# Genomes and Genetics

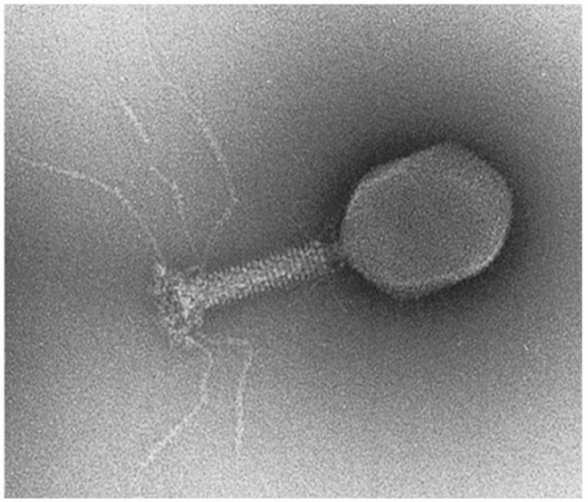
Lecture 3
Biology W3310/4310
Virology
Spring 2016

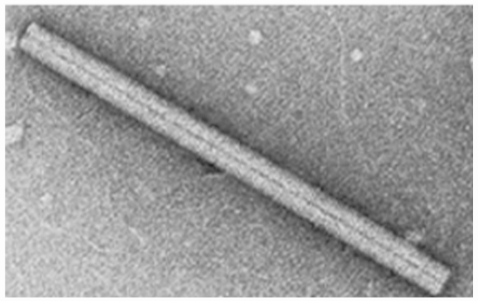
*"...everywhere an interplay between nucleic acids and proteins; a spinning wheel in which the thread makes the spindle and the spindle the thread"*
--ERWIN CHARGAFF

# Virology breakthrough in the 1950's:

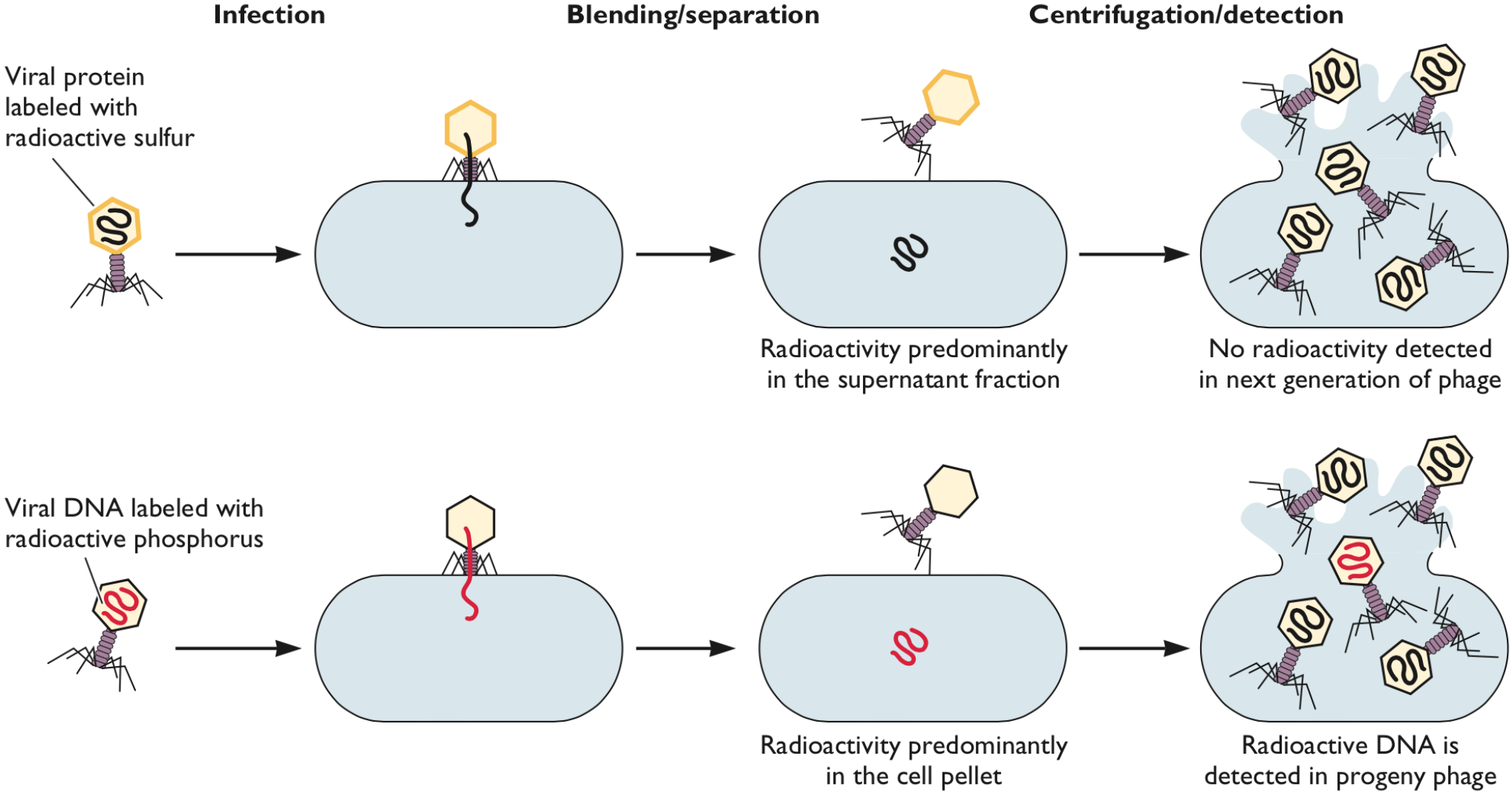
***The viral nucleic acid genome*** *is the genetic code*

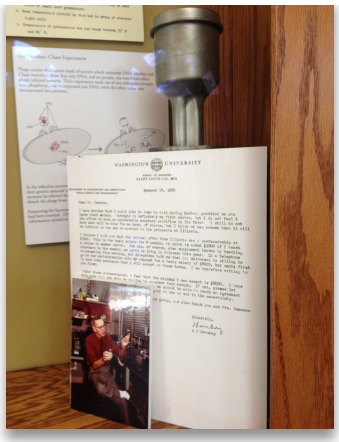
*Hershey-Chase experiment with phage T4*




*Fraenkel-Conrat's work with TMV*



Alfred Hershey & Martha Chase, 1952

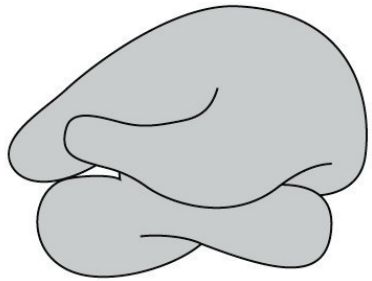
**The bigger surprise: thousands of different virions, seemingly infinite complexity of infections**

***But a finite number of viral genomes***

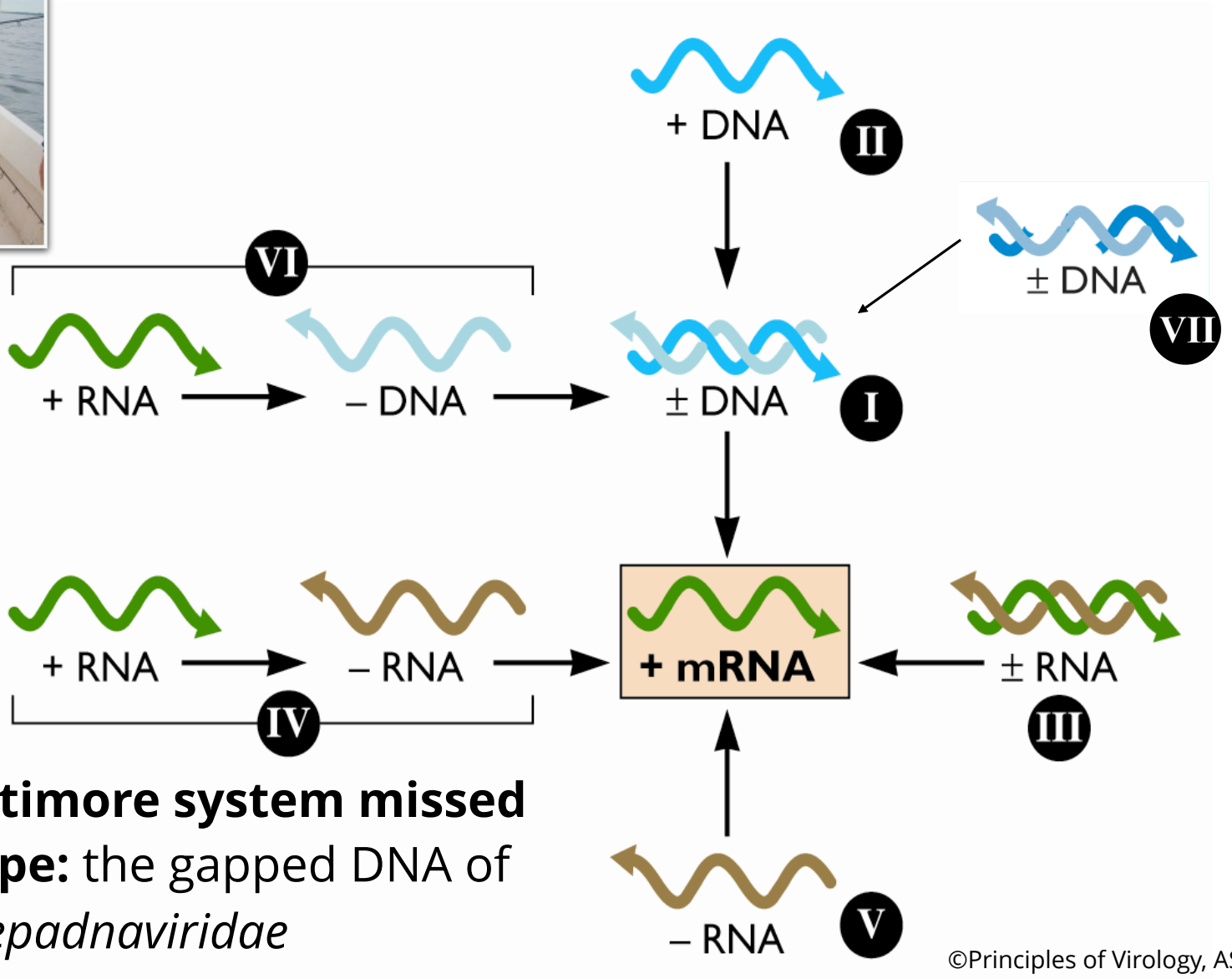
**Key fact makes your life easier:**

***Viral genomes must make mRNA that can be read by host ribosomes***

**All viruses on the planet follow this rule, no exception to date**



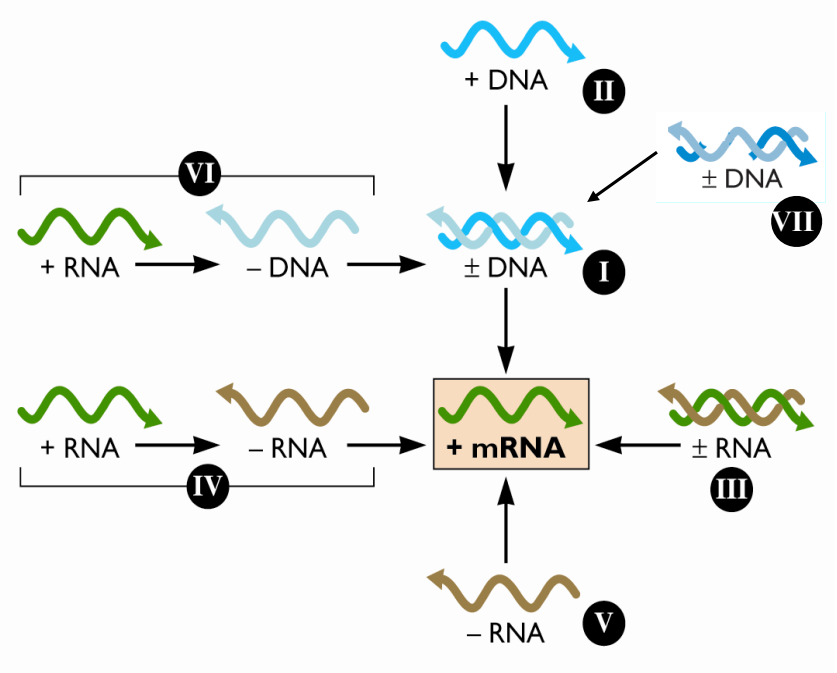
# David Baltimore (Nobel laureate) used this insight to describe a simple way to think about virus genomes




**The original Baltimore system missed one genome type:** the gapped DNA of the *Hepadnaviridae*

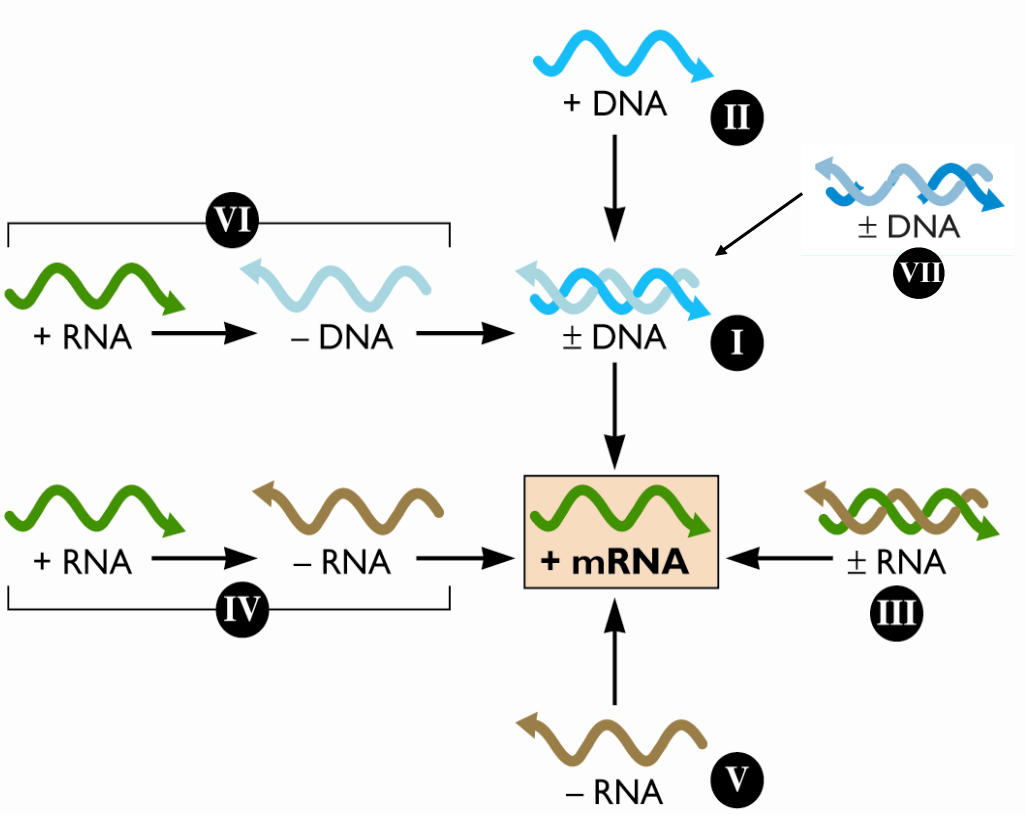
# Definitions

- mRNA (ribosome ready) is always the plus (+) strand
- DNA of equivalent polarity is also the (+) strand
- RNA and DNA complements of (+) strands are negative (-) strands
- Not all (+) RNA is mRNA!



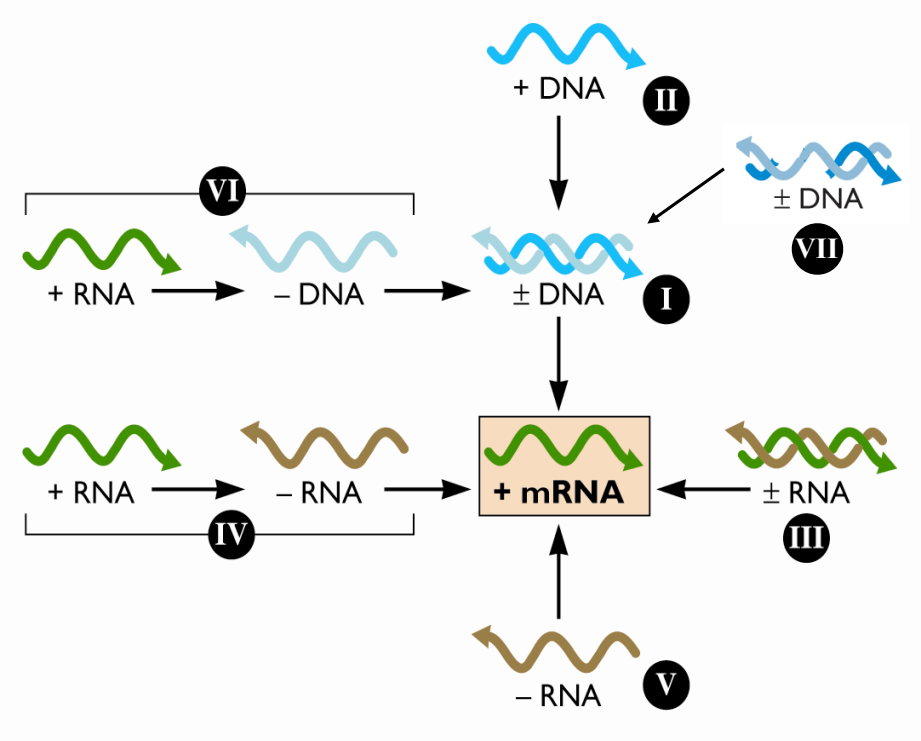
# The elegance of the Baltimore system

*Knowing only the nature of the viral genome,*
*one can deduce the basic steps that must take place*
*to produce mRNA*



# The seven classes of viral genomes

- dsDNA
- gapped dsDNA
- ssDNA
- dsRNA
- ss (+) RNA
- ss (-) RNA
- ss (+) RNA with DNA intermediate



Go to:

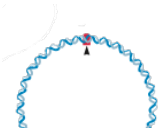
m.socrative.com
room number: virus

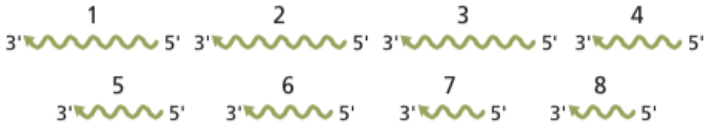
**Why is mRNA placed at the center of the Baltimore scheme?**

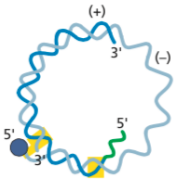
1. Because all virus particles contain mRNA
2. There is no specific reason
3. Because all viral genomes are mRNAs
4. Because mRNA must be made from all viral genomes
5. Because Baltimore studied mRNA

# Viral DNA or RNA genomes are structurally diverse

- Linear
- Circular
- Segmented
- Gapped
- Single-stranded (+) strand
- Single-stranded (-) strand
- Single stranded, ambisense
- Double-stranded
- Covalently attached proteins
- Cross-linked ends of double-stranded DNA
- DNA with covalently attached RNA

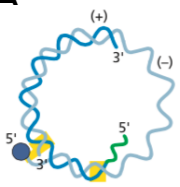







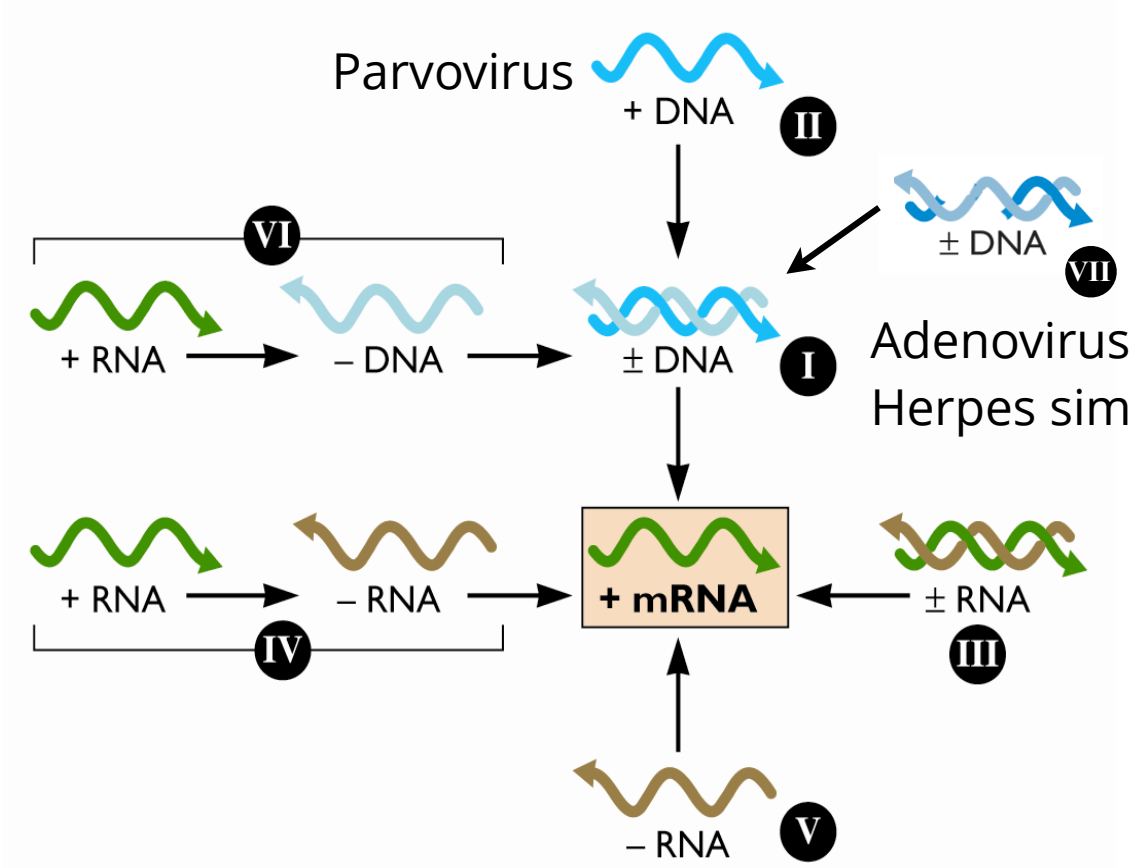


# What is the function of genome diversity?

- DNA *and* RNA based
  - RNA genomes appeared first in evolution (RNA World)
  - Switch to DNA genomes
  - Only RNA genomes on planet today are viral
  - Viroids: Relics of RNA world?
- Linear, circular, segmented, ds, ss, (+), (-)

# Memorize 7 genome types and key virus families




*If you know the genome structure you should be able to deduce:*

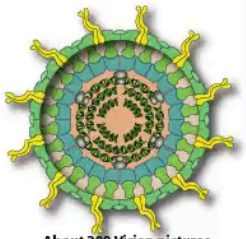
How mRNA is made from the genome

How the genome is copied to make more genomes

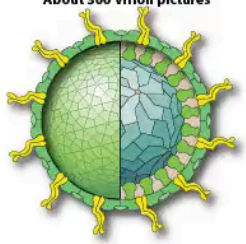
DNA viruses | RNA viruses | Retro-transcribing viruses

ds DNA | ss DNA | dsRNA | ss RNA (+) | ss RNA (-) | ss RNA (RT) | ds DNA (RT)





About 300 Virion pictures



## VIROLOGY NEWS

TWIV:This Week in Virology is a weekly podcast animated by professors Vincent Racaniello and Dick Despommier from the Columbia University, USA.

Virology and Bioinformatics from Virology.ca
*Virus and bioinformatics articles with some microbiology and immunology thrown in for good measure* Curated by Dr. Chris Upton

## VIRALZONE NEWS

**ICTV taxonomy update** September 2014 The International Committee on Taxonomy of Viruses has released Virus Taxonomy: 2013, voted and accepted in 2014. 7 new *Families*, and 47 new *Genera*. Viralzone taxonomy has been updated accordingly, new taxonomic pages are marked by the New logo.

**Ebolavirus and measles updated** August 2014

The structure of filoviruses nucleocapsid have been refined in 2012(PubMed), and morbiliviruses nucleocapsid as well (PubMed), and the ViralZone figures have been updated accordingly. Moreover an ebola virus cycle is available.

>>News archive

**Publication**: ViralZone: a knowledge resource to understand virus diversity. Hulo C, de Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I, Le Mercier P. Nucleic Acids Res. 2011 Jan;39(Database issue):D576-82.

### ViralZone current statistics

October, 2014
**526** Virus description pages:

**97** *Families*
**420** *Genera*
**9** individual *Species*

**211** Viral molecular biology pages

Linking to:
(UniProt release 2014_08)

**401** reference strains
**16,479** manually reviewed proteins
**2,075,771** unreviewed viral proteins

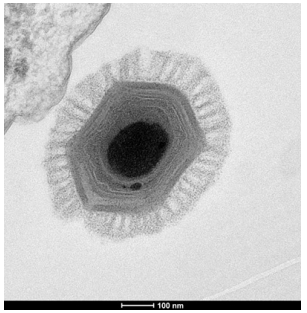
# What information is encoded in a viral genome?

*Gene products and regulatory signals for:*

- Replication of the viral genome
- Assembly and packaging of the genome
- Regulation and timing of the replication cycle
- Modulation of host defenses
- Spread to other cells and hosts

# Information NOT contained in viral genomes

- No genes encoding the *complete* protein synthesis machinery (AARS, eIFs, tRNAs)
- No genes encoding proteins involved in energy production or membrane biosynthesis
- No classical centromeres or telomeres found in standard host chromosomes
- Probably we haven't found them yet - 90% of giant virus genes are novel



# Largest known viral genomes

| Virus | Length | Protein |
|---|---|---|
| Pandoravirus salinus | 2,473,870 | 2,541 |
| Pandoravirus dulcis | 1,908,524 | 1,487 |
| Megavirus chilensis | 1,259,197 | 1,120 |
| Mamavirus | 1,191,693 | 1,023 |
| Mimivirus | 1,181,549 | 979 |
| Moumouvirus | 1,021,348 | 894 |
| Mimivirus M4 | 981,813 | 620 |
| *C. roenbergensis* virus | 617,453 | 544 |
| Mollivirus sibericum | 651,000 | 523 |
| Pithovirus sibericum | 610,033 | 467 |

# Smallest known viral genomes

| Virus | Length | Protein |
|---|---|---|
| Viroid | 120 | none |
| Satellite | 220 | none |
| Hepatitis delta satellite | 1,700 | 1 |
| Circovirus | 1,759 | 2 |
| Anellovirus | 2,170 | 4 |
| Geminivirus | 2,500 | 4 |
| Hepatitis B virus | 3,200 | 7 |
| Levivirus | 3,400 | 4 |
| Partitivirus | 3,700 | 2 |
| Barnavirus | 4,000 | 7 |

Go to:

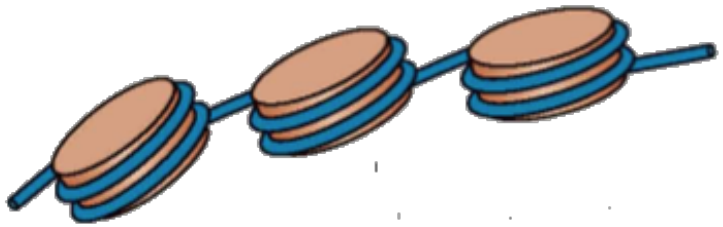
m.socrative.com
room number: virus

**What information may be encoded in a viral genome?**

1. Gene products that catalyze membrane biosynthesis
2. Gene products that catalyze energy production
3. Complete protein synthesis systems
4. Centromeres or telomeres
5. Enzymes to replicate the viral genome

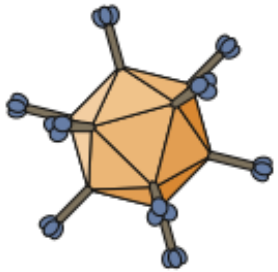
# Viral DNA genomes

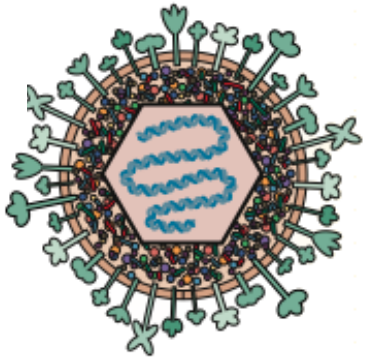
- The host genetic system is based on DNA
- Many DNA viruses emulate the host
- However, almost all viral DNA genomes are NOT like cell chromosomes
- Unexpected tricks have evolved

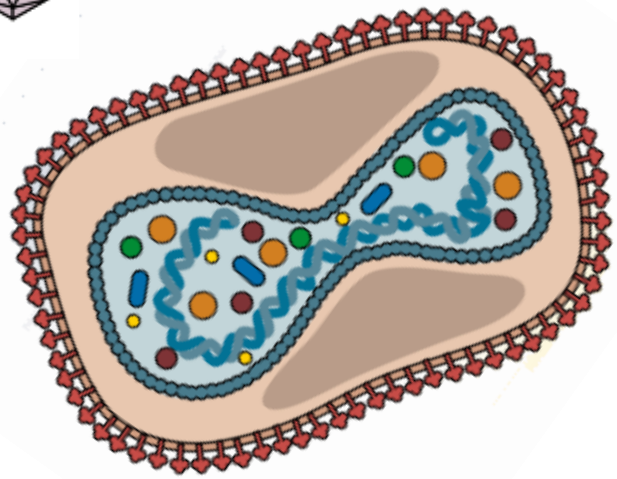


# dsDNA genomes

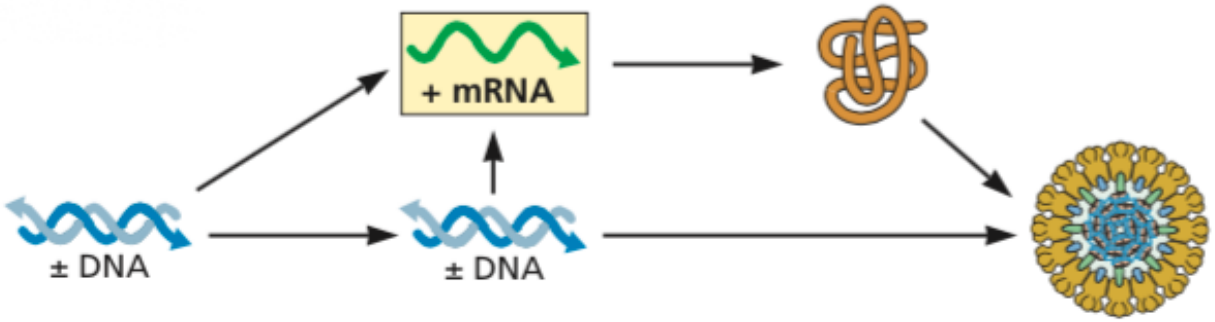
- *Adenoviridae*
- *Herpesviridae*
- *Papillomaviridae*
- *Polyomaviridae*
- *Poxviridae*





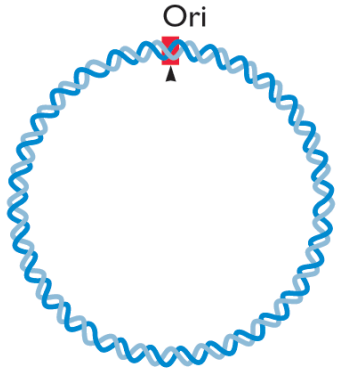


# dsDNA genomes




## Genomes copied by host DNA polymerase

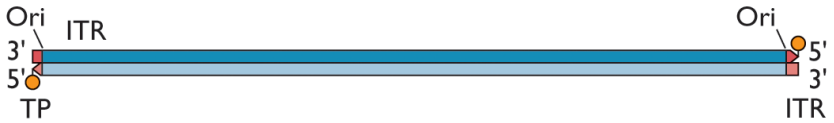
***Polyomaviridae* (5 kbp)**



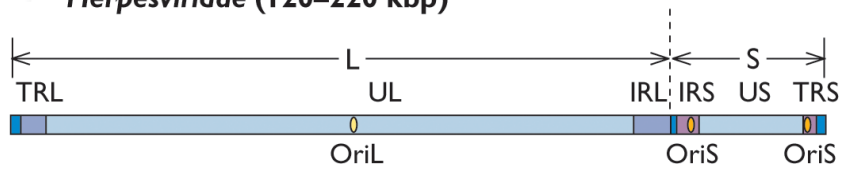

*Papillomaviridae* (8 kbp)

## Genomes encode DNA polymerase

***Adenoviridae* (36–48 kbp)**

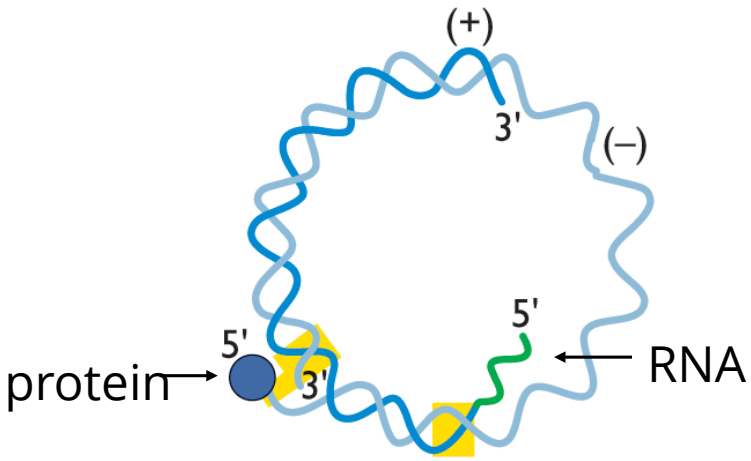



***Herpesviridae* (120–220 kbp)**

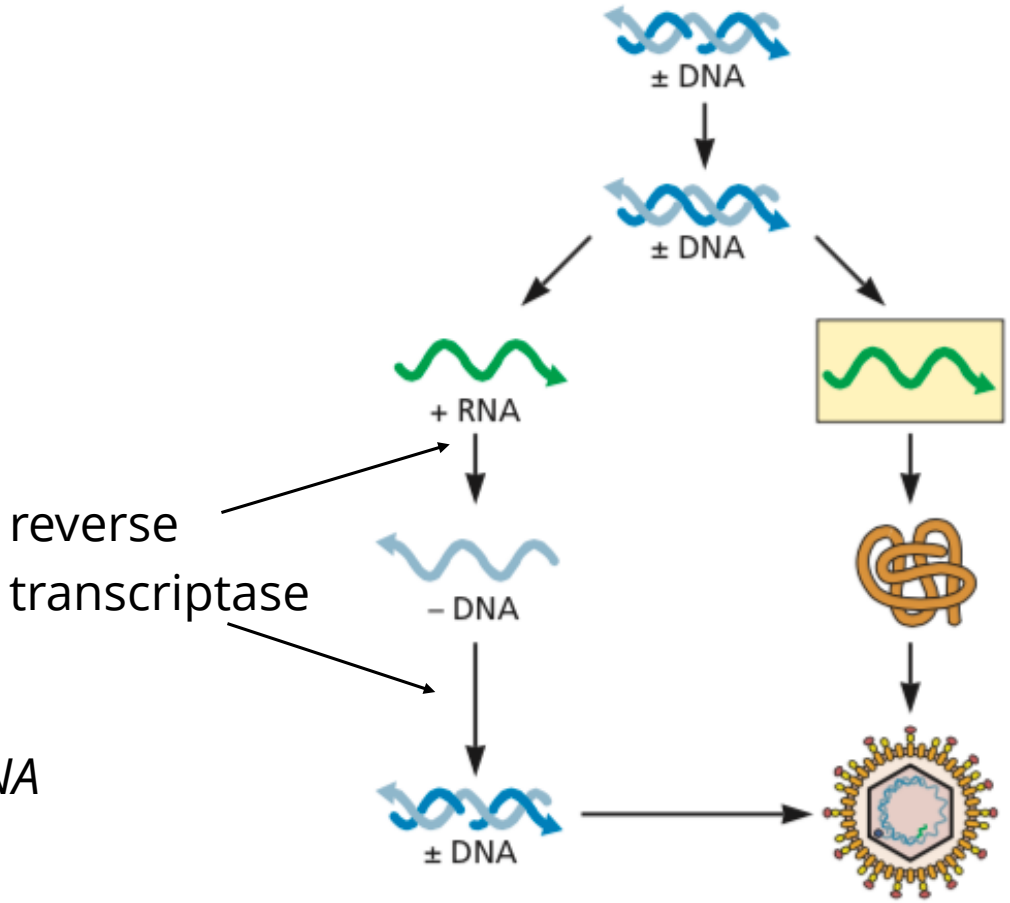



***Poxviridae* (130–375 kbp)**

# Gapped dsDNA genomes



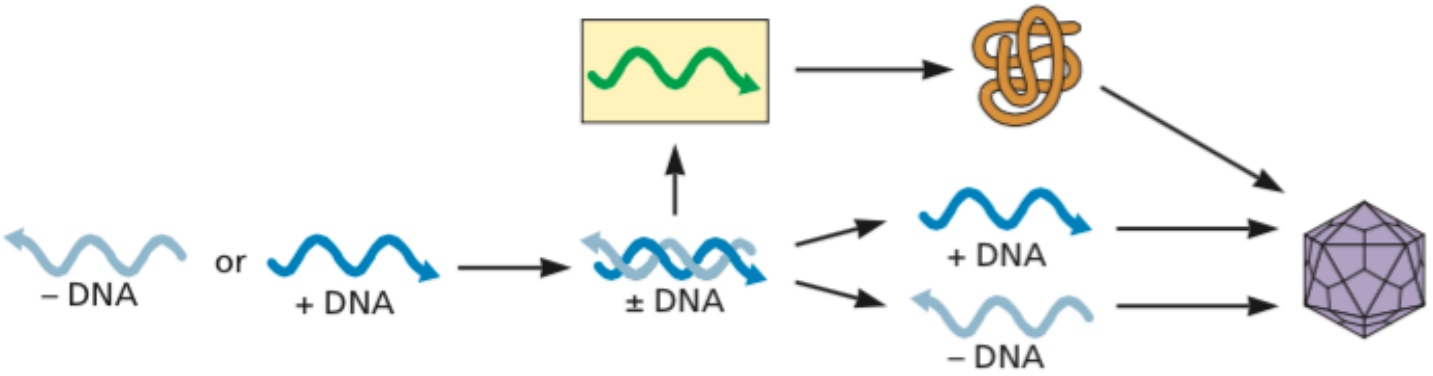

*This genome cannot be copied to mRNA*




*Hepadnaviridae*

Hepatitis B virus

# ssDNA genomes




*Circoviridae* (1.7–2.2 kb)

TT virus (ubiquitous human virus)

*Parvoviridae* (4–6 kb)


B19 parvovirus (fifth disease)

Go to:

m.socrative.com
room number: virus

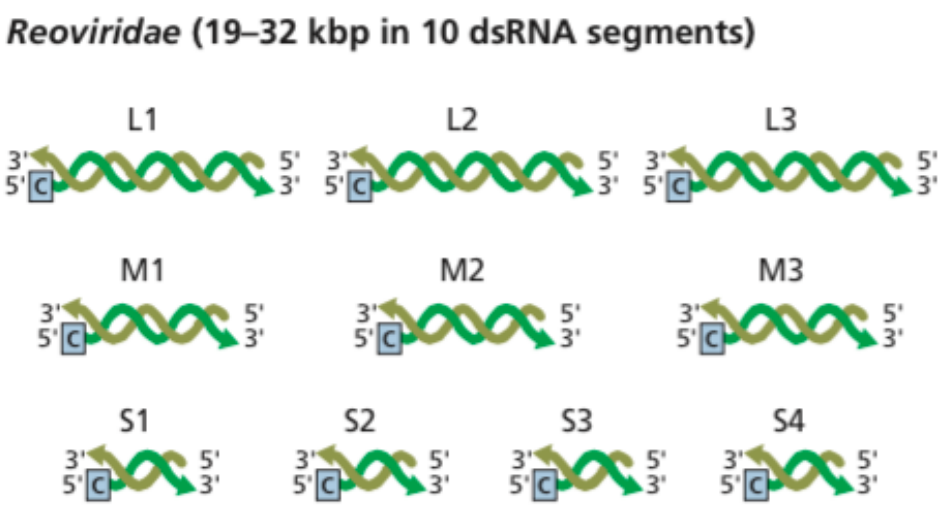
**Which DNA genome, on entry into the cell, can be immediately copied into mRNA?**

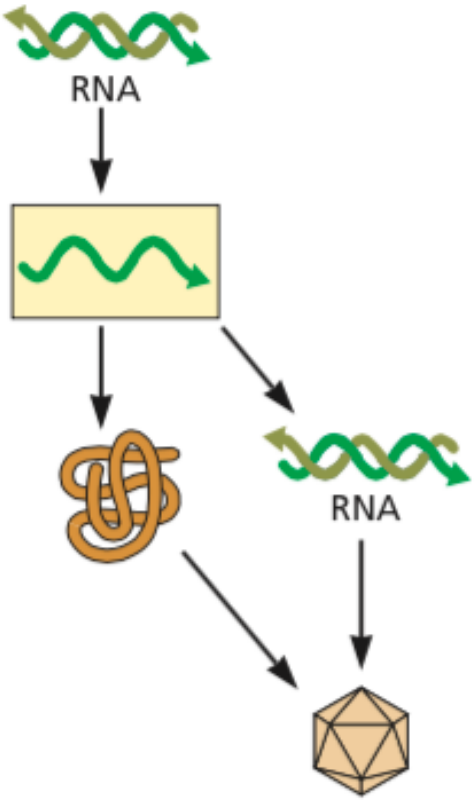
1. dsDNA
2. gapped dsDNA
3. circular ssDNA
4. linear ssDNA
5. All of the above

# RNA genomes

- Cells have no RNA-dependent RNA polymerase (RdRp)
- RNA virus genomes encode RdRp
- RdRp produce RNA genomes and mRNA from RNA templates

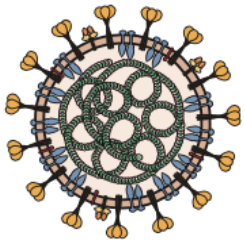
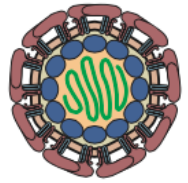
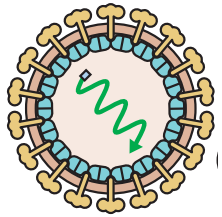
# dsRNA genome



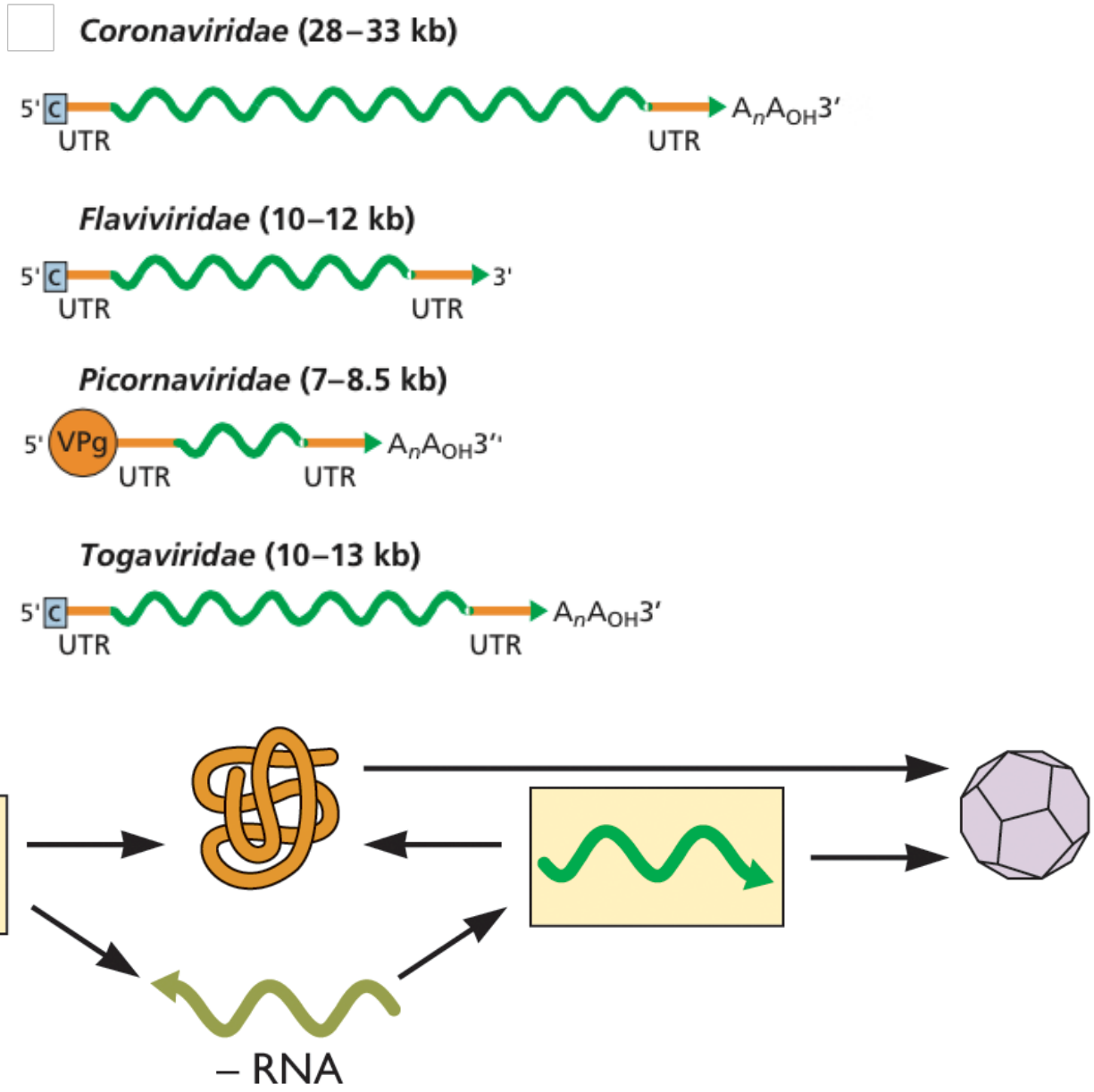




Rotavirus (human gastroenteritis)

# ssRNA: (+) sense

- *Picornaviridae* (*Poliovirus*, *Rhinovirus*) 
- *Caliciviridae* (gastroenteritis) 
- *Coronaviridae* (SARS) 
- *Flaviviridae* (*Yellow fever virus*, *West Nile virus*, *Hepatitis C virus*) 
- *Togaviridae* (*Rubella virus*, *Equine encephalitis virus*) 

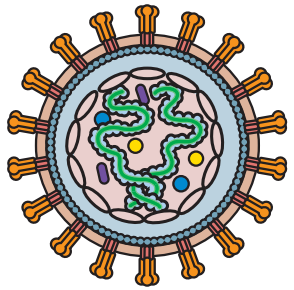
# ssRNA: (+) sense



# ssRNA(+) sense with DNA intermediate

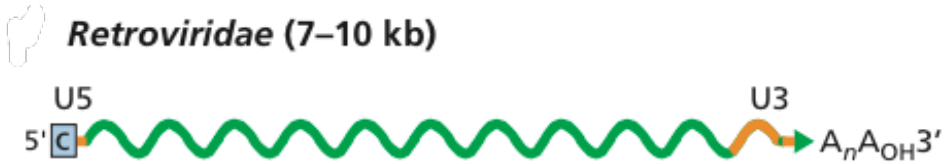
**One viral family: *Retroviridae***

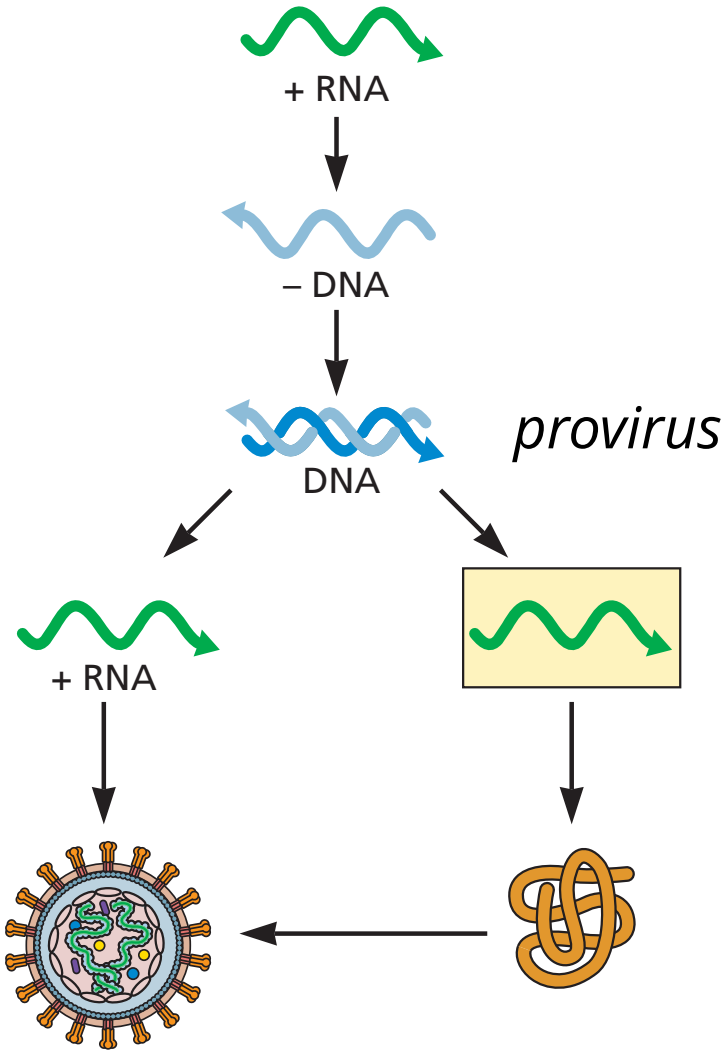
Two human pathogens:



Human immunodeficiency virus (HIV)
Human T-lymphotropic virus (HTLV)

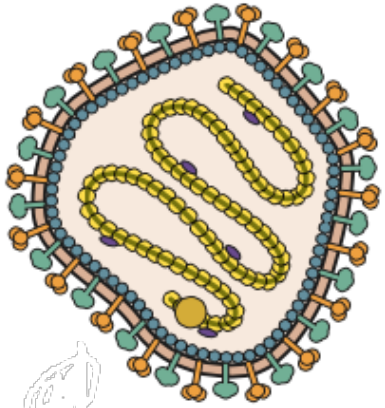
# The remarkable retroviral genome strategy

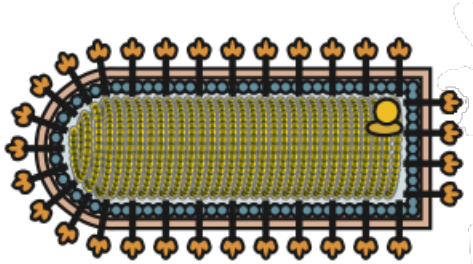


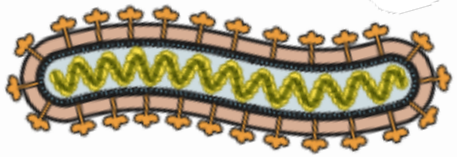



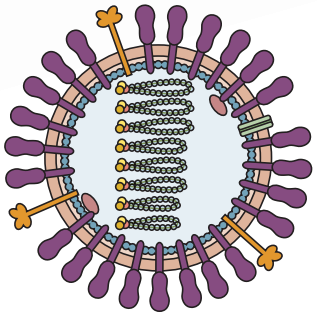
# ssRNA, (-) sense

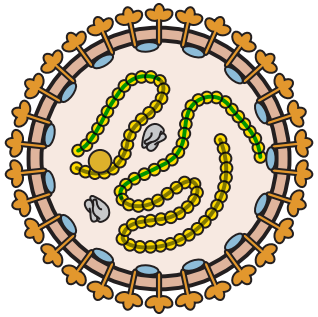
- *Paramyxoviridae* (Measles virus, Mumps virus)
- *Rhabdoviridae* (Rabies virus)
- *Filoviridae* (*Ebola virus, Marburg virus*)
- *Orthomyxoviridae* (Influenza virus)
- *Arenaviridae* (Lassa virus)

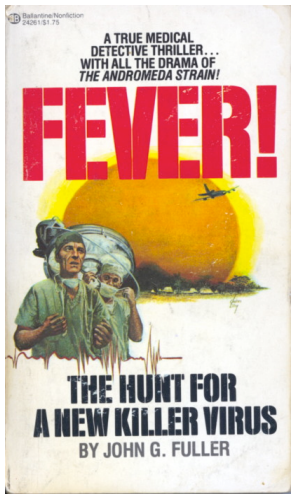




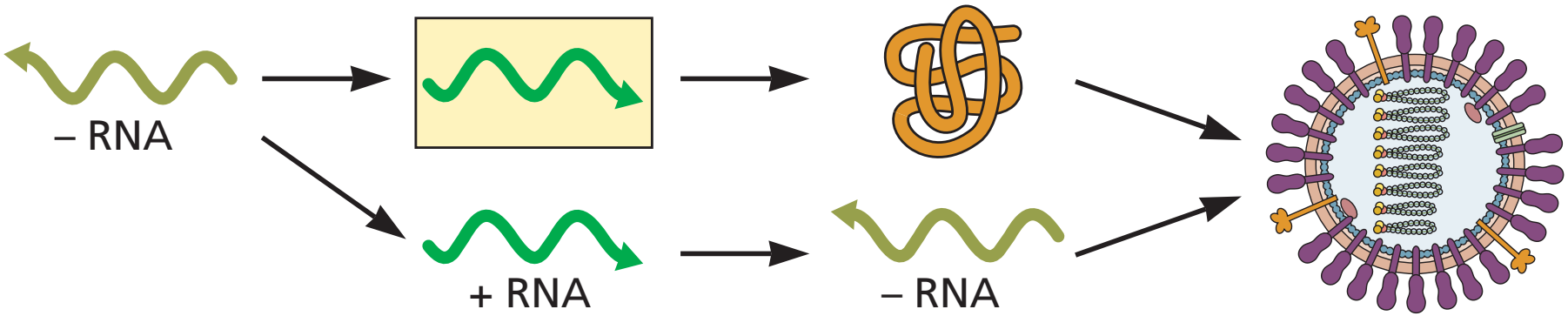






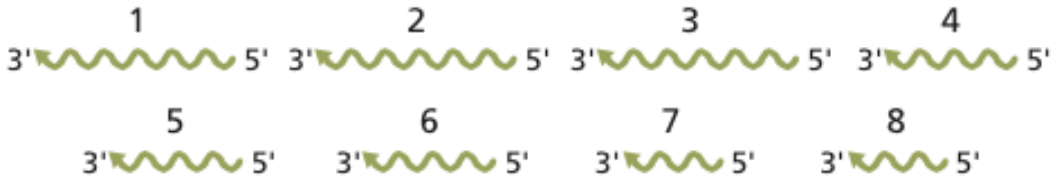


# ssRNA, (-) sense




**Segmented genomes: *Orthomyxoviridae* (10–15 kb in 6–8 RNAs)**

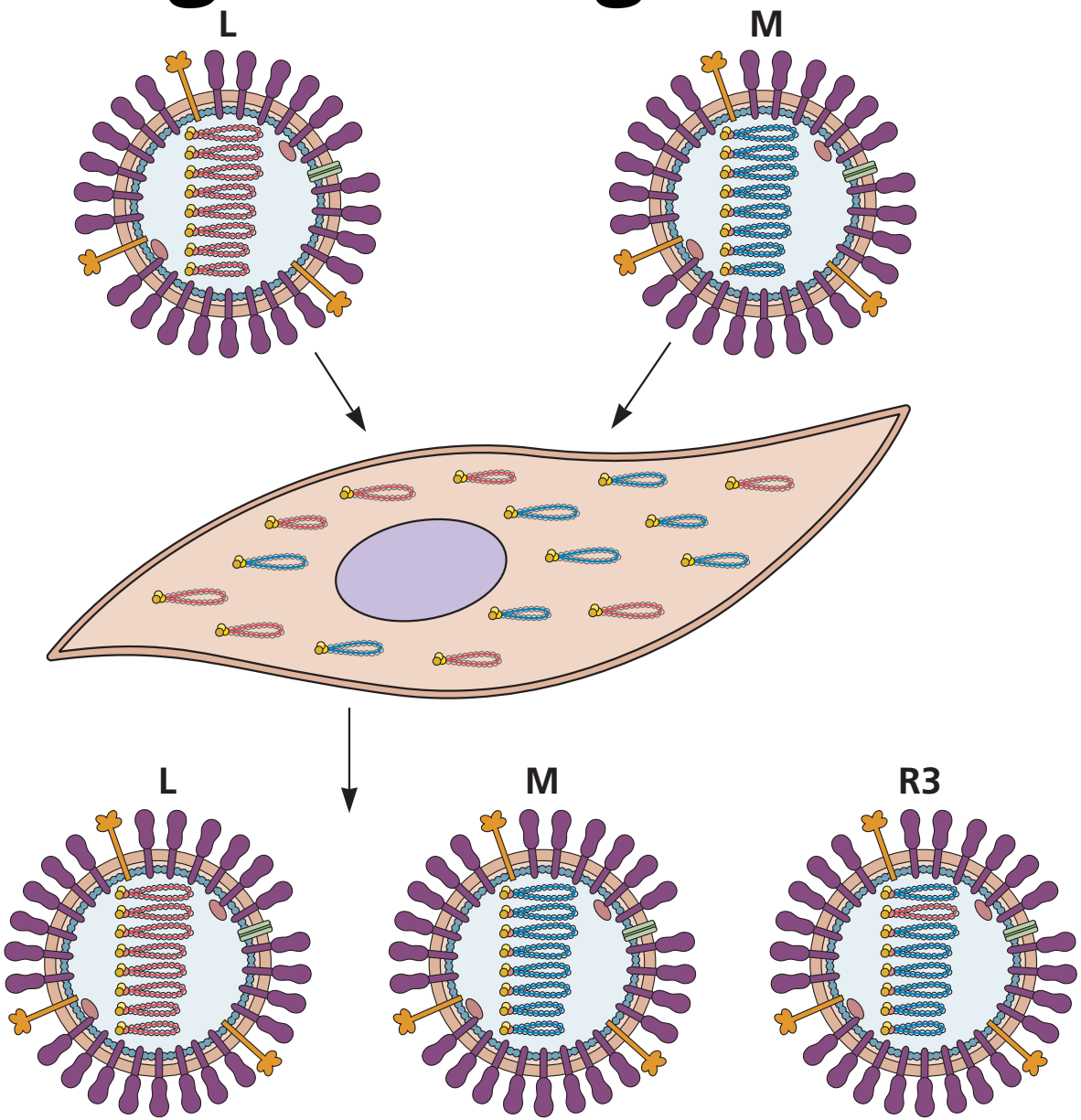
(–) strand RNA segments



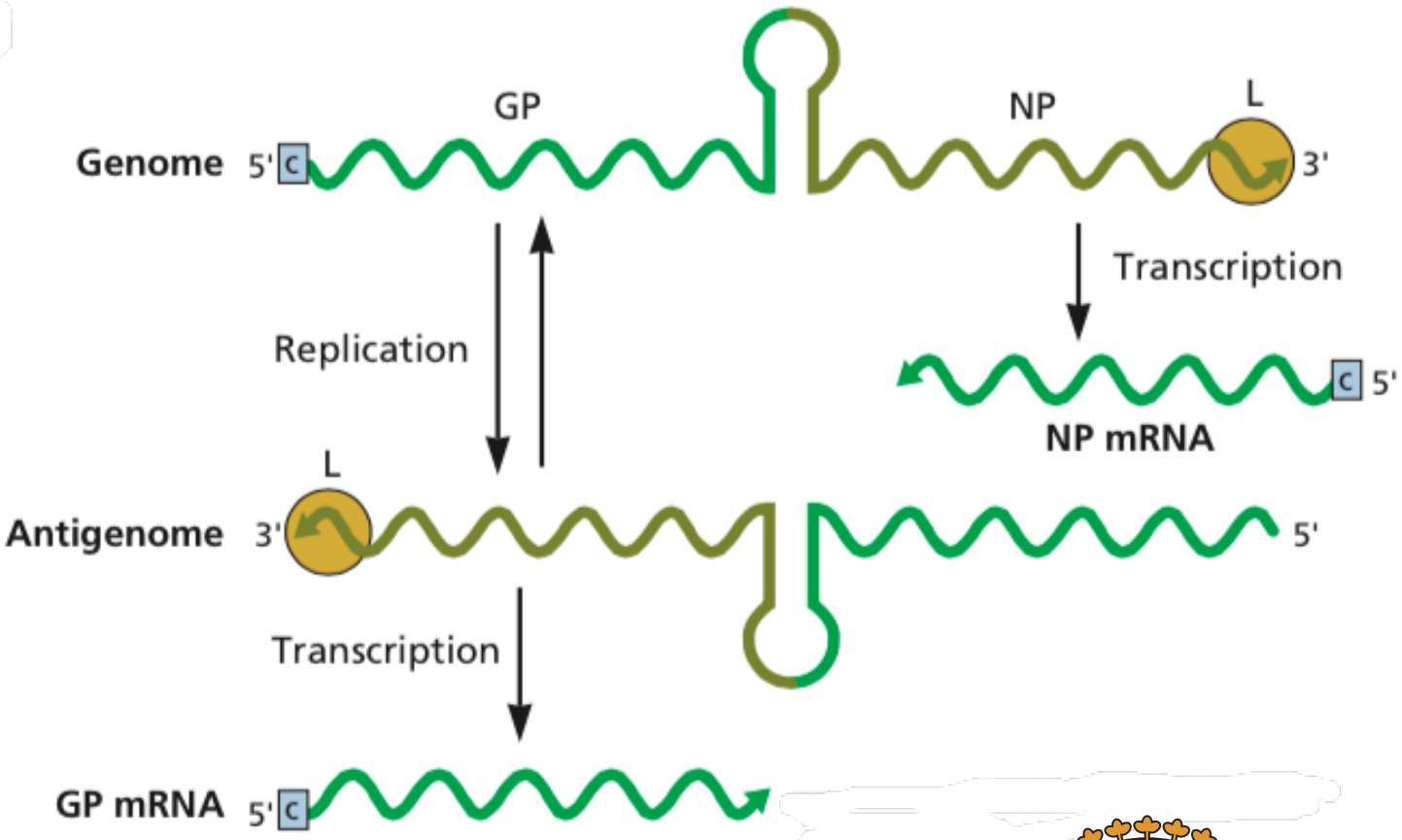

**Nonsegmented genomes: *Paramyxoviridae* (15–16 kb)**

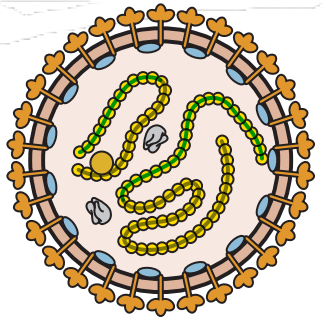

***Rhabdoviridae* (13–16 kb)**

# Reassortment: Consequence of segmented genome



# Ambisense RNA genomes



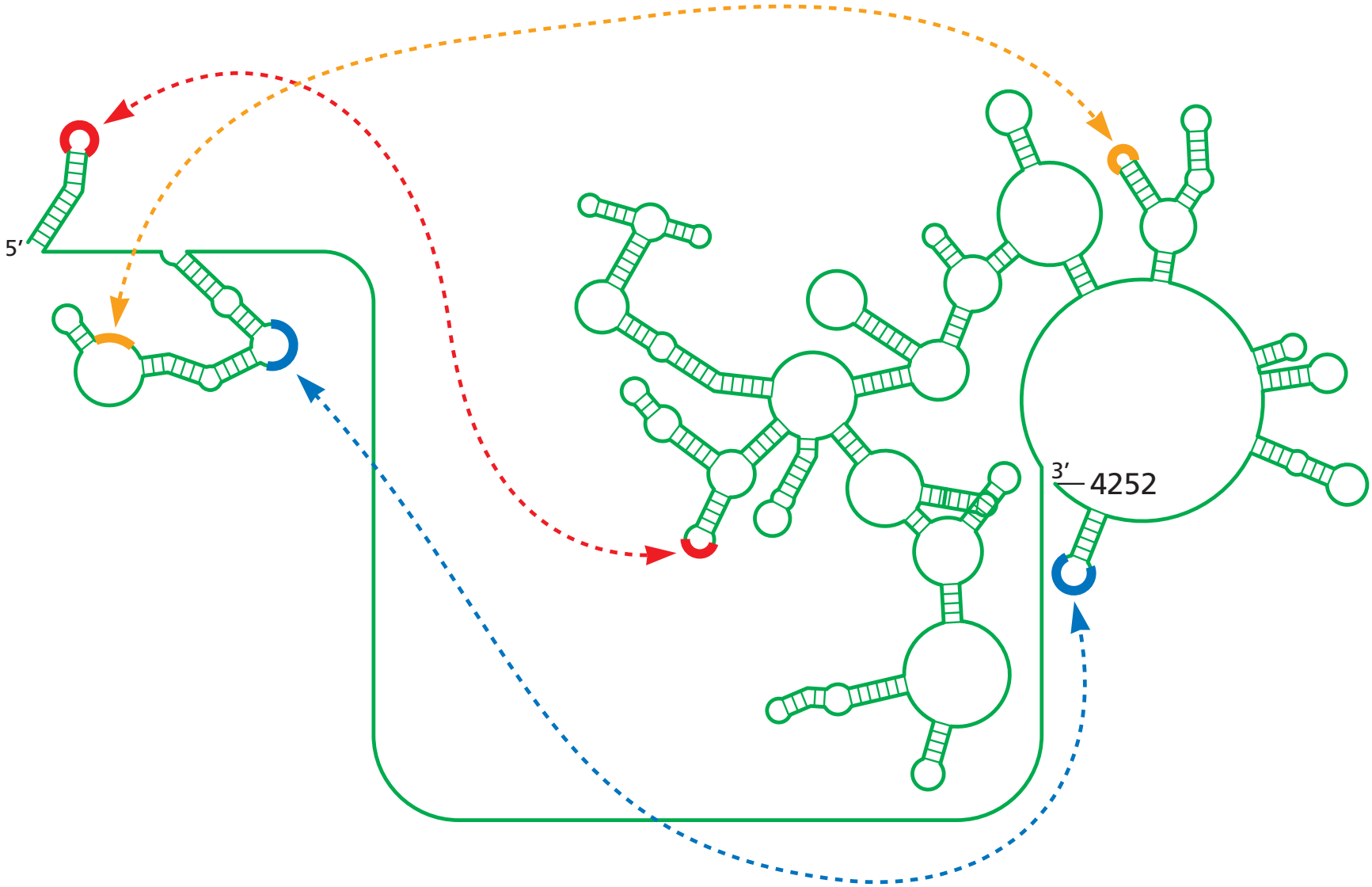



*Arenaviridae*
RNA pol in virion

**A**

Linear (+) strand RNA genome of a picornavirus


**B**



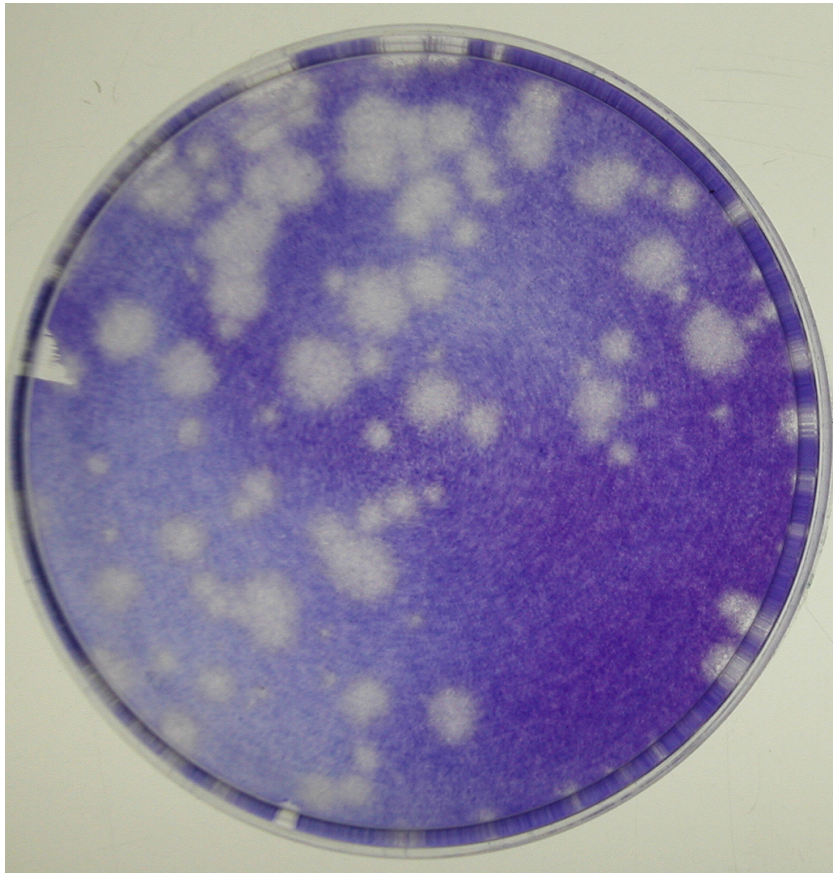
Go to:

m.socrative.com
room number: virus

**Which statement about viral RNA genomes is correct?**

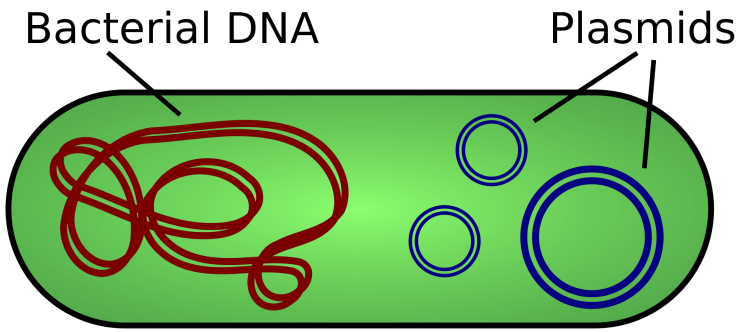
1. (+)ssRNA genomes may be translated to make viral protein
2. dsRNA genomes can be directly translated to make viral protein
3. (+)ssRNA virus replication cycles do not require a (-) strand intermediate
4. RNA genomes can be copied by host cell RNA-dependent RNA polymerases
5. All of the above

# This method allowed the application of genetic methods to animal viruses

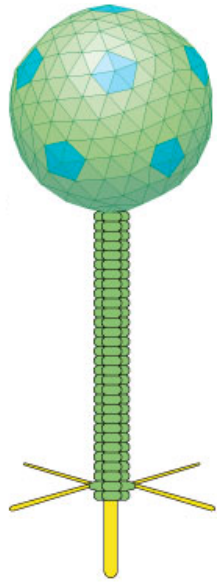


# Engineering mutations into viral genomes - the modern way

- Infectious DNA clone: transfection
- A modern validation of the Hershey-Chase experiment (1952)
- Deletion, insertion, substitution, nonsense, missense

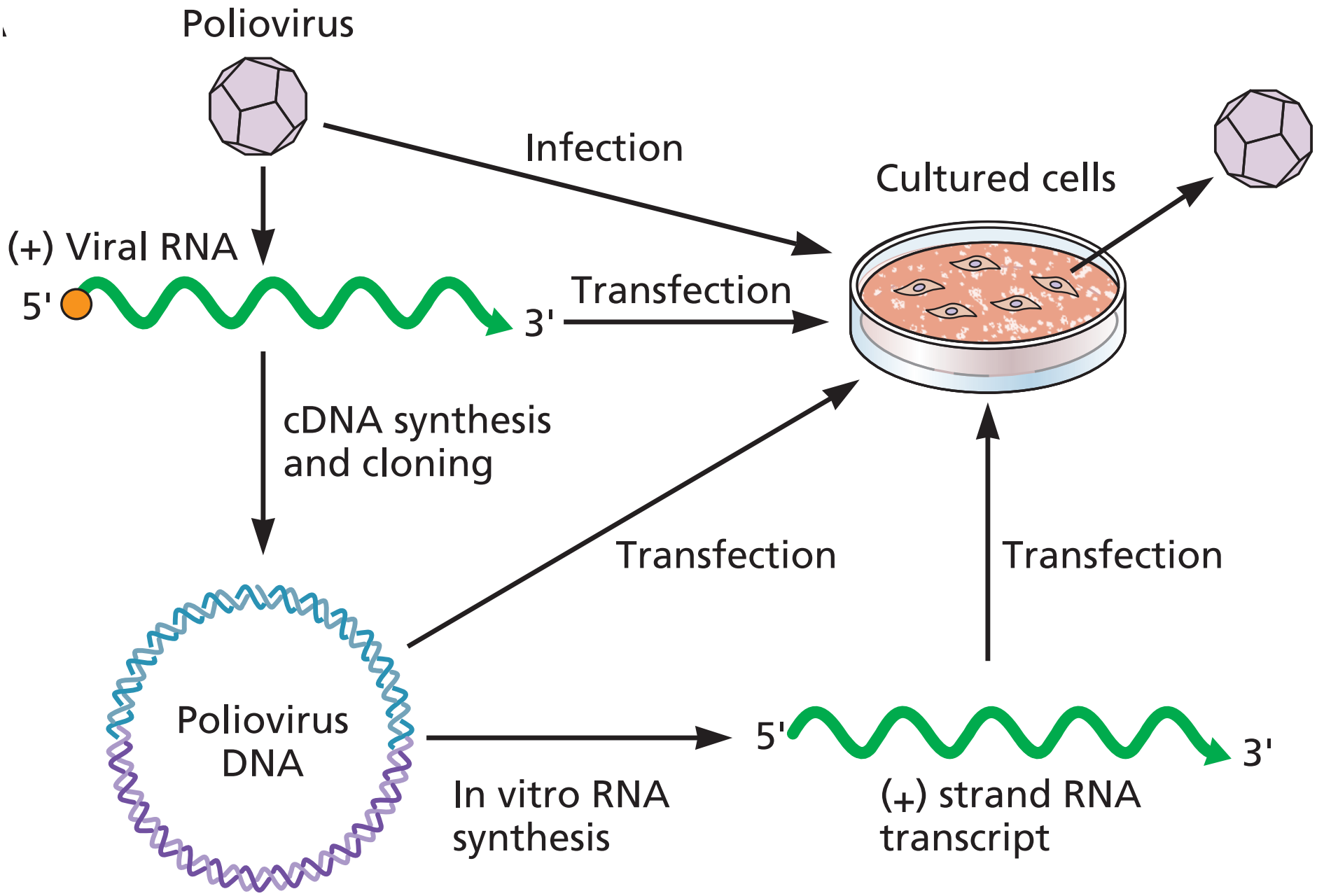


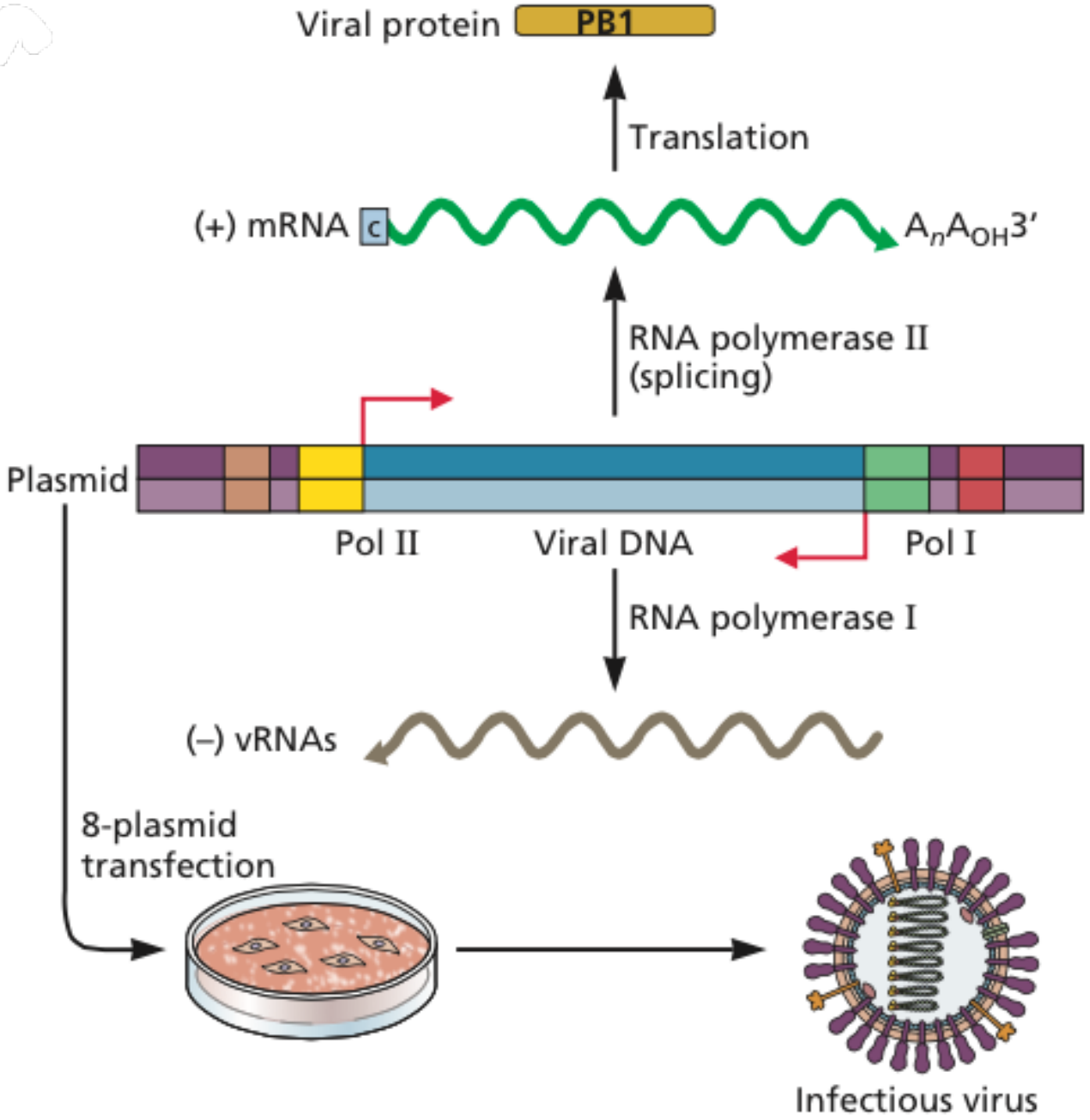
# Genetic methods



***Transfection***

- Production of infectious virus after transformation of cells by viral DNA, first done with bacteriophage lambda
- Trans<u>formation-in</u>fection





# Resurrecting the 1918 influenza virus

- Influenza virus was not identified until 1933
- In 2005, influenza RNA was isolated from formalin-fixed, paraffin-embedded lung tissue sample from autopsy of victim of influenza in 1918
- Influenza RNA also isolated from frozen sample obtained by in situ biopsy of the lung of a victim buried in permafrost since 1918
- Complete nucleotide sequence of all 8 RNA segments determined
- Virus was recovered by transfection of cells with 8 plasmids containing genome sequences

http://www.sciencemag.org/content/310/5745/77.long

# Synthetic Virology and Biosecurity

NSABB: National Science Advisory Board for Biosecurity

- Federal advisory committee to provide advice, guidance, and leadership regarding biosecurity oversight of *dual use research* to all Federal departments and agencies with an interest in life sciences research
- Advises on and recommends specific strategies for the efficient and effective oversight of federally conducted or supported dual use biological research, taking into consideration national security concerns and the needs of the research community.

http://osp.od.nih.gov/office-biotechnology-activities/biosecurity/nsabb