Mercredi 15 septembre 2004

Étude des systèmes de recombinaison chez les bactéries par génomique comparative

Stage de maîtrise — Emmanuel Cornet

Dirigé par Eduardo Rocha

Été 2004

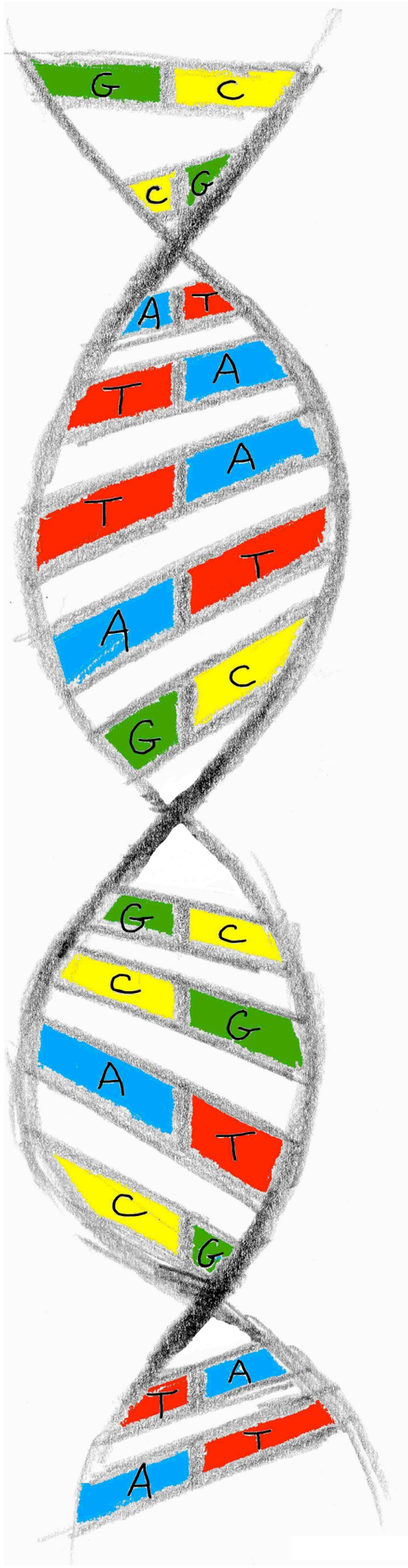
Plan

- Théorie
  - Biologie
  - Algorithme d'alignement de deux séquences
- Recherches
  - Démarche
  - Résultats : quelques commentaires
  - Utilité et perspectives

Théorie

L'ADN

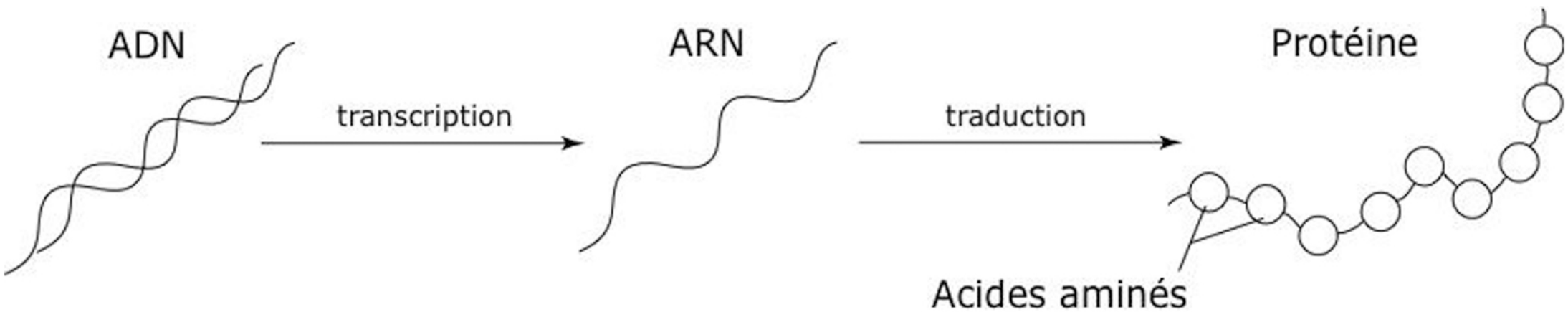
(Acide désoxyribonucléique)



- Structure en double hélice
- Bases azotées A, T, G, C
- Complémentarité A/T et G/C
- Exemple de séquence :

GCTAATCGCACAT

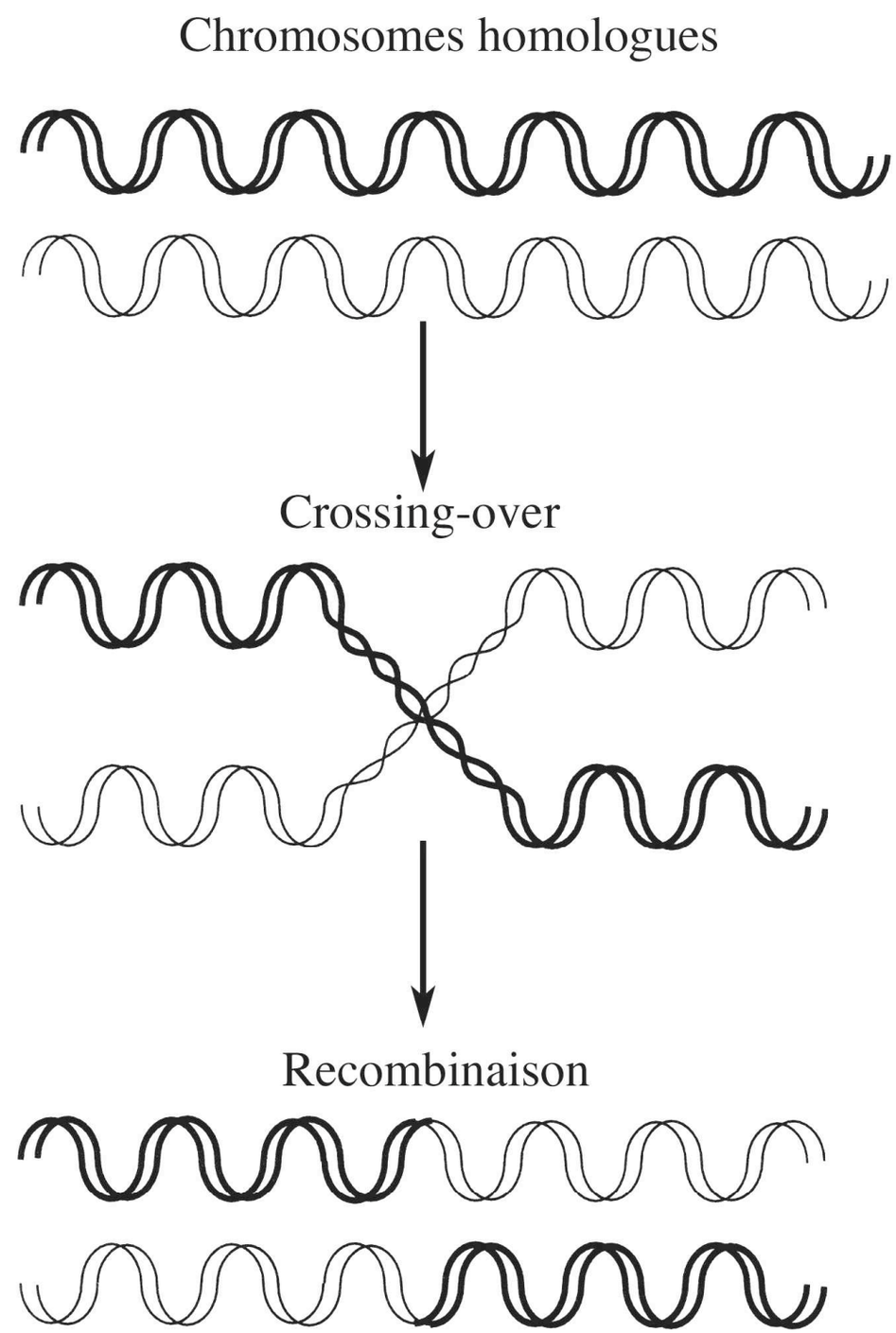
De l'ADN à la protéine



GATCGATTGATCGGC → GAUCGAUUGAUCGGC → YRLIC

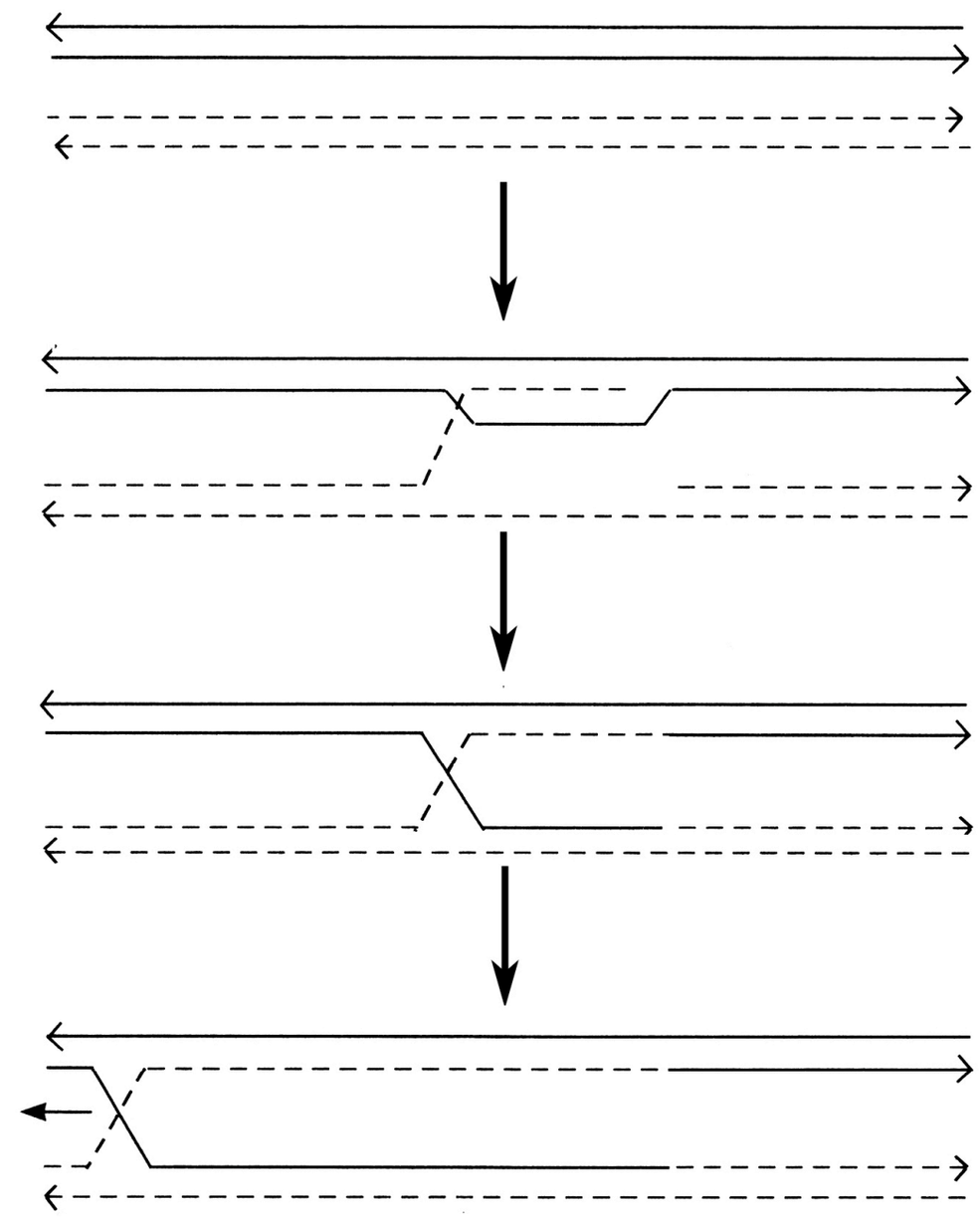
La recombinaison homologue

- Brassage génétique
- Réparation de l'ADN
- Expression de certains gènes
- Transfert horizontal



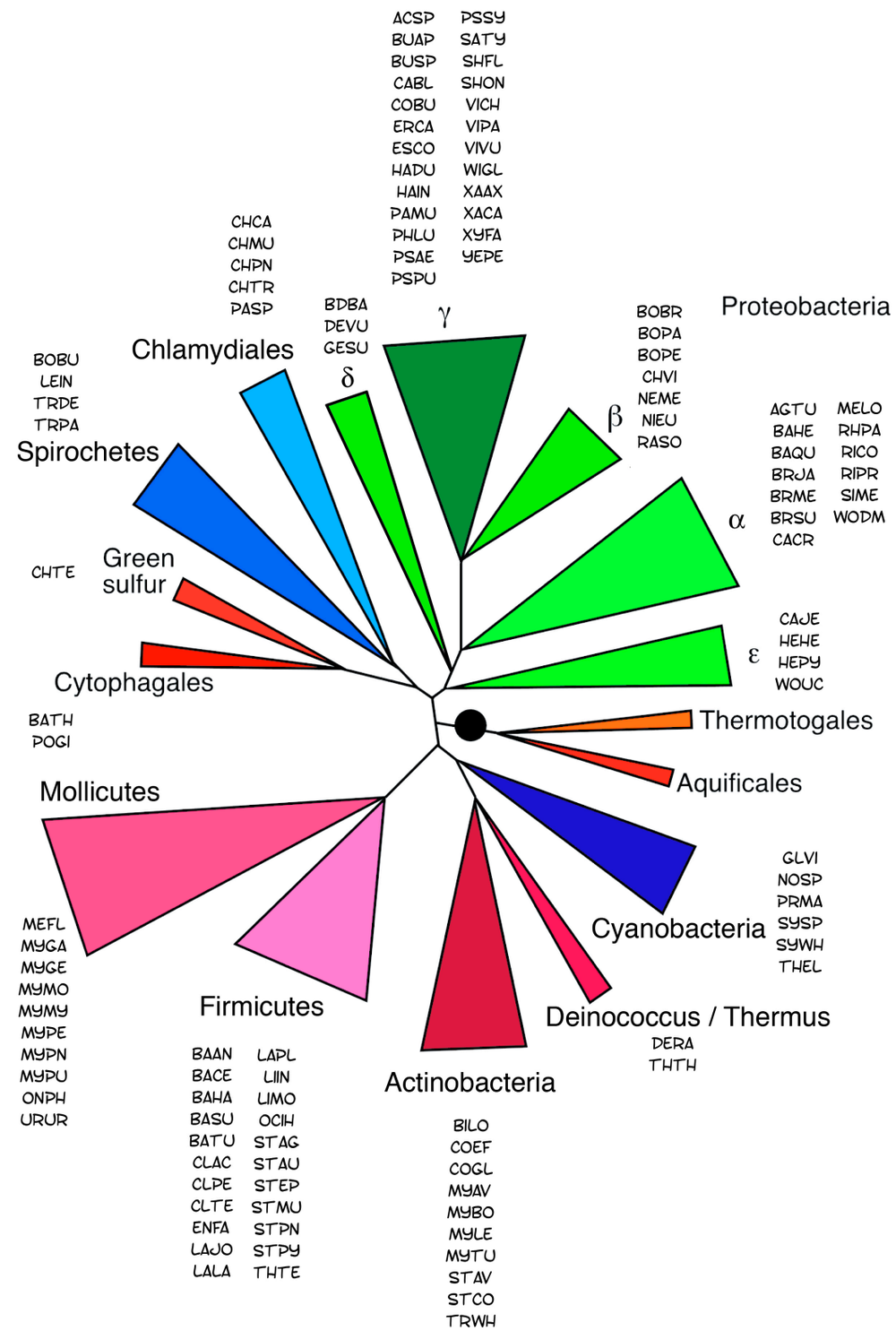
Quelques protéines impliquées

- Hélicase et nucléase : RecBCD
- Hybridation : RecA
- Déplacement de jonction : RuvAB
- Résolution : RuvC, RecG, RusA
- ...



Objet du stage

- Extension des connaissances sur *E. coli*
- Contribution à la cartographie d'autres génomes (120 génomes ~ 10^6 pb, 25 gènes ~ 10^3 pb)
- Étude des vitesses d'évolution et de la répartition des gènes de la recombinaison (confrontation avec les connaissances en biologie moléculaire)



Alignement de séquences

(acides aminés)

```
RGPSSTWA          -RGPSST--WA
        →
ERGSTLKVA         ERG--STLKVA
```

Une matrice de *scoring*

s(R,P)

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	−1	−1	−2	0	−1	−1	0	−2	−1	−1	−1	−1	−2	−1	1	0	−3	−2	0
R	−1	5	0	−1	−3	1	0	−2	0	−3	−2	2	−1	−3	$\boxed{-2}$	−1	−1	−3	−2	−2
N	−1	0	6	1	−2	0	0	0	1	−3	−3	0	−2	−3	−2	1	0	−4	−2	−3
D	−2	−1	1	6	−3	0	2	−1	−1	−3	−3	−1	−3	−3	−1	0	−1	−4	−3	−3
C	0	−3	−2	−3	9	−3	−3	−2	−3	−1	−1	−3	−1	−2	−3	−1	−1	−2	−2	−1
Q	−1	1	0	0	−3	5	2	−2	1	−3	−2	1	0	−3	−1	0	−1	−2	−1	−2
E	−1	0	0	2	−3	2	5	−2	0	−3	−3	1	−2	−3	−1	0	−1	−3	−2	−2
G	0	−2	0	−1	−2	−2	−2	6	−2	−3	−4	−1	−2	−3	−2	0	−2	−2	−3	−3
H	−2	0	1	−1	−3	1	0	−2	7	−3	−3	−1	−1	−1	−2	−1	−2	−2	2	−3
I	−1	−3	−3	−3	−1	−3	−3	−3	−3	4	2	−3	1	0	−3	−2	−1	−2	−1	3
L	−1	−2	−3	−3	−1	−2	−3	−4	−3	2	4	−2	2	0	−3	−2	−1	−2	−1	1
K	−1	2	0	−1	−3	1	1	−1	−1	−3	−2	4	−1	−3	−1	0	−1	−3	−2	−2
M	−1	−1	−2	−3	−1	0	−2	−2	−1	1	2	−1	5	0	−2	−1	−1	−1	−1	1
F	−2	−3	−3	−3	−2	−3	−3	−3	−1	0	0	−3	0	6	−4	−2	−2	1	3	−1
P	−1	−2	−2	−1	−3	−1	−1	−2	−2	−3	−3	−1	−2	−4	7	−1	−1	−4	−3	−2
S	1	−1	1	0	−1	0	0	0	−1	−2	−2	0	−1	−2	−1	4	1	−3	−2	−2
T	0	−1	0	−1	−1	−1	−1	−2	−2	−1	−1	−1	−1	−2	−1	1	4	−2	−2	0
W	−3	−3	−4	−4	−2	−2	−3	−2	−2	−2	−2	−3	−1	1	−4	−3	−2	10	2	−3
Y	−2	−2	−2	−3	−2	−1	−2	−3	2	−1	−1	−2	−1	3	−3	−2	−2	2	6	−1
V	0	−2	−3	−3	−1	−2	−2	−3	−3	3	1	−2	1	−1	−2	−2	0	−3	−1	4

+ *gaps*

Algorithme de Needleman-Wunsch

- Alignement de deux séquences
- Programmation dynamique
- Donne l'alignement optimal (pas forcément unique)

Algorithme de Needleman-Wunsch

$i-1$

x ...ESDELKD-EAIS

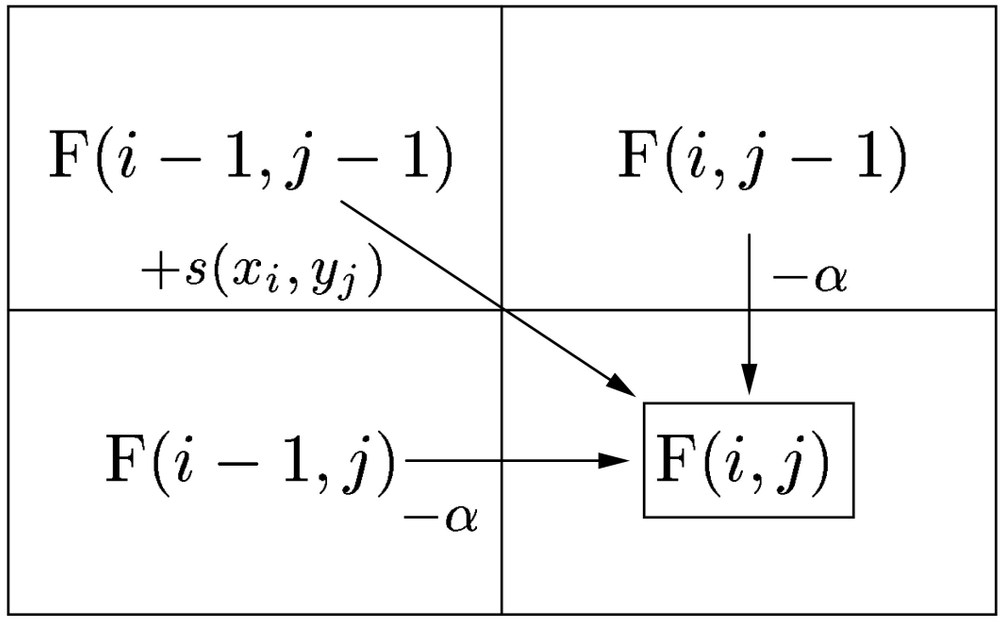
y ...ES---FDNQVIK

$j-1$

- ...EKDELKD-EAIST
  ...ES---FDNQVIKS

- ...EKDELKD-EAIST
  ...ES---FDNQVIK-

- ...EKDELKD-EAIS-
  ...ES---FDNQVIKS



Algorithme de Needleman-Wunsch

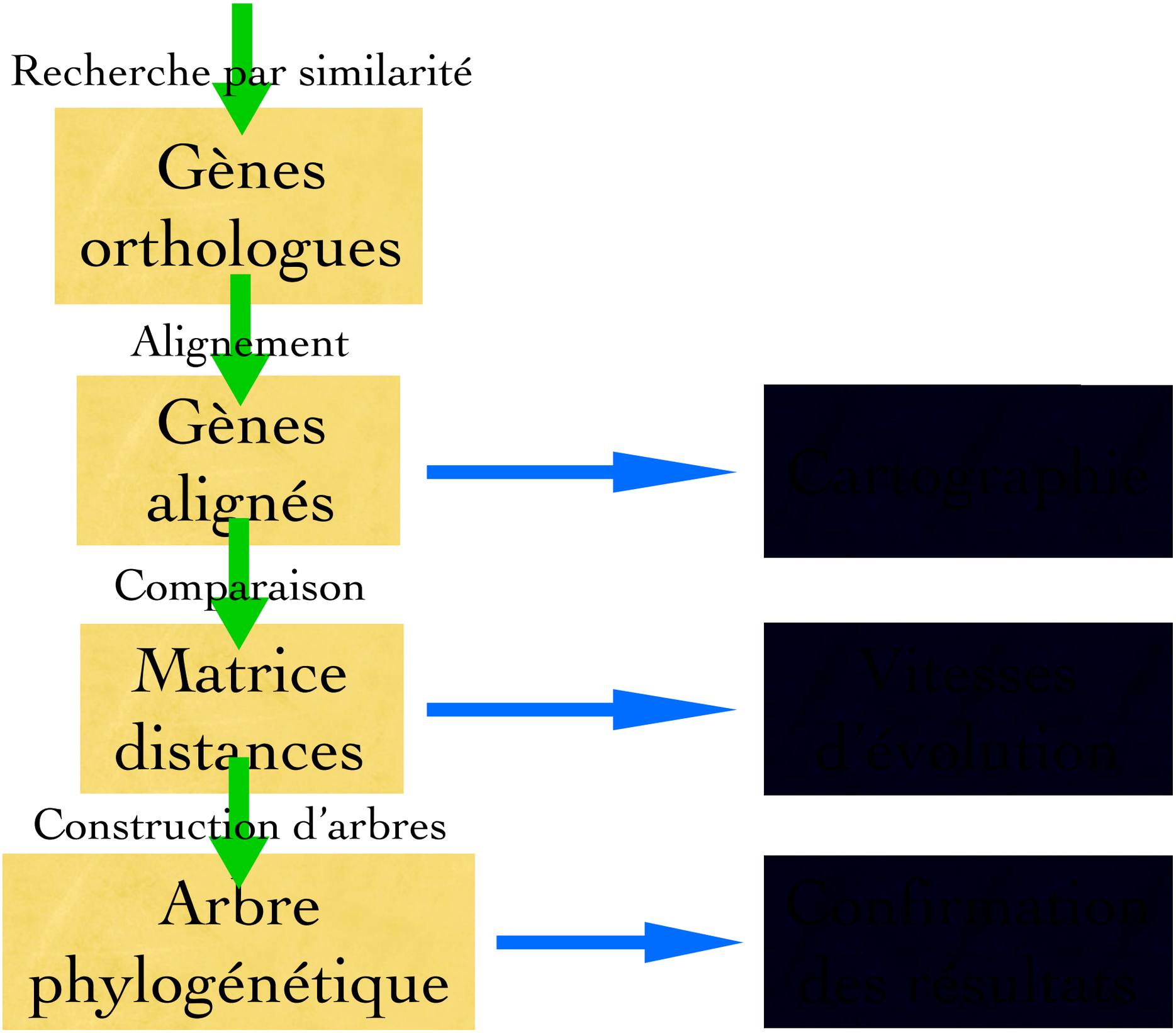
		H	E	A	G	A	W	G	H	E	E
	0 ←	**−8** ←	**−16** ←	−24 ←	−32 ←	−40 ←	−48 ←	−56 ←	−64 ←	−72 ←	−80
	↑ ↖	↖	↖		↖				↖	↖	
P	−8	−2	−9	**−17** ←	**−25**	−33 ←	−42 ←	−49 ←	−57	−65	−73
	↑	↑ ↖	↖		↖						
A	−16	−10	−3	−4 ←	−12	**−20** ←	−28 ←	−36 ←	−44 ←	−52 ←	−60
	↑	↑	↑ ↖	↖	↖	↖					
W	−24	−18	−11	−6	−7	−15	**−5** ←	**−13** ←	−21 ←	−29 ←	−37
	↑ ↖	↖	↖	↖	↖		↑ ↖	↖			
H	−32	−14	−18	−13	−8	−9	−13	−7	**−3** ←	−11 ←	−19
	↑	↑ ↖			↑ ↖	↖		↑ ↖	↖	↖	
E	−40	−22	−8 ←	−16	−16	−9	−12	−15	−7	**3**	−5
	↑	↑	↑ ↖			↖	↖	↖	↑	↑ ↖	
A	−48	−30	−16	−3 ←	−11	−11	−12	−12	−15	**−5**	2
	↑	↑	↑	↑ ↖	↖	↖	↖	↖	↖	↖	
E	−56	−38	−24	−11	−6	−12	−14	−15	−12	−9	**1**

```
HEAGAWGHE-E
--P-AW-HEAE
```

Complexité : $O(n^2)$

Recherches

Démarche


Recherche par similarité
Gènes orthologues
Alignement
Gènes alignés
Cartographie
Comparaison
Matrice distances
Vitesses d'évolution
Construction d'arbres
Arbre phylogénétique
Confirmation des résultats

Résultats : commentaires

- Quelques gènes essentiels (*recA*, *recR*, *ruvAB*)
- Quelques gènes « surestimés » (*recBCD*)
- Cohérence des complexes (présence et vitesse d'évolution)
- Gènes « rares » (*recE*, *recT*, *rusA*)
- Arbres phylogénétiques cohérents
- Les gènes sont d'autant plus conservés qu'ils sont répandus (tendance globale)
- Tests annexes (*Evolver*) : cette dernière observation n'est pas due à une trop grande vitesse d'évolution

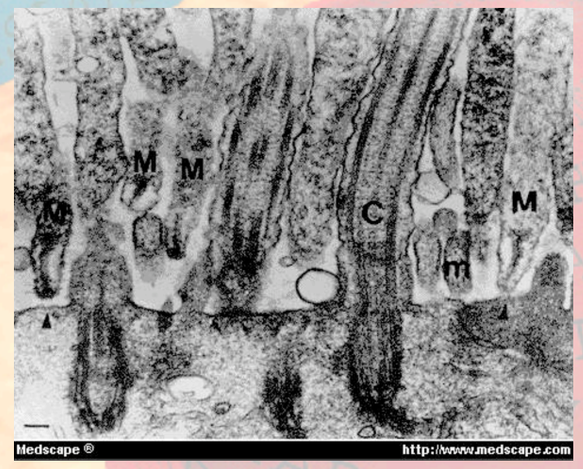
Utilité et perspectives

- Cartographie
- Étude comparative des mécanismes (biologie moléculaire)
- Meilleure connaissance du mécanisme de recombinaison en général
- Histoire des gènes de la recombinaison

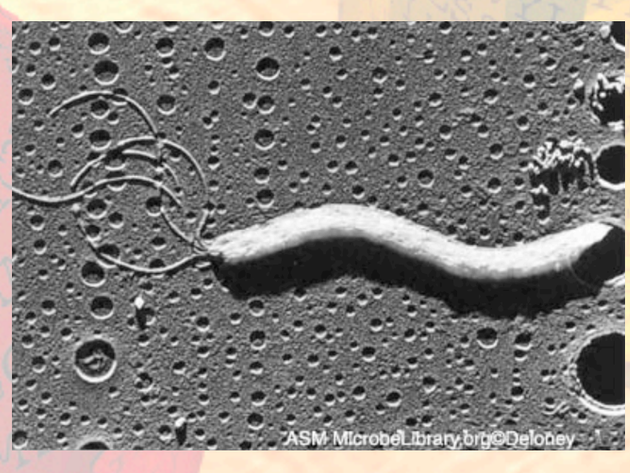
Cette présentation, ainsi que le rapport de stage, sont disponibles en ligne (formats PDF et PostScript) à l'adresse

`http://www.eleves.ens.fr/home/cornet/Stage`

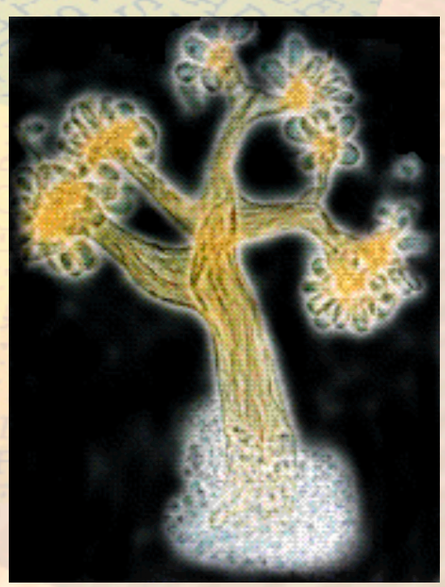
E. coli



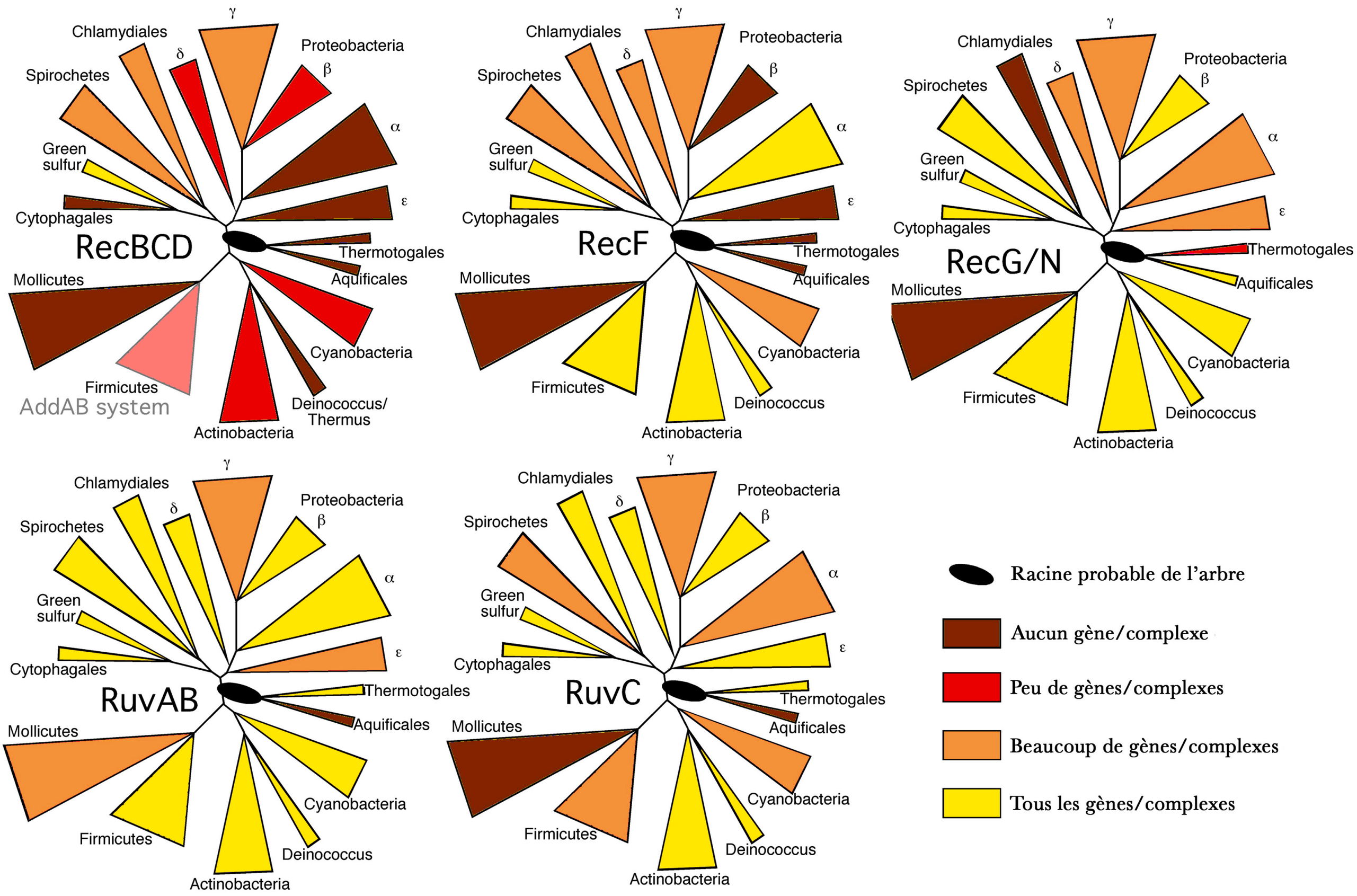
Mycoplasma



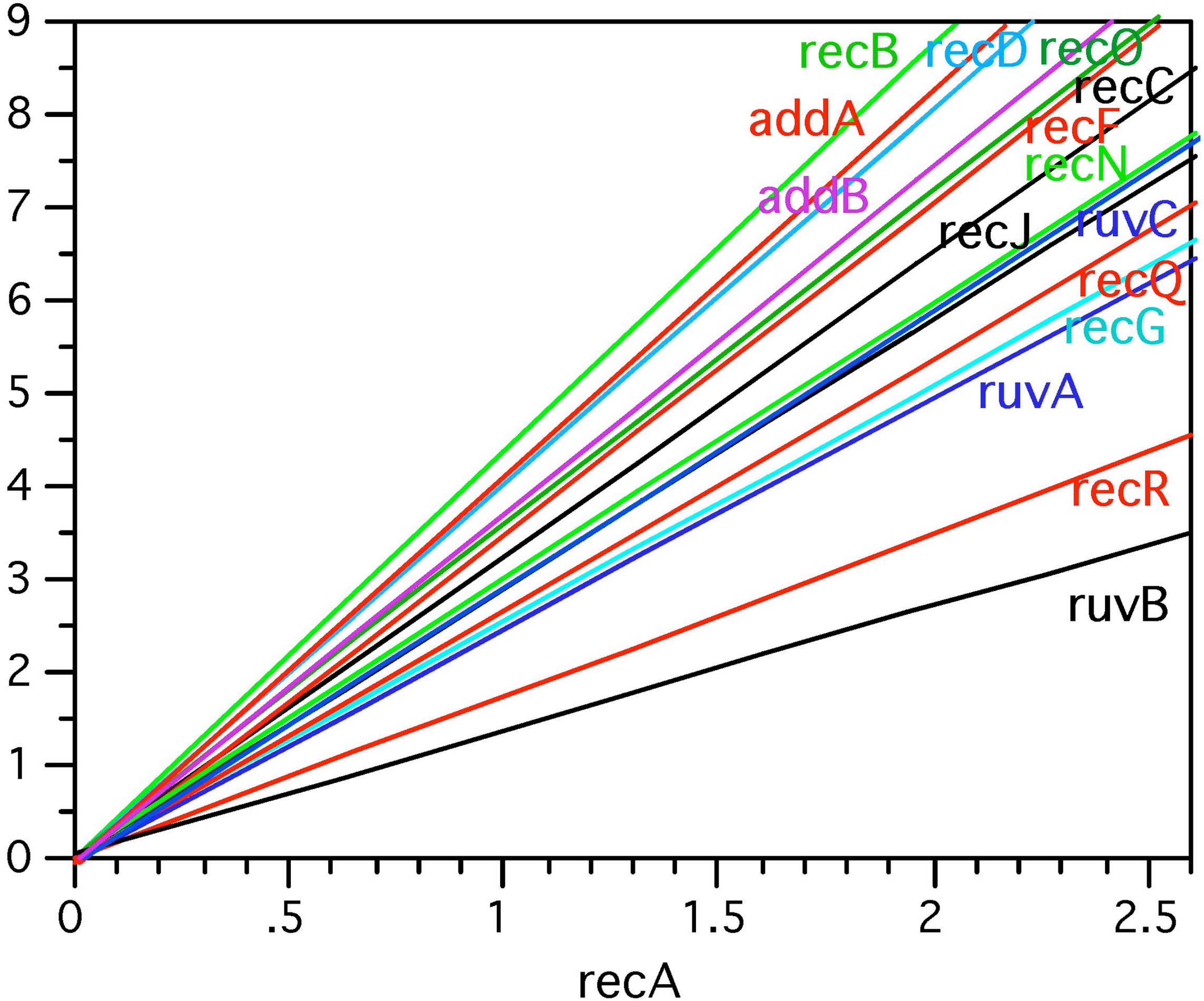
Helicobacter



Myxobacteria


RecBCD
RecF
RecG/N
RuvAB
RuvC
γ
δ
β
α
ε
Chlamydiales
Proteobacteria
Spirochetes
Green sulfur
Cytophagales
Thermotogales
Aquificales
Mollicutes
Cyanobacteria
Firmicutes
Deinococcus/ Thermus
Deinococcus
Actinobacteria
AddAB system
Racine probable de l'arbre
Aucun gène/complexe
Peu de gènes/complexes
Beaucoup de gènes/complexes
Tous les gènes/complexes

BILDA01 01
TRWHB01 00
TRWHA01 00
STCDA01 05
STAVA01 02
MYBOA01 02
MYAYA01 02
COGLA01 01
COIPA01 01
DIEVUA01 01
CHTEA01 01
GLYIA01 03
SYSPA01 03
THILA01 02
NOSPA01 03
PRMAB01 01
PRMAA01 01
SYWHA01 02
PRMAC01 01
POGIA01 00
BATHA01 04
CAJIA01 01
WDUCA01 01
HEBEA01 00
HEPYB01 00
HEPYA01 00
WODMA01 00
RIPRA01 00
RICOA01 01
CACRA01 01
BHPAA01 03
BRJAA01 05
BACUA01 00
BAHEA01 01
SIMEA01 01
BRSUA01 01
BRMEA01 00
MELDA01 00
AGTUB01 01
AGTUA01 01
CHVIA01 01
NEMEB01 01
NEMEA01 01
NIEUA01 01
BASCA01 00
BOPEA01 02
BOPAA01 02
BOBRA01 02
COBUA01 01
XYPAB01 00
XYPAA01 00
XACAA01 01
XAAXA01 01
ACSPA01 01
PSSYA01 03
PSPUA01 01
PSAEA01 05
HADUA01 00
PAMUA01 01
HAINA01 00
WIGLA01 00
VIPAA01 02
VICHA01 00
VIVUB01 02
VIVUA01 01
SHONA01 03
PHLUA01 01
ERCAA01 03
YEPEC01 00
YEPEB01 00
YEPEA01 03
SATYC01 02
SATYB01 02
SATYA01 02
ESCOA01 02
ESCOC01 03
SHFLA01 02
SHFLB01 02
ESCOD01 03
ESCOB01 03
GESUA01 00
BEBAA01 00
THTHA01 01
DERAA01 02
PISPA01 06
AQAEA01 01
PASPA01 02
CHTBA01 00
CHMUA01 00
CHCAA01 00
CHPND01 00
CHPNF01 00
CHPNB01 00
CHPNA01 01
LHINA01 02
BOBUA01 00
TRPAA01 00
TRDEA01 00
THTEA01 01
CLTEA01 01
CLPEA01 01
CLACA01 01
THMAA01 01
FUNUA01 01
MYMYA01 00
MEPLA01 00
MYPUA01 00
MYMOA01 00
MYPLA01 00
MYGAA01 00
URURA01 00
MYPNA01 00
MYGLA01 00
LALAA01 00
STPYD01 01
STPYB01 01
STPYA01 01
STPYC01 01
STPNB01 01
STPNA01 01
STMRUA01 01
STAGB01 02
STAGA01 02
STEPA01 00
STAUB01 01
STAUB01 01
STAUC01 01
STAUB01 01
STAUA01 01
OCBIA01 01
BASUA01 01
BAHAA01 02
BACEA01 03
EMPAA01 03
LAPLA01 01
LAKIA01 00
LIMOA01 01
LIMOB01 01
LHINA01 01
0.1


recB
recD
recO
recC
addA
recF
recN
addB
recJ
ruvC
recQ
recG
ruvA
recR
ruvB
0
1
2
3
4
5
6
7
8
9
0
.5
1
1.5
2
2.5
recA