

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

* et autres hiérarchies



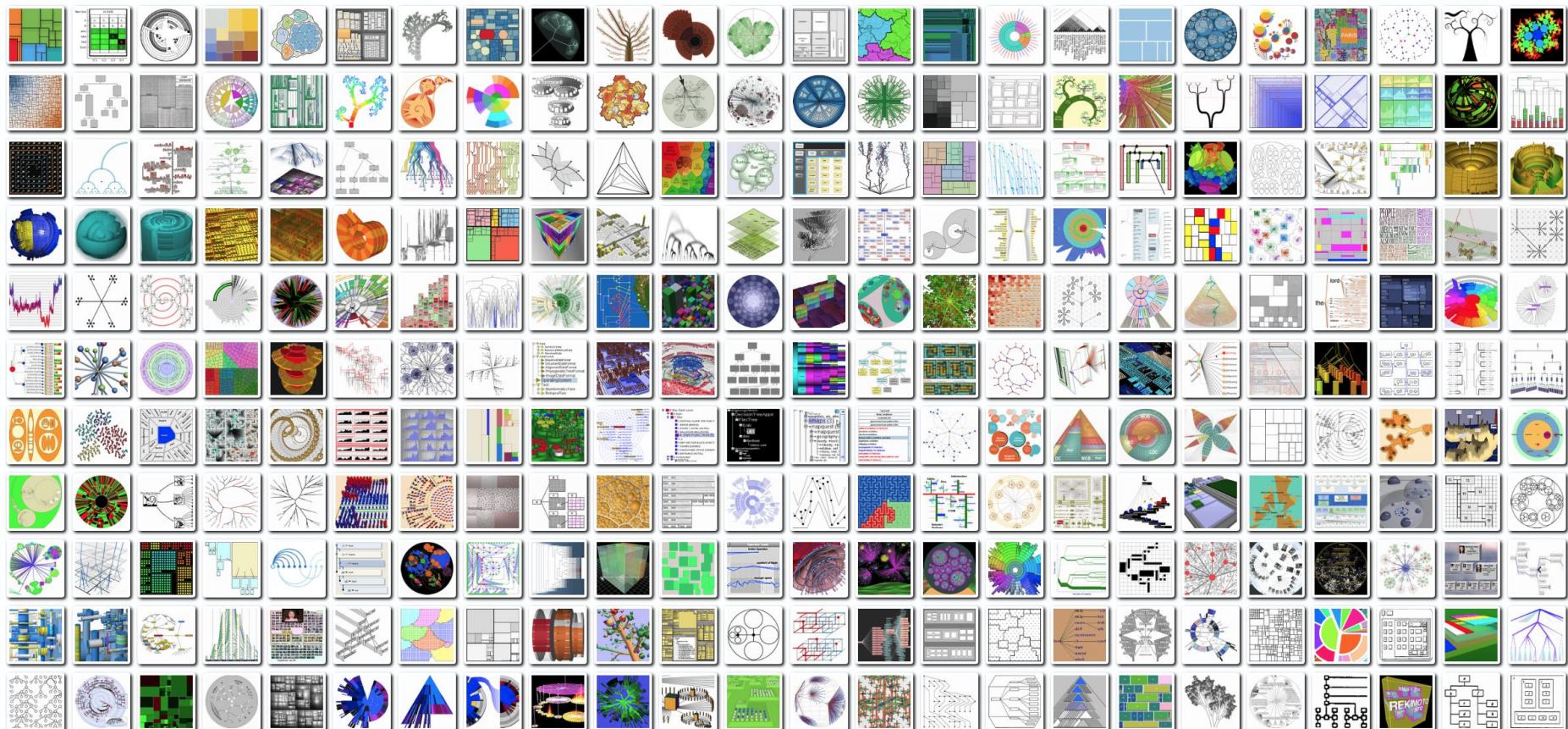
@damdevienne
@lifemap_tol

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



LBBE





<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
- Au delà de Lifemap
- Limites et perspectives de développement
- TP

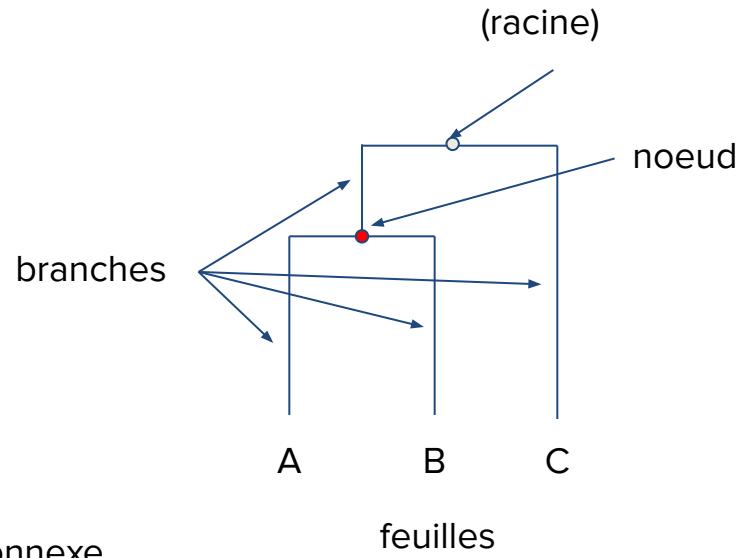
Qu'est-ce qu'un arbre ?

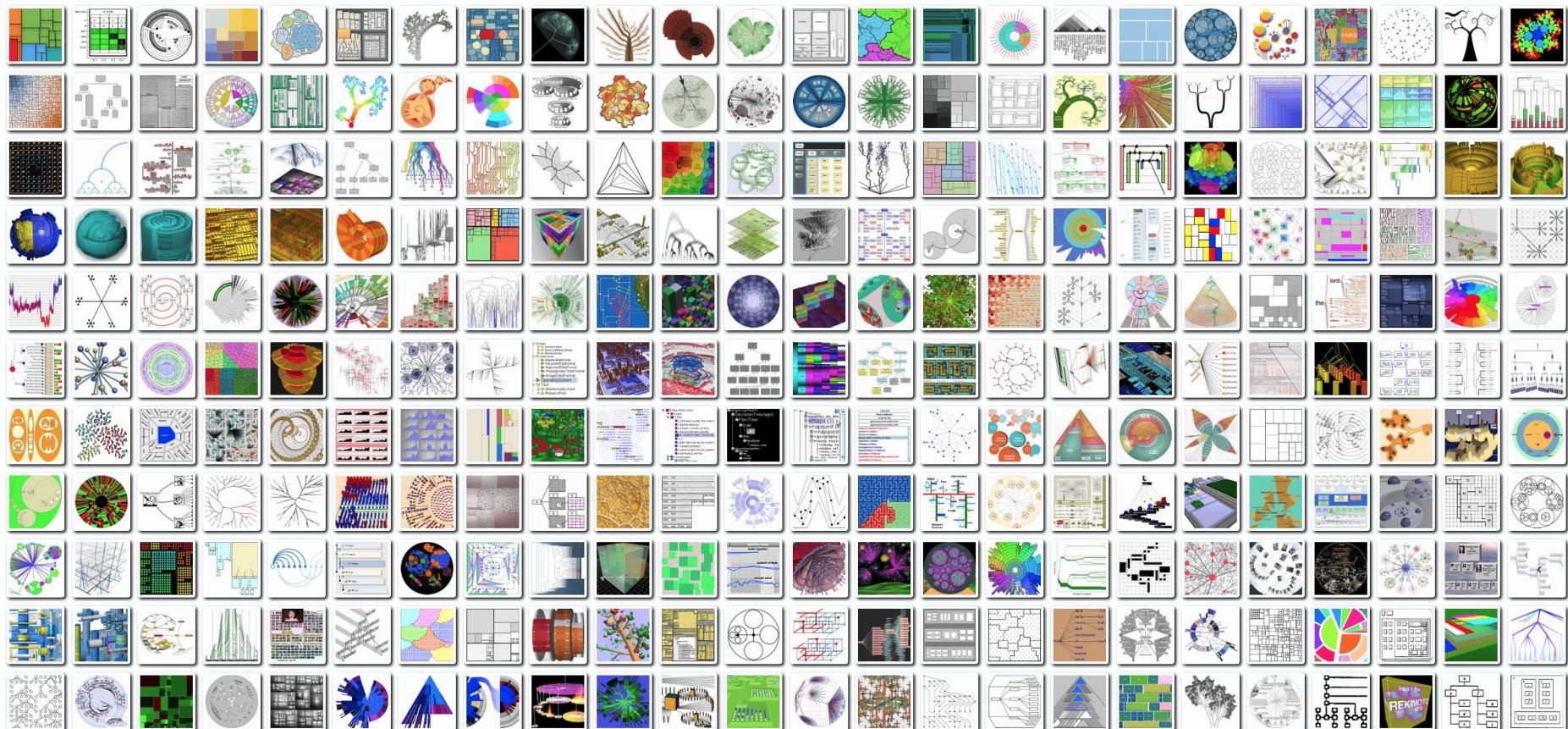


phylogénie
généalogie
système de fichiers
clustering
classification
arbres de décision
mindmap

...

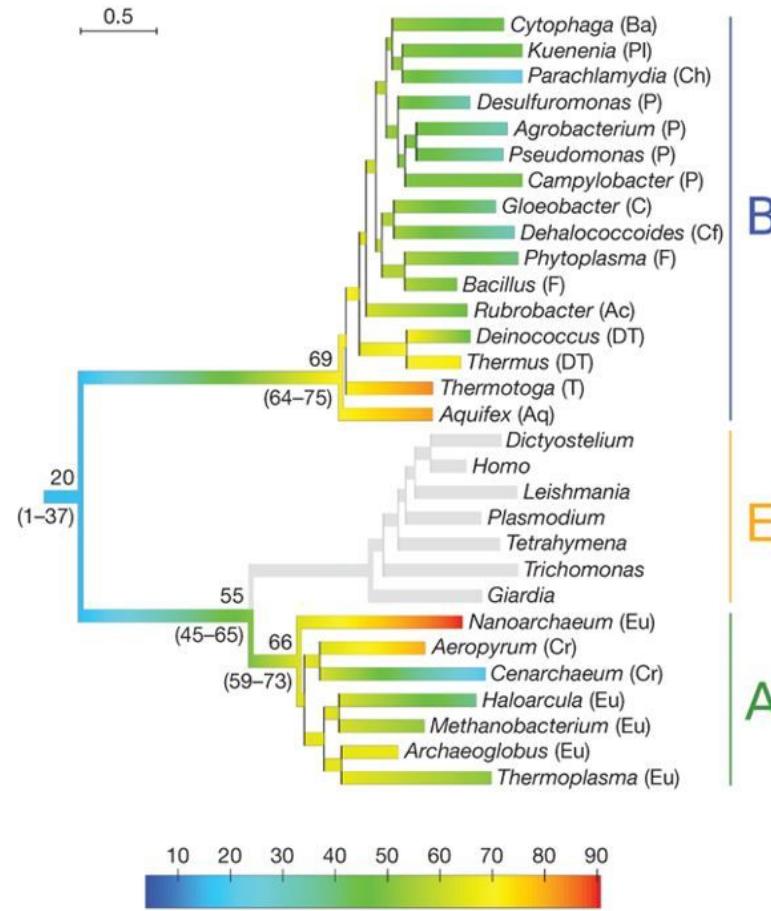
graphe acyclique et connexe





<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

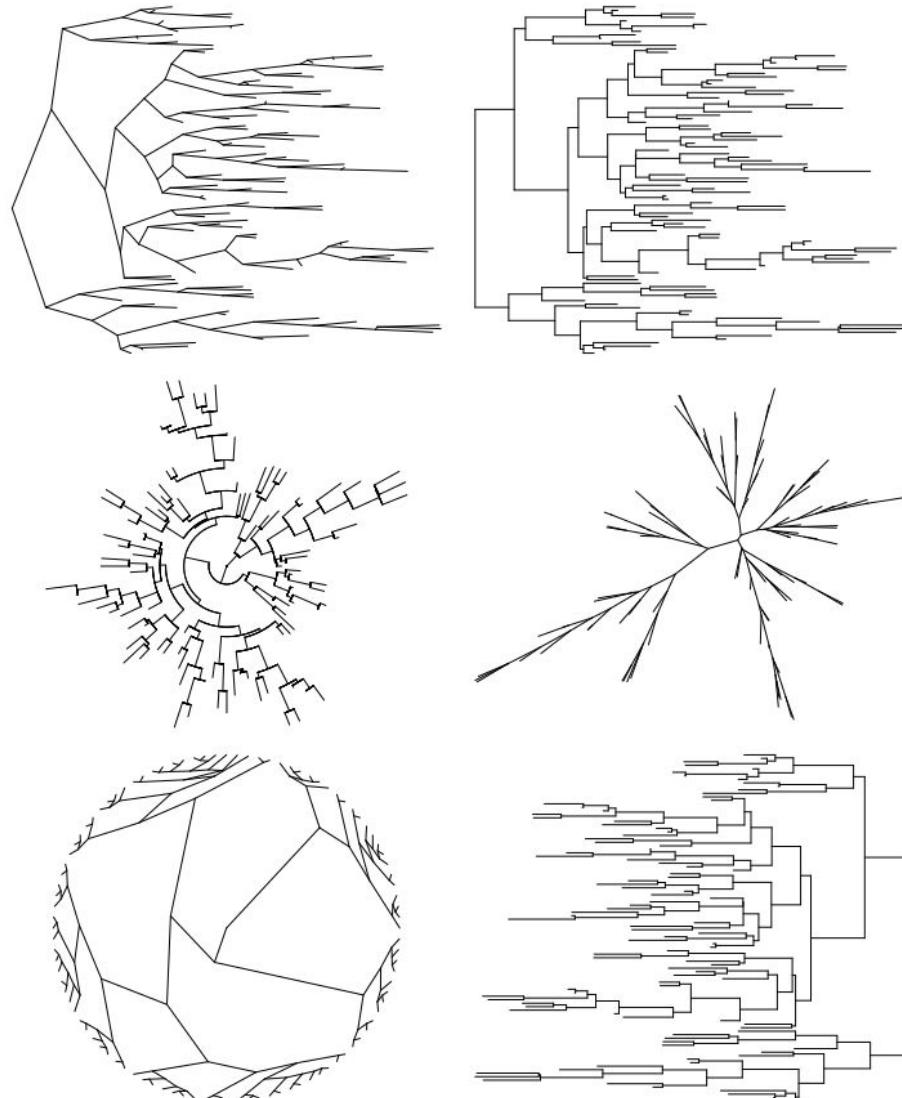
Ceci forme la donnée biologique visualisée

Qu'est-ce qu'un arbre ?

L'arbre lui-même qui est une donnée biologique

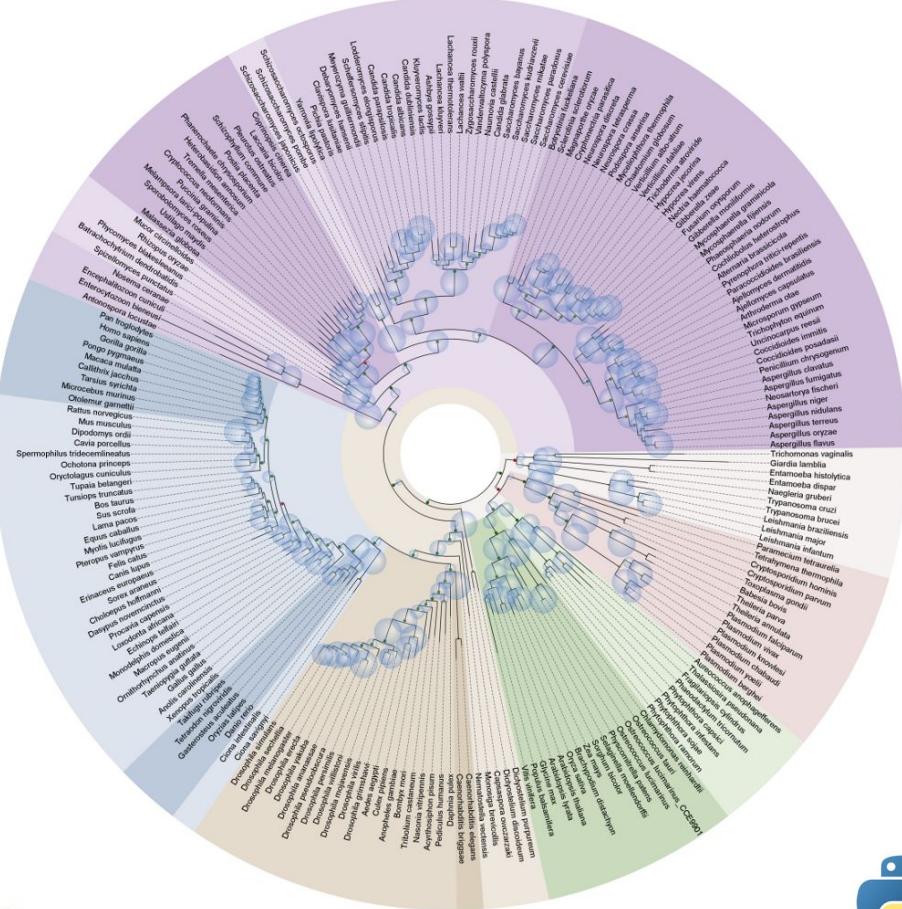
- racine
- noeuds
- feuilles
- branches

```
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 ?

Gene tree consistency support



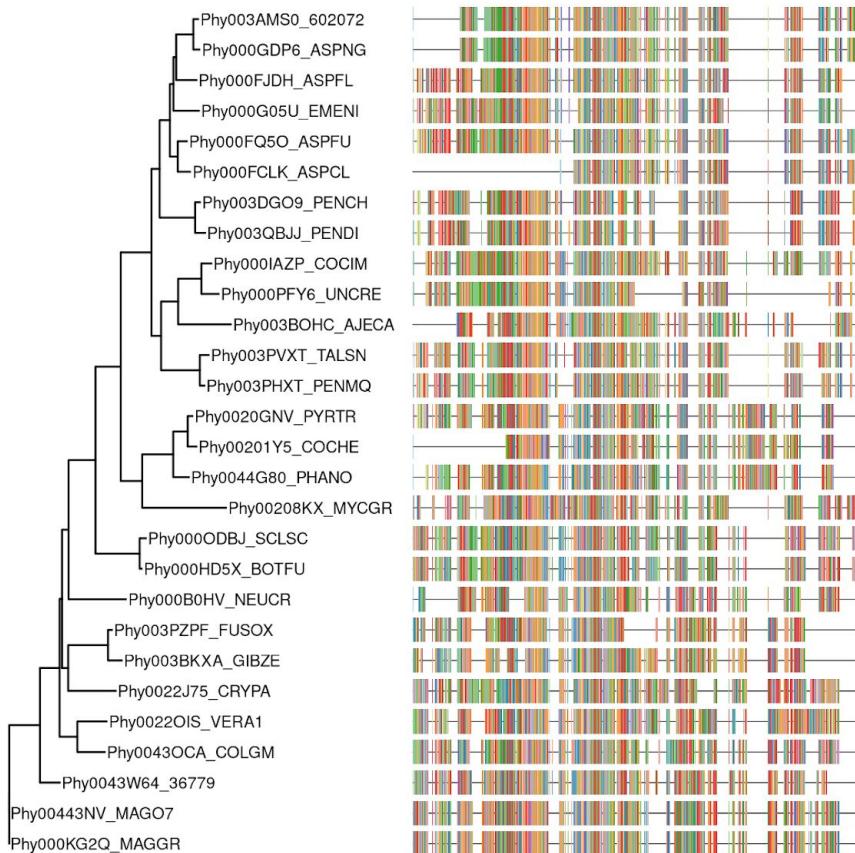
Les informations associées

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



Qu'est-ce qu'un arbre ?

A



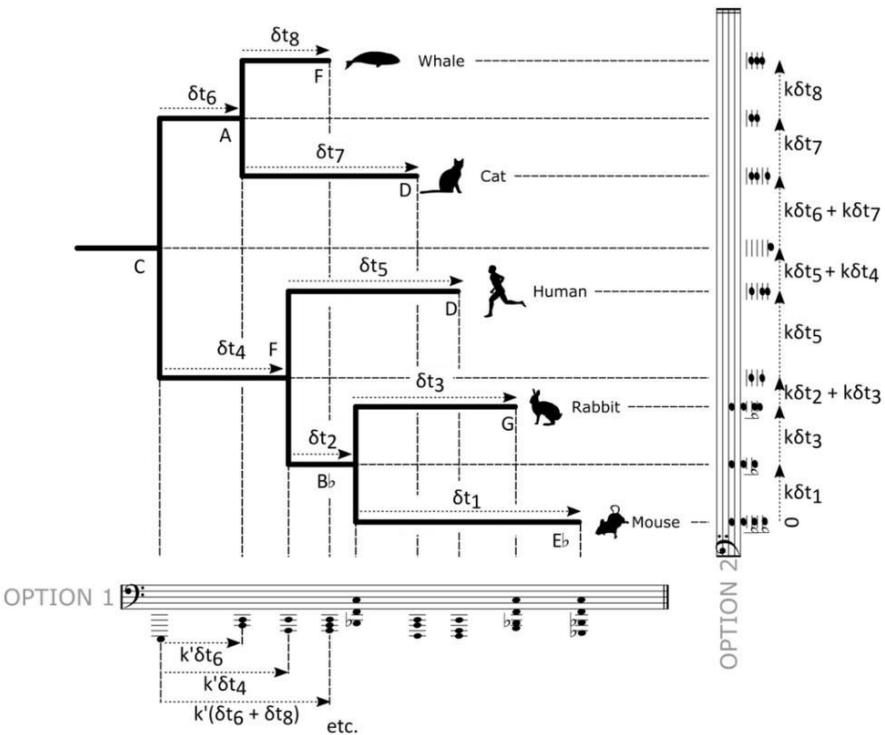
Les informations associées

- 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 ?

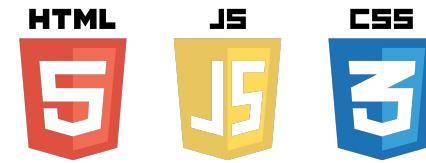


Les informations associées

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

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/>

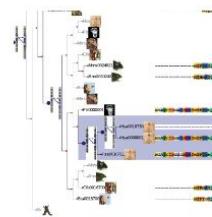
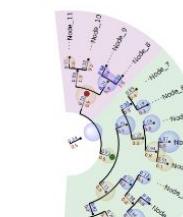
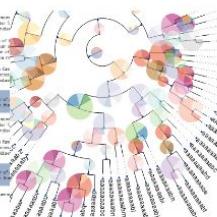
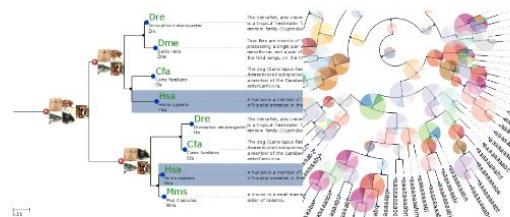
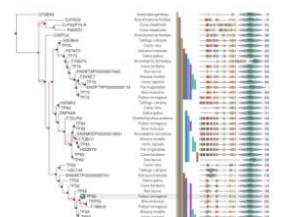
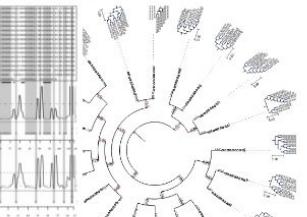
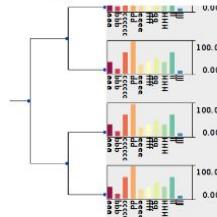
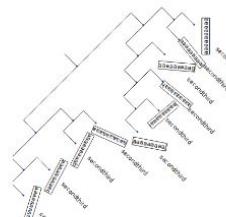
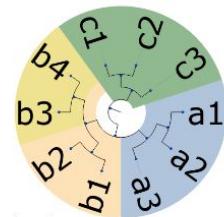
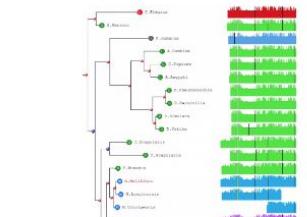
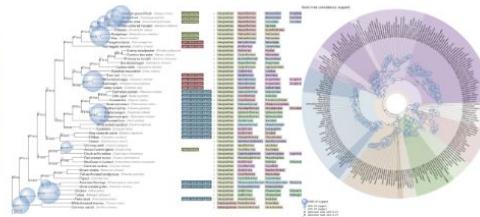
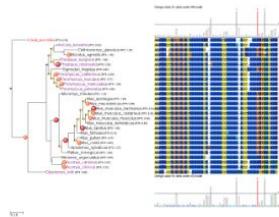
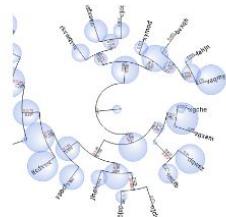
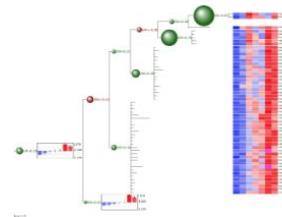
ITOL INTERACTIVE
TREE OF LIFE
<https://itol.embl.de/>

Quelques outils utiles pour visualiser les arbres et les infos associées

ETE Toolkit

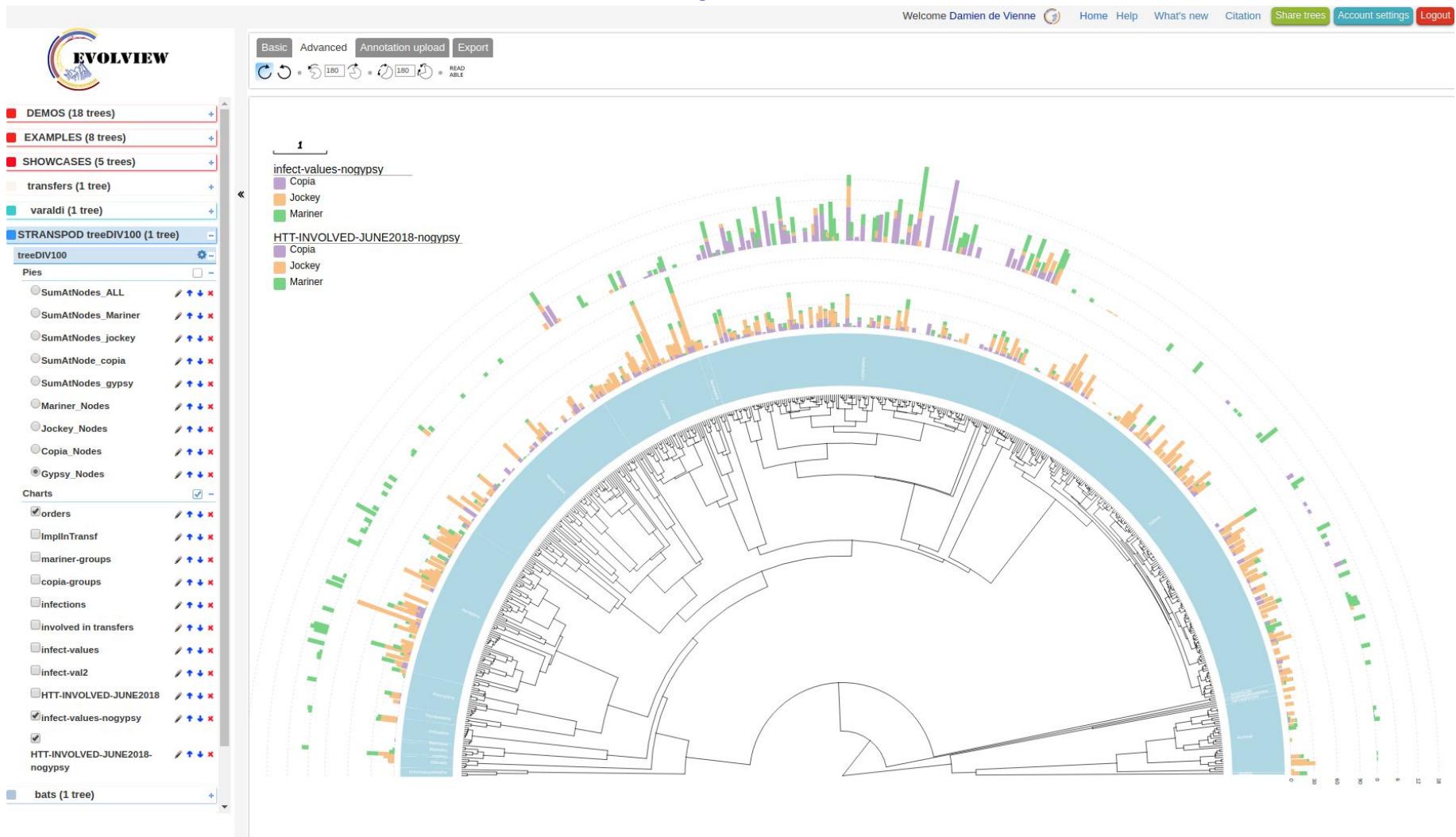
A Python framework to work with trees

<http://etetoolkit.org/>



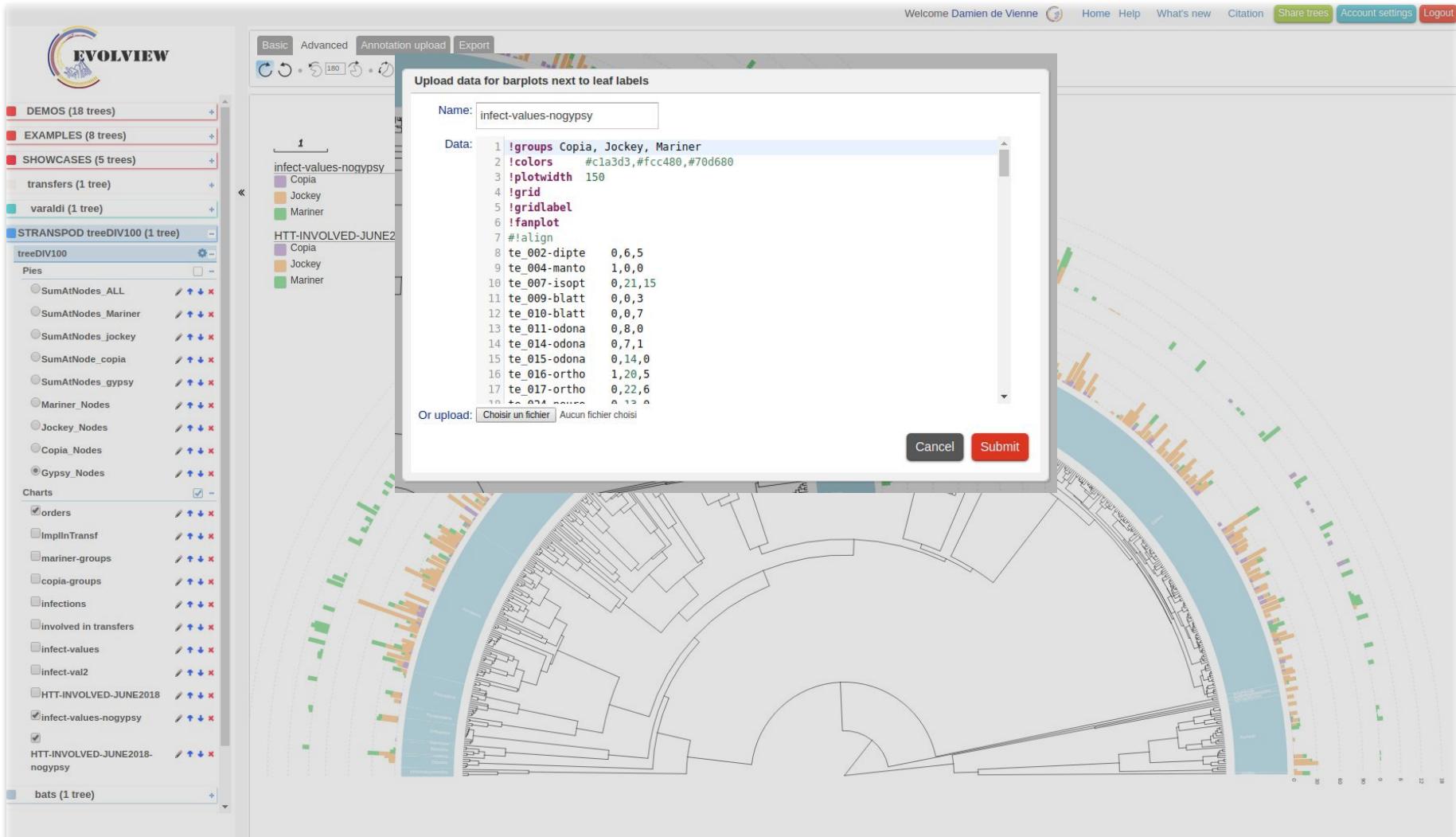
Quelques outils utiles pour visualiser les arbres et les infos associées

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



Quelques outils utiles pour visualiser les arbres et les infos associées

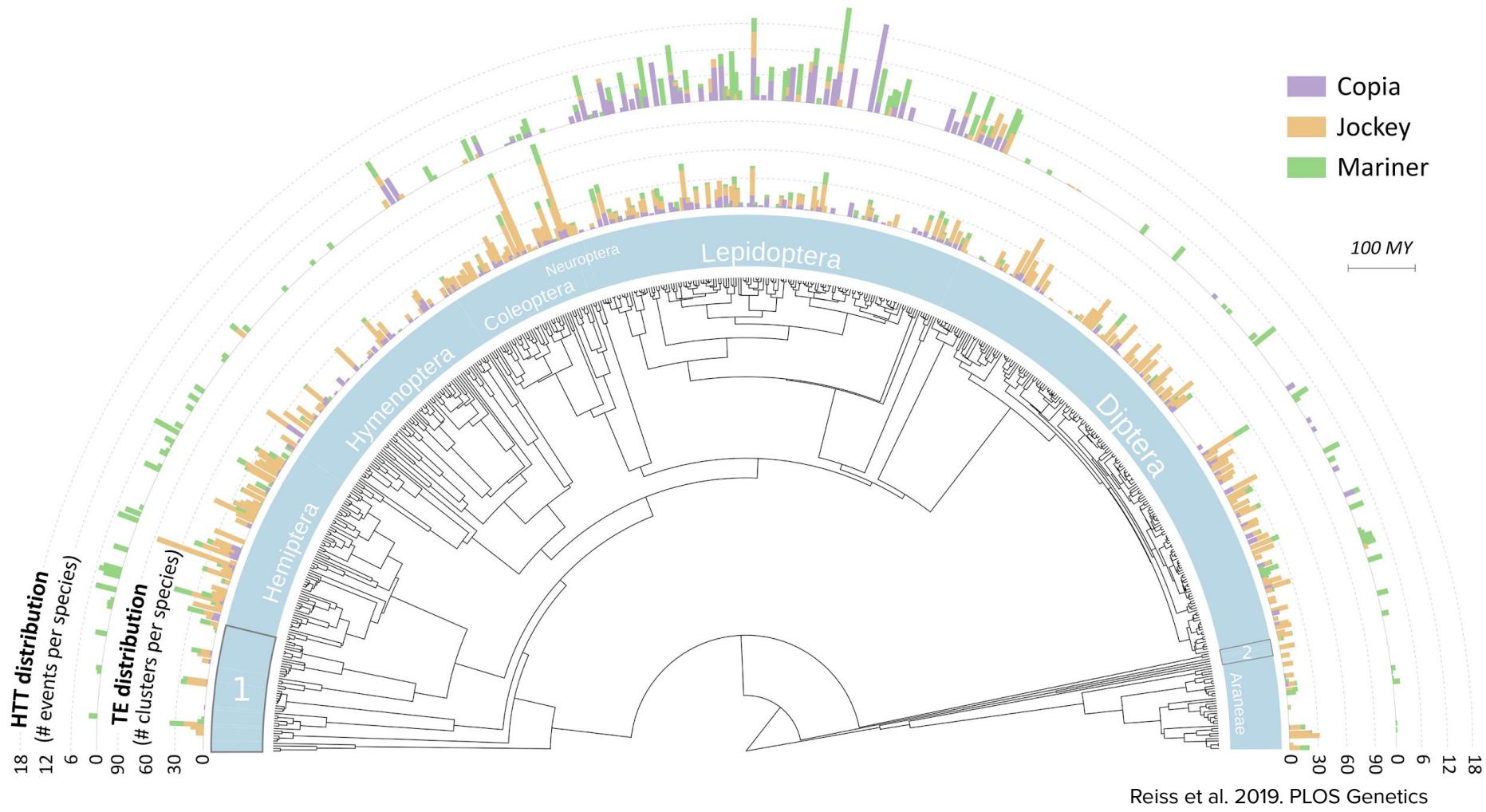
<https://www.evolgenius.info/evolview/>



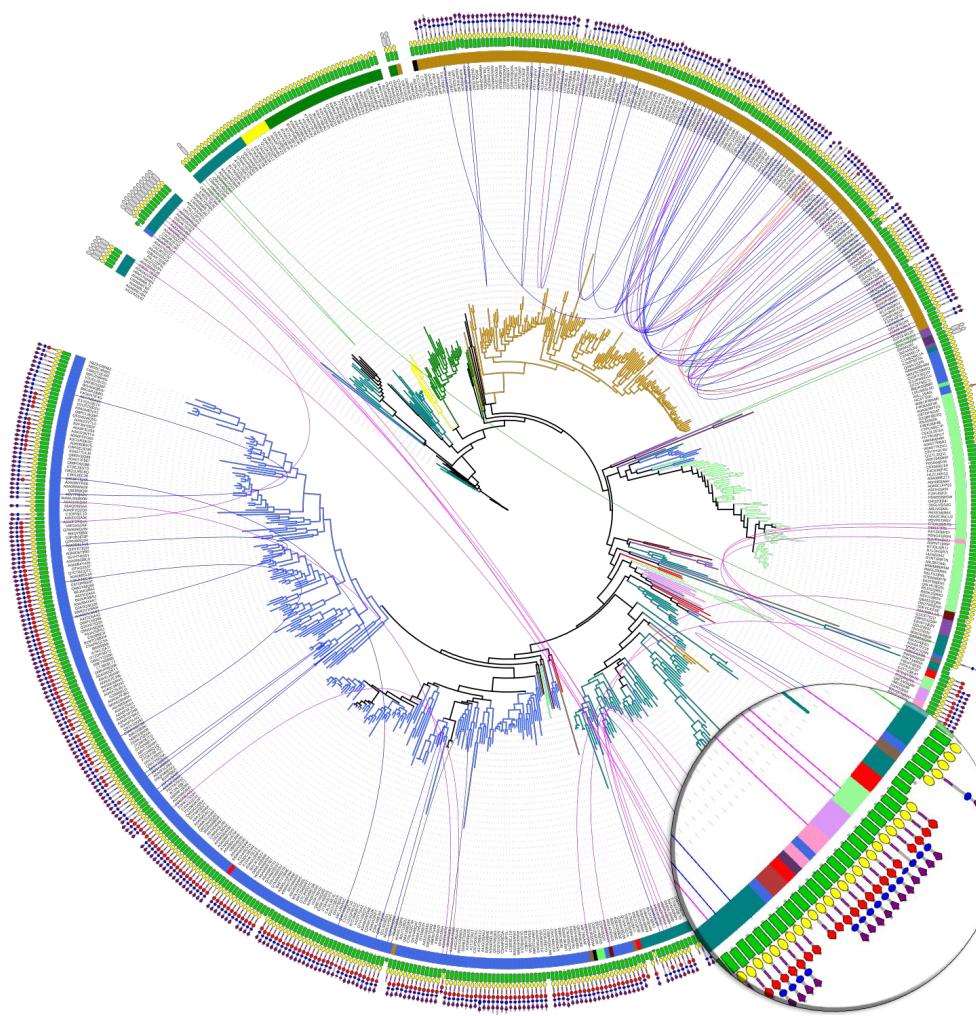
À chaque fois

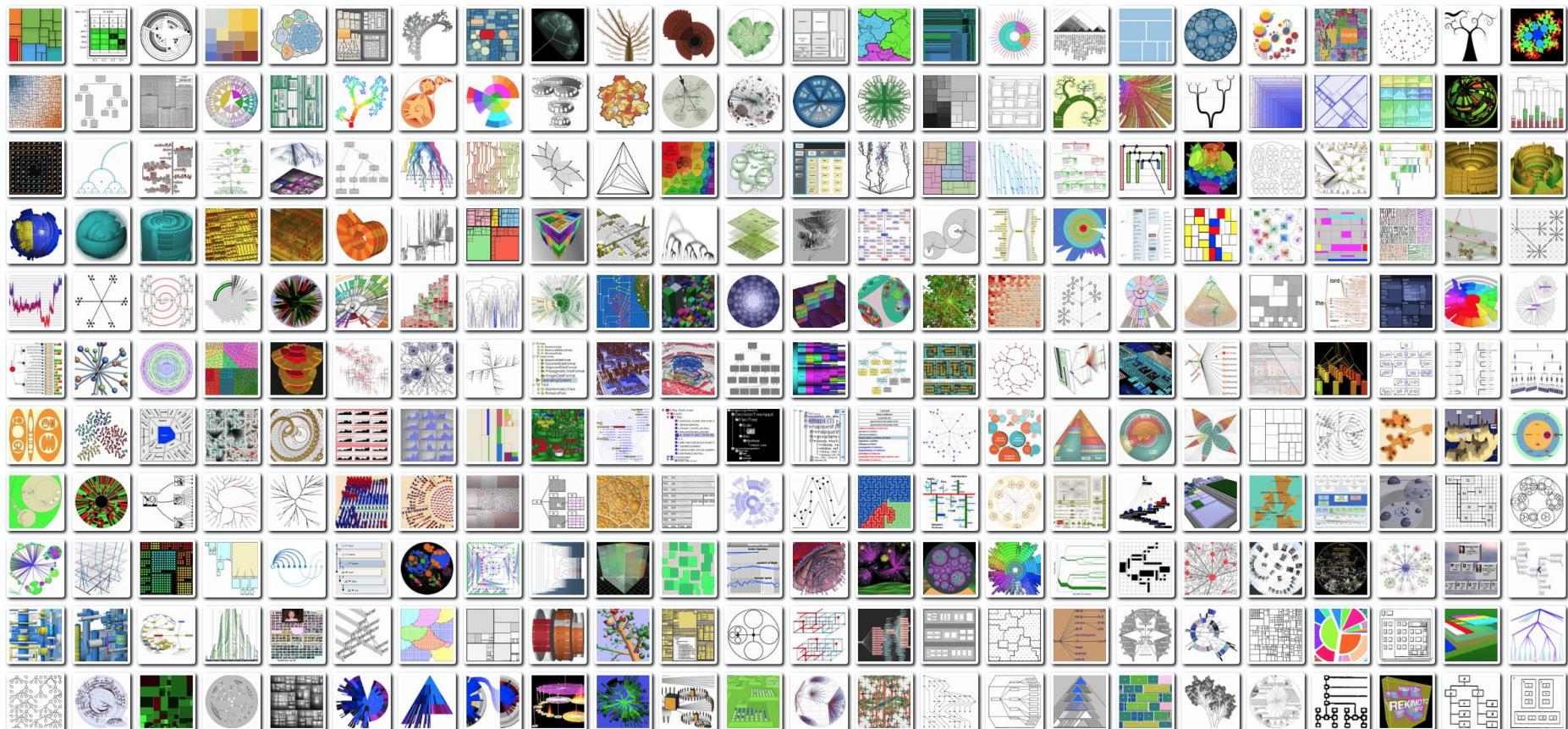
Le même outil pour représenter l'arbre et pour représenter les données associées.

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



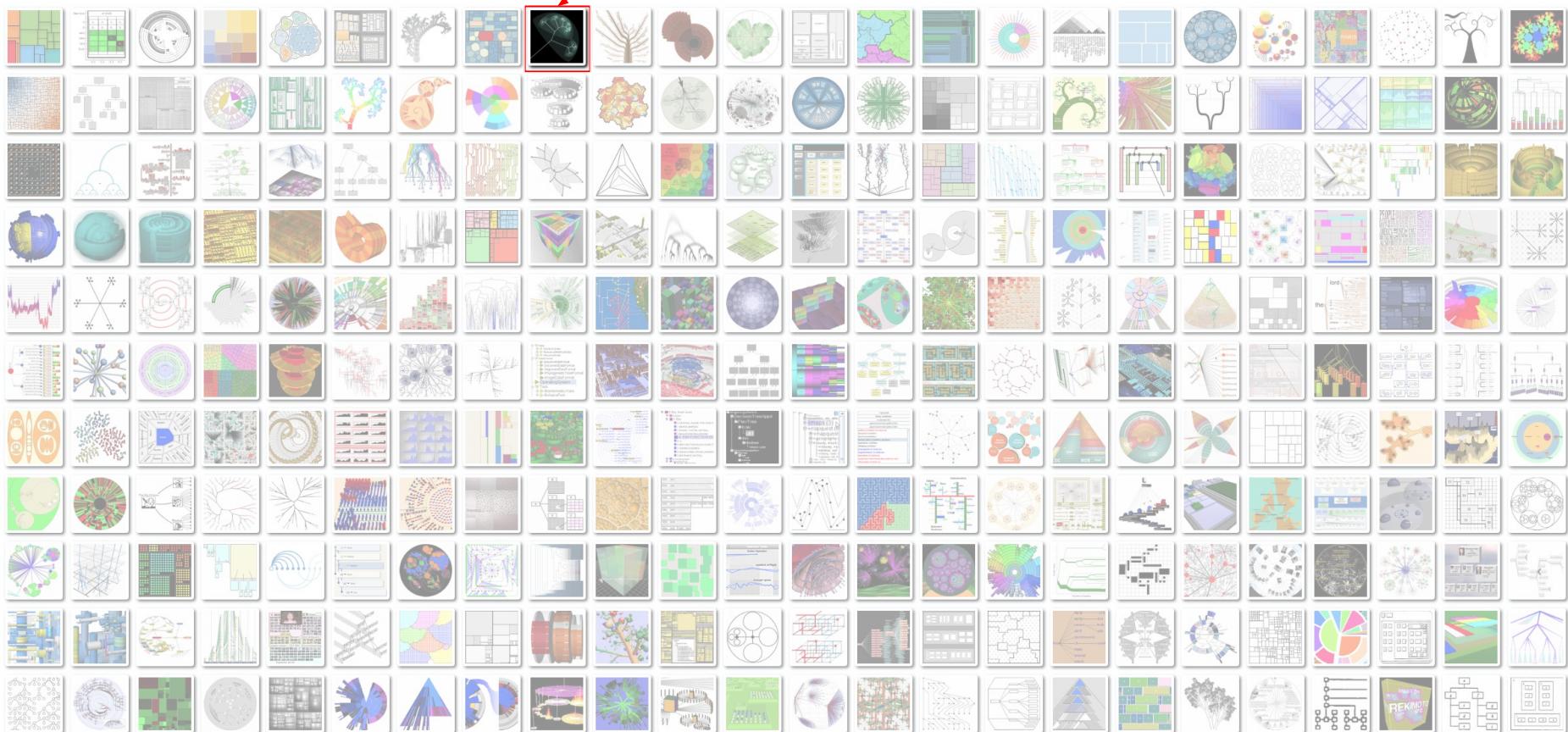
Le “problème” posé par l’arbre de la vie





<https://treevis.net/>

Lifemap (lifemap.univ-lyon1.fr)



<https://treevis.net/>

Lifemap

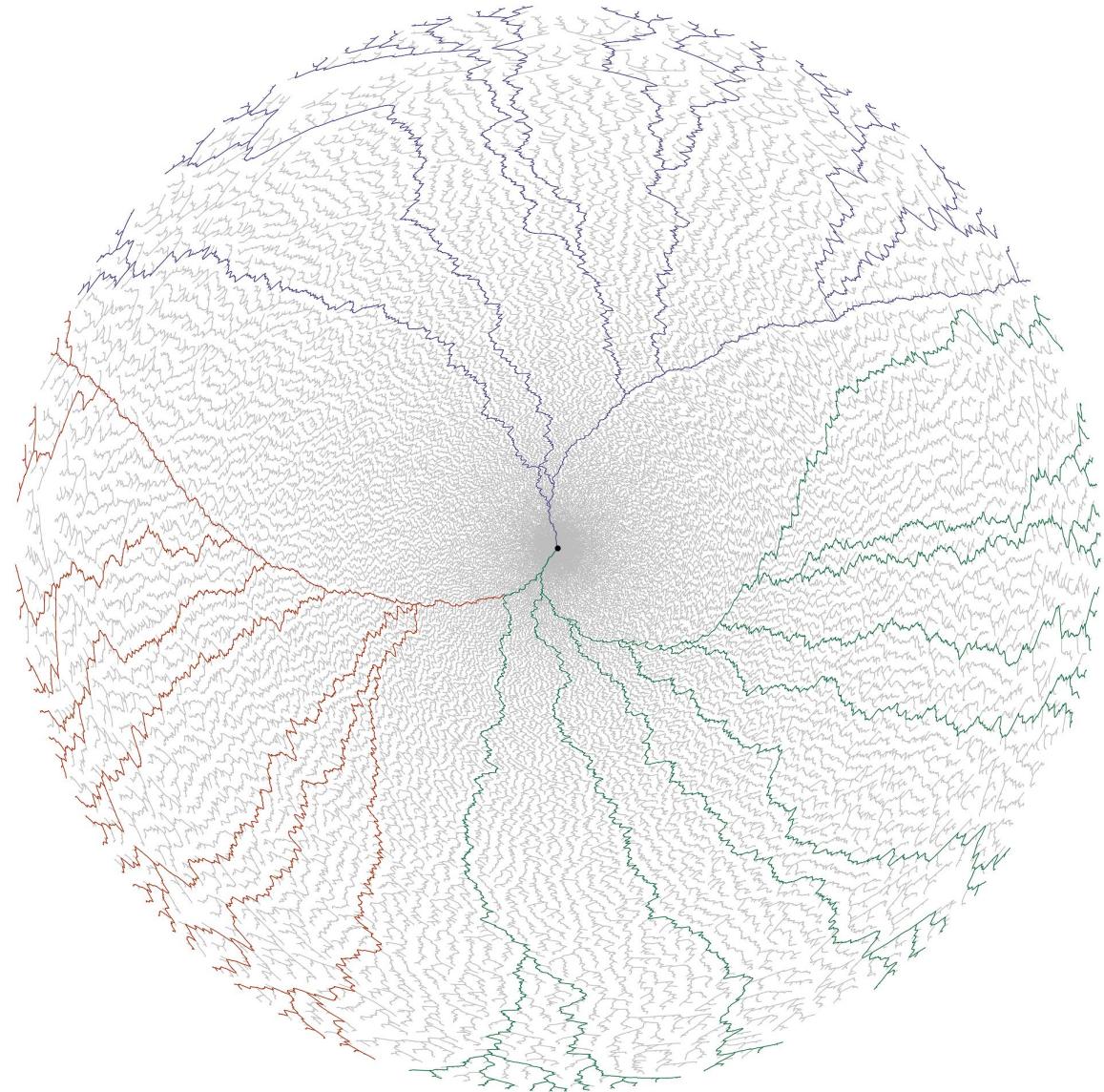


Deep Zoom

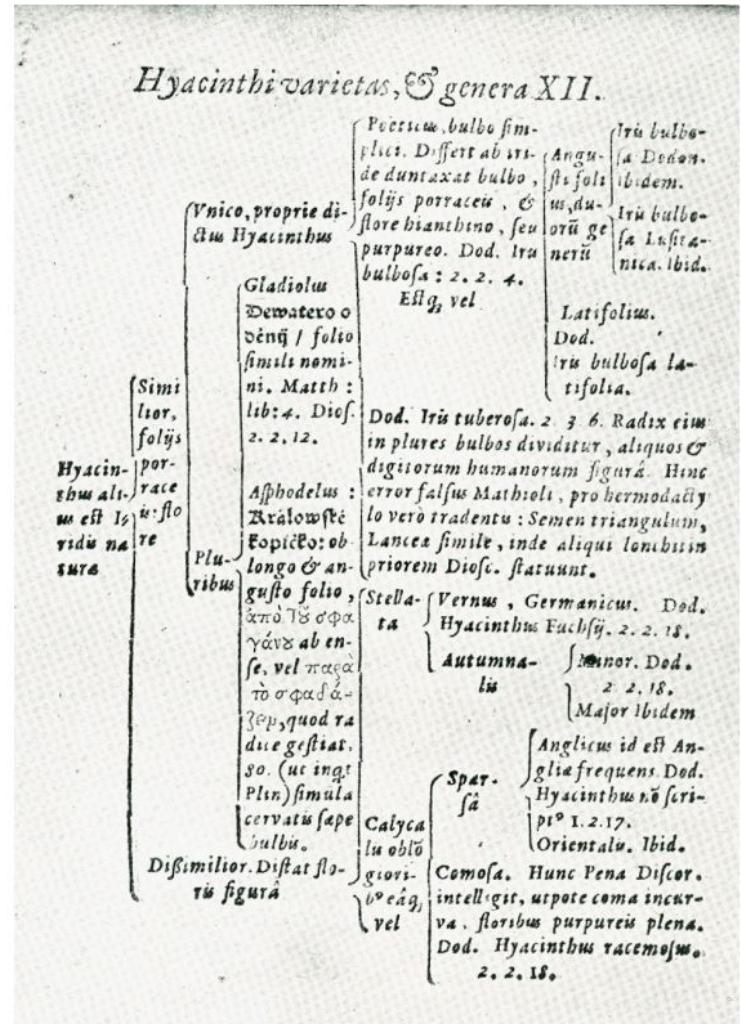
Dissocier l'arbre et les données additionnelles

Arbre du vivant

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



L'arbre du vivant avant Darwin



Review > Biol Direct. 2009 Nov 16;4:43; discussion 43. doi: 10.1186/1745-6150-4-43.

Trees and networks before and after Darwin

Mark A Ragan ¹

Affiliations + expand

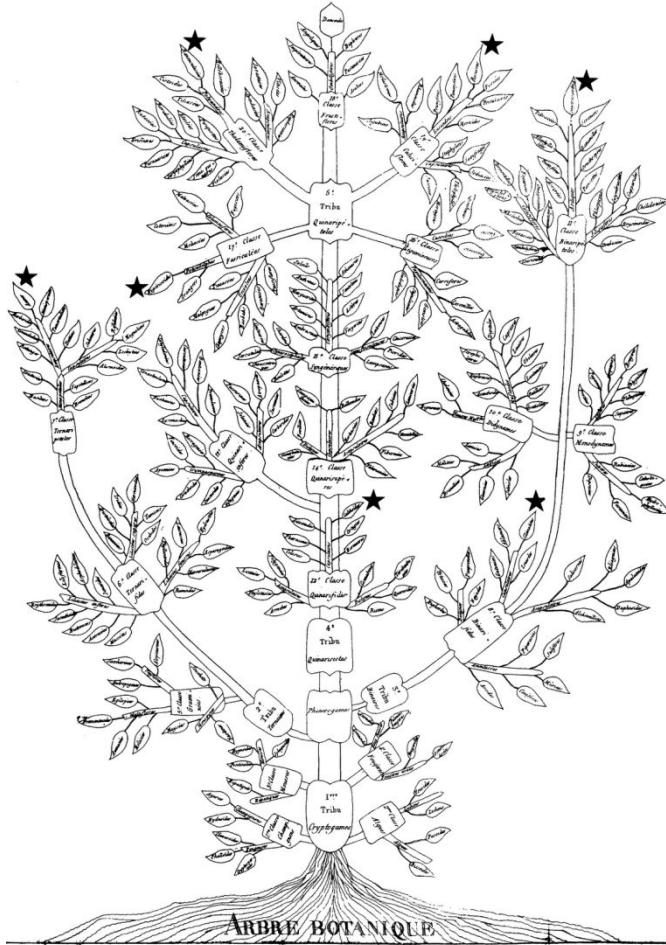
PMID: 19917100 PMCID: PMC2793248 DOI: 10.1186/1745-6150-4-43

Free PMC article

Figure 3
Bifurcating key of hyacinth species, from the *Methodi herbariae libri tres* of Adam Zaluziansky à Zaluzian (1592).



L'arbre du vivant avant Darwin



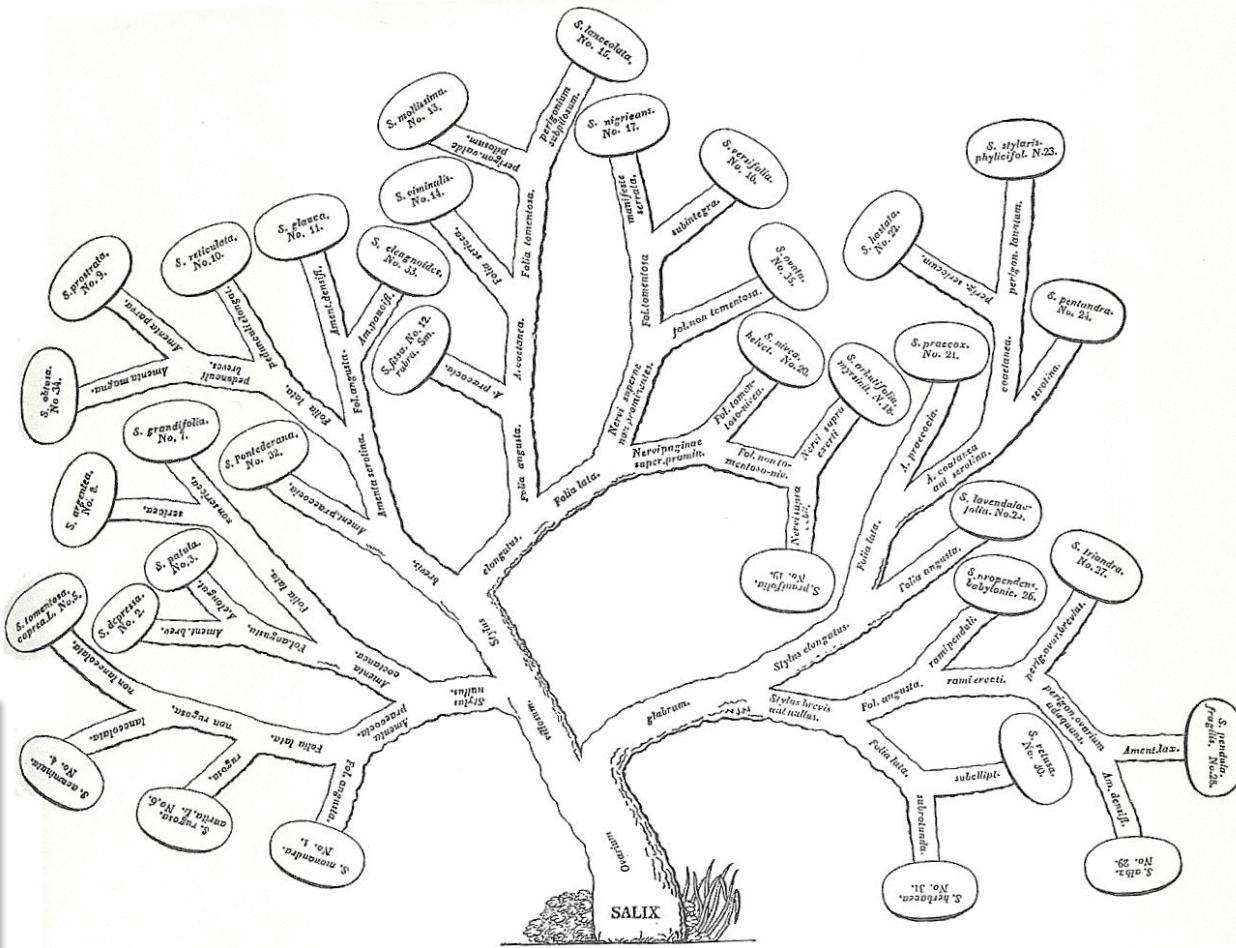
Plant families

“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”

Augustin Augier, 1801



L'arbre du vivant avant Darwin



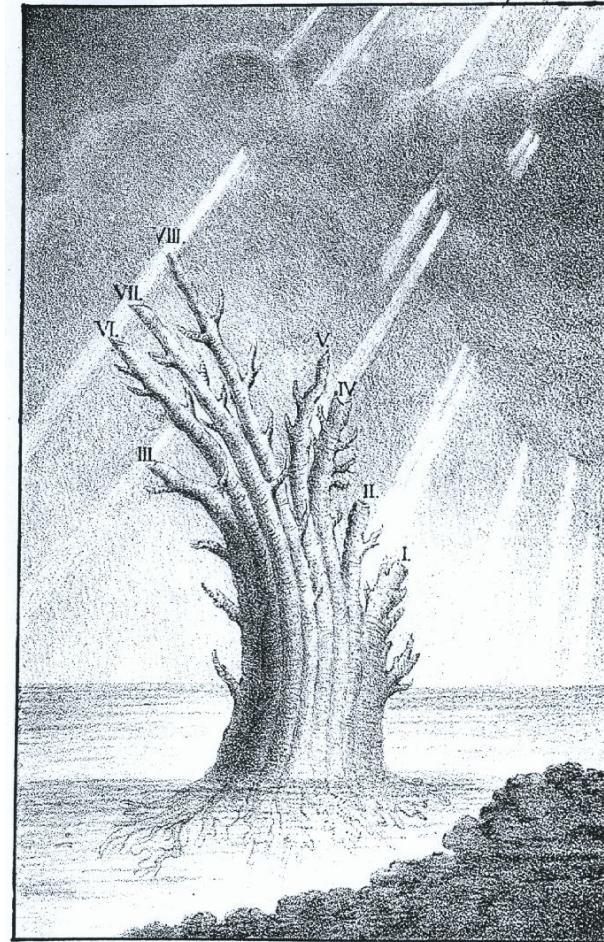
Saules

Nicolas Charles Seringe, 1815



L'arbre du vivant

avant Darwin

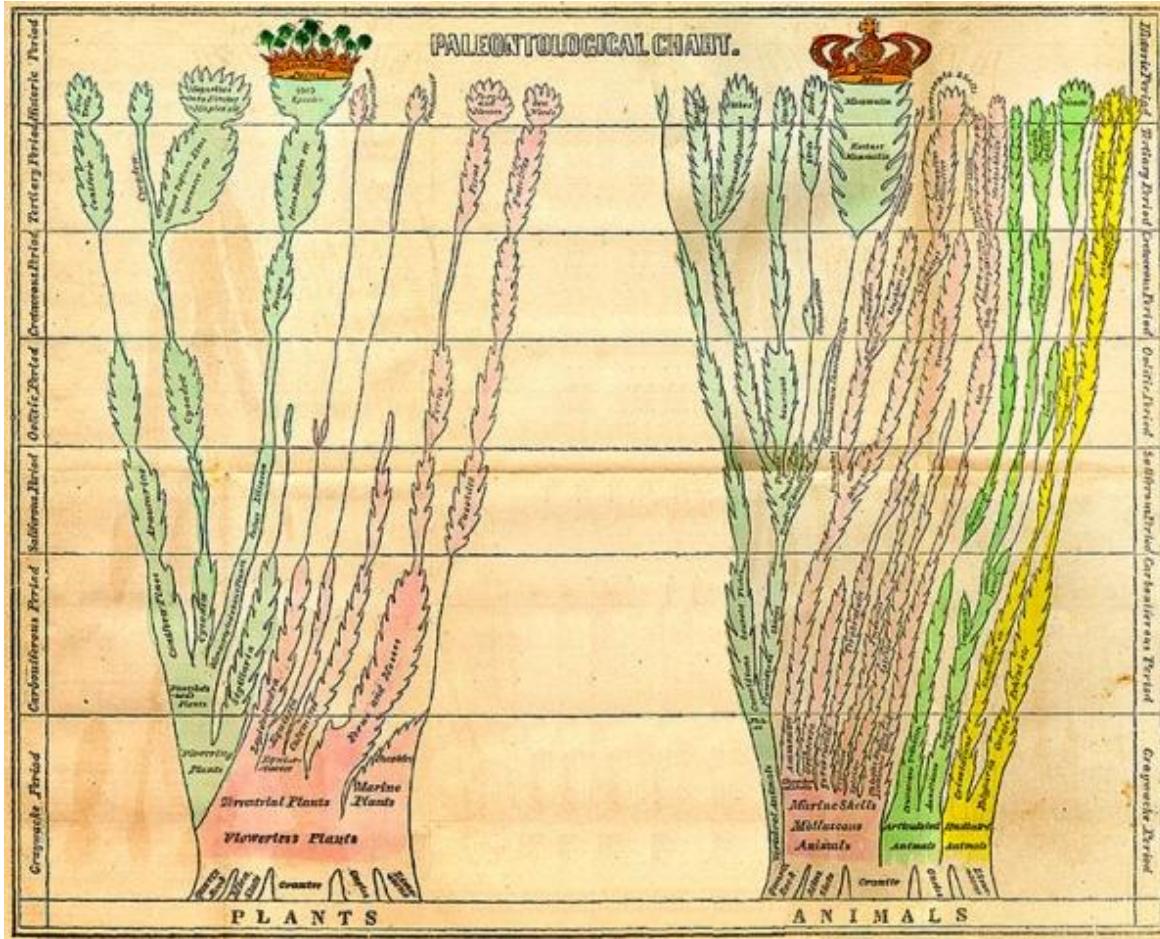


Animal ‘types’

Eichwald, 1829



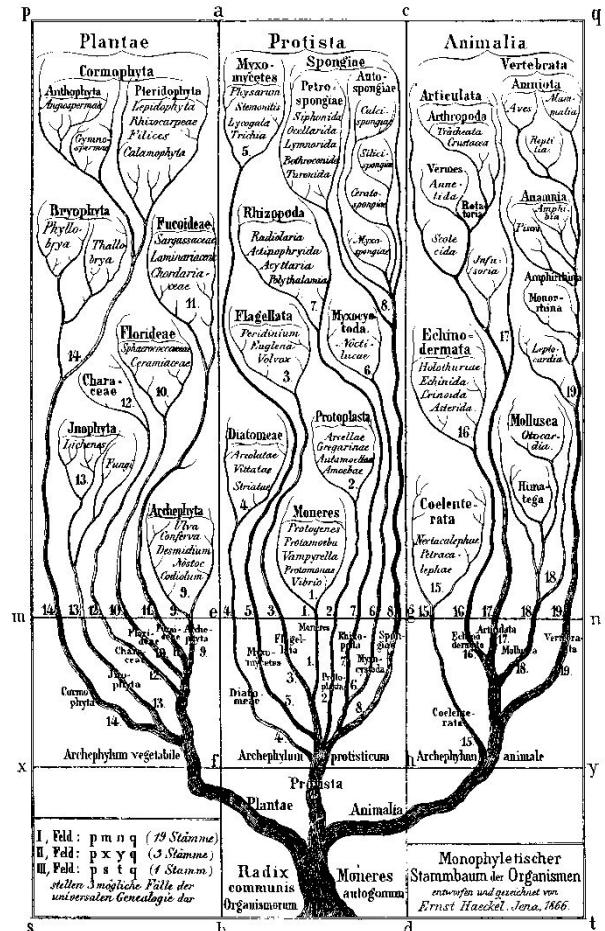
L'arbre du vivant avant Darwin



Fossil times, extinctions, abundance, ‘relationships’

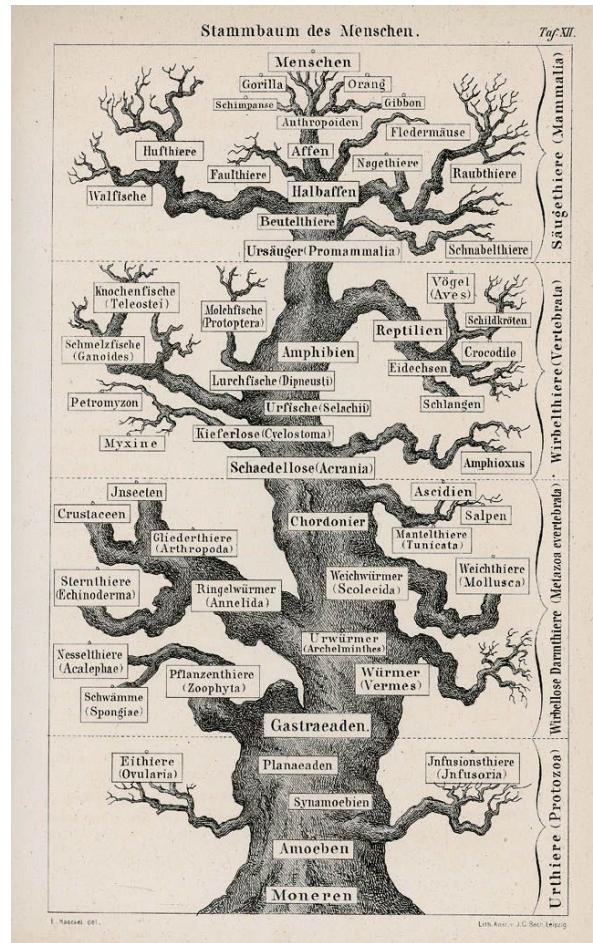
Hitchcock, 1840 (ed. 1852)

L'arbre du vivant après Darwin



Haeckel et al., 1866

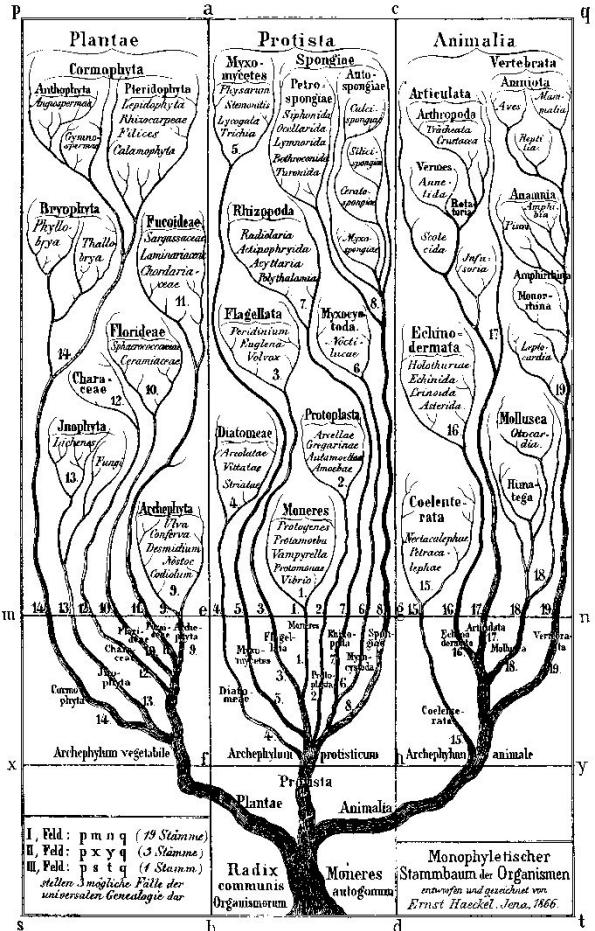
Animals, plants, protists



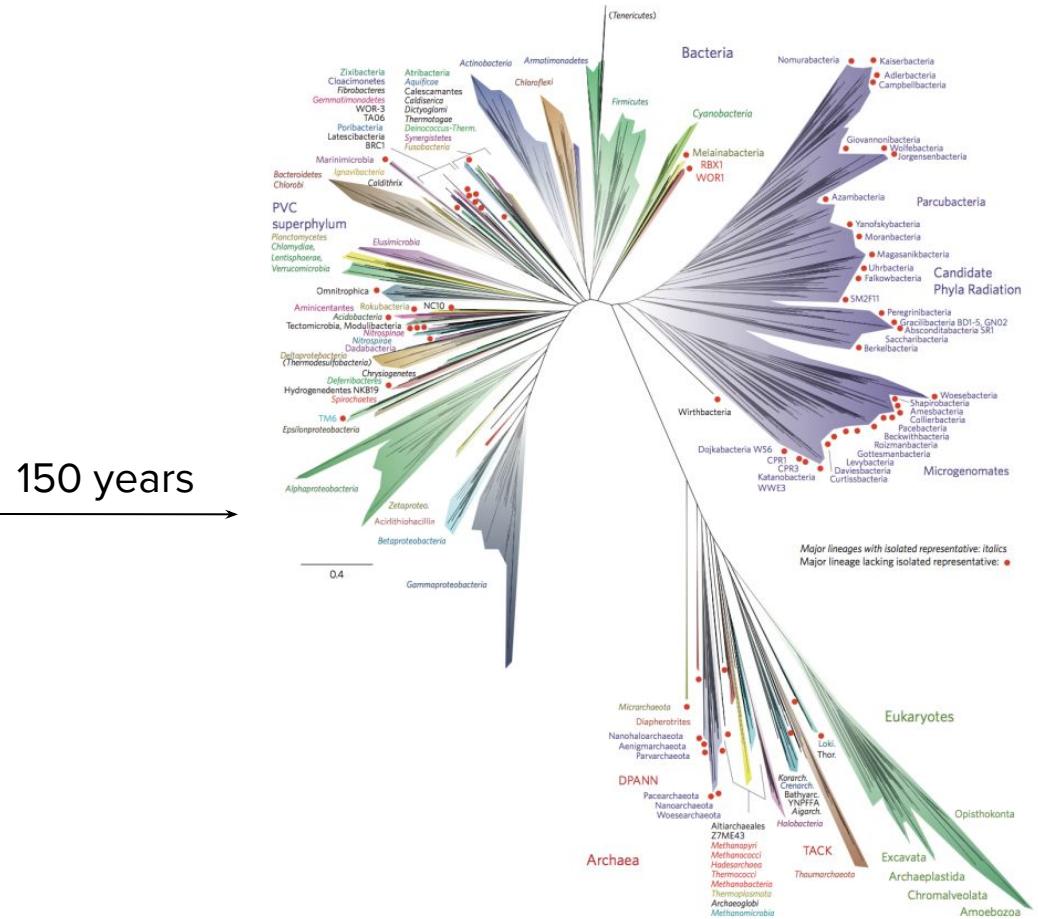
Haeckel, 1874 (?)



L'arbre du vivant après Darwin



Haeckel et al., 1866



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



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

Yarza *et al.* 2008, 2010



Wilson, 2003



Hilnchliff *et al.*, 2015



reviewed in Madisson *et al.*, 2007



WIKISPECIES
free species directory





The Tree of Life



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

Yarza *et al.* 2008, 2010



Wilson, 2003



Hilnchliff *et al.*, 2015



reviewed in Madisson *et al.*, 2007



WIKISPECIES
free species directory





The Tree of Life

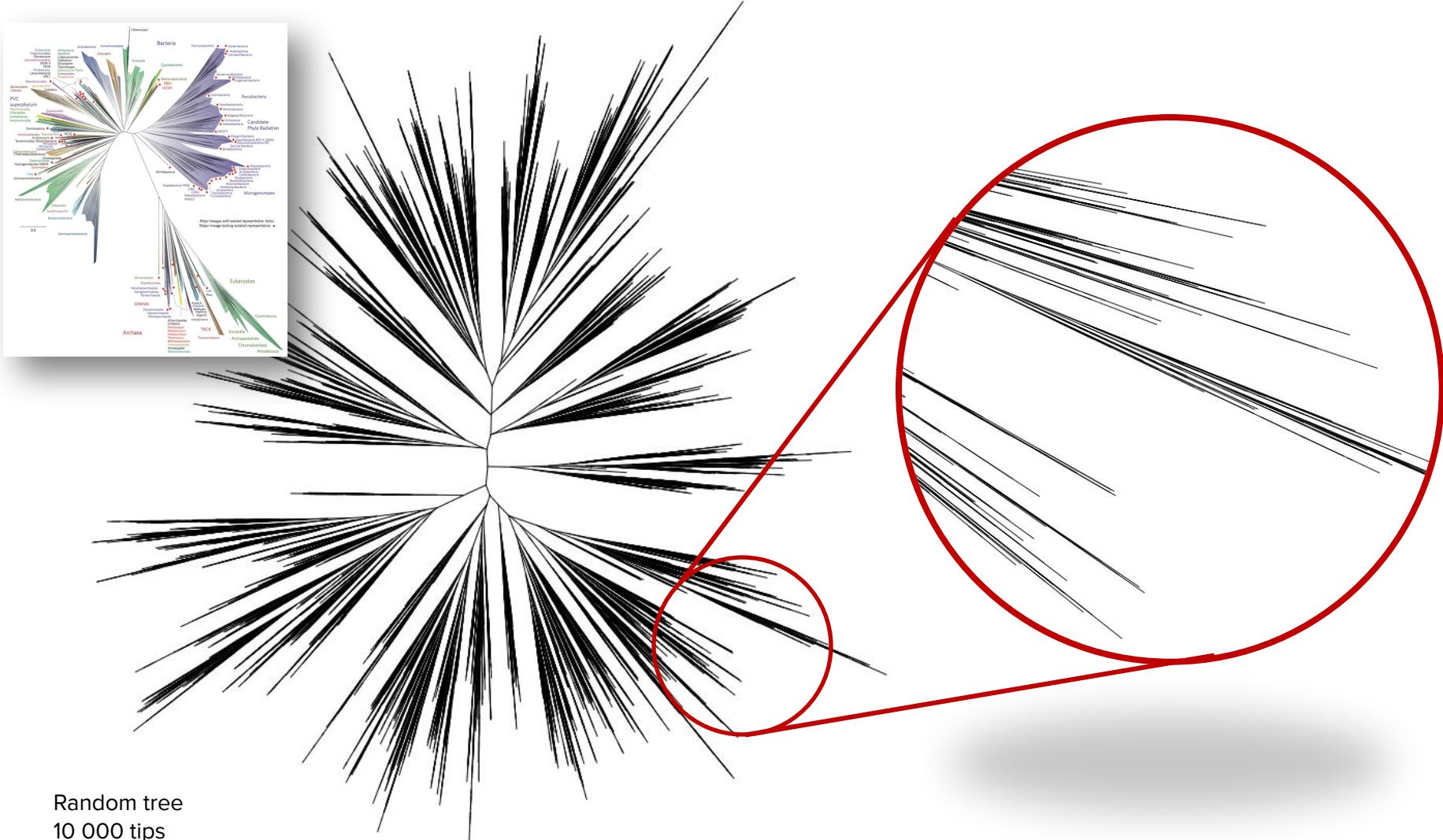
Huge efforts to assemble a comprehensive
Tree of Life...

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

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

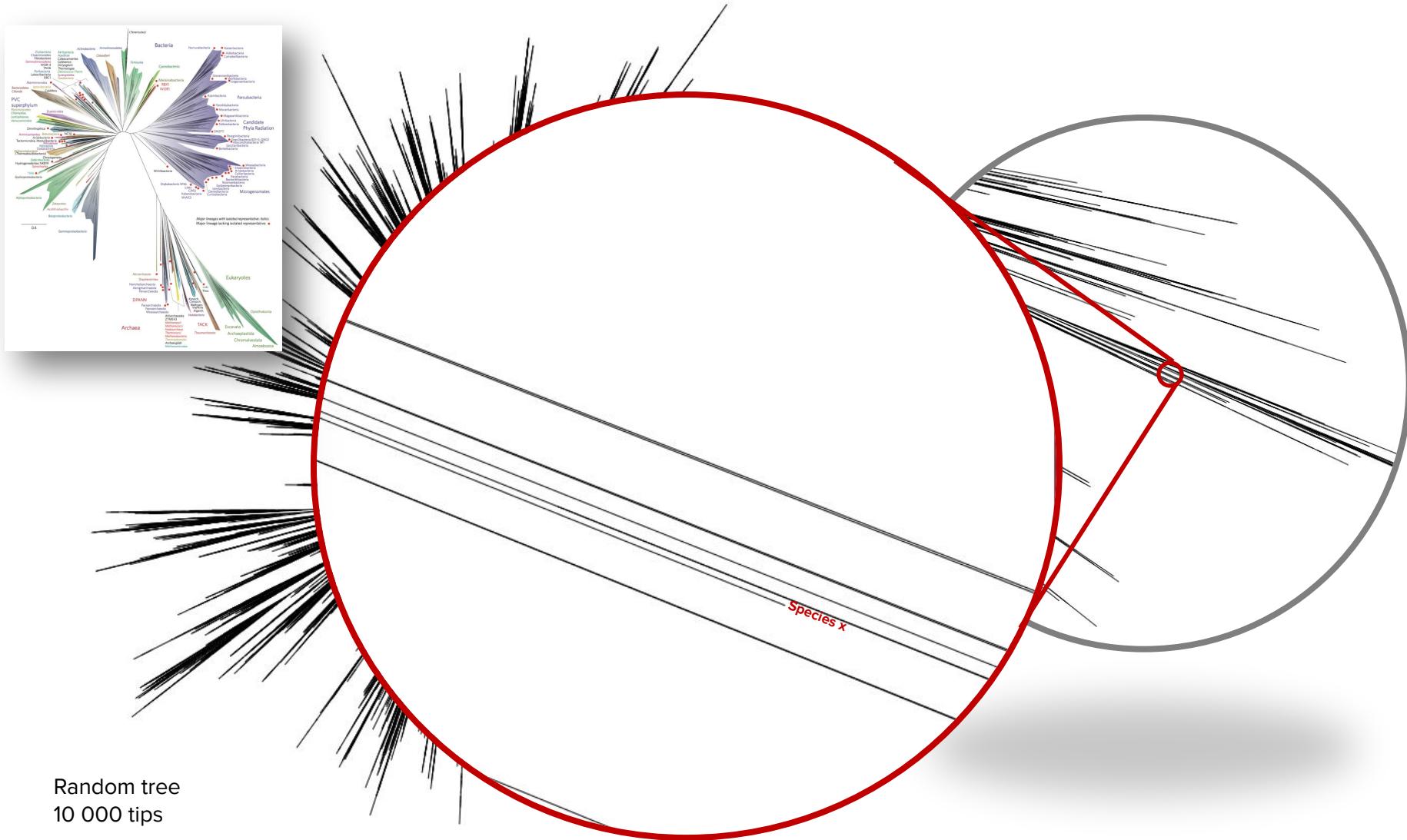
It is difficult to visualize large trees

zooming in is not enough



It is difficult to visualize large trees

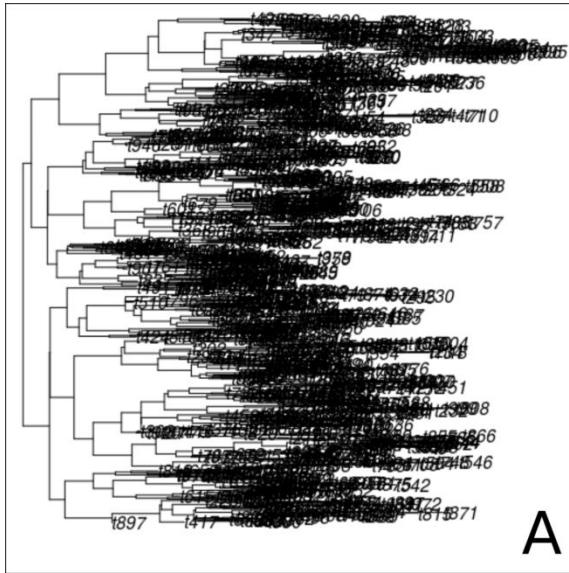
zooming in is not enough





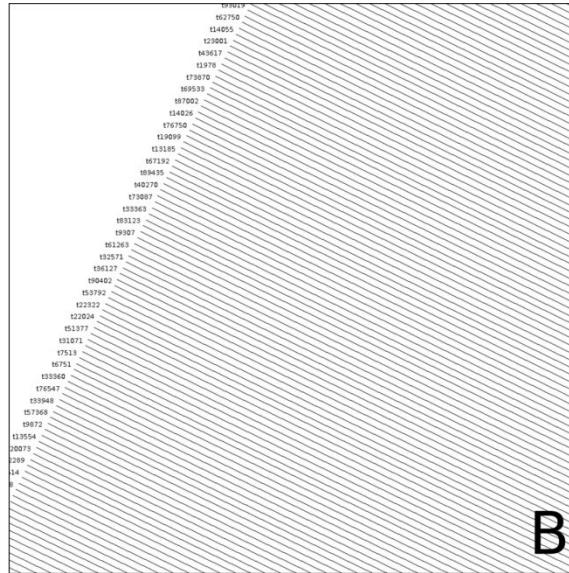
It is difficult to visualize large trees

Labels are problematic...



'ape' package (Paradis et al. 2004)

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



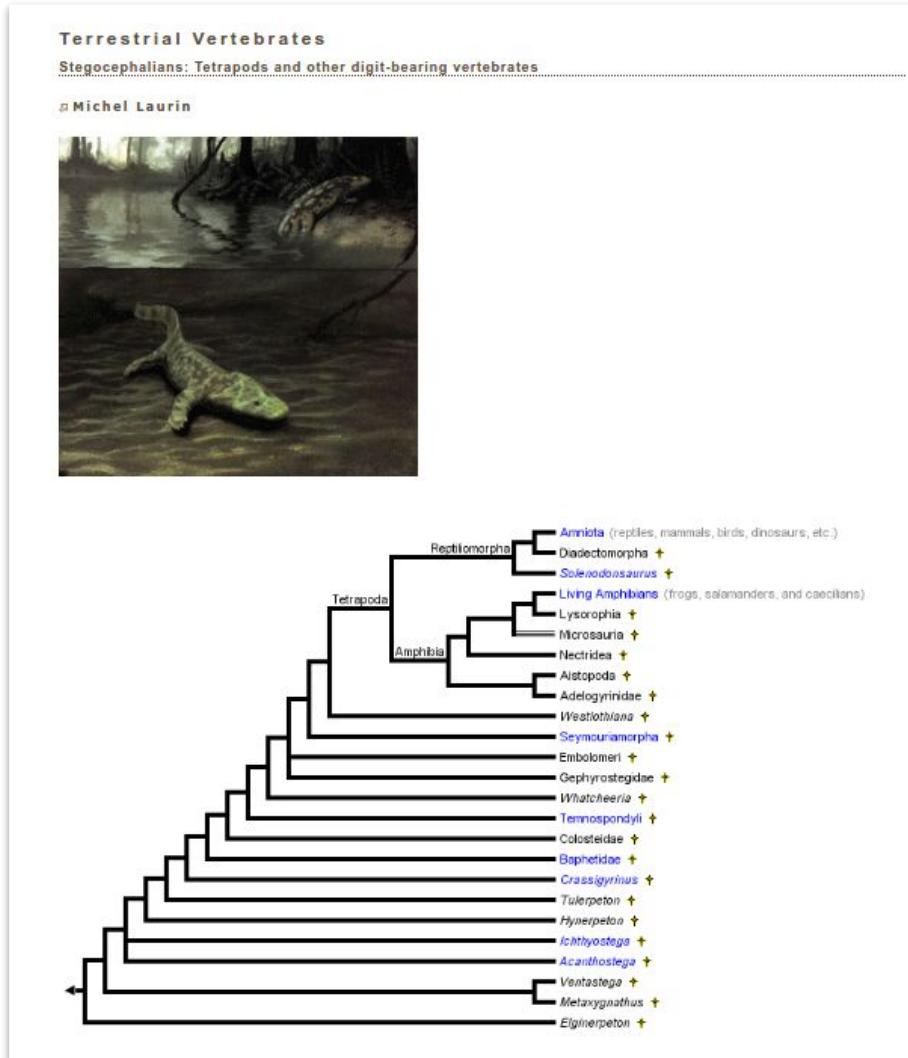
Dendroscope (Huson et al. 2007)

**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





The ‘revolution’ in tree visualization parallel with cartography



Gotlib, 1996, Rubrique-à-Brac, tome 2

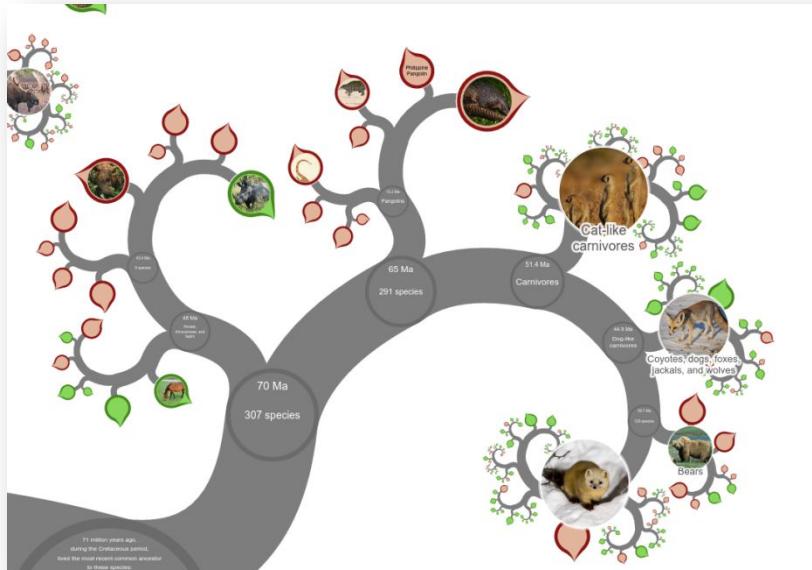
Suggested by Westneat (2009)
and Page (2012) for trees

Deep-zooming





The ‘revolution’ in tree visualization

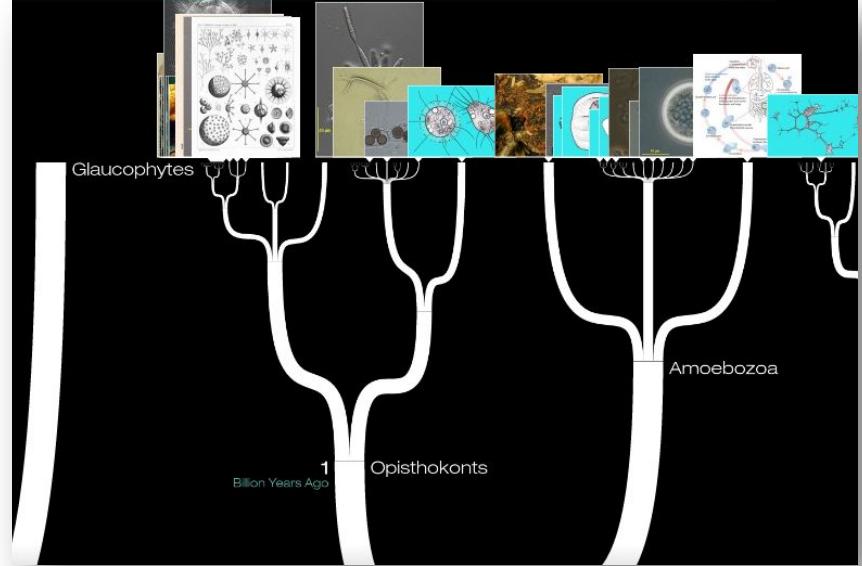


OneZoom
Rosindell and Harmon, 2012

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

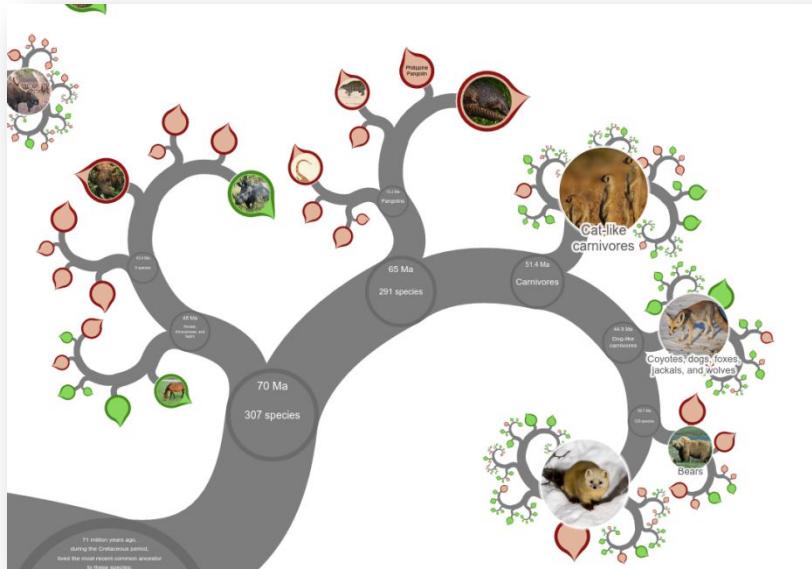


DeepTree
Block et al.,
2012

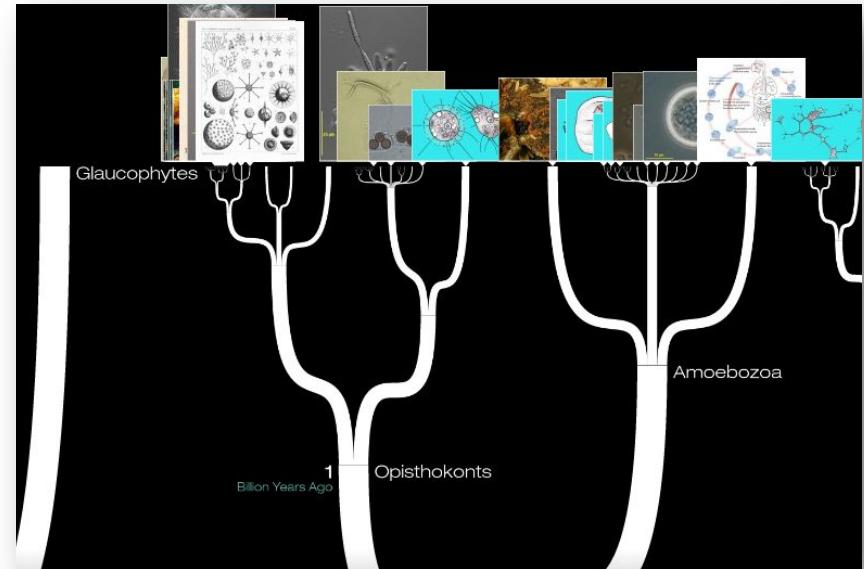
Limited number of species
(70 000 vs 2 Million in the ToL)
Lots of computation while navigating
Fixed (unique) tree



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 ‘revolution’ in tree visualization



OneZoom
Rosindell and Harmon, 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~~



DeepTree
Block et al.,
2012

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

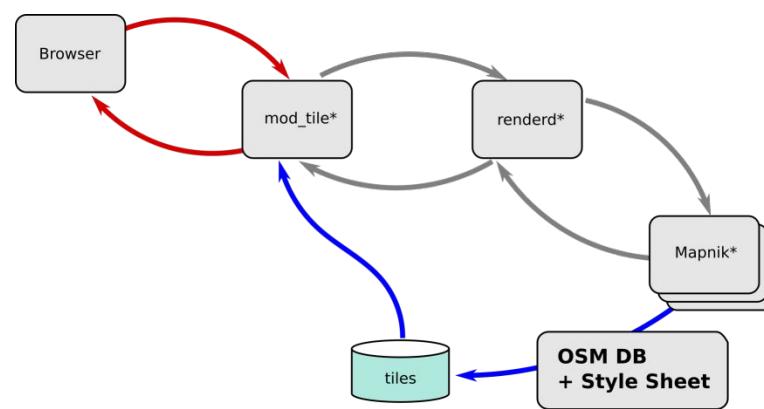


The need for a new visualization

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



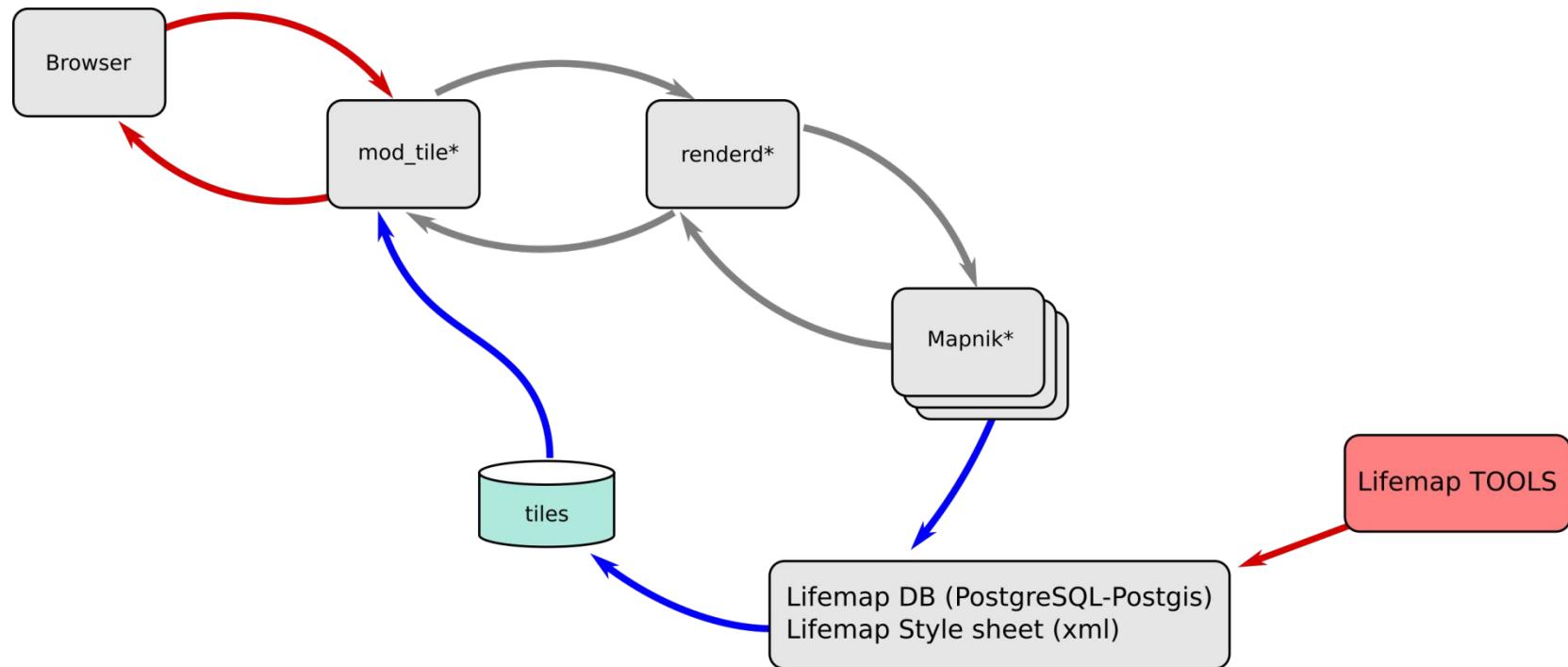
OpenStreetMap





The need for a new visualization

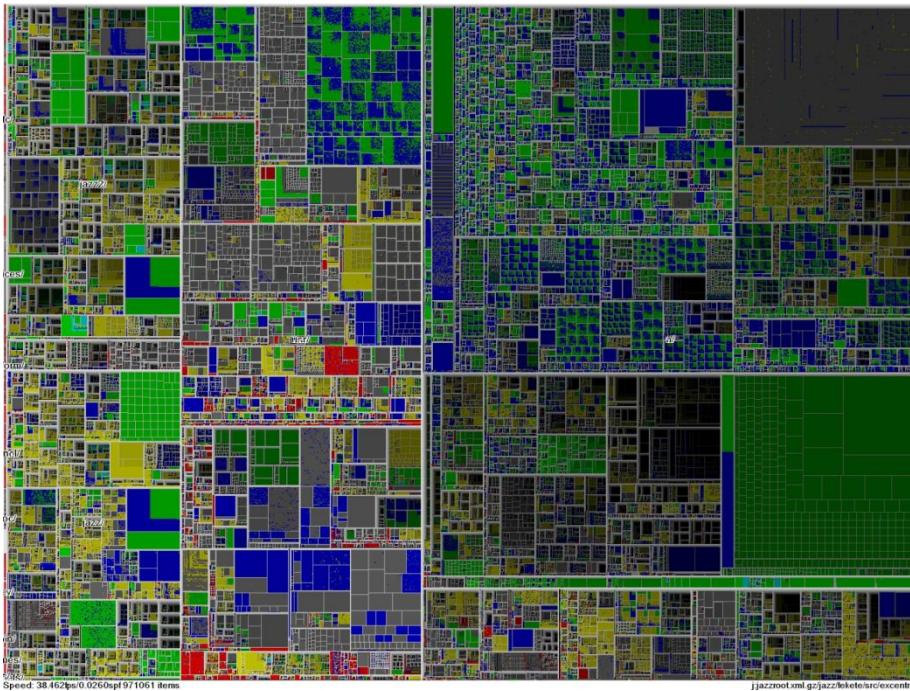
- No limitation in the amount of data (# species)
- Smooth and fast





The need for a new visualization

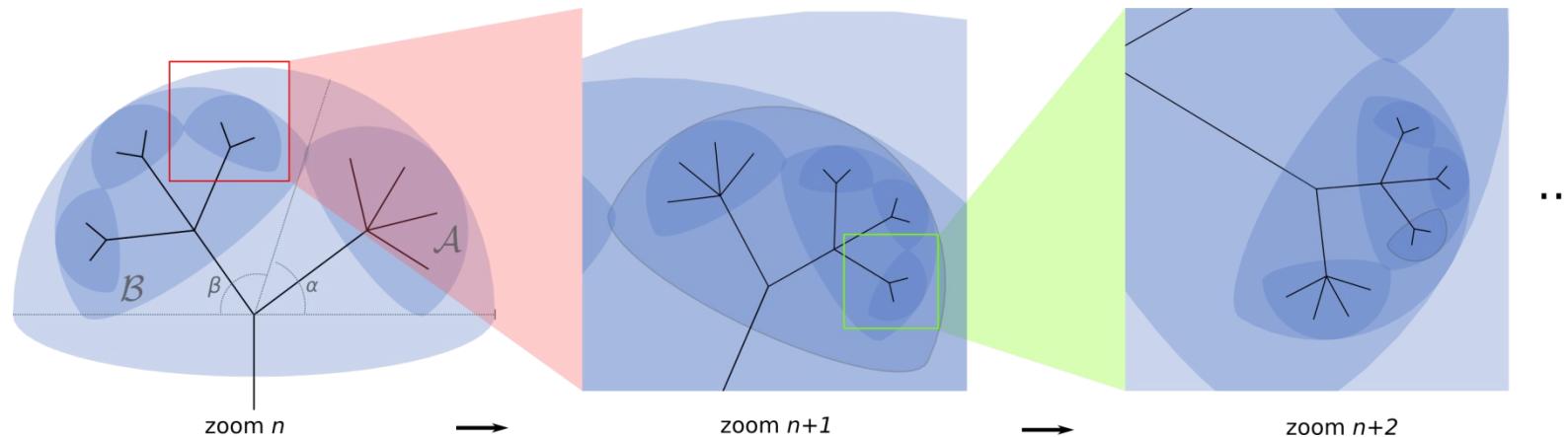
- No limitation in the amount of data (# species)
- Smooth and fast
- Allow multifurcations
- Tree-looking





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

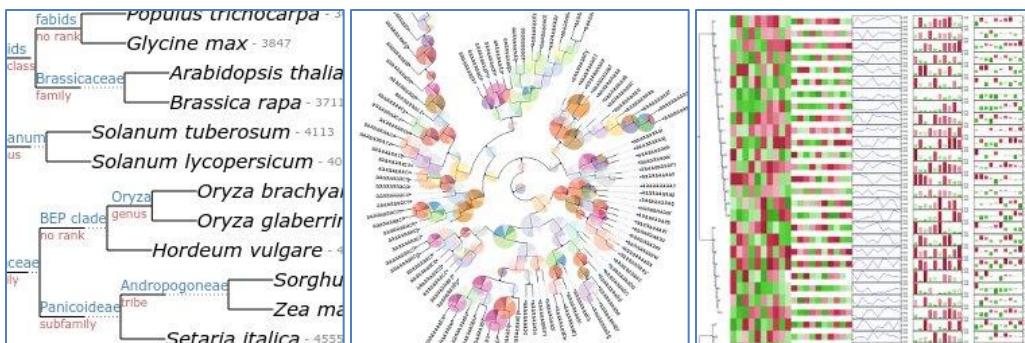
- Get the tree (NCBI, any tree)
- Traverse the tree
 - Compute coordinates of all elements at all zoom levels
 - Get additional information for nodes and tips
 - Write in PostgreSQL/PostGIS database
 - Write in Solr Search Engine

The screenshot shows a Solr search interface with the query 'hol' entered. The results are displayed in a table with three columns: rank, name, and type. The first result is 'Holacanthida ORDER'. The second result is 'Hoplonymphidae FAMILY'. The third result is 'Nutomonas howeae SPECIES'. The fourth result is 'Honigbergiellida ORDER'. The fifth result is 'Honigbergiellidae FAMILY'. The sixth result is 'Haloferax sp. HOB3 SPECIES'.

rank	name	type
1	Holacanthida	ORDER
2	Hoplonymphidae	FAMILY
3	Nutomonas howeae	SPECIES
4	Honigbergiellida	ORDER
5	Honigbergiellidae	FAMILY
6	Haloferax sp. HOB3	SPECIES

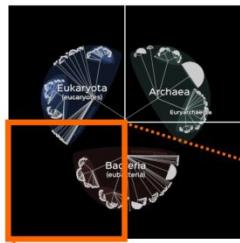
ETE Toolkit

Huerta-Cepas *et al.* 2016

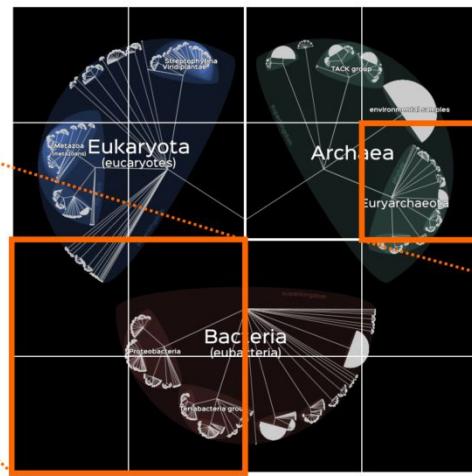


Lifemap

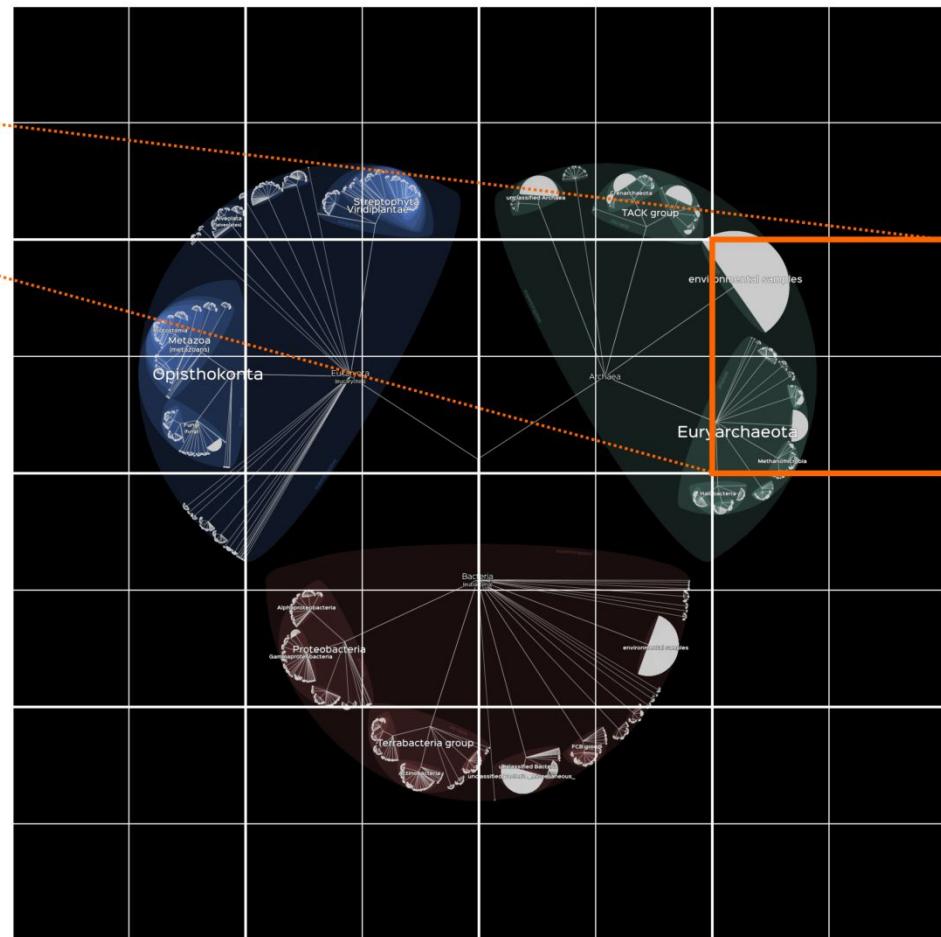
zoom 0



zoom 1



zoom 2

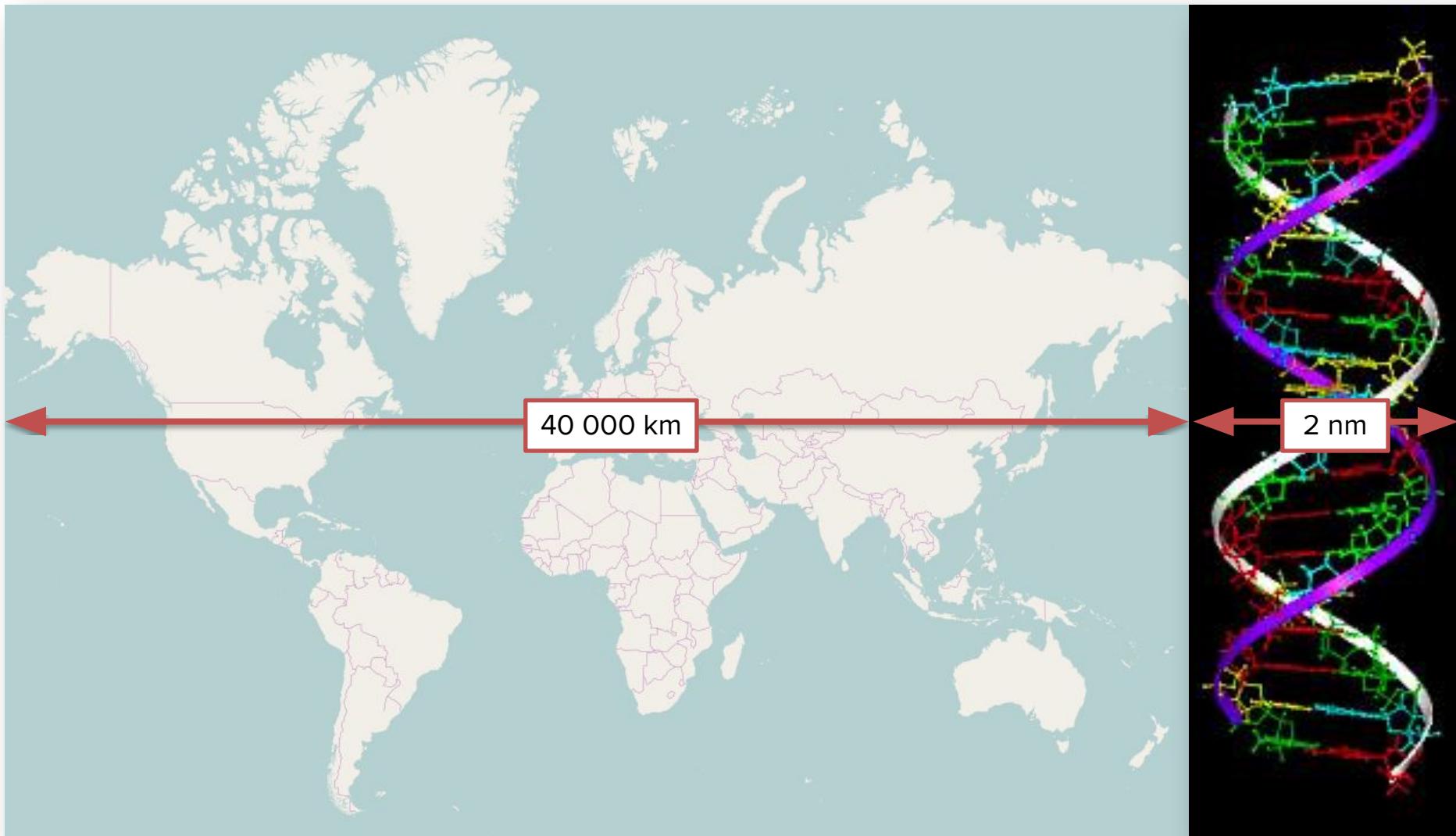


Q : combien de tuiles au zoom 3 ?



Lifemap

52 niveaux de zoom





Lifemap

52 niveaux de zoom

Q : combien de tuiles en tout dans Lifemap ?



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 ?



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

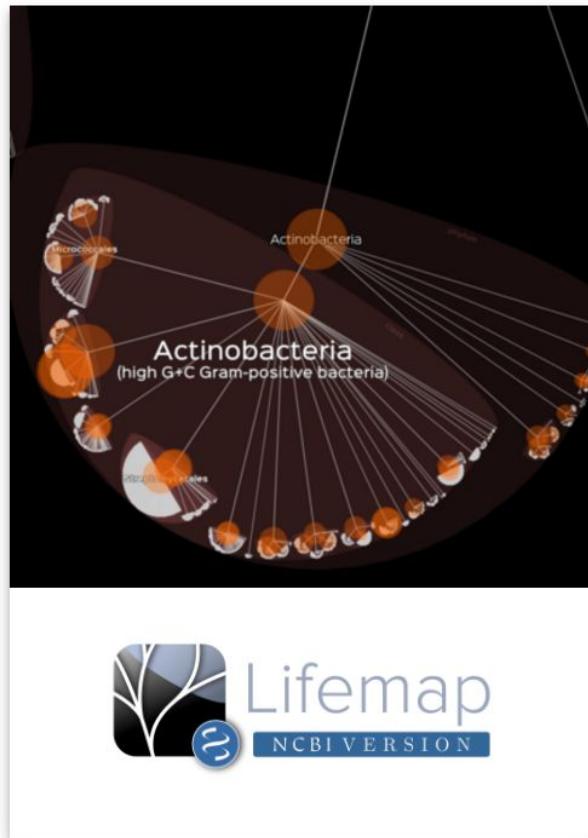
 - For large public, education and research



Lifemap versions



Différentes versions de Lifemap



Full NCBI taxonomy (~ 1 400 000 tips)

Updated weekly

Additional
information:

Opisthokonta NO RANK
NCBI taxid: 33154

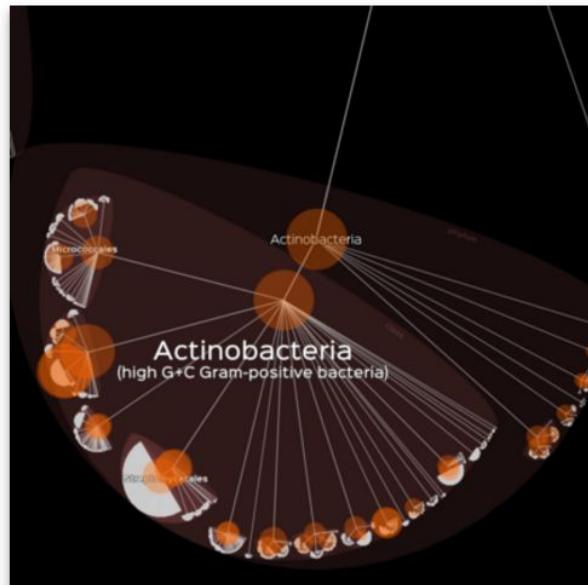
- Number of descendants: 574592
- Explore [Opisthokonta](#) on NCBI
- [Download](#) the complete NCBI subtree of [Opisthokonta](#) in Newick Extended format (NHX)

Explore/edit the wikipedia page of Opisthokonta

[View full ancestry](#) [Close](#)



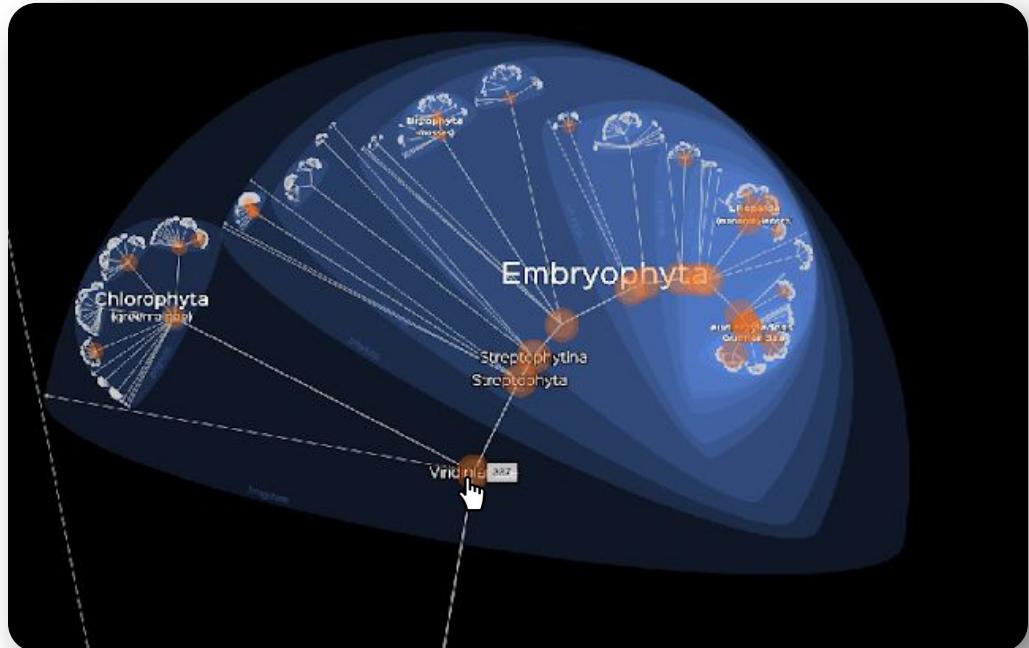
Différentes versions de Lifemap



Lifemap
NCBI VERSION

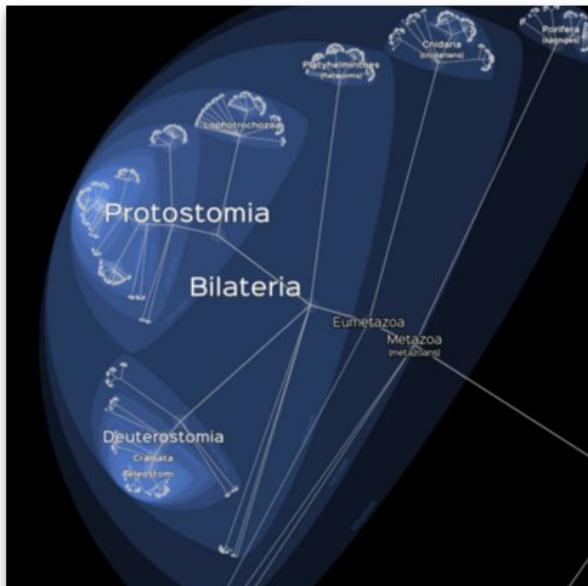
Full NCBI taxonomy (1 220 620 tips)

Updated weekly





Différentes versions de Lifemap



Lifemap
general public

Version grand public

Infos additionnelles de Wikipedia

FR/EN



Eukaryota (eucaryotes) SUPERKINGDOM

A **eukaryote** (ju:'kærɪ.oʊt/ or ju:'kærɪət/ yoo-KARR-ee-oht or yoo-KARR-ee-ət) is any organism whose cells contain a nucleus and other organelles enclosed within membranes. Eukaryotes belong to the taxon **Eukarya** or **Eukaryota**. The defining feature that sets eukaryotic cells apart from prokaryotic cells (Bacteria and Archaea) is that they have membrane-bound organelles, especially the nucleus, which contains the genetic material and is enclosed by the nuclear envelope.

[more on Wikipedia](#)



[View full ancestry](#)

[Close](#)



Différentes versions de Lifemap

The image shows a composite screenshot of the Lifemap mobile application interface. On the left, a large phylogenetic tree diagram is displayed against a dark blue background. The tree highlights various taxonomic groups, including Eukaryota (eucaryotes), Archaea, Bacteria (eubacteria), and several primate families like Haplorrhini, Catarrhini, and Tarsiiformes. A search bar at the top left says "Search species, clade, ...". A modal window in the center shows a detailed view of the "Homo sapiens (human)" clade. Below it, a list of species under the "Laurasiatheria" superorder is shown, including *Homo heidelbergensis*, *Cinara cf. cembrae* (two entries), and *Trichuris sp.*. To the right, another modal window displays images of a giraffe, a bat, and a hedgehog, with a "View full ancestry" button. At the bottom right, a separate phylogenetic tree diagram focuses on the Euarchontoglires clade, specifically highlighting the Rodents and rabbits group. A red line traces the ancestry from the MRCA of Euarchontoglires back to the common ancestor of Homo sapiens and Mus musculus. The bottom of the screen features download links for Google Play and the App Store.

Search species, clade, ...

Homo sapiens (human)

homo s

Homo sapiens (human)
SPECIES

Homo heidelbergensis (Heidelberg man)
SPECIES

Cinara cf. cembrae **Homo_H1**
SPECIES

Cinara cf. cembrae **Homo_H2**
SPECIES

Trichuris sp. ex **Homo sapiens JP-2011**
SPECIES

Laurasiatheria SUPERORDER

Laurasiatheria is a superorder of placental mammals believed to have originated on the northern supercontinent of Laurasia.
more on [Laurasiatheria](#)

Homo sapiens (human)

Mus musculus (house mouse)

View full ancestry

Close

Euarchontoglires
(Rodents and rabbits)

Primates

Homo sapiens

MRCA Euarchontoglires

Mus musculus

GET ON Google Play

Download on the App Store



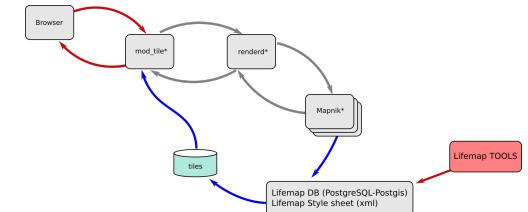
Et une version pour visualiser ses propres arbres



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 × localhost:5580

Welcome to Lifemap server

- 1 Upload a tree file (newick)
- 2 Prepare the tree on the server
- 3 EXPLORE YOUR TREE

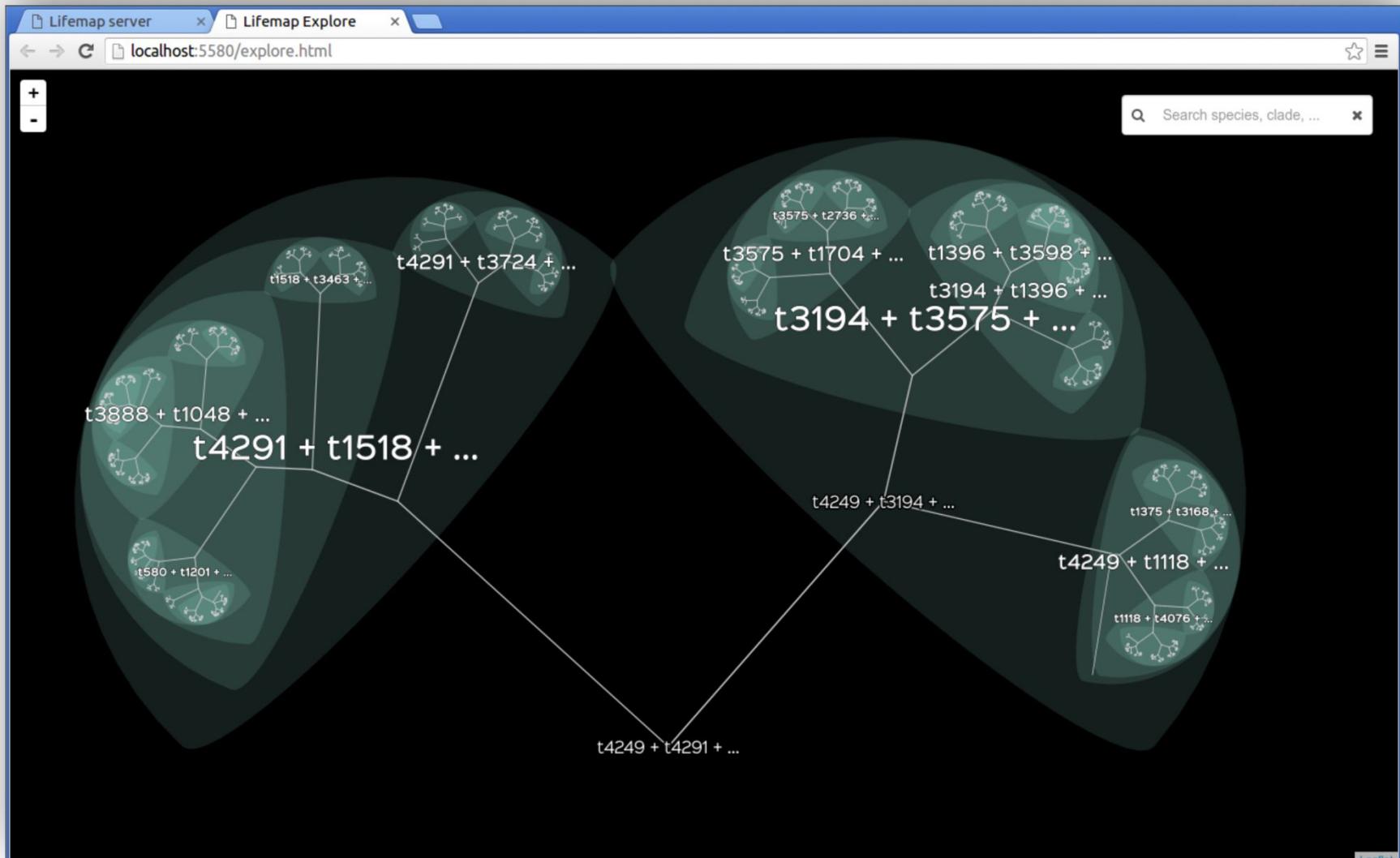
i Select the file on your computer that contains the tree that you want to visualize and explore. The tree must be in **Newick** (parenthetic) format. The extension of the file (.tre, .nw, .txt, ...) does not matter.

...or explore the last tree uploaded

This is Lifemap-server version 1.0.0, © 2016-2020
If you use Lifemap for your communication or publication, please cite the following paper:
D.M. de Vienne. 201x. Lifemap: Exploring the Entire Tree of Life. *Journal.issue.pages*
For any comment or bug report, please send an email to the author at dmdevienne@biomimicry.org



Lifemap-server



ANDROID APP
Google play

Télécharger dans
l'App Store

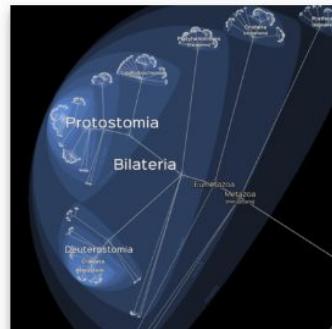
[Tweet](#)

[Share](#)



Lifemap

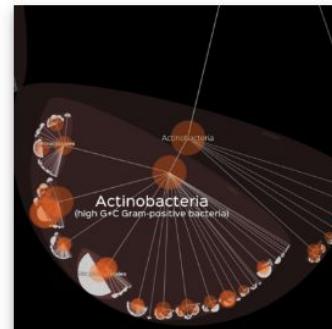
Exploring the Entire Tree of Life



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 description and picture retrieved from Wikipedia when clicking on the tips and nodes.



Complete NCBI taxonomy

~14 Million species

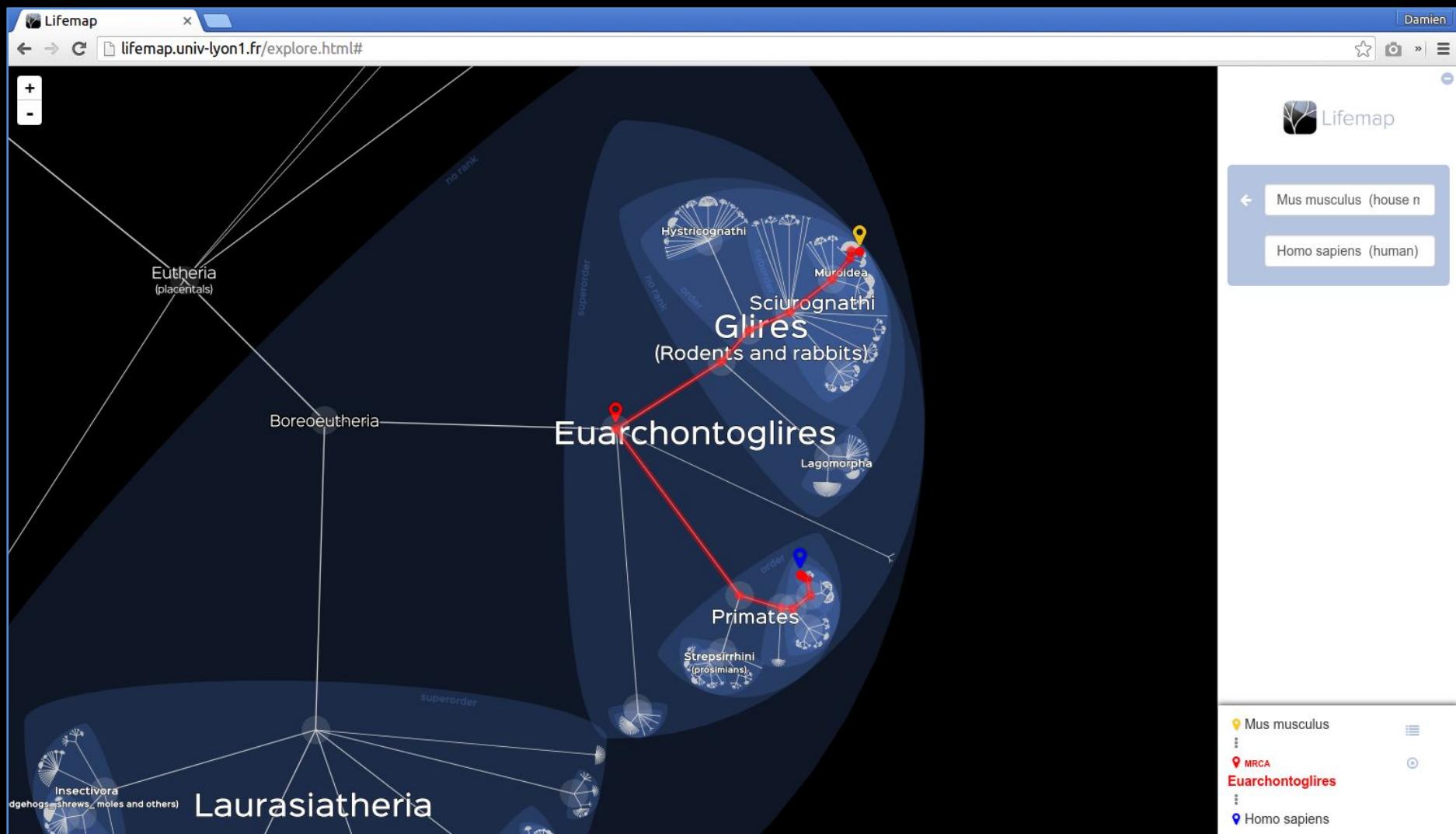
The NCBI taxonomic tree represents the relationships between all species/strains that have at least one sequence in the NCBI/Entrez database. It is curated and validated by expert taxonomists and is updated regularly.



Lifemap by Damien de Vienne / CNRS is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

If you use Lifemap for your publications, please cite:

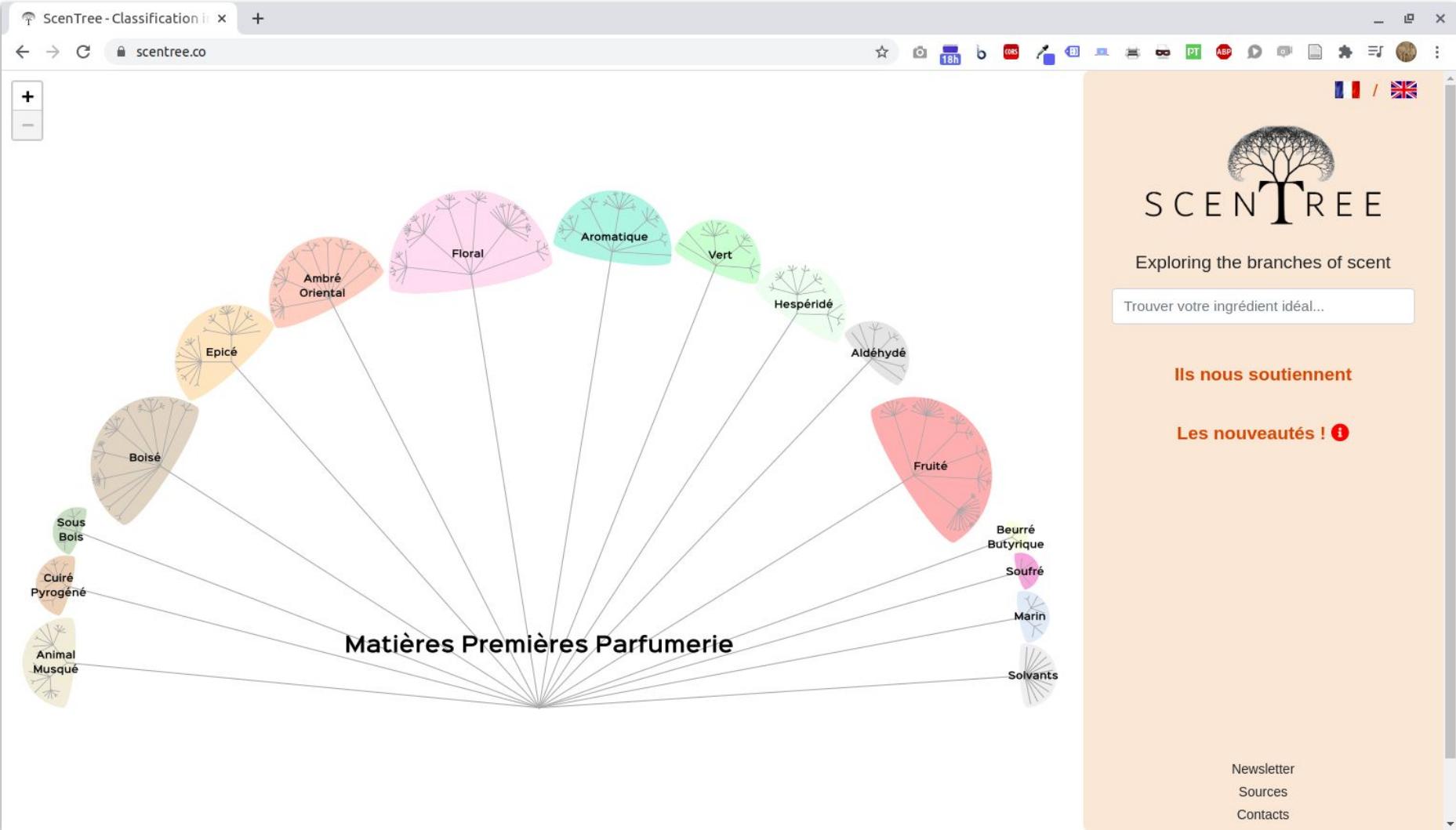
de Vienne DM (2016) Lifemap: Exploring the Entire Tree of Life. *PLOS Biology* 14(12): e2001624. doi: 10.1371/journal.pbio.2001624.





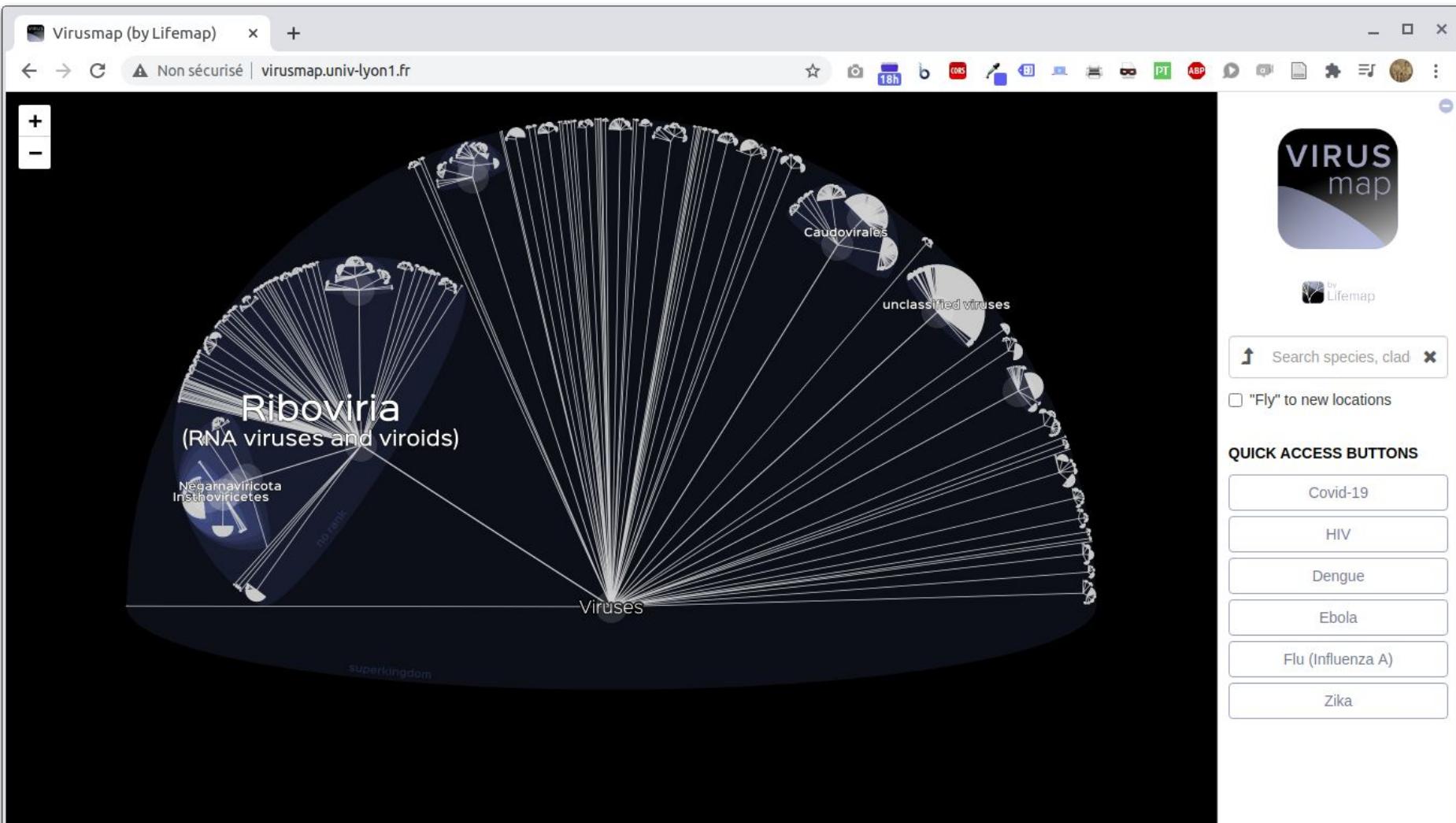
Lifemap et projets associés

- Lifemap a été détourné pour visualiser les "senteurs" en parfumerie
- Lifemap devrait rejoindre le futur musée des sciences de Moscou (2 millions de visiteurs par an).
- Lifemap va rejoindre une exposition itinérante sur 'les intelligences' des musées des sciences de Nantes, Rennes et Québec
- Un travail avec l'ONG Eudaimon (<https://eudaimon-ecp.org/>) est en cours pour utiliser Lifemap pour visualiser une Mind-Map.
- Discussions en cours avec le comité international de taxonomie des virus (ICTV, <https://talk.ictvonline.org/>) pour qu'ils proposent une visualisation de leur taxonomie des virus
- Projet avec une société pharmaceutique anglaise pour visualiser les plantes thérapeutiques



Lifemap et projets associés

- Lifemap a été détourné pour visualiser les "senteurs" en parfumerie
- Lifemap devrait rejoindre le futur musée des sciences de Moscou (2 millions de visiteurs par an. Nous avons chiffré avec le CNRS à ~54000 € la prestation, nous attendons la réponse)
- Lifemap va rejoindre une exposition itinérante sur 'les intelligences' des musées des sciences de Nantes, Rennes et Quebec (chiffré 3000 €)
- Un travail avec l'ONG Eudaimon (<https://eudaimon-ecp.org/>) est en cours pour utiliser Lifemap pour visualiser une Mind-Map. Recherche de financement par l'ONG. Rien de signé avec eux pour l'instant mais ils vont bientôt vouloir un devis.
- Discussions en cours avec le comité international de taxonomie des virus (ICTV, <https://talk.ictvonline.org/>) pour qu'ils proposent une visualisation de leur taxonomie des virus (référence mondiale en la matière, très à la mode en ce moment !)
- J'ai été contacté récemment par une société pseudo-pharmaceutique anglaise (<https://www.sofi.health/>) qui s'intéresse à visualiser les plantes thérapeutiques qu'ils vendent sur la taxonomie présenté dans Lifemap. À suivre

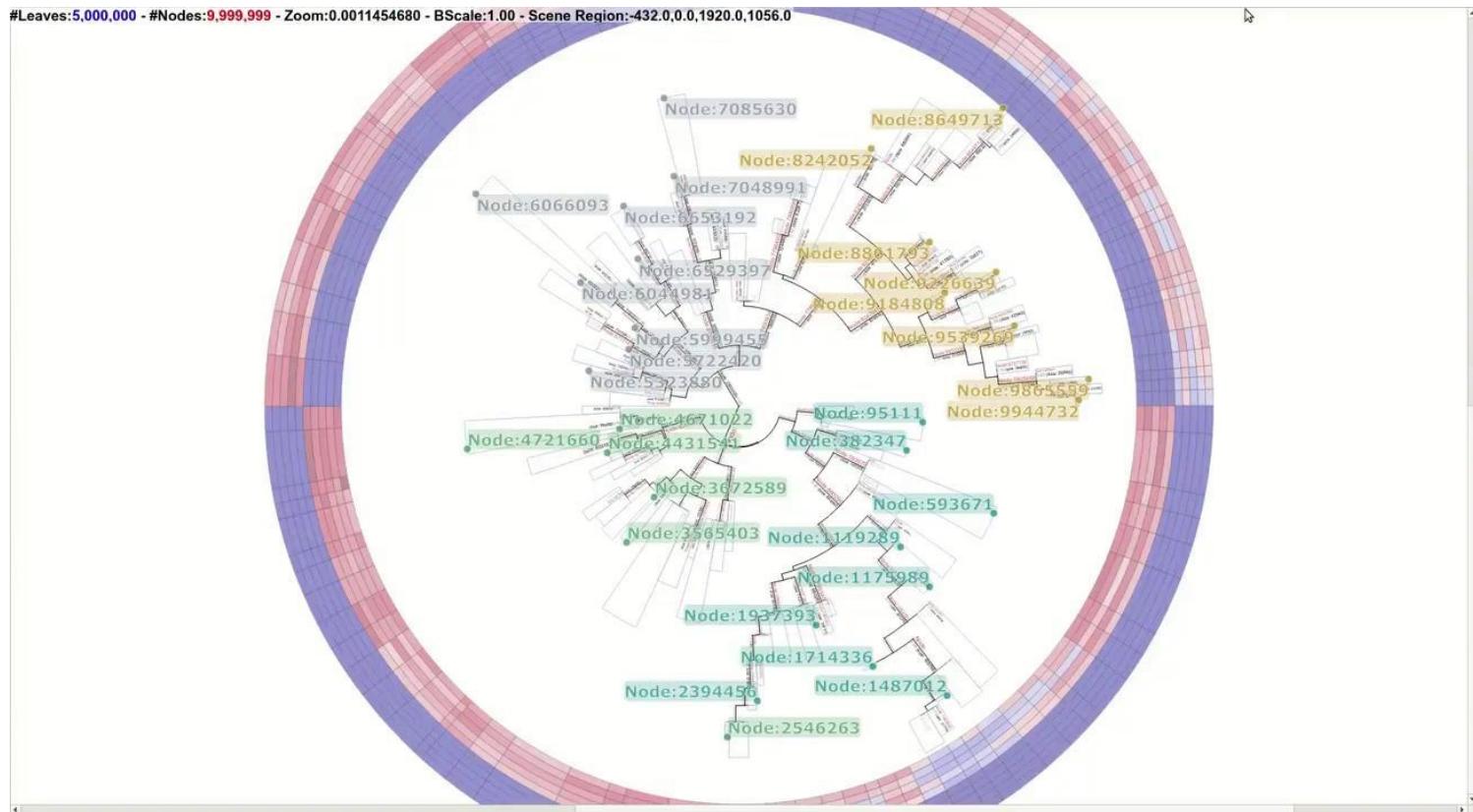


Limites de Lifemap et projets en cours

- **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. [Passage à des tuiles vectorielles \(en cours\)](#)
- **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 en cours

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



Limites de Lifemap et projets en cours

- **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. [Passage à des tuiles vectorielles \(en cours\)](#)
- **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 en cours

- **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 \(R, Shiny, leaflet\)](#)

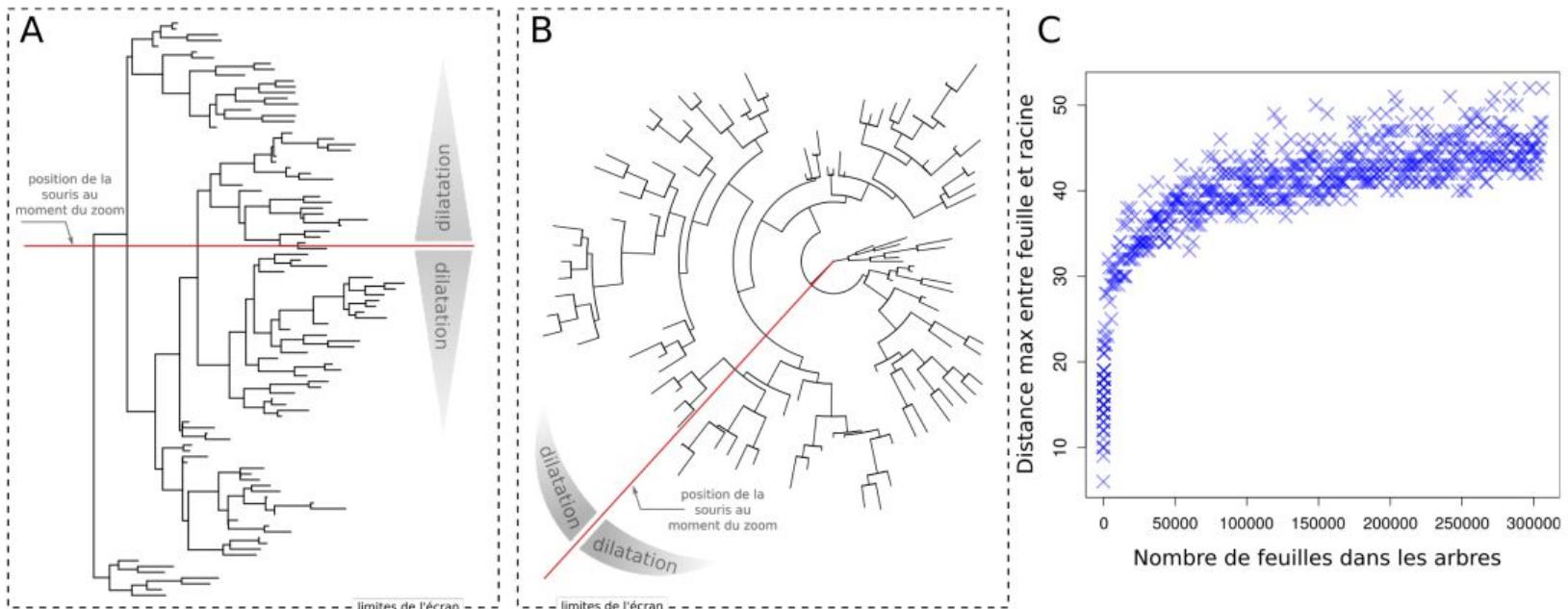


Figure 2. Proposition d'une nouvelle façon de visualiser les grands arbres. Que la représentation soit un phylogramme (A) ou une représentation radiale (B), l'idée est la même. Si un zoom est engagé avec la souris positionnée sur la ligne rouge, alors toutes les coordonnées des nœuds et des feuilles sont recalculées pour éloigner les points de la ligne rouge. L'éloignement sera d'autant plus fort que le point sera proche 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.

Limites de Lifemap et projets en cours

- **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. [Passage à des tuiles vectorielles \(en cours\)](#)
- **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 \[TP\]](#)

TP

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

...depuis



...en utilisant le package



<https://leafletjs.com/>

