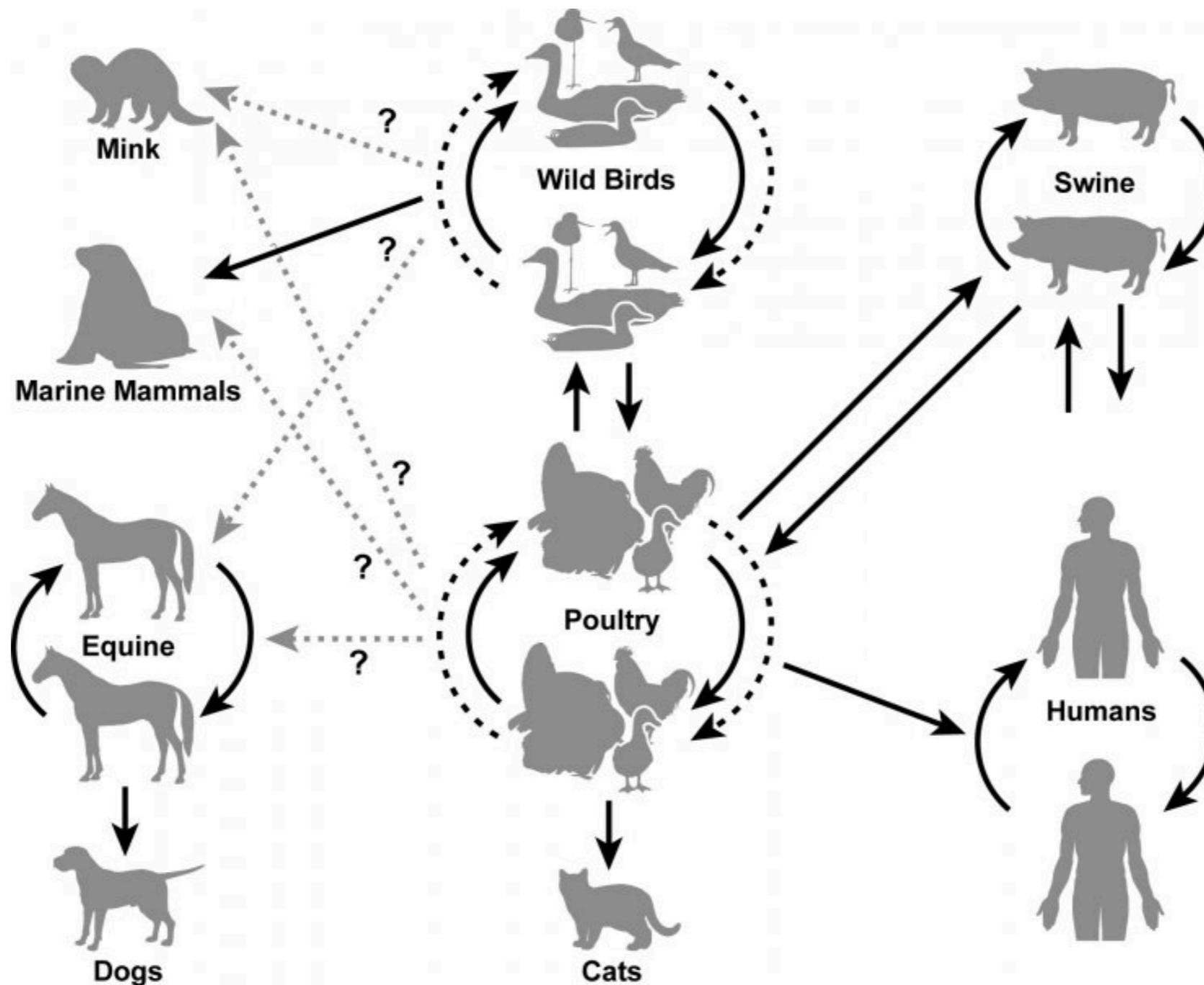


Identification of Reassortant Influenza Viruses at Scale: Algorithm & Applications

Eric J. Ma
27 April 2017

The influenza A virus is a highly successful virus.

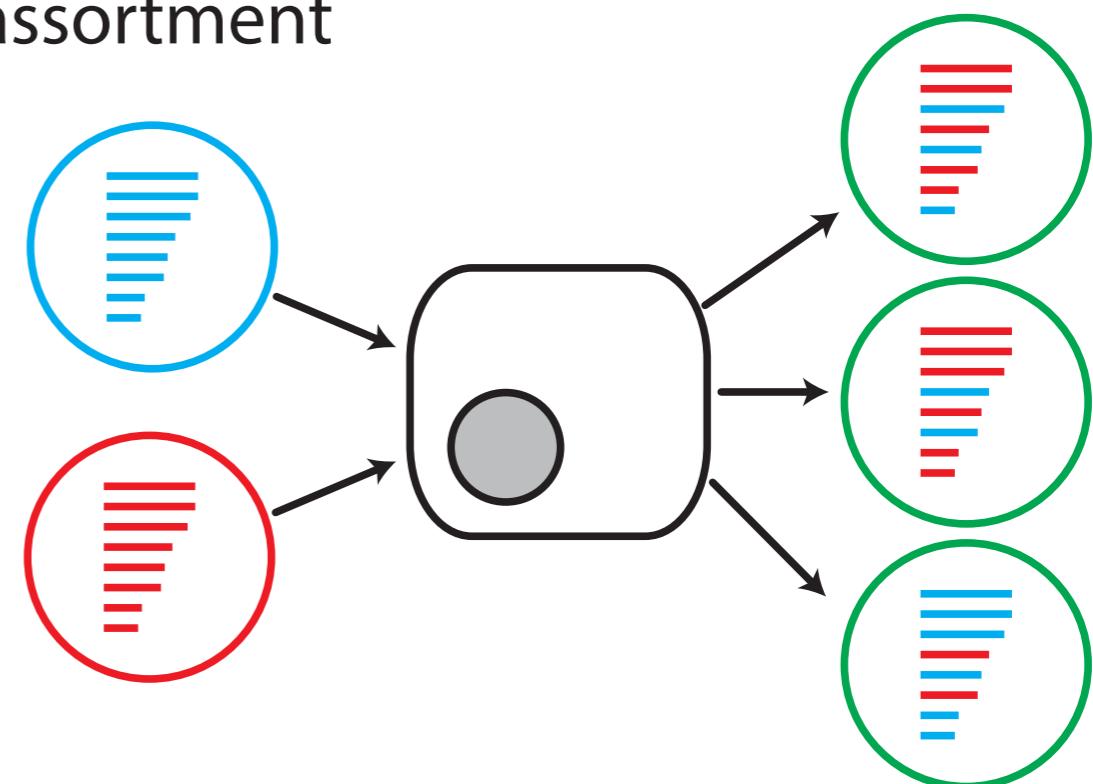


The influenza A virus has a segmented RNA genome that enables reassortment.

Influenza Genome Structure



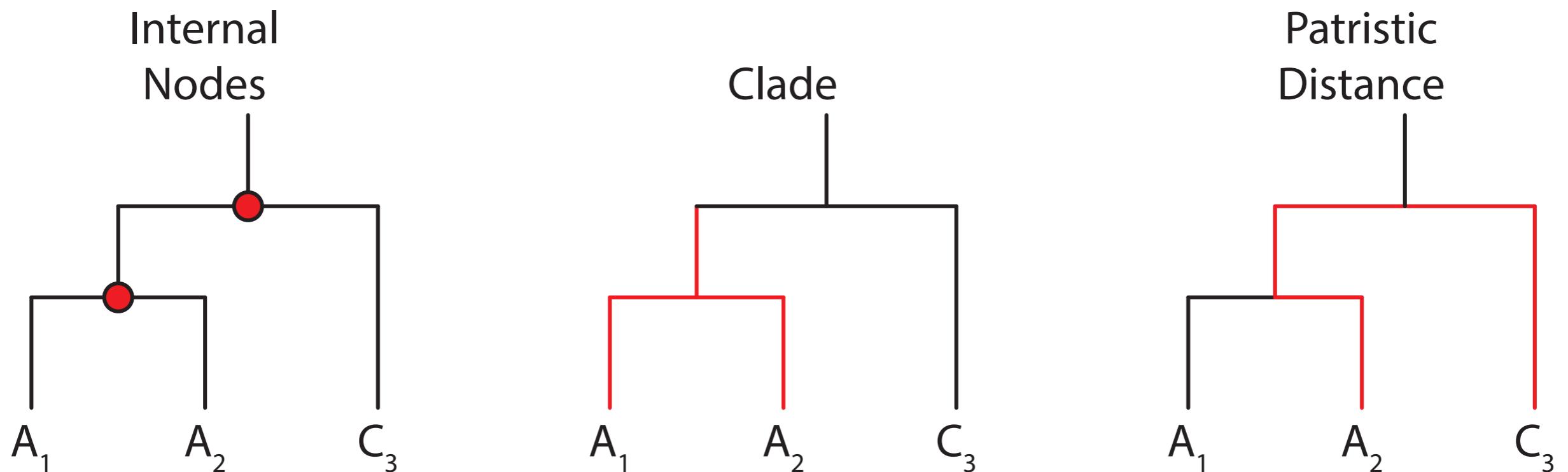
Reassortment



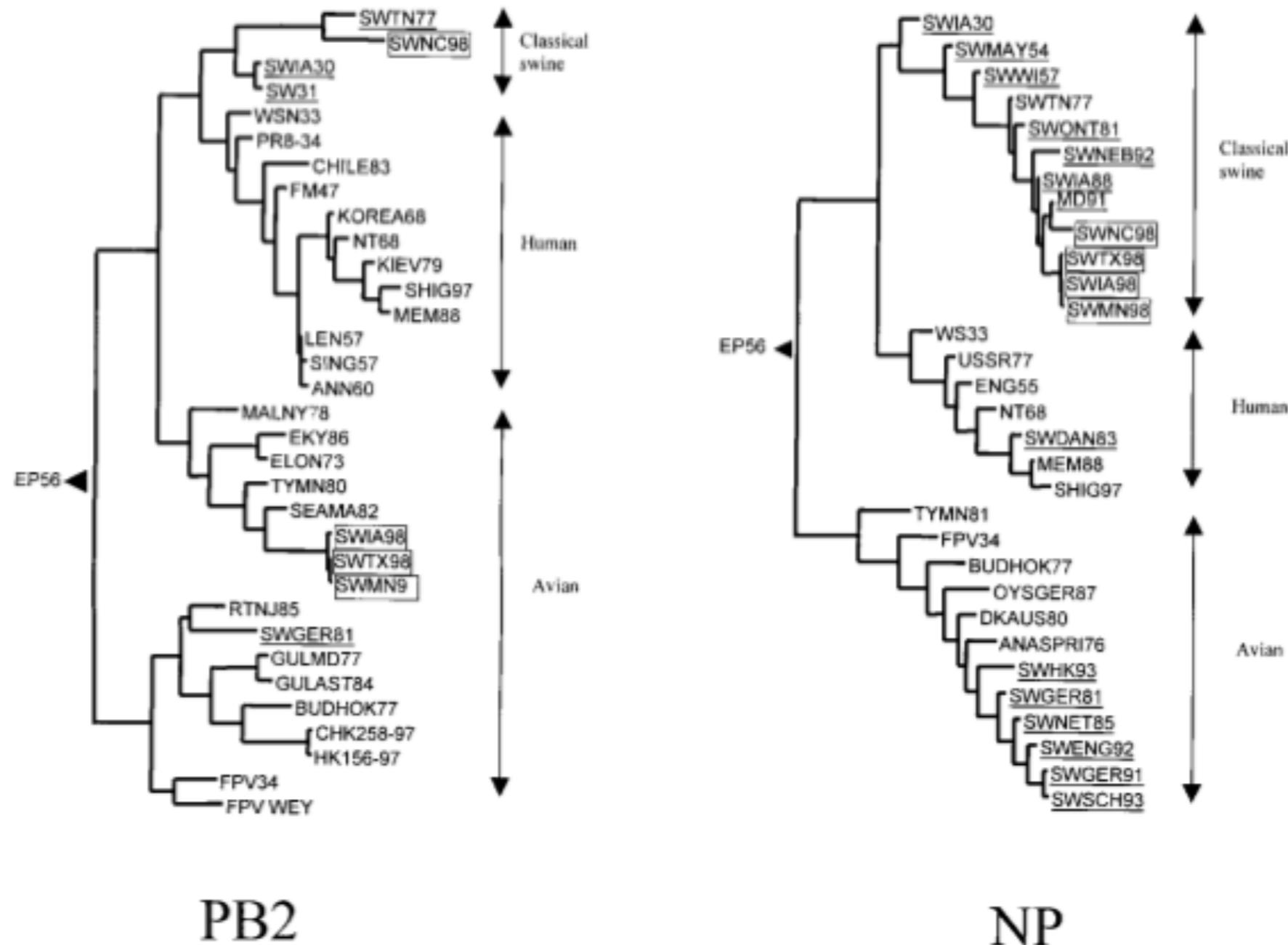
Reassortment is implicated
in pandemics... and host
switches?

And is this really true when tested systematically?

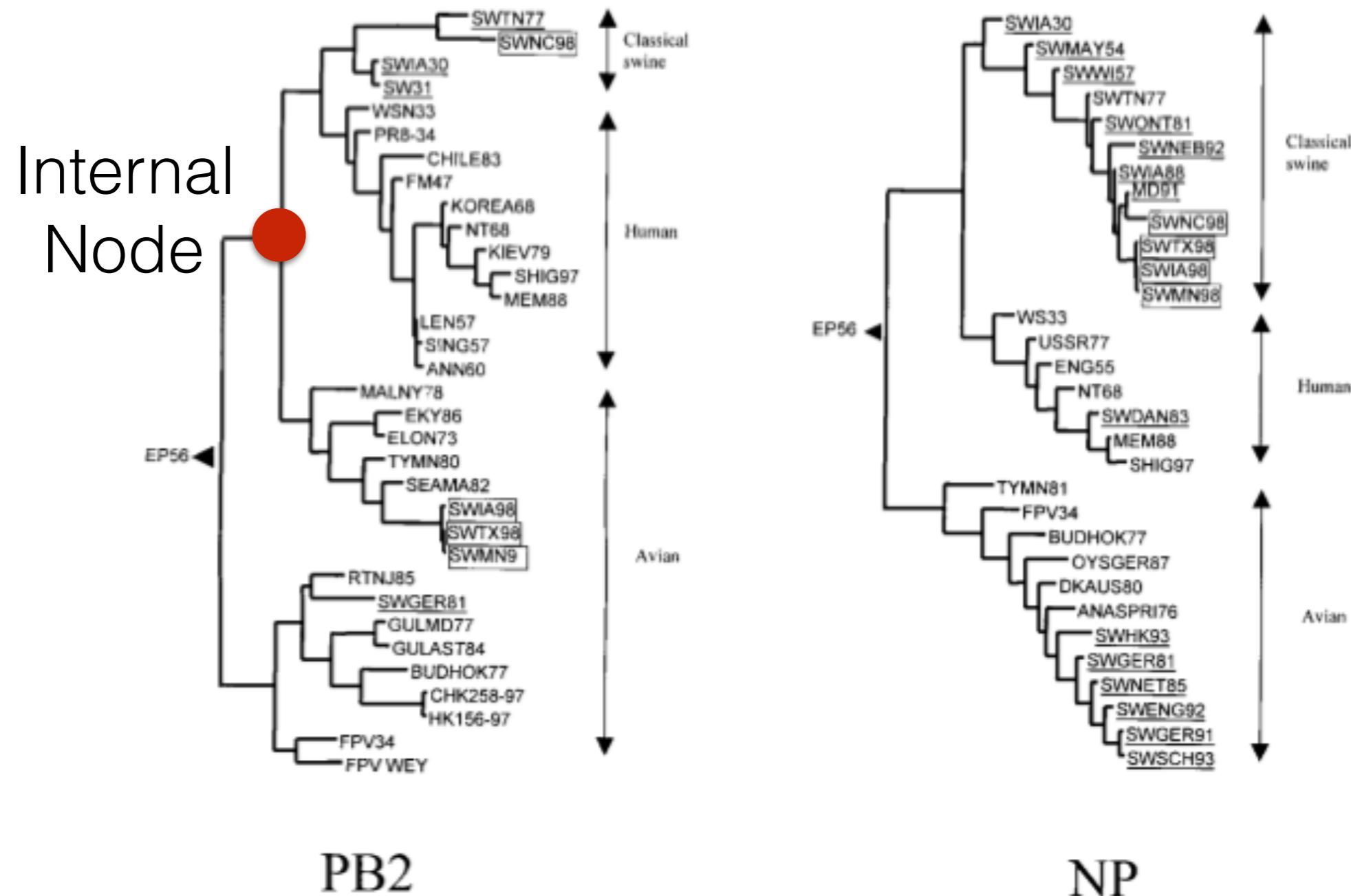
Phylogenetic trees are the dominant tool for understanding viral evolution.



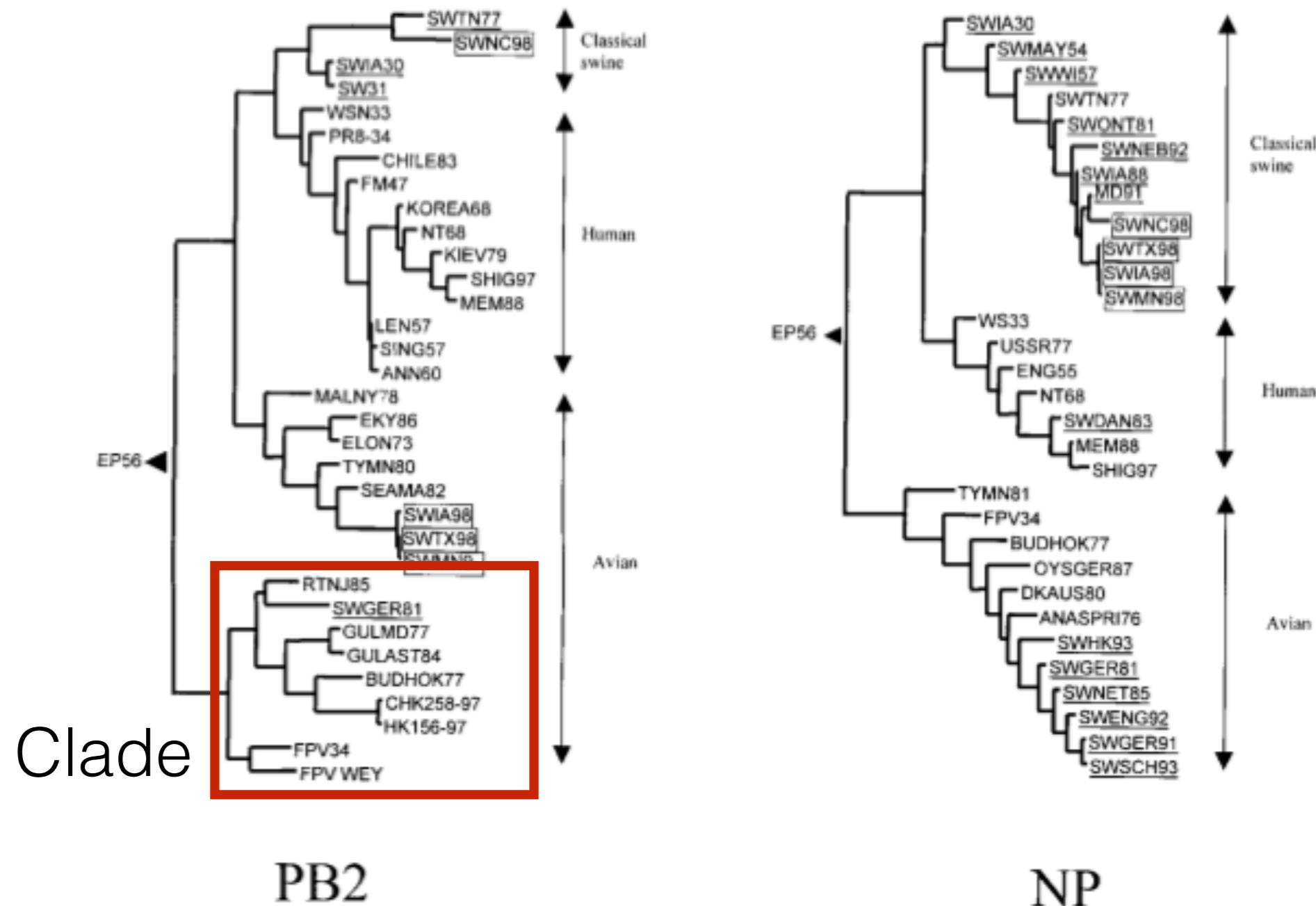
Phylogenetic trees are the dominant tool for understanding viral evolution.



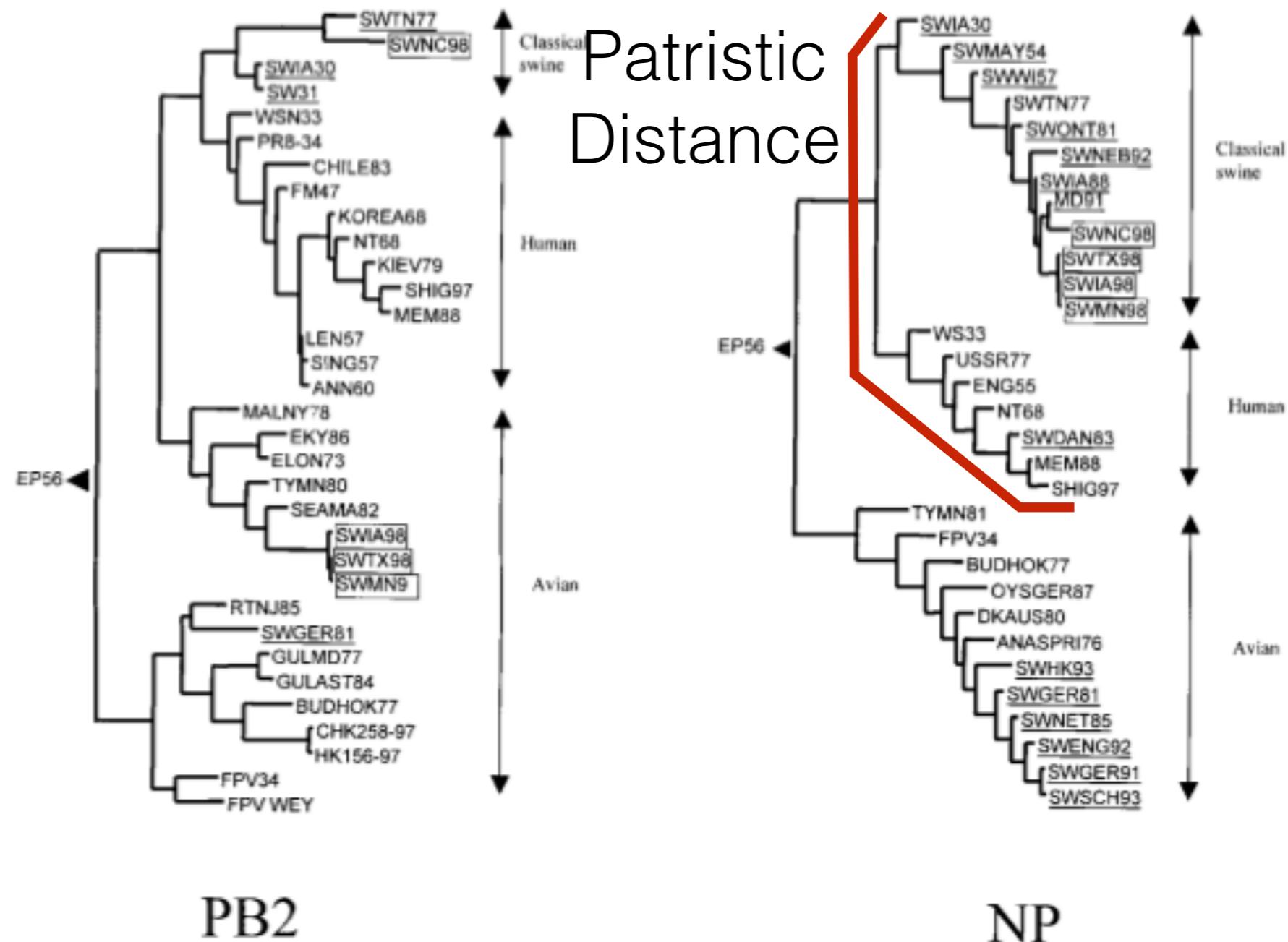
Phylogenetic trees are the dominant tool for understanding viral evolution.



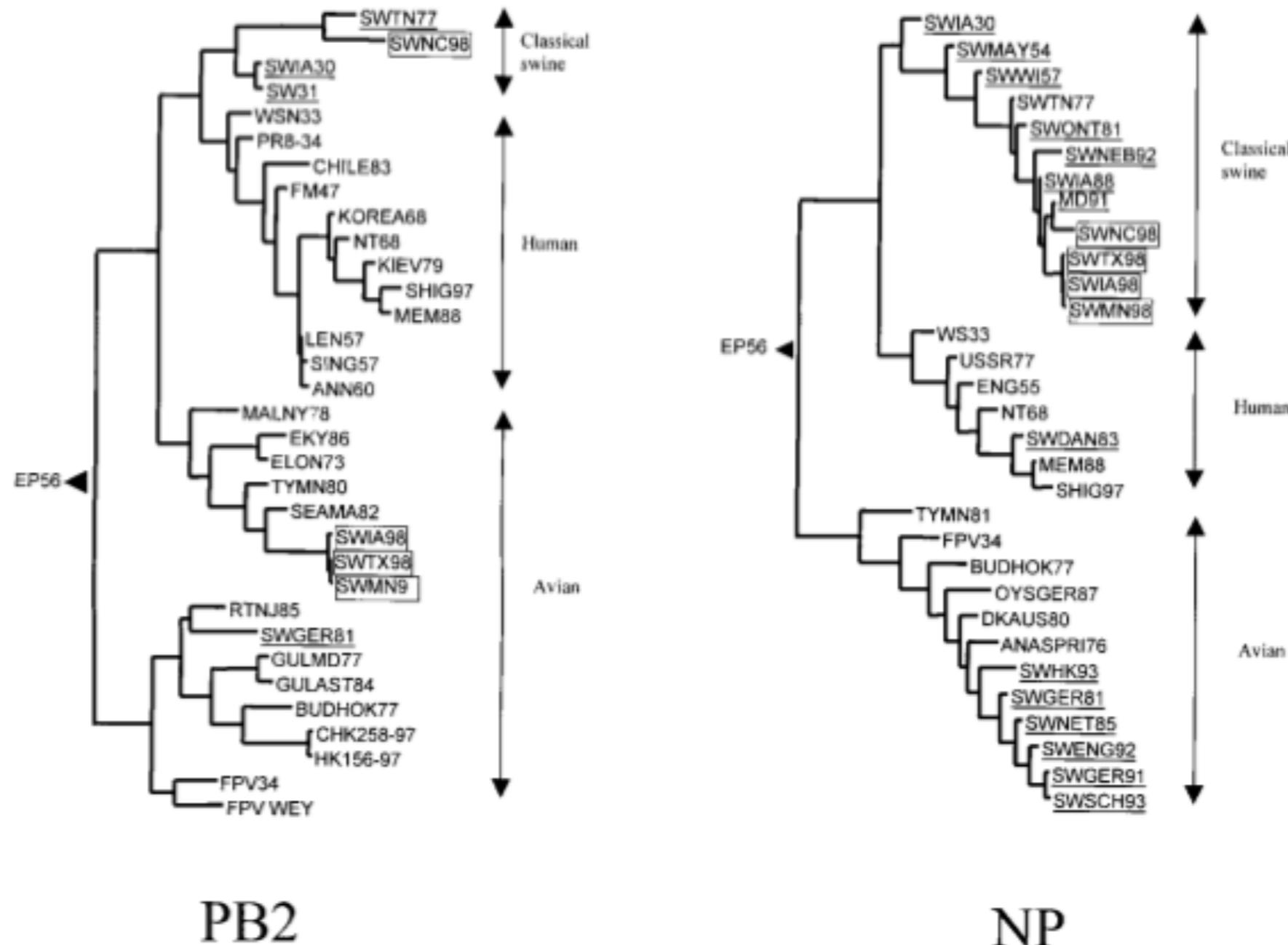
Phylogenetic trees are the dominant tool for understanding viral evolution.



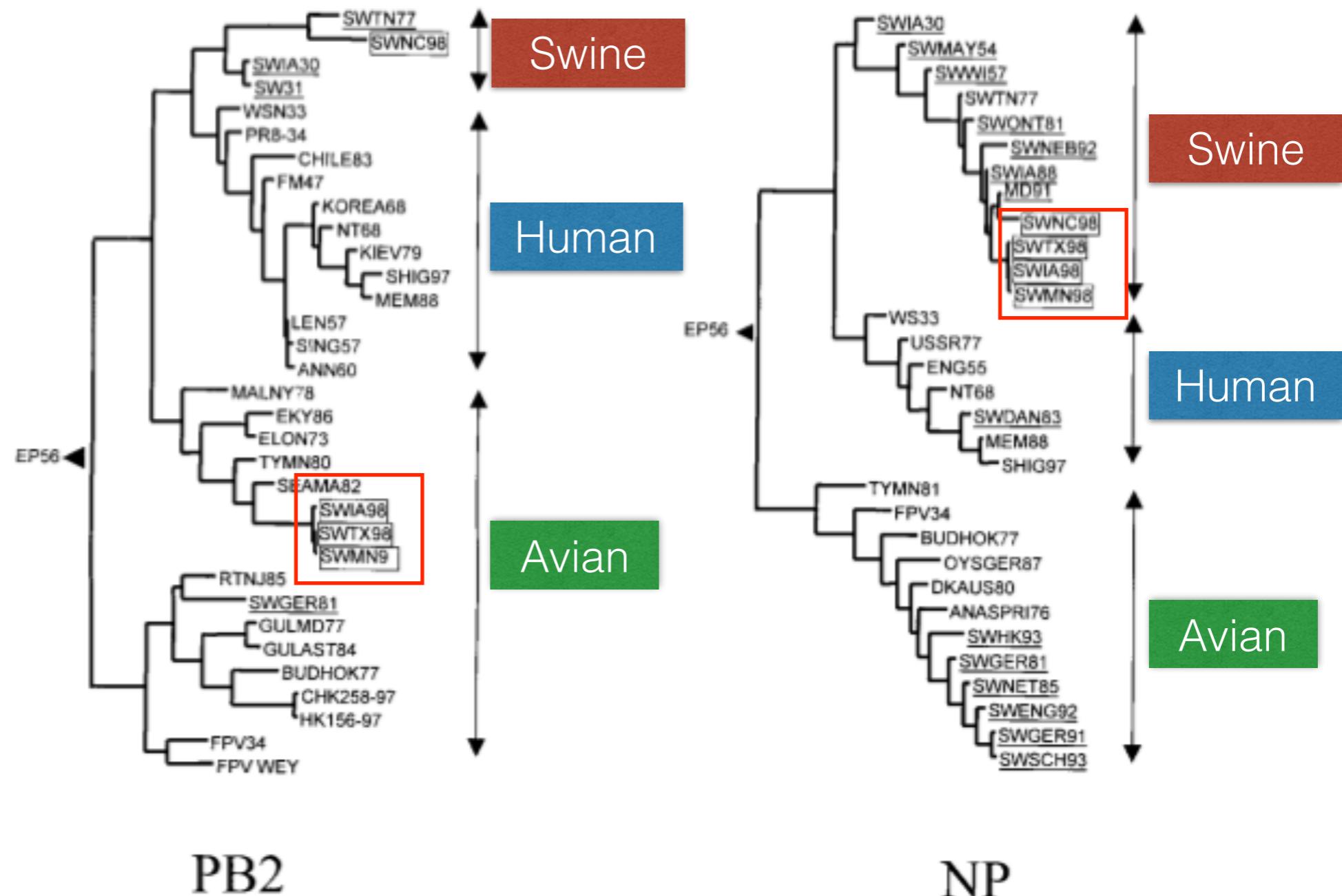
Phylogenetic trees are the dominant tool for understanding viral evolution.



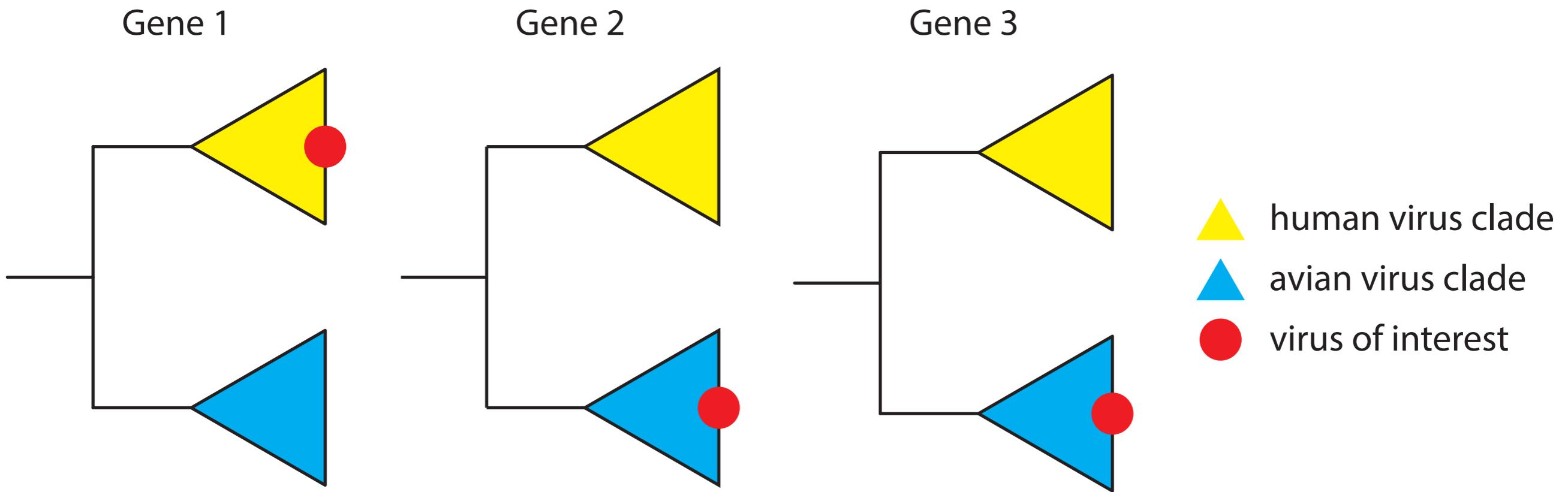
Phylogenetic trees are the dominant tool for understanding viral evolution.



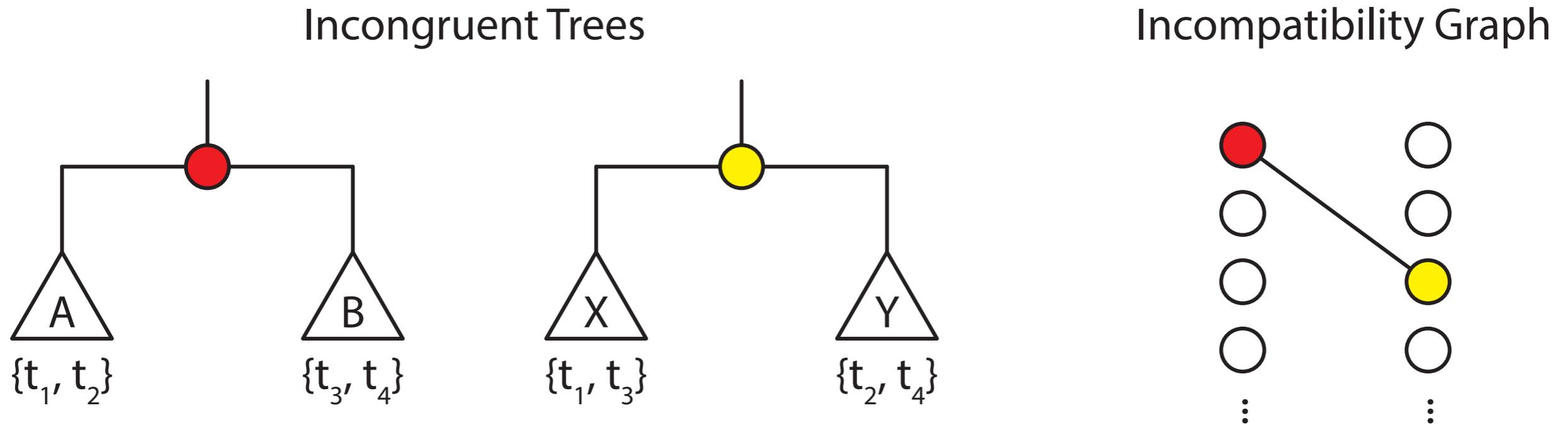
Reassortment is traditionally identified by phylogenetic tree incongruence.



Phylogenetic tree incongruence, or “promiscuous clustering”



Tree splits generalize phylogenetic tree incongruence.



Problem: The best-in-class reassortant-finding algorithms spend most of their time in tree construction. This makes systematically studying reassortment infeasible.

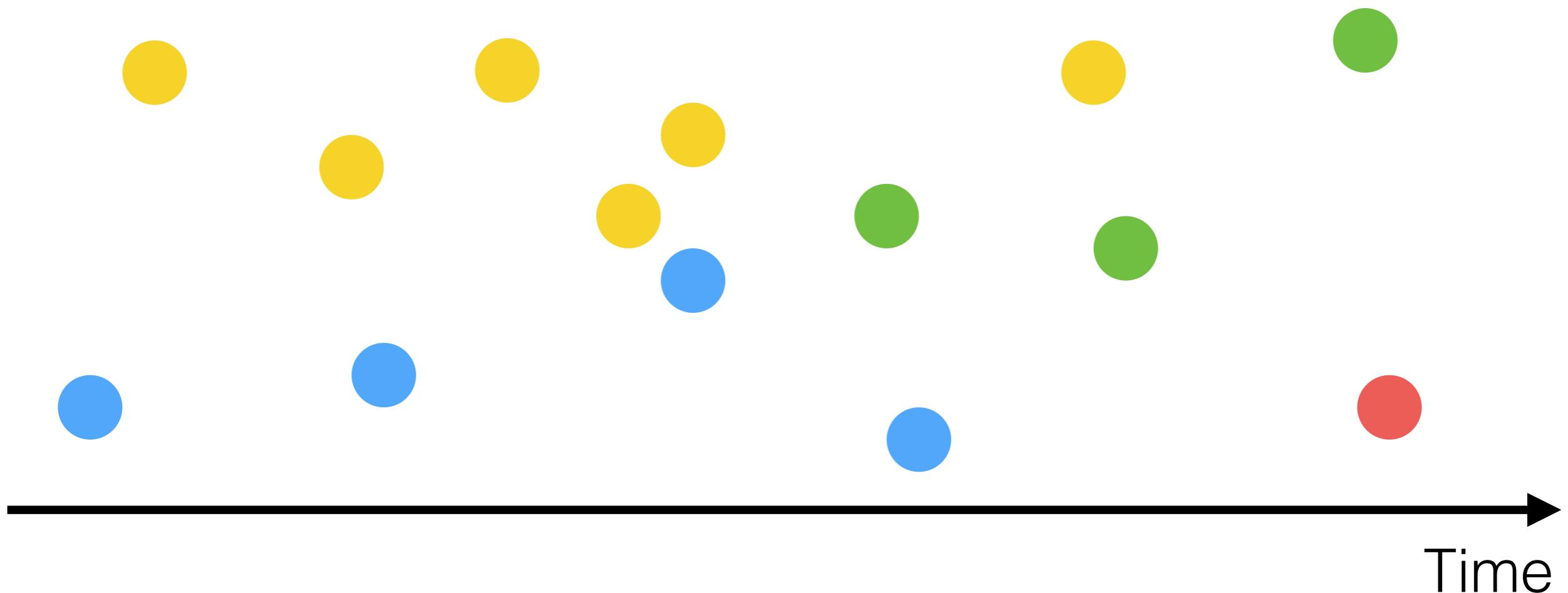
Algorithm for finding reassortant viruses at scale.

Inspiration: Jombart et. al., 2011, Heredity

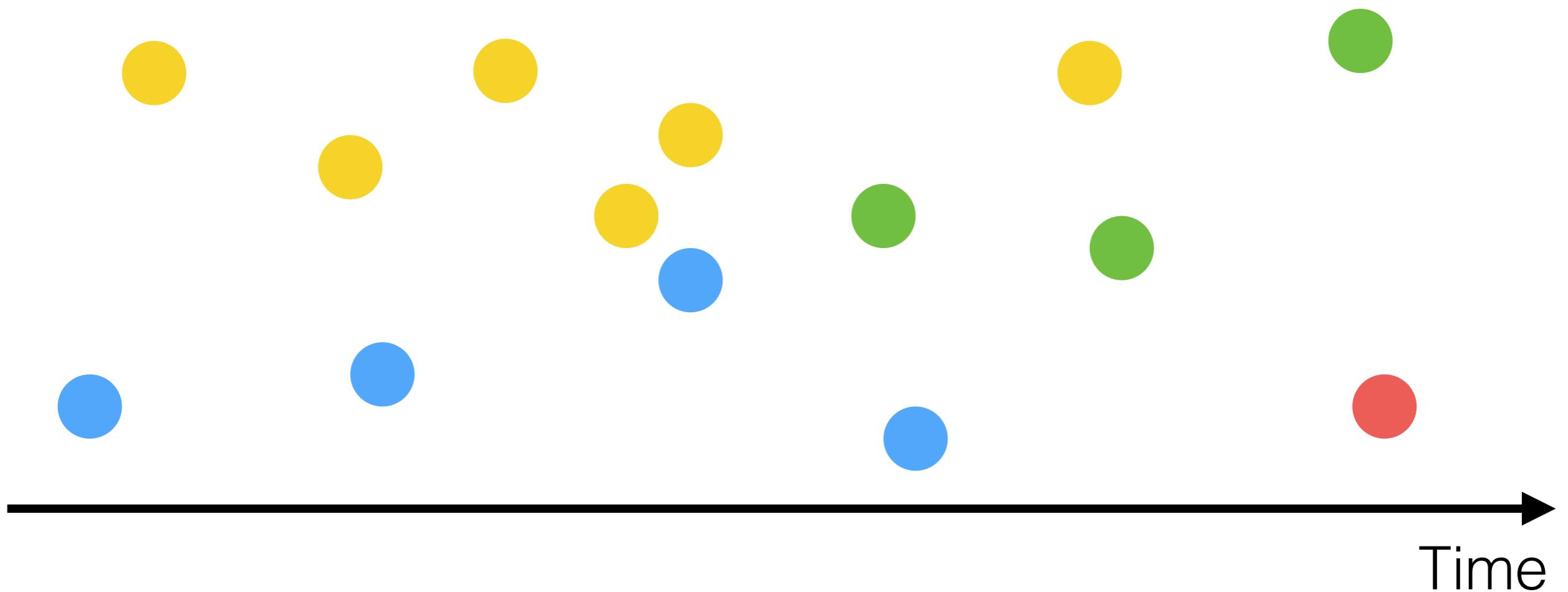


Algorithm for finding reassortant viruses at scale.

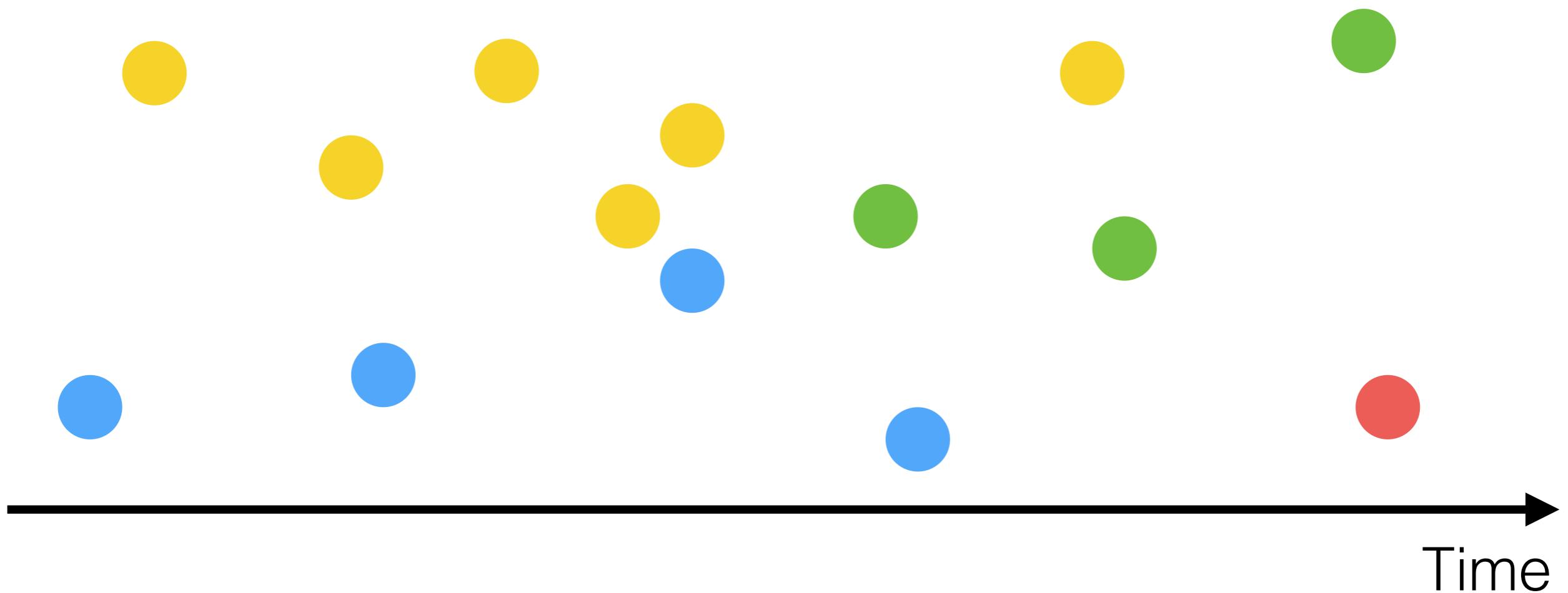
Inspiration: Jombart et. al., 2011, Heredity



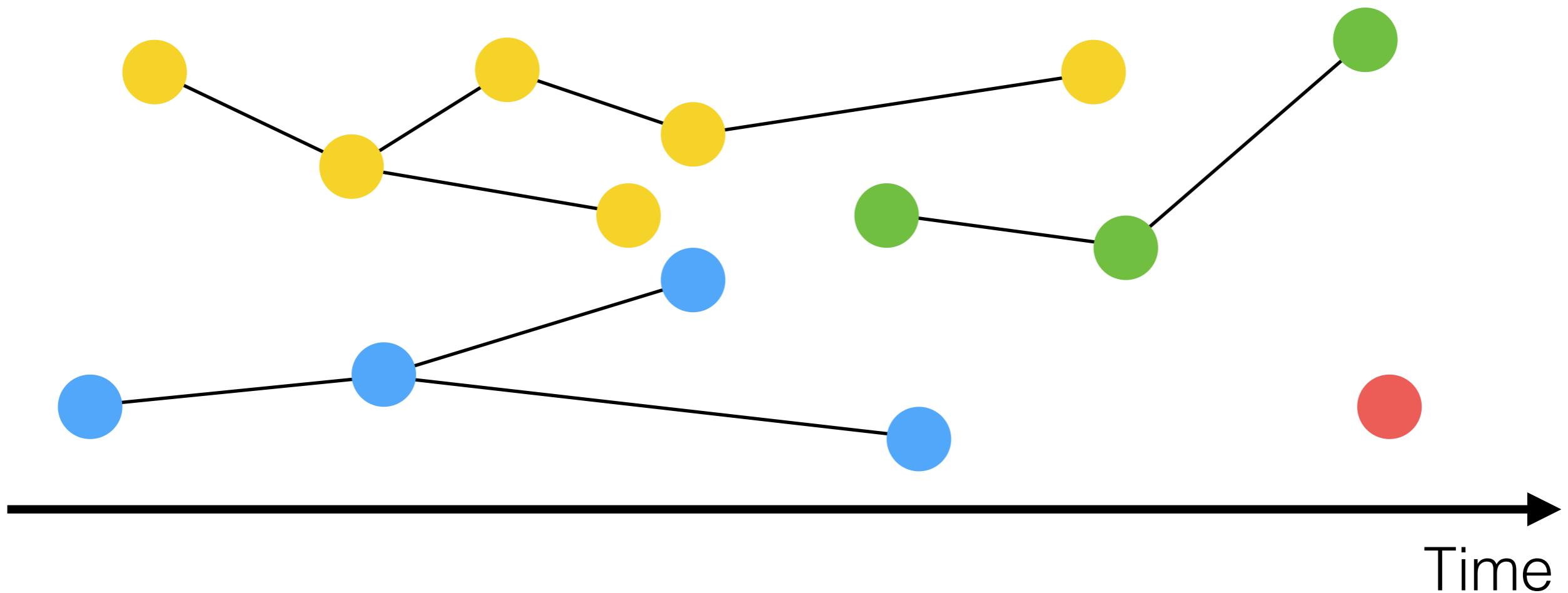
Algorithm for finding reassortant viruses at scale.



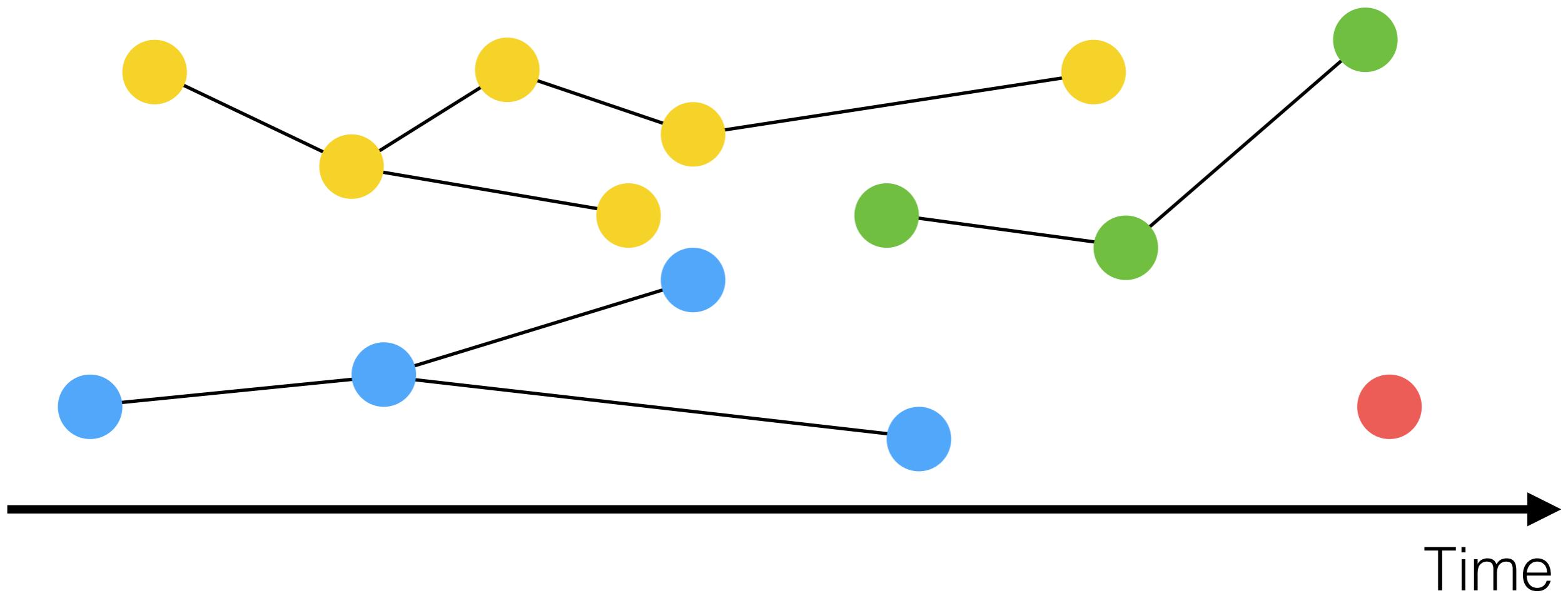
Algorithm for finding reassortant viruses at scale.



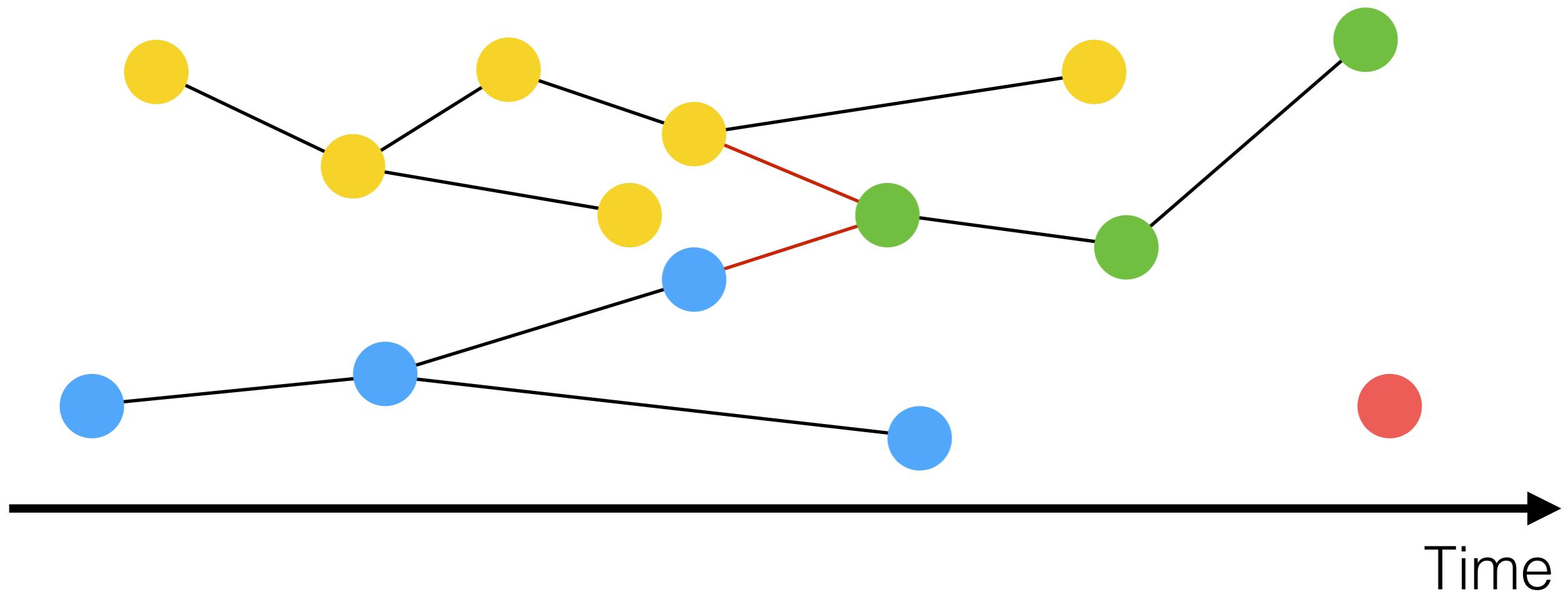
Algorithm for finding reassortant viruses at scale.



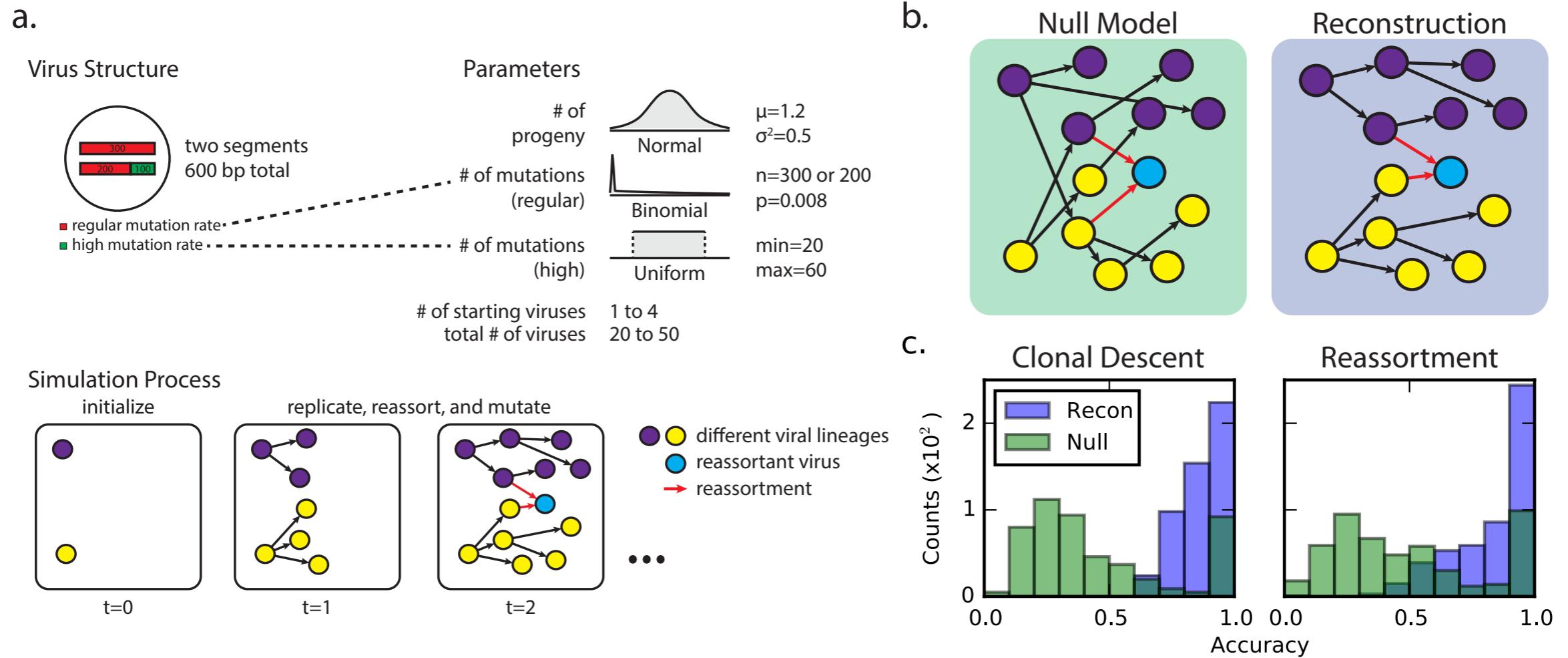
Algorithm for finding reassortant viruses at scale.



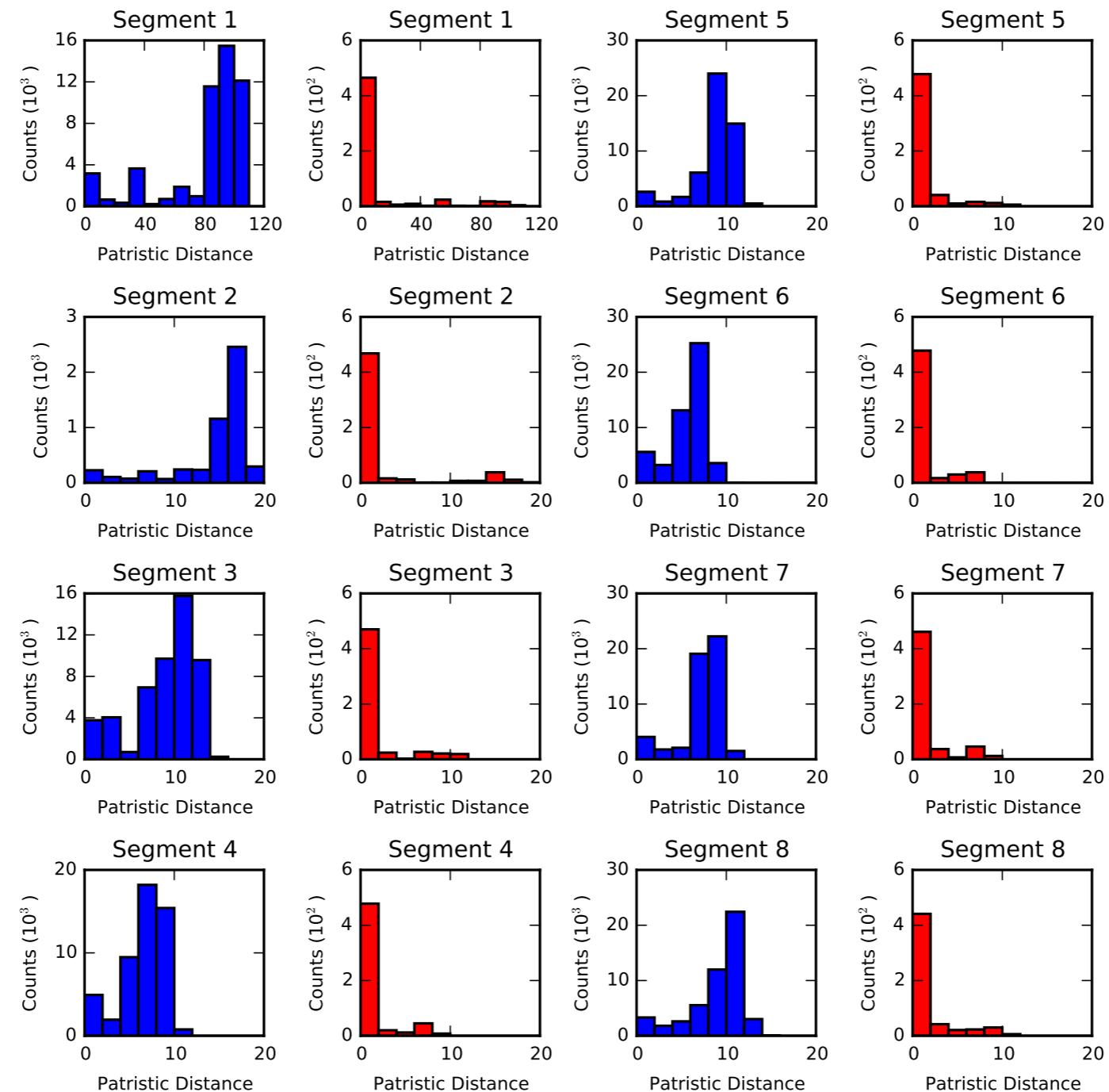
Algorithm for finding reassortant viruses at scale.



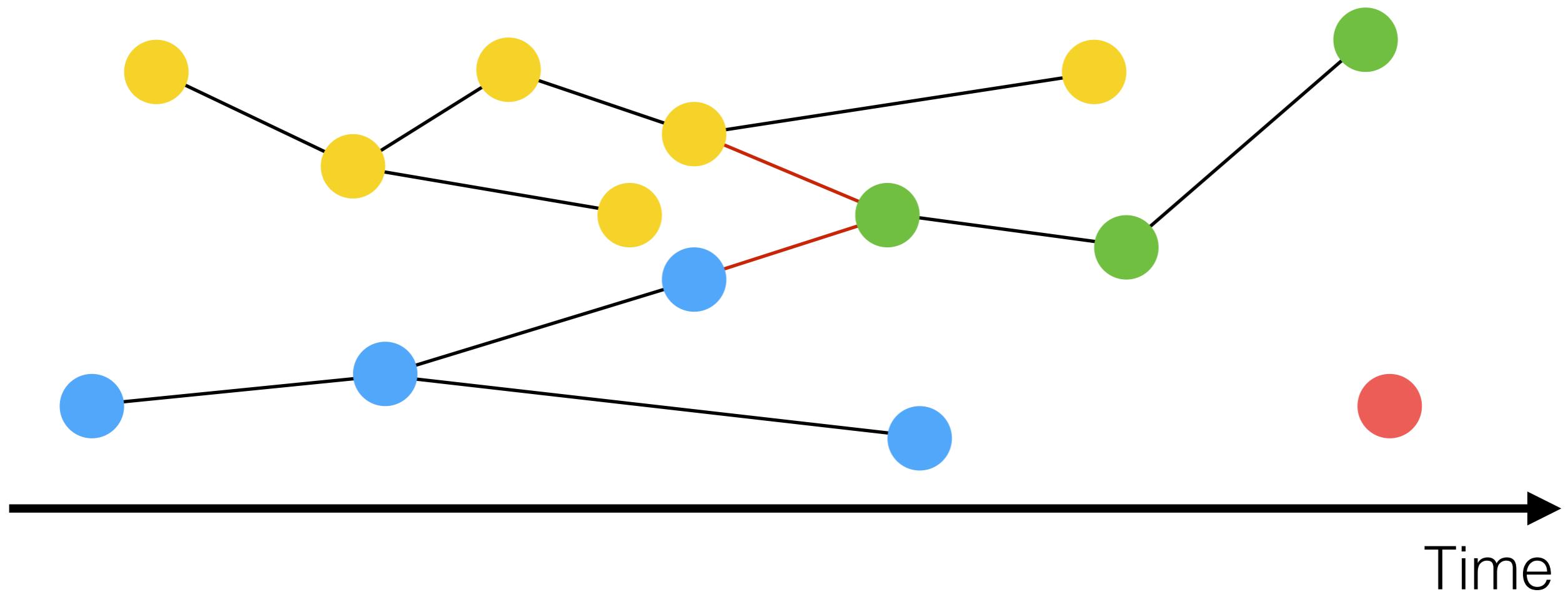
Simulation studies validated algorithm accuracy vs. null.



Algorithm accurately captures phylogenetic relationships.

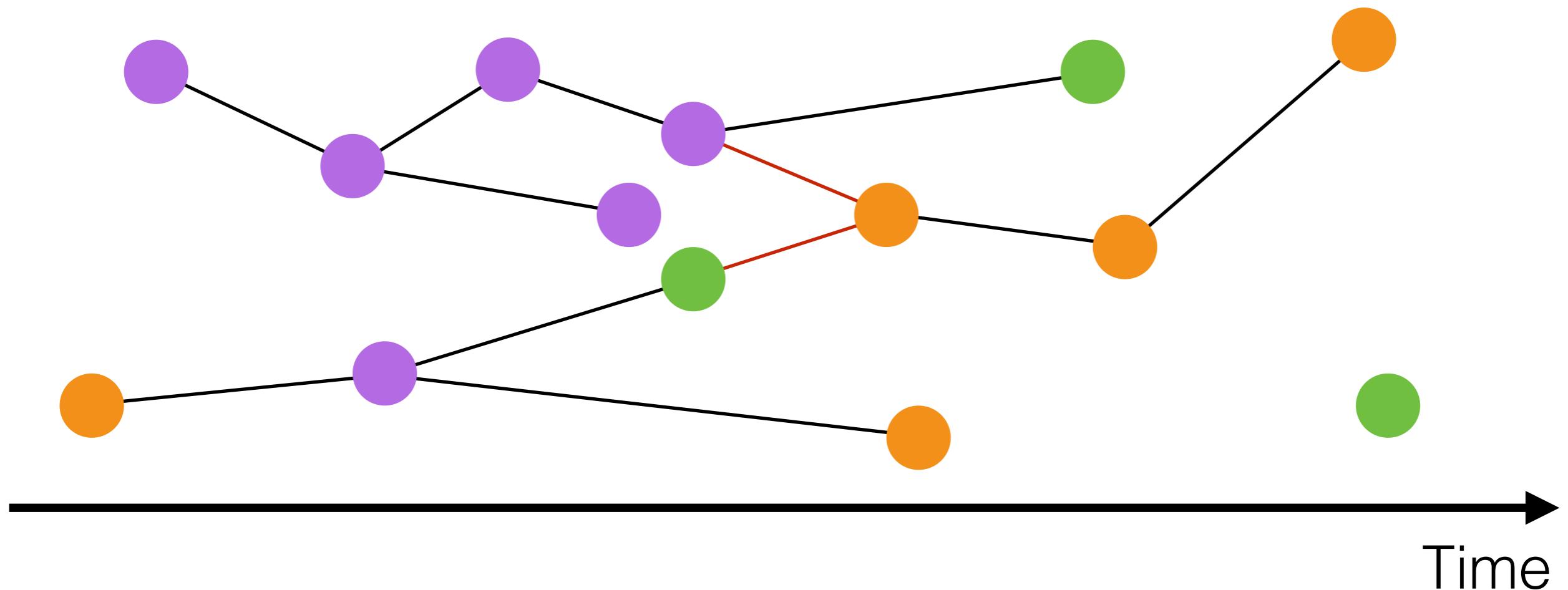


Algorithm for finding reassortant viruses at scale.



Deeper analysis enabled by metadata overlaid on graph.

(coloured by host species)



Application 1: Global reticulate evolution & ecology study.

“is reassortment important for switching hosts?”

Dataset

Dataset

1980-2015 Years of viral isolation.

Dataset

1980-2015 Years of viral isolation.

18,632 Fully sequenced viruses from the IRD

Dataset

1980-2015 Years of viral isolation.

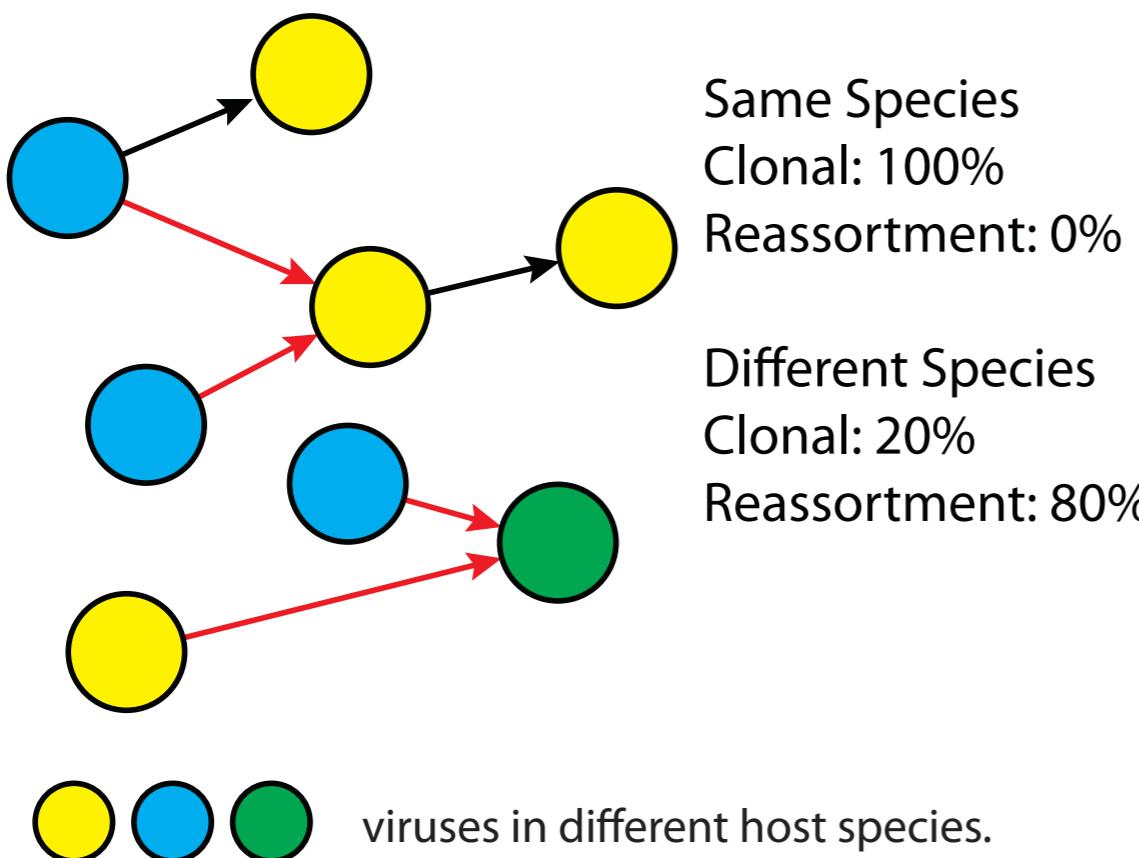
18,632 Fully sequenced viruses from the IRD

2352 Identified reassortant viruses.

Empirical association test of node label pairs with edges.

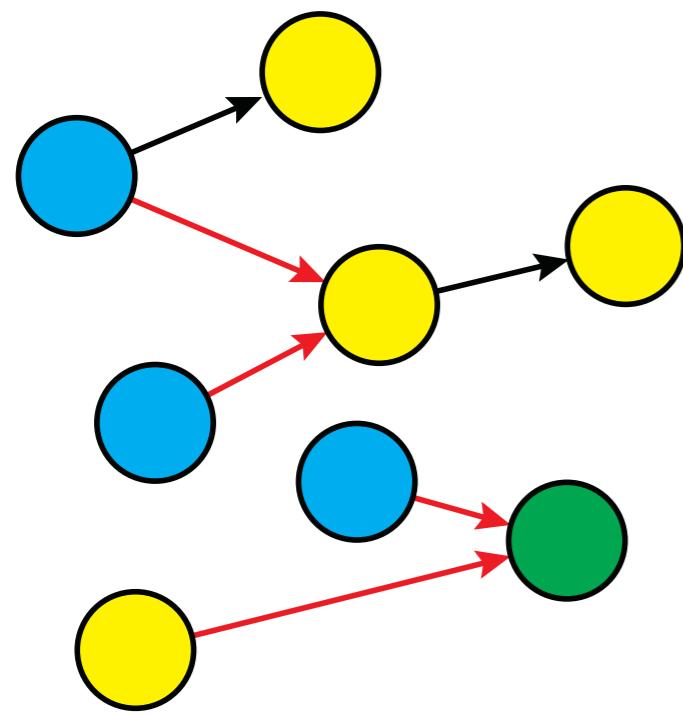
Empirical association test of node label pairs with edges.

Data



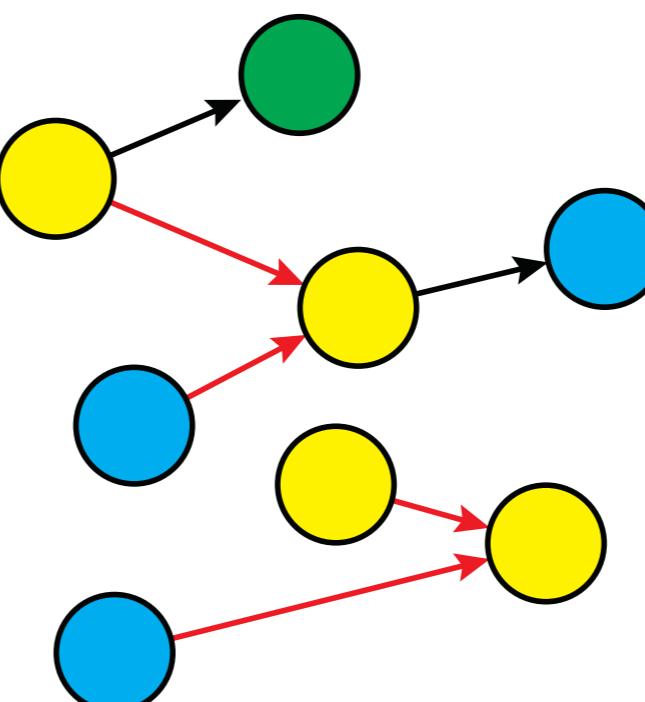
Empirical association test of node label pairs with edges.

Data



viruses in different host species.

Null Distribution



Same Species
Clonal: 100%
Reassortment: 0%

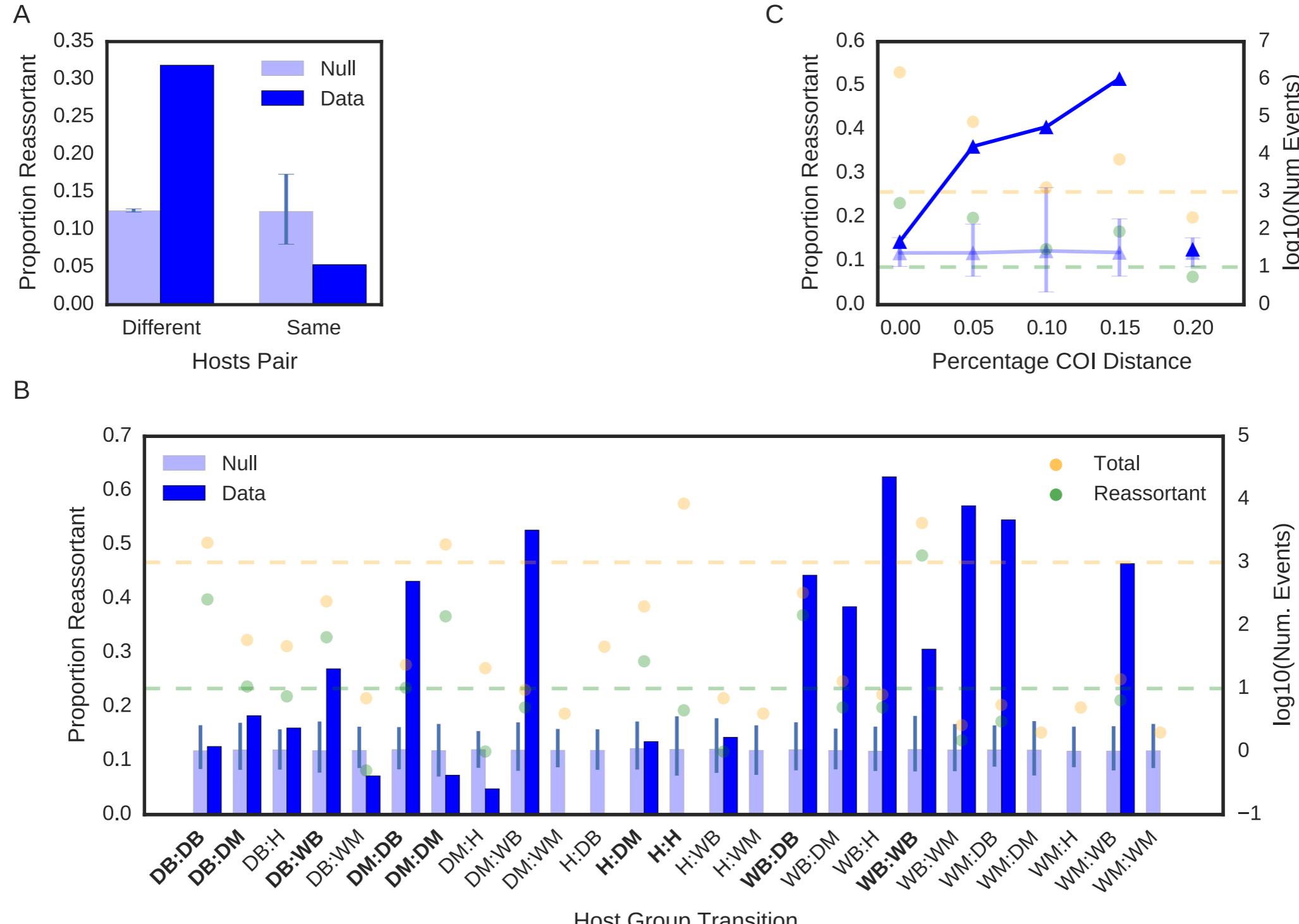
Different Species
Clonal: 20%
Reassortment: 80%

Same Species
Clonal: 0%
Reassortment: 100%

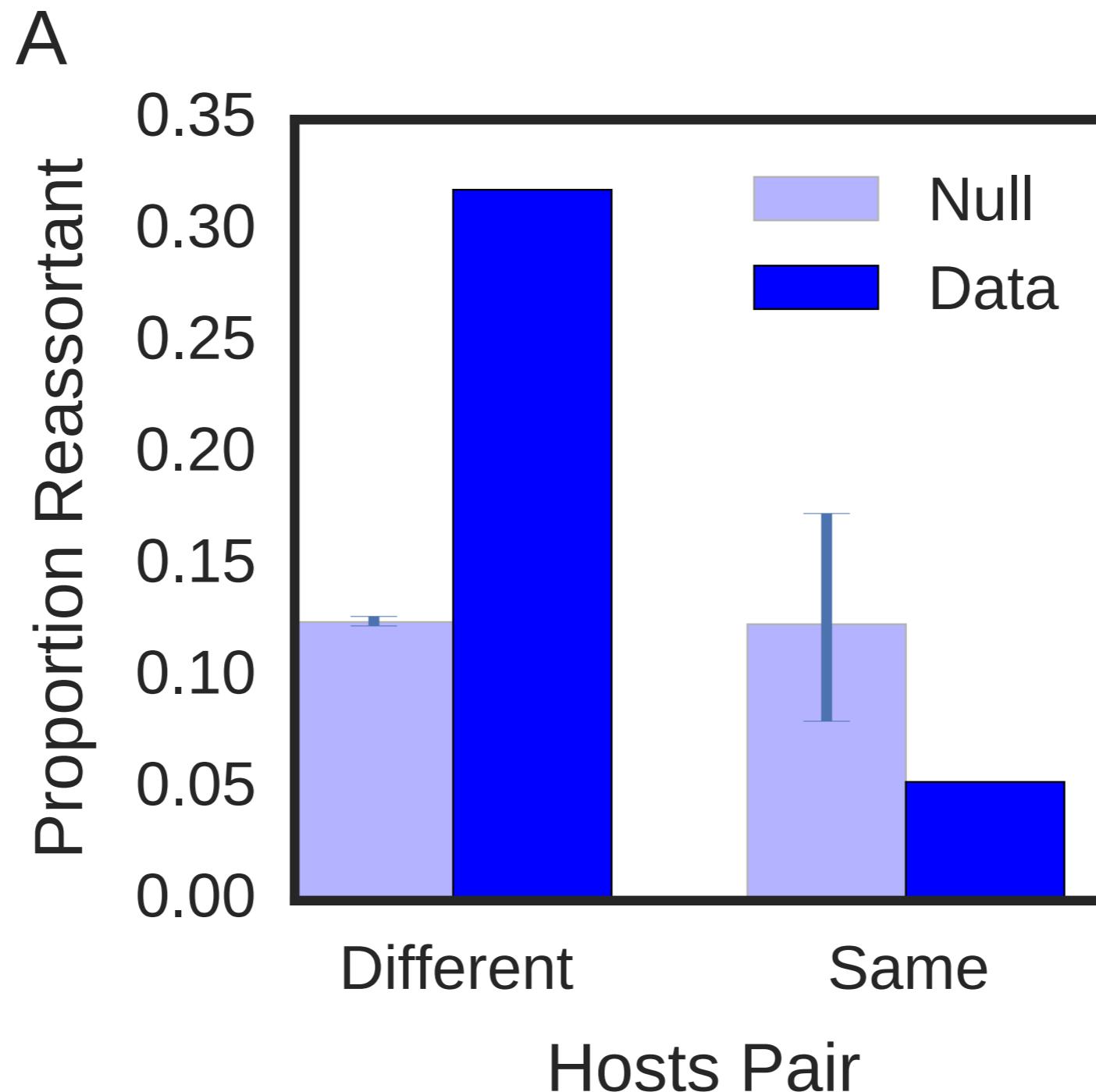
Different Species
Clonal: 50%
Reassortment: 50%

→ clonal descent → reassortment descent

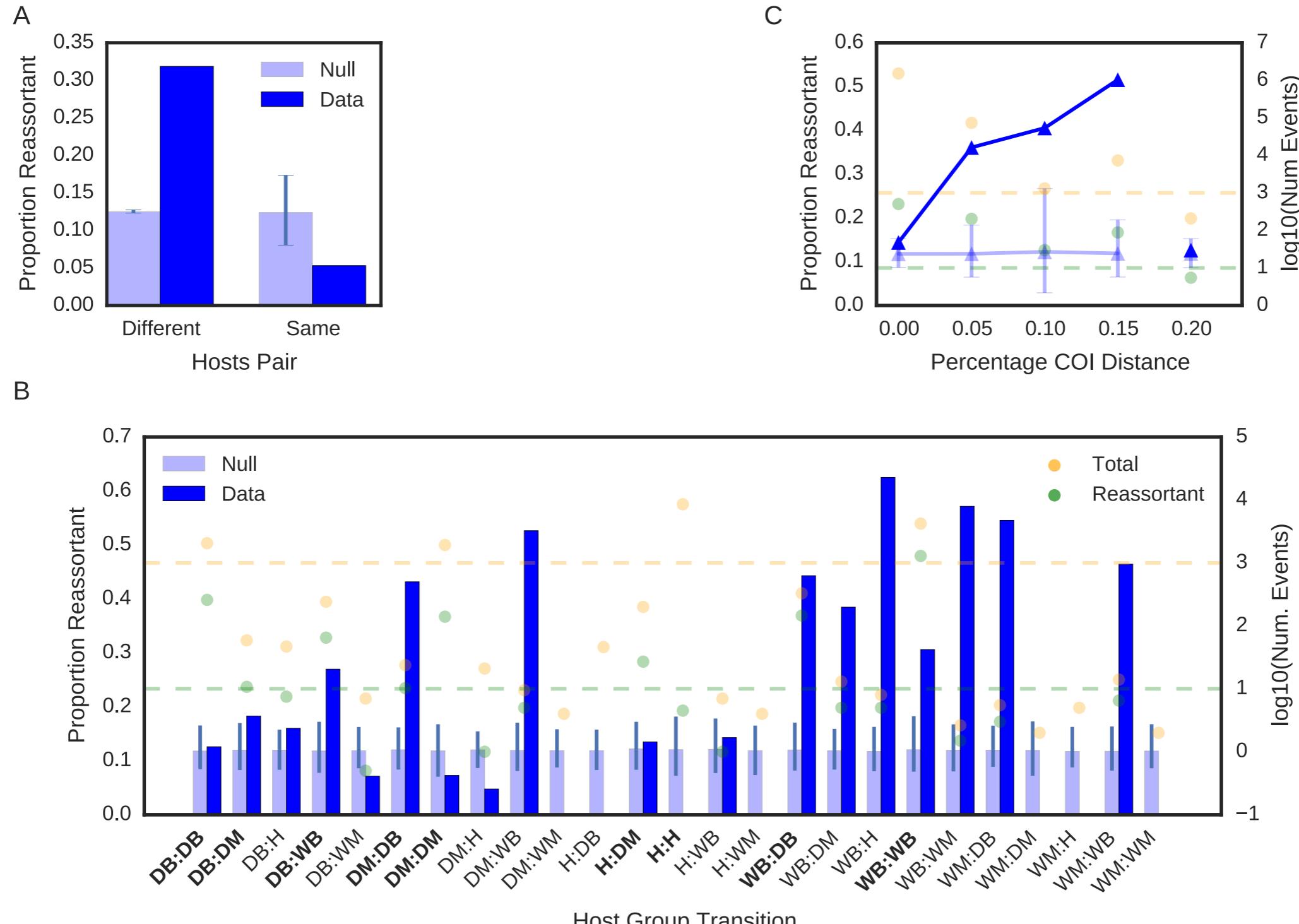
Reassortment is over-represented when crossing between different host species.



Reassortment is over-represented when crossing between different host species.

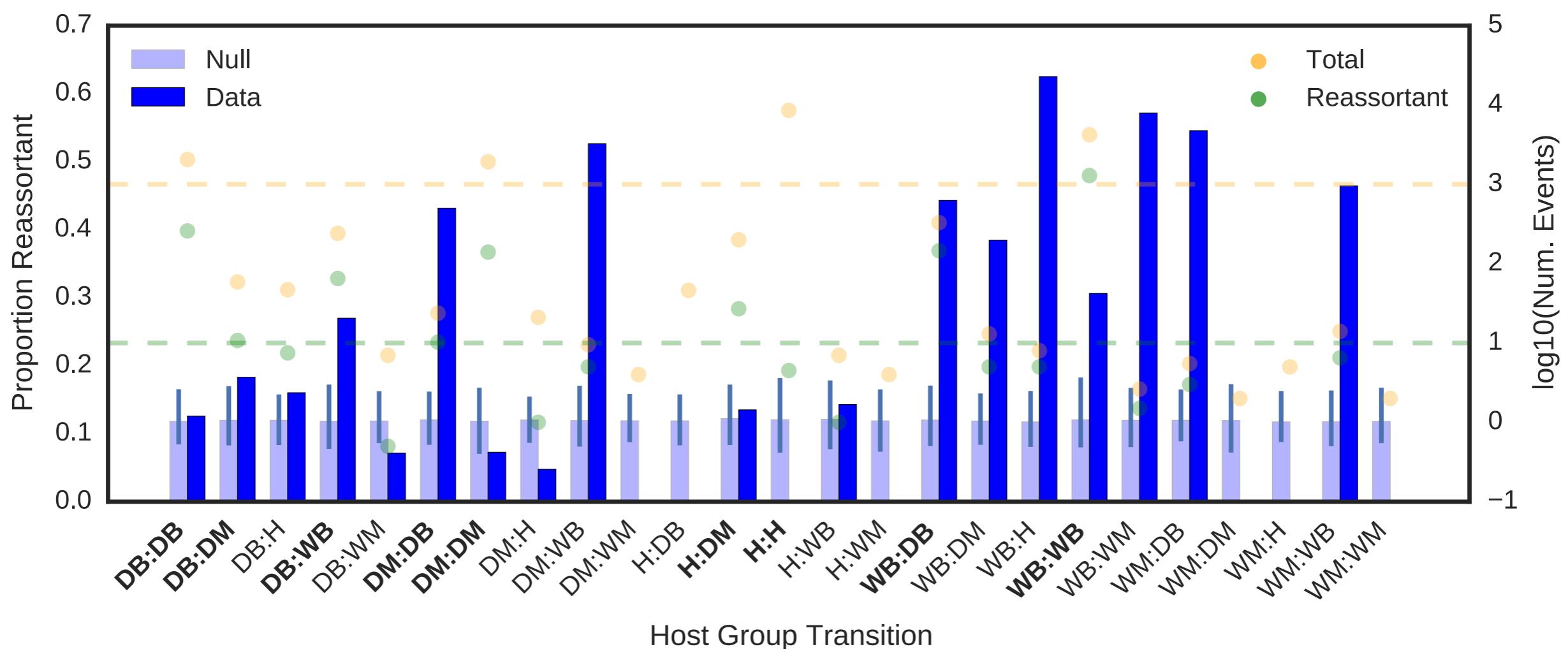


Reassortment is over-represented when crossing between different “ecotypes”.



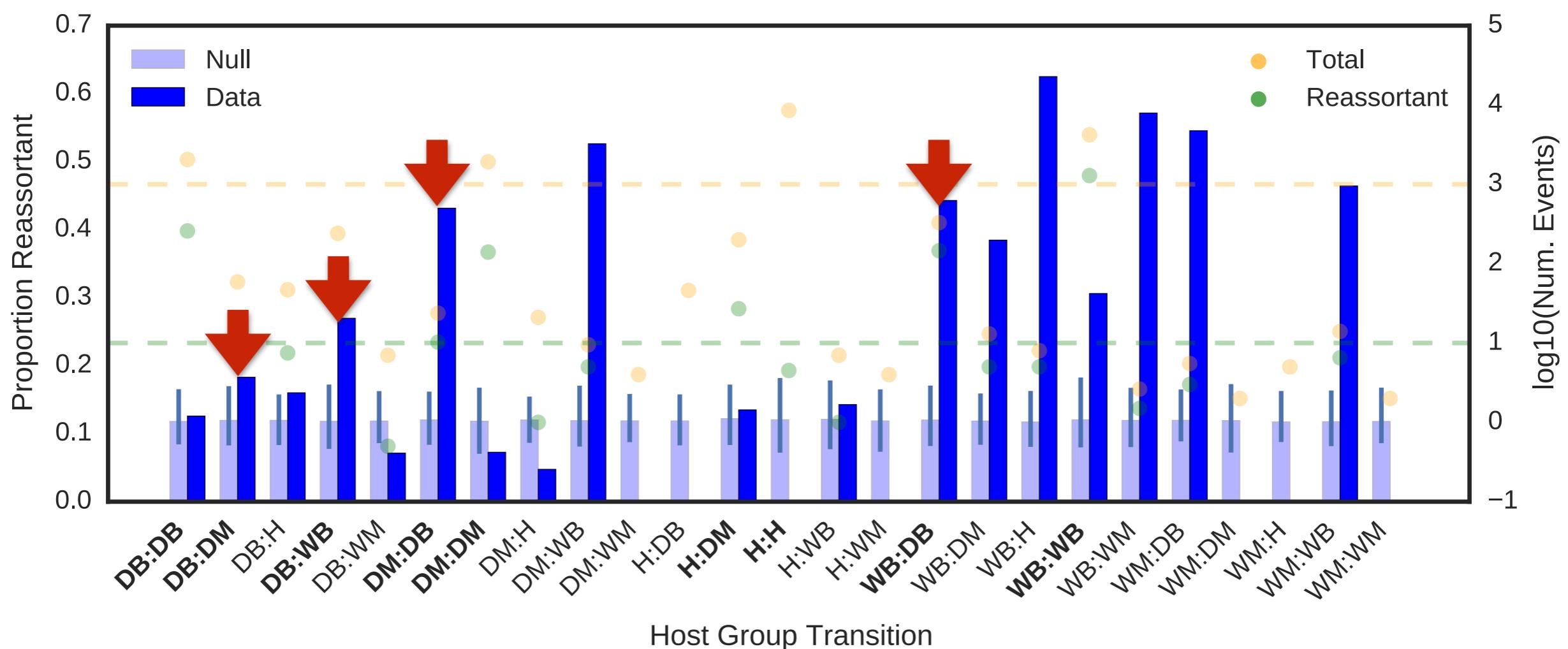
Reassortment is over-represented when crossing between different “ecotypes”.

B



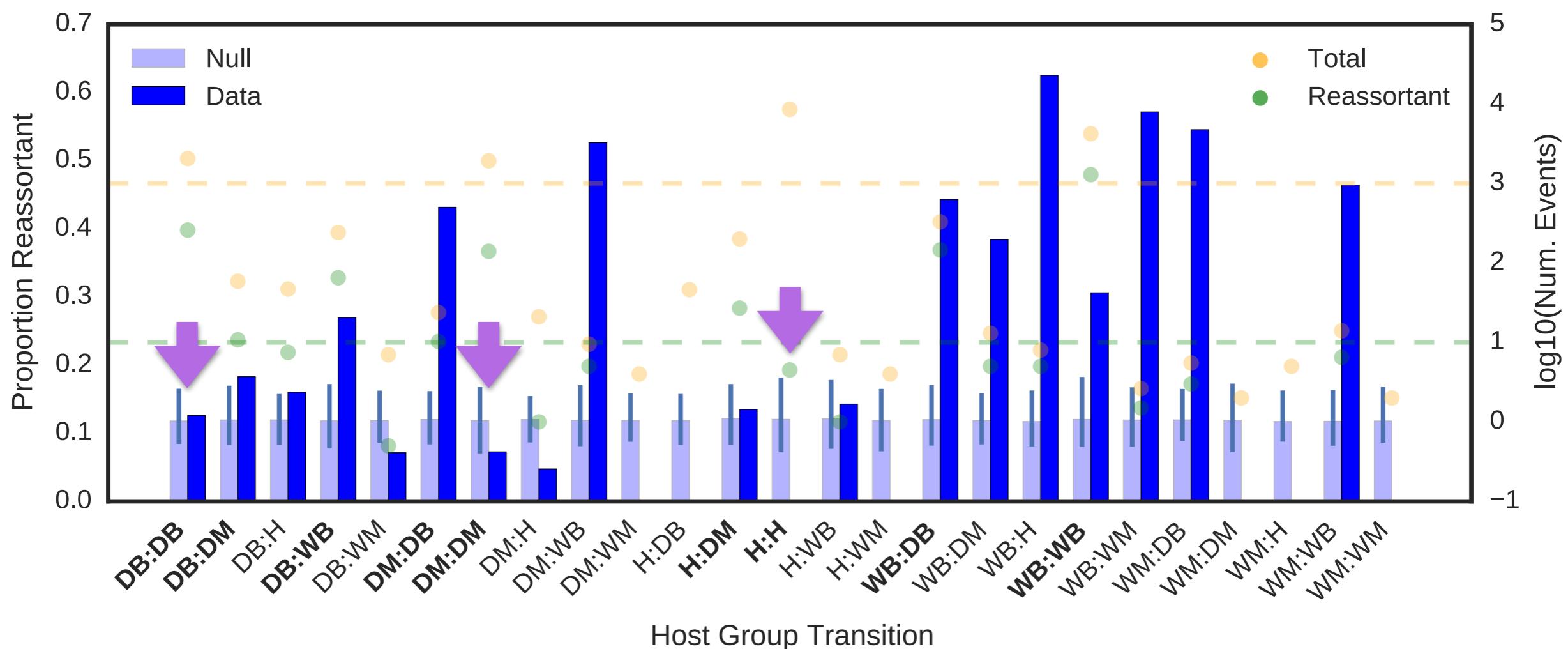
Reassortment is over-represented when crossing between different “ecotypes”.

B



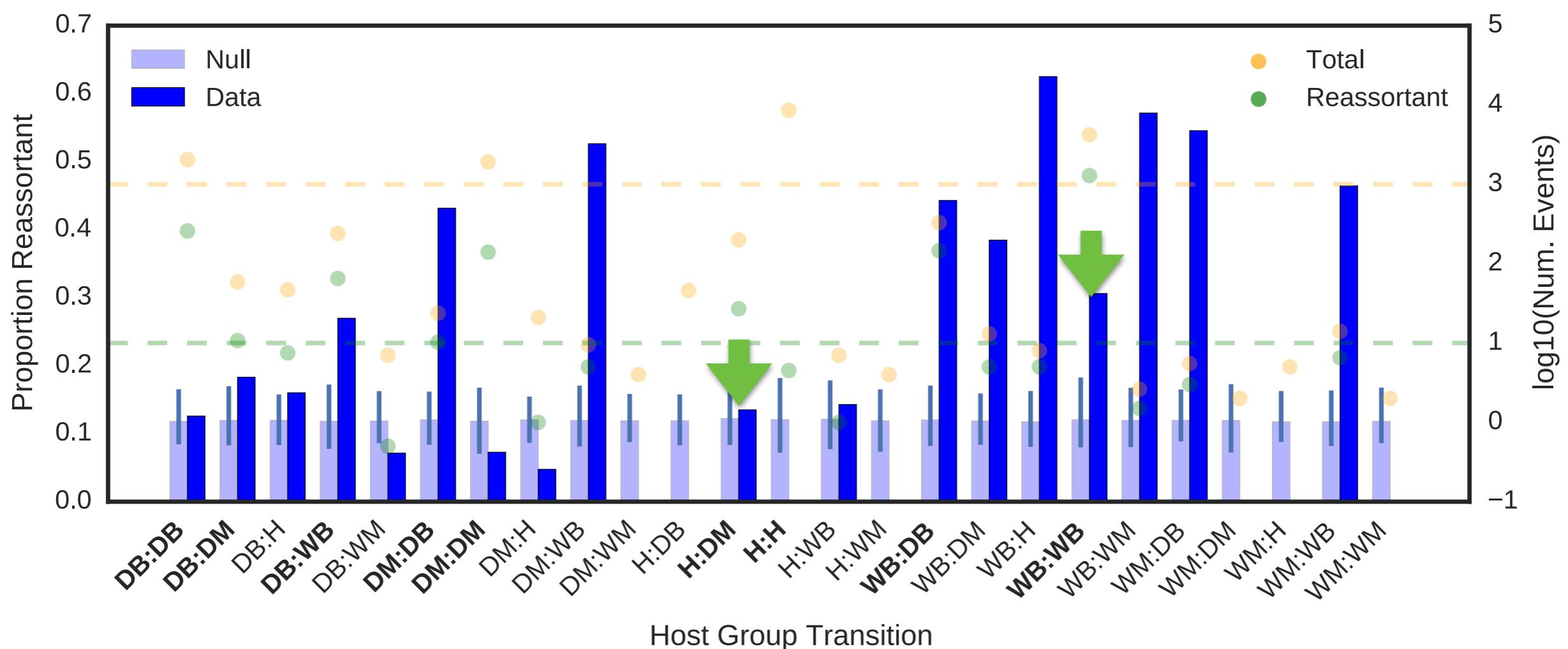
Reassortment is over-represented when crossing between different “ecotypes”.

B



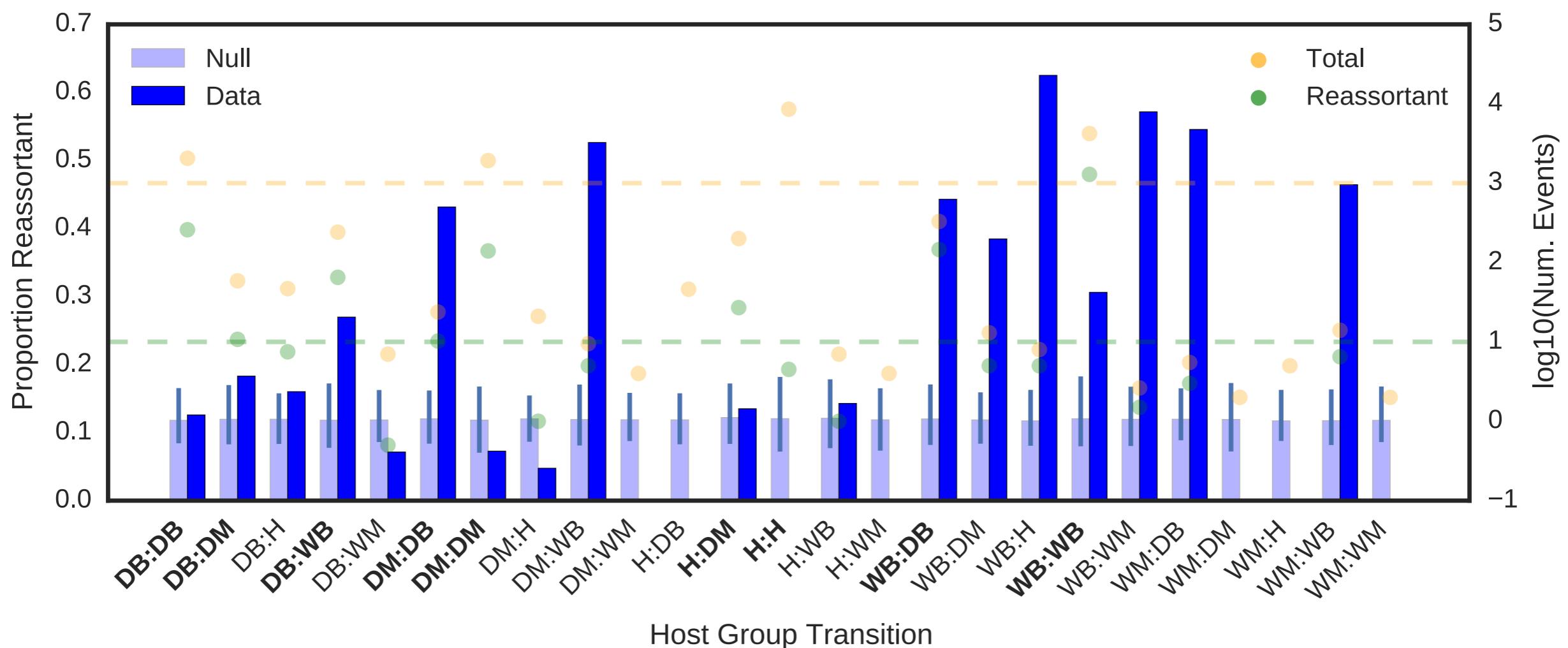
Reassortment is over-represented when crossing between different “ecotypes”.

B

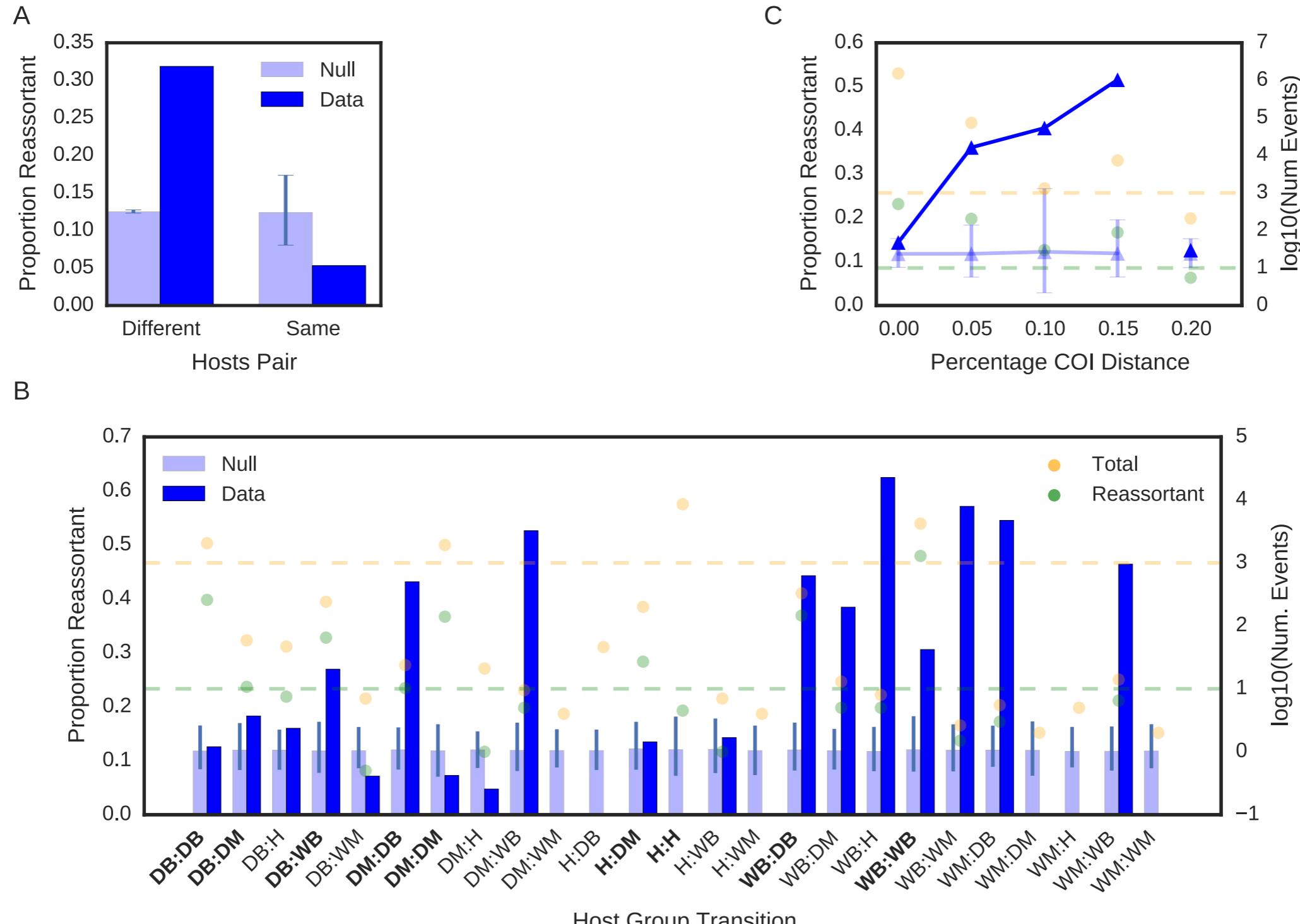


Reassortment is over-represented when crossing between different “ecotypes”.

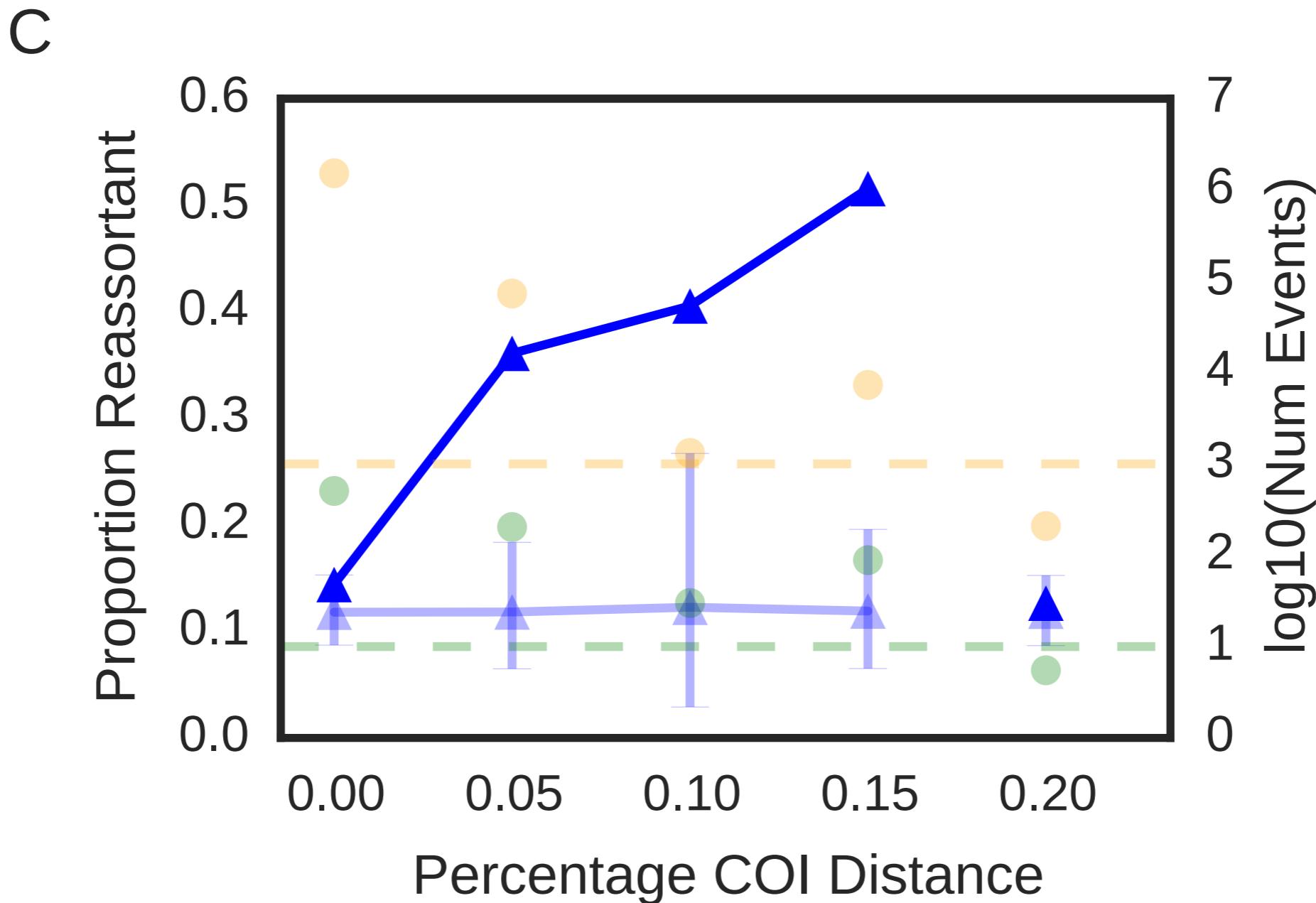
B



As host evolutionary distance increases, reassortment importance increases.



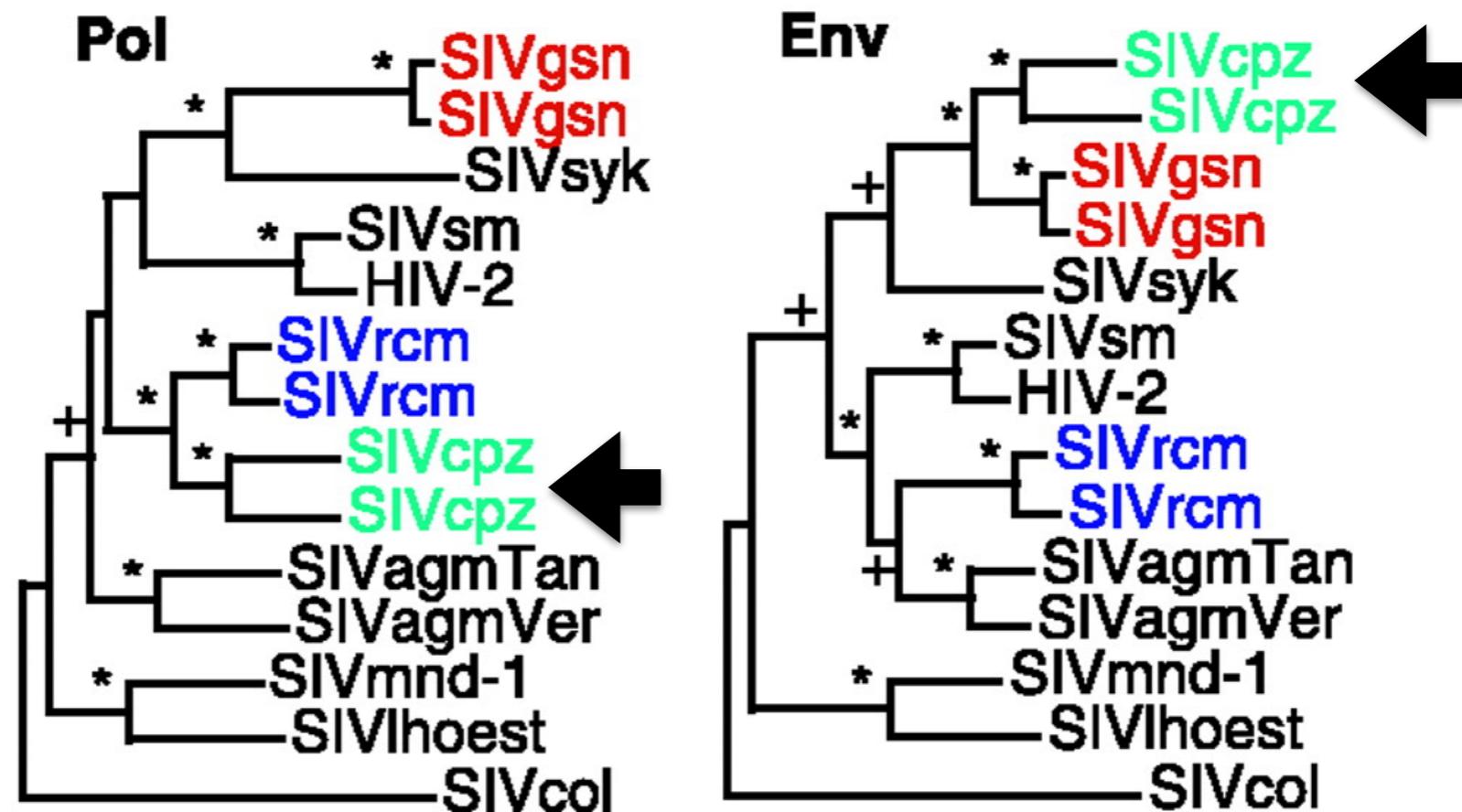
As host evolutionary distance increases,
reassortment importance increases.



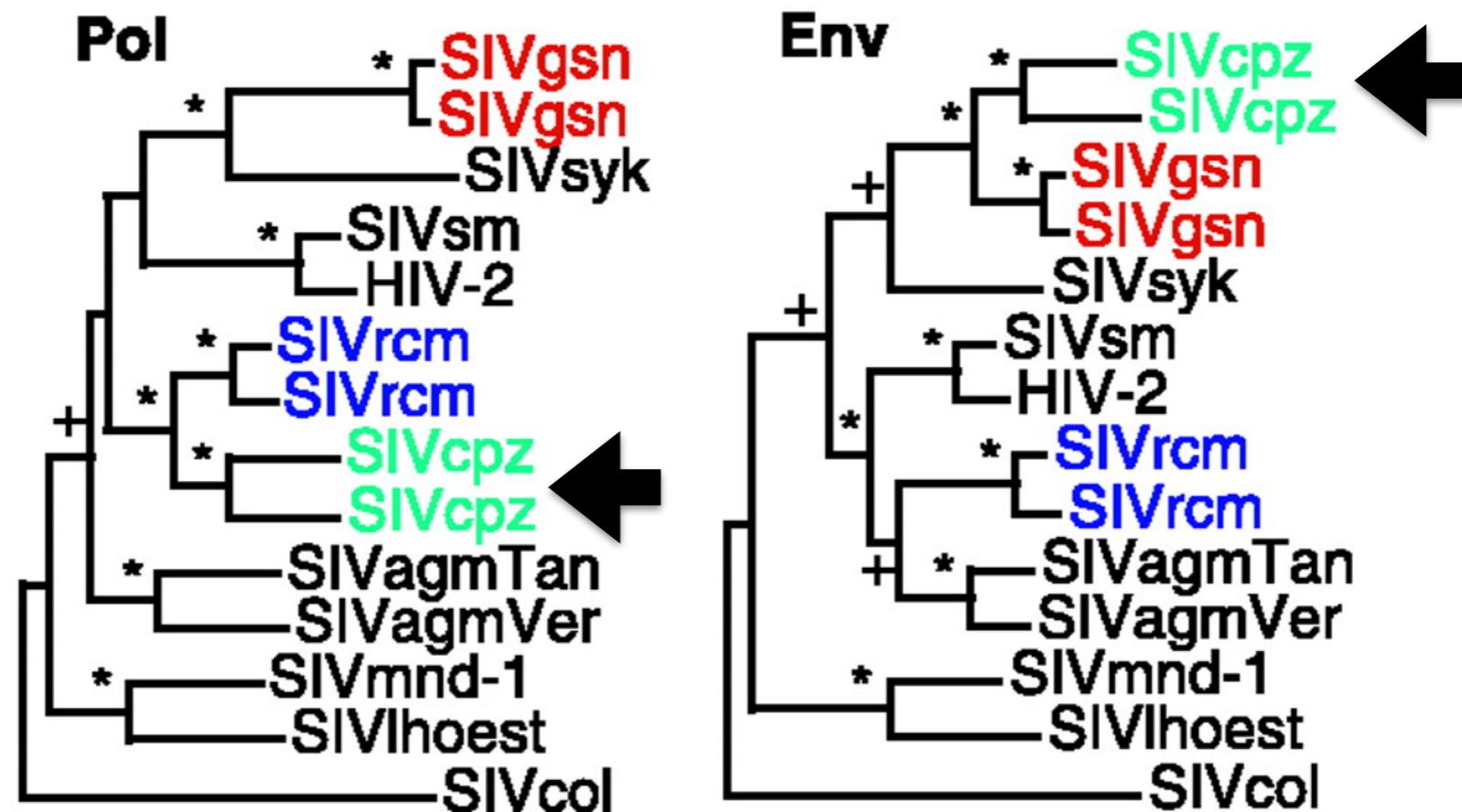
Limitations

- Reassortment before or after host switching?
- Sampling biases?

Parallel: HIV also had a “reticulate” history prior to spillover into humans.



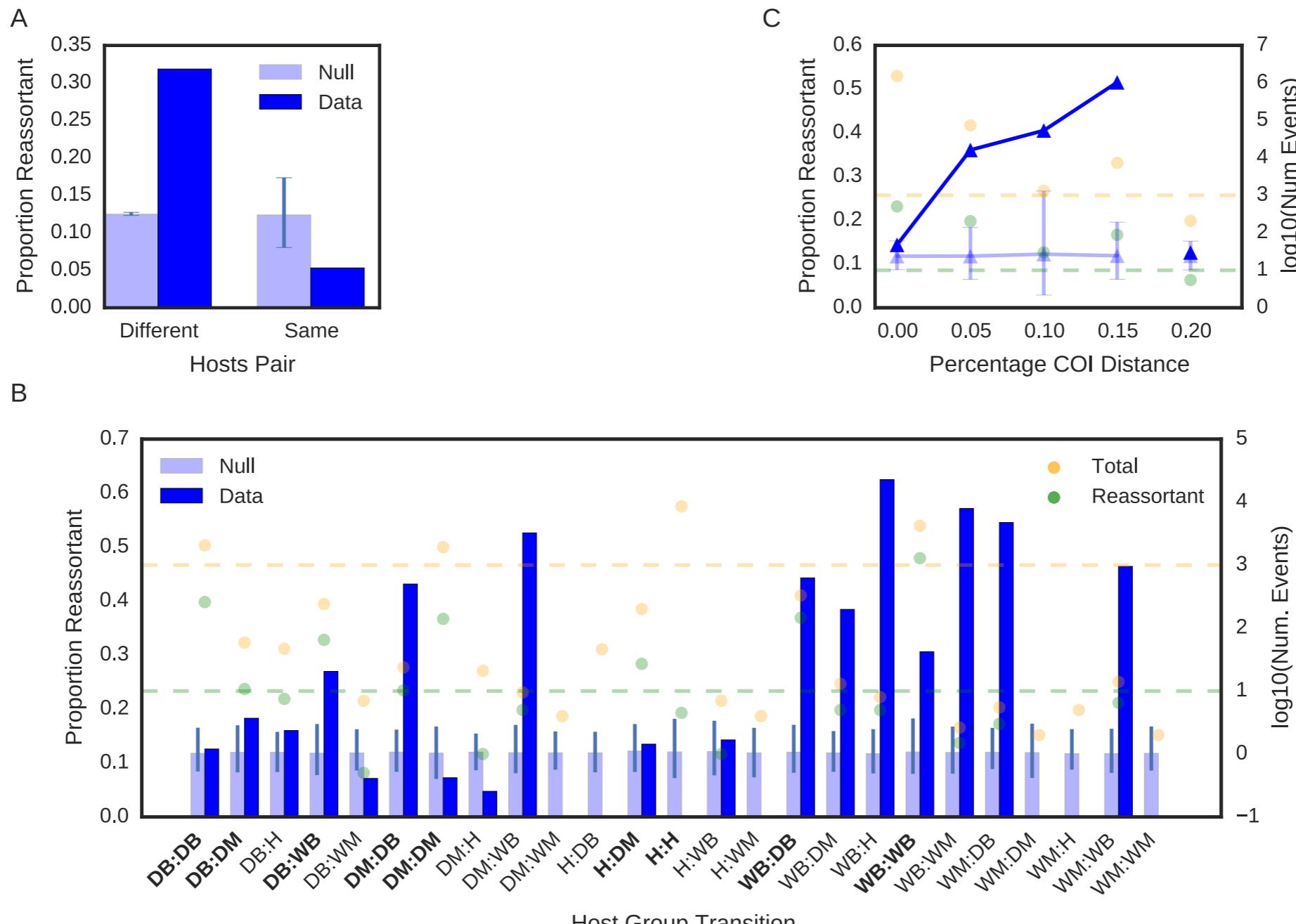
Parallel: HIV also had a “reticulate” history prior to spillover into humans.



Reticulate evolution is favored in influenza niche switching

Eric J. Ma^{a,1}, Nichola J. Hill^a, Justin Zabilansky^a, Kyle Yuan^a, and Jonathan A. Runstadler^{a,b,1}

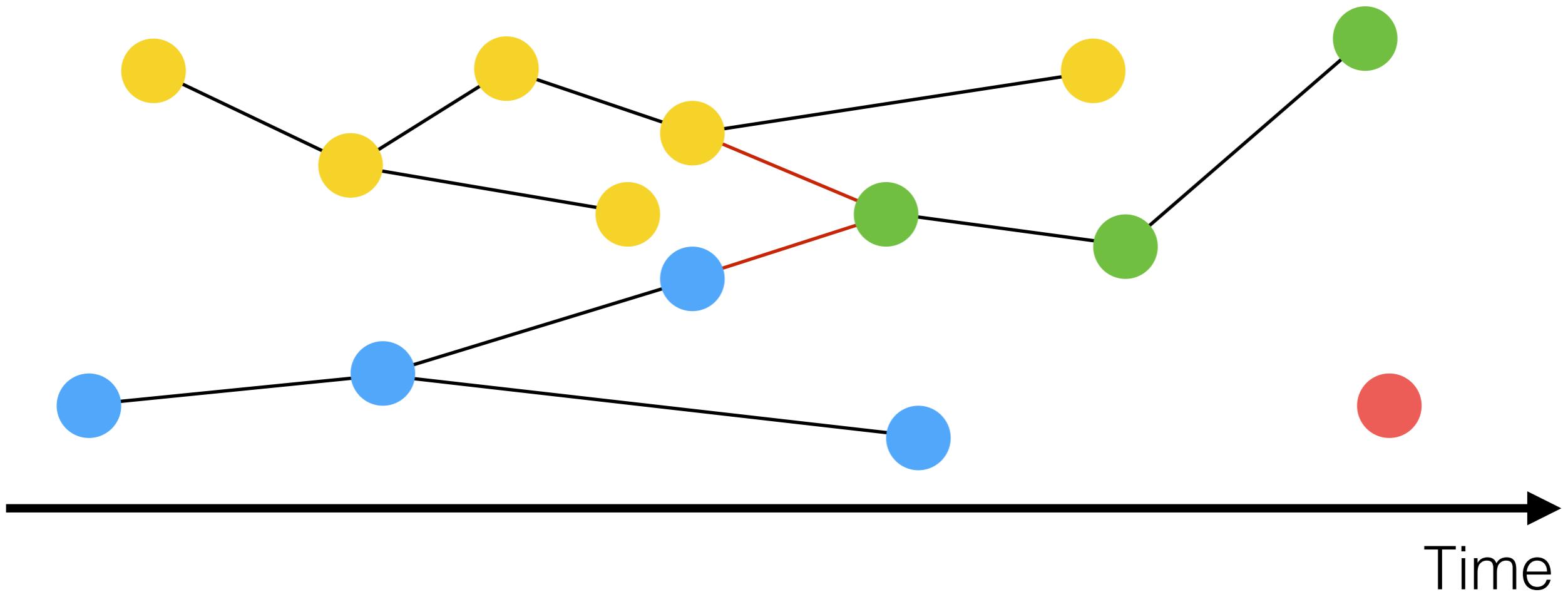
^aDepartment of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; and ^bDivision of Comparative Medicine, Massachusetts Institute of Technology, Cambridge, MA 02139



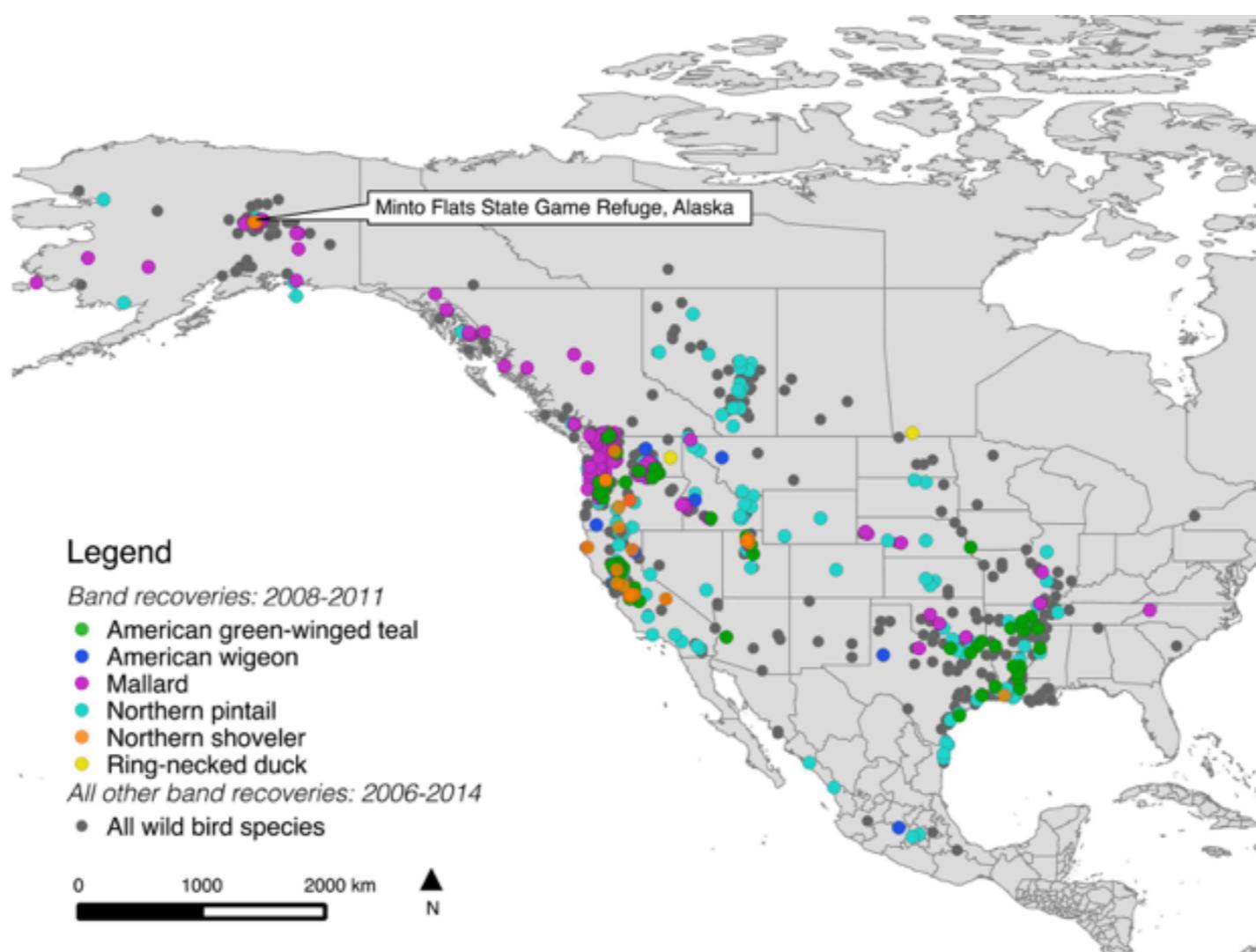
Application 2: Viral transmission and host migration dynamics.

“do viral transmission strategies follow wild bird host seasonality?”

Clonal transmissions and reassortment are complementary transmission modes.



When do clonal transmission and reassortment dominate?
Does it follow the breeding-feeding cycle?



Dataset

Dataset

2009-2012 Breeding ground sampling years.

Dataset

- 2009-2012** Breeding ground sampling years.
- 14,004** Sampled birds.

Dataset

2009-2012 Breeding ground sampling years.

14,004 Sampled birds.

Minto Flats, AK Location of sampling

Dataset

2009-2012 Breeding ground sampling years.

14,004 Sampled birds.

Minto Flats, AK Location of sampling

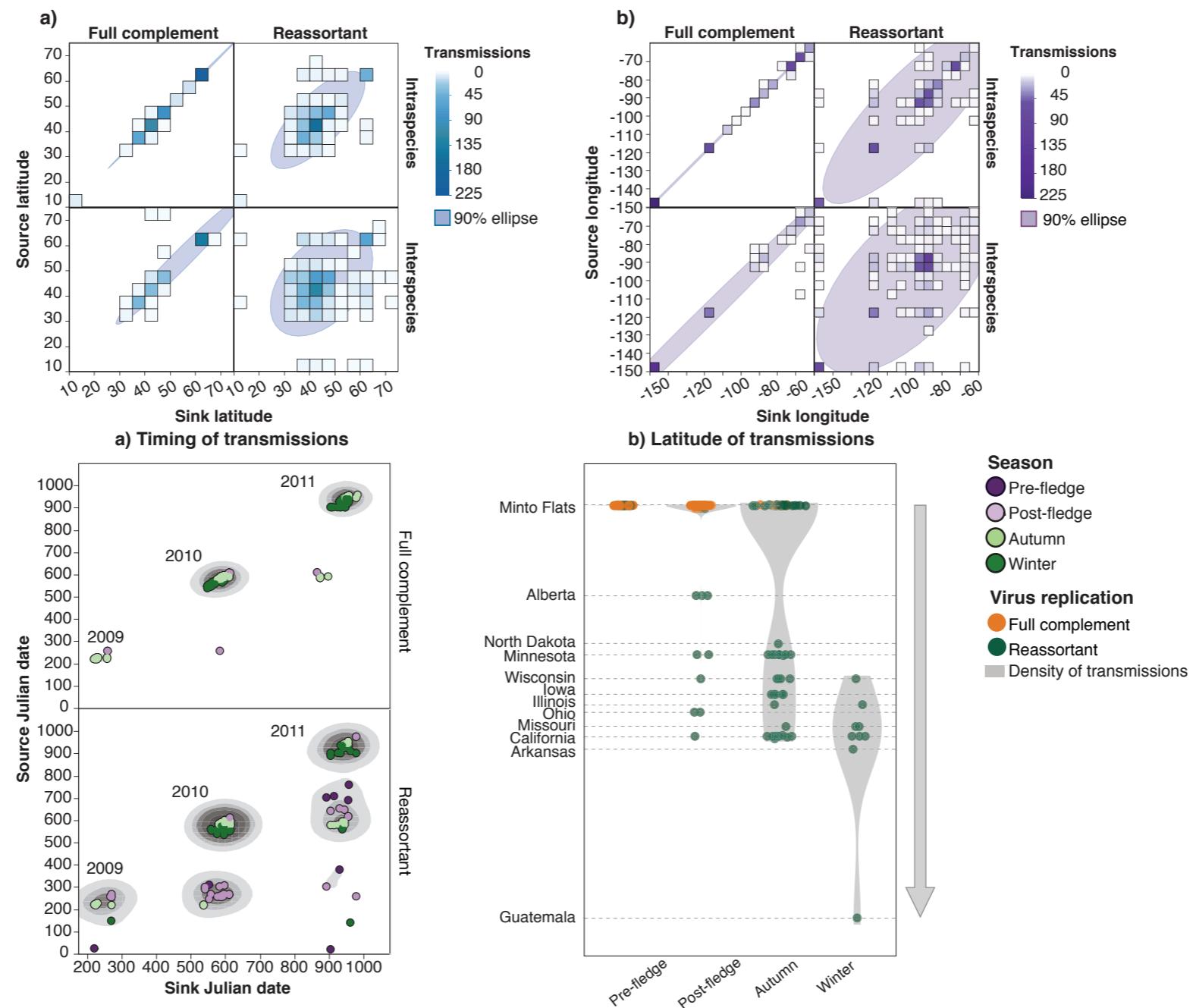
1,787 Fully sequenced viruses.

Sample collection is non-trivial.

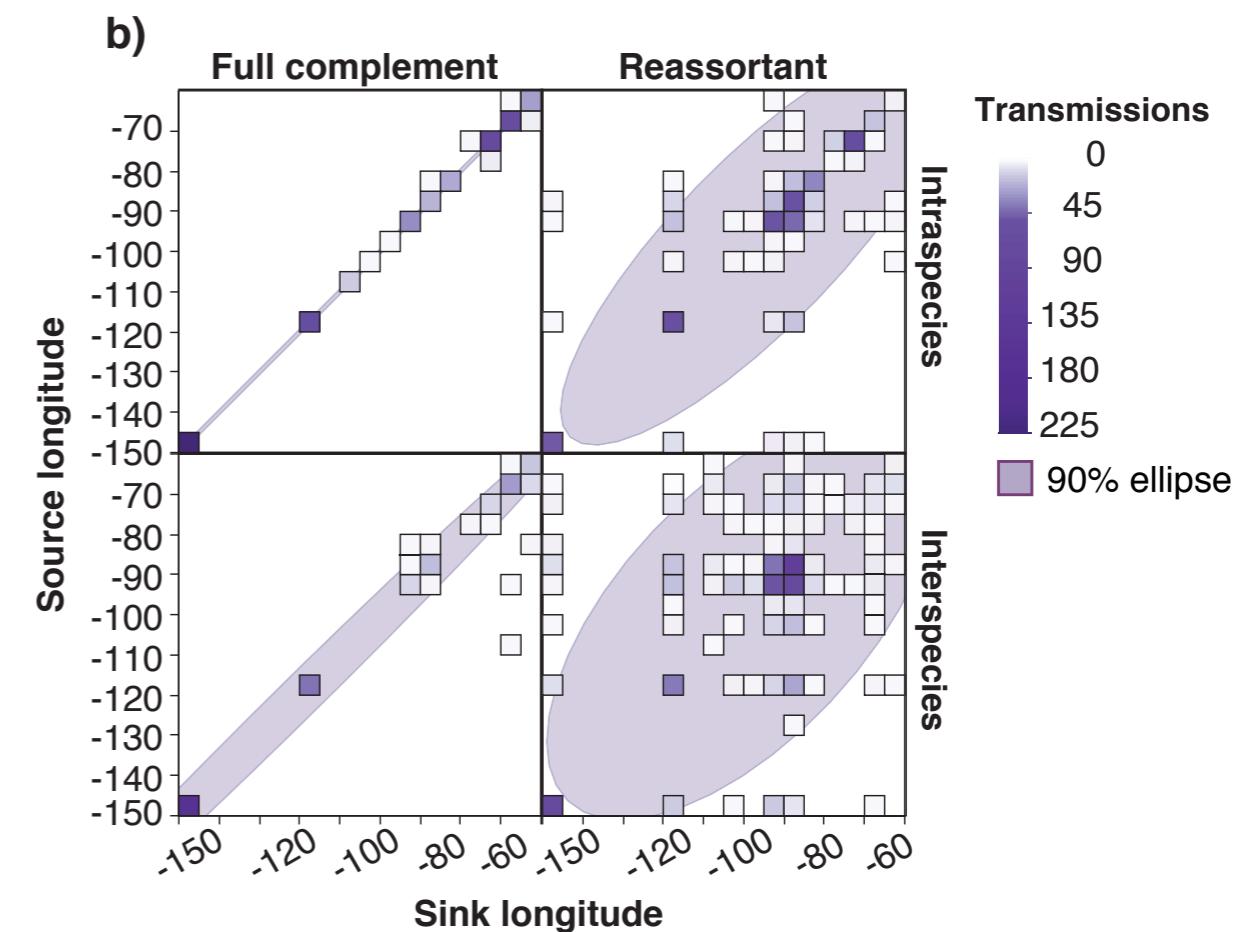
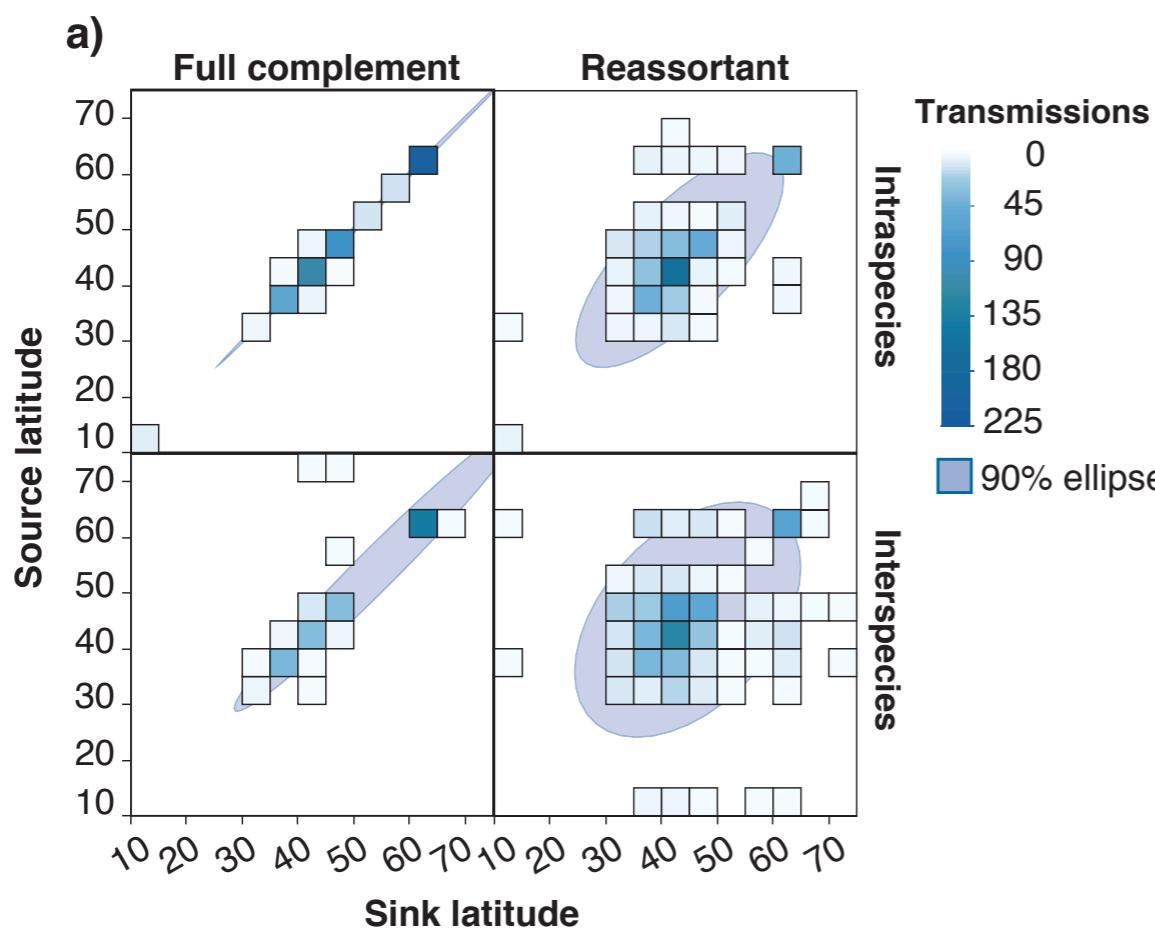


Original photos: Nichola Hill.

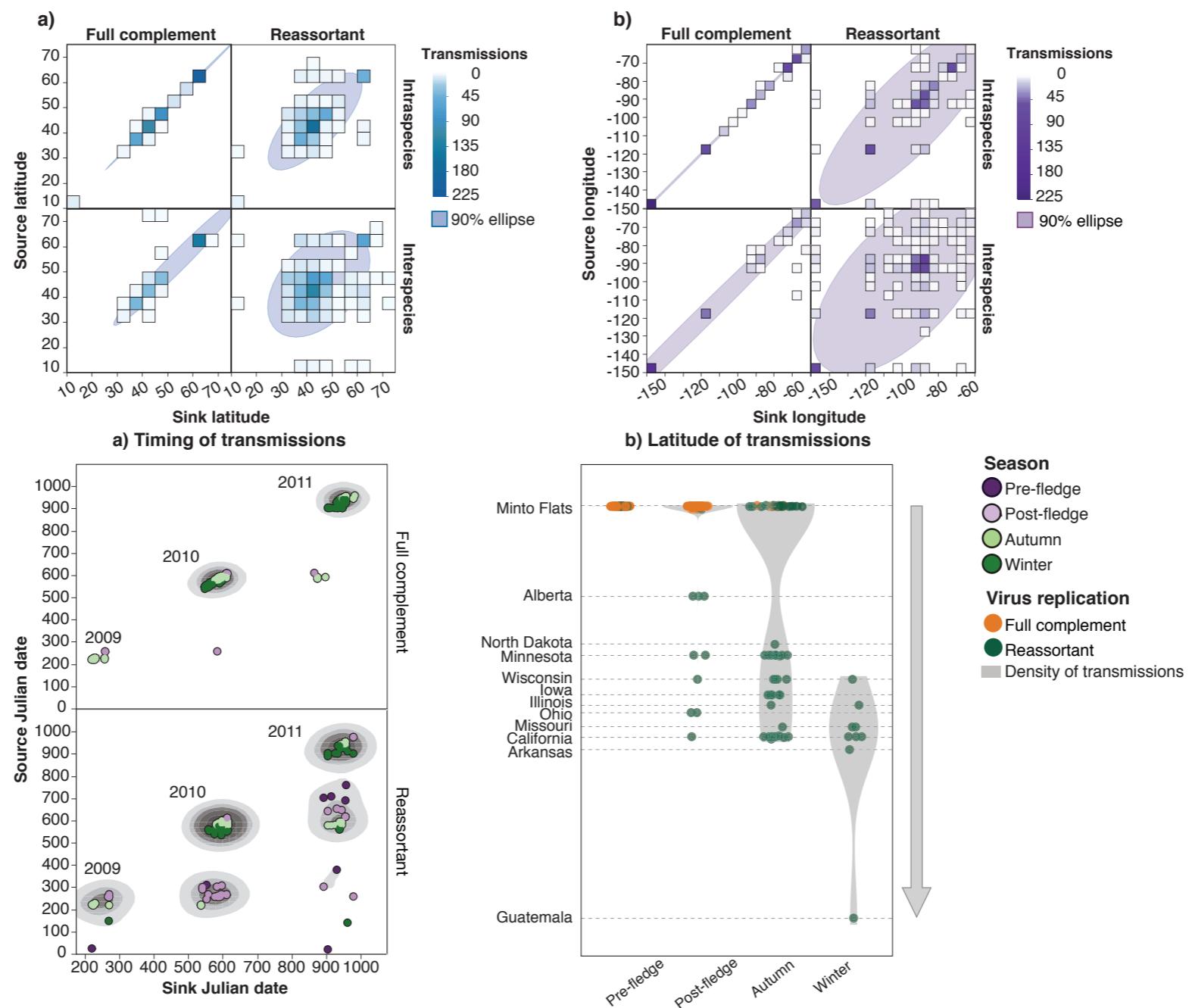
Clonal transmission events are **spatially** local, reassortment events are dispersed.



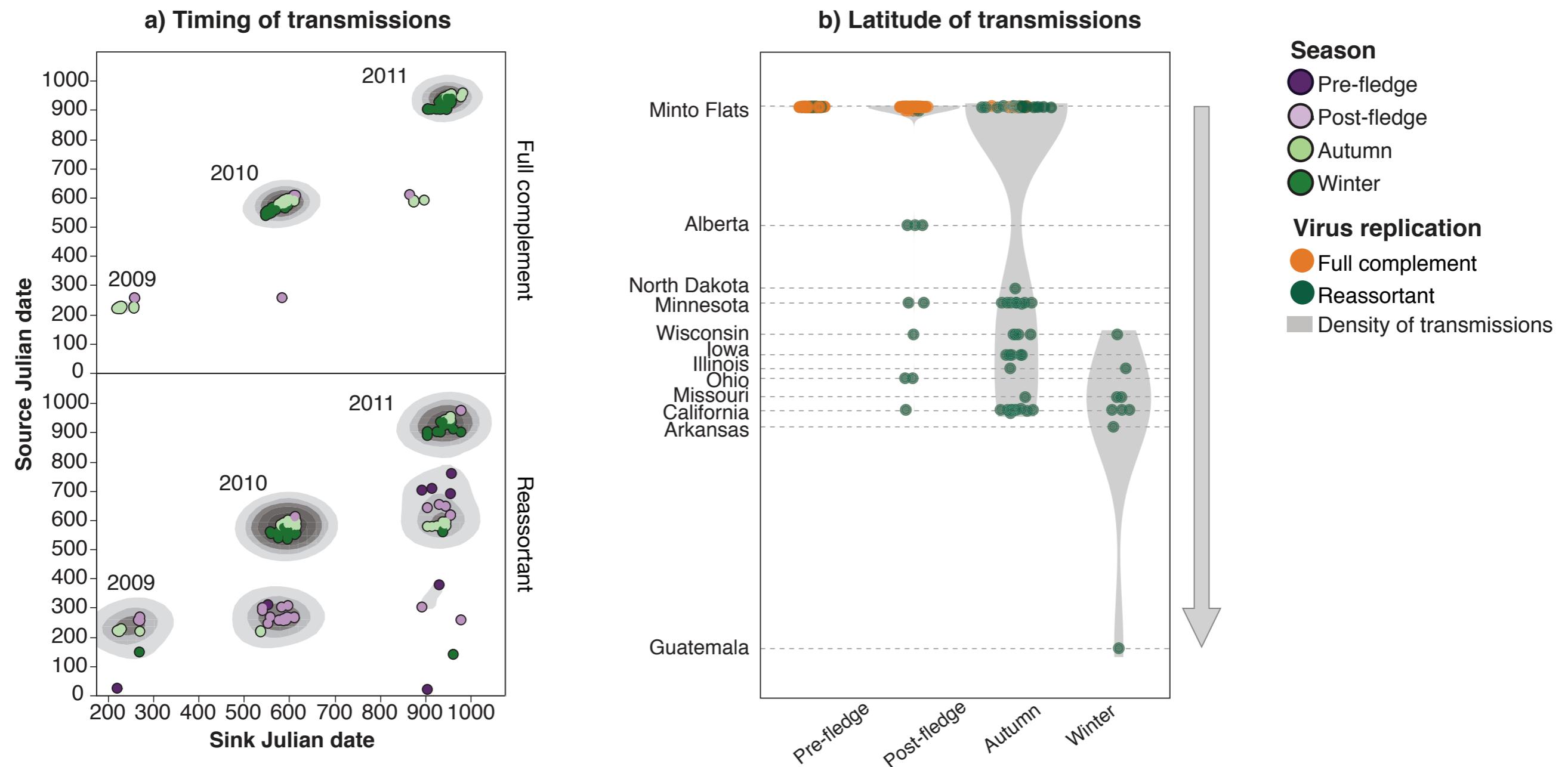
Clonal transmission events are **spatially** local, reassortment events are dispersed.



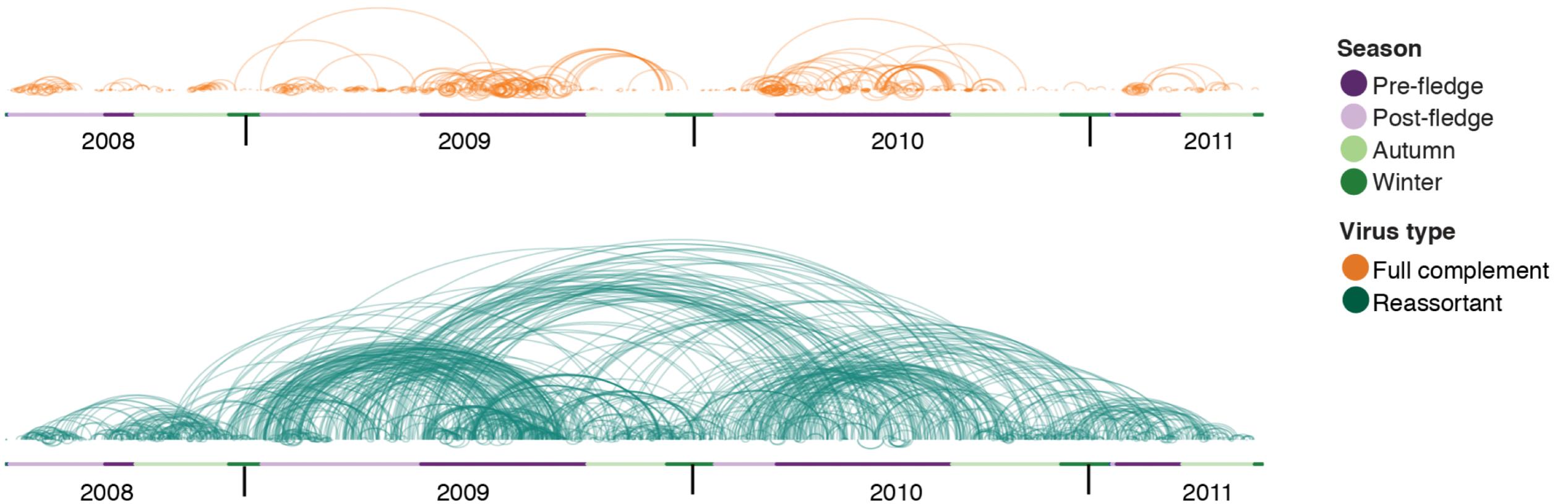
Clonal transmission events are temporally local, reassortment events are dispersed.



Clonal transmission events are **temporally** local, reassortment events are dispersed.



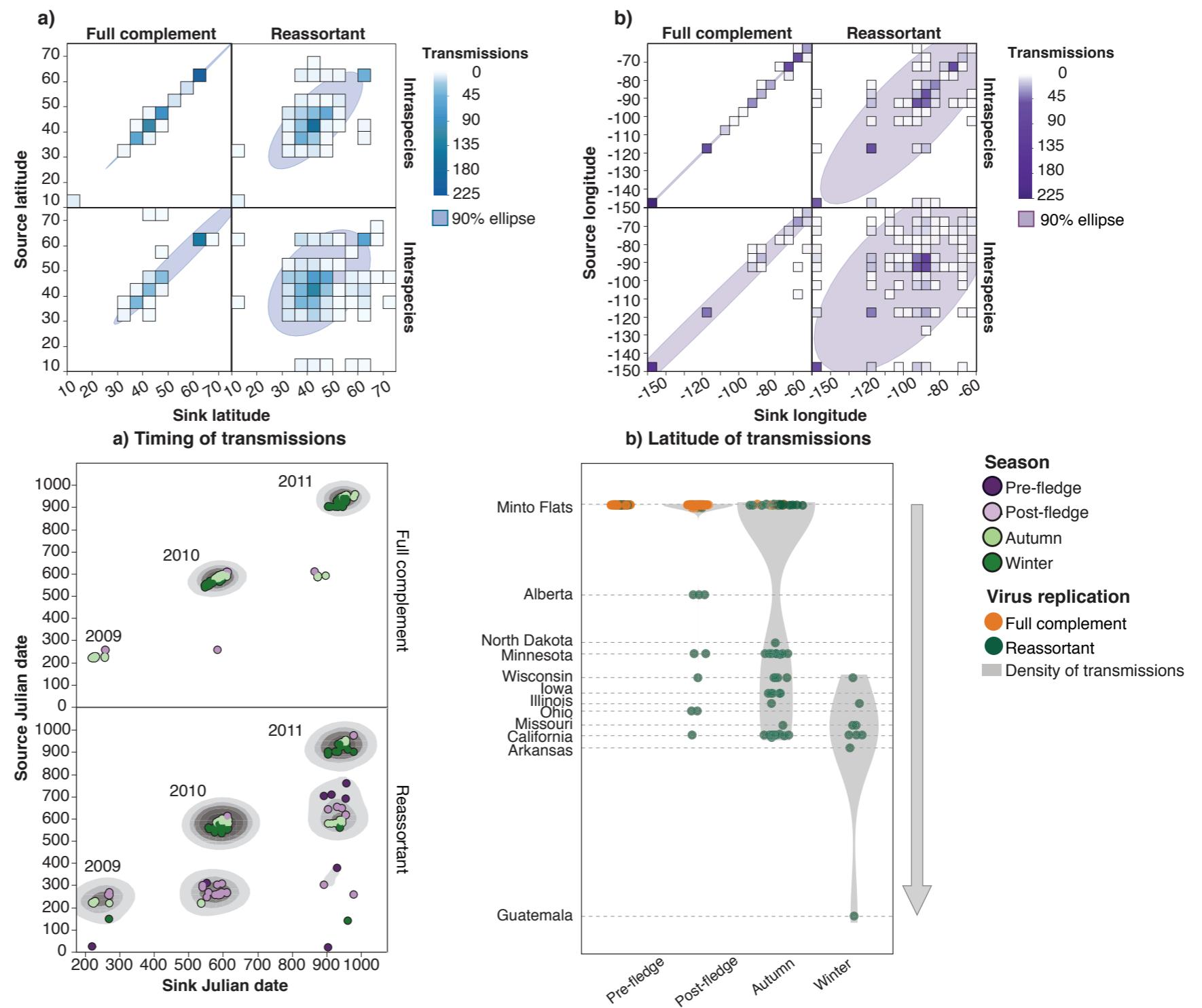
Clonal transmission events are **temporally** local, reassortment events are dispersed.



LETTER

Transmission of influenza reflects seasonality of wild birds across the annual cycle

Nichola J. Hill,¹ Eric J. Ma,¹
 Brandt W. Meixell,^{2,3} Mark S.
 Lindberg,⁴ Walter M. Boyce⁵ and
 Jonathan A. Runstadler^{1*}



Open Question Highlights

- Deeper quantification of ecological niche differences.
- Forecasting emergence of new viral subtypes.
- Phasing of “mixed viral samples”.

Philosophical thoughts on software development.

- Algorithms = Ways of thinking (about problems)
- Good software gets used.

With thanks



Collaborators & Colleagues

Nichola J. Hill (MIT)
Islam T. M. Hussein (MIT)
Christopher Bandoro (MIT)
Wendy Puryear (MIT)
Kimberly R. Davis (MIT)
David K. Duvenaud (UoT)
Mia T. Lieberman (MIT)

Advisor

Jonathan A. Runstadler (MIT)

Committee

Mark Bathe (MIT)

Jukka-Pekka Onnela (HSPH)

Students

Justin Zabilansky (MIT)

Kyle Yuan (MIT)

Andrea Nickerson (MIT)

Vivian Zhong (MIT)

