

PHYLOGÉNIE APPROFONDIE : MÉTHODES & APPLICATIONS EN ÉVOLUTION

Module Master M2 - HMBE 377 M

**Intervenants : Emmanuel J. P. DOUZERY
& Benoît NABHOLZ**

**Institut des Sciences de l'Évolution de Montpellier (ISE-M)
Université de Montpellier**

<https://moodle.umontpellier.fr/course/view.php?id=7843>

I. LA PHYLOGÉNIE : UNE QUÊTE D'INDICES ÉVOLUTIFS

II. LES MÉTHODES DE DISTANCE

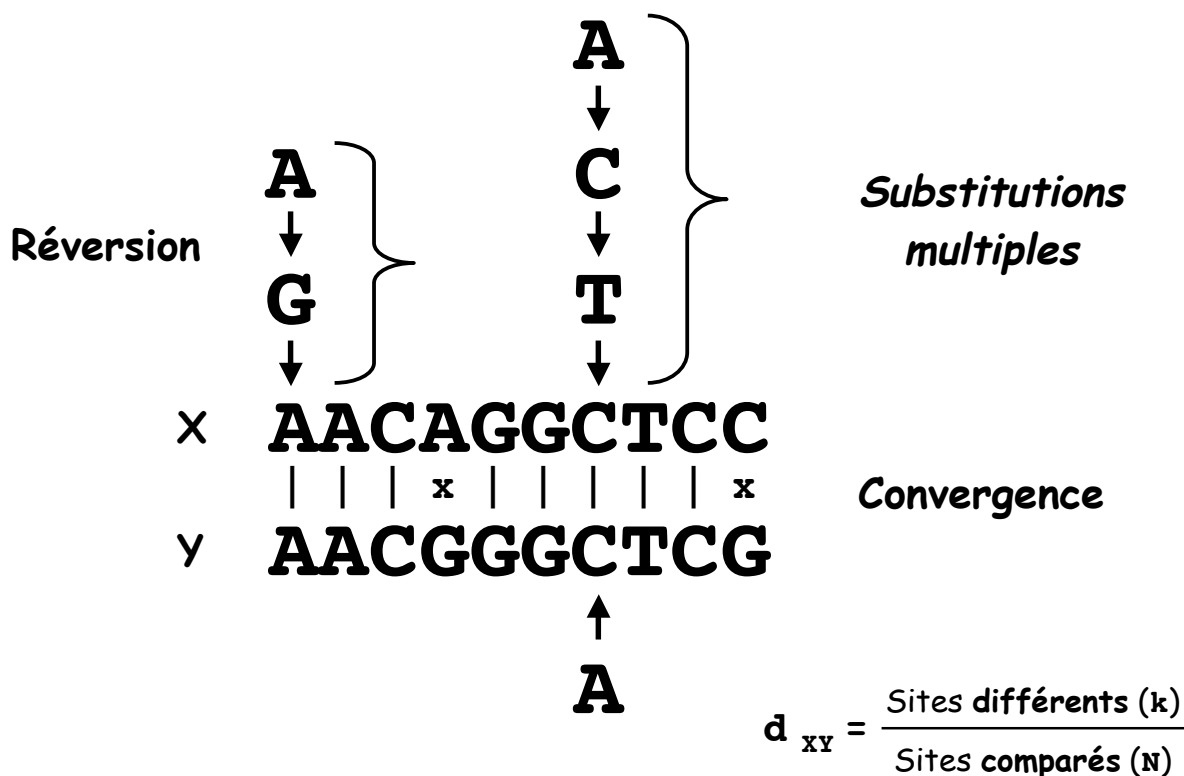
Comparaison par paires de séquences et calcul de distances.

Glaucomys	GTCACTGTGGGCCCCAGGGCTCCTGGCGGTTTCACCATCGGGACCCAAGAGGAACTCCATGGT
Marmota	-----T--T-----T--A-----
Petaurista	-----C--T--G-----A-----
Aplodontia	-----C--T--G-----A-----
Castor	A--A-C--T--A-----G-AC--T--C-A-----
Anomalurus	A-----T-----T--A-----T--CCT--T--G--GG
Idiurus	-----T-----T--A--A-----T--CCT--G--G--GG
Pedetes	A-----C--C--A-----A--G--C-----CCT--G--A--GG
Dipus	---G-C--A-T--C--TC-----G-AGC--T--C-----G--A--
Allactaga	-----C--A-T-----GC--T--G-A-----T--GC--T-----A--A--
Jaculus	-----C--C--A-T--G--TC-----AGC--T--GC--CGG-----G--A--
Sicista	A-----A-T-----T--G-A-----T--C-T-----T--A--
Dryomys	-----CACCA--TT-----T--G--CC--TT--GGT-----G--C--
Glis	-----CAC--T-----T--GAA-----T--G-----A--
Spalax	A-----T-----TT--T--GAA-----T--C-A-----G--
Tachyoryctes	-----CACCA--T-----T--T--G--C-----T--CC-----G--C--
Mus	-----C--A-T-----A--GCT--G-A-C--GT--C-----A--AC--G--
Rattus	A-----A--T-----TA--TCT--G-----T--C-A-----A--TC--
Ctenodactylus	-----C--A--TGG--A--T--A--A--AGC--T--C-A-C-----A--G--T--
Massoutiera	-----C--A--TGG--A--T--A--A--AGC--T--C-A-C-T-----G--T--
Trichys	A-----C--T-----T--A--AC--TT--TC-A-----T--G--A--G--A--
Atherurus	A-----Y--T-----T--A--A-----TT--CC-A-----T--A--G--A--
Cryptomys	-----CA--G-----A-----TGT--CC-----G--G--A--
Heterocephalus	-----T-----T--A-----TGT--CC-A-----A--GC--
Bathyergus	-----G-----T--A-----TGT--CCT-----G--G--
Heliophobius	-----C--G-----TC--AA-----TGT--CC-----G--G--
Thryonomys	-----CAA-----G-----TGT--TC--C-----G--T--
Petromus	-----A--CA--*****
Coendou	-----T--TCA--A--A--TGT--CC-A--A-----A--G--

$d_{A/B} = 4/60 = 6,6\%$

$d_{A/C} = 12/60 = 20,0\%$

SOUS-ESTIMATION DES DISTANCES RÉELLES EN RAISON DE L'HOMOPLASIE

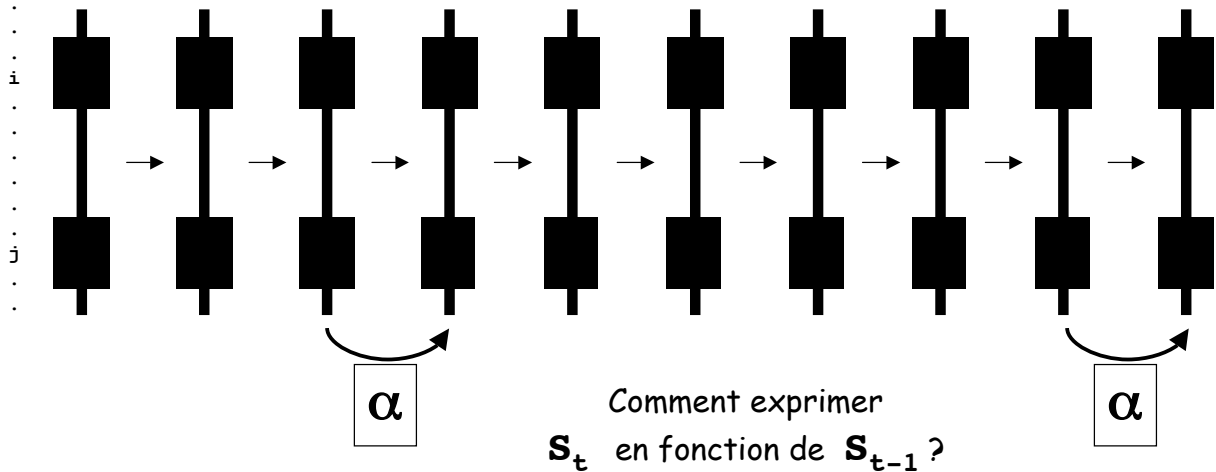


Comment gérer le problème des substitutions multiples ?

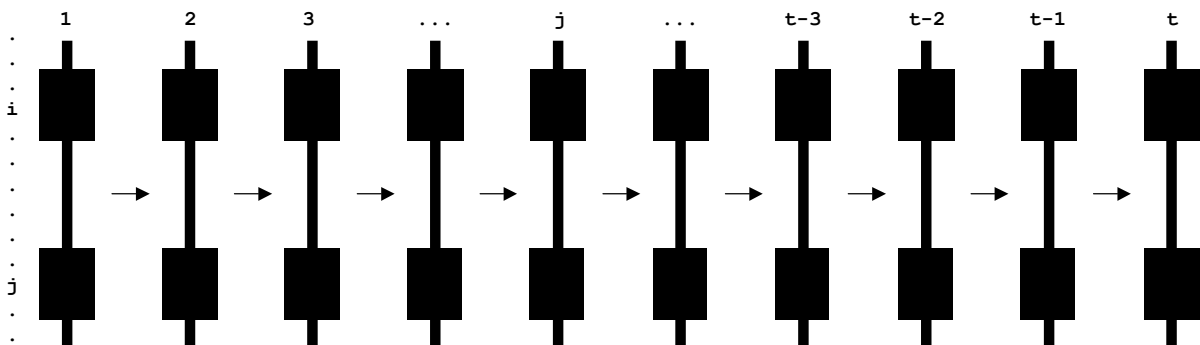
Soit une séquence d'ADN dans laquelle les 4 nucléotides sont équiprobables.

Soit un site donné i dans cette séquence,
occupé par un nucléotide A donné à la génération 0.

Quelle est la probabilité S_t que ce même site i
soit occupé par le même nucléotide A à la génération t ?



CONSTRUCTION D'UN MODÈLE SIMPLE (JC69)



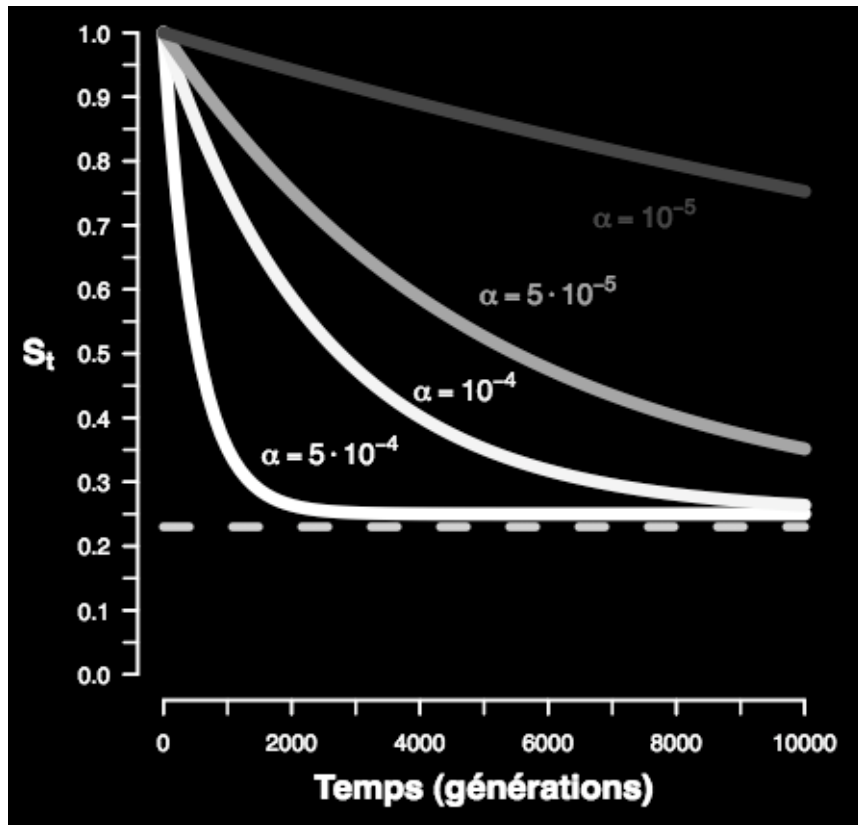
$$S_t = (1 - 3\alpha) \times S_{t-1} + \alpha \times (1 - S_{t-1})$$

Signal :
HOMOLOGIE

Bruit de fond :
HOMOPLASIE

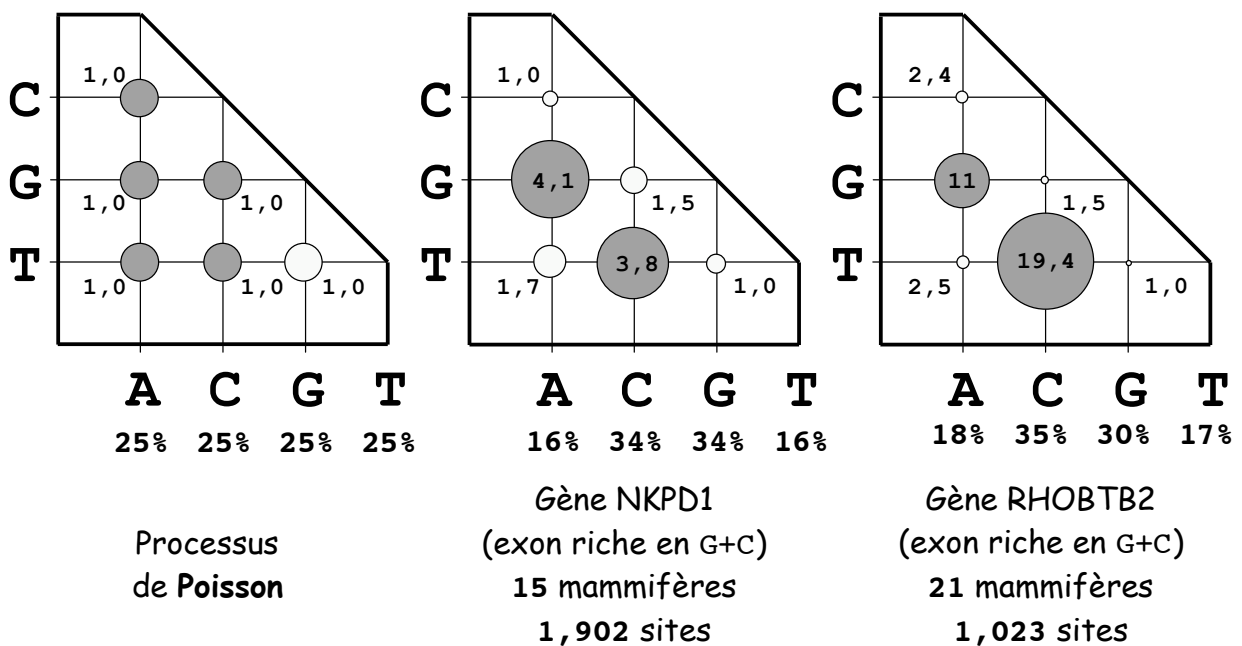
$$S_t = (1 - 4\alpha) \times S_{t-1} + \alpha \quad \Rightarrow \quad S_t = e^{-4\alpha t} \times 3/4 + 1/4$$

CONSTRUCTION D'UN MODÈLE SIMPLE (JC69)



*Pouvons-nous donner le même poids
à tous les changements d'états de caractères ?*

1) ADN & ARN : le taux de changement
des transitions est plus rapide que celui des transversions.



[Source = Base de données OrthoMaM v06 : <http://www.orthomam.univ-montp2.fr>]

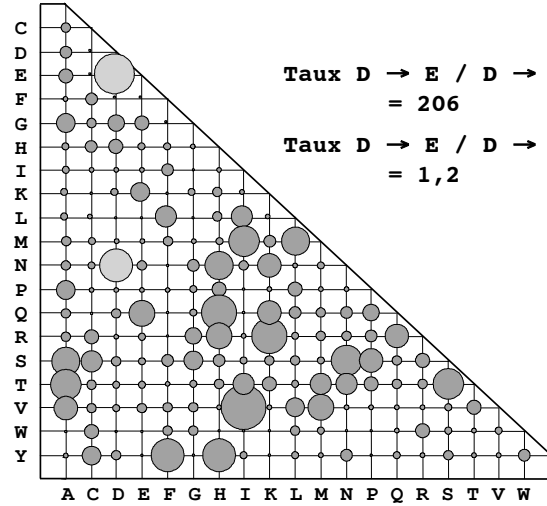
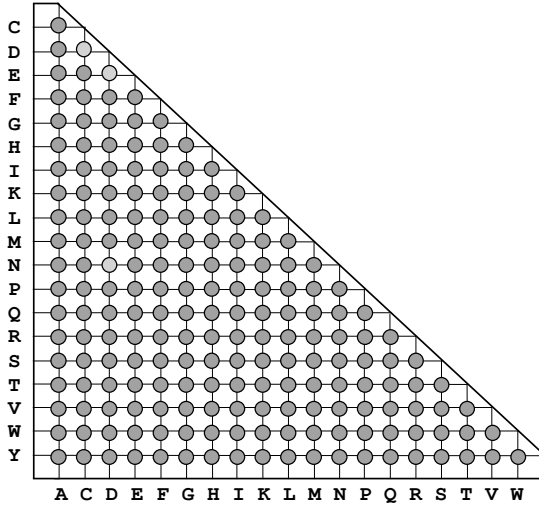
Pouvons-nous donner le même poids
à tous les changements d'états de caractères ?

2) Protéines :
Matrices d'échangeabilité (ex. : JTT ou WAG).

D = Acide aspartique

E = Acide glutamique

N = Asparagine



[Jones, Taylor & Thornton 1992 ; Whelan & Goldman 2001]

Tous les sites sont-ils libres de varier ?

Alignement pour les 3 domaines du monde vivant
de la protéine ribosomale RPS5.

	1	2	3	4	5	6	7	8	9	0	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	
E	P	R	E	D	S	C	R	I	G	K	G	G	Q	V	R	R	Q	S	V	D	V	S	P	L	R	R
U	A	R	E	D	S	T	R	I	G	V	G	G	T	V	R	R	Q	A	C	D	V	S	P	L	R	R
K.	P	R	E	D	S	T	R	I	G	S	A	G	V	V	R	R	Q	A	V	D	V	S	P	L	R	R
	P	R	E	D	S	T	R	I	G	R	A	G	T	V	R	R	Q	A	V	D	V	S	P	L	R	R
A	P	R	E	E	T	T	T	I	E	Y	G	G	A	R	Y	P	K	A	V	D	C	S	P	Q	R	R
R	P	R	E	D	T	T	K	I	T	Y	G	G	I	T	Y	R	V	S	V	D	V	A	P	Q	R	R
C.	P	R	E	E	T	T	R	I	K	Y	G	G	I	G	Y	Q	V	A	V	D	I	S	P	Q	R	R
	P	R	E	E	V	T	R	L	K	Y	G	G	I	A	V	P	K	S	V	D	V	S	P	S	R	R
B	P	E	W	E	V	R	P	T	R	V	G	G	A	T	Y	Q	V	P	I	E	V	P	E	R	R	Q
A	P	R	V	E	V	R	S	R	R	V	G	G	S	T	Y	Q	V	P	V	E	V	G	P	R	R	Q
C.	P	V	L	E	V	K	A	R	R	V	G	G	A	N	Y	Q	V	P	V	E	V	R	P	E	R	R
	P	H	V	E	V	R	S	R	R	V	G	G	A	T	Y	Q	V	P	V	D	V	R	P	E	R	R

Certains sites sont hautement conservés,
tandis que d'autres sont éminemment variables.

[Douzery 2011]

Comment convertir une matrice de distances en arbre ?

	<u>Bsu</u>	<u>Bst</u>	<u>Lvi</u>	<u>Amo</u>	<u>Mlu</u>
<i>Bacillus subtilis</i> (Bsu)	0	17 %	21	31	23
<i>Bacillus stearothermophilus</i> (Bst)		0	30	34	21
<i>Lactobacillus viridescens</i> (Lvi)			0	28	39
<i>Acholeplasma modicum</i> (Amo)				0	43
<i>Micrococcus luteus</i> (Mlu)					0

1. Utilisation de l'horloge moléculaire : UPGMA / WPGMA

<https://en.wikipedia.org/wiki/WPGMA>

2. Utilisation du critère des moindres-carrés : FITCH.

3. Utilisation d'une méthode par agglomération : NJ / BioNJ.

https://en.wikipedia.org/wiki/Neighbor_joining

RAPIDITÉ DES MÉTHODES DE DISTANCE : EXEMPLE D'APPLICATION



Utiliser les séquences d'ADN variables
comme **codes-barres** pour l'identification des organismes.

➔ Bactérie X : ARNr 16S

GGACGAAGCG TTTTGTGACG GTACCTGGAG AAGAAGCACC GGCTAACTAC GTGCCAGCAG

<i>Corynebacterium canis</i>	G G A G G G A C C G G G C A A C C C C G G C A A C A G C G C C G G G
<i>Flexivirga endophyti</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Rudaeiccoccus suwonens</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium dese</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium minu</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium pela</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium frei</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium phoc</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Cutibacterium granul</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium pseu</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Acidipropionibacteri</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium timo</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium glut</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Pseudonocardia antit</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium rena</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Acidipropionibacteri</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium pilb</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium kefi</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Flexivirga alba-v-TT</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium tube</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium effi</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium pseu</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium camp</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium sing</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium afer</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Pseudonocardia endop</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium mass</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Luteipulveratus mong</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Acidipropionibacteri</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium conf</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium muc</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Pseudonocardia kongj</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium capi</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium tapi</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium afer</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium phoc</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium thum</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>QRY Unknown</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium four</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium acet</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Gordonia araii-v-TT</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium tusc</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium urei</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium prep</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium pseu</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium lubr</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Pseudonocardia soli-</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium auri</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium acco</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium epid</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G
<i>Corynebacterium auri</i>	G G A G G G A C C G G G A G A G A C A C C G G C A A C A G C G C C G G G

Requête ➔

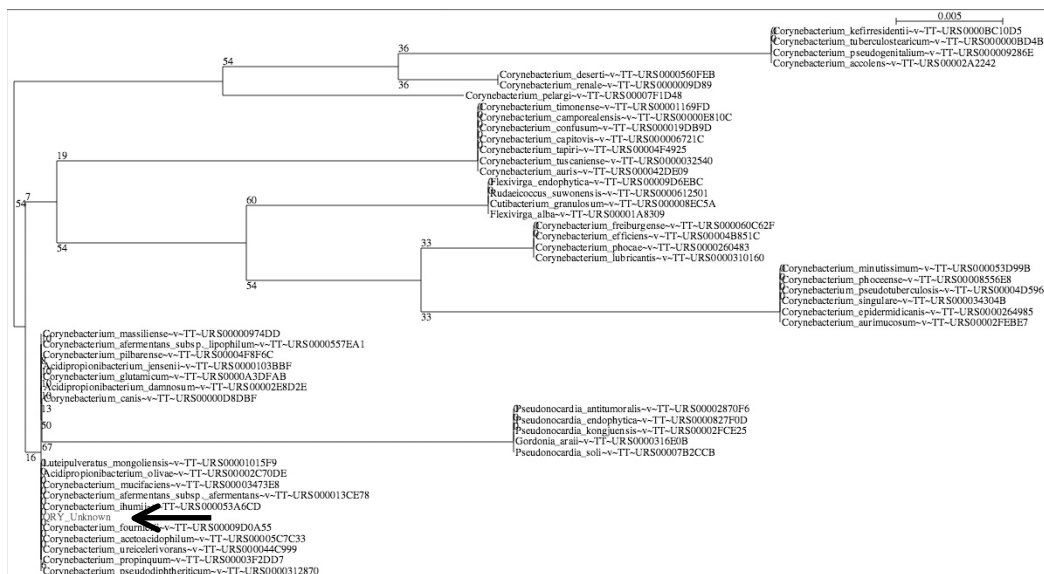
Pipeline :

BLAST
 + Alignement
 + Nettoyage des sites
 + Reconstruction
 de la phylogénie
 par méthode de distance.

<http://umr5558-sud-str1.univ-lyon1.fr/lebibi/lebibi.cgi>

➔ Bactérie X : ARNr 16S

GGACGAAGCG TTTTGTGACG GTACCTGGAG AAGAAGCACC GGCTAACTAC GTGCCAGCAG



<http://umr5558-sud-str1.univ-lyon1.fr/lebibi/lebibi.cgi>

