

Classification of endogenous viral elements

Welkin Johnson

Boston College, Biology Dept., 140 Commonwealth Ave., Chestnut Hill, MA 02467, USA
welkin.johnson@bc.edu

Pierre-Yves Teycheney

CIRAD UMR AGAP, Station de Neufchâteau, F-97130 Capesterre-BE, France-FWI
teycheney@cirad.fr



Endogenous viral elements (EVEs)

Endogenous viral elements are viral nucleotidic sequences that integrated into hosts' genomes by **horizontal gene transfer**:

- **actively** through the action of a virus-encoded **integrase**: retroviruses (family *Retroviridae*) - integration is an obligatory step in the virus' replication cycle
- **passively** by non-homologous end-joining (illegitimate recombination): *Geminiviridae* and *Caulimoviridae* sequences in plant genomes e.g – integration is not an obligatory step in the replication cycle

When integration occurs in **germline cells**, the proviral DNA is passed to future generations like a normal cellular gene and becomes **endogenous** (as opposed to transmission between individuals following conventional infection pathways)

Some endogenous viral elements are replication competent, however most of them are representative of extinct viral taxa



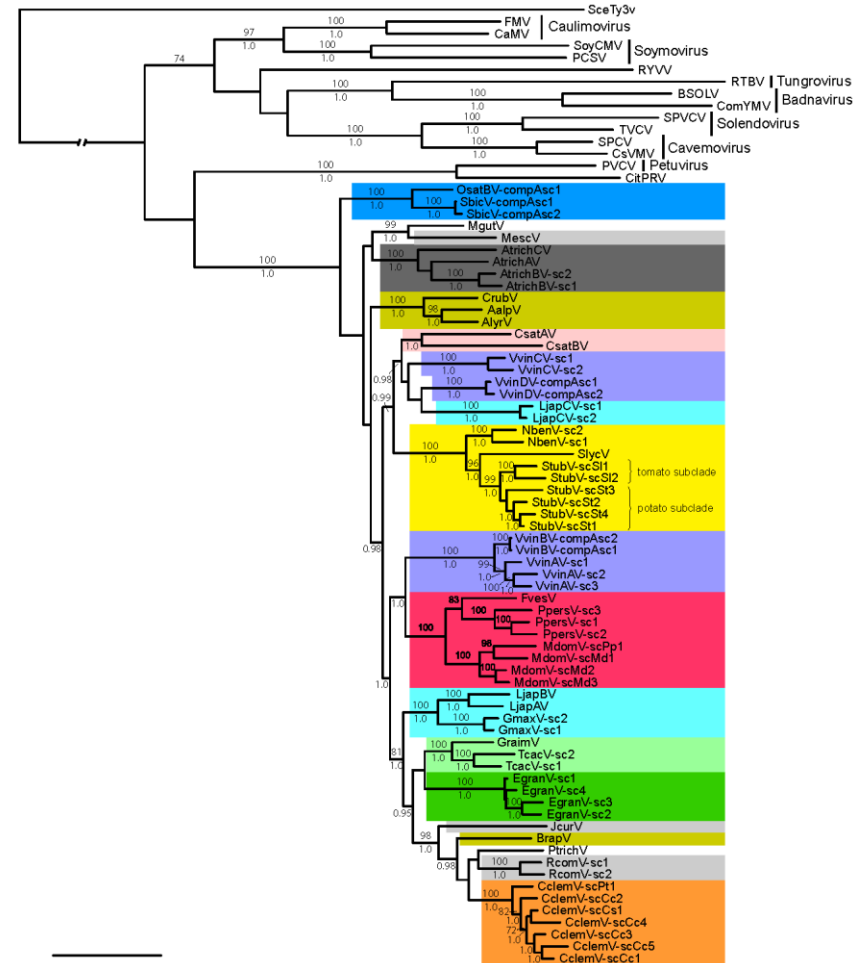
Endogenous loci as fossils representing extinct viral taxa

- Heritable, genetic loci (sequences) representing extinct viruses
- Degraded by substitutions, indels
- Subject to fixation/loss due drift & selection
- Can interleave with extant viruses in taxonomy
- Copy numbers vary :
 - up to thousands of copies in plant genomes
 - Up to hundreds of thousands of ERV copies per vertebrate genome (= billions of loci in the databases)
- 4 of 7 current *Caulimoviridae* genera represented
- 6 of 7 *Retroviridae* genera represented (ERV related to deltaretroviruses have not been discovered, yet)



EVEs provide more resolution to phylogenies and insight into viral origins - *Caulimoviridae*: Reverse Transcriptase (RT) Phylogeny

- Entire viral genomes of extinct viruses can be reconstituted in silico from endogenous sequences:
 - 76 entire or nearly full-length genomes assembled from endogenous viral sequences, corresponding to **34 distinct species defining a new genus in the family *Caulimoviridae*** tentatively named Florendovirus
 - Provide access to fossil sequences of whole virus genera that are now extinct : rice florendovirus sequences are at least 1.8 Myo



Geering et al. (2014) Nature Comms Nature Communications 5: 5269



EVEs provide more resolution to phylogenies and insight into viral origins - *Retroviridae*: Reverse Transcriptase (RT) Phylogeny

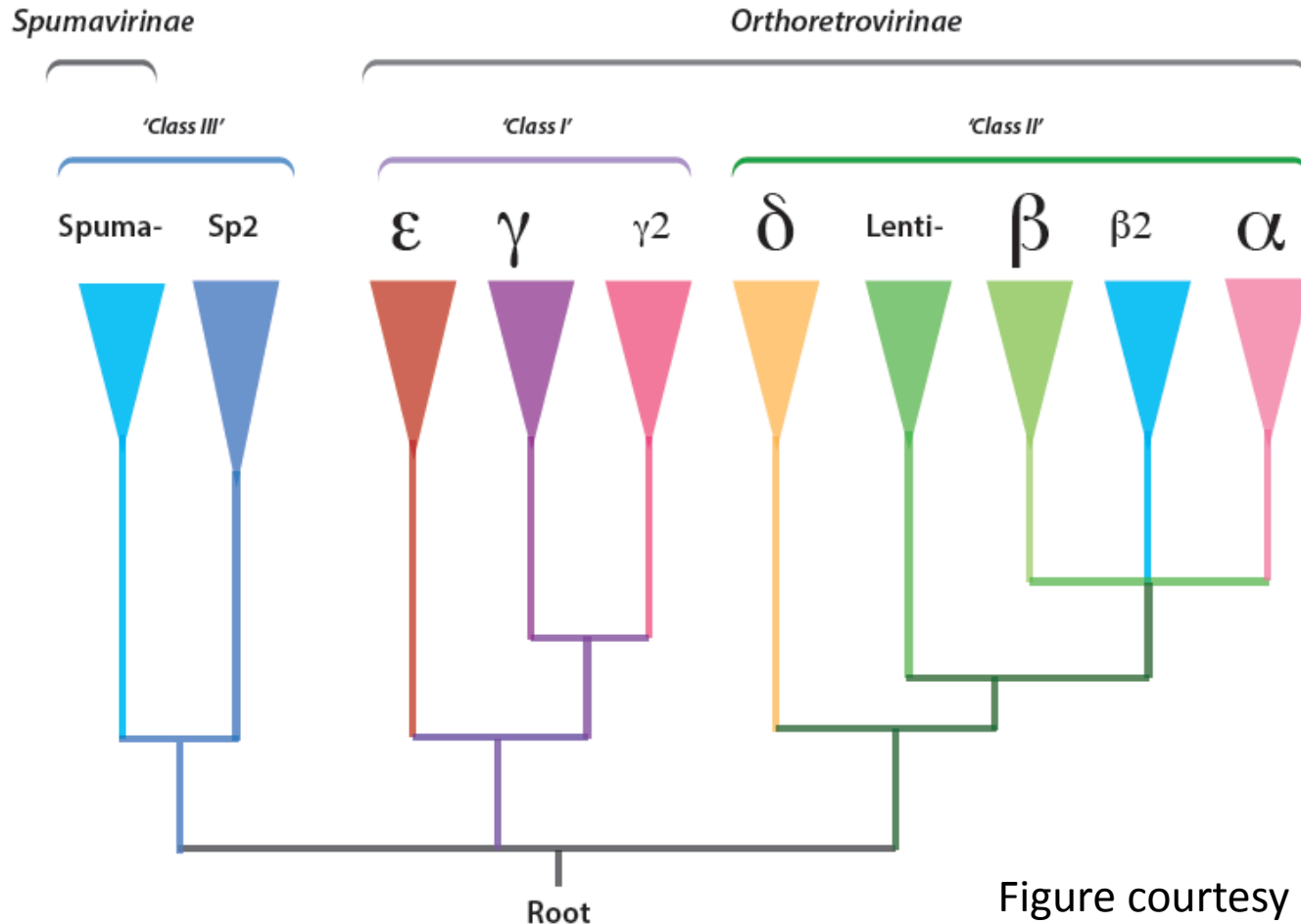


Figure courtesy of R. Gifford

Why it is important to tackle the issue of the taxonomy of EVEs

- To avoid confusion that would result from unorganized naming, such as different names being given to similar sequences
- **Need to provide single governing rules guiding how sequences should be named**
- To prevent taxonomic terms such as family being used to describe different sequences that are no more divergent than members of a virus genus
- **Need to include EVEs in the taxonomy**



Why it is important to tackle the issue of the taxonomy of EVEs



Lucy

Two nomenclatures:

- As a fossil specimen – **AL 288-1**
- As evidence for an extinct hominin: *Australopithecus afarensis*

Points for further discussion

- Endogenous loci are “fossils” representing extinct viral taxa
- Need to be incorporated into viral taxonomy just as real fossils are used to classify extinct organisms
- Incorporation of ERV and EVE into taxonomy provides more resolution to phylogenies, insight into viral origins, natural history and macroevolution
- Endogenous loci require a nomenclature that simultaneously reflects their viral origins AND their existence as genetic loci
- Additionally, these provide evidence for extinct species which will require a nomenclature and incorporation into viral taxonomy
- ERV data are vast, and functional nomenclature is urgently needed and should be in place to facilitate incorporation into taxonomic proposals

