VIRUS STRAINS, VARIABILITY, AND EVOLUTION

Objectives:

1. Know the molecular mechanisms that generate variation in virus sequences

2. Be familiar with principals, approaches, and techniques used in molecular phylogeny of plant viruses

3. Understand how viruses evolve (selection pressures, constraints)

4. Understand the different theories on the origin(s) of viruses

Within a virus species:

A dynamic equilibrium is maintained by the opposing process of **mutation**, which tends to increase the range of variants,
and **negative selection** which tends to eliminate those that are least fit.

Changes in the viruses environment may positively select some variants over others, and this can lead to changes in the consensus sequence.

Process by which the genetic structure of the population of an organism changes with time - **evolution**.

Viruses may follow Darwin's original concepts of survival of the fittest.

Directed Creation/Selection of Variants

Virologists/cell biologists deliberately create variants in the lab in order to study virus/cell functions. Creation of variants is relatively easy

- Select for variants by **sequential passage** of a virus through different hosts.
- Virus cultured under **high** or **low temperatures** may result in variants with different sequences and traits
- In the field, select for resistance-breaking variants by **sequential passage through virus-resistant** hosts.
- In the laboratory variants can be created by:
  nitrous acid treatment of viral RNA,
  ionizing radiation, oxidative damage, and UV irradiation
  Site-directed mutagenesis in infectious cDNA clones of viral genomes

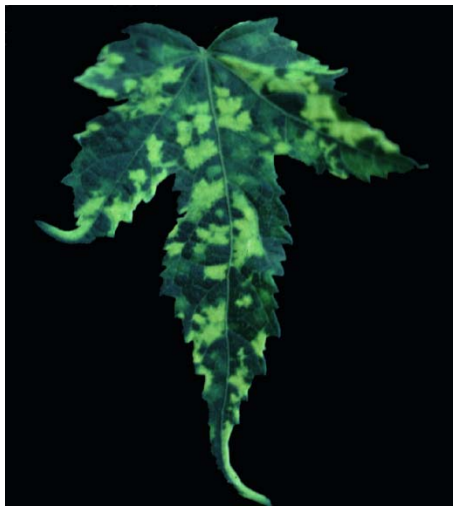
Examples of Variants Occurring "Naturally"

Example 1:

- Repeated maintenance by grafting resulted in loss of vector transmission

  Ex. Abutilon mosaic virus (*Begomovirus*) maintained in *Abutilon spp*. is no longer whitefly transmissible

  (3 mutations in the coat protein have rendered the virus non-transmissible)

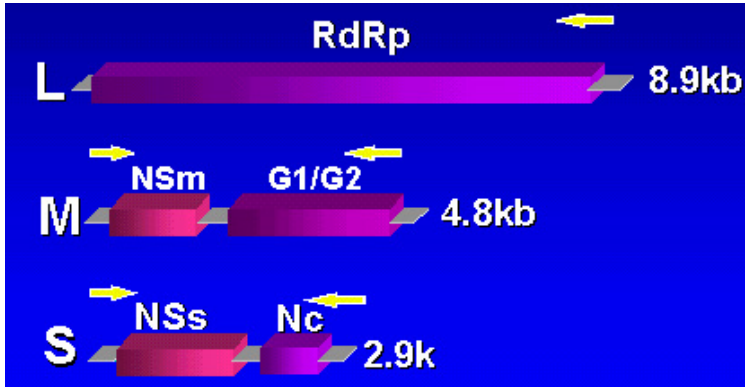


Examples of Variants Occurring "Naturally"

Example 2:

- Repeated mechanical transmission of tospoviruses results in a loss of their ability to be transmitted by thrips.

  Ex. Mutations accumulate in the *Gc* and *Gn* (*G1/G2*) genes on the M-RNA which result in the inability of the proteins to infect the thrips

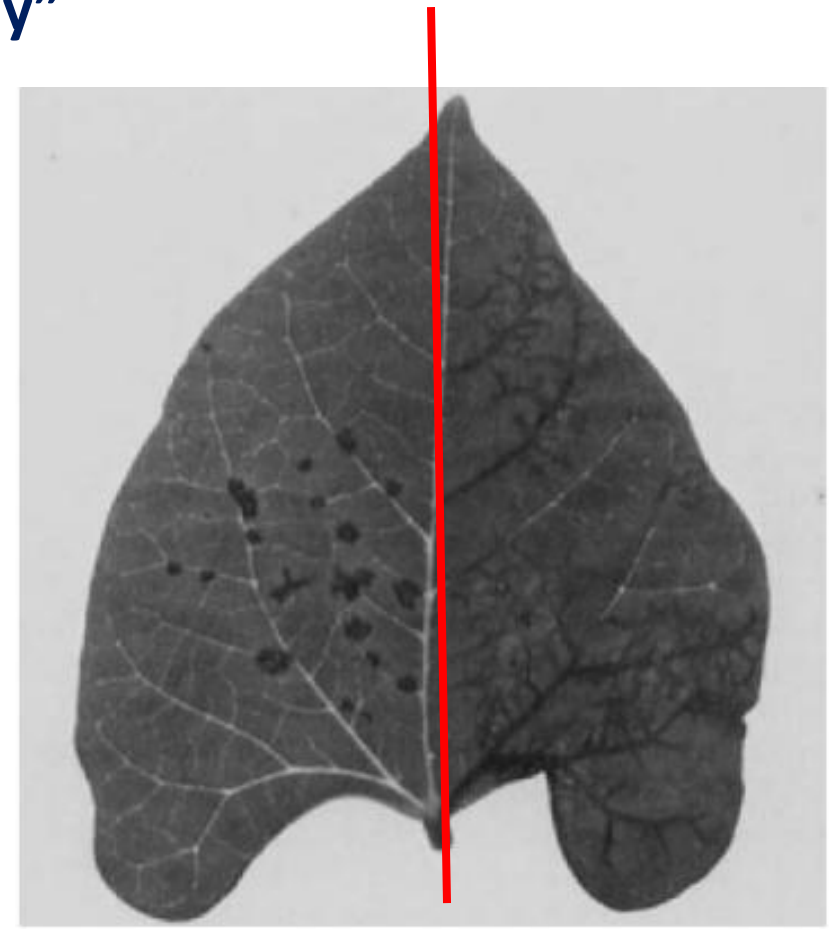


Examples of Variants Occurring "Naturally"

Example 3:

- Passage and maintenance in one host can rapidly select for variants ("host passage effects"

Ex. Symptoms of *Tobacco necrosis virus* maintained in Pinto bean (left side of leaf) compared to the same culture (right side of leaf) maintained in petunia for 6 weeks before inoculation back to bean (6 days PI).

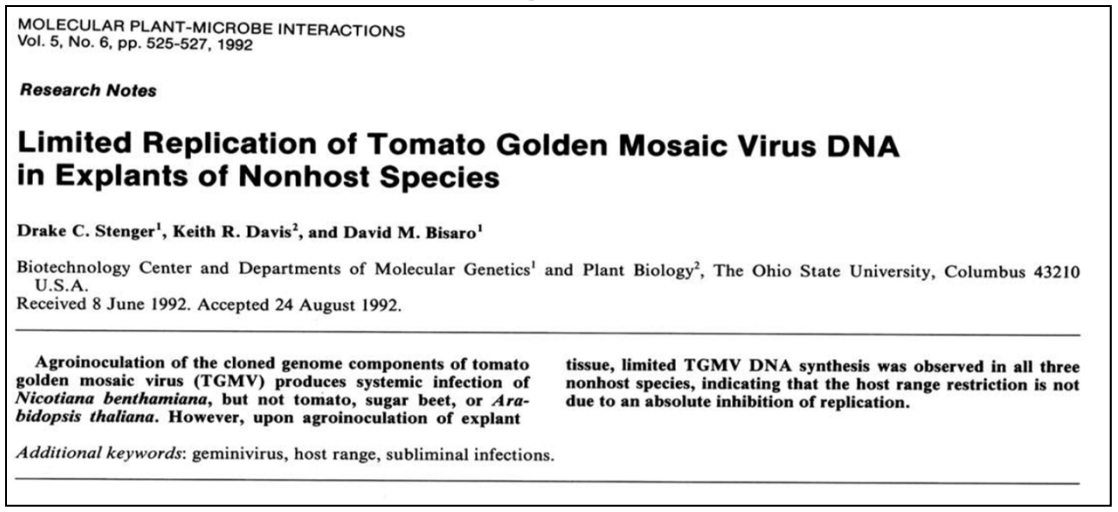


Examples of Variants Occurring "Naturally"

Example 4:

- Repeated maintenance of virus cultures in one host may result in the loss of ability to infect another host(s).

Ex. *Tomato golden mosaic virus* maintained in *N. benthamiana* for 10 years, can no longer infect tomato - its host of origin



MOLECULAR PLANT-MICROBE INTERACTIONS
Vol. 5, No. 6, pp. 525-527, 1992

Research Notes

Limited Replication of Tomato Golden Mosaic Virus DNA in Explants of Nonhost Species

Drake C. Stenger[1], Keith R. Davis[2], and David M. Bisaro[1]

Biotechnology Center and Departments of Molecular Genetics[1] and Plant Biology[2], The Ohio State University, Columbus 43210 U.S.A.

Received 8 June 1992. Accepted 24 August 1992.

Agroinoculation of the cloned genome components of tomato golden mosaic virus (TGMV) produces systemic infection of *Nicotiana benthamiana*, but not tomato, sugar beet, or *Arabidopsis thaliana*. However, upon agroinoculation of explant tissue, limited TGMV DNA synthesis was observed in all three nonhost species, indicating that the host range restriction is not due to an absolute inhibition of replication.

Additional keywords: geminivirus, host range, subliminal infections.

Viruses need variation in order to survive.

What mechanisms do they use to induce variation?

Mechanisms that Generate Variation:

A. **Mutation** (change in nucleotides in genome sequence)

B. **Recombination** (exchange of sequences)

   Additions (new sequence)
   Deletions (loss of sequence)
   Gene Duplications (duplication of sequences)

C. **Reassortment of segmented genomes** (exchange of DNA but not by genetic recomb.; possible in multipartite genomes)

D. Multiple combinations of the above

Mechanisms that Generate Variation:

A. Mutation (change in nucleotides in genome sequence)
Can arise due to a number of mechanisms

1. Natural variants arise from **copying errors** made during replication of the viral genome

2. In nature: **Ionizing radiation, oxidative damage, and UV irradiation**

3. Artificial mutants can be generated in the laboratory (nitrous oxide, site directed mutagenesis, etc..)

Mechanisms that Generate Variation - Mutation

Cellular DNA genomes vs Viral RNA genomes:

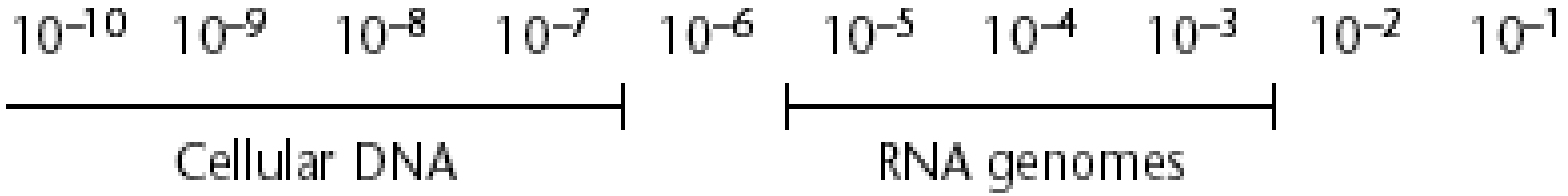
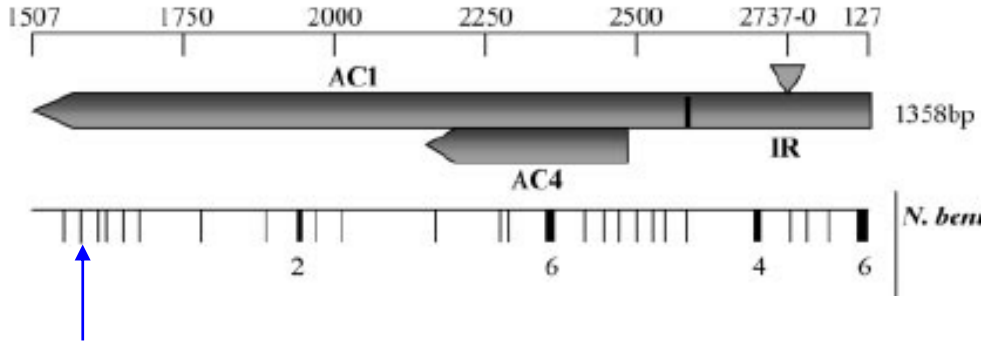


Figure 2 Mutation rates and frequencies. Cellular DNA replication is several orders of magnitude more accurate than RNA virus genome replication or retrotranscription. An average mutation rate of 10^{-4} means that one misincorporation error at a given nucleotide template will occur once every 10 000 times that the polymerase copies this specific nucleotide. For both cellular DNA and viral RNA, mutation rates and frequencies much higher than average have been described (hypermutability). For RNA genomes mutation rates and frequencies lower than average may also occur (hypomutability).

Example: Measuring the rate and location of mutations

Host	No. of Days Post Inoc.	Mutation Frequency
Nicotiana benthamiana	60	3.1 – 4.1 x10^{-4}
tomato	60	4.1 x 10^{-4}



Each bar is a mutation

Results of a study with TYLCCV (begomovirus, ssDNA genome):

- Mutation rates were not dependent upon host

- Mutations were distributed across the genome

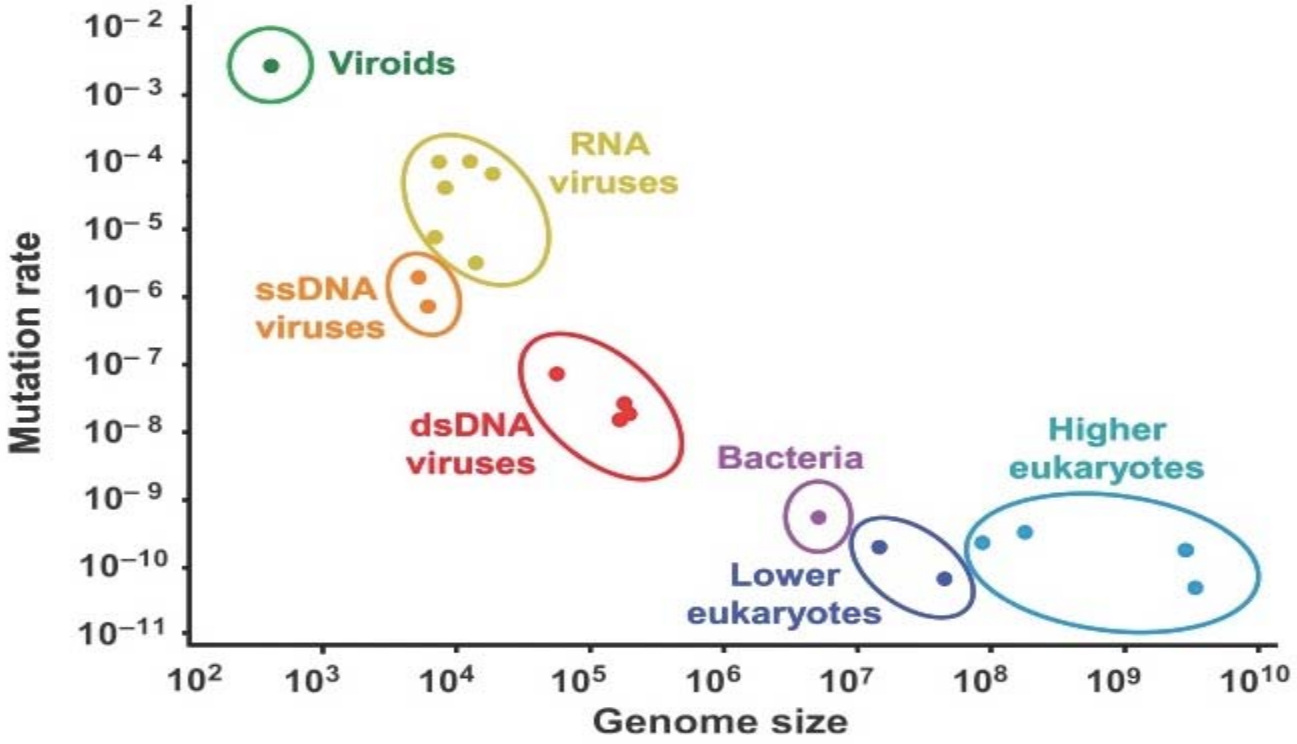
JOURNAL OF VIROLOGY, June 2007, p. 5902–5907
0022-538X/07/$08.00+0 doi:10.1128/JVI.02431-06

Vol. 81, No. 11

Genetic Structure and Population Variability of Tomato Yellow Leaf Curl China Virus[▽]

Linmei Ge,† Jiangtao Zhang, Xueping Zhou,* and Hongye Li*
Institute of Biotechnology, Zhejiang University, Hangzhou, Zhejiang 310029, China

Comparison of Mutation Rates vs Genome Sizes



Viruses and viroids have high rates of mutation relative to their genome size

“RNA viruses operate close to the error threshold that allows maximum exploration of sequence space while conserving the information content of the genotype”.

E. Koonin Nature Reviews 2008

This statement is true for both RNA and DNA viral genomes

Mechanisms that Generate Variation:

A. **Mutation** (change in nucleotides in genome sequence)

B. **Recombination** (exchange of sequences)

Additions (new sequence)
Deletions (loss of sequences)
Gene Duplications (duplication of sequences)

C. **Reassortment of segmented genomes** (exchange of DNA but not by genetic recomb.; possible in multipartite genomes)

D. Multiple combinations of the above

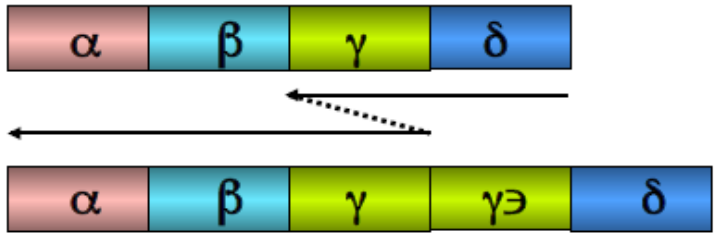
Mechanisms that Generate Variation:

B. Recombination

the process by which segments of genetic information are switched between the nucleotide strands of different genetic variants during the process of replication

Recombination results in deletions, insertions, substitutions, duplications, etc…

Occurs in both DNA and RNA viruses



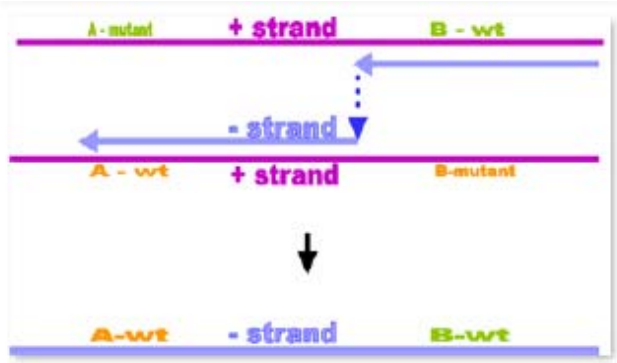
Sequence duplication

Mechanisms that Generate Variation: Recombination

How does recombination occur?

3 Basic Models:

1. Copy Choice mechanism - the viral replicase switches from one template (template switching) to an alternate location on the same or a different template where polymerization of the nascent strand continues.
(there are **hot spots** for template switching.)

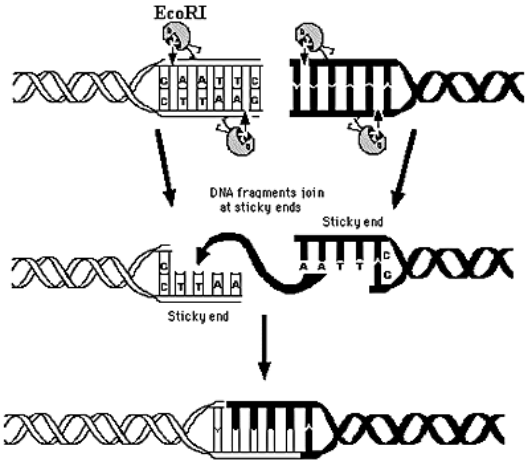


Mechanisms that Generate Variation: Recombination

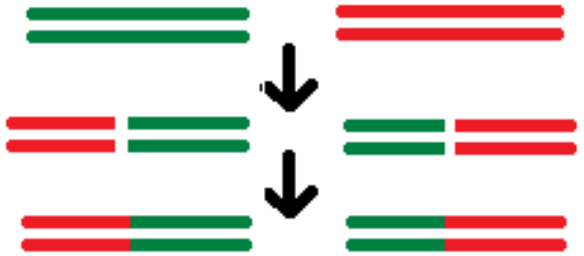
How does recombination occur?

3 Basic Models:

2. Enzymatic cutting and re-ligation – genome may be cut enzymatically, and then ligate with another genome or to another place in the same genome.

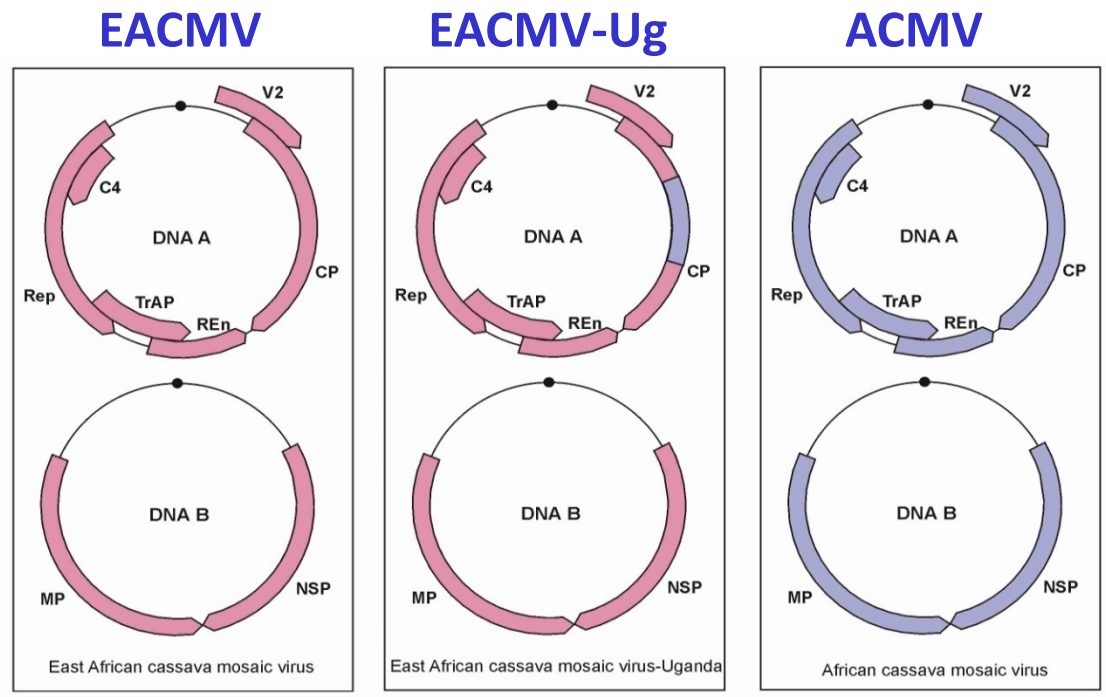


3. Breakage-induced template switching – genome breaks and ligates with another genome or another place in the genome.



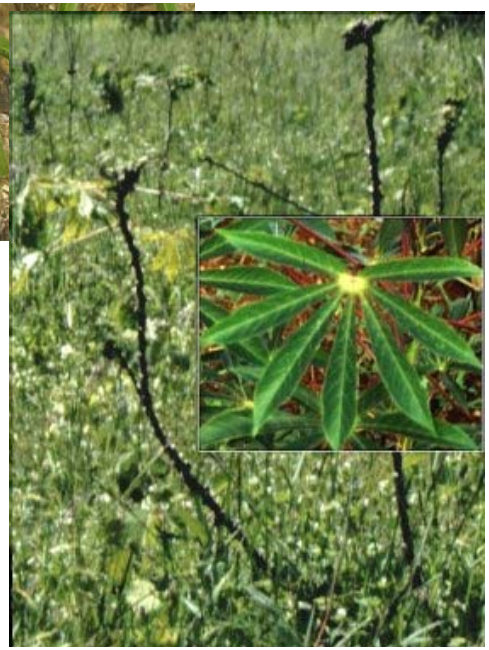
Importance of Recombination to Agriculture:

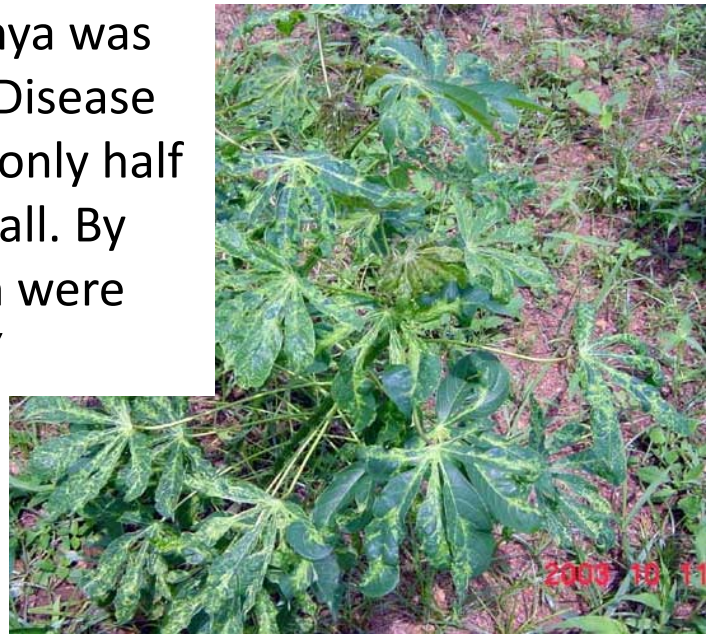
East African Cassava mosaic virus; the first evidence for rapid evolution through recombination



Healthy Cassava

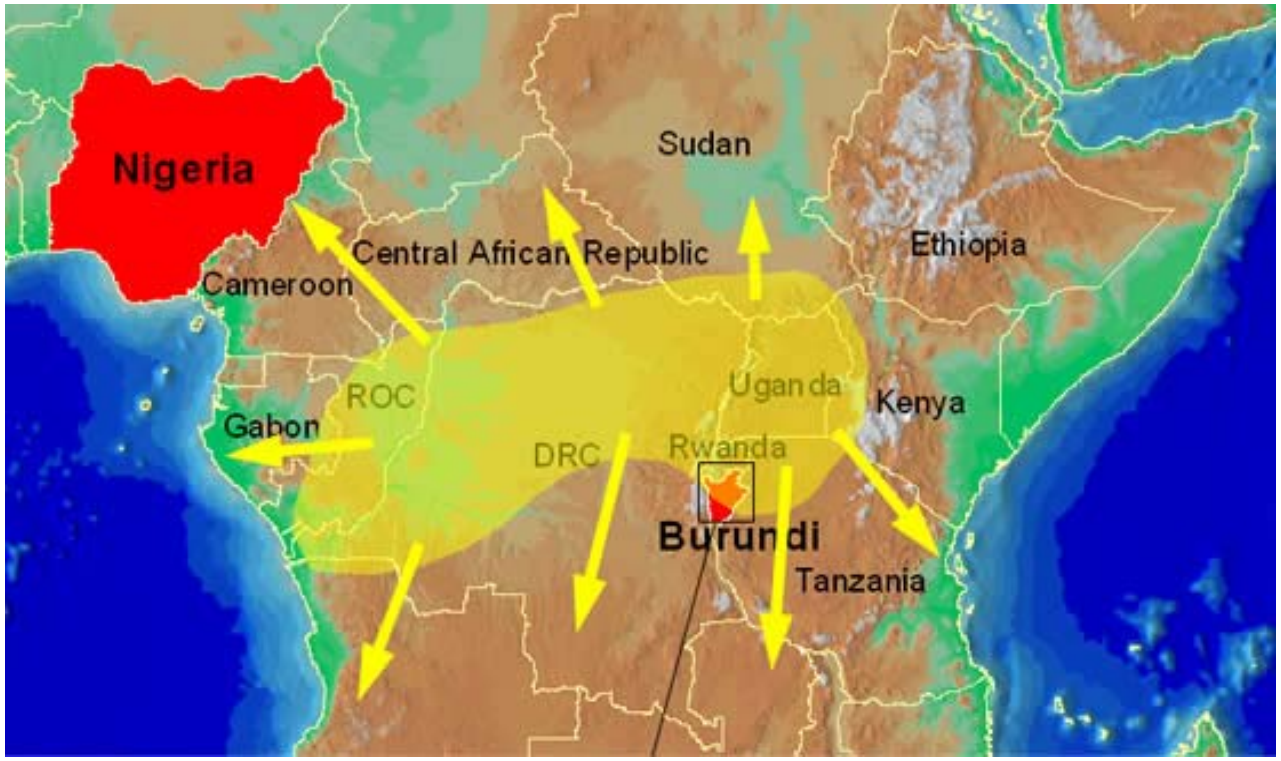
"In 1997 the crop in Western Kenya was devastated by a Cassava Mosaic Disease (CMD). Many farmers harvested only half of their crop, while others lost it all. By 1998 …. yield losses in the region were estimated to be US $10 million. "





ACMV-infected Cassava

The recombinant virus spread from Uganda all over central and South Africa wiping out cassava crops as it spread.



Mechanisms that Generate Variation:

A. **Mutation** (change in nucleotides in genome sequence)

B. **Recombination** (exchange of sequences)

Additions (new sequence)

Deletions (loss of sequences)

Gene Duplications (duplication of sequences)

C. **Reassortment of segmented genomes** (exchange of DNA but not by genetic recomb.; possible in multipartite genomes)

D. Multiple combinations of the above

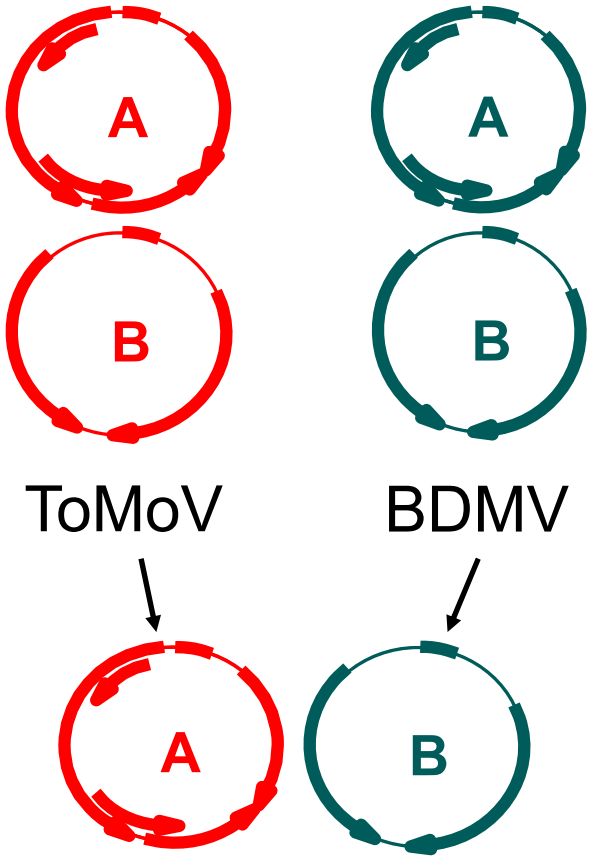
Mechanisms that Generate Variation:

C. Reassortment of segmented genomes.

Infectious **pseudorecombinants** can be artificially generated

The pseudorecombinants produced different symptoms than either of the homologous combinations.

Postulated - the exchange of components between geminiviruses in nature may be a mechanism for generation of new geminiviruses.



Gilbertson et al (J. Gen. Virology 74:23-31)

Mechanisms that Generate Variation:

A. **Mutation** (change in nucleotides in genome sequence)

B. **Recombination** (exchange of sequences)

Additions (new sequence)

Deletions (loss of sequences)

Gene Duplications (duplication of sequences)

C. **Reassortment of segmented genomes** (exchange of DNA but not by genetic recomb.; possible in multipartite genomes)

D. Multiple combinations of the above

D. Viruses use combinations of mechanisms to create/maintain variability

Example 1:

Viruses can replicate for years in a perennial wild plant species,
Giving rise to many different sequences
Some of those sequences will infect crop plants if there is a vector to move them

Example: *Tomato mottle virus* in Florida (causes disease in tomato)

Variant of *Sida golden mosaic virus* which infects *Sida spp.* In Florida

Example 2:

Nine new emergent species believed to be present in tomato in Brazil (Ribeiro et al 2007)

Mutation and recombination combined to create new virus species:

- Found evidence for 3 recent recombination events in *Tomato rugose mosaic virus* (ToRMV) A component, but none in the B component
- Multiple recombination events detected in *Tomato chlorotic mottle virus* (ToCMoV) but all were determined to be ancient events

Ribeiro, S. G., Martin, D. P., Lacorte, C., Simões, I. C., Orlandini, D. R. S., and Inoue-Nagata, A. K. 2007. Phytopathology 97:702-711.

VARIABILITY

How do we measure diversity?

How do we use this information?

Distinguishing And Identifying Strains

Quick Summary:

Two types of criteria are used:

A. Biological criteria

B. Structural criteria

A. Biological Criteria:

Differences in:

- Symptom expression,
- Cross protection,
- Host range,
- Crop varietal reaction,
- Virus-vector relationships (ex. vector specificity).

Variants that differ biologically can be known as **(biotypes, pathotypes).**

Distinguishing And Identifying Variants

Quick Summary:

Two types of criteria are used:

A. Biological criteria

B. Structural criteria

B. Structural Criteria

- Differences in viral genome sequences –
  detected by hybridization, presence or absence of restriction endonuclease sites, etc..

- Differences in viral proteins -
  Detected by differences in amino acid sequence, physical characteristics (particle stability), serological characteristics (serotypes)

B. Structural Criteria

Since 1990's - nucleotide sequence of the viral genomic nucleic acid and amino-acid sequence of viral proteins are the basis for phylogenetic analyses.

Computer-based algorithms –
Calculate the degrees of **similarity** or **dissimilarity** between sequences of nucleotides or amino acids.

Programs align the sequences of nucleotides or amino acids side by side for pair-wise or stepwise multiple comparison.

Produce **tabular distance matrices** and **phylogenetic trees** (dendrograms) as a quantitative indication of phylogenetic relationship.

Phylogenetic Trees: Presenting Evolutionary Relationships

Systematics: area of research that describes the pattern of relationships among taxa and is intended to help us understand the history of all life.

In phylogenetic studies, the most convenient way of visually presenting evolutionary relationships among a group of organisms is through illustrations called **phylogenetic trees**.

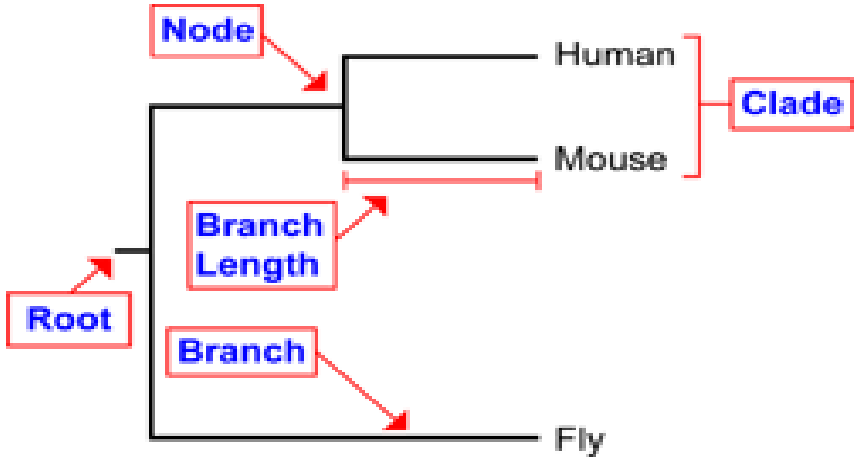
Molecular Phylogenetic Analysis:
Can generate trees from:
Nucleic acid sequences, amino acid sequences, whole genome, single gene or region, etc..

DNA, RNA, and protein sequences can be considered as phenotypic traits.

Phylogenetic Tree Features Defined:

Node – represents a taxonomic unit. This can be an existing species or an ancestor.

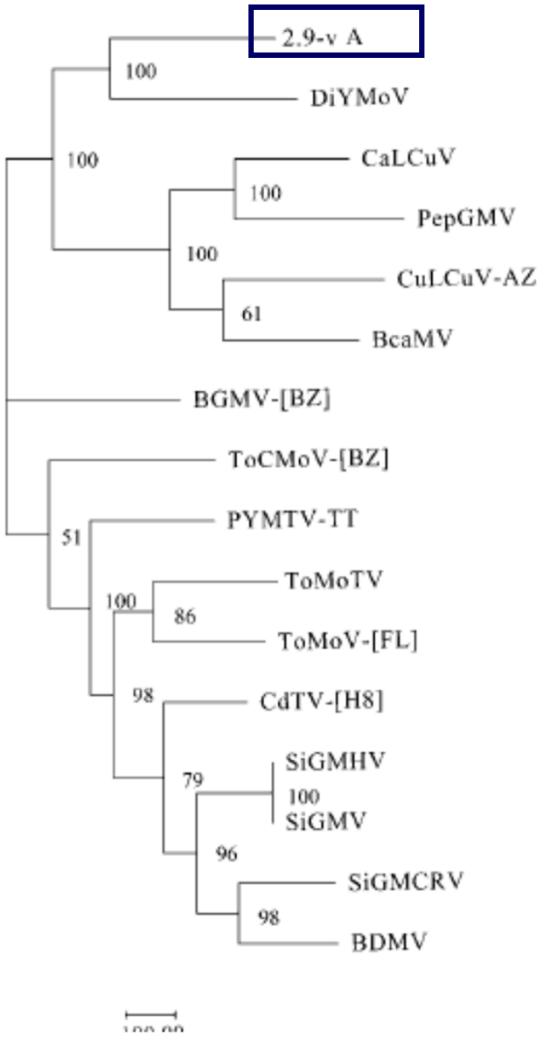
Branch – Defines the relationship between the taxa in terms of descent and ancestry.



Branch Length – Represents the number of changes that have occurred in the branch.

Clade – A group of two or more taxa or sequences that includes both their common ancestor and all their descendents.

Root – The common ancestor of all taxa.



Phylogenetic tree of the A component of isolate 2.9-v

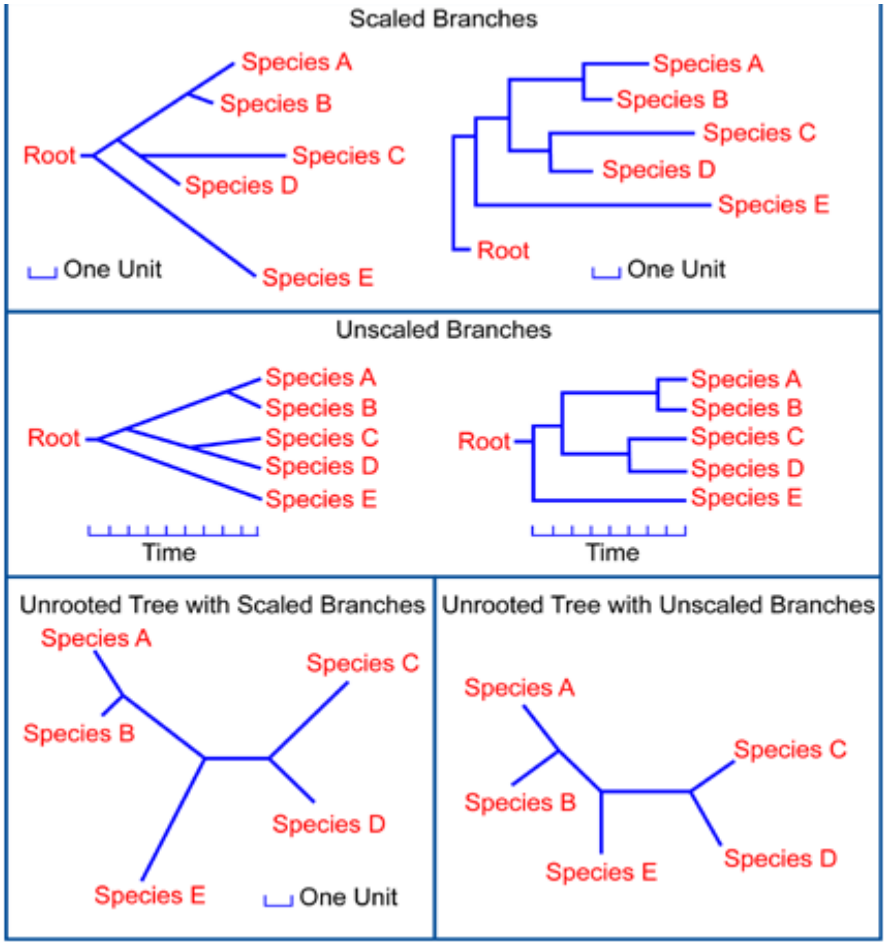
2.9v - new virus sequence from tomato in Venezuela…. new strain or new virus?

- **Vertical distances are arbitrary**
- **Horizontal distances are proportion to the number of nucleotide differences**

New name:
Tomato yellow margin leaf curl virus

Nava (2013) Arch. Virol. 158:399-406

Possible ways of presenting a phylogenetic tree:



Phylogenetic trees …. relatedness ….. common ancestors ….. evolution

Viruses have many mechanisms to adapt to changes in their circumstances –

But can we trace these changes back to tell us something about their origins?

What we know:

Evolution: Relationship of Plant Viruses to Animal Viruses

Ex. Replication in Plants and Animals:

A number of plant viruses, including species in 3 genera in the *Reoviridae*, 2 genera in the *Rhabdoviridae*, and the genera *Tospovirus* (*Bunyaviridae*), *Marafivirus* (*Tymoviridae*) and *Tenuivirus* (family unassigned) replicate both in plants and in their insect vector.

These plant viruses are closely related to their animal counterparts, indicating they have common ancestory.

What we know:

Evolution: Relationship of Plant Viruses to Animal Viruses

Viruses certainly have played an important role in evolution of their hosts because of their ability to pick up gene sequences and carry them to different cells or organisms

- Retroviruses and retro-transposons share several similar characteristics
  - integration into host genome (being one)

Where did viruses come from?

Do they have a common origin?
Or did they arise in multiple events?

There are 5 theories as to the origin(s) of viruses

Theory 1.

Viruses are **degenerate micro-organisms** that gradually developed through parasitic adaptation from bacteria or from phytoplasmas. They that have lost most of their own functions during increasing dependence on their hosts.

Theory 2.

Viruses are **fragments of host nucleic acid that have become autonomous.**
ie the dsDNA of *Caulimoviridae and Geminiviridae* occur inside the host nucleus as a *'minichromosome'*, fully double-stranded, supercoiled, and associated with histones.

Theory 3.

They were **predecessors of micro-organisms** and have played a role in prebiotic chemical evolution.

Theory 4.

Viruses may have evolved from **genetic elements called LTR (long terminal repeat) retrotransposons (**found in the genomes of many eukaryotic species, especially plants). LTRs produce virion-like ovoid to spheroid particles that are not infectious, but are essential in the life cycle of the element; LTRs make up large % of plant genomes (48-69% of maize genome is retrotransposon sequences.)

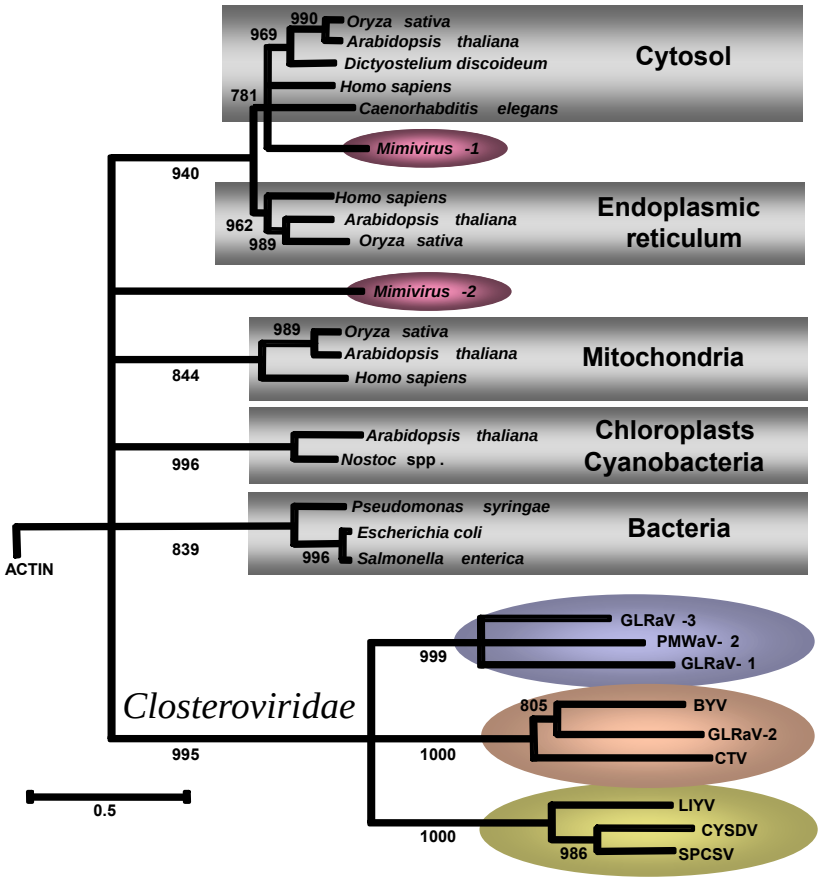
Theory 5.

Viruses could have **evolved from autonomous self-replicating host nucleic-acid molecules** such as plasmids and transposons by acquiring genes that code for encapsidating protein, further supporting independence.

Where do we find evidence for the origins of viruses?

Look in their genes

Category 1: Some virus genes have homologs in eucaryotic cells



Case in point:

There are *Hsp70* homologs in the genomes of viruses in *Mimivirus* and *Closteroviridae*

HSP 70:
Conserved ubiquitously expressed heat shock proteins. Proteins with similar structure exist in virtually all living organisms. The Hsp70s are an important part of the cell's machinery for protein folding, and help to protect cells from stress.

Dolja *et al*. (2006) *Virus Res.* **117**:38

Category 2: Viral Hallmark Genes

These are genes shared by many diverse groups of viruses

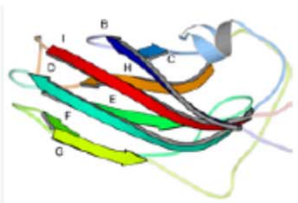
These are genes that have only distant homologs in cellular organisms

These genes play major roles in genome replication, packaging and assembly

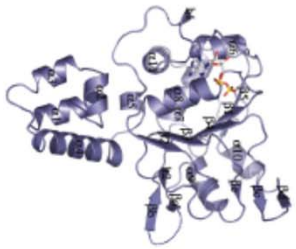
These genes provide strong support for monophyly for all viral members of their respective gene families

Can be viewed as distinguishing characters of the “virus state”

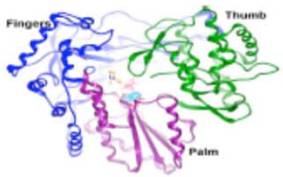
Proteins encoded by viral hallmark genes



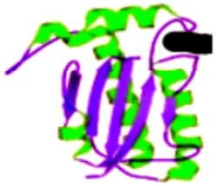
1. Jelly-roll capsid protein



2. Superfamily 3 helicase



3. RNA-dependent RNA polymerase and Reverse transcriptase



4. Rolling circle replication initiation endonuclease

5. Viral archaeo-eukaryotic DNA primase

6. UL9-like superfamily 2 helicase

7. Packaging ATPase of the FtsK family

8. ATPase subunit of terminase

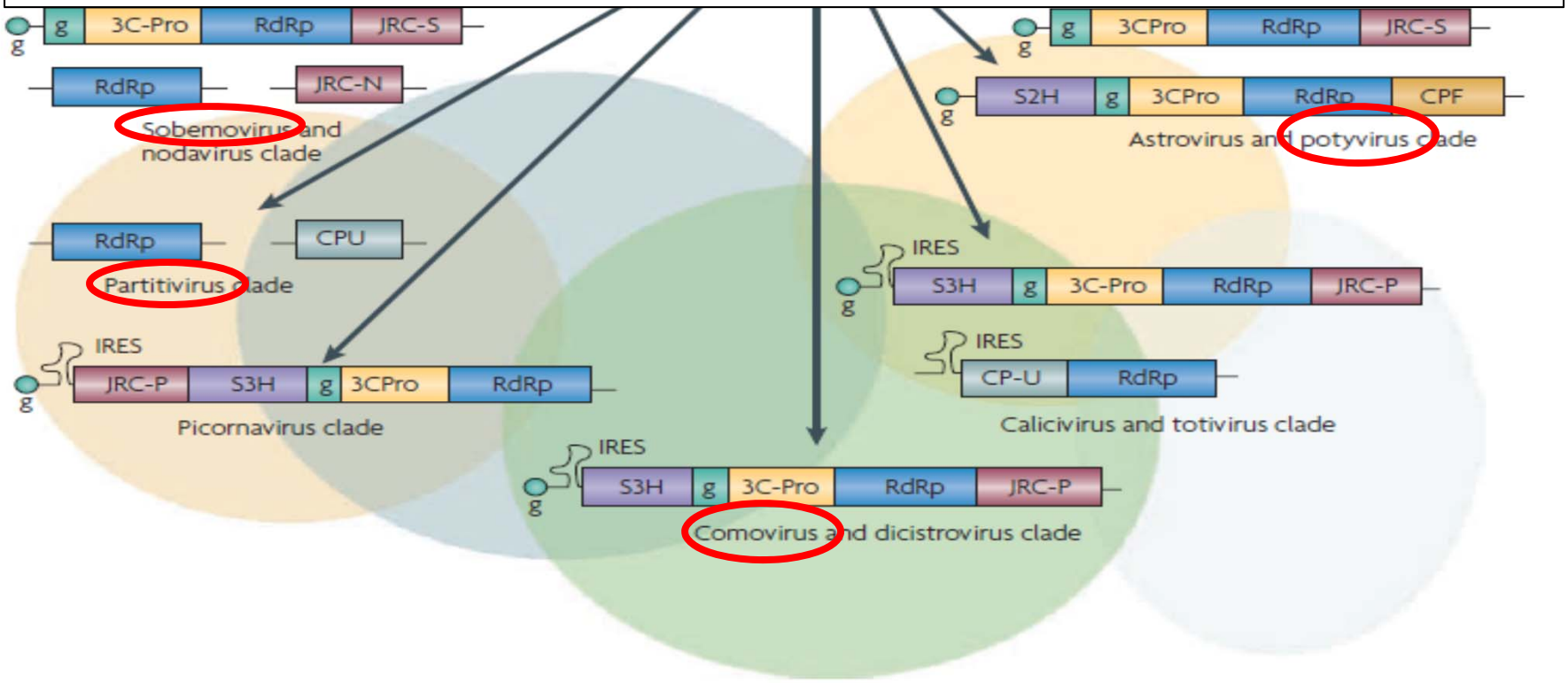
The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups

Eugene V. Koonin, Yuri I. Wolf*, Keizo Nagasaki‡ and Valerian V. Dolja§*

Abstract | The recent discovery of RNA viruses in diverse unicellular eukaryotes and developments in evolutionary genomics have provided the means for addressing the origin of eukaryotic RNA viruses. The phylogenetic analyses of RNA polymerases and helicases presented in this Analysis article reveal close evolutionary relationships between RNA viruses infecting hosts from the Chromalveolate and Excavate supergroups and distinct families of picorna-like viruses of plants and animals. Thus, diversification of picorna-like viruses probably occurred in a 'Big Bang' concomitant with key events of eukaryogenesis. The origins of the conserved genes of picorna-like viruses are traced to likely ancestors including bacterial group II retroelements, the family of HtrA proteases and DNA bacteriophages.

ORFs for equivalent proteins are present and organized in similar manners across very diverse RNA viruses:

RdRp (RNA-dependent RNA polymerase, **s3H** (superfamily 3 helicase),
CPro or sPro (chymotrypsin-like cysteine or serine proteases), **VPg or g** (viral protein, genome-linked), **JRC** (jelly-roll capsid protein).



Evolution: Origin of plant viruses

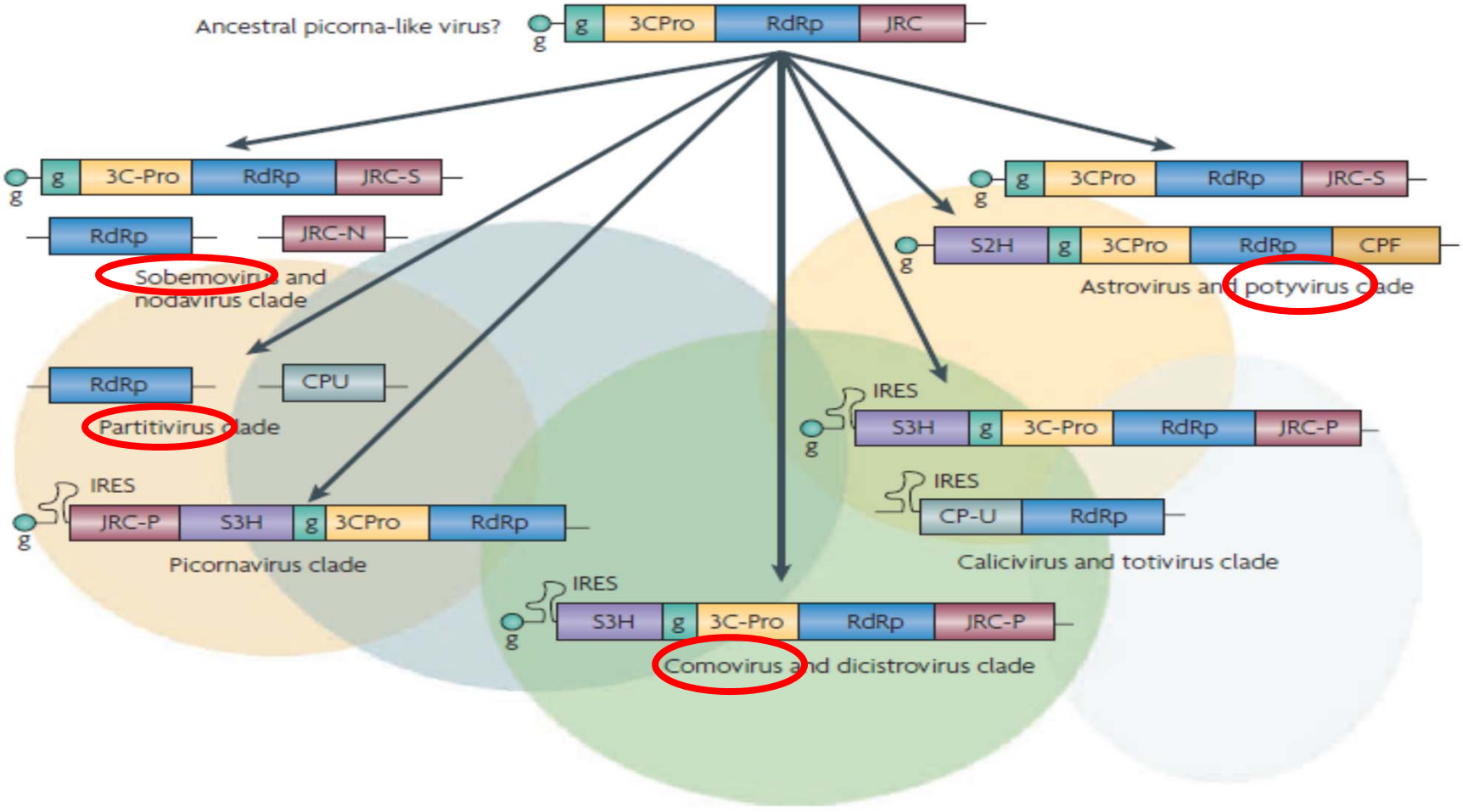
- Gene arrangement of core genes is similar among plant and animal viruses with the same type of genome

- Genome sequencing has shown that there are certain virus groups that share conserved domains (sequences within genes) in genes which encode similarly functioning proteins (ex. RNA polymerases)

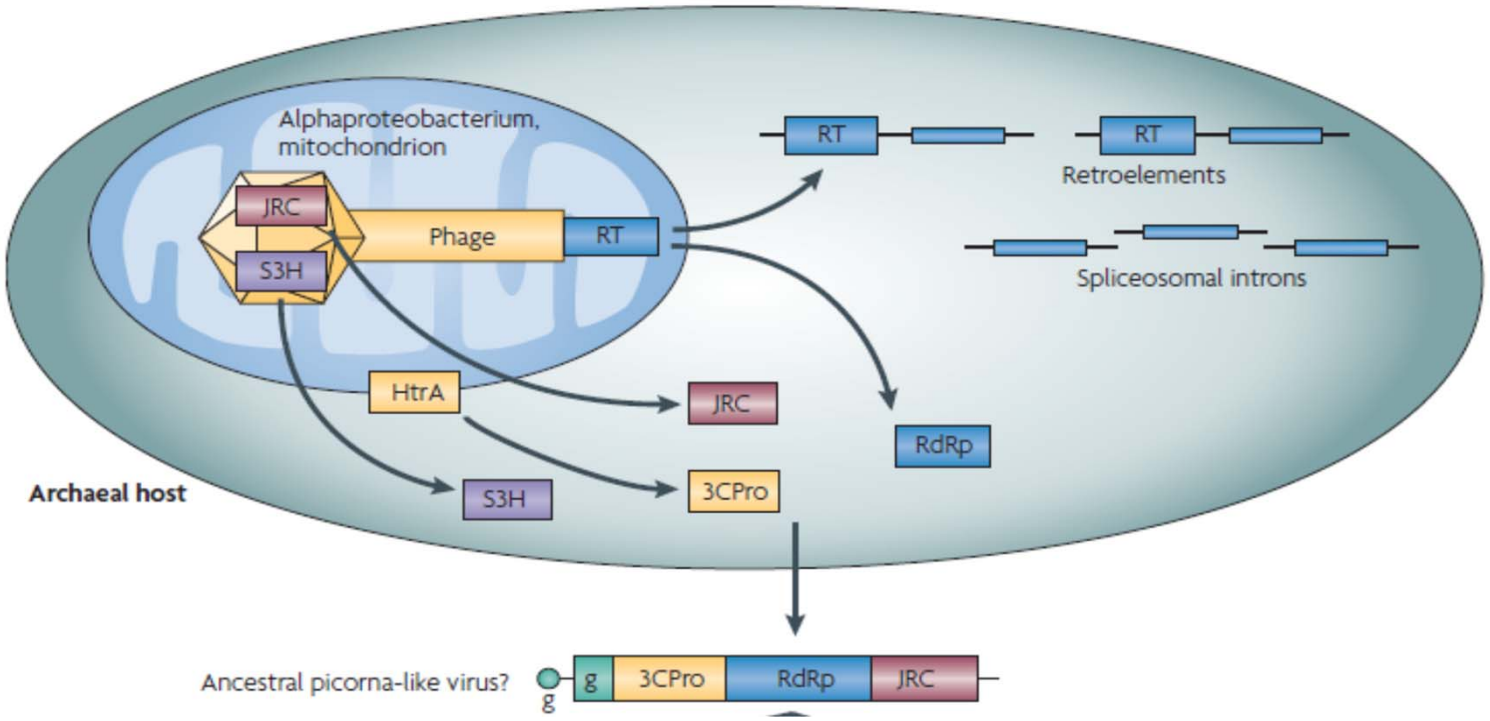
- Attempts have been made to assemble certain families of plant viruses and of animal viruses that share such feature into large supergroups or **superfamilies.**

  Ex. **Picorna-like viruses**

Example: Proposed Evolution of Six Clades of Picorna-like viruses



A proposed evolution of picorna-like viruses:

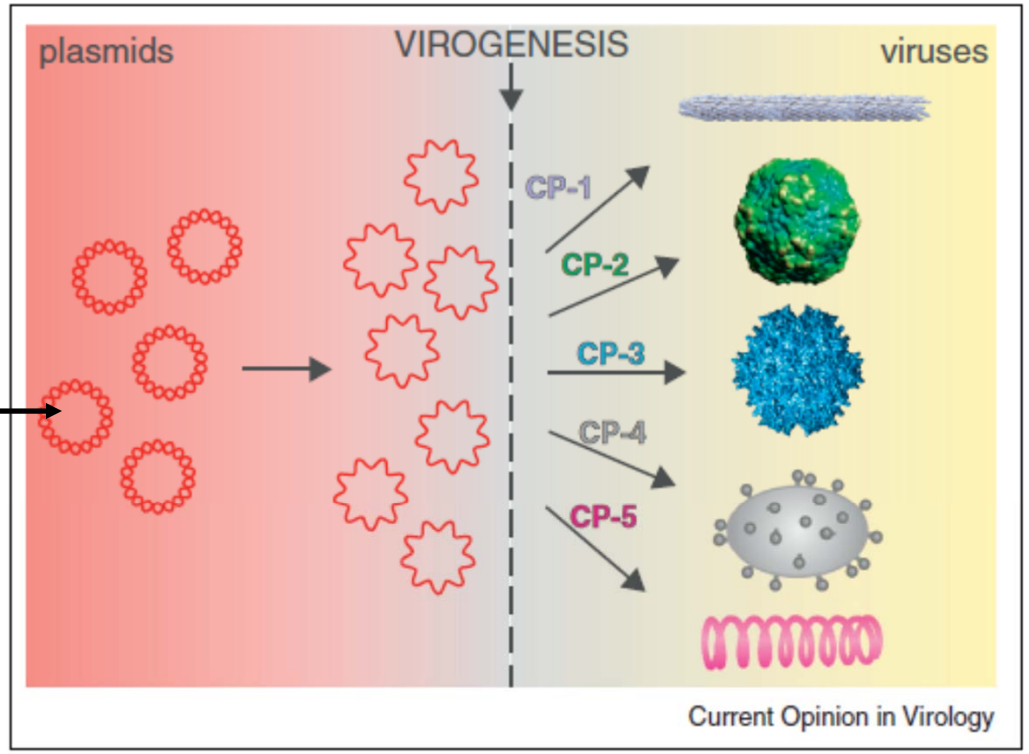


Viral genes were derived from multiple sources: bacterial retro-elements, DNA bacteriophages, HtrA proteins

HtrA: protease whose role is to degrade misfolded proteins in the cell periplasm (matrix between outer and inner bacterial membranes)

ssDNA viruses:
Proposed - Polyphyletic origin

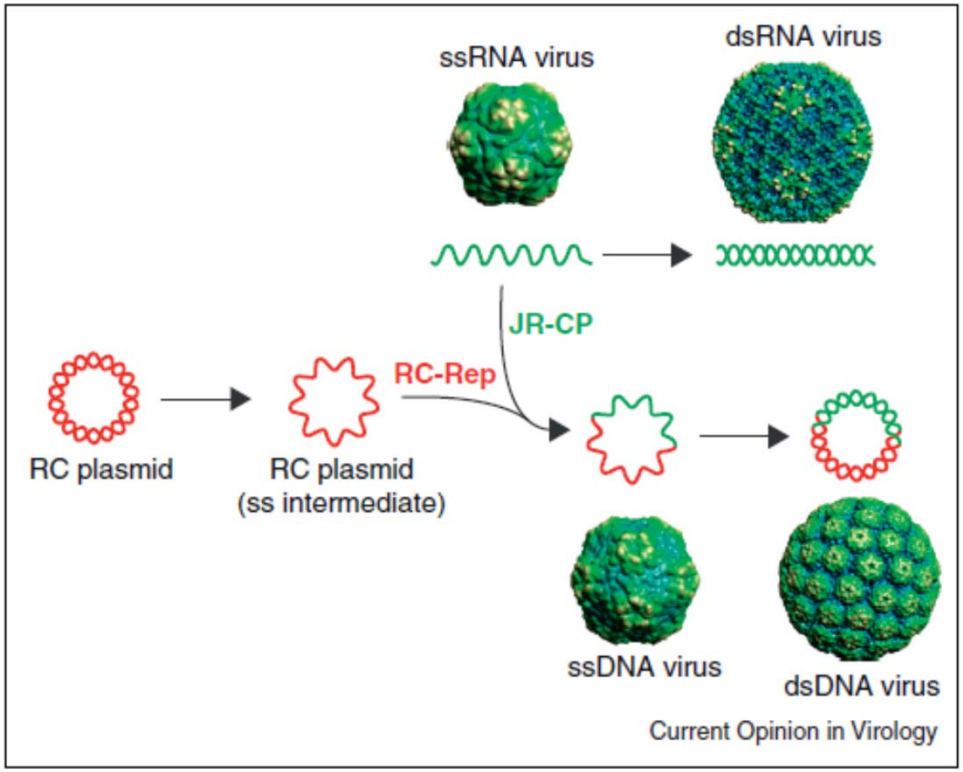
Plasmids that use RC = rolling circle Rep



Polyhyletic origin of ssDNA viruses. The scheme depicts independent origin of different lineages of ssDNA viruses from plasmids upon acquisition of virion formation modules (i.e. capsid proteins, CP1-5). Note that the five viral lineages do not share a common viral ancestor, even though the plasmids from which they derive might be related.

Krupovic 2013 Current Opinion in Virology 3:578–586

Proposed: RNA to DNA jump for the origin of ssDNA viruses:



A scheme depicting the 'RNA-to-DNA jump' scenario for the origin of ssDNA viruses. Emergence of dsRNA and dsDNA viruses from ssRNA and ssDNA viruses, respectively, is also shown. Images of the depicted virions were downloaded from the VIPER database (viperdb.scripps.edu/).

RC = rolling circle Rep

Krupovic 2013 Current Opinion in Virology 3:578–586