UNIVERSITÉ PARIS SUD

FACULTÉ DES SCIENCES D'ORSAY

Analyse de Séquences

M1 BIBS

http://rna.igmors.u-psud.fr/gautheret/cours/

V. 2012.2

Deux mots de Bioinformatique

Données

- Génomique
- Protéomique
- Données d'Interaction
- Cristallographie / RMN
- Données cliniques
- chimiothèques

Bioinformatique

- Analyse de séquences
- Imagerie
- Modélisation moléculaire
- Modélisation des systèmes
- Data mining

Applications

- Drug design
- Pharmaco-genomics
- Diagnostic médical
- Amélioration plantes
- Biologie fondamentale
- Biologie synthétique

Bio-informatique et bioinformatique

Biologiste + Informaticien =

Biologiste =

Informaticien =

Ne pas perdre de vue notre objet d'étude

« Bioinformatics is just a word to describe part of modern data-intensive molecular biology"

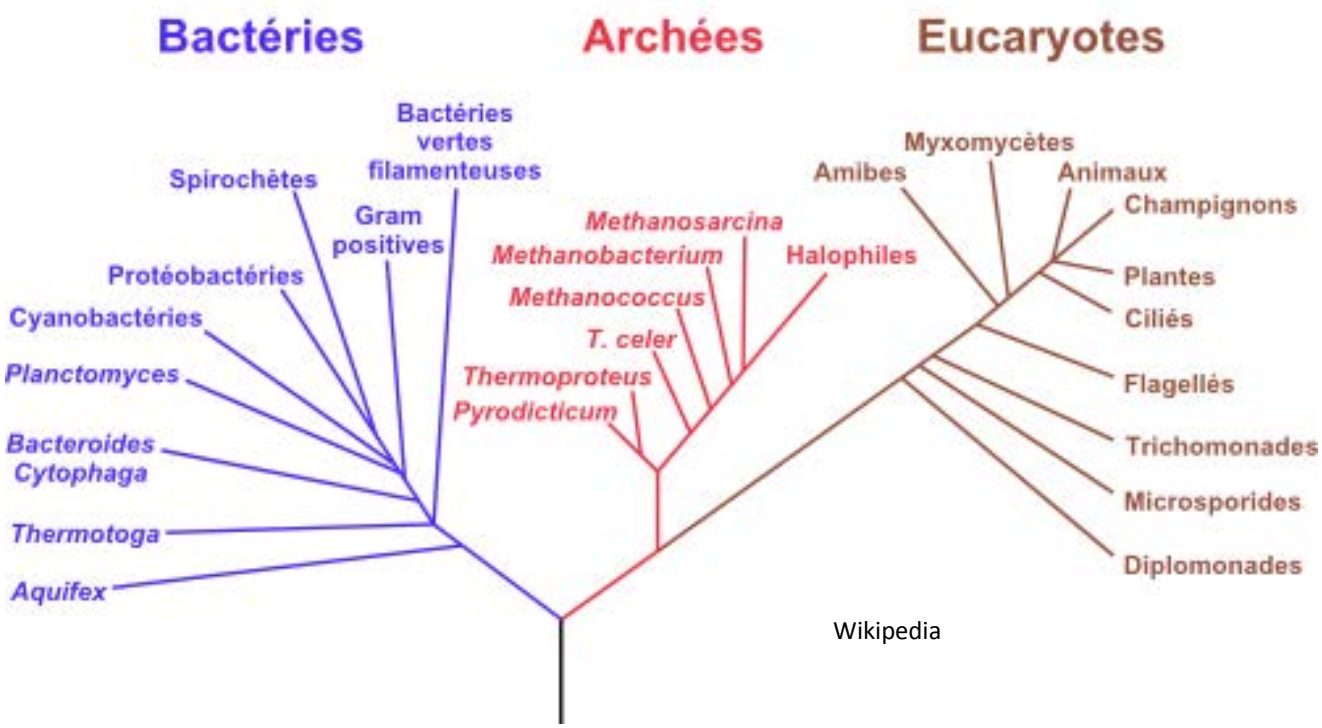
Ewan Birney, 2012

Contenu général de l'UE

- D.Gautheret
  - Les données de séquence: génomes modèles, séquençage, structure des génomes
- A. Lopes
  - Alignement de séquence, phylogénie moléculaire
- D.Gautheret – C. Pereira
  - Homologie, annotation
  - TD Annotathon

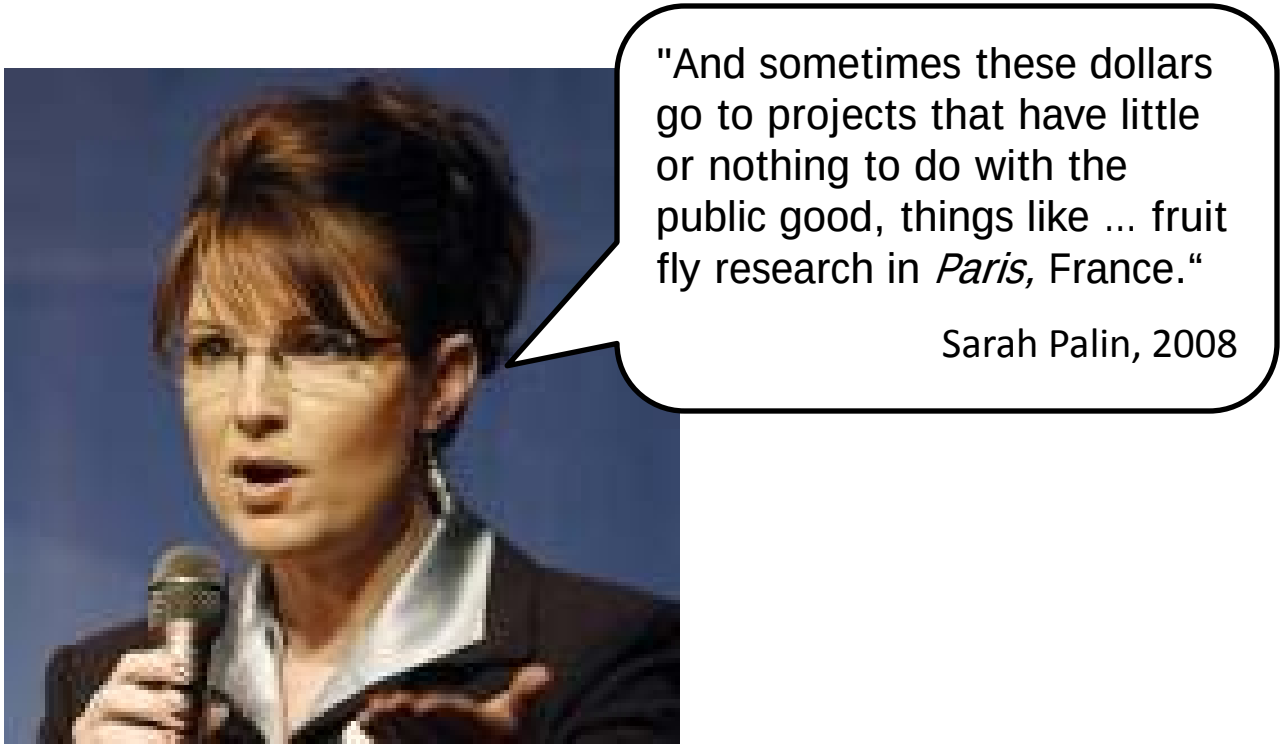
1: Les données

Rappel: l'arbre du vivant

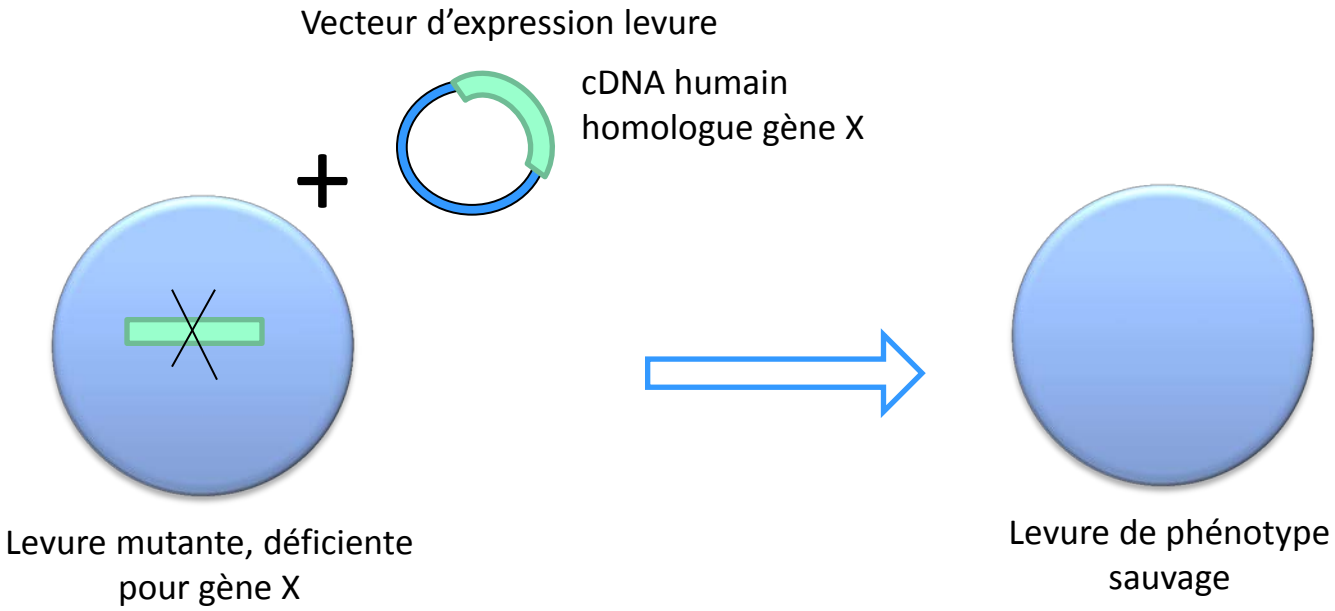


Wikipedia

LUCA


"And sometimes these dollars go to projects that have little or nothing to do with the public good, things like … fruit fly research in *Paris,* France.“
Sarah Palin, 2008

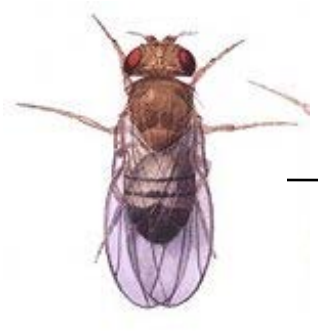
La puissance du raisonnement évolutif…



- Des dizaines de gènes humains sont capables de complémenter des mutations de levure

Pourquoi étudier la « mouche à fruits » ?

Banfi et al. *Nature Genetics*, 1996



Gènes causant des phénotypes mutants

Recherche de similitudes

Milliers de transcrits humains connus en 1996

```
ttgtagctactgat
cgatcgtagctagc
gggcatgcat

acatgcatgctagc
tgactagctagcta
gtcgatgc

gcgatgctagctag
tcgatcggactgat
cgatcgcagatcg

gcatgctagctagt
cgatgctagctgat
cgtagt
```

66 gènes candidats pour des maladies génétiques ou des fonctions essentielles chez l'homme

- **Toutes les connaissances structurales et fonctionnelles acquises chez la drosophile sur ces gènes deviennent applicables.**

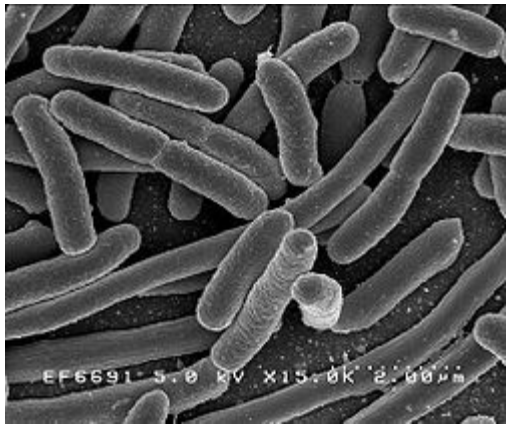
Et l'on s'est mis à séquencer des génomes-modèles!

Différentes raisons de choisir un modèle

- Facilité expérimentale
  - Grosses cellules, corps transparent, temps de génération court, taux de reproduction élevé
- Représentant d'une fonction
  - Différenciation cellulaire, Système immunitaire, Photosynthèse, symétrie bilatérale…
- Génome dense
- Données accumulées
  - Mutants, génétique, biochimie…

Escherichia coli

- Phylum
  - Bactéries gram -, protéobactéries
- Modèle pour:
  - La microbiologie
  - Réseaux d'interaction dans la cellule
  - Voies métaboliques
  - Bactéries pathogènes (certaines souches)
  - Formation de biofilms
- Avantages:
  - Culture aisée
  - Grande connaissance de la biologie de l'espèce



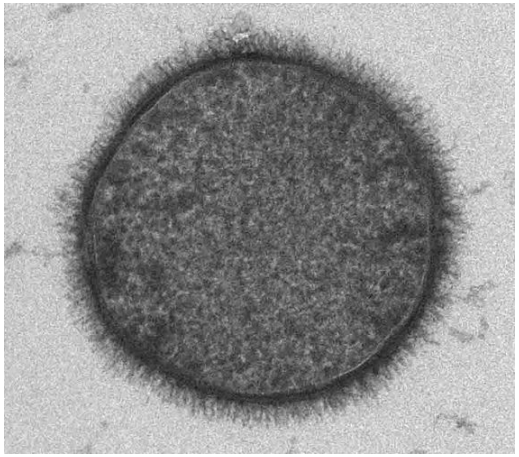
Wikipedia

4.5 Mb

4200 gènes

Bacillus subtilis

- Phylum
  - Bactéries gram +
- Modèle pour:
  - Sporulation (différenciation cellulaire élémentaire)
  - Flagelles
- Avantages:
  - Grande connaissance de la biologie de l'espèce



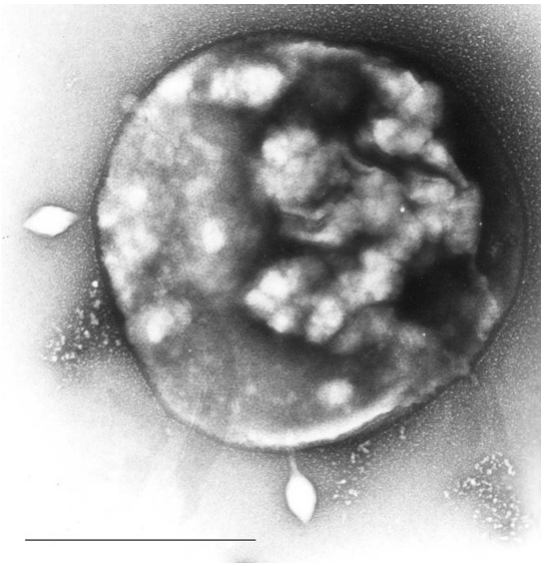
Wikipedia

4.2 Mb

4100 gènes

Sulfolobus acidocaldarius

- Phylum
  - Archaes, Chrenarchaeotes
- Modèle pour:
  - Biologie des Archées
  - Réplication de l'ADN
- Avantages:
  - Systèmes moléculaires proches des eucaryotes mais plus simples



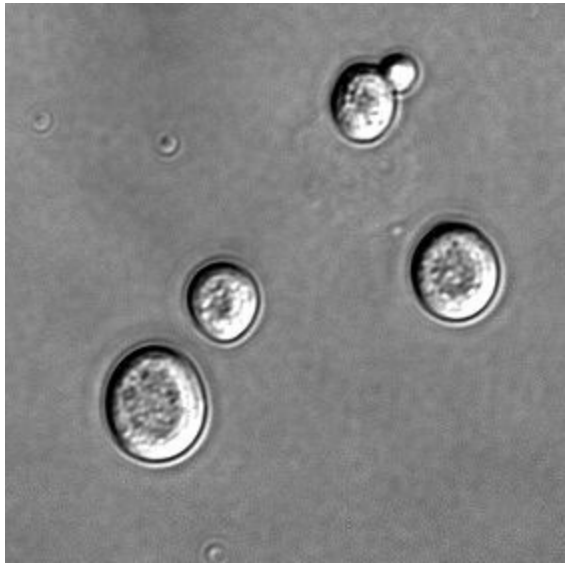
Wikipedia

2.2 Mb

2300 gènes

Saccharomyces cerevisiae

- Phylum
  - Eucaryotes, Fungi, Ascomycetes
- Modèle pour
  - Les eucaryotes
  - Le cycle cellulaire
  - Différentiation sexuelle primitive
  - Fermentation
- Avantages
  - Temps de génération court
  - Transformation facile pour introduire de nouveaux gènes
  - Culture haploide possible: KO

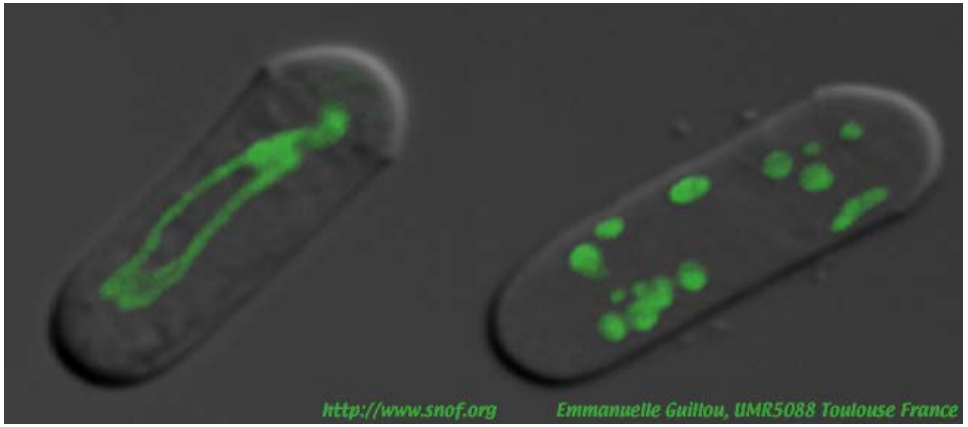


Wikipedia

13 Mb

6000 gènes

Schizosaccharomyces pombe



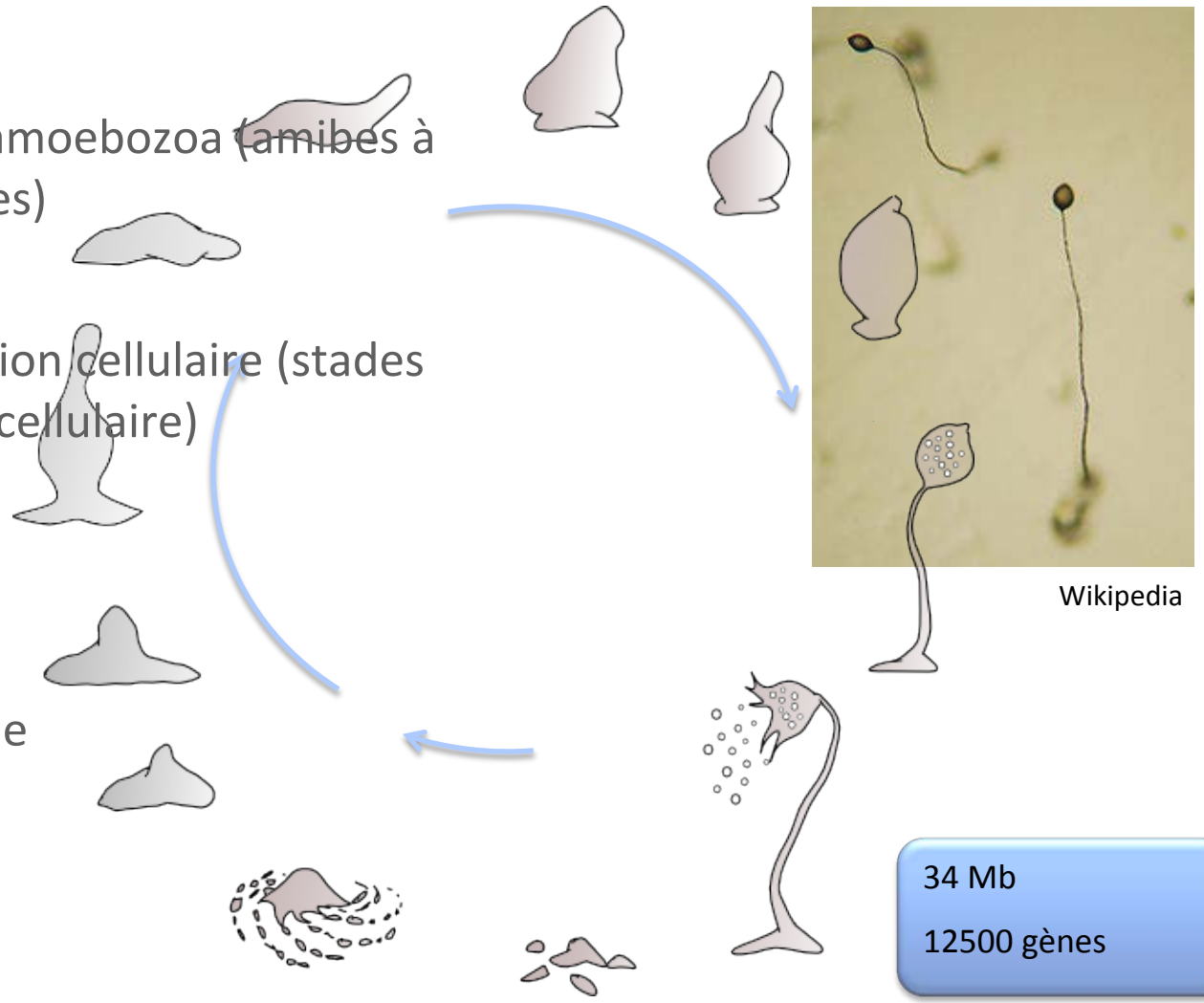
- Phylum
  - Eucaryotes, Fungi, Ascomycetes
- Modèle pour
  - Génétique, biologie cellulaire et moléculaire
  - Le cycle cellulaire, la chromatine
  - Certaines maladies humaines
- Avantages
  - Phase haploide dominante

14 Mb

4800 gènes

Dictyostelium discoidum

- Phylum
  - Eucaryote/amoebozoa (amibes à pseudopodes)
- Modèle pour
  - Différenciation cellulaire (stades uni- et pluricellulaire)
  - Chimiotaxie
  - Apoptose
- Avantages
  - Culture facile



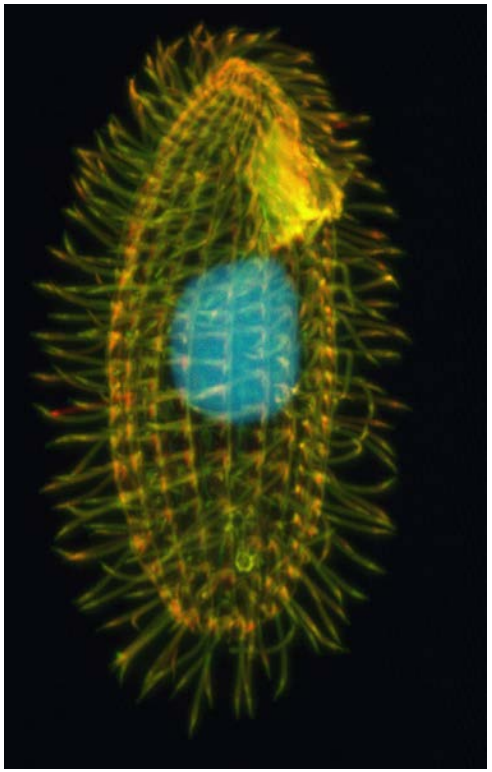
Wikipedia

34 Mb

12500 gènes

Tetrahymena thermophyla

- Phylum
  - Cilié (protozoaire)
- Modèle pour
  - Compartiments cellulaires, microtubules
  - Activités enzymatiques
  - Recherche biomédicale
- Avantages
  - Culture facile en grande quantité



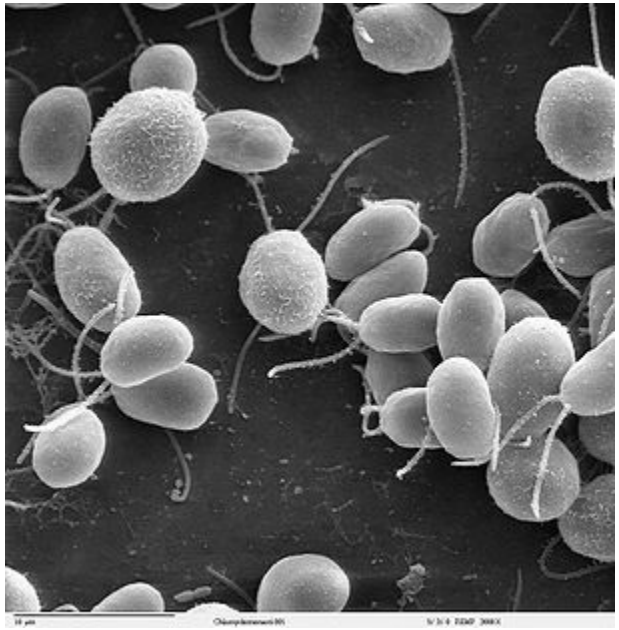
Wikipedia

104 Mb (MAC génome)

27000 gènes

Chlamydomonas reinhardtii

- Phylum :
  - Algue verte (chlorophyte) unicellulaire
- Modèle pour
  - Eucaryotes photosynthétiques, plantes
  - Mouvement, réponse à la lumière
  - Biologie cellulaire et moléculaire
- Avantages
  - Nombreux mutants connus
  - Nombreux outils biologiques disponibles



Wikipedia

14 Mb

4800 gènes

Arabidopsis thaliana

- Phylum
  - Plantes, Angiosperme, dicotyledone
- Modèle pour
  - Biologie végétale, agronomie
- Avantages
  - Petite taille de la plante
  - Petit génome pour une plante
  - Cycle de développement court (2 mois)
  - 40000 graine/plan

Wikipedia

157 Mb

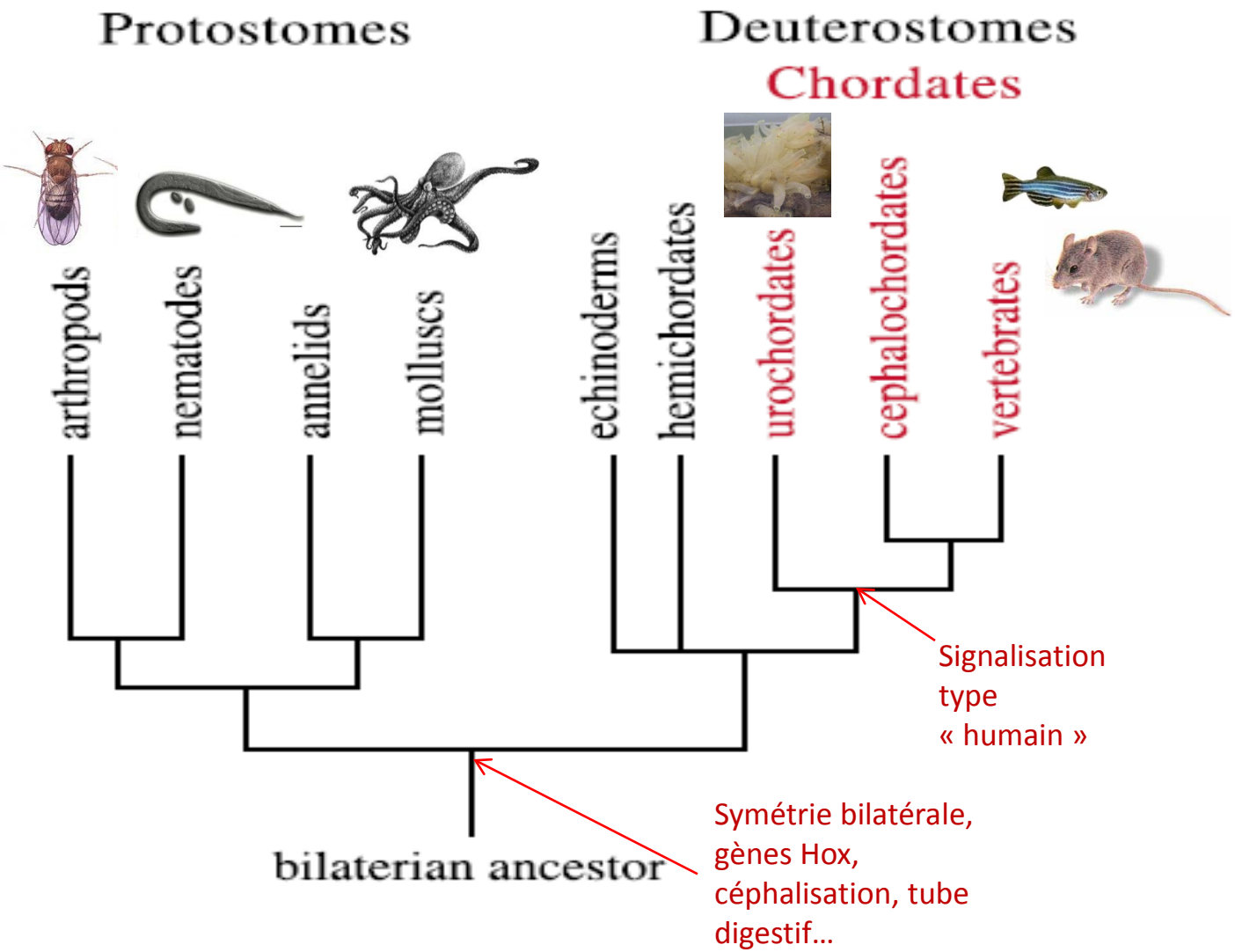
26000 gènes

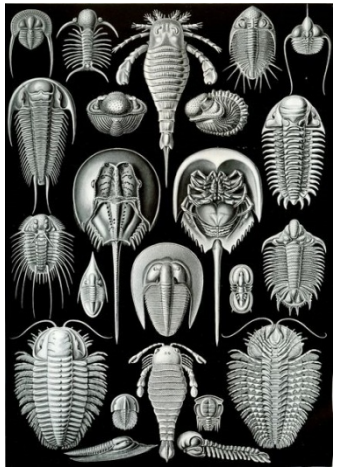
Les métazoaires (animaux pluricellulaires)

Source: http://tolweb.org

Pluricellularité hétérotrophe, différentiation, formation des tissus animaux

Les bilatériens





Ernst Haeckel

Dehal et al. Science, 298:2002

Coenorhabditis elegans

Source: www.wormatlas.org

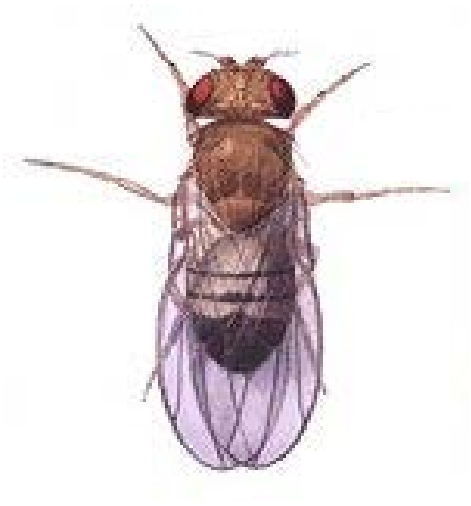
- Phylum
  - Metazoaire, Nématode
- Modèle pour:
  - Développement/différenciation
  - RNA interférence
- Avantages:
  - Pharynx, uterus, oocytes, oviducte, intestin, ovaire et même un comportement social en *959 cellules*.
  - Descendance: environ 300
  - Génome disponible depuis 1999: le premier génome animal complet.
  - Pour chaque gène humain d'intérêt, l'homologue dans C. elegans peut être identifié en quelques minutes. Trouver sa fonction dans le ver est ensuite relativement facile par K.O. (Knock Out).

100 Mb

14000 gènes

Drosophila melanogaster

- Phylum:
  - Metazoaire, Insecte
- Modèle pour
  - Génétique
  - Développement
  - Biochimie
  - Biologie cellulaire
- Avantages
  - Facilité d'élevage
  - Génétique bien connue



Wikipedia

170 Mb

12000 gènes

Ciona intestinalis

- Phylum
  - Urochordés (chordés primitifs), Tuniciers
- Modèle pour
  - Gènes Hox, développement
  - Contient la plupart des gènes de signalisation et de développement de l'homme
  - Origine des chordés/vertébrés
- Avantages
  - Le plus petit génome parmi les chordés pouvant être manipulés expérimentalement
  - Le système experimental le plus simple chez les chordés

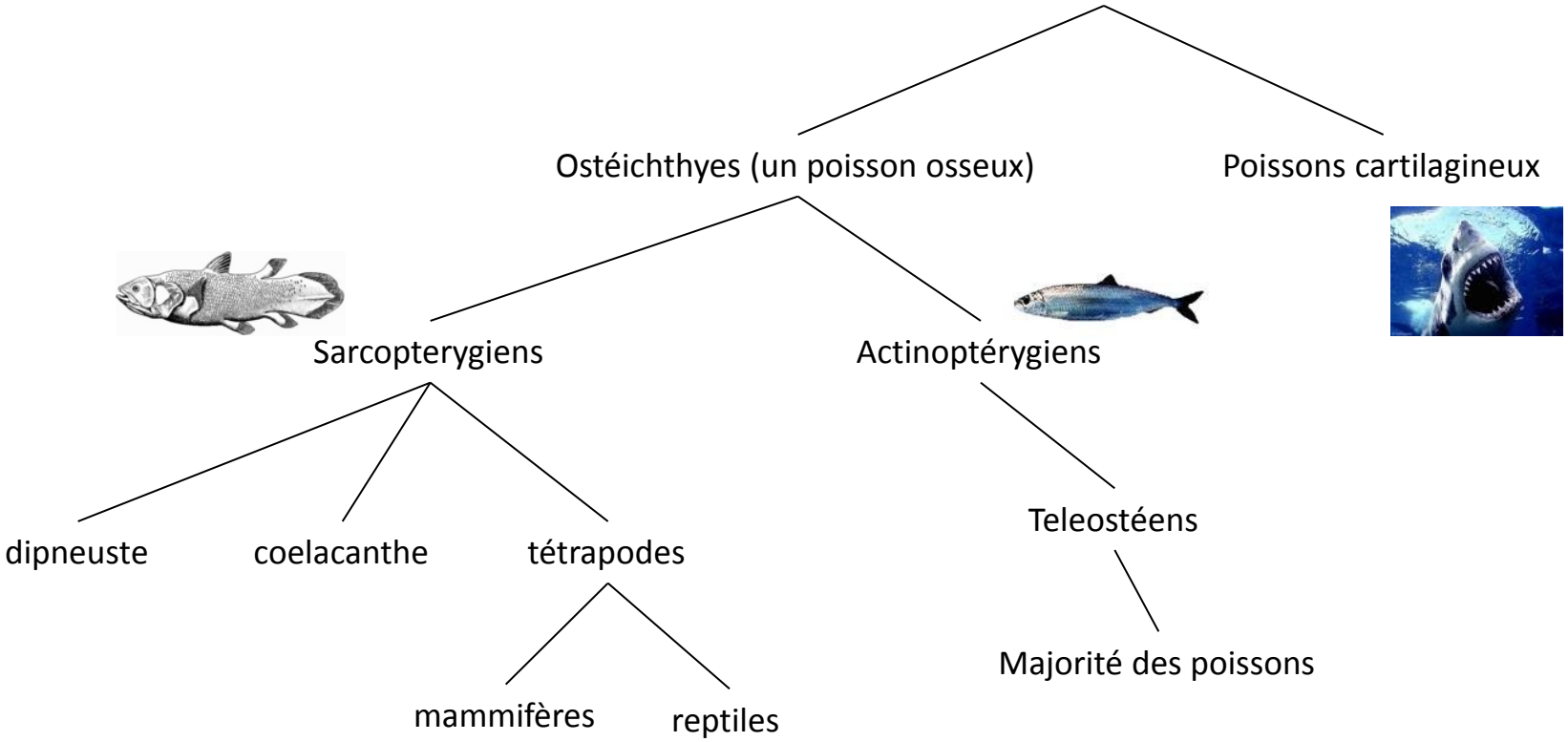
Wikipedia

160 Mb

16000 gènes

Nous sommes des poissons!

Les poissons ne forment pas un groupe monophylétique.



Danio rerio

- Phylum
  - Métazoaire, Vertébré, Téléostéens
- Modèle pour
  - Evolution des vertébrés
  - Développement
  - Génétique
  - Régénération des organes
- Avantages
  - Développement rapide de l'embryon
  - Embryons grands, transparents, resistants

Ensembl

1,5 Gb

47000 gènes

Xenopus

- Phylum
  - Métazoaire, Vertébré, Amphibien
- Modèle pour
  - Développement des vertébrés
  - Processus cellulaires chez les vertébrés
- Avantages
  - Grands oocytes (1mm): facilité d'expression d'ARN exogène + suivi embryon, ou utilisation d'extraits cytoplasmiques
  - Pb avec *X. laevis*: génome tetraploïde -> utilisation de *X. tropicalis (diploïde)*

X. laevis, X. tropicalis, Photo of Enrique Amaya

1.5 Gb

18 000 gènes

Mus musculus

- Phylum
  - Métazoaire, Mammifère
- Modèle pour
  - Maladies humaines, cancers
  - Essais pharmaceutiques
- Avantages
  - Reproduction relativement rapide
  - KO/clones

Wikipedia

3 Gb

22000 gènes

Les génomes non modèles mais « intéressants »

- Pathogènes procaryotes
  - *Staphilococcus aureus, Yersinia pestis …*
- Pathogènes eucaryotes
  - *Plasmodium falsciparum, …*
- Cultures
  - *Oryza sativa, Zea mays…*
- Bétail
  - *Sus scrofa, Bos taurus…*
- Biotechnologie
  - *Lactobacillus, Rhyzobium…*
- Génomique comparative
  - *Pan troglodytes, Ciona savignyi…*

Homo sapiens

- Phylum
  - Métazoaire, Mammifère, Primate
- Modèle pour
  - Mauvais modèle, mais intéressant à plus d'un titre
- Inconvénients
  - Expérimentation impossible sauf cellules somatiques en culture (Hela..)
  - Temps de génération lent

Raël

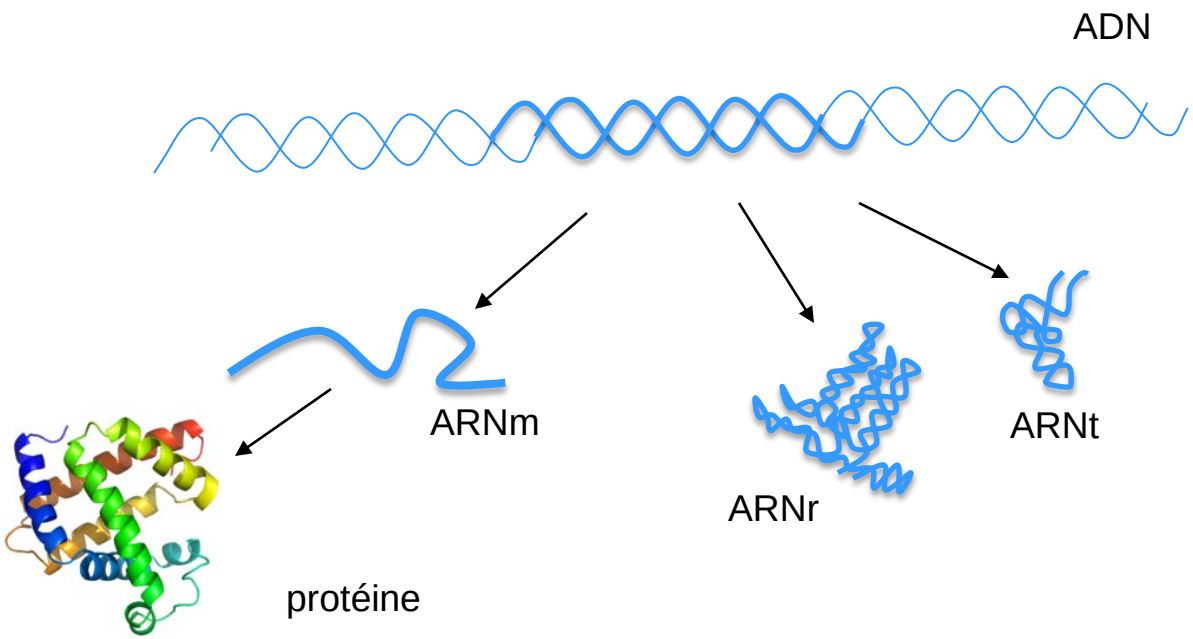
3 Gb

22000 gènes

2. Structure des génomes

Le gène

- Un gène est une séquence d'ADN qui spécifie la synthèse d'une protéine ou d'un ARN fonctionnel (définition ~Wikipedia)
- Un gène peut donc coder pour un ARN messager ou pour un ARN non-messager (ARNr, ARNt, …)



Caractéristiques des génomes procaryotes

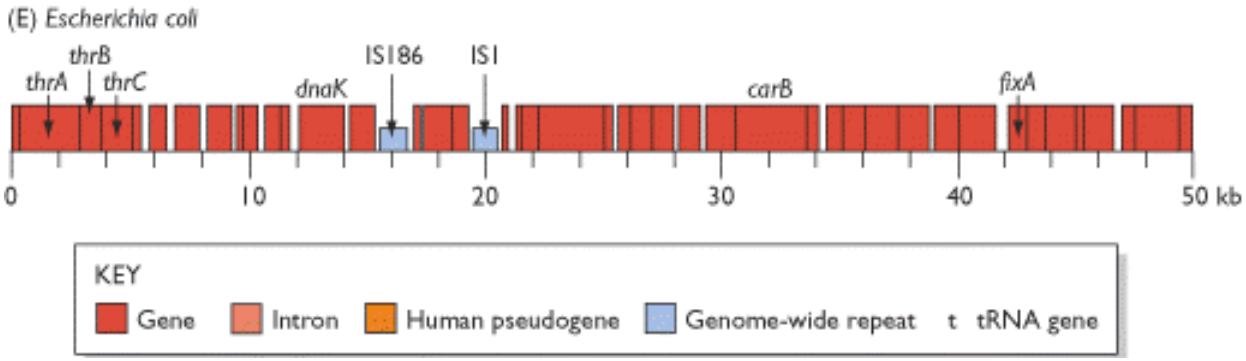
- Chromosome circulaire unique
- Présence possible de petites séquences d'ADN circulaires indépendantes : **les plasmides.**
- Contenu en G+C variable selon les espèces.
    - Ex : 22% chez un parasite *Wigglewortbia glossinidia*
    - 67% chez *deinicoccus radiodurans*

Les gènes procaryotes

- Fraction codante des génomes élevée.
  - > 90% codant
  - Peu de séquences intergéniques
  - Génome « compact »
- Chez les procaryotes **la séquence des gènes est continue**. **Pas d'intron**
- Gènes organisés en opérons. 600 opérons dans le génome de *Escherichia coli*.

Densité des gènes procaryotes

- Longueur gène 950 nt. en moyenne (coli)
- Haute densité en gènes: 95% du génome est transcrit chez E. coli.

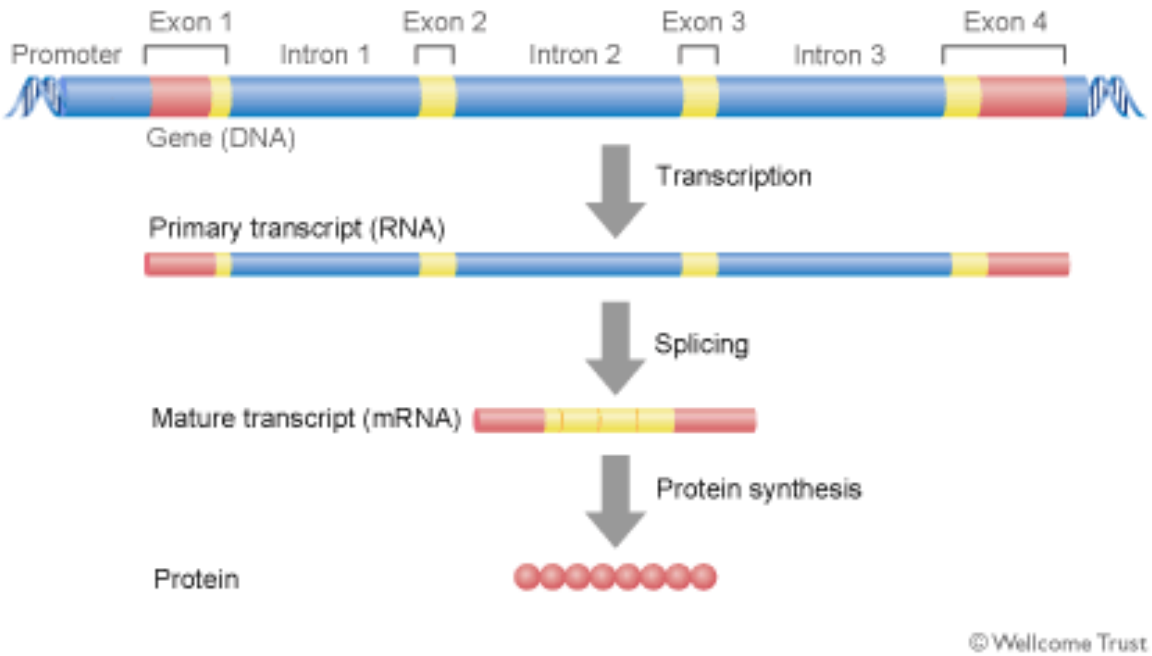


From « Genomes 2 », T.A. Brown

Caractéristiques des génomes eucaryotes

- Dans le noyau
- Taille >> procaryote
- Plusieurs chromosomes (homme 23, cheval 32, levure 16, drosophile 4…)
- Gènes « disloqués » (exons, introns)
- Grandes régions intergéniques de fonction inconnue

Le gène de mammifère



transcription

maturation

traduction

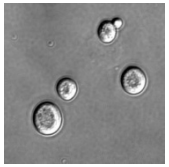
–Gène humain moyen: 27kb, 9 introns, codant:1,3kb , exon moyen: 145 bp, intron moyen:3365 bp.

–Gènes "monstres": dystrophine: 2,4 Mb; Facteur de coagulation VIII: 186 kb, 26 exons; Tinine: codant: 80kb, 178 exons

Densité des gènes eucaryotes

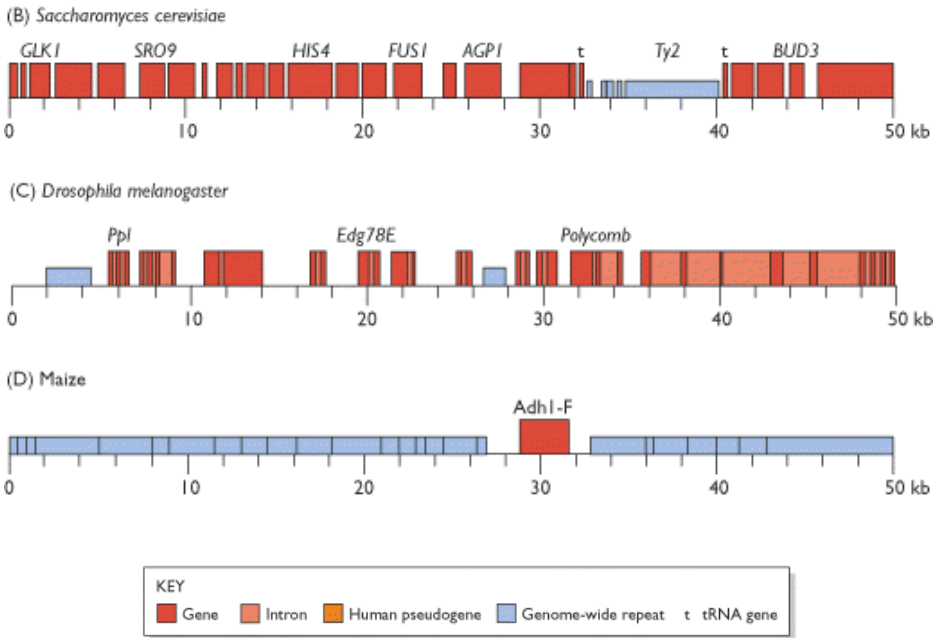
Densité moyenne:

- S. Cerevisiae: 1 gène/2kb.



- Drosophila: 1gène/10kb

- Maïs: 1 gène tous les 70kb
- Humain: 1 gène tous les 100kb



From « Genomes 2 », T.A. Brown

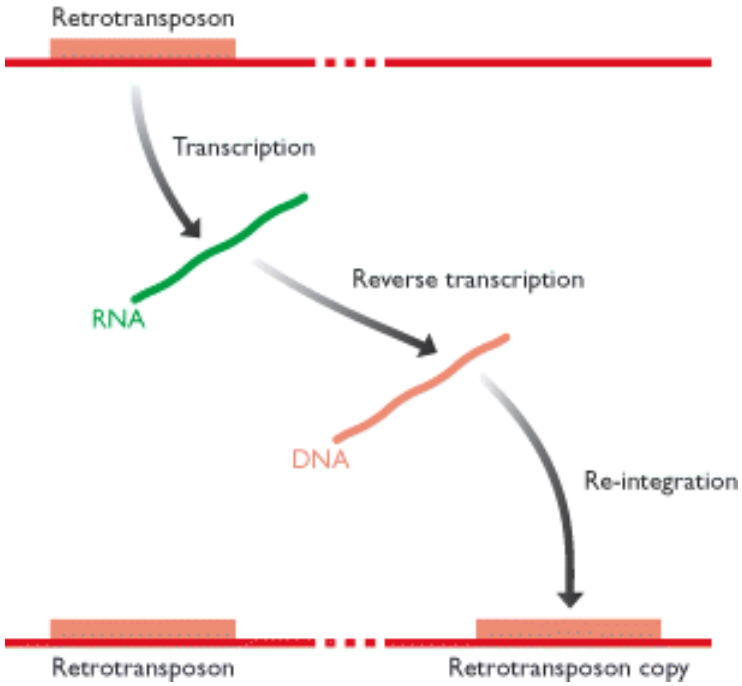
Junk DNA: les séquences répétées dans le génome humain

- 5 classes de séquences répétées
  - Répétition de type transposon (ou interspersed repeats)
  - copie rétro-transposées inactives de gènes (protéines ou ARN) (processed pseudogenes)
  - Répétition simples de k-mères courts, p. ex. (A)n, (CA)n ou (CGG)n
  - Segments dupliqués: blocs de 10–300 kb copiés d'une région à l'autre
  - Blocs de séquences répétes en tandems (centromères, télomères, clusters de gènes ribosomiques)

Les plus fréquentes

- Un ADN pas si "poubelle" que ça qui joue un grand rôle dans la transformation des gènes et l'apparition de nouveaux gènes.
- Il y a des régions pauvres en répétitions (p. ex. région des gènes HOX) et des régions riches (région de 500kb du chr. 11 contenant 89% d'éléments transposables)
- Les répétitions humaines sont relativement anciennes comparées à celles qu'on trouve dans le génome de drosophile. Notre génome a des difficultés pour se débarasser des répétitions.

Retroposons: les principales séquences répétées chez l'homme



From « Genomes 2 », T.A. Brown

Séquences répétées de type transposon

- Les séquences répétées de type transposon représentent plus de 1/3 du génome des vertébrés
- Génome humain: 45% !!

Classes of interspersed repeat in the human genome

			Length	Copy number	Fraction of genome
LINEs	Autonomous	ORF1 ORF2 (pol) AAA	6–8 kb	850,000	21%
SINEs	Non-autonomous	A B AAA	100–300 bp	1,500,000	13%
Retrovirus-like elements	Autonomous	gag pol (env)	6–11 kb	450,000	8%
	Non-autonomous	(gag)	1.5–3 kb		
DNA transposon fossils	Autonomous	transposase	2–3 kb	300,000	3%
	Non-autonomous		80–3,000 bp		

En quoi nos génomes diffèrent-ils?

- Les individus d'espèces différentes ont des génomes différents par la taille, l'ordre et la nature des informations qu'ils contiennent.
- On considère souvent que 2 individus de la même espèce possèdent le « même » génome.
- En fait, le génome de chaque individu est unique. Chez l'homme le génome diffère de 0.5% entre 2 personnes non apparentées

Les variations du génome dans une population

- **Très importantes médicalement**
  - Pharmacogénomique: comment chaque patient répond aux drogues
  - Marqueurs de susceptibilité aux maladies

- **Polymorphismes dans le génome humain**
  - Insertions, délétions, duplications, réarrangements
    - rares et peu étudiés
  - Microsatellites etc..
  - Single Nucleotide Polymorphism (SNP)

Single Nucleotide Polymorphism (SNP)

- Le polymorphisme le plus commun chez l'homme.
- La plupart n'ont pas d'implications fonctionnelles
- 1 SNP tous les ~1000 bases chez l'homme

Applications

- Les SNP constituent une trace historique pour l'étude de la phylogénie humaine: ils mutent lentement et ont peu de chance de réapparaître de façon récurrente.
- Les SNP sont à l'origine de susceptibilité ou de résistance à de nombreuses maladies.
- Cartographie de maladies à caractères complexes (cancers, diabète, maladies mentales)
- Prédiction des réponses aux drogues.

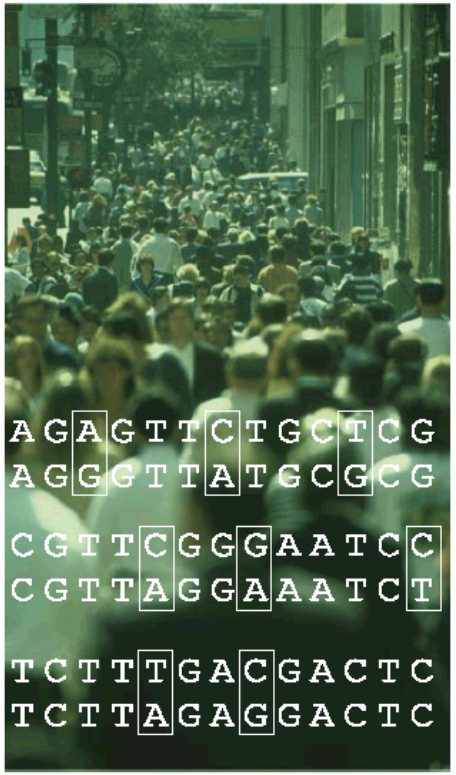
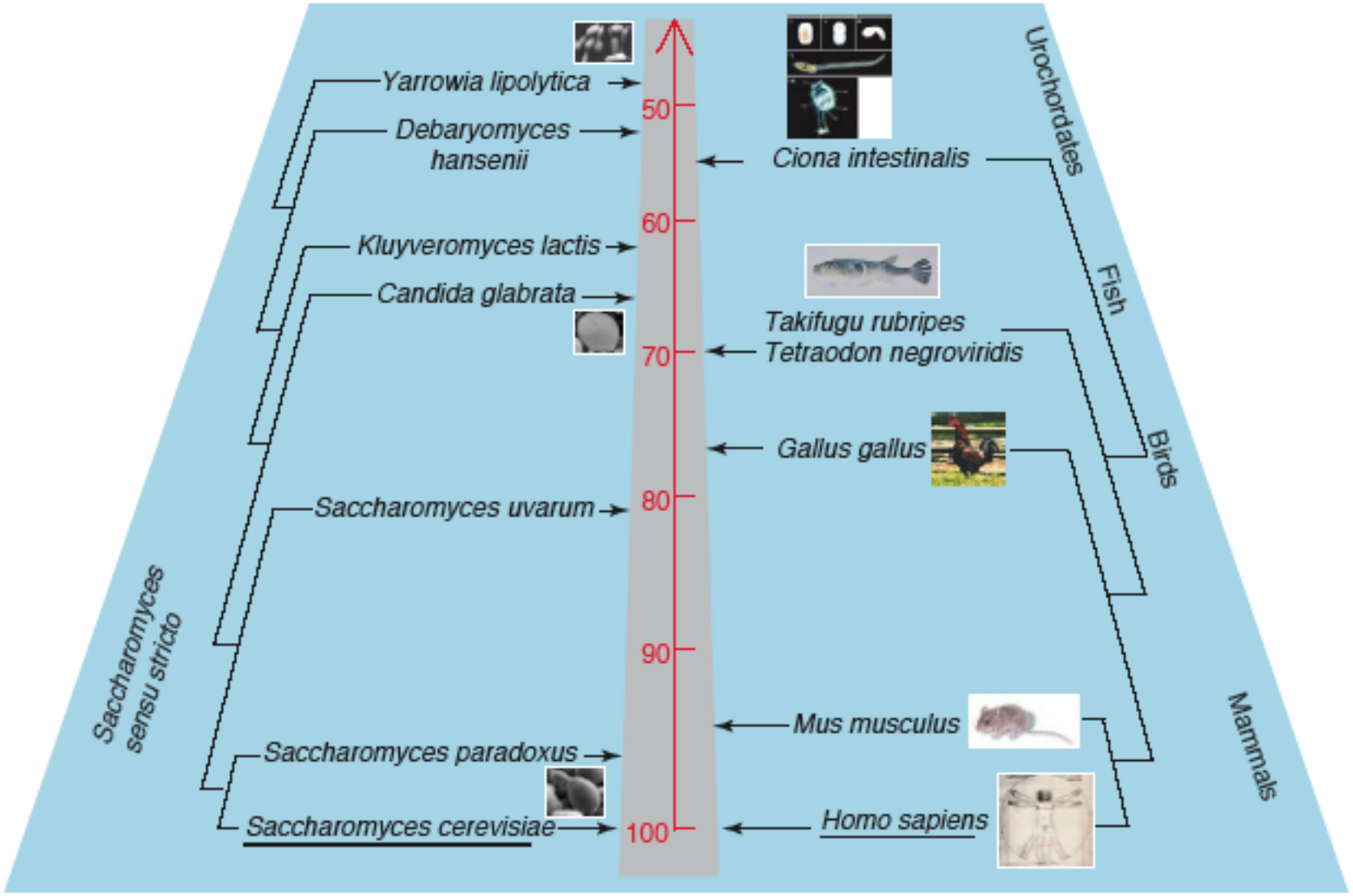


Figure 1 **The most common sources of variation between humans are single nucleotide polymorphisms (SNPs) — single base differences between genome sequences. Fragments of two sequences, with eight SNPs, are shown.**

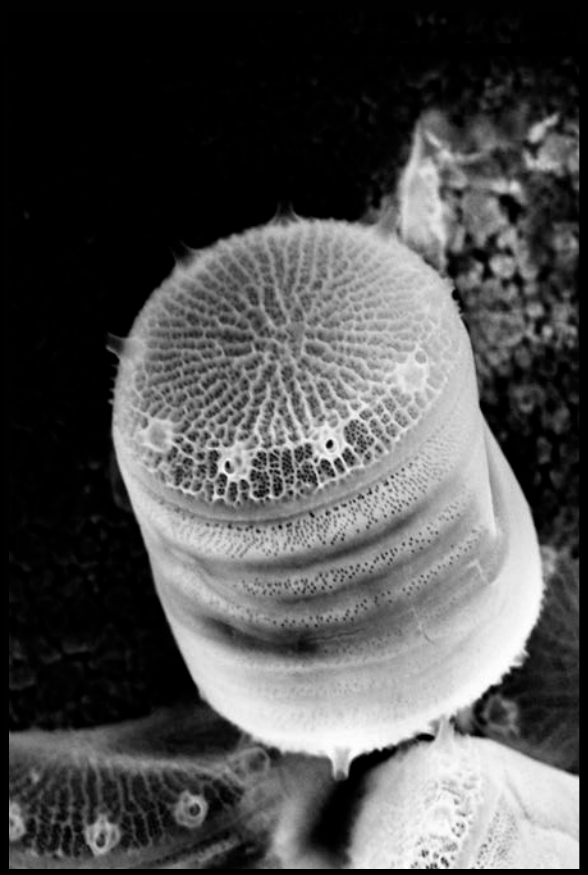
La ressemblance entre génomes

- Homme A / homme B
  - 99,5% identique (0,5% différence)
- Homme/chimpanzé
  - Codant: 98,5% identique
  - Non codant: ~96% identique
  - 90Mb d'insertions/délétions et 35 millions de différences ponctuelles
- Homme/souris
  - Codant: 90% identique
  - Non codant: la majorité est sans identité apparente, mais on trouve quand même de nombreux segments semblables (« conservés »)
- Homme/poulet
  - Codant: 80% identique
- Homme/poisson
  - Codant: 70% identique



From B. Dujon, Trends in Genetics, 2006
Diatomées: Chris Bowler

Diatomées

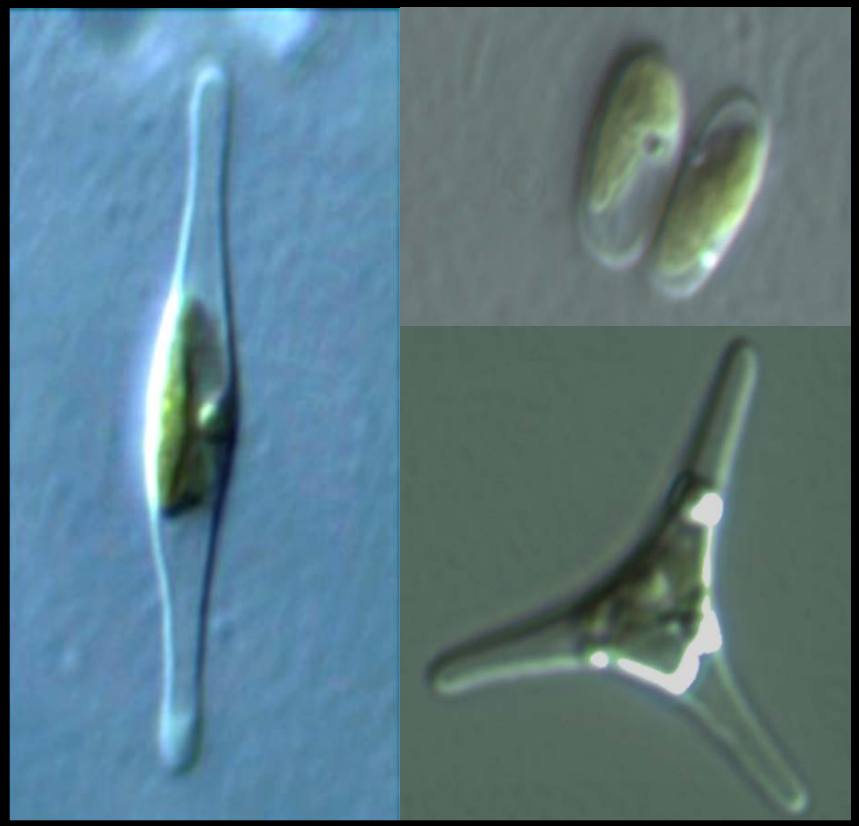


Thalassiosira pseudonana

32 Mb, ~ 11,000 genes

Armbrust et al.

Science (2004)

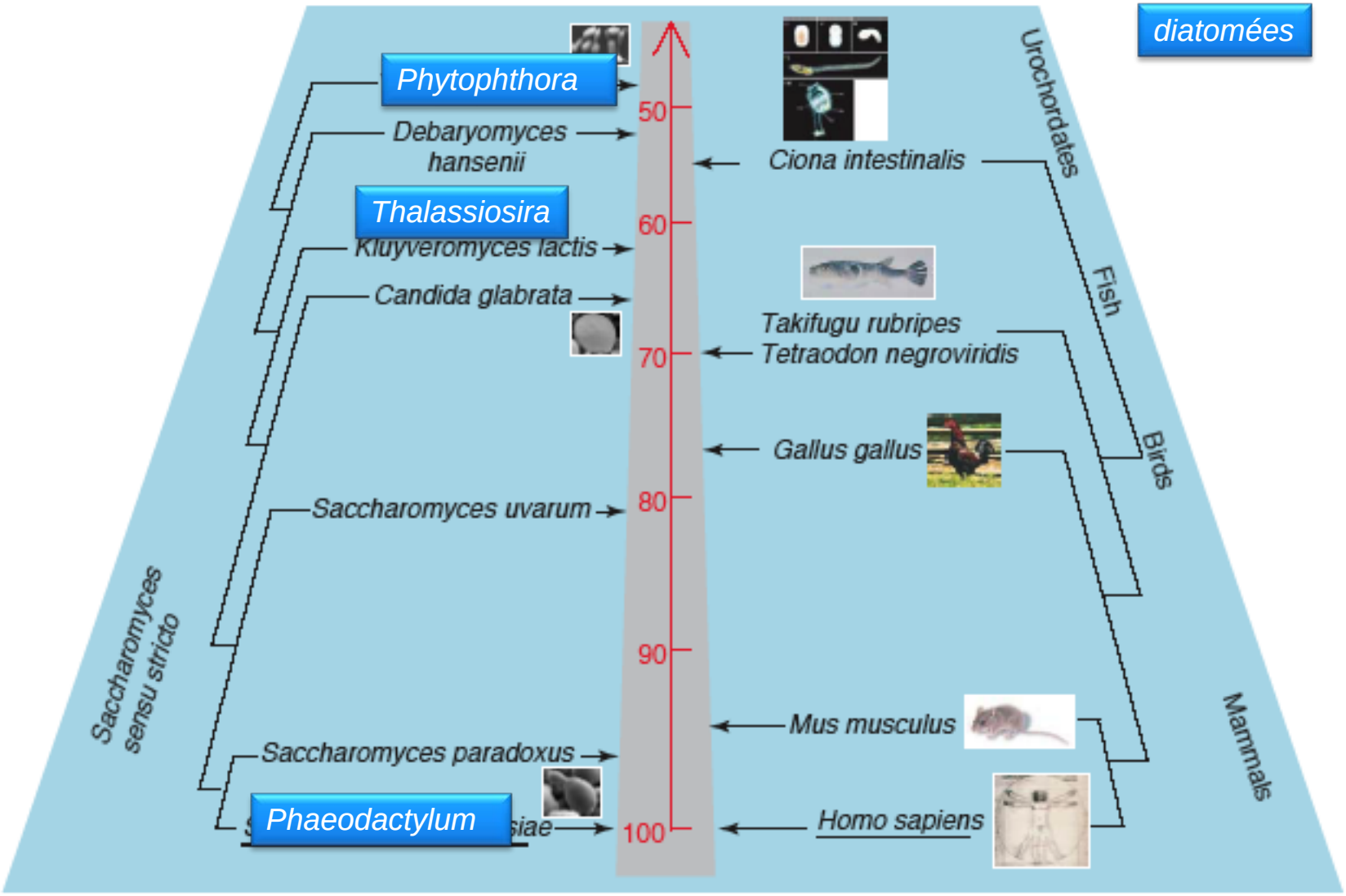


Phaeodactylum tricornutum

27 Mb, ~10,000 genes

Bowler et al.

Nature (2008)



From B. Dujon, Trends in Genetics, 2006
Diatomées: Chris Bowler

Exemples de conservation de séquence entre espèces

RPLP0: un gène de protéine ribosomale universellement conservé. mRNA humain (exons seulement)

- >ENST00000392514 cdna:KNOWN_protein_coding
  GTCTGACGGGCGATGGCGCAGCCAATAGACAGGAGCGCTATCCGCGGTTTCTGATTGGCT
  ACTTTGTTCGCATTATAAAAGGCACGCGCGGGCGCGAGGCCCTTCTCTCGCCAGGCGTCC
  TCGTGGAAGTGACATCGTCTTTAAACCCTGCGTGGCAATCCCTGACGCACCGCCGTGATG
  CCCAGGGAAGACAGGGCGACCTGGAAGTCCAACTACTTCCTTAAGATCATCCAACTATTG
  GATGATTATCCGAAATGTTTCATTGTGGGAGCAGACAATGTGGGCTCCAAGCAGATGCAG
  CAGATCCGCATGTCCCTTCGCGGGAAGGCTGTGGTGCTGATGGGCAAGAACACCATGATG
  CGCAAGGCCATCCGAGGGCACCTGGAAAACAACCCAGCTCTGGAGAAACTGCTGCCTCAT
  ATCCGGGGGAATGTGGGCTTTGTGTTCACCAAGGAGGACCTCACTGAGATCAGGGACATG
  TTGCTGGCCAATAAGGTGCCAGCTGCTGCCCGTGCTGGTGCCATTGCCCCATGTGAAGTC
  ACTGTGCCAGCCCAGAACACTGGTCTCGGGCCCGAGAAGACCTCCTTTTTCCAGGCTTTA
  GGTATCACCACTAAAATCTCCAGGGGCACCATTGAAATCCTGAGTGATGTGCAGCTGATC
  AAGACTGGAGACAAAGTGGGAGCCAGCGAAGCCACGCTGCTGAACATGCTCAACATCTCC
  CCCTTCTCCTTTGGGCTGGTCATCCAGCAGGTGTTCGACAATGGCAGCATCTACAACCCT
  GAAGTGCTTGATATCACAGAGGAAACTCTGCATTCTCGCTTCCTGGAGGGTGTCCGCAAT
  GTTGCCAGTGTCTGTCTGCAGATTGGCTACCCAACTGTTGCATCAGTACCCCATTCTATC
  ATCAACGGGTACAAACGAGTCCTGGCCTTGTCTGTGGAGACGGATTACACCTTCCCACTT
  GCTGAAAAGGTCAAGGCCTTCTTGGCTGATCCATCTGCCTTTGTGGCTGCTGCCCCTGTG
  GCTGCTGCCACCACAGCTGCTCCTGCTGCTGCTGCAGCCCCAGCTAAGGTTGAAGCCAAG
  GAAGAGTCGGAGGAGTCGGACGAGGATATGGGATTTGGTCTCTTTGACTAATCACCAAAA
  AGCAACCAACTTAGCCAGTTTTATTTGCAAAACAAGGAAATAAAGGCTTACTTCTTTAAA
  AAGTCTCTGGACTCTTAA

Alignement avec mRNA RPLP0 de *Pan troglodytes*

```
GENE ID: 452301 RPLP0 | ribosomal protein, large, P0 [Pan troglodytes]

Score = 2161 bits (2396),  Expect = 0.0
Identities = 1211/1218 (99%), Gaps = 1/1218 (0%)
Strand=Plus/Plus

Query  1     GTCTGACGGGCGATGGCGCAGCCAATAGACAGGAGCGCTATCCGCGGTTTCTGATTGGCT  60
             ||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     GTCGGATGGGCGATGGCGCAGCCAATAGACAGGAGCGCTATCCGCGGTTTCTGATTGGCT  60

Query  61    ACTTTGTTCGCATTATAAAAGGCACGCGCGGGCGCGAGGCCCTTCTCTCGCCAGGCGTCC  120
             ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61    ACTTTGTTCGCAT-ATAAAAGGCACGCGCGGGCGCGAGGCCCTTCTCTCGCCAGGCGTCC  119

Query  121   TCGTGGAAGTGACATCGTCTTTAAACCCTGCGTGGCAATCCCTGACGCACCGCCGTGATG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  120   TCGTGGAAGTGACATCGTCTTTAAACCCTGCGTGGCAATCCCTGACGCACCGCCGTGATG  179

Query  181   CCCAGGGAAGACAGGGCGACCTGGAAGTCCAACTACTTCCTTAAGATCATCCAACTATTG  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  180   CCCAGGGAAGACAGGGCGACCTGGAAGTCCAACTACTTCCTTAAGATCATCCAACTATTG  239

Query  241   GATGATTATCCGAAATGTTTCATTGTGGGAGCAGACAATGTGGGCTCCAAGCAGATGCAG  300
             ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  240   GATGATTATCCGAAATGTTTCATCGTGGGAGCAGACAATGTGGGCTCCAAGCAGATGCAG  299

Query  301   CAGATCCGCATGTCCCTTCGCGGGAAGGCTGTGGTGCTGATGGGCAAGAACACCATGATG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  300   CAGATCCGCATGTCCCTTCGCGGGAAGGCTGTGGTGCTGATGGGCAAGAACACCATGATG  359

Query  361   CGCAAGGCCATCCGAGGGCACCTGGAAAACAACCCAGCTCTGGAGAAACTGCTGCCTCAT  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  360   CGCAAGGCCATCCGAGGGCACCTGGAAAACAACCCAGCTCTGGAGAAACTGCTGCCTCAT  419

Query  421   ATCCGGGGGAATGTGGGCTTTGTGTTCACCAAGGAGGACCTCACTGAGATCAGGGACATG  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  420   ATCCGGGGGAATGTGGGCTTTGTGTTCACCAAGGAGGACCTCACTGAGATCAGGGACATG  479

Query  481   TTGCTGGCCAATAAGGTGCCAGCTGCTGCCCGTGCTGGTGCCATTGCCCCATGTGAAGTC  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  480   TTGCTGGCCAATAAGGTGCCAGCTGCTGCCCGTGCTGGTGCCATTGCCCCATGTGAAGTC  539

Query  541   ACTGTGCCAGCCCAGAACACTGGTCTCGGGCCCGAGAAGACCTCCTTTTTCCAGGCTTTA  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  540   ACTGTGCCAGCCCAGAACACTGGTCTCGGGCCCGAGAAGACCTCCTTTTTCCAGGCTTTA  599

Query  601   GGTATCACCACTAAAATCTCCAGGGGCACCATTGAAATCCTGAGTGATGTGCAGCTGATC  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  600   GGTATCACCACTAAAATCTCCAGGGGCACCATTGAAATCCTGAGTGATGTGCAGCTGATC  659
```

Alignement avec mRNA RPLP0 de *Danio rerio*

```
>ref|NM_131580.2| G Danio rerio ribosomal protein, large, P0 (rplp0), mRNA
 gb|BC062854.1| UGM Danio rerio ribosomal protein, large, P0, mRNA (cDNA clone MGC:77791
IMAGE:7001273), complete cds
Length=1105

 GENE ID: 58101 rplp0 | ribosomal protein, large, P0 [Danio rerio]
(Over 10 PubMed links)

 Score =  888 bits (984),  Expect = 0.0
 Identities = 801/1000 (80%), Gaps = 10/1000 (1%)
 Strand=Plus/Plus

Query  137   GTCTTTAAACC--CTGCGTGGCAATCCCTG-ACGCACCGCCGTGATGCCCAGGGAAGACA  193
             |||||||||||  |||      | |||| || | |||| ||   |||||||||||||||||
Sbjct  31    GTCTTTAAACCGGCTGTTCACCGATCCTTGGAAGCACTGCAAAGATGCCCAGGGAAGACA  90

Query  194   GGGCGACCTGGAAGTCCAACTACTTCCTTAAGATCATCCAACTATTGGATGATTATCCGA  253
             |||| || |||||||||||||||||| || || ||||||||||  ||||||| || || |
Sbjct  91    GGGCCACGTGGAAGTCCAACTACTTTCTGAAAATCATCCAACTGCTGGATGACTACCCCA  150

Query  254   AATGTTTCATTGTGGGAGCAGACAATGTGGGCTCCAAGCAGATGCAGCAGATCCGCATGT  313
             | |||||||| ||||| ||||||||||| |||||||||||||||||||    |||||  |||
Sbjct  151   AGTGTTTCATCGTGGGCGCAGACAATGTCGGCTCCAAGCAGATGCAGACCATCCGTCTGT  210

Query  314   CCCTTCGCGGGAAGGCTGTGGTGCTGATGGGCAAGAACACCATGATGCGCAAGGCCATCC  373
             |||| || || ||||| || ||||| ||||| || |||||||||||| | ||||||||||
Sbjct  211   CCCTGCGGGGCAAGGCCGTCGTGCTCATGGGGAAAAACACCATGATGAGGAAGGCCATCC  270

Query  374   GAGGGCACCTGGAAAACAACCCAGCTCTGGAGAAACTGCTGCCTCATATCCGGGGGAATG  433
             | || |||||||||||||||||||||||||||  ||||| || || ||||| ||||| |
Sbjct  271   GTGGCCACCTGGAAAACAACCCAGCTCTGGAGAGGCTGCTTCCCCACATCCGCGGGAACG  330

Query  434   TGGGCTTTGTGTTCACCAAGGAGGACCTCACTGAGATCAGGGACATGTTGCTGGCCAATA  493
             ||||||| || |||||||||||||| || |||||| || | ||| || ||||||| || |
Sbjct  331   TGGGCTTCGTCTTCACCAAGGAGGATCTGACTGAGGTCCGAGACCTGCTGCTGGCAAACA  390

Query  494   AGGTGCCAGCTGCTGCCCGTGCTGGTGCCATTGCCCCATGTGAAGTCACTGTGCCAGCCC  553
             | ||||| |||||||||||||||||||||||| ||||| ||||| || || ||||| ||||
Sbjct  391   AAGTGCCCGCTGCTGCCCGTGCTGGTGCCATCGCCCCCTGTGAGGTGACCGTGCCGGCCC  450

Query  554   AGAACACTGGTCTCGGGCCCGAGAAGACCTCCTTTTTCCAGGCTTTAGGTATCACCACTA  613
             ||||||| || ||||| || ||||||||||| || ||||||||||| || |||||||| |
Sbjct  451   AGAACACCGGGCTCGGTCCTGAGAAGACCTCTTTCTTCCAGGCTTTGGGAATCACCACCA  510

Query  614   AAATCTCCAGGGGCACCATTGAAATCCTGAGTGATGTGCAGCTGATCAAGACTGGAGACA  673
             | |||||||| || ||||||||||| ||||||| || |||||||||||  |||||||||
```

Alignement avec mRNA RPLP0 de *Drosophia melanogaster*

```
>ref|NM_079487.3| UGM Drosophila melanogaster ribosomal protein LP0 (RpLP0), mRNA
Length=1261

 GENE ID: 40451 RpLP0 | Ribosomal protein LP0 [Drosophila melanogaster]
(Over 10 PubMed links)

 Score =  340 bits (376),  Expect = 2e-90
 Identities = 423/577 (73%), Gaps = 2/577 (0%)
 Strand=Plus/Plus

Query  184  AGGGAAGACAGGGCGACCTGGAAGTCCAACTACTTCCTTAAGATCATCCAACTATTGGAT  243
            |||||  ||| |||  | |||||| |  | |||||| | ||| |  |  |||| || |||
Sbjct  142  AGGGAGAACAAGGCAGCGTGGAAGGCTCAGTACTTCATCAAGGTTGTGGAACTGTTCGAT  201

Query  244  GATTATCCGAAATGTTTCATTGTGGGAGCAGACAATGTGGGCTCCAAGCAGATGCAGCAG  303
            || |  || || || ||||| ||||| || ||||| ||||||||||||||||||||| |
Sbjct  202  GAGTTCCCAAAGTGCTTCATCGTGGGCGCCGACAACGTGGGCTCCAAGCAGATGCAGAAC  261

Query  304  ATCCGCATGTCCCTTCGCGGGAAGGCTGTGGTGCTGATGGGCAAGAACACCATGATGCGC  363
            ||||| |     ||| || ||   ||| || ||||| ||||||||||||||||||||||||
Sbjct  262  ATCCGTACCAGCCTGCGTGGACTGGCCGTCGTGCTTATGGGCAAGAACACCATGATGCGC  321

Query  364  AAGGCCATCCGAGGGCACCTGGAAAACAACCCAGCTCTGGAGAAACTGCTGCCTCATATC  423
            ||||||||||| || || ||||| ||||||||    |||||||| ||||| || || |||
Sbjct  322  AAGGCCATCCGCGGTCATCTGGAGAACAACCCGCAGCTGGAGAAGCTGCTACCCCACATC  381

Query  424  CGGGGGAATGTGGGCTTTGTGTTCACCAAGGAGGACCTCACTGAGATCAGGGACATGTTG  483
              ||| || ||||| || |||||||||||||  || ||| | ||| |   | |||| | ||
Sbjct  382  AAGGGCAACGTGGGATTCGTGTTCACCAAGGGCGATCTCGCCGAGGTGCGCGACAAGCTG  441

Query  484  CTGGCCAATAAGGTGCCAGCTGCTGCCCGTGCTGGTGCCATTGCCCCATGTG-AAGTCAC  542
            ||||     ||||||||  ||  | ||||||| | || || |||||||| | || | ||||
Sbjct  442  CTGGAGTCCAAGGTGCGCGCCCCCGCCCGTCCCGGCGCTATTGCCCC-TCTGCACGTCAT  500

Query  543  TGTGCCAGCCCAGAACACTGGTCTCGGGCCCGAGAAGACCTCCTTTTTCCAGGCTTTAGG  602
              | || || |||||||| ||  | || |||||||||||   || ||||||||  |
Sbjct  501  CATCCCGGCGCAGAACACCGGCTTGGGACCCGAGAAGACCAGTTTCTTCCAGGCCCTGTC  560

Query  603  TATCACCACTAAAATCTCCAGGGGCACCATTGAAATCCTGAGTGATGTGCAGCTGATCAA  662
             ||| | || ||||| |||| ||| || ||||||||| | |  |||||||   |  | ||
Sbjct  561  CATCCCGACCAAAATTTCCAAGGGAACAATTGAAATCATCAACGATGTGCCCATCCTGAA  620

Query  663  GACTGGAGACAAAGTGGGAGCCAGCGAAGCCACGCTGCTGAACATGCTCAACATCTCCCC  722
            | |||| ||||| || || |||  ||| || || ||||| |||||| | ||||||||  ||
Sbjct  621  GCCTGGCGACAAGGTCGGCGCCTCCGAGGCGACACTGCTCAACATGTTGAACATCTCGCC  680
```

Alignement avec mRNA RPLP0 de *Arabidopsis thaliana*

```
>ref|NM_129559.2| UEGM Arabidopsis thaliana 60S acidic ribosomal protein P0-1 (AT2G40010)
mRNA, complete cds
Length=1153

 GENE ID: 818589 AT2G40010 | 60S acidic ribosomal protein P0-1
[Arabidopsis thaliana] (10 or fewer PubMed links)

 Score = 66.2 bits (72),  Expect = 6e-08
 Identities = 261/402 (65%), Gaps = 14/402 (3%)
 Strand=Plus/Plus

Query  409  CTGCTGCCTCATATCCGGGGGAATGTGGGCTTTGTGTTCACCAAGGAGGACCTCACTGAG  468
            || || |||| | | | ||||||||| || ||  | ||||| ||||  ||| | |  ||
Sbjct  241  CTCCTTCCTCTTCTTCAGGGGAATGTTGGGTTGATCTTCACTAAGGGTGACTTGAAGGAA  300

Query  469  ATCAGGGAC-ATGTTGCTGGCCAATAAGGTGCCAGCTGCTGCCCGTGCTGGTGCCATTGC  527
             |||| ||  | ||||||    |  |||||   |||| |||| ||||  |||    | ||
Sbjct  301  GTCAGTGAAGAGGTTGCTAAGTAC-AAGGTTGGAGCTCCTGCTCGTGTAGGTTTAGTCGC  359

Query  528  CCCATGTGAAGTCACTGTGCCAGCCCAG--AACACTGGTCTCGGGCCCGAGAAGACCTCC  585
             |||  ||| ||    |||| |  ||||  |||||||||| |  ||     ||||||||
Sbjct  360  TCCAATTGATGTGGTCGTGCAA--CCAGGCAACACTGGTCTTGACCCTTCACAGACCTCC  417

Query  586  TTTTTCCAG--GCTTTAGGTATCACCACTAAAATCTCCAGGGGCACCATTGAAATCCTGA  643
            || ||||||  |||| |   ||  | || ||||||  ||  || ||  |||| ||| | |
Sbjct  418  TTCTTCCAGGTGCTTAA--CATTCCAACCAAAATCAACAAAGGTACGGTTGAGATCATAA  475

Query  644  GTGATGTGCAGCTGATCAAGACTGGAGACAAAGTGGGAGCCAGCGAAGCCACGCTGCTGA  703
                |||| |||| |||||||  || |||||||| ||  |   ||| ||  |||| ||
Sbjct  476  CCCCTGTGGAGCTCATCAAGAAAGGCGACAAAGTCGGTTCATCCGAGGCTGCGCTTCTTG  535

Query  704  ACATGCTCAACATCTCCCCCTTCTCCTTTGGGCTGGTCATCCAGCAGGTGTTCGACAATG  763
             || |||    |||   ||||| || | ||| || ||  |  ||   || | ||| ||||
Sbjct  536  CCAAGCTTGGAATCAGGCCCTTTTCGTATGGTCTCGTTGTTGAGTCAGTCTACGATAATG  595

Query  764  G--CAGCATCTACAACCCTGAAGTGCTTGATATCACAGAGGA  803
            |  |||  | |  |||||||||||||||| |  |||| || ||
Sbjct  596  GGTCAG--TGTTTAACCCTGAAGTGCTTAACCTCACTGAAGA  635
```

Alignement avec mRNA RPLP0 de *Schizosaccharomyces pombe*

```
>ref|NM_001023549.1| GM Schizosaccharomyces pombe 972h- 60S acidic ribosomal protein
Rpp1-3 (rpp103), mRNA
Length=330

 GENE ID: 2539598 rpp103 | 60S acidic ribosomal protein Rpp1-3
[Schizosaccharomyces pombe 972h-] (10 or fewer PubMed links)

 Score = 53.6 bits (58),  Expect = 4e-04
 Identities = 77/108 (71%), Gaps = 3/108 (3%)
 Strand=Plus/Plus

Query  1023  tgctgccaccacagctgctcctgctgctgctgcagccccagctAAGGTTGAAGCCAAGGA  1082
             ||||| | || | ||| || |||  ||||| | ||    |   ||||  ||||||||||
Sbjct  225   TGCTGGCGCCGCCGCTCCTGCTGAAGCTGCCGAAGAAGAAAAGAAGGAAGAAGCCAAGGA  284

Query  1083  AGAGTCGGAGGAGTCGGACGAGGATATGGGATTTGGTCTCTTTGACTA  1130
              ||   ||| ||||| || ||||| ||||| |||||  | ||||||||
Sbjct  285   GGA---GGAAGAGTCTGATGAGGACATGGGTTTTGGCTTGTTTGACTA  329
```

RPLP0: la protéine

>sp|P05388|RLA0_HUMAN 60S acidic ribosomal protein P0 OS=Homo sapiens GN=RPLP0 PE=1 SV=1

MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGKAVVLMGKNTM
MRKAIRGHLENNPALEKLLPHIRGNVGFVFTKEDLTEIRDMLLANKVPAAARAGAIAPCE
VTVPAQNTGLGPEKTSFFQALGITTKISRGTIEILSDVQLIKTGDKVGASEATLLNMLNI
SPFSFGLVIQQVFDNGSIYNPEVLDITEETLHSRFLEGVRNVASVCLQIGYPTVASVPHS
IINGYKRVLALSVETDYTFPLAEKVKAFLADPSAFVAAAPVAAATTAAPAAAAAPAKVEA
KEESEESDEDMGFGLFD

```
>sp|P19889.1|RLA0_DROME RecName: Full=60S acidic ribosomal protein P0; AltName: Full=Apurinic-
endonuclease; AltName: Full=DNA-(apurinic
or apyrimidinic site) lyase
Length=317

 Score =  417 bits (1072),  Expect = 2e-146, Method: Compositional matrix adjust.
 Identities = 210/317 (66%), Positives = 255/317 (80%), Gaps = 0/317 (0%)

Query  1    MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGKAVVLMGKNTM  60
            M RE++A WK+ YF+K+++L D++PKCFIVGADNVGSKQMQ IR SLRG AVVLMGKNTM
Sbjct  1    MVRENKAAWKAQYFIKVVELFDEFPKCFIVGADNVGSKQMQNIRTSLRGLAVVLMGKNTM  60

Query  61   MRKAIRGHLENNPALEKLLPHIRGNVGFVFTKEDLTEIRDMLLANKVPAAARAGAIAPCE  120
            MRKAIRGHLENNP LEKLLPHI+GNVGFVFTK DL E+RD LL +KV A AR GAIAP
Sbjct  61   MRKAIRGHLENNPQLEKLLPHIKGNVGFVFTKGDLAEVRDKLLESKVRAPARPGAIAPLH  120

Query  121  VTVPAQNTGLGPEKTSFFQALGITTKISRGTIEILSDVQLIKTGDKVGASEATLLNMLNI  180
            V +PAQNTGLGPEKTSFFQAL I TKIS+GTIEI++DV ++K GDKVGASEATLLNMLNI
Sbjct  121  VIIPAQNTGLGPEKTSFFQALSIPTKISKGTIEIINDVPILKPGDKVGASEATLLNMLNI  180

Query  181  SPFSFGLVIQQVFDNGSIYNPEVLDITEETLHSRFLEGVRNVASVCLQIGYPTVASVPHS  240
            SPFS+GL++ QV+D+GSI++PE+LDI  E L ++F +GV N+A+VCL +GYPT+AS PHS
Sbjct  181  SPFSYGLIVNQVYDSGSIFSPEILDIKPEDLRAKFQQGVANLAAVCLSVGYPTIASAPHS  240

Query  241  IINGYKRVLALSVETDYTFPLAEKVKAFLADPSAFVAAAPVAAATTAAPAAAAAPAKVEA  300
            I NG+K +LA++  T+  F  A  +K ++ DPS F AAA  +AA  A  A        +
Sbjct  241  IANGFKNLLAIAATTEVEFKEATTIKEYIKDPSKFAAAASASAAPAAGGATEKKEEAKKP  300

Query  301  KEESEESDEDMGFGLFD  317
            + ESEE D+DMGFGLFD
Sbjct  301  ESESEEEDDDMGFGLFD  317
```

```
>sp|P57691.1|RLA03_ARATH  RecName: Full=60S acidic ribosomal protein P0-3
Length=323

 GENE ID: 820296 AT3G11250 | 60S acidic ribosomal protein P0-3
[Arabidopsis thaliana] (10 or fewer PubMed links)

 Score =  311 bits (798),  Expect = 1e-104, Method: Compositional matrix adjust.
 Identities = 168/321 (52%), Positives = 220/321 (69%), Gaps = 4/321 (1%)

Query  1    MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGKAVVLMGKNTM  60
            M +  +A  K  Y  K+ QL+D+Y +  +V ADNVGS Q+Q IR  LRG +VVLMGKNTM
Sbjct  1    MVKATKAEKKIAYDTKLCQLIDEYTQILVVAADNVGSTQLQNIRKGLRGDSVVLMGKNTM  60

Query  61   MRKAIRGHLEN--NPALEKLLPHIRGNVGFVFTKEDLTEIRDMLLANKVPAAARAGAIAP  118
            M++++R H EN  N A+  LLP ++GNVG +FTK DL E+ + +   KV A AR G +AP
Sbjct  61   MKRSVRIHSENSGNTAILNLLPLLQGNVGLIFTKGDLKEVSEEVAKYKVGAPARVGLVAP  120

Query  119  CEVTVPAQNTGLGPEKTSFFQALGITTKISRGTIEILSDVQLIKTGDKVGASEATLLNML  178
             +V V   NTGL P +TSFFQ L I TKI++GT+EI++ V+LIK GDKVG+SEA LL  L
Sbjct  121  IDVVVQPGNTGLDPSQTSFFQVLNIPTKINKGTVEIITPVELIKQGDKVGSSEAALLAKL  180

Query  179  NISPFSFGLVIQQVFDNGSIYNPEVLDITEETLHSRFLEGVRNVASVCLQIGYPTVASVP  238
             I PFS+GLV+Q V+DNGS+++PEVLD+TE+ L  +F  G+  V S+ L + YPT+A+ P
Sbjct  181  GIRPFSYGLVVQSVYDNGSVFSPEVLDLTEDQLVEKFASGISMVTSLALAVSYPTLAAAP  240

Query  239  HSIINGYKRVLALSVETDYTFPLAEKVKAFLADPSAFVAAAPVAAATTAAPAAAAAPAKV  298
            H  IN YK  LA++V TDYTFP AEKVK FL DPS FV AA   +A     +A A  A
Sbjct  241  HMFINAYKNALAIAVATDYTFPQAEKVKEFLKDPSKFVVAAAAVSADAGGGSAQAGAAAK  300

Query  299  EAKEESEESDEDM--GFGLFD  317
              +++ E  +ED   GFGLFD
Sbjct  301  VEEKKEESDEEDYEGGFGLFD  321
```

```
>sp|O74864.1|RLA0_SCHPO  RecName: Full=60S acidic ribosomal protein P0
Length=312

 GENE ID: 2538893 rpp0 | 60S acidic ribosomal protein Rpp0 (predicted)
[Schizosaccharomyces pombe 972h-] (10 or fewer PubMed links)

 Score =  318 bits (816),  Expect = 1e-107, Method: Compositional matrix adjust.
 Identities = 170/308 (55%), Positives = 219/308 (71%), Gaps = 3/308 (1%)

Query  10   KSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGKAVVLMGKNTMMRKAIRGHL  69
            K+ YF K+  L + Y   F+V  DNV S+QM  +R  LRG A ++MGKNTM+R+A+RG +
Sbjct  8    KAQYFEKLRSLFEKYNSLFVVNIDNVSSQQMHTVRKQLRGTAELIMGKNTMIRRAMRGII  67

Query  70   ENNPALEKLLPHIRGNVGFVFTKEDLTEIRDMLLANKVPAAARAGAIAPCEVTVPAQNTG  129
             + P LE+LLP +RGNVGFVFT  DL E+R+ ++AN + A AR  AIAP +V VPA NTG
Sbjct  68   NDMPELERLLPVVRGNVGFVFTNADLKEVRETIIANVIAAPARPNAIAPLDVFVPAGNTG  127

Query  130  LGPEKTSFFQALGITTKISRGTIEILSDVQLIKTGDKVGASEATLLNMLNISPFSFGLVI  189
            + P KTSFFQALGI TKI+RGTIEI SDV L+    KVG SEATLLNMLNISPF++G+ +
Sbjct  128  MEPGKTSFFQALGIPTKITRGTIEITSDVHLVSKDAKVGPSEATLLNMLNISPFTYGMDV  187

Query  190  QQVFDNGSIYNPEVLDITEETLHSRFLEGVRNVASVCLQIGYPTVASVPHSIINGYKRVL  249
              ++D G++++PE+LD++EE L    L     + ++ L   YPT+ SV HS++N YK ++
Sbjct  188  LTIYDQGNVFSPEILDVSEEDLIGHLLSAASIITAISLGANYPTILSVMHSVVNAYKNLV  247

Query  250  ALSVETDYTFPLAEKVKAFLADPSAFVAAAPVAAATTAAPAAAAAPAKVEAKEESEESDE  309
            A+S+ T+YTF   E+ KAFLADPSAFV A   AA   AA   A APA   A EE EESDE
Sbjct  248  AVSLATEYTFEGTEQTKAFLADPSAFVVA---AAPAAAAGGEAEAPAAEAAAEEEEESDE  304

Query  310  DMGFGLFD  317
            DMGFGLFD
Sbjct  305  DMGFGLFD  312
```

```
>sp|B6YSX9.1|RLA0_THEON  RecName: Full=Acidic ribosomal protein P0 homolog; AltName: Full=L10E
Length=339

 GENE ID: 7017837 rplP0 | acidic ribosomal protein P0
[Thermococcus onnurineus NA1] (10 or fewer PubMed links)

 Score =  131 bits (330),  Expect = 4e-38, Method: Compositional matrix adjust.
 Identities = 77/257 (30%), Positives = 130/257 (51%), Gaps = 3/257 (1%)

Query  7    ATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGKAVVLMGKNTMMRKAIR  66
            A WK     ++  ++  YP   +V   NV +  + ++R  LRGKA++ + +NT++  AI+
Sbjct  5    AEWKKKEVEELTNIIKSYPVIALVDVANVPAYPLSKMREKLRGKALLRVSRNTLIELAIK  64

Query  67   GHLEN--NPALEKLLPHIRGNVGFVFTKEDLTEIRDMLLANKVPAAARAGAIAPCEVTVP  124
               +    P LEKL+ HI+G  G + T+ +  ++  +L  +K PA A+ G   P +V +P
Sbjct  65   RAAQELGKPELEKLIDHIQGGAGILATEMNPFKLYKLLEESKTPAPAKPGVPVPRDVVIP  124

Query  125  AQNTGLGPEK-TSFFQALGITTKISRGTIEILSDVQLIKTGDKVGASEATLLNMLNISPF  183
            A  T + P       QALGI  +I +G + I  D  ++K G+ +    A +LN L I P
Sbjct  125  AGPTSISPGPLVGEMQALGIPARIEKGKVSIQKDYTVLKAGEVITEQLARILNALGIEPL  184

Query  184  SFGLVIQQVFDNGSIYNPEVLDITEETLHSRFLEGVRNVASVCLQIGYPTVASVPHSIIN  243
              GL +   +++G +Y PEVL I EE   +   +   +  ++ +   YPT  ++   I
Sbjct  185  EVGLNLLAAYEDGIVYTPEVLAIDEEEYINLLQQAYMHAFNLSVNTAYPTSQTIEAIIQK  244

Query  244  GYKRVLALSVETDYTFP  260
             Y     ++VE  Y  P
Sbjct  245  AYLGAKNVAVEAGYITP  261
```

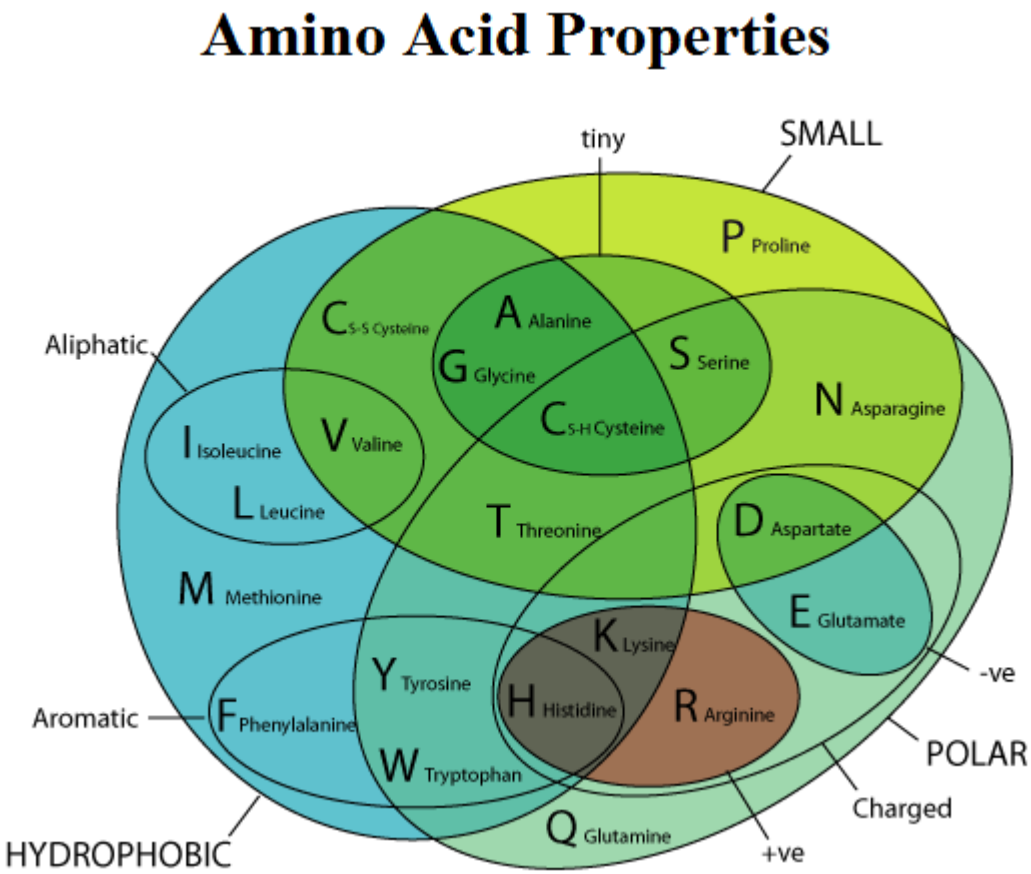
Pourquoi la recherche protéique est-elle plus sensible?

le code génétique

Première lettre (côté 5')	Deuxième lettre: U		C		A		G		Troisième lettre (côté 3')
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop	A
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

codon d'initiation — codon de terminaison

Deuxième raison (plus importante):



From Livingstone, C. D. and Barton, G. J. (1993), "Protein Sequence Alignments: A Strategy for the Hierarchical Analysis of Residue Conservation", Comp. Appl. Bio. Sci., 9, 745-756.

-> nécessité de capturer ces propriétés dans un score

Alignement multiple de la protéine ribosomale L0

```
                                   *   .     :     .         .   *       : : :     .
Q5E940_BOVIN -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE    76
  RLA0_HUMAN -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE    76
  RLA0_MOUSE -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE    76
    RLA0_RAT -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE    76
  RLA0_CHICK -----------MPREDRATWKSNYFMKIIQLLDDYPKCFVVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE    76
  RLA0_RANSY -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQQIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--SALE    76
Q7ZUG3_BRARE -----------MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKQMQTIRLSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE    76
  RLA0_ICTPU -----------MPREDRATWKSNYFLKIIQLLNDYPKCFIVGADNVGSKQMQTIRLSLRGK-AIVLMGKNTMMRKAIRGHLENN--PALE    76
  RLA0_DROME -----------MVRENKAAWKAQYFIKVVELFDEFPKCFIVGADNVGSKQMQNIRTSLRGL-AVVLMGKNTMMRKAIRGHLENN--PQLE    76
  RLA0_DICDI -----------MSGAG-SKRKKLFIEKATKLFTTYDKMIVAEADFVGSSQLQKIRKSIRGI-GAVLMGKKTMIRKVIRDLADSK--PELD    75
Q54LP0_DICDI -----------MSGAG-SKRKNVFIEKATKLFTTYDKMIVAEADFVGSSQLQKIRKSIRGI-GAVLMGKKTMIRKVIRDLADSK--PELD    75
  RLA0_PLAF8 -----------MAKLSKQQKKQMYIEKLSSLIQQYSKILIVHVDNVGSNQMASVRKSLRGK-ATILMGKNTRIRTALKKNLQAV--PQIE    76
  RLA0_SULAC -----MIGLAVTTTKKIAKWKVDEVAELTEKLKTHKTIIIANIEGFPADKLHEIRKKLRGK-ADIKVTKNNLFNIALKNAG-----YDTK    79
  RLA0_SULTO ----MRIMAVITQERKIAKWKIEEVKELEQKLREYHTIIIANIEGFPADKLHDIRKKMRGM-AEIKVTKNTLFGIAAKNAG-----LDVS    80
  RLA0_SULSO ----MKRLALALKQRKVASWKLEEVKELTELIKNSNTILIGNLEGFPADKLHEIRKKLRGK-ATIKVTKNTLFKIAAKNAG-----IDIE    80
  RLA0_AERPE MSVVSLVGQMYKREKPIPEWKTLMLRELEELFSKHRVVLFADLTGTPTFVVQRVRKKLWKK-YPMMVAKKRIILRAMKAAGLE---LDDN    86
  RLA0_PYRAE -MMLAIGKRRYVRTRQYPARKVKIVSEATELLQKYPYVFLFDLHGLSSRILHEYRYRLRRY-GVIKIIKPTLFKIAFTKVYGG---IPAE    85
  RLA0_METAC ------MAEERHHTEHIPQWKKDEIENIKELIQSHKVFGMVGIEGILATKMQKIRRDLKDV-AVLKVSRNTLTERALNQLG-----ETIP    78
  RLA0_METMA ------MAEERHHTEHIPQWKKDEIENIKELIQSHKVFGMVRIEGILATKIQKIRRDLKDV-AVLKVSRNTLTERALNQLG-----ESIP    78
  RLA0_ARCFU ------MAAVRGS---PPEYKVRAVEEIKRMISSKPVVAIVSFRNVPAGQMQKIRREFRGK-AEIKVVKNTLLERALDALG-----GDYL    75
  RLA0_METKA MAVKAKGQPPSGYEPKVAEWKRREVKELKELMDEYENVGLVDLEGIPAPQLQEIRAKLRERDTIIRMSRNTLMRIALEEKLDER--PELE    88
  RLA0_METTH -------------MAHVAEWKKKEVQELHDLIKGYEVVGIANLADIPARQLQKMRQTLRDS-ALIRMSKKTLISLALEKAGREL--ENVD    74
  RLA0_METTL -------MITAESEHKIAPWKIEEVNKLKELLKNGQIVALVDMMEVPARQLQEIRDKIR-GTMTLKMSRNTLIERAIKEVAEETGNPEFA    82
  RLA0_METVA -------MIDAKSEHKIAPWKIEEVNALKELLKSANVIALIDMMEVPAVQLQEIRDKIR-DQMTLKMSRNTLIKRAVEEVAEETGNPEFA    82
  RLA0_METJA --------METKVKAHVAPWKIEEVKTLKGLIKSKPVVAIVDMMDVPAPQLQEIRDKIR-DKVKLRMSRNTLIIRALKEAAEELNNPKLA    81
  RLA0_PYRAB -------------MAHVAEWKKKEVEELANLIKSYPVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSRNTLIELAIKKAAQELGKPELE    77
  RLA0_PYRHO -------------MAHVAEWKKKEVEELAKLIKSYPVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSRNTLIELAIKKAAKELGKPELE    77
  RLA0_PYRFU -------------MAHVAEWKKKEVEELANLIKSYPVVALVDVSSMPAYPLSQMRRLIRENNGLLRVSRNTLIELAIKKVAQELGKPELE    77
  RLA0_PYRKO -------------MAHVAEWKKKEVEELANIIKSYPVIALVDVAGVPAYPLSKMRDKLR-GKALLRVSRNTLIELAIKRAAQELGQPELE    76
  RLA0_HALMA -----MSAESERKTETIPEWKQEEVDAIVEMIESYESVGVVNIAGIPSRQLQDMRRDLHGT-AELRVSRNTLLERALDDVD-----DGLE    79
  RLA0_HALVO -----MSESEVRQTEVIPQWKREEVDELVDFIESYESVGVVGVAGIPSRQLQSMRRELHGS-AAVRMSRNTLVNRALDEVN-----DGFE    79
  RLA0_HALSA -----MSAEEQRTTEEVPEWKRQEVAELVDLLETYDSVGVVNVTGIPSKQLQDMRRGLHGQ-AALRMSRNTLLVRALEEAG-----DGLD    79
  RLA0_THEAC -------------MKEVSQQKKELVNEITQRIKASRSVAIVDTAGIRTRQIQDIRGKNRGK-INLKVIKKTLLFKALENLGD----EKLS    72
  RLA0_THEVO -------------MRKINPKKKEIVSELAQDITKSKAVAIVDIKGVRTRQMQDIRAKNRDK-VKIKVVKKTLLFKALDSIND----EKLT    72
  RLA0_PICTO -------------MTEPAQWKIDFVKNLENEINSRKVAAIVSIKGLRNNEFQKIRNSIRDK-ARIKVSRARLLRLAIENTGK----NNIV    72
       ruler 1.......10........20........30........40........50........60........70........80........90
```