ARCHAEA

Laboratoire de Biologie Moléculaire du Gène chez les Extremophiles

Institut de Génétique et Microbiologie

INSTITUT PASTEUR

Patrick Forterre

forterre@pasteur.fr

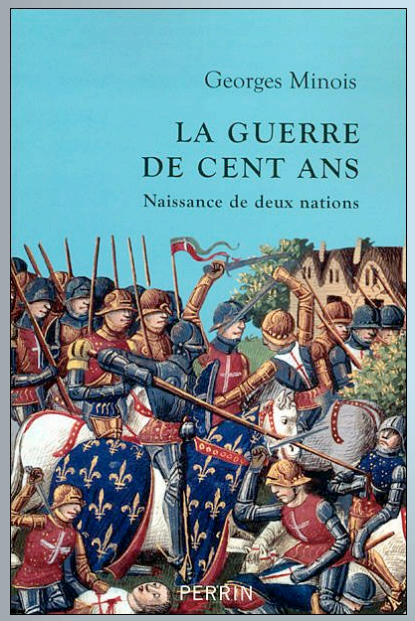
Molecular Biology of hyperthermophilic archaea
DNA topoisomerases
DNA replication, DNA repair

Phylogenomics
Genome evolution
Phylogeny of Archaea
Detection of viruses in archaeal and bacterial genomes

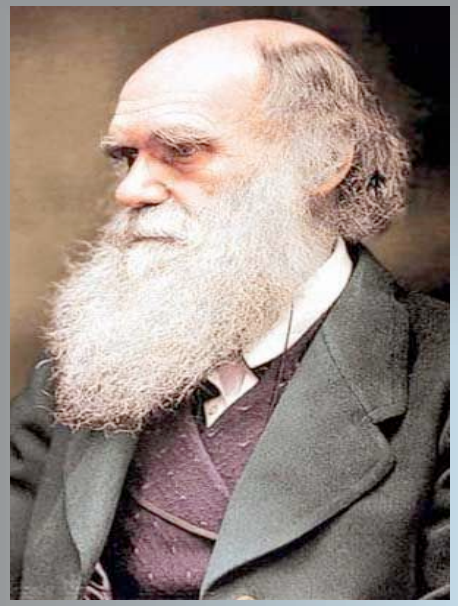
Viruses and plasmids from hyperthermophilic archaea

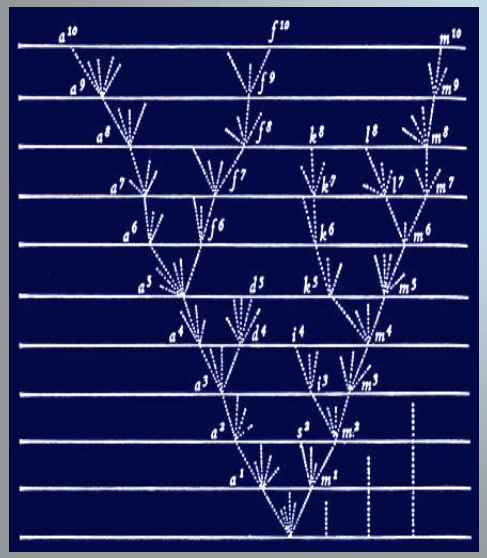
David Prangishvili

The three (at least) billions years war between ribosomes and capsids encoding organisms (cells and viruses)



Patrick Forterre
Institut Pasteur
University Paris-Sud



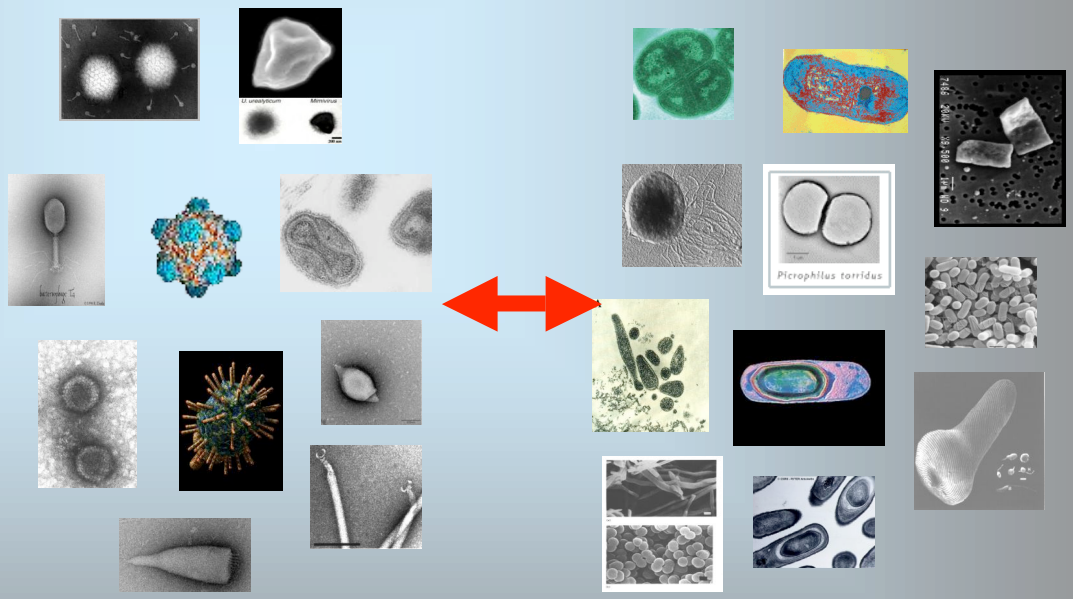


Thesis: the conflict between cellular and viral organisms as the major engine of biological evolution

Evolution: variation + natural selection

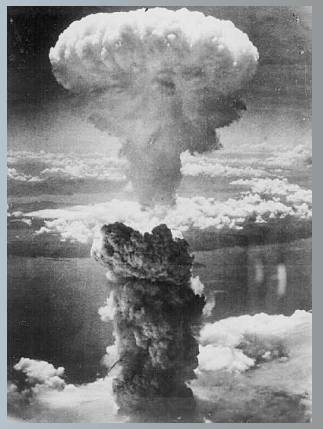
Virus

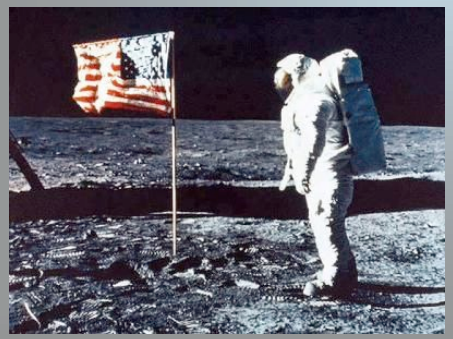
Cells



mutations + survive to the virus (or to the cell counter-attack)

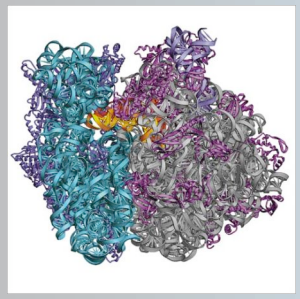
Scientists have underestimated the role played by the cells/viruses war in biological evolution because most of them despite war!!

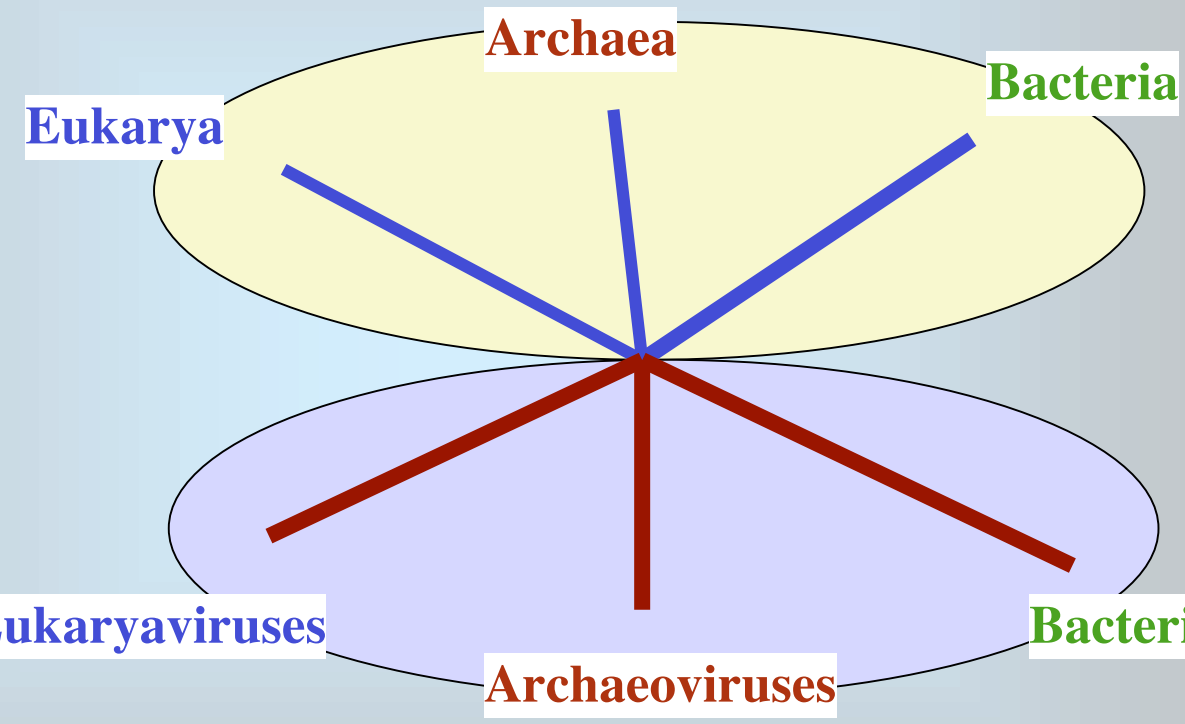


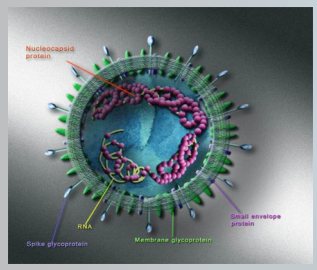


However, we know that wars have been the source of great progress (not politically correct!)

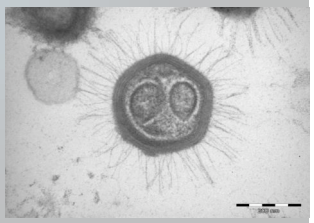
The living world can be divided in two major categories: ribosomes and capsids encoding organisms (cells and viruses)







Didier Raoult et Patrick Forterre
Nature Reviews Microbiol., 6 :315-319, 2008

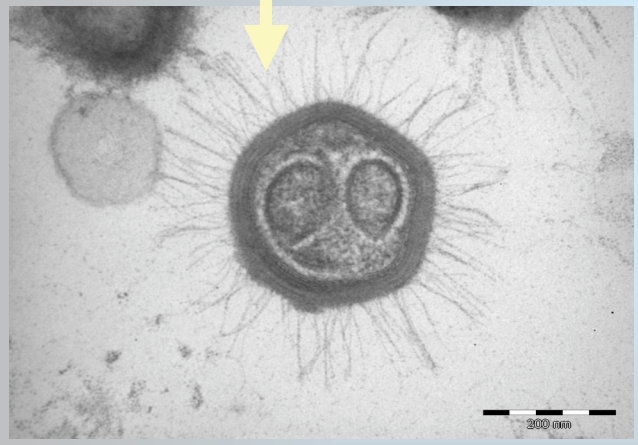


Assumption n°1 : viruses are living organisms

The viral organism correspond to the viral factory

Jean-Michel Claverie, Genome Biology, 2006

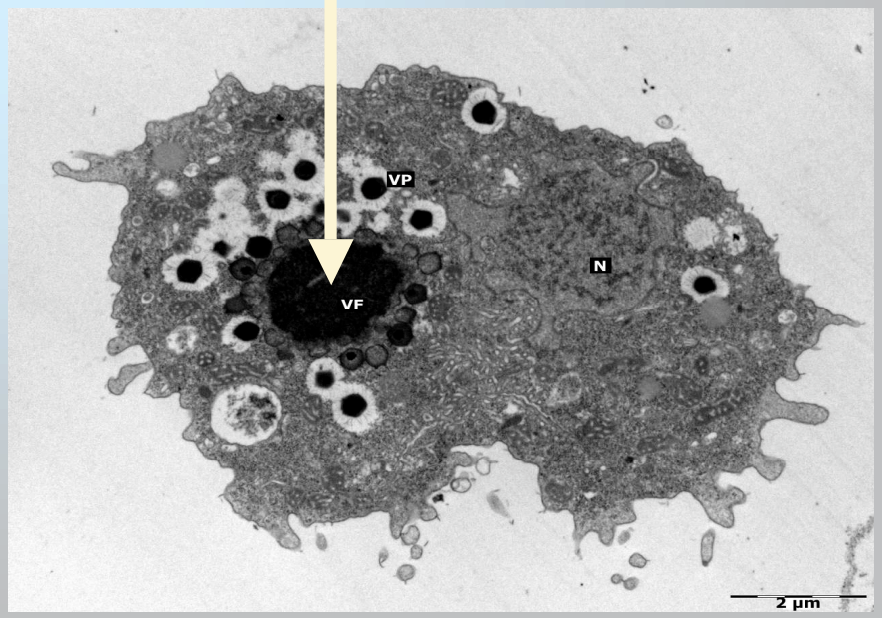
This is not a virus

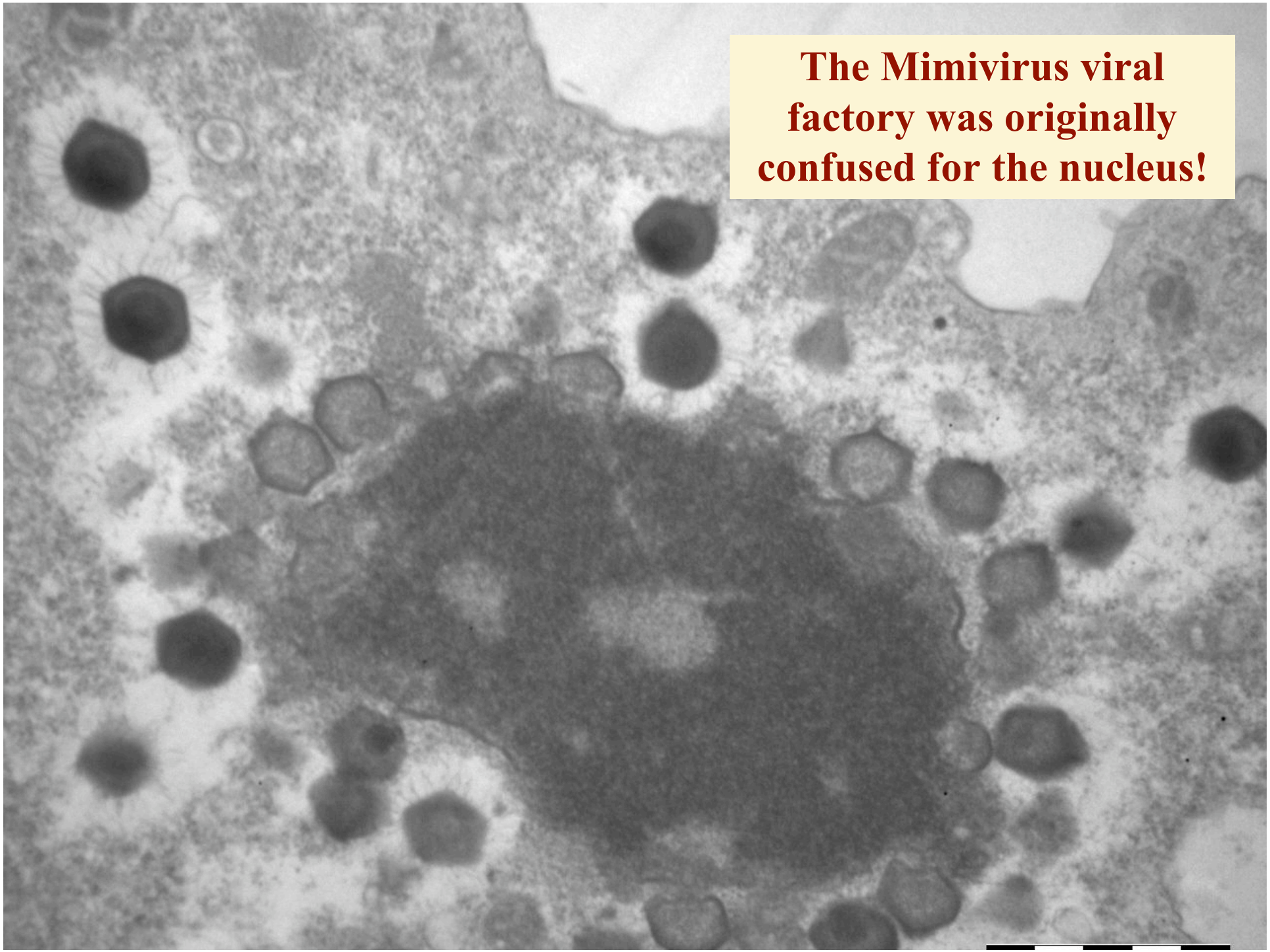


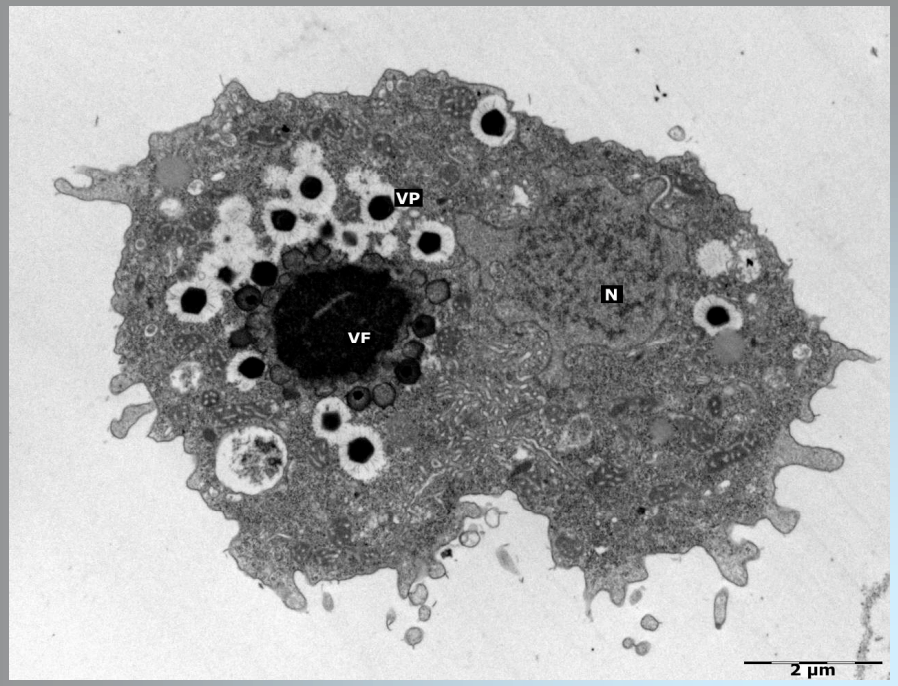
The mimivirus
A genome of 1.2 Mb

This is not a man

This is a virus

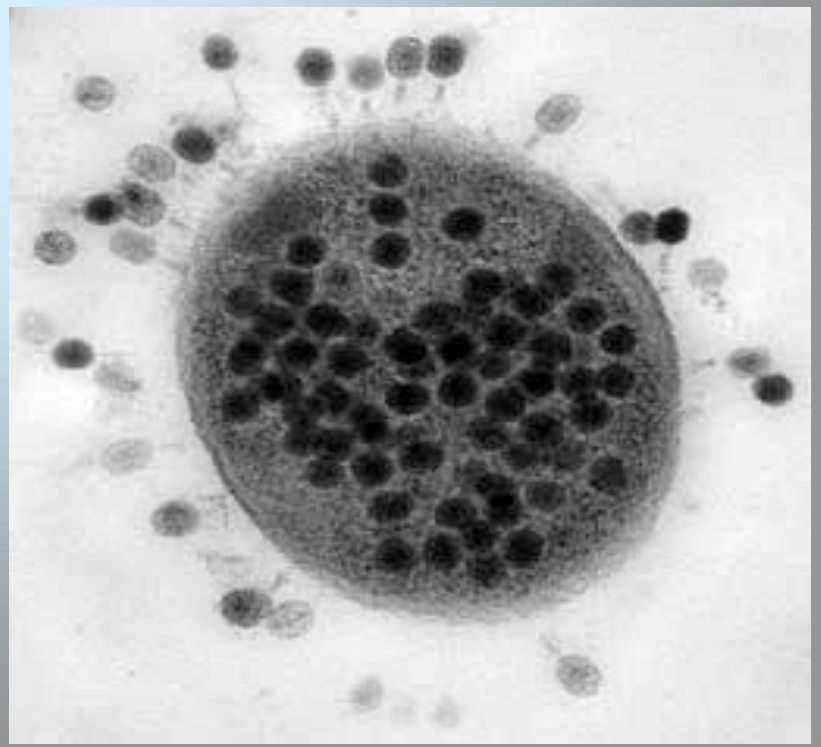



The Mimivirus viral factory was originally confused for the nucleus!



A viral factory is a cellular organism!

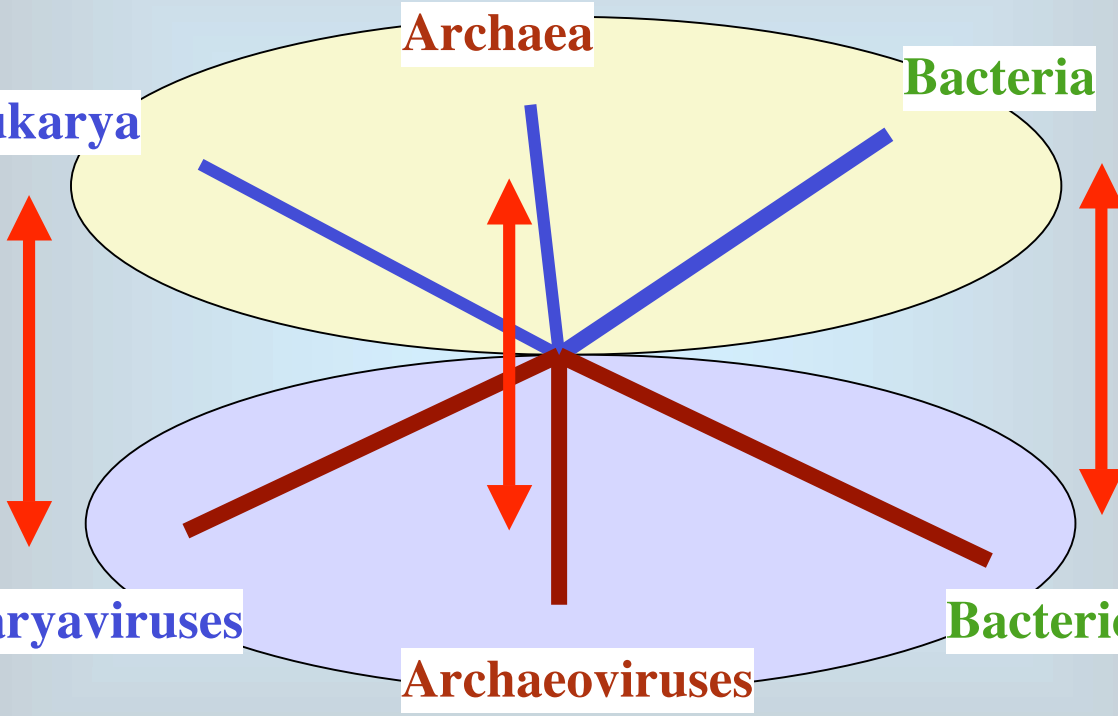
A bacteriovirus converts the bacterial cell into a viral factory

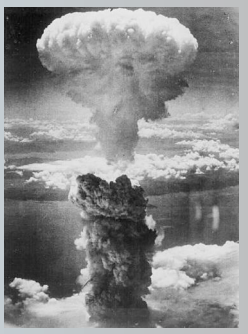


The fact to consider viruses as living organisms has major psychological consequences for biologists

Viruses can now be viewed as major positive actors in the history of life

If the two major categories of the living world are ribosomes and capsids encoding organisms, one indeed expect that their conflict will be a major factor of life evolution

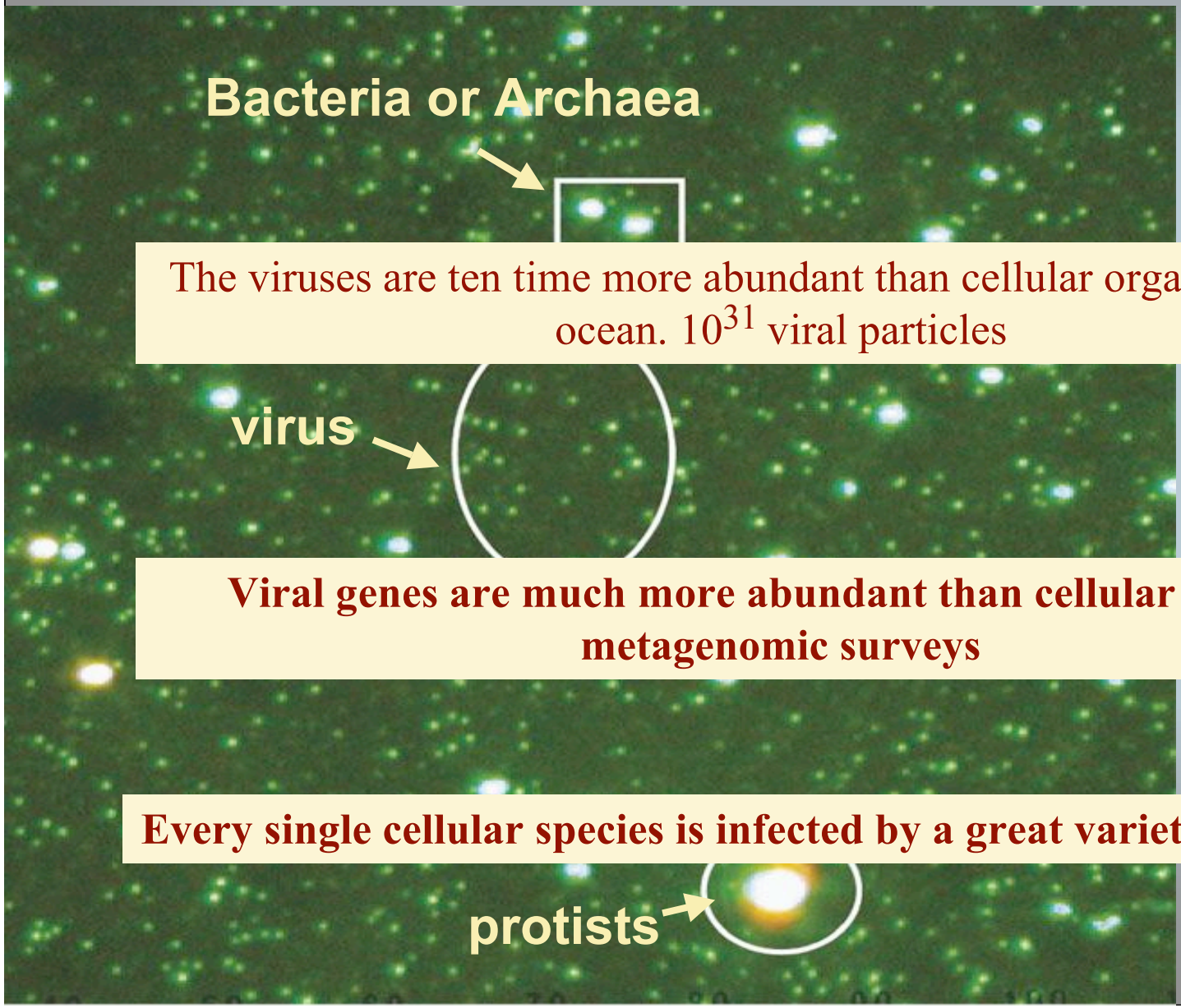




mutation + survive to the virus (or to the cell counter-attack)

Alliance or cooperation (against a third party!) hot and cold war, altruism, traitors, complex strategies)

Viruses are by far the most abundant biological entities on our planet



The viruses are ten time more abundant than cellular organisms in the ocean. 10^{31} viral particles

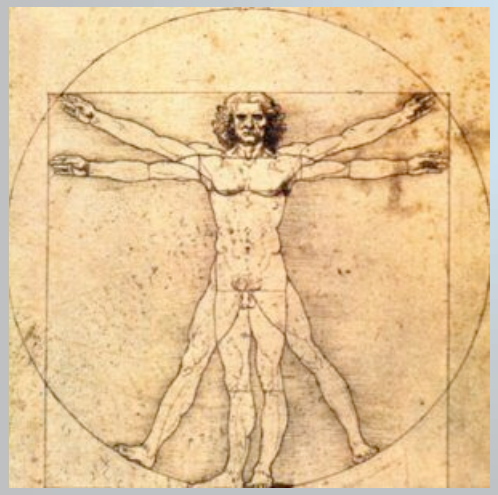
Viral genes are much more abundant than cellular genes in metagenomic surveys

Every single cellular species is infected by a great variety of viruses

Assumption n°1 : viruses are living organisms

Assumption n°2 : viruses are more abundant than cells

They will put a selection pressure on their cellular « hosts » or potential victims.



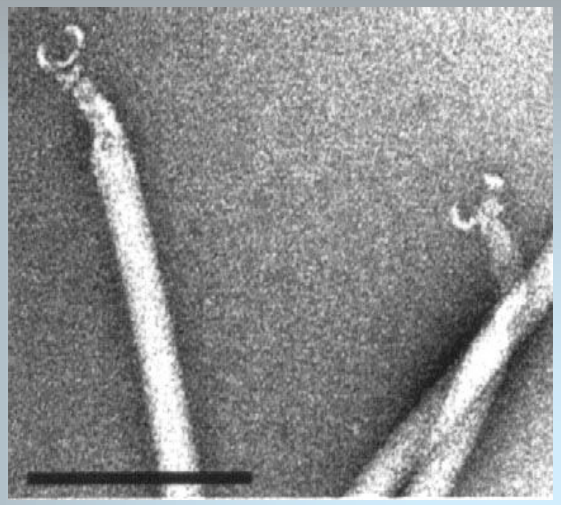
Assumption n°1 : viruses are living organisms

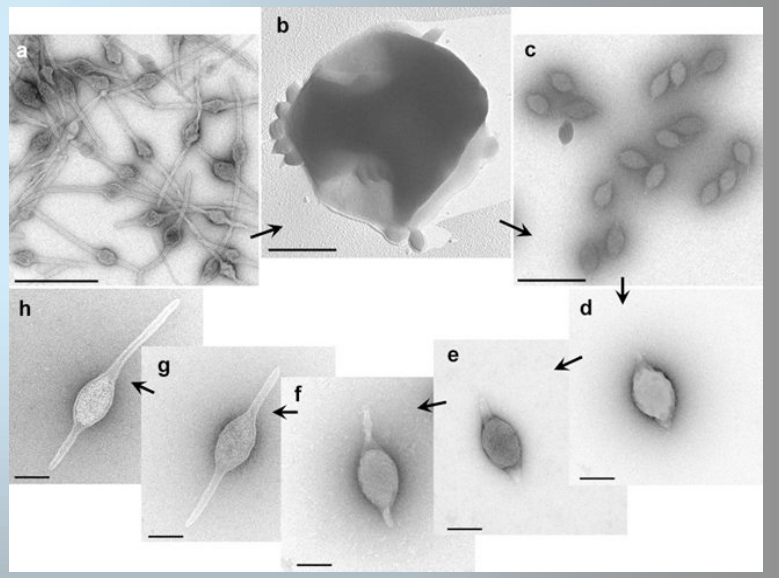
Assumption n°2 : viruses are more abundant than cells

Assumption n°3 : viruses are more diverse than cells

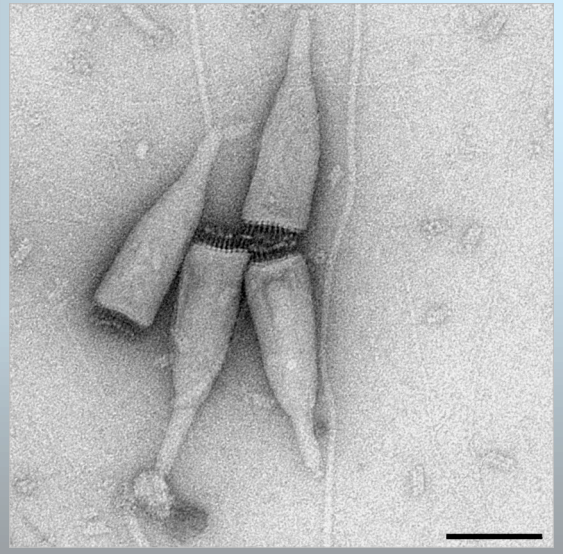
DNA or RNA Genomes (or both)
Linear or circular
Single or double-standed
Various strategies to initiate genome replication

Viruses can also have very diverse morphologies





David Prangishvili
Institut Pasteur



Prangishvili, Forterre and Garrett
Nature Rev Microbiol, 4, 837-848 2006

Recently discovered viruses infecting hyperthermophilic archaea are especially spectacular

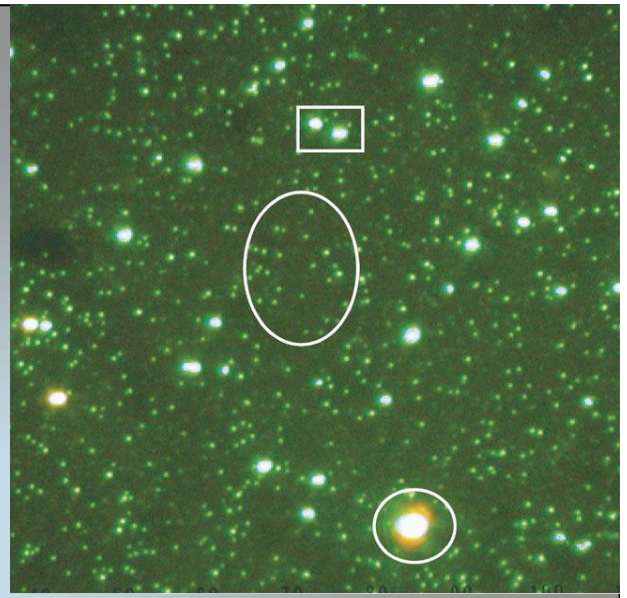
Assumption n°1 : viruses are living organisms

Assumption n°2 : viruses are more abundant than cells

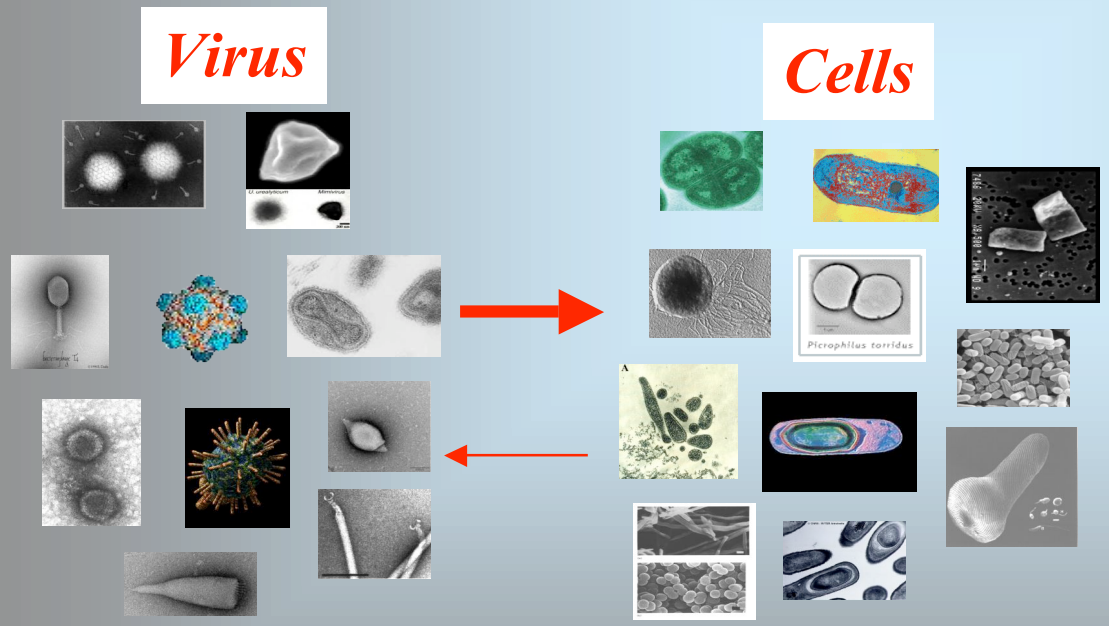
Assumption n°3 : viruses are more diverse than cells

Assumption n°4 : many cellular genes are of viral origin

One aspect of the war game, gene transferts occur continuously between cells and viruses



It is usually assumed that transfers occur mainly from cell to viruses (viruses as pick-pockets)



But assuming an equal rate of transfer, they should occur mainly from viruses to cells, since they are many more viruses than cells!!

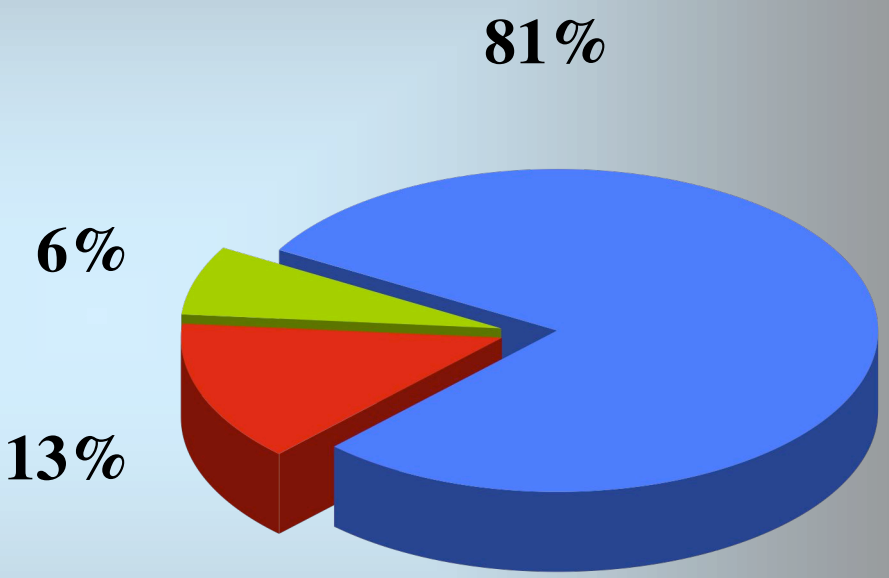
Furthermore, it is easier for a cell than for a virus to accomodate new genes

Indeed, they are few cellular genes in viral genomes

In contrast, they are many genes of viral (plasmid) origin in cellular genomes

About 20% of all genes in bacterial and archaeal genomes have an atypical dinucleotide composition and are likely of viral origin

Diego Cortez, Simonetta Gribaldo, Patrick Forterre

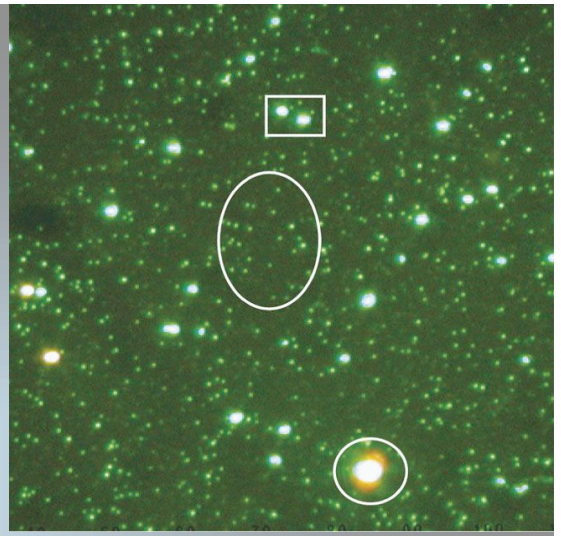


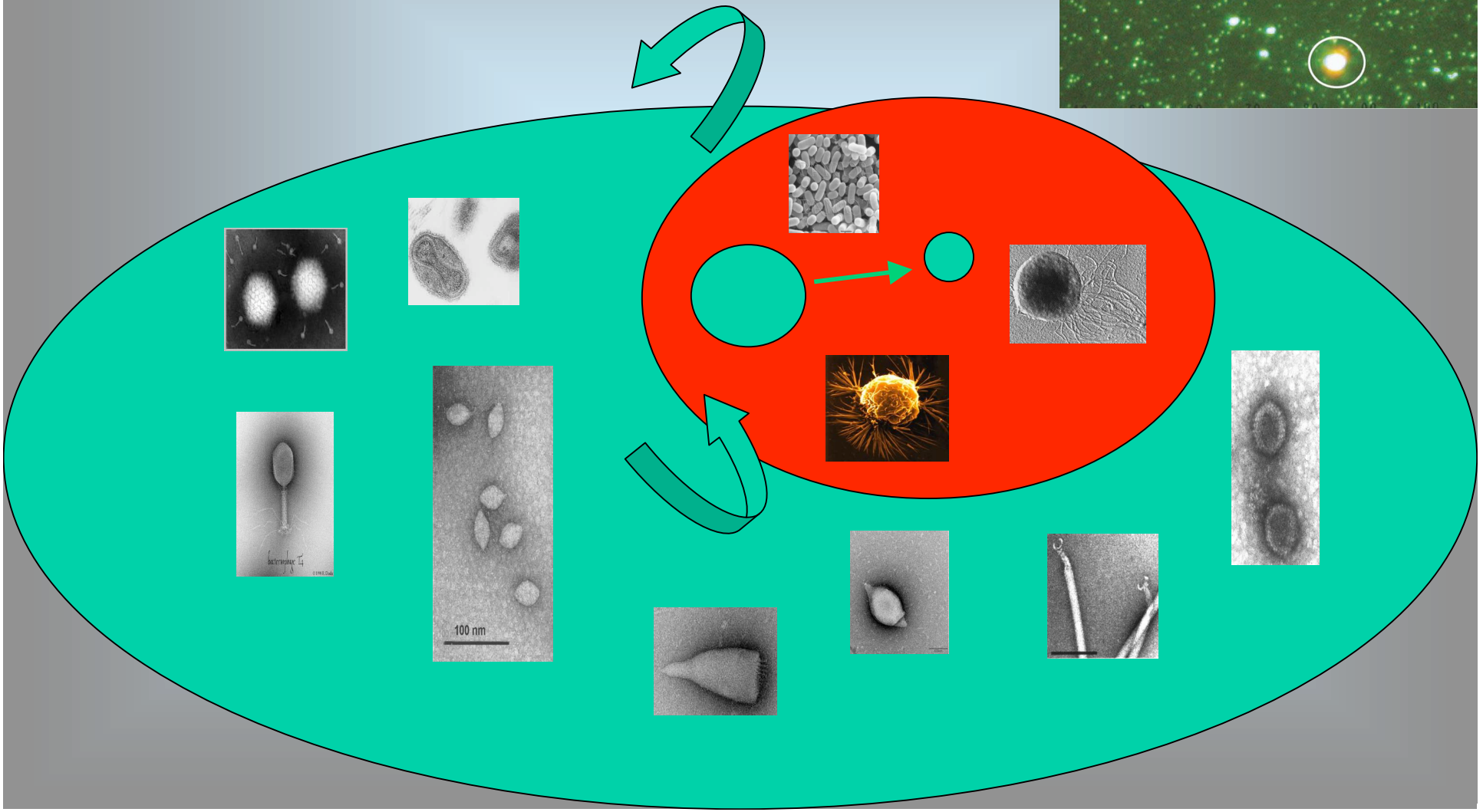
13 % are clearly of viral origin

The majority of ORFans are located in viral-related elements

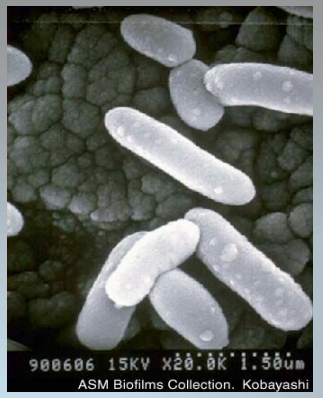
The viral reservoir hypothesis

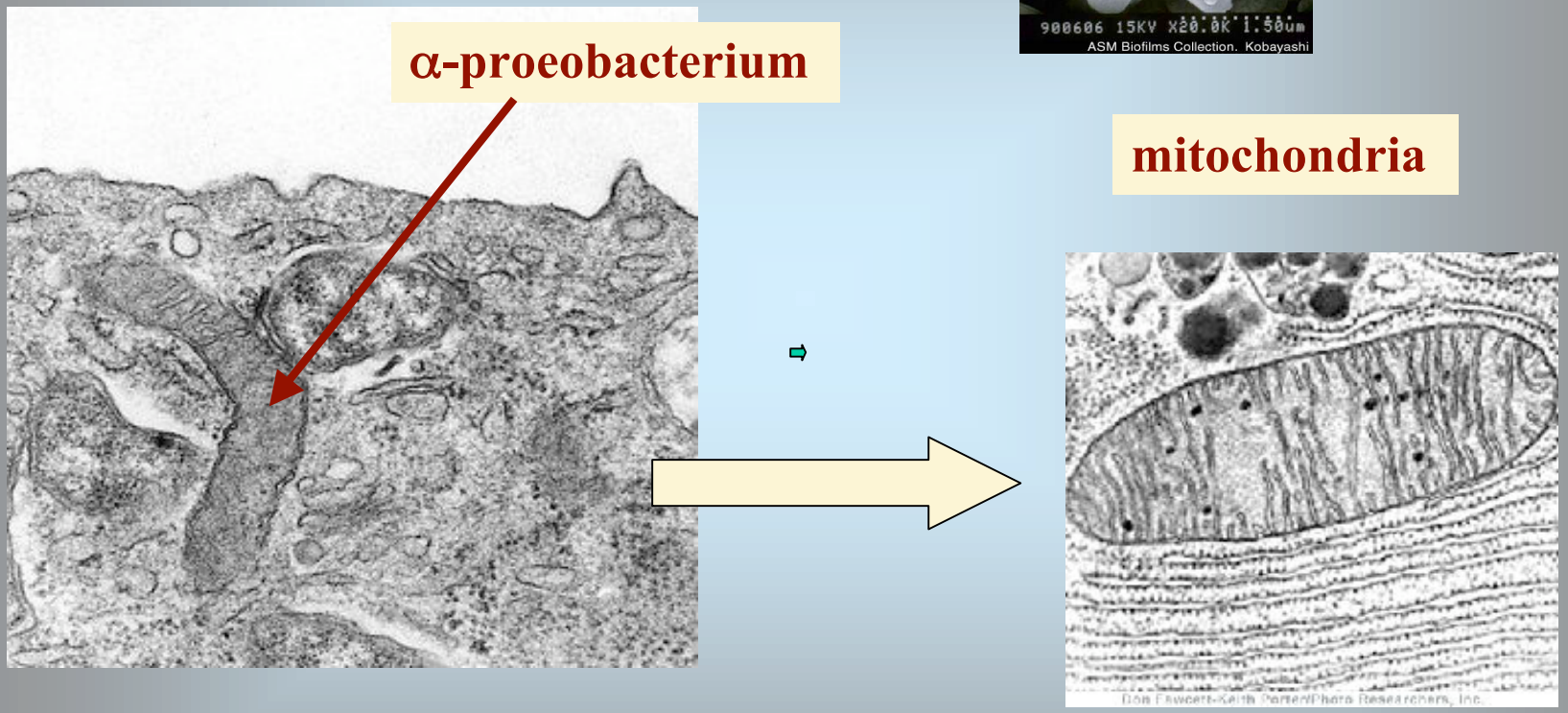
Cells are giant pick-pockets of viral genes





An ancient case of tranfer of viral genes into cellular genome:
The viral origin of the mitochondrial DNA replication and transcription machineries

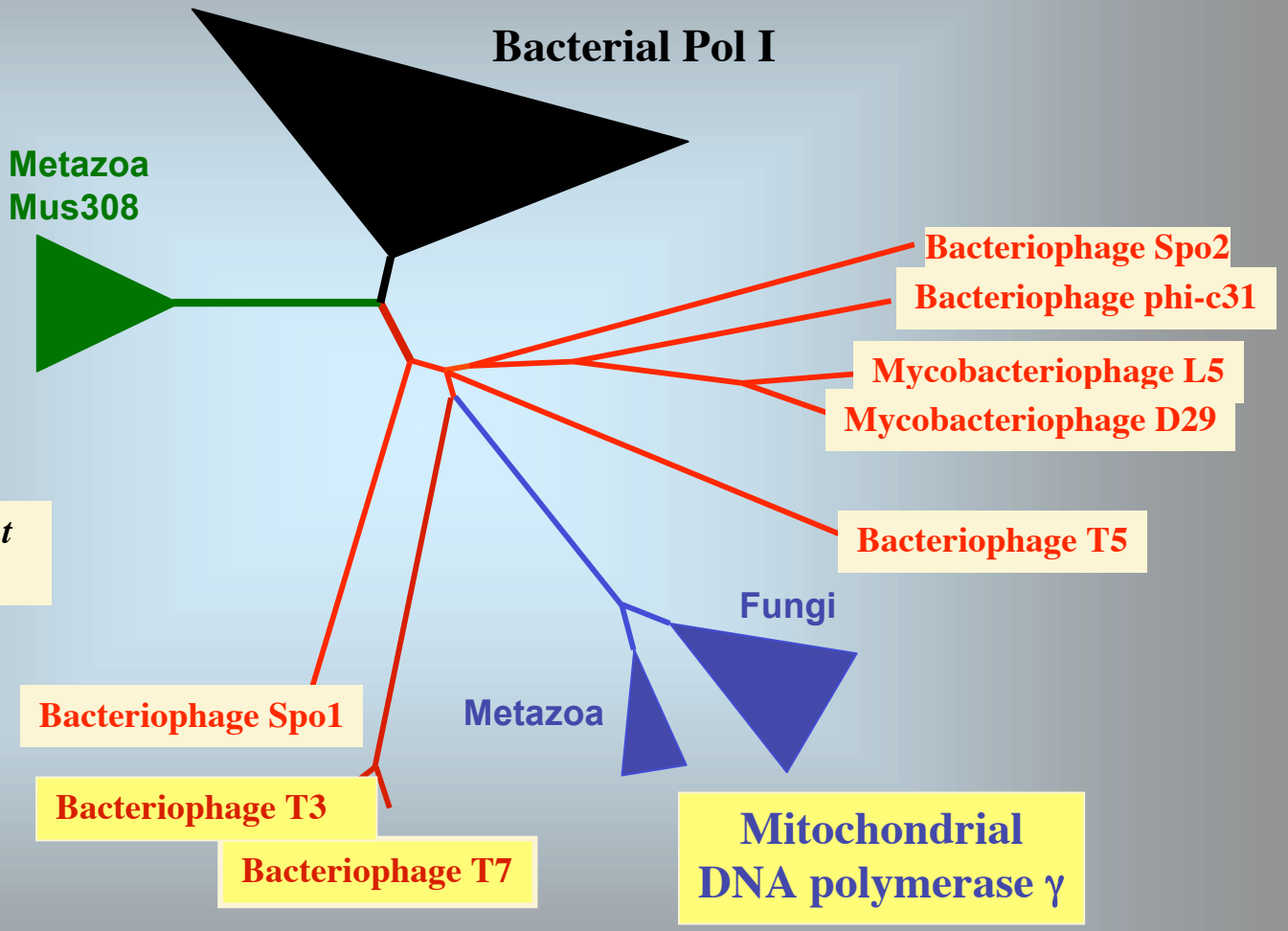




It has been known for twenty years that mitochondrial RNA polymerase is monomeric and looks like T3/T7 RNA polymerase !!!

More recently, it was shown that mitochondrial DNA polymerase gamma is also related to T3/T7 DNA polymerase

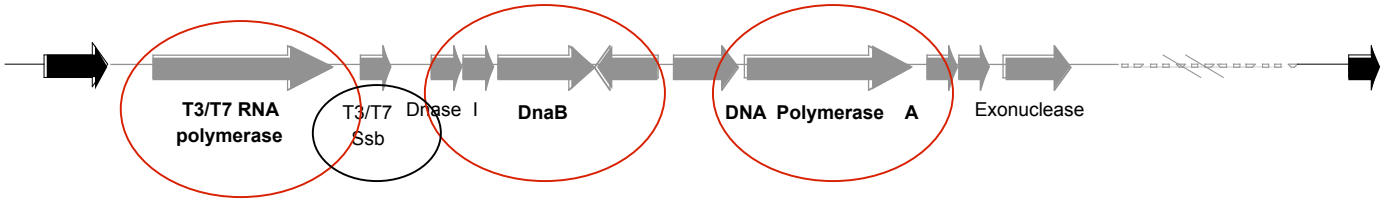
Filée, Forterre and Laurent
J. Mol. Evol. 2002



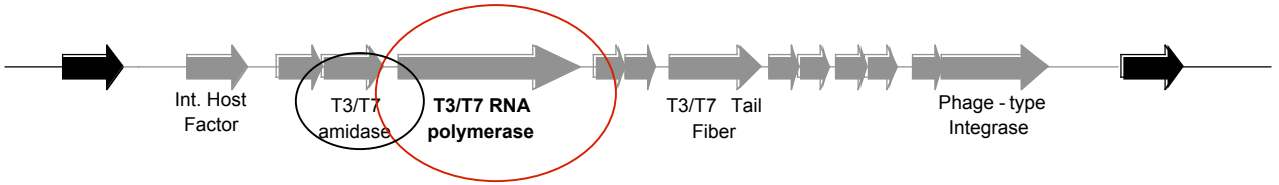
idem for the mitochondrial helicase/primase

Genes encoding proteins homologous to mitochondrial enzymes are present in « virus islands » in the genomes of several Proteobacteria (*Filée and Forterre, TIM, 2005*)

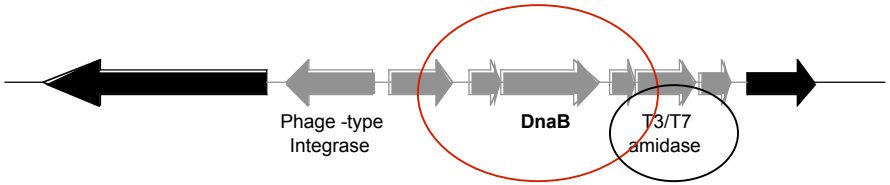
Pseudomonas putida KT2440



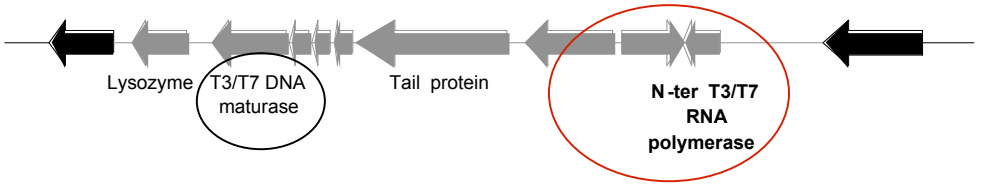
Agrobacterium tumefaciens C58



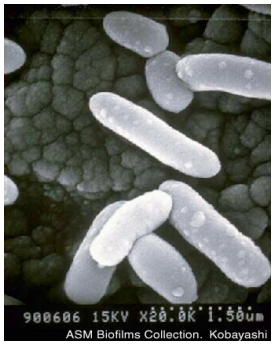
Bacteroides thetaiotamicron VPI-5482



Xanthomonas axonopodis pv. citri



The alpha proteobacterium at the origin of mitochondria should have harbour such integrated provirus



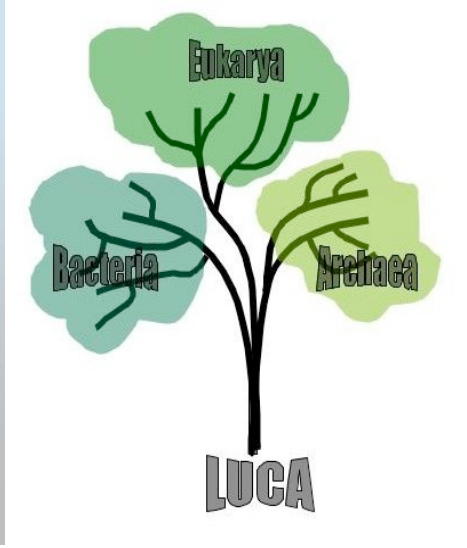
Assumption n°1 : viruses are living organisms

Assumption n°2 : viruses are more abundant than cells

Assumption n°3 : viruses are more diverse than cells

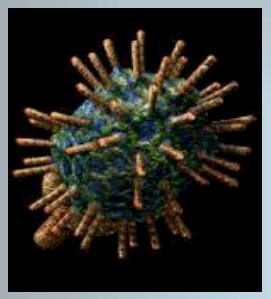
Assumption n°4 : many cellular genes are of viral origin

Assumption n°5 : viruses are ancient

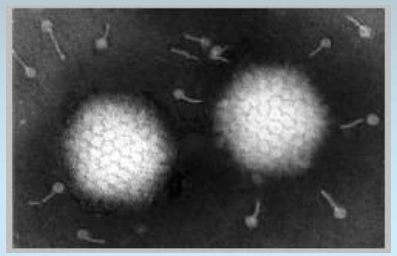


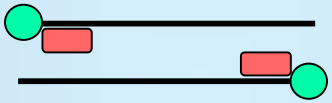
Viruses infecting different domains share similar atypical DNA replication mechanisms

Homologous protein-primed DNA polymerases of the B family

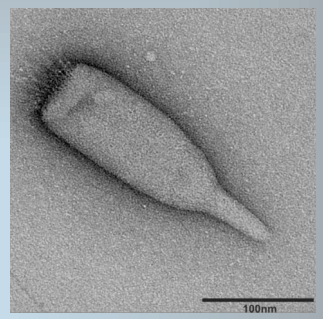


Génome du **bactériophage Φ29** (*B. subtilis*)



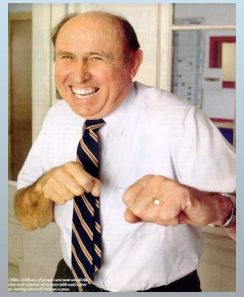


Génome des **Adenovirus** (*Homo sapiens*)



Génome de **l'Ampullavirus** (*Acidianus infernus*)

Bacteria



Eukarya

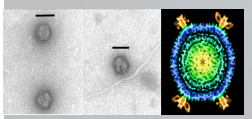
Archaea

Cells from different domains can be infected by DNA viruses sharing homologous capsid proteins ***(Denis Bamford, Helsinki)***

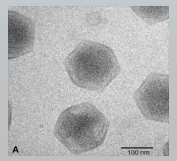
Two non-homologous types of capsid proteins have been indentified until now

Hypothesis: viruses are very ancients, they probably predate the last Universal Cellular Ancesor (LUCA)

Archaea



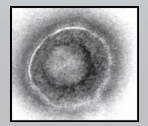
Eucarya



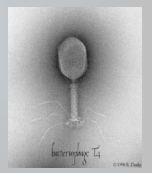
Bacteria

Archaea

Eucarya



Bacteria



Capsid proteins seem to co-evolve with infected hosts/victims:

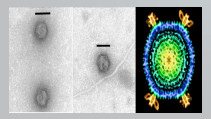
homology between capsid proteins can be detected by sequence similarity inside domain but only by structural similarity between domains

Hypothesis: viruses are very ancients, they probably predate the last Universal Cellular Ancesor (LUCA)

Eucarya (adenovirus)

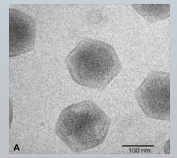
Archaea

Crenarchaea



Euryarchaea

Eucarya

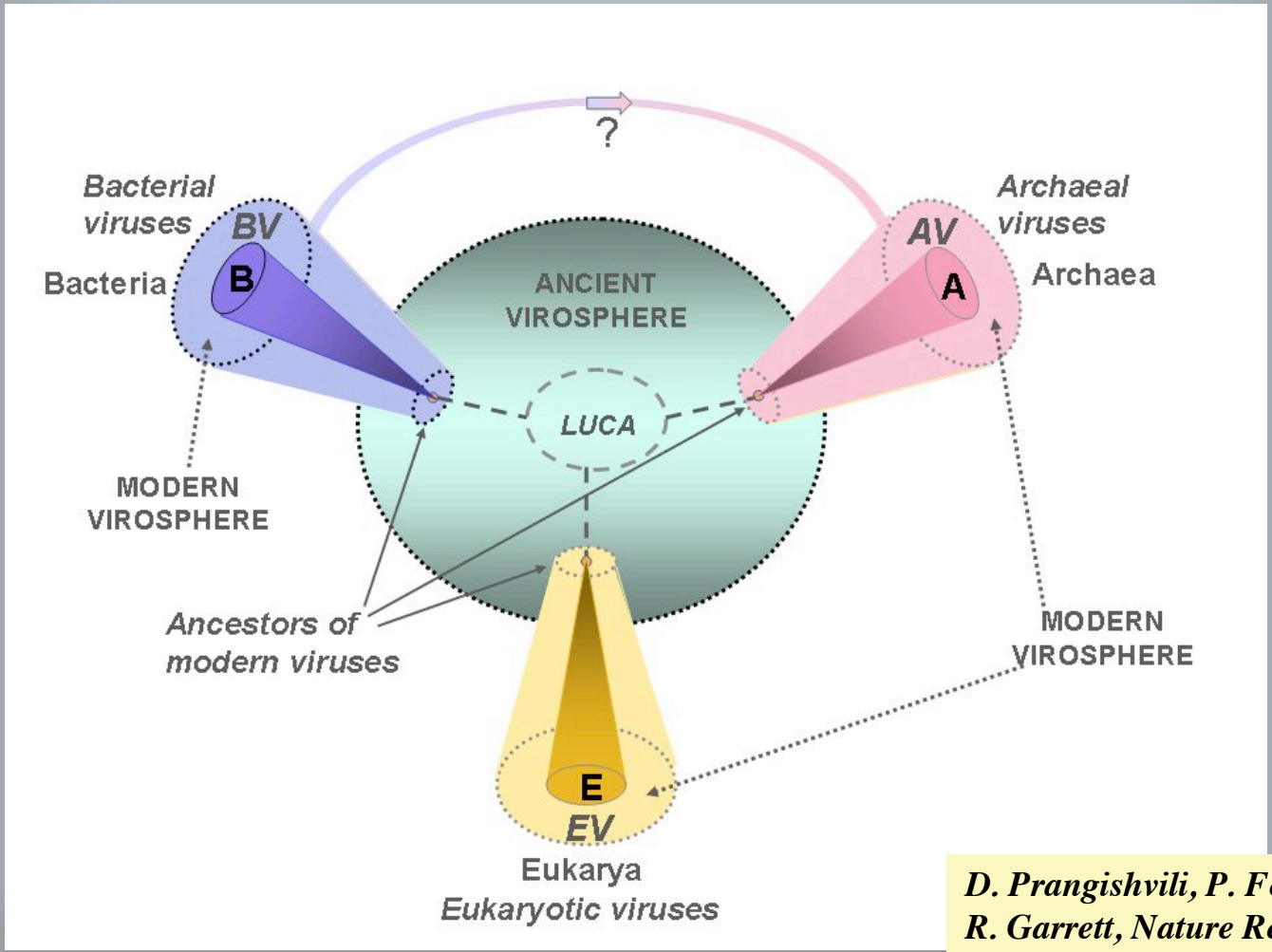


Bacteria

Firmicutes

Proteobacteria

Hypothesis: viruses are very ancients, they probably predated the last Universal Cellular Ancesor (LUCA)



D. Prangishvili, P. Forterre R. Garrett, Nature Reviews Microbiology, 2006

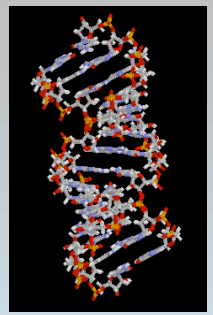
Hypothesis: viruses originated before modern DNA-cells

Zillig ***Forterre*** ***Koonin and Dolja***

The first viruses probably infected ancient RNA-cells

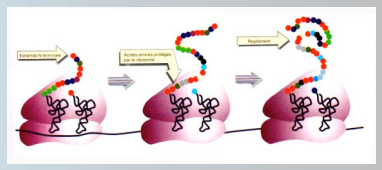
Forterre

First age of the RNA world



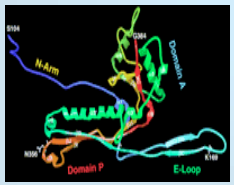
Cells with RNA genomes and ribozymes

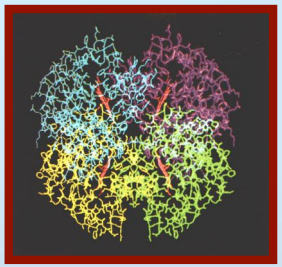
« invention » of the ribosome

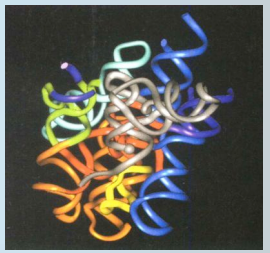


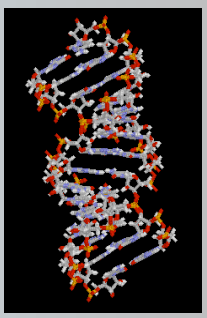
Second age of the RNA world

Cells with RNA genomes and proteins

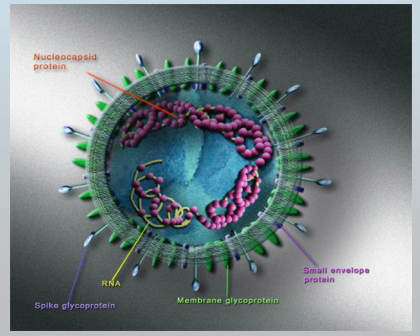




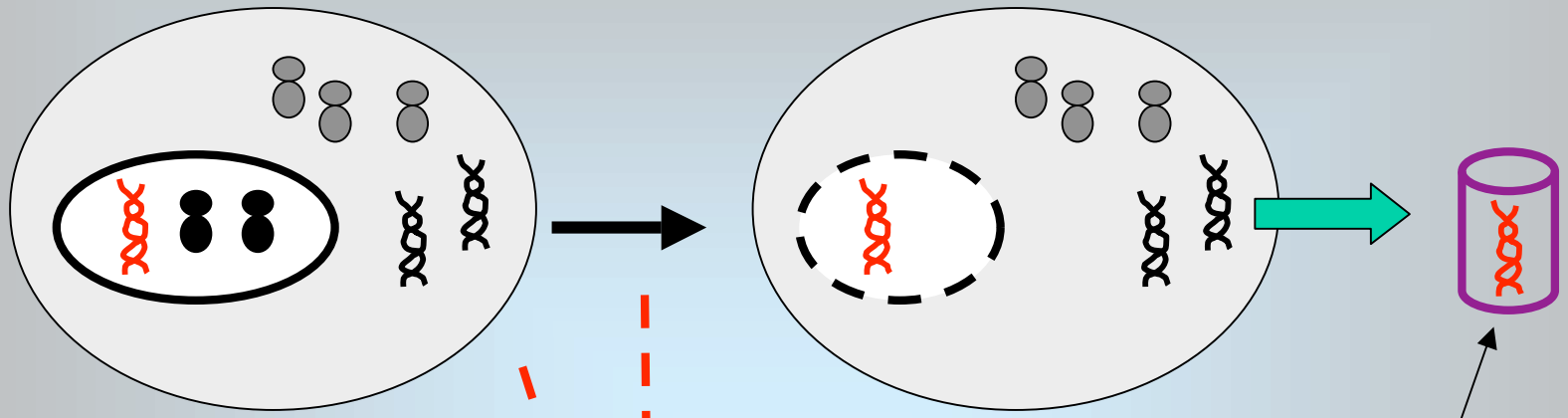




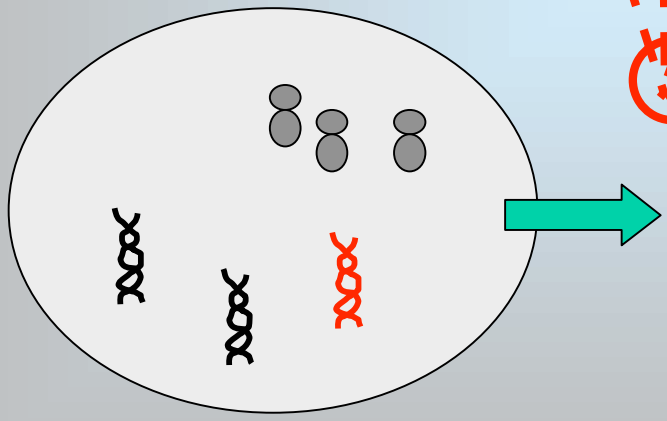
Hypothesis: viruses originated in a world of RNA-cells



Hypothesis: viruses originated in a world of RNA-cells



The reduction hypothesis



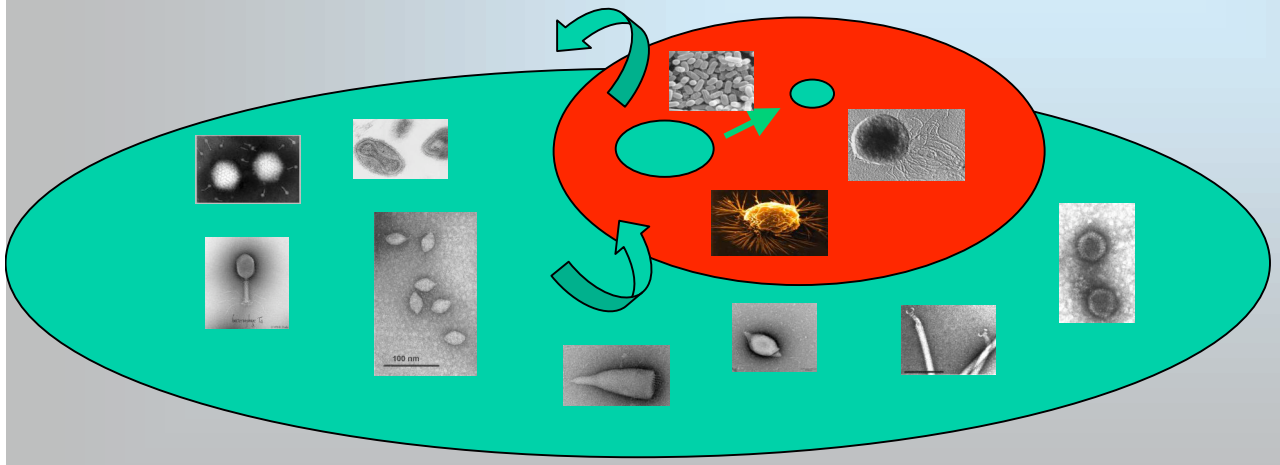
The escape hypothesis

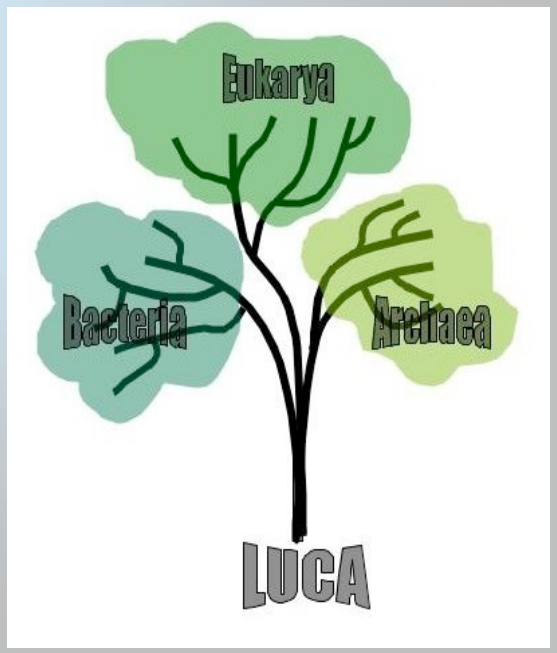
Forterre, P. Virus Res. 2006

Assumption n°5 : viruses are ancient

The incoming flux of viral genes into cellular genomes started before LUCA and has been continuous for more than 3 billions years

This war game probably started in a cellular RNA world





viruses are living organisms

viruses are more abundant than cells

viruses are more diverse than cells

many cellular genes are of viral origin

viruses are ancient

Conclusion: the conflict between cellular and viral organisms has been the major engine of biological evolution

Impact of viruses on cellular evolution

Introduction in cellular organisms of innovations that first originated in the viral world

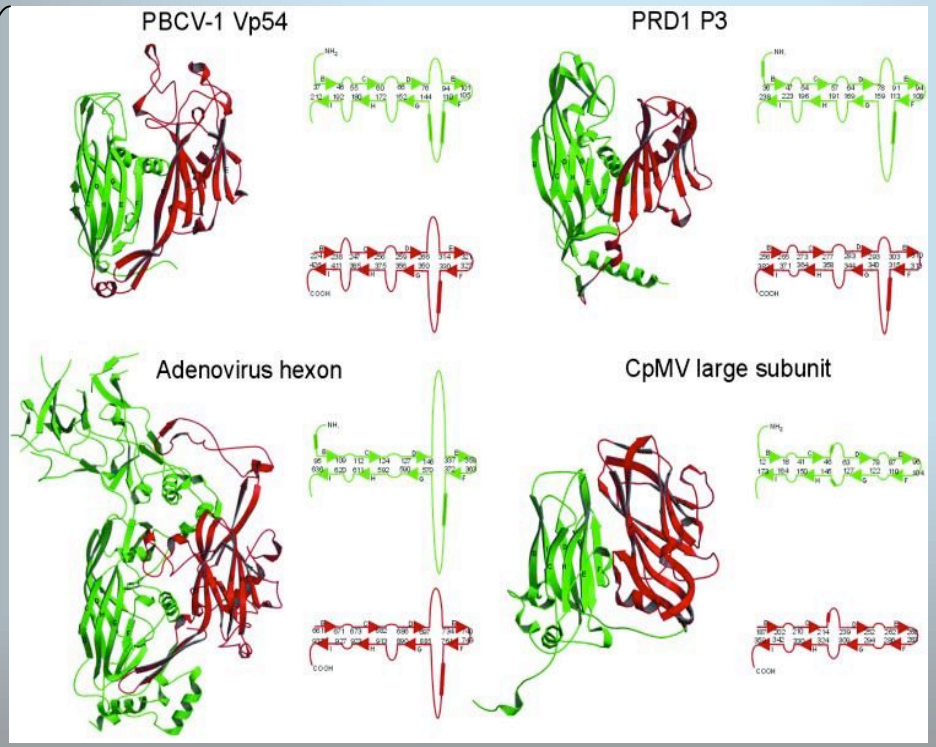
Invention of new mechanisms to fight back against viruses (arm race)

Viruses and the origin of DNA genomes?

Viruses and the origin of DNA genomes?

The double-jelly rolls fold is present in the capsid of both DNA and RNA viruses

Nandhagopal, N…..Rossman, M.G. al., PNAS, 99, 2002



Eukarya dsDNA

Bacteria dsDNA

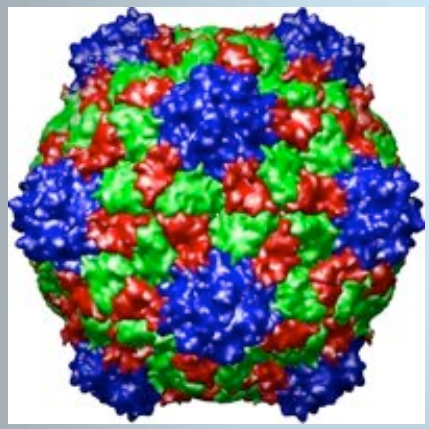
Eukarya ssRNA

« Thus, the capsid proteins of large dsDNA icosahedral viruses …and the small ssRNA icosahedral viruses are likely to have evolved from a common ancestor »

Nandhagopal, N…..Rossman, M.G. al., PNAS, 99, 2002

RNA virus (CpMV)

DNA virus (PRD1)



But, if DNA viruses originated from RNA viruses, this seems to imply that the origin of DNA from RNA occurred in the viral world

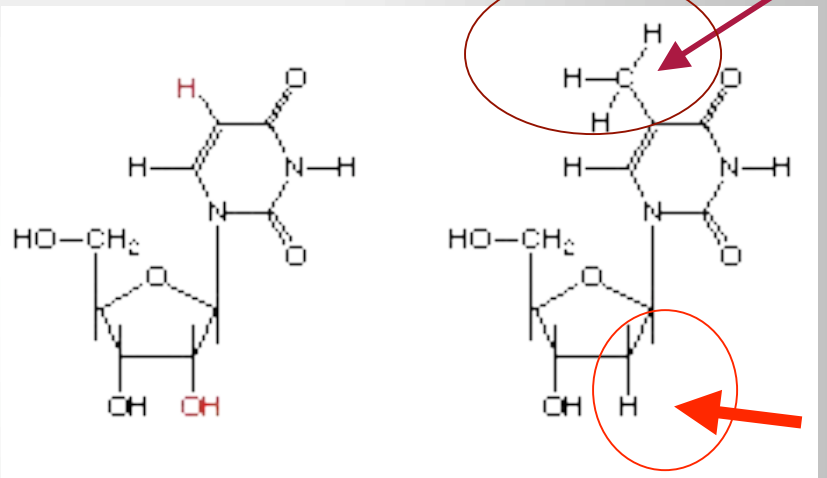
Why RNA was replaced by DNA?

Why RNA was replaced by DNA?

The traditional answer for the origin of DNA

DNA is a modified form of RNA

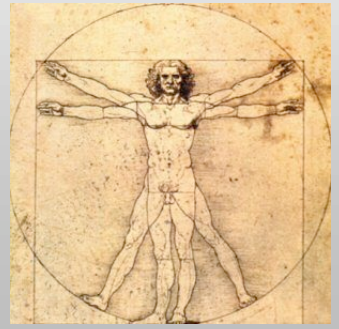
DNA replaced RNA because it is more stable and spontaneous cytosine deamination (C to U) can be repaired in DNA but not in RNA

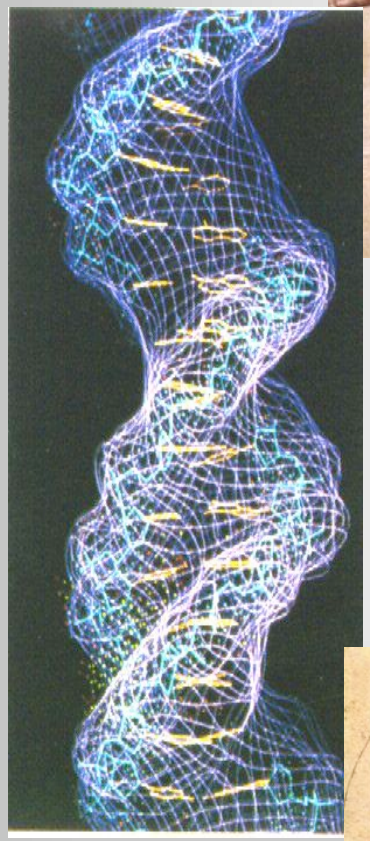


uracil (RNA)

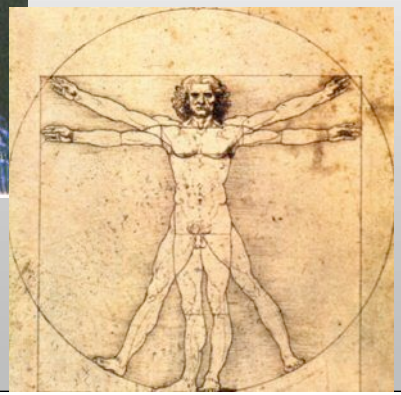
Thymine (RNA)

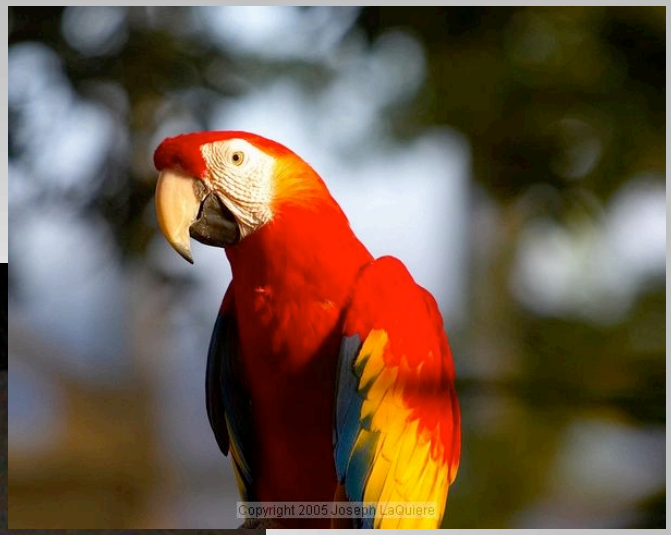
These properties allowed the increase in genome size *required for evolution towards more complexity??*



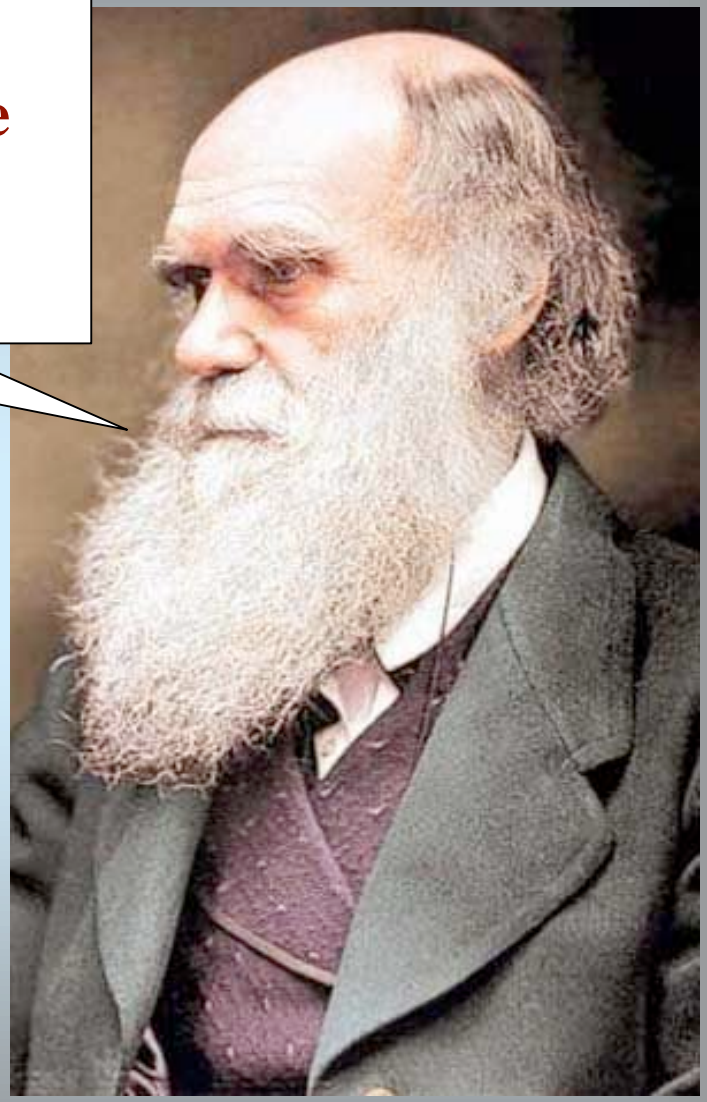


Yes, but only God
(Intelligent Design)
could have predict it !

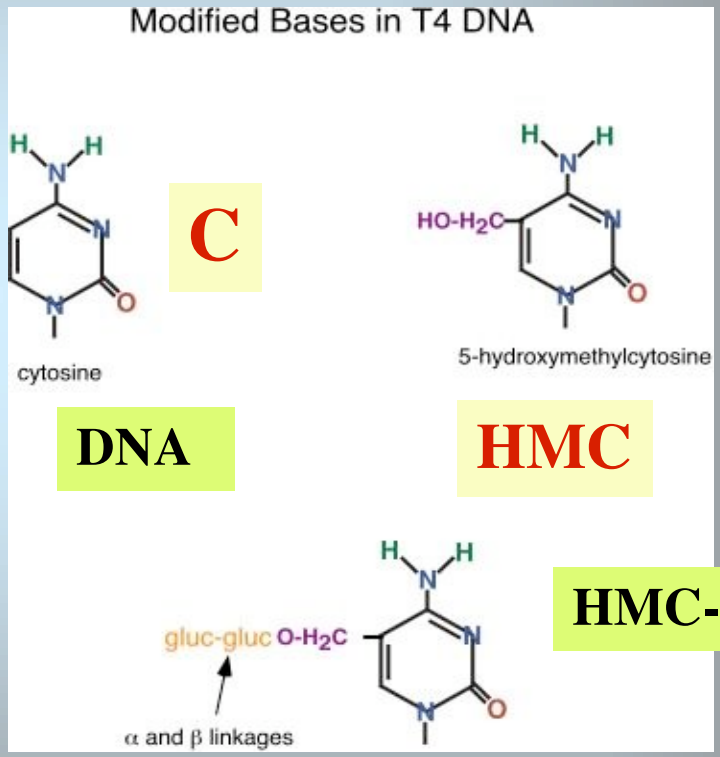



Copyright 2005 Joseph LaQuiere

One should understand the immediate selective advantage for the first organism with a DNA genome

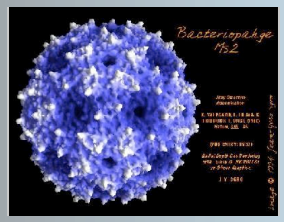


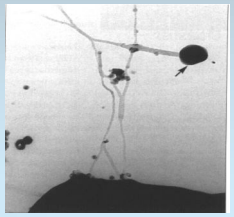
We know that many viruses have « invented » new forms of DNA

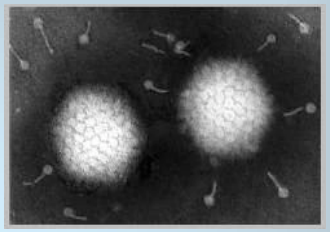


To protect their genomes against nucleases encoded by their cellular « hosts

Hypothesis: RNA viruses have invented DNA to protect their genomes. exactly as some modern viruses have modified their DNA.







RNA → **U-DNA** → **T-DNA** → **HMC-DNA**

ThyA

RNR

ThyX

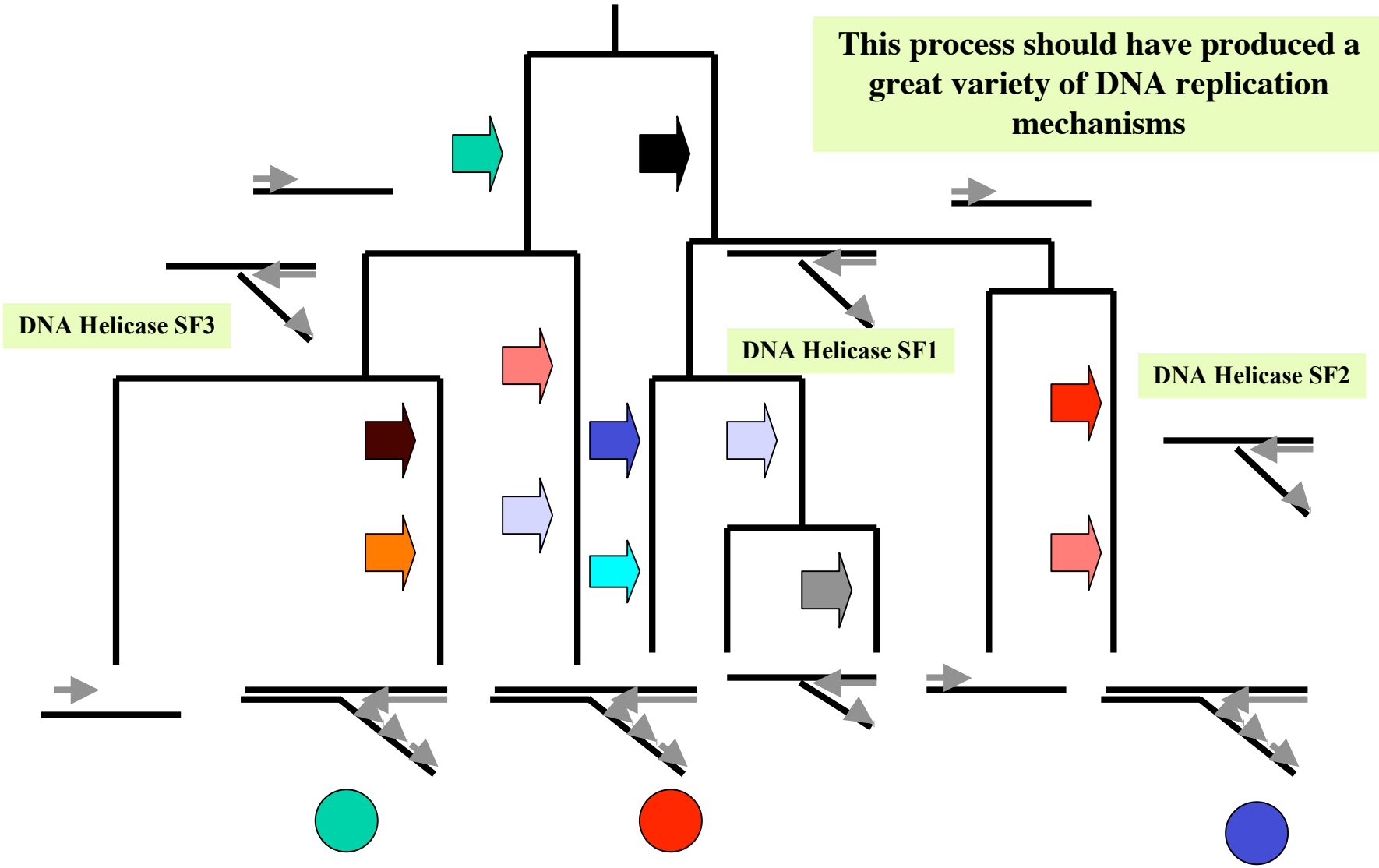
ThyA-like

Forterre, P.
Current Opinion in Microbiol. 2002

Discovery of a new family of thymidylate synthase
(Myllykallio et al., Science, 2002)

If DNA was first selected in the viral world, different lineades of DNA viruses might have « invented » more and more complex mechanisms for DNA replication repair and recombination by recruting proteins previously used for RNA metabolism

This process should have produced a great variety of DNA replication mechanisms



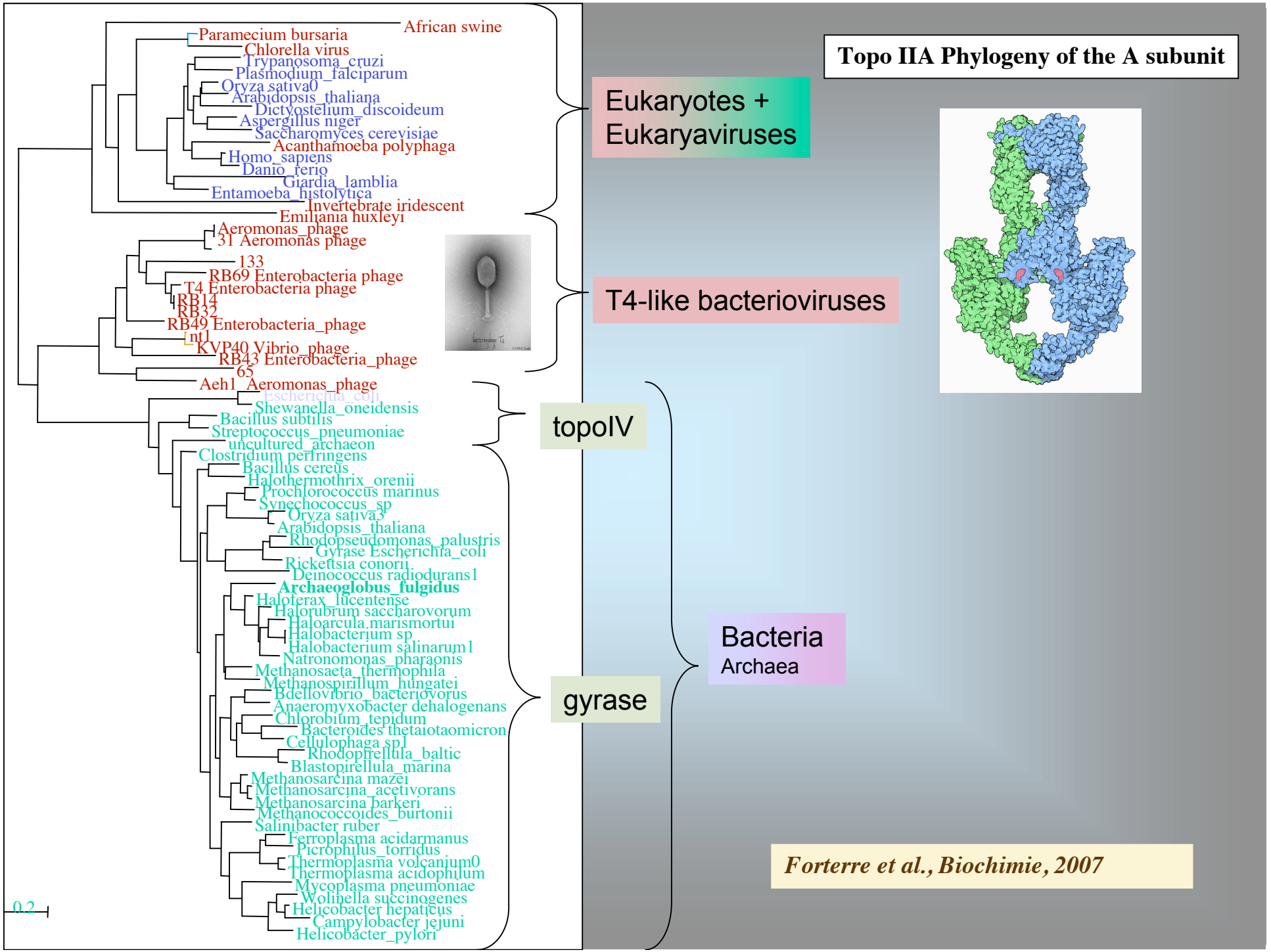
Many DNA viruses encode proteins involved in DNA metabolism that are very divergent from their cellular homologues

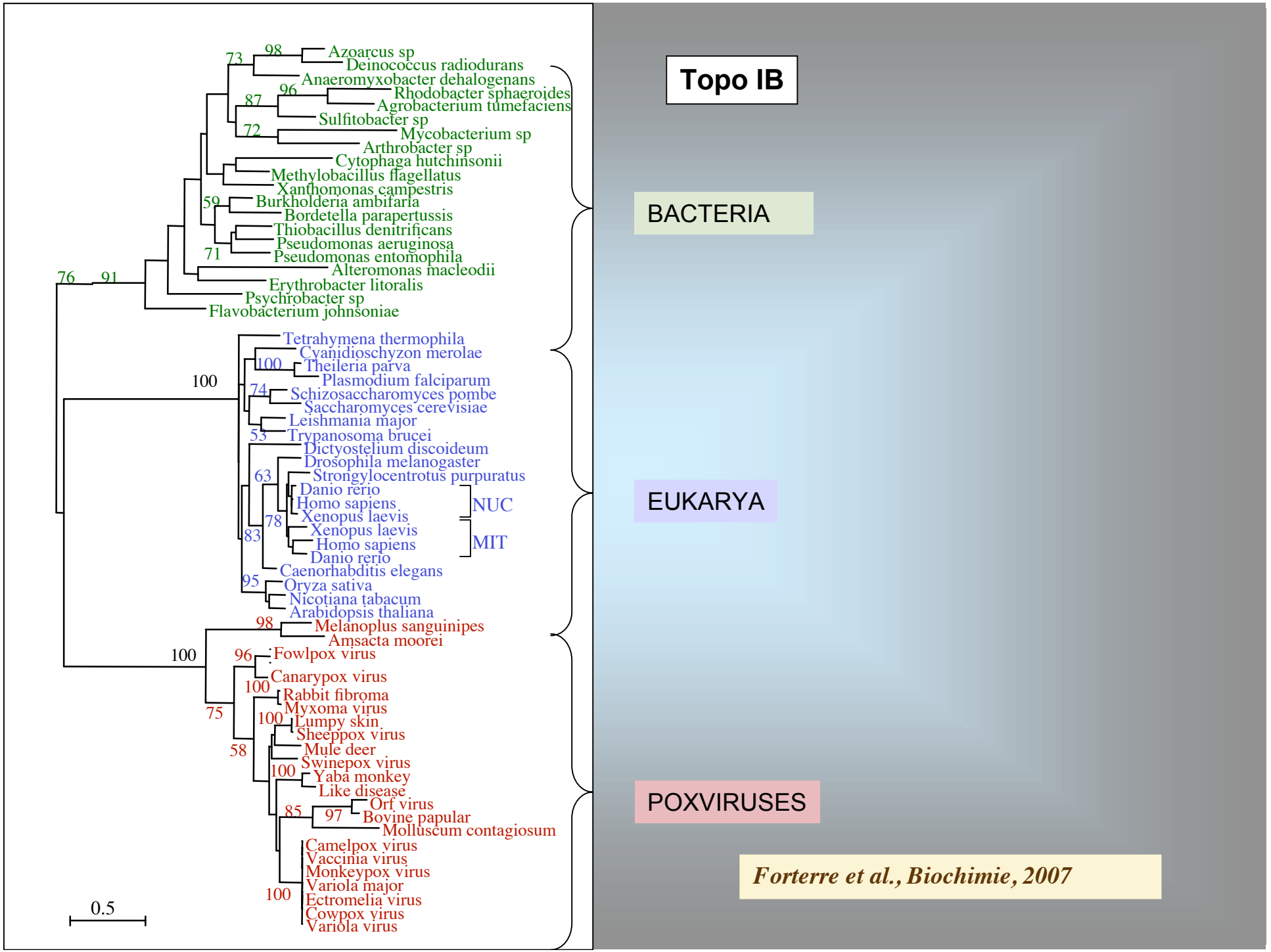
Protein-primed DNA polymerases
Type IIA DNA topoisomerases, ⇐
Type IB DNA topoisomerases, ⇐
DNA ligase,
Thymidylate synthases,
Ribonucleotide reductases,

Many DNA viruses encode proteins involved in DNA metabolism which have no cellular homologues

DNA helicase superfamily III,
Rep initiator protein for Rolling-circle replication,
DNA polymerase family E,
T3/T7 monomeric RNA polymerase

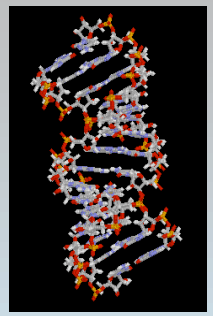
Hypothesis: DNA replication mechanisms and associated proteins originated in the viral world and were later on transferred (or not) to cells


Topo IIA Phylogeny of the A subunit
African swine
Paramecium bursaria
Chlorella virus
Trypanosoma_cruzi
Plasmodium_falciparum
Oryza sativa0
Arabidopsis_thaliana
Dictyostelium_discoideum
Aspergillus niger
Saccharomyces cerevisiae
Acanthamoeba polyphaga
Homo_sapiens
Danio_rerio
Giardia_lamblia
Entamoeba_histolytica
Invertebrate iridescent
Emiliania huxleyi
Aeromonas_phage
31 Aeromonas phage
133
RB69 Enterobacteria phage
T4 Enterobacteria phage
RB14
RB32
RB49 Enterobacteria_phage
nt1
KVP40 Vibrio_phage
RB43 Enterobacteria_phage
65
Aeh1_Aeromonas_phage
Escherichia_coli
Shewanella_oneidensis
Bacillus subtilis
Streptococcus_pneumoniae
uncultured_archaeon
Clostridium perfringens
Bacillus cereus
Halothermothrix_orenii
Prochlorococcus marinus
Synechococcus_sp
Oryza sativa3
Arabidopsis_thaliana
Rhodopseudomonas_palustris
Gyrase Escherichia_coli
Rickettsia conorii
Deinococcus radiodurans1
Archaeoglobus_fulgidus
Haloferax_lucentense
Halorubrum saccharovorum
Haloarcula marismortui
Halobacterium sp
Halobacterium salinarum1
Natronomonas_pharaonis
Methanosaeta_thermophila
Methanospirillum_hungatei
Bdellovibrio_bacteriovorus
Anaeromyxobacter dehalogenans
Chlorobium_tepidum
Bacteroides thetaiotaomicron
Cellulophaga sp1
Rhodopirellula_baltic
Blastopirellula_marina
Methanosarcina mazei
Methanosarcina_acetivorans
Methanosarcina barkeri
Methanococcoides_burtonii
Salinibacter ruber
Ferroplasma acidarmanus
Picrophilus_torridus
Thermoplasma volcanium0
Thermoplasma acidophilum
Mycoplasma pneumoniae
Wolinella succinogenes
Helicobacter hepaticus
Campylobacter jejuni
Helicobacter_pylori
0.2
Eukaryotes +
Eukaryaviruses
T4-like bacterioviruses
topoIV
gyrase
Bacteria
Archaea
Forterre et al., Biochimie, 2007



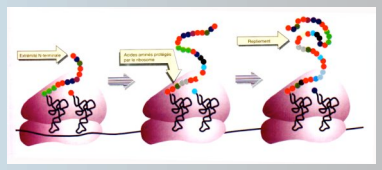
Forterre et al., Biochimie, 2007

First age of the RNA world



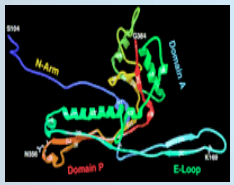
Cells with RNA genomes and ribozymes

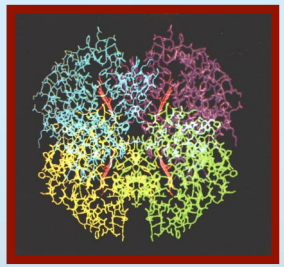
« invention » of the ribosome

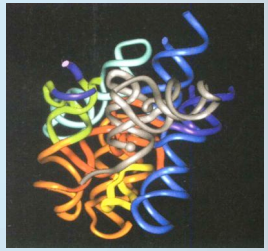


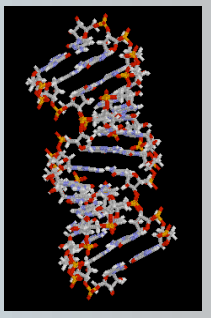
Second age of the RNA world

Cells with RNA genomes and proteins

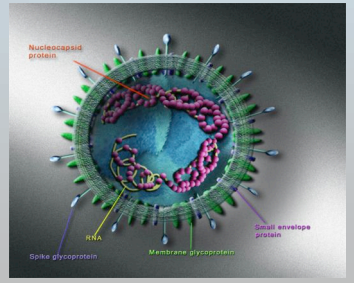


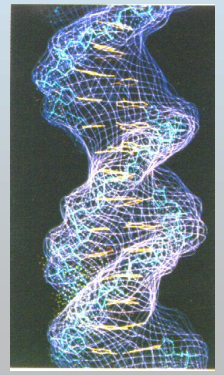






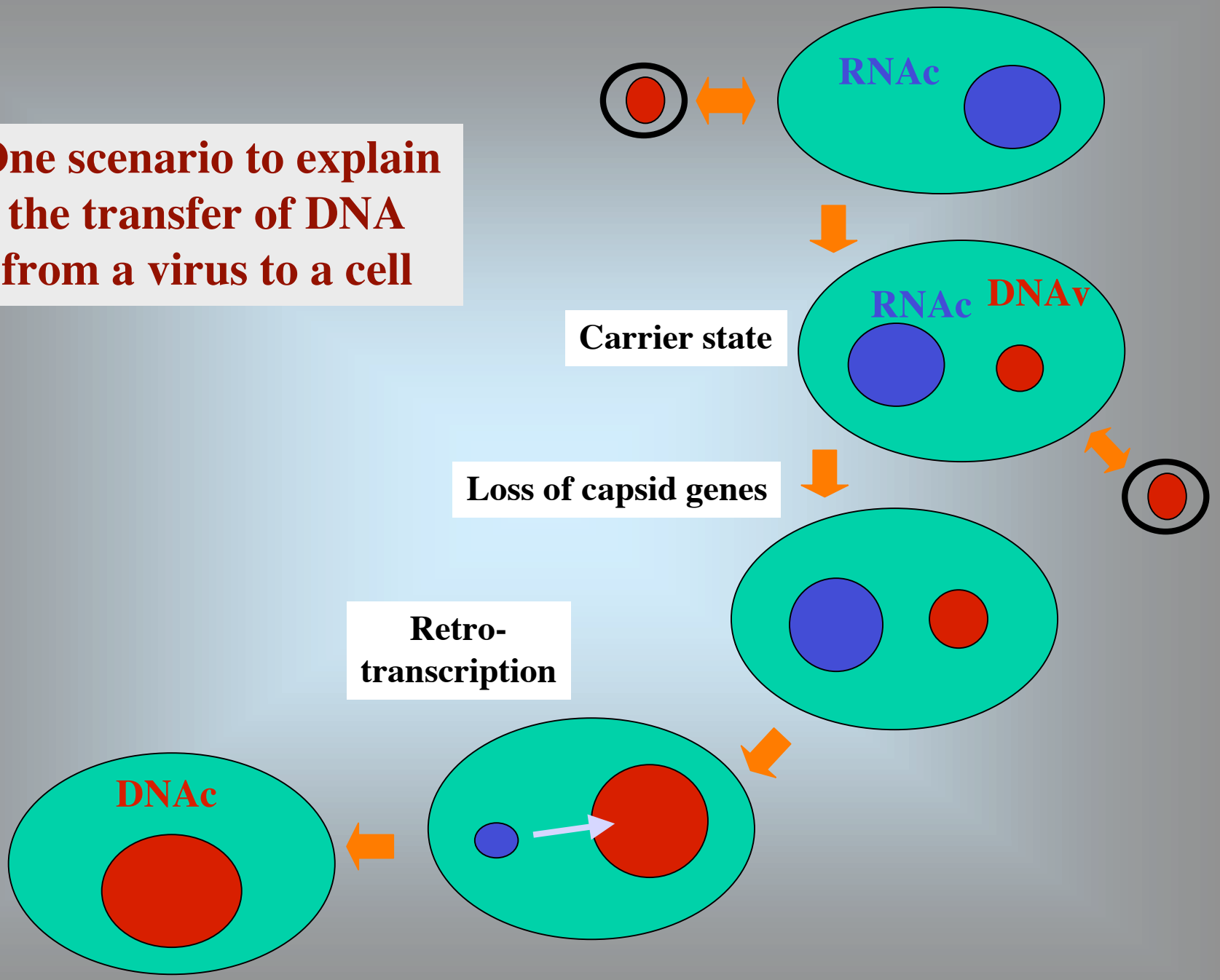
Third age: RNA, Proteins and DNA

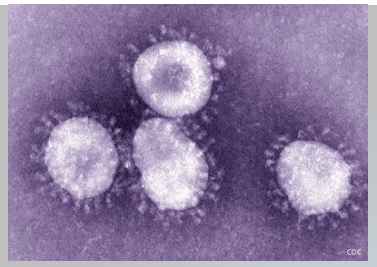


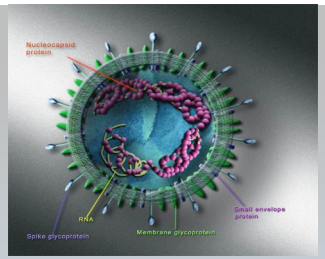


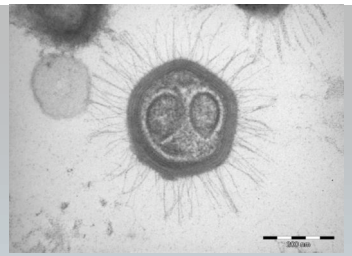
?

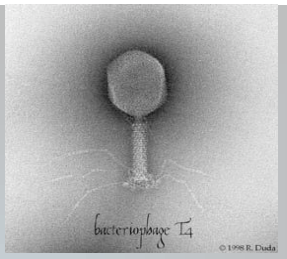
One scenario to explain the transfer of DNA from a virus to a cell



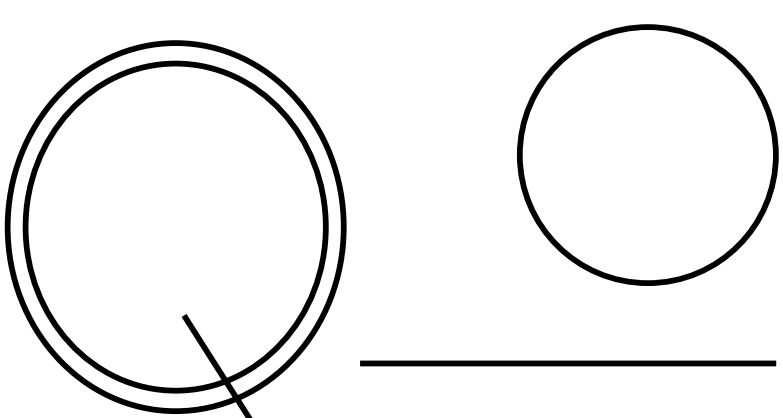


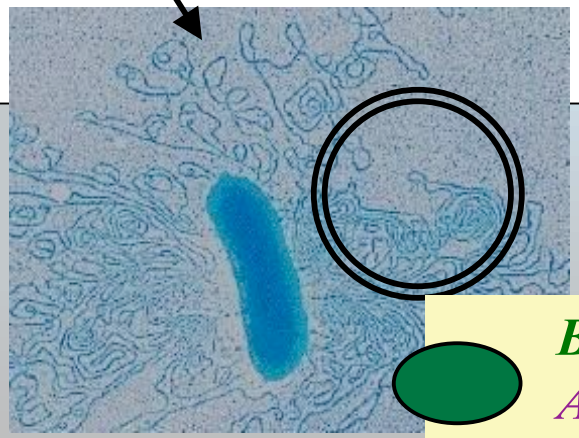






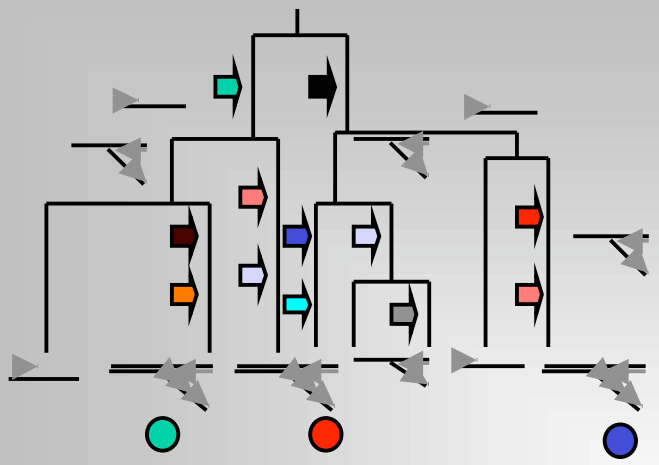
Only a subset of viral replication proteins and replication mechanisms were transferred from viruses to cells





Bacteria
Archaea

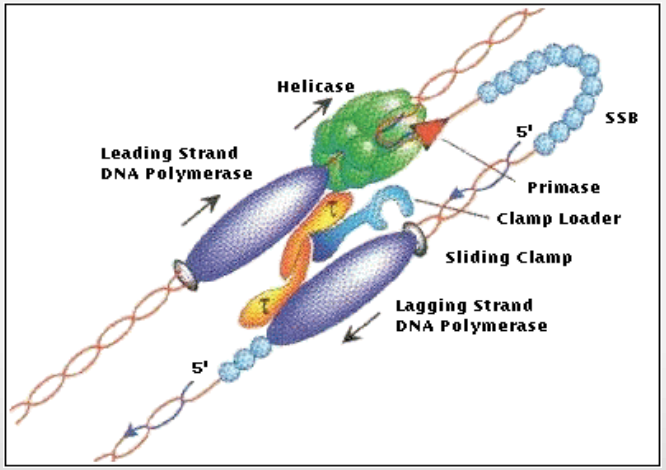
Eucaryotes



The viral origin of DNA replication mechanisms could explain why there are two non-homologous protein sets for DNA replication in cellular organisms

Archaea/Eukarya

Mcm
DNA pol (family B)
Primase



Bacteria

DnaB
DNA pol III (family C)
DnaG

A/E Mcm and DnaB belong to different families of AAA+ ATPases
A/E and bacterial replicases belong to different non homologous DNA polymerase families
A/E Primase and DnaG belong to different protein superfamilies

A viral origin for bacterial DNA replication proteins?

(Mol micro, 1999)

Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins

Patrick Forterre
Institut de Génétique et Microbiologie, Bat 409, CNRS, UMR 8621, Université Paris-Sud, 91405 Orsay Cedex, France.

A viral origin for eukaryotic DNA replication proteins?

(J. virol, 2000)

A Hypothesis for DNA Viruses as the Origin of Eukaryotic Replication Proteins

LUIS P. VILLARREAL[1]* AND VICTOR R. DEFILIPPIS[2]
Departments of Molecular Biology and Biochemistry[1] and Ecology and Evolutionary Biology,[2] University of California, Irvine, California 92697

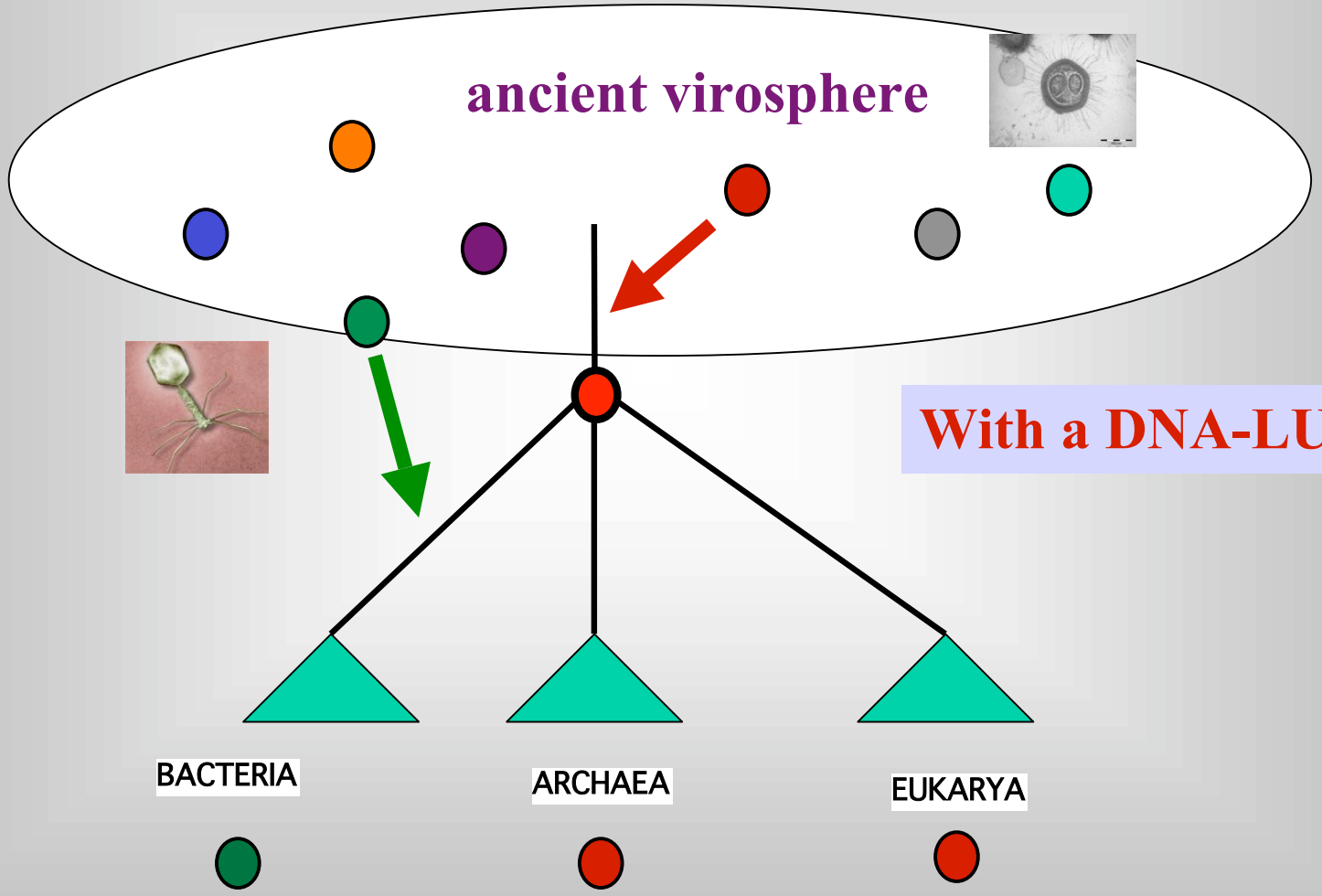
A viral origin for all DNA replication proteins?

(Current Opin Microbiol, 2002)

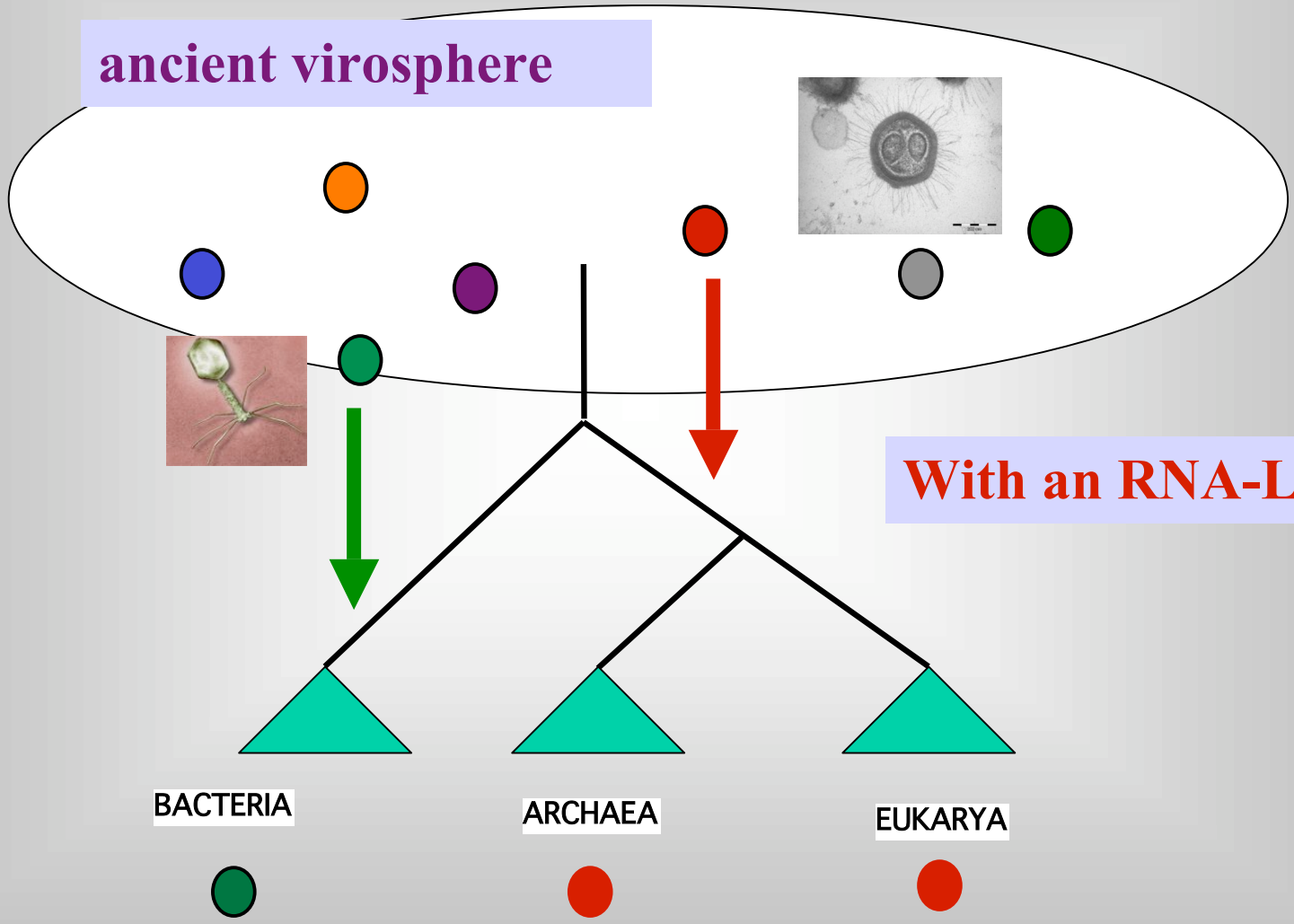
One transfer before LUCA followed by a non-orthologous replacement in bacteria

Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins

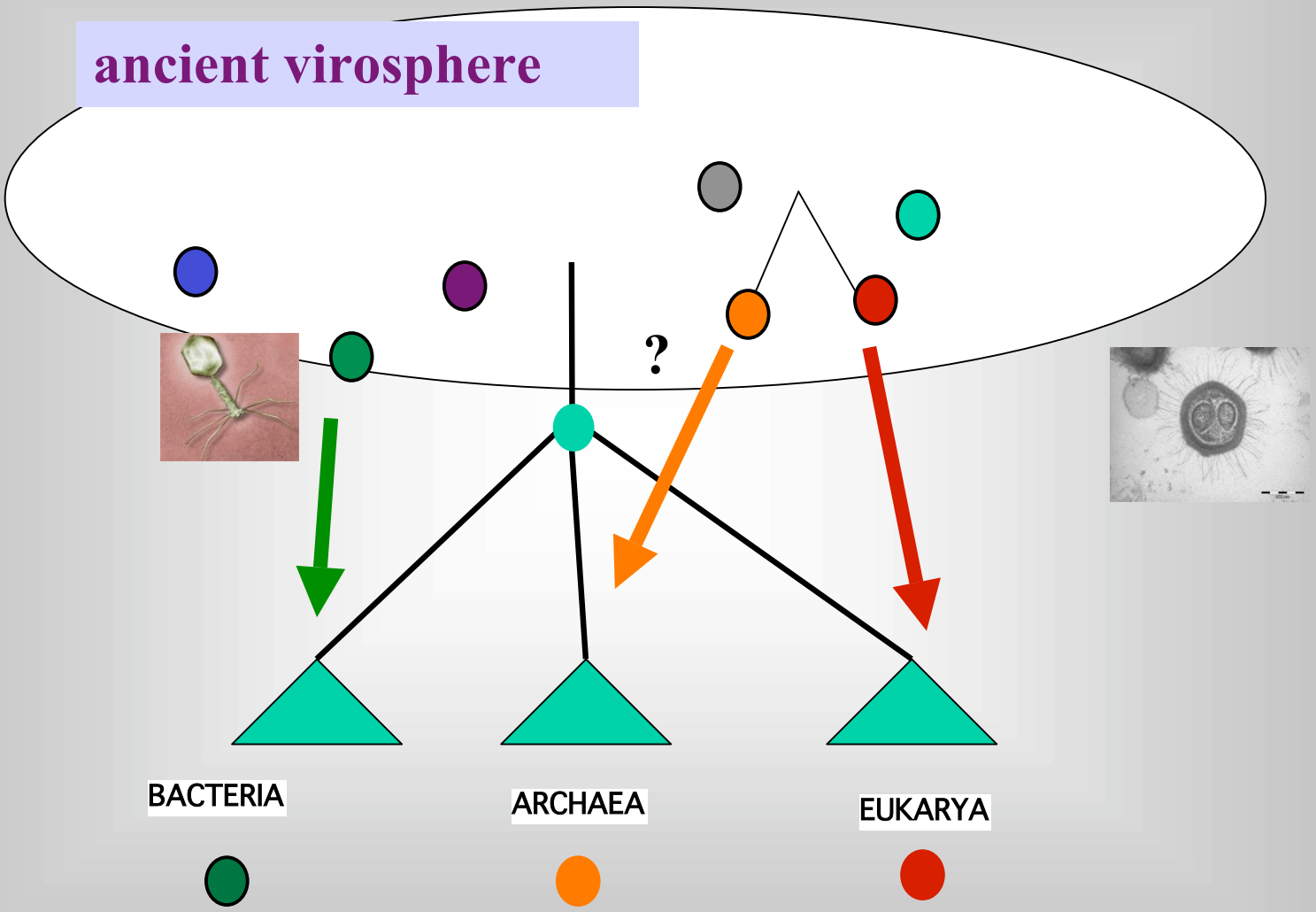
Patrick Forterre
Institut de Génétique et Microbiologie, Bat 409, CNRS, UMR 8621, Université Paris-Sud, 91405 Orsay Cedex, France.



Two times ? *(Forterre, 2002, Current Opin. Microbiol.)*



Why not three times ? *(Forterre, 2006, PNAS)*

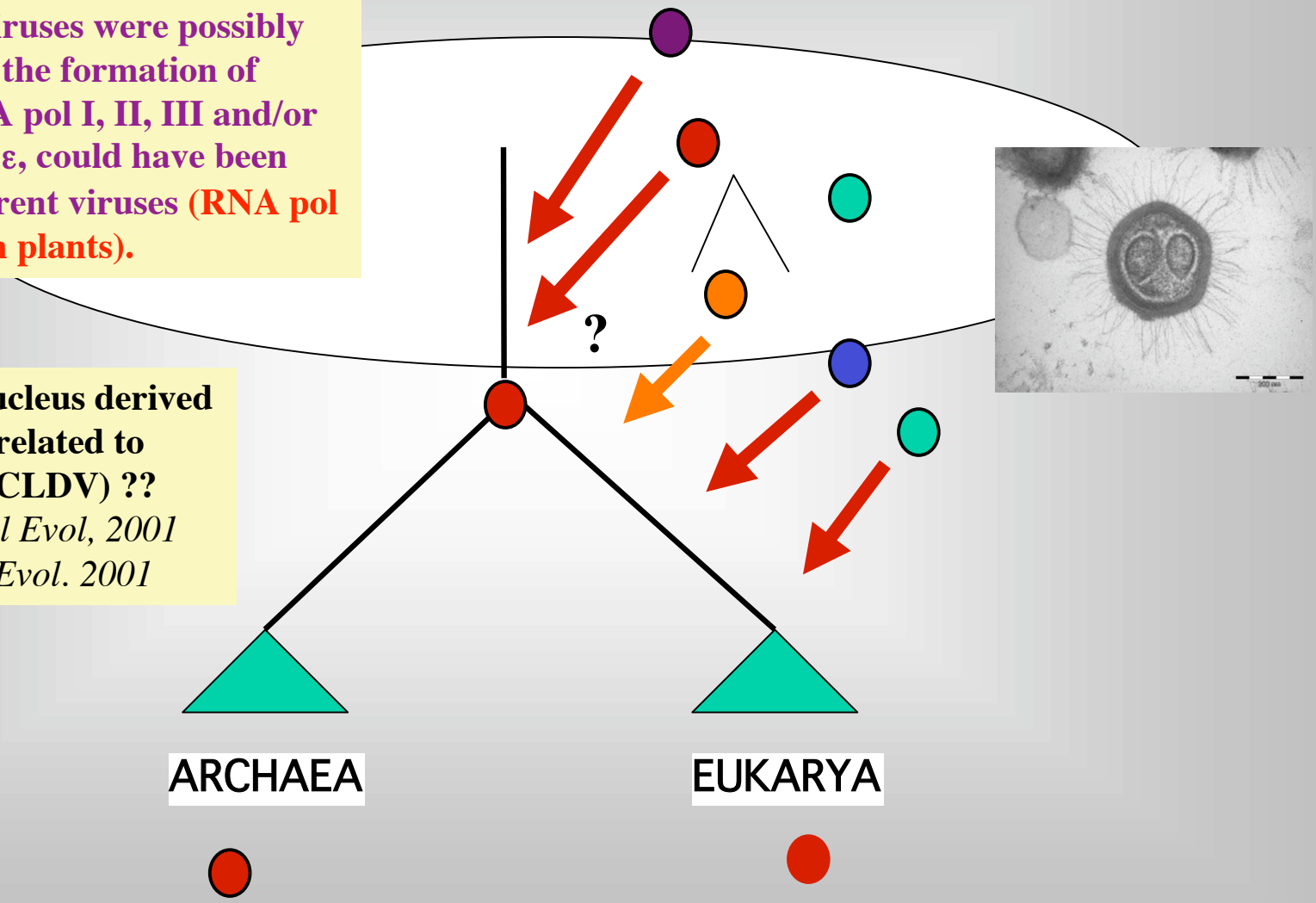


The three viruses, three domains theory

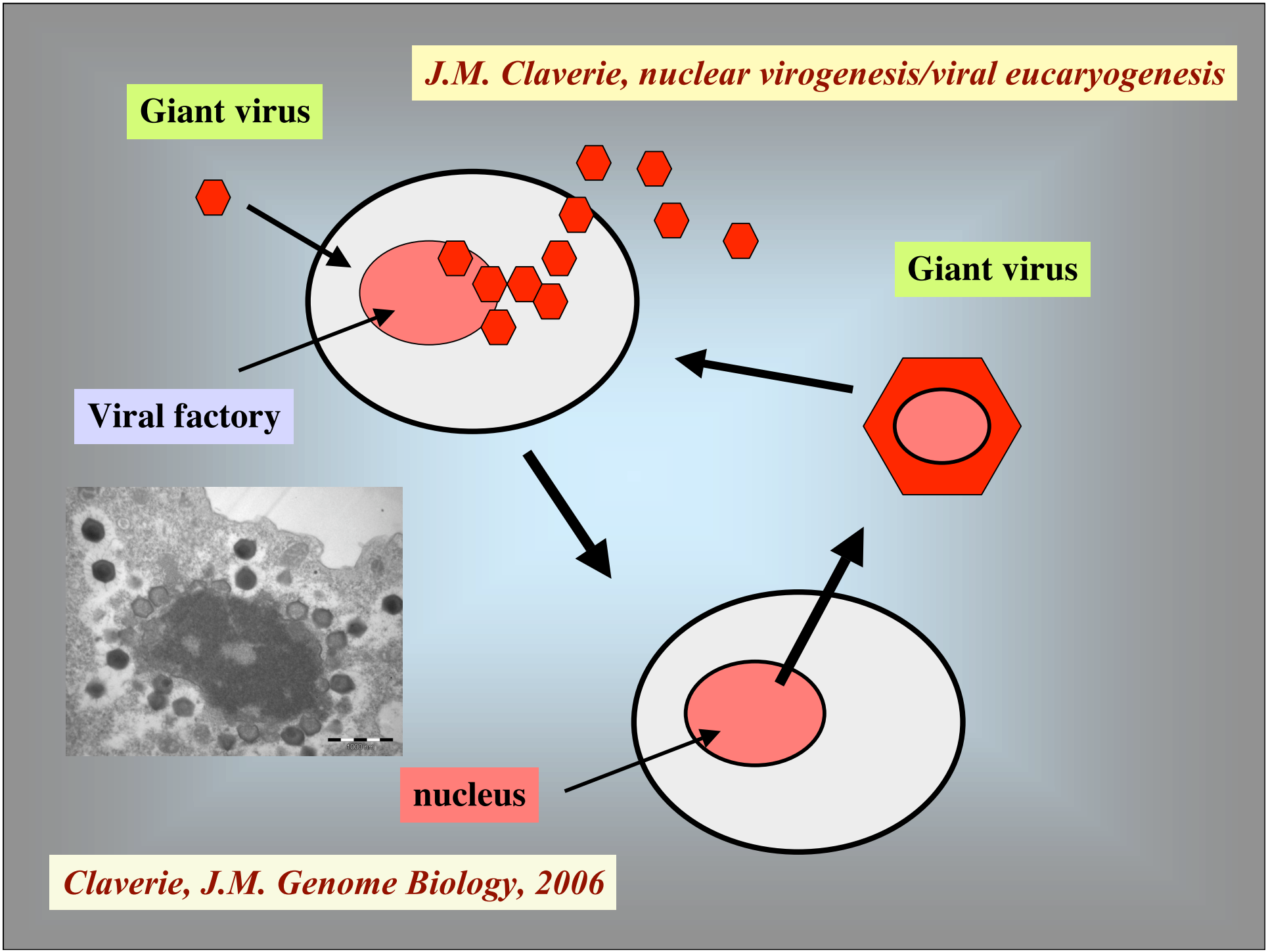
Several viruses at the origin of the eukaryotic nucleus?

Several DNA viruses were possibly involved for the formation of eucaryotes (RNA pol I, II, III and/or DNA pol α, δ, ε, could have been provided by different viruses (RNA pol IV in plants).

The eucaryotic nucleus derived from a virus related to poxviruses (NCLDV) ??
Takemura, J Mol Evol, 2001
Bell, LJ. Mol. Evol. 2001

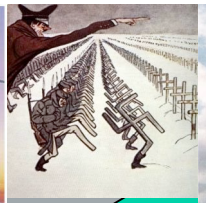


J.M. Claverie, nuclear virogenesis/viral eucaryogenesis

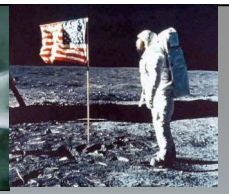


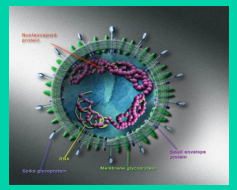
Claverie, J.M. Genome Biology, 2006












DNA

Placenta (Syncitin)

Cellular 3R mechanisms

The three cellular domains

The nucleus of Eucaryotic cells

Mitochondrial replication and transcription Machineries

Formation of cell walls (bacteria, some archaea)

Formation of transport systems (bacteria)

Formation of regulatory systems (RNAi) (eukaryotes)

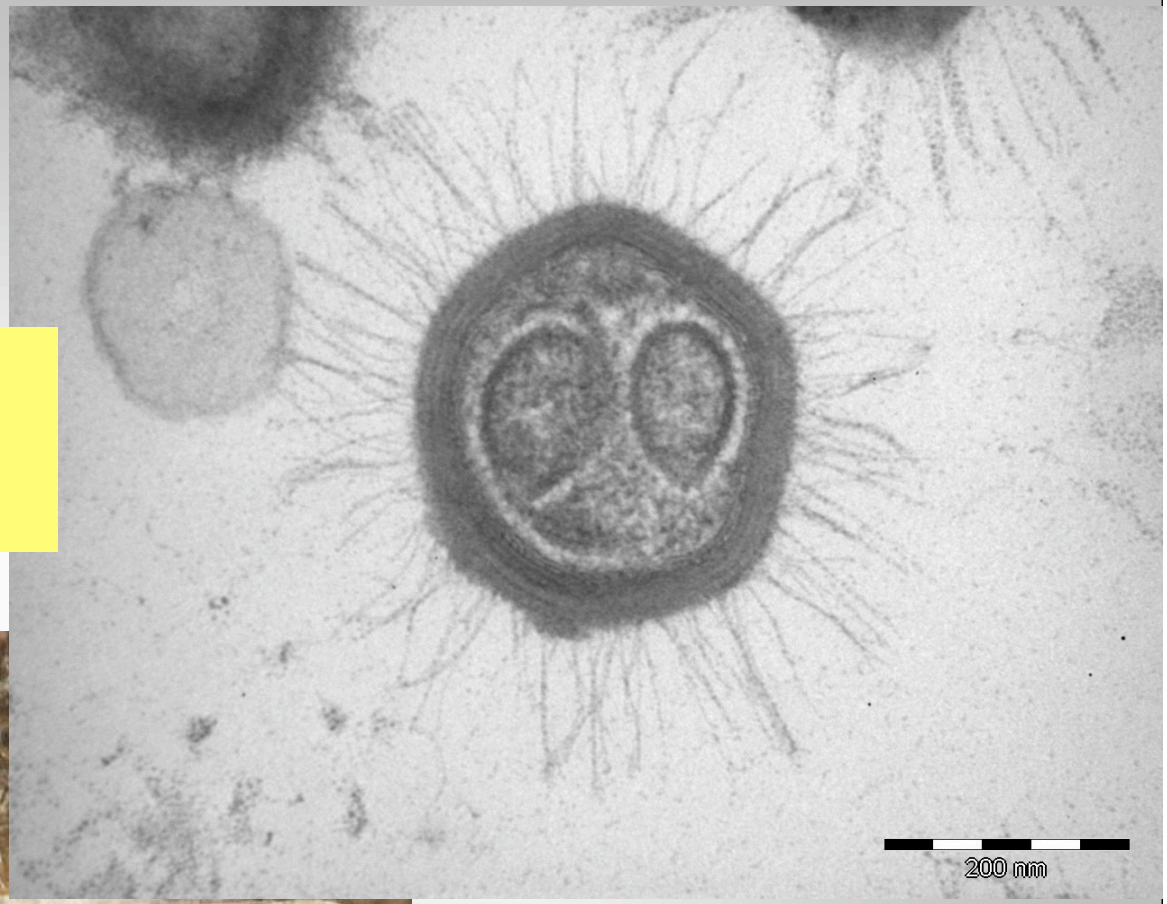
Formation of the immune system (eukaryotes)

Capping of mRNA (eukaryotes)

The spliceosome??

complexity

Viruses instead of God?



Deus ex machina?

viruses are living organisms

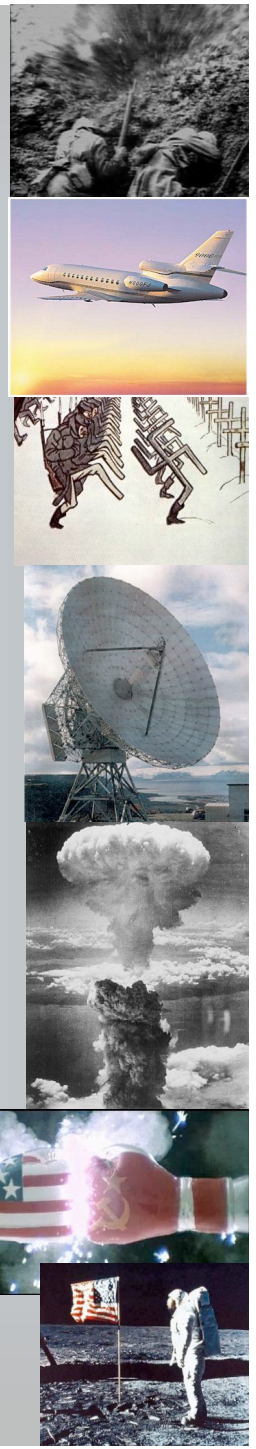
viruses are more abundant than cells

viruses are more diverse than cells

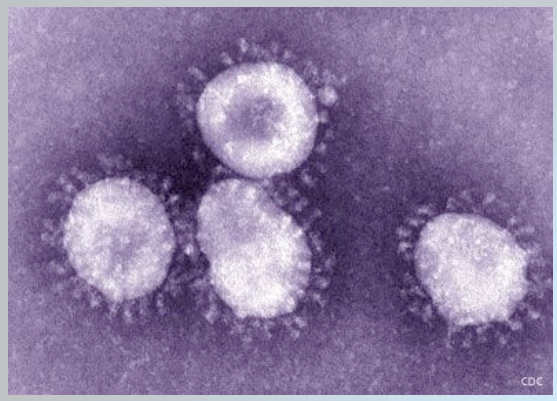
many cellular genes are of viral origin

viruses are ancient

Conclusion:the billions years war between cells and viruses has been the major factor of natural genome formation, evolution, editing, manpulation, and so on……



Can we imagine relatively complex RNA cells?



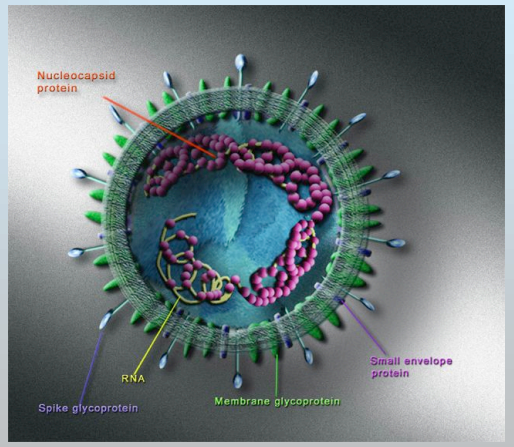
RNA viruses are usually underestimated.
Proteins produced by RNA viruses are as complex as those produced by DNA viruses/cells

RNA viruses recruit RE membranes to form complex viral factories

RNA can be repaired and replicated faithfully

Coronaviruses (30 kb) encodes both an RNA primase and a faithfull RNA polymerase using a 3' to 5' proofreading activity

Eckerle et al., J. Virol, 2007,
Imbert et al., EMBO J, 2006



Some RNA viruses (arenavirus) harbor ribosomes in their capsids

Reminder: all human proteins are encoded by RNA « genes » carry on by the messenger RNA!