

# Viral diversity

Duncan Golicher

26 March 2023

# What is known about viruses?

- ▶ Example textbook. Field's virology
- ▶ Well over 2000 pages per volume (2 volumes)
- ▶ Each chapter apologises for what is **not** included
- ▶ Uncertainties, known unknowns, unknown unknowns: but very few known knowns
- ▶ Most chapters focus on pathology and molecular biology
- ▶ **Ecology** of viruses remains poorly understood

## Web of science

- ▶ The pandemic has brought viruses into the news
- ▶ However there is a shortage of literature on the **ecology** of viruses
- ▶ Search terms “virus” + “ecology”
- ▶ Search term “coronavirus”

**Results: 4,998**

*(from Web of Science Core Collection)*

**You searched for: TOPIC: (virus) AND TOPIC: (ecology) ...More**

---

**Results: 57,628**

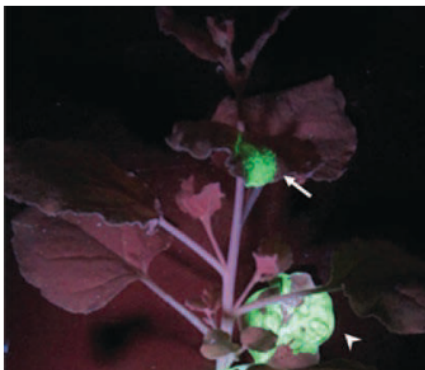
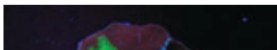
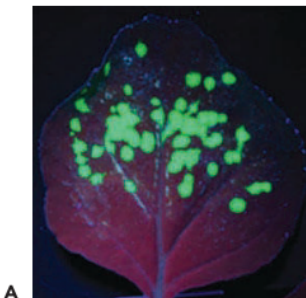
*(from Web of Science Core Collection)*

## Quick history

- ▶ Viruses not recognised until 1880s
- ▶ Adolf Meyer working on tobacco at Wageningen identified TMV
- ▶ Clearly an infectious agent, but could not fulfill Koch's postulates as only cultivable in living cells:
- ▶ The issue with Koch's postulate continues to muddy some debates to this day
- ▶ Beijerinck named the agent a "contagium vivum luidum" or a contagious living liquid.
- ▶ d'Herelle developed the plaque assay in 1917
- ▶ First electron micrographs were taken of tobacco mosaic virus (TMV) in 1939.

## Tobacco mosaic virus

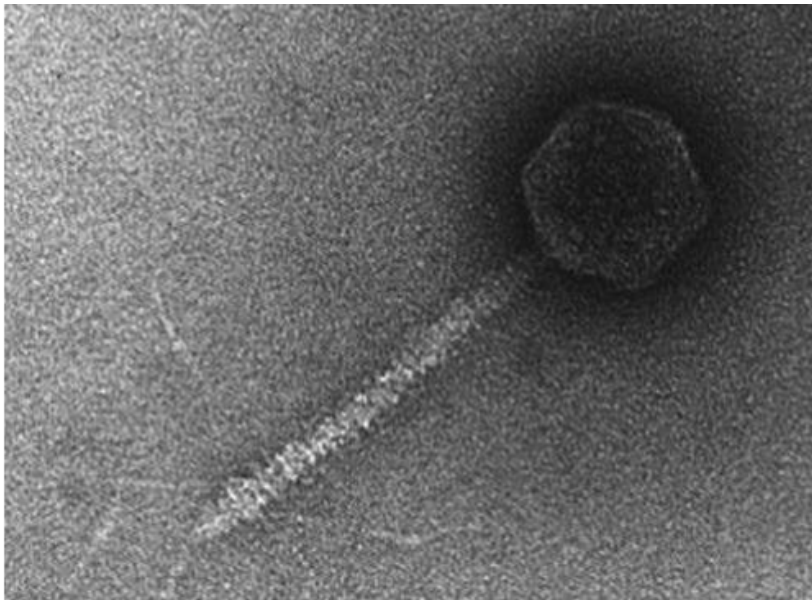
- ▶ Remains a key experimental organism
- ▶ Plant viruses are possibly more diverse than higher animal viruses (more species of plant)
- ▶ Plant viruses have a different set of challenges (movement through an organism with cell walls)
- ▶ Plant viruses are not thought to infect animals directly (possible exceptions)



# Bacteriophages

- ▶ Lytic (pathogenic) and lysogenic (non pathogenic) infections identified in 1900s
- ▶ Experiments with viruses resulted in the discovery of mRNA and the role of DNA
- ▶ First complete nucleotide sequence of an RNA virus genome was the bacteriophage MS2 (1976)
- ▶ Most abundant “organisms” on the planet
- ▶ Present at roughly  $10^7/mL$  in coastal seawater
- ▶ Around  $10^{31}$  individual tailed phage virions on the planet
- ▶ If all the tailed phages were laid end to end, they would extend for 200 million light years into intergalactic space.

## Phage seen through electron microscope

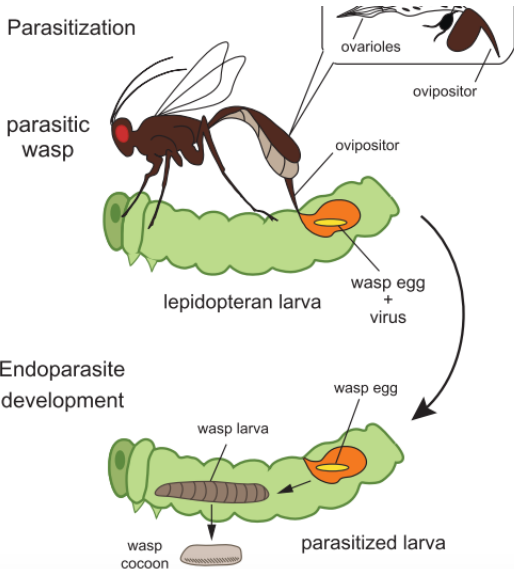


## Insect viruses

- ▶ Abundant, ubiquitous and highly diverse
- ▶ Mammalian and avian viruses may have evolved in insects
- ▶ Insects are also vectors for both animal and plant viruses
- ▶ Can play a role in very complex life histories
- ▶ Some extremely complex symbiotic relationships



# Parasitic wasps use viruses as “venom”

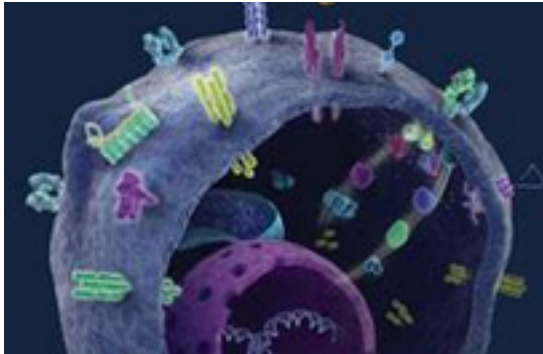


**FIGURE 73.8. Polydnavirus transmission.**

Polydnavirus particles form in the epithelial cells of adult female ovaries and are released into the calyx fluid (*top inset*) where they accumulate to high densities. In the first step of parasitization (*top*), the wasp egg is injected along with virus and venom into a lepidopteran larva during oviposition. During endoparasitic development (*bottom*), polydnavirus-encoded gene products suppress the larval immune response allowing the wasp egg to develop. After parasitization, the wasp larva spins a cocoon, pupates, and emerges as an adult to repeat the process. (Courtesy of Dr. Bruce A. Webb, University of Kentucky; modified with permission.)

## Mammalian viruses

- ▶ Experiments on viruses of animals used immortalised human cell lines and monkey kidney cells
- ▶ Polio virus was an early focal virus
- ▶ Virology allowed key insights into oncology and cancer
- ▶ Cellular receptors began to be identified and studied in depth
- ▶ Attention mainly on pathogenic viruses of humans (and some of rodents and domestic animals)
- ▶ Full diversity of mammalian viruses remains unknown



# Viral evolution

- ▶ Fundamental aspects of viral evolution remains unknown, contentious, or both
- ▶ Precise estimates of mutation rates unknown
- ▶ Deep origins of viruses still uncertain
- ▶ Did viruses predate cellular organisms?
- ▶ “Escaped gene theory” vs “RNA first” theory
- ▶ Did the earliest replicators resemble what we now know as viruses?
- ▶ Should viruses even be classified as life?

# Evolution and phylogeny

- ▶ Viruses generally evolve at high rates, following basic and extended Darwinian principles
- ▶ Evolutionary time frame much shorter than that of any other organism.
- ▶ Species concept cannot be uncontentiously applied.
- ▶ True antiquity of currently circulating viruses still vague.
- ▶ No fossil record

# Baltimore classification

Highest level of classification of viruses

- ▶ Split into 7 groups based on type of genetic material
- ▶ DNA viruses (single and double stranded)
- ▶ RNA viruses (double stranded, single stranded positive and negative sense )
- ▶ Reverse transcribing viruses

## Number of viral species

- ▶ Perhaps over one million viruses of vertebrates
- ▶ ~20 different viruses in each of the 50,000 vertebrates
- ▶ “Minimum of 320,000 mammalian viruses awaiting discovery”  
(Anthony et al. 2013)

## Species concept as applied to viruses

- ▶ Importance of species in biology derives from their importance in systematics
- ▶ Prior to the 1980s the notion of viral “species” was rejected by most of the scientific community
- ▶ Clearly could not apply Ernst Mayr’s definition(“*species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups*”)
- ▶ Concepts of species, strains, variants and viral isolates caused confusion (Alimpiev 2019)
- ▶ 1982 International Committee on Taxonomy of Viruses(ICTV): species = “a cluster of strains from a variety of sources, or a population of strains from a particular source, which **have in common a set of stable properties** that separate **the cluster** from other **clusters of strains.**”

## Modified definitions

- ▶ Van Regenmortel: “a polythetic class of viruses (*strains*) that constitute a replicating lineage and occupy a particular ecological niche”
- ▶ Polythetic is defined as “characteristics which occur commonly in members of a group or class, **none of which is essential** for membership of that group or class.”
- ▶ 2012 ICTV viral species definition based on shared ancestry. “A monophyletic **group of viruses (strains)** whose properties can be distinguished from those of other species by multiple criteria”
- ▶ Can be used to define any virus **taxon**. (I.e. not just a species)
- ▶ Current definition essentially defines virus species (clades) based **only on their genomic sequence**
- ▶ However genetic sequences are a continuum, with no naturally defined breaks



# Time since divergence

Knowing time since divergence key to understanding viral evolution and viral “novelty.”

▶ Three approaches

1. Find a good match between the phylogenetic tree of viruses and that of their hosts. Assume co-divergence (not usually possible as few viruses are narrowly host specific .. can work for DNA viruses .. not generally RNA)
2. Find endogenous genome copies within the host: Use the substitution rate of the host (similar to 1)
3. **Measure current evolution (i.e mutations fixed during time frame of human observation)**

# RNA viruses

- ▶ Particularly difficult to produce phylogenies
- ▶ Rarely strictly host specific
- ▶ Very fast rates of mutation
- ▶ Many RNA virus populations thought to exist as **quasispecies**
- ▶ Swarms of individual viral genomes where every member is unique.

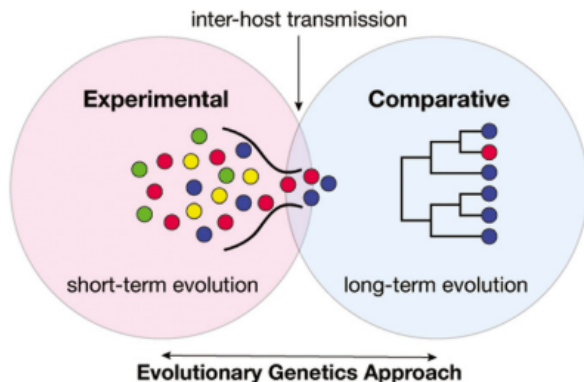
## Molecular clock

- ▶ Rate of evolution has been estimated for polio (short RNA ~7500 base pairs)
- ▶ Estimated that there is one synonymous substitution per 100 base pairs per year (Jorba et al. 2008)
- ▶ Non-synonymous substitutions constrained by purifying selection
- ▶ However, these measurements were on a population with low prevalence
- ▶ Hard to generalise, as rate of evolution depends on number of infections

## Evolution involves the host

- ▶ Experimental approaches using in vitro and in vivo **laboratory models** focus on short-term **intrahost evolutionary mechanisms**
- ▶ These may not always be relevant to natural systems.
- ▶ Should consider data collected over multiple cycles of virus–host transmission. (Geoghegan and Holmes 2018)
- ▶ Evolution can take place at different rates and follow different lines in different host populations

# Geoghegan and Holmes (2018)

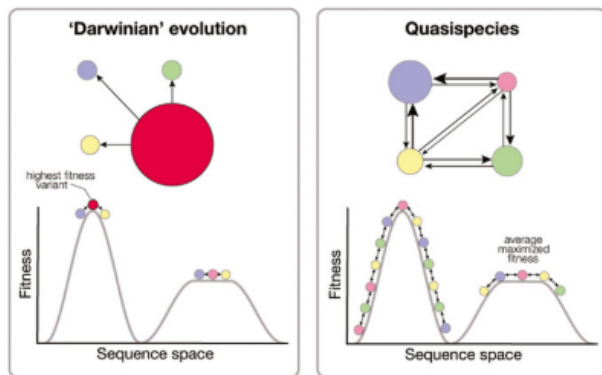


**Figure 1** Approaches to studying RNA virus evolution. The Venn diagram illustrates the two historical, and largely parallel, strands of research in virus evolution—the experimental and the comparative—that arose in the late 1970s. They generally only overlap in the study of a limited number of interhost virus transmission events that often involve a substantial population bottleneck. Through the use of *in vitro* or *in vivo* model systems, experimental studies largely focus on evolution in the short-term, particularly that which occurs within individual hosts. In contrast, comparative approaches deal with interhost, epidemiological-scale dynamics that entail multiple rounds of interhost transmission and are usually based on phylogenetic analysis. We argue that the two evolutionary approaches

## Quasispecies evolution

- ▶ Mutation–selection balance in which a distribution of variant viral genomes is ordered around the fittest, or “master,” sequence.
- ▶ Mutation rates in RNA viruses are so high that the frequency of any variant is not only a function of its own replication rate (fitness), but also of the mutation of other variants in the population derived from the original sequences
- ▶ “Mutational coupling” leads to a distribution of evolutionarily interlinked viral genomes
- ▶ Entire mutant distribution behaves as a single unit
- ▶ Natural selection acts on the mutant distribution as a whole, rather than on individual variants.
- ▶ The “quasispecies” as a whole evolves to maximize its average fitness.
- ▶ Has been applied to SARS-Cov-1 (Xu, Zhang, and Wang 2004) and SARS-CoV-2 (Jary et al. 2020)

# Quasispecies evolution



**Figure 2** "Darwinian" vs. quasispecies models of RNA virus evolution. In the Darwinian virus population, natural selection favors the variant with the highest individual fitness (circle shown in red), with lower-fitness variants (blue, green, and yellow) produced by mutation at a relatively low rate. Under the quasispecies model, very high mutation rates lead to a mutational coupling among variants (of different colors). This, in turn, means that the viral population evolves as a single unit, with the mutational landscape greatly impacting virus evolution and natural selection acting on the population as a whole, maximizing mean fitness. In the top part of the figure, the circle sizes represent relative fitness values, whereas

# Recombination

- ▶ Many, many examples of hybridisation in plants
- ▶ Some examples in well known animals (e.g edible frogs)
- ▶ Can be very difficult to establish as hybrids unless the sequences of all species involved are known.
- ▶ The equivalent of Tignons, Ligers and edible frogs also exist in the world of viruses:
- ▶ Known as **Recombinants**
- ▶ The result of simultaneous infection with different viral sequences

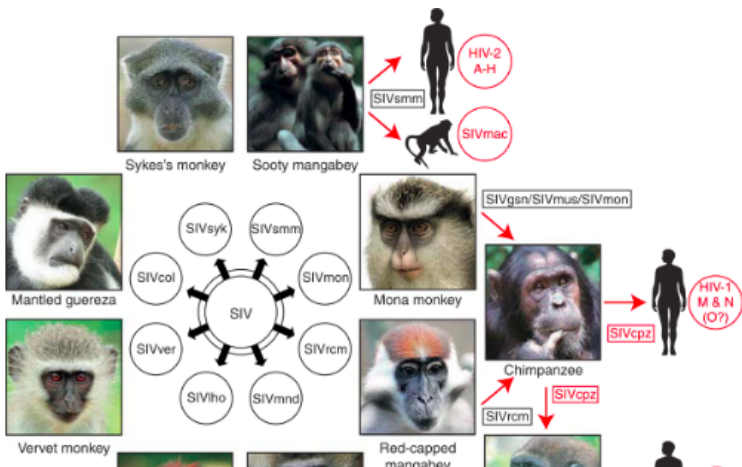


## “Hybrid viruses”: Influenza

- ▶ Influenza A, B, C and D are genera
- ▶ Enveloped negative sense single strand RNA viruses with a segmented genome
- ▶ Cell entry through glycoproteins: haemagglutinin (HA) and neuraminidase (NA)
- ▶ 16 antigenically different HA and 9 antigenically different NA serotypes
- ▶ Interchange, of genomic RNA segments occurs when two viruses of the same type infect the same cell.

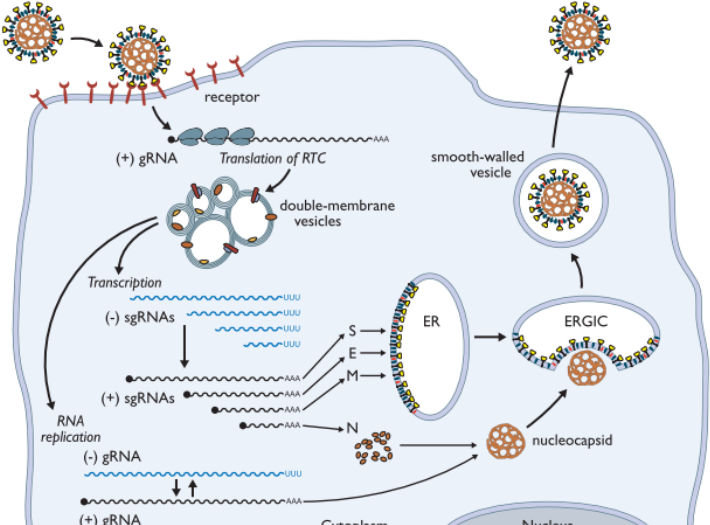
# Recombination is SIV (simian IV) and Ebola

- ▶ SIVcpz represents a complex mosaic, generated by **recombination** of two lineages of SIVs that infect monkeys (Sharp and Hahn 2011)
- ▶ Multiple ebola recombinants in simians (Wittmann et al. 2007)



# Coronavirus subgenomic RNA

- ▶ Coronavirus viral RNA synthesis involves the transcription of multiple sub-genomic RNAs
- ▶ Potential for recombination unclear, but highly likely



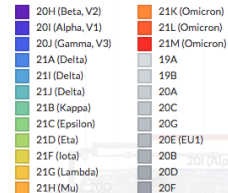
## Putative phylogeny of SARS-CoV-2 clades

- ▶ Divergence based on assumptions regarding conserved characteristics of the sequence.
- ▶ Dates rely on assumed molecular clock
- ▶ Most sequences have gone extinct
- ▶ Extant (still circulating) sequences appear to have reduced virulence

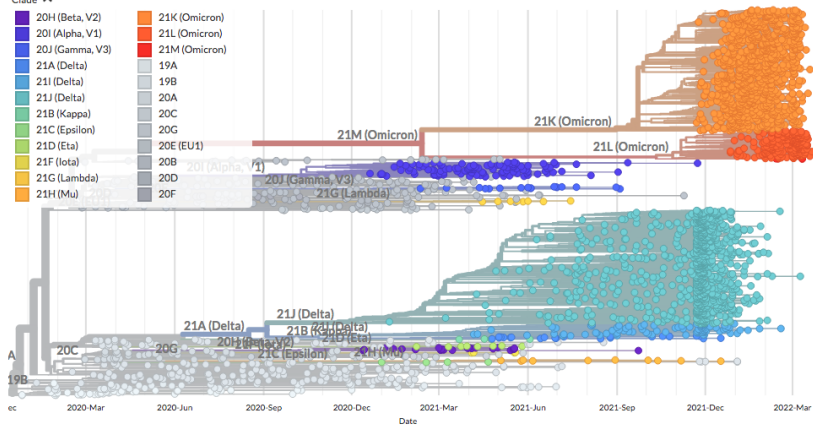
# Putative phylogeny of SARS-CoV-2 clades

## Phylogeny

Clade ^



🔍 ZOOM TO SELECTED | RESET LAYOUT



## References

Alimpiev, Egor. 2019. "Rethinking the Virus Species Concept" 25.

Anthony, Simon J., Jonathan H. Epstein, Kris A. Murray, Isamara Navarrete-Macias, Carlos M. Zambrana-Torrel, Alexander Solovyov, Rafael Ojeda-Flores, et al. 2013. "A strategy to estimate unknown viral diversity in mammals." *mBio* 4 (5): 1–15. <https://doi.org/10.1128/mBio.00598-13>.

———, et al. 2013. "A strategy to estimate unknown viral diversity in mammals." *mBio* 4 (5): 1–15. <https://doi.org/10.1128/mBio.00598-13>.

Geoghegan, Jemma L., and Edward C. Holmes. 2018. "The phylogenomics of evolving virus virulence." *Nature Reviews Genetics* 19 (12): 756–69. <https://doi.org/10.1038/s41576-018-0055-5>.

———. 2018. "The phylogenomics of evolving virus virulence." *Nature Reviews Genetics* 19 (12): 756–69. <https://doi.org/10.1038/s41576-018-0055-5>.