

# *Topology of Viral Evolution*

---

JOSEPH MINHOW CHAN, GUNNAR CARLSSON, AND RAUL RABADAN

CENTER OF COMPUTATIONAL BIOLOGY, BIOINFORMATICS

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY

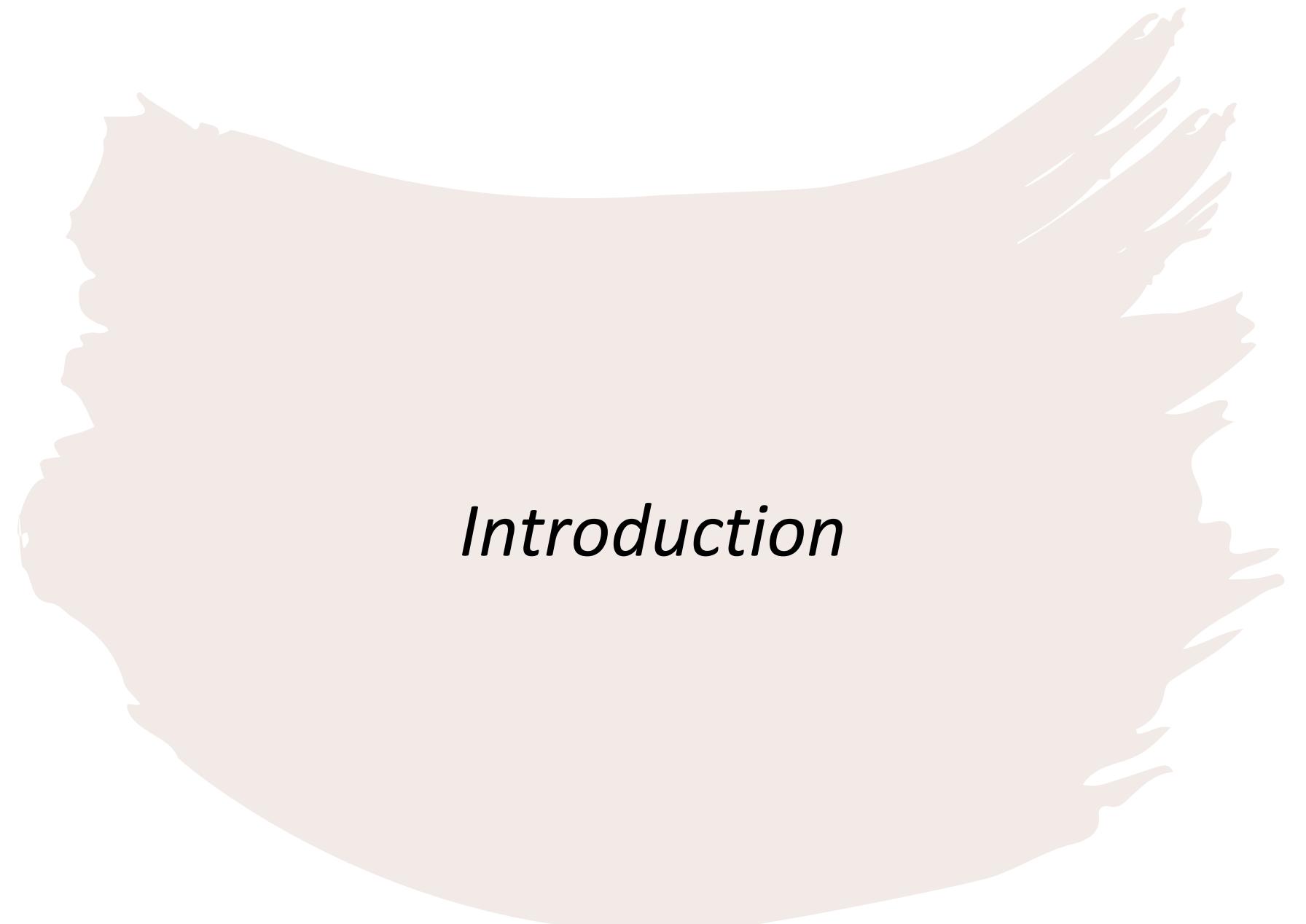
REPORTER: TI-WEN LU

SUPERVISOR: DR. RIVER CHIANG

**PNAS**  
www.pnas.org

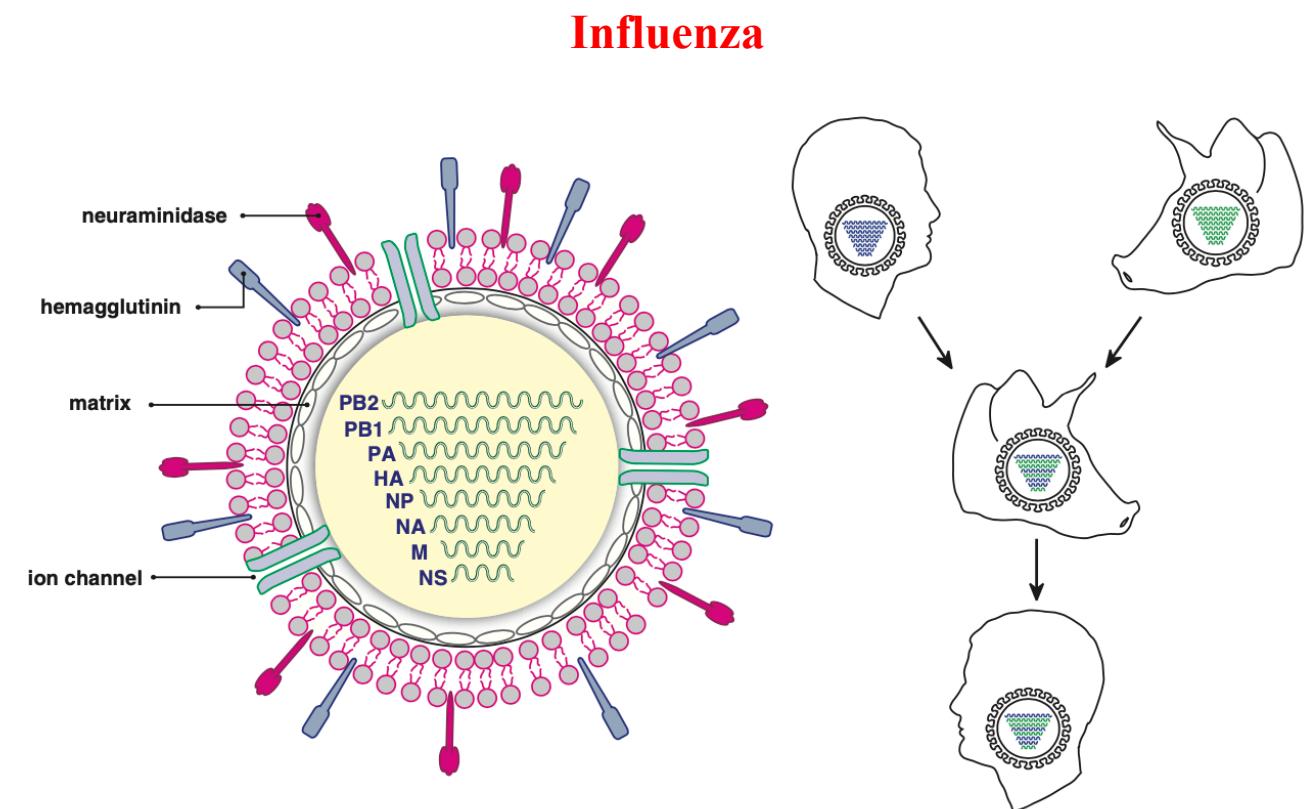
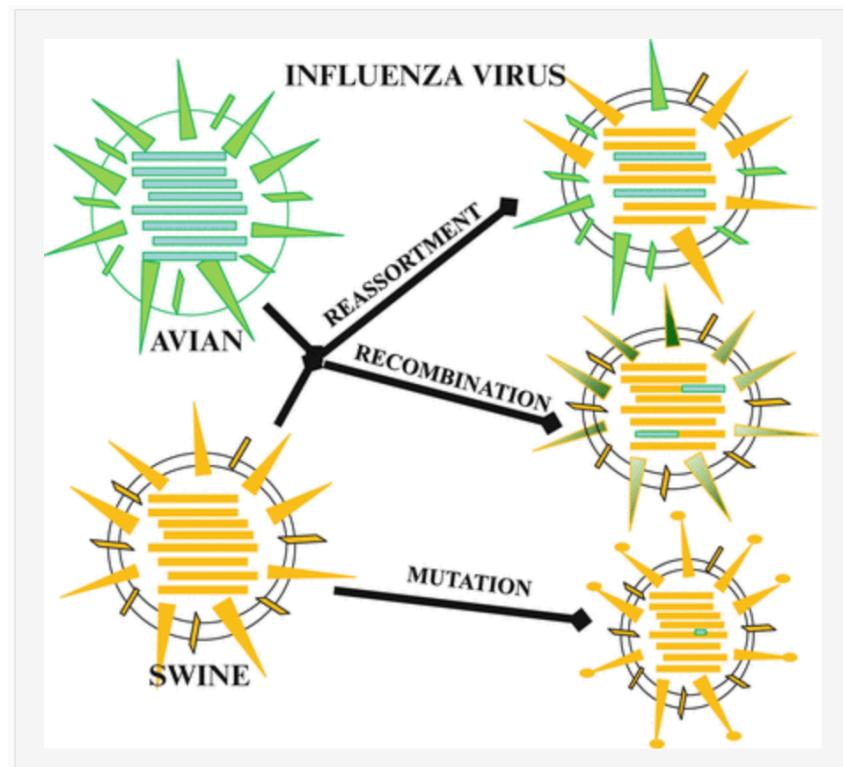
# *Table of Contents*

- Introduction
  - Viral Reassortment and Recombination
  - Horizontal Evolution
  - Representation of Horizontal Evolution
    - Persistent Homology & Barcode Diagram
- Results
- Conclusion

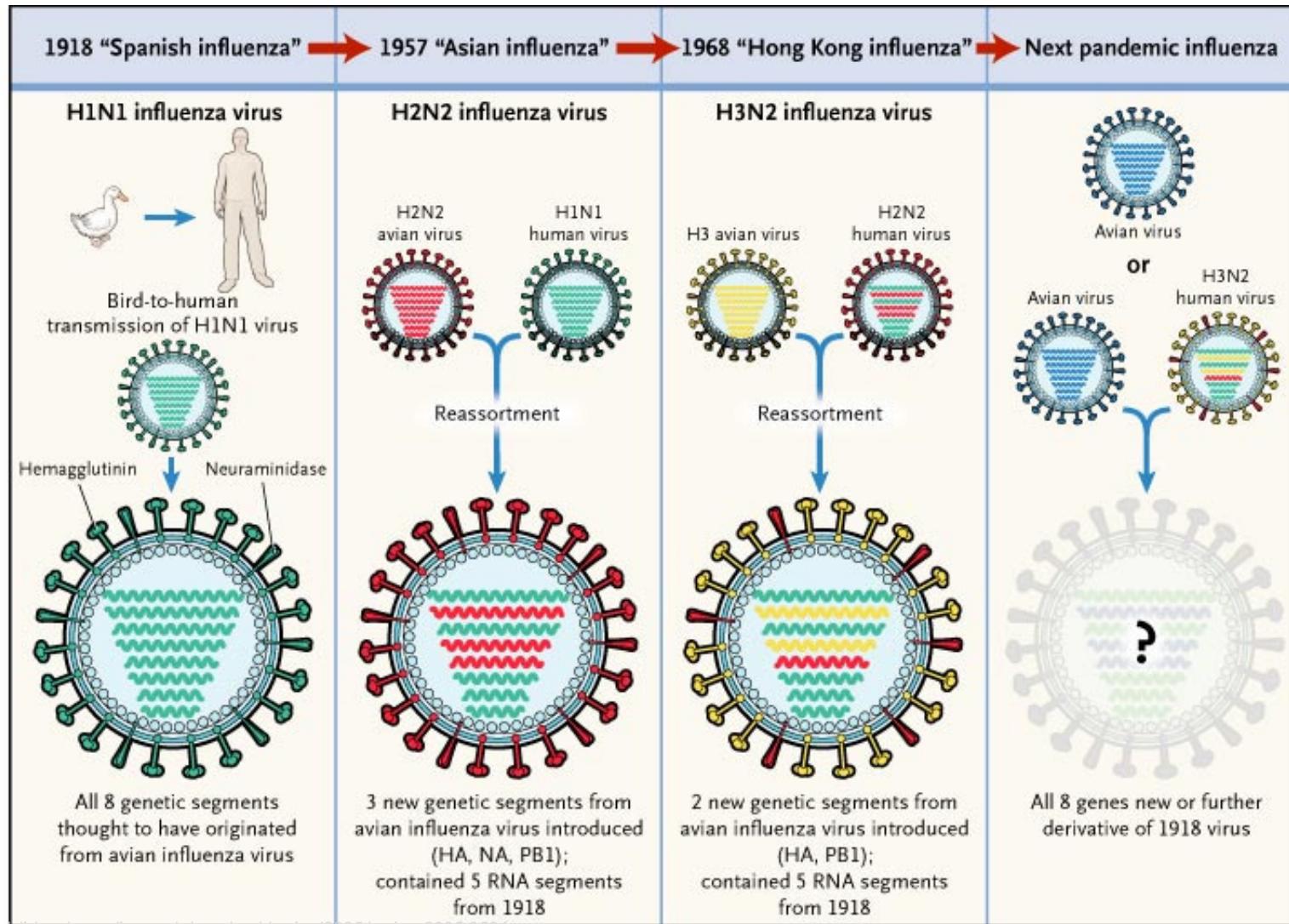


# *Introduction*

# Viral Reassortment and Recombination

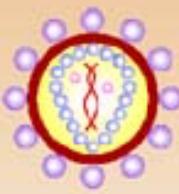


# Viral Reassortment and Recombination

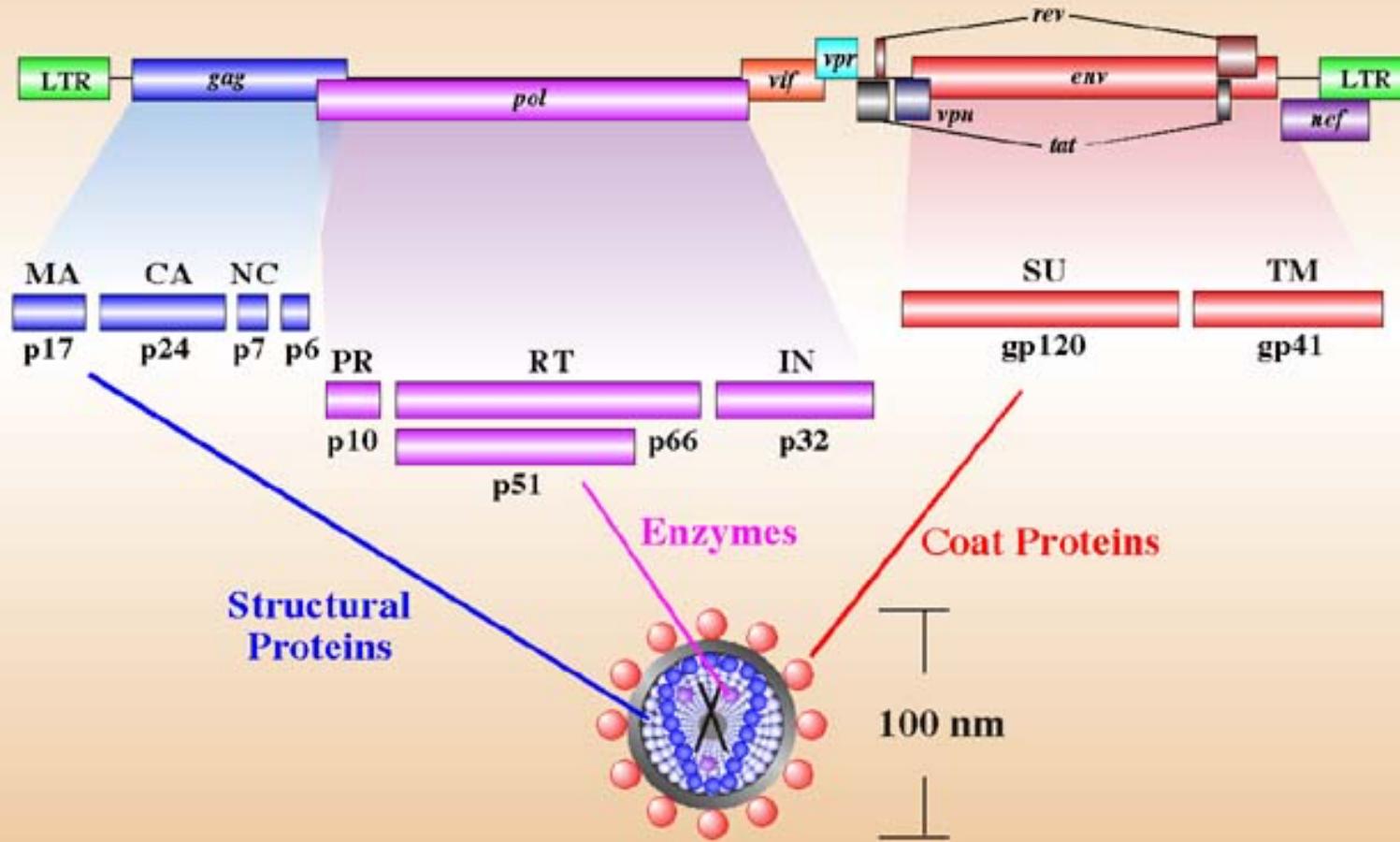


# HIV

- A kind of retrovirus
- Notorious for high diversity mediated by a high mutation rate and frequent homologous recombination.
- Leading to antiretroviral resistance and immune evasion



# Genomic Organization of HIV-1



Volume 3, Number 3—September 1997

*Perspective*

## Recombination in HIV: An Important Viral Evolutionary Strategy

Donald S. Burke✉

Author affiliation: Walter Reed Army Institute of Research, Rockville, Maryland, USA

[Cite This Article](#)

### Abstract

Human immunodeficiency virus (HIV) is a diploid virus: each virion carries two complete RNA genomic strands. Homologous recombination can occur when a cell is coinfecte<sup>d</sup> with two different but related strains. Naturally occurring recombinant HIV strains have been found in infected patients in regions of the world where multiple genotypic variants cocirculate. One recombinant HIV strain has spread rapidly to millions of persons in Southeast Asia. Recombination is a mechanism whereby high level and multidrug-resistant strains may be generated in individual treated patients. Recombination also poses theoretical problems for the development of a safe HIV vaccine. Certain features of HIV replication, such as syncytium formation and transactivation, may be best understood as components of a sexual reproductive cycle. Recombination may be an important HIV evolutionary strategy.

Human immunodeficiency virus (HIV)-1, like all retroviruses, is "diploid." Each viral particle contains two RNA strands of positive polarity, each full length and potentially able to replicate (1). No other virus families, RNA or DNA, are diploid. Typically both RNA strands in a retroviral particle derive from the same parent provirus. However, if an infected cell simultaneously harbors two different proviruses, one RNA transcript from each provirus can be encapsidated into a single "heterozygous" virion. When this virion subsequently infects a new cell, the reverse transcriptase may jump back and forth between the two RNA templates so that the newly synthesized retroviral DNA sequence is recombinant between that of the two parents (2). All subsequent progeny virions will be of this recombinant genotype. HIV-1 strains with chimeric genomes thought to have arisen through homologous recombination have recently been discovered in nature (3).

Temin observed that the replication strategy of HIV-1 suggests a form of primitive sexual reproduction (4), which is apparently genderless but sexual in that 1) two parental gametes must fuse into a single progeny, 2) the genetic information of the parental strains is recombined, and 3) subsequent offspring carry genetic information from both parents.

### On This Page

[Theoretical Advantages and Disadvantages of Recombination](#)

[Evidence of HIV Recombination in Nature](#)

[Mechanisms of Retroviral Recombination](#)

[HIV as a Primitive Sexual Organism](#)

[The Lentivirus Gene Pool and Origins of Contemporary HIV Clades](#)

[Barriers to HIV Recombination](#)

[Consequences of Recombination for Prevention and Treatment of HIV Infections](#)

[Conclusion](#)

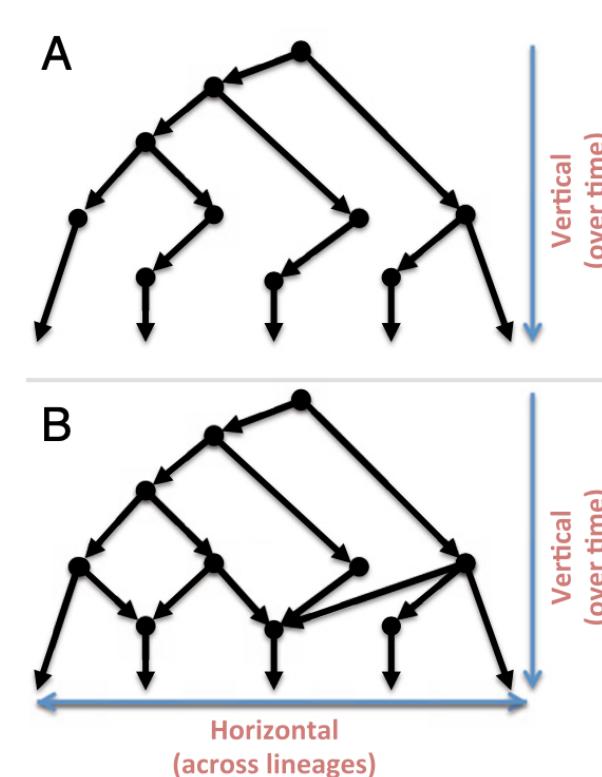
[Cite This Article](#)

### Downloads

[Article](#) 

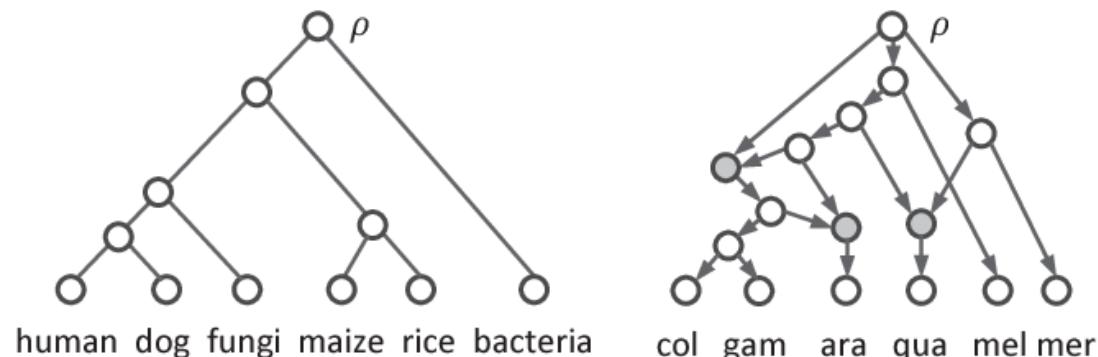
# *Horizontal Evolution*

- Horizontal evolution occurs when distinct clades merge together to form a new hybrid lineage.
- For example:
  - Species hybridization in eukaryotes
  - Lateral gene transfer in bacteria
  - Recombination and reassortment in viruses
  - Viral integration in eukaryotes
  - Fusion of genomes of symbiotic species



# *Representation of Horizontal Evolution*

- Phylogenetic tree is unable to characterize horizontal evolution.
  - The tree structure models vertical evolution mediated by random mutations over multiple generations.
- Phylogenetic networks

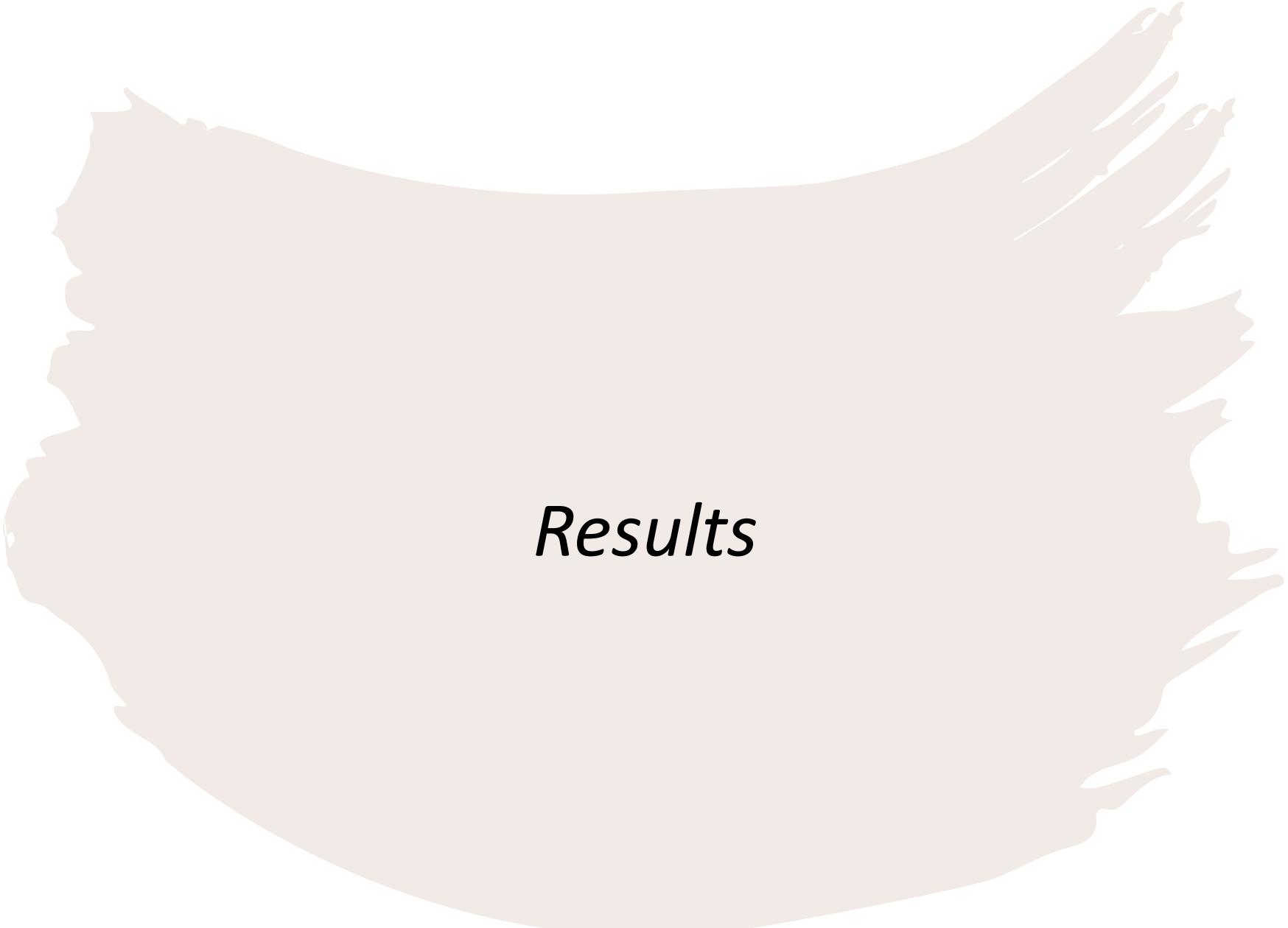


- Cons:
  - Results depend on the ordering of samples in the data matrix.
  - NP-hard to determine whether a tree exists in an evolutionary network



## *Aim*

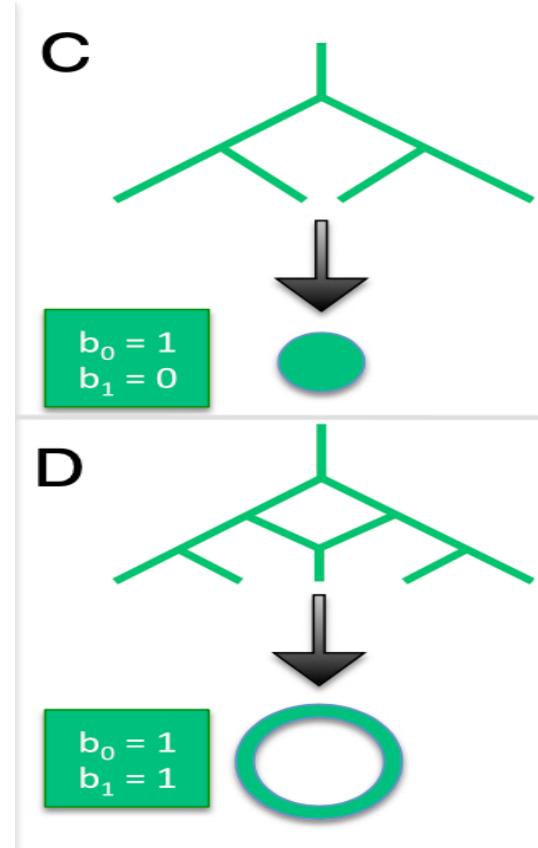
To propose a comprehensive and fast method of extracting large scale patterns from genomic data that captures both vertical and horizontal evolution at the same time.



# *Results*

# Persistent Homology in Evolution

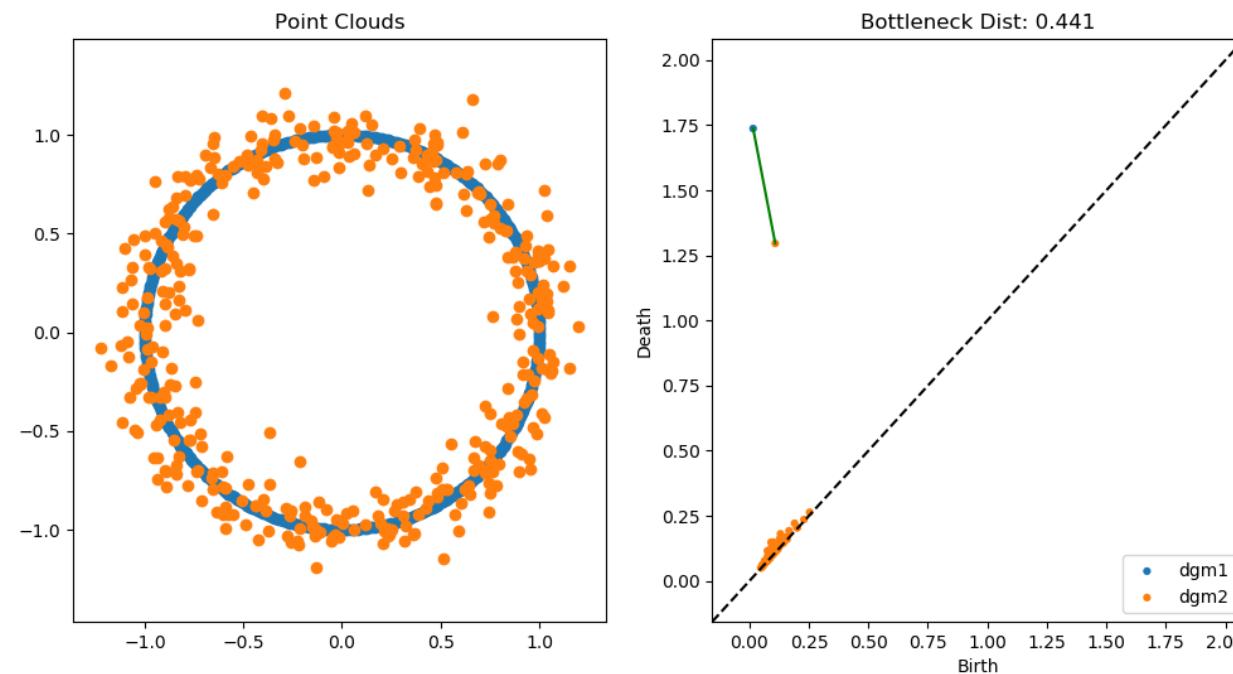
**Theorem 2.1** Let  $M$  be any tree-like finite metric space, and let  $r \geq 0$ . Then the complex  $V(M, r)$  is a disjoint union of acyclic complexes. In particular,  $H_i(V(M, r)) = \{0\}$  for  $i \geq 1$ .



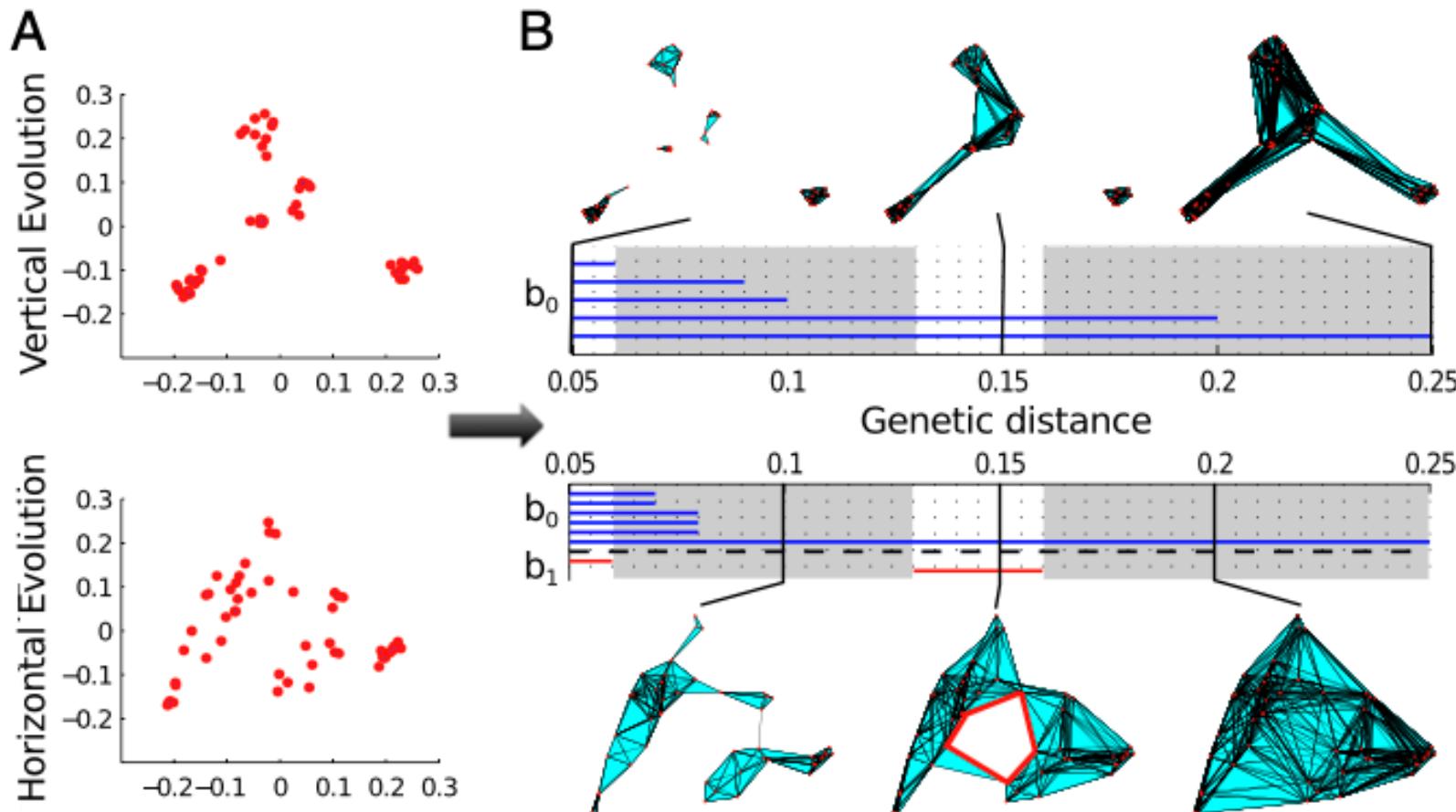
# Persistent Homology in Evolution

**Theorem 2.2** *Each of the assignments are distance non-increasing, so for any two compact metric spaces  $X, Y$ , we have*

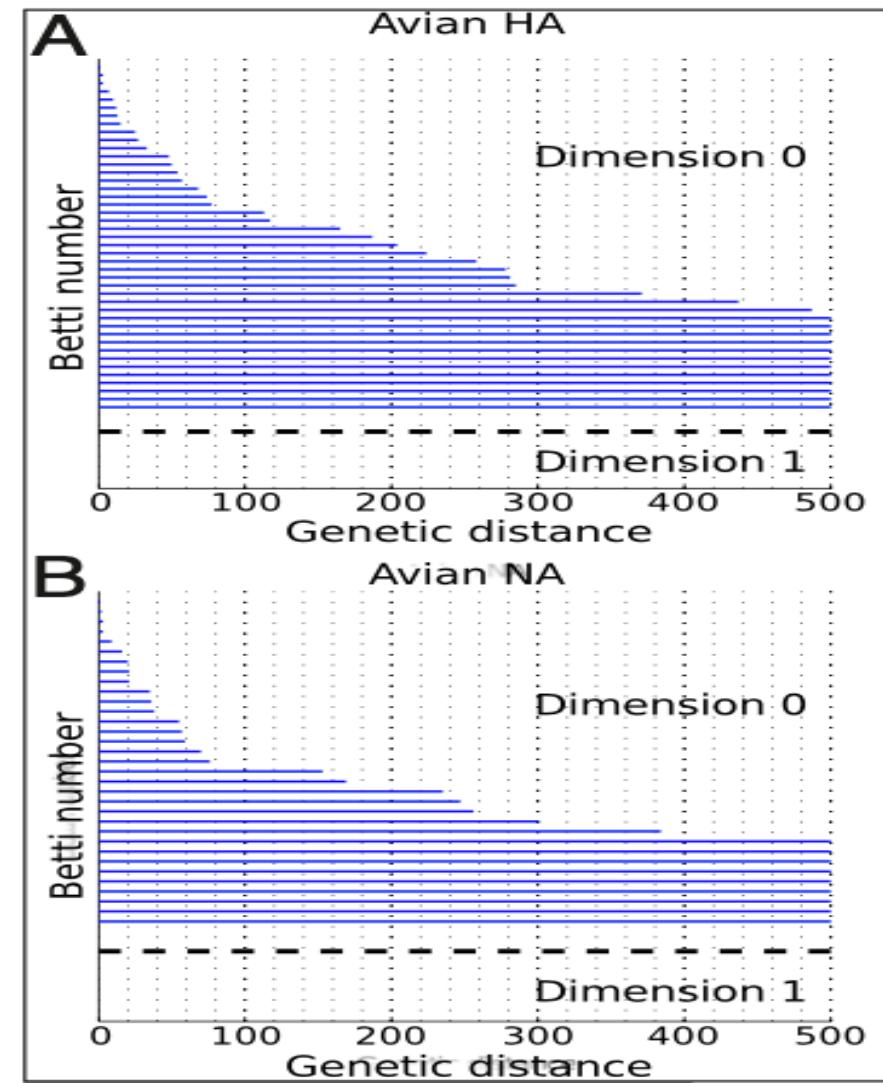
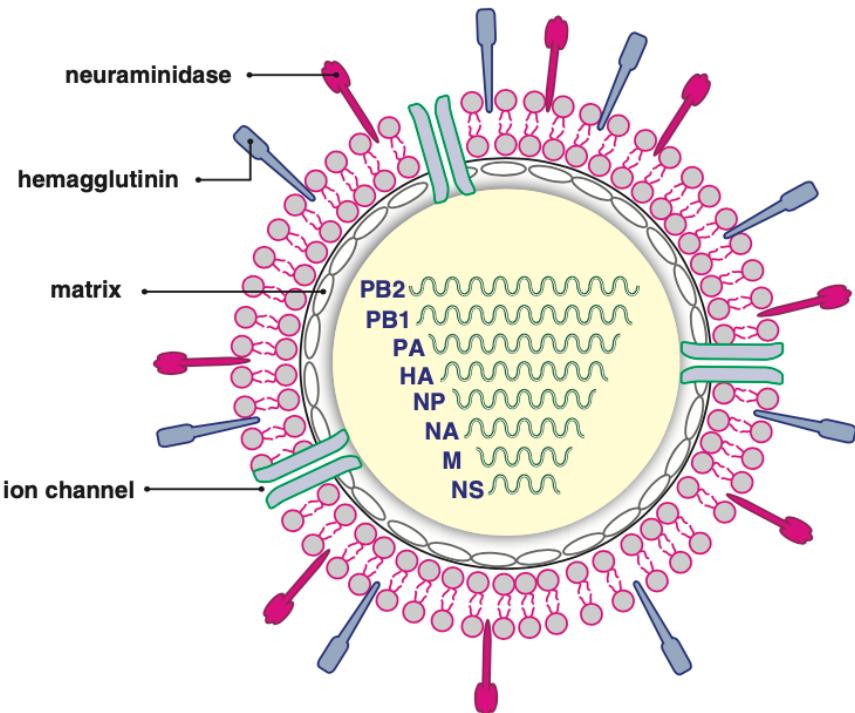
$$d_B(H_k(X), H_k(Y)) \leq d_{GH}(X, Y)$$



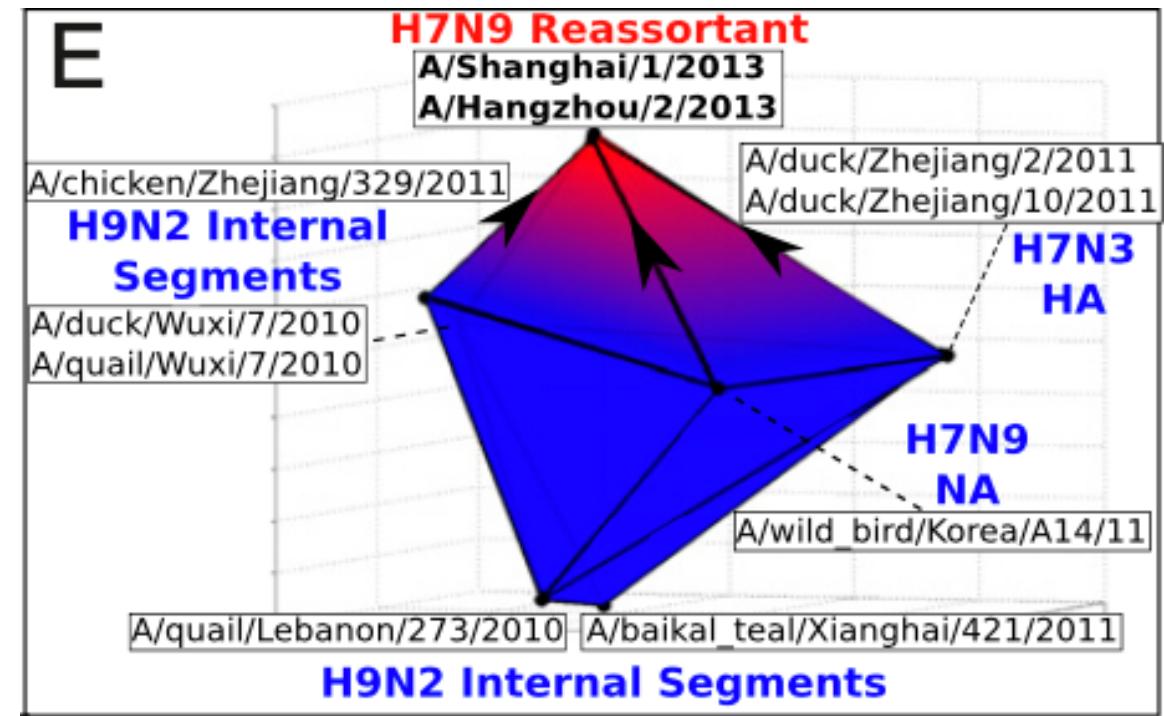
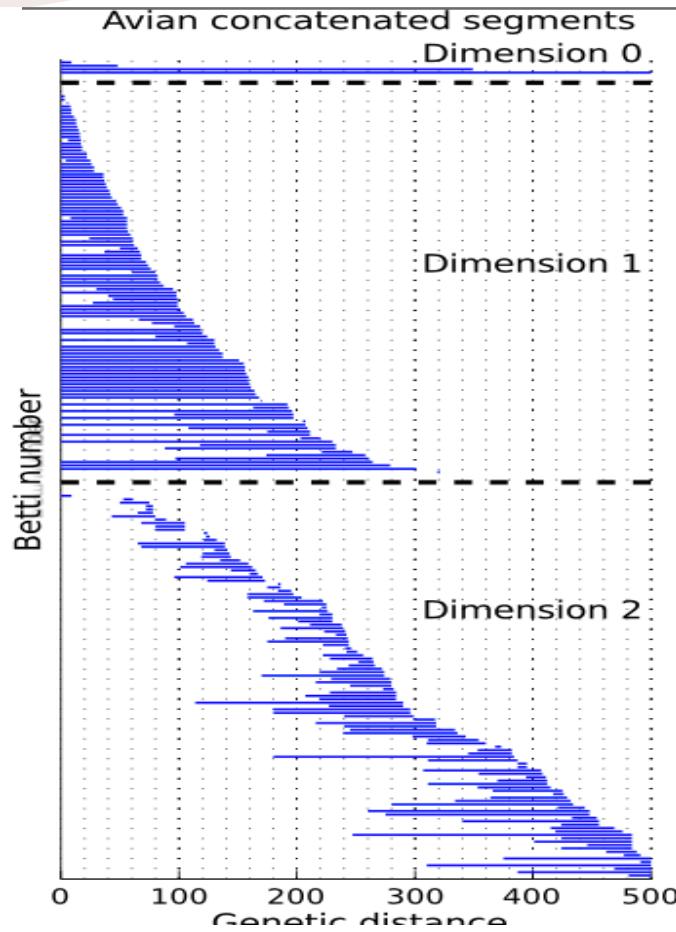
*Persistent homology characterizes topological features of vertical and horizontal evolution.*



# *Vertical Evolution in Influenza*

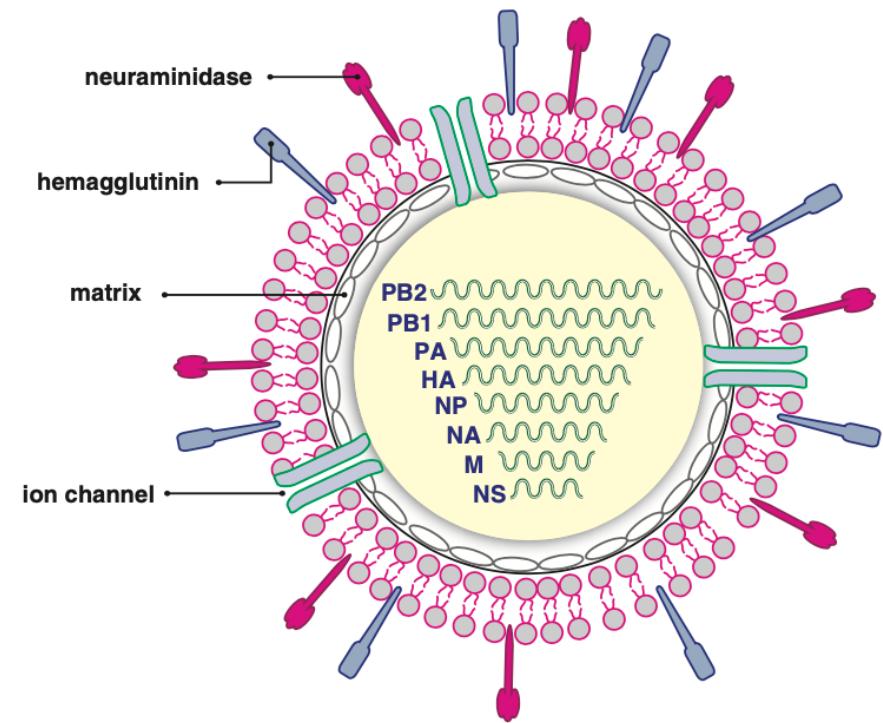
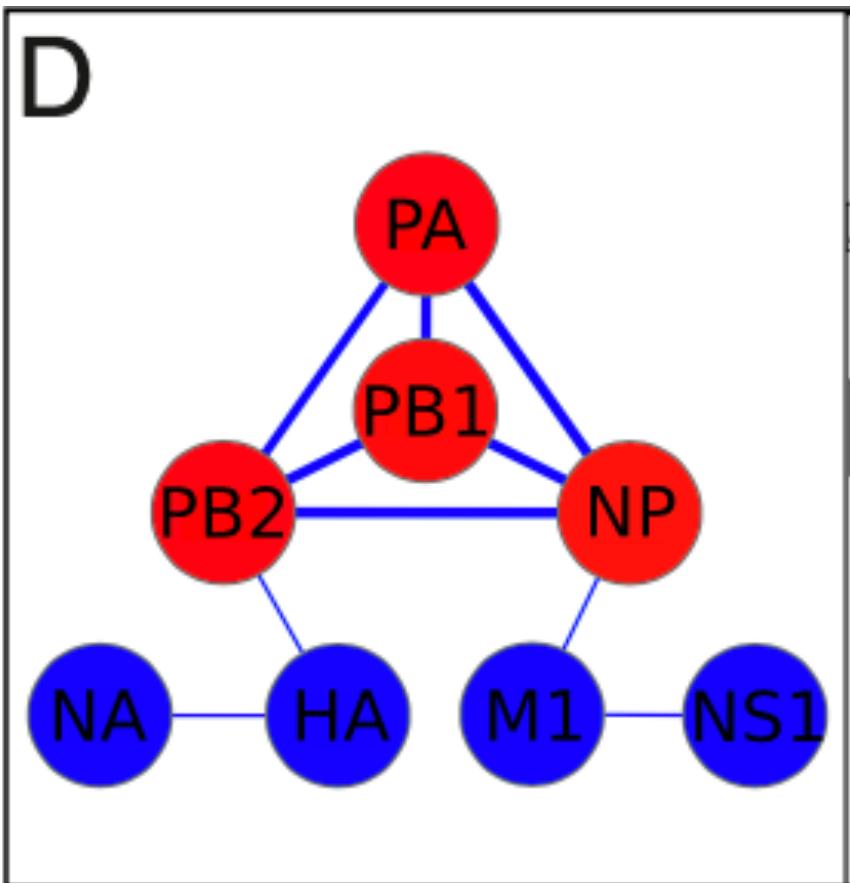


# *Reassortment in Influenza Evolution*

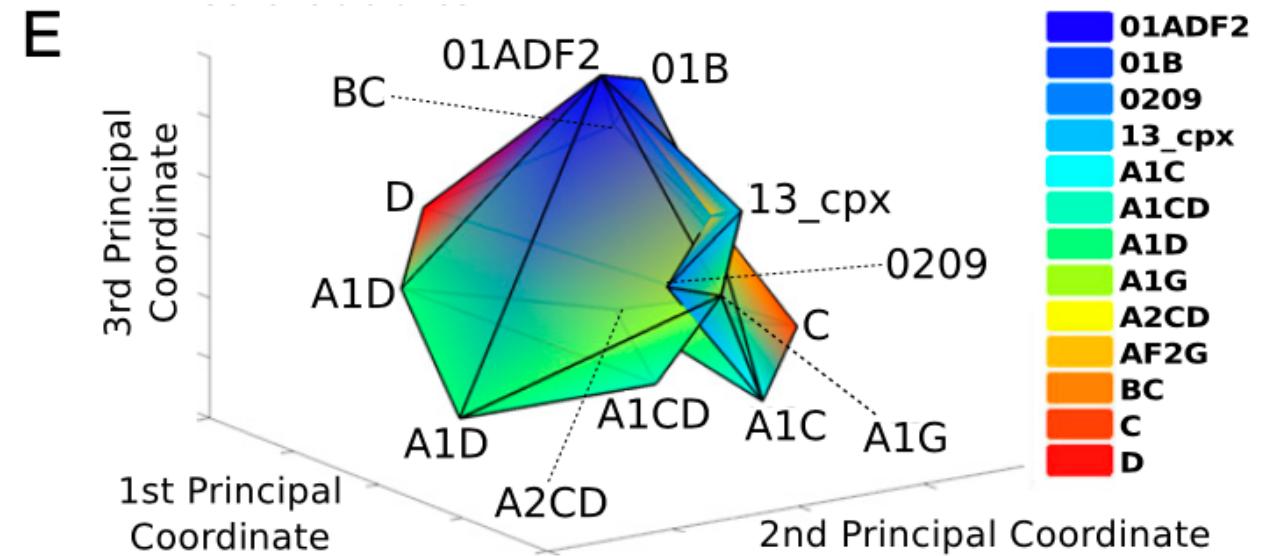
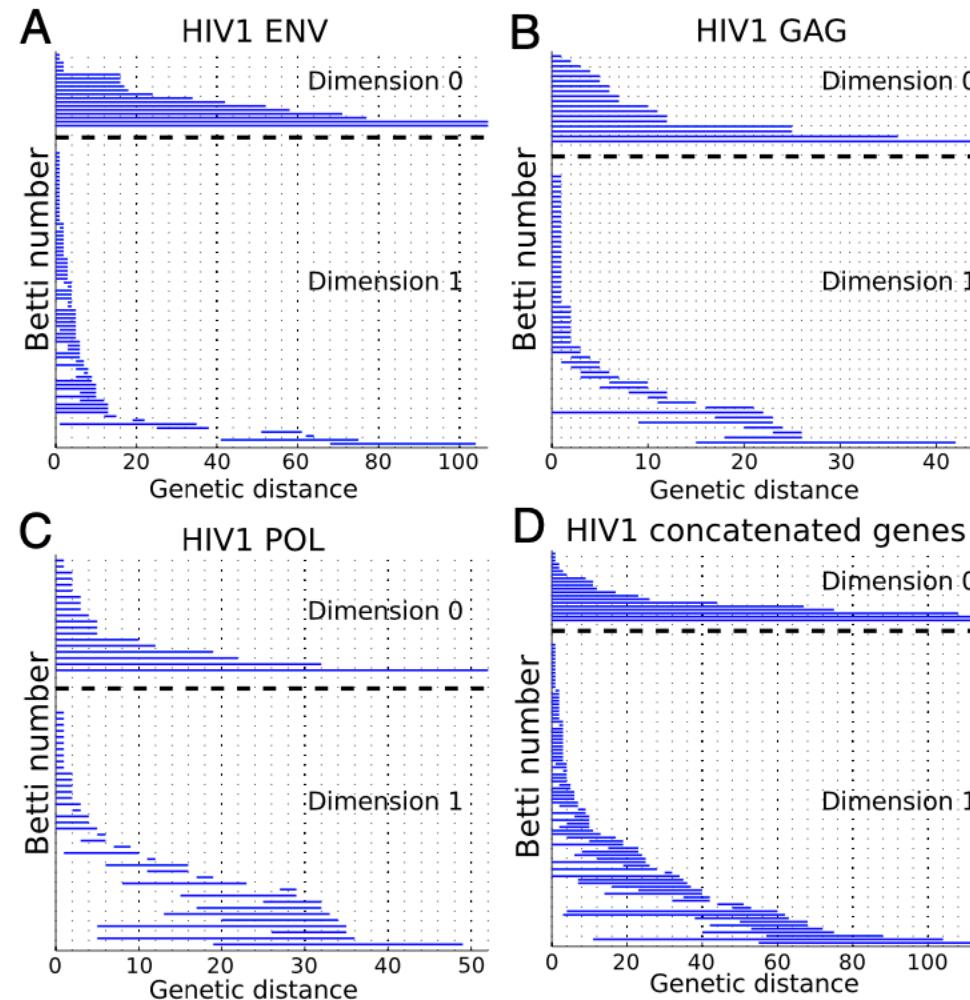


Recapitulating the H7N9 Avian Influenza Triple Reassortment

# *Nonrandom Reassortment Pattern in Avian Influenza*



# Topology of HIV Recombination Evolution



# *Conclusion*

- Pros:
  - **Quickly** determine **robust** topological properties of evolution from big genomic datasets.
- Cons:
  - In general, in order to make the result explainable, designing or determining the data input such as “distance” or “data points” are the most difficult.

# *Reference*

- Raul Rabadan, Andrew J. Blumberg - Topological Data Analysis for Genomics and Evolution,  
**Cambridge University Press**

## *Supplementary*

- For ordinary topological spaces  $X$ , the (mod-2) homology of  $X$  in dimension  $k$  is a Boolean vector space whose dimension counts the number of independent occurrences of certain patterns thought of as  $k$ -dimensional holes.
- It is computed using matrix operations applied to a  $k$ -dimensional boundary matrix.

