

Bioinformatics

CS300

Chap 3

Sequence Alignment:

Investigating an Influenza Outbreak

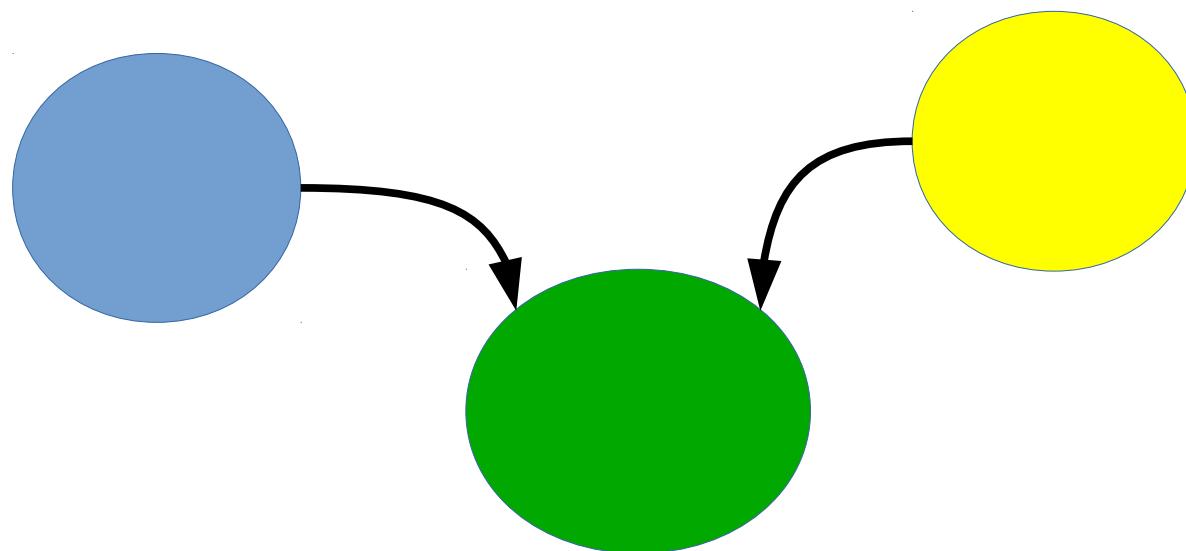
Fall 2017

Oliver Bonham-Carter



Descent with Modification

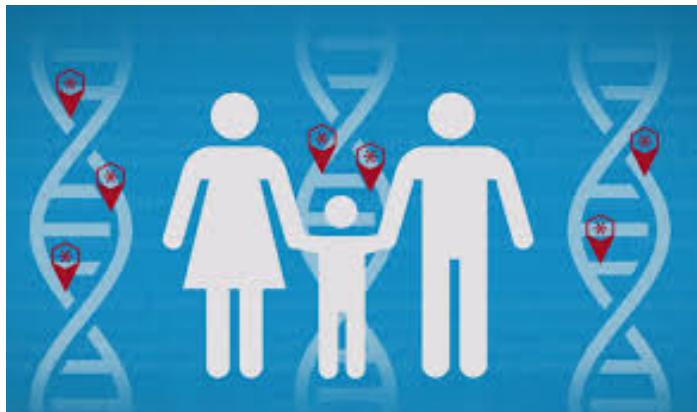
- ***Descent with modification*** is simply a passing traits from parent to offspring.
- One of the fundamental ideas behind Charles Darwin's theory of evolution.
- Traits are passed on to children in a process known as heredity.





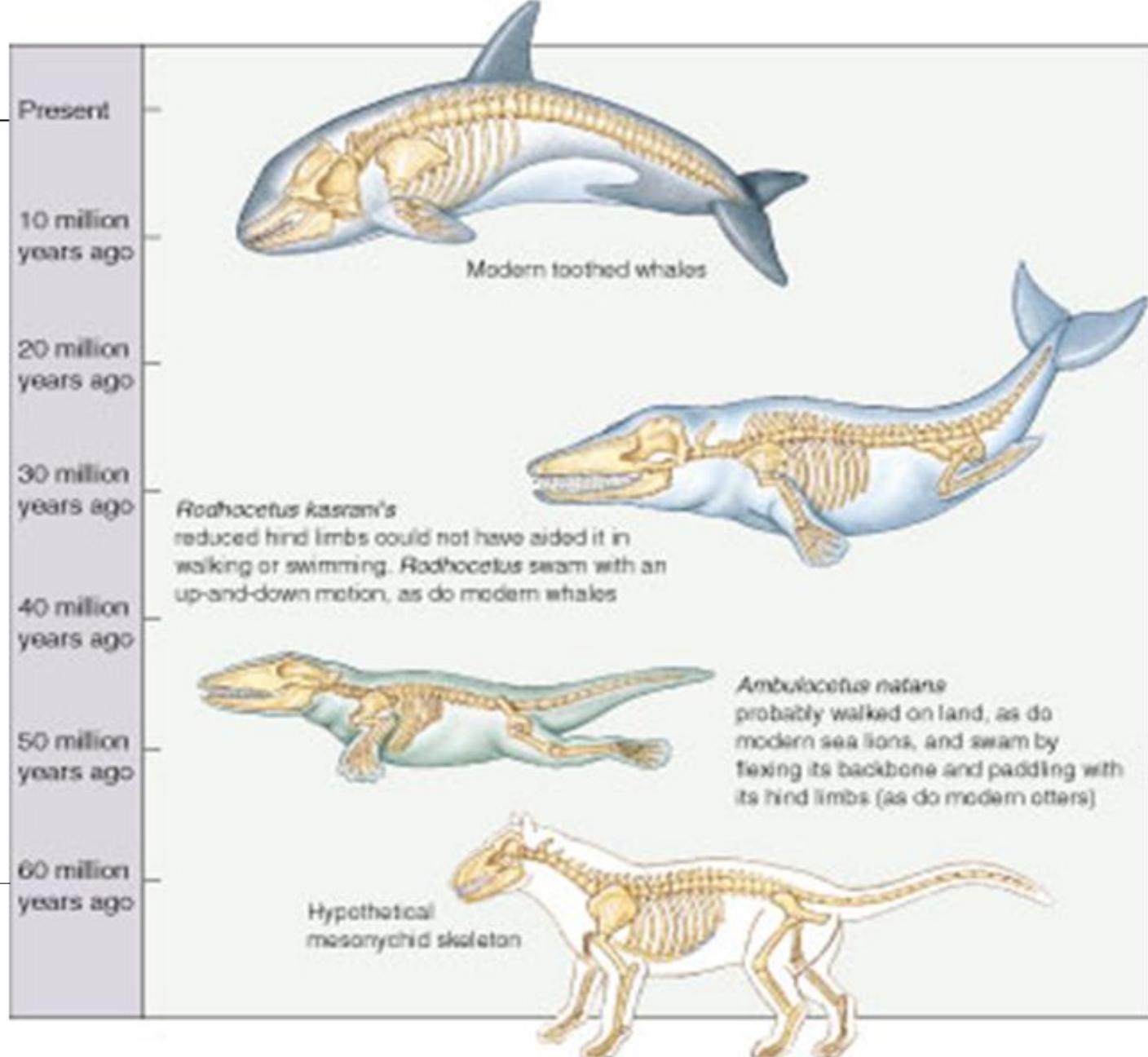
Passed Down From Genome to Genome

- DNA replication ensures a **mostly** faithful passing of the genome to progeny
- What would be the consequence of 100% accurate replication?
- Is that high similarity really desirable for a species?
- How does decent with modification happen?



Descent With Modification

Present Day

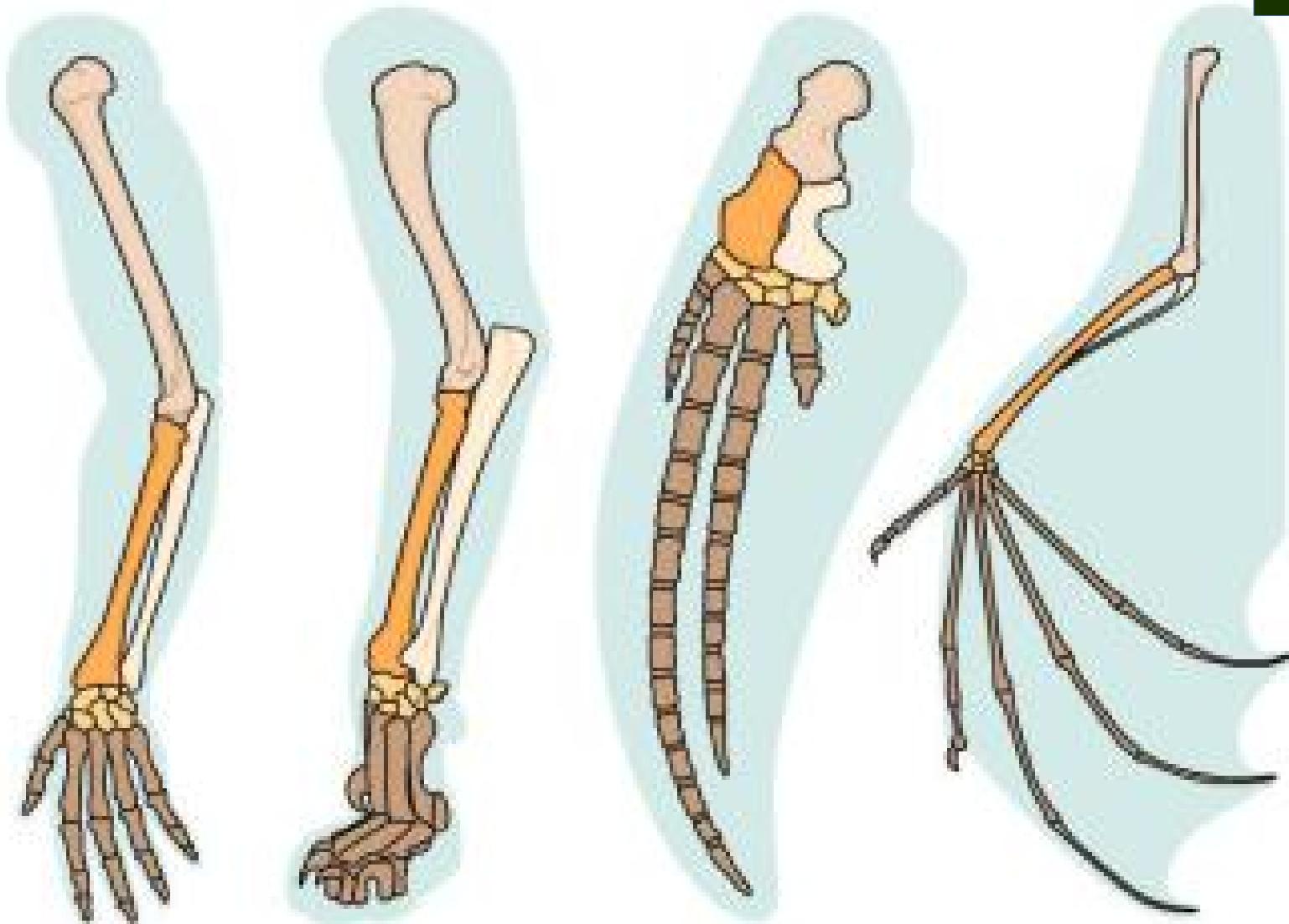


Long Ago



ALLEGHENY
COLLEGE

Same Bone, Different Day



Human

Cat

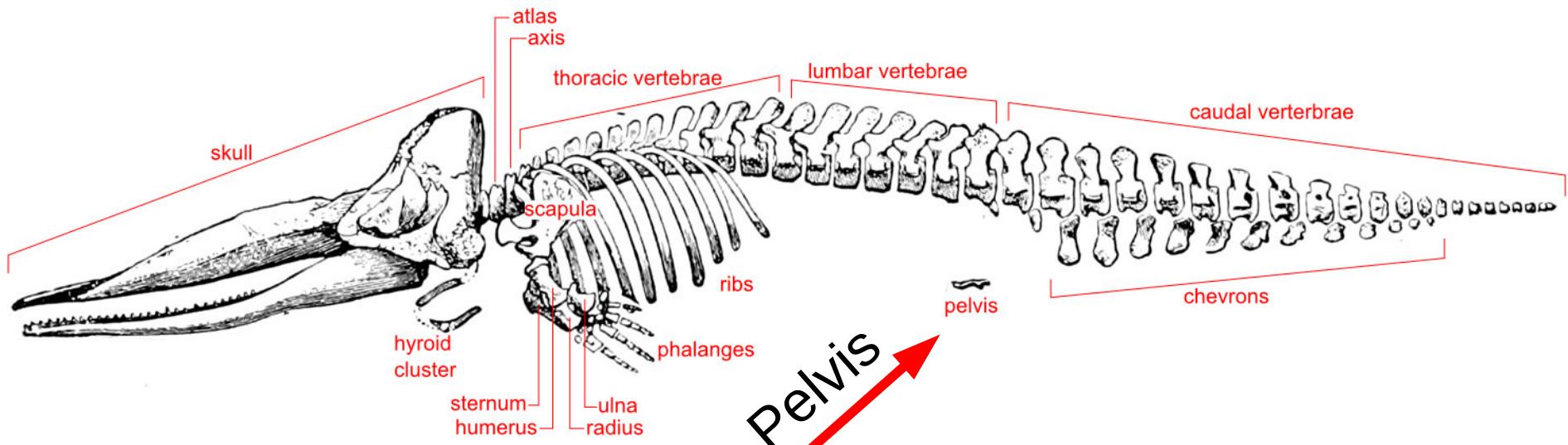
Whale

Bat



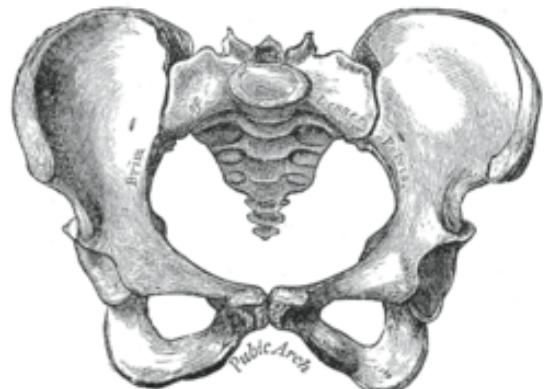
ALLEGHENY
COLLEGE

Same Bone, Different Day



Whale's Pelvis

Human's
Pelvis





ALLEGHENY
COLLEGE

How Does Descent With Modification Happen?

Mutation

- A change in a DNA sequence
- Results from errors in replication or repair
- Mutation is the ultimate source of genetic variation

Domestic Dog – *Canis lupus familiaris*
All descendant from the grey wolf
All the same species (sub-species)
Breeds – variation within the species





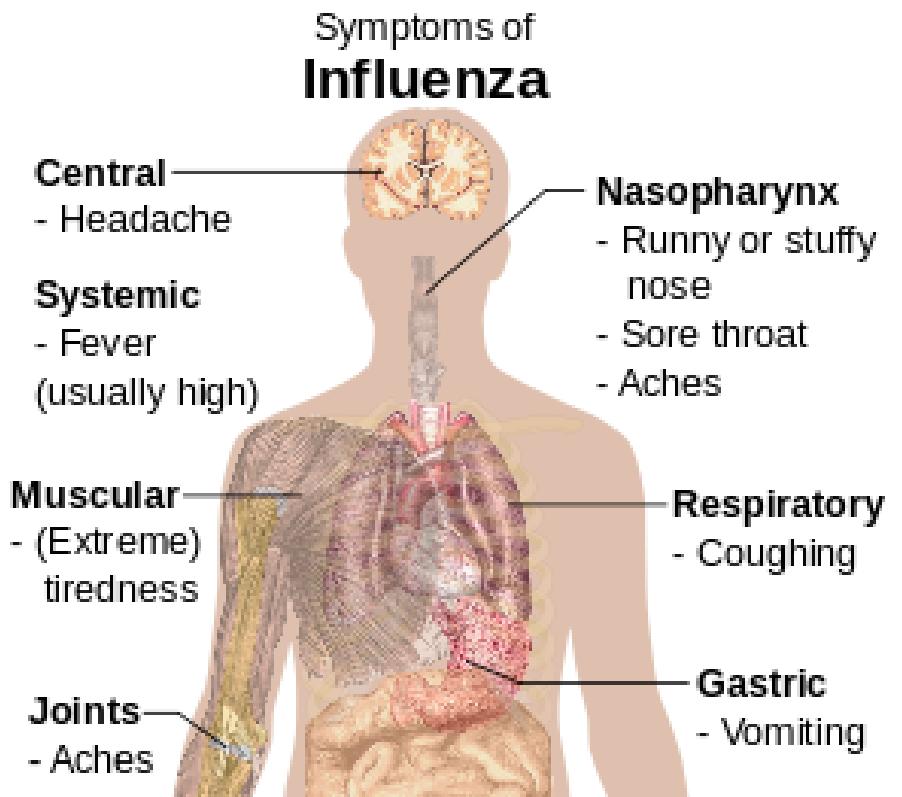
Sequence Variations

- Sequences may have diverged from a common ancestor through various types of mutations:
- substitutions (ACGA ➔ AGGA)
- insertions (ACGA ➔ ACCGGAGA)
- deletions (ACGGAGA ➔ AGA)
- You are UNIQUE and SLIGHTLY GENETICALLY DIFFERENT from each of your parents, grand parents, great grand parents ...
- Retro Viruses (Influenza, HIV, etc) are also unique and slightly genetically different from their ancestors



What Is Influenza?

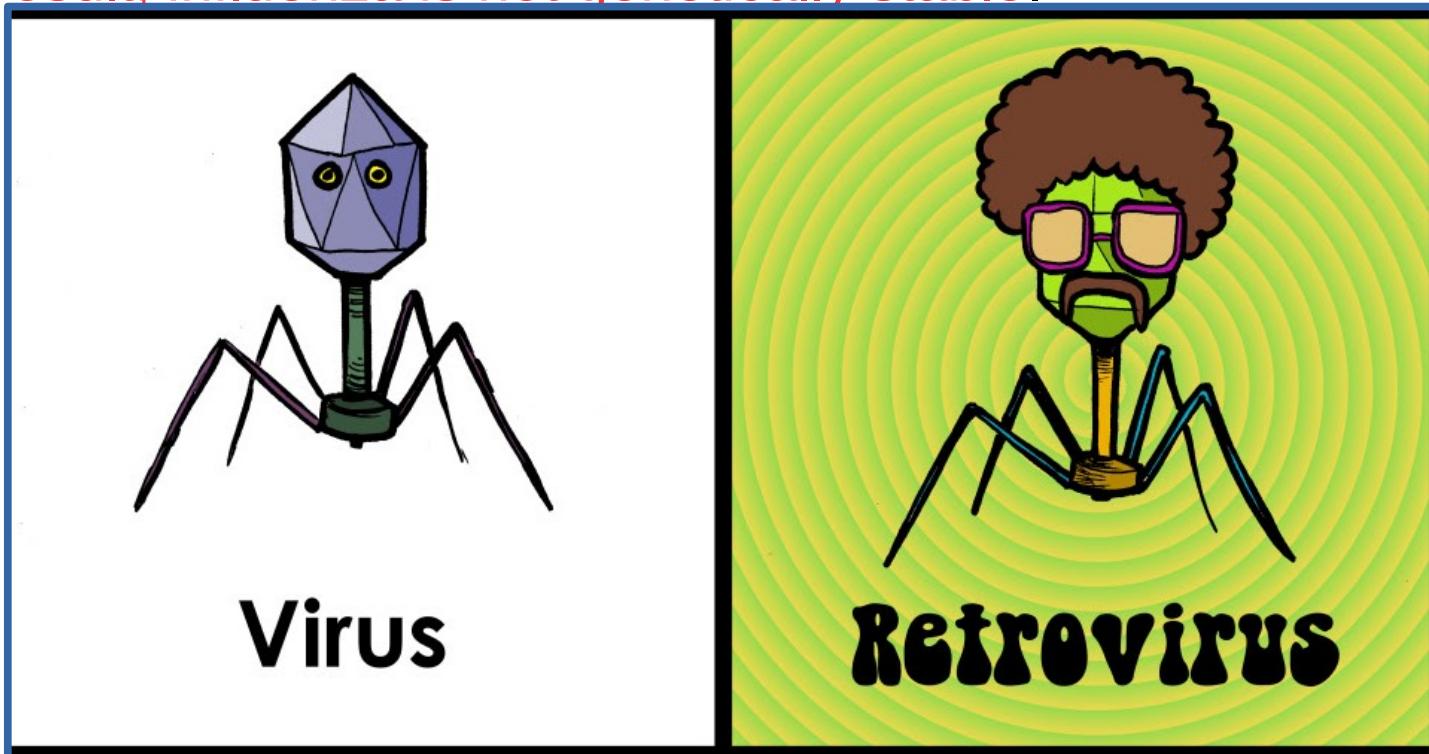
- Flu, also known as the influenza, is a contagious disease that is caused by the flu **virus**. It attacks the respiratory tract in humans (nose, throat, and lungs). The flu is different from a cold. Flu usually comes on suddenly and may include these symptoms:
- Fever
- Headache
- Tiredness (can be extreme)
- Dry cough
- Sore throat
- Nasal congestion
- Body aches





The Influenza Virus: an RNA Virus (or retro virus)

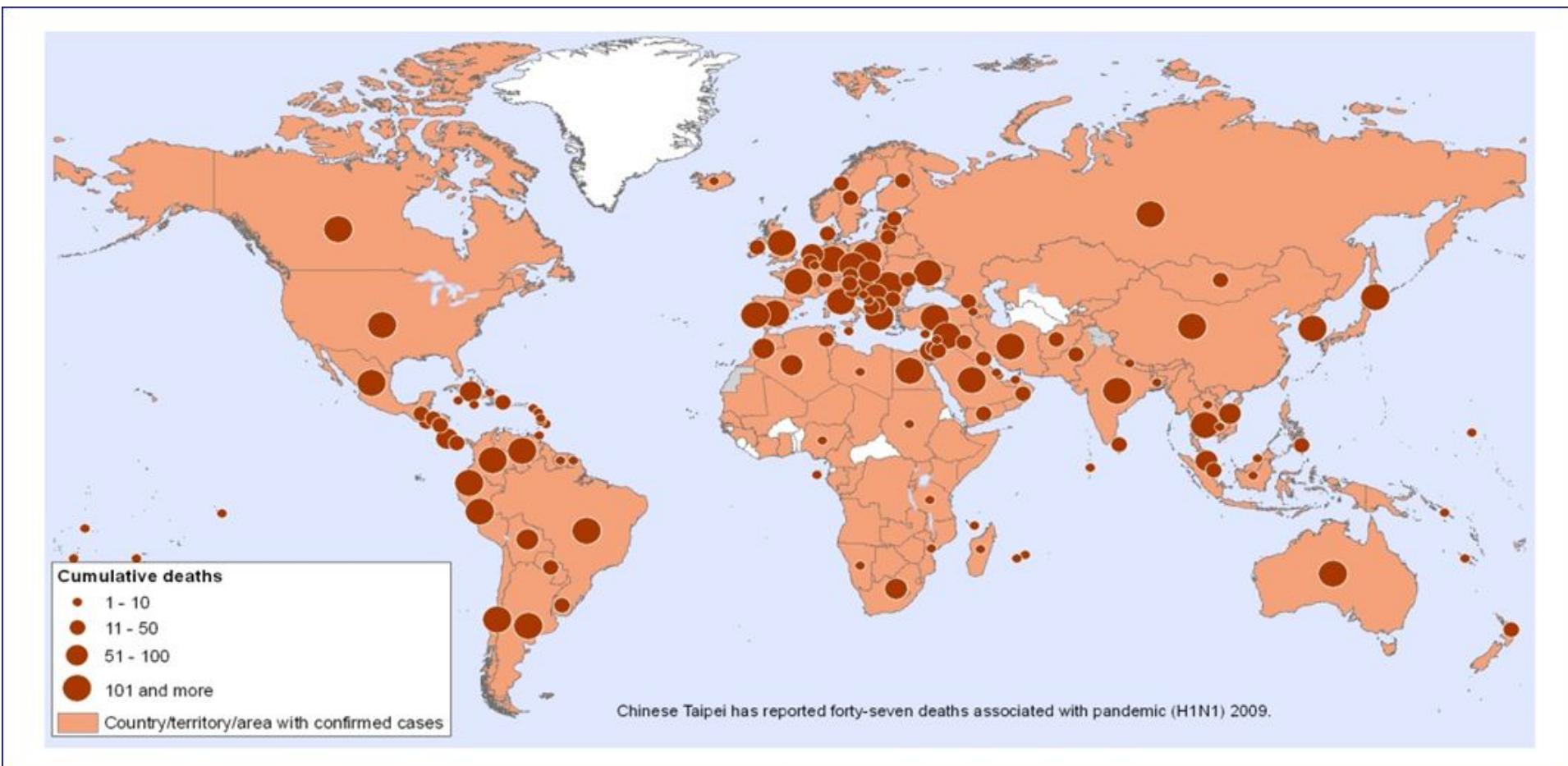
- Like all living things, influenza makes small errors—mutations—when it copies its genetic code during reproduction.
- Influenza lacks the ability to repair those errors, because it is an RNA virus; RNA, unlike DNA, lacks a self-correcting mechanism.
- As a result, influenza is not genetically stable.

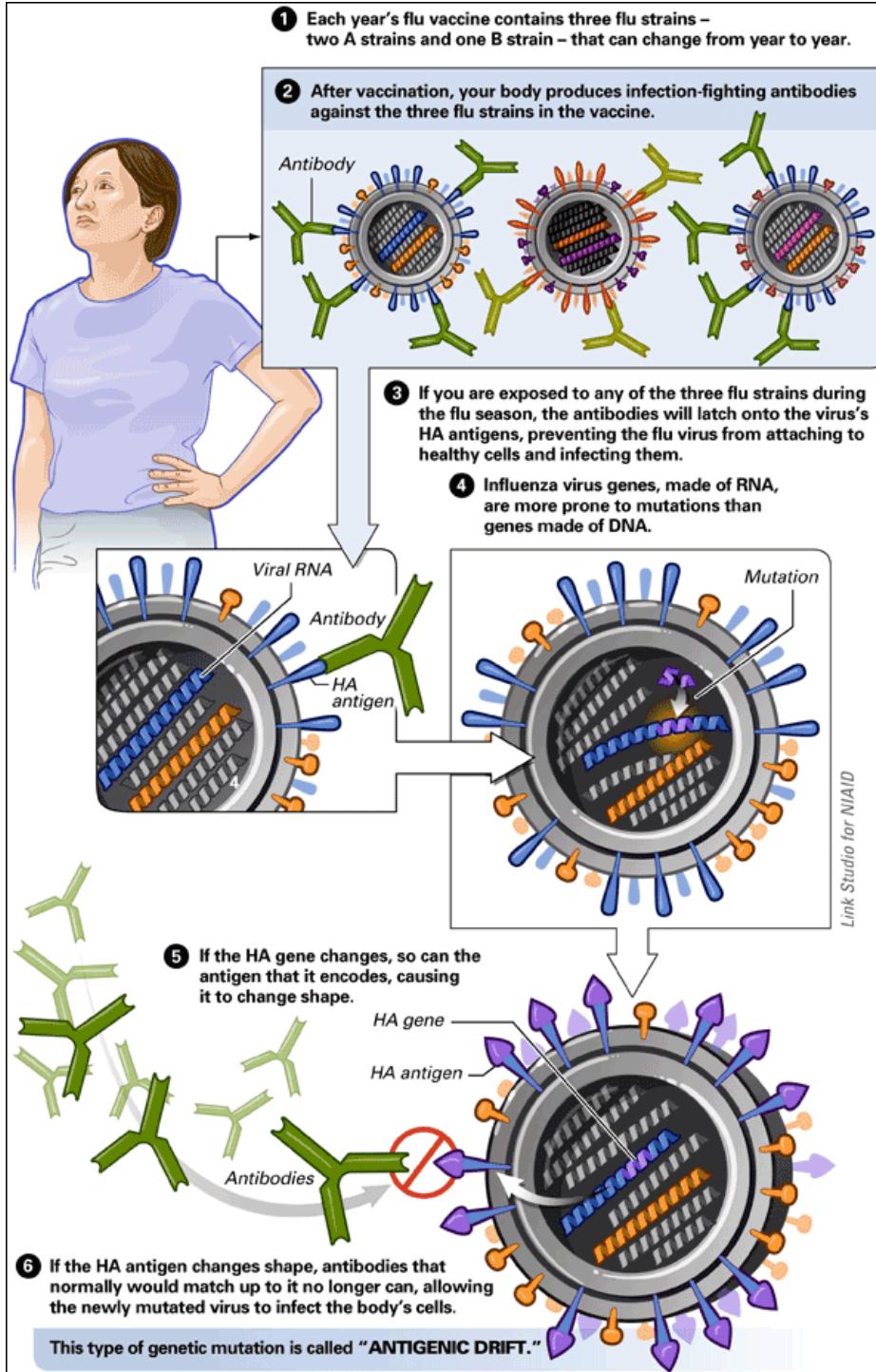




ALLEGHENY
COLLEGE

Tracking Infectious Disease 2009 H1N1 Influenza Pandemic

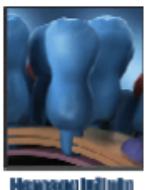
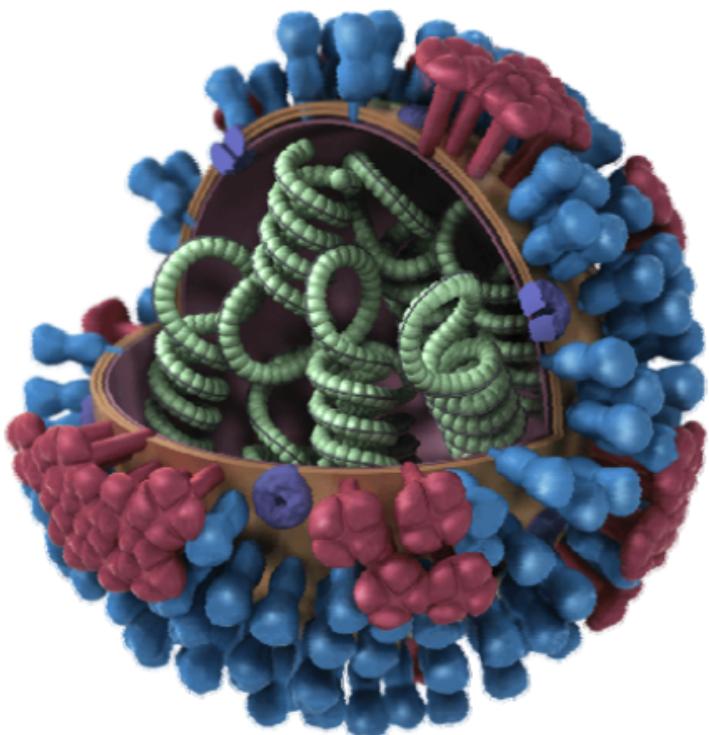




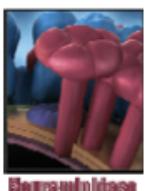


H1N1 Virus

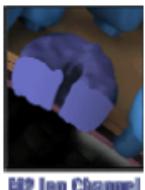
H1N1 Influenza Virus



Hemagglutinin- protein the virus uses to attach to the host cells



Neuraminidase- enables the virus to be released from the host cell



M2 Ion Channel- allows protons to move through the viral envelope and is essential for the virus replication process



RNP- Ribonucleoprotein containing the virus RNA genome



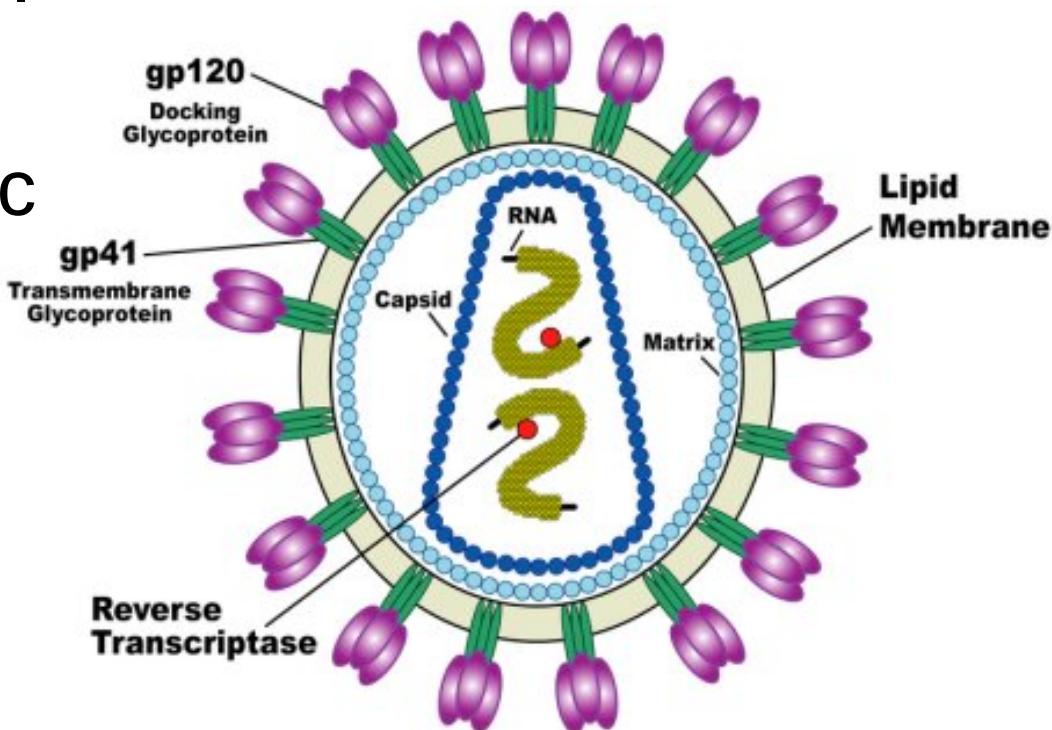
Viral Evolution

- Viruses evolve very quickly
 - Some of the highest mutation rates known
 - Arms race with immune system
 - Viruses – mutation rate 0.0001 - 0.000001 mutations per base per generation
 - One mutation every 10,000 – 1,000,000 nucleotides
 - Influenza genome size = ~14,000 nucleotides
 - Entire genome is coding regions (genes)
- Humans – 0.00000001 mutations per base per generation
 - One mutation every 100,000,000 nucleotides
 - Human genome size 3 billion nucleotides
 - Only 1.5% of genome is coding regions (genes)



Viruses Using DNA vs RNA

- DNA viruses are mostly double-stranded while RNA viruses are single-stranded.
- RNA mutation rate is higher than DNA mutation rate.
- Mutation causes major changes in virus genetic code





Viruses Using DNA vs RNA

- DNA replication takes place in the nucleus while RNA replication takes place in the cytoplasm.
- DNA viruses are stable while RNA viruses are unstable.
- In DNA viruses, viral genetic code is injected in the host DNA for duplication and decoding.
- RNA viruses skip DNA for duplication and decoding.



Reverse Transcription

- Retroviruses do not kill the host cell initially because they can insert their genome into the host genome.
- This process is called reverse transcription and is done by the viral protein reverse transcriptase. In the case of HIV, viral protein integrase then inserts the HIV DNA into host DNA.
- High mutation rate of genetic material from one organism to another.
- These mutations can be used to track virus spread.



Molecular Epidemiology of HIV Transmissions in a Dental Practice

How related is the DNA of each virus infection?

HIV-positive persons included in study of HIV isolates from a Florida dental practice

Person	Sex	Known risk factors	<i>Average differences in DNA sequences (%)</i>	
			From HIV from dentist	From HIV from controls
Dentist	M	Yes		11.0
Patient A	F	No	3.4	10.9
Patient B	F	No	4.4	11.2
Patient C	M	No	3.4	11.1
Patient E	F	No	3.4	10.8
Patient G	M	No	4.9	11.8
Patient D	M	Yes	13.6	13.1
Patient F	M	Yes	10.7	11.9

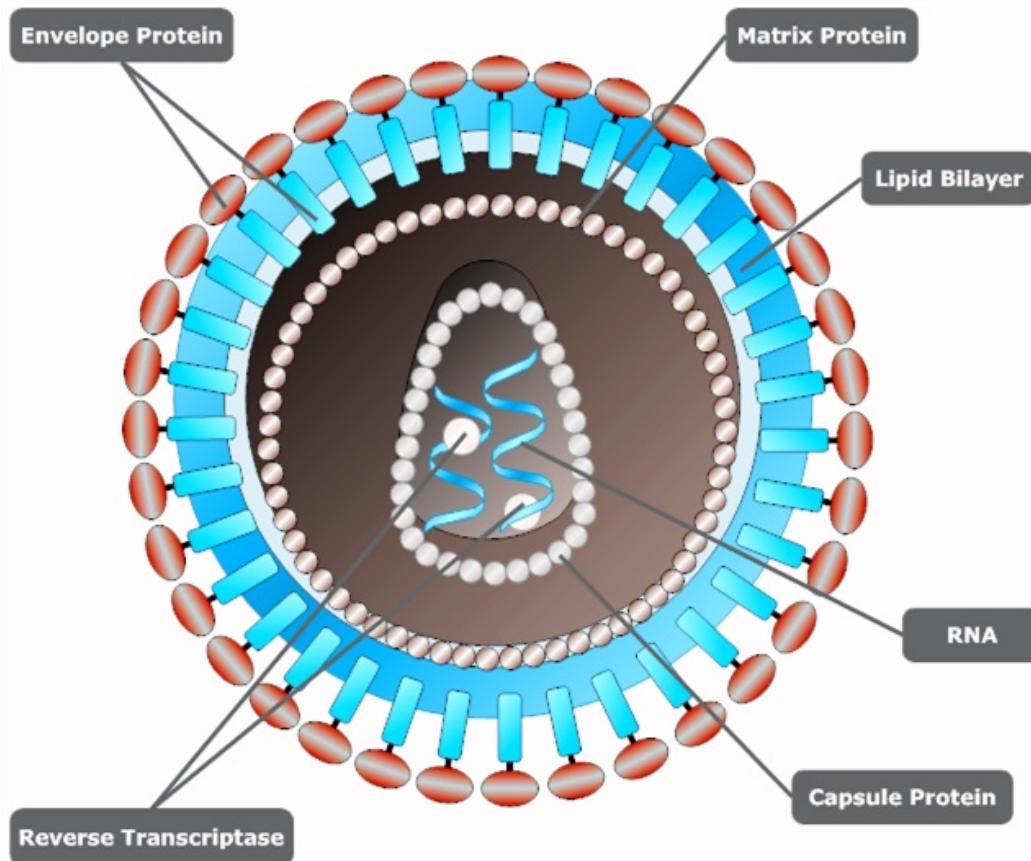
Source: After C. Ou et al., *Science* 256(1992):1165-1171, Table I.

Similar DNA implies closer relationship between organisms



Supporting Videos

- Retrovirus reverse transcripton
- <https://www.youtube.com/watch?v=eS1GODinO8w>





Pairwise Alignment Similarity and Relatedness

Alignment of a gene from two closely related viruses

Hemagglutinin gene from virus A: ATGAACGCAATACTCGTAGTT...

||||| ||||||||| |||||

Hemagglutinin gene from virus B: ATGAAGGCAATACTAGTAGTT...

Few Mismatches



Alignment of a gene from two distantly related viruses

Hemagglutinin gene from virus A: ATGAACGCAATACTCGTAGTT...

||| ||| ||| ||| | |

Hemagglutinin gene from virus C: ATGCACGAAATGCTCGGACCT...

Lots of Mismatches





ALLEGHENY
COLLEGE

Concept Questions: Discuss With Your Group

- How is similarity between genes related to the biological concept of descent from a common ancestor?
- The influenza virus mutates so rapidly that you would likely be able to identify at least a couple of mutations over the length of the complete virus genome even if you sequenced two viruses from two different patients within the same influenza outbreak. What do you think might be some considerations in deciding whether two viruses with different genome sequences actually represent two different strains?
- RNA viruses (retro viruses) are prone to genetic mutations during replication. After a mutation, parts of their genetics have been completely changed. Is it still possible to study their genetics (over long periods of time) given their code will quickly change? Why or why not?

THINK