



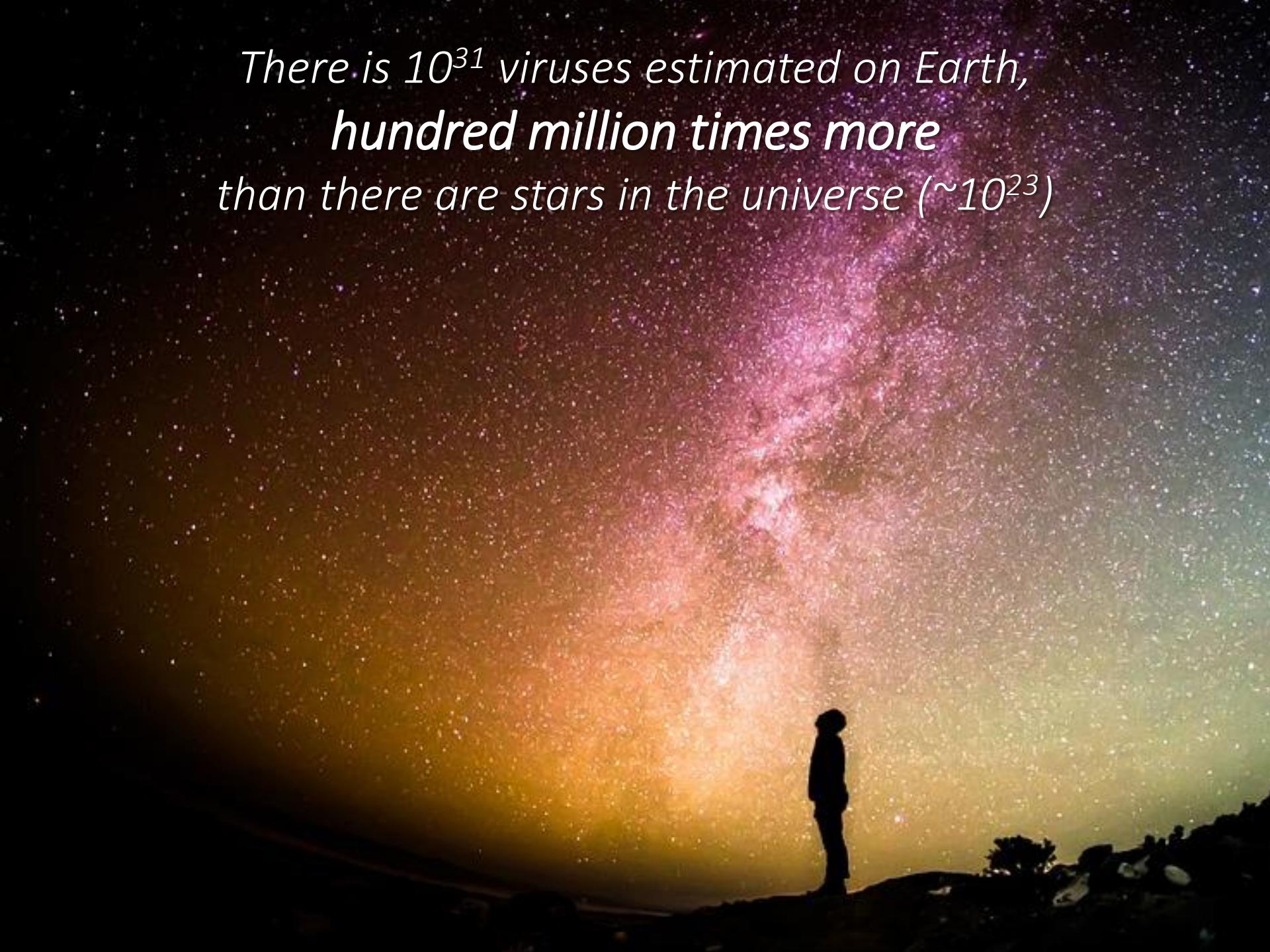
# *Découverte de nouveaux virus : l'apport des données de séquençage haut débit d'animaux non modèles pour comprendre l'évolution virale*

---

Philippe Gayral

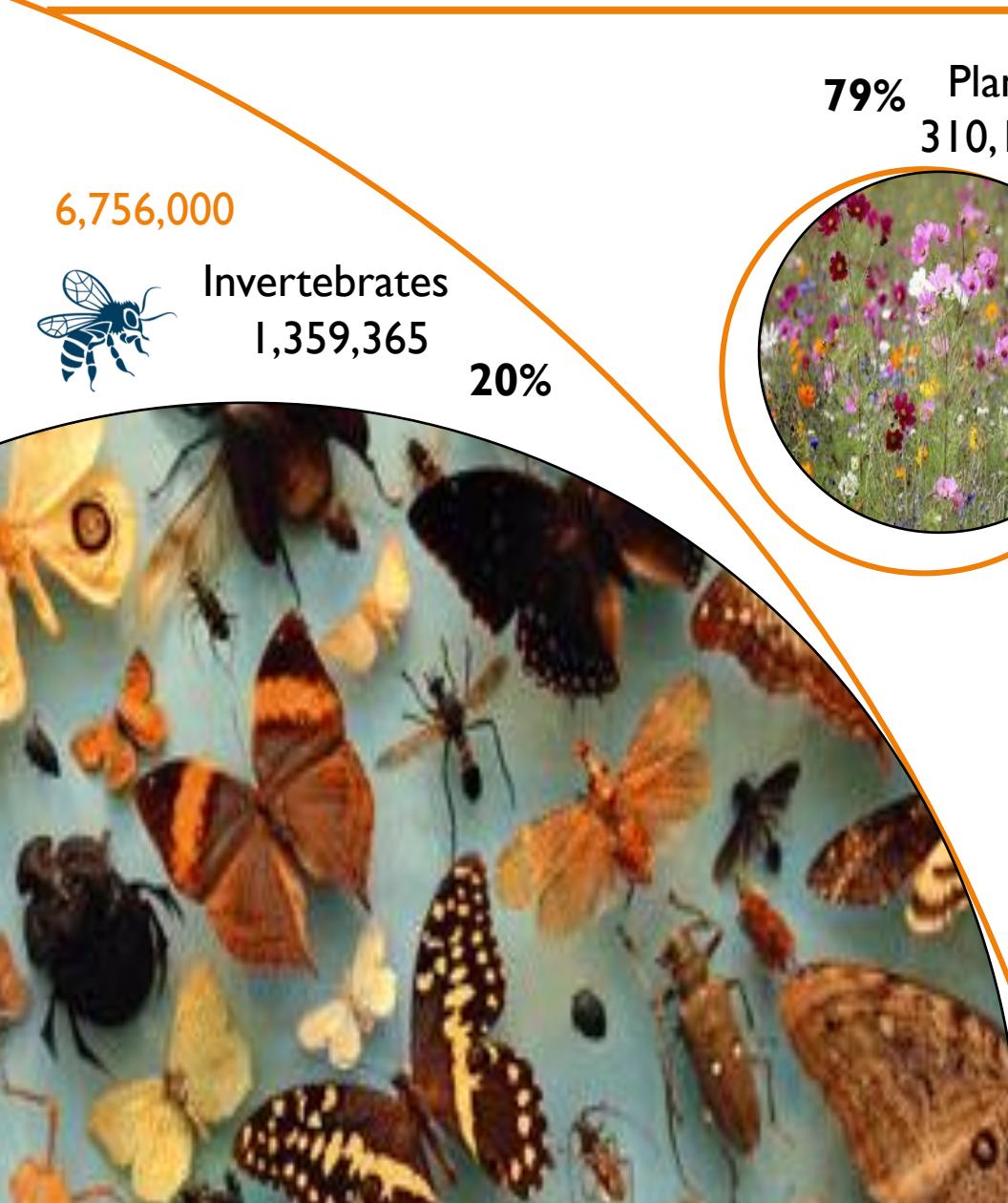
INSTITUT DE RECHERCHE SUR LA BIOLOGIE DE L'INSECTE  
UMR 7261 CNRS-Université François-Rabelais de Tours

*There is  $10^{31}$  viruses estimated on Earth,  
hundred million times more  
than there are stars in the universe ( $\sim 10^{23}$ )*



# World diversity

largely  
unknown



6,756,000



Invertebrates  
1,359,365

20%

79% Plants

310,129



390,800

7%

Fungi

98,998

1,500,000

Estimated



Chordata

80,500

64,788

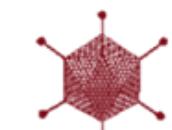


Bacteria

7,643

1,000,000

# of species



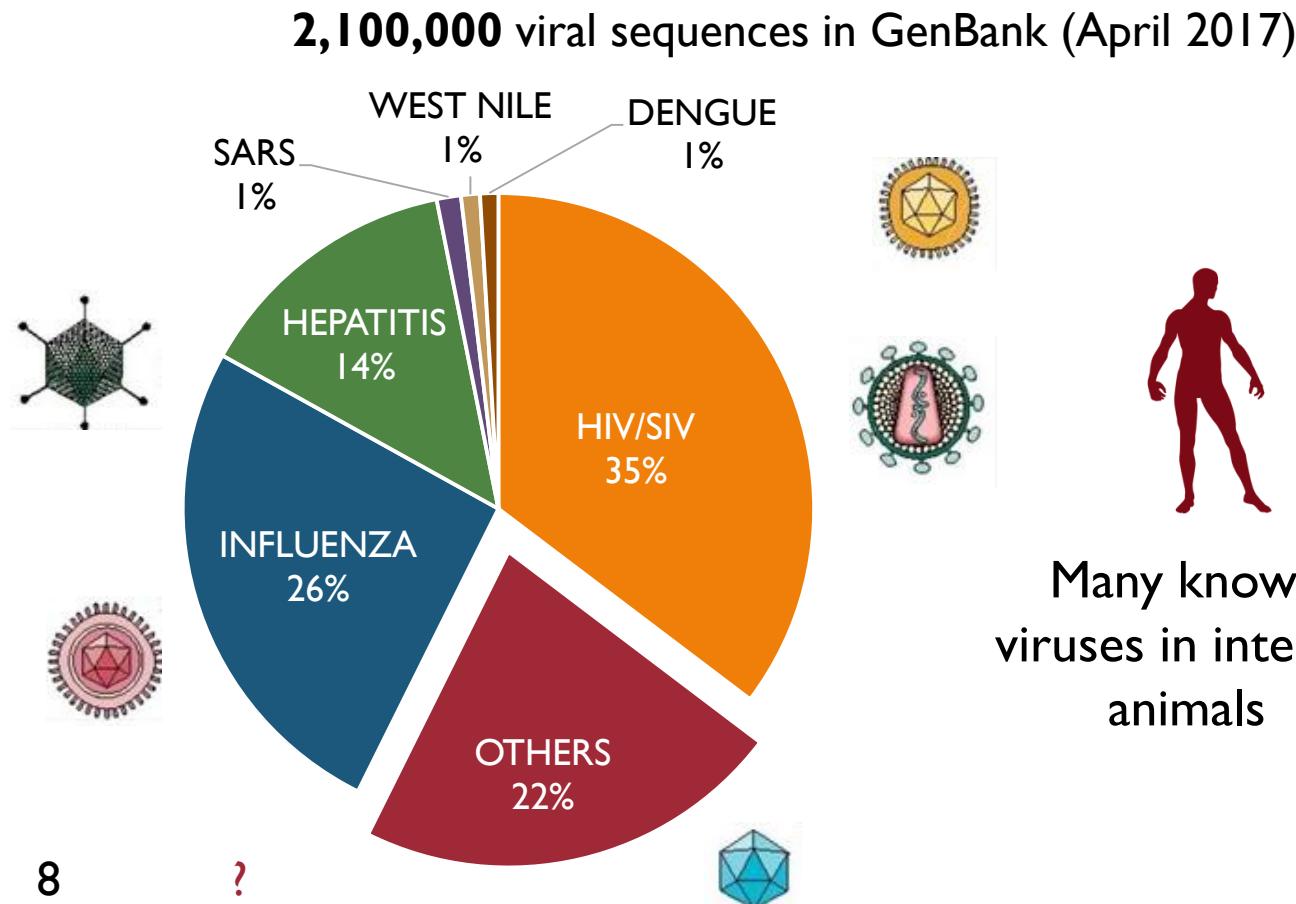
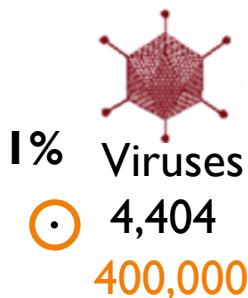
1%

Viruses

4,404

400,000

# Biased knowledges of viral diversity



Order	8	?
Family	122	?
Genus	735	?
Species	4,404	400,000



Many known viruses in interest animals



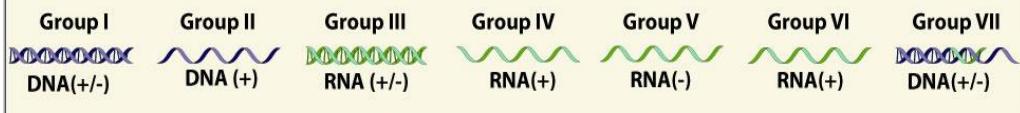
Viral biodiversity relatively unknown

# Virus definition

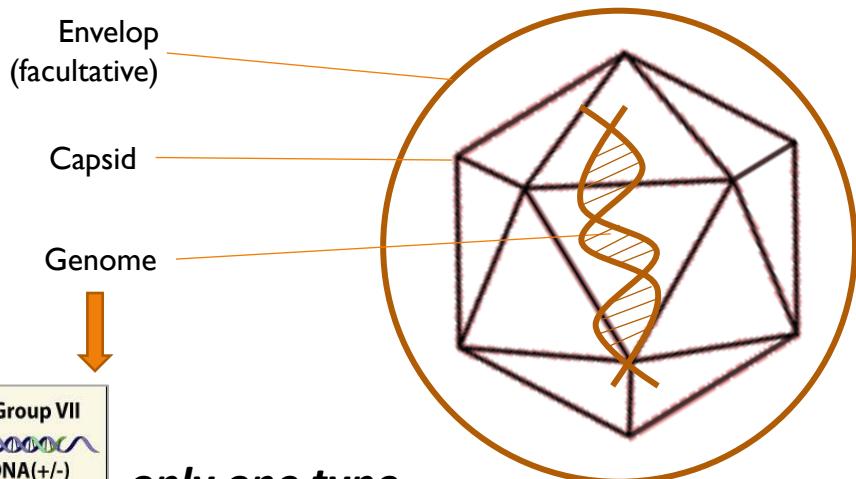
*"Viruses should be considered as viruses because viruses are viruses."*

Lwoff, 1957, J Gen Microb

## Virus typical structure



7 classes Baltimore classification (1971)



***only one type  
transmissible form***

**strictly intracellular and potentially pathogenic entities with an infectious phase**

**unable to grow and to undergo binary fission**

→ Different from bacteria

→ Use cellular components  
(e.g. ribosomes & energy)

# Ubiquitous viruses

All type of environment

- Marin
- Terrestrial

Extreme conditions

- pH
- Temperature
- Pression
- Salinity ...

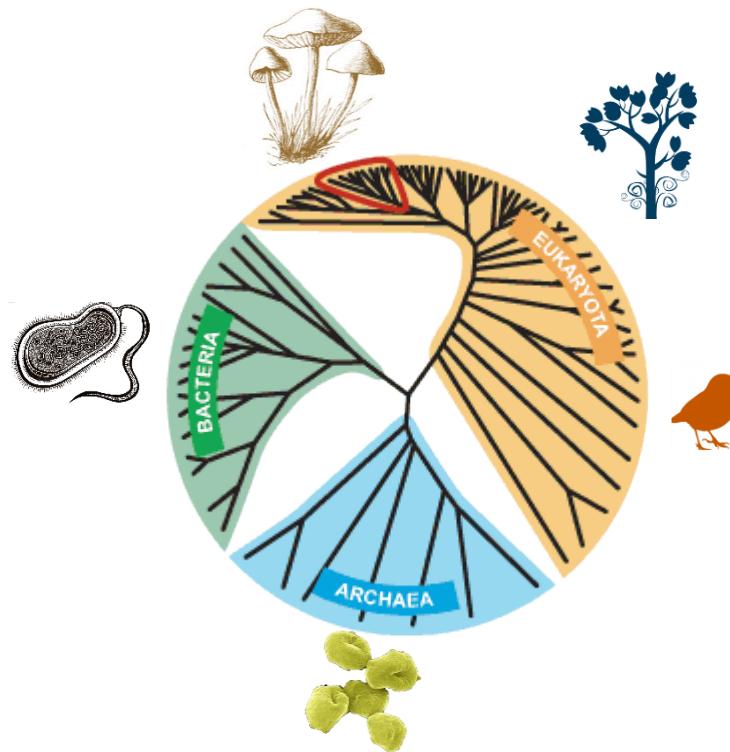
Santos, F., et al, 2011, ISME

Prangishvili, D, 2003, Res Microbiol

Boujelben, et al, 2012, Appl Environ Microbiol

Santos, et al, 2010, Environ Microbiol

Suttle, C. A., 2007, Nat Rev Microbiol

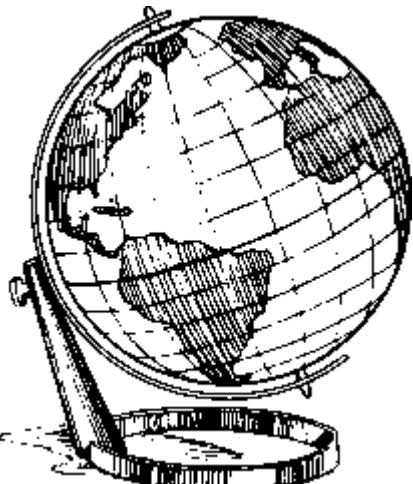


Abundant  $10^{31}$  viral particles  
on Earth

Play a major role in  
ecosystems equilibrium

(e.g. in Oceans)

Breitbart & Rohwer, 2005, Trends Microbiol  
Suttle, C. A., 2007, Nat Rev Microbiol



All type of organisms

- Archaea
- Bacteria
- Eukaryota
- Viruses

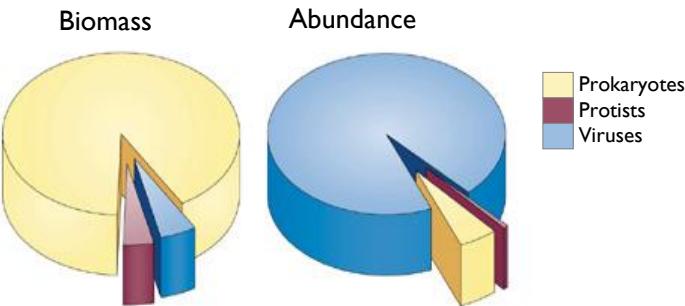
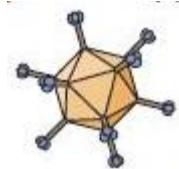
Prangishvili, 2013, Annu Rev Microbiol

Ackermann, 2003, Res Microbiol

Koonin, et al, 2015, Virology

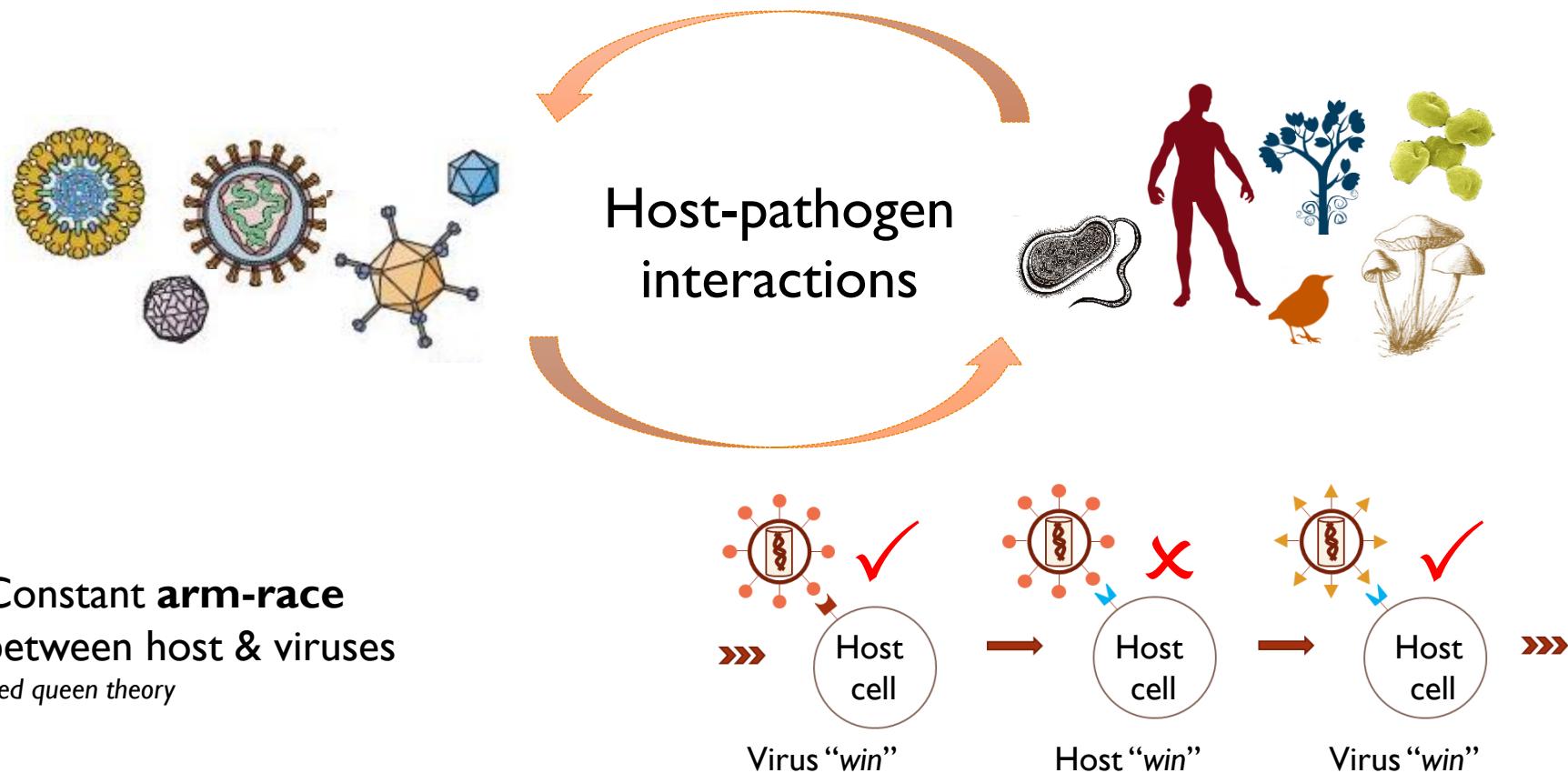
Sun , et al, 2010, J Virol

La Scola, et al, 2008, Nature



Suttle, 2007

# Long-term viral evolution



## Host-virus co-evolution

Viruses involved in **host genome evolution** & in **horizontal gene transfers**

Numerous integrated viruses

Drezen, et al, 2016, Pichon, et al, 2016  
Bézier et al, 2013, Herniou et al, 2013  
Valles et al., 2012, Wu et al., 2012  
Feschotte and Gilbert, 2012, Thézé et al., 2013

# Environmental genomics & viral diversity

**Genomics**  
1 individual / 1 species



**Environmental genomics**  
global scale (water, soil, feces ...)

## How study viral diversity ?

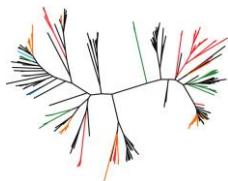


- DNA Barcoding      e.g. 16S for bacteria

NO common gene shared by all viruses !



- Metaviromic      Option choose by numerous studies



Viral diversity study with discovery of 99% unknown viruses  
Paez-Espino, et al, Nature, 2016

But NO  
information on  
each viral-host  
interactions !



Studies on groups of hosts (Arthropods)  
Viral diversity of specific hosts  
Expand of host range  
Li, et al, eLife, 2015; Shi, et al, Nature, 2016



Viral purification from pool of “things”  
e.g. TARA Oceans (20L of seawater per samples)  
Brum, et al, Science, 2015

## Our approach – Meta-transcriptomic

Transcriptomes – used to functional studies, gene expression ...

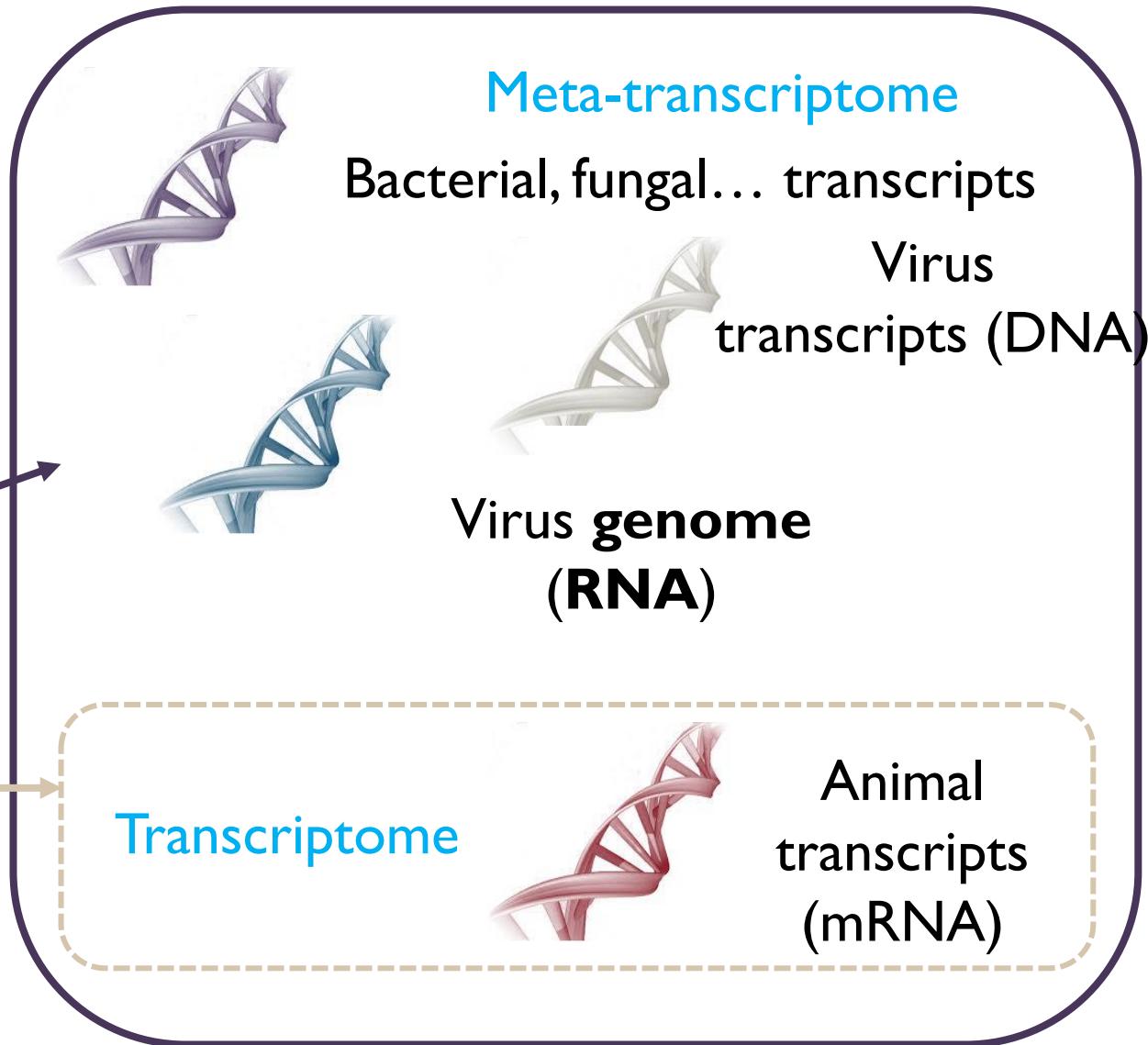
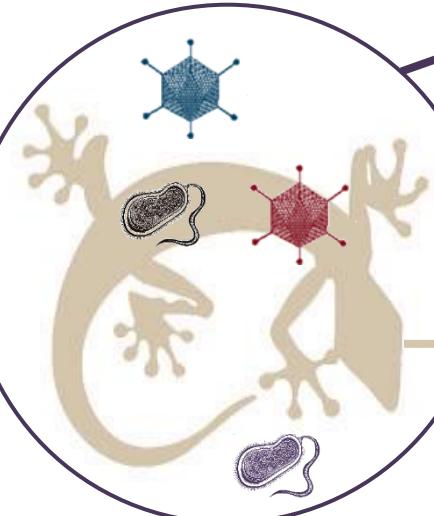
Lot of data **wasted** (~10%)

While **great** source of information !



# Meta-transcriptomic approach

Organism  
=  
Complex  
environment





# Objectives

How viral diversity is represented in animal species ?

## Part I

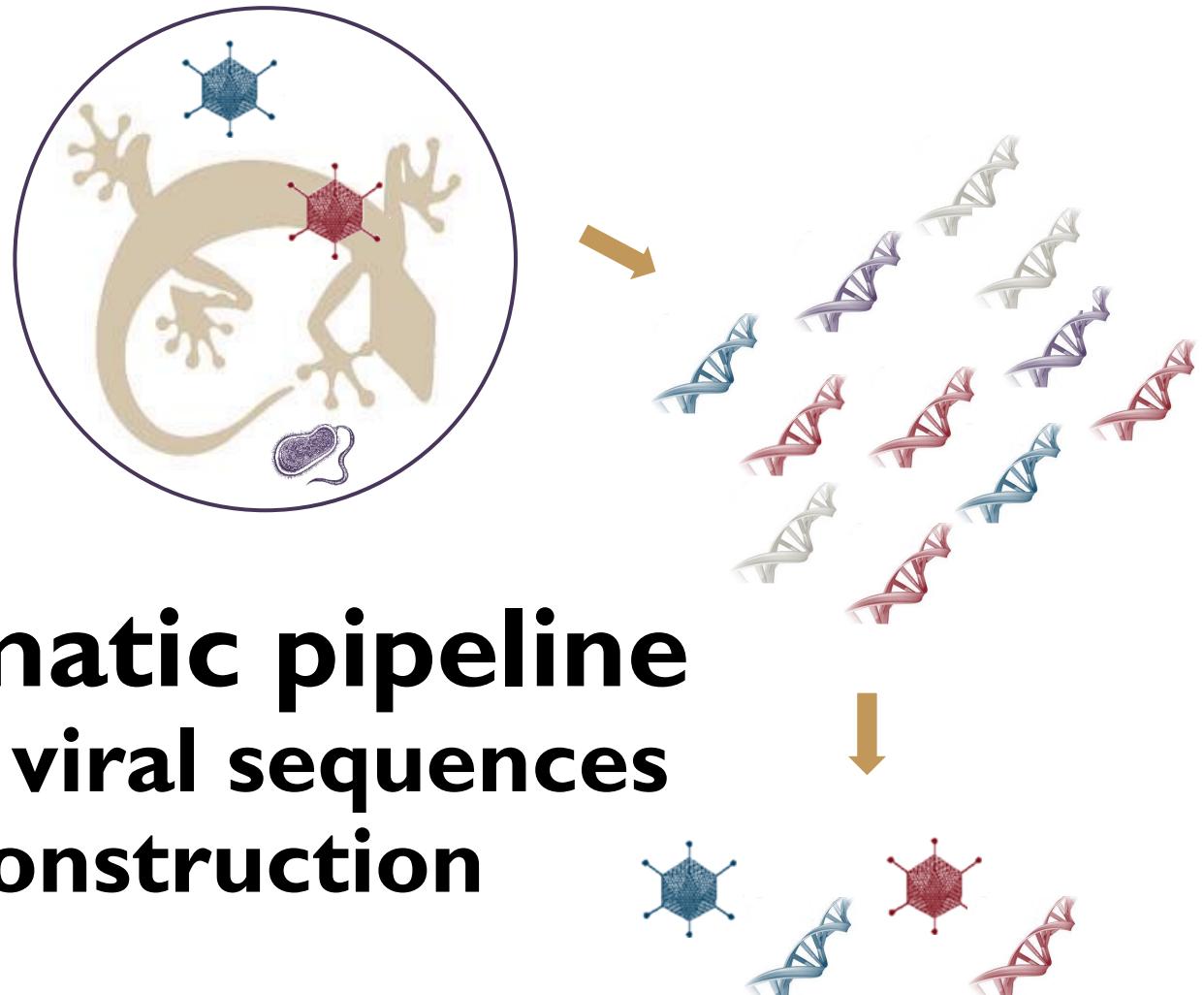
Methodological aspects

How to find viruses ?

## Part 2

Discovery of new viruses

Host shifts  
Virus evolution  
Viral systematic, taxonomy

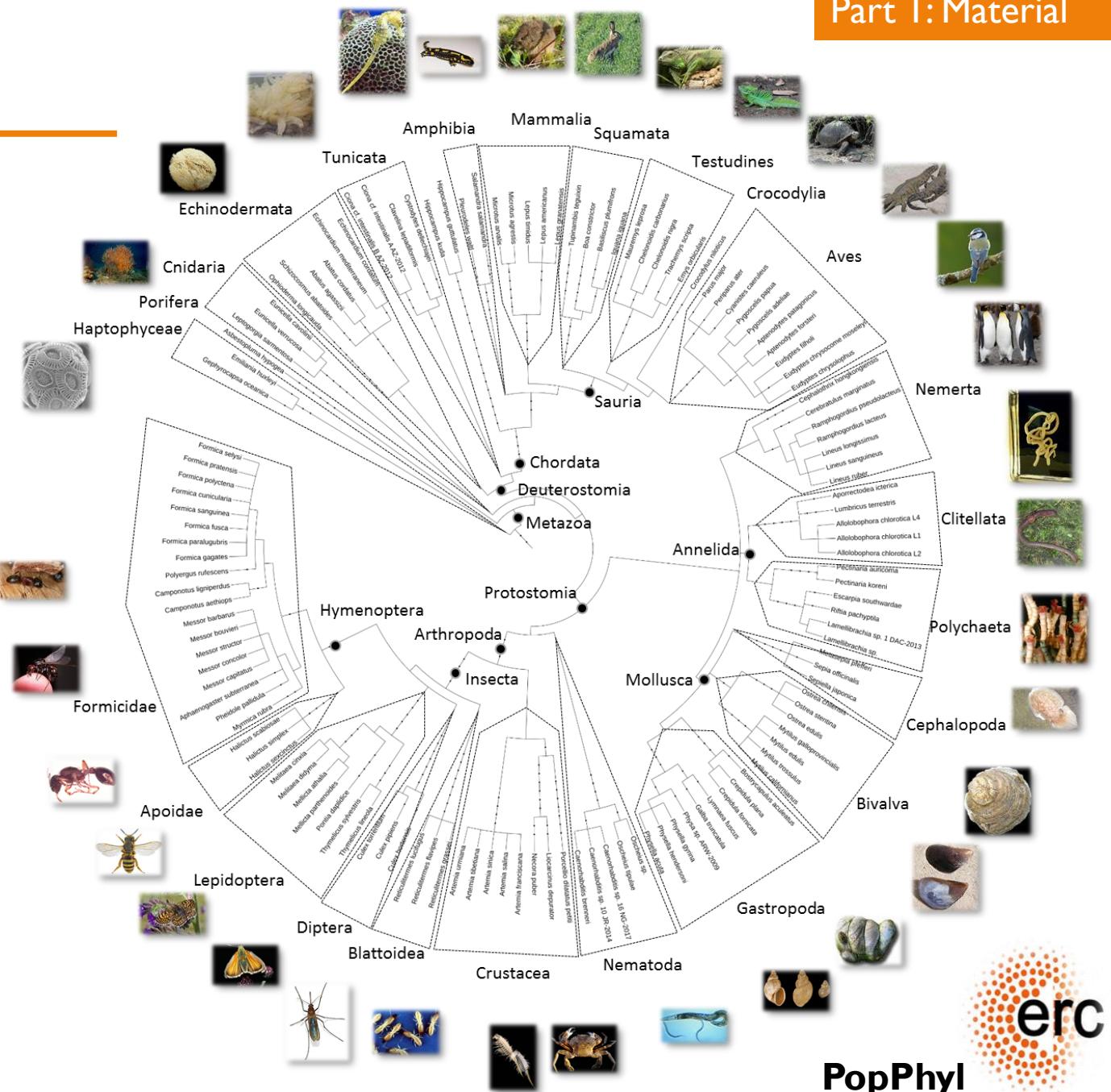


# Dataset

523 individual transcriptomes

135 non model species

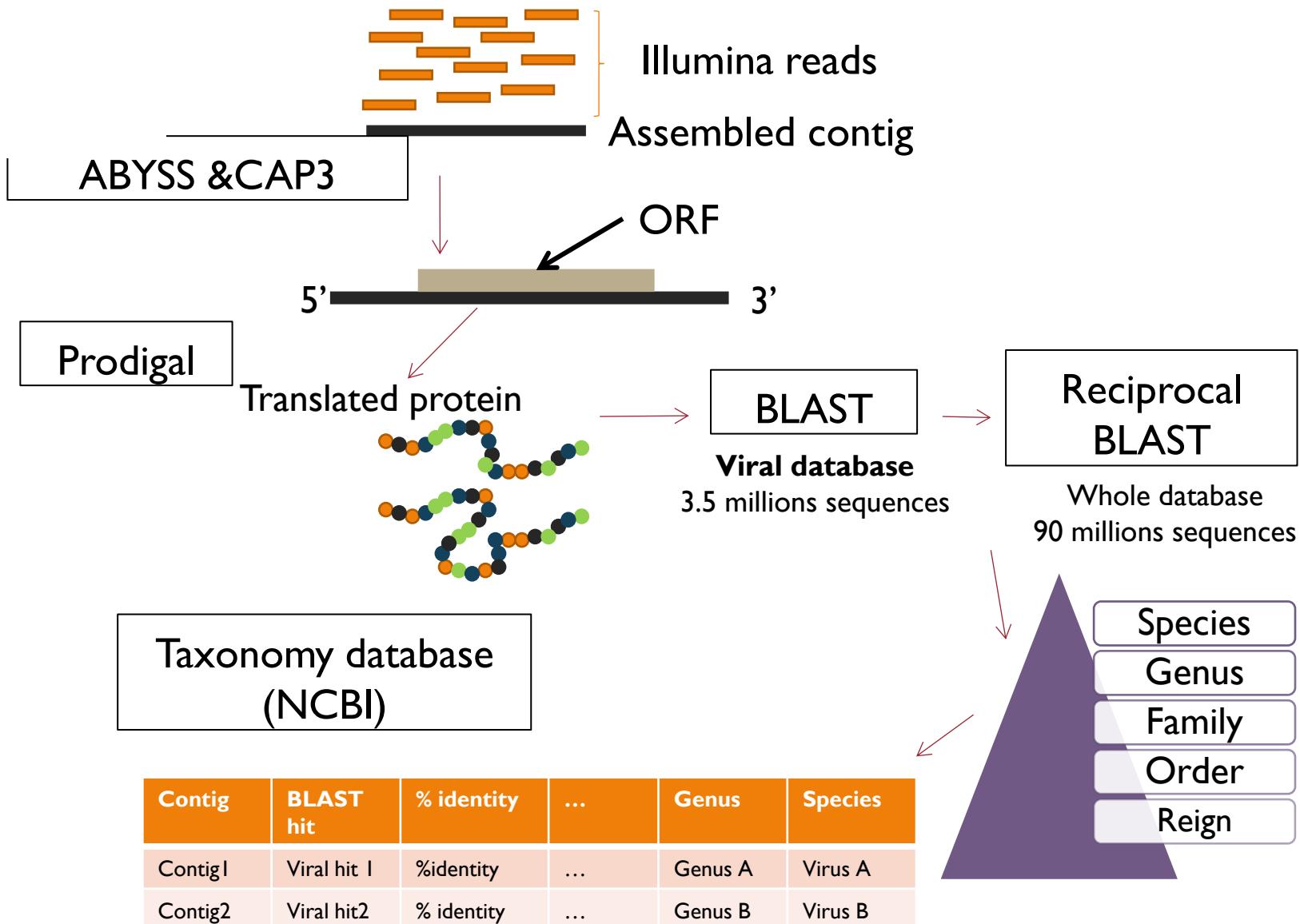
Representatives of metazoan diversity



- Gayral et al, Mol Ecol Res, 2011  
 Cahais, Gayral et al, Mol Ecol Res, 2012  
 Gayral et al, PLoS Genet, 2013  
 Romiguier et al, J Evol Biol, 2014  
 Romiguier, Gayral et al, Nature, 2014

# Pipeline 1/2 :Automatic phase

From transcriptomes to list of viral protein



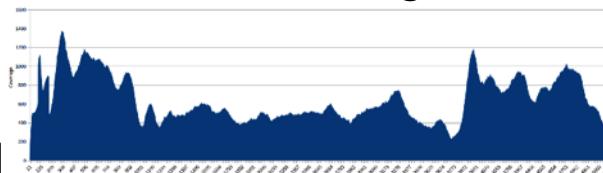
# Pipeline 2/2 : Manual phase

From list of viral protein to viral genomes and phylogenies

List of viral proteins

Mapping

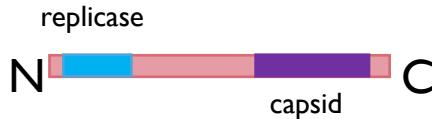
Genome coverage



Genomic organization



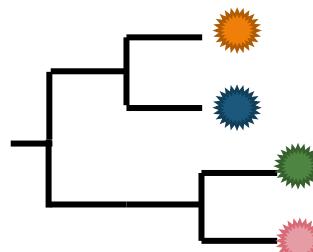
InterProScan



MAFFT  
GBlocks

ModelTest  
ProtTest

PhyML  
MrBayes



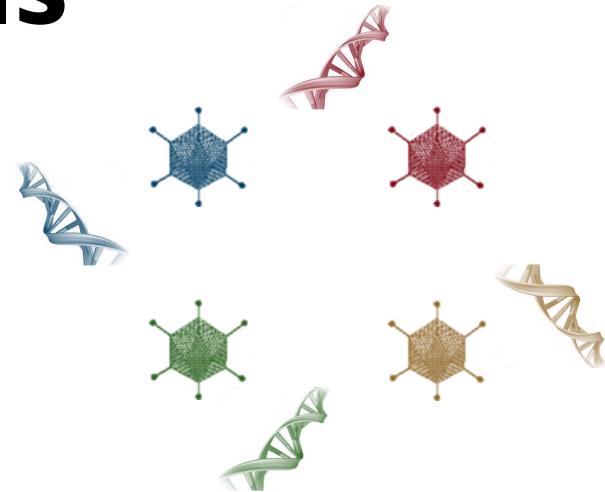
**Maximum likelihood**

Bootstrap statistics aLRT

**Bayesian method**

$10^6$  generations sampling every 100 generations

# **II Description of new viruses in animals**

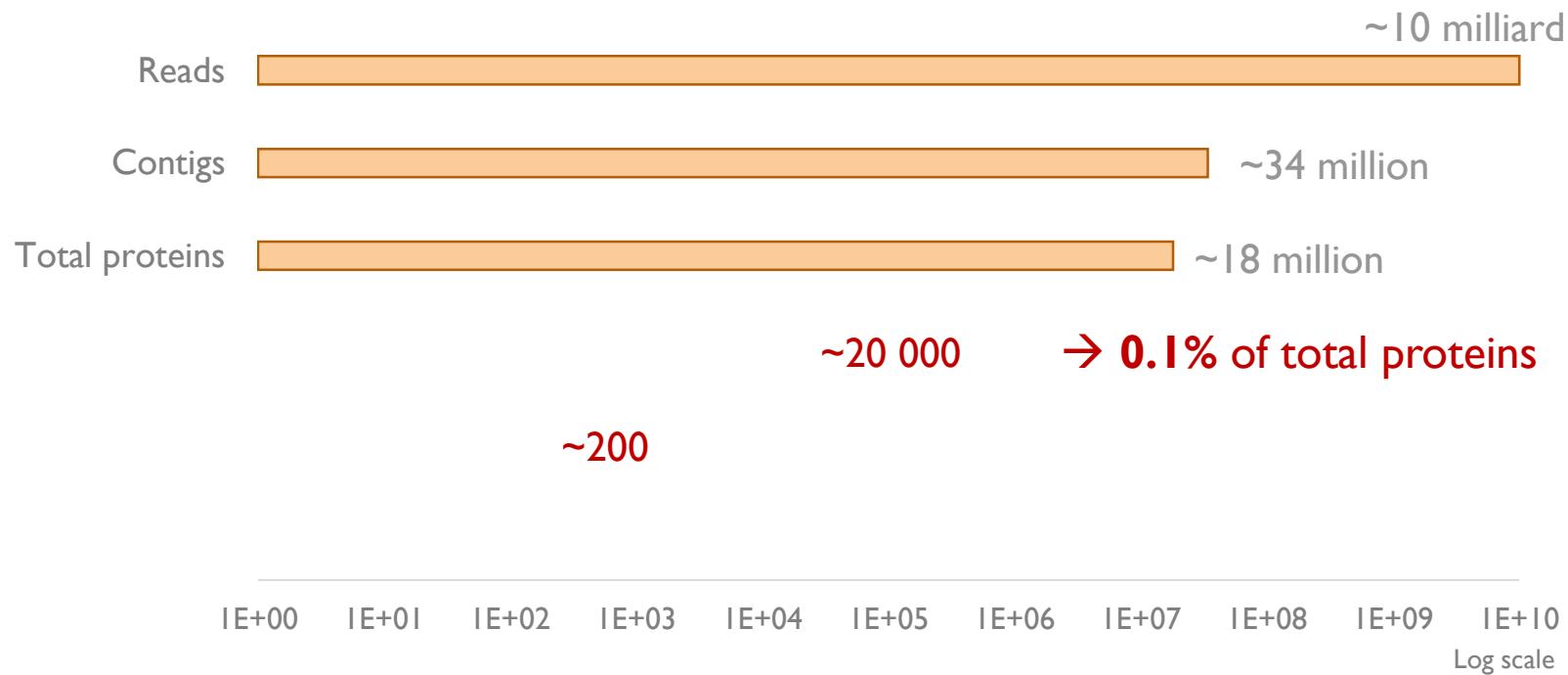




# Data

**523 individual  
transcriptomes**

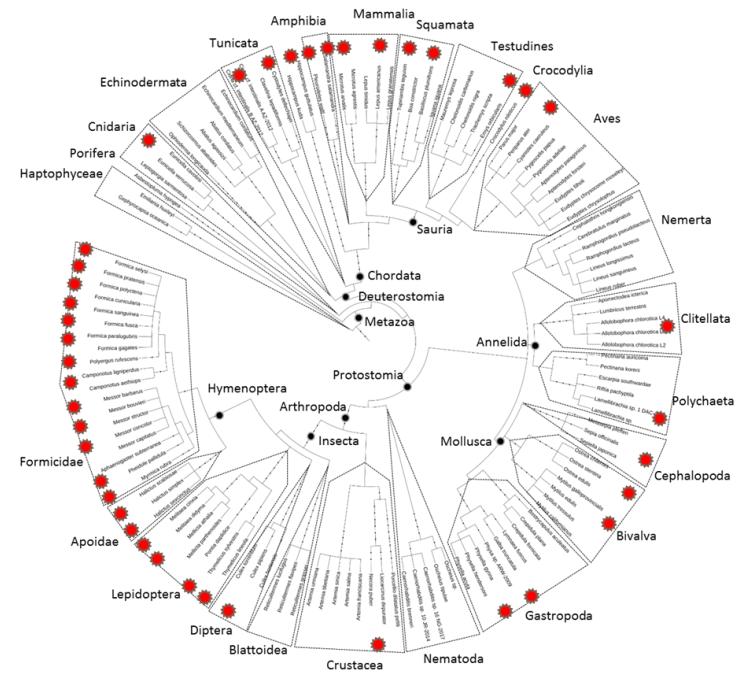
**135 non model  
species**



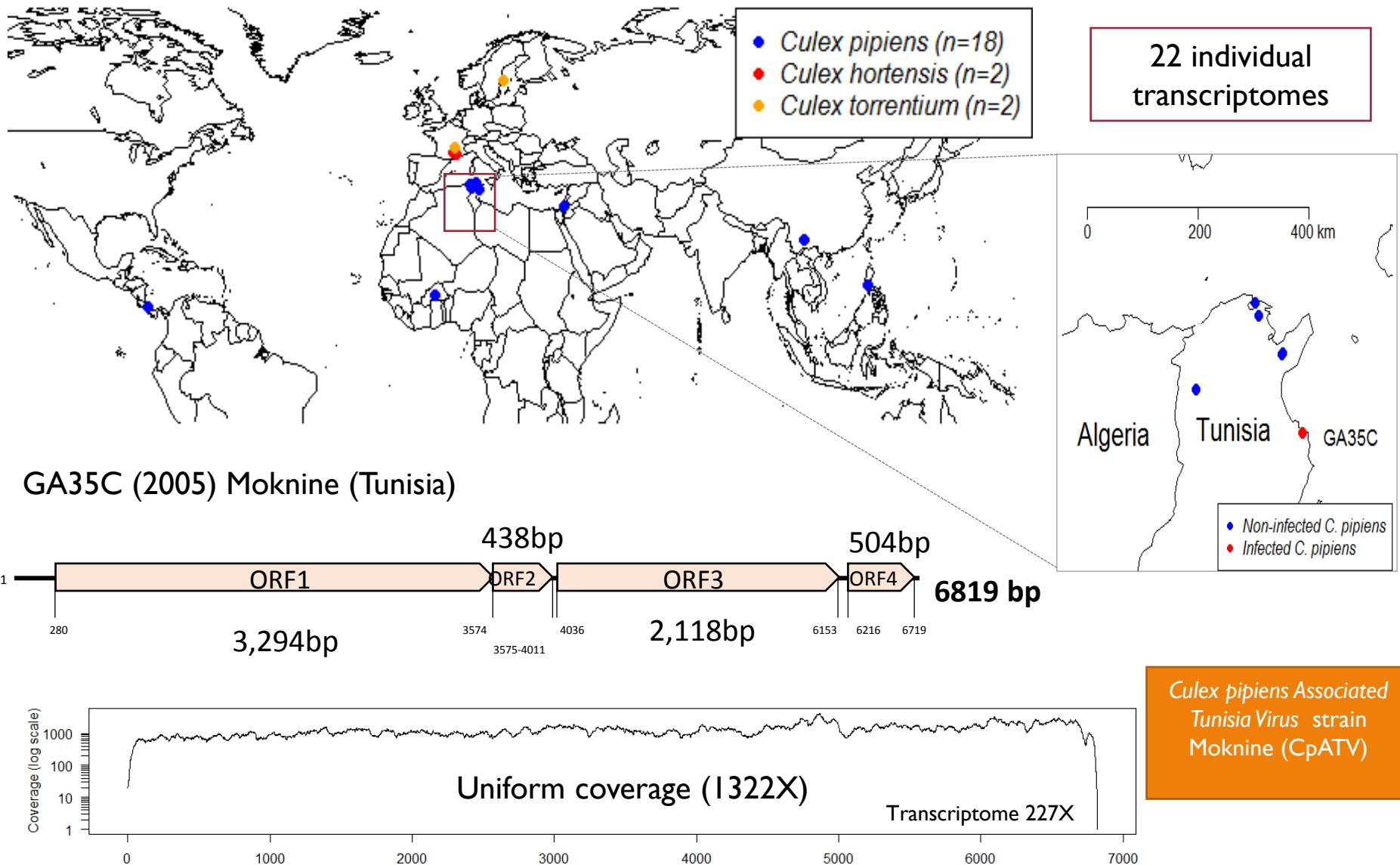
# ~50 new full-length viruses



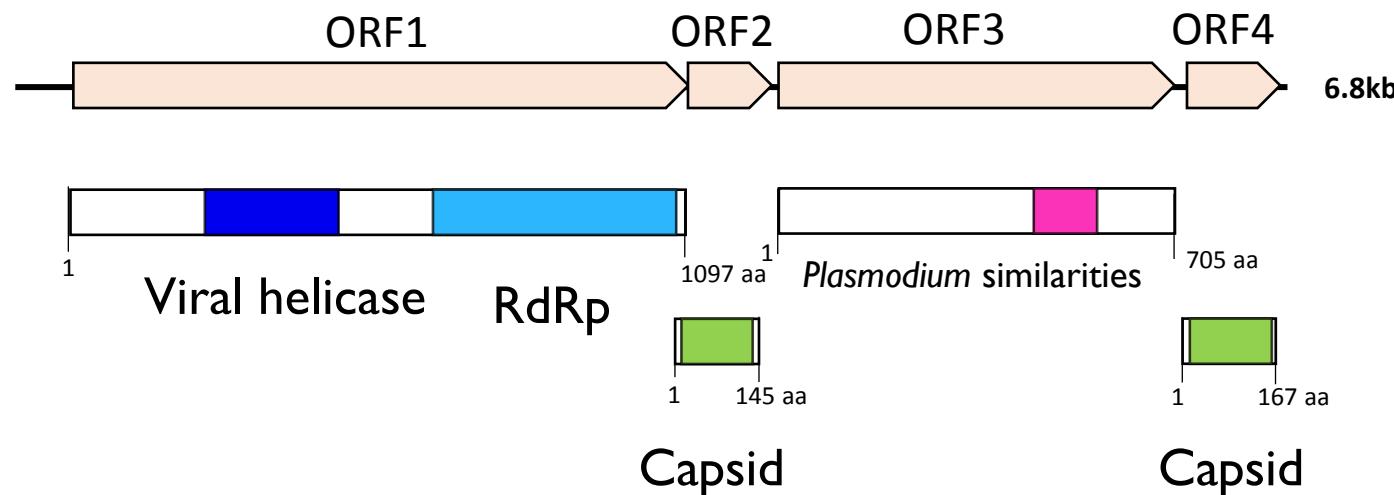
TAXONOMY	POTENTIEL HOST
Parvoviridae	Lamellibrachia
Reoviridae	Thymelicus, Myrmica, Halictus
Partitiviridae	Messor
Totiviridae	Formica
Mononegavirales	<i>Microtus, Hippocampus, Myrmica, Formica, Camponotus, Armadilidium</i>
Arenaviridae	<i>Basilicus</i>
Bunyaviridae	<i>Messor, Eunicella, Formica, Melitaea, Necora</i>
Orthomyxoviridae	<i>Hippocampus</i>
Rhabdoviridae	<i>Formica, Thymelicus, Camponotus, Culex, Sepia</i>
Unassigned	<i>Ciona</i>
Retroviridae	<i>Crocodylus / Emys / Lepus / Microtus / Tupinambis</i> <i>Pleurodeles / Salamandra</i>
Arteriviridae	<i>Microtus</i>
Coronaviridae	<i>Hippocampus</i> <i>Melitaea / Messor / Halictus</i>
Picornavirales	<i>Ostrea, Rhizaria, Formica, Myrmica, Mytilus, Galba</i> <i>Pheidole / Polyergus / Allolobophora / Cystodites / Formica</i>
Betaflexiviridae	<i>Allolobophora</i>
Benyviridae	<i>Basilicus, Parus</i>
Flaviviridae	<i>Physa</i>
Nodaviridae	<i>Melitaea</i>
Tombusviridae	<i>Physa</i>
Virga-Negevirus	<i>Culex</i>
Virgaviridae	<i>Armadilidium / Ciona / Formica</i> <i>Necora, Polyergus, Armadilidium</i>
Unassigned	<i>Halictus, Messor</i>



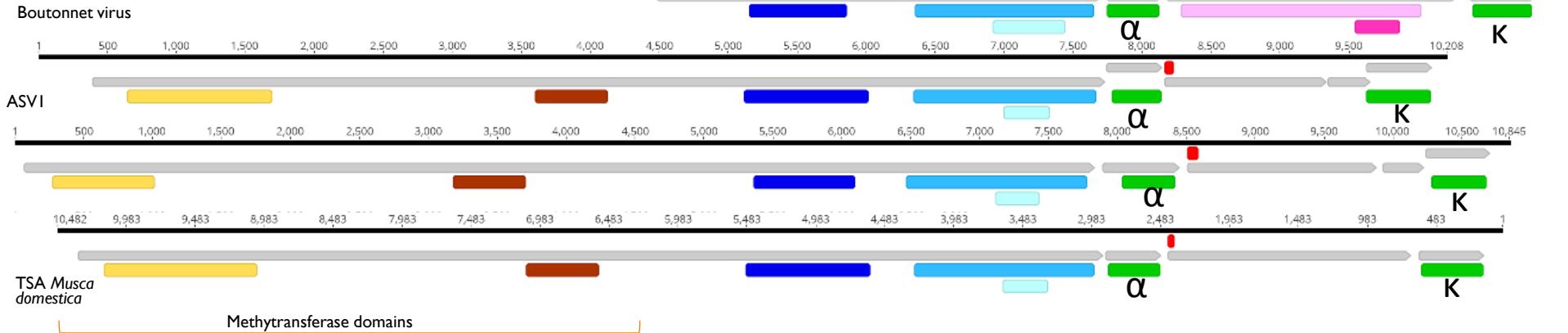
# What was found in mosquitoes ?



# CpATV genomic organization

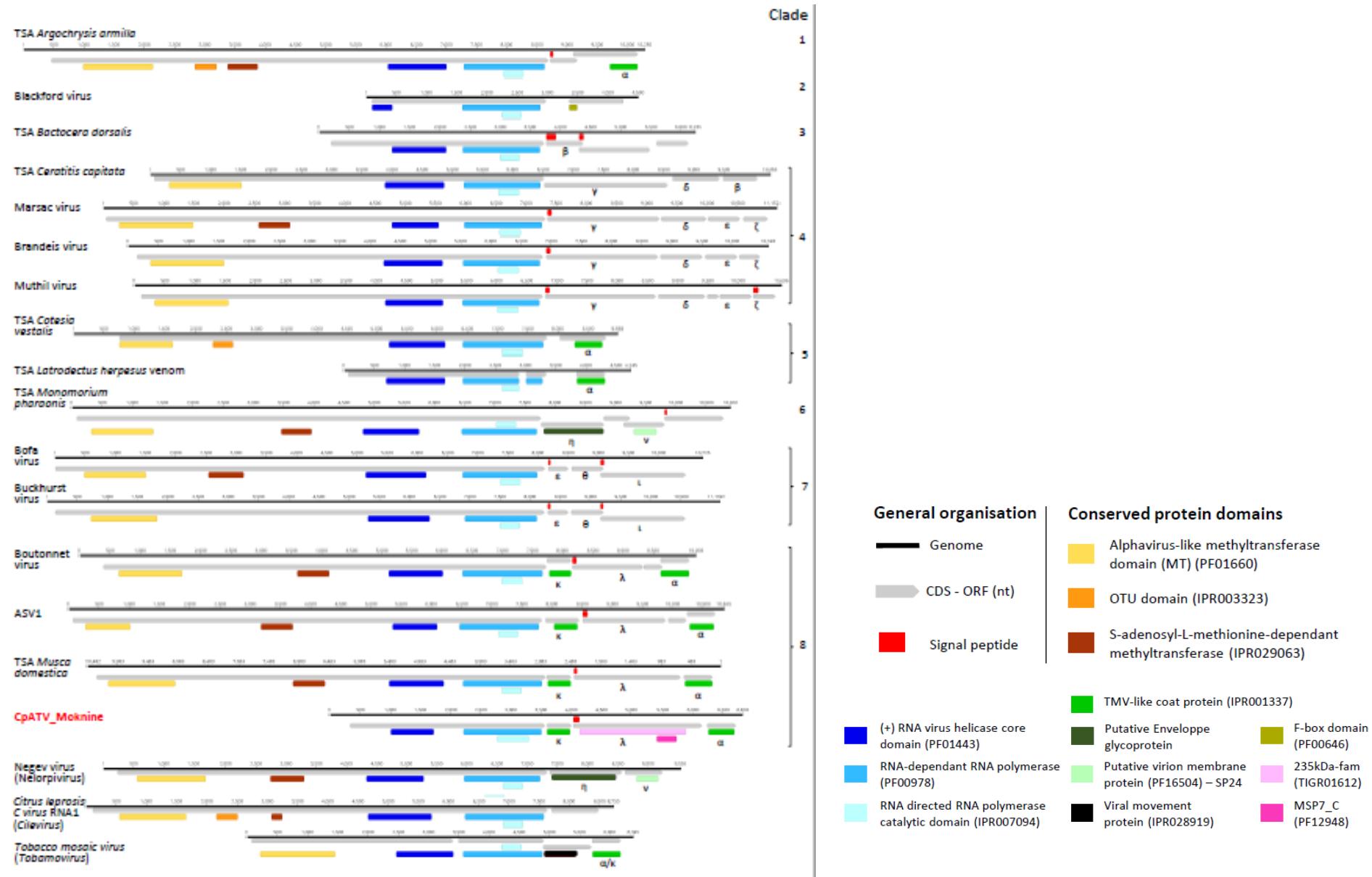


CpATV\_Moknike

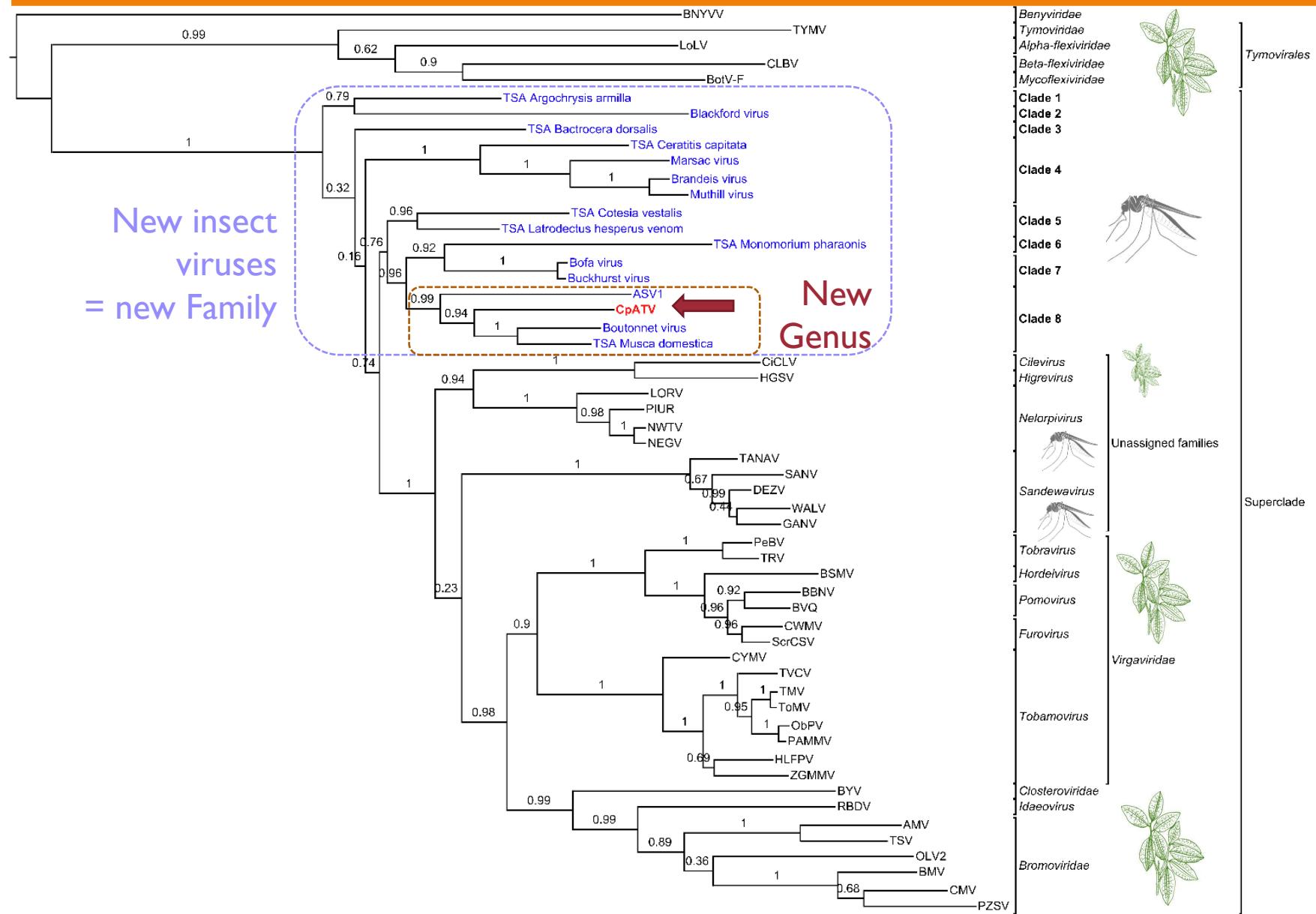


Same genomic organization as new insect viruses  
Except methyltransferase

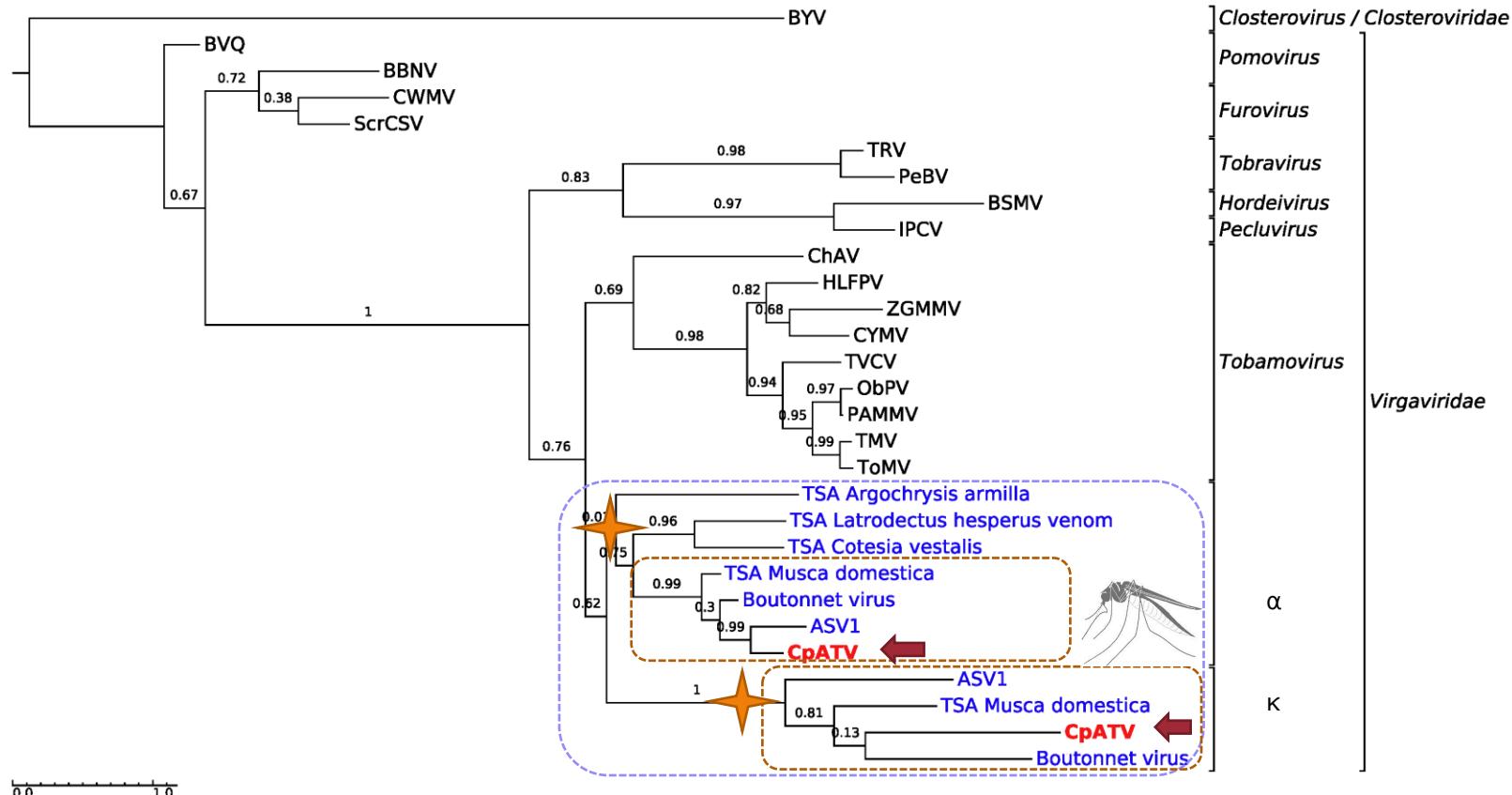
# Comparative genomics



# Replicase phylogeny



# Capsids phylogeny

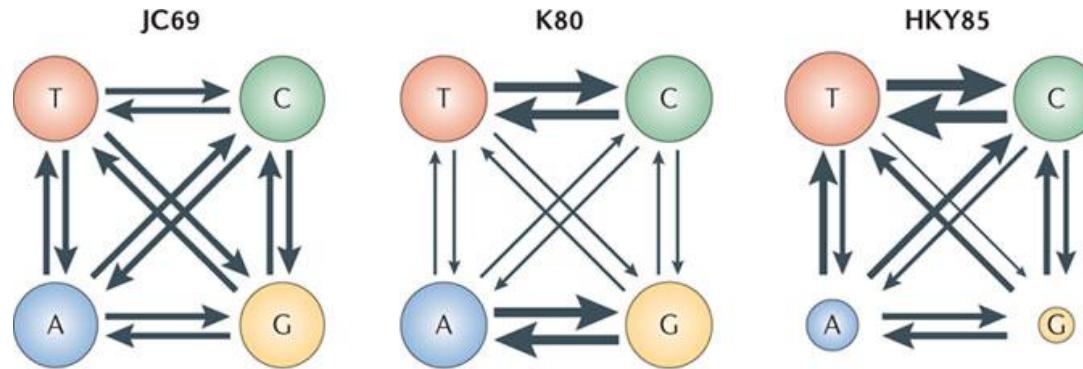


Phylogenetically closely related to plant viruses *Virgaviridae*

Origins for the two capsids :  
**two independent events of acquisition from plant viruses**

**In silico virus detection : still functional?**

# ML models characterize gene evolution



$$dN/dS = \frac{\text{non-synonymous substitution rate}}{\text{synonymous substitution rate}}$$

Nonsynonymous / Synonymous substitution

TCCGAT	ATA	TGG	CAA	CCC	GAC	AAA	
S	D	I	W	Q	P	D	K

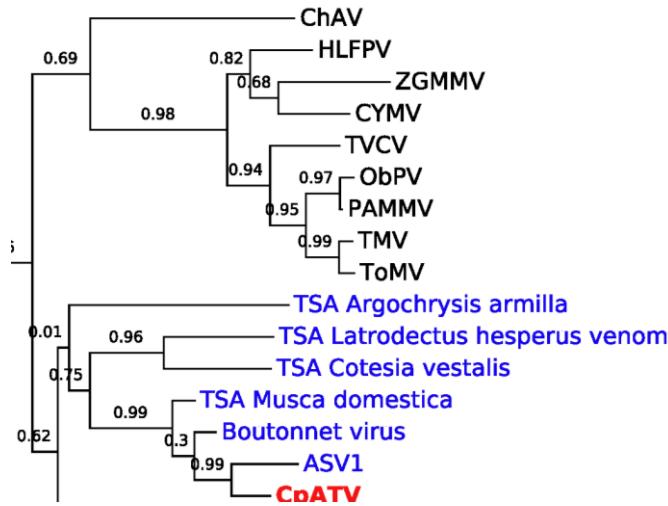
TCA	GAT	CTA	TGG	CAG	CCC	CAC	AAA
S	D	L	W	Q	P	R	K

$dN/dS < 1$  evidence of **stabilizing (=negative) selection**  
gene with (important and) **conserved function**

$dN/dS \approx 1$  evidence of **neutral evolution**  
gene with **no function** (loss)

$dN/dS > 1$  evidence of **positive selection**  
gene with important function, new variants are selected for = **adaptation**

# ML models characterize gene evolution : CpATV



- **Nested models** assuming one or several dN/dS ratio for the branches of interest
- Compared with **Likelihood Ratio Tests**

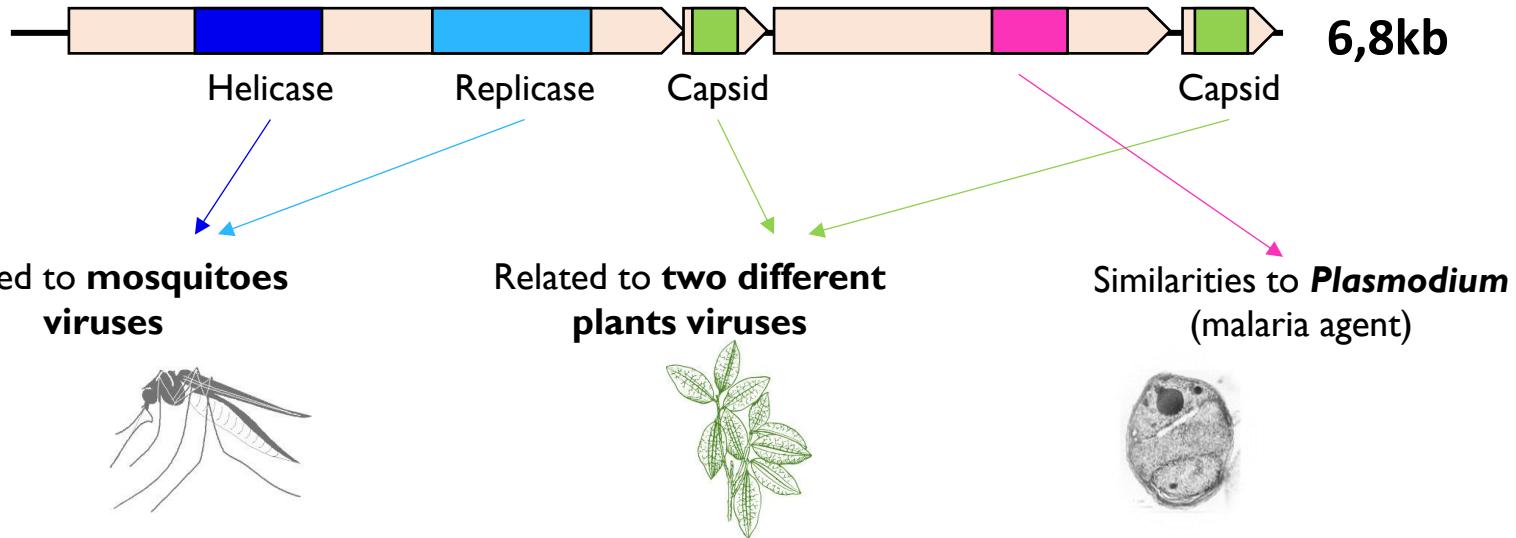
	<b>ORF</b> <b>Branch or subtree of interest</b>	<b>Helicase ORFI</b>	<b>RdRp ORFI</b>	<b>Capsids α and κ</b>
dN/dS	<b>CpATV</b>	0.004	0.003	ORF2 (κ): 0.10 ORF4 (α): 0.08
	<b>New insect viruses</b>	0.01	0.008	0.01
	<b>Closely-related infectious viruses</b>	0.004	0.006	0.02
	<b>p-value</b>	0.93	0.56	0.95

# To sum up



Small  
RNA  
virus

## Culex pipiens Associated Tunisia Virus (CpATV)



Mosaic virus / Multiple origins

**Functional** virus (under purifying selection)

Two strains with **conserved** indel **polymorphism**

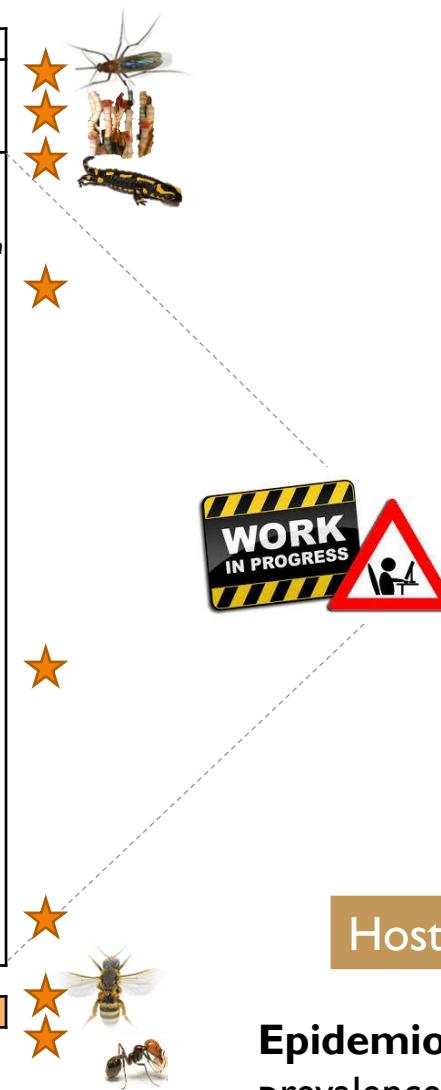
Conserved genomic organization compared to other new insect viruses

New viral genus or family

Is CpATV infectious for mosquitoes and/or plants ?  
Can CpATV modulate arbovirus transmission?

# What about other viruses ?

TAXONOMY	POTENTIEL HOST
Virga-Negevirus	<i>Culex</i>
Parvoviridae	<i>Lamellibrachia</i>
Retroviridae	<i>Pleurodeles / Salamandra</i>
Partitiviridae	<i>Messor</i>
Totiviridae	<i>Formica</i>
Mononegavirales	<i>Microtus, Hippocampus, Myrmica, Formica, Camponotus, Armadilidium</i>
Arenaviridae	<i>Basiliscus</i>
Bunyaviridae	<i>Messor, Eunicella, Formica, Melitaea, Necora</i>
Orthomyxoviridae	<i>Hippocampus</i>
Rhabdoviridae	<i>Formica, Thymelicus, Camponotus, Culex, Sepia</i>
Unassigned	<i>Ciona</i>
Retroviridae	<i>Crocodylus / Emys / Lepus / Microtus / Tupinambis</i>
Arteriviridae	<i>Microtus</i>
Coronaviridae	<i>Hippocampus</i>
Picornavirales	<i>Melitaea / Messor / Halictus</i>
Betaflexiviridae	<i>Ostrea, Rhizaria, Formica, Myrmica, Mytilus, Galba</i>
Benyviridae	<i>Pheidole / Polyergus / Allolobophora / Cystodites / Formica</i>
Flaviviridae	<i>Allolobophora</i>
Nodaviridae	<i>Basilicus, Parus</i>
Tombusviridae	<i>Physa</i>
Virgaviridae	<i>Melitaea</i>
Unassigned	<i>Physa</i>
Sinavirus	<i>Armadilidium / Ciona / Formica</i>
	<i>Necora, Polyergus, Armadilidium</i>
	<i>Halictus, Messor</i>



Virus-centered

Host-centered

**Epidemiology:**  
prevalence, transmission,  
host range, geographic distribution...

# Conclusions - perspectives

---

## ✓ Innovative methodology

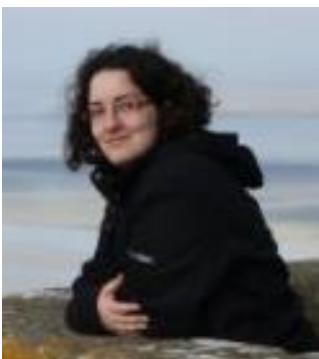
From transcriptomes to viral evolution, genomes & phylogenies

## ✓ Toward global viral diversity pictures in non-model animals

NGS 3<sup>rd</sup> generation: virome analysis, complex trophic interactions, gene networks, community ecology...

# Acknowledgments

**Diane BIGOT**  
PhD student



Equipe EVE / EVIL

**Elisabeth HERNIOU**

Annie Bézier

Karine Musset

Jean-Philippe Vernadet (M2)

Michèle Germain (M2)

Manon Romary (DUT)

Sélim Arhror (M1)



## *Transcriptomes*

Nicolas Galtier

Jonathan Romiguier

Marion Ballenghiens

Vincent Cahais

## *Parvoviridae*

Mylène Ogliastro

Sarah François

## *Mosquitoes*

Mylène Weill

Anna Rivero

## *Retroviridae*

Marc Sitbon

Valérie Courgnaud

Gilles Labesse

Jean-Luc Battini

## *Bee viruses*

Anne Dalmon



## *Bee viruses*

James Cook

Bronwen Roy

**WESTERN SYDNEY**  
UNIVERSITY



Chunsheng Hou

Qingyun Diao

Shuai Deng



Institut of  
Apicultural  
Research



Jenny Cory



SIMON FRASER  
UNIVERSITY



Bourse de recherche Jean  
& Marie-Louise  
DUFRENOY



Thanks for your attention

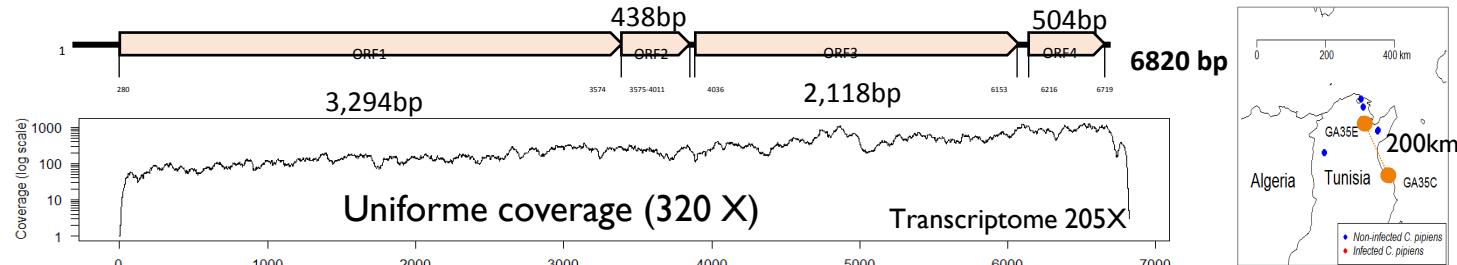
# Two strains – one virus

GA35C (2005) Moknine (Tunisia)



GA35E (2011) El Habibia (Tunisia)

Same genomic organisation



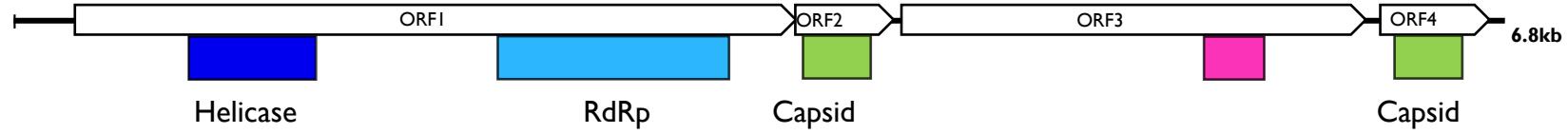
→ CpATV strain El Habibia

26 SNPs (Single Nucleotide Polymorphisms) between the both strains (0,4% variation)

## INDEL 4450

Polymorphism maintained over 6 years

## CpATV

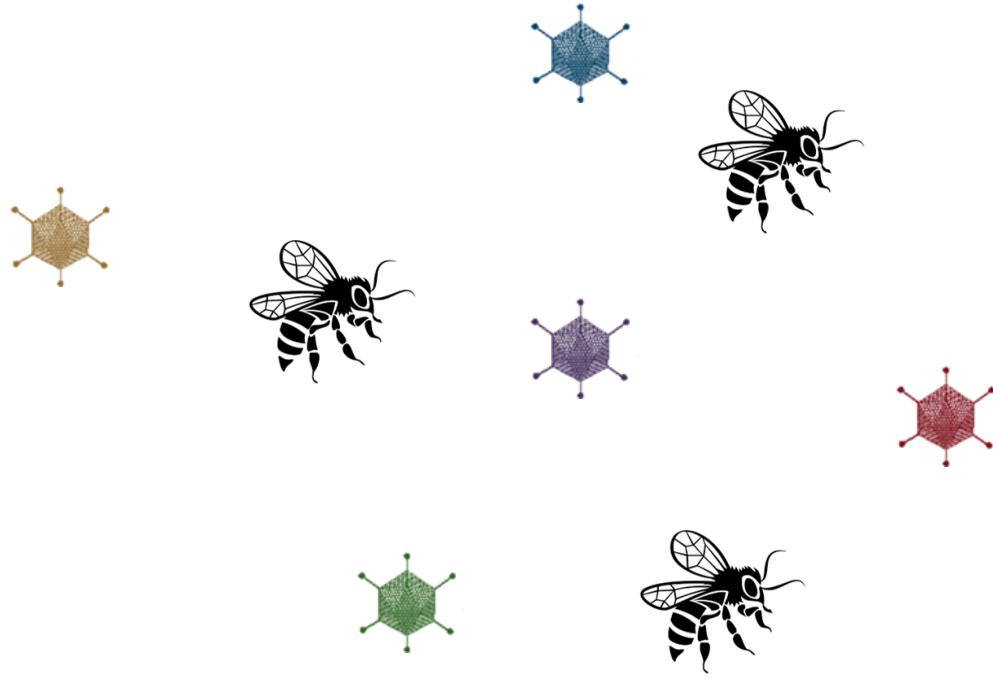


A insertion

- Premature STOP codon
- Two smallest proteins
- Role for functionality?

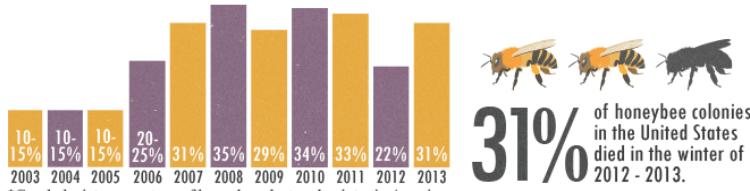
# Part 3

Bee viruses



# Bees disappear ...

COLONY COLLAPSE DISORDER IN AMERICA



## Multifactorial causes

### Nutritional stress



### Mites

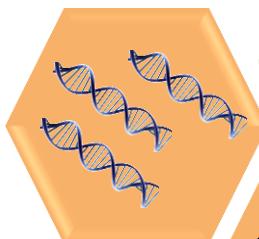


*Varroa destructor*

### Pesticides



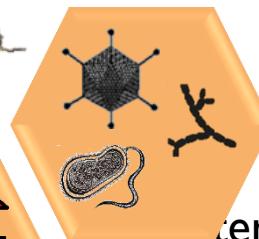
### Lack of genetic diversity



Stress transports  
Climate change  
....



### Loss of natural habitat



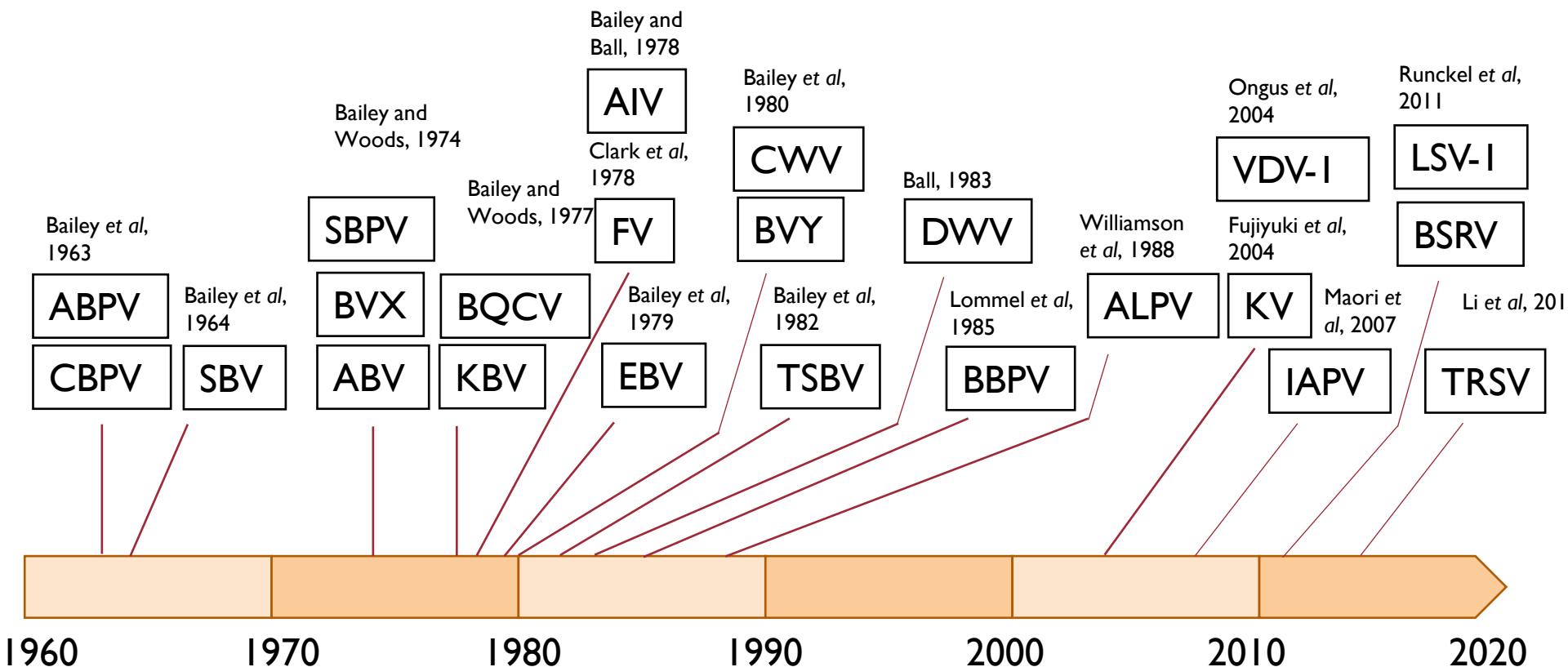
### Bacteria, fungi & viruses

Impact of pathogens correlated with CCD

Cornman et al. 2012

# Emergence of viruses in honeybees

Time of first documentation of major honeybee viruses since 1960



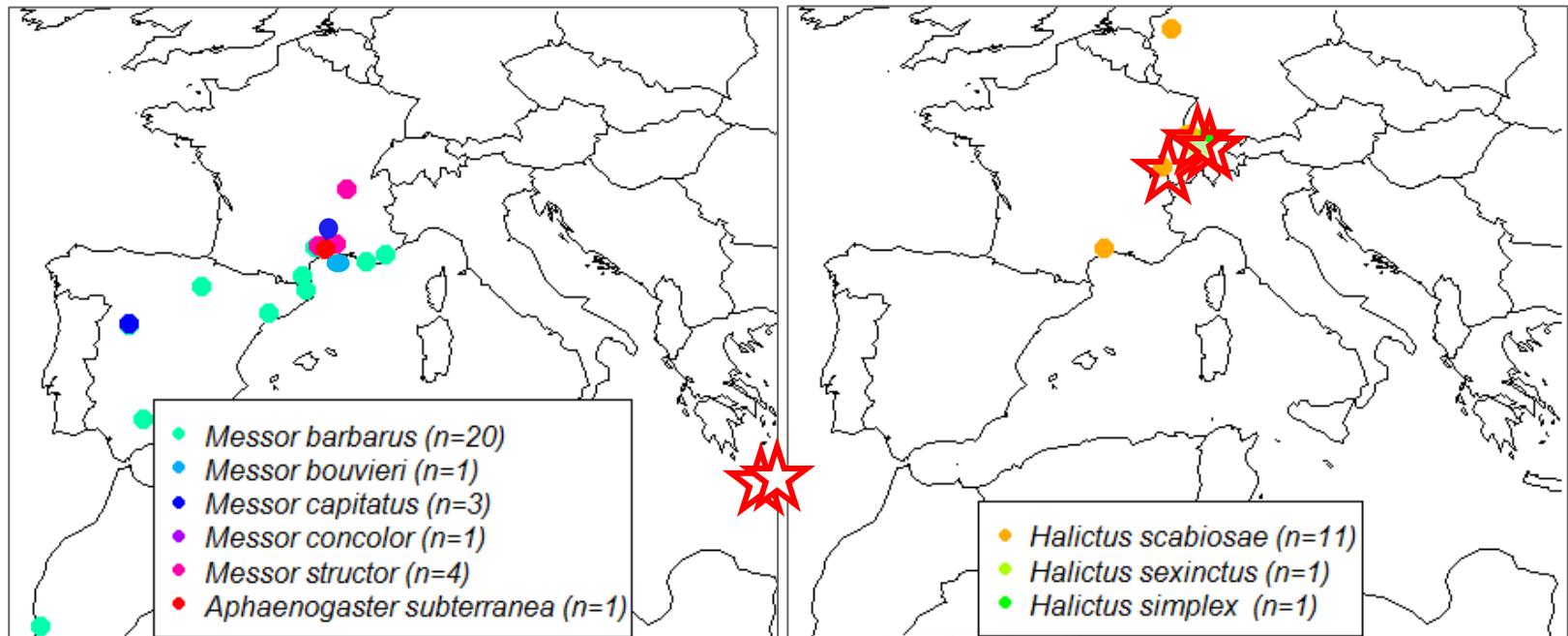
# Biological data

PopPhyl



## 43 individual Illumina transcriptomes

### Sampling



6 species  
30 individuals

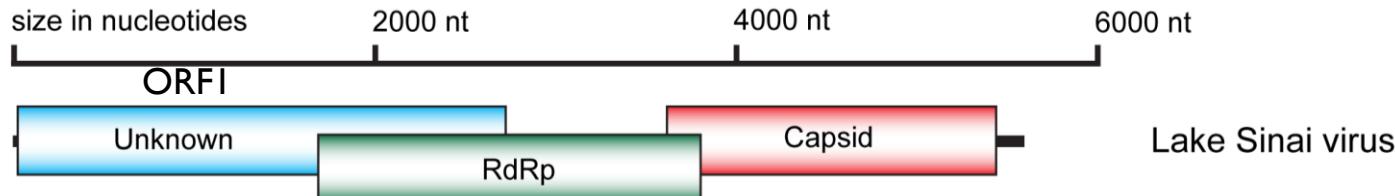


3 species  
13 individuals

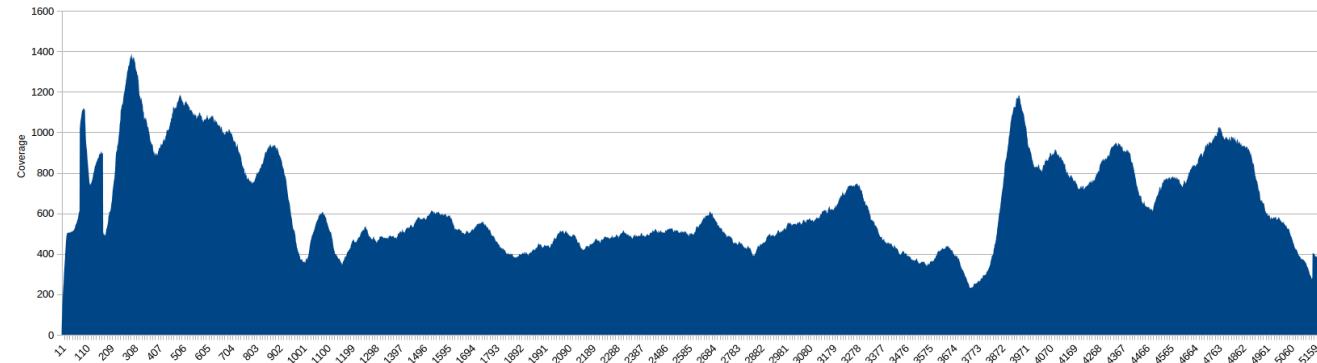
→ 5 new complete viral genomes (ssRNA+ virus)

# Results

Same genomic organization as a bee virus *Lake Sinai Virus*

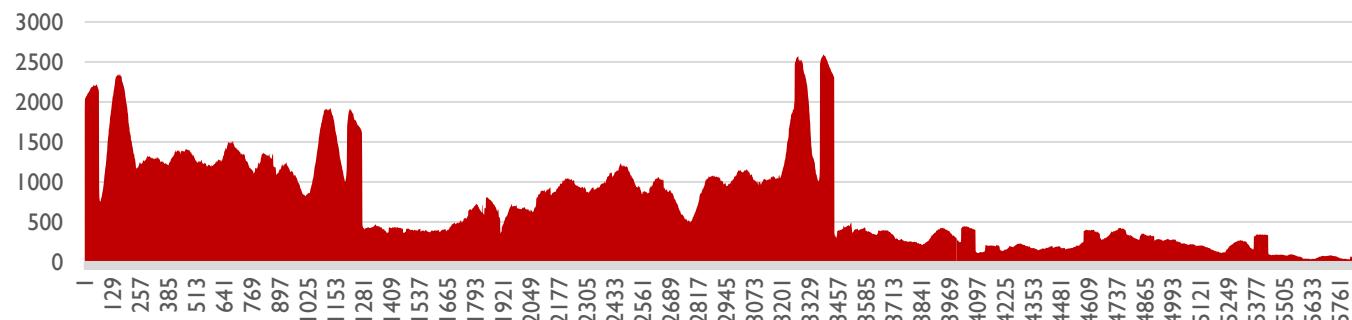


Runckel et al, 2011



640X

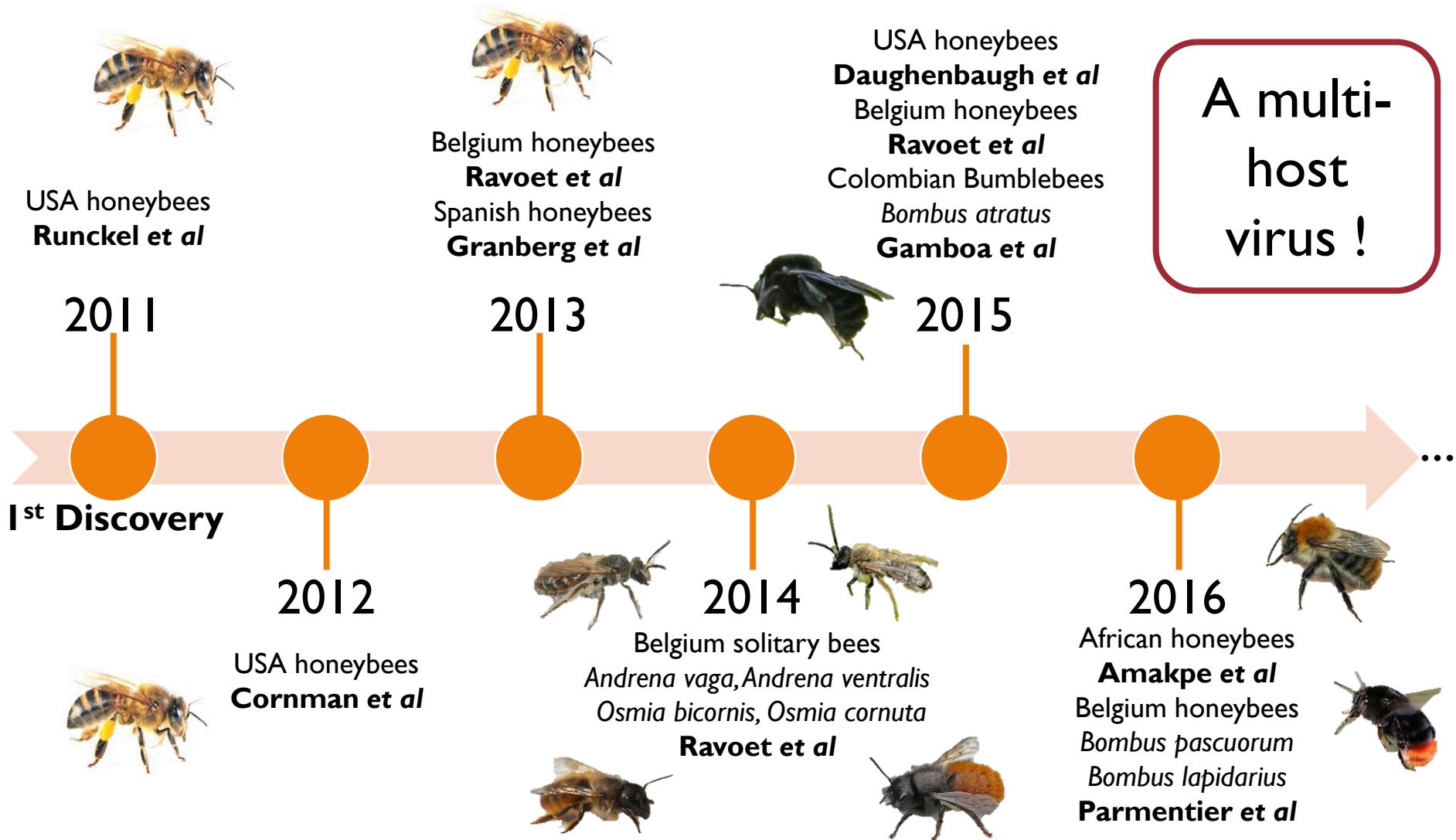
Mapping coverage of a complete genome



329X



# Lake Sinai Virus – History

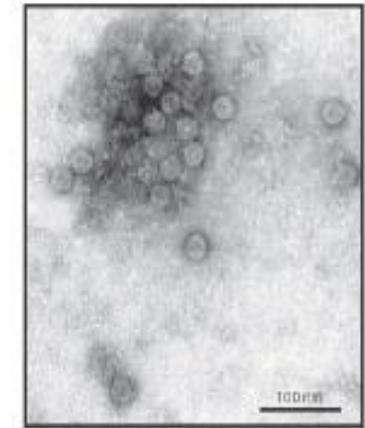


# Lake Sinai Virus – What do we know ?

Is LSV a **harmless virus?**

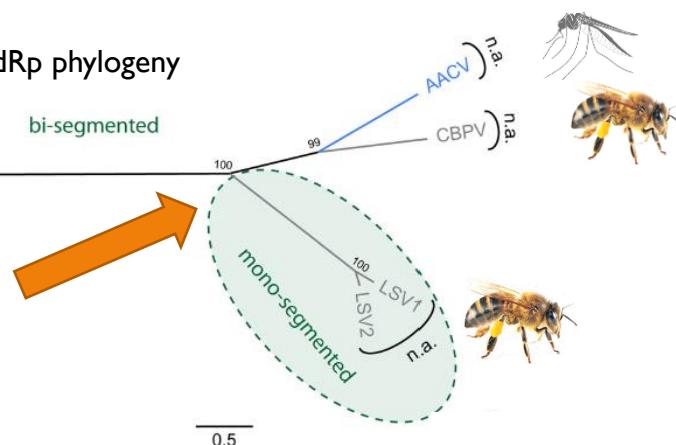
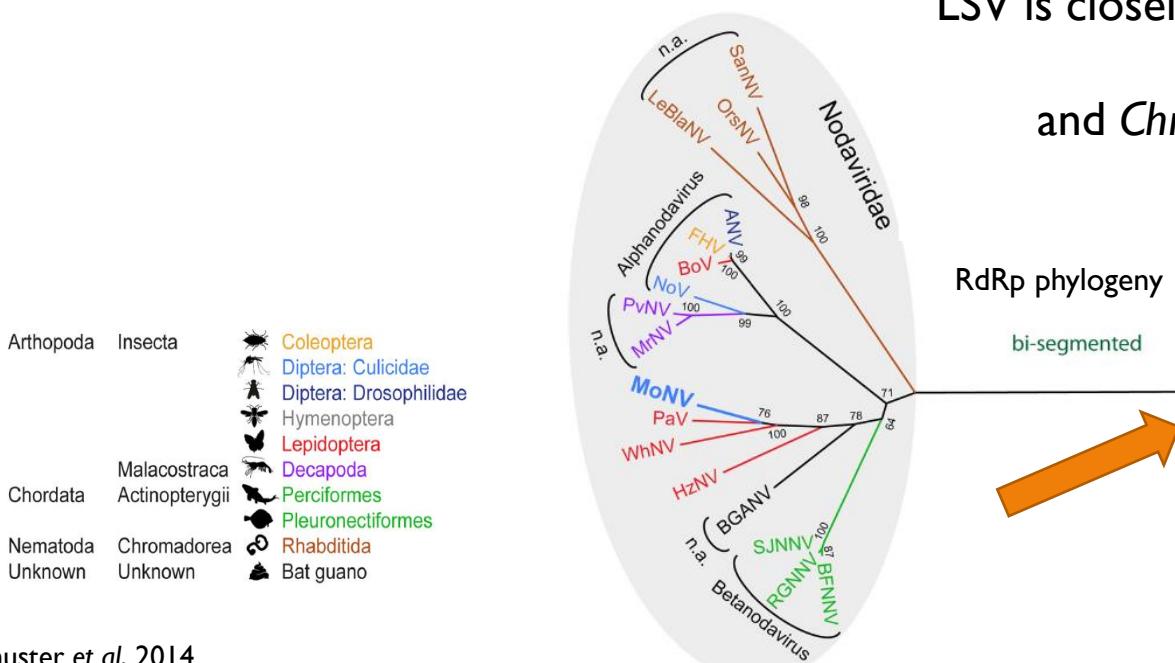
LSV is **infectious** : viral genome replicates in honeybees

Discovered **during CCD** + positive correlation with **weak bee colonies**

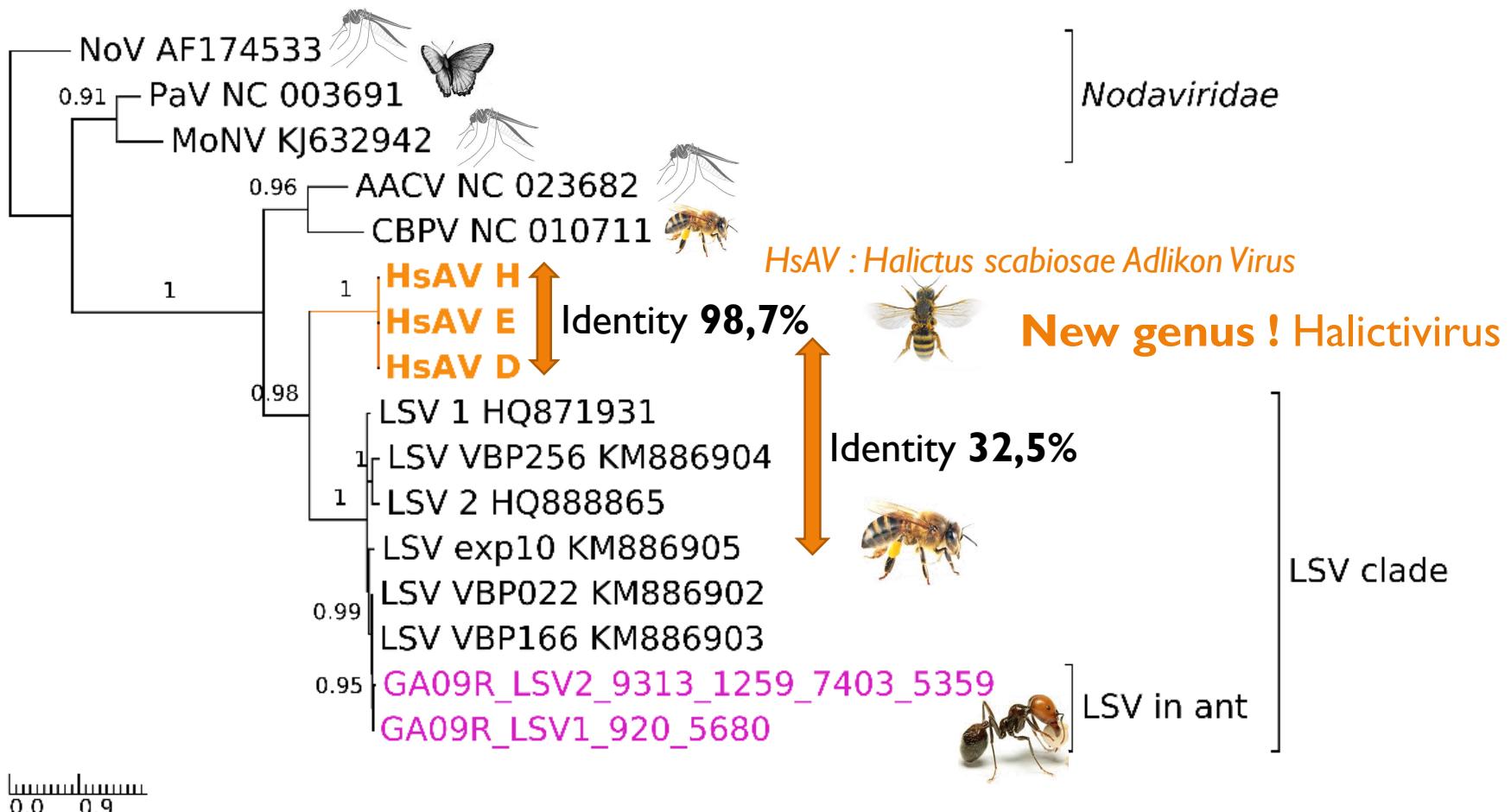
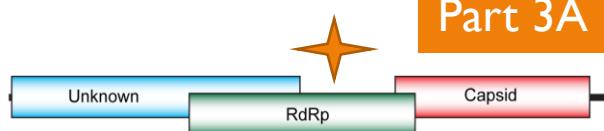


Daughenbaugh et al, 2015

LSV is closely related to **Anopheline Associated C Virus (AACV)** and **Chronic Bee Paralysis Virus (CBPV)**



# RdRp protein ML phylogeny



Discovery of a new genus of wild bee virus !

LSV described for the first time in ants !



# Small scale evolution of *Sinaivirus*

Focus on LSV diversity

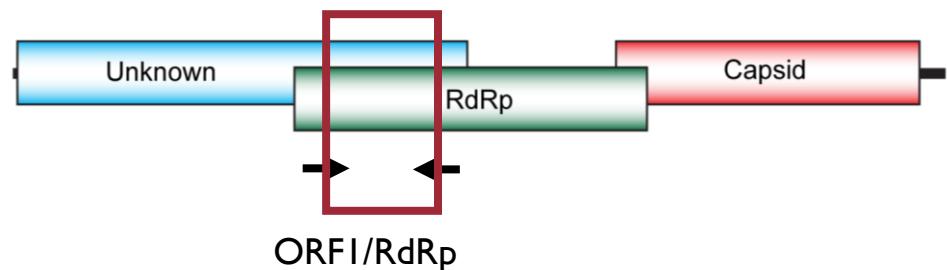
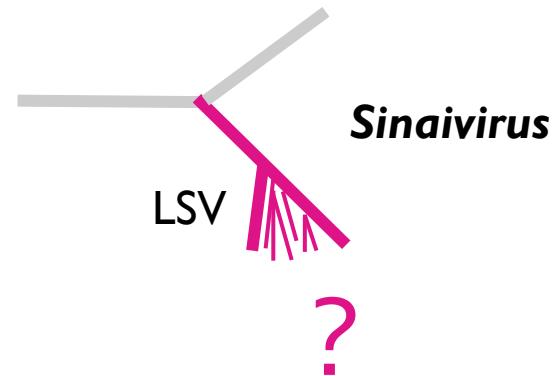
Build a most exhaustive phylogeny  
and resolved ant LSV position

Method

RT-PCR detection with homemade degenerated  
primers for LSV detection

Sanger sequencing

ML phylogeny



Allow amplification of  
all LSV strains

# Focus on LSV diversity

Sampling for ML phylogeny tree building

Build a most exhaustive phylogeny

All known sequences in public database

**6** known LSV genomes

**47** partial LSV published sequences



## New LSV sequences

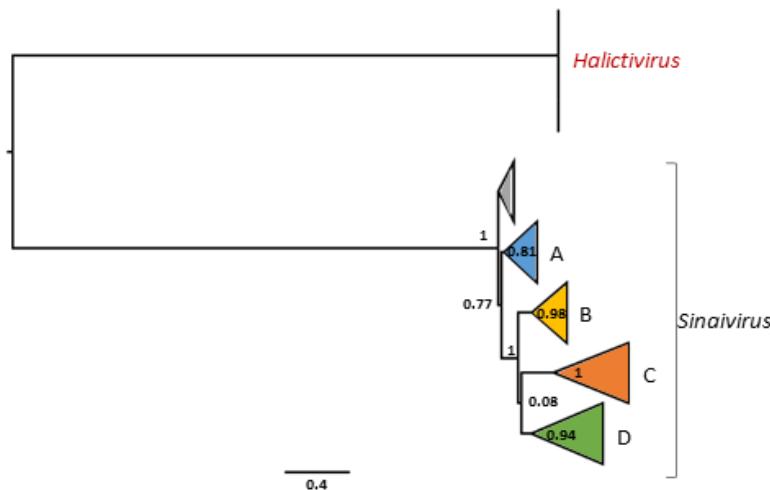
**Bigot et al, in prep**

**4** LSV sequences found in **ants** *in silico*

**33** LSV sequences found in **honeybees** *in vivo*

**14** samples France  
**4** samples Italia  
**3** samples Canada  
**7** samples Australia  
**5** samples China

# LSV diversity



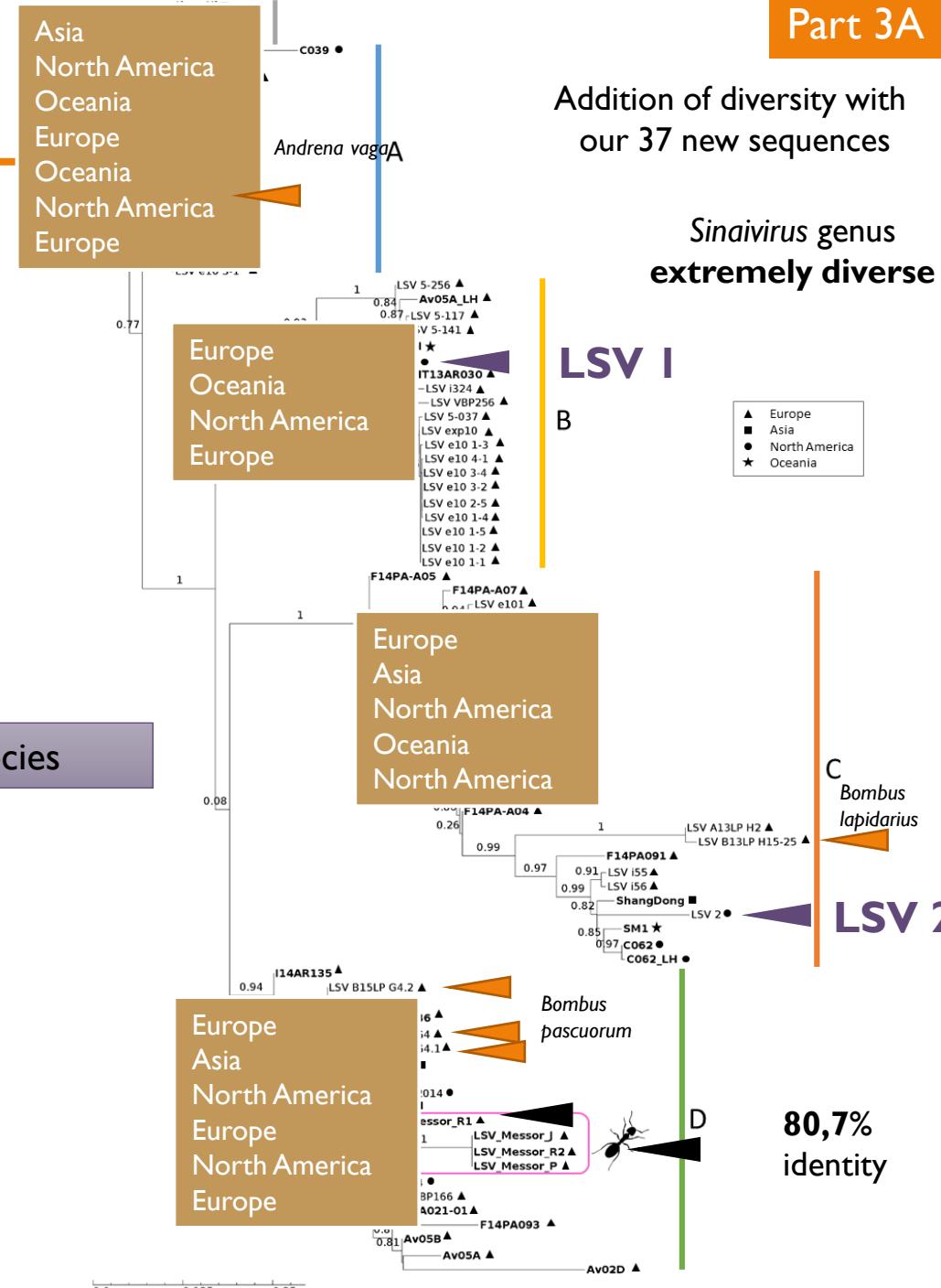
Halictivirus resolves *Sinaivirus* phylogeny

ICTV recognizes only LSV1 and LSV2 as species

LSV in other non-*Apis* bees even in ants

Two LSV strains co-infecting a unique ant  
Once observed in bees  
Increase complexity for management of infection

Diversity not correlated with geographical origins

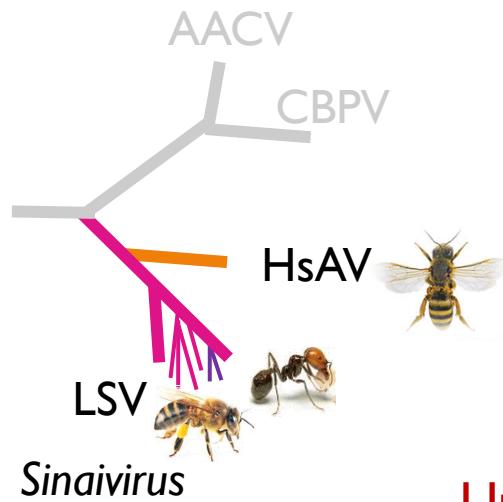


# To sum up

Article 4, Bigot et al, *Virus Evolution*, in prep

43 hymenopteran transcriptomes

→ 5 complete viral genomes + 2 partial genomes in ants  
→ 2 viruses



New HsAV virus – Halictivirus new genus !

LSV described for the first time in ants  
+ co-infection

Underestimated sampling for bee virus discovery ?

Are bee viruses common in ants ?

What about bee viruses in wild hymenoptera ?