

Académie de Pharmacie – 4 Décembre 2019

# **Les Pandoravirus, nouvelle frontière de la Biologie**

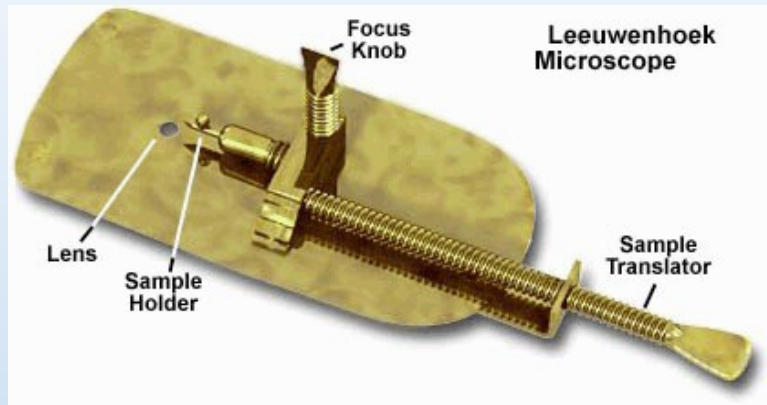
Jean-Michel Claverie

Aix-Marseille université, CNRS

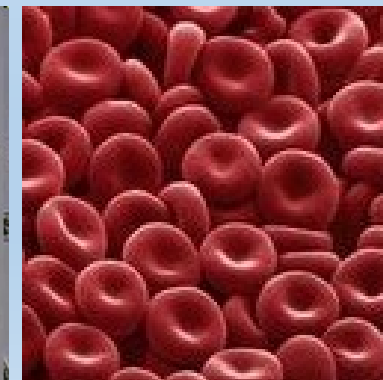
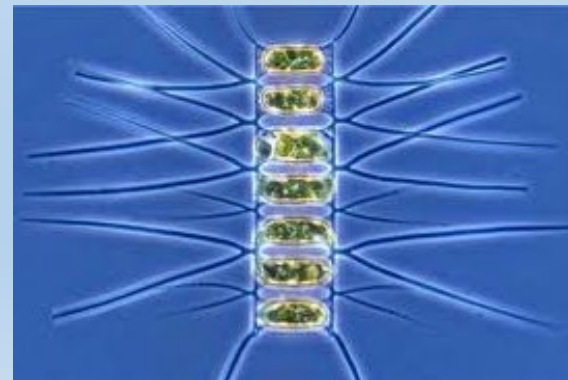
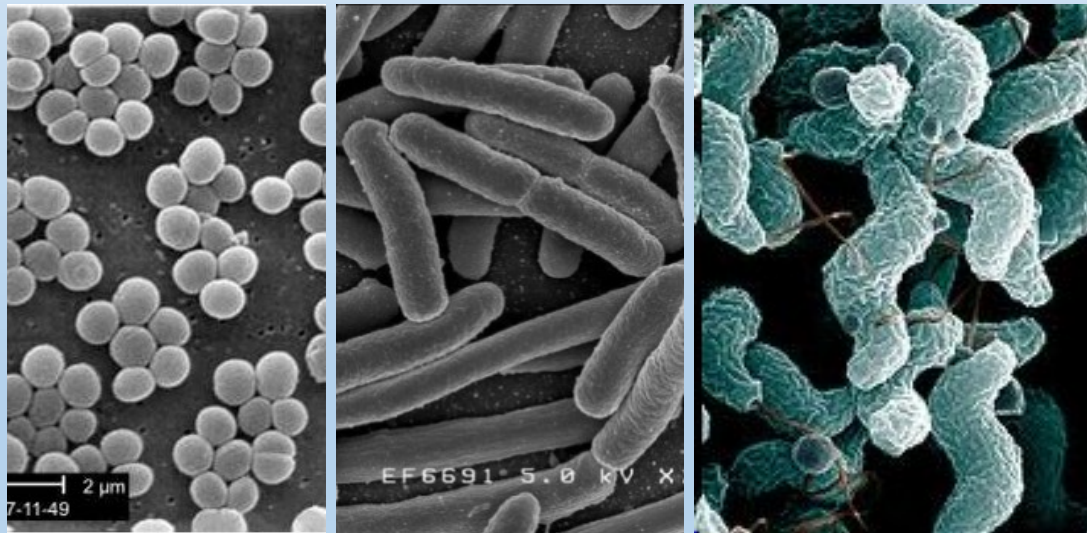
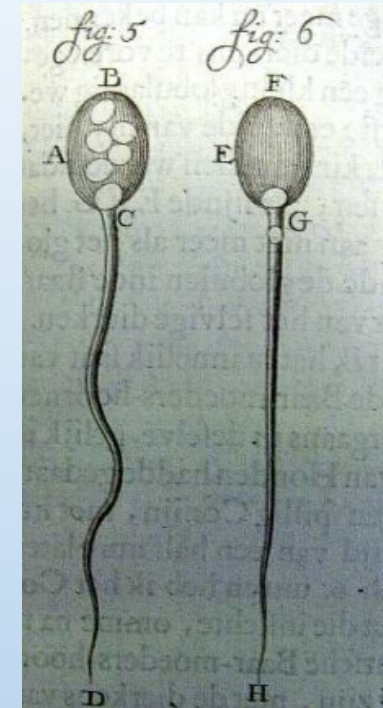
Structural & Genomic Information Laboratory

*Jean-Michel.Claverie@univ-amu.fr*

# Antoni Van Leeuwenhoek (1632 – 1723): "Microbes"



1er microscope 300X (1,4  $\mu$ m)



Doute de la génération spontanée



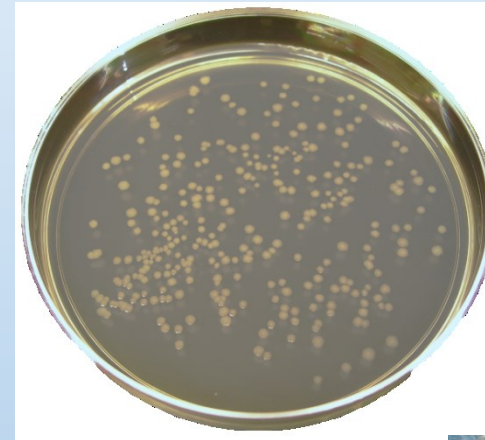
# La théorie des germes (1878)

Fin de la théorie des "miasmes & et de la génération spontanée"

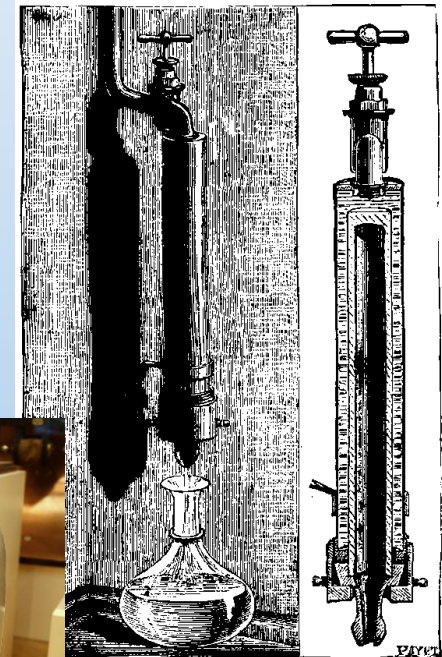
- Pasteur: les maladies transmissibles sont dues à des microbes : maladie du ver à soie (microsporidies), choléra des poules.
- Koch: *Bacillus anthracis* (1877), *Bacillus tuberculosis* (1882) & *Vibrio cholera* (1883)



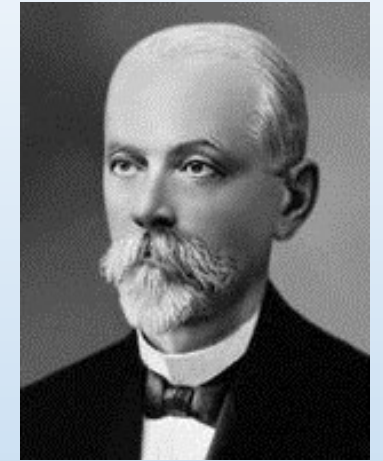
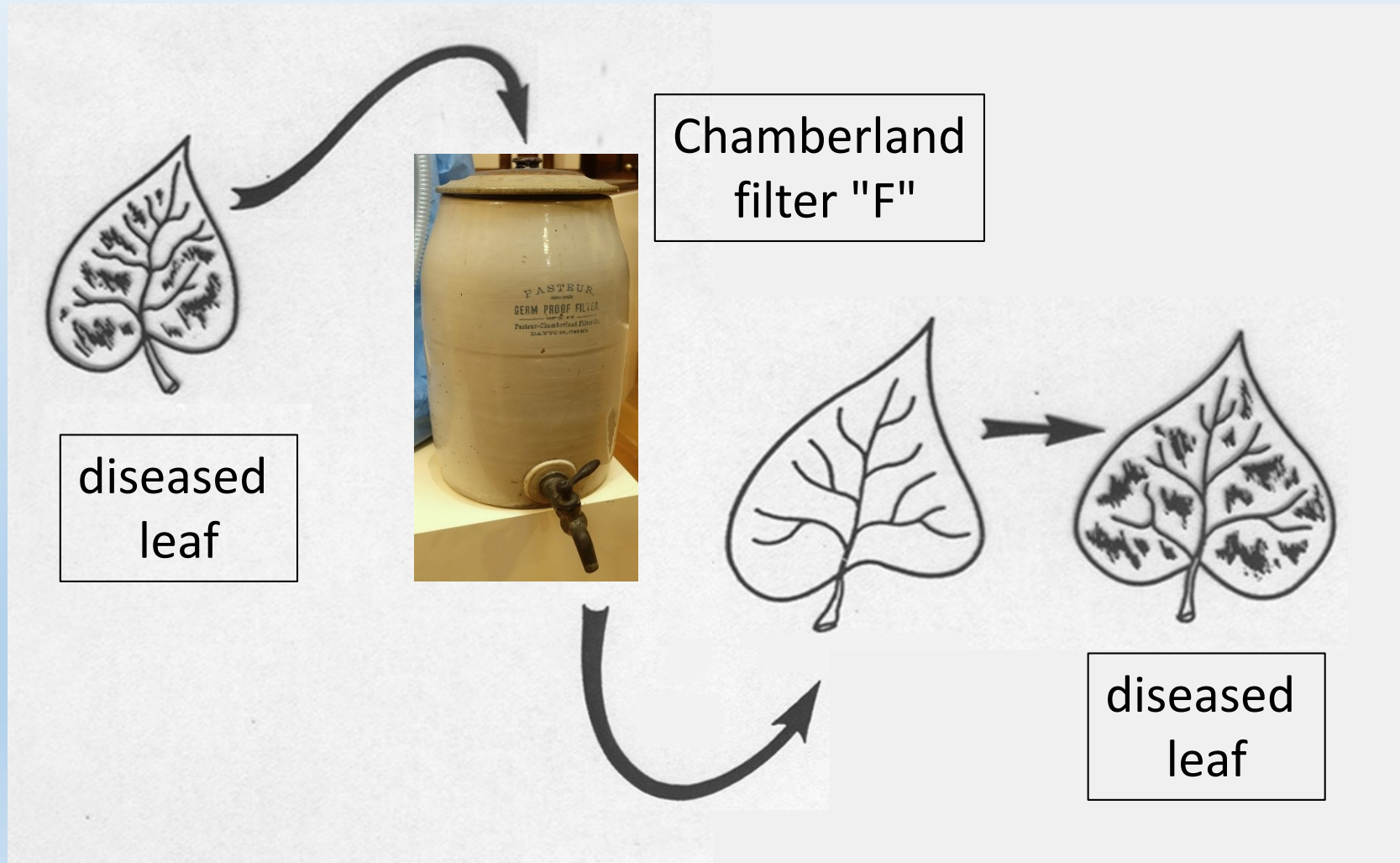
La preuve standard  
en Microbiologie



Le filtre de  
Chamberland  
Stérilisation par la  
rétention des microbes  
(1884)



# La découverte du virus de la mosaïque du tabac (1892)



Dmitri Ivanovski  
(Дмитрий Иванович)

En violation de  
la théorie des  
germes (1878)

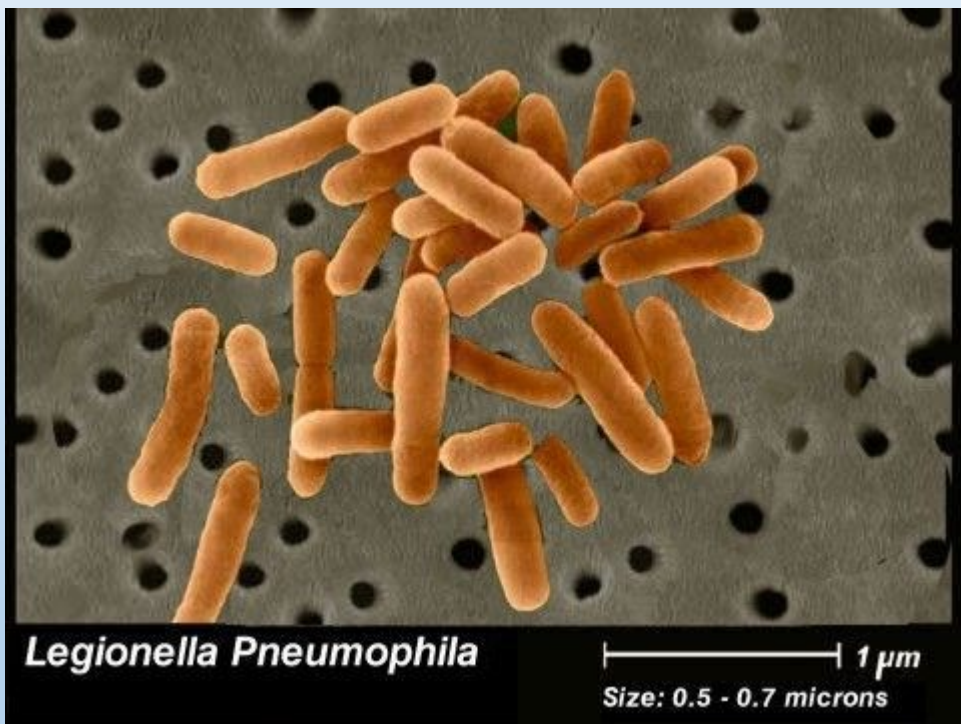
# Le premier siècle de la Virologie

- Les virus sont invisibles en microscopie optique
- Les virus passent à travers les filtres (0.2 $\mu$ m in  $\emptyset$ )(Chamberland "F")
- Les virus sont des parasites intracellulaires obligatoires
- Ils infectent aussi les bactéries (phages, Twort, 1915; d'Hérelle, 1917)
- Certains sont cristallisables (TMV, W. Stanley, 1935)
- Les premières images sont obtenues en microscopie électronique (Kausche, Pfankuch & Ruska 1939)
- Les virus ont beaucoup moins de gènes que les cellules

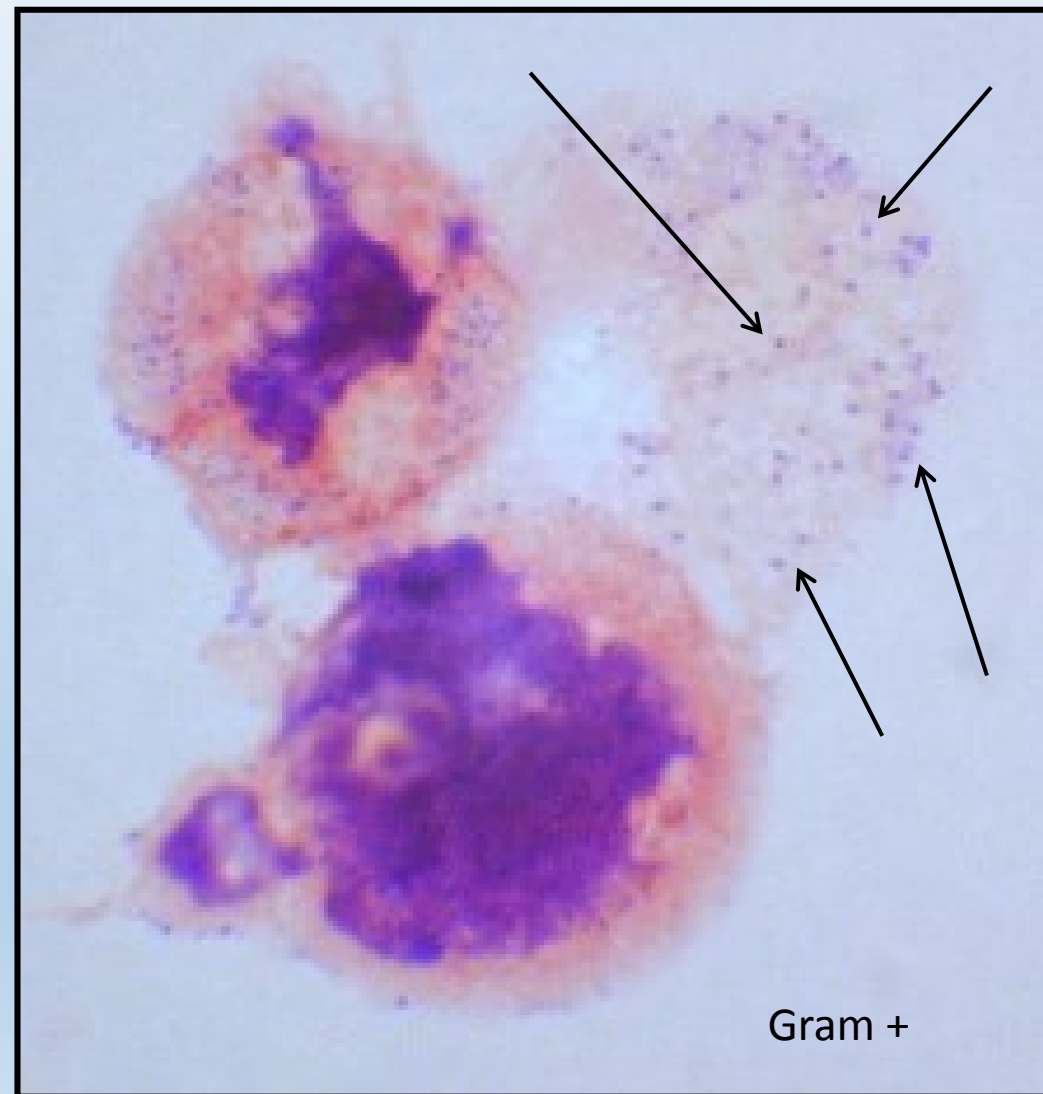


# La découverte de "bradfordcoccus" (1992)

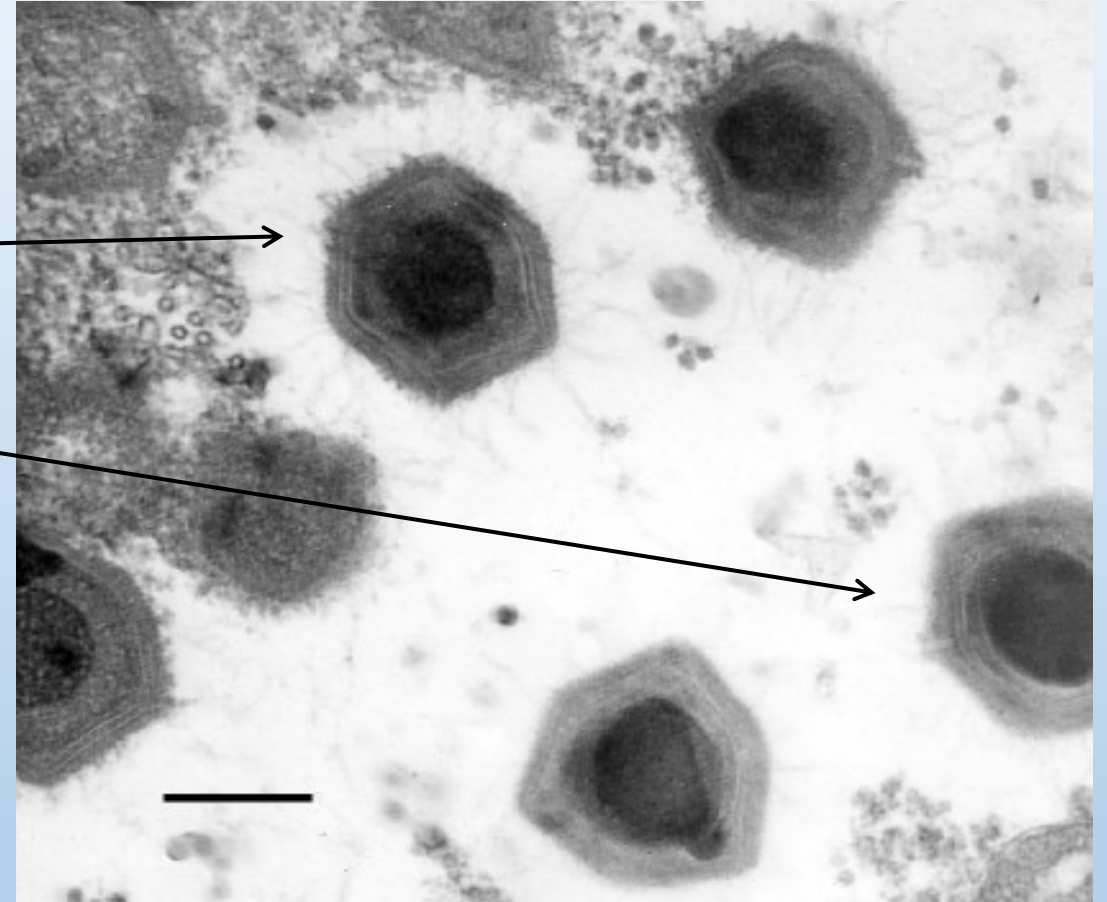
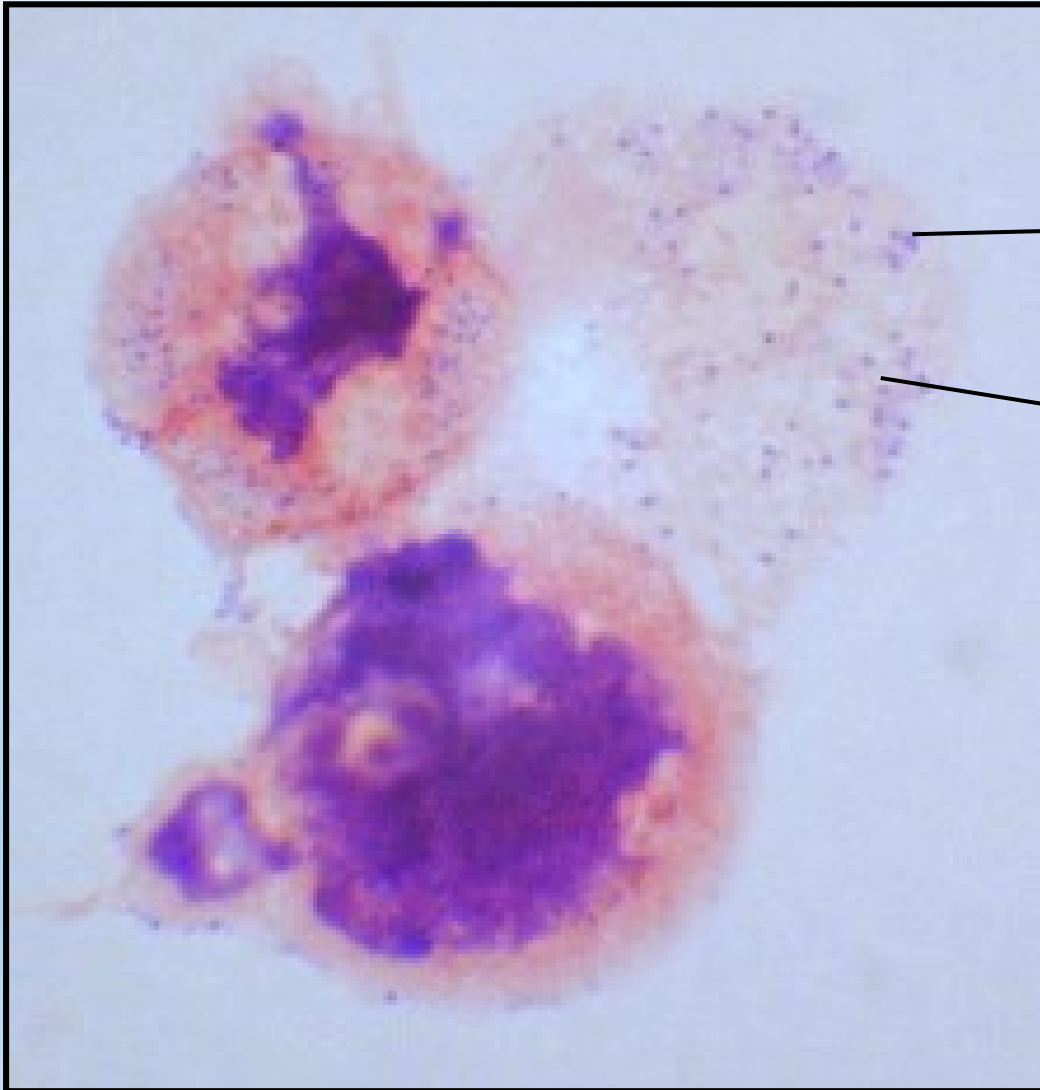
A la recherche de la source d'une épidémie de légionellose à Bradford (UK)



Timothy Robotham (1992)



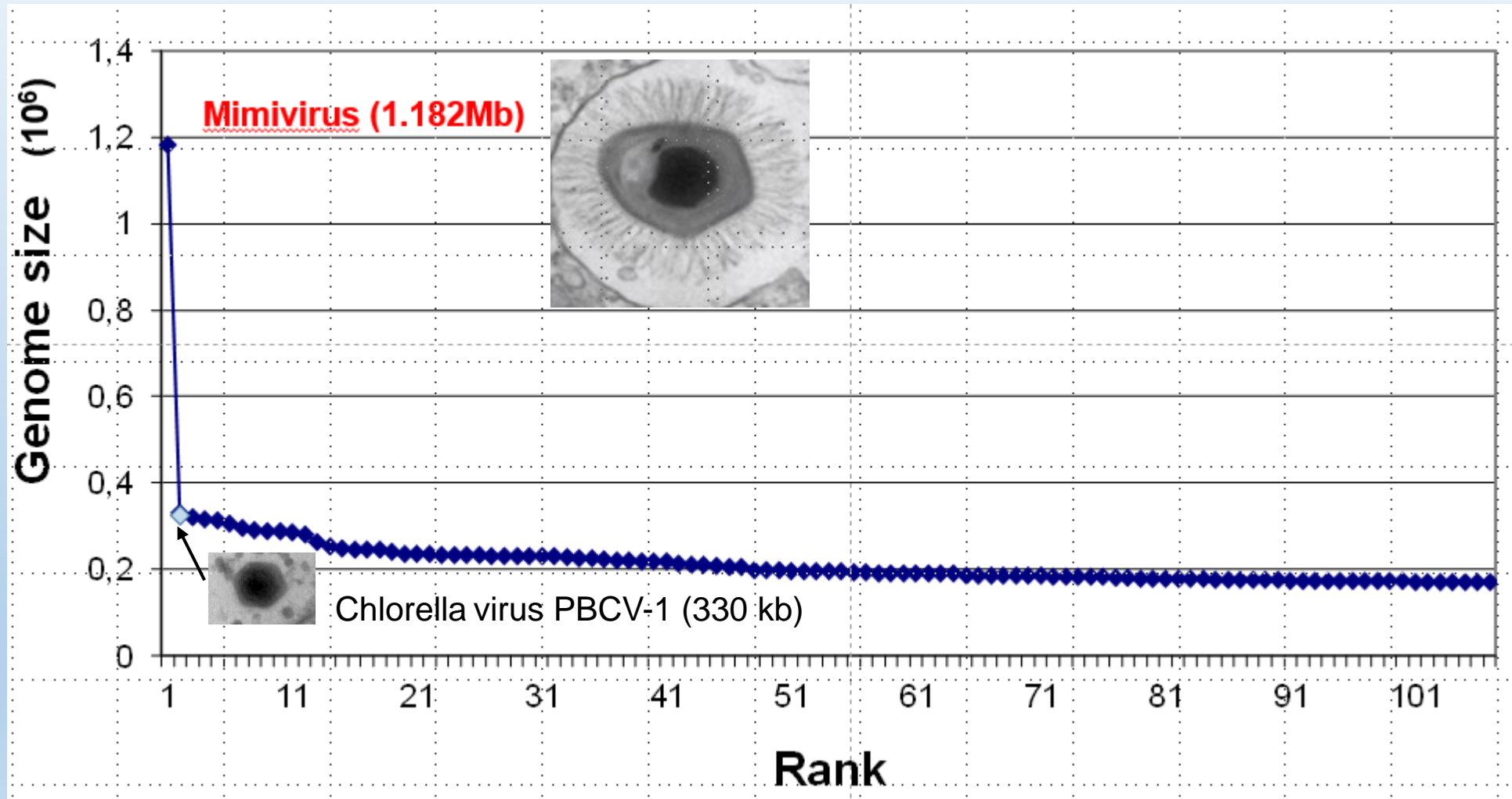
# La découverte de Mimivirus (2002)



La Scola *et al.*, Science 2003

Raoult *et al.*, Claverie. Science 2004

Mimivirus:  $\approx 1000$  genes, plus que de nombreuses bactéries parasites



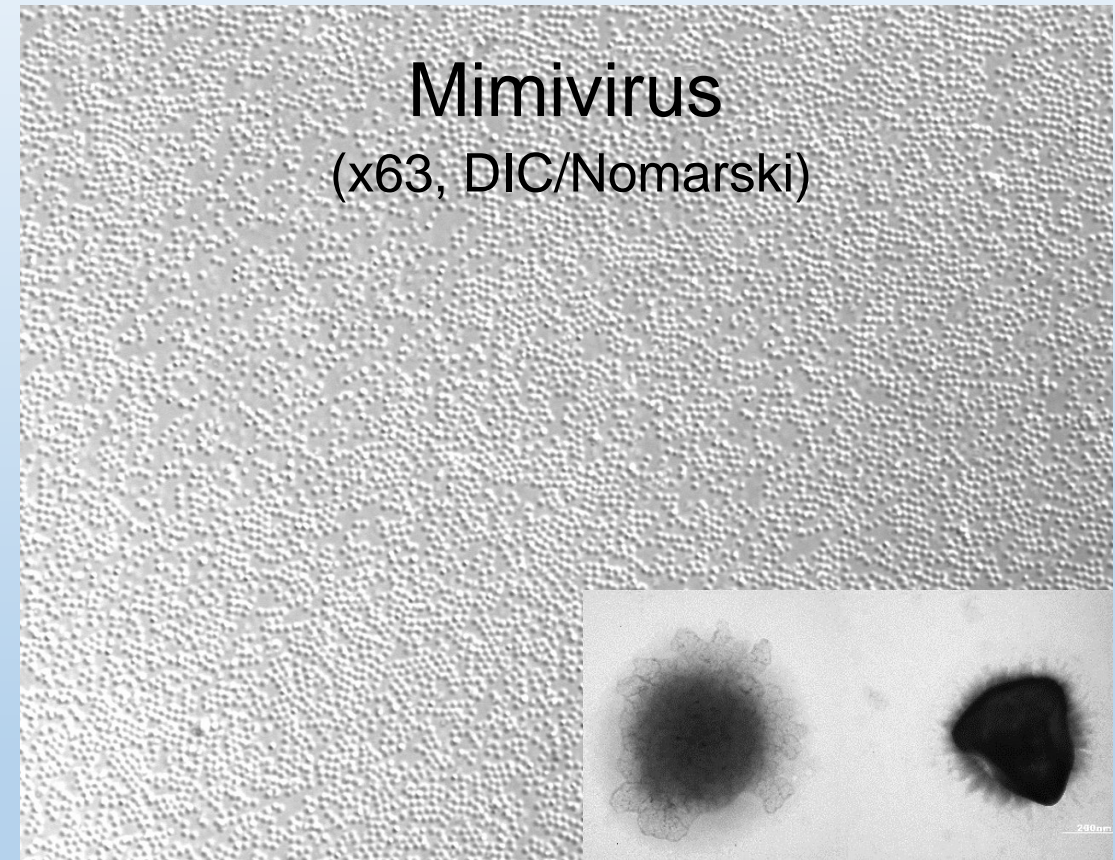


# le concept de virus (Lwoff, 1957, Lwoff & Tournier 1966)

- ~~• Viruses are small (less than  $0.2\mu\text{m}$  in  $\emptyset$ )~~
- ~~• Viruses are invisible by light microscopy~~
- Viruses are obligate intracellular parasites
- Viruses cannot generate ATP
- Viruses do not have ribosomes (don't make proteins)\*
- Viruses do not divide (they multiply otherwise)
- A microbes is either a virus or a cell

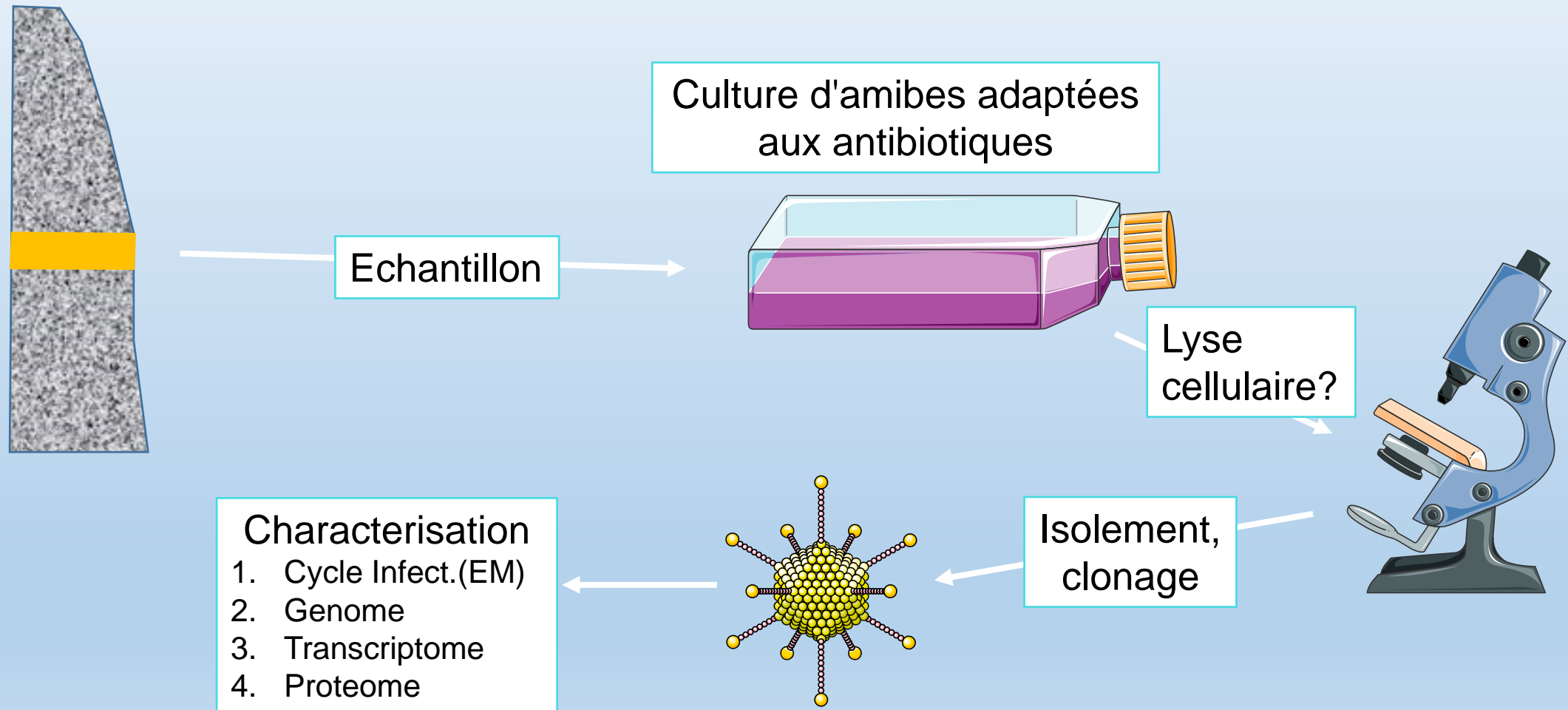
\* Mais que dire de ces amino-acyl tRNA ligases?

# Mimivirus: le premier virus "géant"



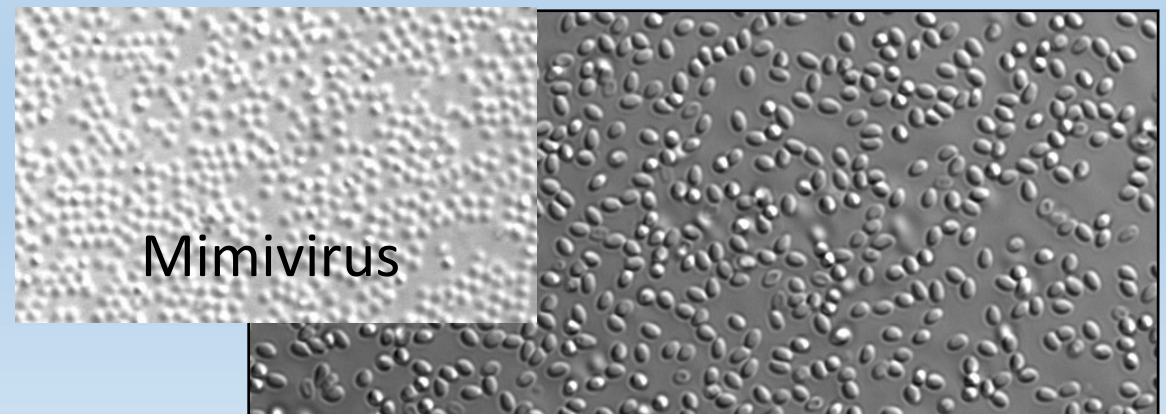
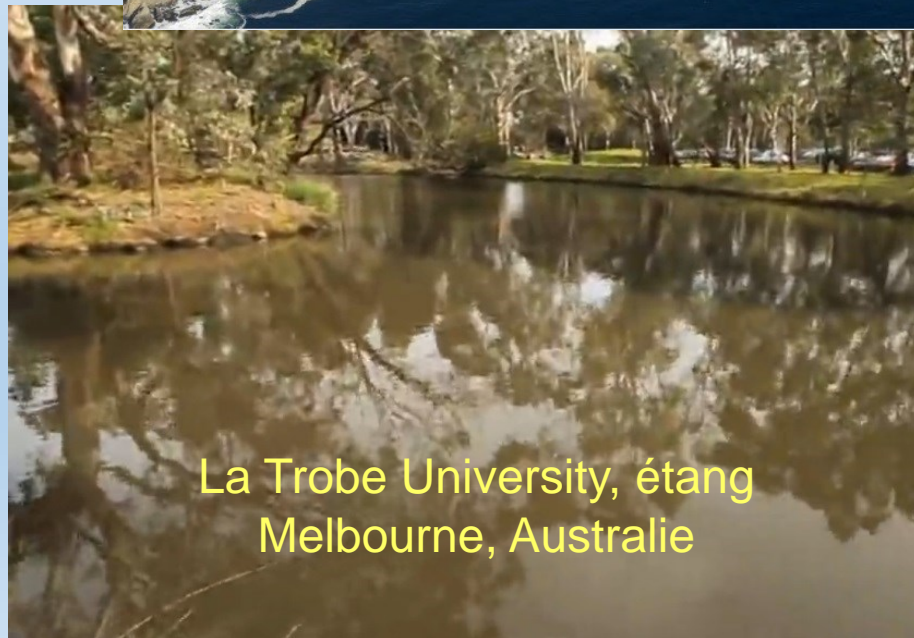
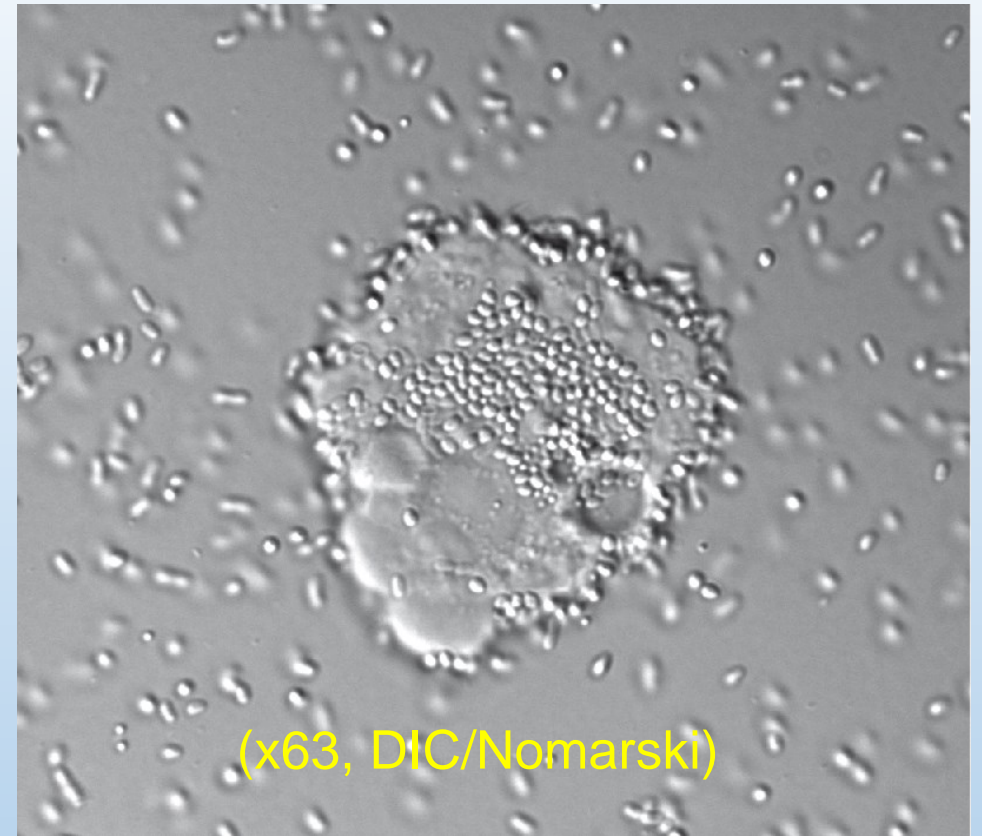
Mimivirus  
(x63, DIC/Nomarski)

# Recherche systématique de virus d'acanthamoeba





# Découvertes des premiers pandoravirus (2009-2011)





# Nouvelle forme de vie?

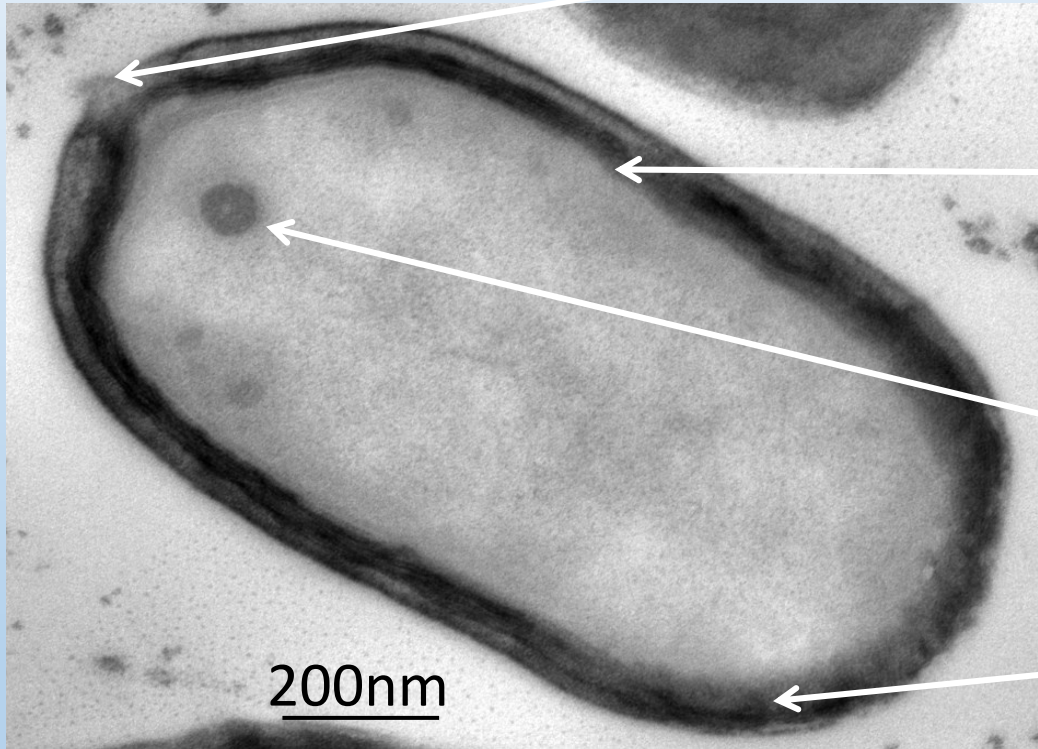


Ostiole-like apex

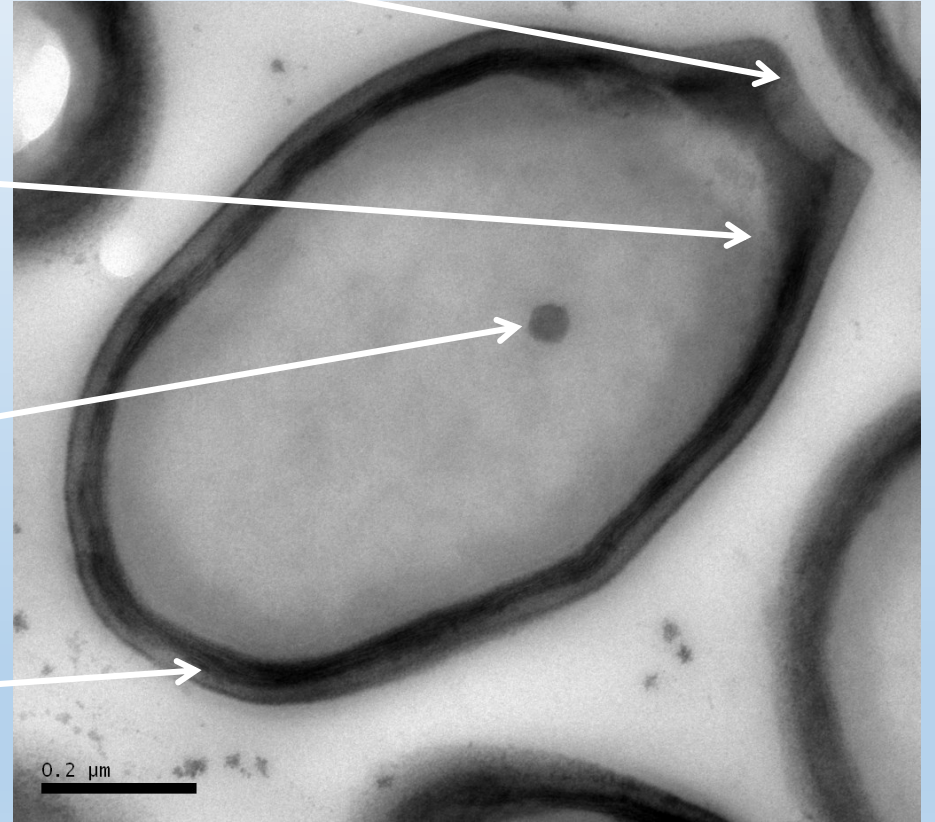
Membrane

Electron dense sphere

Three layer tegument

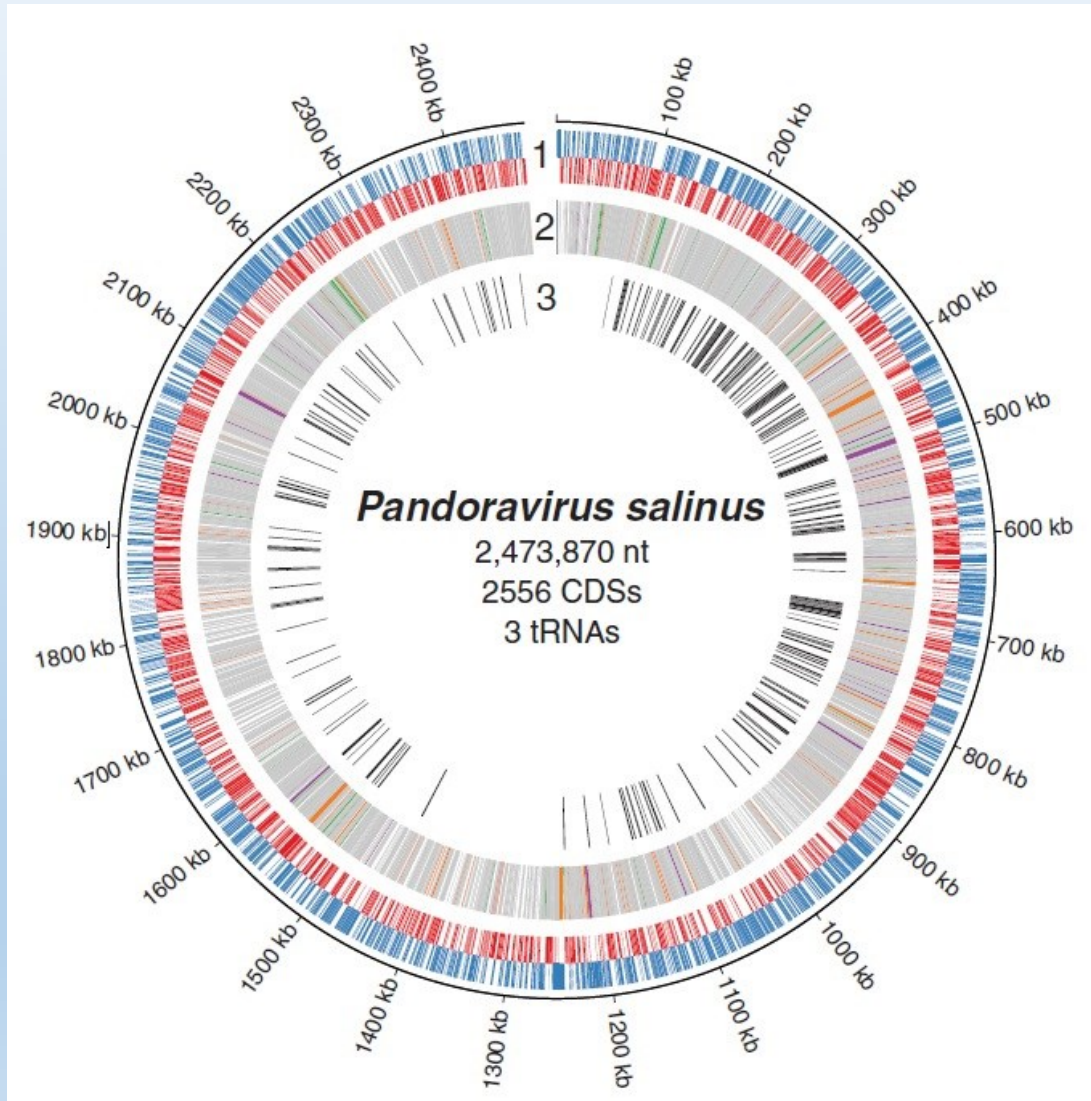


Pandoravirus salinus



Pandoravirus dulcis

# Pandoravirus: une seconde famille de virus géants



- Pas d'appareil de traduction
- Pas de division (FtsZ)
- Pas de synthèse d'ATP

→ Ce doit être un virus  
mais

- Pas de protéine majeure de capsid

Conclusion:

Les *Mimiviridae* ne sont pas la seule famille de virus géants!



# Pandoravirus: la nouvelle frontière

*Encephalitozoon cuniculi*

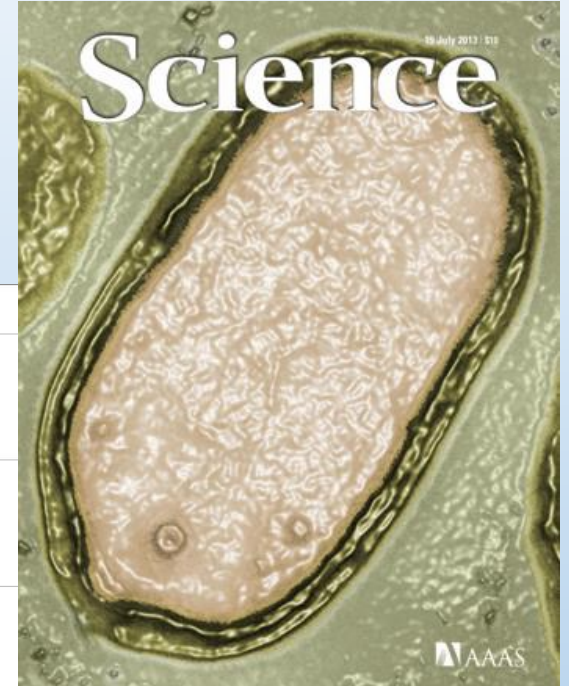
2.5 Mb 1,996 protéines

*Encephalitozoon intestinalis*

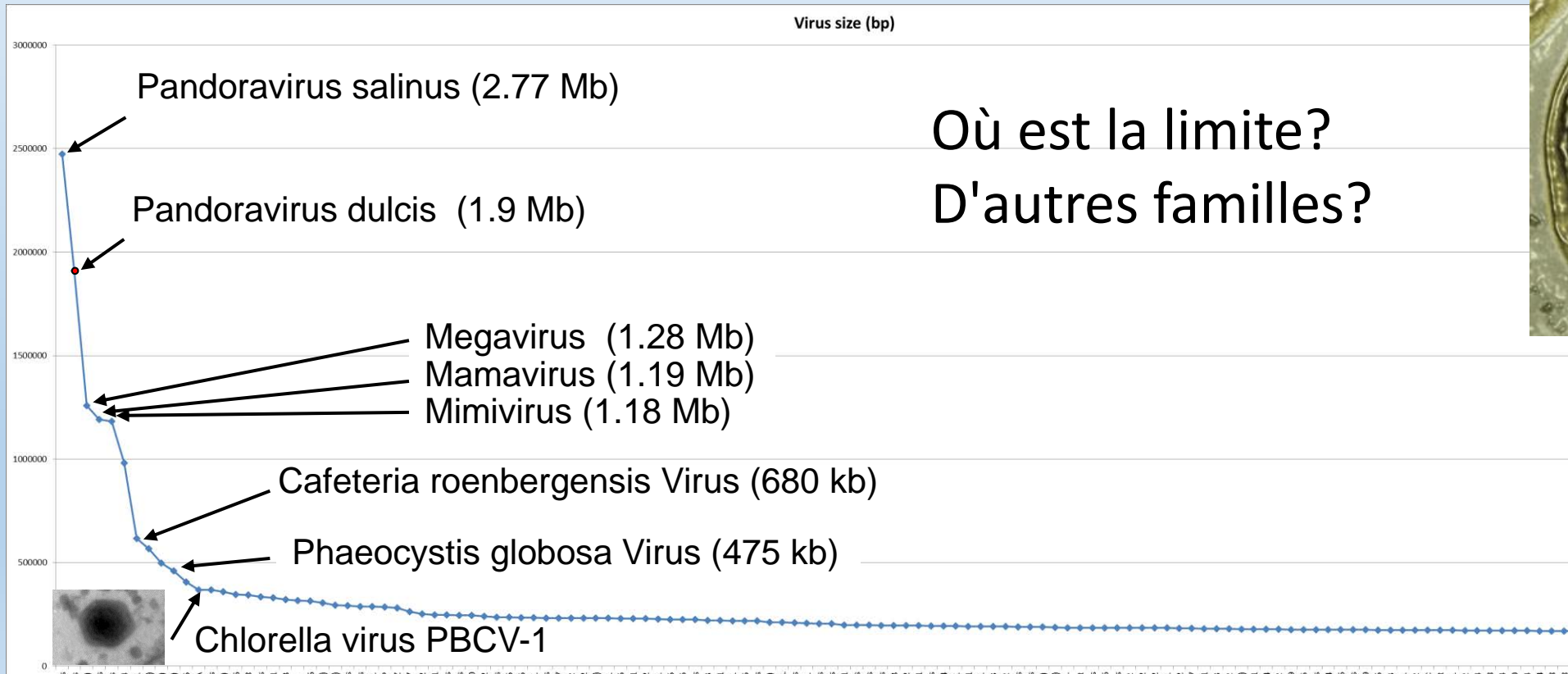
2.22 Mb 1,833 protéines

*Encephalitozoon romaleae*

2.19 Mb 1,831 protéines

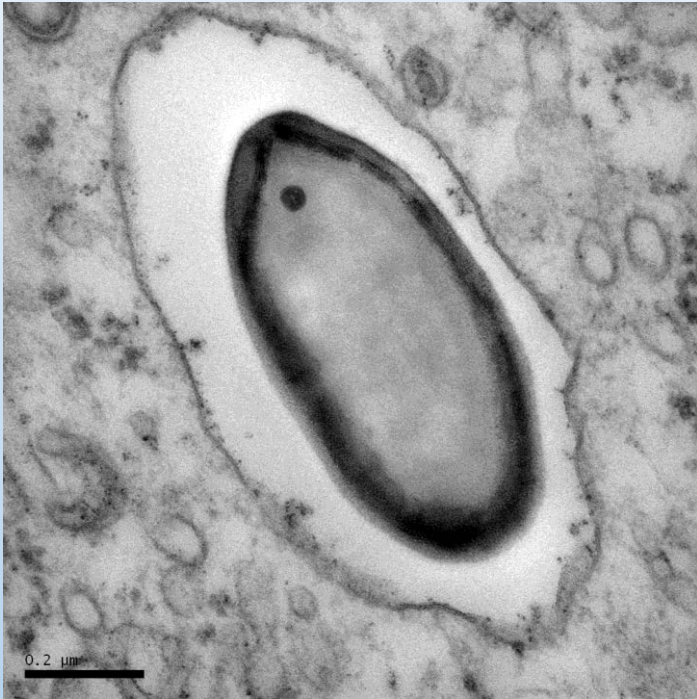


Philippe *et al.*,  
Science 2013



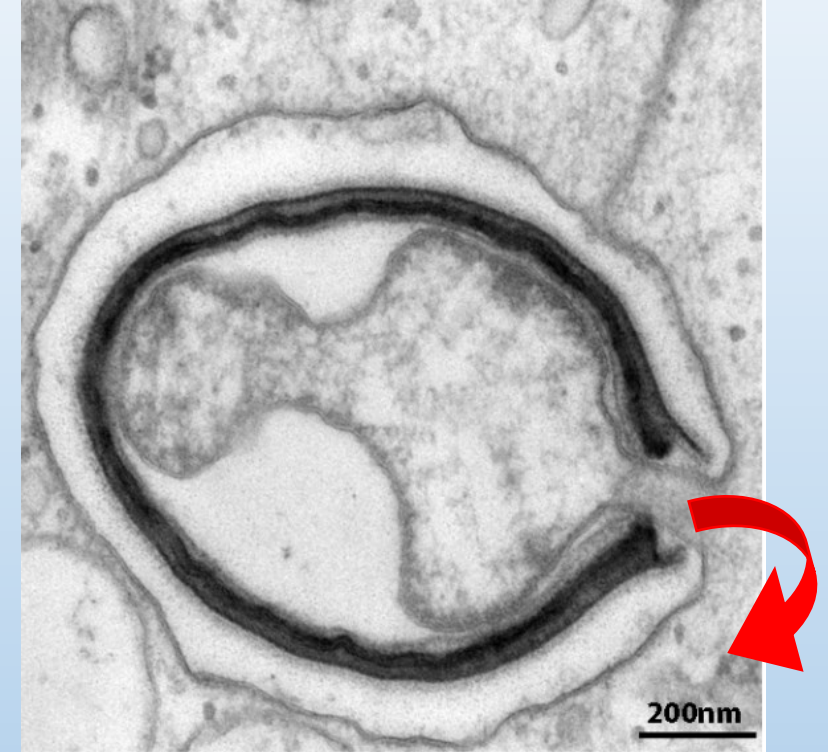
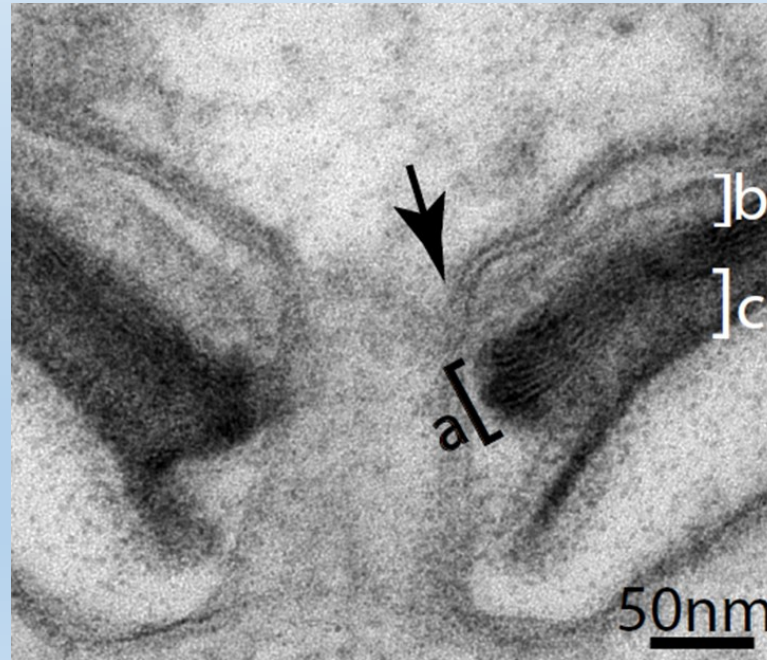
Où est la limite?  
D'autres familles?

# Pandoravirus: cycle infectieux



Step 1:  
phagocytosis

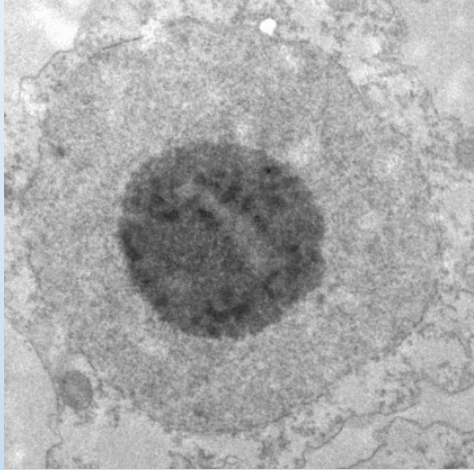
Step 2 :  
membrane fusion



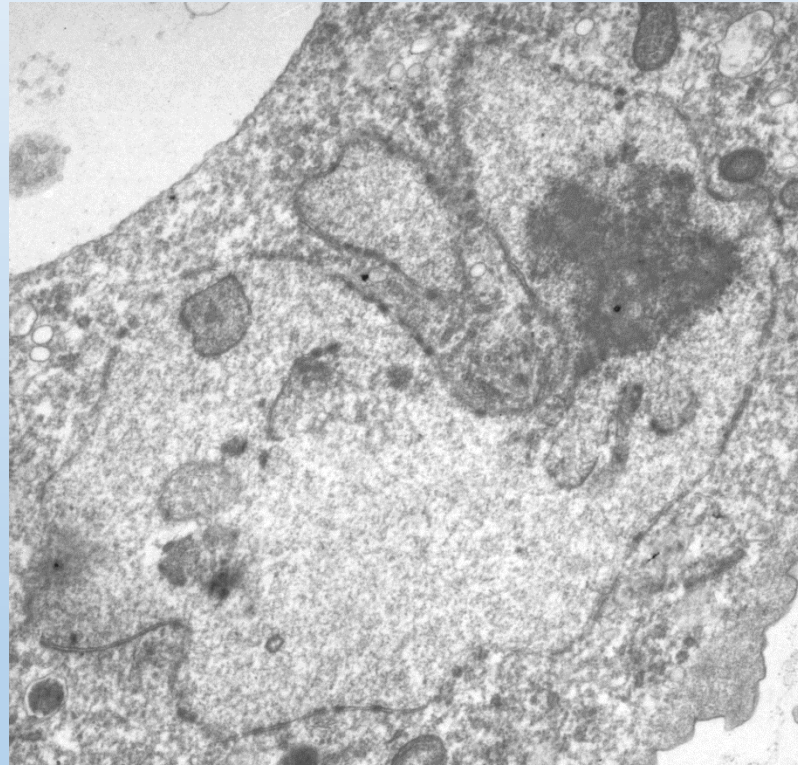
Step 3 :  
« downloading »



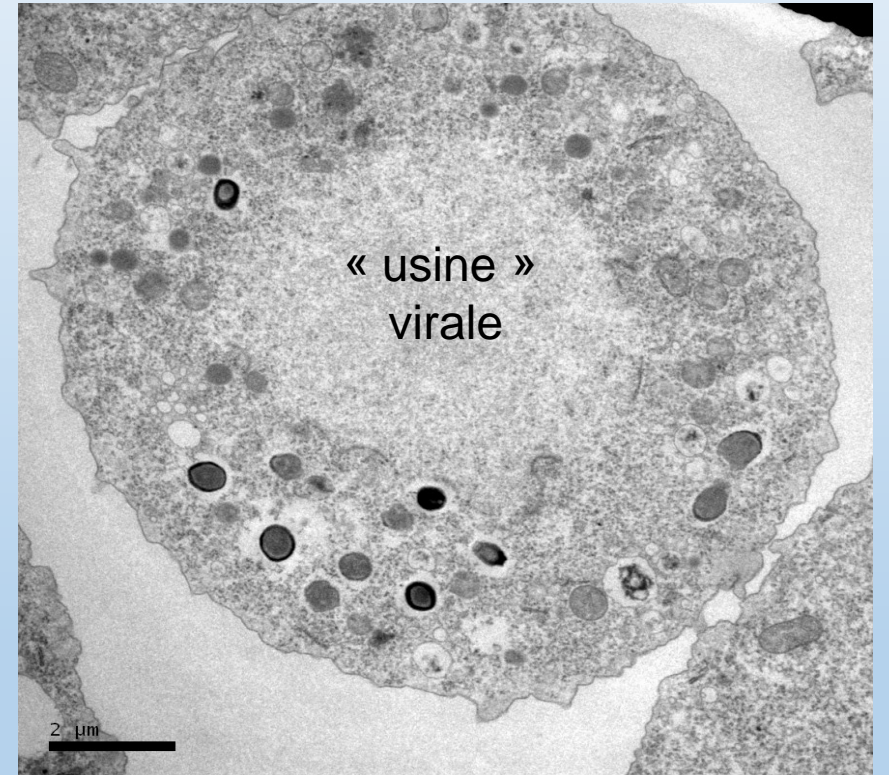
# Phase nucléaire initiale



Acanthamoeba cell  
(non infectée)



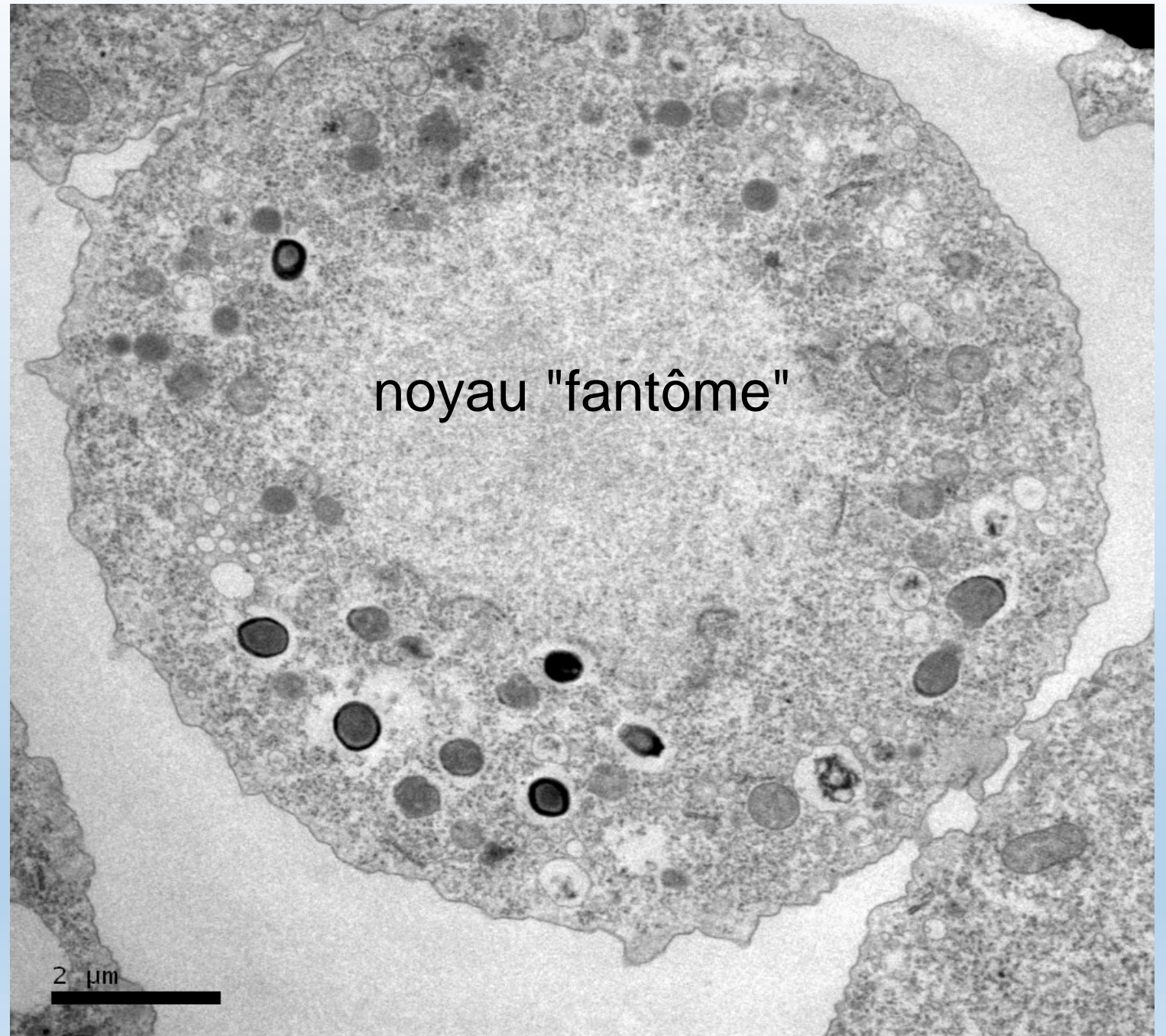
Infectée (3h p.i.)



Formation des premières  
particules (6h p.i.)

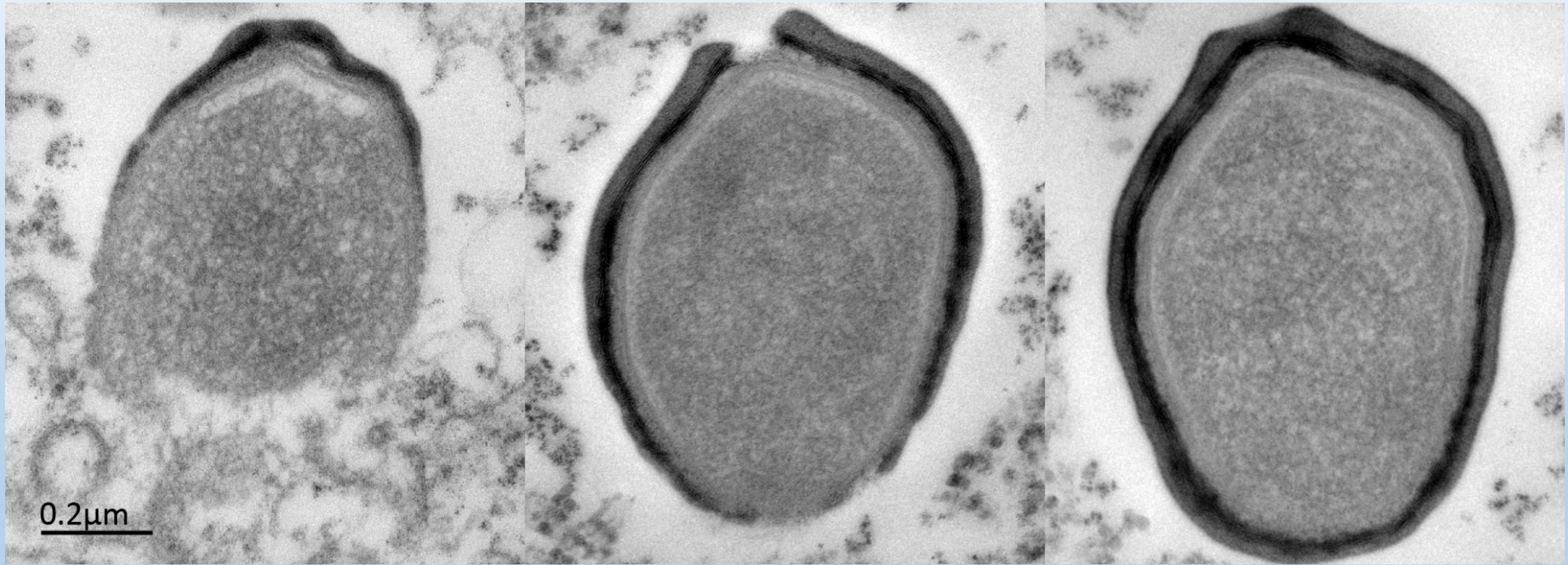


# L'usine virale



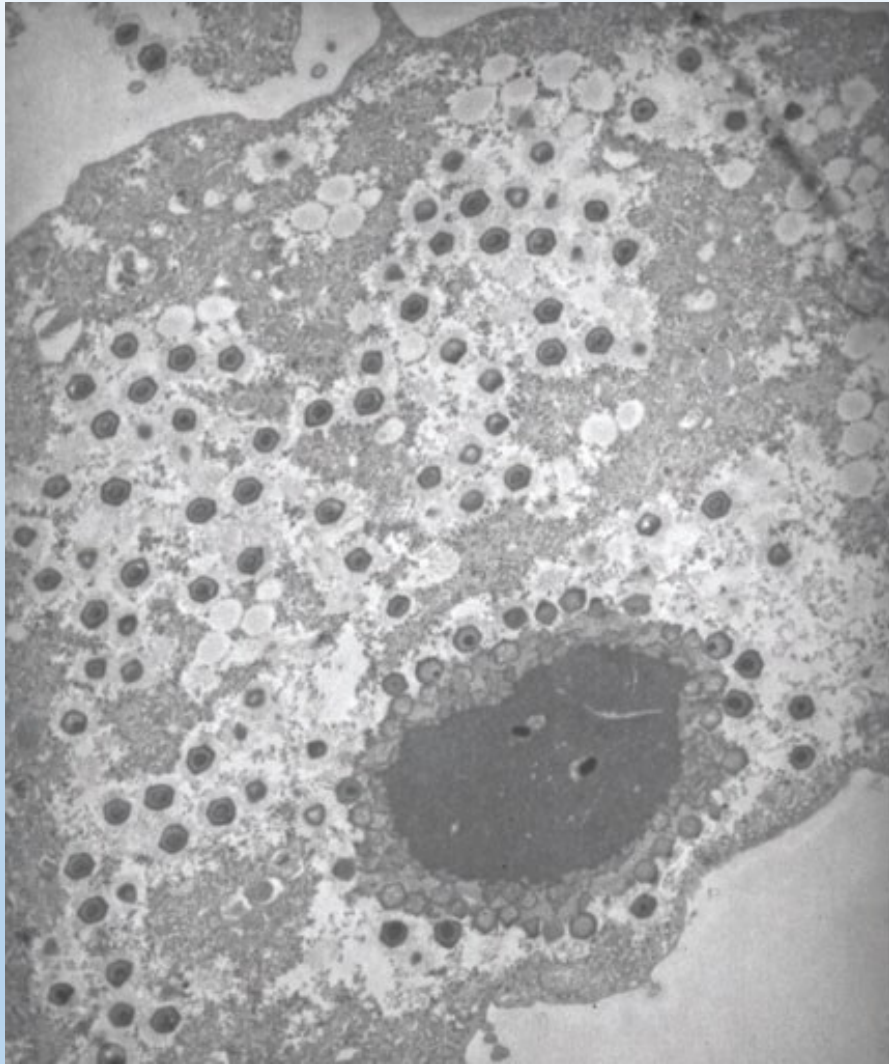


# Formation des particules

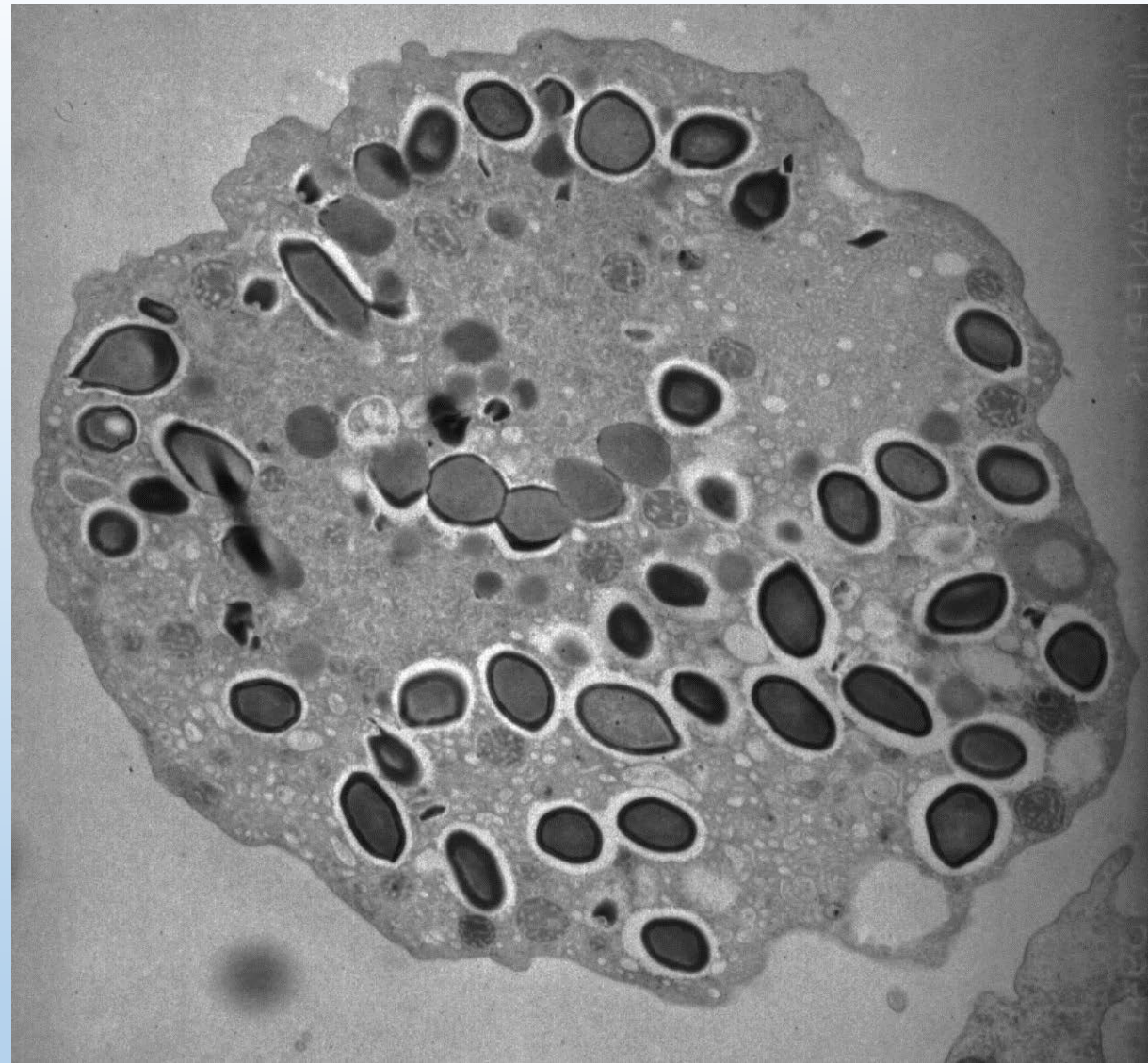




Fin du cycle



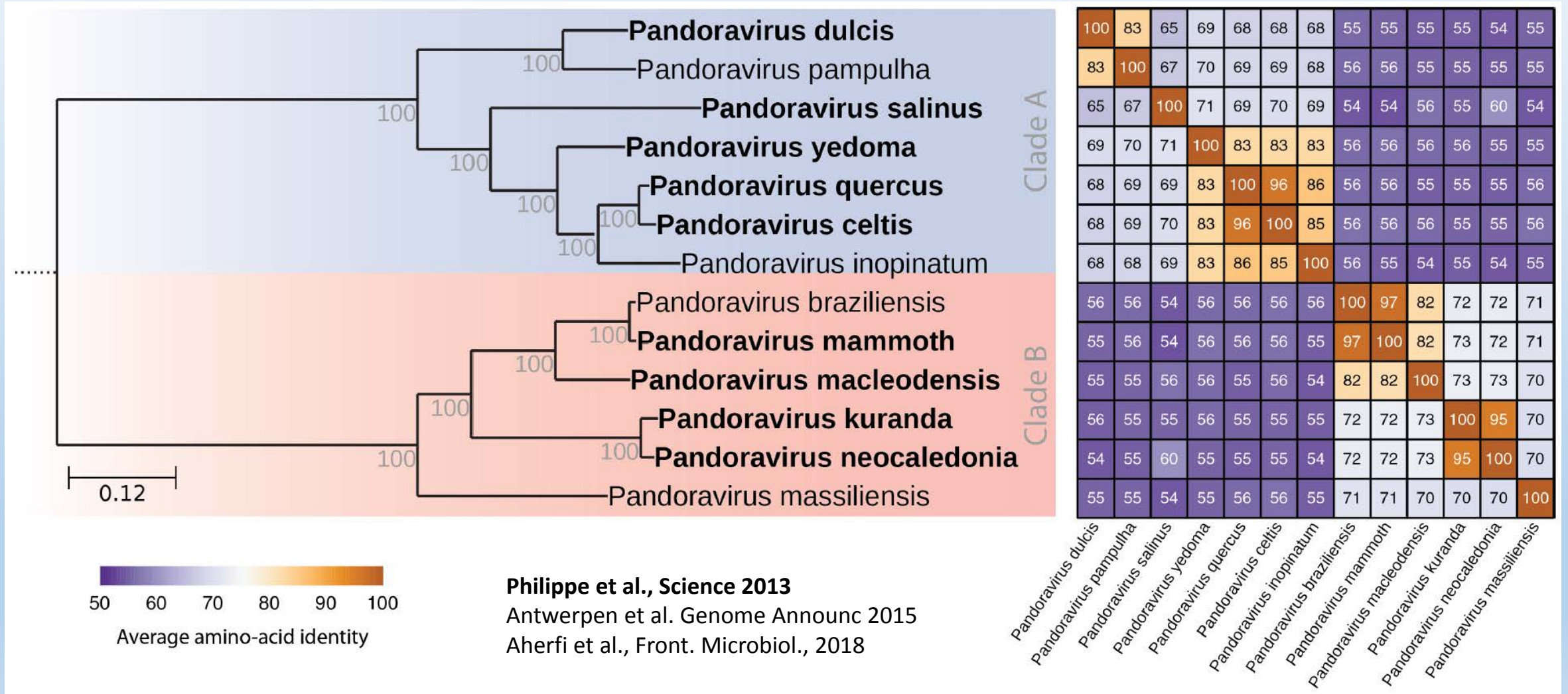
Mimivirus

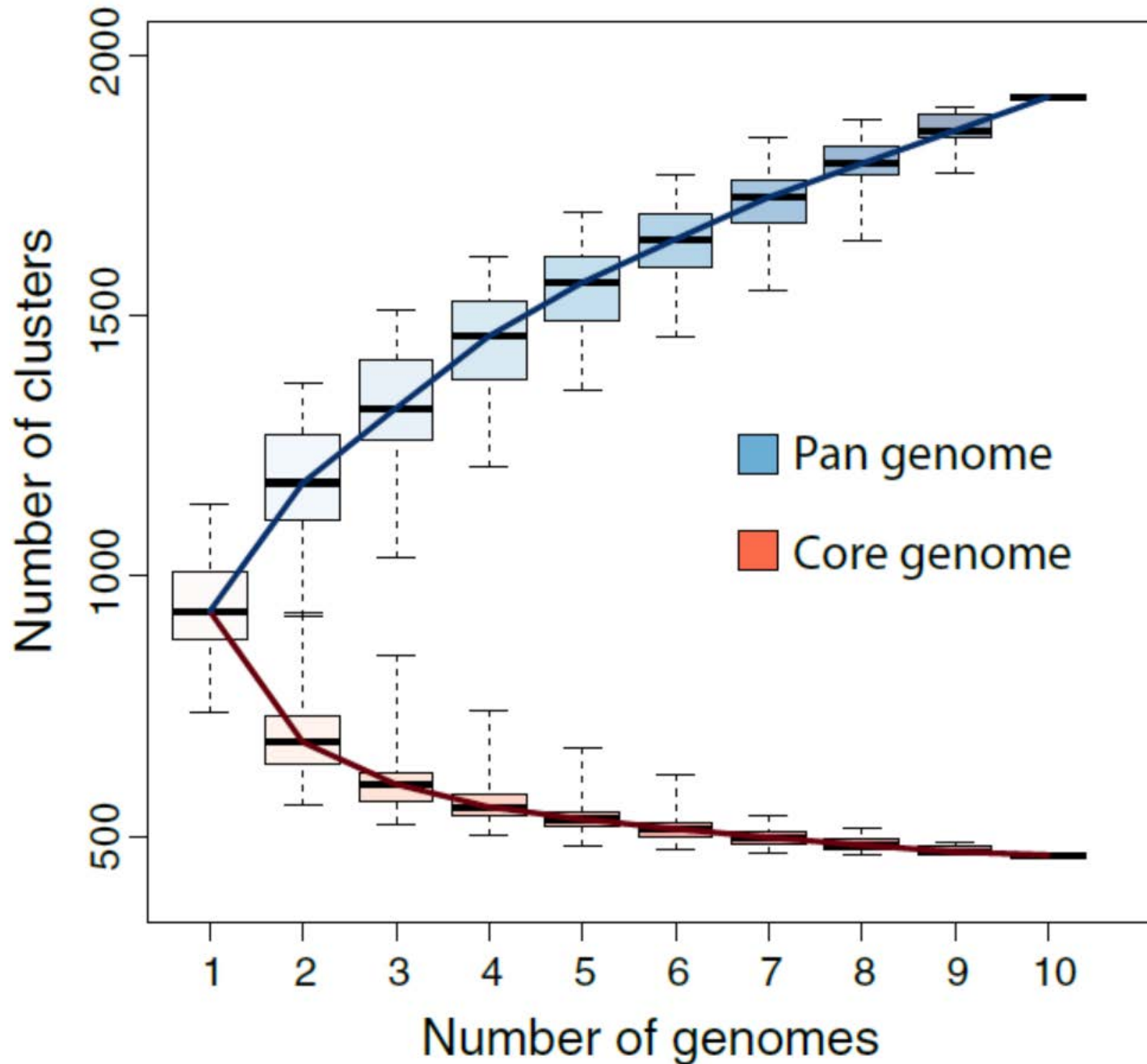


Pandoravirus



# Pandoraviruses: 2 clades distinct





← 1915 (+ 60),  
Open (Heap law  $\alpha < 1$ )

Pandoraviruses:  
de nouvelles  
protéines pour  
chaque nouvel isolat

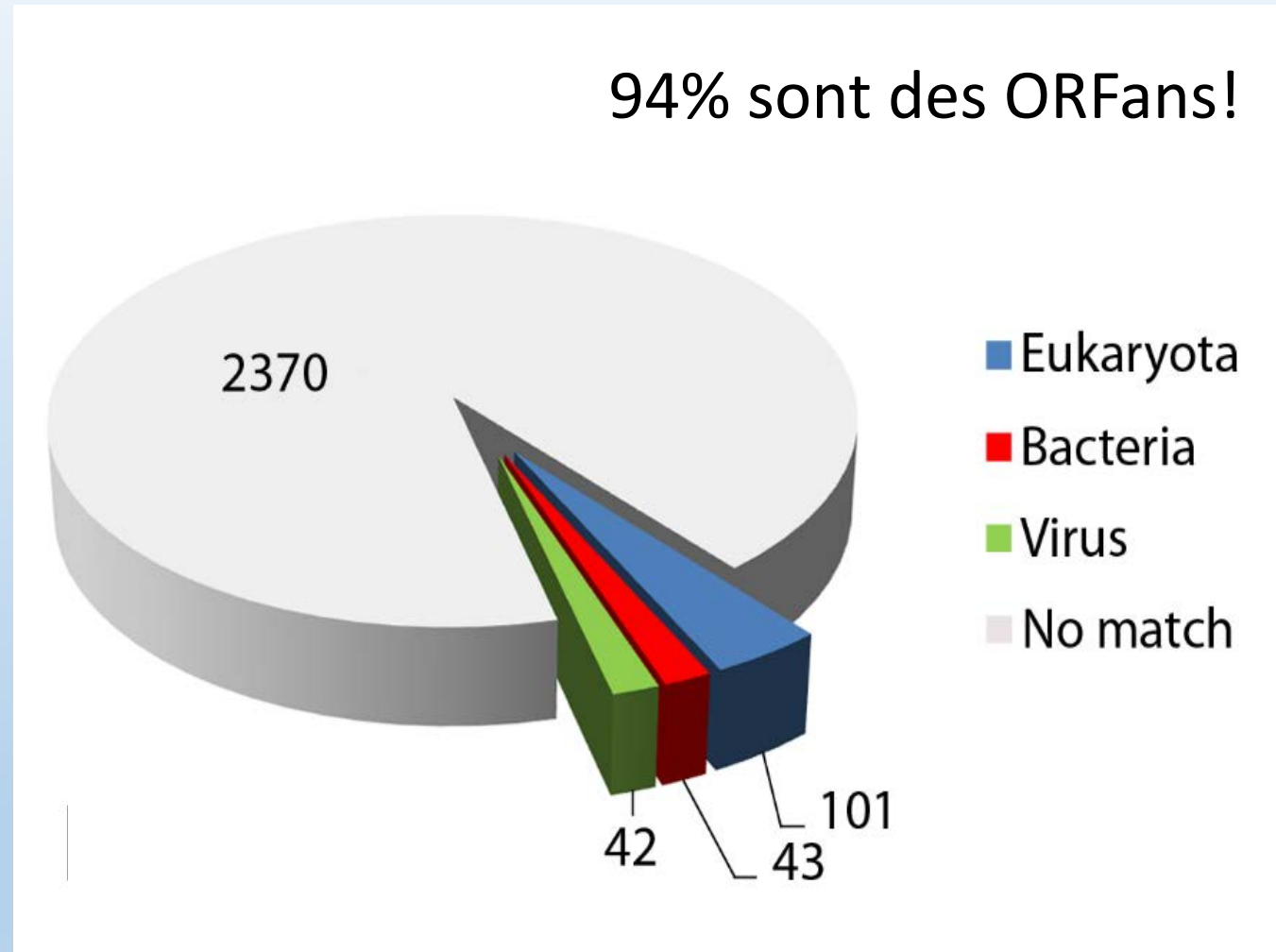
← 464 "core proteins"

Legendre *et al.*, Nat. Comm. 2018;  
Legendre *et al.*, Front. Microb. 2019

# D'où viennent tous ces nouveaux gènes?

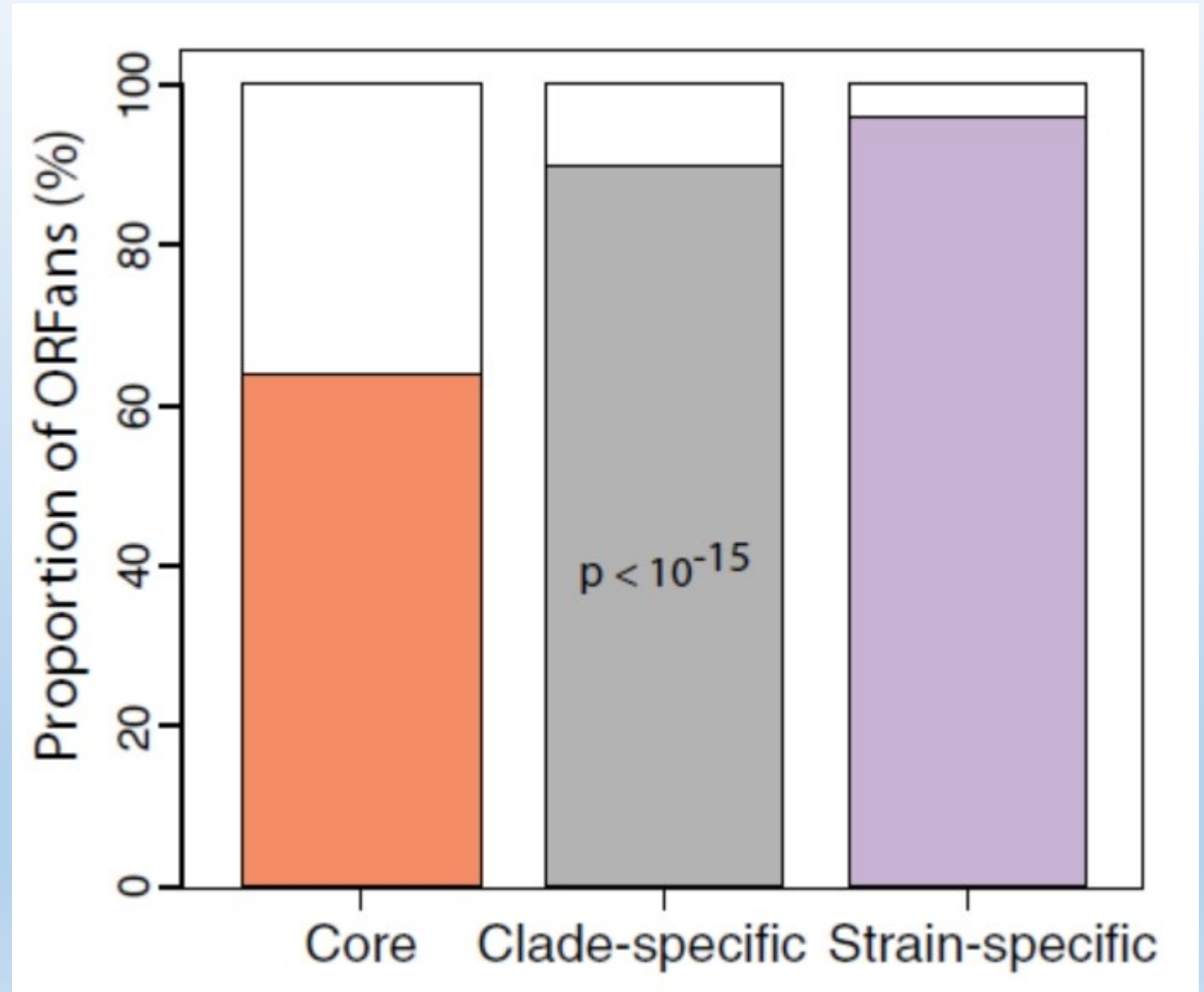
*P. salinus*

2.8 Mb linear DNA  
2,473,870 nt  
2556 proteins

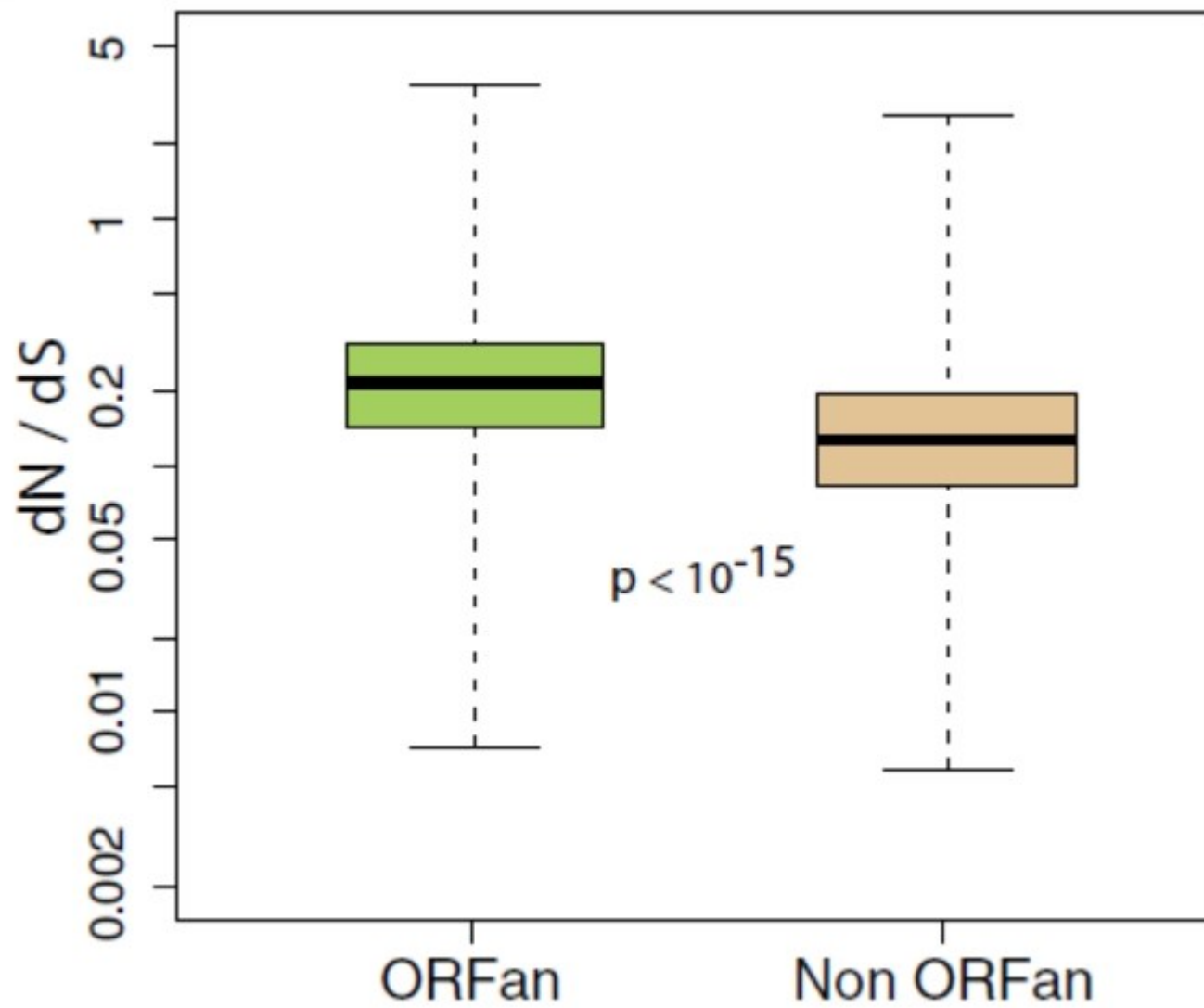




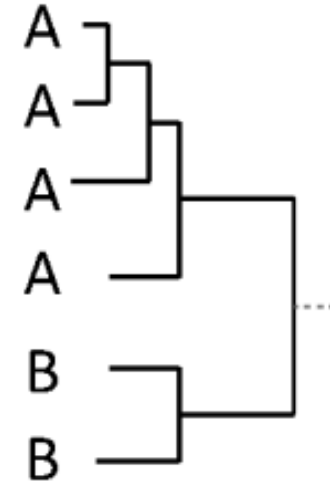
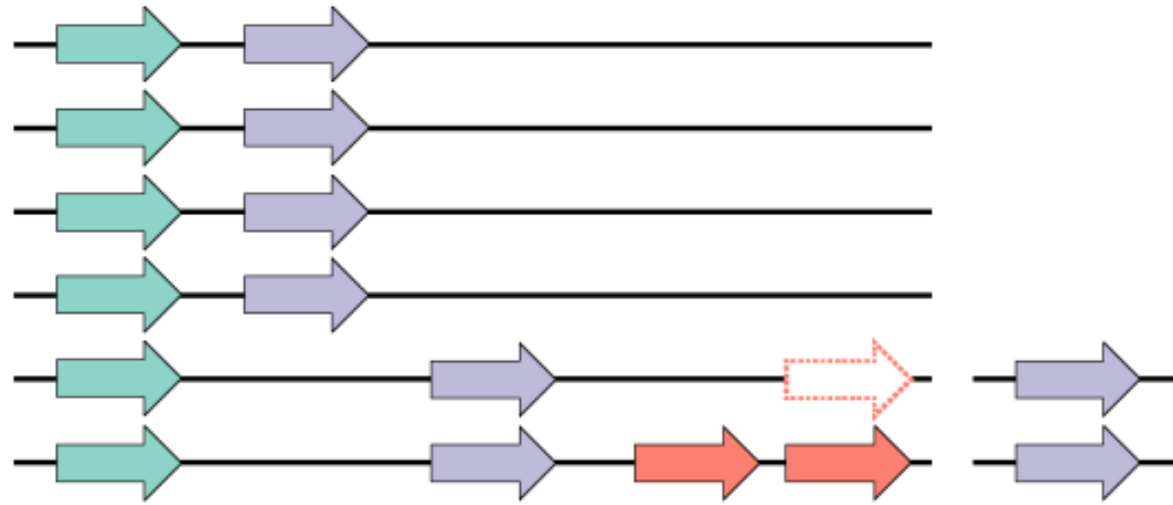
Les protéines les moins partagées sont les plus orphelines



Les protéines  
orphelines sont aussi  
sous pression de  
sélection négative,  
Mais moins que les  
autres



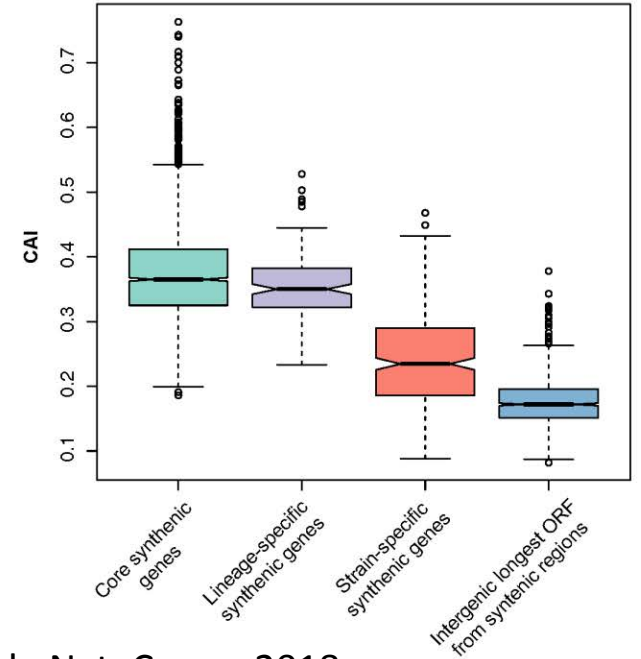
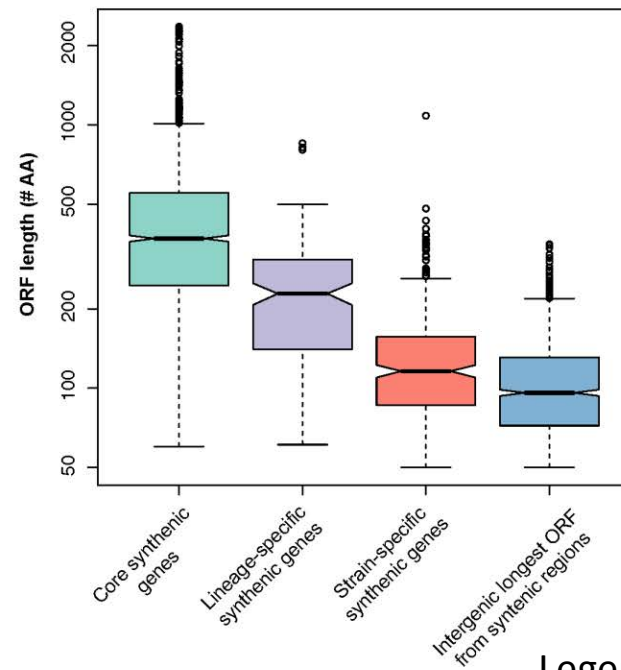
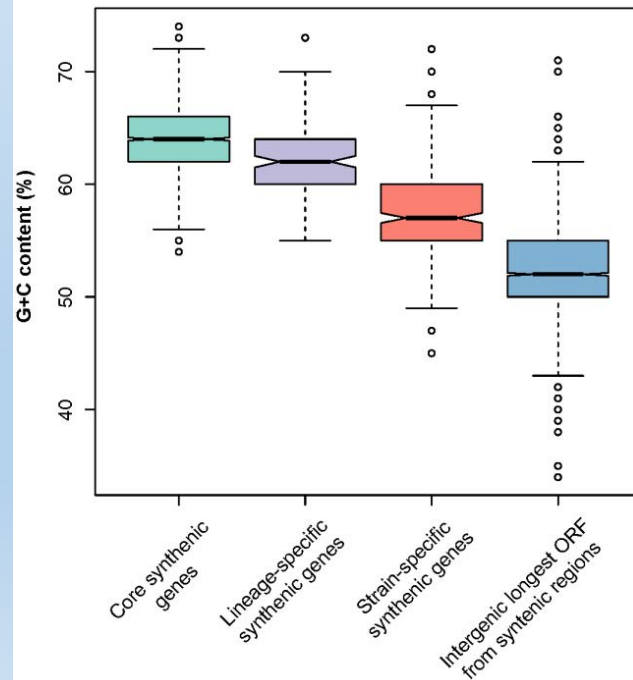
# Création *de novo* de gènes



GC

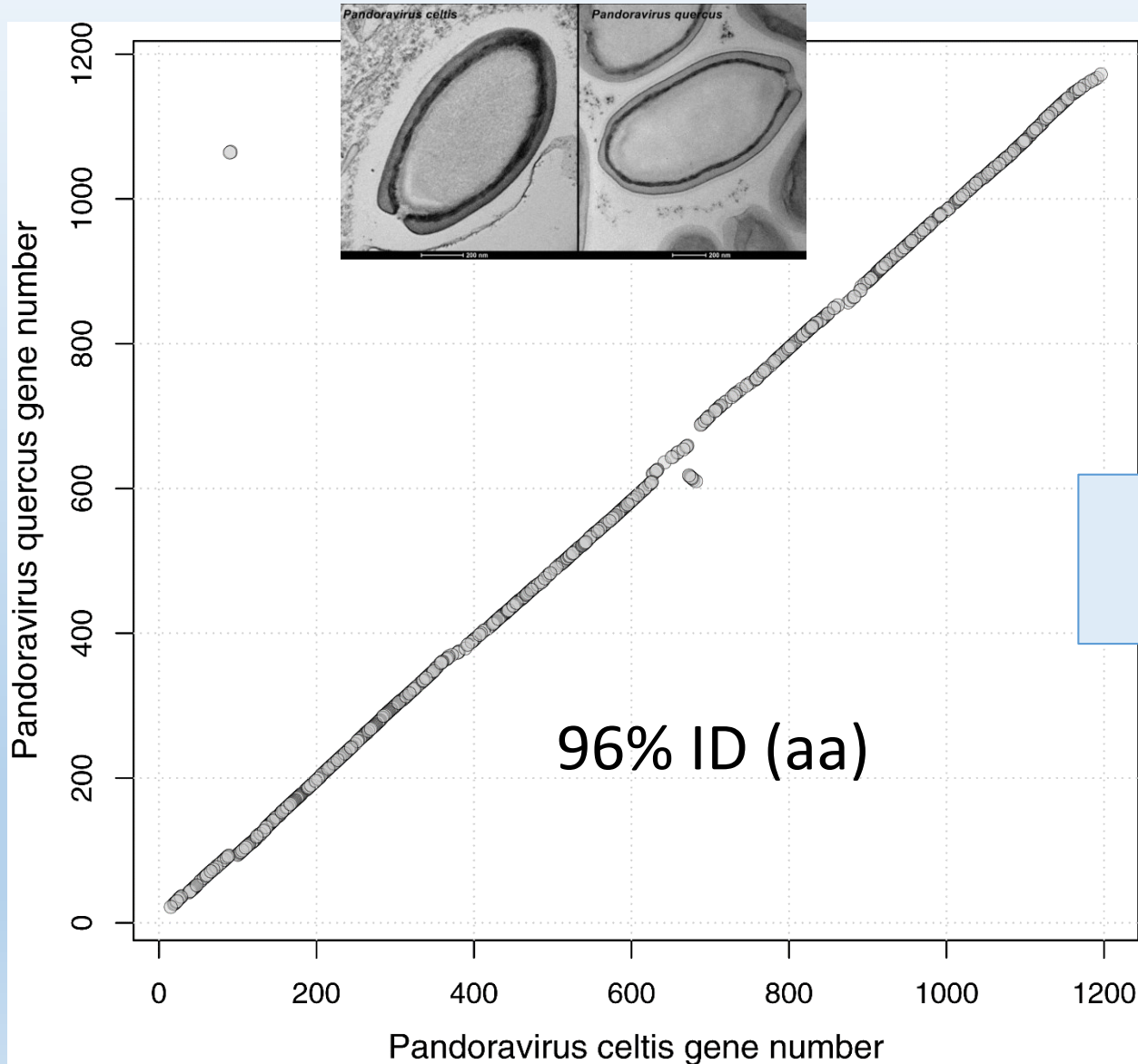
ORF length

CAI





# Micro-évolution: *P. celtis* et *P. quercus*



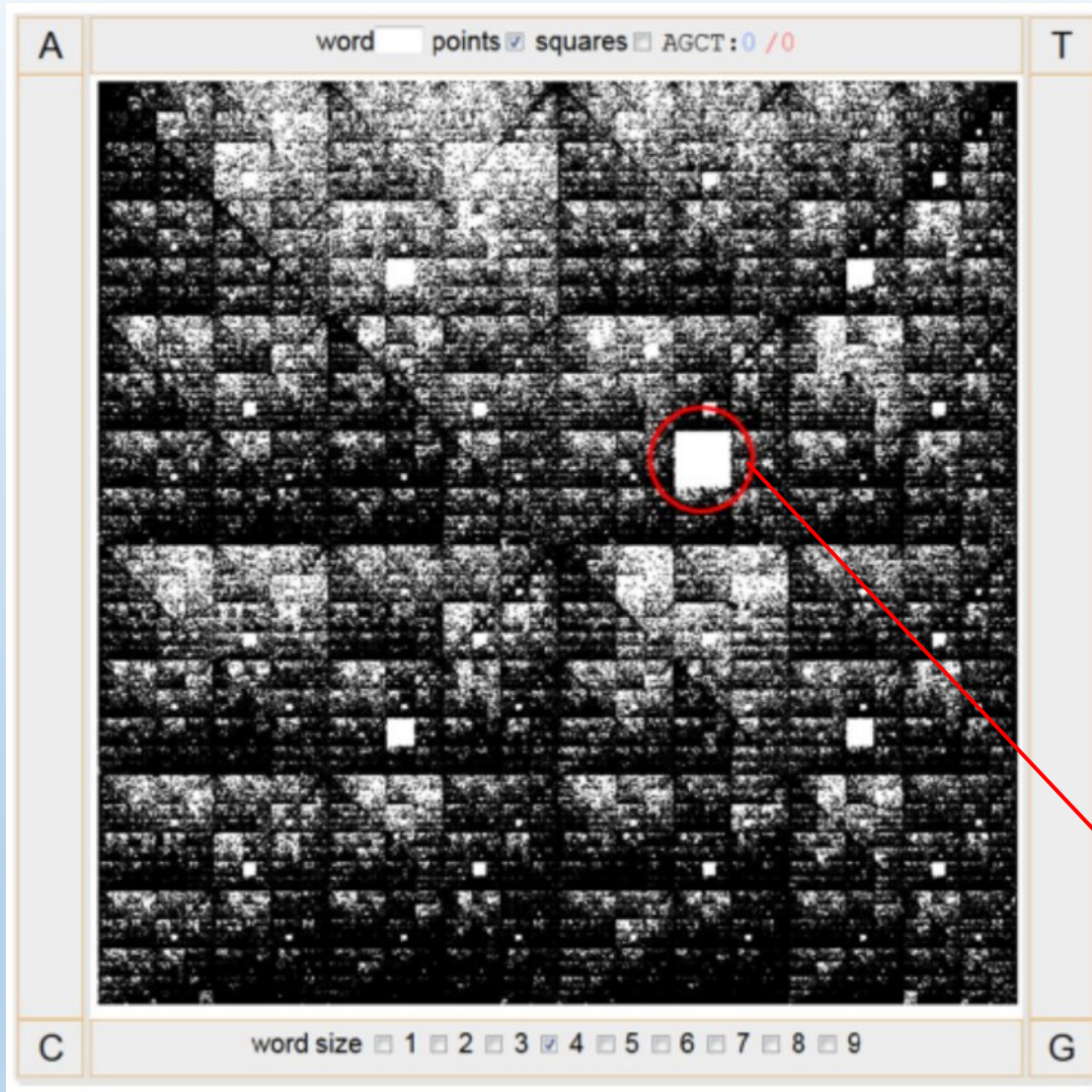
- P. celtis*: 9 unique ORFs
- P. quercus*: 21 unique ORFs
- lower G+C%
- lower CAI

Renforce l'hypothèse de création de gènes *de novo*.

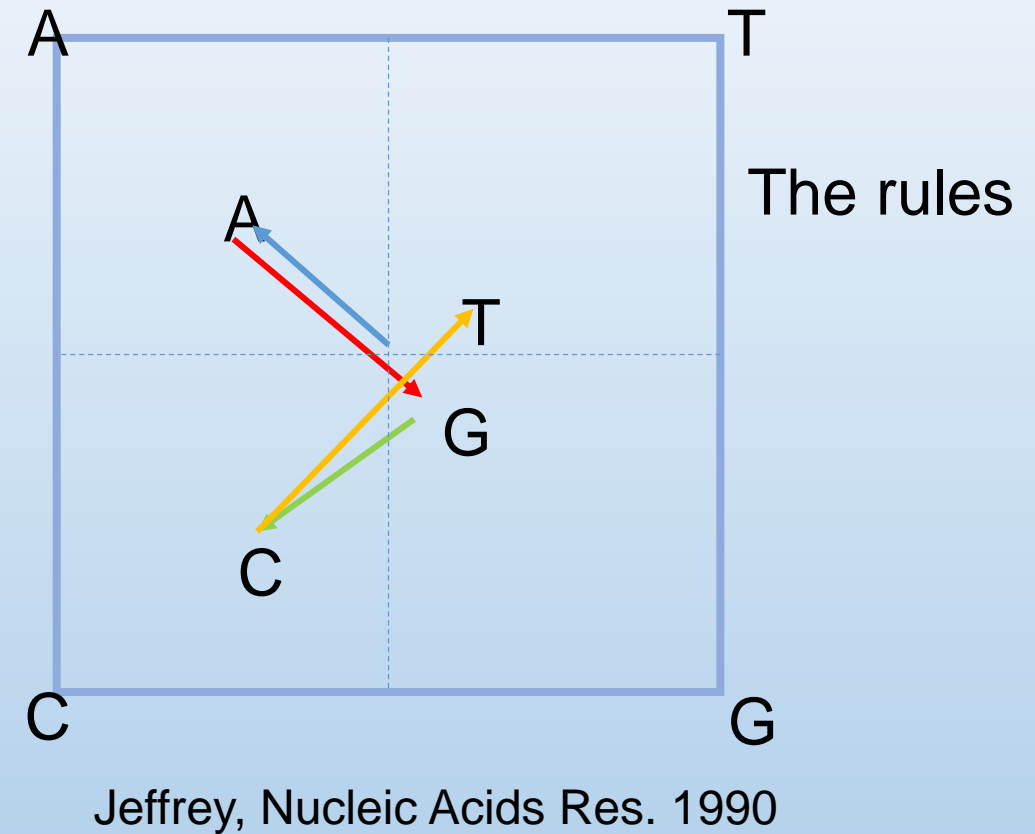
-----  
ATGGCGCAACGACCGGTAA  
ATGGCGCAACGACCGGTGA

ORF, puis transcription

# Le mystère du mot "AGCT" manquant (*P. dulcis*, *P. quercus*)



Chaos Game Representation



Absence of "AGCT"

# Eviter "AGCT" dans une séquence aléatoire de 2 Mb est impossible



$$\% (G+C) = 64\%$$

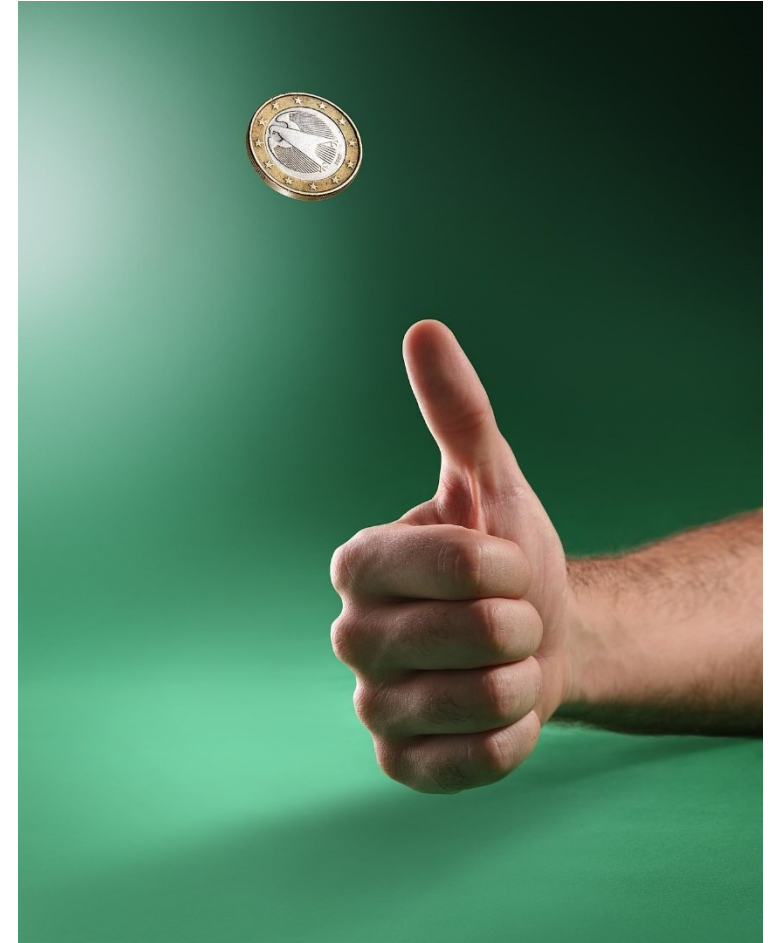
$$P_{AGCT} = 3.3 \cdot 10^{-3}$$

$$(1 - P_{AGCT})^{500.000} \approx 1.7/10^{718}$$

$$(Q_{AGCT})^4 \approx 8.35/10^{2872}$$

Après de multiples corrections de Bonferroni:

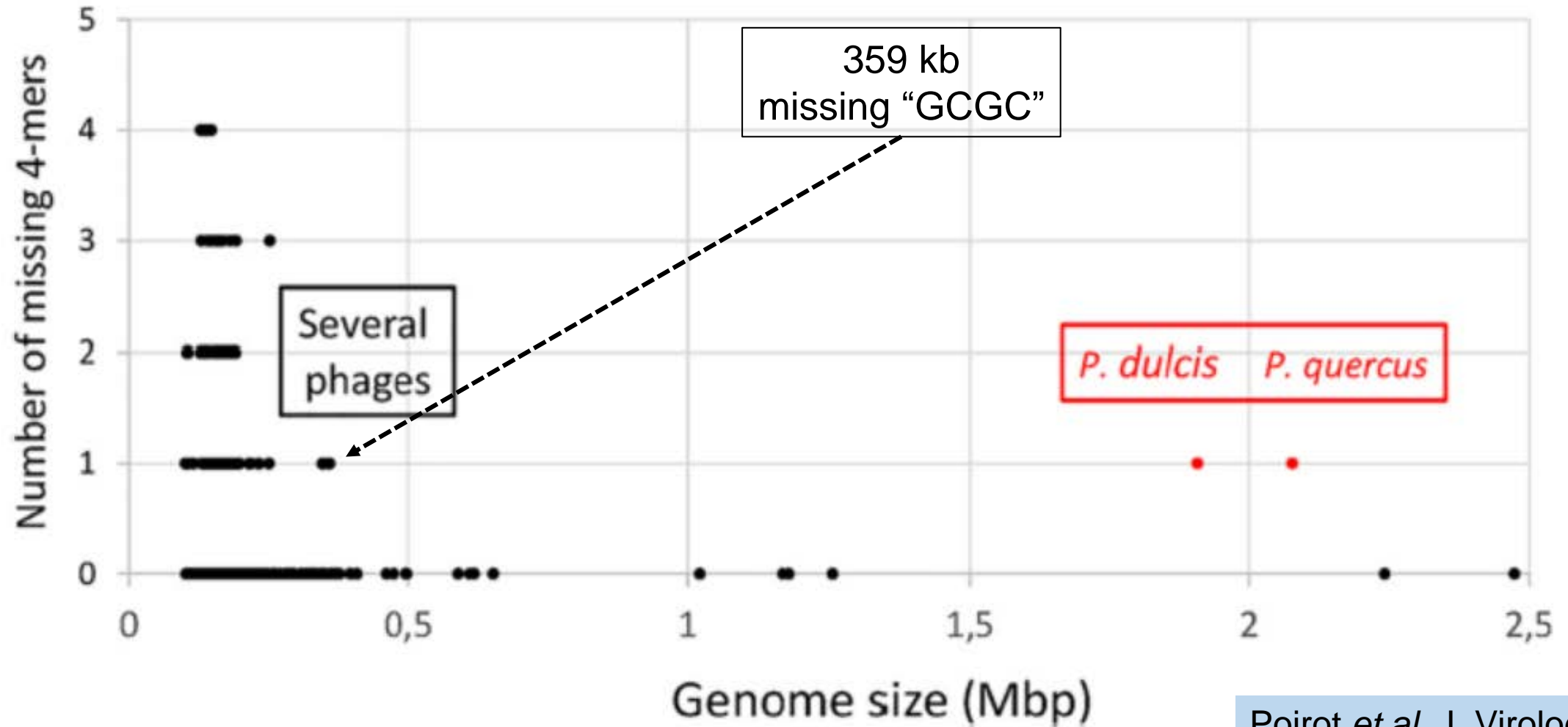
$$Q_{any/GenBank} \approx 4/10^{2864}$$



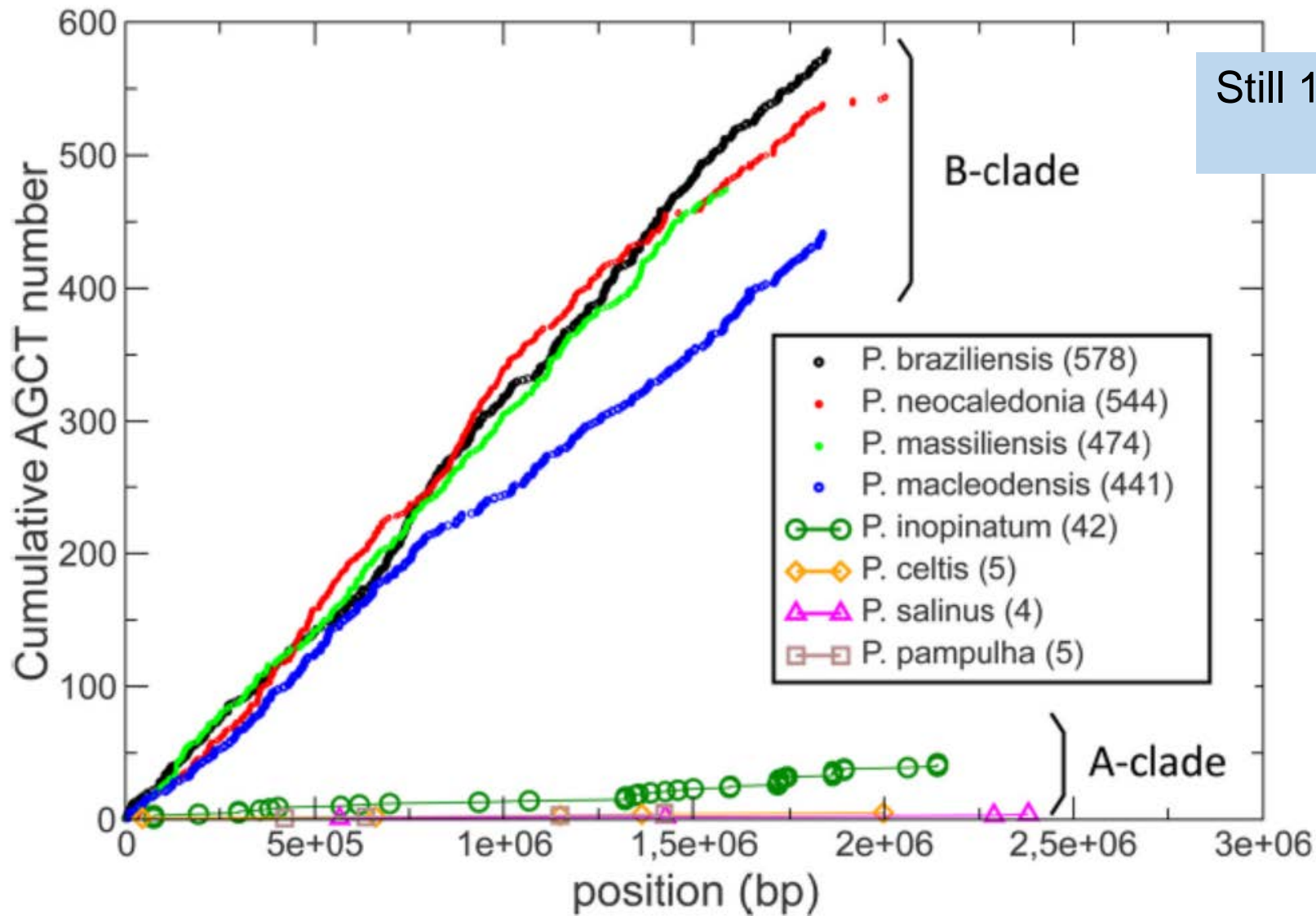
Pas de "face" après 9513 lancers d'une pièce!



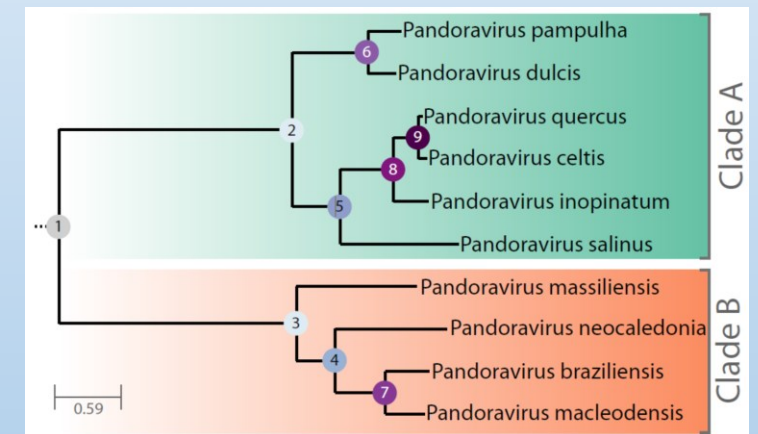
# les plus grand génomes sans un 4-mer



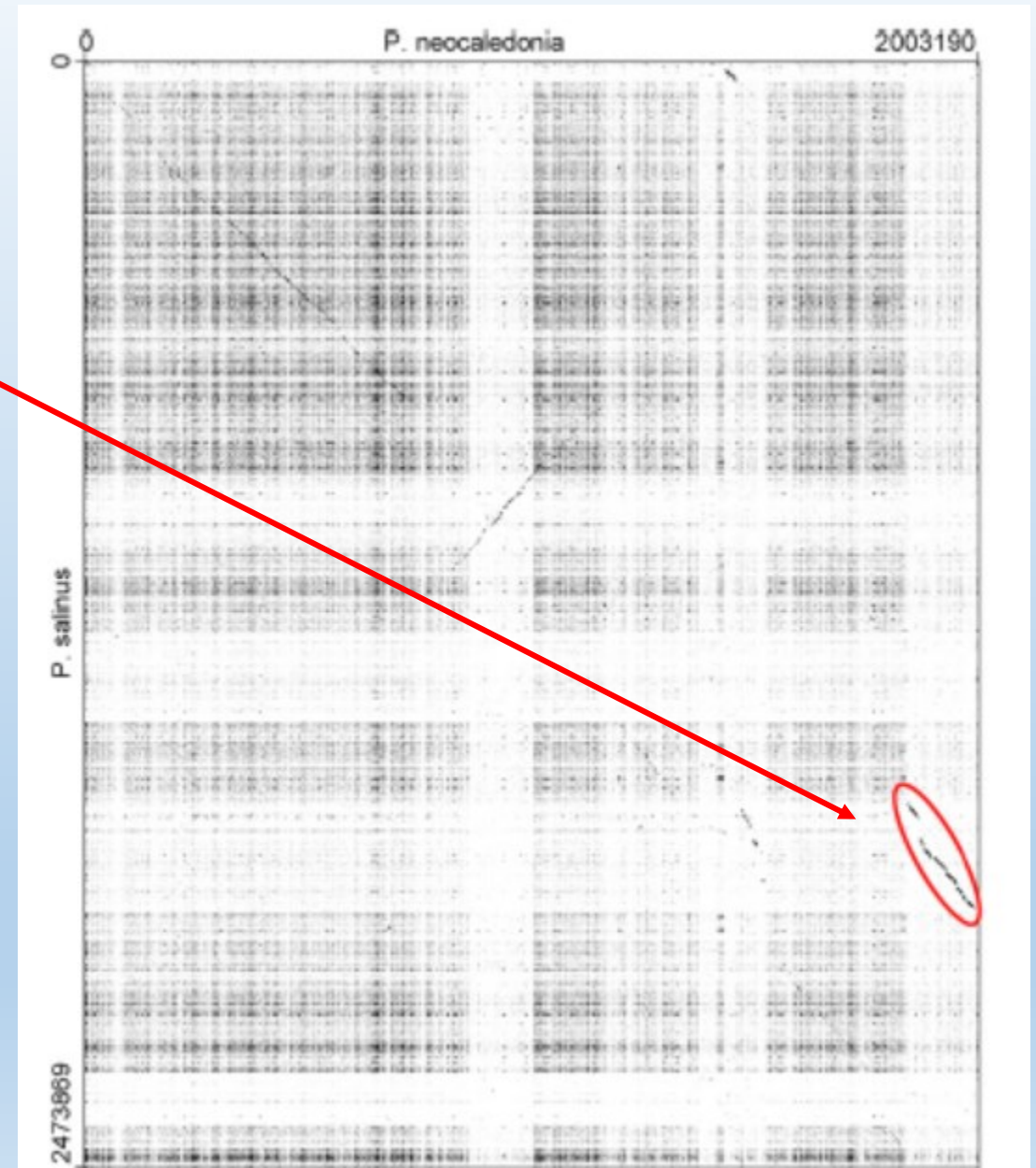
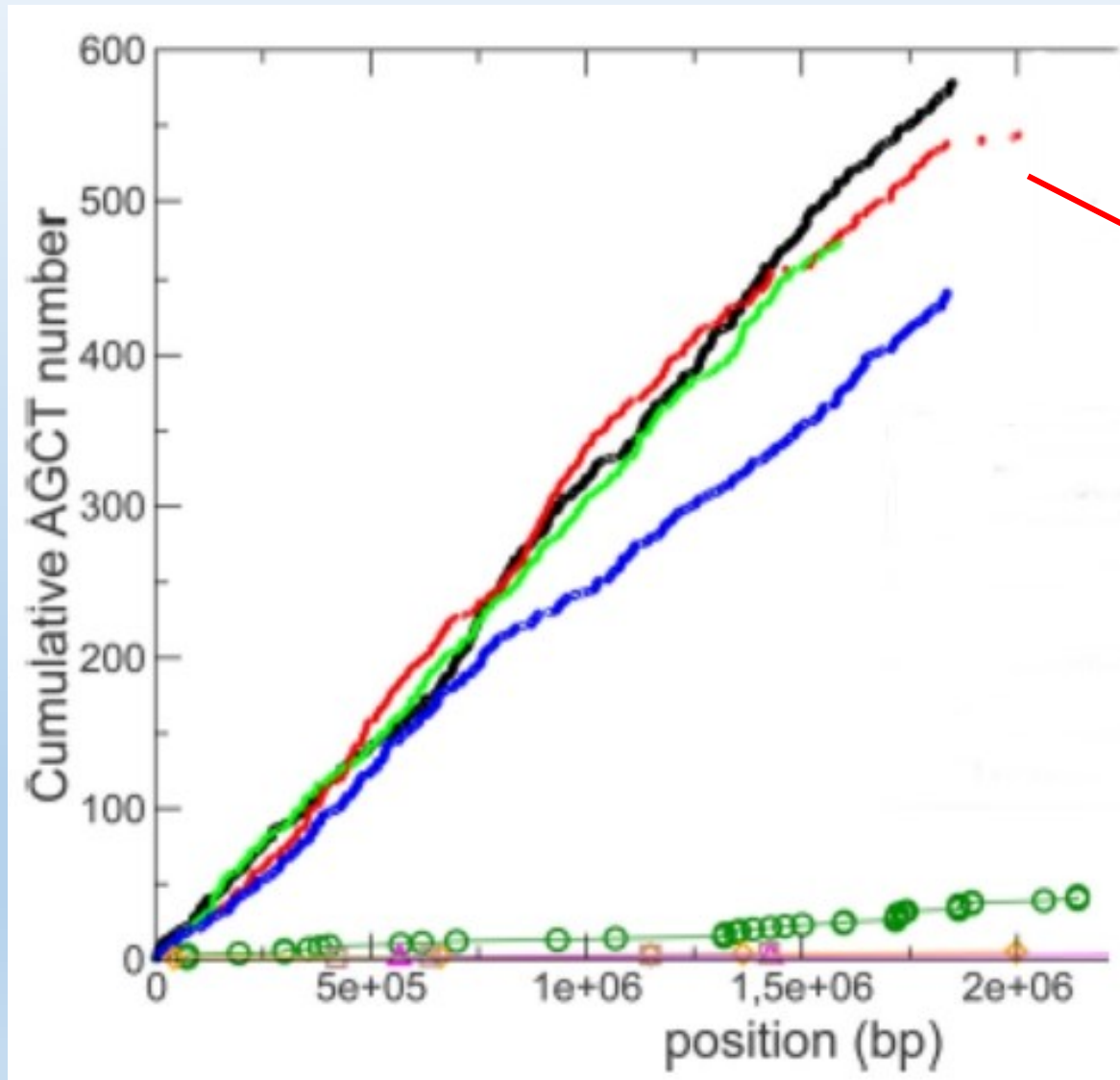
# La fréquence de "AGCT" corrèle avec la phylogénie



Still 10 times less than expected  
( $\approx 6000$ )

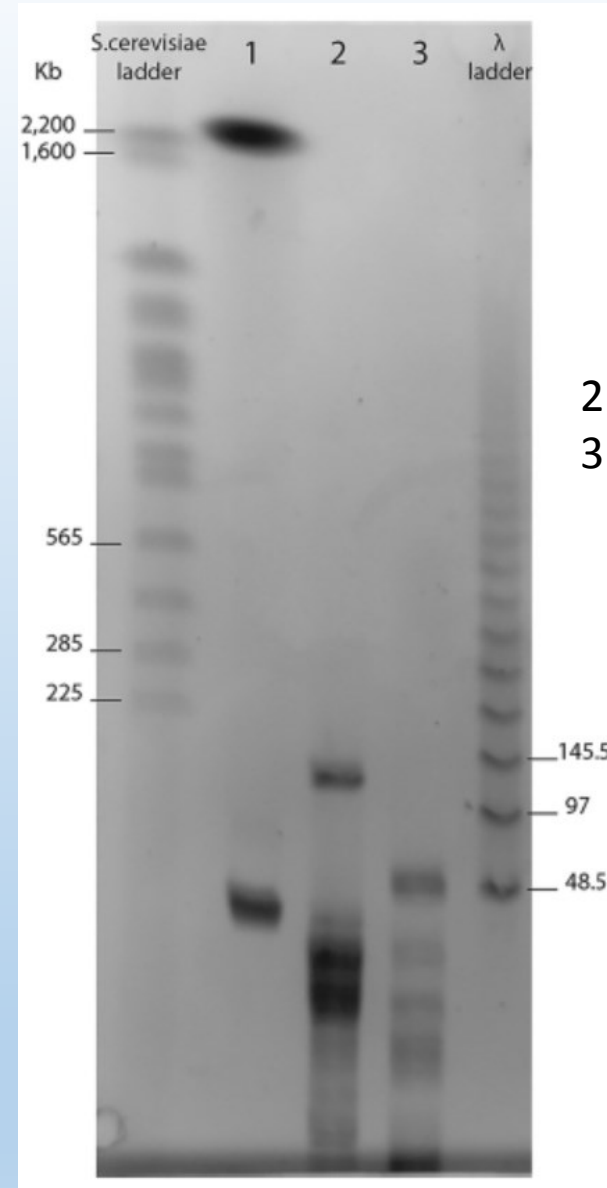


"AGCT" est un signal fort





Les "AGCT" encore  
présents dans le clade B  
ne sont pas modifiés  
(aussi PACBIO)



2: digestion PvuII (CAGCTG)  
3: digestion AluI (AGCT)

## Les autres familles de virus infectant Acanthamoeba

Date [reference]	Family	Virion (shape)	Virion Max size (nm)	Genome (kb)	GC %	Replication
2003 [1]	<u>Mimiviridae</u>	Icosahedron	755	1.500-370	25	Cytoplasm
2013 [12]	<u>Pandoraviridae</u>	Amphora	1200x500	2.800-1.850	61	Nucleus
2014 [13]	<u>Pithoviridae</u>	Amphora	2000x700	685-575	38	Cytoplasm
2015 [14]	<u>Molliviridae</u>	Sphere	600	650	60	Nucleus
2009 [33]	<u>Marseilleviridae</u>	Icosahedron	200	390-360	43	Nucleus/Cytoplasm
2015 [34]	<u>Faustoviridae</u>	Icosahedron	250	465-350	36	Nucleus/Cytoplasm
2017 [35]	<u>Medusaviridae</u>	Icosahedron	200	380	62	?

N'ont aucune anomalie concernant "AGCT"



Depuis 80 ans, nos connaissances bâtissent de nouveaux mondes



Merci pour votre attention.  
des questions ?



Bench:

- Jean-Marie Alempic
- Lionel Bertaux
- Estelle Grossetete
- Sandra Jeudy
- Audrey Lartigue
- Nadège Philippe

Computer:

- Virginie Collomb
- Matthieu Legendre
- Olivier Poirot
- Sebastien Santini



Dr. Chantal Abergel



European Research Council  
Established by the European Commission

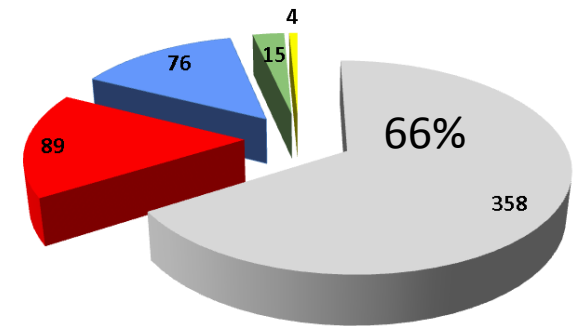
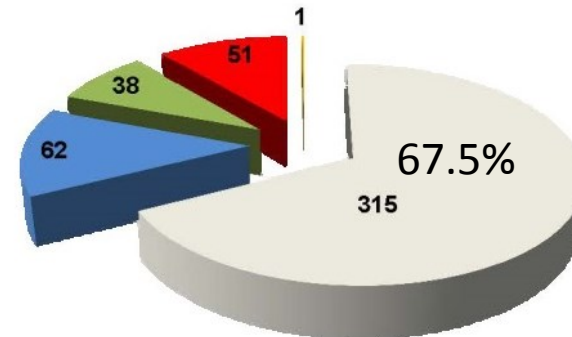
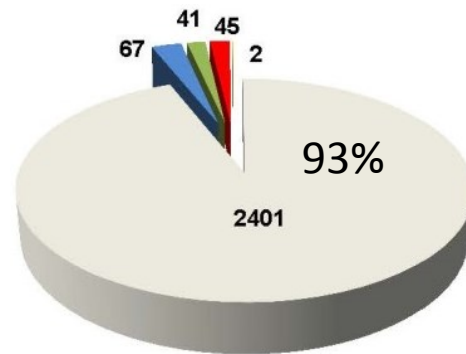
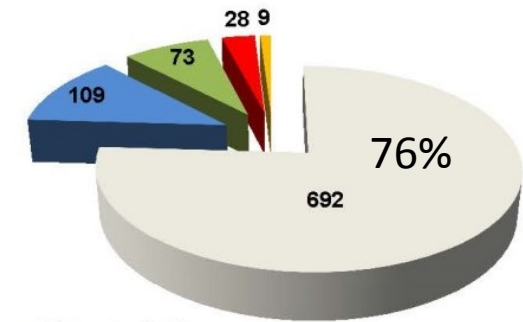
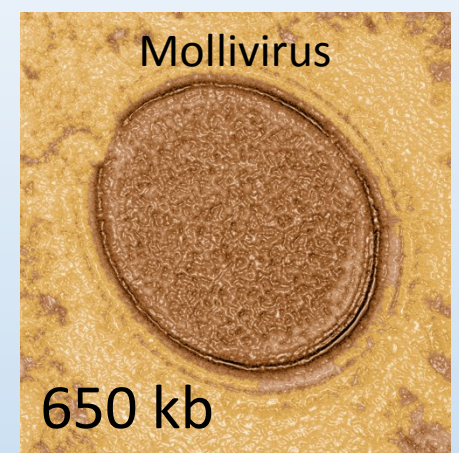
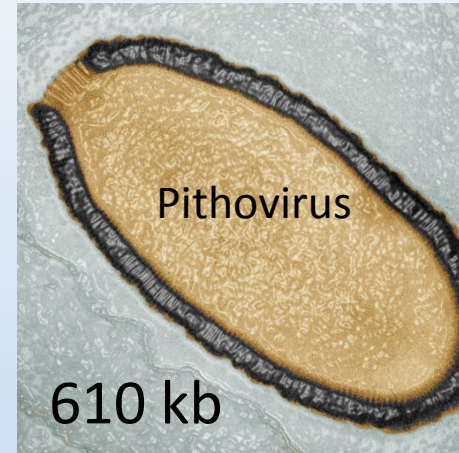
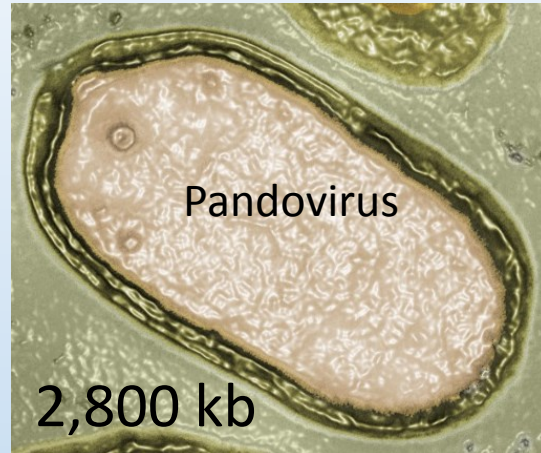
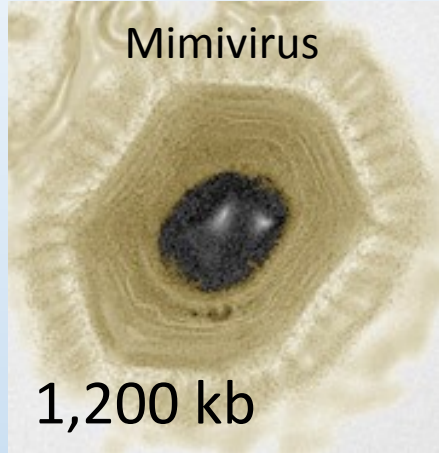


# "AGCT" is specifically under represented

Parameter	Value for:	
	<i>P. dulcis</i>	<i>P. quercus</i>
Genome size (bp)	1,908,524	2,077,288
Frequency (strand 1)		
AGC		
Inter-ORF	0.0101 (1/99)	0.0098 (1/102)
ORF	0.0112 (1/89)	0.0110 (1/90)
Global	0.0109 (1/92)	0.0106 (1/94)
GCT		
Inter-ORF	0.0102 (1/98)	0.0097 (1/103)
ORF	0.0156 (1/64)	0.0145 (1/68)
Global	0.0138 (1/72)	0.0129 (1/77)
AGC/GCT (2 strands, global)	0.0123 (1/81)	0.0118 (1/85)
AGC/GCT overall rank	37/64	43/64
$P_{(AGC)} \times P_{(T)}$	$2.24 \times 10^{-3}$ (1/446)	$2.31 \times 10^{-3}$ (1/432)
AGCT expected no. [one strand $\times P_{(AGC)} \times P_{(T)}$ ]	4,286	4,898
AGCT observed no.	0	0
ACGT expected no. [one strand $\times P_{(ACG)} \times P_{(T)}$ ]	7,884	8,387
ACGT observed no.	5,822	6,165

Expected from 3-mers

# The ORFan enigma



- No match
- Eukaryota
- Bacteria
- Virus
- Archaea

For each new family >2/3 ORFANs  
although they manipulate the same building blocks than cells:  
nucleic acids, nucleotides, sugars, lipids, amino acids...

# Functional oddities:

More than 2000 protein-coding genes, but a strong cell dependency

**Encode a complete transcription machinery,  
but not loaded in the virion -> nucleus dependency**

- No topoisomerase
- No DNA ligase
- No sliding clamp
- No DNA repair

**Almost no nucleotide metabolism**

- No thymidine kinase
- No thymidylate synthase

**Only 4 of the 8 most conserved NCLDV "core genes"\***

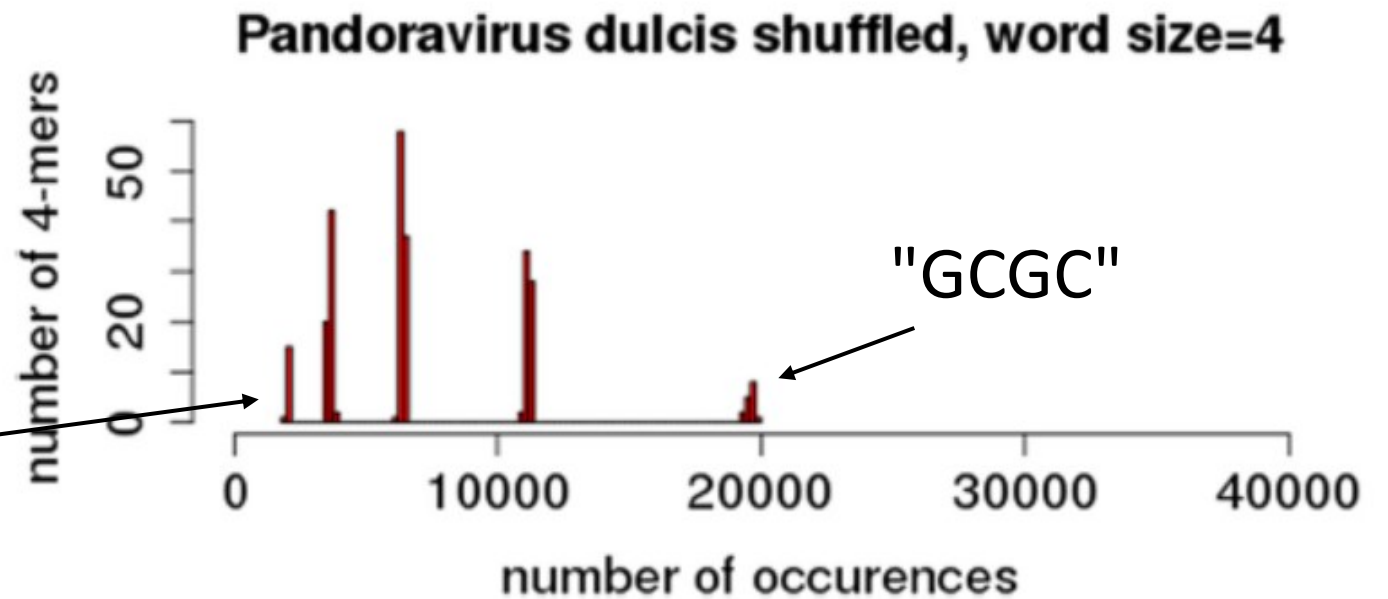
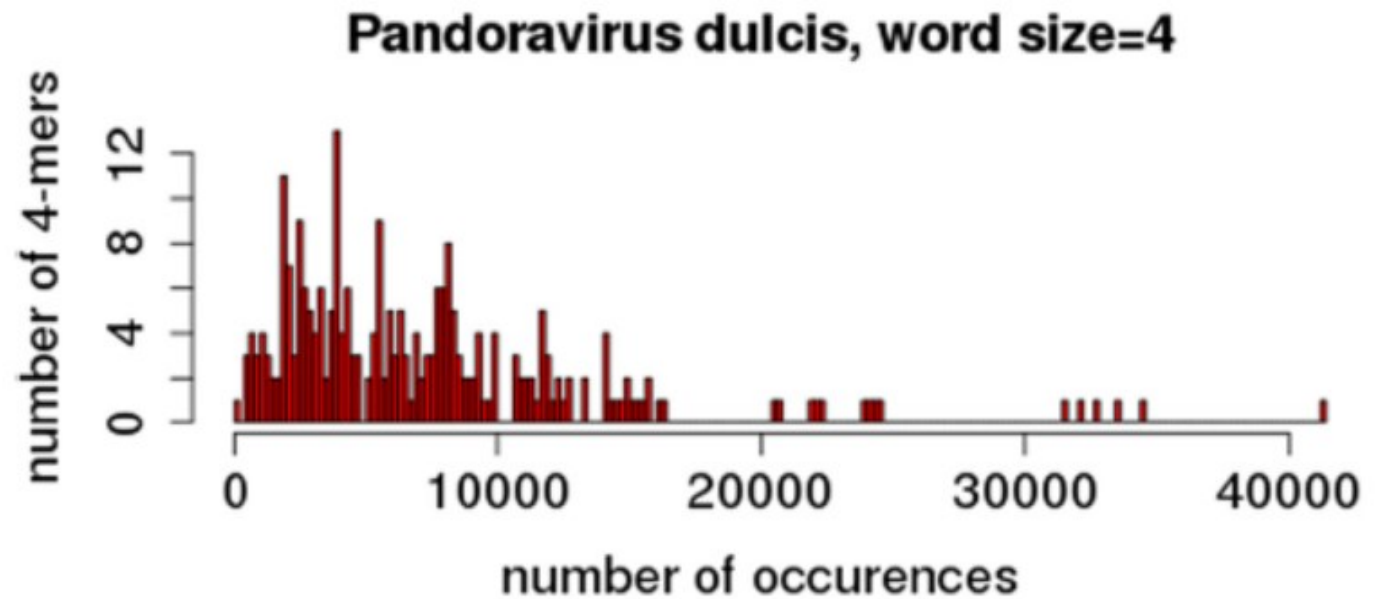
- No major capsid protein (jelly-roll)
- No VV D6R helicase
- No F10L-like protein kinase
- No VLTF2-like transcription factor

\* DNA polB, VV A18 helicase, Thiol oxidoreductase, VV A32 virion packaging ATPase



# Natural vs randomized 4-mer distribution

Multiple evolutionary constraints pictured at once



Pandoraviridae

Mimiviridae

Cladistic tree

Legendre *et al.* 2018

Pithoviridae

