

# Giant Viruses !

Structural & Genomic Information Lab.  
CNRS Aix-Marseille Univ.

J-M. Claverie

[Jean-Michel.Claverie@univ-amu.fr](mailto:Jean-Michel.Claverie@univ-amu.fr)

Cours International de Bioinformatique et  
Analyse de Génomes, Tunis 2017



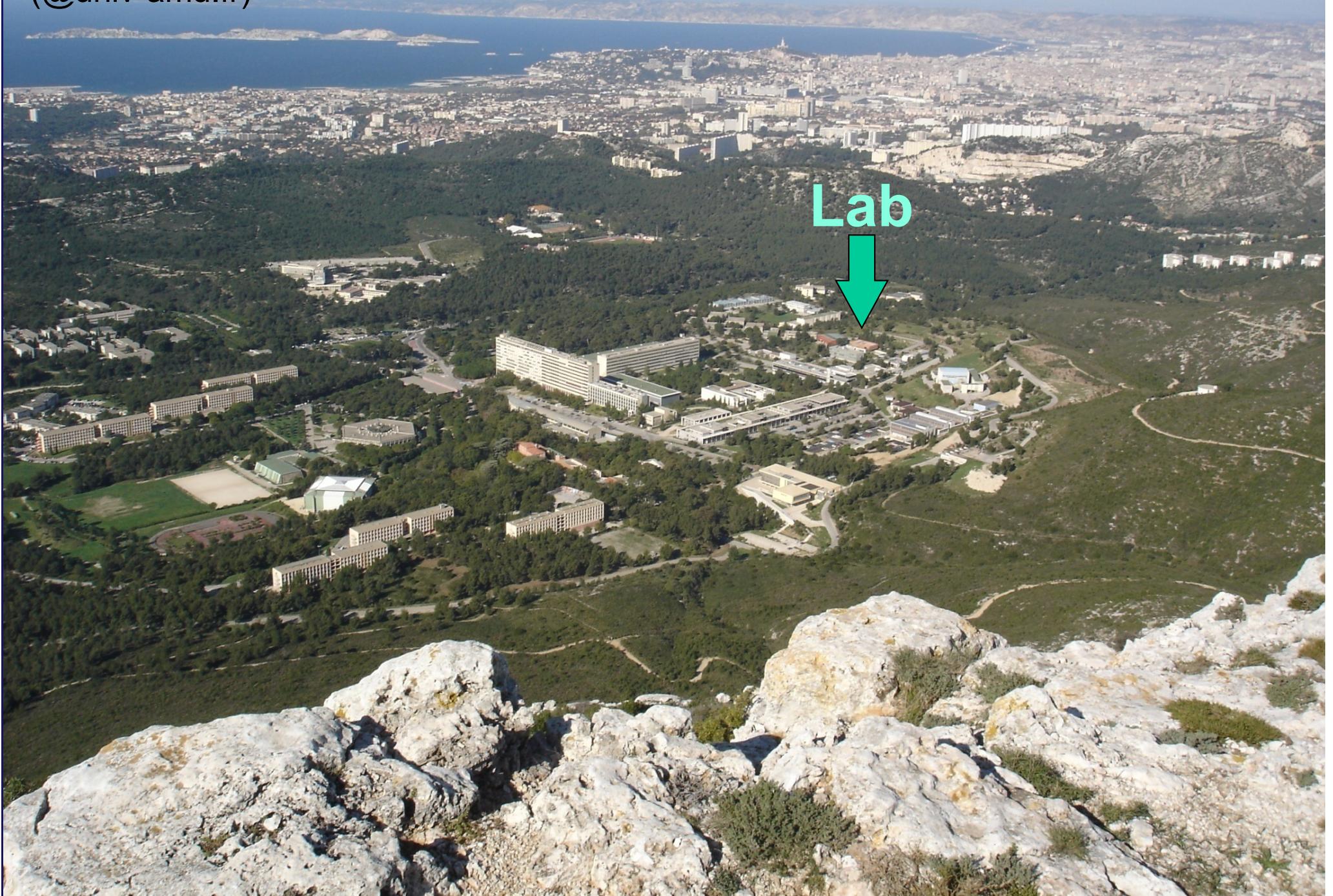
## Marseille



# Parc des Calanques (cliffs of limestone)



# Aix-Marseille University – Luminy campus (@univ-amu.fr)





CENTRE NATIONAL  
DE LA RECHERCHE  
SCIENTIFIQUE

Aix\*Marseille  
université

Information  
Génomique et  
Structurale



28 members (19)

A component of  
the Mediterranean  
Institute of Microbiology  
(IMM)

300 p, 5 CNRS Labs  
12 technical platforms



[Jean-Michel.Claverie@univ-amu.fr](mailto:Jean-Michel.Claverie@univ-amu.fr)  
[Chantal.Abergel@igs.cnrs-mrs.fr](mailto:Chantal.Abergel@igs.cnrs-mrs.fr)

# The Germ theory

(Pasteur 1822-95, Koch 1843-1910)\*

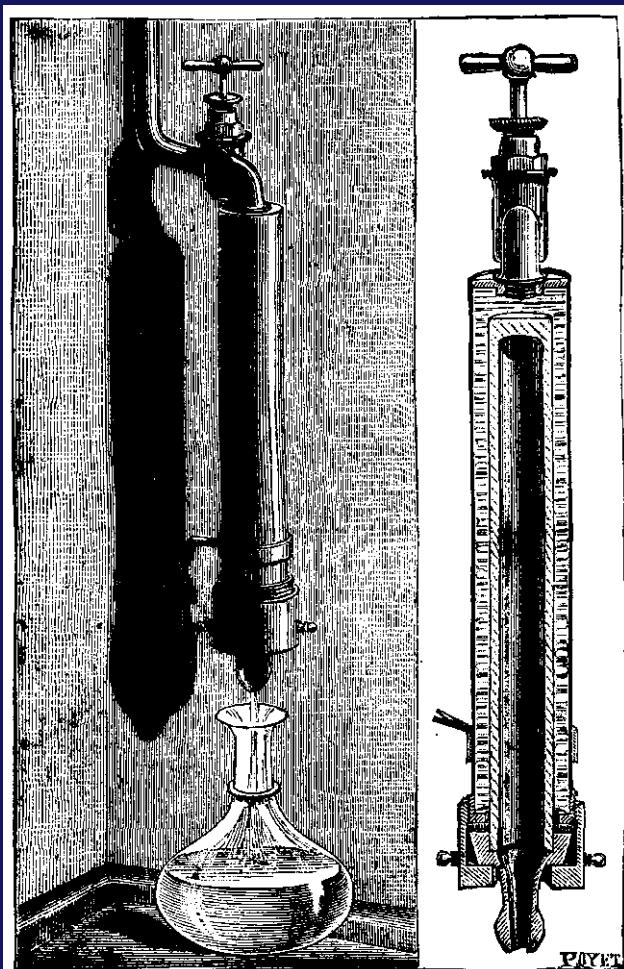
- Pasteur: Infectious diseases are due to microbes (1878): silkworms disease (microsporidie), Fowl cholera.
- Koch: Koch's postulates
  - Isolation of *Bacillus anthracis* (1877), *Bacillus tuberculosis* (1882) & *Vibrio cholera* (1883)(**Petri** dish + agar)

End of the miasma theory

\*[Antoni van Leeuwenhoek (1632 -1723)]

# The Chamberland filter

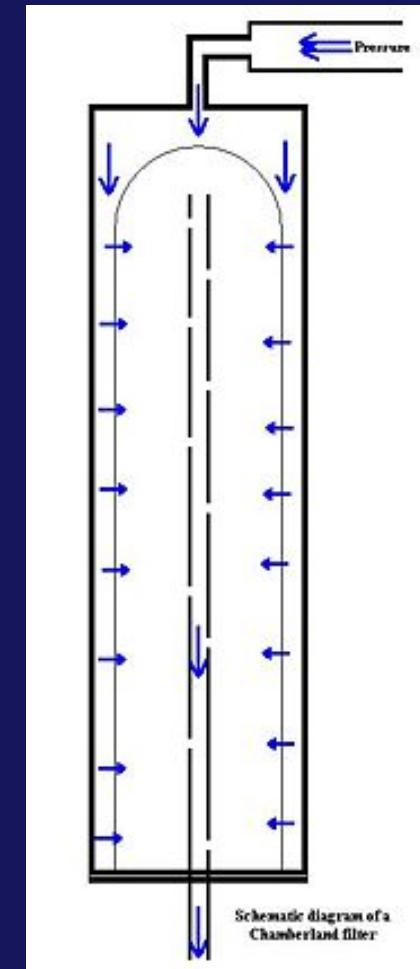
*Microbe retention as the standard protocol (1884)*



Typhoid fever

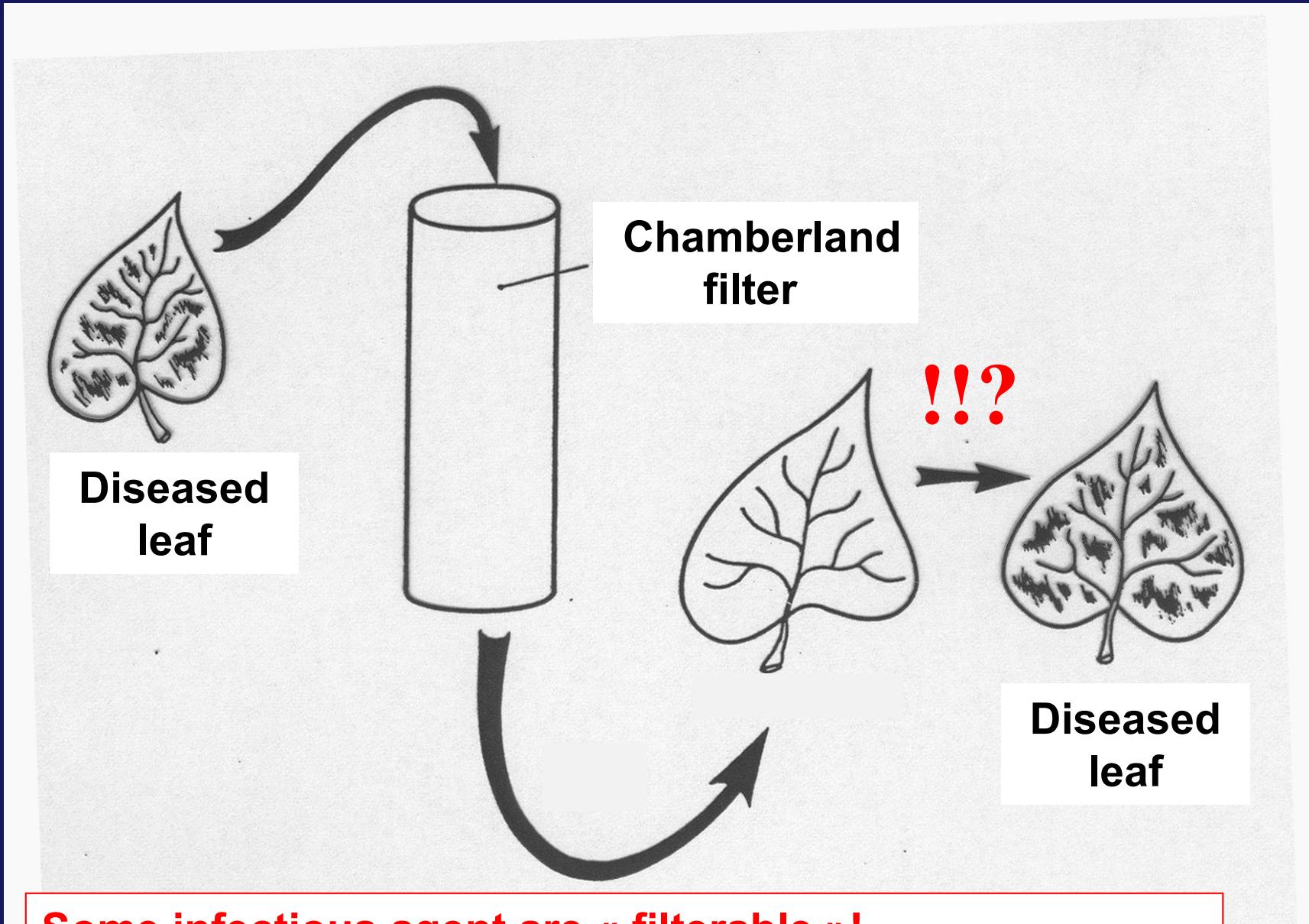
100% > 0.9 micron

98% > 0.5 micron



Porcelain filter

# Dimitri Ivanovski discovers the first “virus”(1892)



**Some infectious agent are « filterable »!**  
Soon after: rabies, foot and mouth disease, myxomatosis, etc

# What is a “virus” ?

First modern definition by A. Lwoff (1957):

Viruses are defined by negative properties:

- 1) Not visible by light microscopy
- 2) Not retained by the Chamberland filter
- 3) Not cultivable
- 4) No energy production
- 5) No translation (no ribosome)
- 6) No division

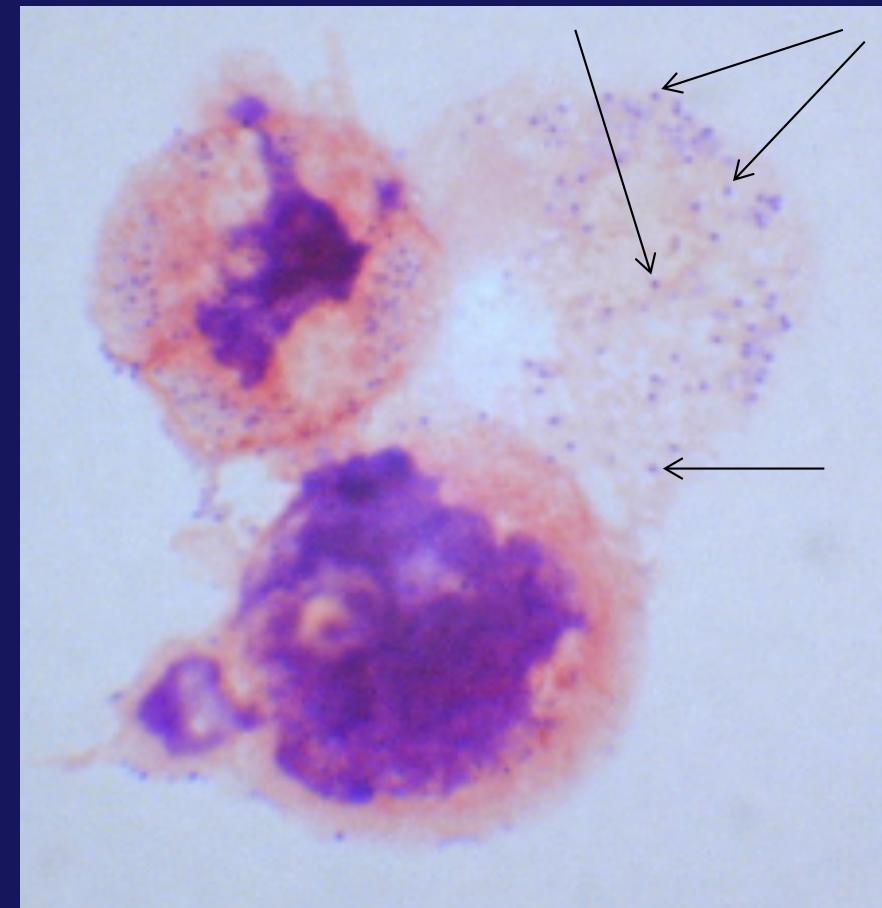
# Mimivirus discovery

(looking for Legionella inside Acanthamoeba)

In a cooling tower in the Bradford hospital, UK  
in search of pneumonia agents



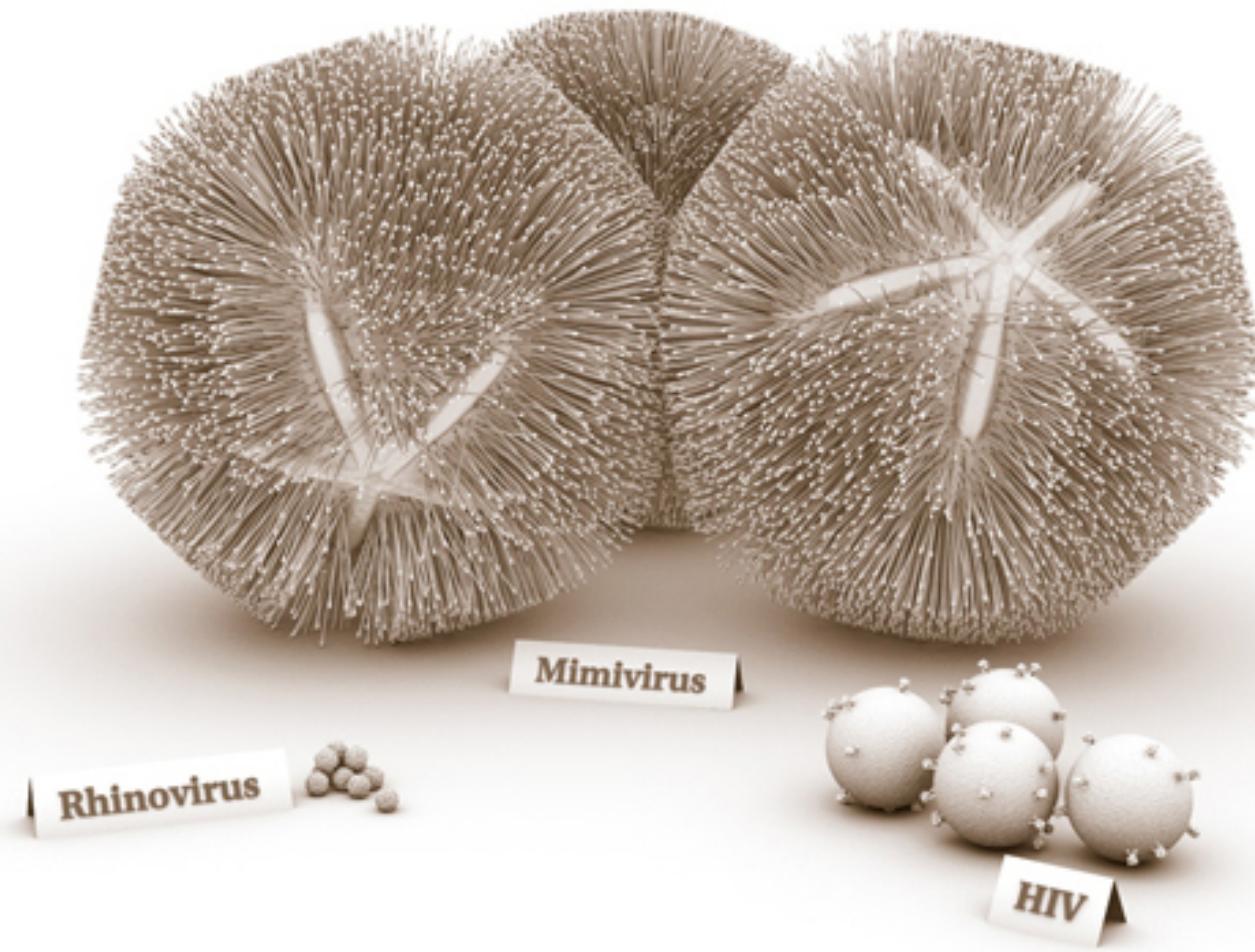
Dr. Tim Rowbotham, 1992  
("Bradfordcoccus")



AMERICAN  
**Scientist**

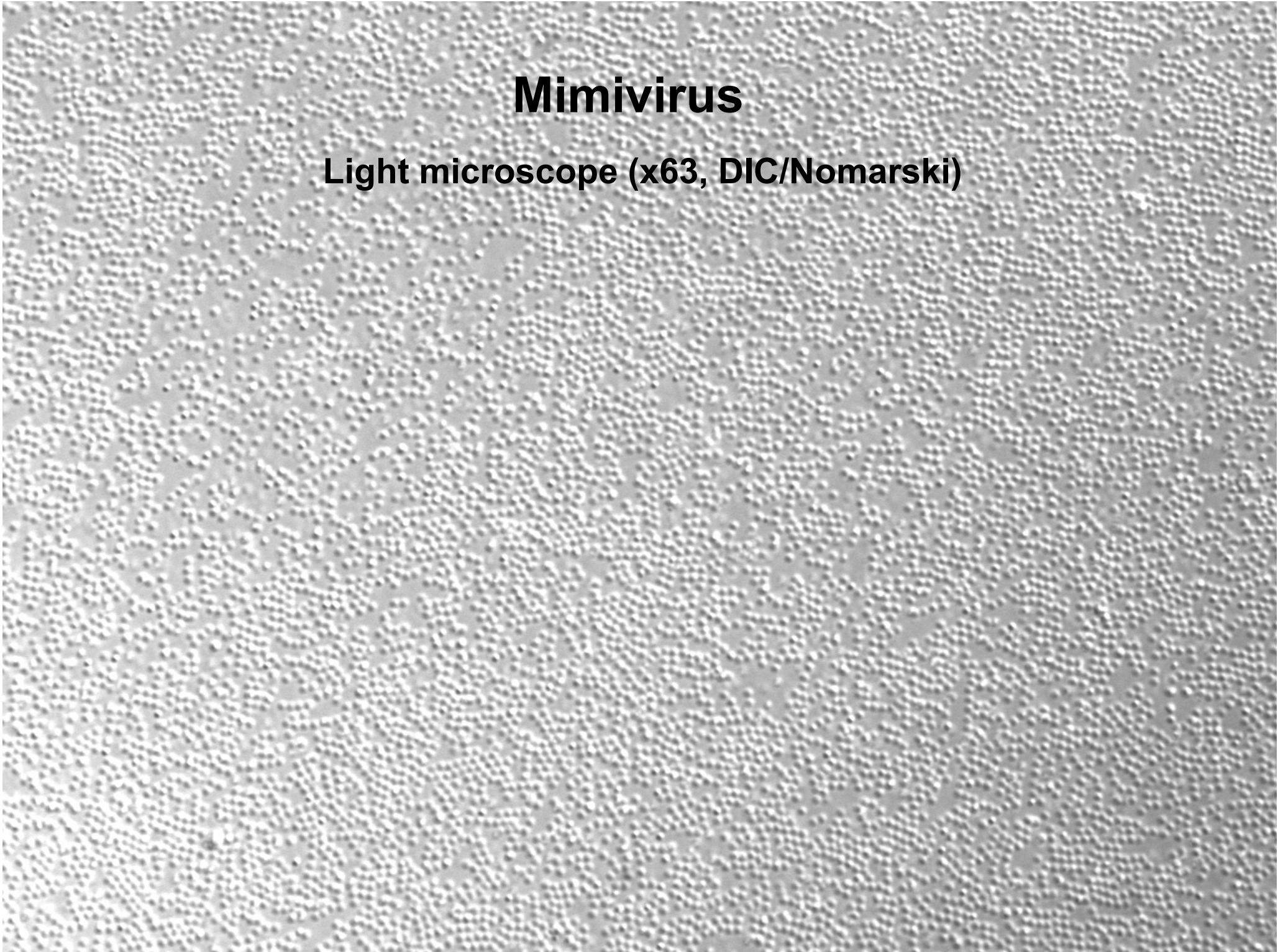
July–August 2011

[www.americanscientist.org](http://www.americanscientist.org)

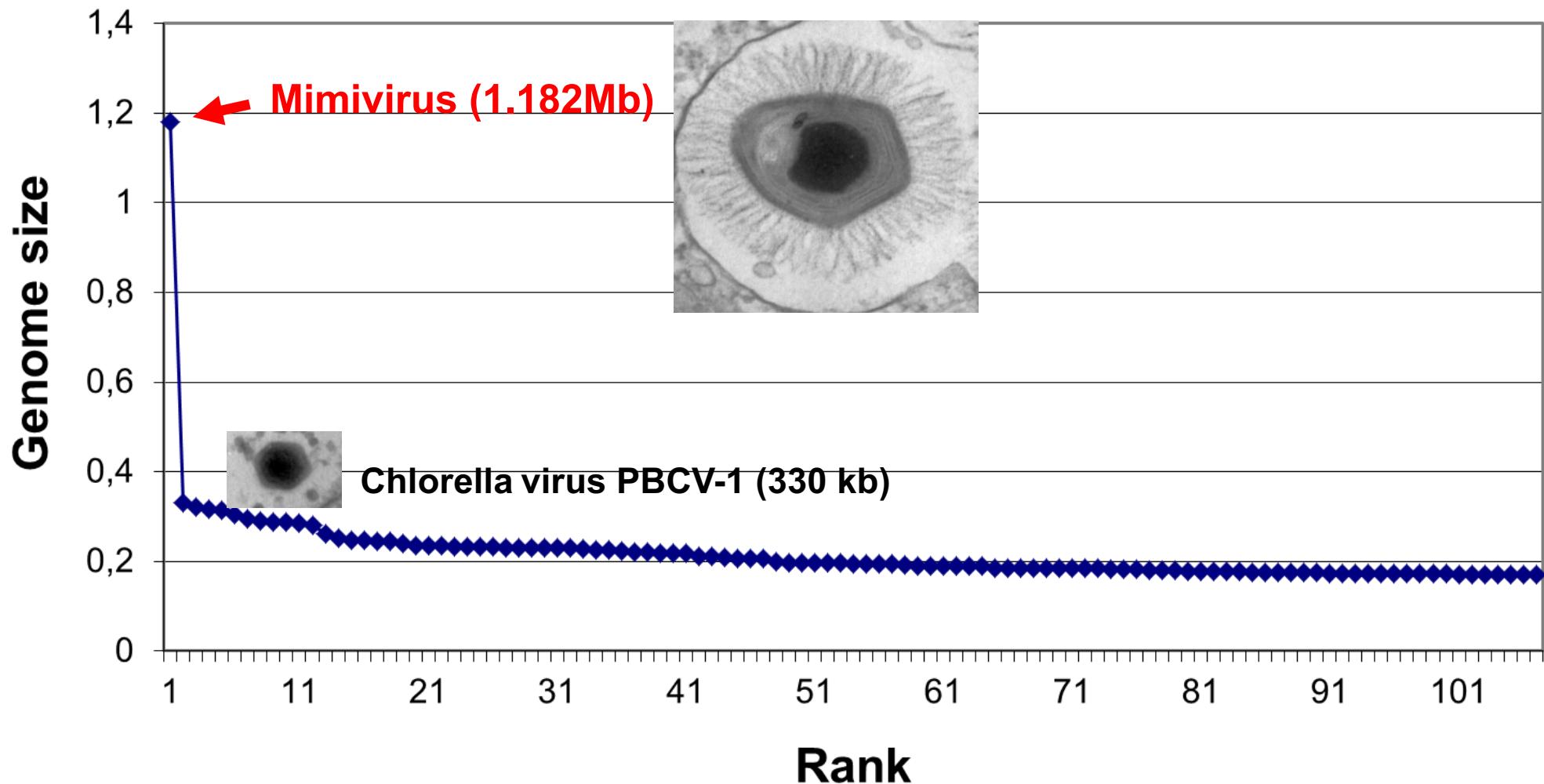


# **Mimivirus**

**Light microscope (x63, DIC/Nomarski)**



# Leaping into the giant virus era (2004-20xx)



# A viral genome larger than that of many bacteria

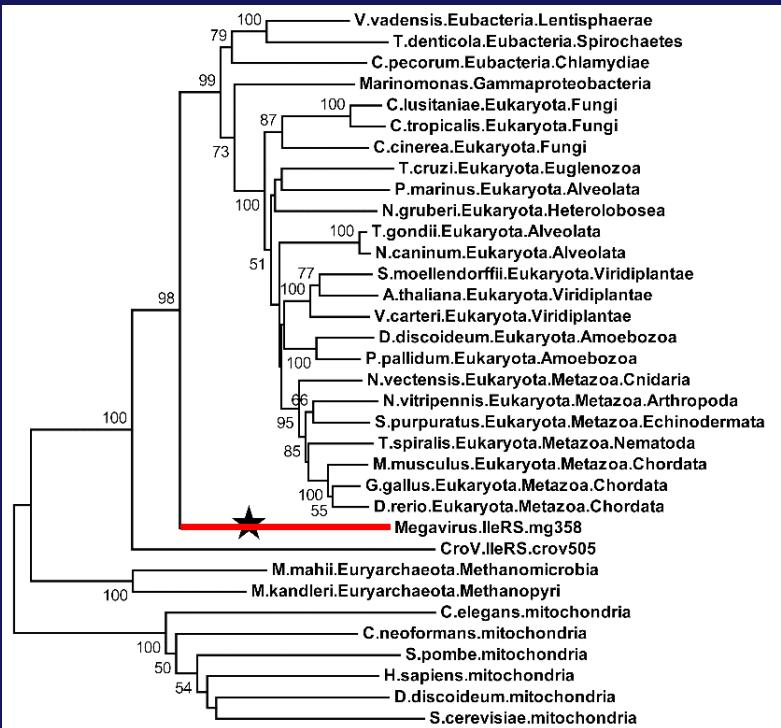


Smaller genomes (cellular organisms)		
<i>Carsonella ruddii</i>	0.16	Proteobacteria: gamma
<i>Buchnera aphidicola str. BCc</i>	0.42	Proteobacteria; gamma
<i>Nanoarchaeum equitans Kin4-M</i>	0.49	Archaea
<i>Mycoplasma genitalium G-37</i>	0.58	Firmicutes; Mollicutes
<i>Buchnera aphidicola str. APS</i>	0.66	Proteobacteria; gamma
<i>Buchnera aphidicola str. Sg</i>	0.64	Proteobacteria; gamma
<i>Wigglesworthia glossinidia</i>	0.70	Proteobacteria; gamma
<i>Candidatus Blochmannia floridanus</i>	0.71	Proteobacteria; gamma
<i>Ureaplasma parvum</i>	0.75	Firmicutes; mollicutes
<i>Mycoplasma mobile 163K</i>	0.78	Firmicutes; mollicutes
<i>Mycoplasma pneumoniae M129</i>	0.82	Firmicutes; mollicutes
<i>Onion yellows phytoplasma</i>	0.86	Firmicutes; mollicutes
<i>Tropheryma whipplei (Whipple)</i>	0.93	Actinobacteria
<i>Mycoplasma pulmonis UAB</i>	0.96	Firmicutes; mollicutes
<i>Mycoplasma gallisepticum</i>	1.00	Firmicutes; mollicutes
<i>Chlamydia trachomatis</i>	1.04	Chlamydiae
<i>Chlamydia muridarum</i>	1.08	Chlamydiae
<i>Rickettsia prowazekii (typhus)</i>	1.11	Proteobacteria; alpha
<i>Treponema pallidum (syphilis)</i>	1.14	Spirochaetes
<i>Chlamydophila caviae</i>	1.173	Chlamydiae
<b>MIMIVIRUS</b>	<b>1.181</b>	<b>ds-DNA virus</b>

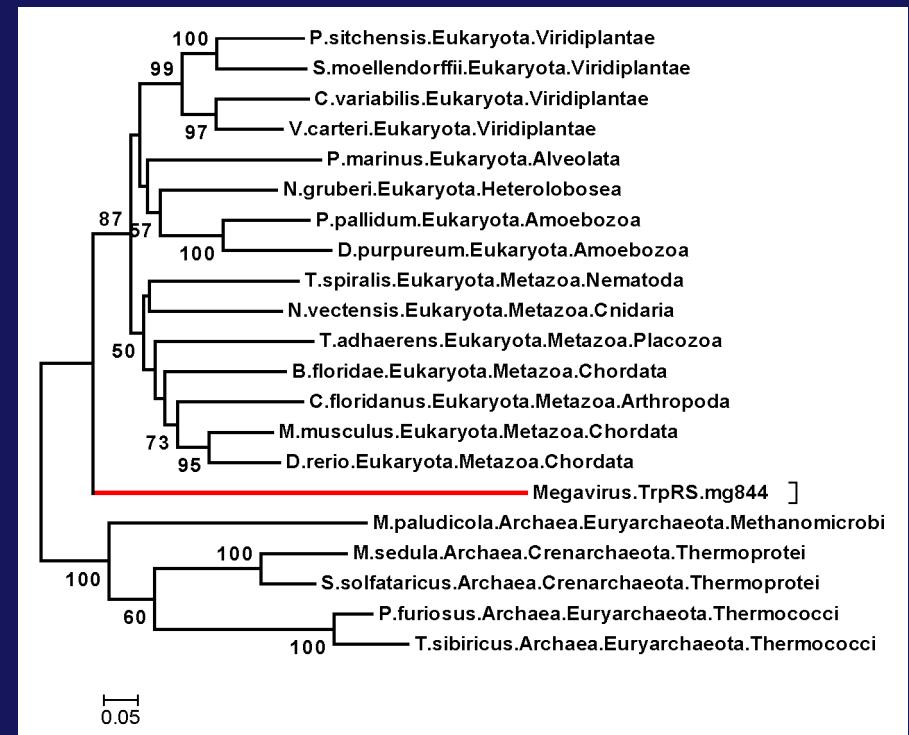
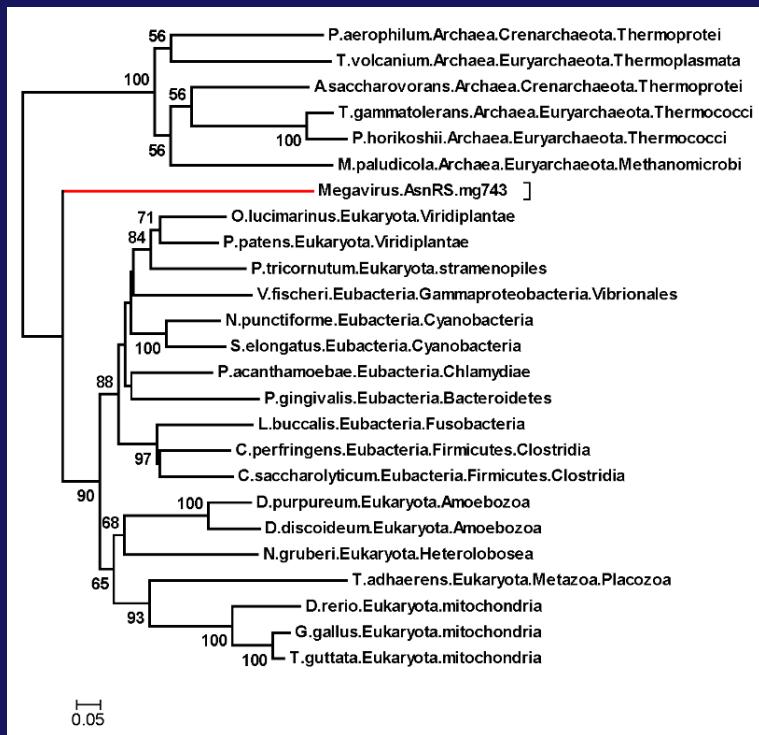
# The most unexpected feature: translation remnants

ORF #	Definition/putative function	Comment
R663	Arginyl-tRNA synthetase	Translation
L124	Tyrosyl-tRNA synthetase	Translation
L164	Cysteinyl-tRNA synthetase	Translation
R639	Methionyl tRNA synthetase	Translation
R726	Peptide chain release factor eRF1	Translation END
R624	GTP binding elongation factor eF-Tu	Translation ELONG
R464	Translation initiation factor SUI1	Translation START
L496	Translation initiation factor 4E (mRNA cap binding)	Translation START
R405	tRNA (Uracil-5-)-methyltransferase	tRNA modification

+ 6 tRNA: 3 Leu, Trp, His, Cys



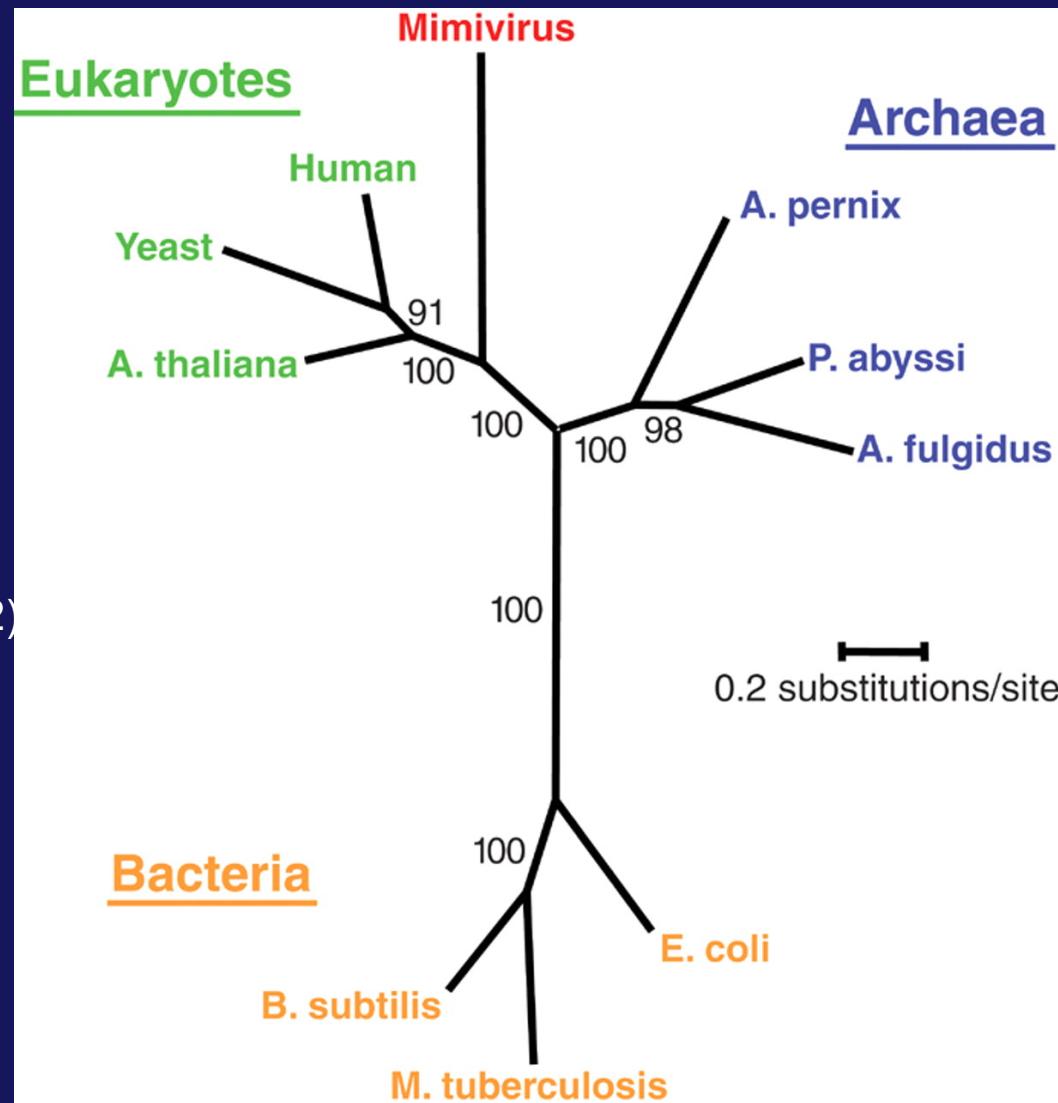
# The viral aa-tRNA ligase are not recent horizontal transfers



# Mimivirus & the three domains of life (Eukaryota, Eubacteria, and Archaea)

(concatenated sequences of seven universally conserved protein sequences)

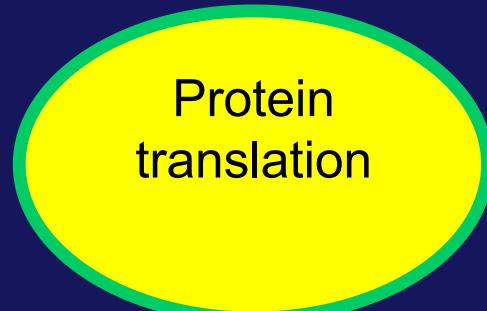
The alignment contains 3164 sites.



Arg-tRNA synthetase  
Met-tRNA synthetase  
Tyr-tRNA synthetase  
RNA pol II (RPB1)  
RNA polymerase II (RPB2)  
PCNA  
5'-3' exonuclease

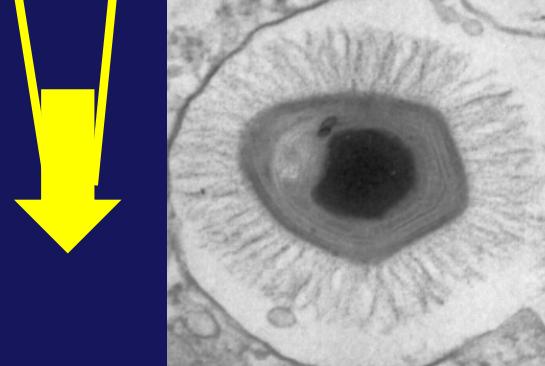
Science 2004;306:1344-1350

# A simple working hypothesis



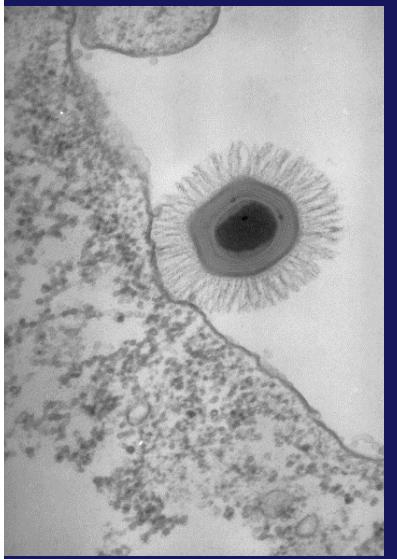
Very ancestral cellular origin

*Genome reduction  
(as for all intracellular parasites)*

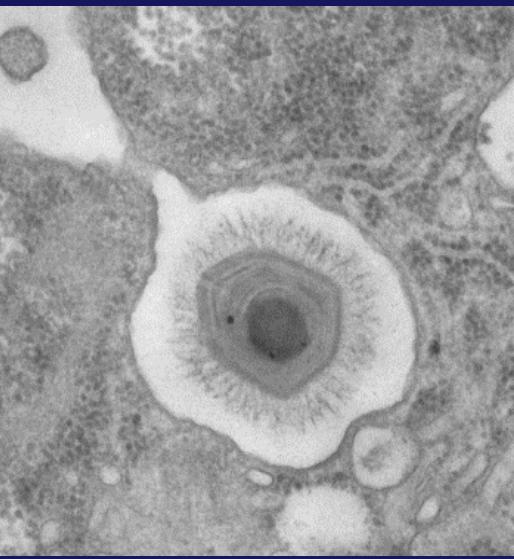


Today's giant viruses

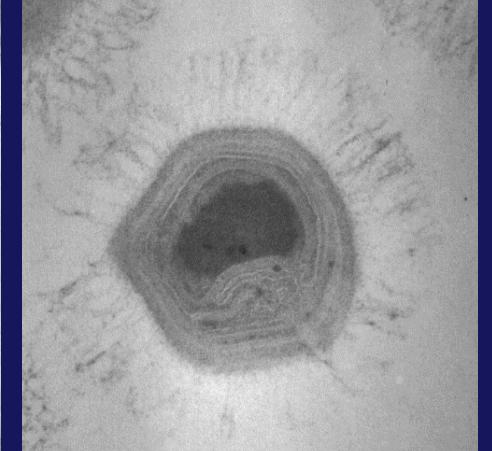
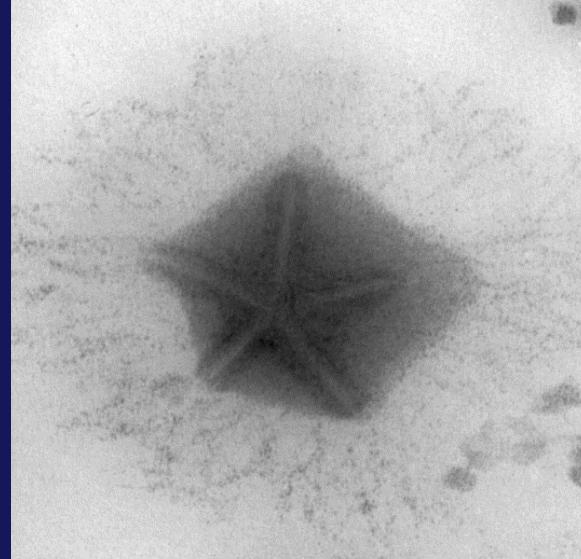
# Mimivirus: Infectious cycle



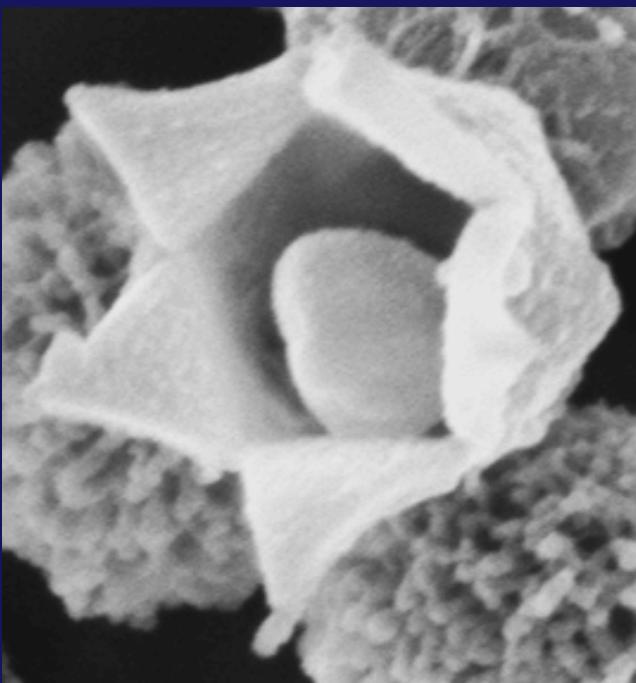
Surface binding



Phagocytosis



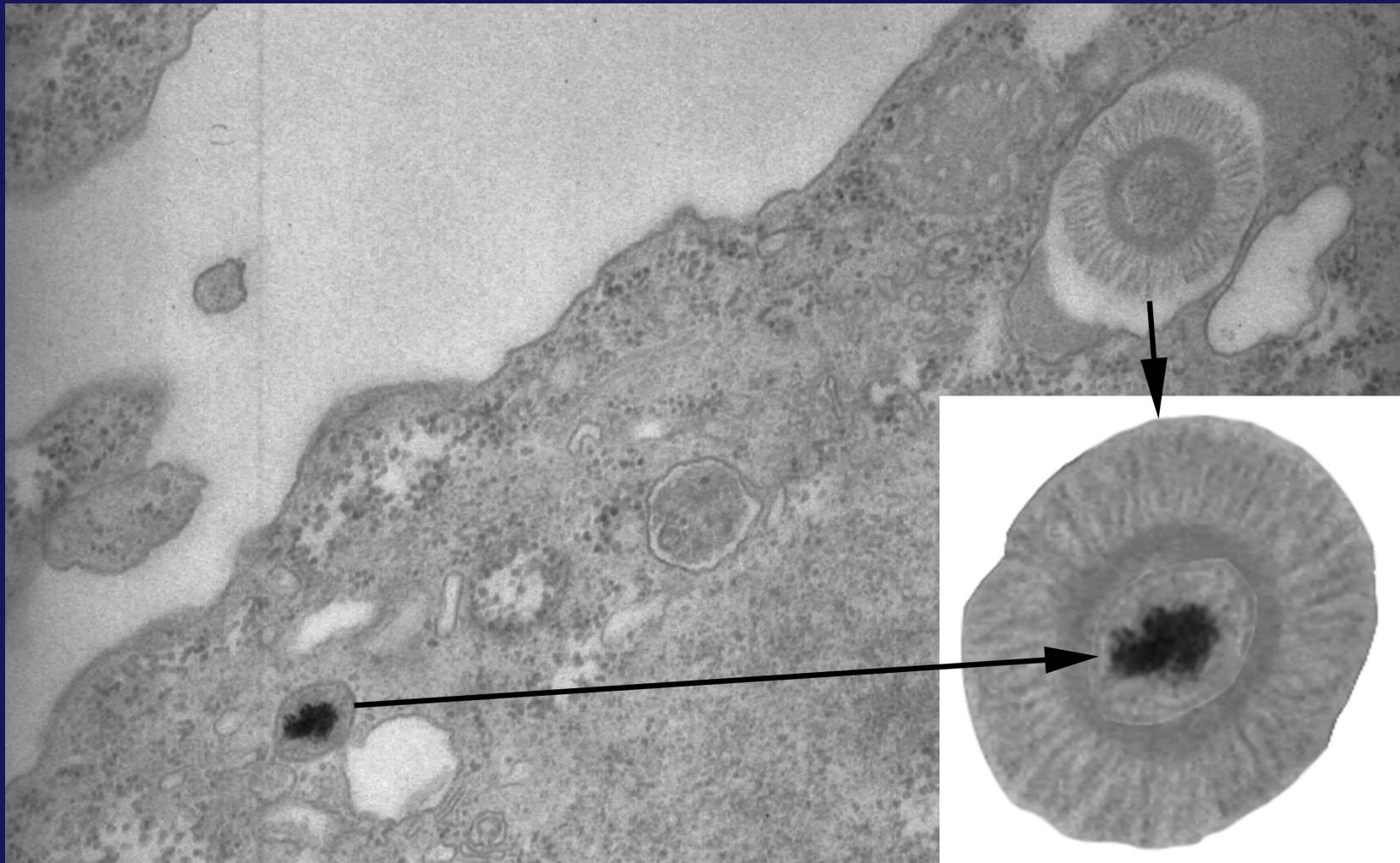
"Stargate"

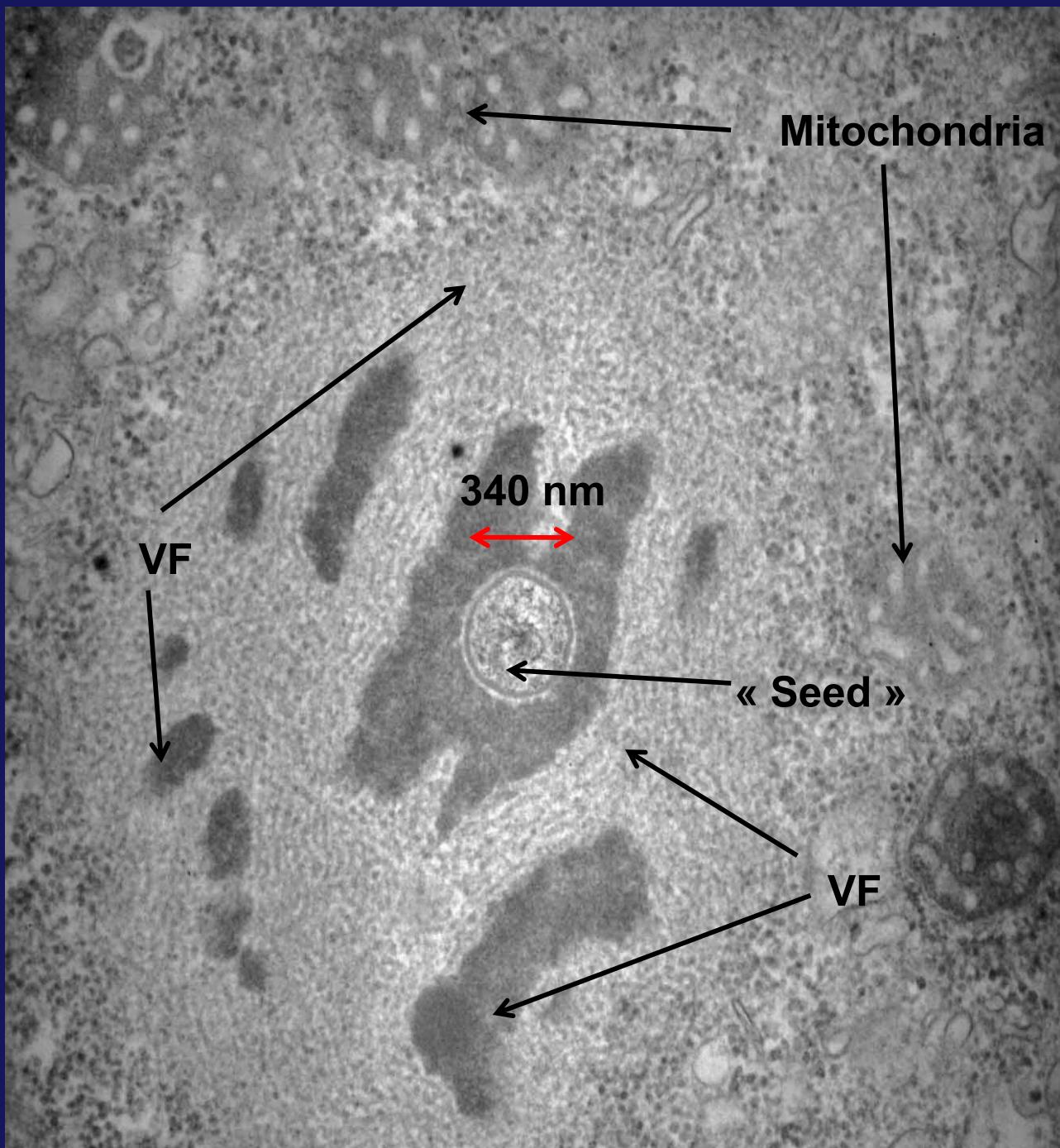


Membrane  
fusion

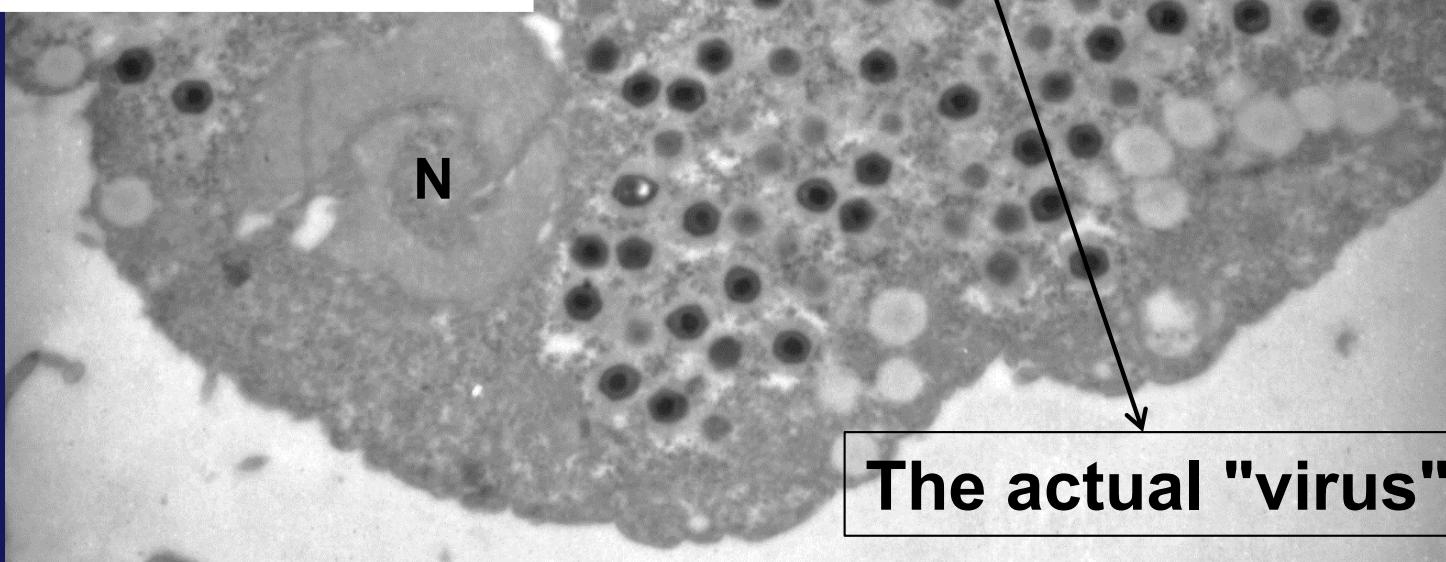
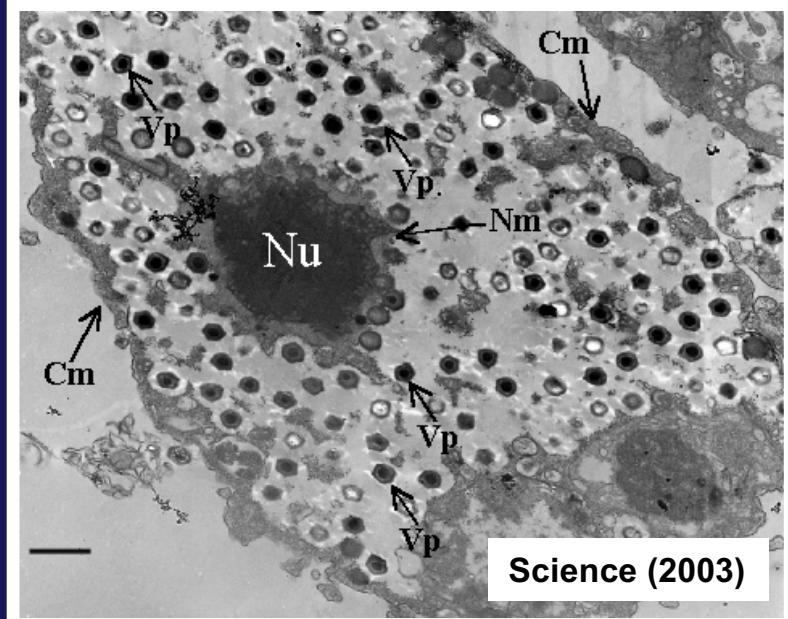


# The virion factory « seed »: the starting point of a developmental program





# A gigantic virion factory



The actual "virus"

# The environmental turn

— SCIENCE, 28 March 2003 —

## BREVIA



### A Giant Virus in Amoebae

Bernard La Scola,<sup>1</sup> Stéphane Audic,<sup>2</sup> Catherine Robert,<sup>1</sup> Liang Jungang,<sup>1</sup> Xavier de Lamballerie,<sup>3</sup> Michel Drancourt,<sup>1</sup> Richard Birtles,<sup>1</sup> Jean-Michel Claverie,<sup>2\*</sup> Didier Raoult<sup>1\*</sup>

similarity to *Baculoviridae* or *Asfarviridae* homologs. These results suggest that Mimivirus occupies an intermediary position between *Poxvirus*, *Iridoviridae*, and *Phycodnaviridae*, which Mimivirus appears to share the Vp54 capsid protein and a glucosamine synthetase similar to the *Paramecium bursaria Chlorella vi-* Mimivirus appears as a deep branch in the phylogenetic tree. The 1.2-Megabase Genome Sequence of Mimivirus

### The 1.2-Megabase Genome Sequence of Mimivirus

Didier Raoult,<sup>1\*</sup> Stéphane Audic,<sup>2</sup> Catherine Robert,<sup>1</sup> Chantal Abergel,<sup>2</sup> Patricia Renesto,<sup>1</sup> Hiroyuki Ogata,<sup>2</sup> Bernard La Scola,<sup>1</sup> Marie Suzan,<sup>1</sup> Jean-Michel Claverie<sup>2\*</sup>

We recently reported the discovery and preliminary characterization of Mimivirus, the largest known virus, with a 400-nanometer particle size comparable to mycoplasma. Mimivirus is a double-stranded DNA virus growing in amoebae. We now present its 1,181,404-base pair genome sequence, consisting of 1262 putative open reading frames, 10% of which exhibit a similarity to proteins of known functions. In addition to exceptional genome size, Mimivirus exhibits many features that distinguish it from other nucleocytoplasmic large DNA viruses. The most unexpected is the presence of numerous genes encoding central protein-translation components, including four amino-acyl transfer RNA synthetases, peptide release factor 1, translation elongation factor EF-TU, and translation initiation factor 1. The genome also exhibits six tRNAs. Other notable features include the presence of both type I and type II topoisomerases, components of all DNA repair pathways, many polysaccharide synthesis enzymes, and one intein-containing gene. The size and complexity of the Mimivirus genome challenge the established frontier between viruses and parasitic cellular organisms. This new sequence data might help shed a new light on the origin of DNA viruses and their role in the early evolution of eukaryotes.

16 August 2005  
**Virology Journal**



Research

**Mimivirus relatives in the Sargasso sea**  
Elodie Ghedin<sup>1,2</sup> and Jean-Michel Claverie<sup>\*3</sup>



J. Craig Venter



The Sorcerer II Global Ocean Sampling

SCIENCE, 19 November 2004

Recently isolated from amoebae growing in viral genes tentatively classified Mimivirus

# ASSEMBLE



Looking for giant viruses

ECIM  
Las Cruces

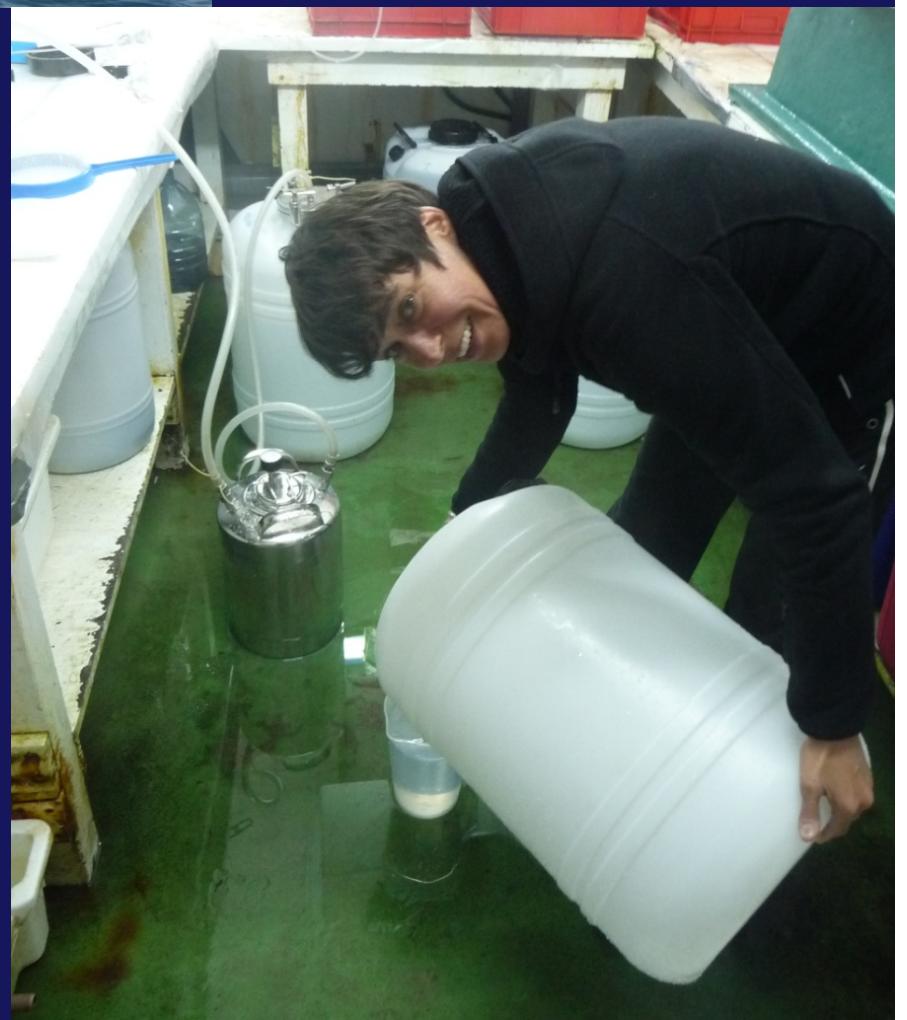
Boating

Diving

Collecting

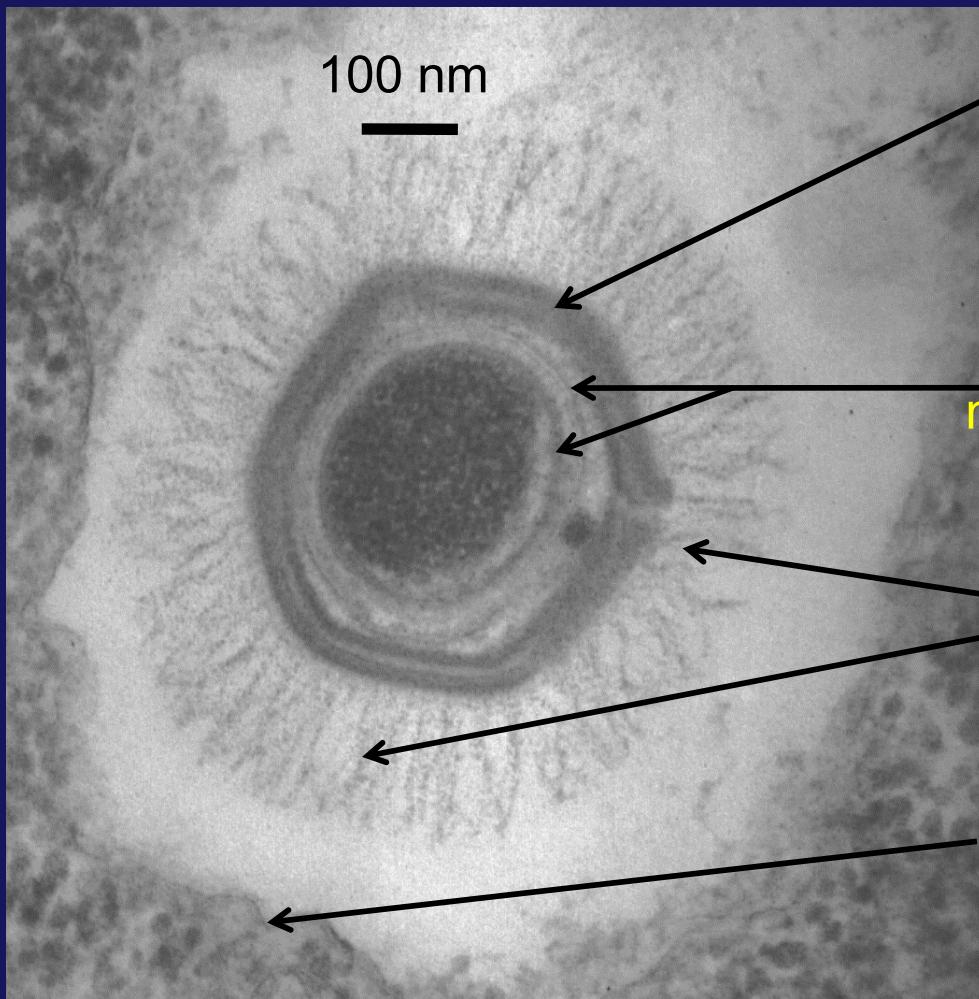
Filtering

Culturing



# Megavirus: 1.28 Mb!

*Mimivirus*



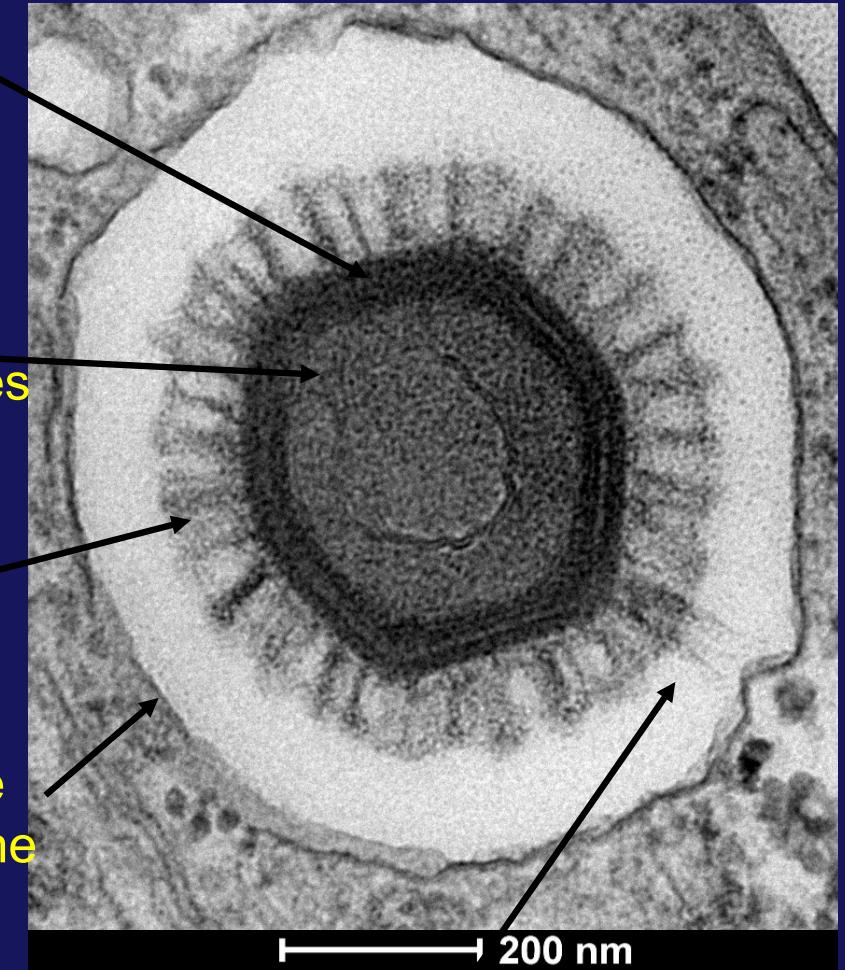
Capsid  
(protein)

Lipid  
membranes

Fiber  
layer

Vacuole  
membrane

*Megavirus chilensis*



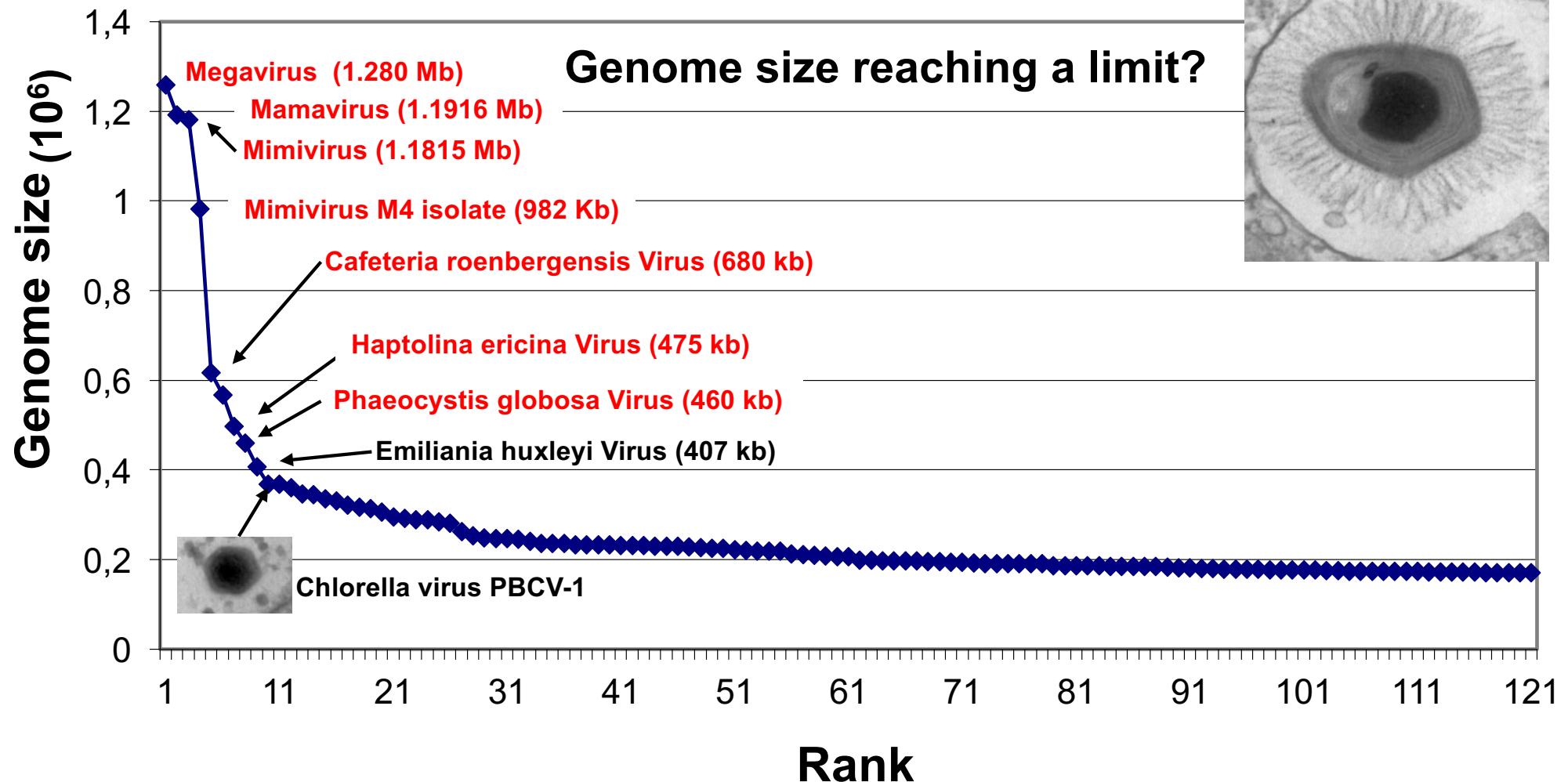
cowlicks !

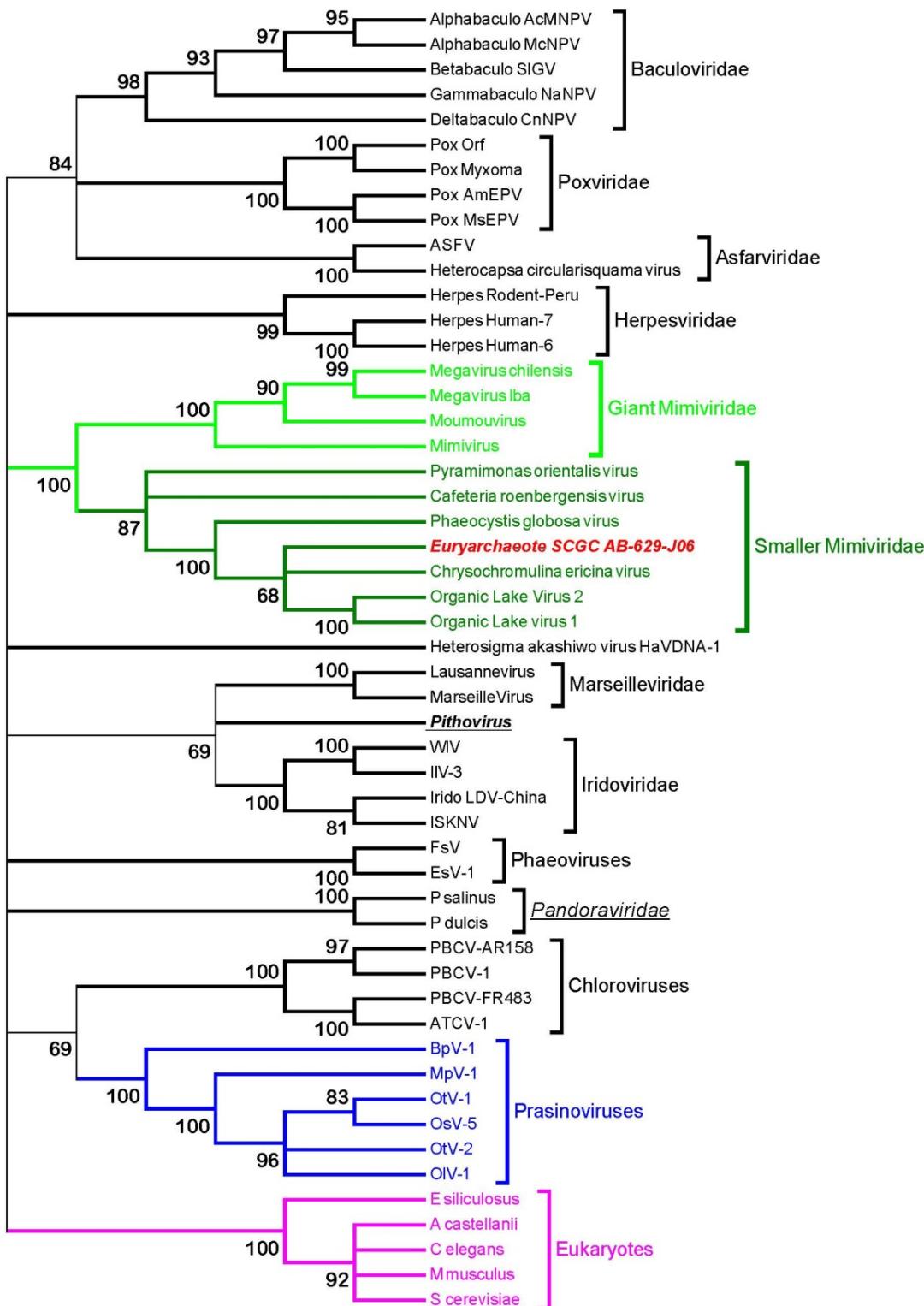
# Megavirus: more remnants of a translation apparatus

- Mimivirus was the first representative of a large family
- These giant viruses could derive from even bigger ancestors

virus	Capsid (nm)	Genes	hair	Stargate	Translation
<b>Megavirus chilensis</b>	440	1120	75 nm	Yes	<b>7 (Arg, Tyr, Met, Cys, Ile, Trp, Asn)</b>

# Giants among viruses (2004-2011)





## Mimiviridae/Megaviridae

A new, clearly distinct family  
of large dsDNA viruses  
infecting unicellular protists

DNA polB phylogeny

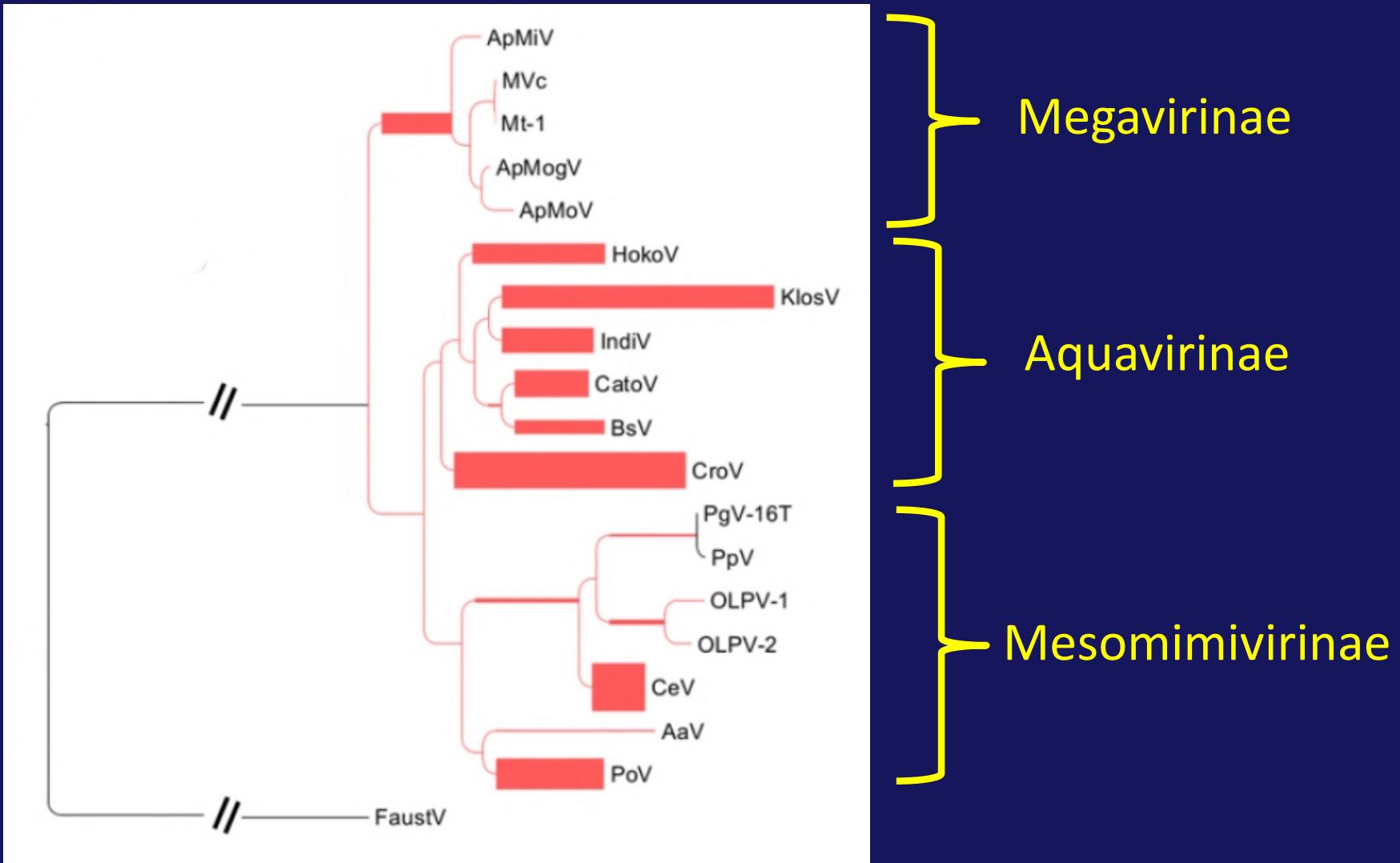
# Bloom-associated Mimiviridae



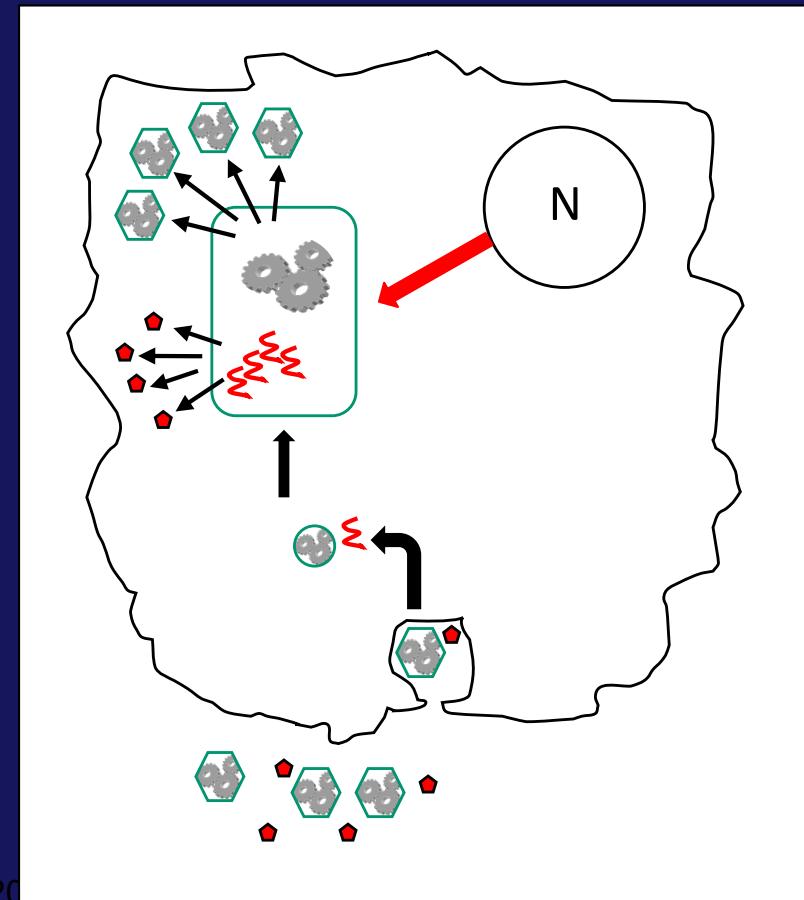
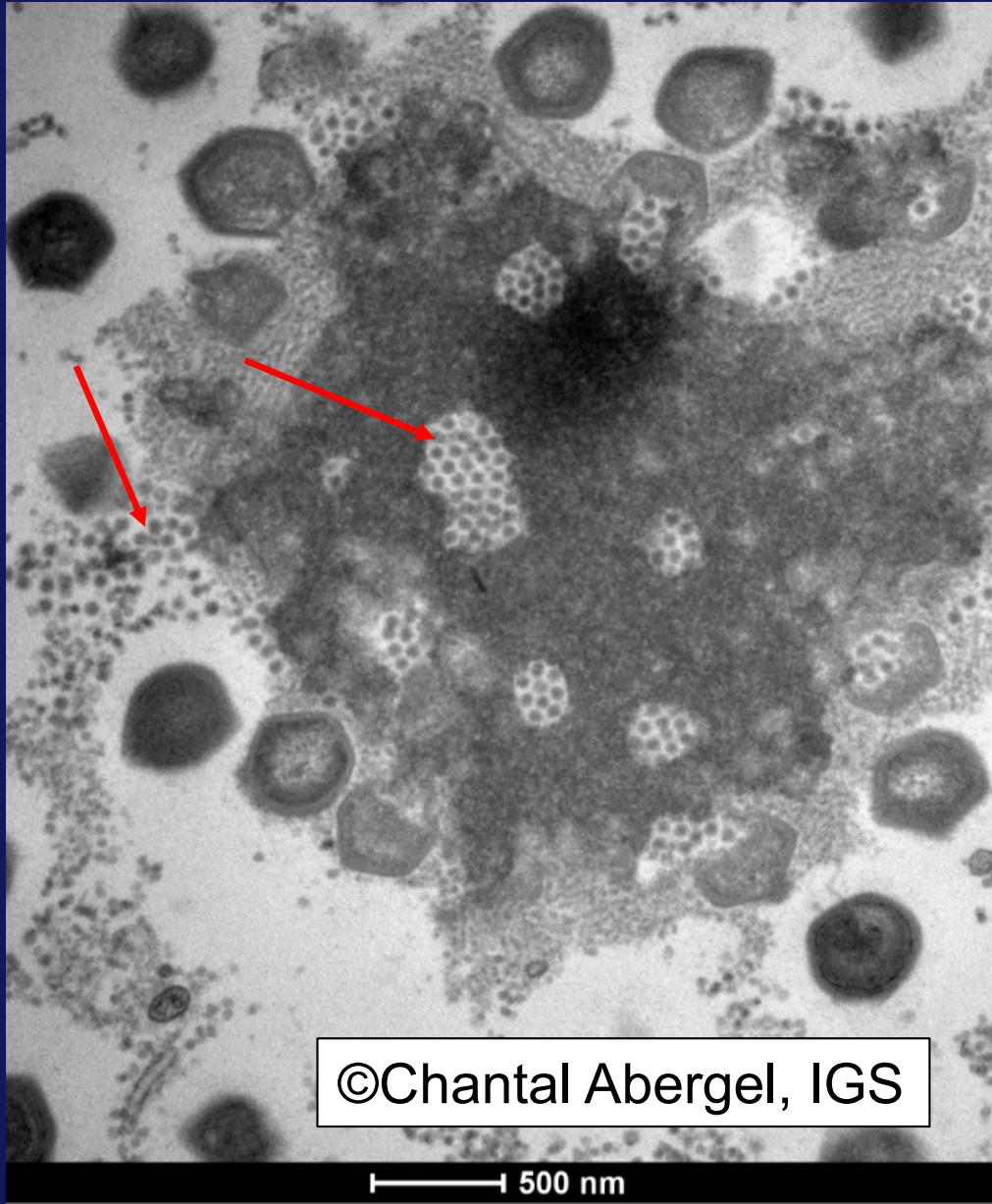
*Haptophyta,  
Prymnesiophyceae*

PgV genome: 460 kb [PNAS 2013]  
Chrysochromulina ericina V: 474 kb [J. Virol. 2017]

# Mimiviridae today



# The Mimiviridae have their own viruses: the « virophage »



50-70 nm  
17-22 kb

# Mimiviridae today

Family	Prototype	Virion type	Dimension	Genome, size, GC%	Specific feature	Host
Mimiviridae		Icosahedral		Linear DNA	MutS, Virophage	Acanthamoeba
Megavirinae					Long fibers	
	<i>Mimivirus</i>	Icosahedral	500+250 nm	1.18 Mb, 28%	4 AaRS	Acanthamoeba
	<i>Moumouvirus</i>	Icosahedral	420+200 nm	1.02 Mb, 25%	5 AaRS	Acanthamoeba
	<i>Megavirus chilensis</i>	Icosahedral	520+150 nm	1.26 Mb, 25%	7 AaRS	Acanthamoeba
	<i>Platanovirus KSL-5</i>	Icosahedral	290+140 nm	?	?	Saccamoeba
Mesomimivirinae					Algae-infecting	
	<i>P. globosa virus (PgV)</i>	Icosahedral	150 nm	460 kb, 32%		Phaeocystis
	<i>H. ericina virus (CeV)</i>	Icosahedral	160 nm	474 kb, 25%		Haptolina
	<i>Aav</i>	Icosahedral	140 nm	371 kb, 28.7%		Aureococcus
Aquavirinae						
	<i>CroV</i>	Icosahedral	300 nm	693 kb, 23%	1 AaRS	Cafeteria
	<i>BsV</i>	Icosahedral	300 nm	1.386 Mb, 25%		<i>Bodo saltans</i>
	<i>Klosneuvirus</i>	No isolate	?	1.57 Mb, 28.6%	19 AaRS	unknown
	<i>Catovirus</i>	No isolate	?	1.53 Mb, 26.4%	15 AaRS	unknown
	<i>Hokovirus</i>	No isolate	?	1.33 Mb, 21.4%	13 AaRS	unknown
	<i>Indivirus</i>	No isolate	?	860 kb, 26.6%	3 AaRS	unknown

# New Life Form: Chile (2009), Australia (2011)



Tunquen, Chili

Shallow sediments



Melbourne, Australia

La Trobe University, pond mud

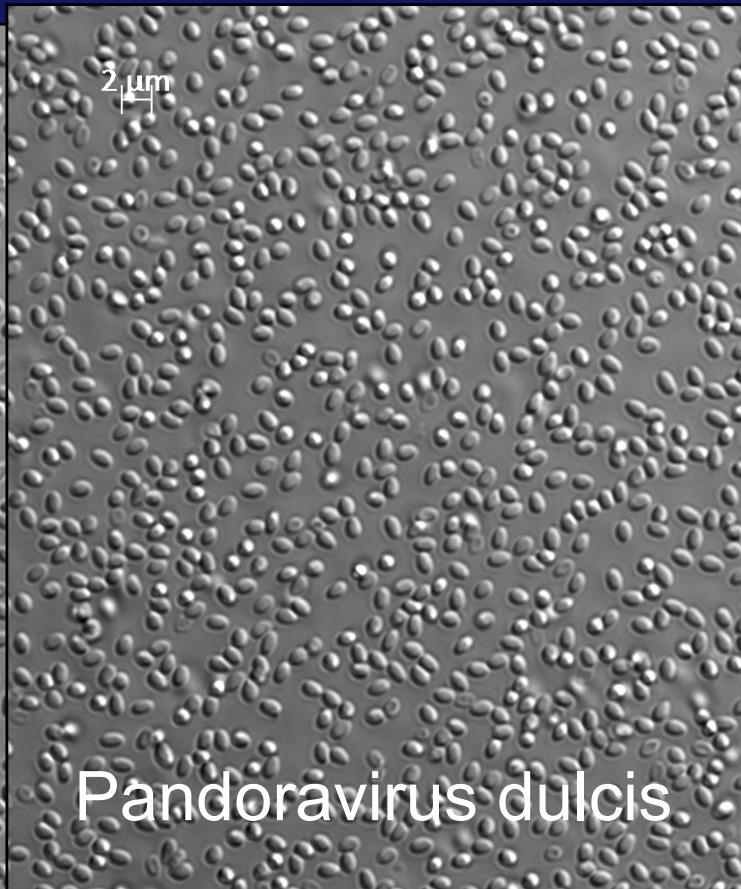


# New Life Form: Chile (2009), Australia (2011)

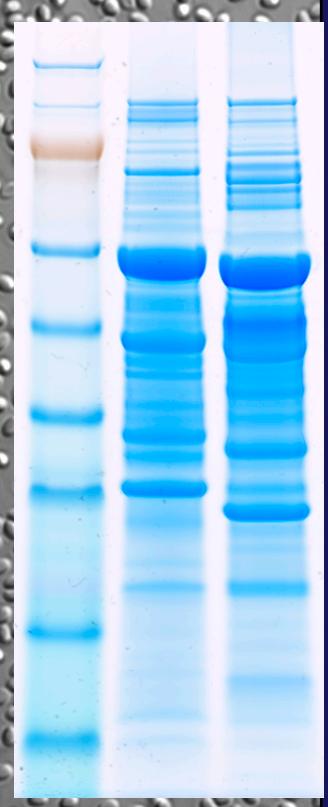
Light microscopy (x63, DIC/Nomarski)



Pandoravirus salinus

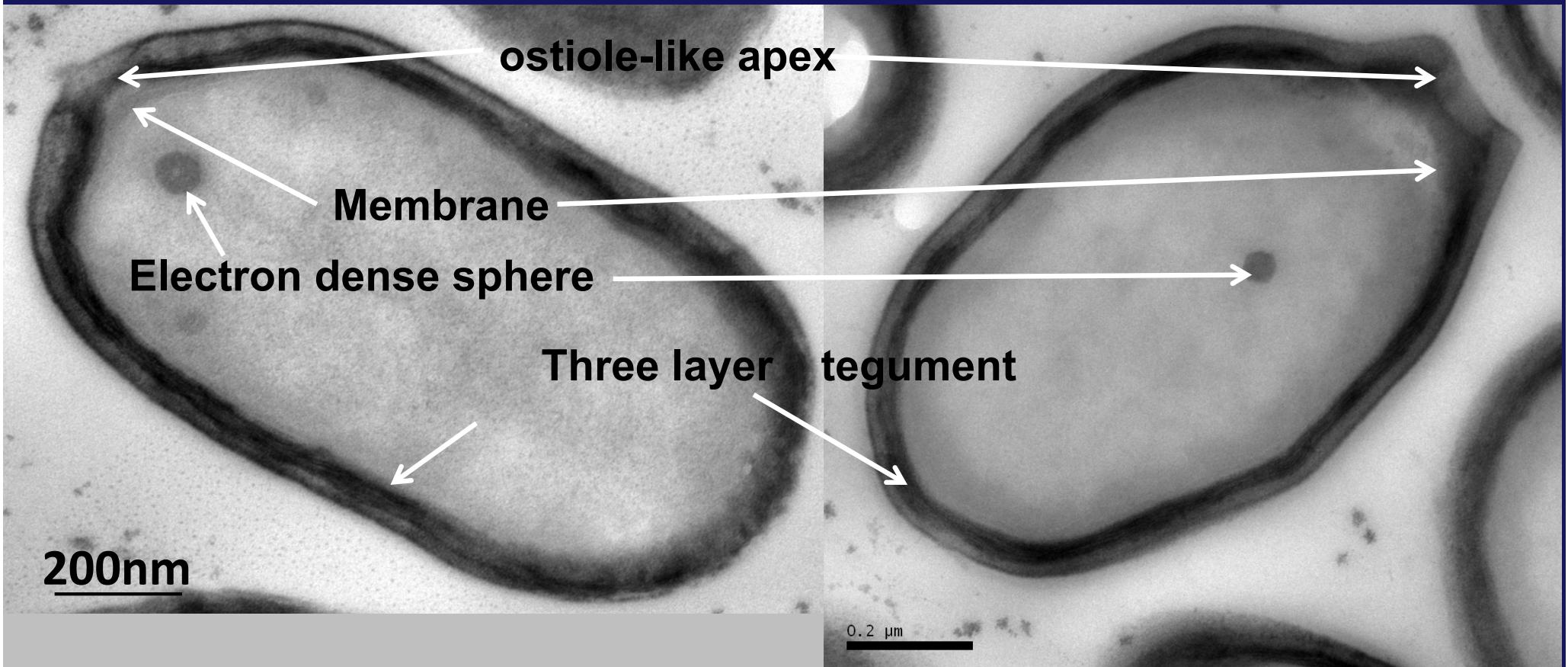


Pandoravirus dulcis



Mimivirus

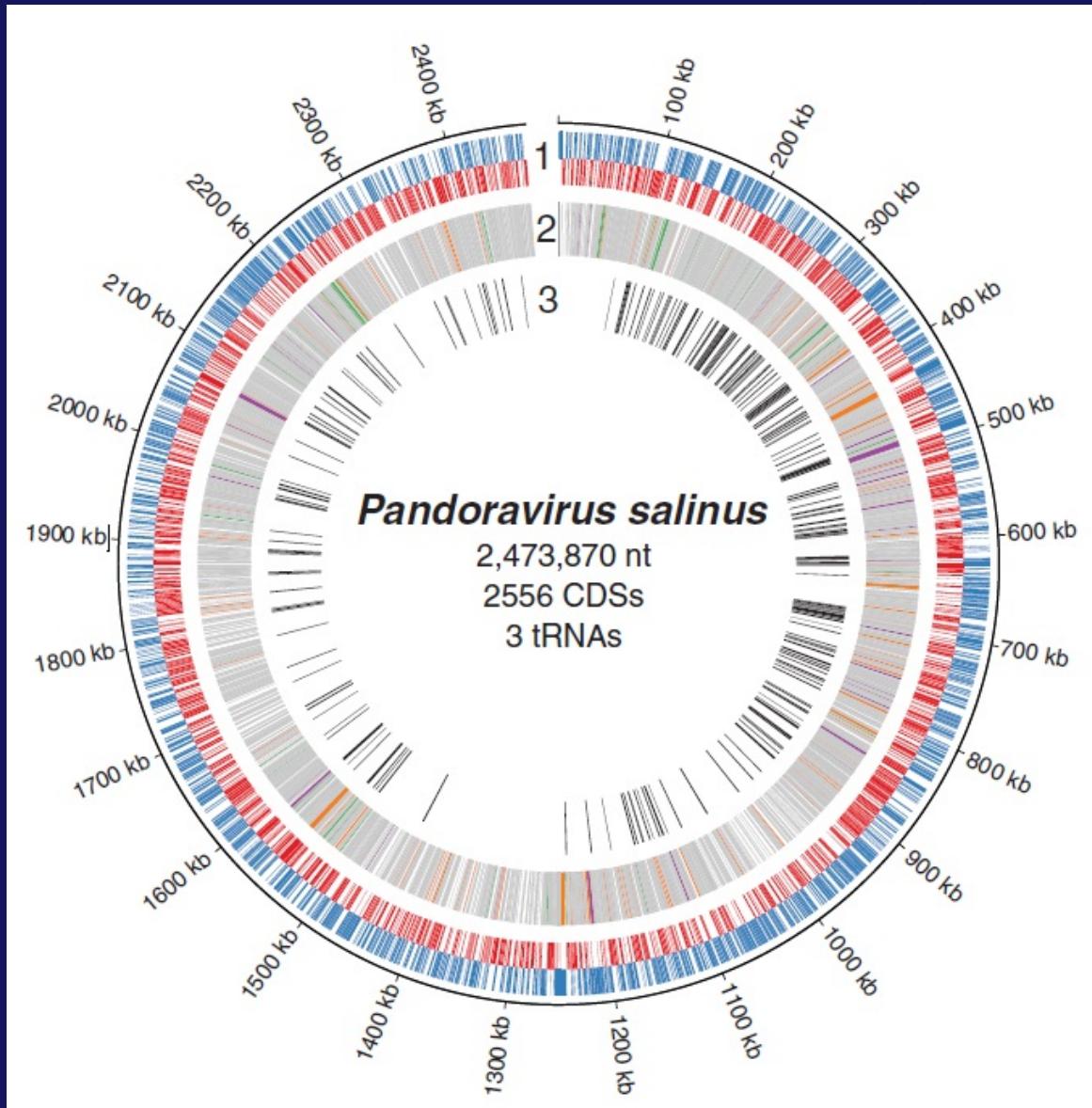
# New Life Form: Chile (2009), Australia (2011)



Pandoravirus salinus

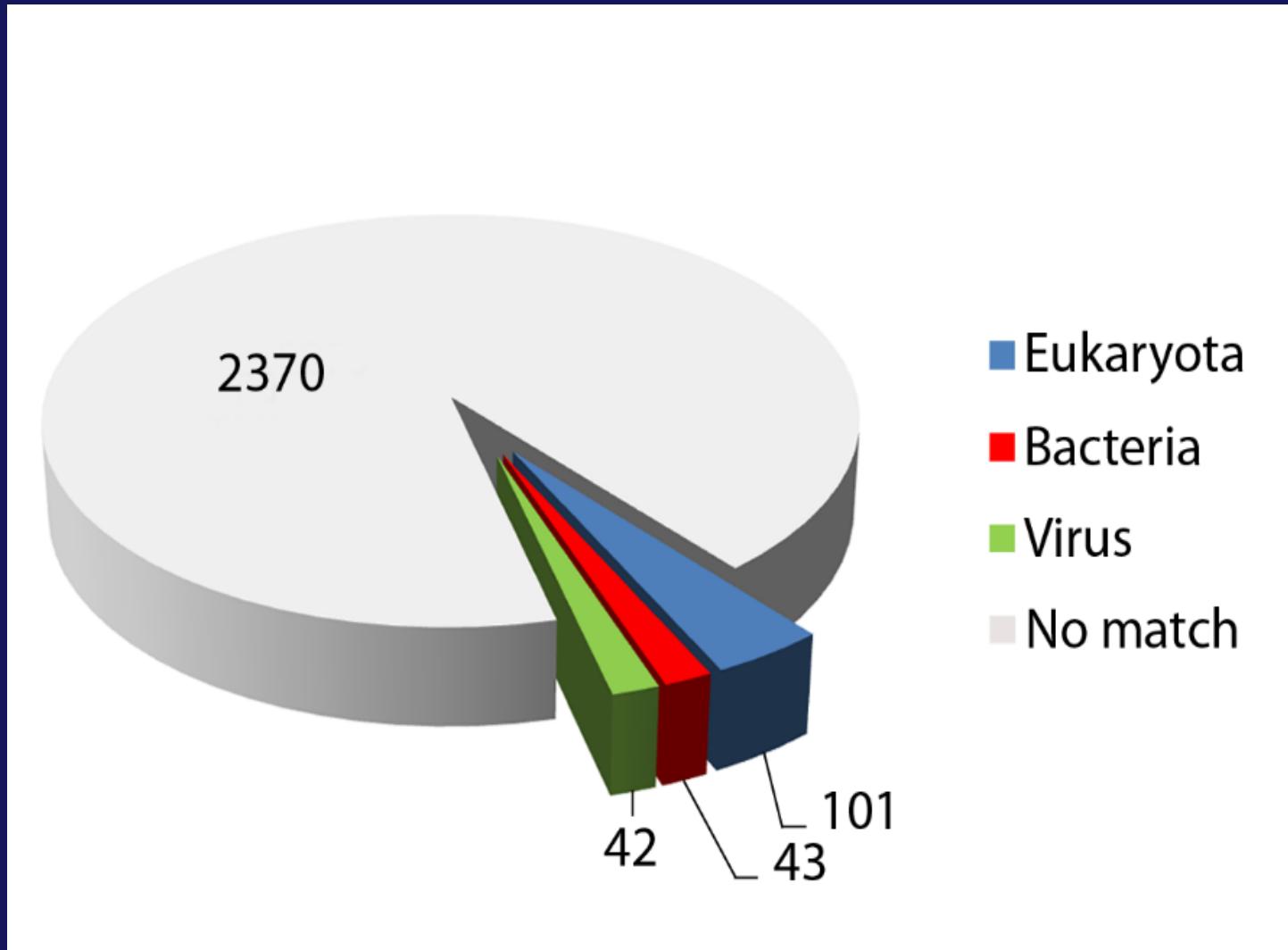
Pandoravirus dulcis

# *Pandoravirus salinus* (Chile): 2.8 Mb, 62%GC

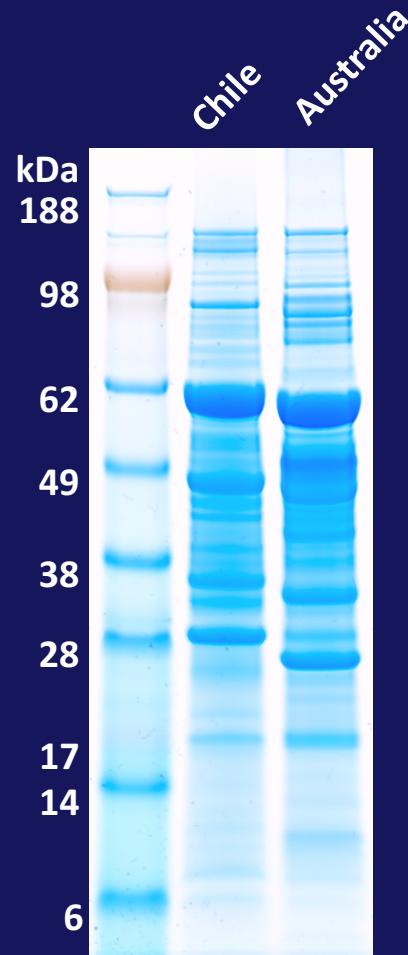


- No ribosomal protein
- No dividing apparatus (FtsZ)
- No ATP production enzyme
- This must be a virus
- But: No trace of Major Capsid Protein

# 94% of the genes encode ORFans !



# *P. salinus* proteomic validation



~ 200 predicted proteins found in the particle

-> **These entities use the standard genetic code**

All of the *P. salinus* particle proteins have Homologs in the *P. dulcis* genome

83% of them have no database homolog !

Particle proteomics

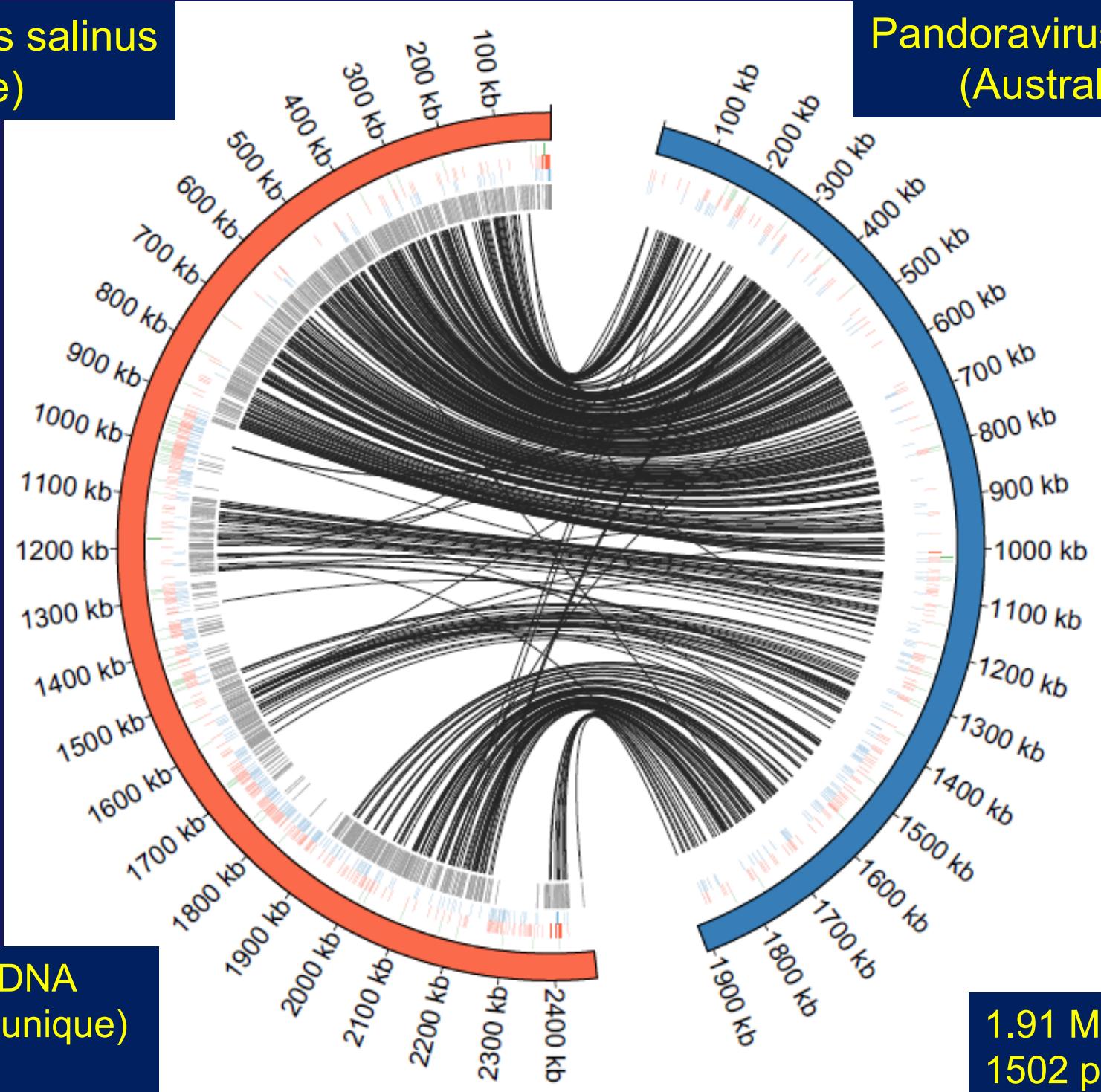
MS-MS spectrometry (Y. Couté, C. Bruley, J. Garin, Grenoble)

Pandoravirus salinus  
(Chile)

2.8 Mb linear DNA  
2,473,870 nt (unique)  
2556 proteins

Pandoravirus dulcis  
(Australia)

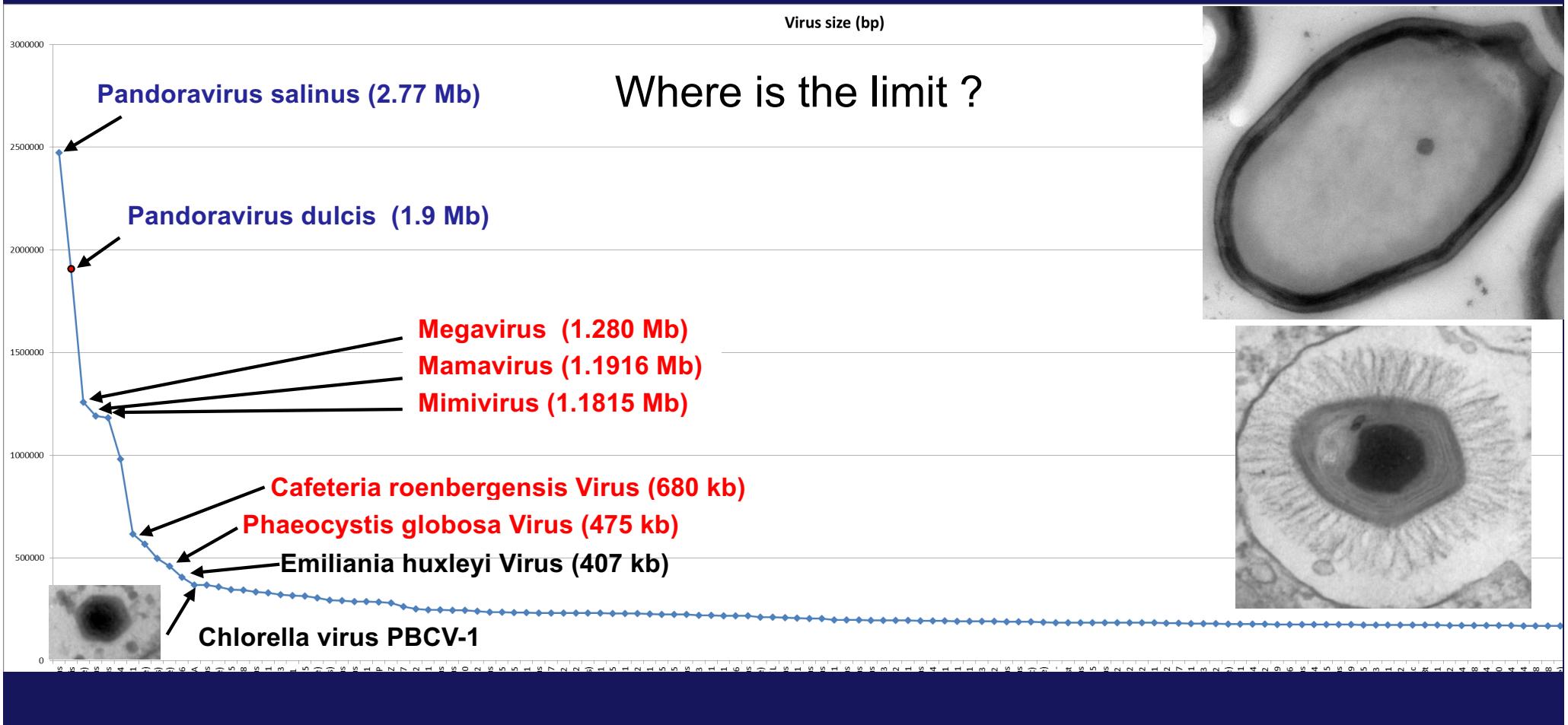
1.91 Mb  
1502 proteins



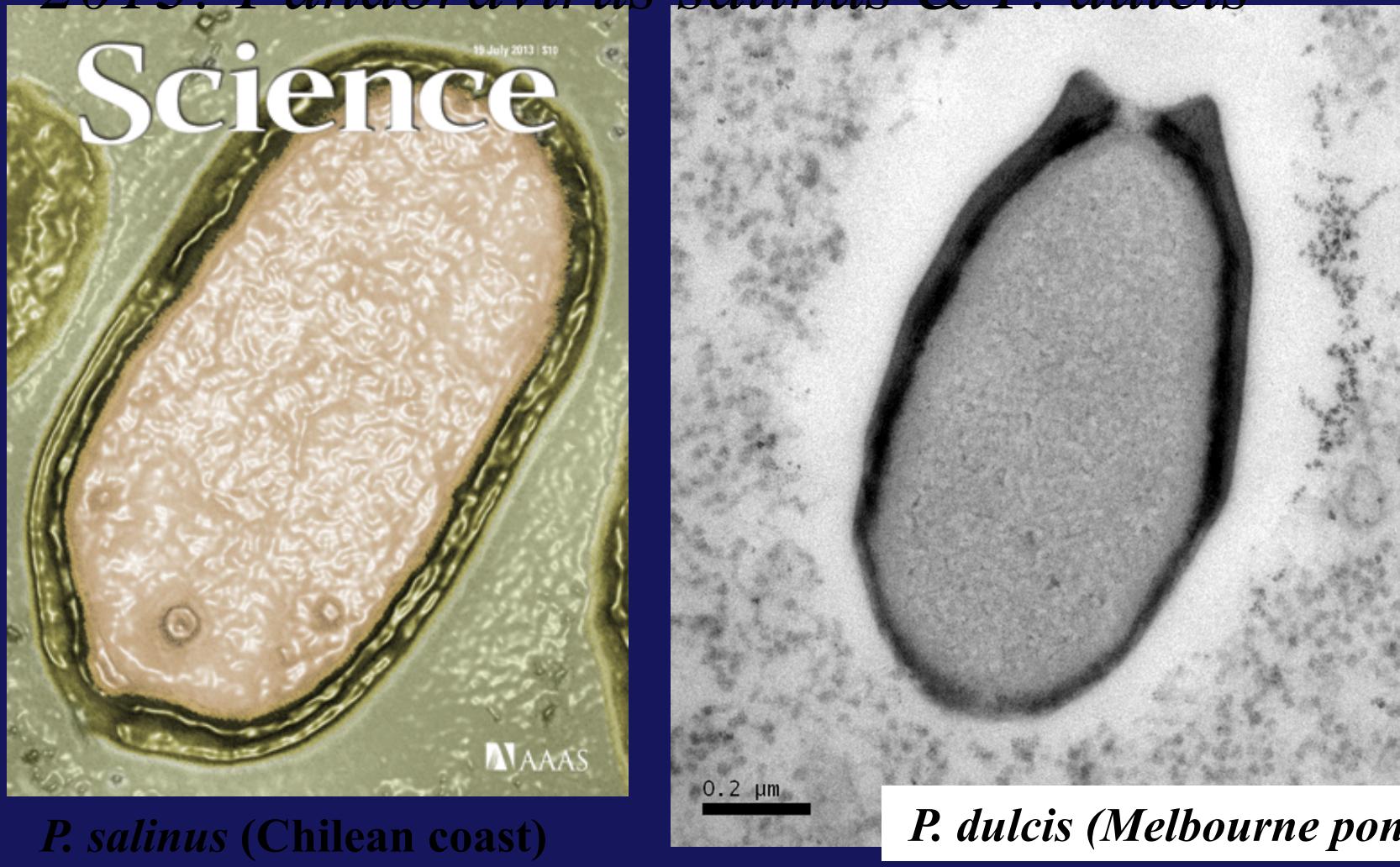
# Giants among giant viruses (2013)

## More genes than parasitic microsporidia (eukaryotes)

<i>Encephalitozoon cuniculi</i>	2.5 Mb	1,996 proteins
<i>Encephalitozoon intestinalis</i>	2.22 Mb	1,833 proteins
<i>Encephalitozoon romaleae</i>	2.19 Mb	1,831 proteins



2013: *Pandoravirus salinus* & *P. dulcis*

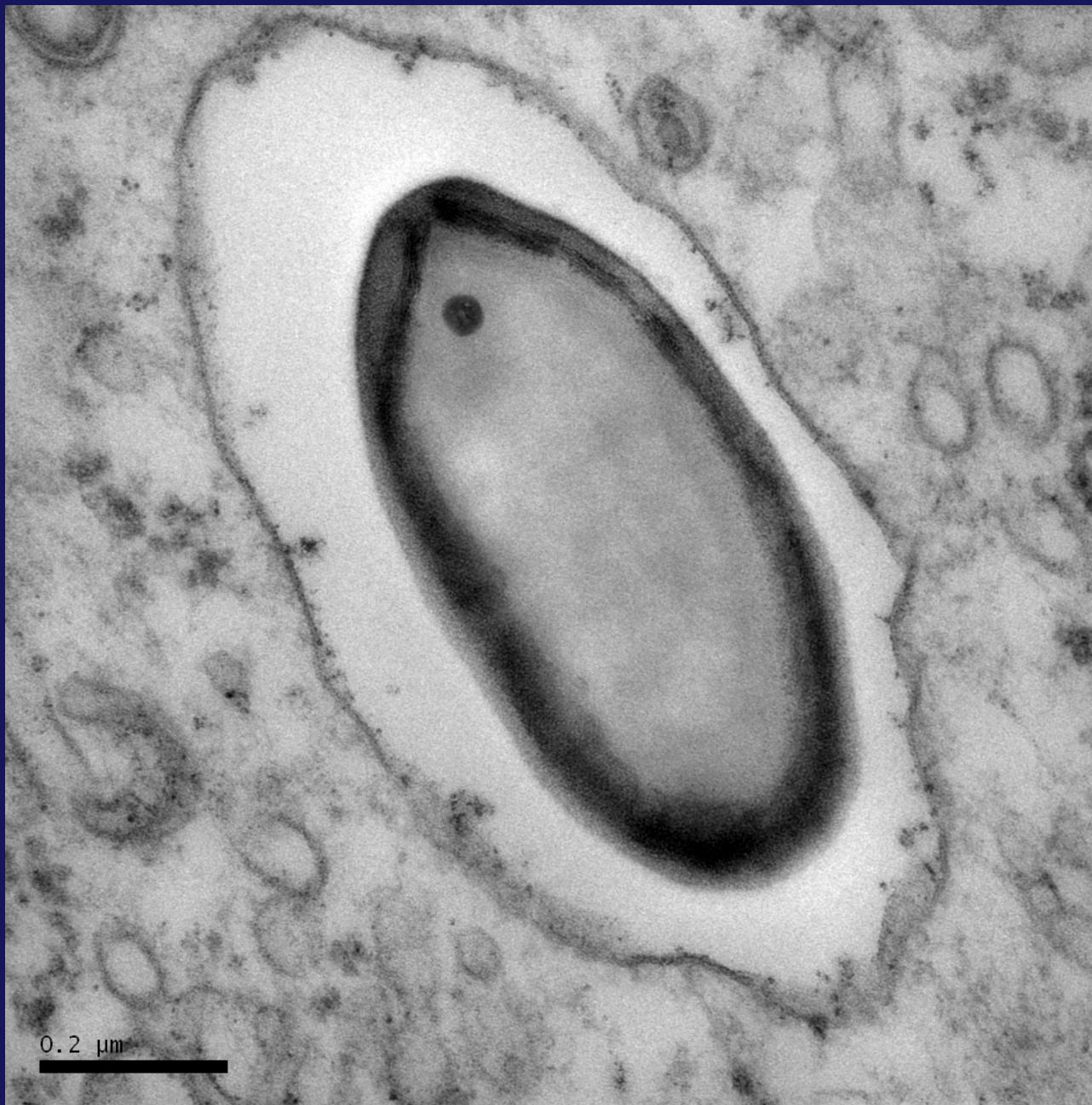


Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb  
reaching that of parasitic

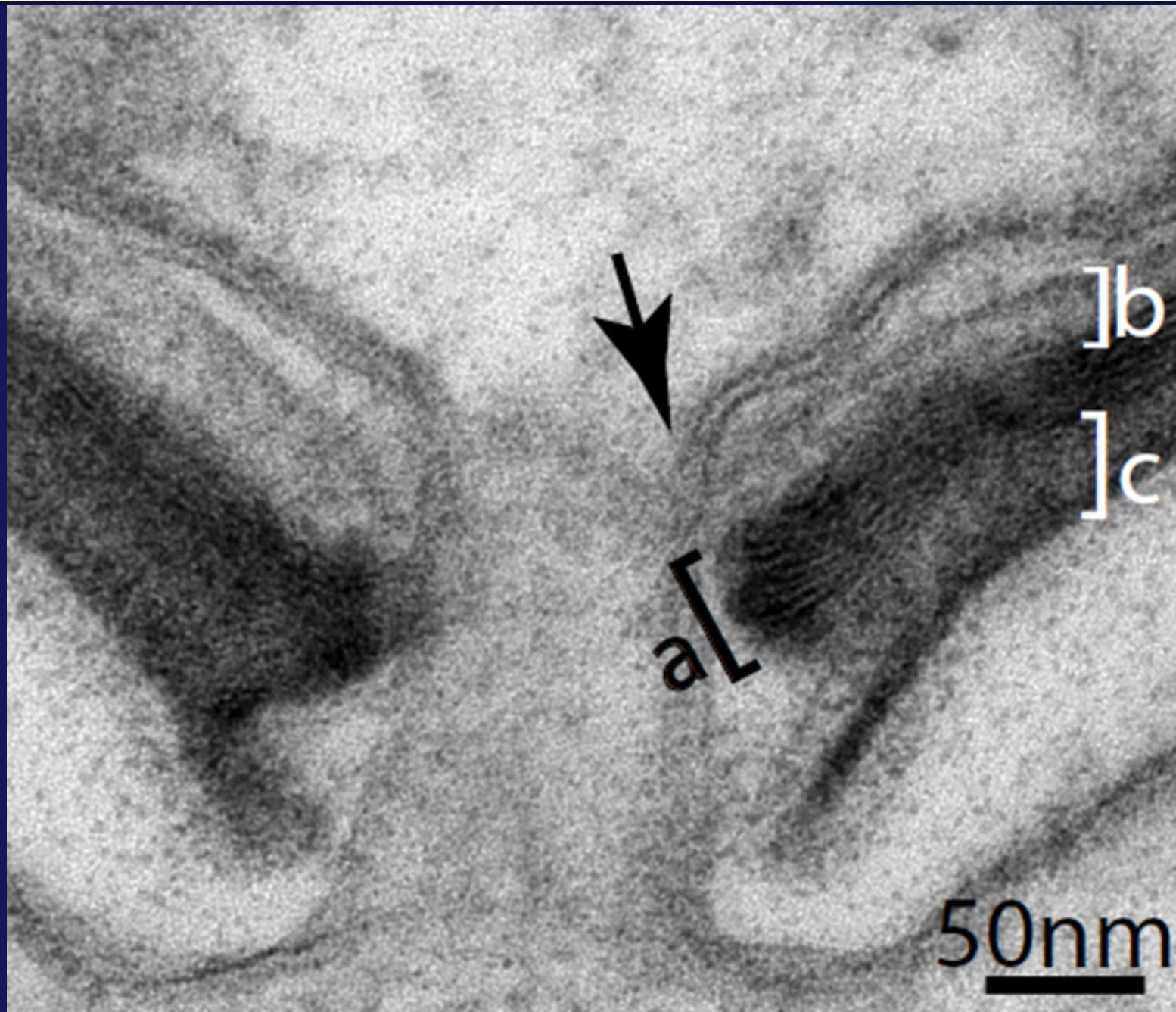
# Pandoravirus: Infectious cycle

# Step 1: phagocytosis

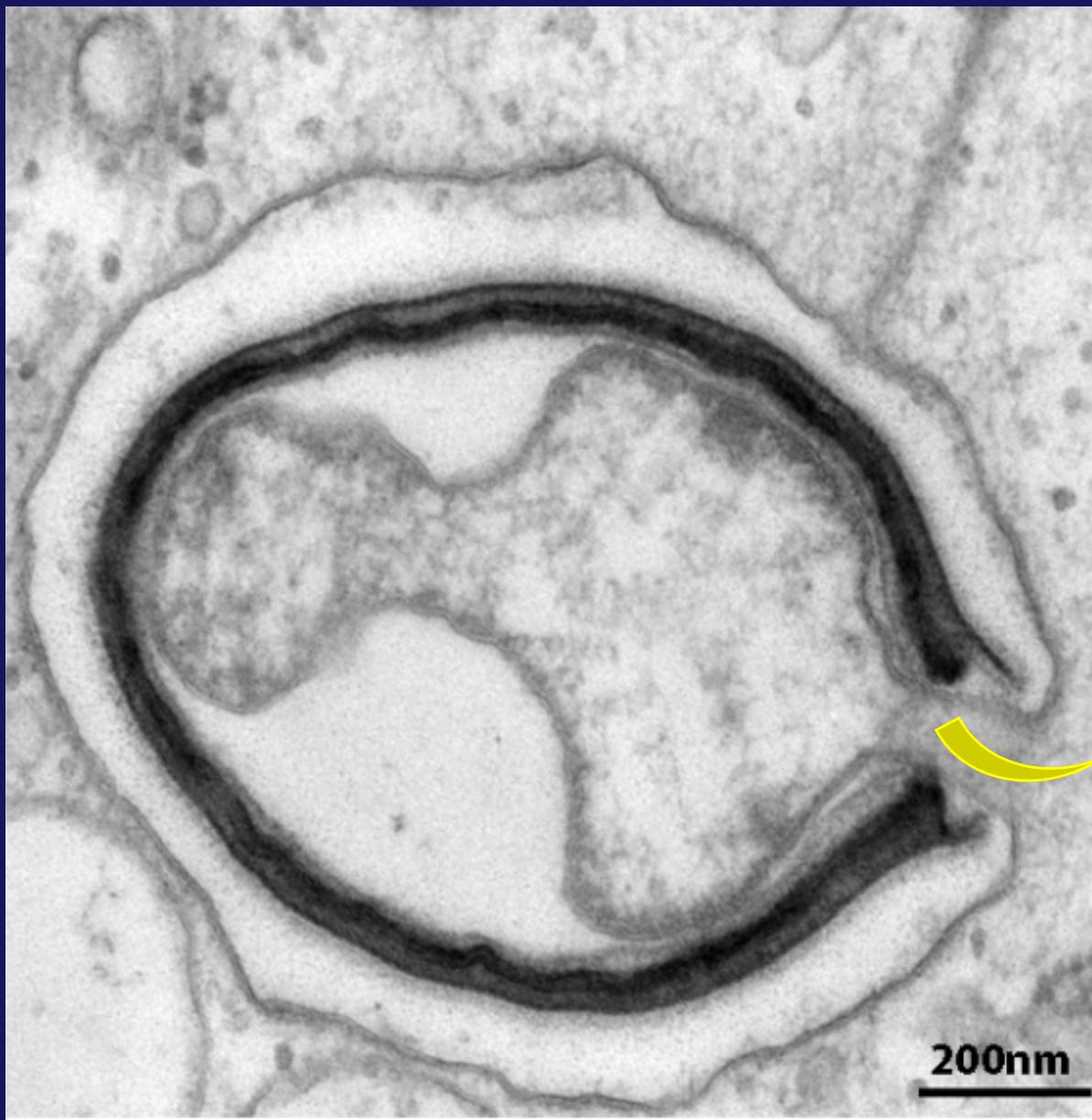
---



## Step 2 : membrane fusion



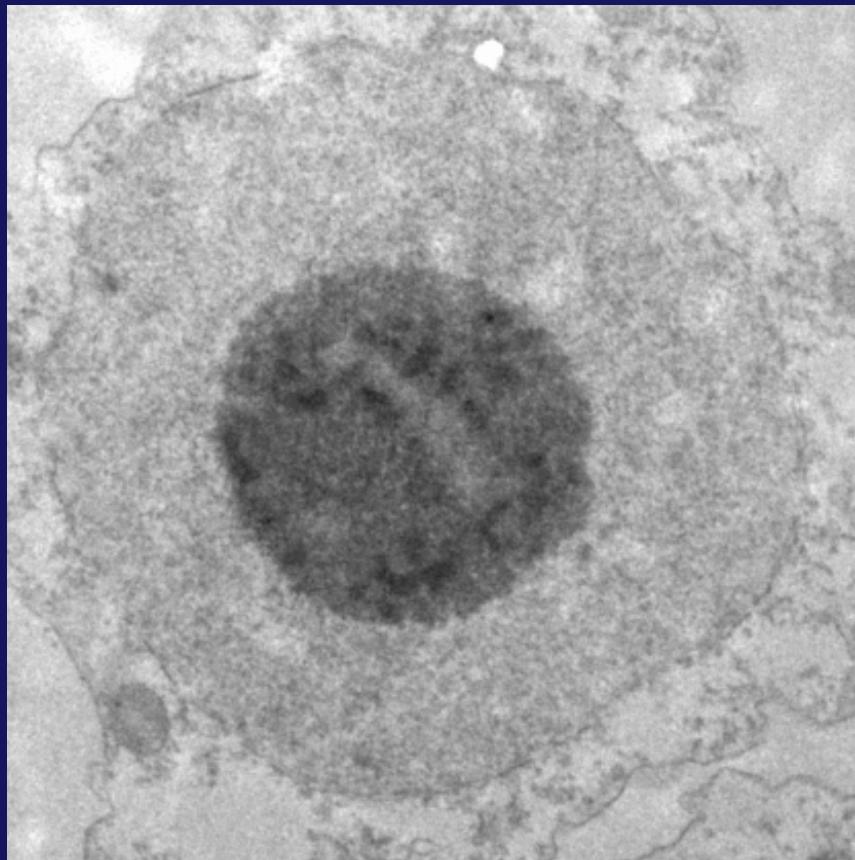
## Step 3 : « downloading »



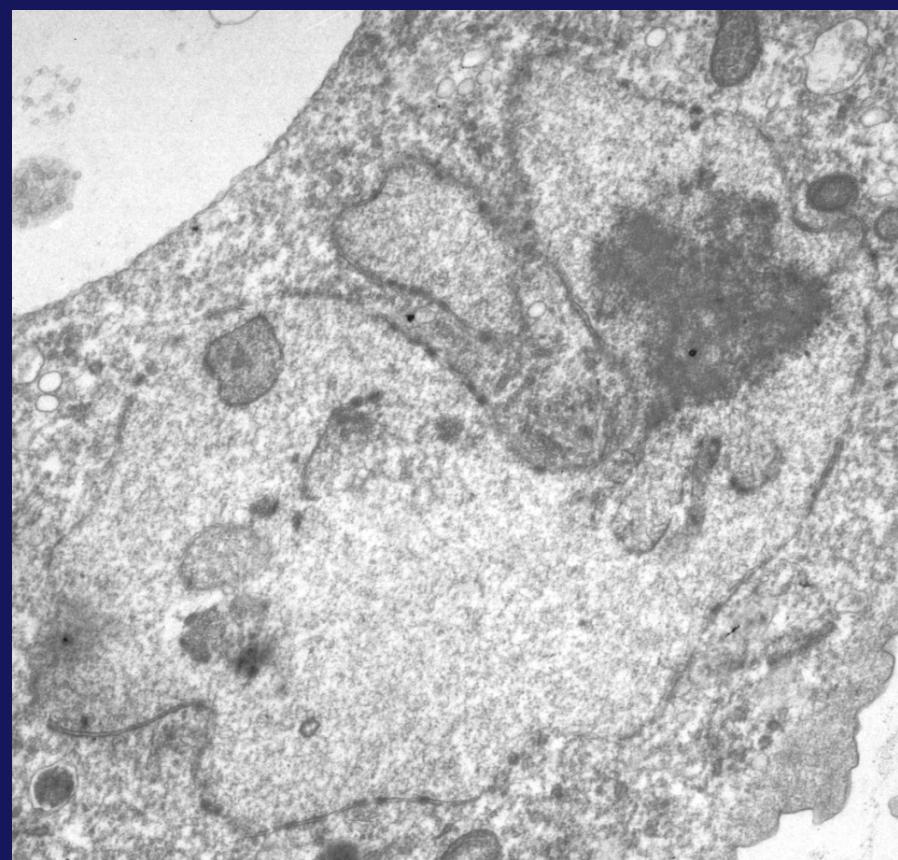
## Step 4: Early nuclear (disruptive) phase

---

Healthy Acanthamoeba cell

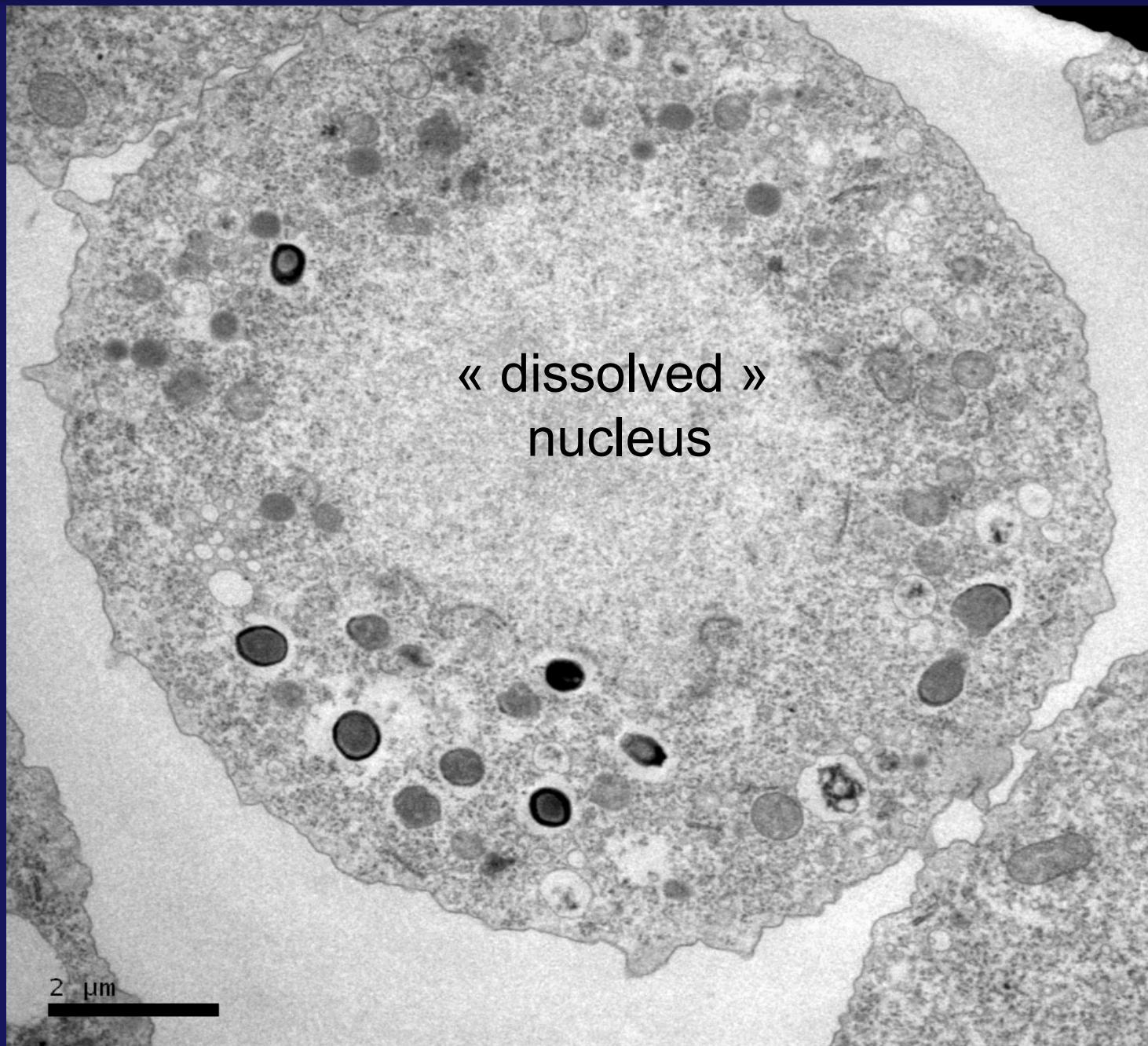


Infected cell (3h p.i.)



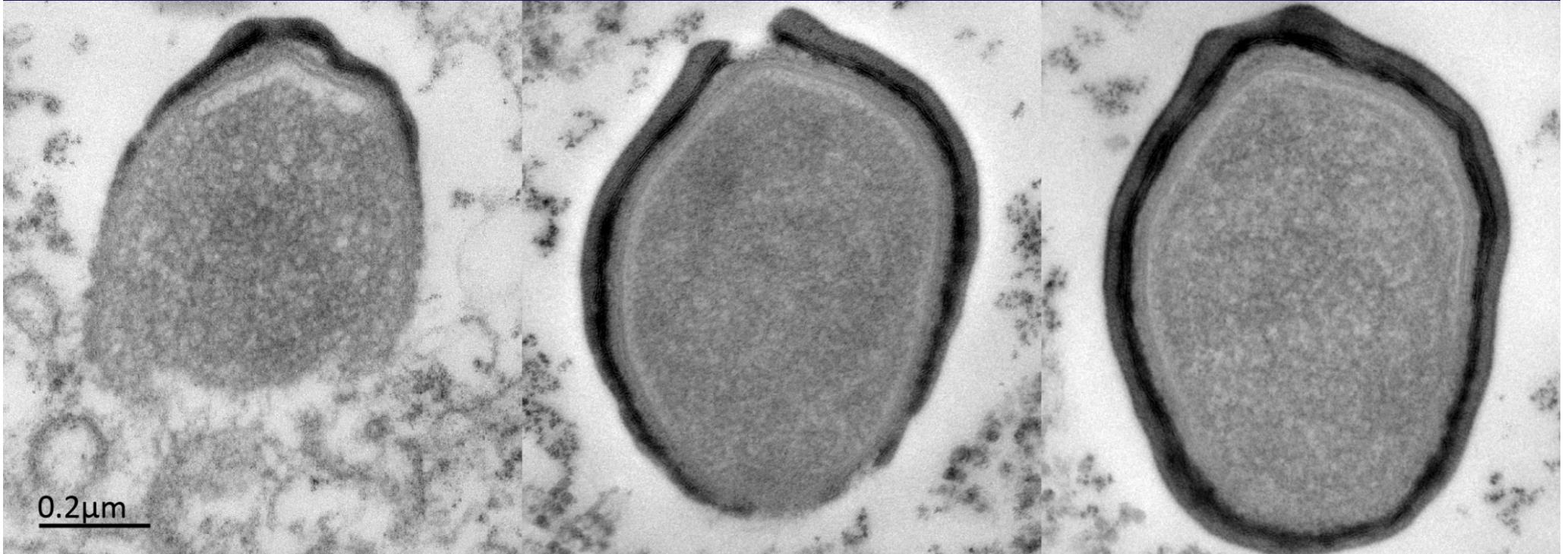
## Step 5: Particle formation

---



# Particle formation: “knitting”

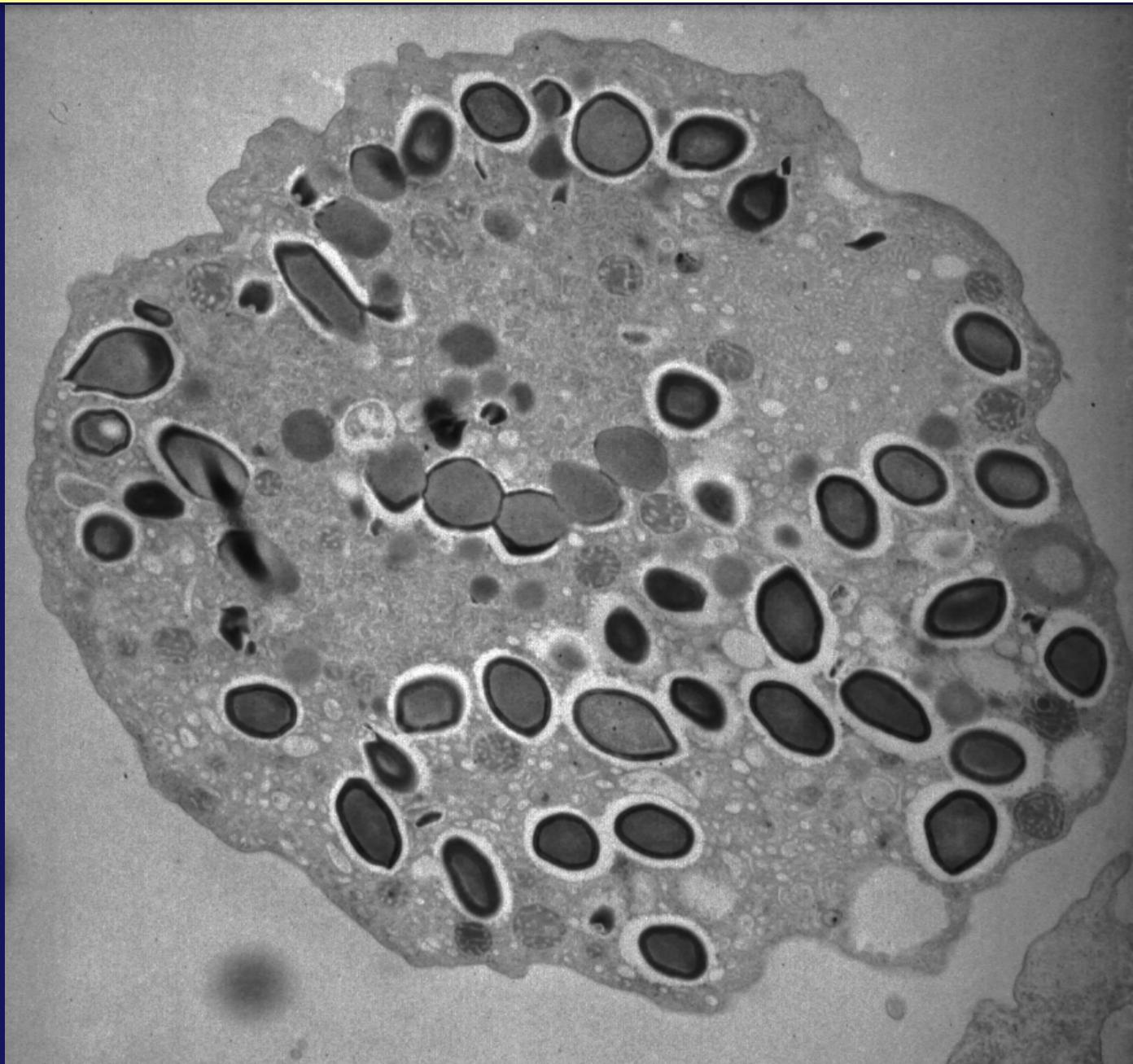
---



No division

# End of cycle

---



# Despite their huge genome Pandoraviruses are nucleus-dependent

---

EM: Cell nucleus is quickly modified after the infection

Genome:

At least 10% of the genes (with database homologs) exhibit spliceosomal introns (U2-dependent, GT-AG)

These introns are short (<200 nt), more than one third remain in phase with the flanking exons

Proteome:

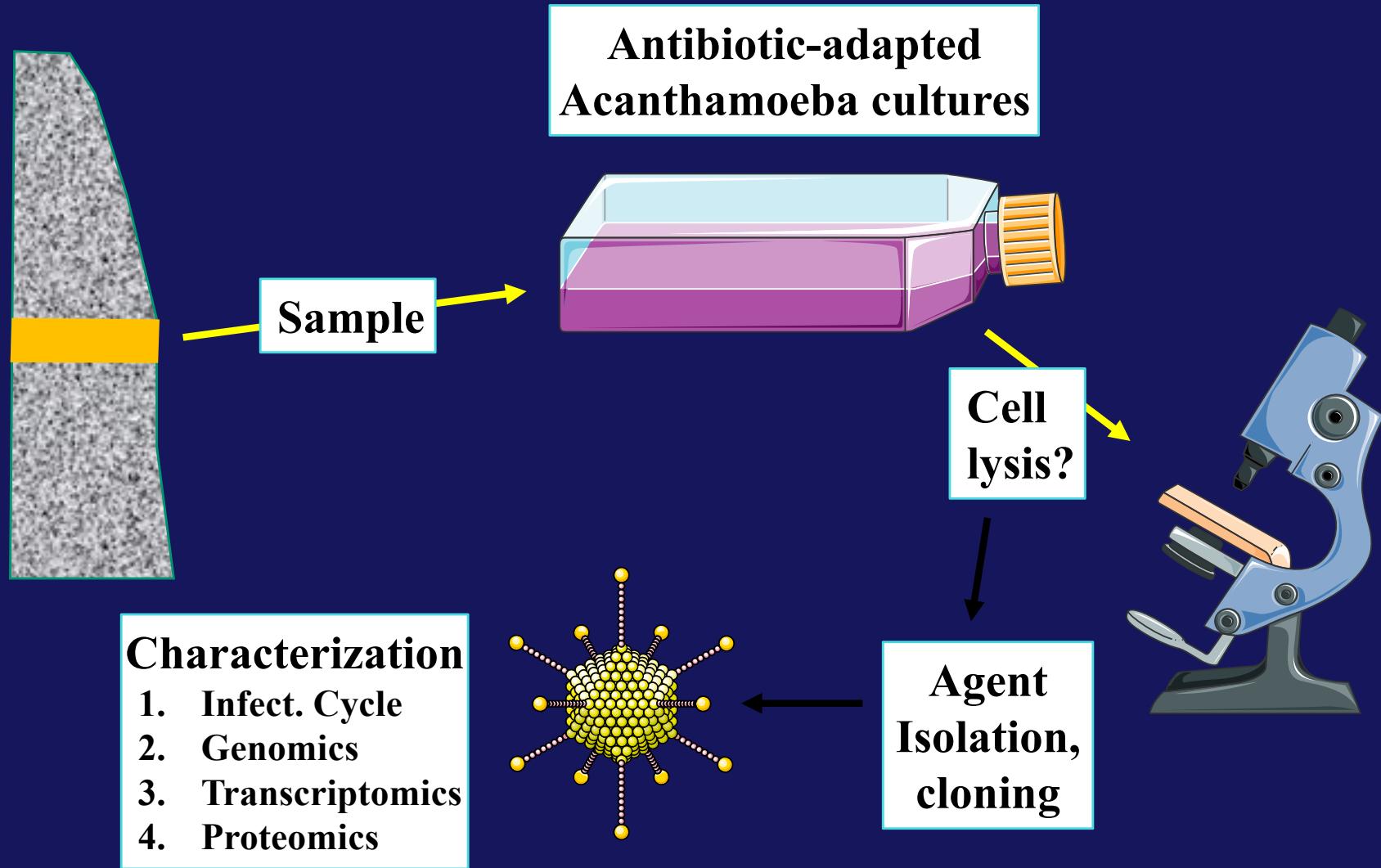
The particles do not incorporate any transcription machinery

# The Emerging Family Pandoraviridae

Jean-Michel Claverie

Matthieu Legendre, Chantal Abergel, *et al.*

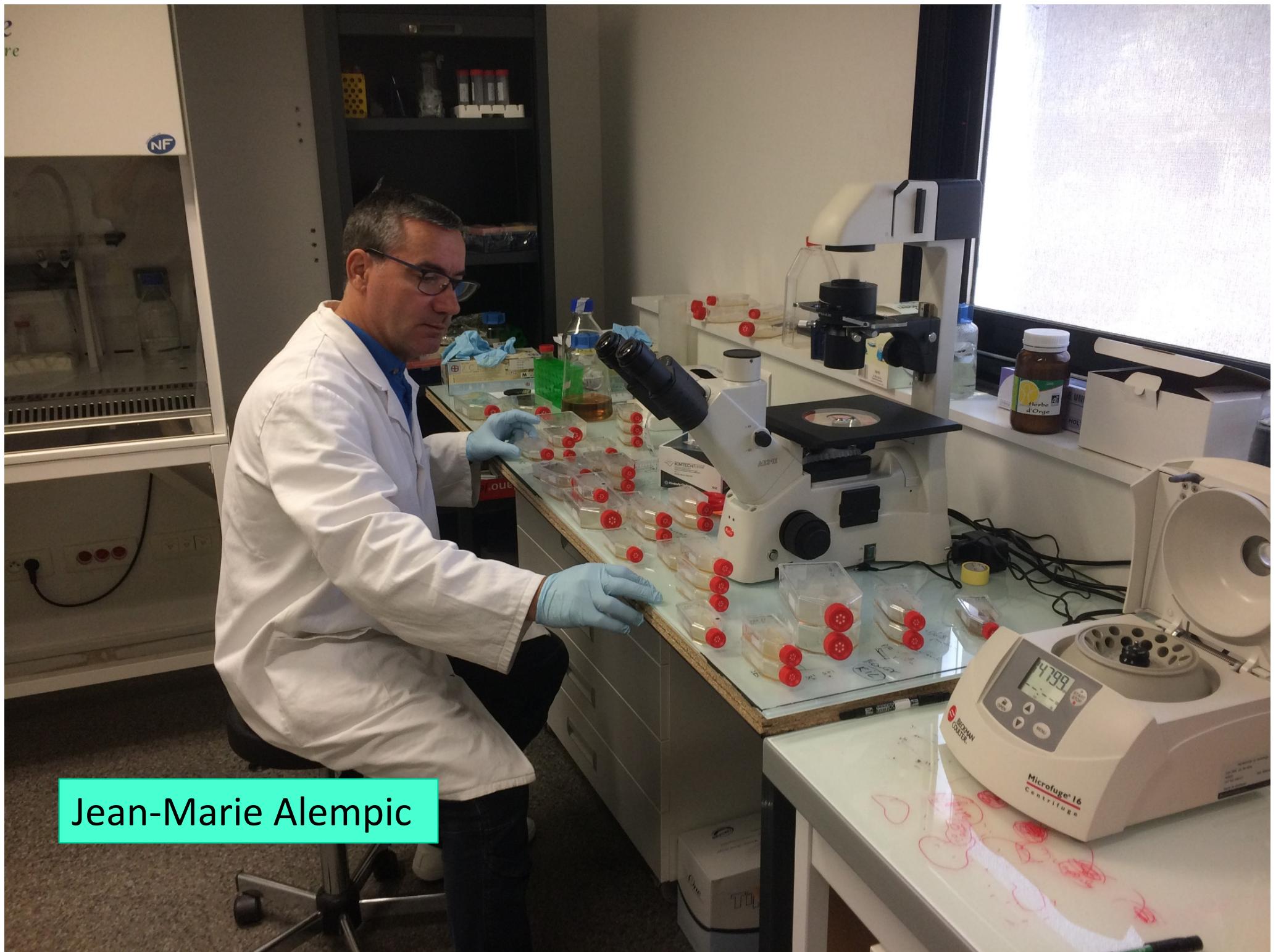
# Protocol: looking for Amoeba-killing viruses

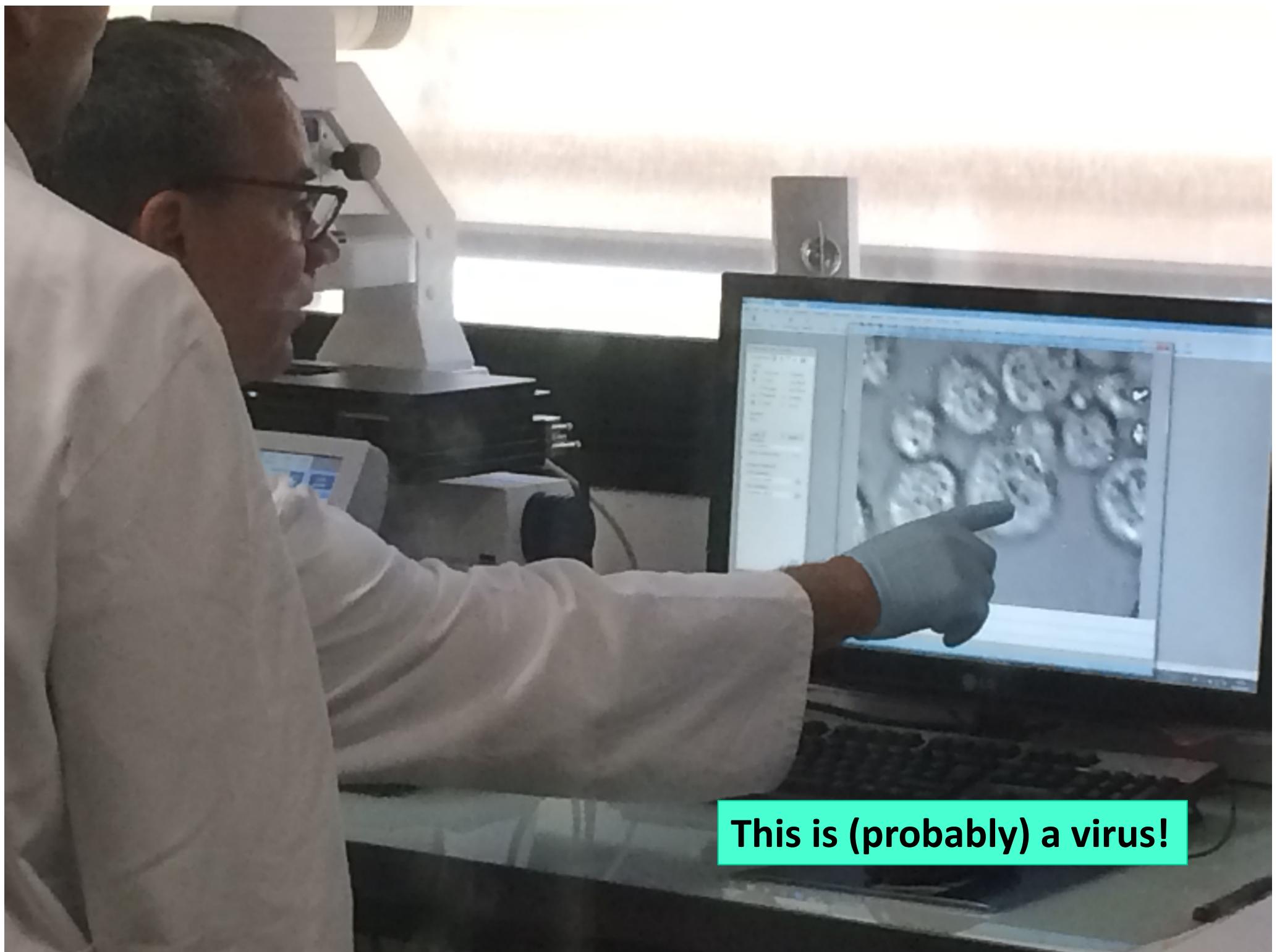


2  
re

NF

Jean-Marie Alempic





This is (probably) a virus!

# 6 isolates from 6 distant locations

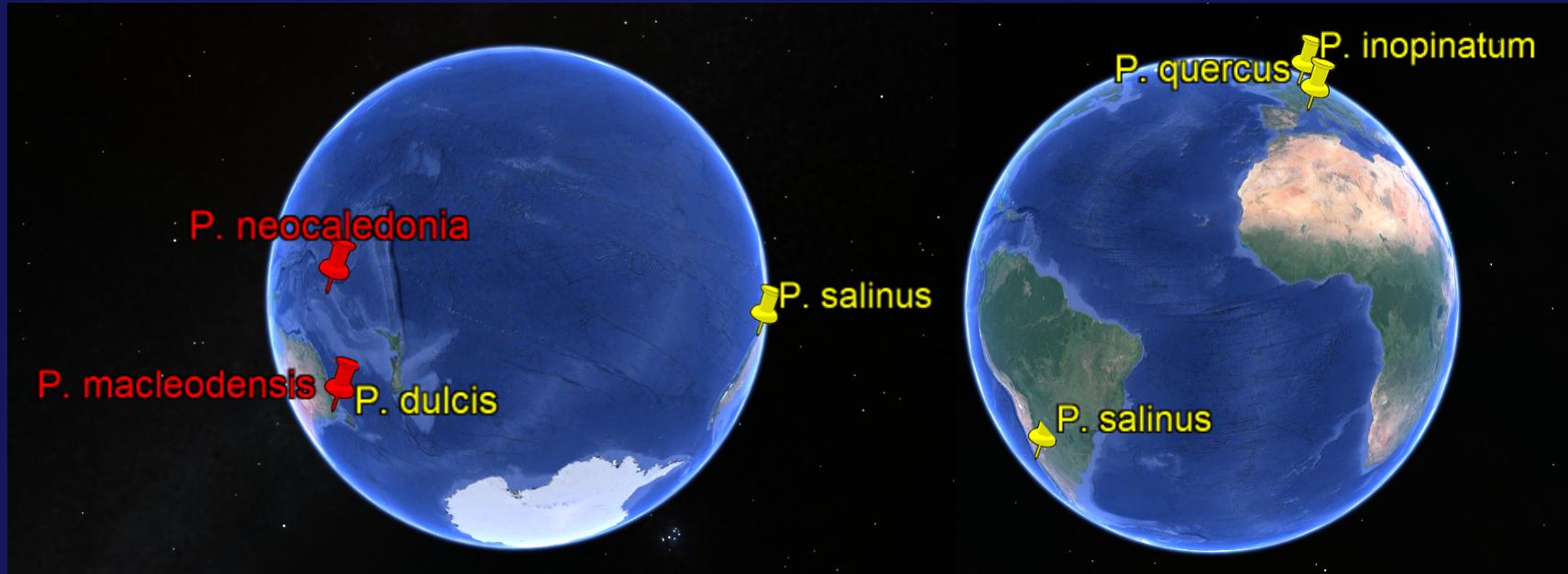
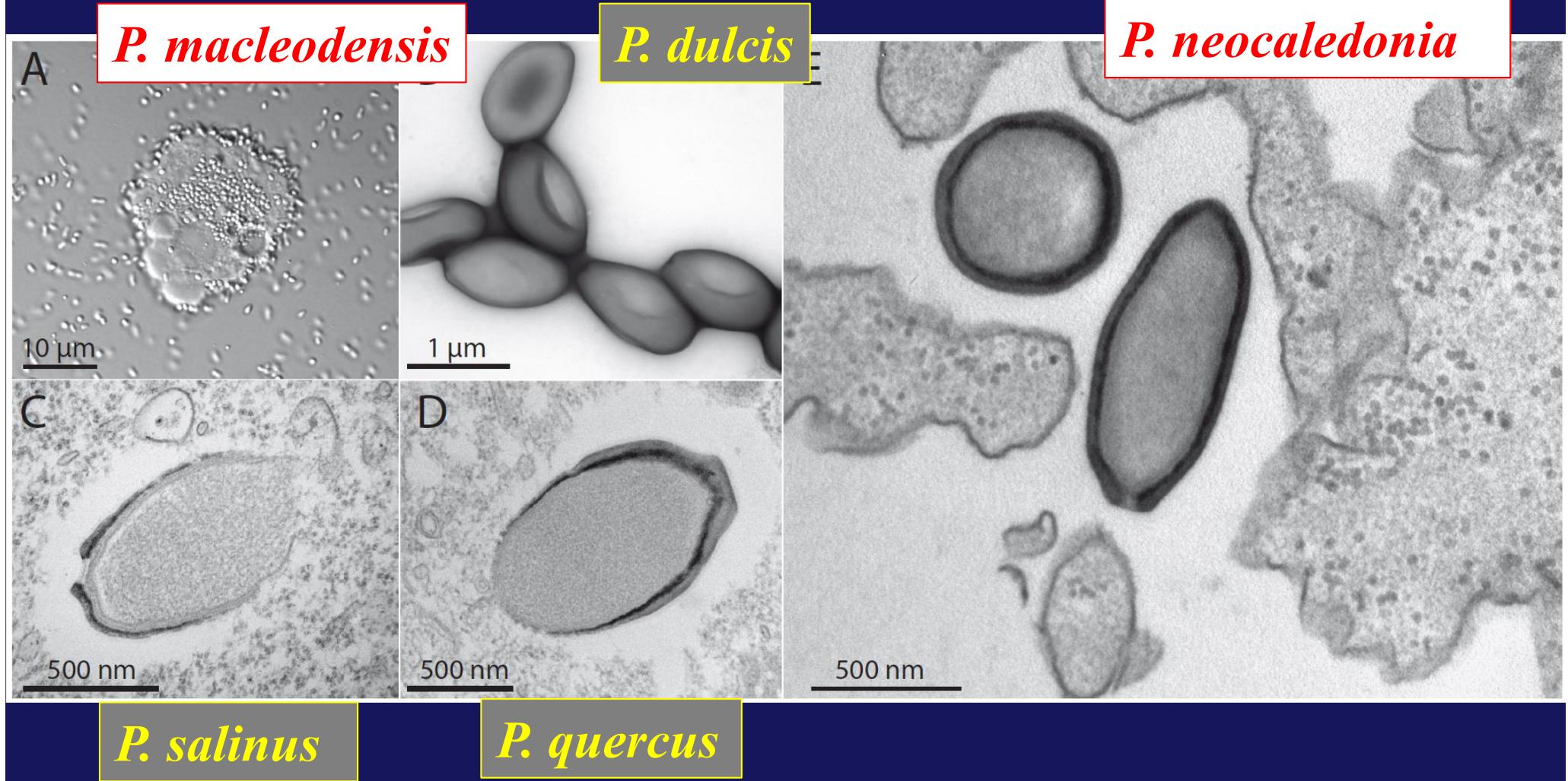


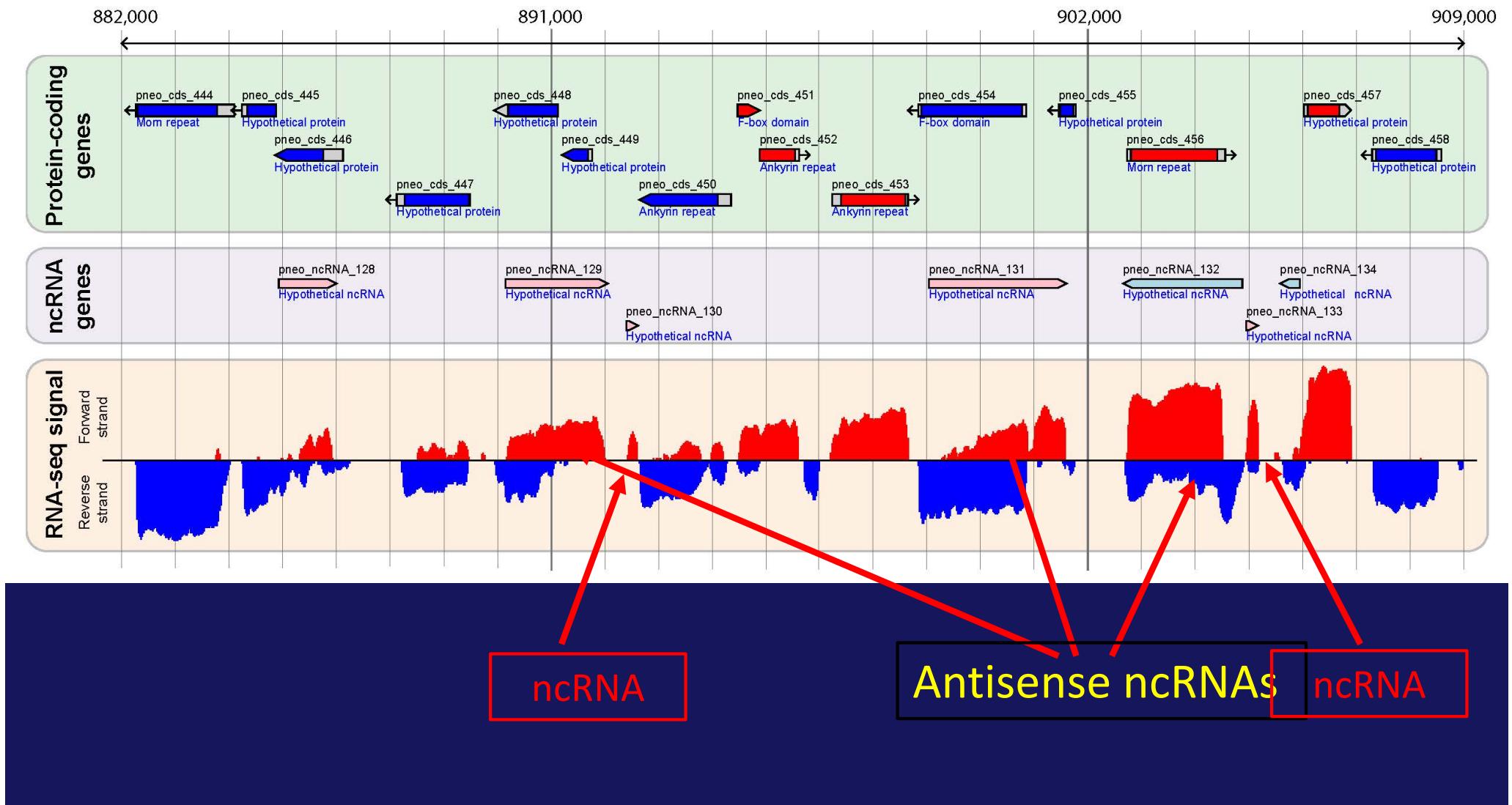
Table S2. Pairwise global protein sequence conservation based on 1:1 orthologs super-alignment.

	<i>P. salinus</i>	<i>P. inopinatum</i>	<i>P. quercus</i>	<i>P. dulcis</i>	<i>P. macleodensis</i>
<i>P. inopinatum</i>	73%				
<i>P. quercus</i>	74%	88%			
<i>P. dulcis</i>	70%	71%	72%		
<i>P. macleodensis</i>	54%	54%	55%	55%	
<i>P. neocaledonia</i>	54%	54%	54%	55%	76%

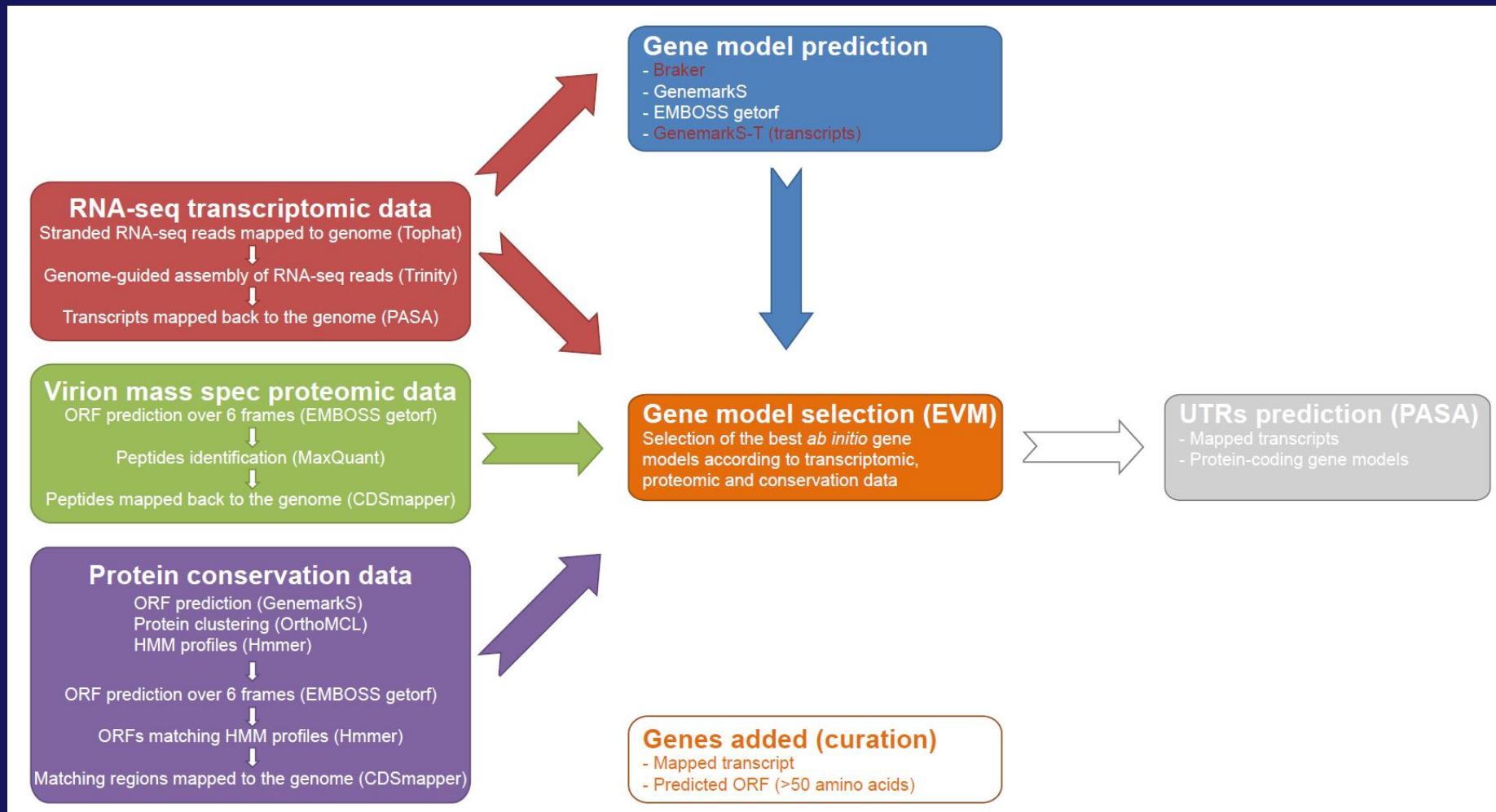
# 6 isolates looking all the same



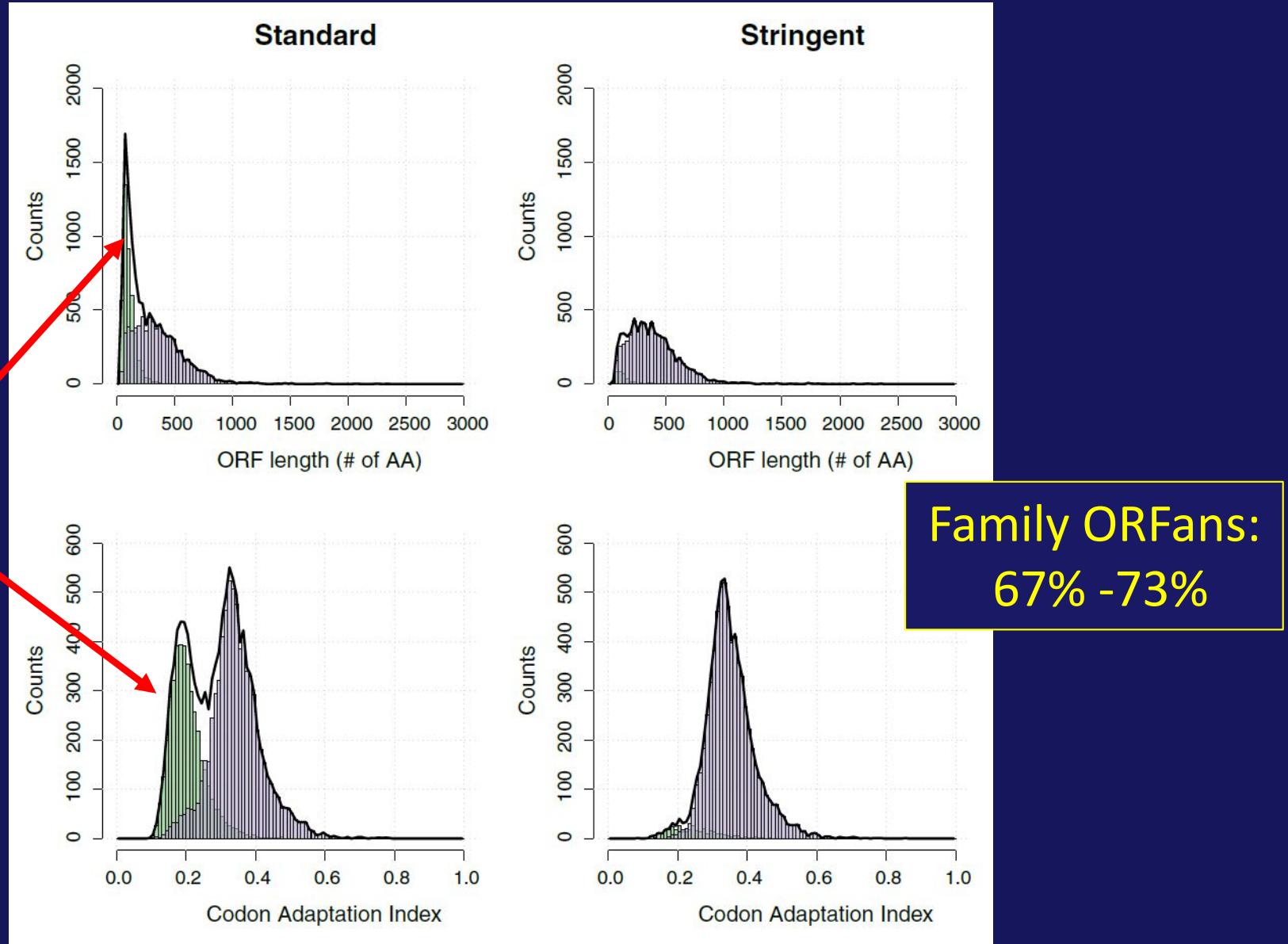
# No good annotation w/o RNA-Seq data



# A stringent reannotation: compensate high GC% - induced artefacts with additional information



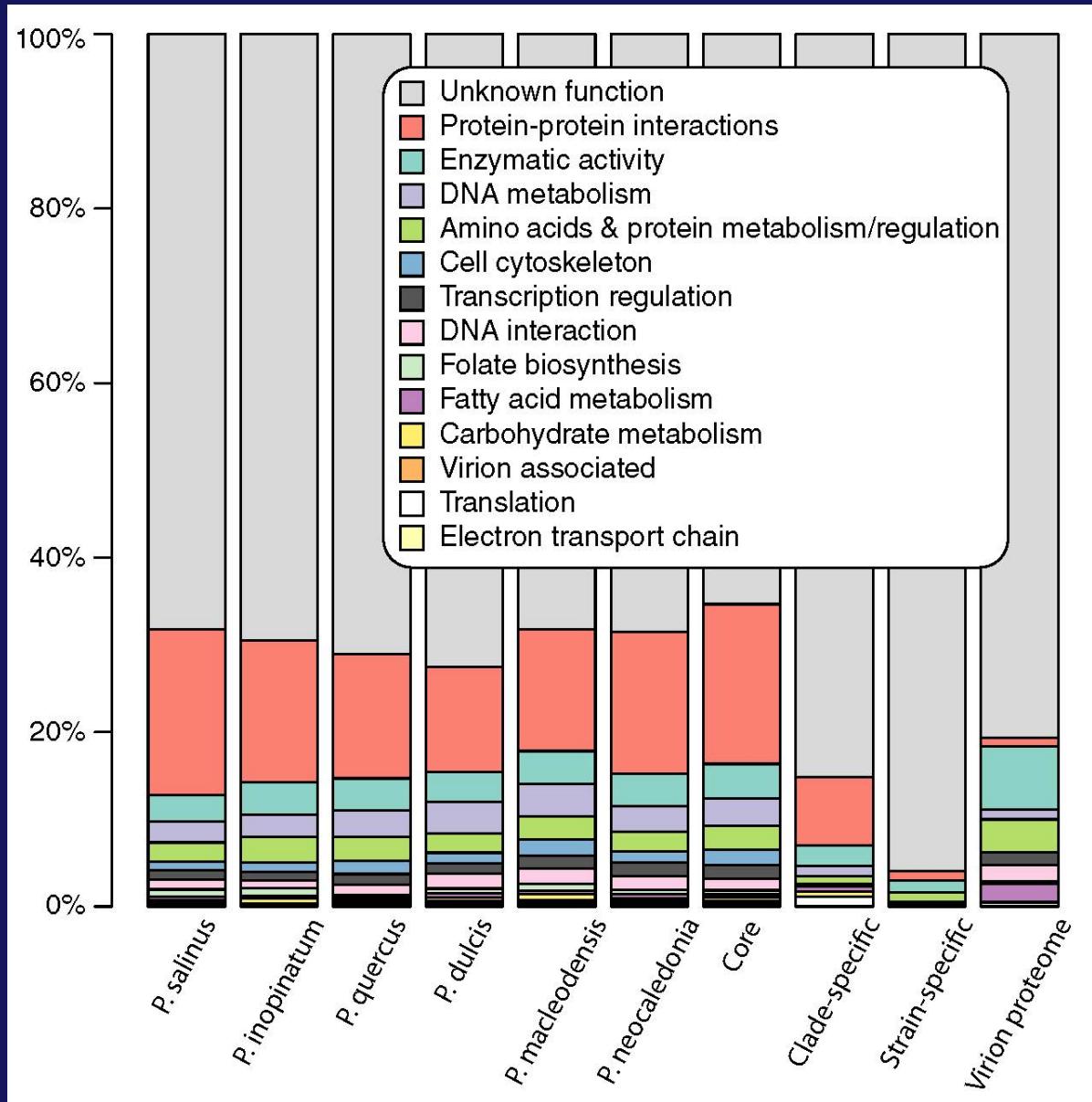
# Stringent annotation: a healthier starting point



# Pandoraviridae today: A stringent reannotation: up to 44% less protein-coding genes

Name	Origin	Genome	RNA-Seq	Particle Proteome	Genome size (bp) (G+C)%	N ORFs* (standard)	N Genes (stringent)
<i>P. salinus</i>	Chile	us	+	+	2,473,870 62%	2394 (2541)*	1430 ORFs 214 NC, 3 tRNA
<i>P. dulcis</i>	Australia	us	+	+	1,908,524 64%	1428 (1487)*	1070 ORFs 268 NC, 1 tRNA
<i>P. quercus</i>	France	+	+	+	2,077,288 61%	1863	1185 ORFs 157 NC, 1 tRNA
<i>P. neocalledonia</i>	New Caledonia	+	+	+	2,003,191 61%	1834	1081 ORFs 249 NC, 3 tRNA
<i>P. macleodensis</i>	Australia	+	-	-	1,838,258 58%	1552	926 ORFs 1 tRNA
<i>P. inopinatum</i>	Germany	Ref (8)	-	-	2,243,109 61%	2397 (1839)*	1307 ORFs 1 tRNA
<i>Megavirus chilensis</i>	Chile	us	us	us	1.26 Mb, 25.2%	1120	1108

# Stringent annotation: functional analysis



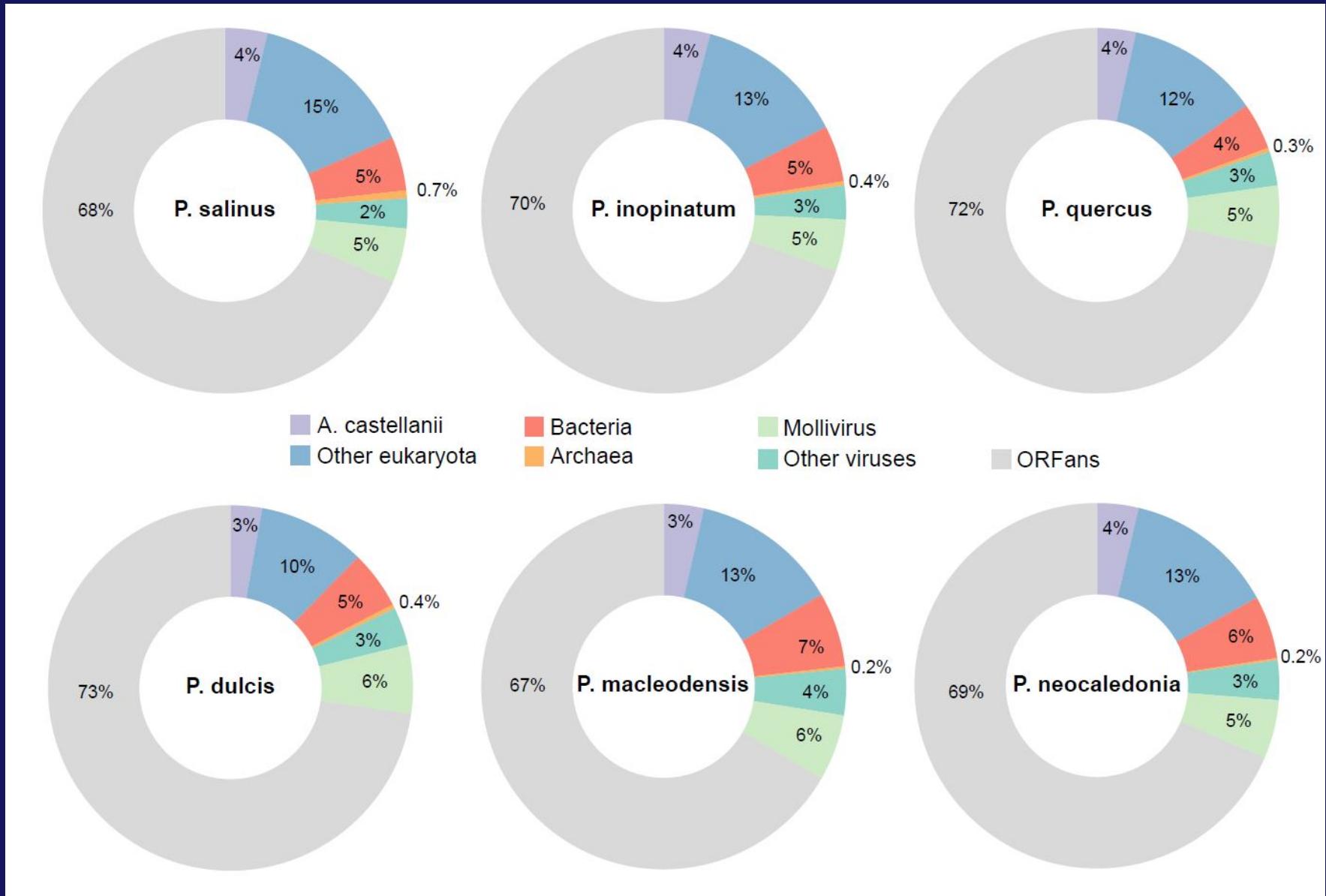
% Functional Attributes:  
strain-specific < clade-specific < Pandoravirions < average < core

What could explain

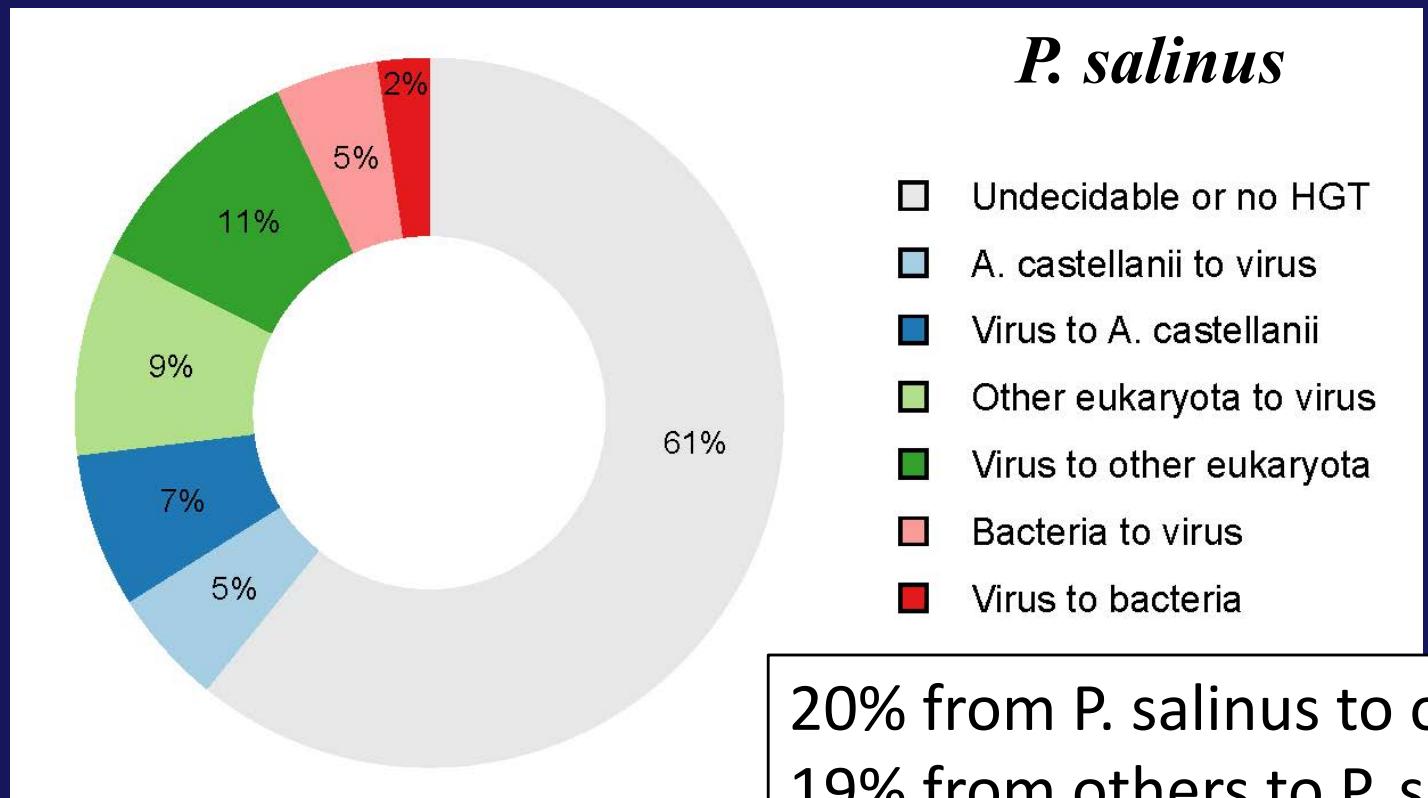
- the uniquely large genome of Pandoraviruses ?
- the large proportion of anonymous proteins
- the large proportion of ORFans ?

- a huge frequency of gene gain through HGT ?
- a huge frequency of gene duplication ?
- a hugely complex ancestor ?
- anything else ?

# Stringent annotation: still a lot of family ORFans

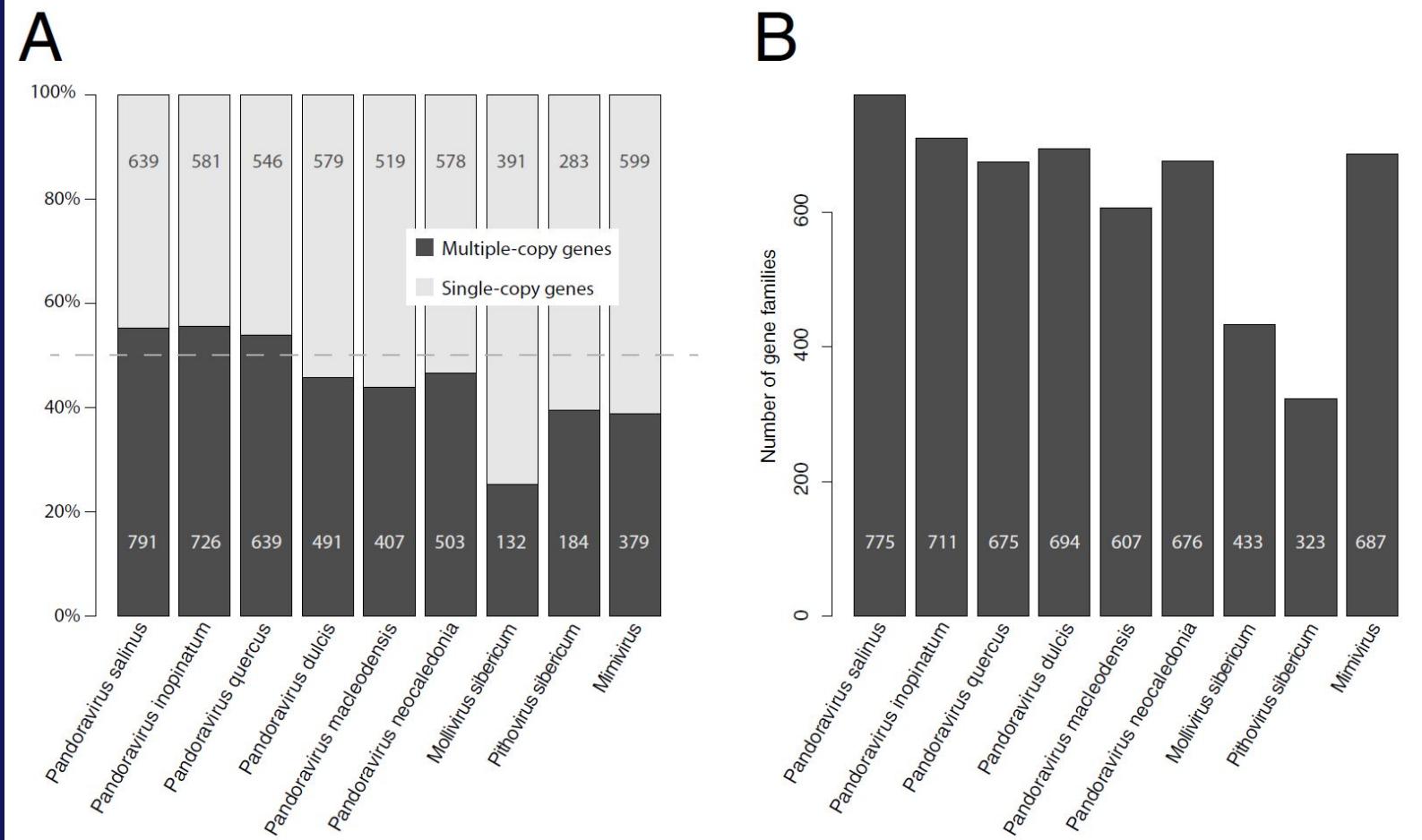


HGTs: contributed at most 15% of the gene content  
(at least) 6%



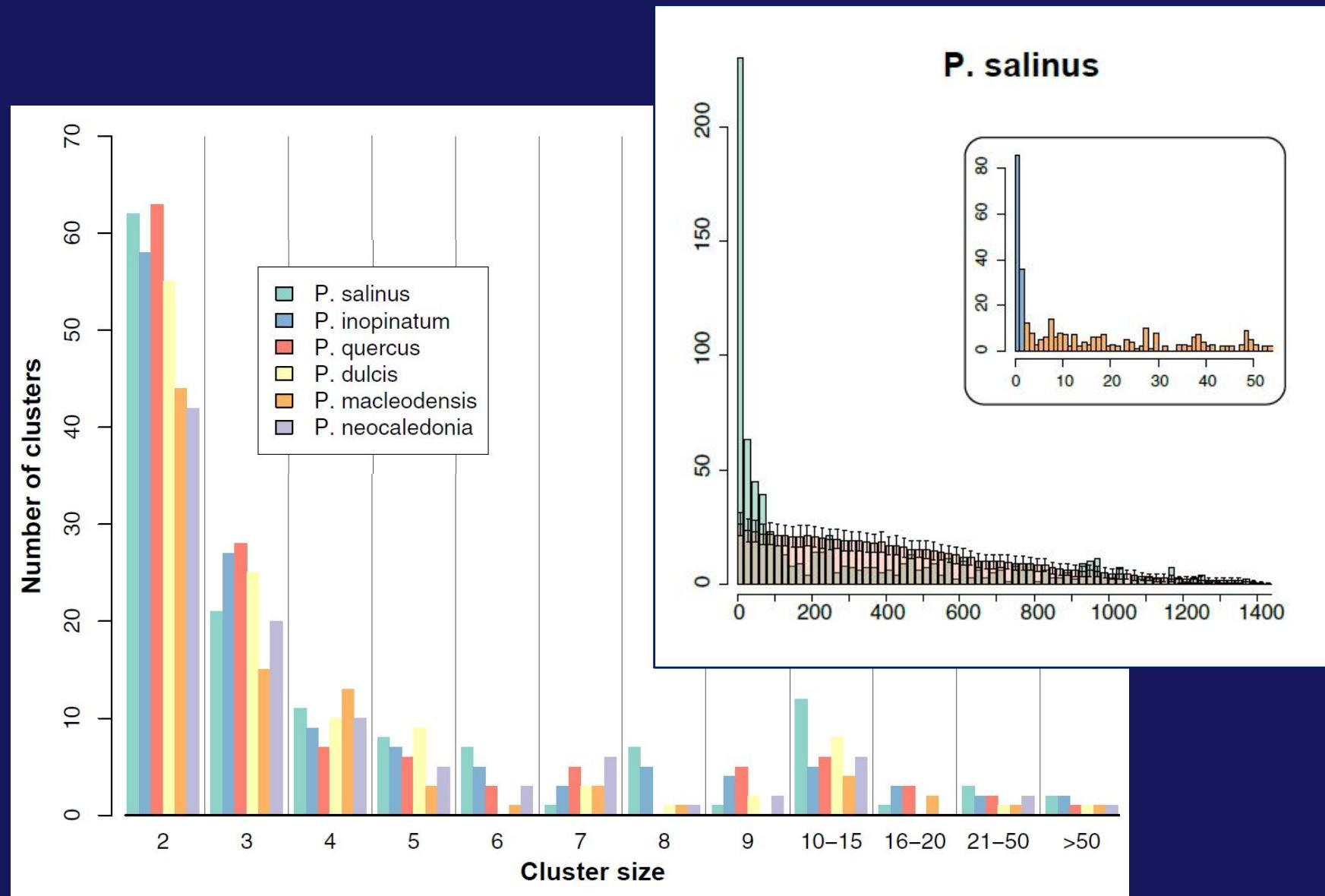
Nothing special compared to other large dsDNA viruses

# Duplication analysis



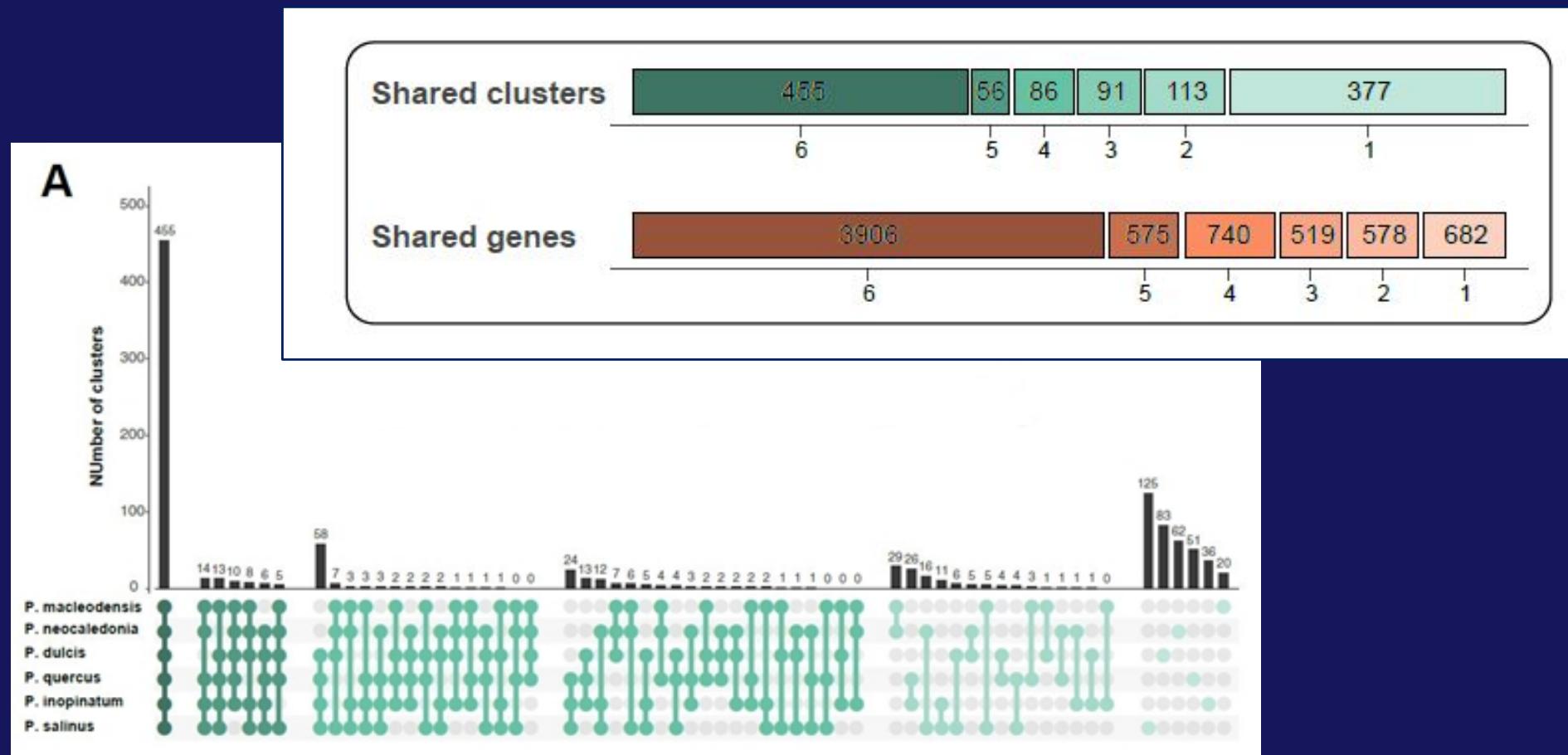
Not so different from Mimivirus (half the size)

# Duplications are mostly tandem repeats



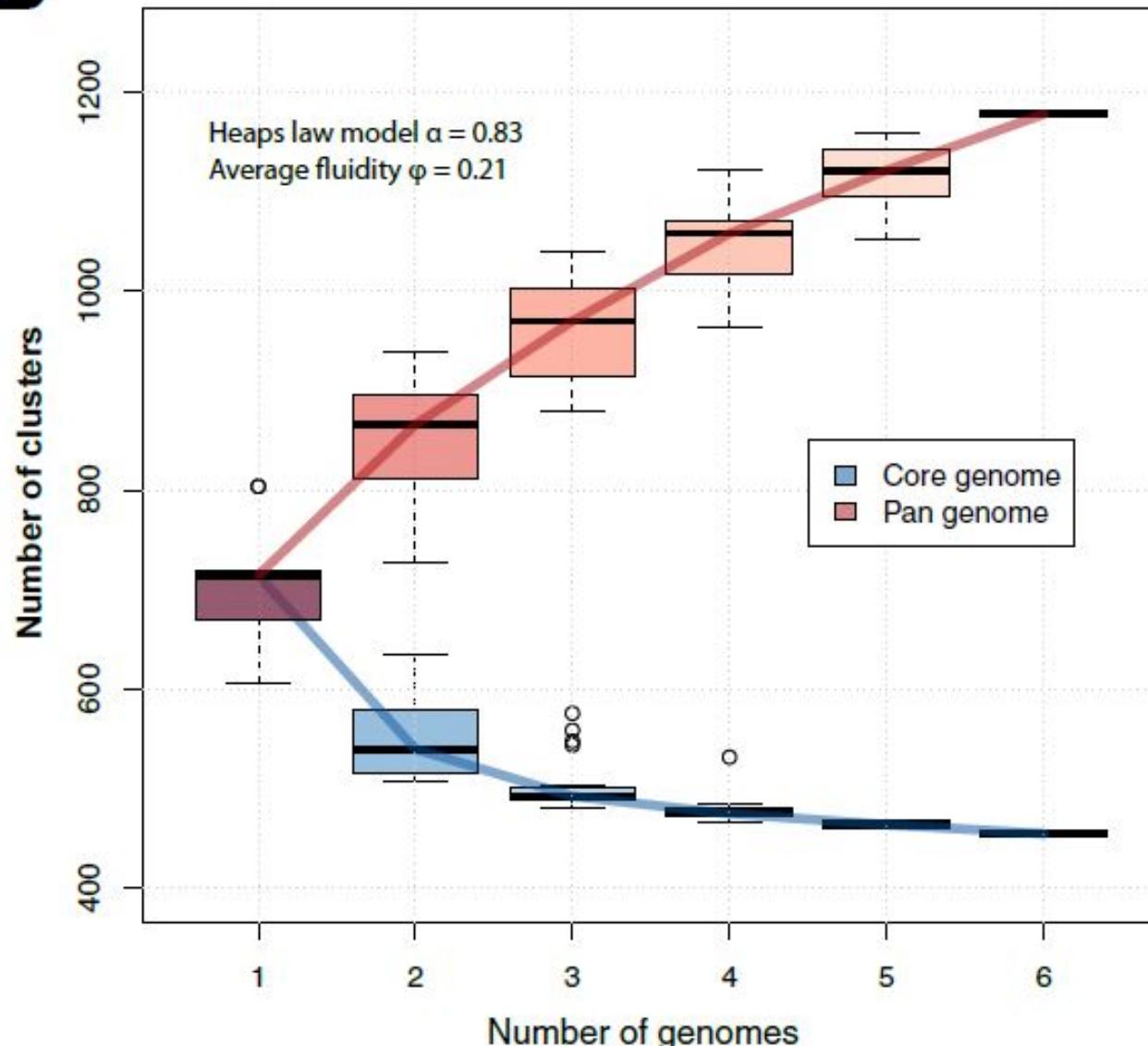
# The Pandoravirus genomes are diverse

- core: 455 clusters
- strain-specific: 377 clusters



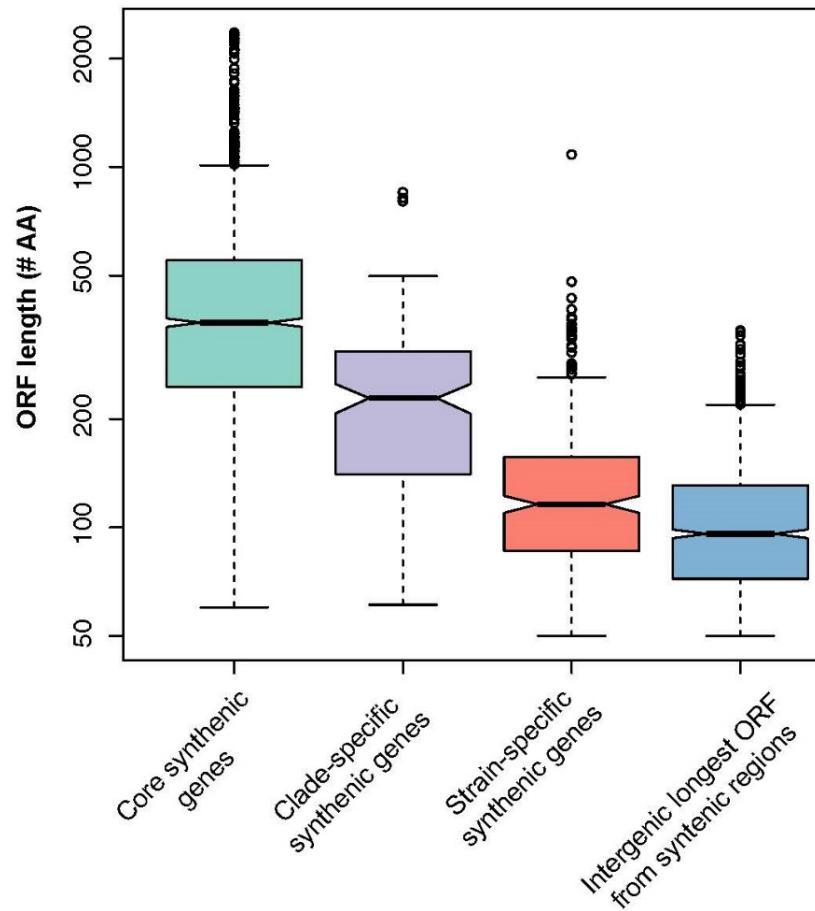
# The Pandoraviridae pan genome is ... open!

B

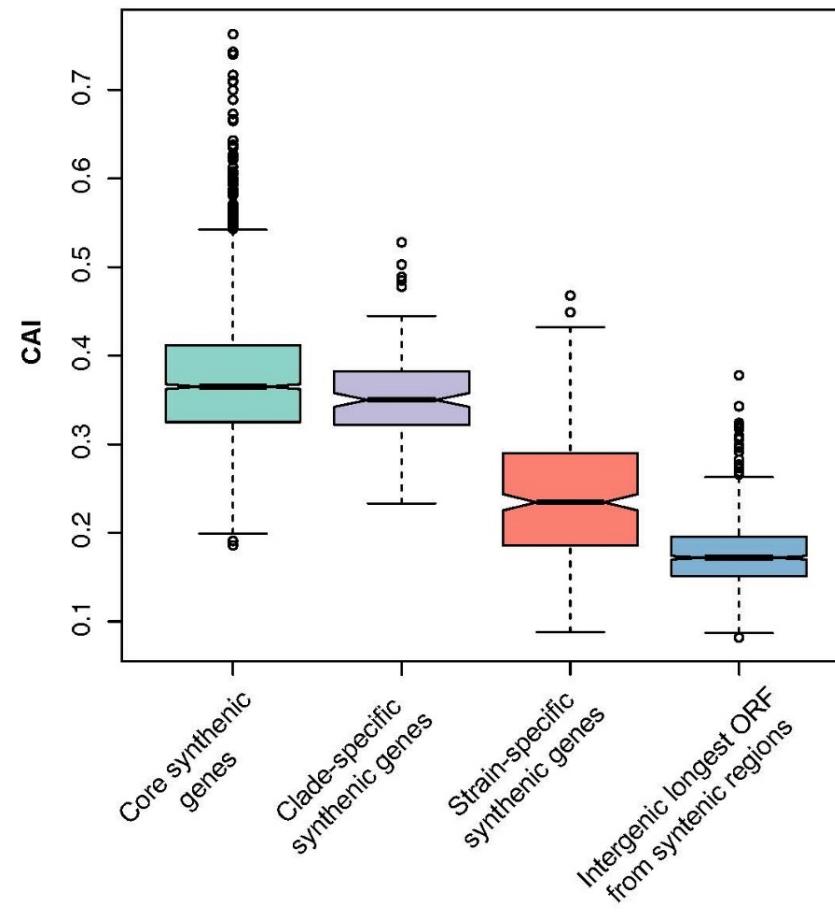


# Strain-specific genes: statistical similarity with intergenic regions ?

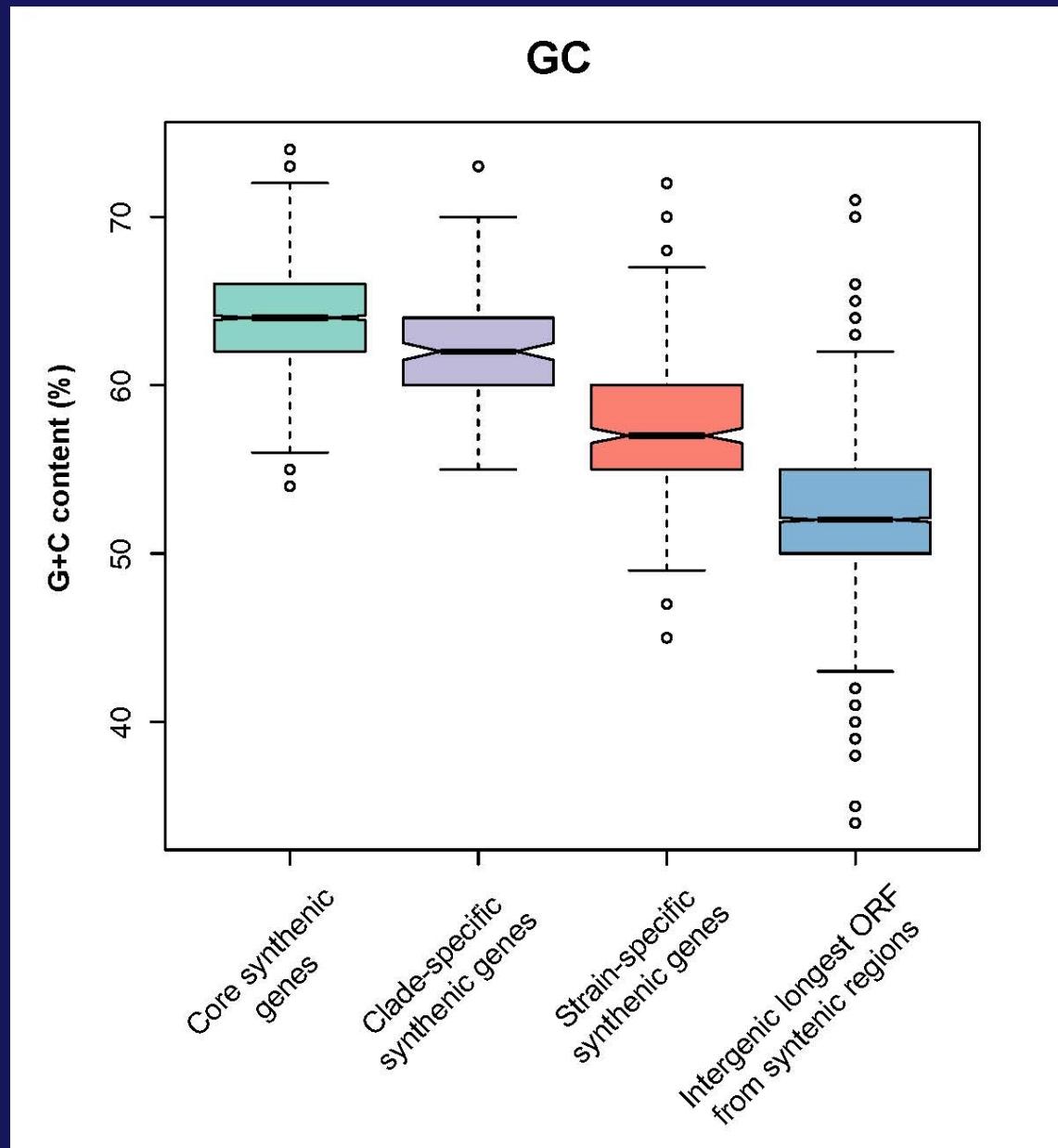
ORF length



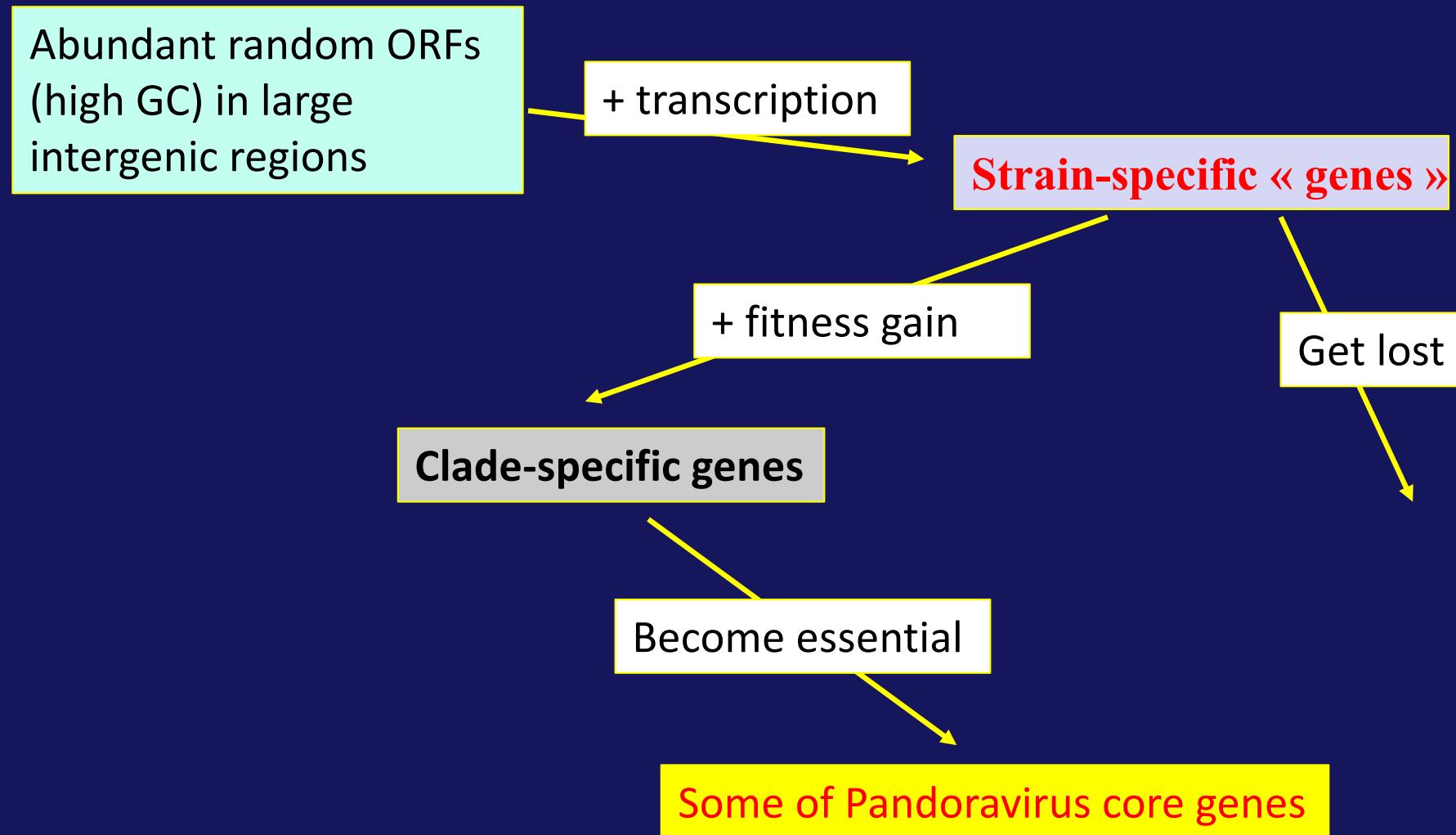
CAI



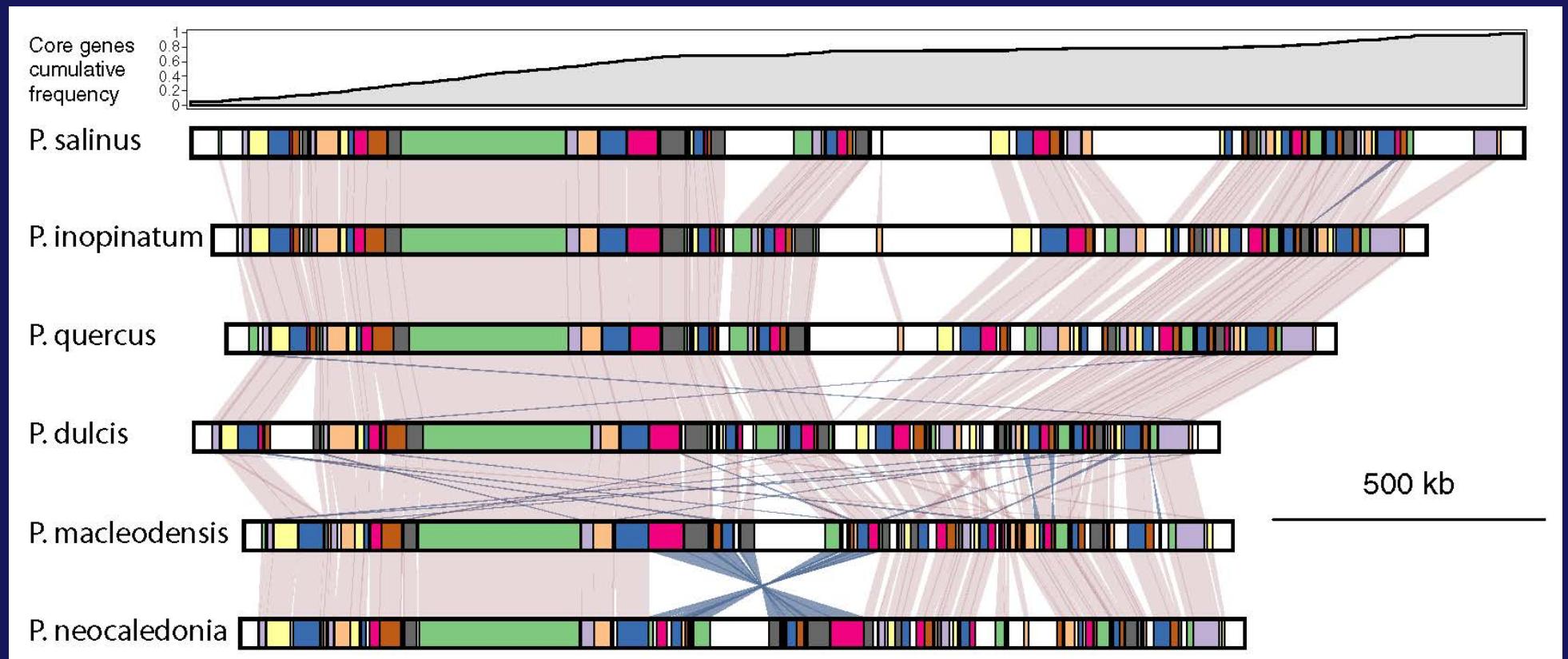
# Strain-specific genes: statistical similarity with intergenic regions ?



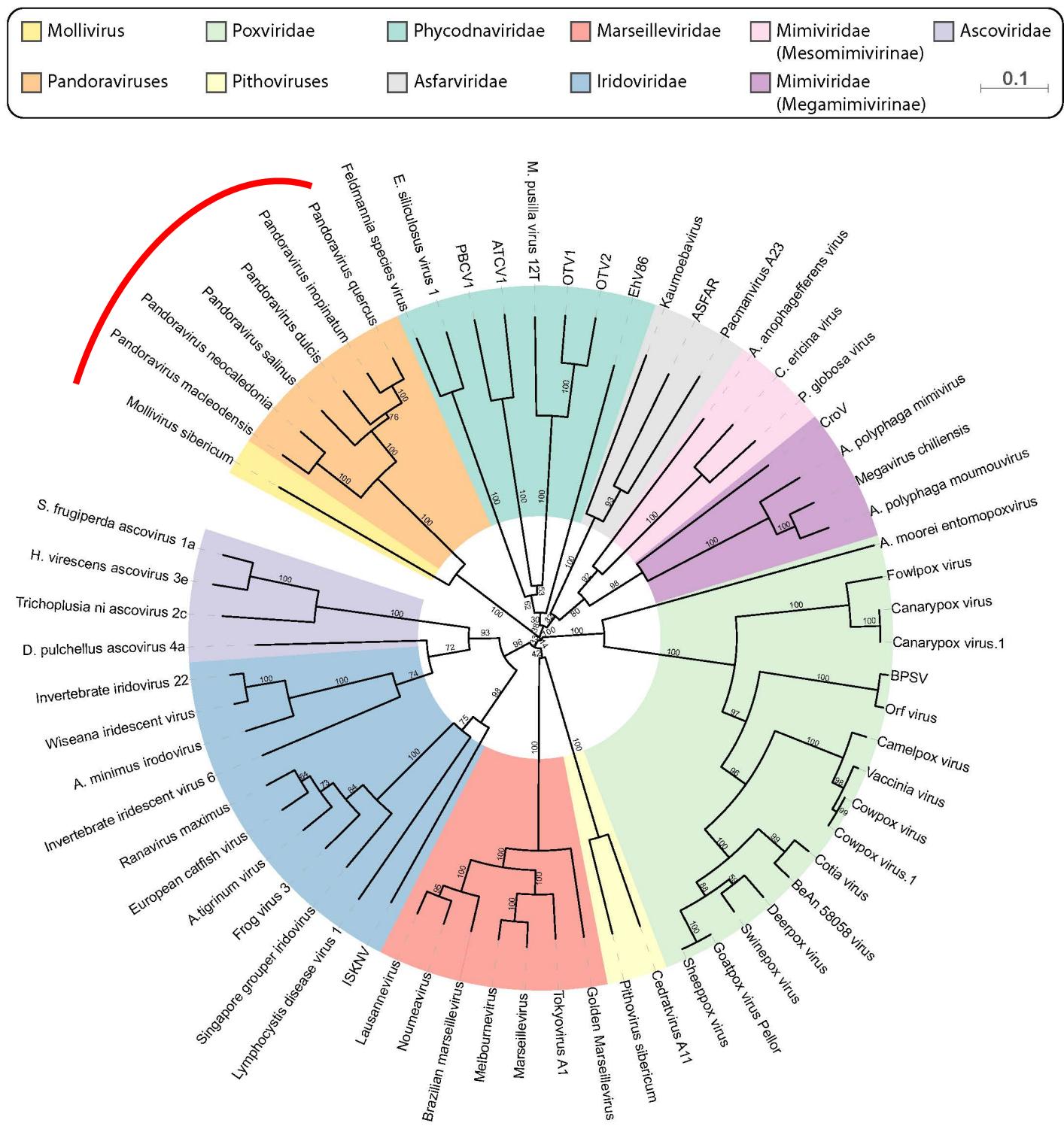
# The *de novo* gene creation hypothesis



# The *de novo* gene creation hypothesis would maintain the overall collinearity



# Despite their divergence, the Pandoraviruses constitute a tight new family

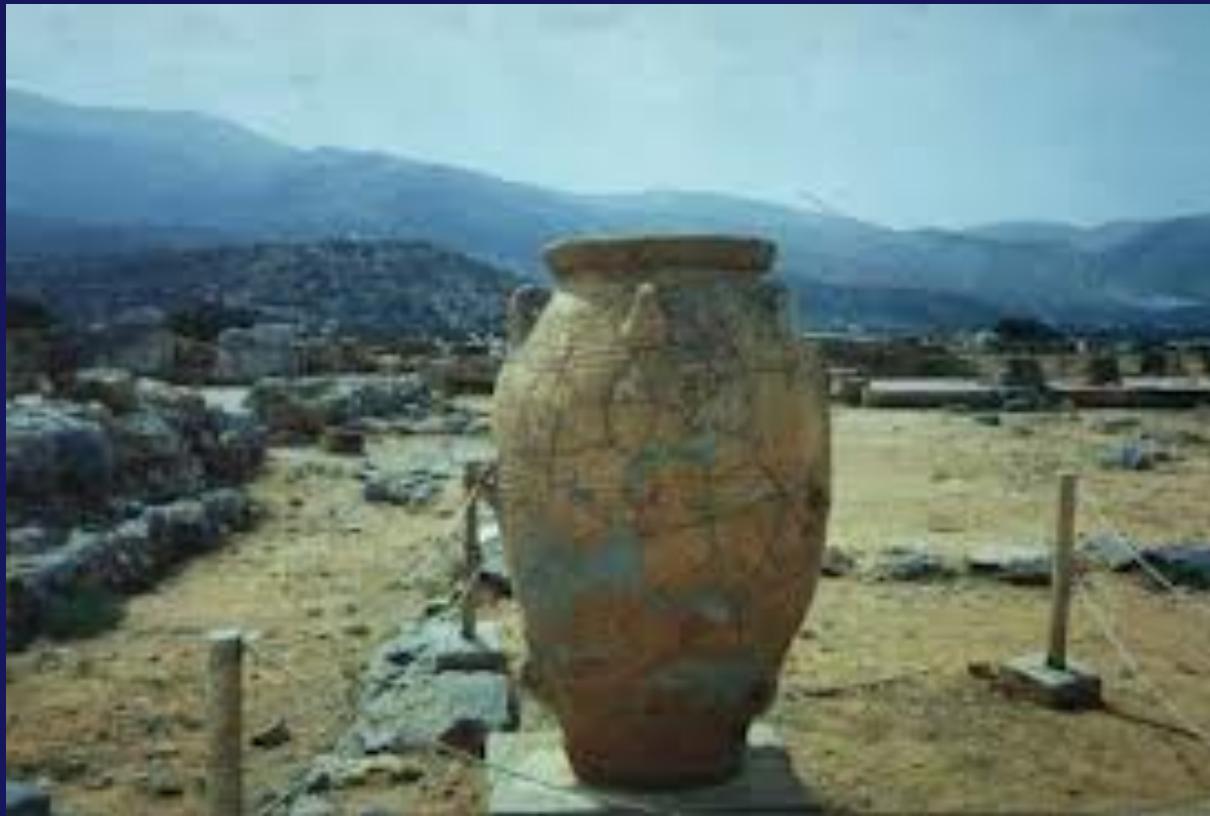


# Pandoravirididae today

(Acanthamoeba-infecting)

Clade	Prototype	Virion type	Dimension	Genome, size, GC%	Specific features
		Amphora		L DNA, term. repeats	Ostiole, tegument
A	<i>P. salinus</i>	Amphora	1000x500 nm	2.77 Mb, 61.7%	
A	<i>P. quercus</i>	Amphora	1000x500 nm	2.07 Mb, 61%	
A	<i>P. inopinatum</i>	Amphora	1000x500 nm	2.24 Mb, 60.6%	
A	<i>P. dulcis</i>	Amphora	1000x500 nm	1.91 Mb, 63.7%	
B	<i>P. neocaledonia</i>	Amphora	1000x500 nm	2 Mb, 61%	
B	<i>P. macleodensis</i>	Amphora	1000x500 nm	1.84 Mb, 58%	

# Opening Pandora's "box"



Philippe et al.  
(July 2013)

# Regeneration of whole fertile plants from 30,000-y-old fruit tissue buried in Siberian permafrost

Svetlana Yashina<sup>a,1</sup>, Stanislav Gubin<sup>b</sup>, Stanislav Maksimovich<sup>b</sup>, Alexandra Yashina<sup>a</sup>, Edith Gakhova<sup>a</sup>, and David Gilichinsky<sup>b,2</sup>

Institutes of <sup>a</sup>Cell Biophysics and <sup>b</sup>Physicochemical and Biological Problems in Soil Science, Russian Academy of Sciences, Pushchino 142290, Russia

Edited\* by P. Buford Price, University of California, Berkeley, CA, and approved January 25, 2012 (received for review November 8, 2011)

Whole, fertile plants of *Silene stenophylla* Ledeb. (Caryophyllaceae) have been uniquely regenerated from maternal, immature

However, to date, no viable flowering plant remains have been discovered from these ancient permafrost sediments.

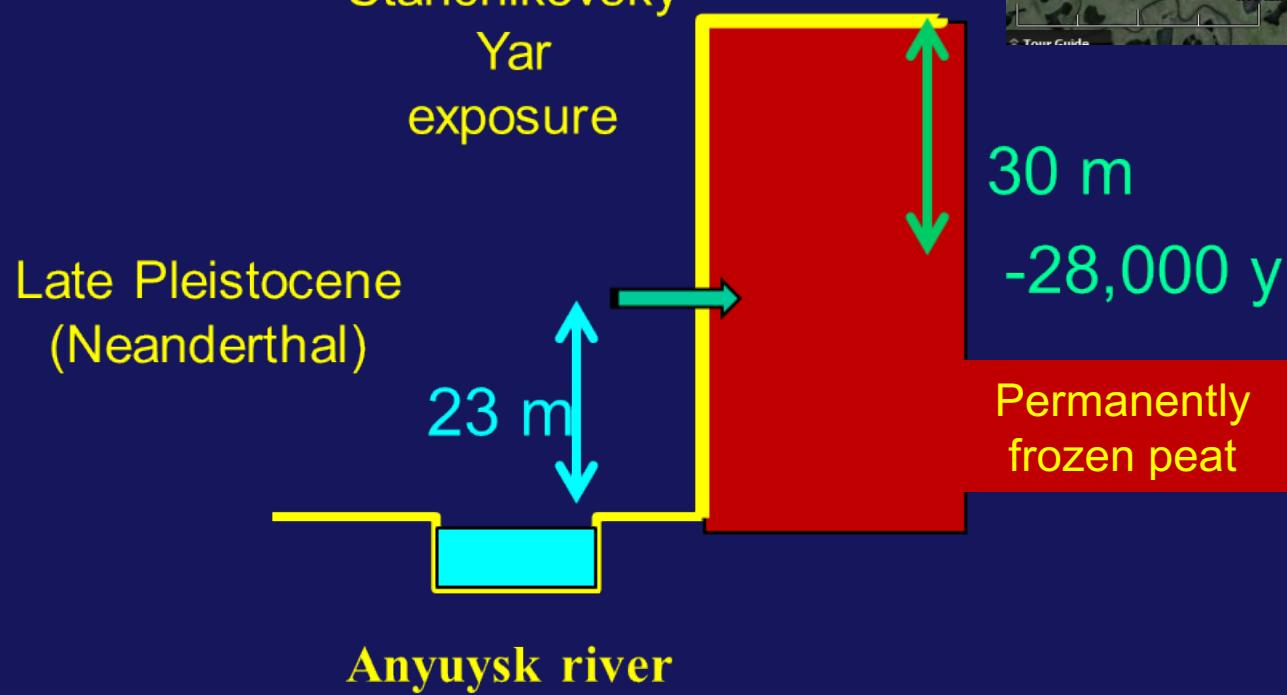
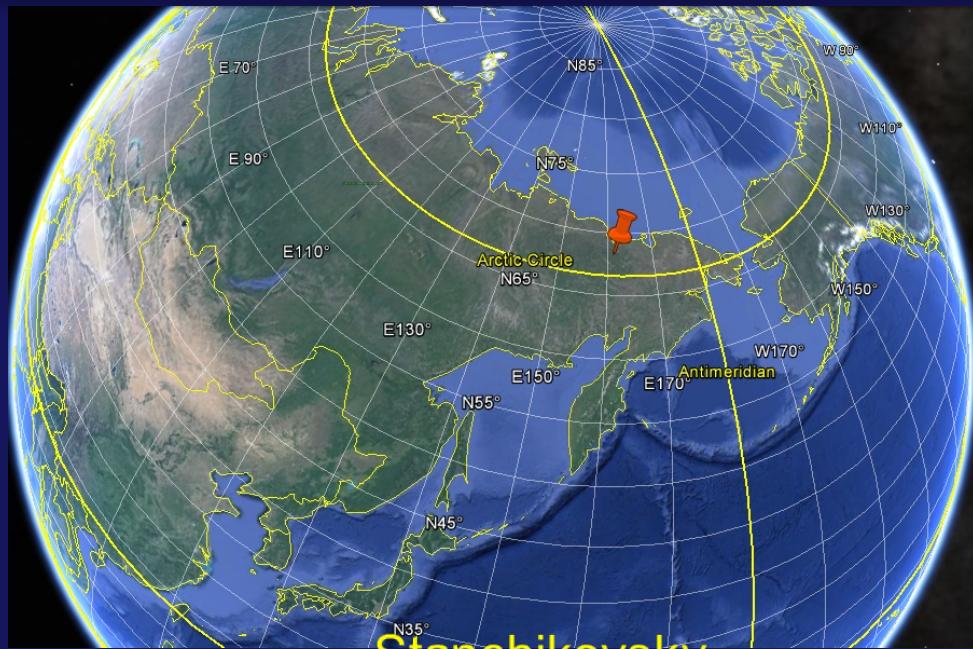
Geocryology Lab (Pushchino)  
studying permafrost microbes  
since 1985

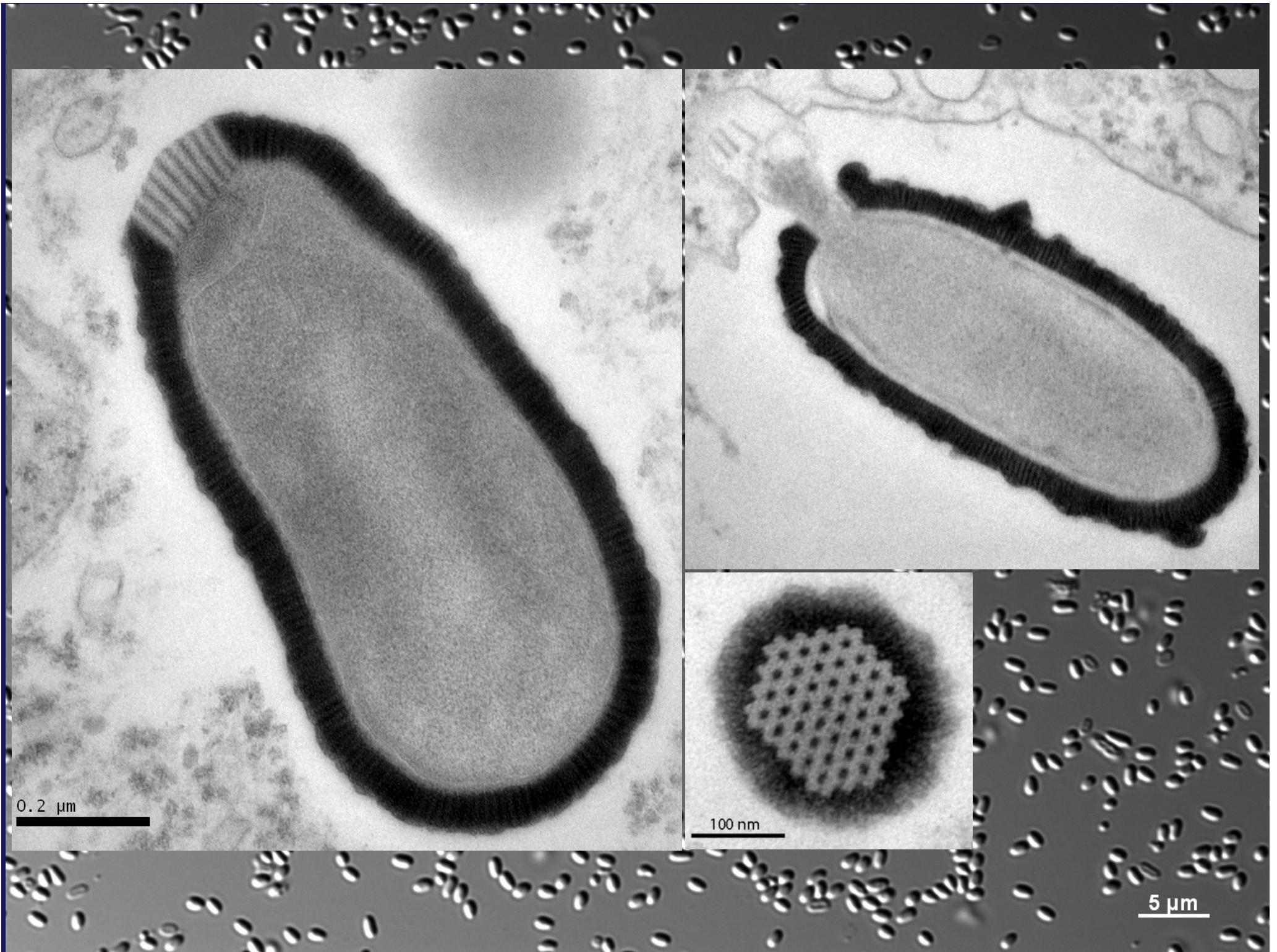
Lyubov Shmakova· Elizaveta Rivkina  
Institute of Physicochemical & Biological Problems in Soil  
Science, Pushchino, Russia.



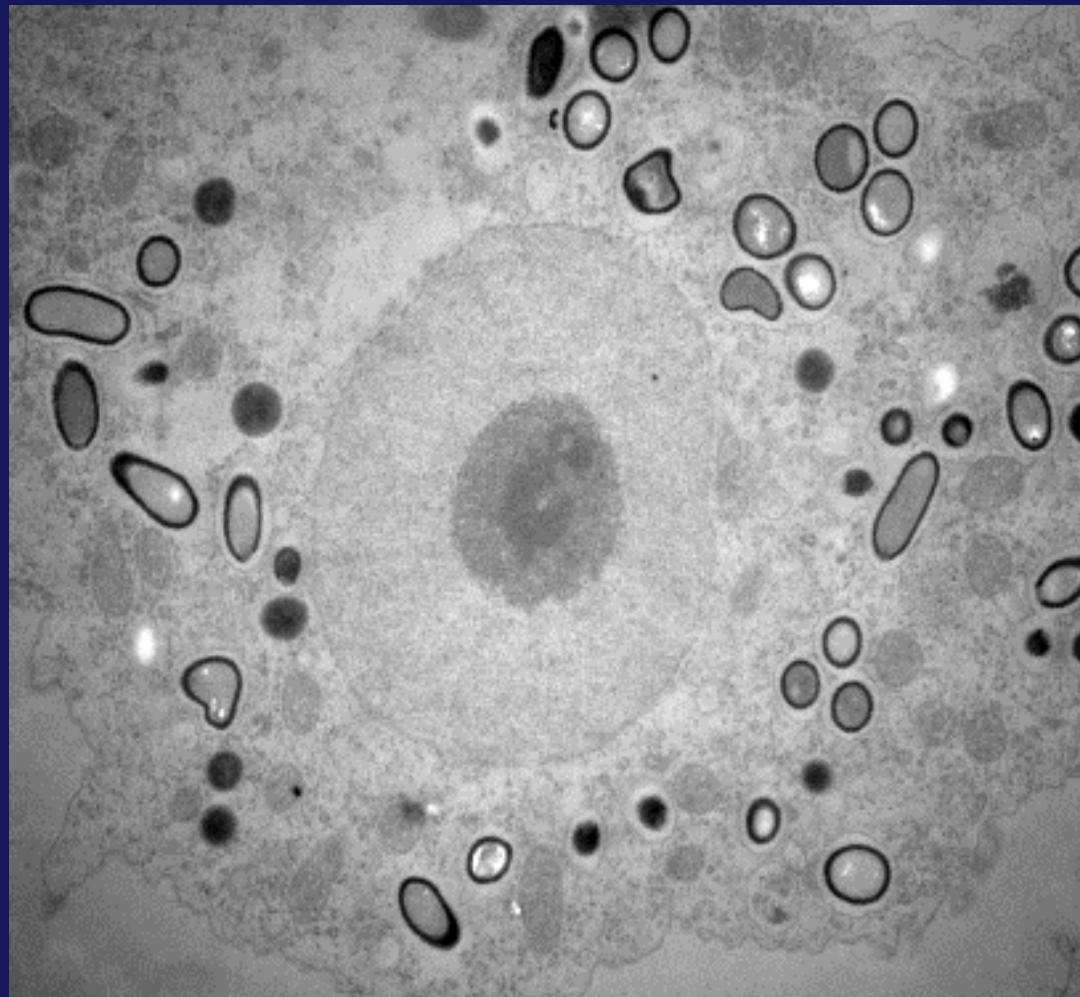
Dr. David Gilichinsky  
(1948 – 2012)

# Looking back in time



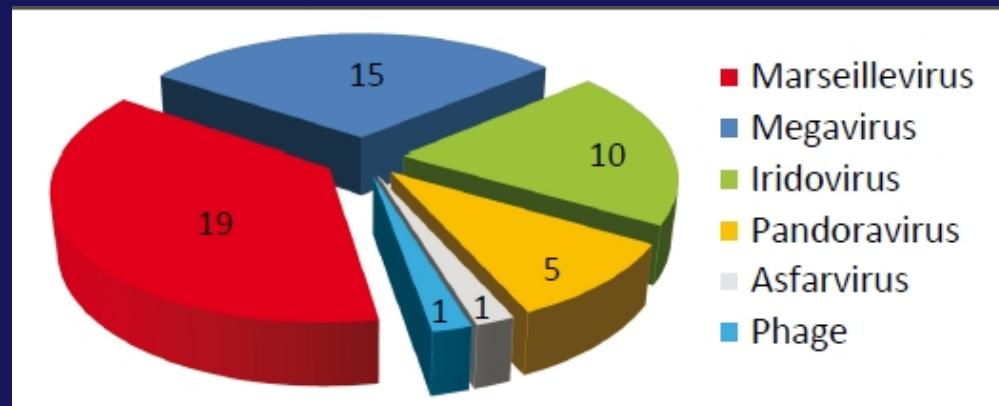
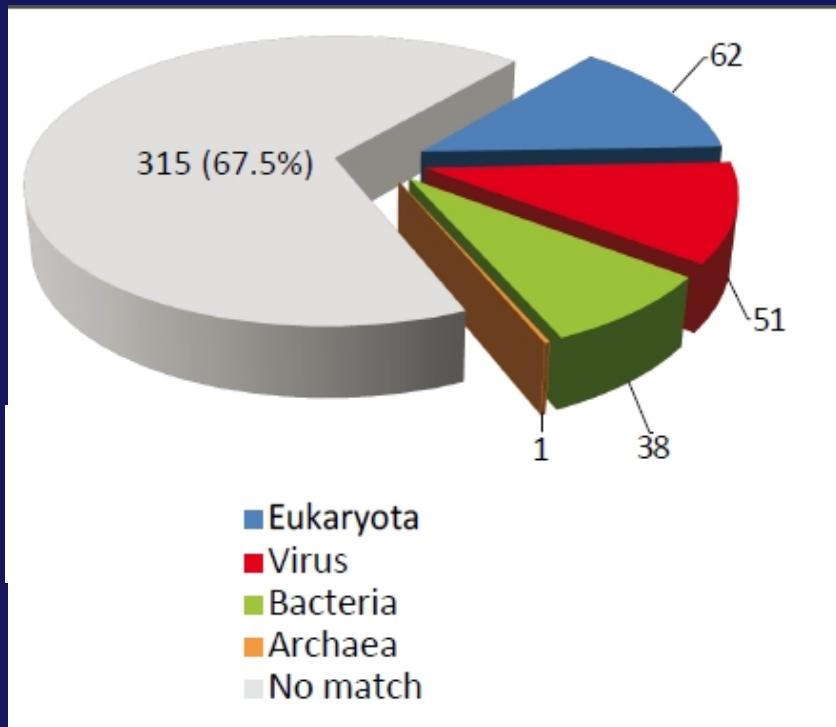


# *Pithovirus sibericum*

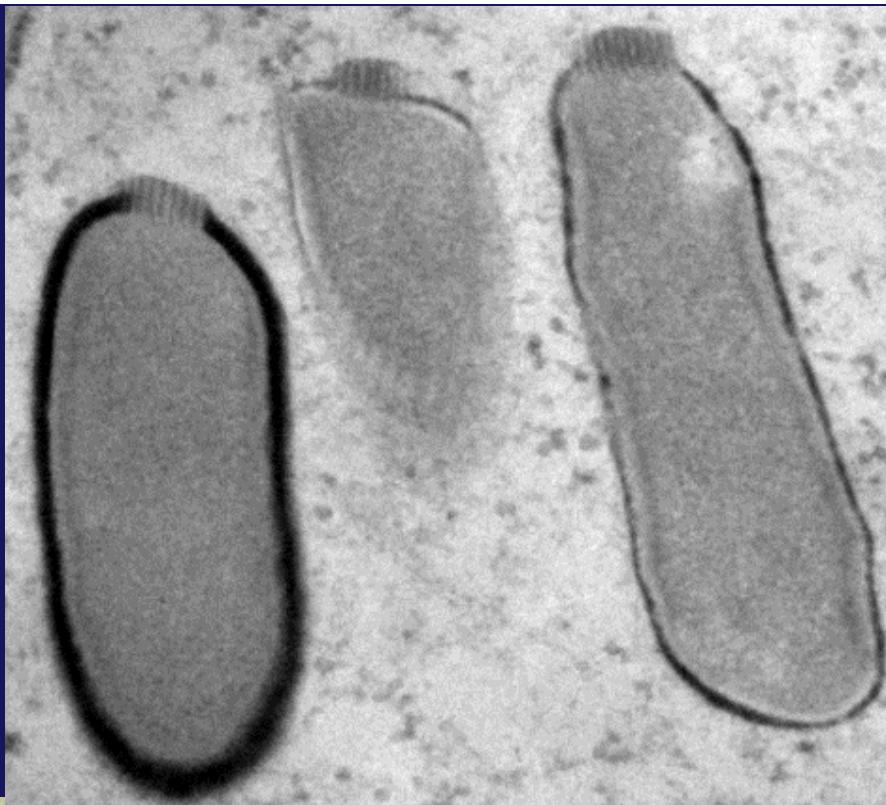


Like Poxviridae or Mimiviridae: intracytoplasmic replication  
(encoded /packaged transcription machinery)

# Pithovirus genome: 610 kb, circular



AT-rich (64%), 467 proteins, 67% ORFans, 73 % unknown  
9% genes similar to Iridoviruses and Marseillevirus



# Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology

March 3rd, 2014

Matthieu Legendre<sup>a,1</sup>, Julia Bartoli<sup>a,1</sup>, Lyubov Shmakova<sup>b</sup>, Sandra Jeudy<sup>a</sup>, Karine Labadie<sup>c</sup>, Annie Adrait<sup>d</sup>, Magali Lescot<sup>a</sup>, Olivier Poirot<sup>a</sup>, Lionel Bertaux<sup>a</sup>, Christophe Bruley<sup>d</sup>, Yohann Couté<sup>d</sup>, Elizaveta Rivkina<sup>b</sup>, Chantal Abergel<sup>a,2</sup>, and Jean-Michel Claverie<sup>a,e,2</sup>

<sup>a</sup>Structural and Genomic Information Laboratory, Unité Mixte de Recherche 7256 (Institut de Microbiologie de la Méditerranée) Centre National de la Recherche Scientifique, Aix-Marseille Université, 13288 Marseille Cedex 9, France; <sup>b</sup>Institute of Physicochemical and Biological Problems in Soil Science, Russian Academy of Sciences, Pushchino, Moscow Region, Russia; <sup>c</sup>CEA, Institut de Radioprotection et de Sûreté Nucléaire, Paris, France; <sup>d</sup>UMR 7253, Institut de Biologie Structurale, Villeurbanne, France; <sup>e</sup>CNRS, Institut de Biologie Structurale, Villeurbanne, France

## EDITORS' CHOICE

EDITED BY KRISTEN MUELLER AND JESSE SMITH

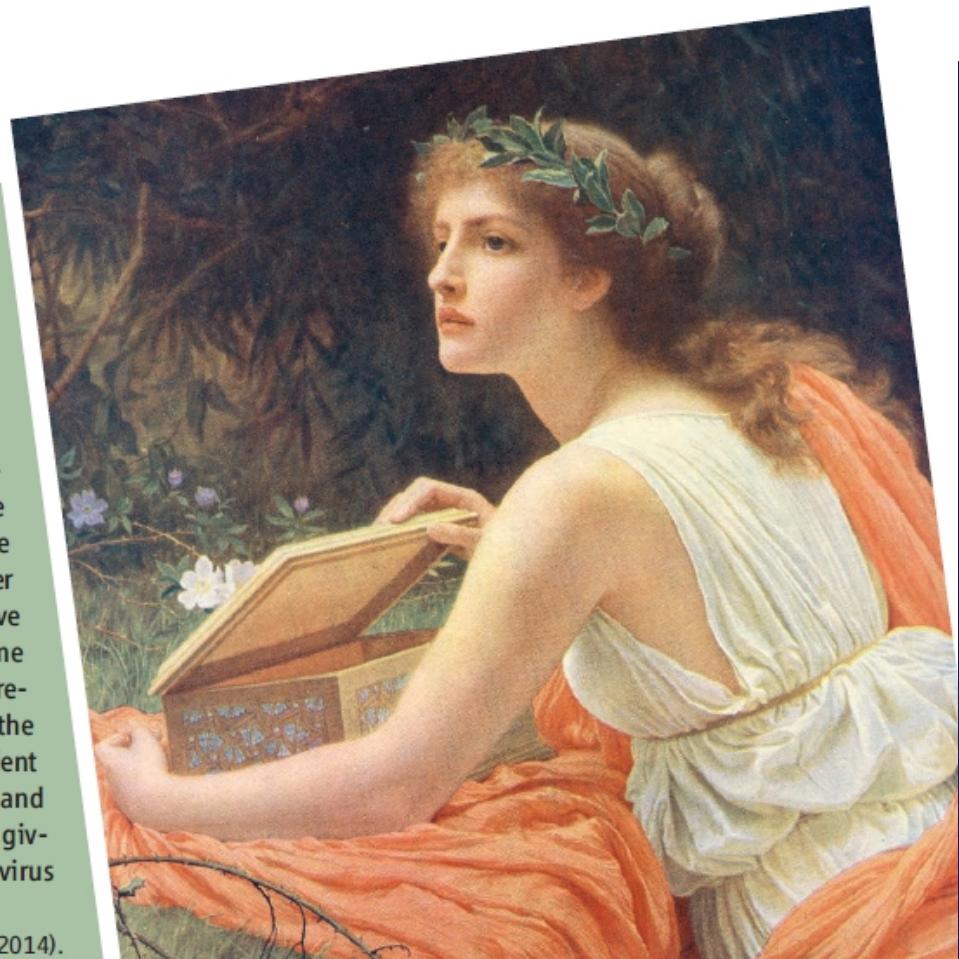
VIROLOGY

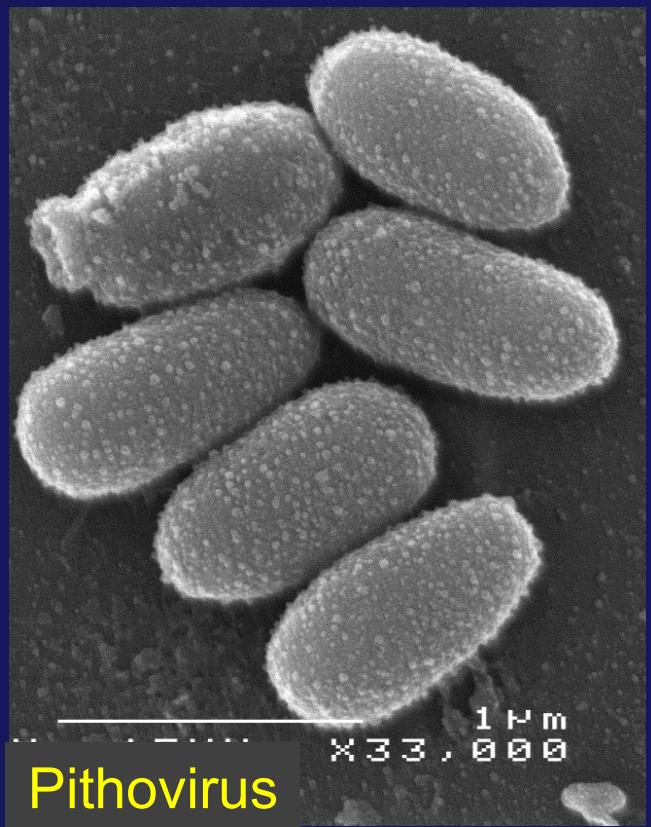
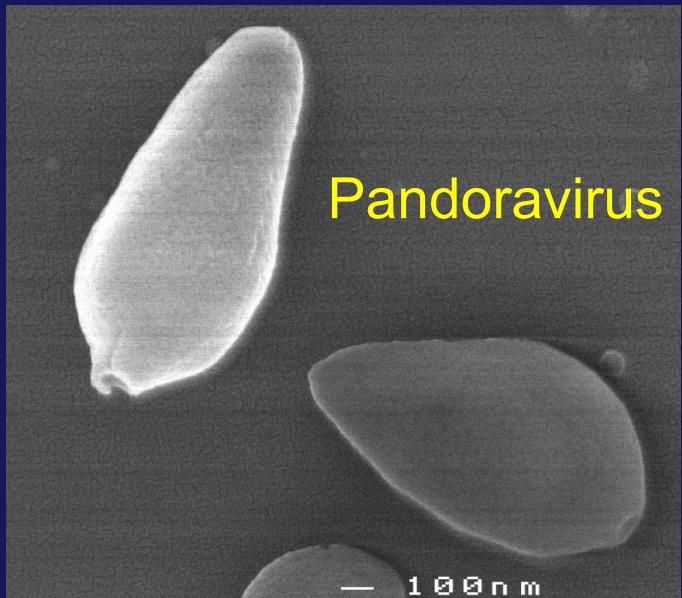
Science, March 6th

### Opening Pandora's Box

The first woman on Earth, Pandora, had a "box," or rather a jar, that Zeus commanded her to safeguard and never open. Of course she opened it, and thus evil spread around the world. Recently, an extraordinarily distinctive group of giant viruses that parasitize amoebas were described and named Pandoravirus, not because they contain all evil but merely because they are jar-shaped. Legendre *et al.* have added to this still-tiny pantheon with another jar-shaped viral particle 1.5  $\mu\text{m}$  long, containing a rather diminutive 600-kb AT-rich genome (as compared to the up to 2.8-Mb genome seen in Pandoraviruses) and a cytoplasmic replication machinery resembling that of the original Megaviridae. The authors named the virus Pithovirus because Pandora's jar was called a "pithos" in ancient Greek. This virus was revived from a Siberian permafrost sample and infects amoebas. Although named for the jar and not its contents, given its origins, this discovery hints that viruses more evil than Pithovirus might be revived as the tundra melts. — CA

Proc. Natl. Acad. Sci. U.S.A. 111, 10.1073/pnas.1320670111 (2014).





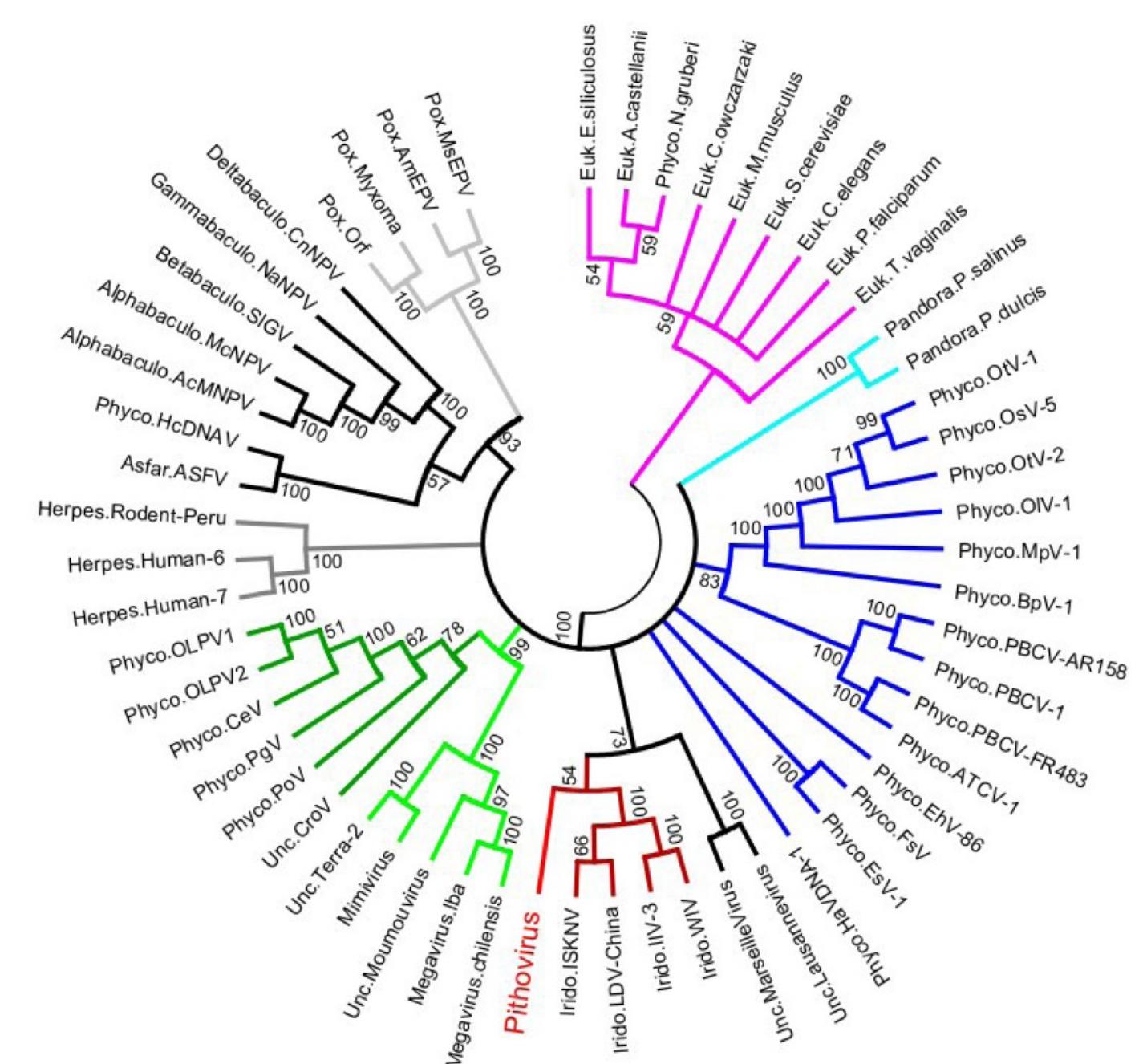
Is the Pandora's box a vehicle for viruses with genomes as diverse as those disseminated by icosahedral particles ?

# Pithoviridae today

Family	Prototype	Virion type	Dimension	Genome, size, GC%	Specific feature	Host
Pithoviridae		Amphora		C-P DNA	Variable size, cork	Acanthamoeba
	<b>P. sibericum</b>	Amphora	1500x500 nm	610 kb, 35.8%	One cork	Acanthamoeba
	<b>P. massiliensis</b>	Amphora	1500x500 nm	683 kb, 35.4%	One cork	Acanthamoeba
	<b>Cedratvirus A11</b>	Amphora	1000x500 nm	589 kb, 42.6%	Two corks	Acanthamoeba
	<b>C. lausannensis</b>	Amphora	1000x500 nm	575 kb, 42.8%	Two corks	Acanthamoeba
???????	Orpheovirus LCC2	Amphora	1000x500 nm	1.45 Mb, 25%	Pandora-like cork	Vermamoeba

# phylogenetic affinity for Irido/ Marseillevirus:

Genuine of  
the result of  
HGT ?



DNA pol B

# Two for the price of one: *Mollivirus sibericum*

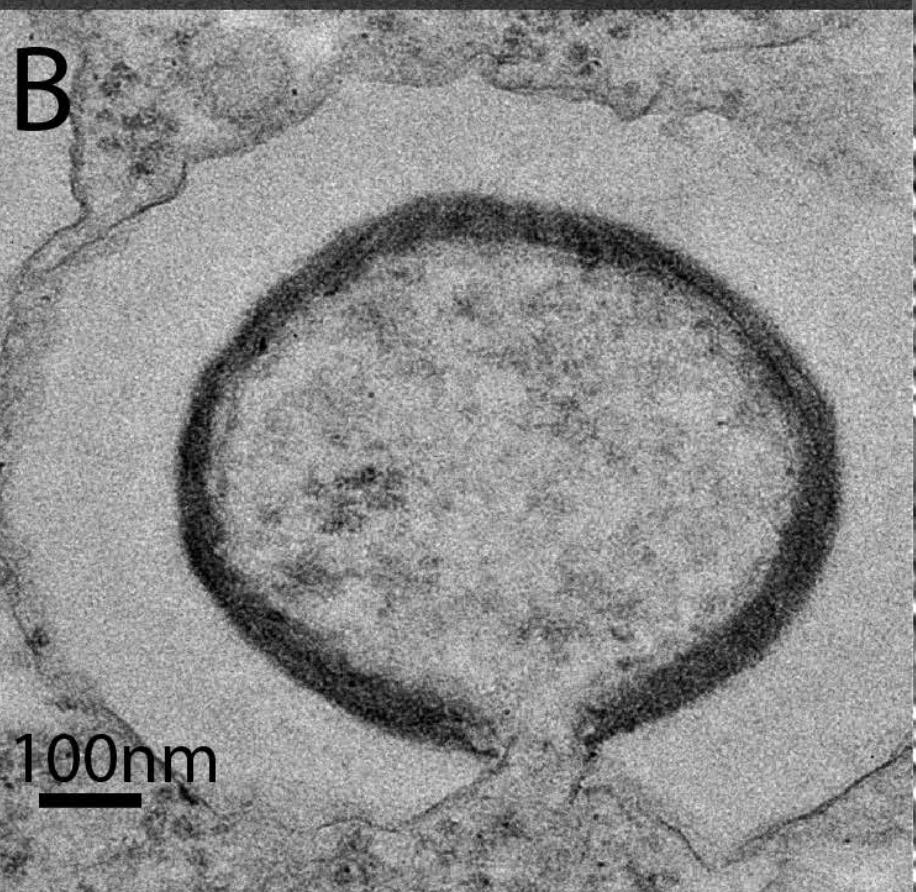
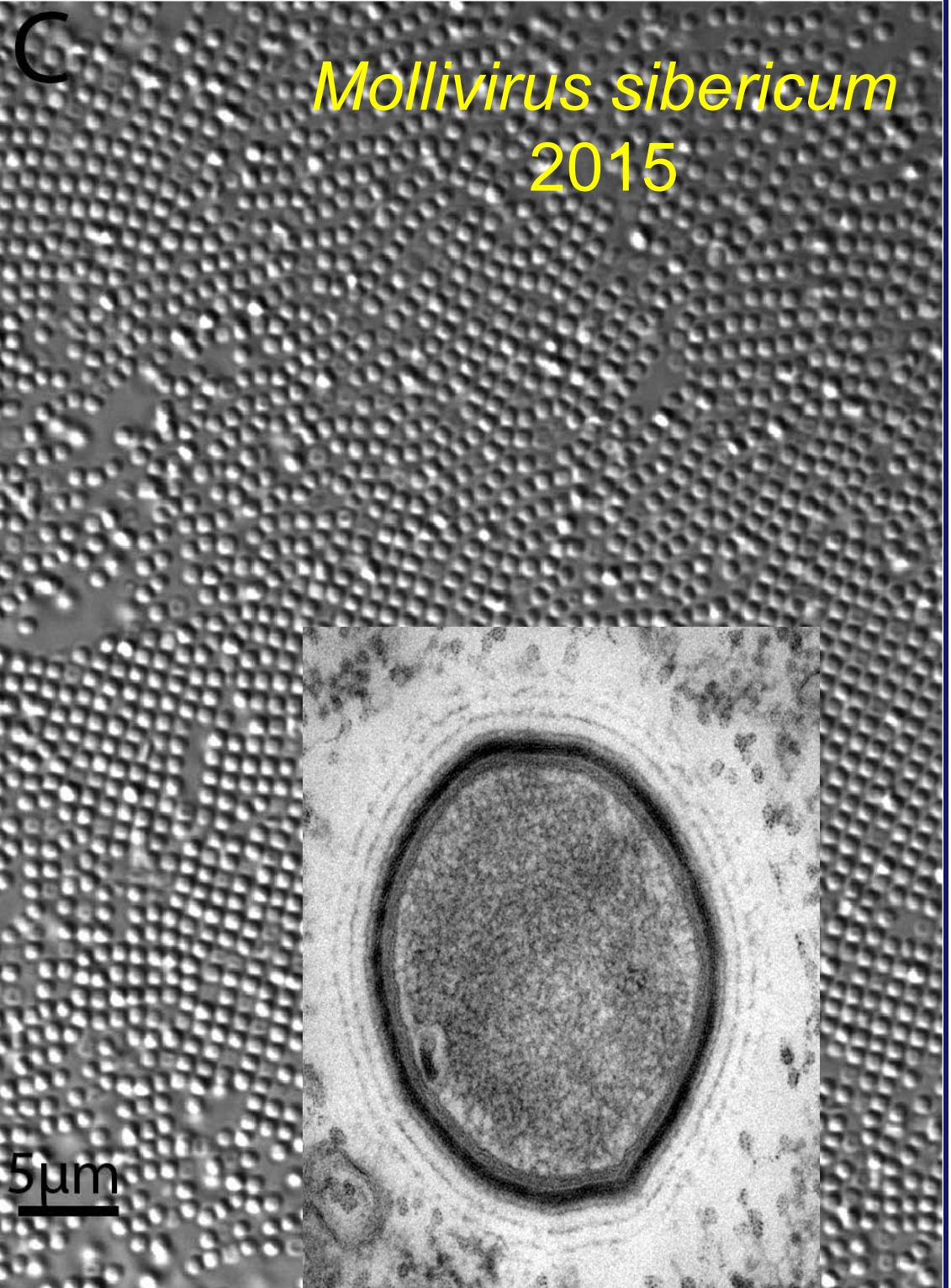
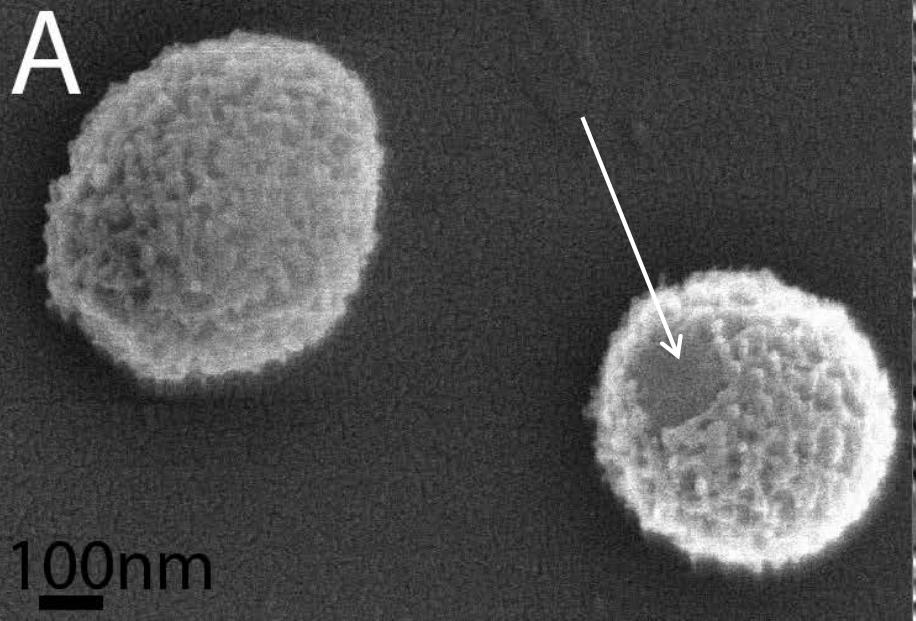


## In-depth study of *Mollivirus sibericum*, a new 30,000-year-old giant virus infecting *Acanthamoeba*

Matthieu Legendre<sup>a,1</sup>, Audrey Lartigue<sup>a,1</sup>, Lionel Bertaux<sup>a</sup>, Sandra Jeudy<sup>a</sup>, Julia Bartoli<sup>a,2</sup>, Magali Lescot<sup>a</sup>, Jean-Marie Alempic<sup>a</sup>, Claire Ramus<sup>b,c,d</sup>, Christophe Bruley<sup>b,c,d</sup>, Karine Labadie<sup>e</sup>, Lyubov Shmakova<sup>f</sup>, Elizaveta Rikina<sup>f</sup>, Yohann Couté<sup>b,c,d</sup>, Chantal Abergel<sup>a,3</sup>, and Jean-Michel Claverie<sup>a,g,3</sup>

- Virion: ~0.6 μm in diameter
- dsDNA genome, linear, 10-kb terminal inverted repeats
- 651,523 base pairs, 60% G+C
- 523 protein-coding genes
- Nucleo-cytoplasmic
- 6h replication cycle

The image shows a screenshot of a CNN news article. The headline reads "Ancient squirrel's nest leads to discovery of giant virus". The article is by Jen Christensen and was published on September 11, 2015. The CNN logo is visible in the top left corner, and there are social media sharing icons at the bottom right.



# 7 families of giant viruses infecting protists

Family	Virion type	Size (nm)	N isolates	Genome size	GC%	Life-style
Mimiviridae	icosahedral	755	>20	1.6Mb-370kb	25	cytoplasmic
Marseilleviridae <sup>1</sup>	icosahedral	200	>20	360kb-390kb	43	Nucleo-cytoplasmic
Pandoraviridae	Amphora	1000x500	6	2.8Mb-1.85Mb	61	Nuclear
Pithoviridae	Amphora	(1000-2000)x500	5	575kb-685kb	38	Cytoplasmic
Molliviridae	Spherical	600	1	650kb	60	Nuclear
Faustoviridae <sup>1</sup>	icosahederal	200-250	3	350kb-465kb	36	Nucleo-cytoplamsic
Medusaviridae <sup>2</sup>	icosahederal	200	1	380kb	62	?

1:

Boyer M, et al., Raoult D. (2009) Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. PNAS USA. 106 :21848-53.

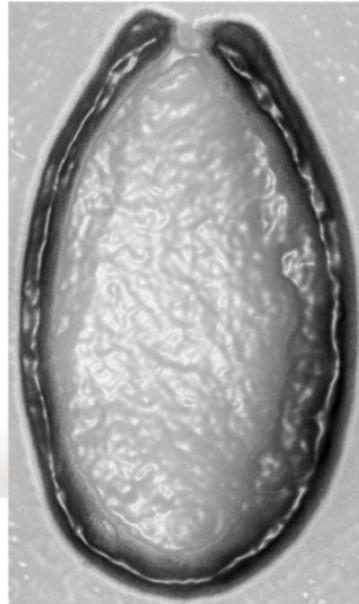
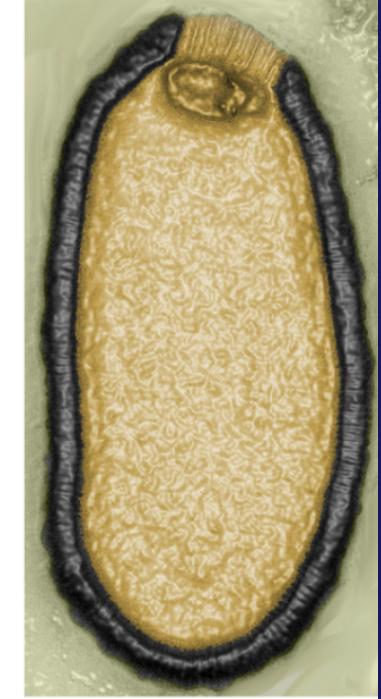
Reteno DG, et al., Raoult D, La Scola B. (2015) Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. J Virol. 89:6585-94.

2: Takemura et al. (Nov. 2017) (unpublished)

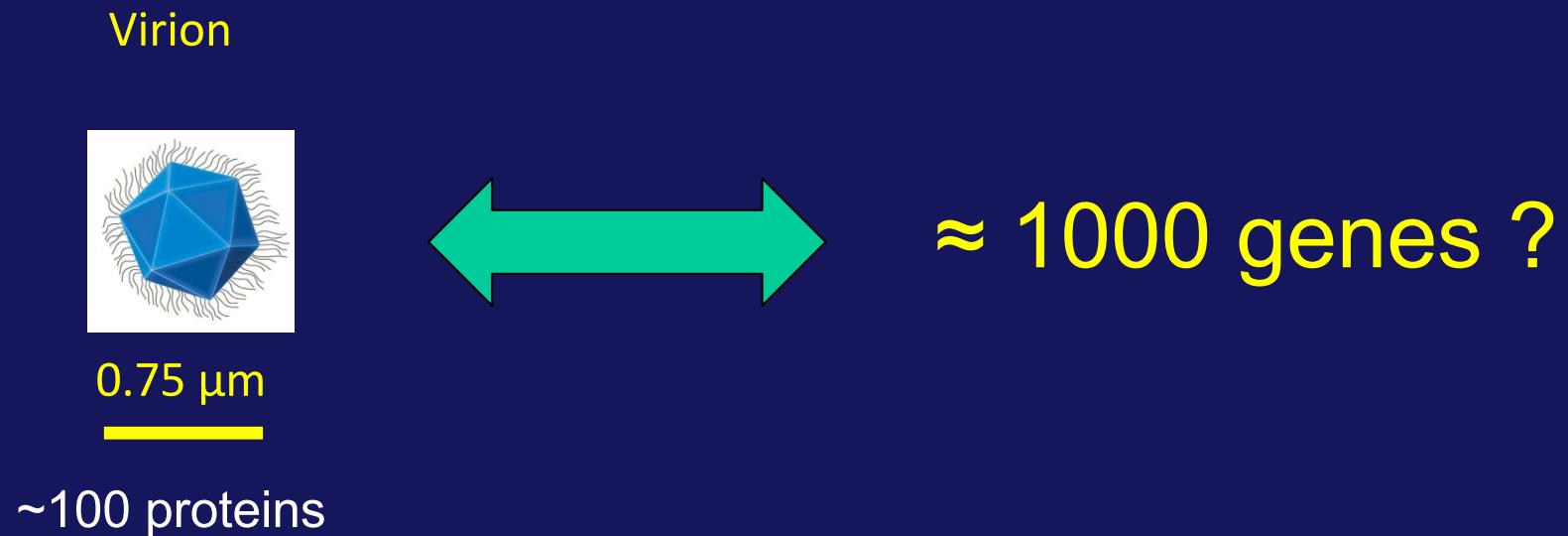
So, what are giant viruses changing  
about our previous conceptions  
of viruses?

“Viruses” (in general) can only be recognized by the way they propagate their genomes,

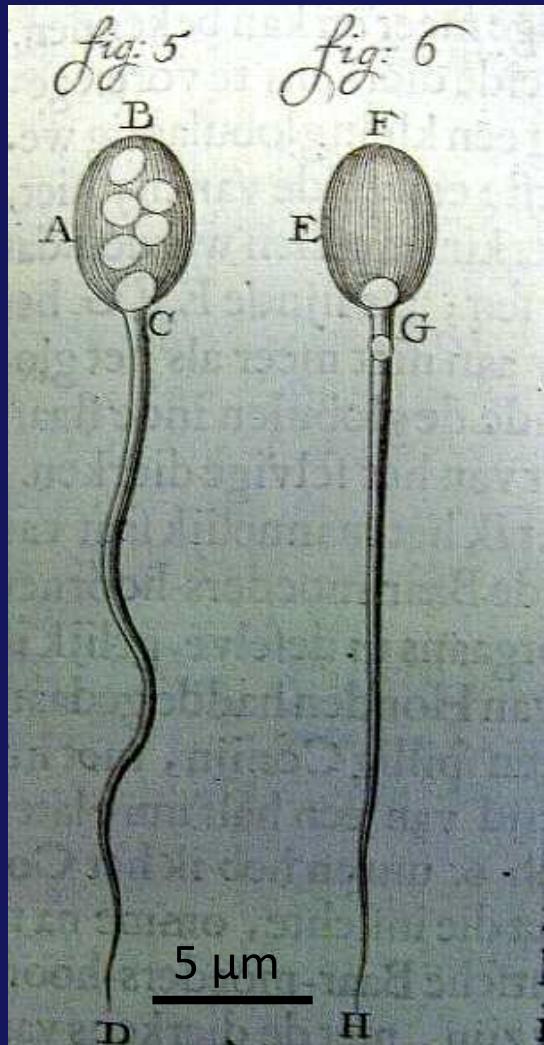
NOT by the structure of their particle  
Nor their gene content

		
<b><i>Megavirus chilensis</i></b>	<b><i>Pandoravirus salinus</i></b>	<b><i>Pithovirus sibericum</i></b>
	<b>Genome (bp)</b>	
1.26 million >70% A+T	2.5 million >60% G+C	600,000 64% A+T
	<b>Diameter</b>	
0.5 µm	0.5 µm	0.5
	<b>Length</b>	
Na	1 µm	1.5 µm

# The gene content is not commensurate to the virion (the « box ») complexity

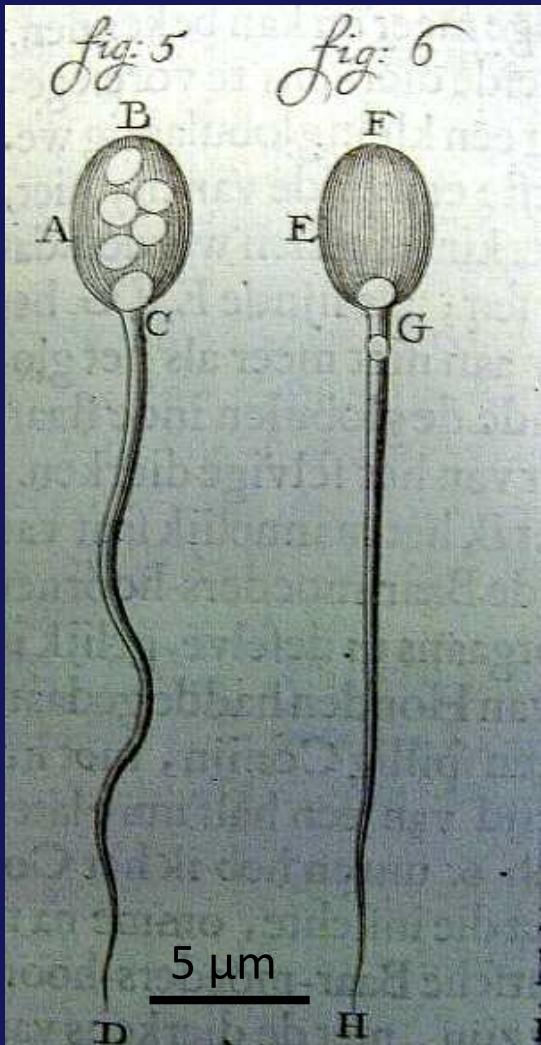


# A surprising genome size for a simple unicellular organism

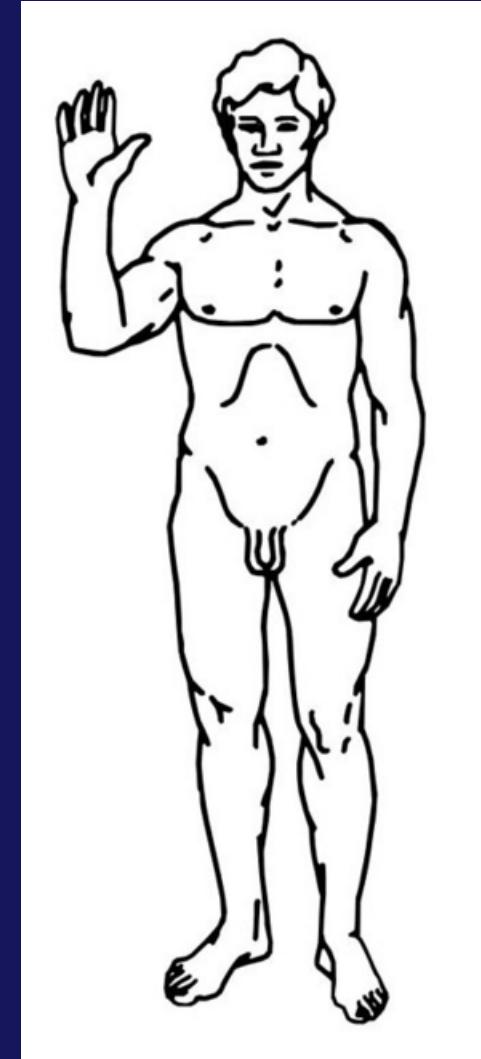


25.000 genes ?  
(3 Gb)

Two ways to pack and express the same  
3 Gb genome encoding 25,000 proteins.

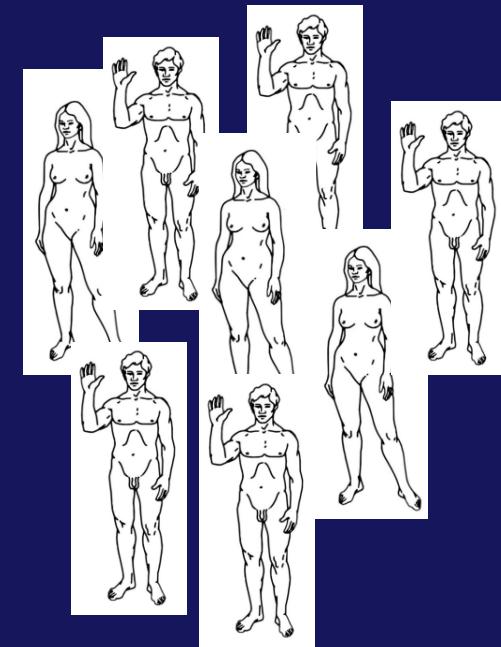
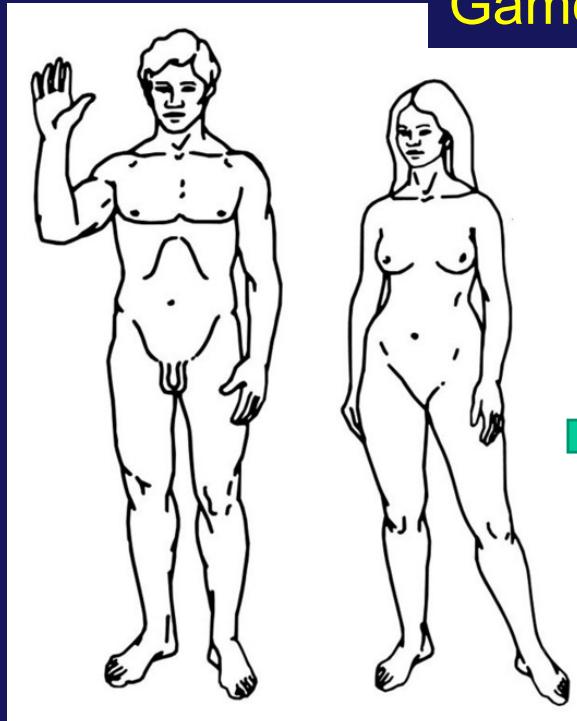


© Antoni Van Leeuwenhoek (~1670)

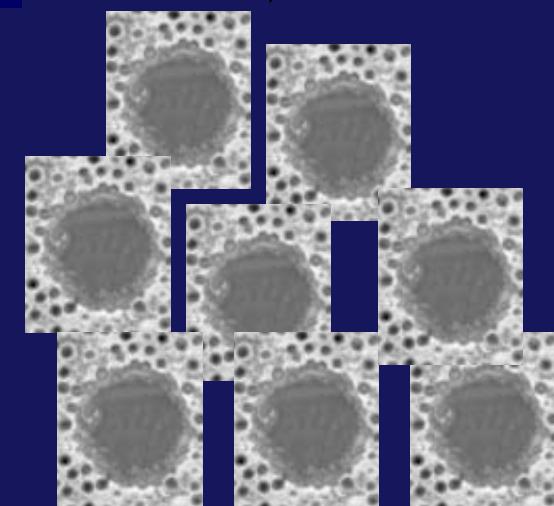
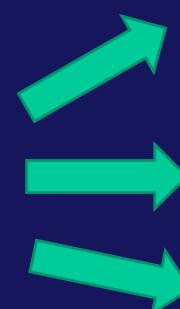
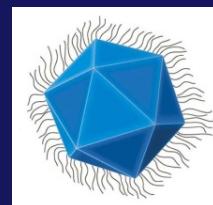
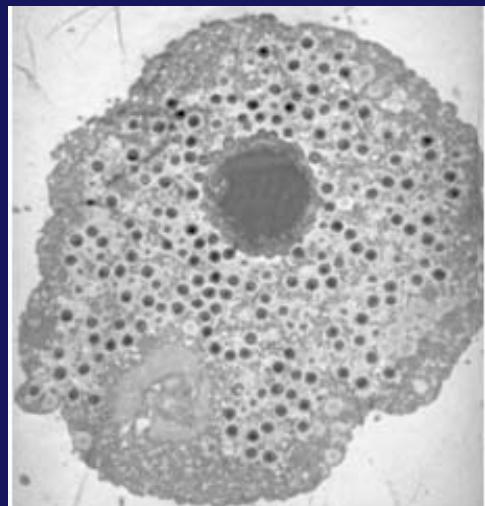


© Pioneer 10 (3 mars 1972)

## Gamete-based gene dissemination



*Developmental program*

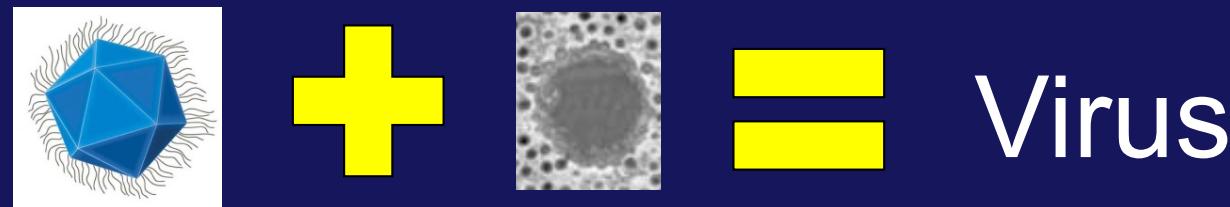


Transitory intracellular  
Parasitic microorganism

virion-based gene dissemination

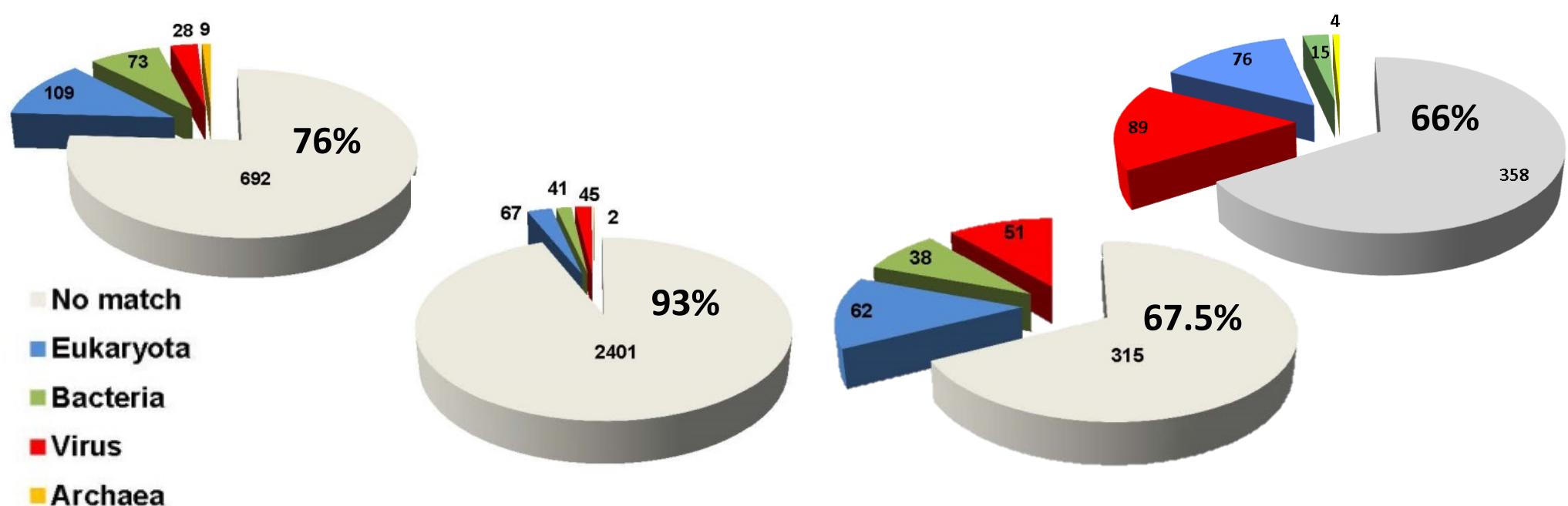
# Virus ≠ Virion

1. The virus is not the virion (the particle)
2. a « virus » is a process, not an object
3. A virus is « alive » during its intracellular phase



A viral genome is both the blueprint of the box +  
(partially) that of the factory needed to build it

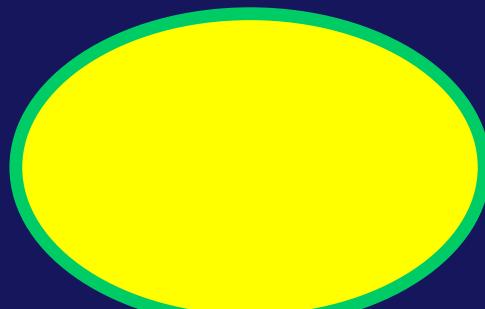
# Why so many “ORFans”?



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

# Hypotheses?

Mimiviridae



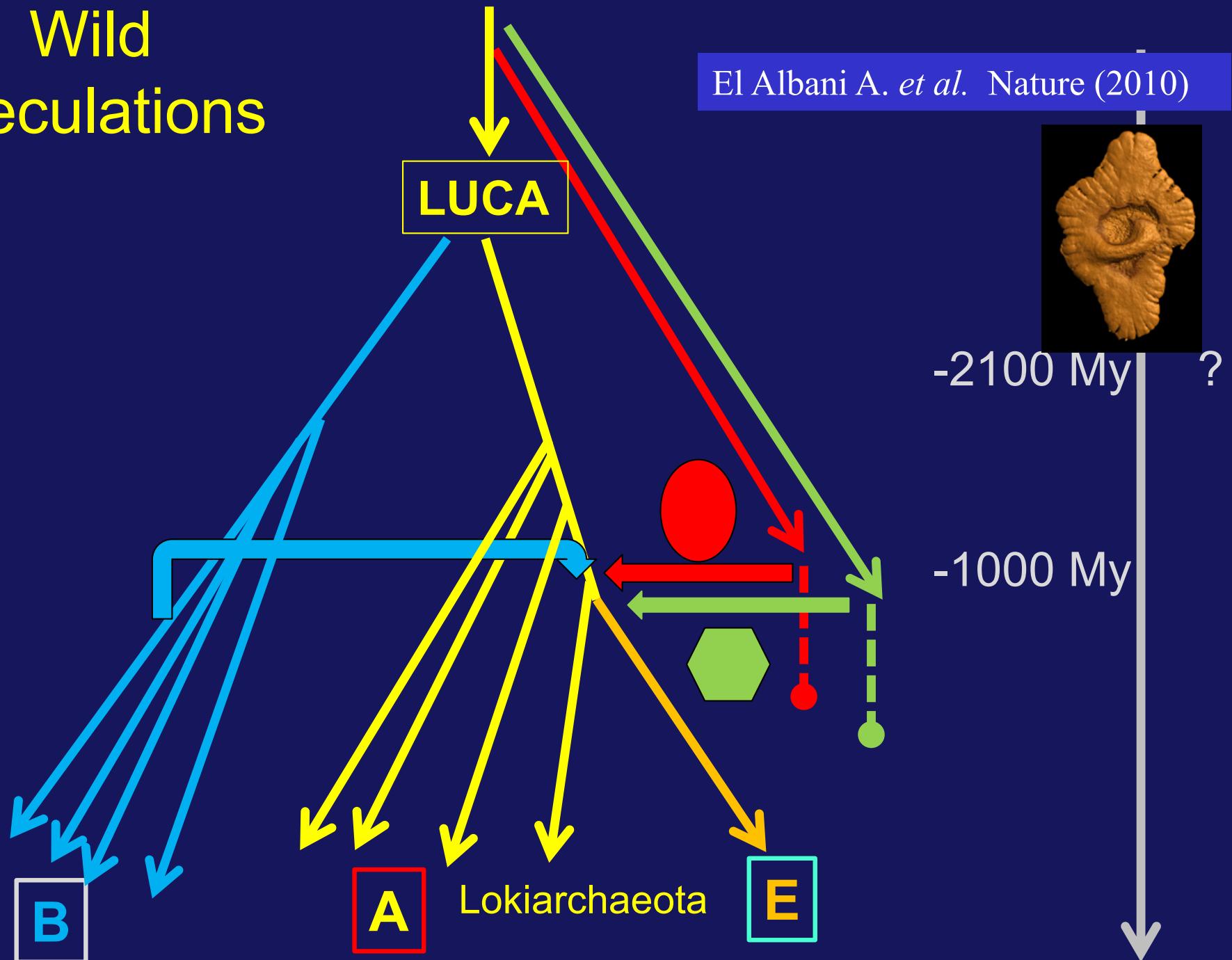
Pandoraviridae



De novo  
gene creation ?

# Wild speculations

El Albani A. et al. Nature (2010)



# A new discipline: “environmental virology”

- Traditionally, viruses were discovered and isolated in the context of diseases
- We can now expand our vision of the “virosphere” much beyond viruses that have a direct impact on human health and human activities.

Thus, more surprises are expected, that will hopefully help clarify the evolutionary position of viruses with respect to the 3 "classical" domains of Life, and maybe the origin of Life itself



**Direction IGS:** Jean-Michel Claverie, Chantal Abergel

**Wet Lab:** Nadège Philippe, Julia Bartoli, Sandra Jeudy, Audrey Lartigue, Virginie Seltzer, Lionel Bertaux, Defne Arslan, Gabriel Doutre

**Bioinfo:** Matthieu Legendre, Olivier Poirot, Magali Lescot, Sébastien Santini, Hiroyuki Ogata

**Proteomic:** IRSTV: Yohann Couté, Christophe Burley, Jérôme Garin

**Permafrost samples :** Lyubov Shmakova, Elizaveta Rivkina

**Microscopy:** A. Bernadac, J-P Chauvin, A Aouane

**Sequencing:** Karine Labadie, France Génomique ()

Parc des Calanques



**Thank you for  
your attention.  
Anything else  
you would like  
to know?**



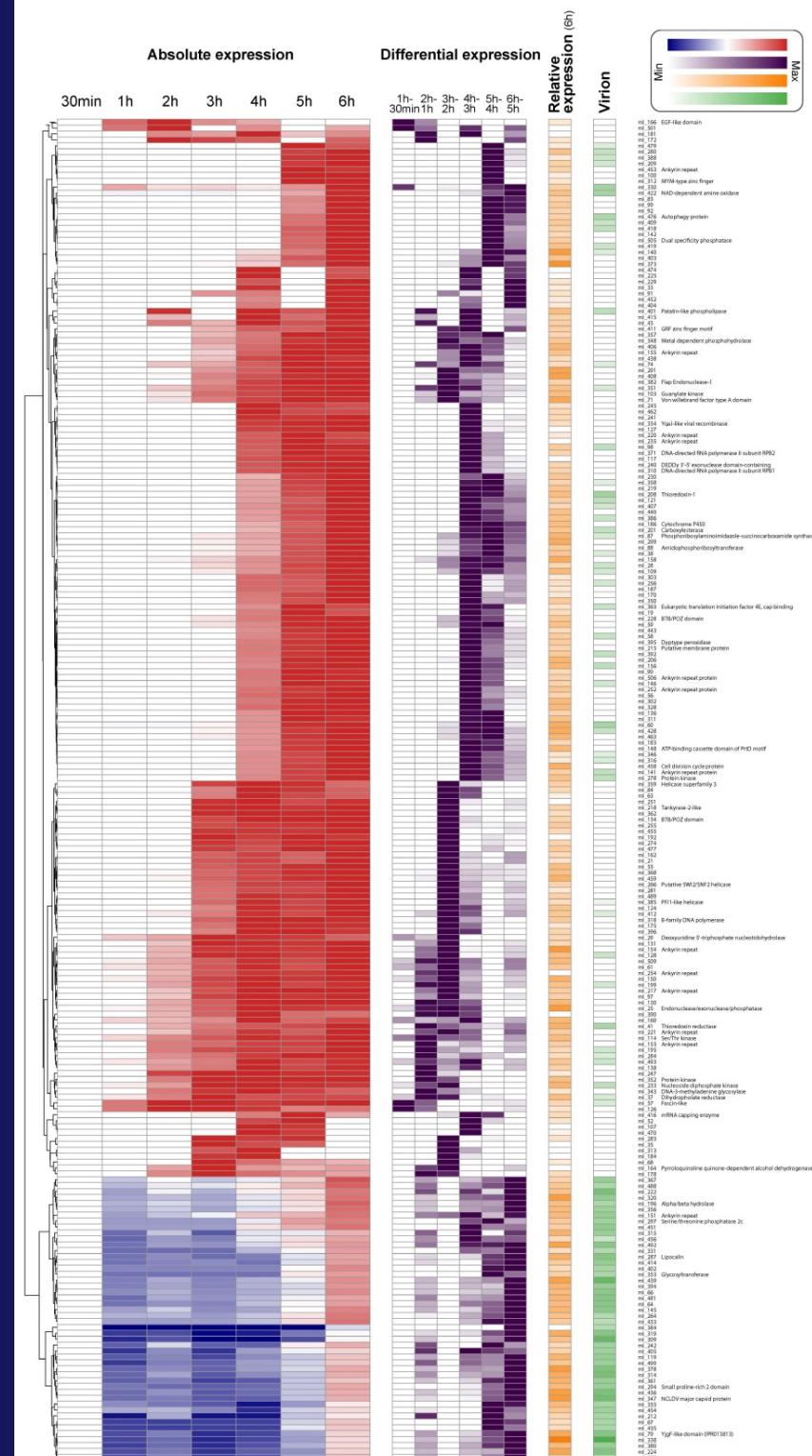
**Dr. Chantal Abergel**

# Our definition

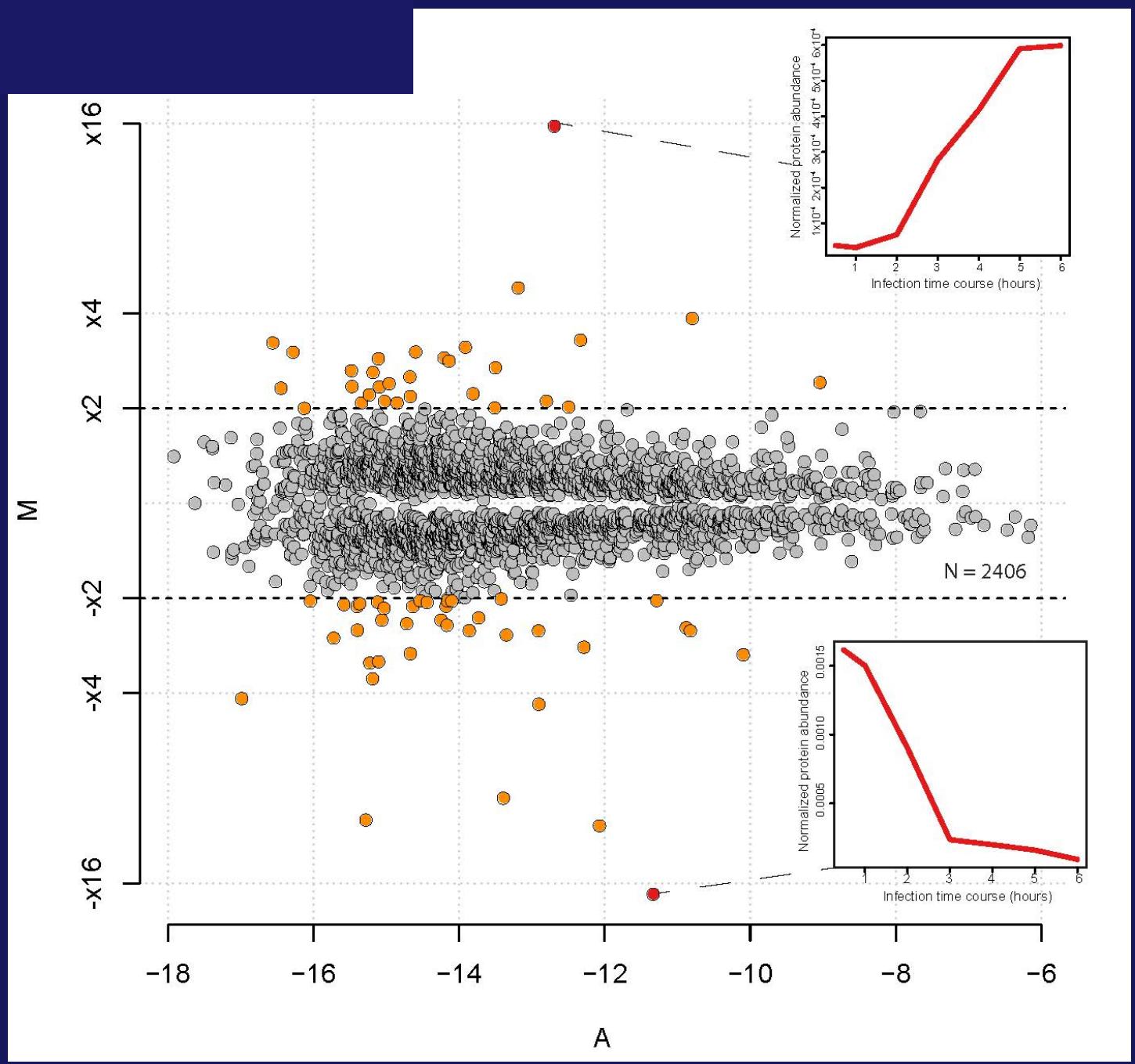
A virus is a biological entity the genome of which is

- 1) replicated by a macromolecular system that it does not entirely encode (absolute parasitism)
- 2) disseminated using a specialized structure the maintenance of which does not require energy.

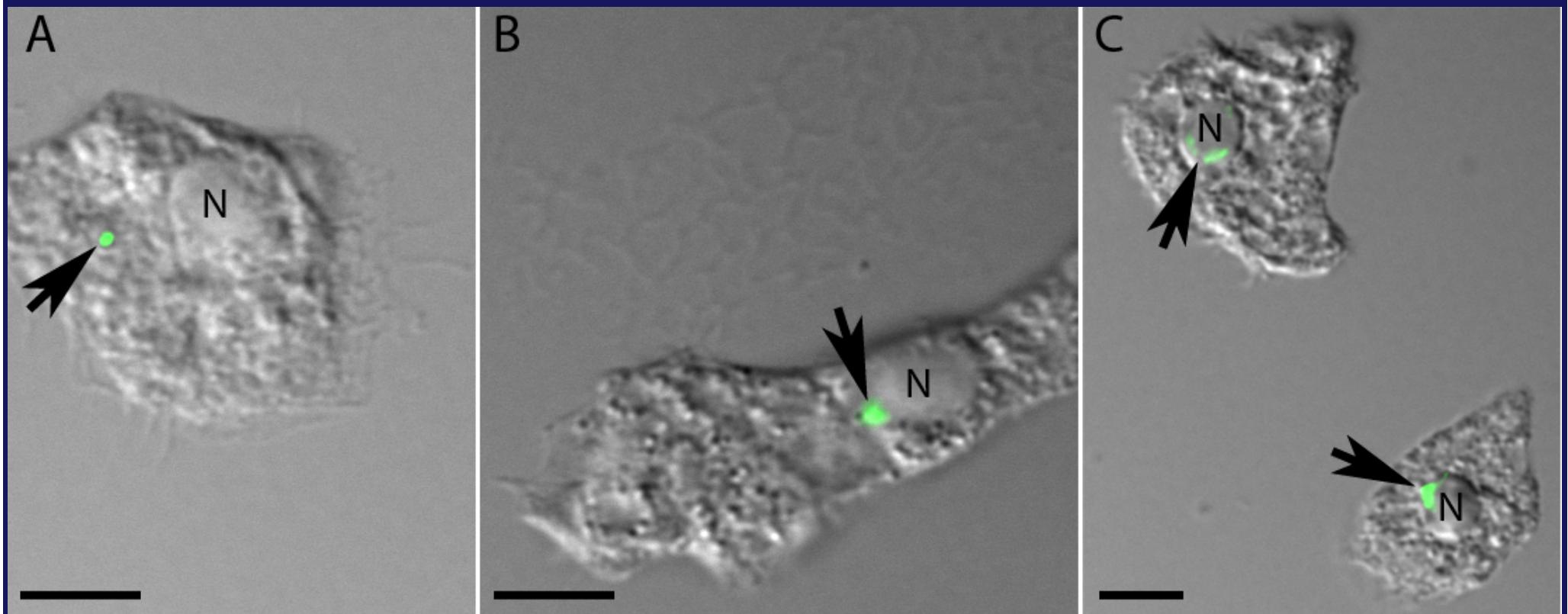
# Mollivirus: Proteomic time-course

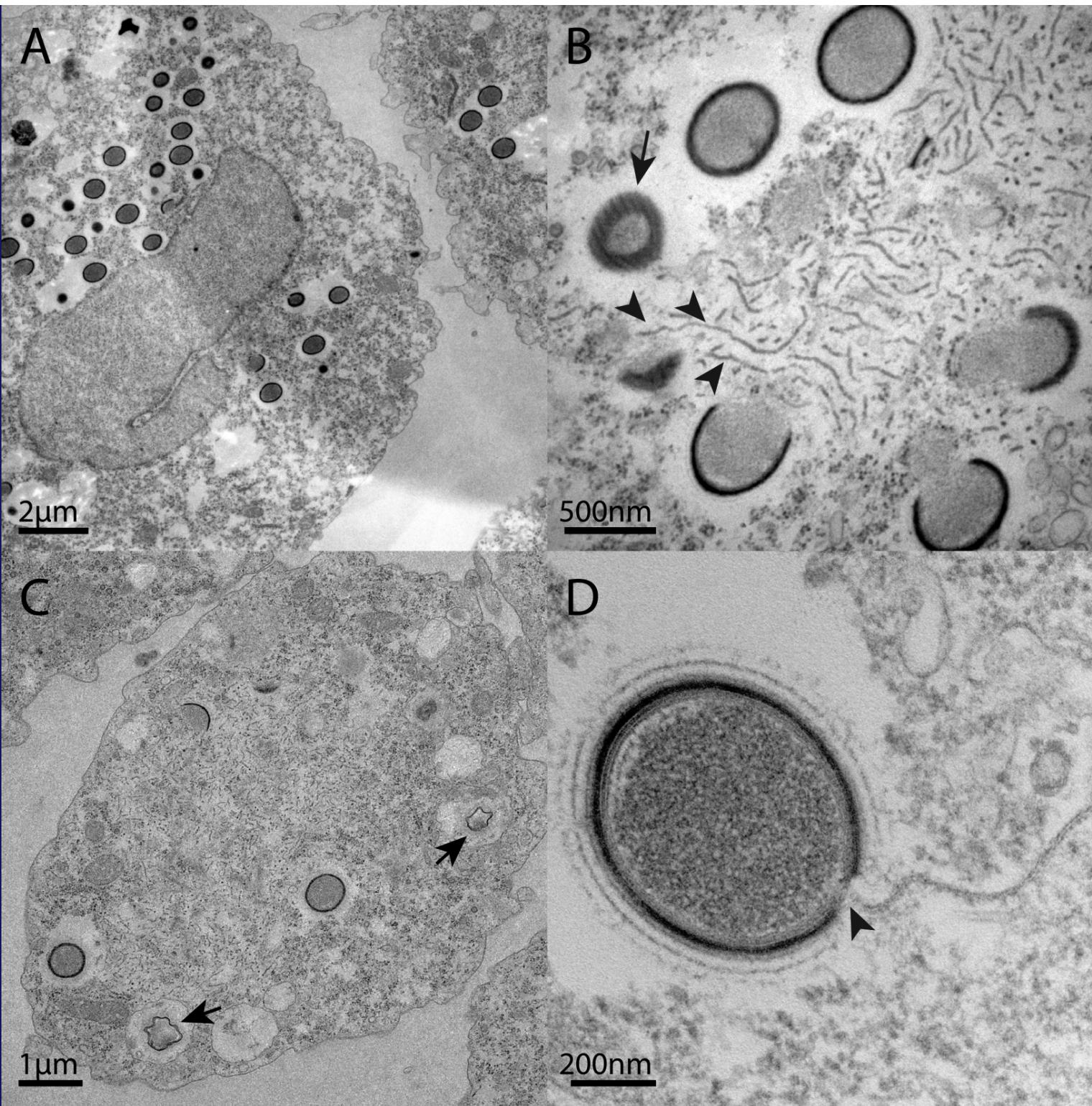


# Mollivirus proteomic time-course: 30 proteins up 38 proteins down

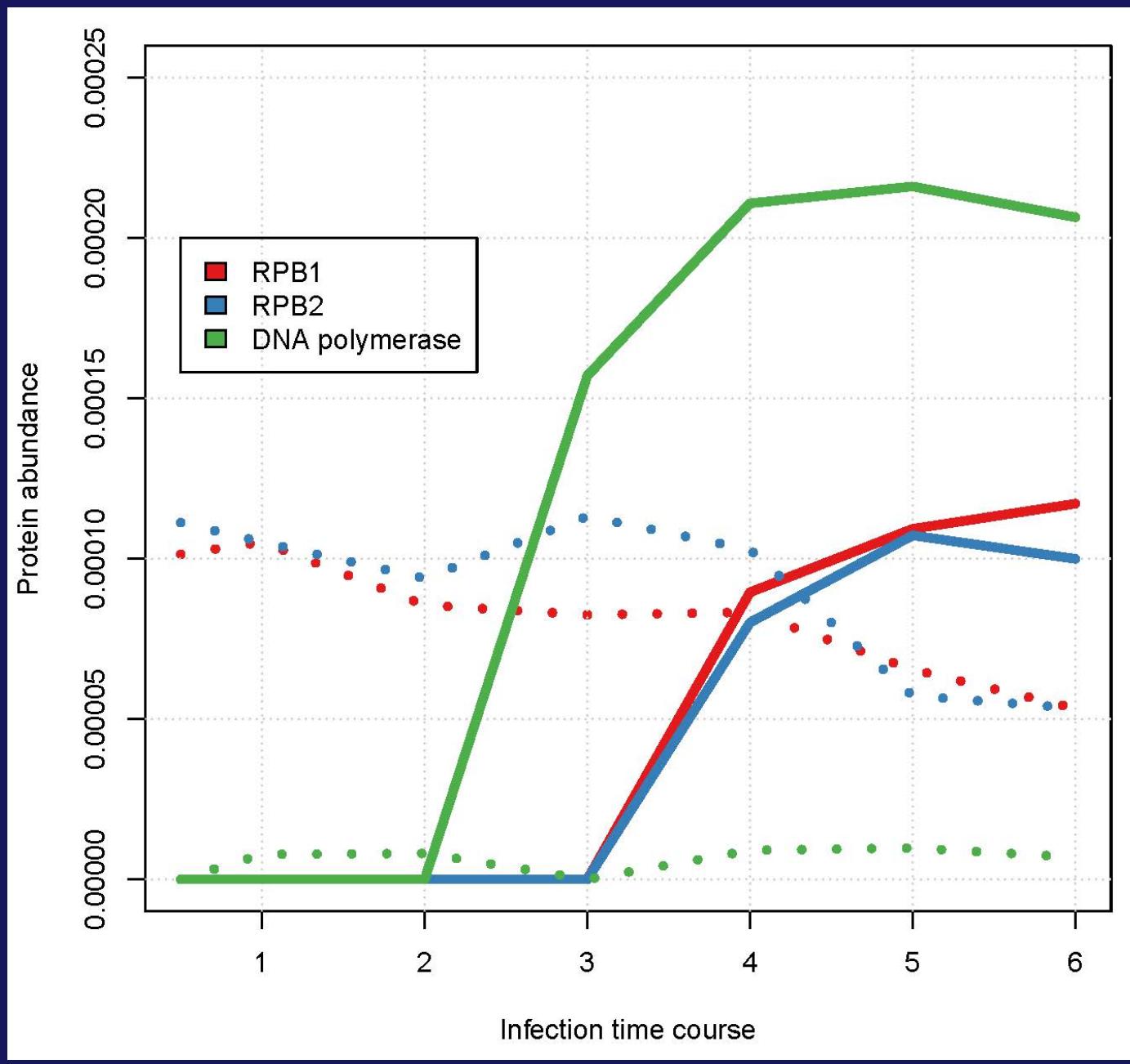


## Mollivirus EdU-labeled DNA visualized in infected *A. castellanii*

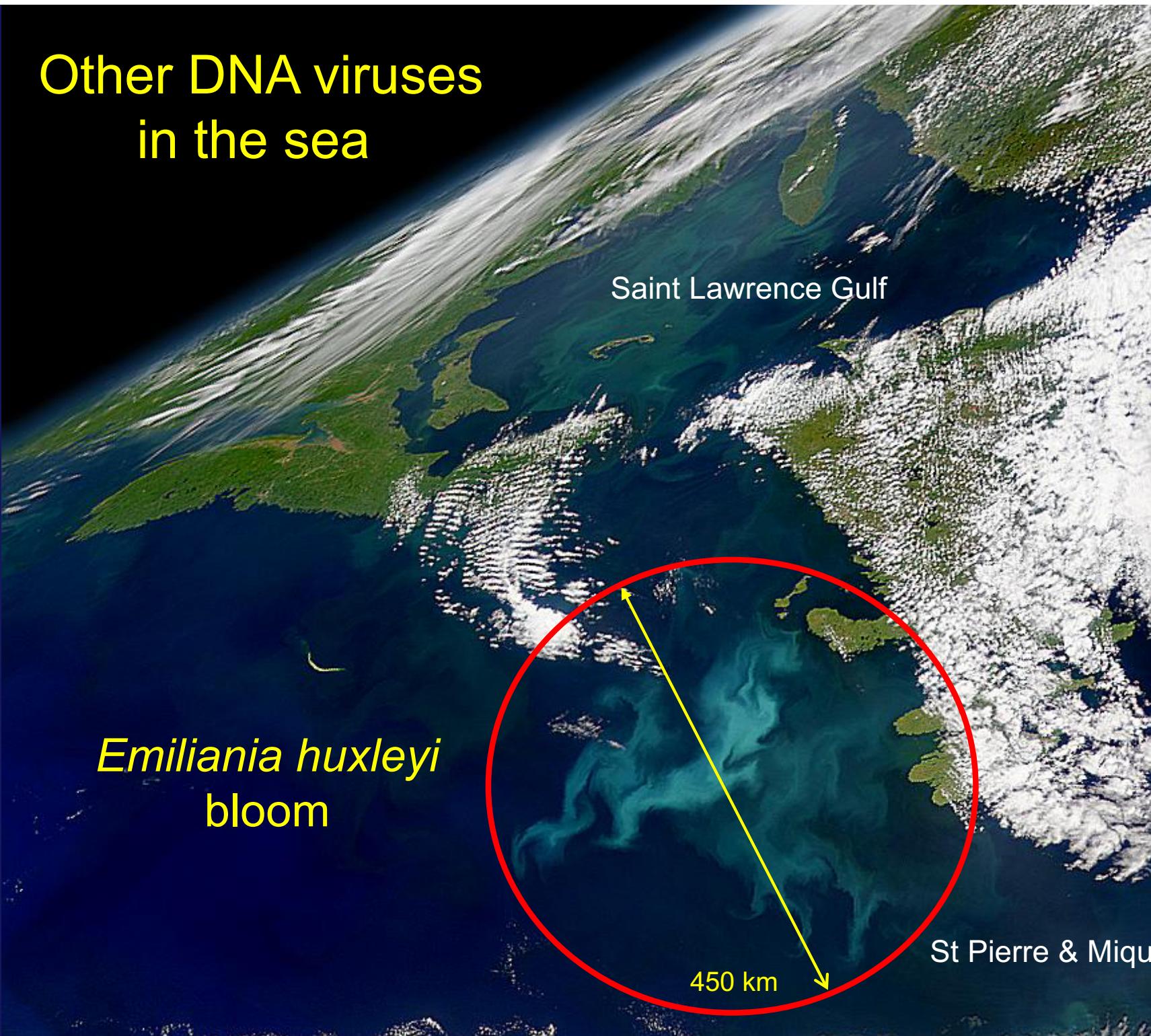


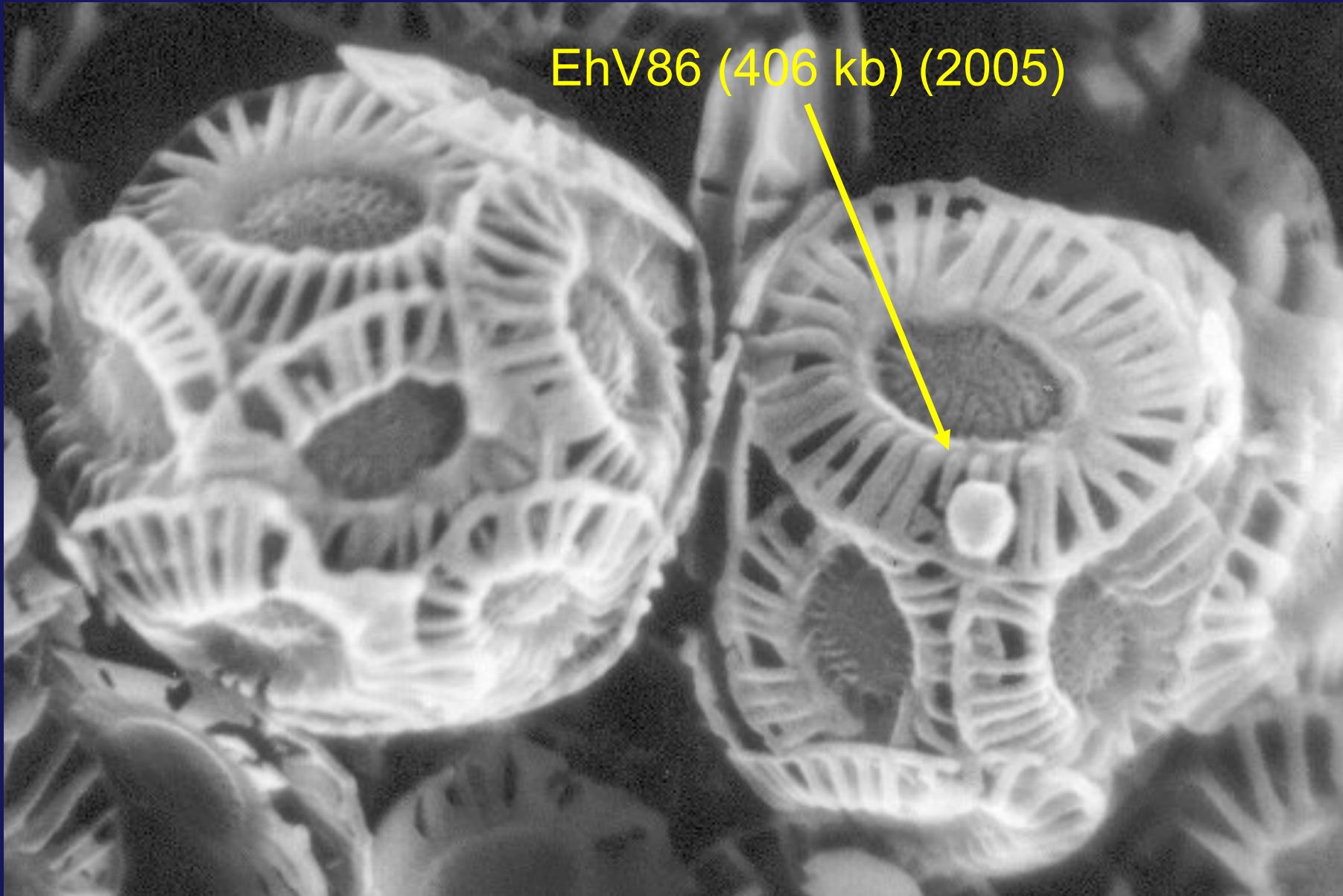


# Mollivirus: DNA pol B synthetized before RNA pol



# Other DNA viruses in the sea

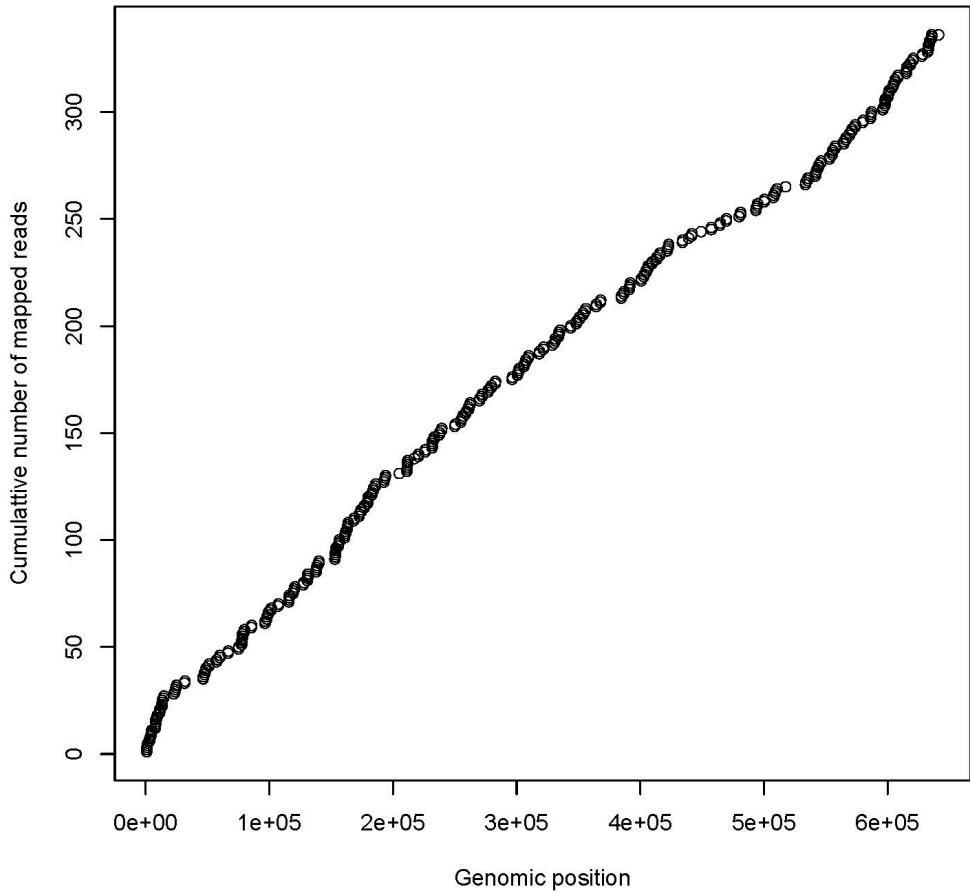




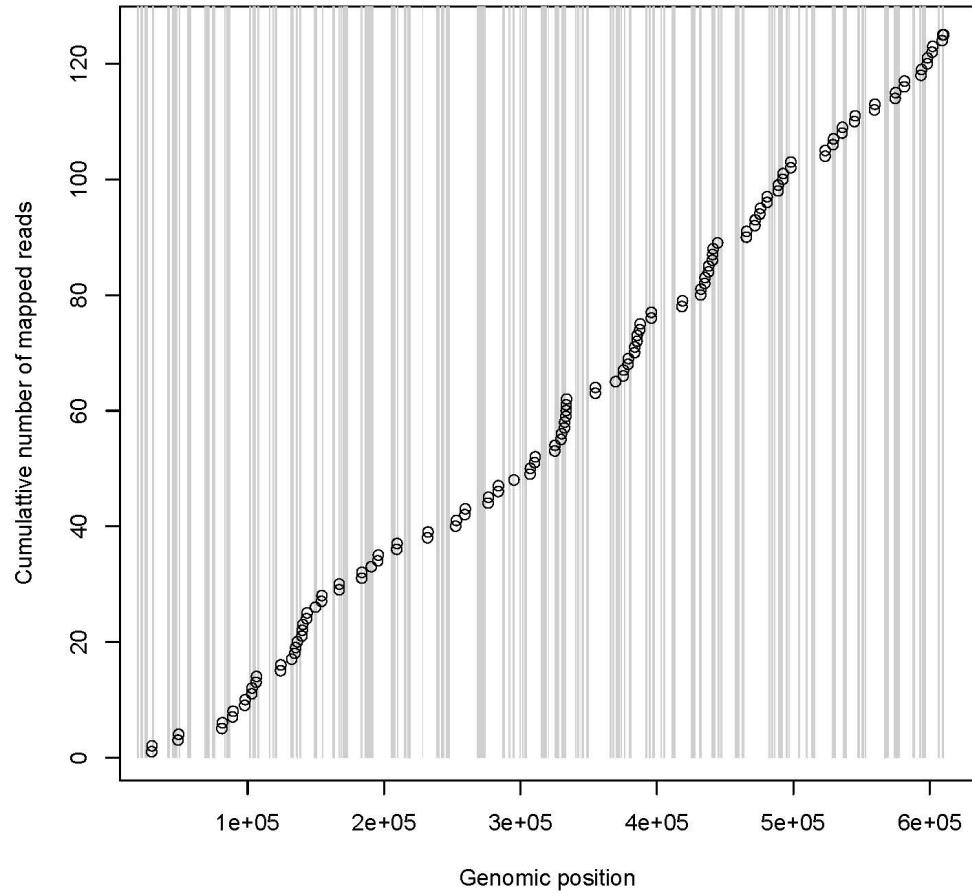
Credit: Dr. Declan Schroeder, Dr. William Wilson, The Plymouth Marine Laboratory, UK

# Original sample metagenomics

Mollivirus



Pithovirus



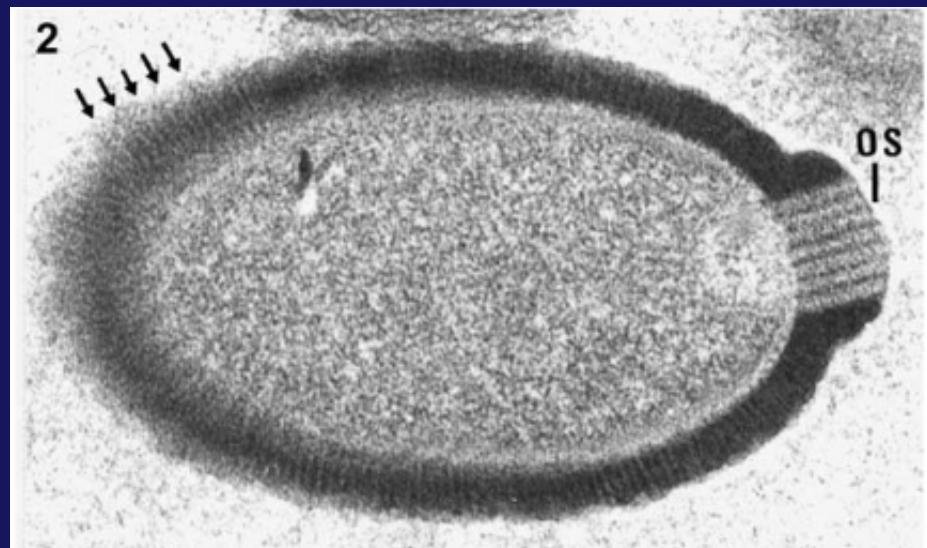
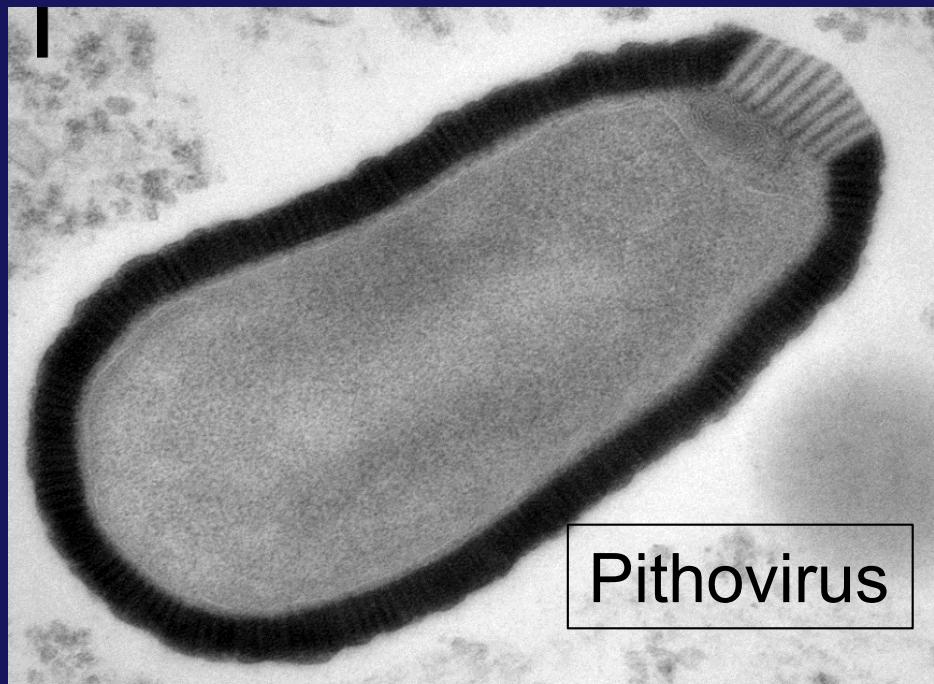
4.8% coverage

2 % coverage

ORIGINAL PAPER

R. Michel · E. N. Schmid · R. Hoffmann · K.-D. Müller

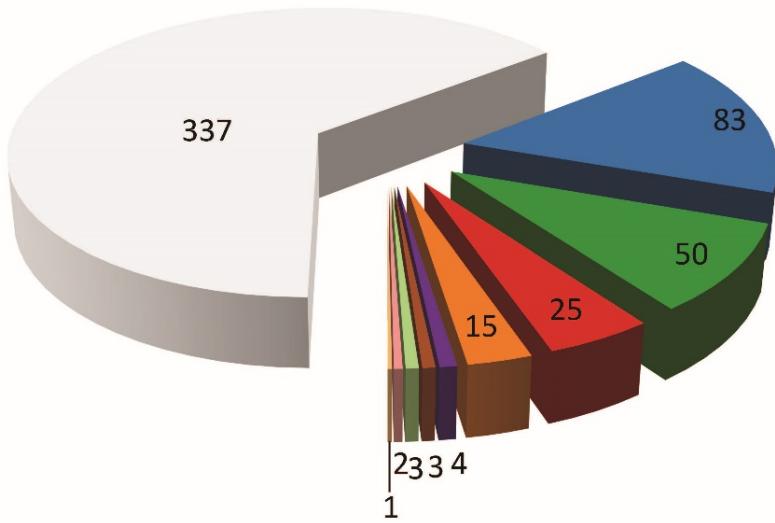
**Endoparasite KC5/2 encloses large areas of sol-like cytoplasm within *Acanthamoebae*. Normal behavior or aberration?**



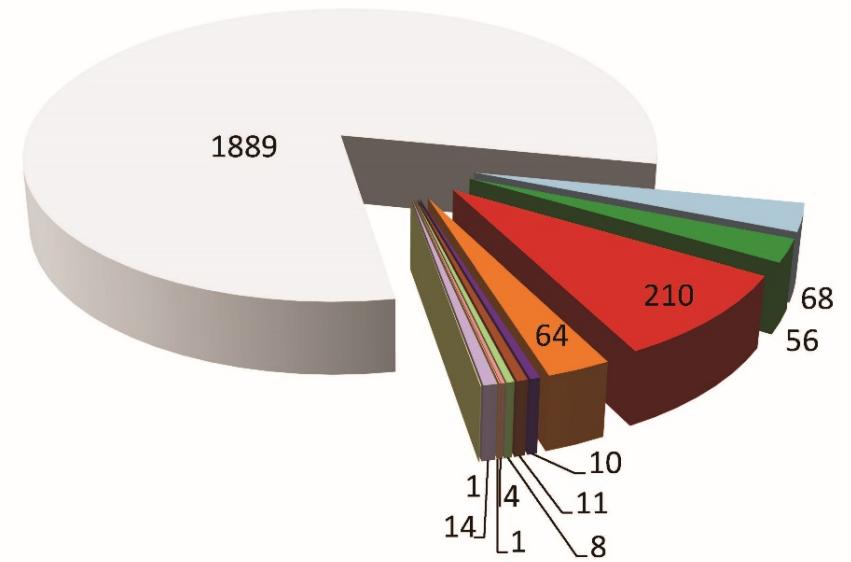
*Pandoravirus inopinatum*  
NC\_026440: 2,243,109 bp  
1839 proteins  
Scheid, Zoeller, Woelfel, Stoecker.,  
Georgi & Antwerpen

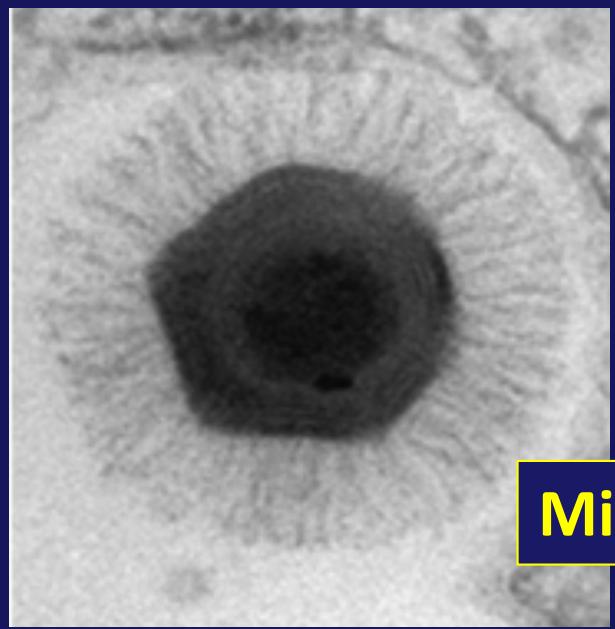
# Mollivirus might be particularly prone to HGT

**Mollivirus**



**Pandoravirus**



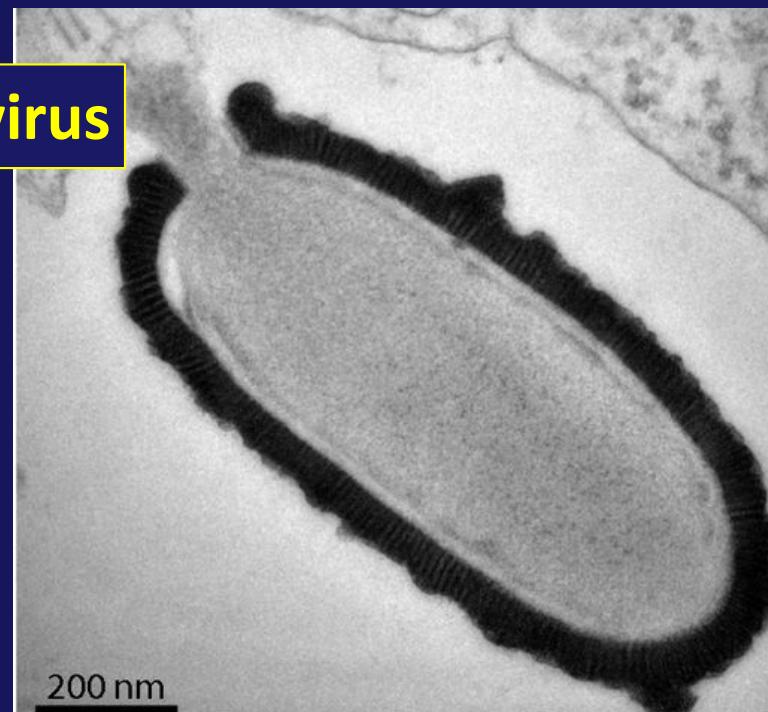


Mimivirus

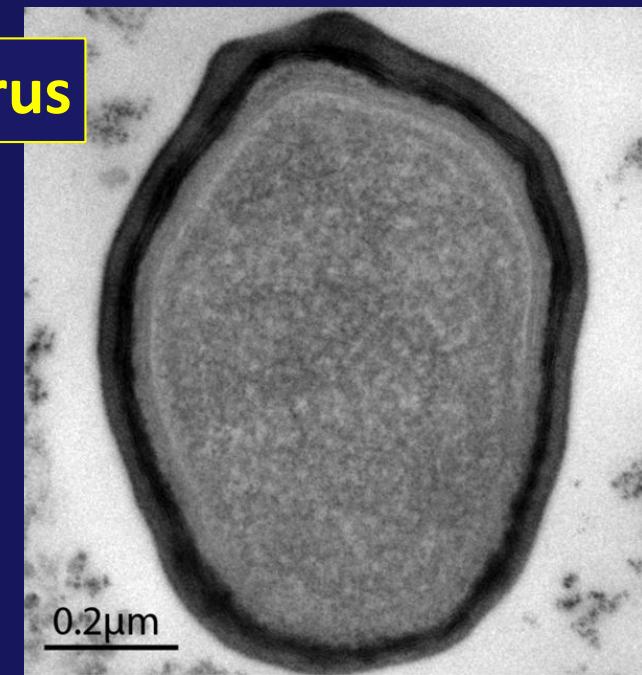


Mollivirus

Pithovirus



Pandoravirus



0.2 μm

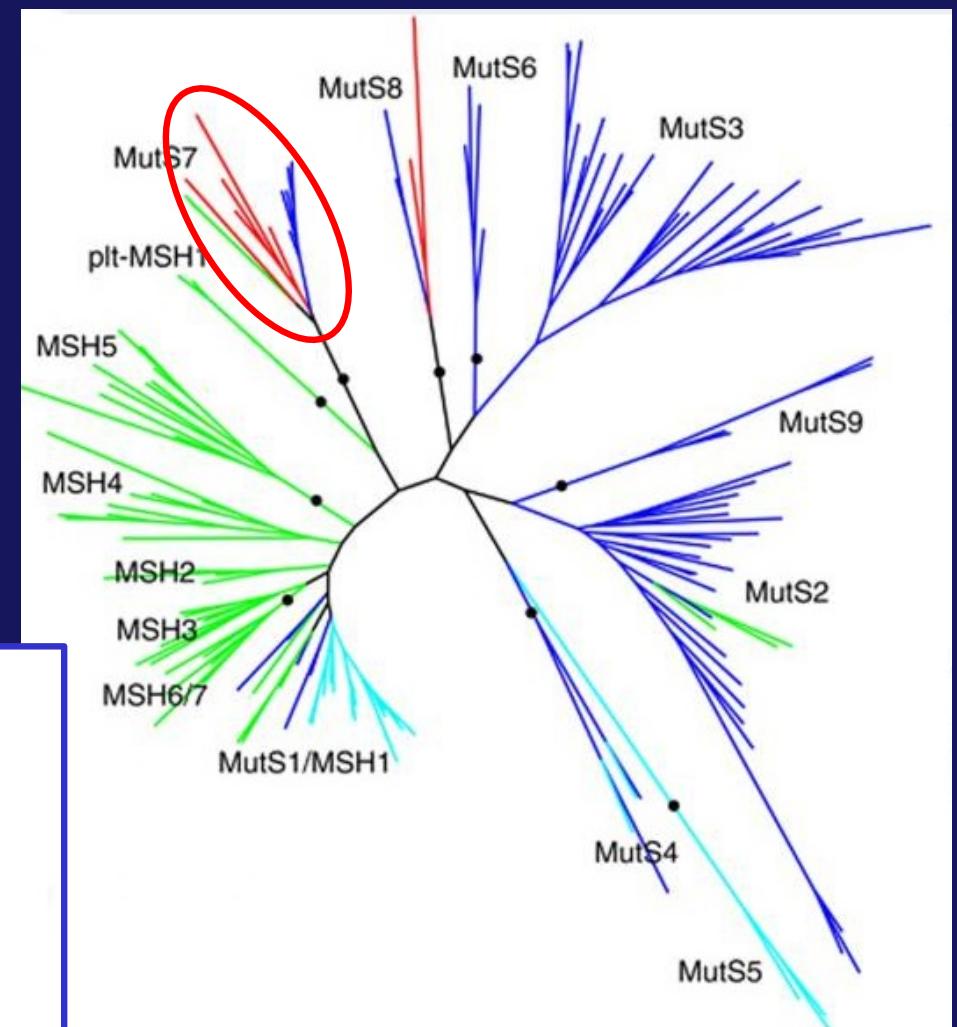
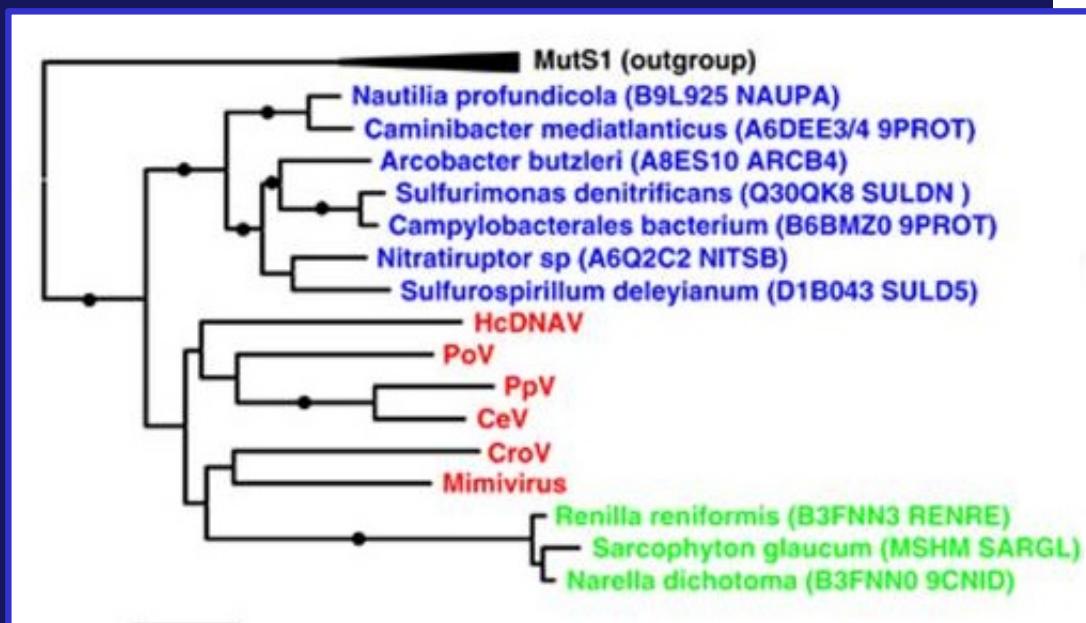
# Mimiviridae: MutS7 ?

(ISME J, 2011)

$\varepsilon$ -proteobacteria  $\leftrightarrow$  Mimivirus

Octocorals

(prior to radiation but after  
branching off from hexacorals)



} Mitochondria