

Recherche *in silico* de protéines mitochondriales

Analyse du protéome mitochondrial à travers l'évolution et les pathologies



La mitochondrie humaine moderne

- Nombreux **métabolismes** (production énergétique, biosynthèses)
- **Mort cellulaire** programmée
- Des **milliers** par cellule
- Impliquées dans un grand nombre de **processus pathologiques** (cancer, diabète, obésité, maladies génétiques et neurodégénératives)

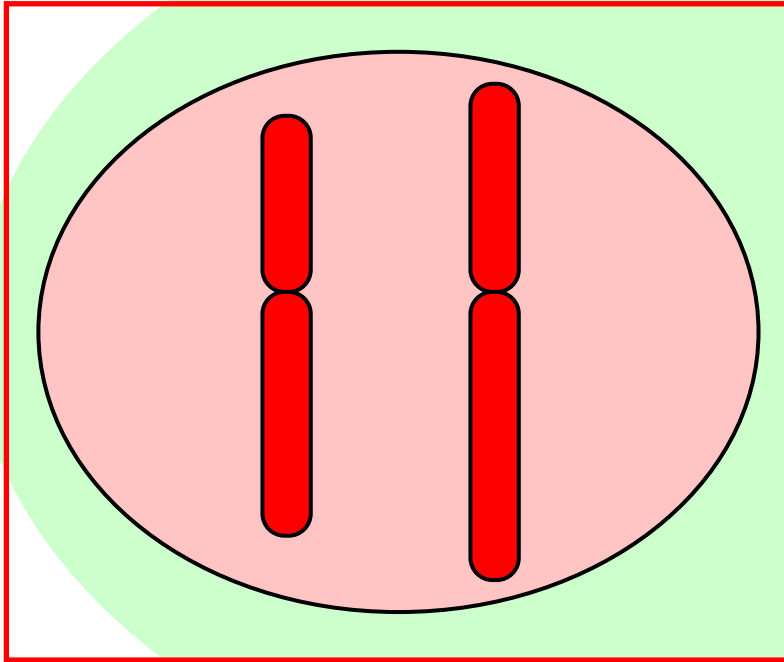


Contexte

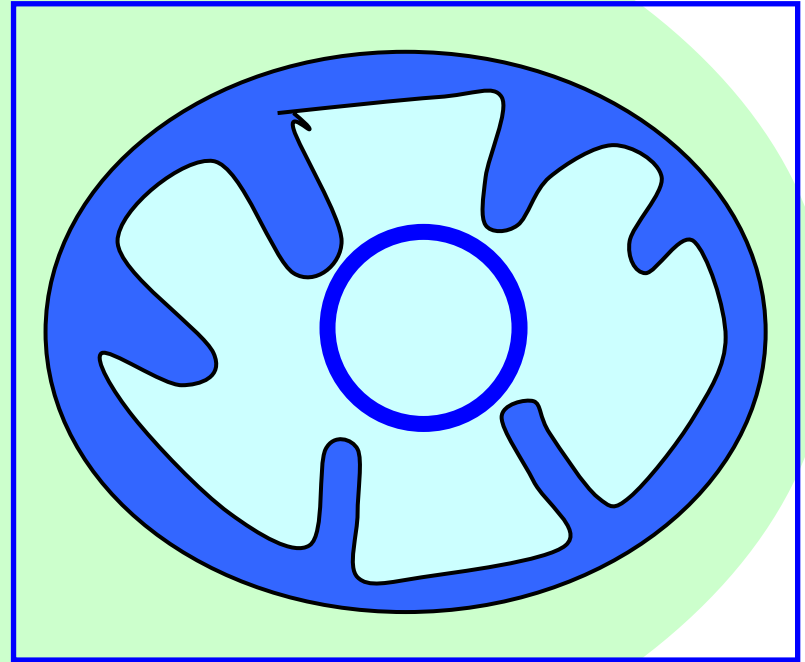
L'origine du « génome mitochondrial nucléaire »



Evolution



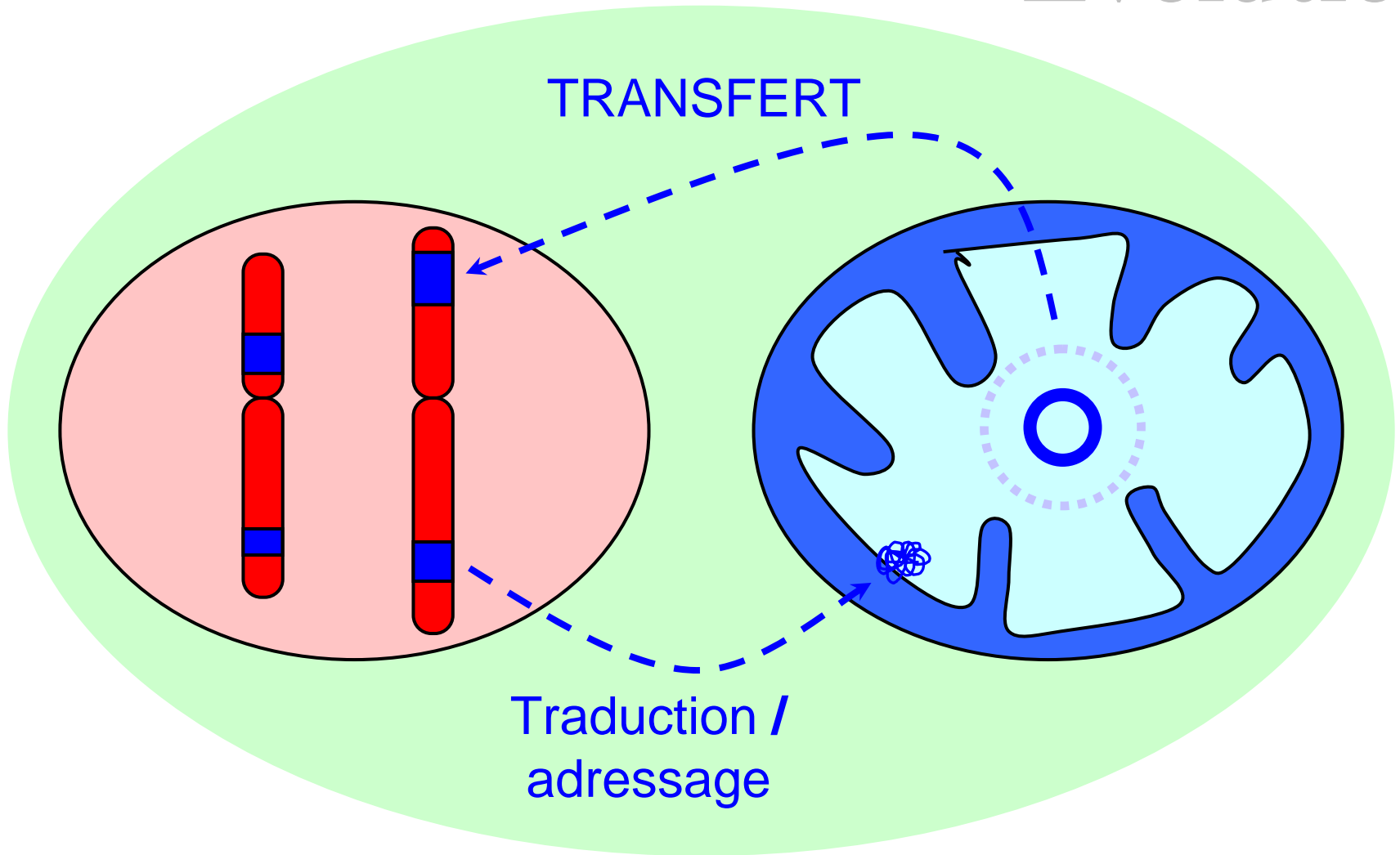
Noyau eucaryote primitif



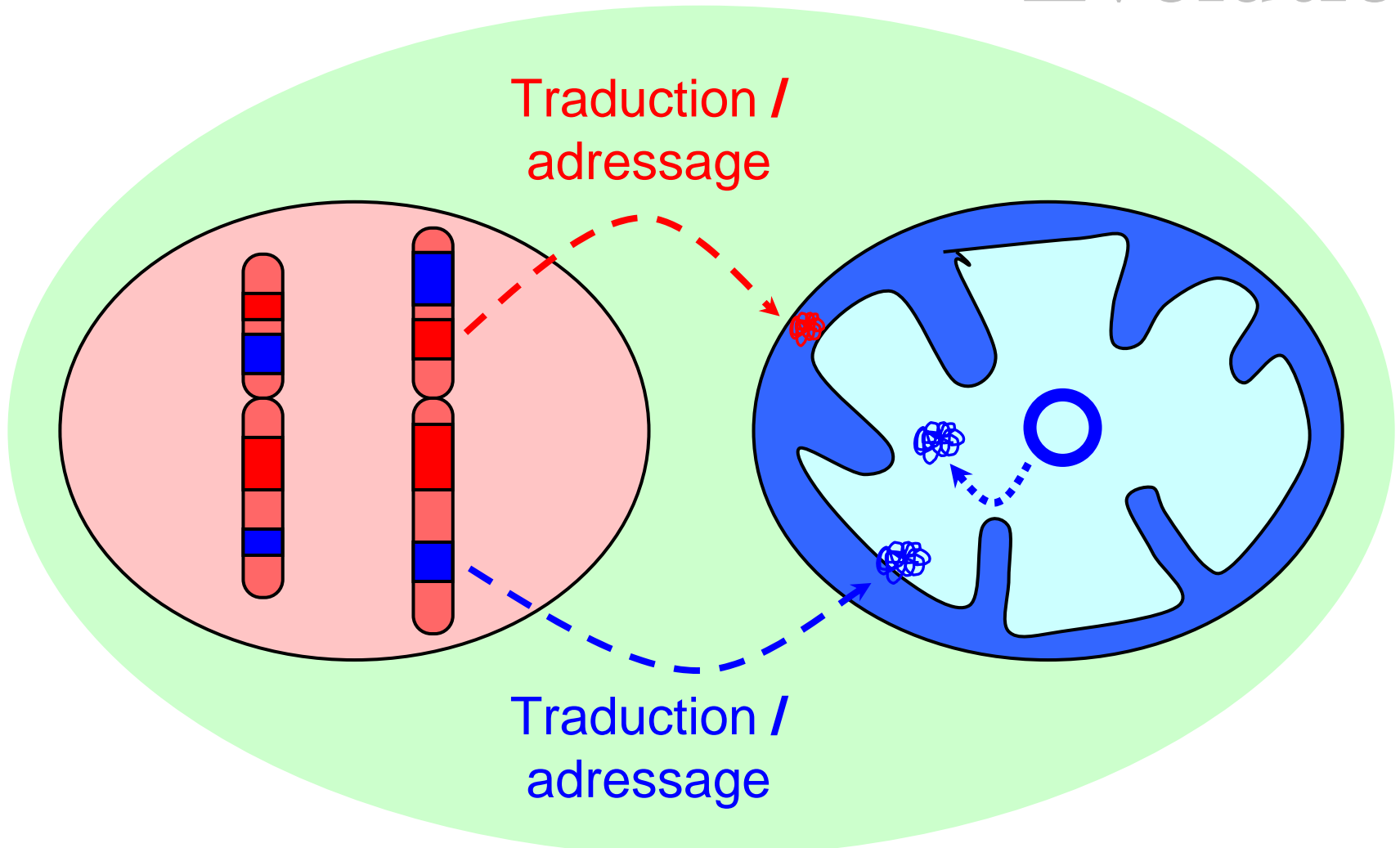
Mitochondrie
(origine endosymbiotique)



Evolution



Evolution

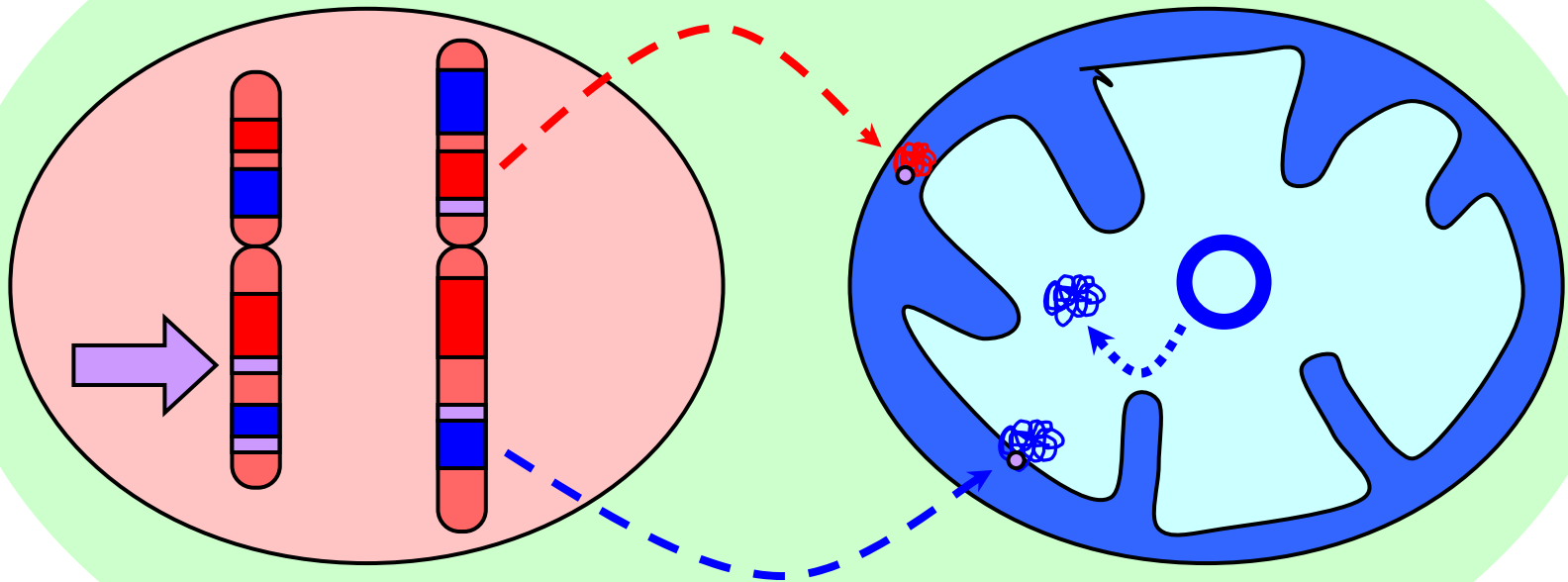


Double origine du « génome mitochondrial nucléaire »



Evolution

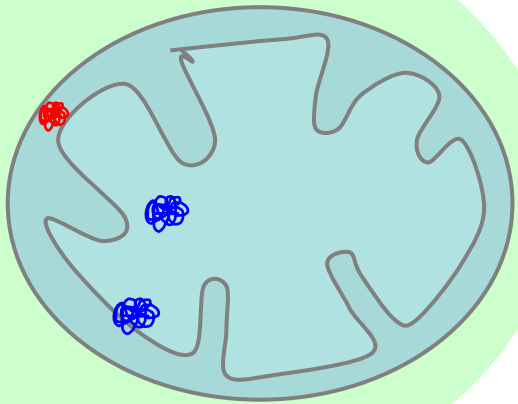
Adressage mitochondrial



Double origine du « génome mitochondrial nucléaire »



Le protéome mitochondrial



- Entre **1000** et **2000** protéines mitochondriales (13 codées par le génome mitochondrial)
- Actuellement, ~ **500 connues**
- Ces protéines ont une **double origine évolutive**:
 - origine de la cellule hôte **eucaryote** primitive
 - origine **procaryote** suite à une endosymbiose ancienne (*Rickettsia prowazekii* plus proche parent connu)

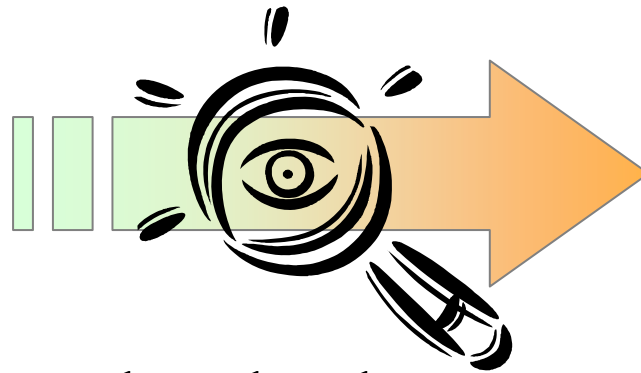
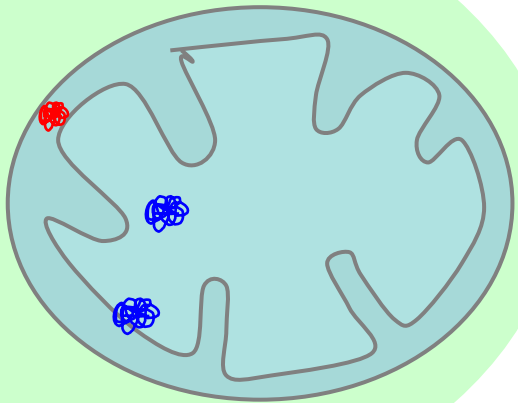


Travail réalisé

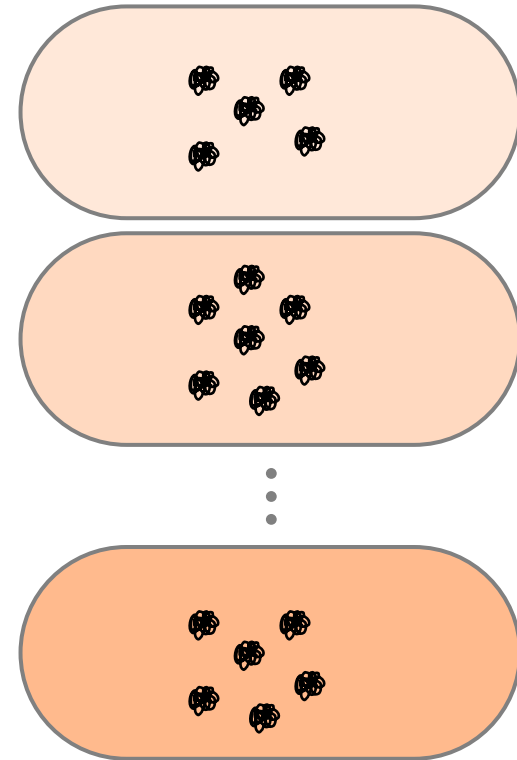
*Etude de l'origine évolutive du
protéome mitochondrial*



Recherche d'homologies



Recherche des
protéines homologues



Protéines mitochondriales humaines
classées par **fonction** et **localisation** dans la
mitochondrie (fév. 2004)

Protéomes bactériens
(fév. 2004)



Point de vue informatique

```
>sp|Q9UQ90|SPG7_HUMAN  
MAVLLLLLLRALGRGPGGPRP  
>sp|Q16595|FRDA_HUMAN  
MWTLGRRVAVAGLLASPSPAQA  
>sp|075027|ABCB7_HUMAN  
MALLAMHSWRWAAAAAAFEKRR  
>sp|060313|OPA1_HUMAN  
MWRLRRAAVACEVCOSLVKHSS  
>sp|P12235|ADT1_HUMAN  
GDHAWSFLKDFLAGGVAAAVSK  
>sp|P54098|DPOG1_HUMAN  
MSRLLWRKVAGATVGGPVPAP
```



Comparaison

- ✓ E-value < 1.10^{-4}
- ✓ Identité > 19%
- ✓ PRDF pour les identités < 23%

```
>protei n1 [bact1]  
MDLKZKLSJKJW  
>protei n2 [bact1]  
MHKJHFDYQUIDLQNLJK  
>protei n3 [bact1]  
MHKJHFDYQUIDLQNLJK  
>protei n4 [bact1]  
MHKJHFDYQUIDLQNLJK  
>protei n1 [bact2]  
MHKJHFDYQUIDLQNLJK  
>protei n2 [bact2]  
MHKJHFDYQUIDLQNLJK  
>protei n1 [bact3]  
MHKJHFDYQUIDLQNLJK
```

393 protéines

avec **fonction** et **localisation**

[Swiss-Prot, fév. 2004]

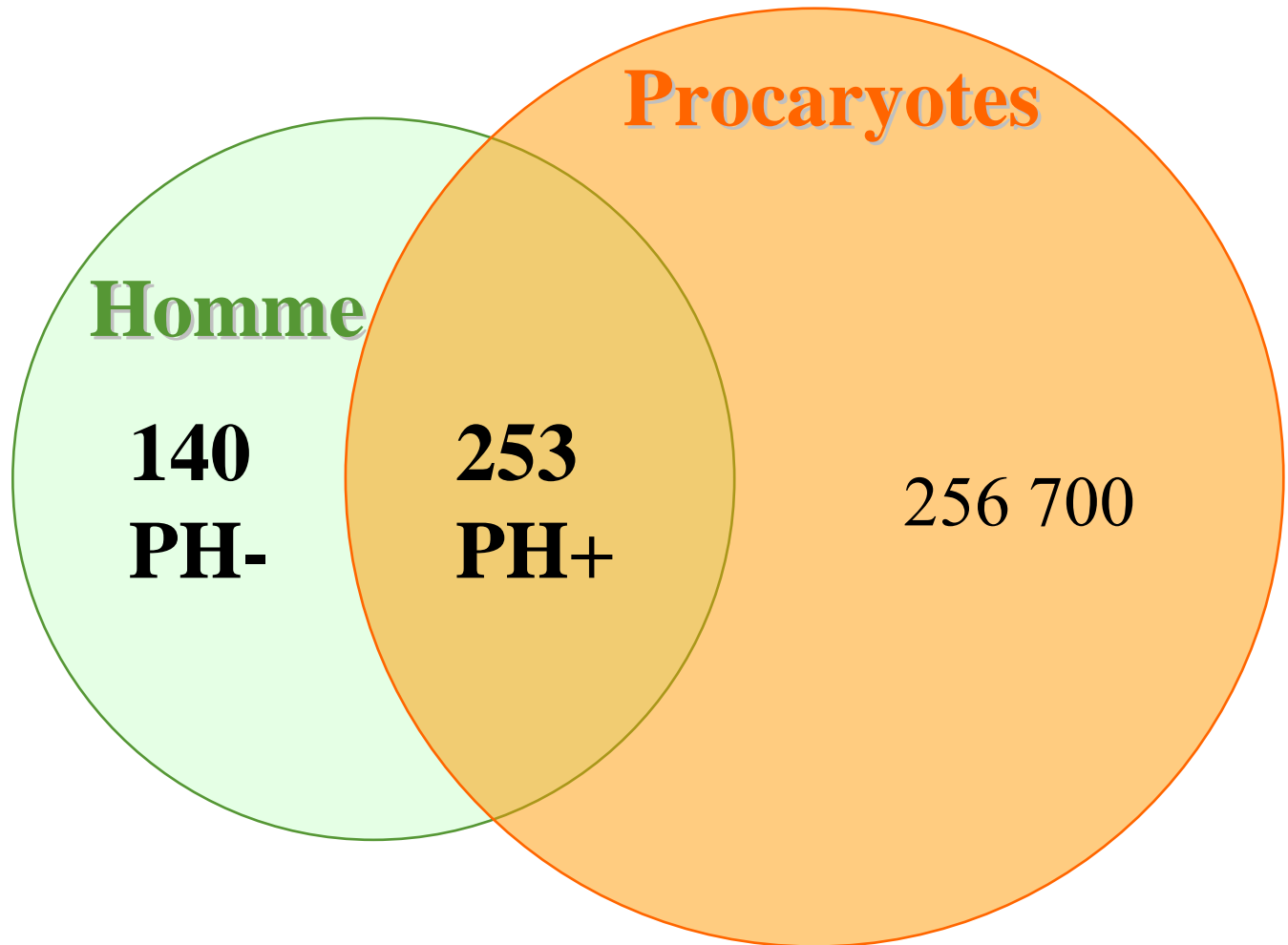
94 protéomes bactériens,

256 953 protéines

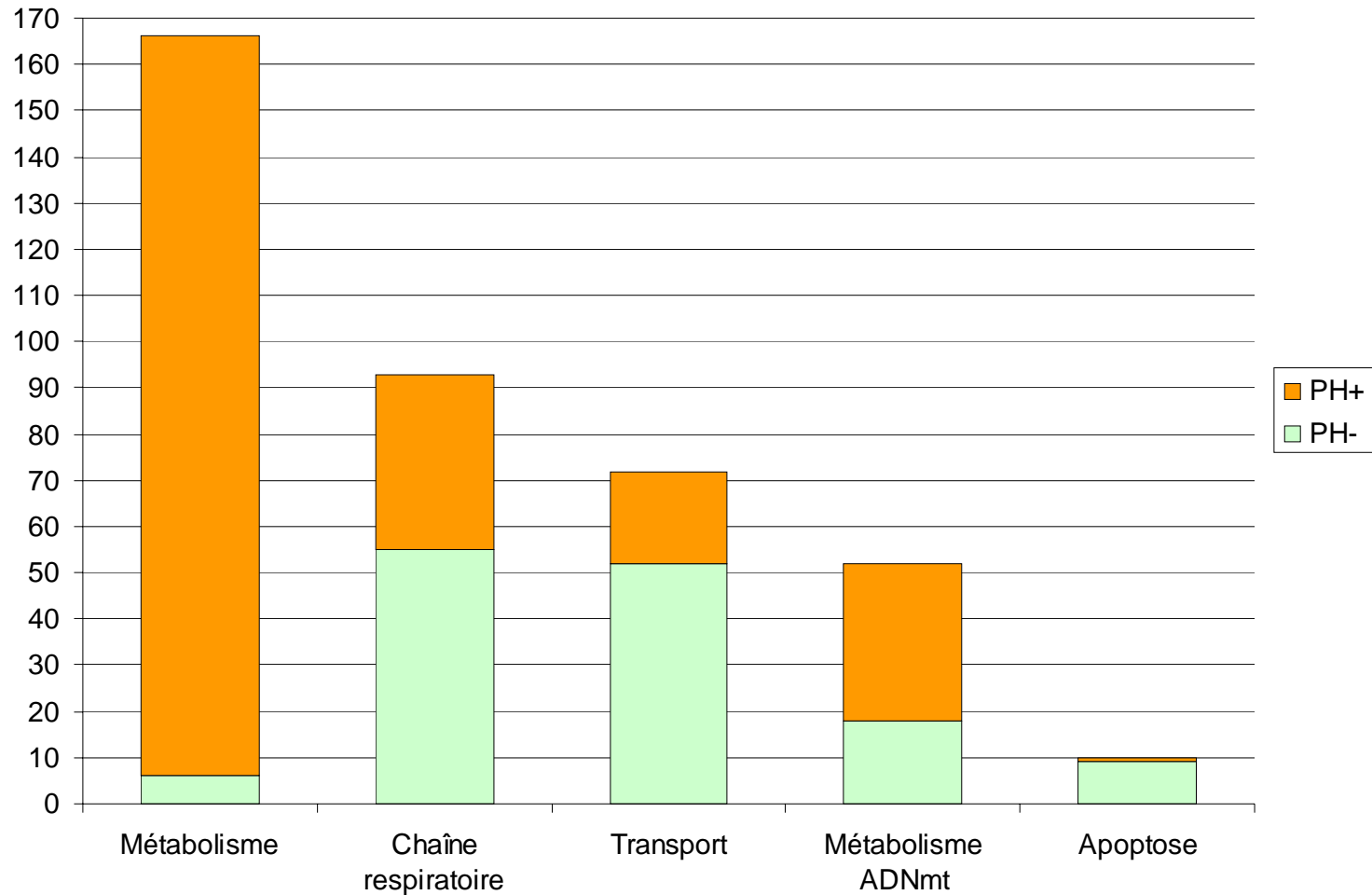
[NCBI, fév. 2004]



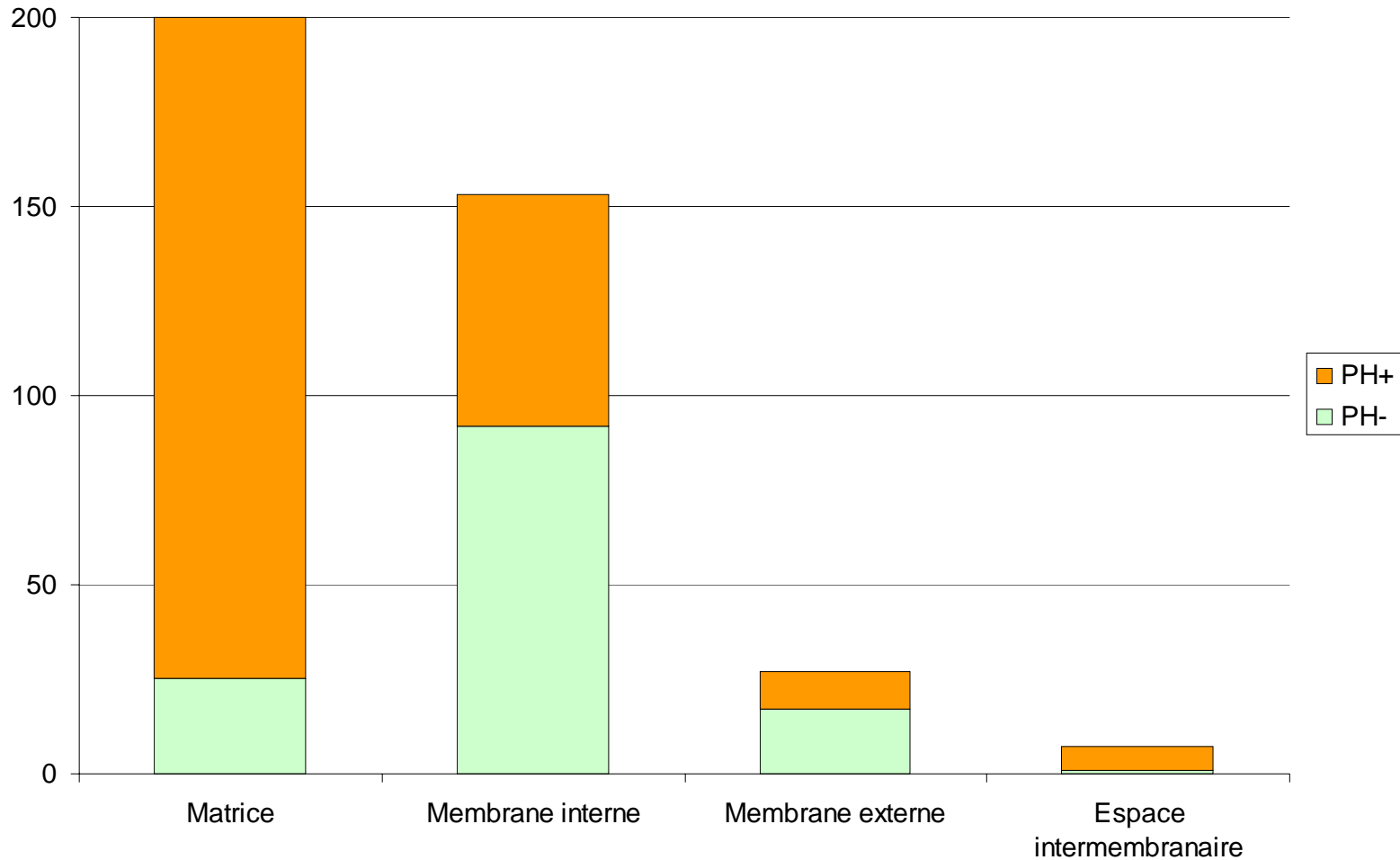
Etude des protéines homologues



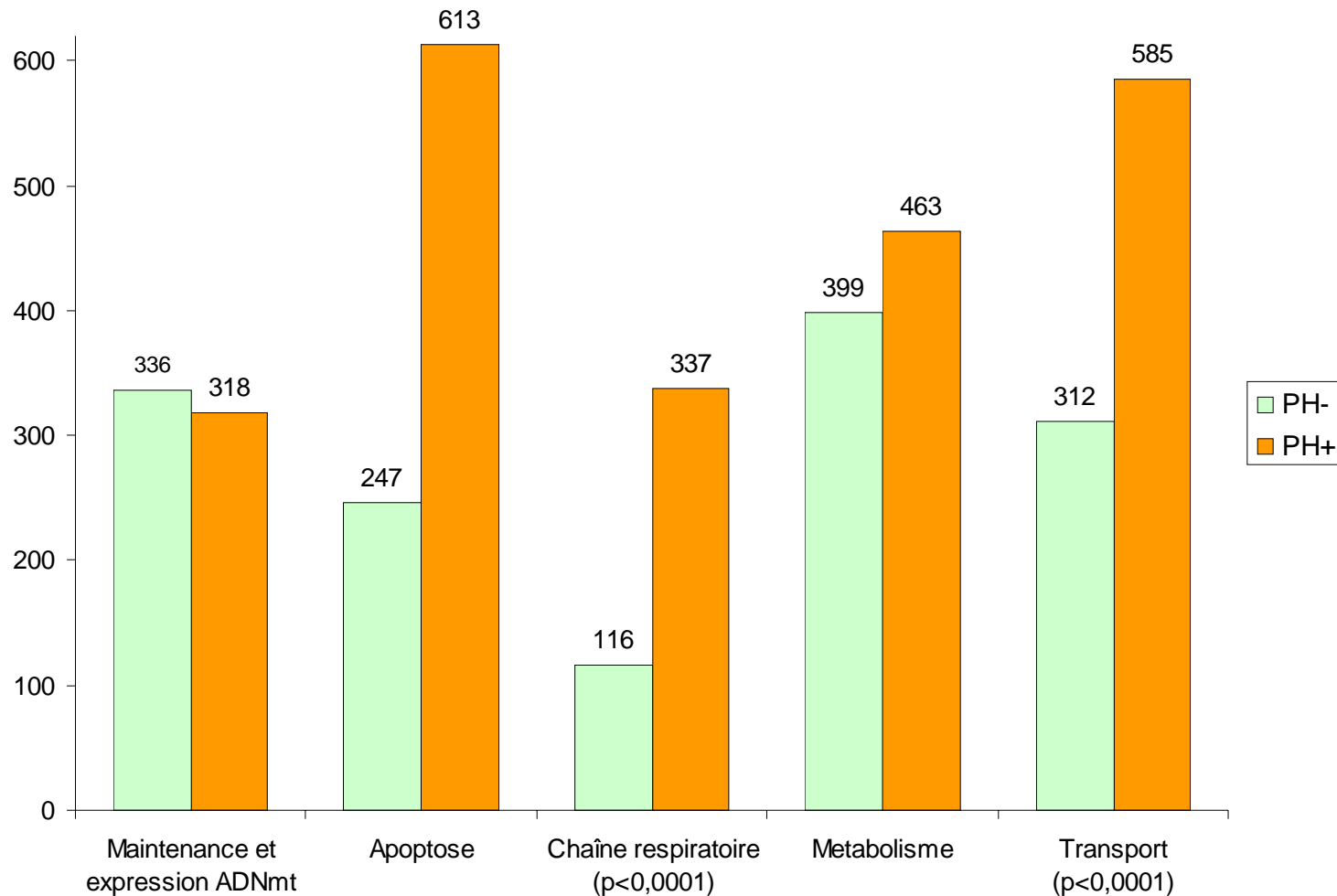
Fonction des protéines mitochondriales



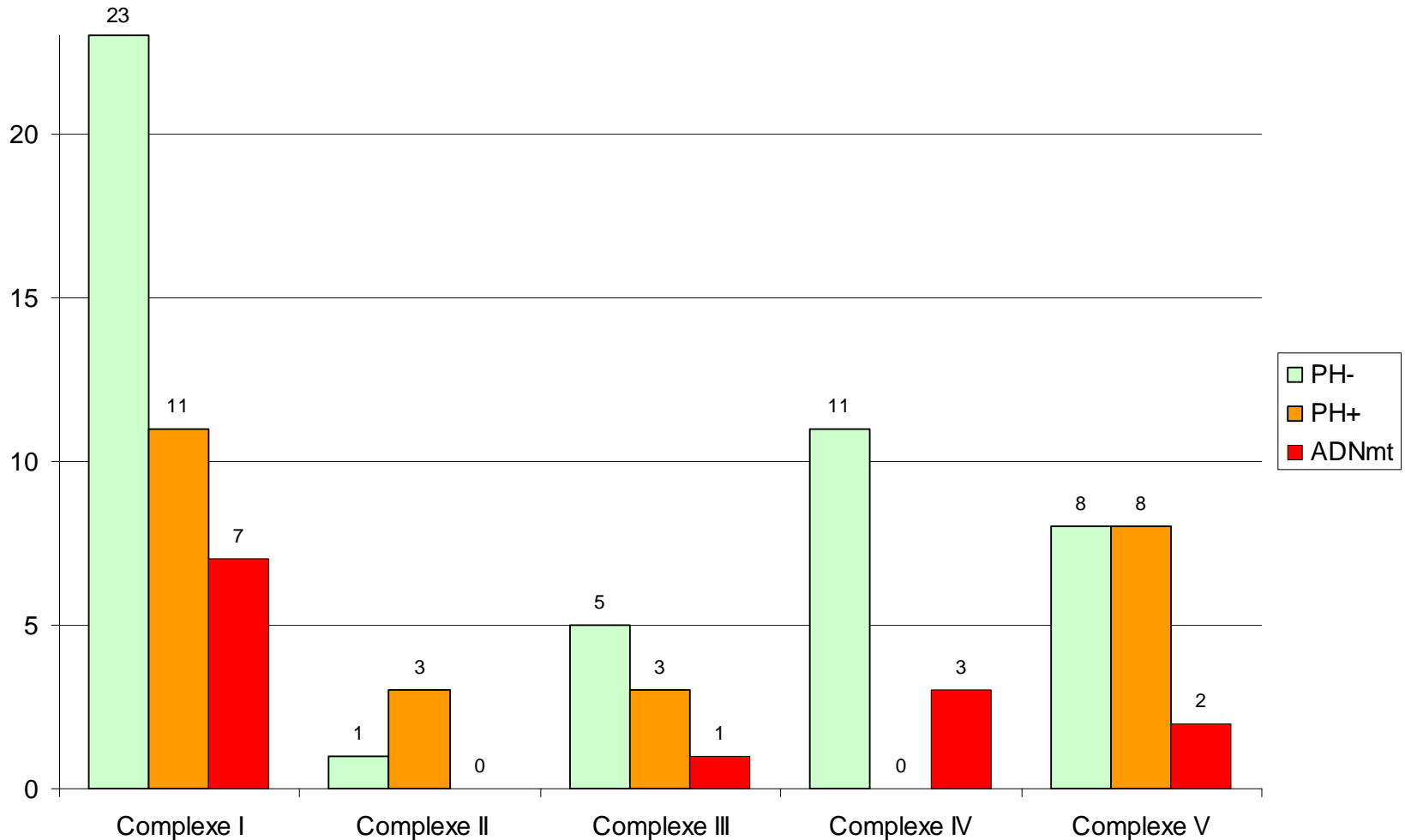
Sous-localisation des protéines mitochondriales



Longueur des protéines vs origine



Protéines de la chaîne respiratoire: nombre vs origine



Les mitochondriopathies impliquent majoritairement des protéines PH+

Chaîne respiratoire et cycle de Krebs

<i>Gene</i>	<i>OMIM</i>	<i>PH</i>
NDUFS1 (Complex I)	157 655	+
NDUFS2 (Complex I)	602 985	+
NDUFS3 (Complex I)	603 846	+
NDUFS4 (Complex I)	602 694	+
NDUFS7 (Complex I)	601 825	+
NDUFS8 (Complex I)	602 141	+
NDUFV1 (Complex I)	161 015	+
NDUFV2 (Complex I)	600 532	+
SDHA (Complex II and Krebs cycle)	600 857	+
SDHB (Complex II and Krebs cycle)	115 310	+
SDHC (Complex II and Krebs cycle)	605 373	+
SDHD (Complex II and Krebs cycle)	168 000	-
HUMQPC (Complex III)	191 330	-
BCS1L (Complex III assembly factor)	603 647	+
SURF1 (Complex IV assembly factor)	185 620	+
SCO1 (Complex IV assembly factor)	603 644	+
SCO2 (Complex IV assembly factor)	604 377	+
COX10 (Complex IV assembly factor)	602 125	+
COX15 (Complex IV assembly factor)	603 646	+
FH (Krebs cycle)	150 800	+

Maintenance et expression de l'ADN mitochondrial

<i>Gene</i>	<i>OMIM</i>	<i>PH</i>
TP (thymidine phosphorylase)	603 041	+
DGUOK (deoxyguanosine kinase)	251 880	+
TWINKLE (DNA helicase)	157 640	+
TK2 (thymidine kinase 2)	251 880	+
DNC (deoxynucleotide carrier)	607 196	-
ANT1 (adenine nucleotide translocator 1)	157 640	-
POLG1 (DNA polymerase γ)	157 640	-

Autres fonctions mitochondriales

<i>Gene</i>	<i>OMIM</i>	<i>PH</i>
SPG7 (paraplegin)	607 259	+
FRDA (frataxin)	229 300	+
ABC7 (ABC transporter)	301 310	+
DDP1 (deafness-dystonia protein)	304 700	-
OPA1 (optic atrophy 1)	165 500	-

OMIM, Online Mendelian Inheritance in Man
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>.



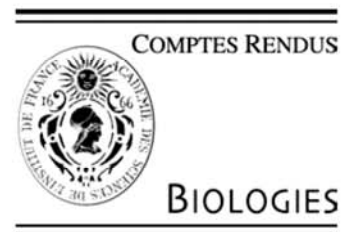
Pour plus d'informations



Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

C. R. Biologies 327 (2004) 1095–1101



<http://france.elsevier.com/direct/CRASS3/>

Genetics / Génétique

Mitochondrial diseases preferentially involve proteins with prokaryote homologues

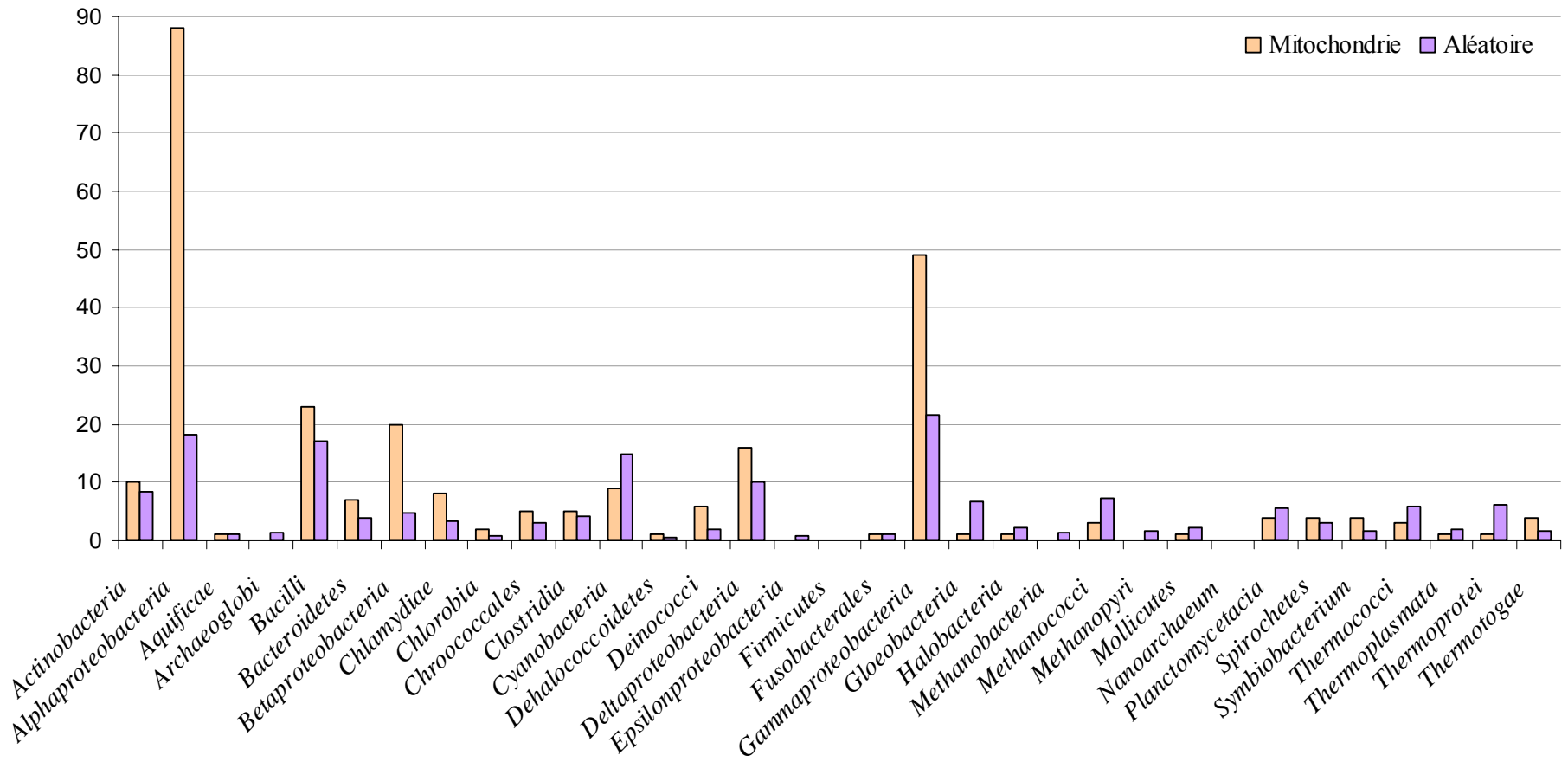
Yves Tourmen^{a,*}, Marc Ferré^a, Yves Malthièry^a, Philippe Dessen^b, Pascal Reynier^a

^a *Laboratoire de biochimie et biologie moléculaire, INSERM E0018, CHU d'Angers, 4, rue Larrey, 49033 Angers, France*

^b *Laboratoire de génétique oncologique, UMR 8125 CNRS, Institut Gustave-Roussy, 94805 Villejuif, France*

Received 5 July 2004; accepted 8 September 2004

Origine du protéome mitochondrial



Projet

*Analyse de nouvelles protéines
mitochondriales par génomique
comparative*



Identification de protéines mitochondriales

Peptide
N-terminal
supplémentaire
d'origine
eucaryote

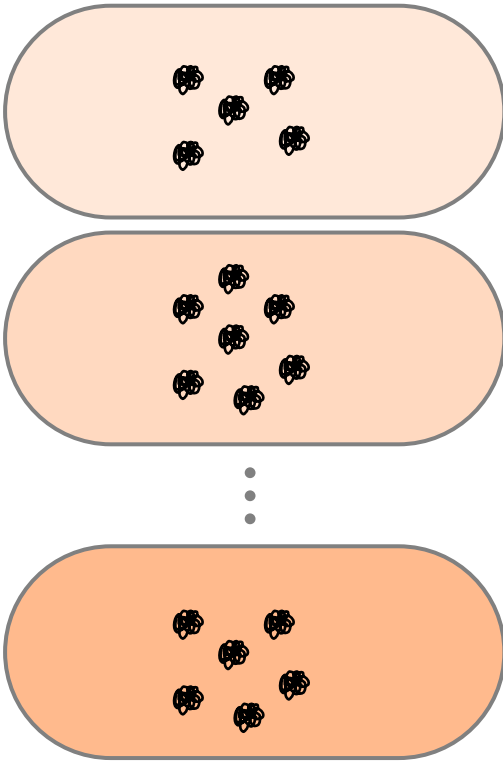
Protéines mitochondriales PH+

Homologue procaryote

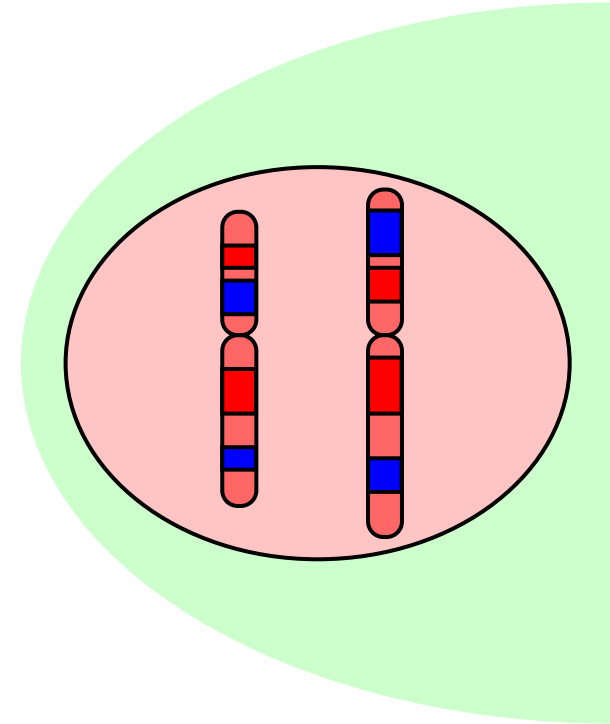
88% des protéines **PH+ plus grandes** (435 vs 400 aa, $p < 0,0001$)

Hypothèse: L'identification de cette **structure chimérique** doit permettre de **découvrir de nouvelles protéines mitochondriales**

Recherche de nouvelles protéines mitochondriales



Recherche des
gènes homologues



Protéomes bactériens
(300 complets)

Génomes eucaryotes
(39 complets)



Point de vue informatique

```
>prote in1 [bact1]
MDLKZKLSJKJW
>prote in2 [bact1]
MHKJHFDYQUDLQNLJK
>prote in3 [bact1]
MHKJHFDYQUDLQNLJK
>prote in4 [bact1]
MHKJHFDYQUDLQNLJK
>prote in1 [bact2]
MHKJHFDYQUDLQNLJK
>prote in2 [bact2]
MHKJHFDYQUDLQNLJK
```

400 000 protéines
($150 \cdot 10^6$ aa)



Comparaison

Filtrage:

- ✓ Distance entre 2 HSP, colinéarité
- ✓ Couverture
- ✓ Présence d'un adressage

Validation:

- ✓ MitoProt / TargetP
- ✓ Publications
- ✓ Étude phylogénétique

```
>chr1
AGGTCAATAAGGTCAAT
attatctagcaGGTCAA
TAAGGTCAATAattatc
tagcaAGGTCAATAAGG
TCAATAattatctagca
AGGTCAATAAGGTCAAT
AattatctagcaGGTCA
ATAAGGTCAATattatc
tagcaGGTCAATAAGGT
CAATAattatctagcaA
GGTCAATAAGGTCAATA
attatctagcaAGGTCA
ATAAGGTCAATAattat
ctagcaattatctagca
```

Génomes eucaryotes

Hs ($3,5 \cdot 10^9$ b / $30 \cdot 10^3$ gènes),
Mm, *Dm*, *Ce*, ...



Problème



Estimation du temps de calcul sur machine locale (SunFire 280R, bi-processeur/4Go RAM)

- 39469 prot. alphaproteo. vs chr. 19 humain:
durée **26 heures**
- Extrapolation pour l'ensemble du génome humain:
 $50 \times 26h = 54$ jours
Pour toutes les bactéries: $54 \text{ jours} \times 10 =$ **540 jours**

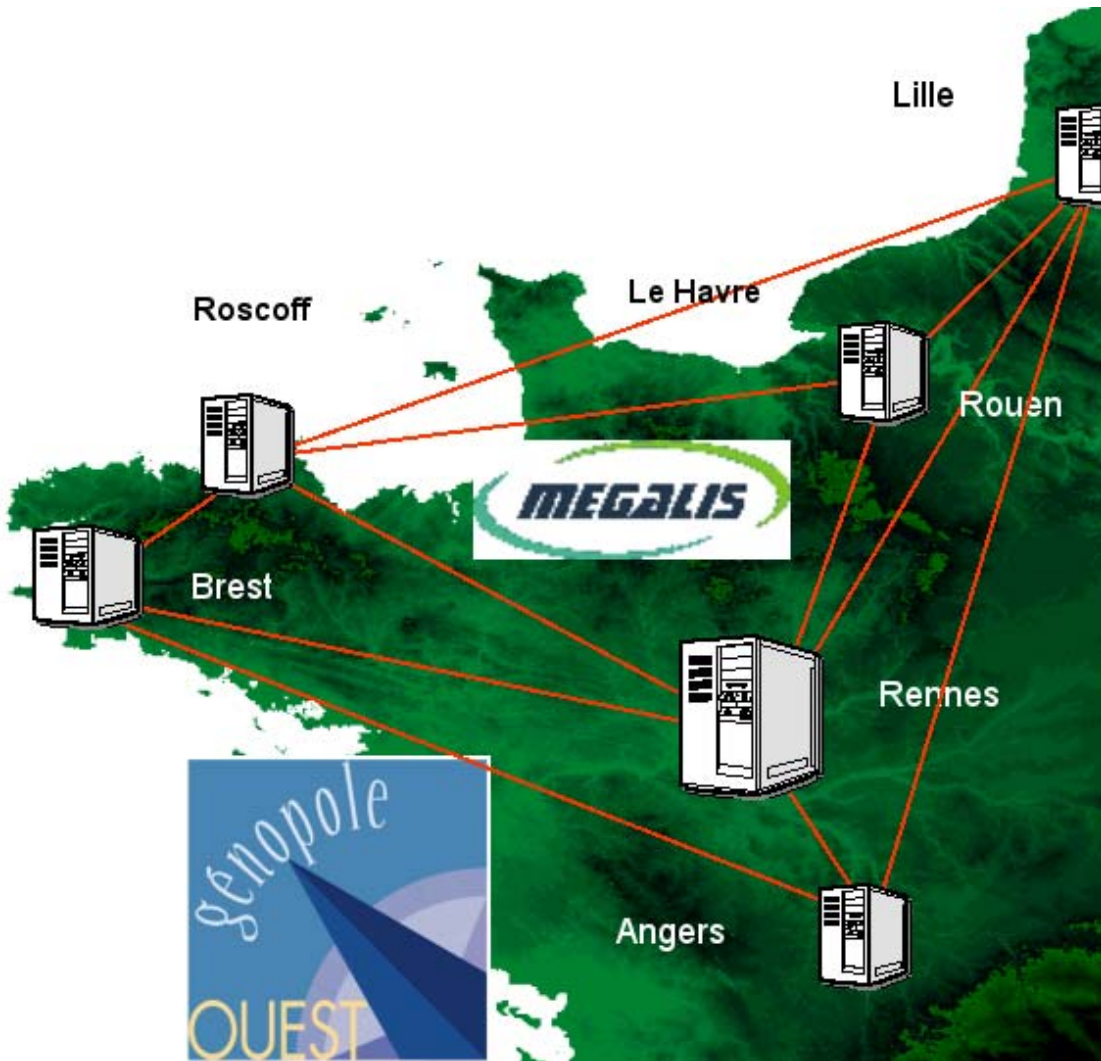


Application mitopid

GénoGRID / DEISA / ReMIX



GénoGRID



- ACI GRID 2001
- La **gridification** de l'algorithme TBLASTN ne pose **pas de problème particulier**
- Problème d'**accès distribué aux bases de données** sur la grille

mitopid sur GénoGRID

Banque protéique

```
>protei n1 [bact1]
ADLKZKLSJKJW
>protei n2 [bact1]
AHKJHFDYQUIDLQNLJK
>protei n3 [bact1]
AHKJHFDYQUIDLQNLJK
>protei n4 [bact1]
AHKJHFDYQUIDLQNLJK
>protei n1 [bact2]
AHKJHFDYQUIDLQNLJK
>protei n2 [bact2]
AHKJHFDYQUIDLQNLJK
```

```
>protei n1 [bact1]
MDLKZKLSJKJW
>protei n2 [bact1]
MHKJHFDYQUIDLQNLJK
```

```
>protei n3 [bact1]
MDLKZKLSJKJW
>protei n4 [bact1]
MHKJHFDYQUIDLQNLJK
```

```
>protei n1 [bact2]
ADLKZKLSJKJW
>protei n2 [bact2]
MHKJHFDYQUIDLQNLJK
```

Génome

```
>chr1
AGGTCAATAAGGTCAAT
attatctagcaGGTCAA
TAAGGTCAATAattatc
tagcaAGGTCAATAAGG
TCAATAattatctagca
AGGTCAATAAGGTCAAT
AattatctagcaGGTCA
ATAAGGTCAATattatc
tagcaGGTCAATAAGGT
CAATAattatctagcaA
GGTCAATAAGGTCAATA
attatctagcaAGGTCA
ATAAGGTCAATAattat
ctagcaattatctagca
```

Tâche principale:

découpe la banque protéique

Tâches élémentaires:

TBLASTN avec **filtrage**



DEISA



- **4 000 processeurs / 24 Teraflops**
- **125 armoires dans 3 pays**
(Allemagne, France, Italie)
- **IBM-SP4 , 12 nœuds SMP P690+**



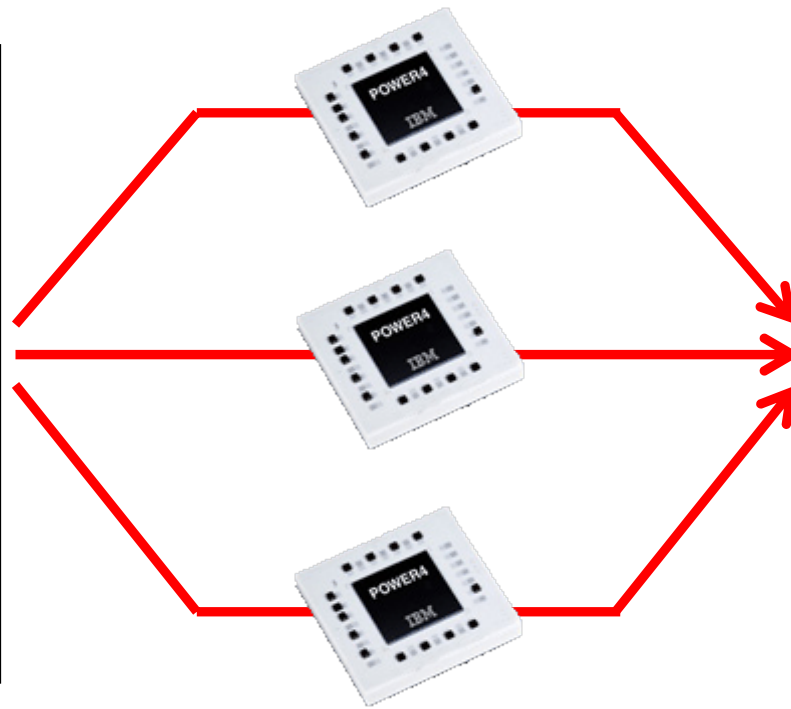
- **32 processeurs / RAM 128 Go**
- **Calcul préliminaire: 8 jours**
- **Temps nécessaire extrapolé:**
10 jours



mitopid sur DEISA

Banque protéique

```
>protei n1 [bact1]
ADLKZKLSJKJW
>protei n2 [bact1]
AHKJHFDYQUDLQNLJK
>protei n3 [bact1]
AHKJHFDYQUDLQNLJK
>protei n4 [bact1]
AHKJHFDYQUDLQNLJK
>protei n1 [bact2]
AHKJHFDYQUDLQNLJK
>protei n2 [bact2]
AHKJHFDYQUDLQNLJK
```

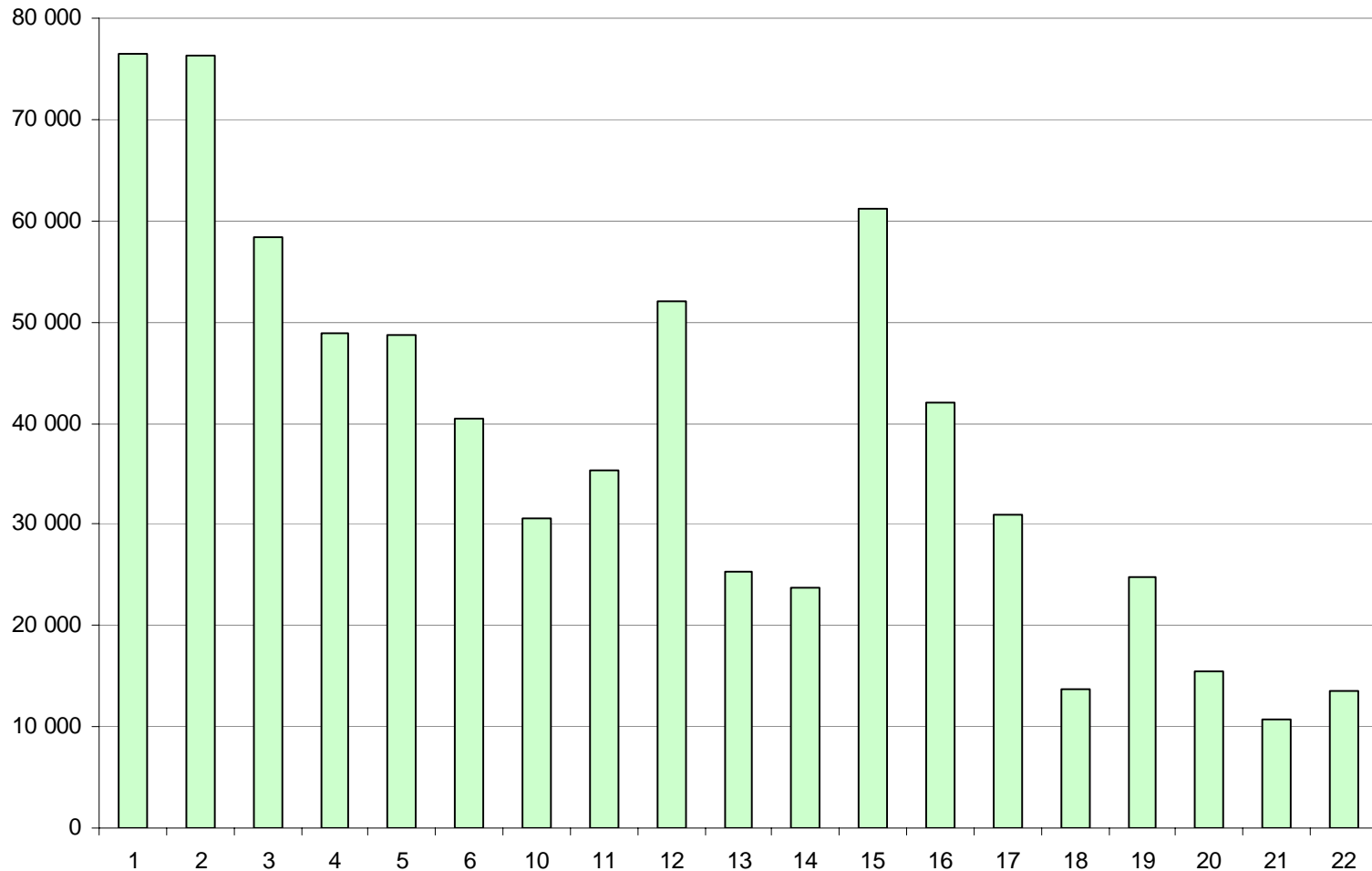


```
>chr1
AGGTCAATAAGGTCAAT
attatctagcaGGTCAA
TAAGGTCAATAattatc
tagcaAGGTCAATAAGG
TCAATAattatctagca
AGGTCAATAAGGTCAAT
AattatctagcaGGTCA
ATAAGGTCAATattatc
tagcaGGTCAATAAGGT
CAATAattatctagcaA
GGTCAATAAGGTCAATA
attatctagcaAGGTCA
ATAAGGTCAATAattat
ctagcaattatctagca
```

TBLASTN multithread avec filtrage local



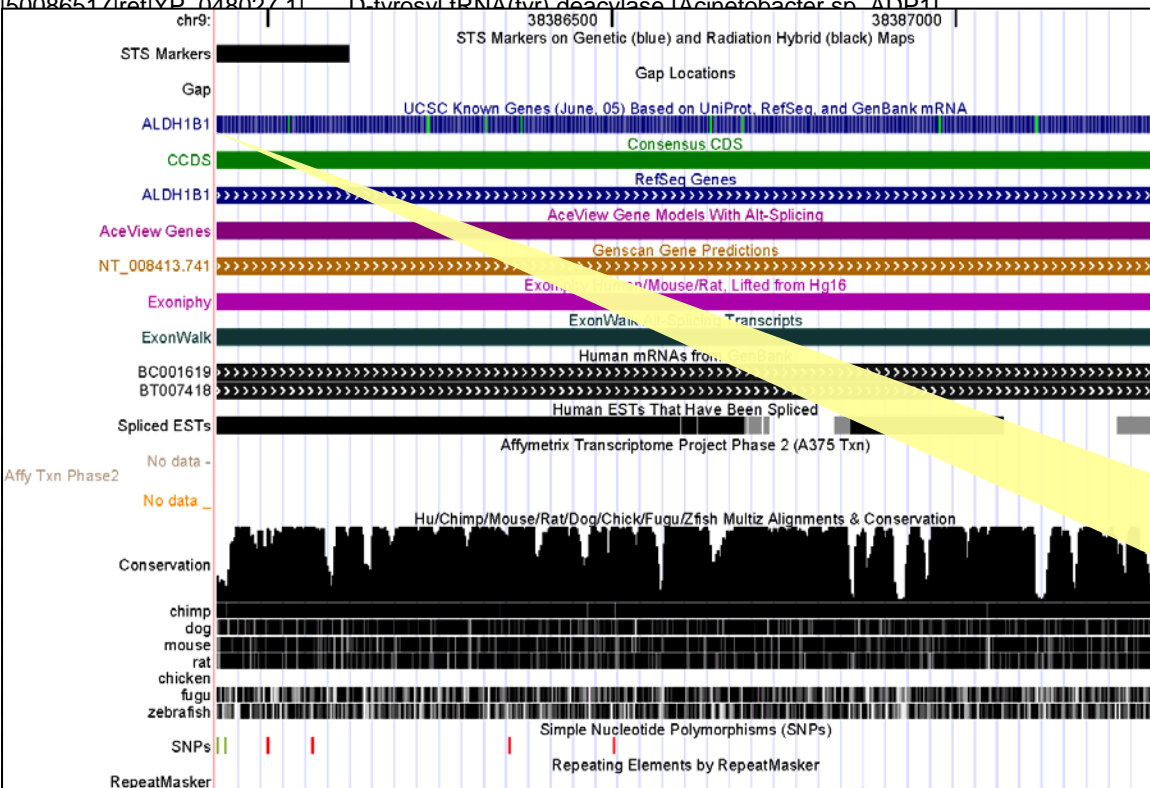
Résultats préliminaires



Test préliminaire

% Parameters: -i=test/prokVShg.mpid1 -o=test/prokVShg.mpid2 -l=test/mitopid.log -n=100 -c=40 -g=50

% query_name	query_description	chr_name	sjct_start	sjct_end
gi 50086409 ref YP_047919.1	NADP+-dependent succinate semialdehyde dehydrogenase [Acinetobacter sp. ADP1]	9	38385928	38387286
gi 50086409 ref YP_047919.1	NADP+-dependent succinate semialdehyde dehydrogenase [Acinetobacter sp. ADP1]	5	76621950	76623347
gi 50086456 ref YP_047966.1	peptide chain release factor 2 [Acinetobacter sp. ADP1]	4	190034915	190035748
gi 50086509 ref YP_048019.1	putative short-chain dehydrogenase [Acinetobacter sp. ADP1]	6	99729340	99730044
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	20	18524650	18524865
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	15	78259517	78259621
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	1	221068906	221069772
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	X	114986665	114987501
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	9	38386099	38387277
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	5	76622130	76623284
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	20	41292858	41293232
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	X	112571945	112572358
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	21	19152015	19152386
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	1	145585729	145586100
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	1	145423725	145424096
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	2	11442398	11442769
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	2	129970144	129970515
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	2	26296702	26297032
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	2	146366198	146366569
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	2	1039964	35104332
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	2	134438706	59579104
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	2	13444366	134444927
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	2	100642453	100642924
gi 50086517 ref YP_048027.1	D-tyrosyl-tRNA(tyr) deacylase [Acinetobacter sp. ADP1]	5	65903690	65904061



Swiss-Prot:
AL1B1_HUMAN
Subcellular location:
Mitochondrial matrix



ReMIX

- ACI MD 2004
- Accélérer la recherche **par le contenu** dans de **grandes** bases de données
- Prototype de **mémoire reconfigurable** de grande capacité
- Basée sur la technologie **FLASH** (très rapide)
- Permet de stocker des index de **plusieurs centaines de Gigaoctets**



mitopid sur ReMIX ?

- Usage de BLAST + filtres maison n'est **pas le plus adapté** aux besoins
- Les gros moyens de calcul **supportent mal** les codes non disponibles en standard
- ✓ Programmer en **adéquation directe** avec notre problématique
- ✓ Mise à **disposition immédiate**



Contributeurs

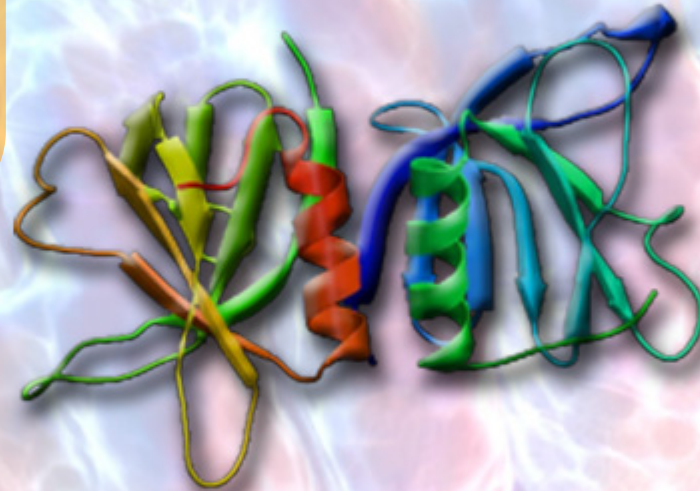


**Inserm U694
(Angers)**

Y. Tourmen,
M. Ferré,
Y. Malthièry,
P. Reynier

IRISA (Rennes)

M. Macwing, H. Leroy,
J. Pley, M. Hurfin,
D. Lavenier



IDRIS (Orsay)

O. Glorieux, A. Fouilloux, I. Dupays,
P-F. Lavallé, V. Alessandrini

