



# Comparaison d'arbres phylogénétiques

Nicolas CASTEJON, Anis DJARI

Master IC spécialité Bioinformatique

Tuteurs: Anne-Muriel Arigon, Vincent Berry

2009

# Plan

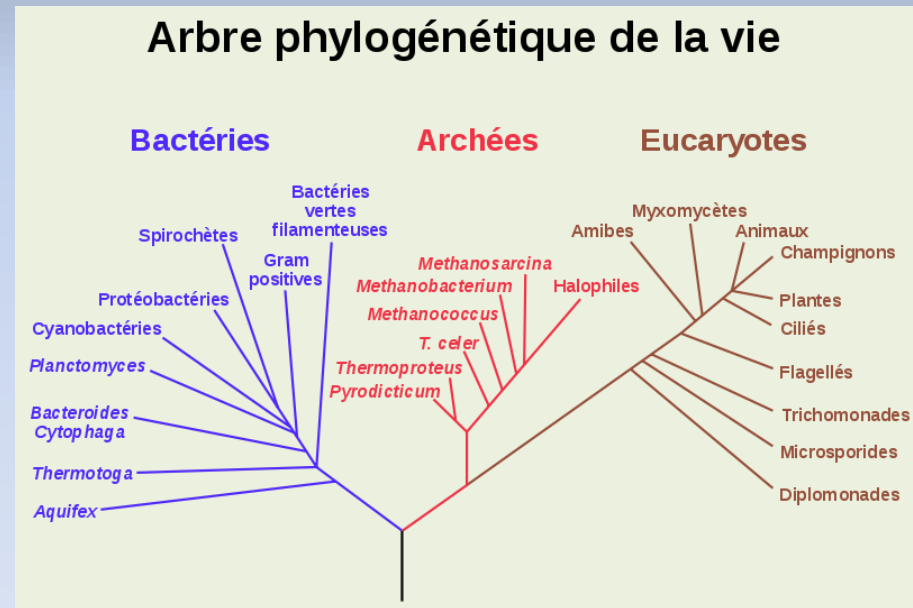
- Introduction
- Pré-analyse et cahier des charges
- Analyse
- Modélisation
- Conception
- Réalisation
- Conclusion et perspectives

# Introduction

La phylogénie est l'étude de la formation et l'évolution des organismes vivants en vue d'établir leur parenté.

Représentation courante :

Arbre phylogénétique basé sur les **données primaires** (séquences génétiques, caractères morpho-anatomiques)



A phylogenetic tree of living things, based on RNA.  
NASA Astrobiology Institute

# Introduction

- Construction des arbres selon 2 types de méthodes:



Méthodes des distances, par exemple :

- **UPGMA** (**U**nweight **P**air **G**roup **M**ethod with **A**rithmetic mean)
- **NJ** (**N**eighbor-**J**oining)



Méthodes basées sur le caractère :

- Maximum de **parcimonie**
- Maximum de **vraisemblance**

# Introduction

- PhySIC\_IST (Scornavacca C. et al. , 2008) une méthode de construction de super arbre en deux étapes:

➡ 1) Prétraitement par STC (Source Trees Correction)

➡ 2) Construction du super arbre par comparaison des arbres sources après le traitement STC

# Interface et fonctionnalités actuelles

## Page de soumission : PhySIC\_IST

### PhySIC\_IST: cleaning source trees to infer more informative supertrees.

Scornavacca C., Berry V., Lefort V., Douzery E.J.P., Ranwez V.  
BMC Bioinformatics. 2008, Oct 4;9:413.

Please cite [THIS](#) paper if you use PhySIC\_IST.

Server load: 96%



### PhySIC\_IST online execution

Source tree file

File  Example file

Backbone tree file (optional)

Outgroup file (optional)

File format

UNIX  Windows  Mac

Bootstrap threshold for source clade selection

Correction threshold used by STC

Your name

Your email

# Interface et fonctionnalités actuelles

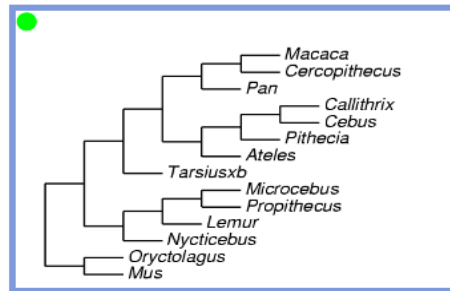
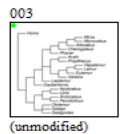
## Page de visualisation : CompPhy

This tool is on development stage, some bugs could remain. Comments much welcome.

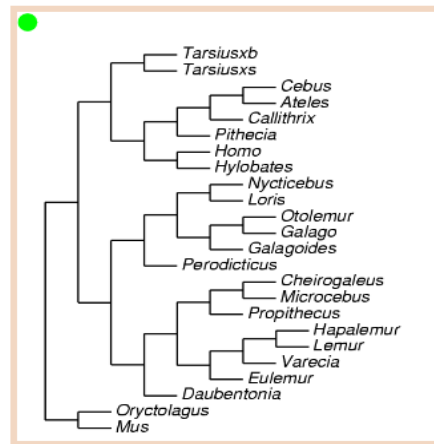
Display tree in:

- Top view  
 Bottom view

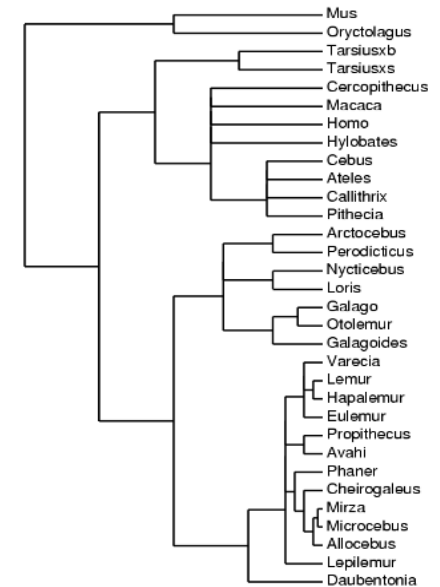
Click on a tree:



Zoom on source tree #001



Zoom on source tree #002



Zoom on supertree  
Correction threshold: 0.9

# Cahier des charges

Extension envisagée des fonctionnalités disponibles sur la page de PhySIC\_IST et CompPhy:

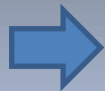
- Comparaison d'arbres 2 à 2
- Colorer les taxons sélectionnés
- Afficher les espèces en plus ou en moins
- Afficher les différences de soutien
- « Ladderizer/Swapper »
- Afficher le MAST
- Modifier Taille des arbres
- Soumettre un super arbre
- Construction de super arbre à partir d'une sélection d'arbres.
- Sessions utilisateurs
- Connexion partagée
- Système de presse-papier
- Système de Post-It
- Mode Expert

# Analyse des fonctionnalités

Fonctionnalités	Technologies	Difficulté de mise œuvre
Afficher espèces en plus ou en moins	Manipulation fichier newick	+
Afficher différences de soutien	ScripTree, fichier d'annotations	+
Mode Expert	Html , CSS, ScripTree	+
Modifier Taille des arbres à l'écran	ScripTree	+
Soumettre un super arbre	Php, ScripTree	+
Système de presse-papier	Manipulation de fichiers sur serveur de LIRMM	+
Comparaison d'arbre 2 à 2	Html , CSS, ScripTree	++
Colorer taxons sélectionnés	Mapping + ScripTree	++
Ladderizer/Swapper	Mapping + ScripTree	++
Afficher le MAST	ScripTree, PhySIC_IST	++
Système de Post-It	Manipulation fichier annotations	+++
Sessions utilisateurs	BDD (SQL, Php)	+++
Système de connexion partagée	BDD + système verrou	++++
Construction de super arbre à partir d'une sélection d'arbres.	Php, ScripTree, fichiers newick	++++

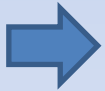
# Fonctionnalités à implémenter

- Articulées autour de 3 axes:



## Visualiser des arbres:

- Ladderizer/swapper
- Colorer taxons
- Soumettre un super arbre
- Modifier taille des arbres à l'écran
- Différence de soutien
- Mode Expert



## Comparer n arbres:

- MAST



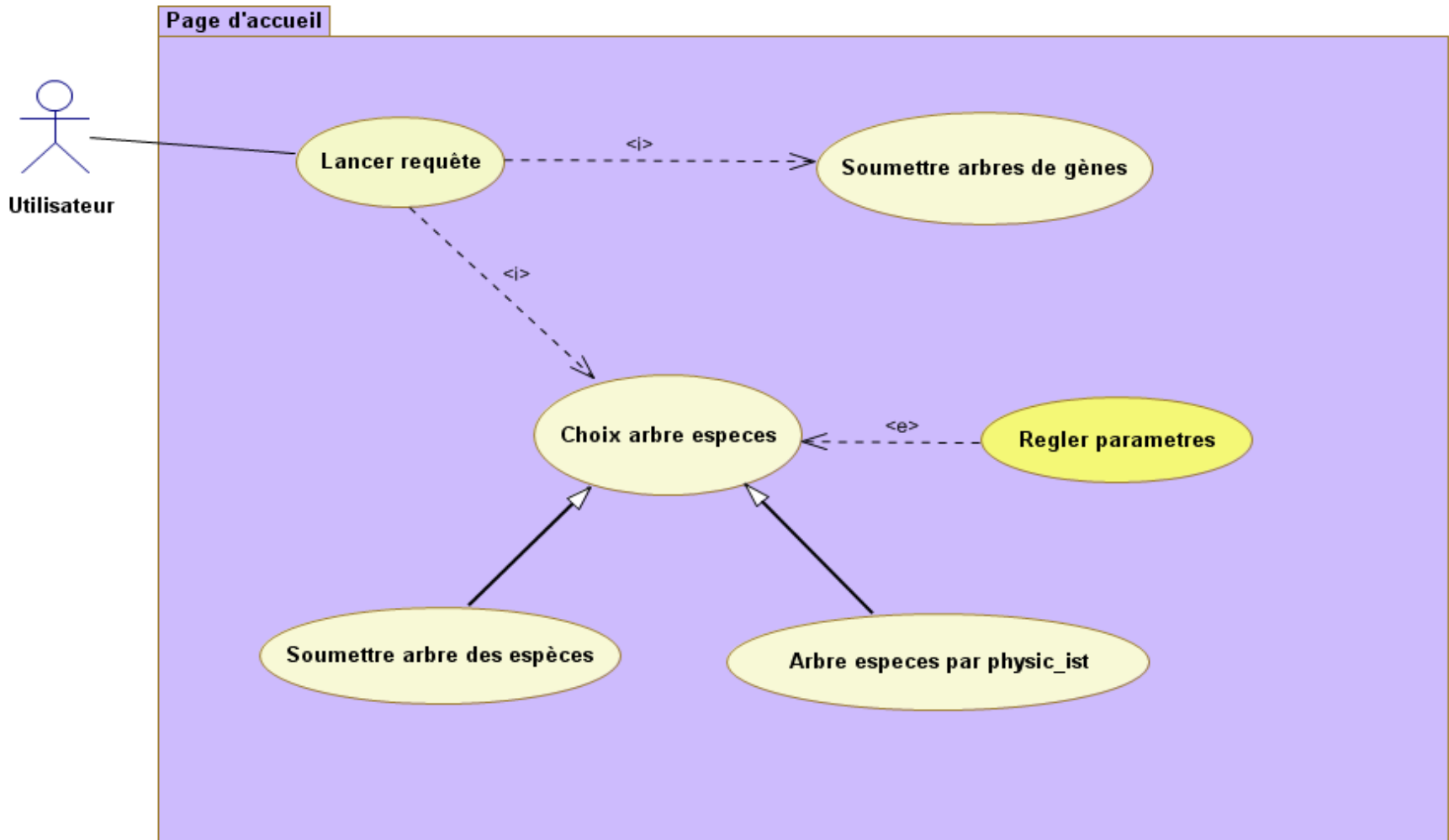
## Comparer 2 arbres:

- Espèces en plus ou en moins

# Modélisation

- Diagrammes de cas d'utilisation:
  - Page d'accueil
  - Page de visualisation
- Diagramme de classes

# Diagrammes de cas d'utilisation







# Conception

## La maquette

Page d'accueil (phySIC\_IST)

Page de visualisation (CompPhy)

# Réalisation

- ❑ Réalisation graphique (site basé sur la maquette)
- ❑ Pas d'implémentation des fonctionnalités (manque de temps, problème scripTree)

# Conclusion

## Bilan

- ❑ Cahier des charges à réaliser: sélection des fonctionnalités à implémenter
- ❑ Analyse, modélisation, conception  
-> travail en amont effectué pour l'implémentation

# Conclusion

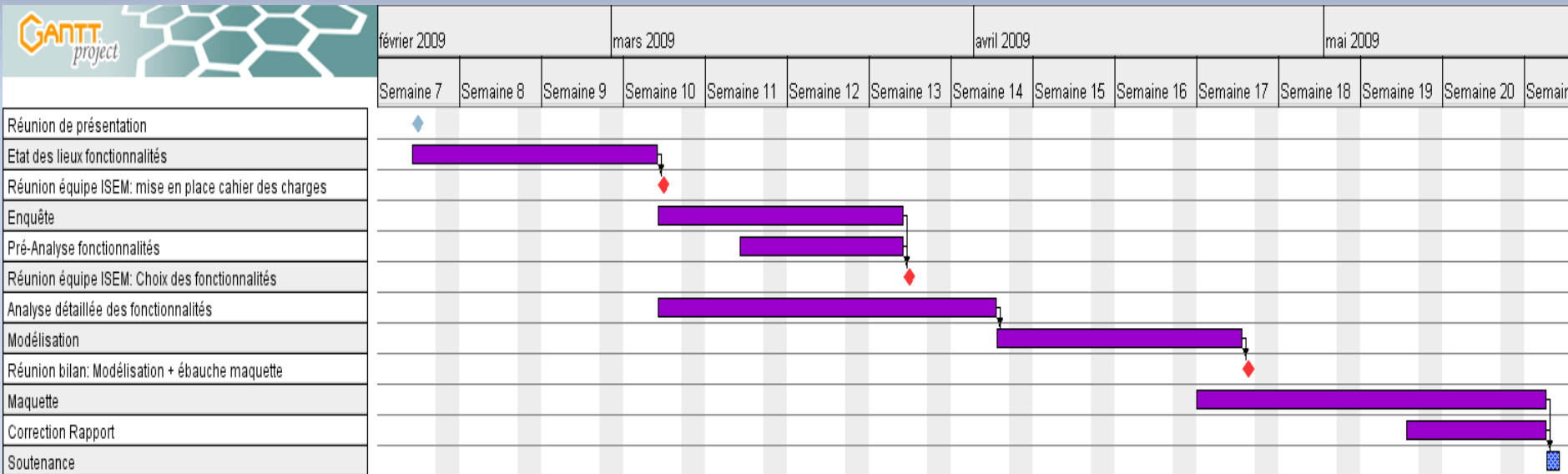
## Critiques

- ❑ Mauvaise organisation du temps

- > lacunes en modélisation

- > premier projet: manque d'expérience

# Planning réel du projet



# Perspectives

- Implémenter les différentes fonctions
- Presse-papier plus développé
- Création d'un système de comptes utilisateur
- Connexion partagée

**Merci de votre attention**

# Enquête ISEM sur les fonctionnalités

	Samuel Blanquart	Juan Escobar	Frédéric Delsuc
Logiciel utilisé pour comparaison	AWTY (bayésien)	CompPhy	CompPhy
Compte utilisateur	OUI	OUI	OUI
Connexion partagée	OUI	OUI	OUI
Coloration des taxons sélectionnés	OUI	OUI	OUI
Système de post-It	OUI	OUI	OUI
Autre fonctionnalité	<ul style="list-style-type: none"><li>• Recherche d'intrus</li><li>• Mise en avant de la différence de topologie</li></ul>	<ul style="list-style-type: none"><li>• Calcule de distance entre les arbres</li></ul>	