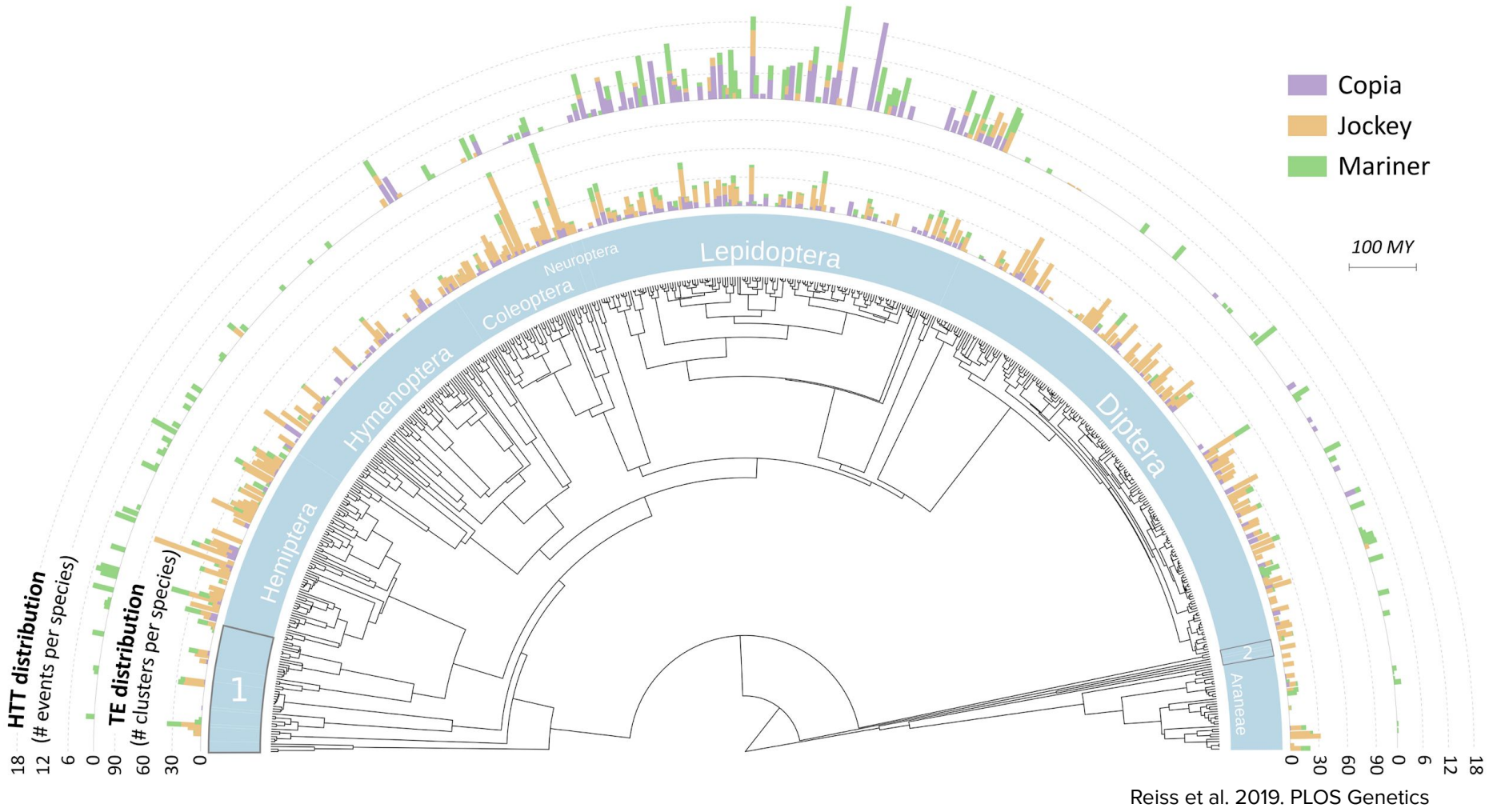


Quelques outils utiles pour visualiser les arbres et les infos associées

https://www.evolgenius.info/evolview/

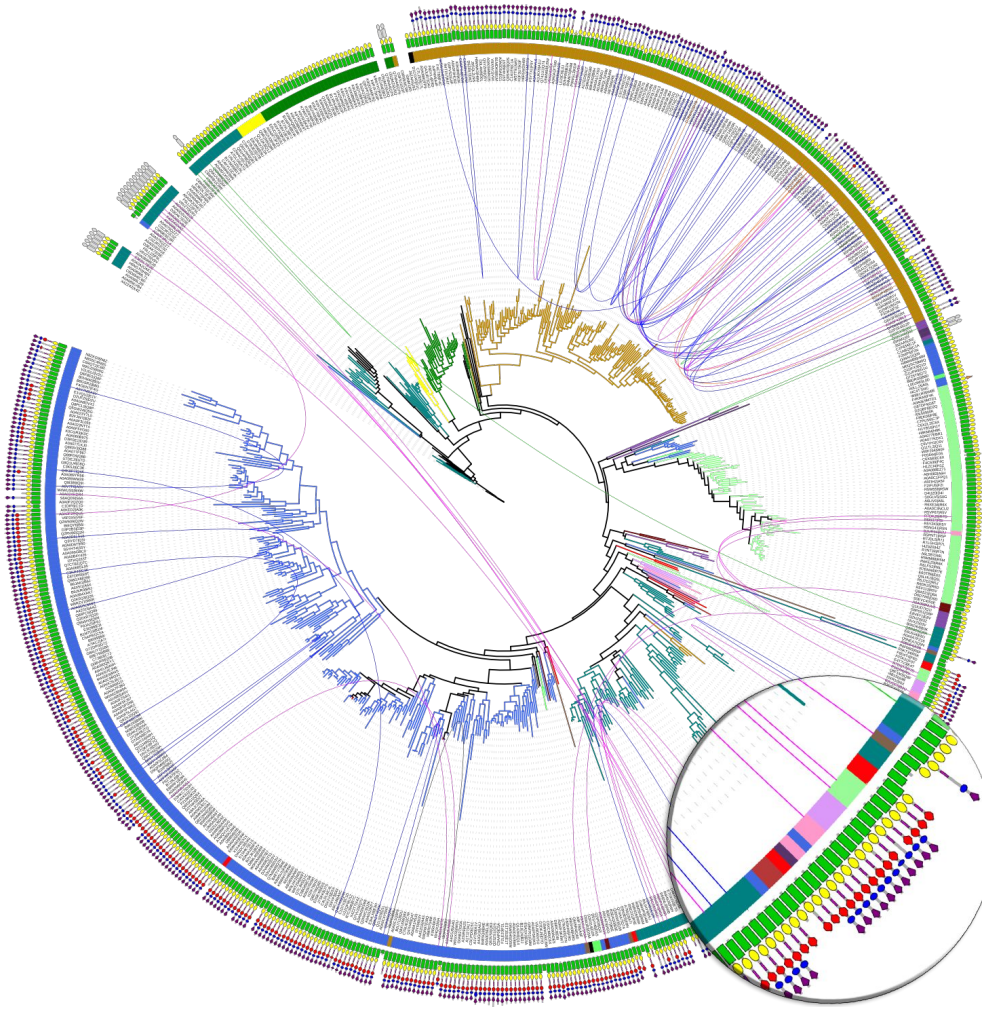


Le “problème” posé par l’arbre de la vie (et tous les grands arbres)

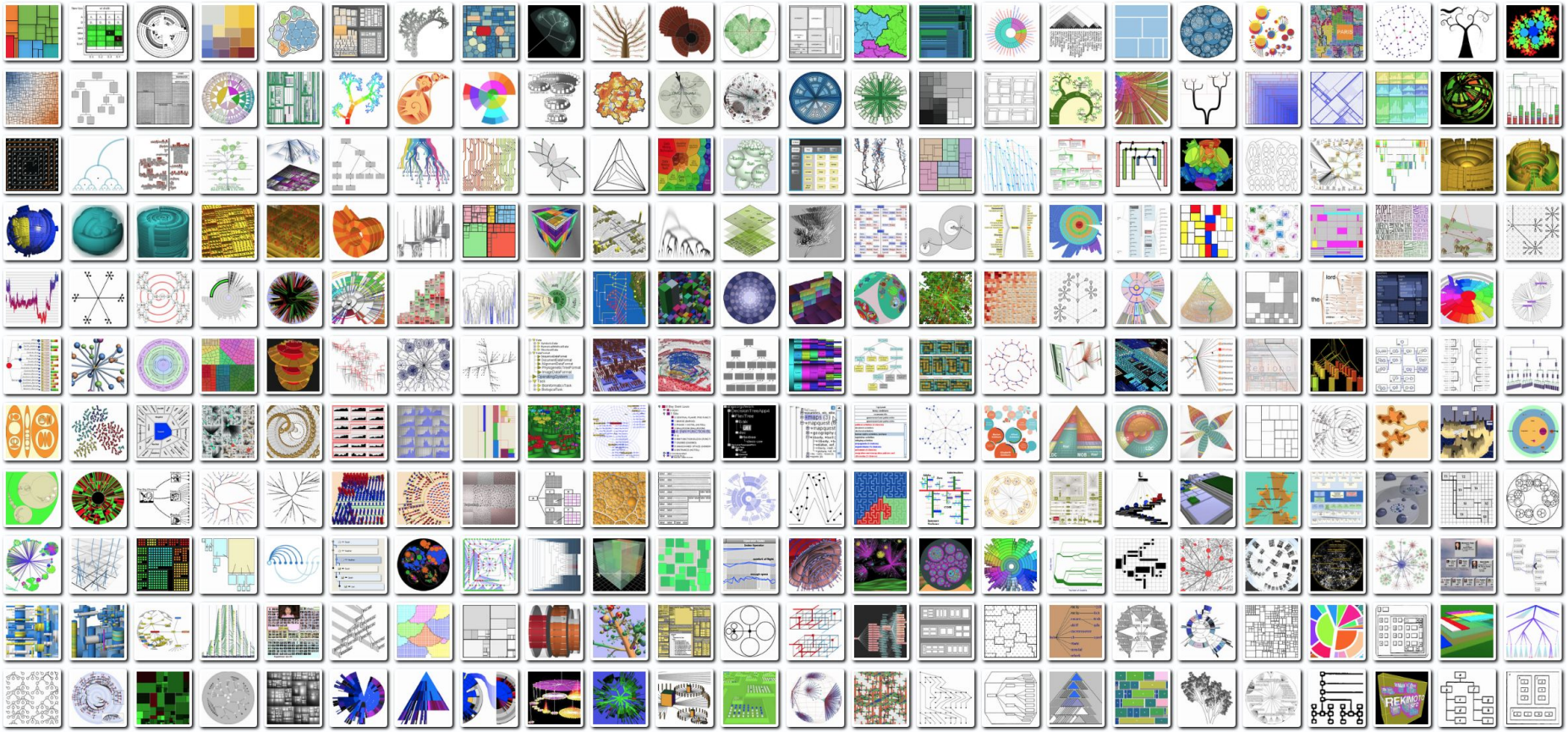


Reiss et al. 2019. PLOS Genetics

Le “problème” posé par l’arbre de la vie



https://itol.embl.de/img/gallery/ex6.png



https://treevis.net/

Lifemap (lifemap.univ-lyon1.fr)

https://treevis.net/

Lifemap

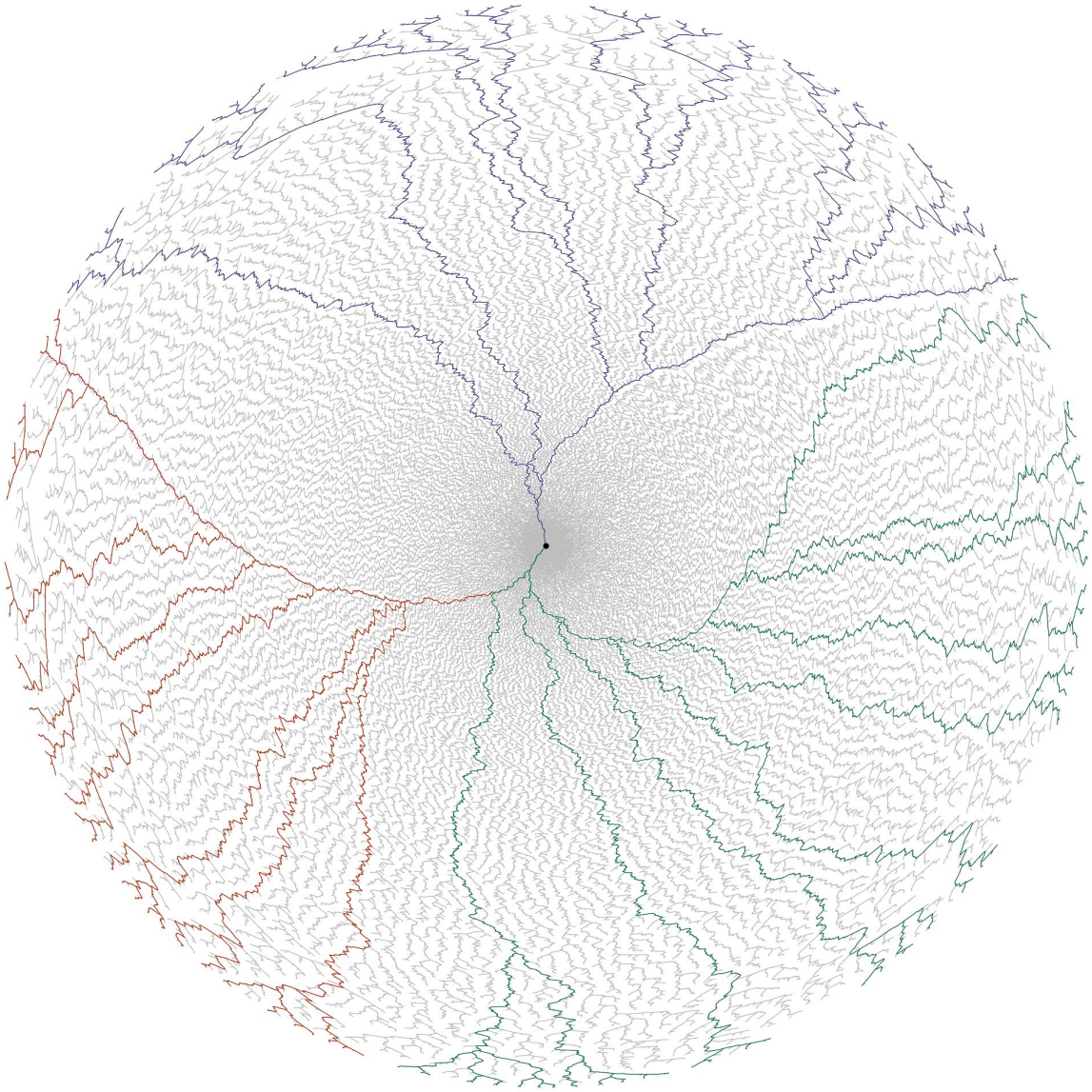
Deep Zoom

Dissocier l’arbre et les données additionnelles

LBBE
BIOMETRIE ET BIOLOGIE EVOLUTIVE
cnrs

Arbre du vivant

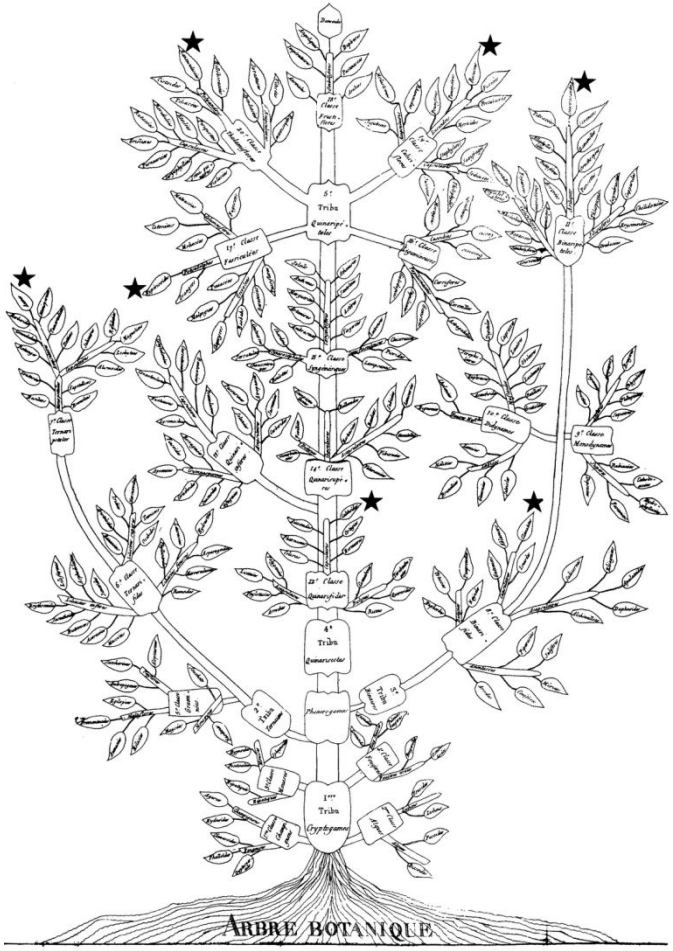
Une représentation unique des liens évolutifs entre toutes les espèces de la planète, espèces éteintes et espèces vivantes



Duchemin W, de Vienne DM. Exposition SCUBE. 2017

L’arbre du vivant

avant Darwin



“I worked for a long time trying to fit families into a continuous series but I found great difficulty… finally I solved it by separating the branches in two and was successful at least at joining them at their bases. At that point I became convinced that plants form different series united at the base, a gradation as that of tree branches”

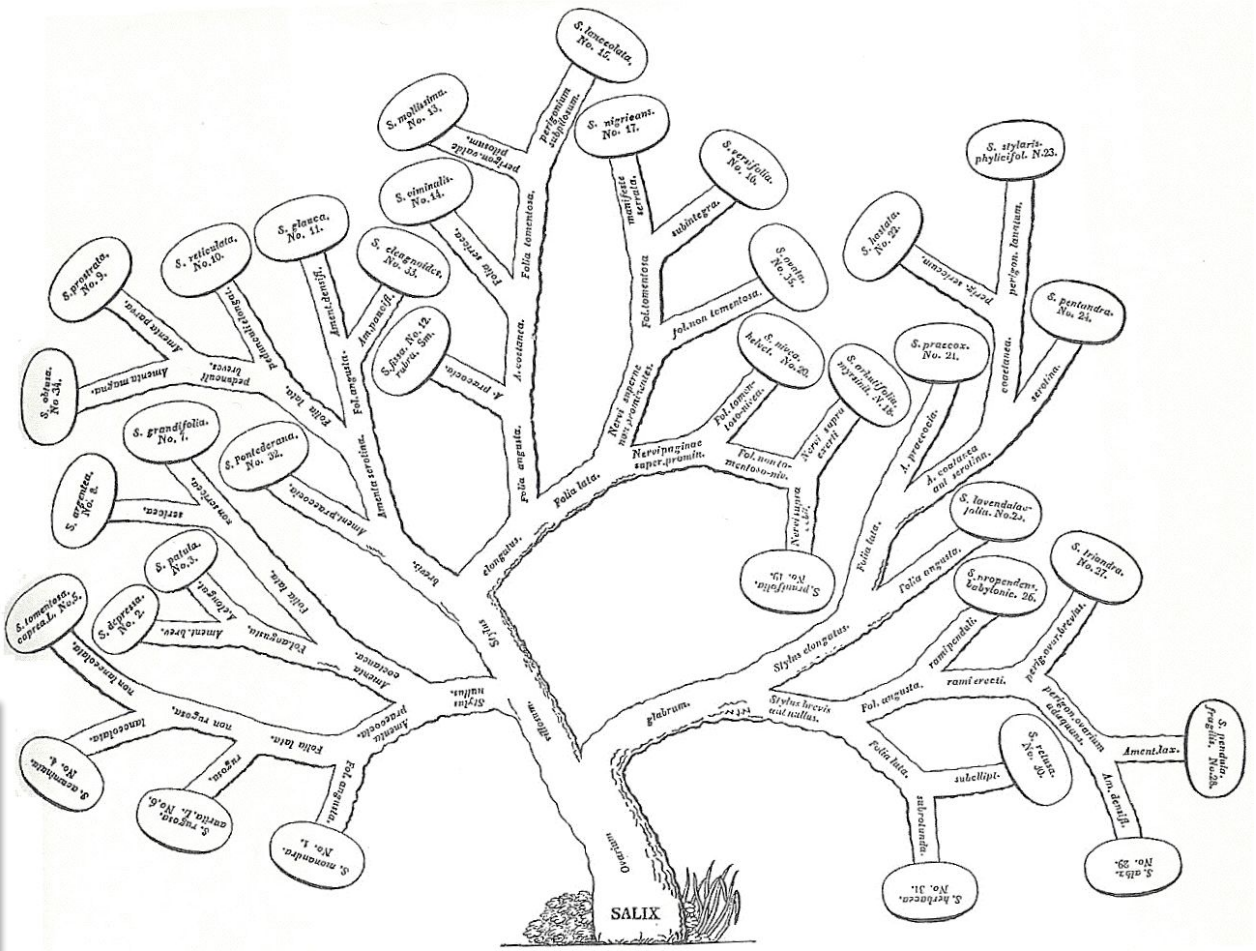
Plant families

Augustin Augier, 1801

L’arbre du vivant

avant Darwin

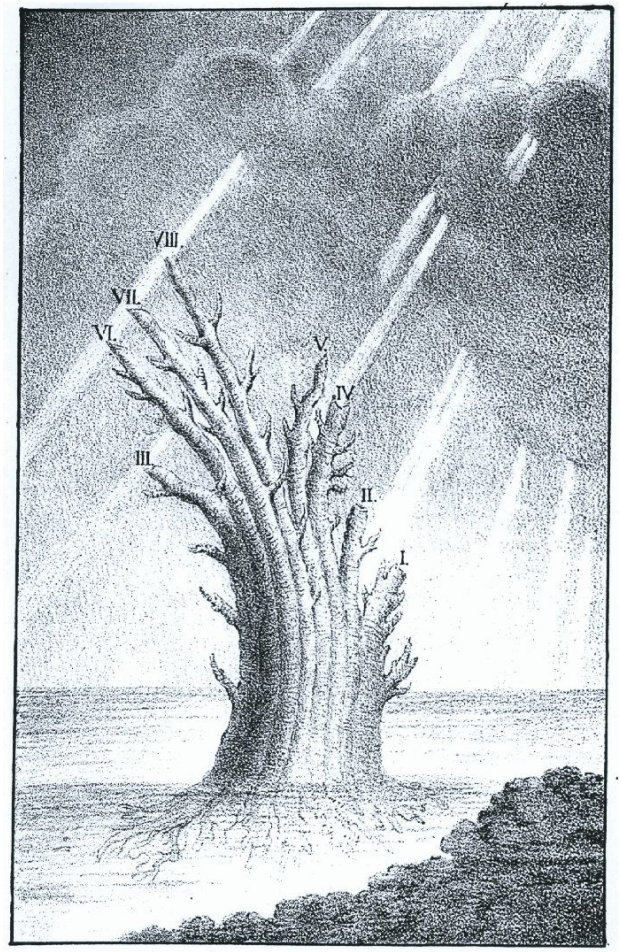
Saules



Nicolas Charles Seringe, 1815

L’arbre du vivant

avant Darwin

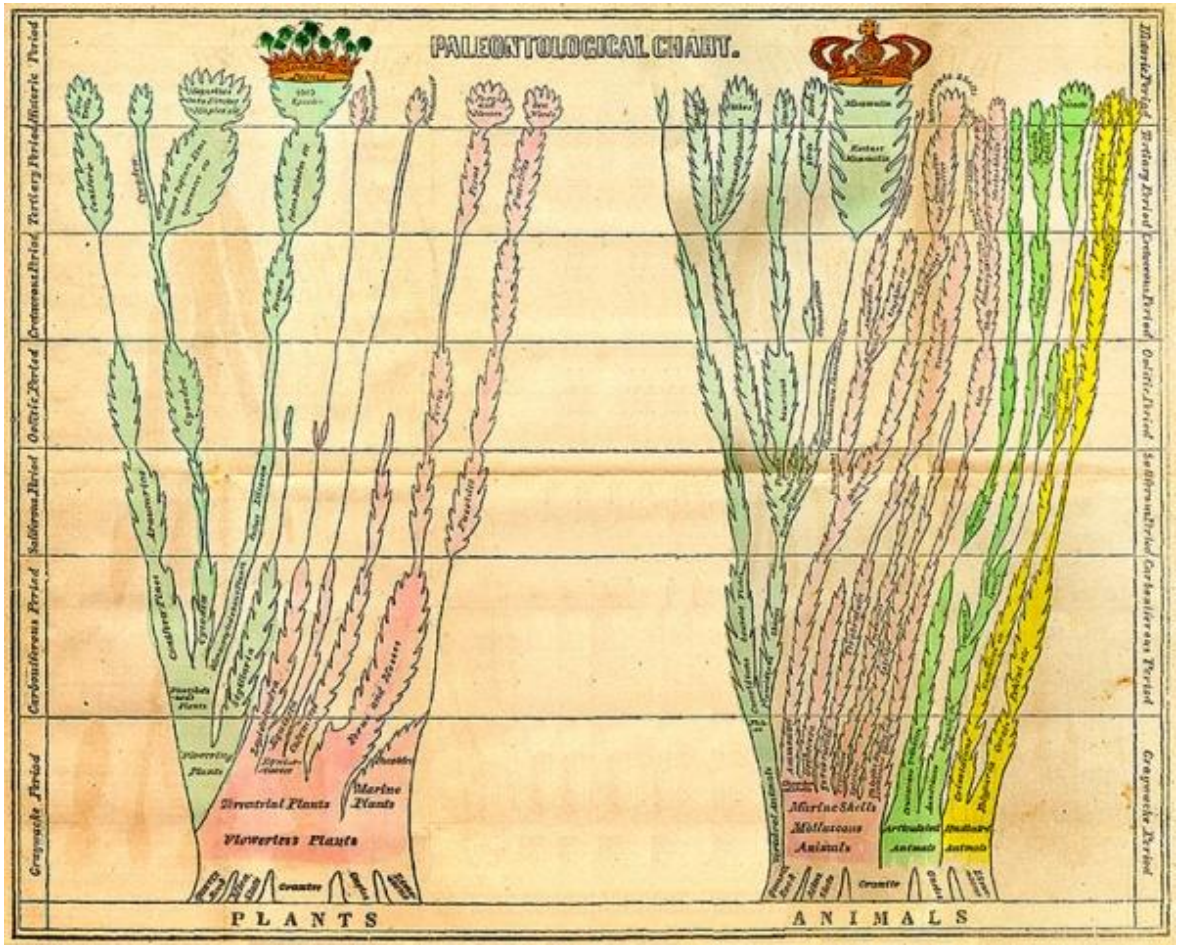


Animal ‘types’

Eichwald, 1829

L’arbre du vivant

avant Darwin

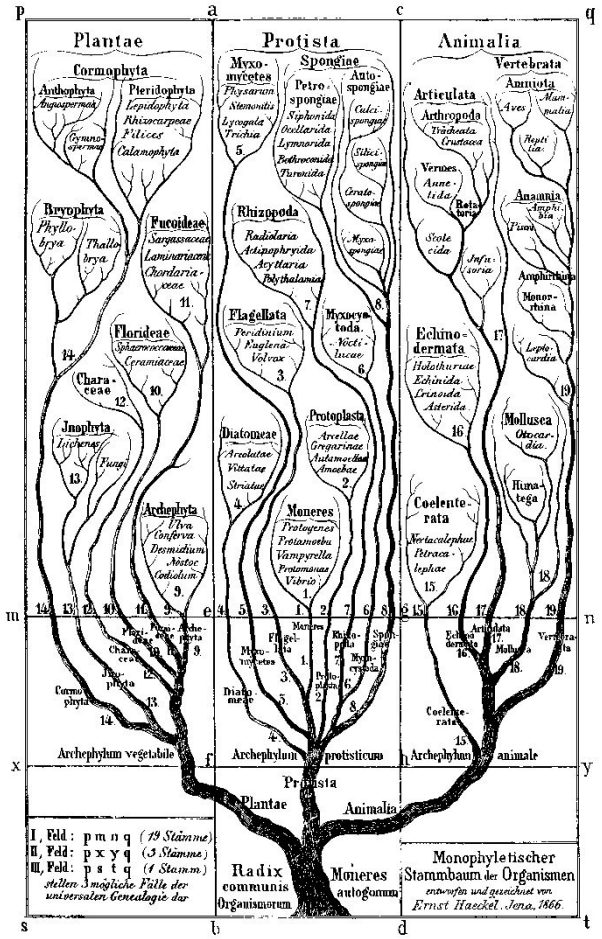


Fossil times, extinctions, abundance, ‘relationships’

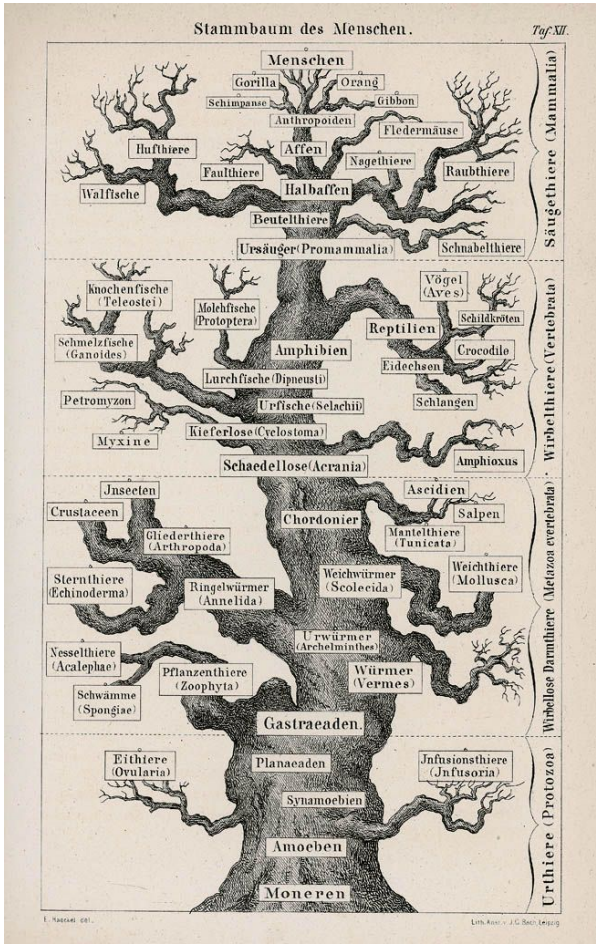
Hitchcock, 1840 (ed. 1852)

L’arbre du vivant

après Darwin



Haeckel *et al.*, 1866

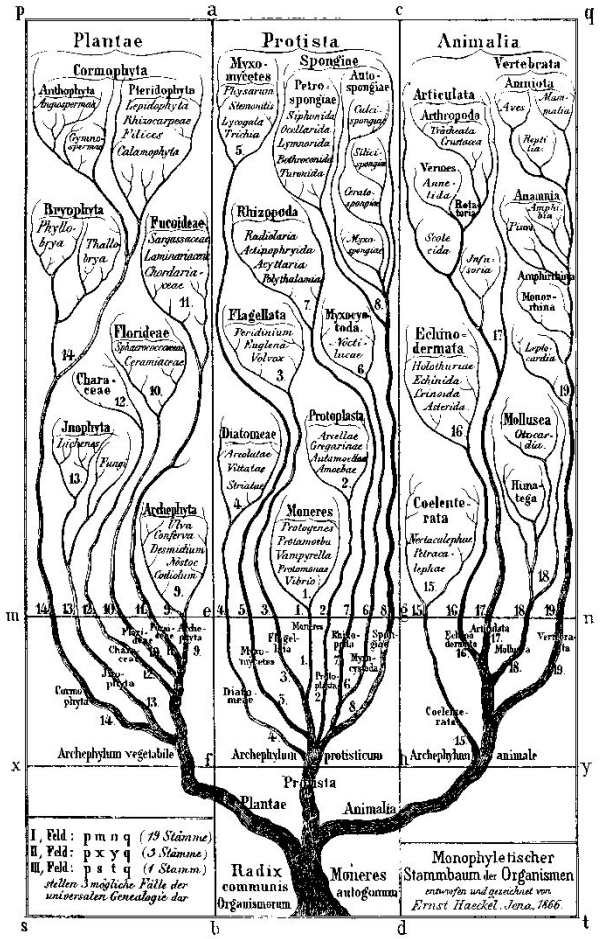


Haeckel, 1874 (?)

Animals, plants, protists

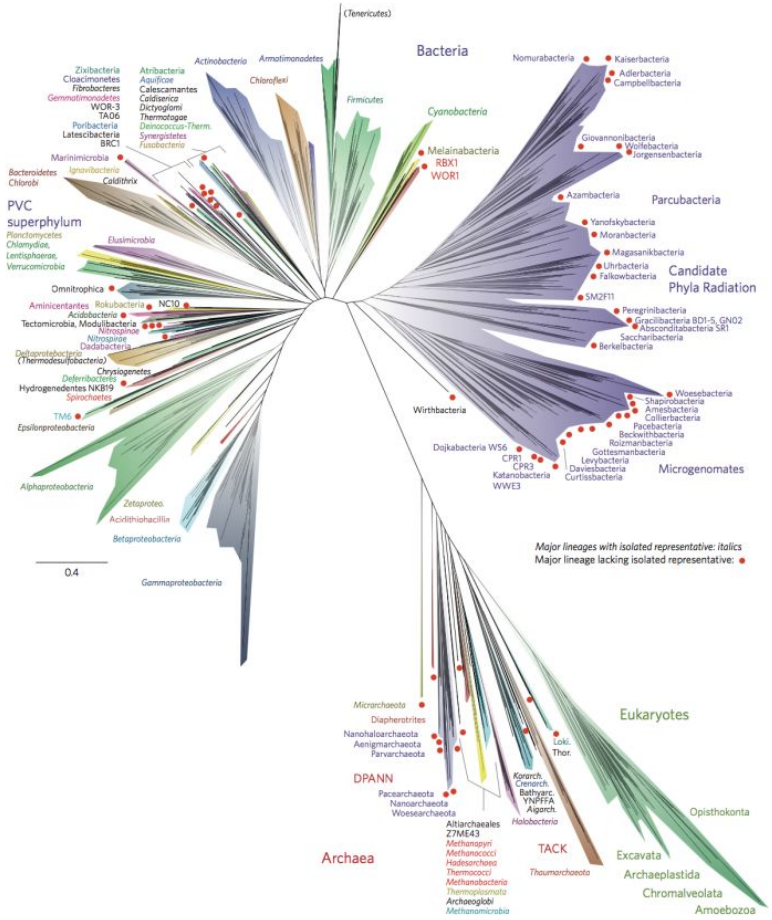
L’arbre du vivant

après Darwin



Haeckel *et al.*, 1866

150 years



Hug *et al.*, 2016

The Tree of Life

- Huge efforts to assemble a comprehensive Tree of Life
  - More sequences
  - Metagenomics
  - Improved reconstruction methods
  - Improved aggregation methods
  - …

The Tree of Life

OPEN Tree of Life

Hilnchliff *et al.*, 2015

Catalogue of Life

WIKISPECIES
free species directory

All-species Living Tree Project by silva
high quality ribosomal RNA databases

Yarza *et al.* 2008, 2010

NCBI Taxonomy

eol
Encyclopedia of Life

Wilson, 2003

reviewed in Madisson *et al.*, 2007

The Tree of Life

OPEN Tree of Life

Hilnchliff *et al.*, 2015

Catalogue of Life

WIKISPECIES
free species directory

All-species Living Tree Project by silva
high quality ribosomal RNA databases

Yarza *et al.* 2008, 2010

NCBI Taxonomy

eol
Encyclopedia of Life

Wilson, 2003

reviewed in Madisson *et al.*, 2007

The Tree of Life

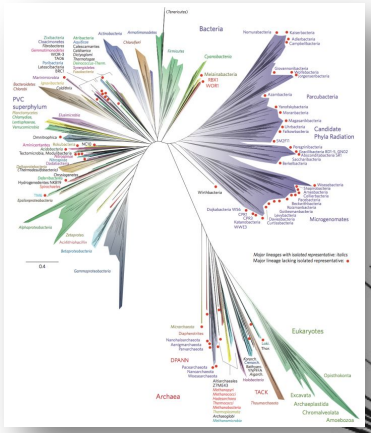
Huge efforts to assemble a comprehensive Tree of Life...

- More sequences
- Metagenomics
- Improved reconstruction methods
- Improved aggregation methods
- ...

...But almost none to be able to visualize and explore it

It is difficult to visualize large trees

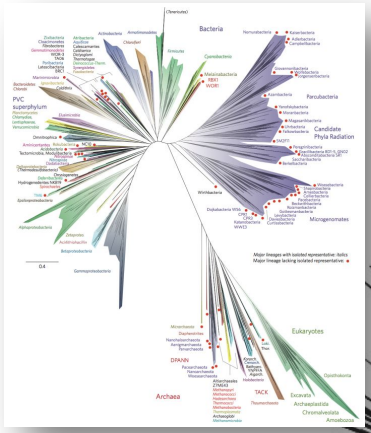
zooming in is not enough



Random tree
10 000 tips

It is difficult to visualize large trees

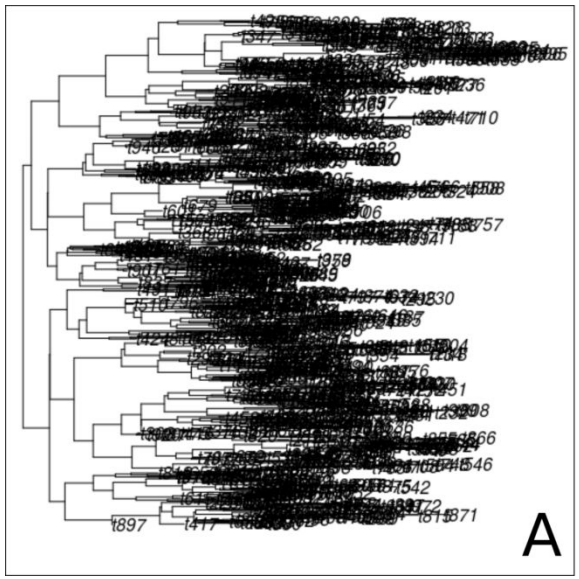
zooming in is not enough



Random tree
10 000 tips

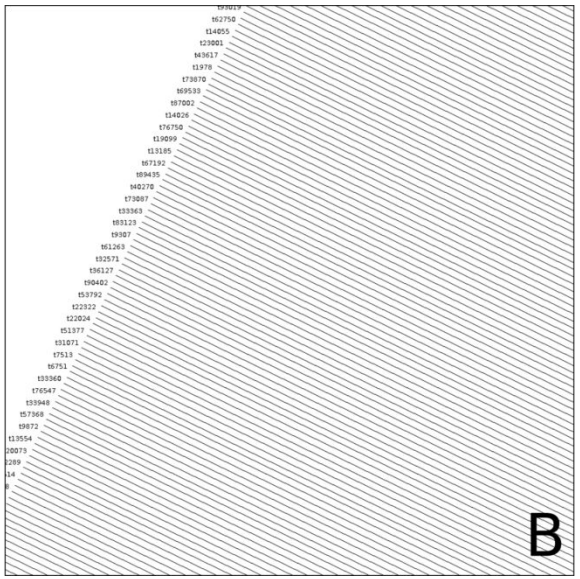
It is difficult to visualize large trees

Labels are problematic…



‘ape’ package (Paradis et al. 2004)

Labels overlap
Not interactive
No additional information
No links to external resources

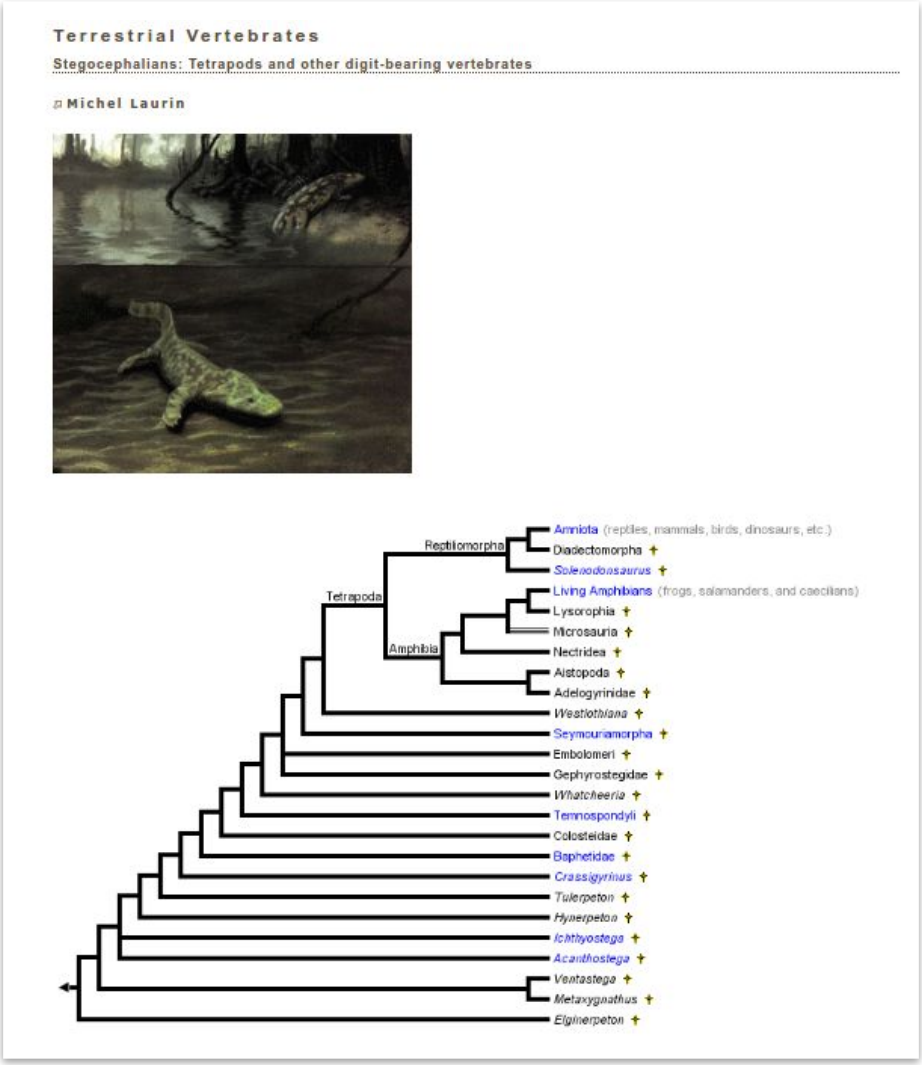


Dendroscope (Huson *et al.* 2007)

Loss of branching pattern when labels are visible
No additional information
No links to external resources

It is difficult to visualize large trees

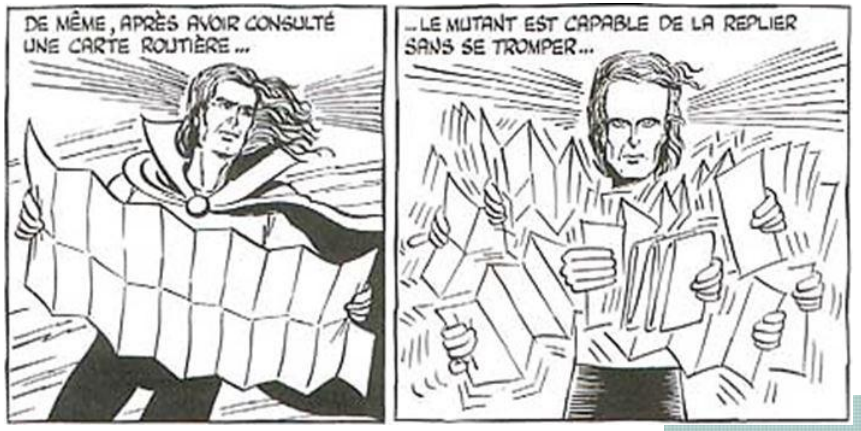
Multi-page approaches are not convenient



www.tolweb.org

The ‘revolution’ in tree visualization

parallel with cartography

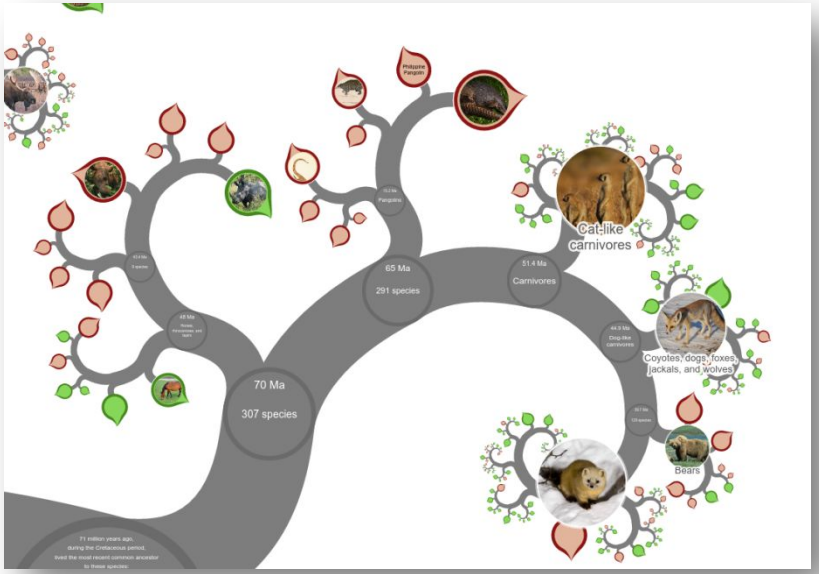


Gotlib, 1996, Rubrique-à-Brac, tome 2

Suggested by Westneat (2009) and Page (2012) for trees

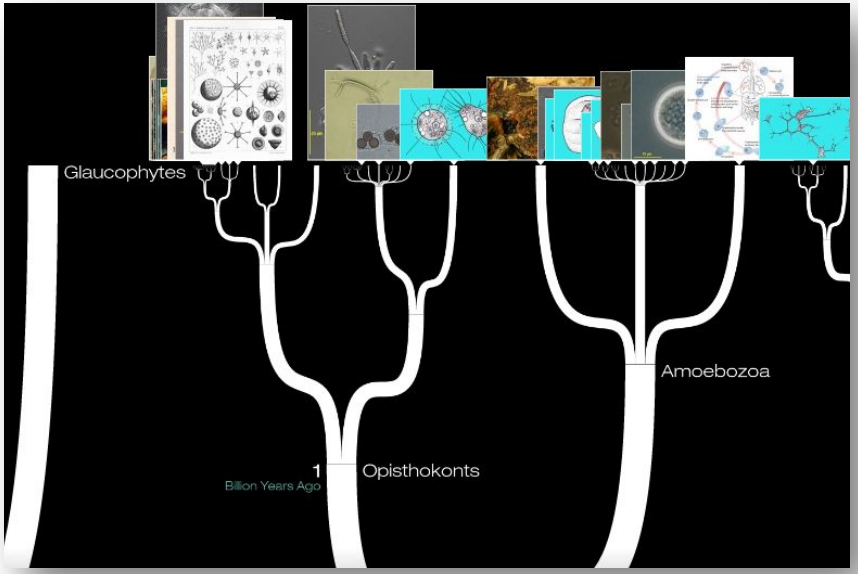
Deep-zooming

The ‘revolution’ in tree visualization



OneZoom

Rosindell and Harmon, 2012



DeepTree

Block *et al.*, 2012

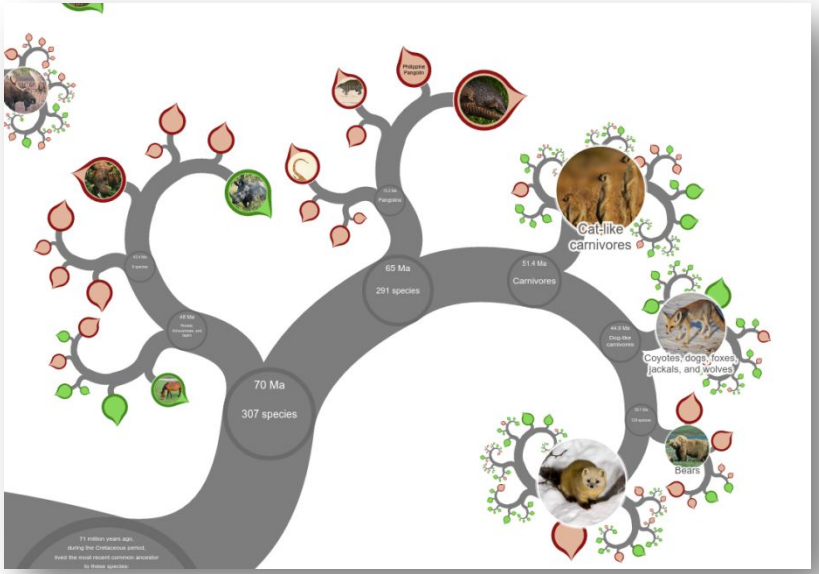
Only for education / communication purpose (not a tool for researchers)

No multifurcations
(80% of the nodes in the ToL)
Easy to get lost

Cannot handle more than 400,000 species

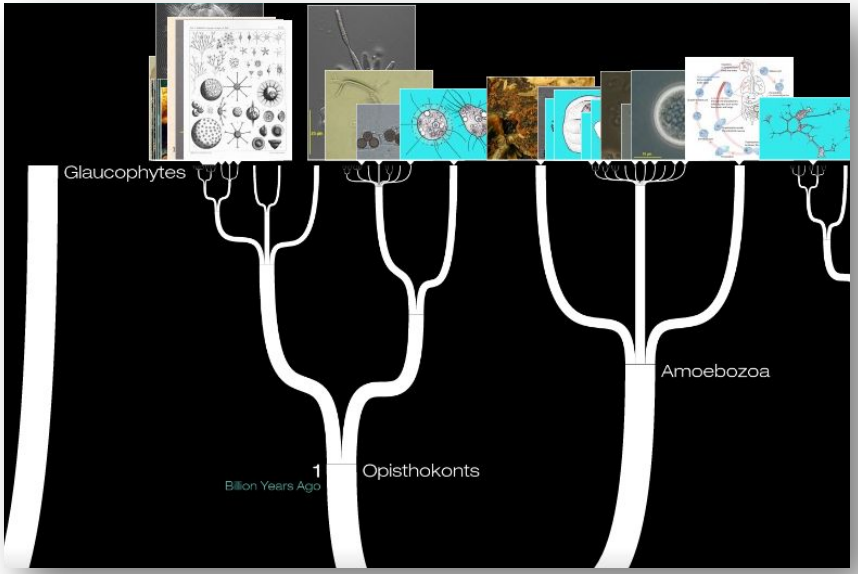
Limited number of species
(70 000 *vs* 2 Million in the ToL)
Lots of computation while navigating
Fixed (unique) tree

The ‘revolution’ in tree visualization



OneZoom

Rosindell and Harmon, 2012



DeepTree

Block *et al.*, 2012

Only for education / communication purpose (not a tool for researchers)

No multifurcations
(80% of the nodes in the ToL)
Easy to get lost

~~**Cannot handle more than 400,000 species**~~

Limited number of species
(70 000 *vs* 2 Million in the ToL)
Lots of computation while navigating
Fixed (unique) tree

The need for a new visualization

- No limitation in the amount of data (# species)
- Smooth and fast
- Allow multifurcations
- Tree-looking

- Links to additional resources
- Serve as a background for additional information (as layers) TP

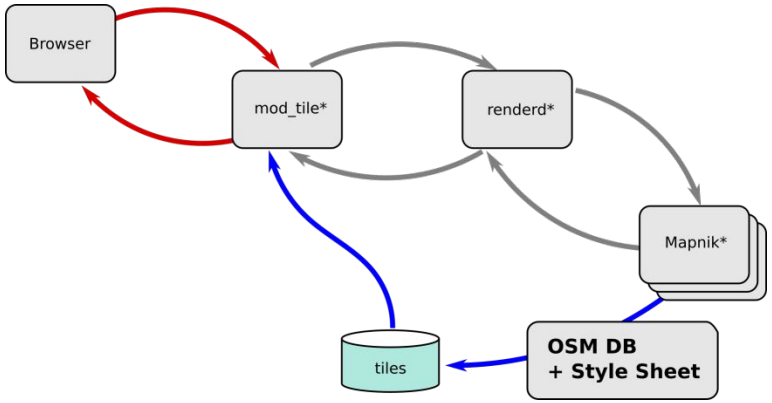
- Possibility to visualize your own trees

- For large public, education and research

The need for a new visualization

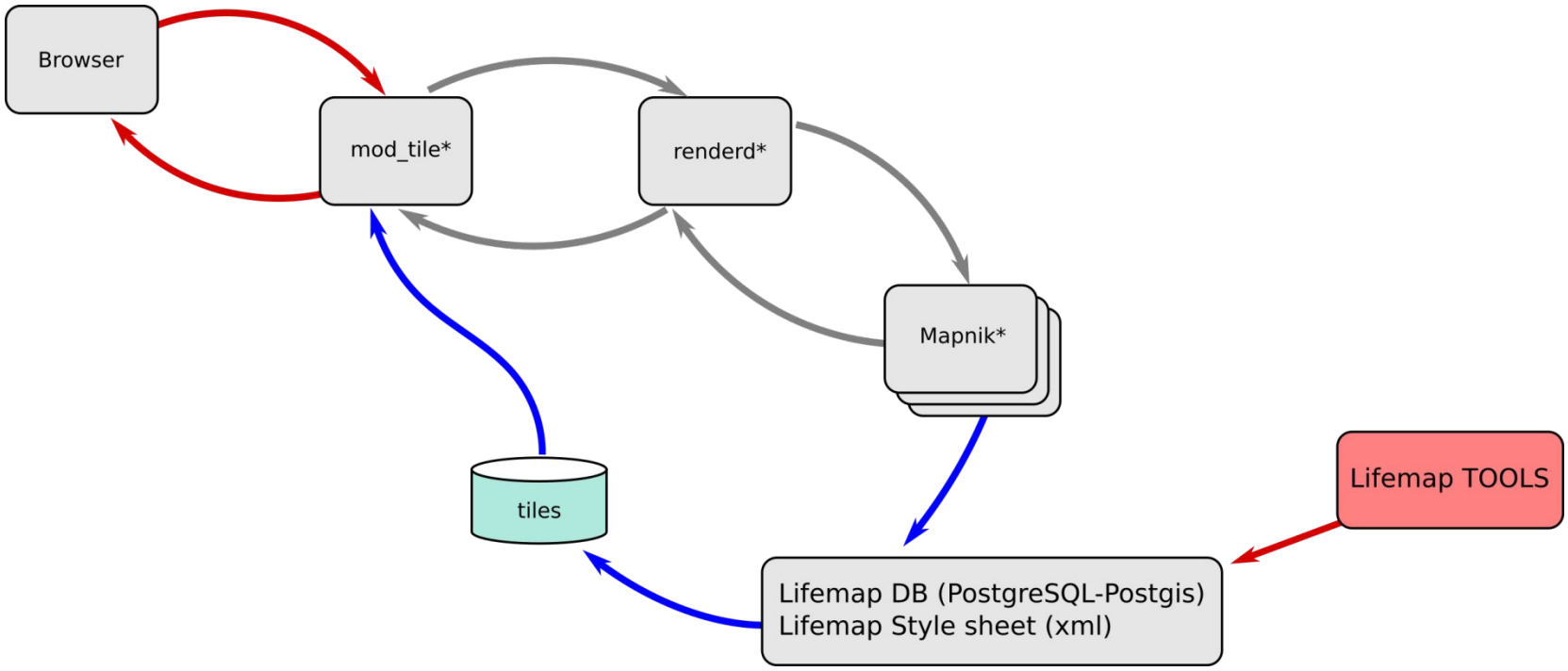
- No limitation in the amount of data (# species)
- Smooth and fast

OpenStreetMap



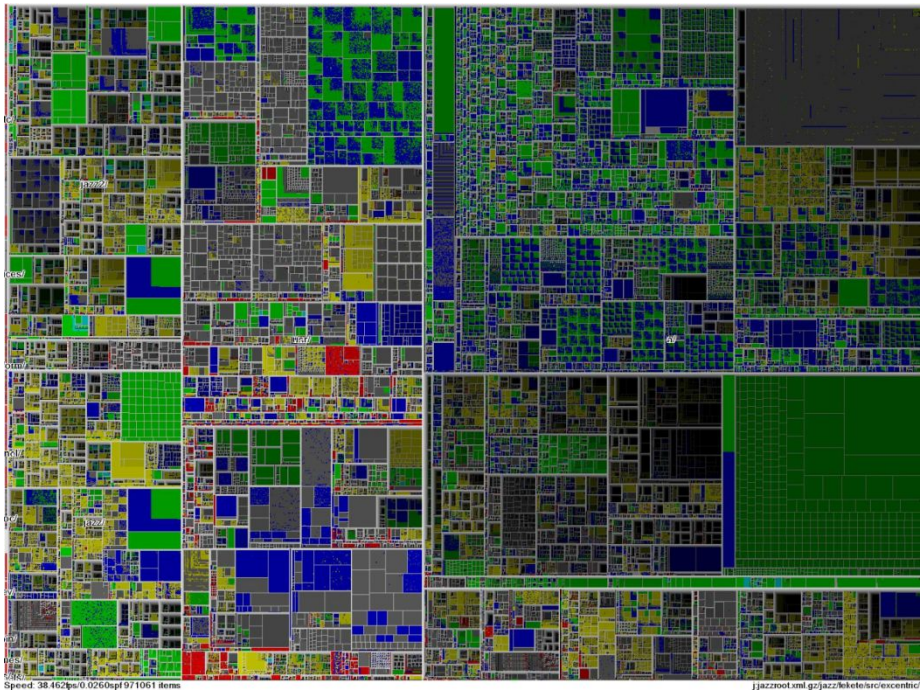
The need for a new visualization

- No limitation in the amount of data (# species)
- Smooth and fast



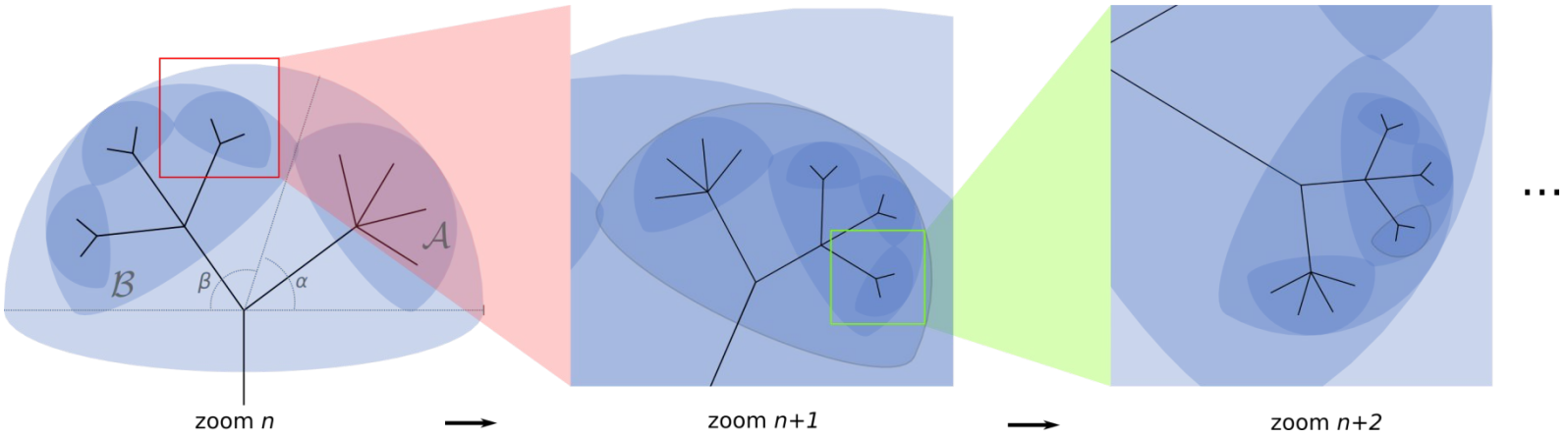
The need for a new visualization

- No limitation in the amount of data (# species)
- Smooth and fast
- Allow multifurcations
- Tree-looking



The need for a new visualization

- No limitation in the amount of data (# species)
- Smooth and fast
- Allow multifurcations
- Tree-looking



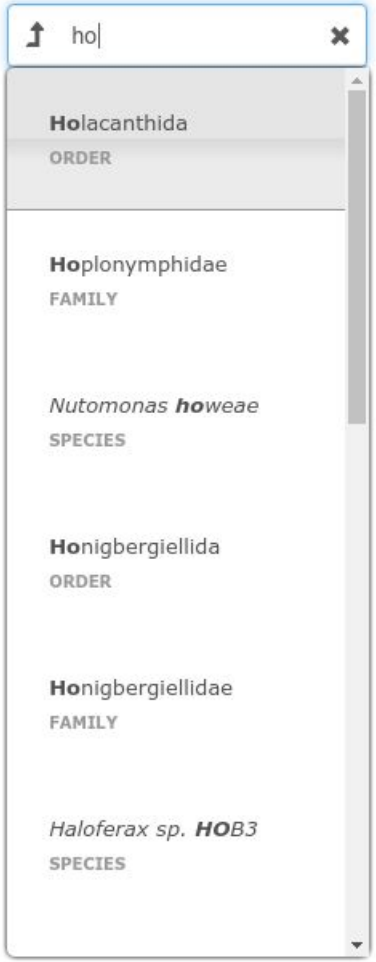
Lifemap tools / pipeline

- Get the tree (NCBI, any tree)
- Traverse the tree
  - Compute coordinates of all elements at all zoom levels
  - Get additional information for nodes and tips
  - Write in PostgreSQL/PostGIS database
  - Write in Solr Search Engine

Lifemap tools / pipeline

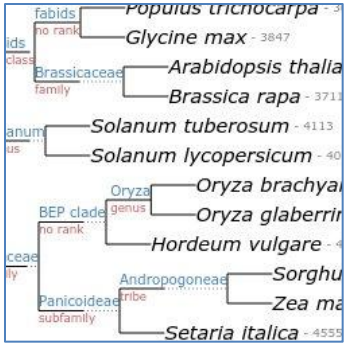
- Get the tree (NCBI, any tree)
- Traverse the tree
  - Compute coordinates of all elements at all zoom levels
  - Get additional information for nodes and tips
  - Write in PostgreSQL/PostGIS database
  - Write in Solr Search Engine

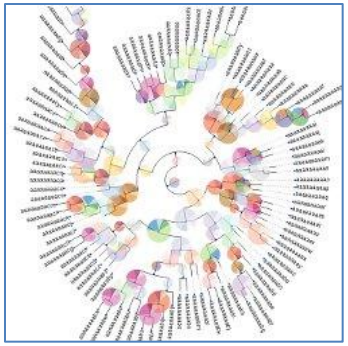
Solr

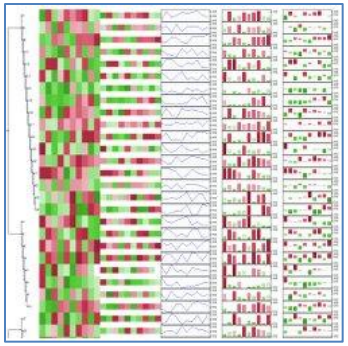


ETE Toolkit

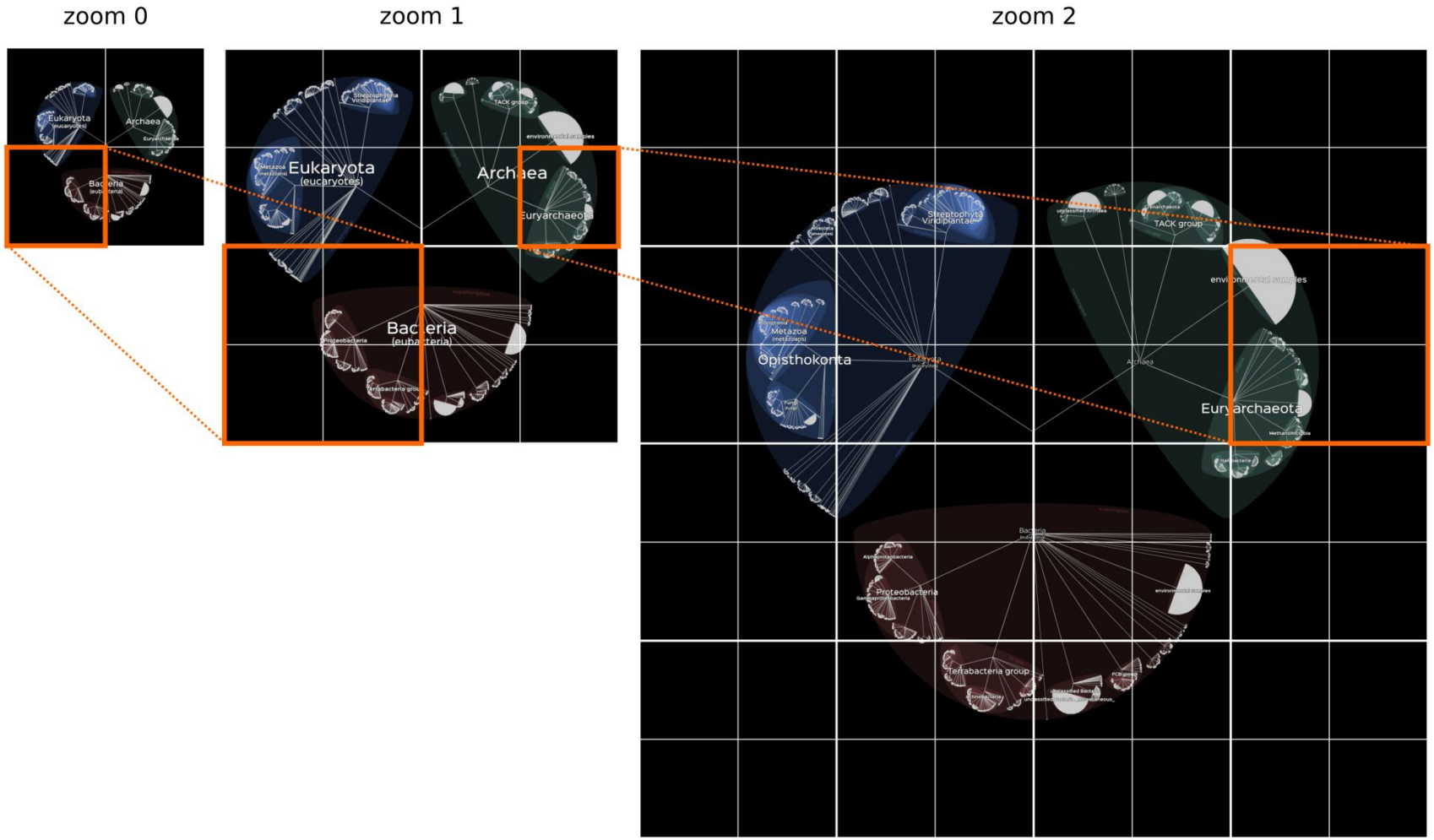
Huerta-Cepas *et al.* 2016







Lifemap



Q : combien de tuiles au zoom 3 ?

The need for a new visualization

- No limitation in the amount of data (# species)
- Smooth and fast
- Allow multifurcations
- Tree-looking

- Links to additional resources
- Serve as a background for additional information (as layers)

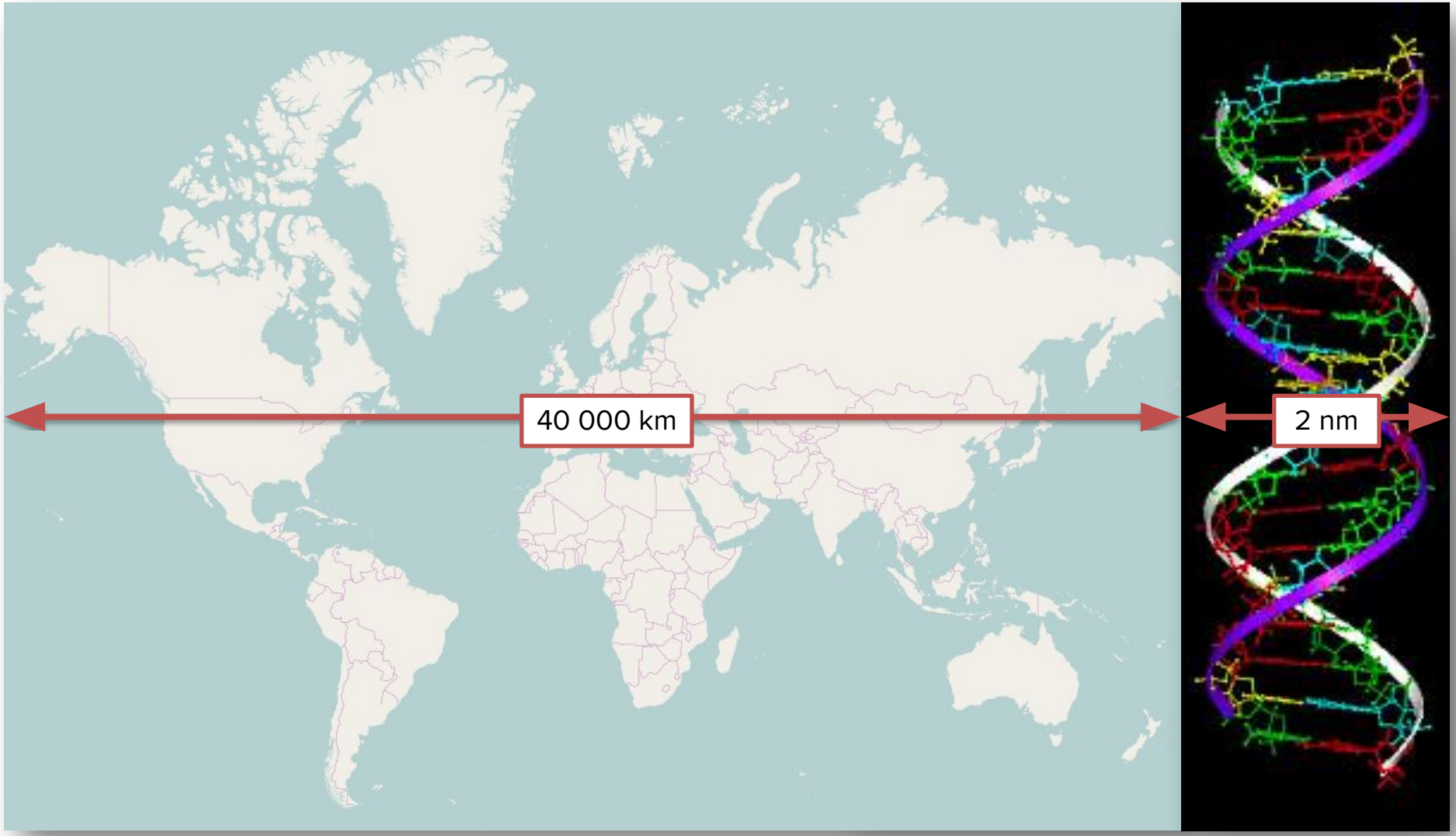
- Possibility to visualize your own trees

- For large public, education and research

Lifemap versions

Deux versions de Lifemap

52 niveaux de zoom



Deux versions de Lifemap

52 niveaux de zoom

Q : combien de tuiles en tout dans Lifemap ?

Deux versions de Lifemap

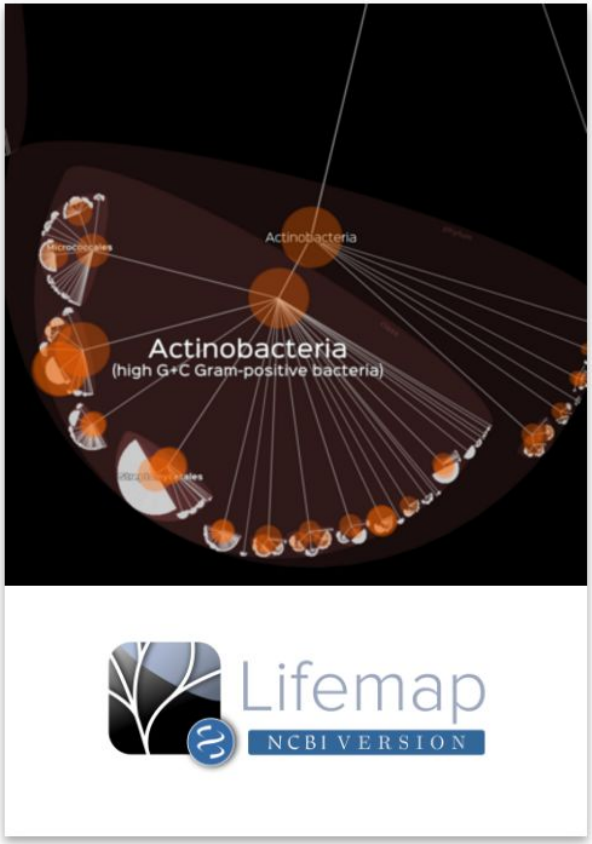
52 niveaux de zoom

Q : combien de tuiles en tout dans Lifemap ?

R :
4^53 = 8.112964e+31 tuiles
6Ko par tuile en moyenne
4.867778e+32 Ko en tout
4.867778e+23 To
...un million de milliard de fois plus que la quantité de données générées dans le monde entier chaque année.

Solutions ?

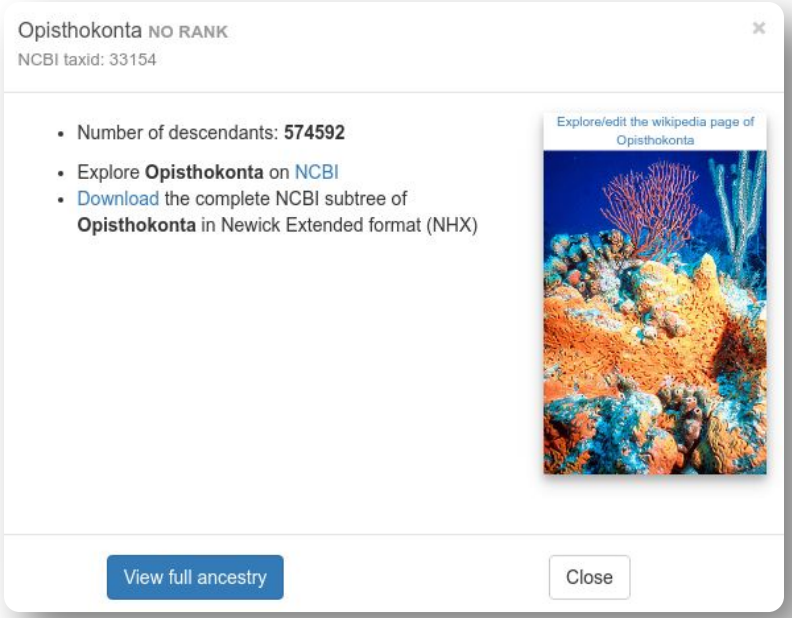
Deux versions de Lifemap



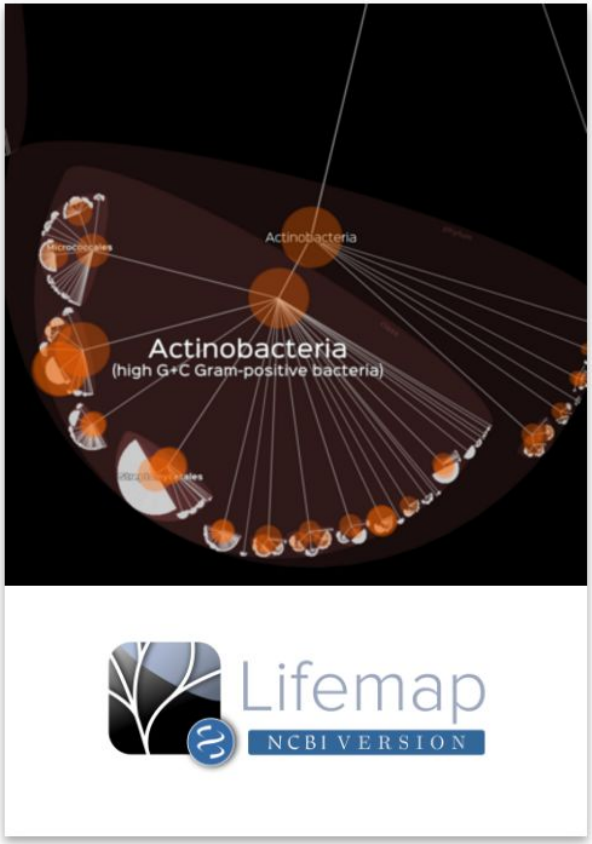
Full NCBI taxonomy (1 220 620 tips)

Updated weekly

Additional information:

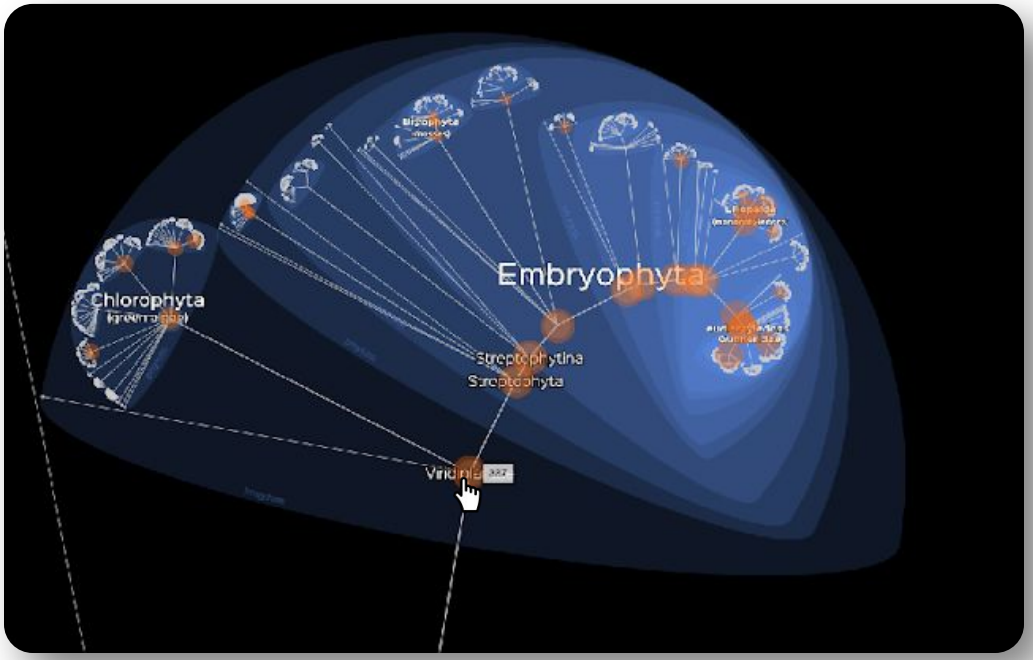


Deux versions de Lifemap

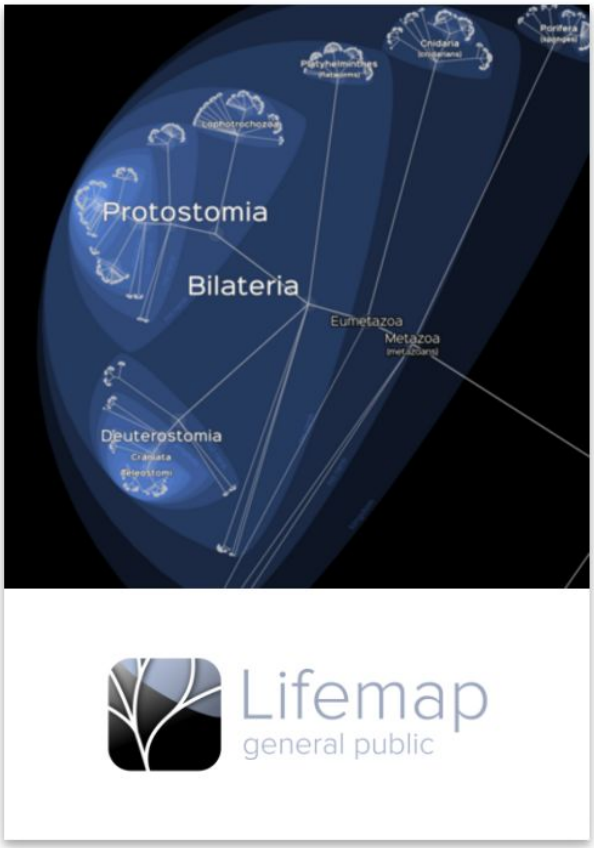


Full NCBI taxonomy (1 220 620 tips)

Updated weekly

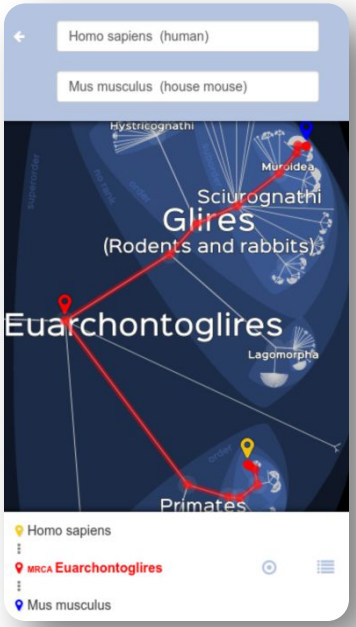


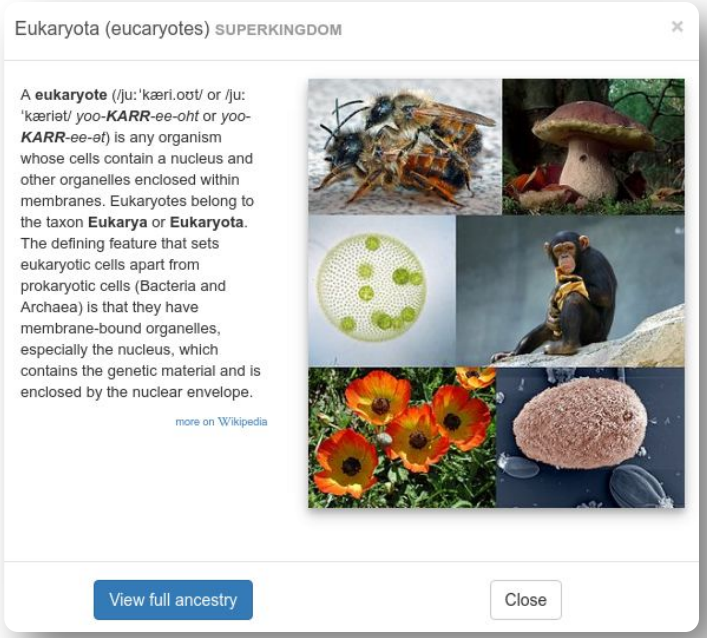
Deux versions de Lifemap



Simplified NCBI taxonomy (802639 tips)

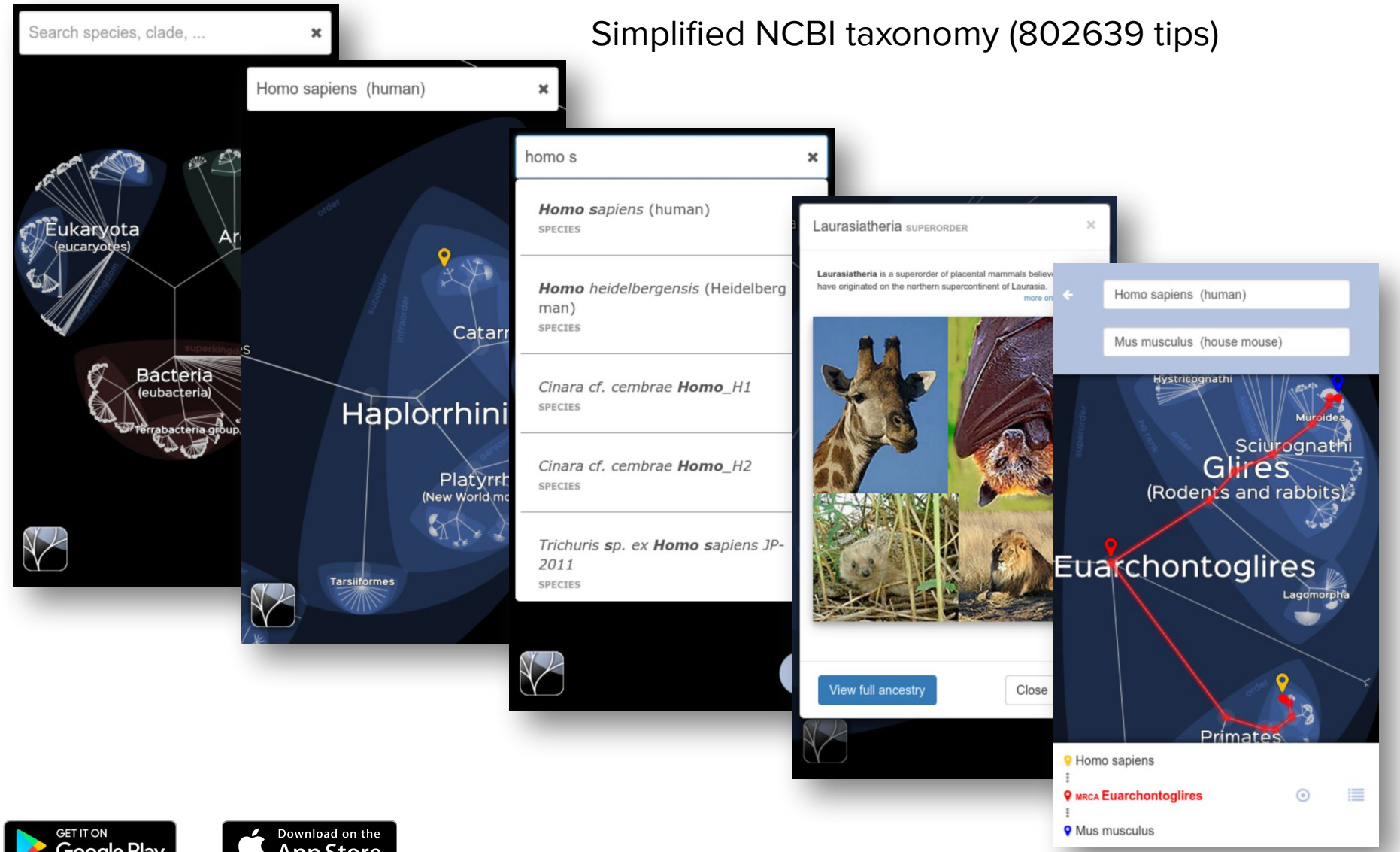
Removed subspecies level as well as ‘unidentified’, ‘environmental’, ‘unclassified’, ‘uncultured’ taxa.





Deux versions de Lifemap

Simplified NCBI taxonomy (802639 tips)



GET IT ON Google Play

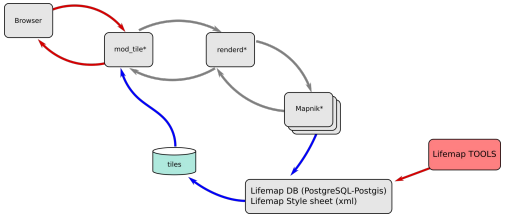
Download on the App Store

Et une version pour visualiser ses propres arbres

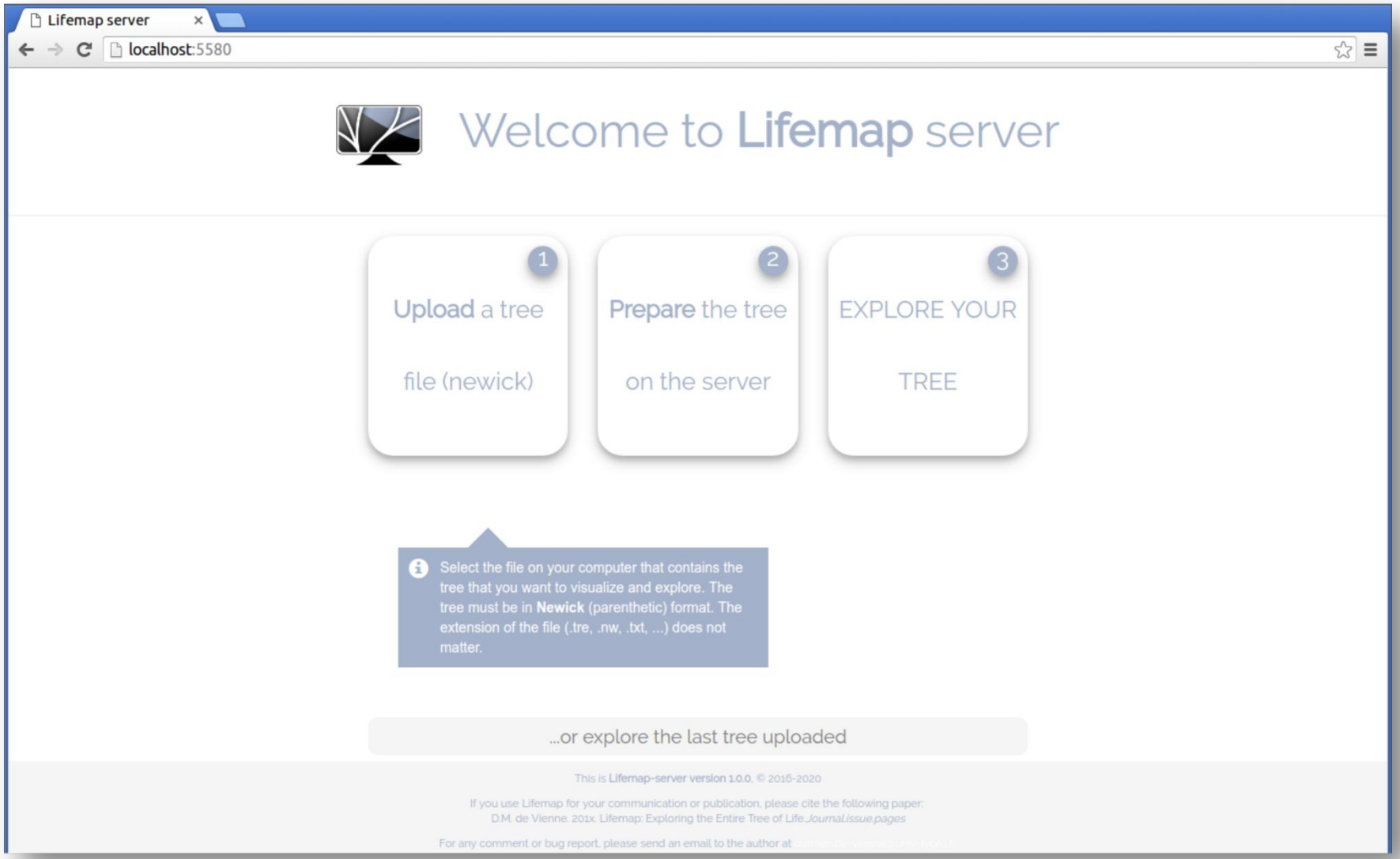
Lifemap-server

http://lifemap.univ-lyon1.fr/download.html

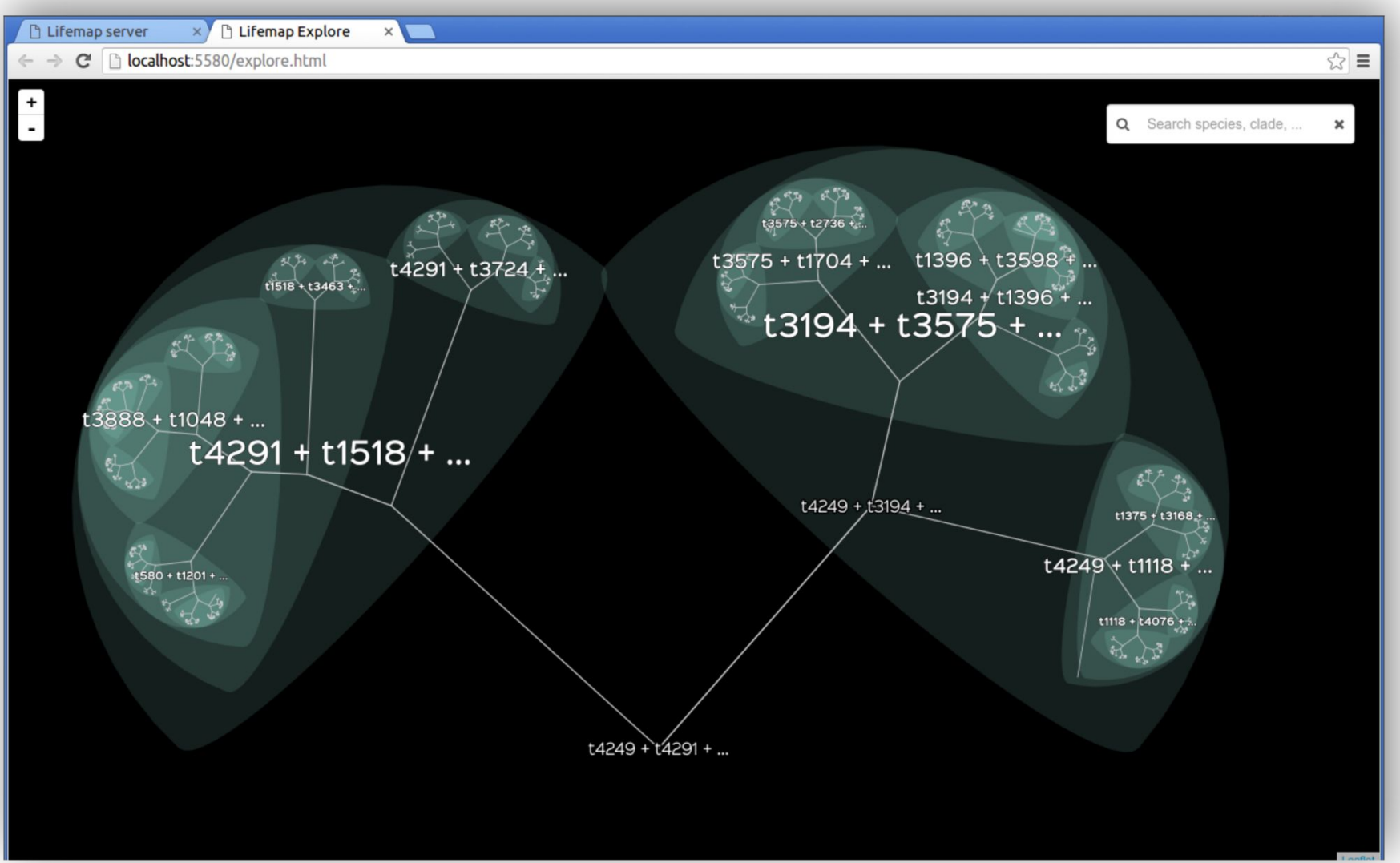
- Upload your own (Newick) trees
- Nodes are named automatically if necessary
- Explore the tree, search nodes and tips with an efficient search tool (Solr)



Lifemap-server



Lifemap-server



Lifemap - Welcome

lifemap.univ-lyon1.fr

Damien

Overleaf project | iOS Platform Gui | Google Play Dev | ColorBrewer: Co | iWantHue

Google play

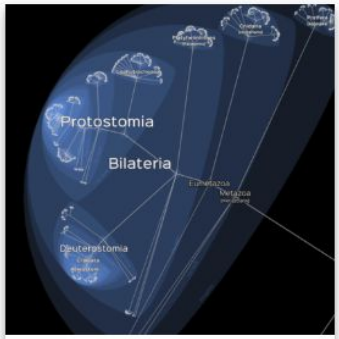
Télécharger dans l'App Store

Tweet

Share

Lifemap

Exploring the Entire Tree of Life

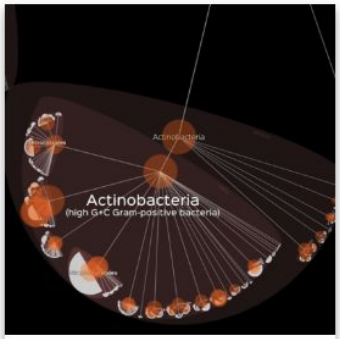


Lifemap general public

Tree simplified from NCBI taxonomy

~800,000 species

This 'large public' tree is a reduction of the whole NCBI taxonomy. It only shows species (no strains, environmental samples, unidentified species, etc.). It displays a descritption and picture retrieved from Wikipedia when clicking on the tips and nodes.



Lifemap NCBI VERSION

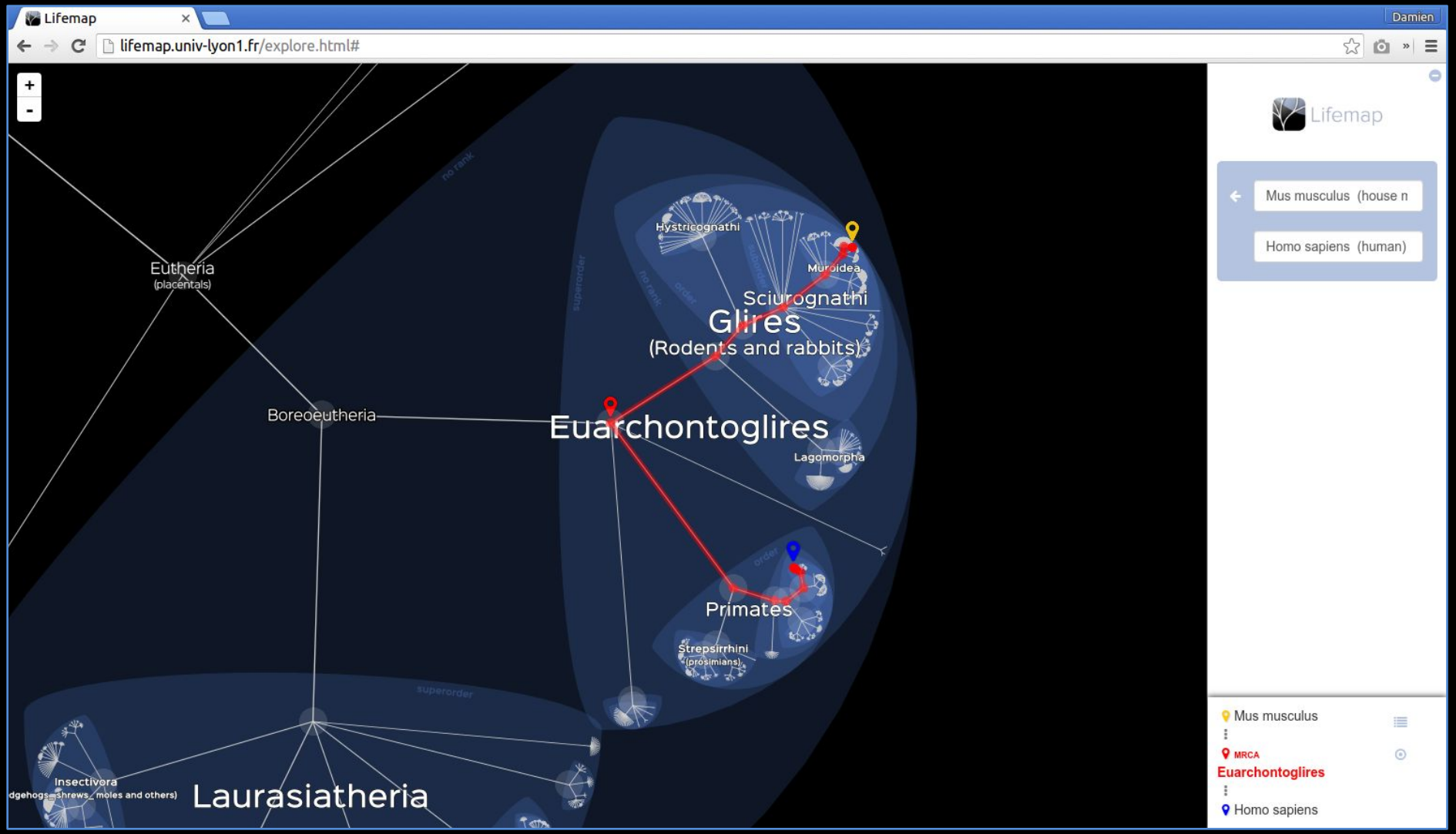
Complete NCBI taxonomy

~1.4 Million species

The NCBI taxonomic tree represents the relationships between all species/strains that have at least one sequence in the NCBI/Entrez database. It is curated and validated by expert taxonomists and is updated regularly.

If you use Lifemap for your publications, please cite:

de Vienne DM (2016) Lifemap: Exploring the Entire Tree of Life. *PLOS Biology* 14(12): e2001624. doi: 10.1371/journal.pbio.2001624.


Lifemap
Damien
lifemap.univ-lyon1.fr/explore.html#
+
-
Eutheria
(placentals)
Boreoeutheria
Laurasiatheria
Insectivora
(hedgehogs, shrews, moles and others)
Hystricognathi
Muroidea
Sciurognathi
Glires
(Rodents and rabbits)
Euarchontoglires
Lagomorpha
Primates
Strepsirrhini
(prosimians)
no rank
superorder
order
Lifemap
Mus musculus (house n
Homo sapiens (human)
Mus musculus
MRCA
Euarchontoglires
Homo sapiens

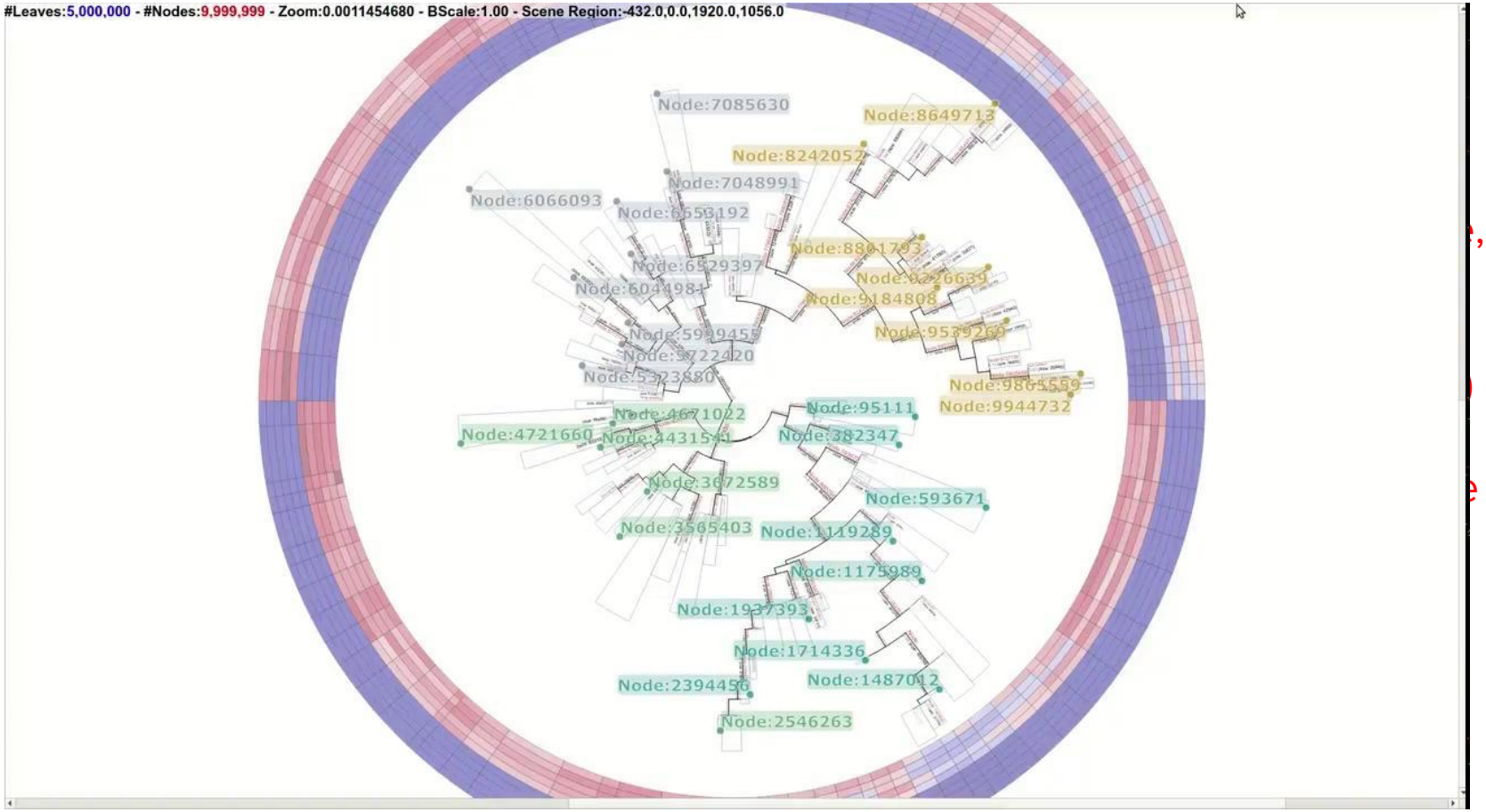

Craniata
Gnathostomata
(jawed vertebrates)
Teleostomi
(bony vertebrates)

Limites de Lifemap et projets associés

- **Longueurs de branches**. Comment combiner le zoom profond et le respect des longueurs de branche. Question ouverte
- **Lourdeur du processus**. Il faut plusieurs heures chaque semaine pour mettre à jour Lifemap : 20 minutes pour le calcul, qq heures pour la génération des tuiles aux 10 premiers niveaux de zoom. Trop long
- **Manque de souplesse.** Et si nous souhaitions modifier l'arbre à la volée pour le construire de façon collaborative ? Projet wiki Tree of Life (JS, node, D3)
- **Grands arbres dans R avec Shiny.** Rien n'existe pour zoomer dans un grand arbre (N>1000) et l'explorer dans R. Projet R-BigTree (R, Shiny, leaflet)
- **Faciliter l'utilisation de Lifemap comme fond de carte.** Interopérabilité entre R, Lifemap et NCBI et visualisation grâce à leaflet (depuis R). Projet de création du package R-lifemap [**TP**]

Limites de Lifemap et projets associés

- **Longueurs de branches**. Comment combiner le zoom profond et le respect des longueurs de branche. Question ouverte



Limites de Lifemap et projets associés

- **Longueurs de branches**. Comment combiner le zoom profond et le respect des longueurs de branche. Question ouverte
- **Lourdeur du processus**. Il faut plusieurs heures chaque semaine pour mettre à jour Lifemap : 20 minutes pour le calcul, qq heures pour la génération des tuiles aux 10 premiers niveaux de zoom. Trop long
- **Manque de souplesse.** Et si nous souhaitions modifier l'arbre à la volée pour le construire de façon collaborative ? Projet wiki Tree of Life (JS, node, D3)
- **Grands arbres dans R avec Shiny.** Rien n'existe pour zoomer dans un grand arbre (N>1000) et l'explorer dans R. Projet R-BigTree (R, Shiny, leaflet)
- **Faciliter l'utilisation de Lifemap comme fond de carte.** Interopérabilité entre R, Lifemap et NCBI et visualisation grâce à leaflet (depuis R). Projet de création du package R-lifemap [**TP**]

Limites de Lifemap et projets associés

- **Grands arbres dans R avec Shiny.** Rien n'existe pour zoomer dans un grand arbre (N>1000) et l'explorer dans R. Projet R-BigTree (R, Shiny, leaflet)

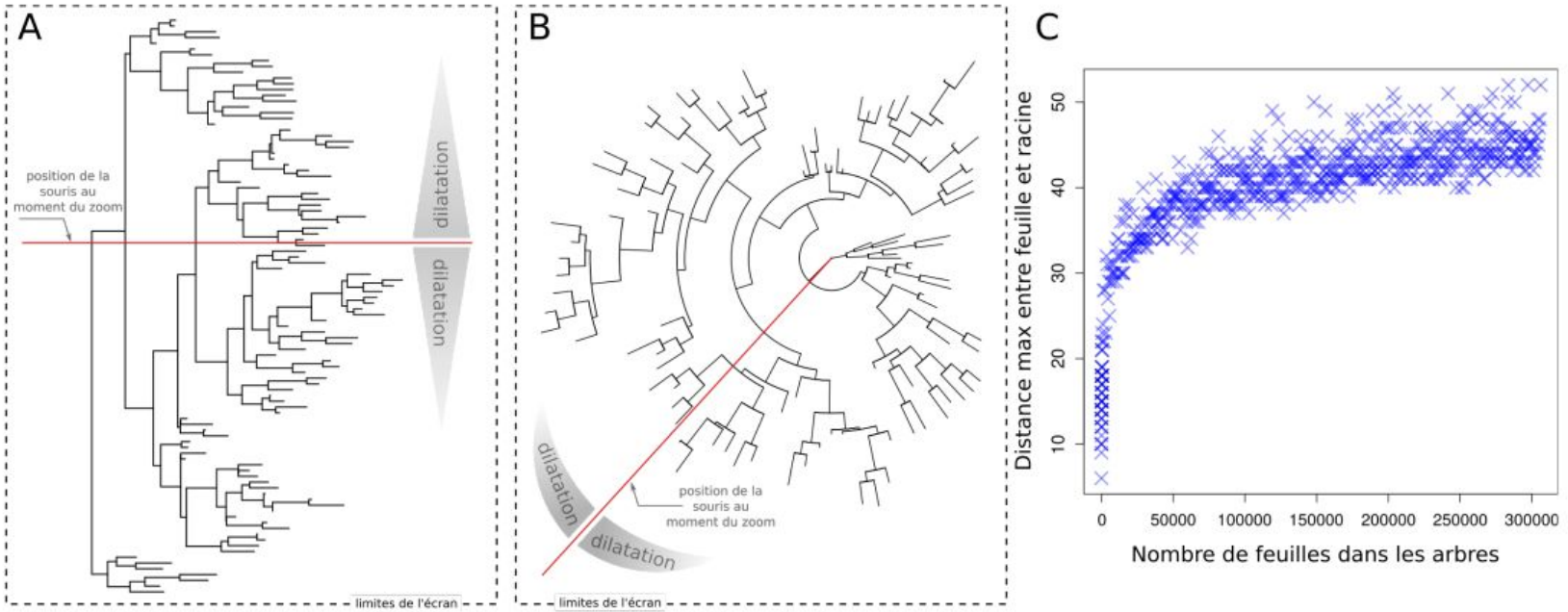


Figure 2. Proposition d'une nouvelle façon de visualiser les grands arbres. Que la représentation soit un phylogramme (A) ou une représentation radiale (B), l'idée est la même. Si un zoom est engagé avec la souris positionnée sur la ligne rouge, alors toutes les coordonnées des nœuds et des feuilles sont recalculées pour éloigner les points de la ligne rouge. L'éloignement sera d'autant plus fort que le point sera poche de cette ligne. Il s'agit donc d'une dilatation, selon l'axe y pour A ou selon un angle pour B. C) La représentation proposée fonctionne car la profondeur de l'arbre (distance maximum entre une feuille et la racine) reste faible même pour des nombres importants de feuilles dans les arbres.

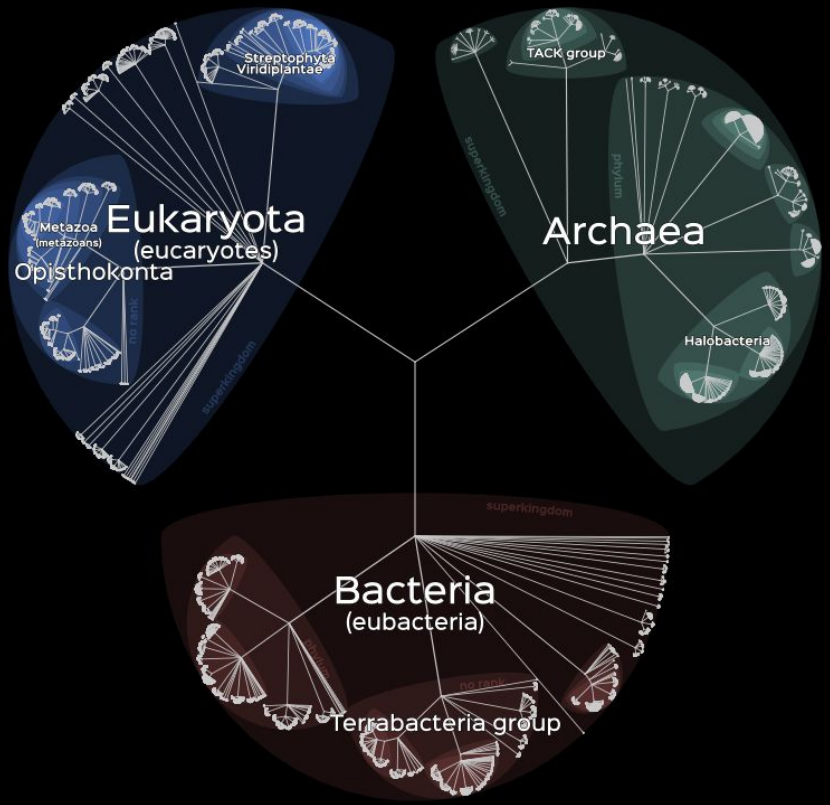
TP

- Interagir **avec** Lifemap
- Visualiser des données biologiques **sur** Lifemap
- Créer une App Shiny pour visualiser ces données.

...depuis 

...en utilisant le package Leaflet

https://leafletjs.com/


Streptophyta
Viridiplantae
Metazoa
(metazoans)
Eukaryota
(eucaryotes)
Opisthokonta
no rank
superkingdom
TACK group
superkingdom
phylum
Archaea
Halobacteria
superkingdom
Bacteria
(eubacteria)
no rank
Terrabacteria group