

Molecular evolution of FMDV

Using data to provide insights into tracing and disease prevalence

Dan Haydon (et al)

Collaborators & Funders

Graham Freimanis

Antonello Di Nardo

Don King

Paul Johnson

Mafalda Viana

Richard Orton

Abdelkader Behdenna

Eleanor Cottam

David Paton

Samuel Soubeyrand

Marco Morelli

Gael Thebaud

David King

Valerie Mioulet/WRLFMD



EMIDA



Overarching motivation

Can we use viral genetic data to :

- Trace who-infected-who?
- Identify missing cases?
- Estimate the case detection rate from a partially observed sample?

Bayesian Reconstruction of Disease Outbreaks by Combining Epidemiologic and Genomic Data

Thibaut Jombart*, Anne Cori, Xavier Didelot, Simon Cauchemez, Christophe Fraser*, Neil Ferguson

MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London, United Kingdom

Foot-and-Mouth Disease virus

A Bayesian Inference Framework to Reconstruct Transmission Trees Using Epidemiological and Genetic Data

Marco J. Morelli^{1†}, Gaël Thébaud², Joël Chadœuf³, Donald P. King⁴, Daniel T. Haydon^{1*},
Samuel Soubeyrand³

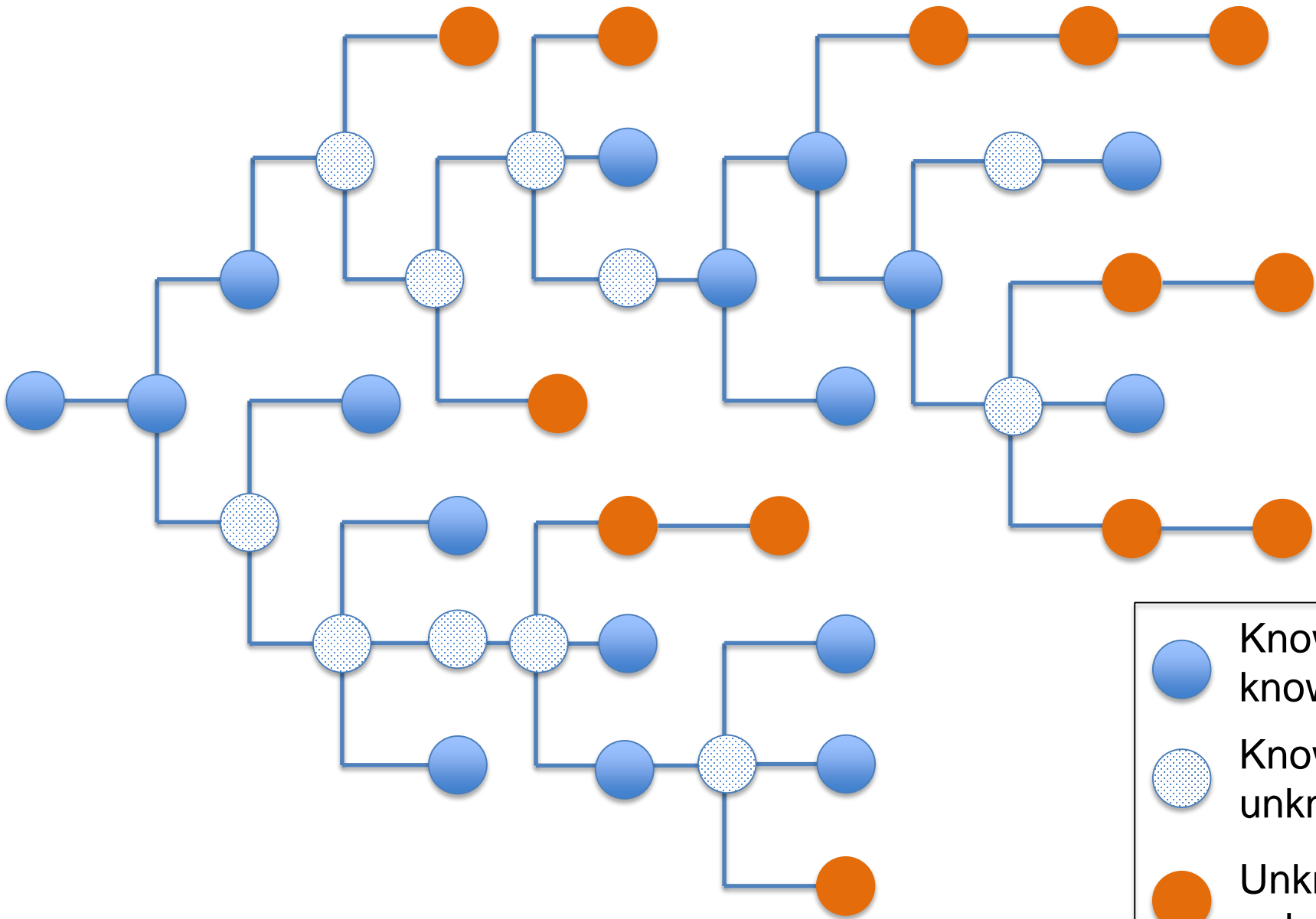
¹ Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom, ² INRA, UMR BGPI, Cirad TA A-54/K, Montpellier, France, ³ INRA, UR546 Biostatistics and Spatial Processes, Avignon, France, ⁴ Institute for Animal Health, Pirbright, United Kingdom




[†] Department of Zoology, University of Oxford, Oxford, UK
¹ Centre for Tropical Veterinary Medicine, University of Liverpool, Neston, UK
² Eosene Corporation, 700, 5th Avenue, New York, NY, USA
³ Veterinary Laboratories Agency, New Haw, Surrey, UK
⁴ Department of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK

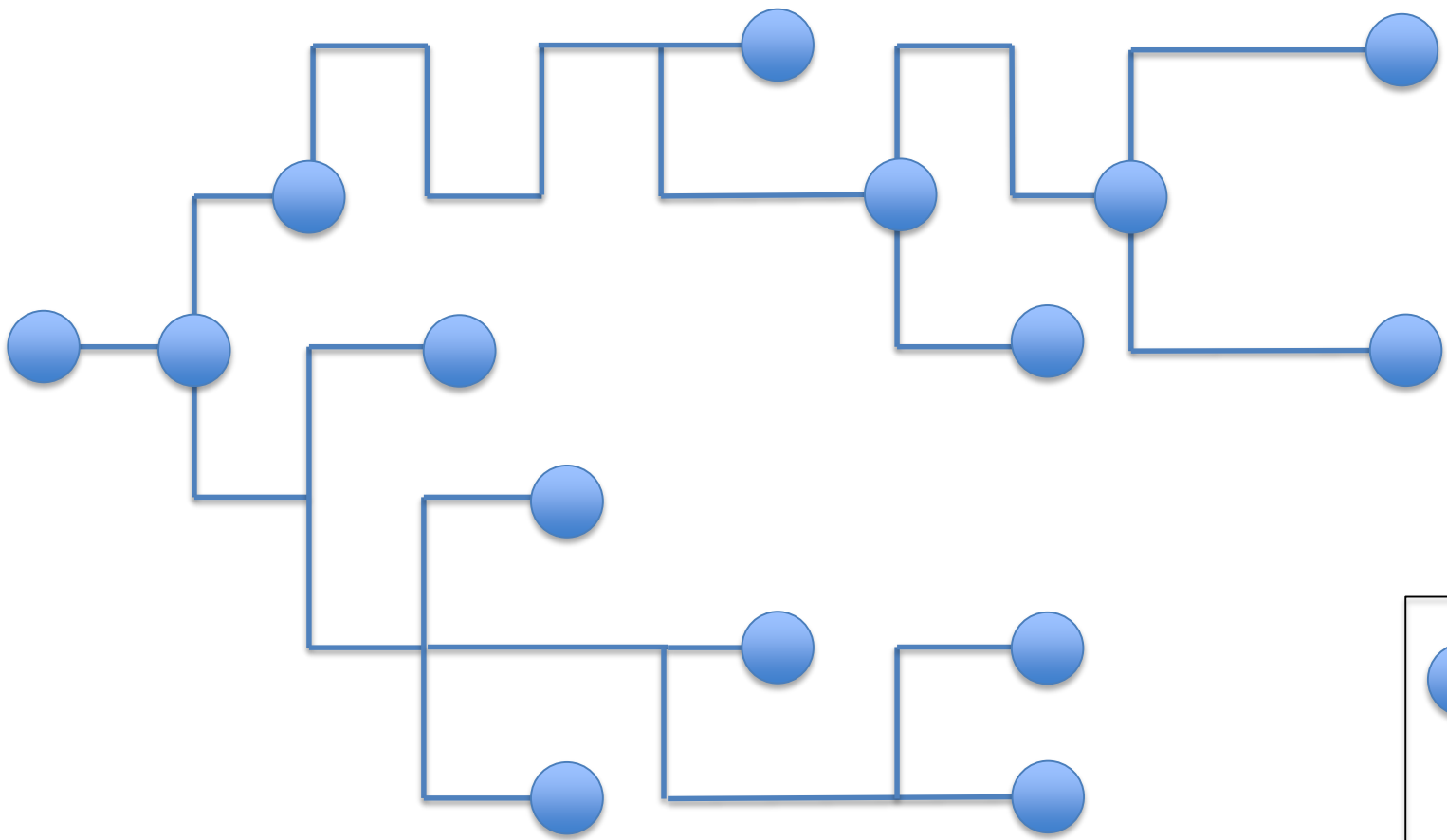
Overarching motivation


Can we use viral genetic data to

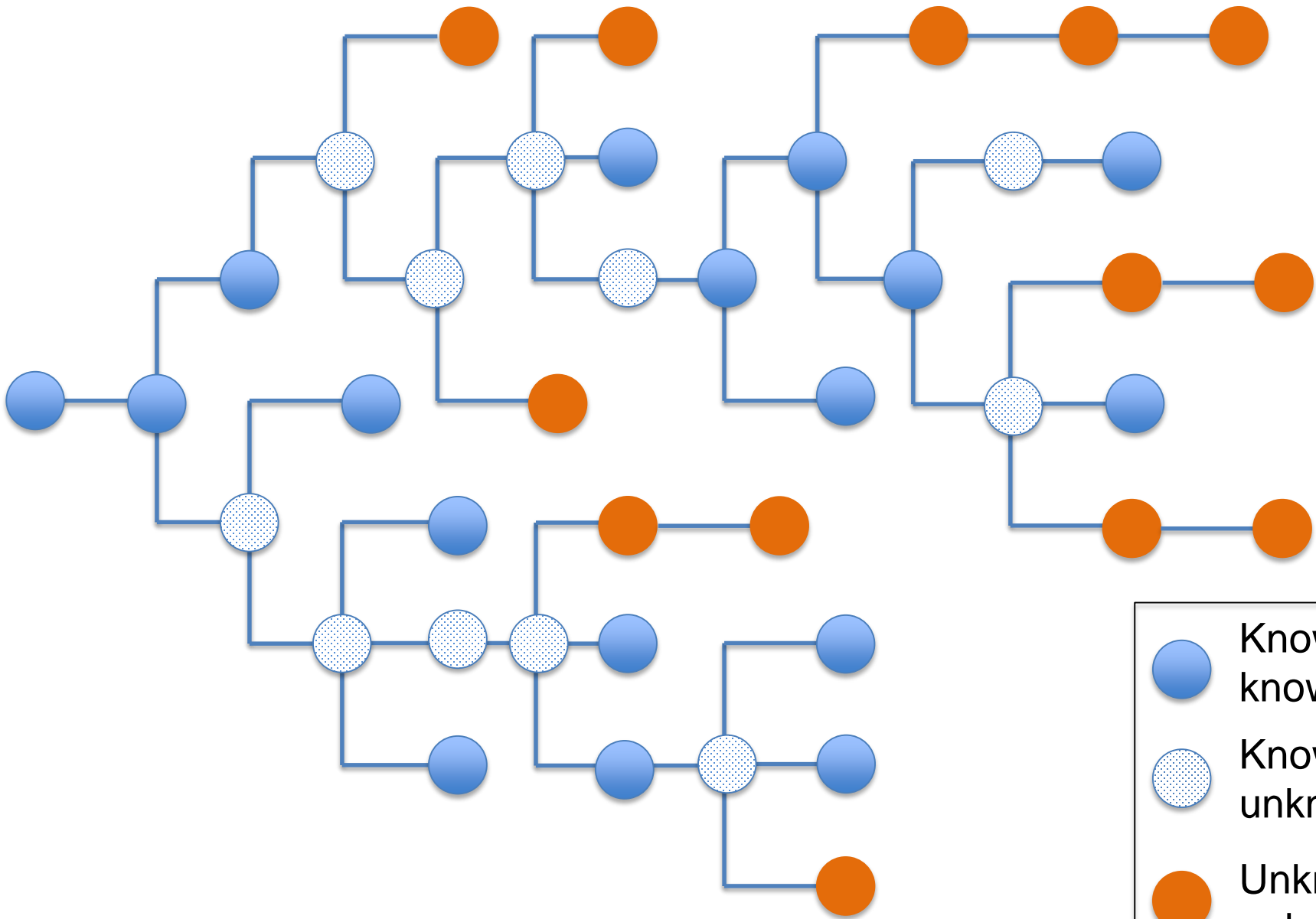
- Trace who-infected-who?
- Identify missing cases?
- Estimate the case detection rate from a partially observed sample?






	Known knows
	Known unknowns
	Unknown unknowns



 Known
knowns



	Known knows
	Known unknowns
	Unknown unknowns

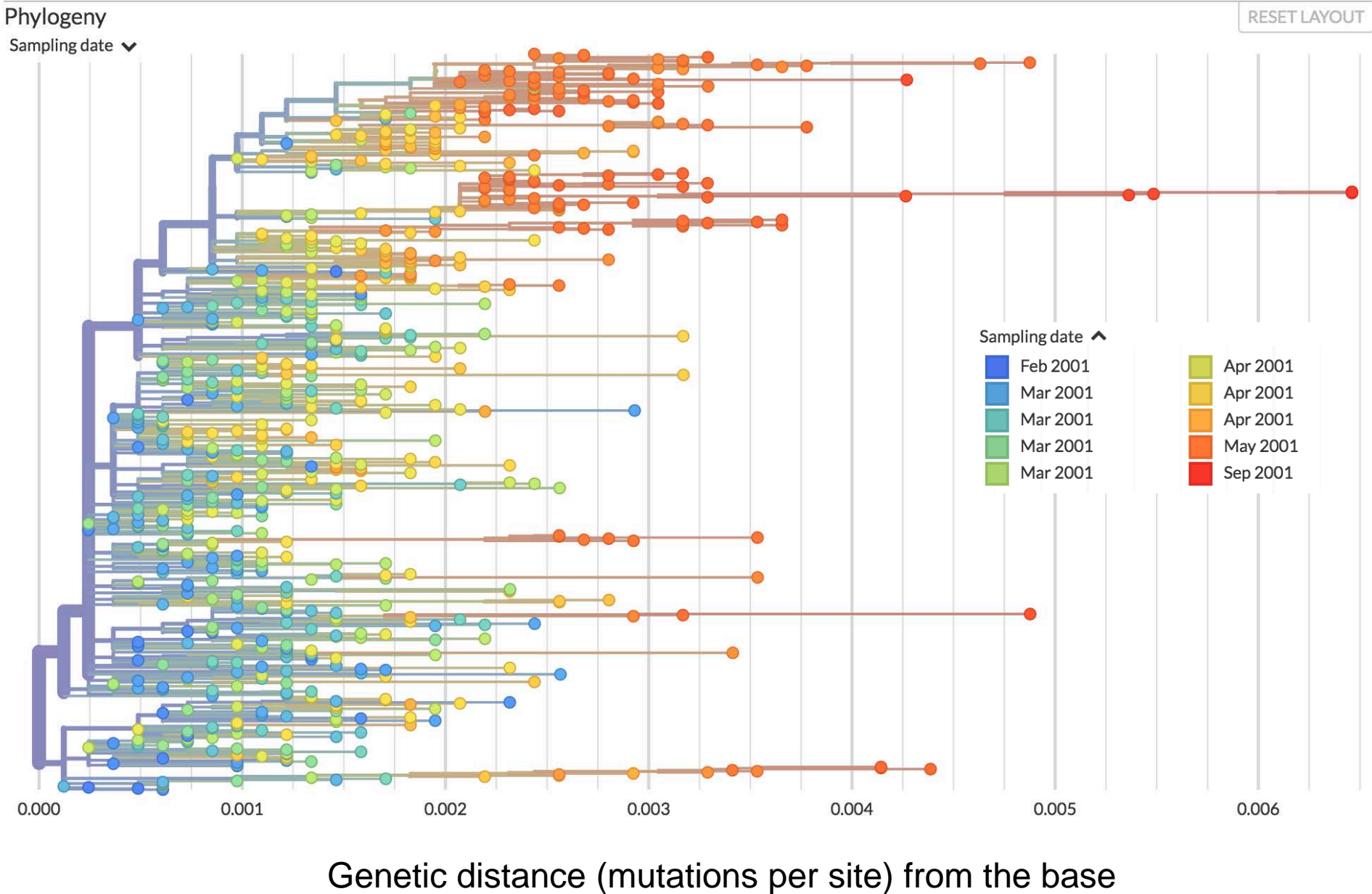
Key Questions

- How many mutations typically arise on an infected premise (IP)?
- How many mutations typically arise post transmission?
- What proportion of sampled IPs left descendants that we sampled?

The 2001 UK FMDV outbreak

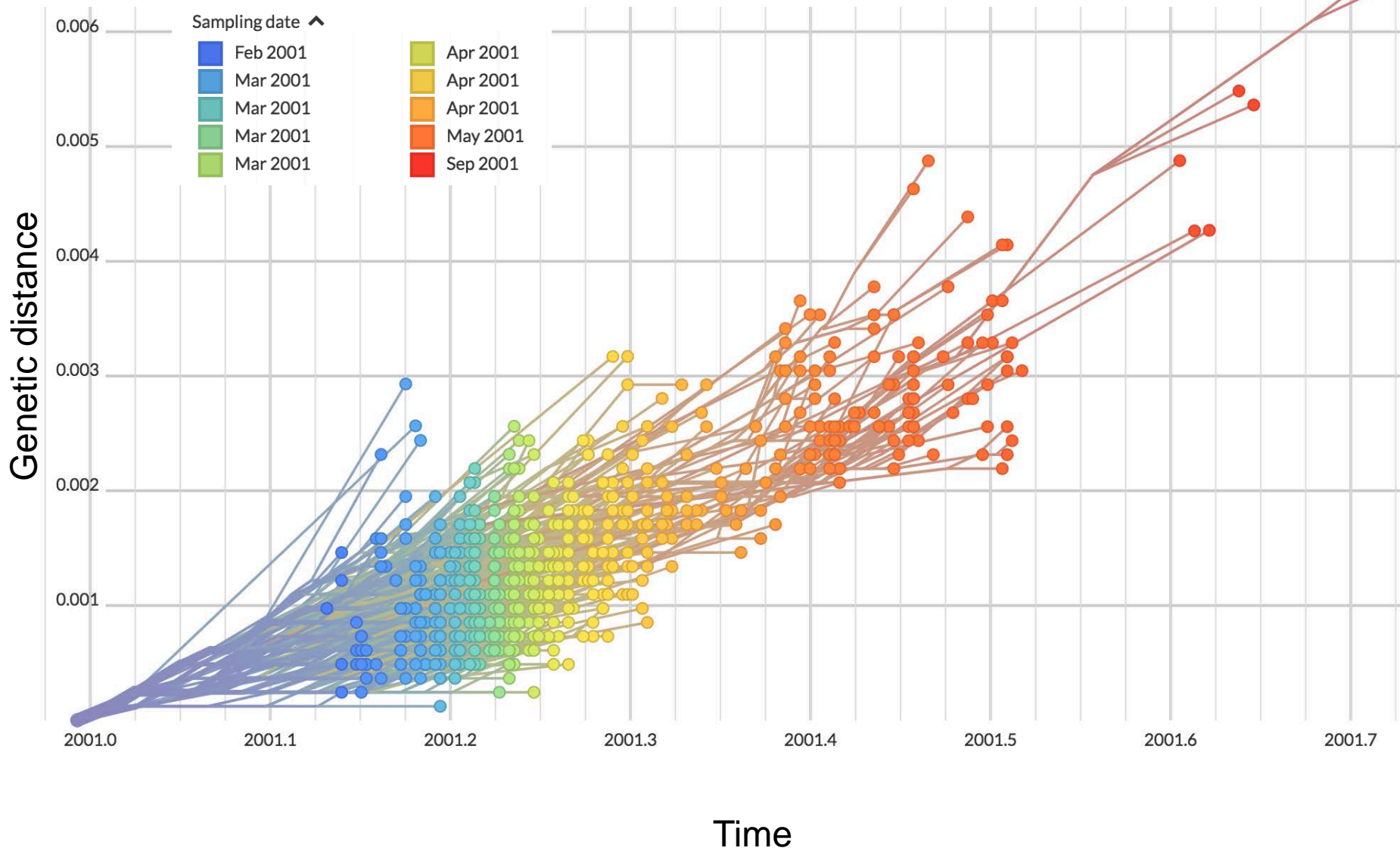
- First case reported 23rd Feb. 2001
- Last case reported 30th Sep. 2001
- 2026 officially recognized IPs
- IPs culled asap following reporting
- 7455 farms pre-emptively culled
- Whole genome sequences from 684 IPs

Length of the phylogeny = 3058 mutations



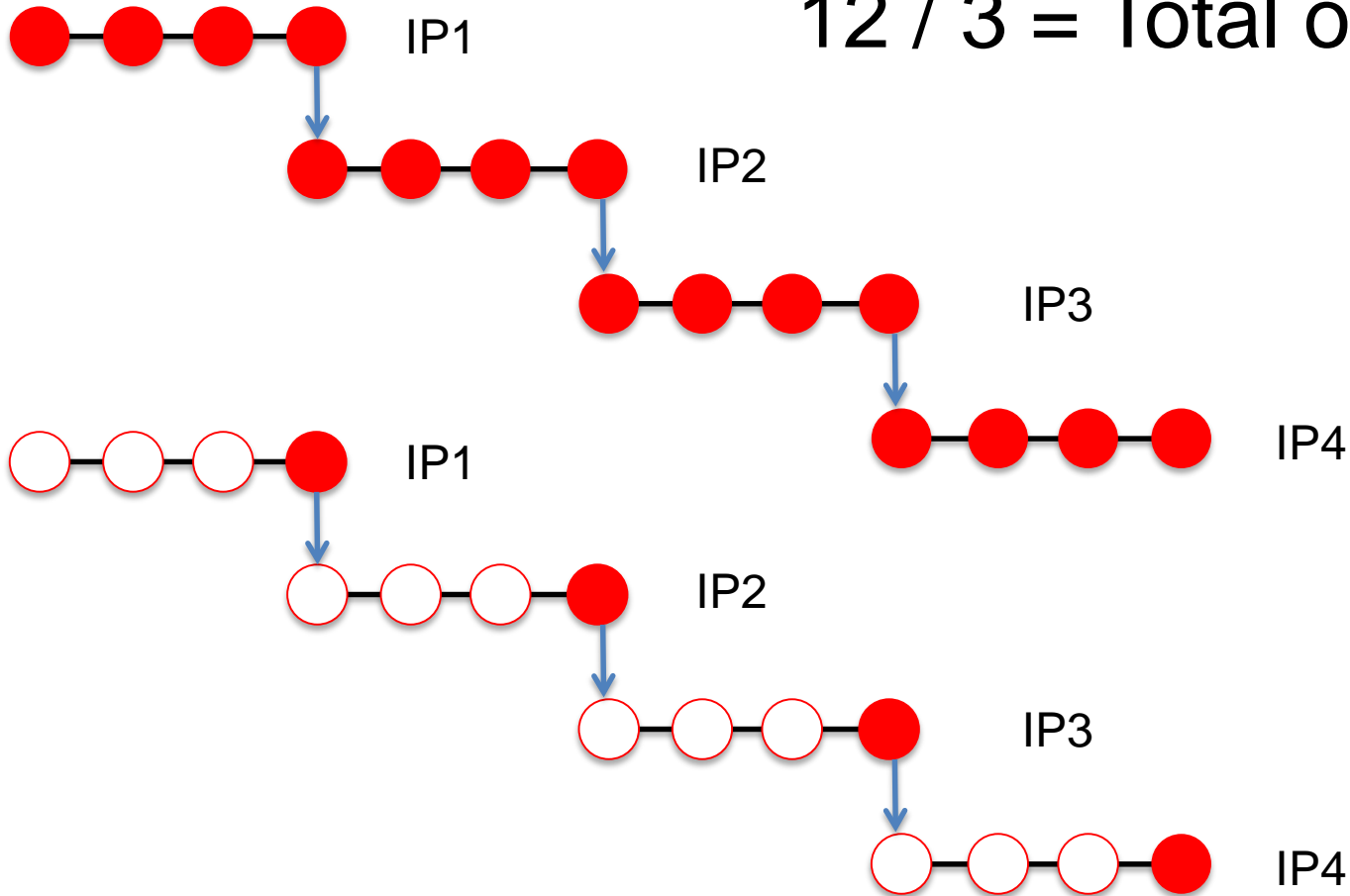
Clock ~ just over 1 mutation fixed/week

RESET LAYOUT

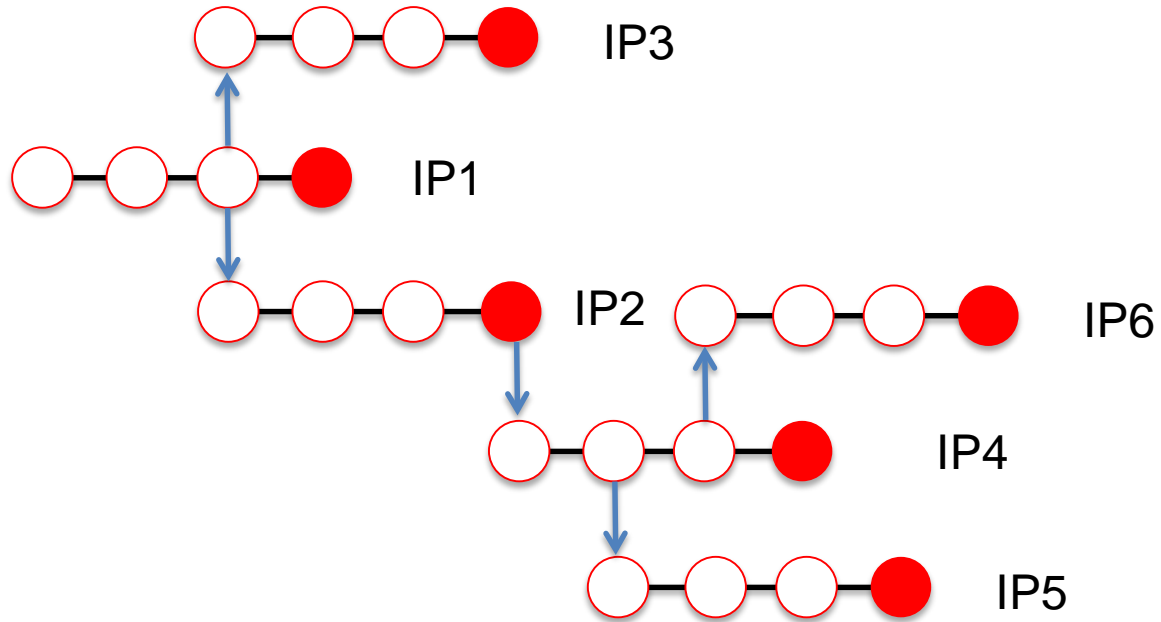


D mutations per IP

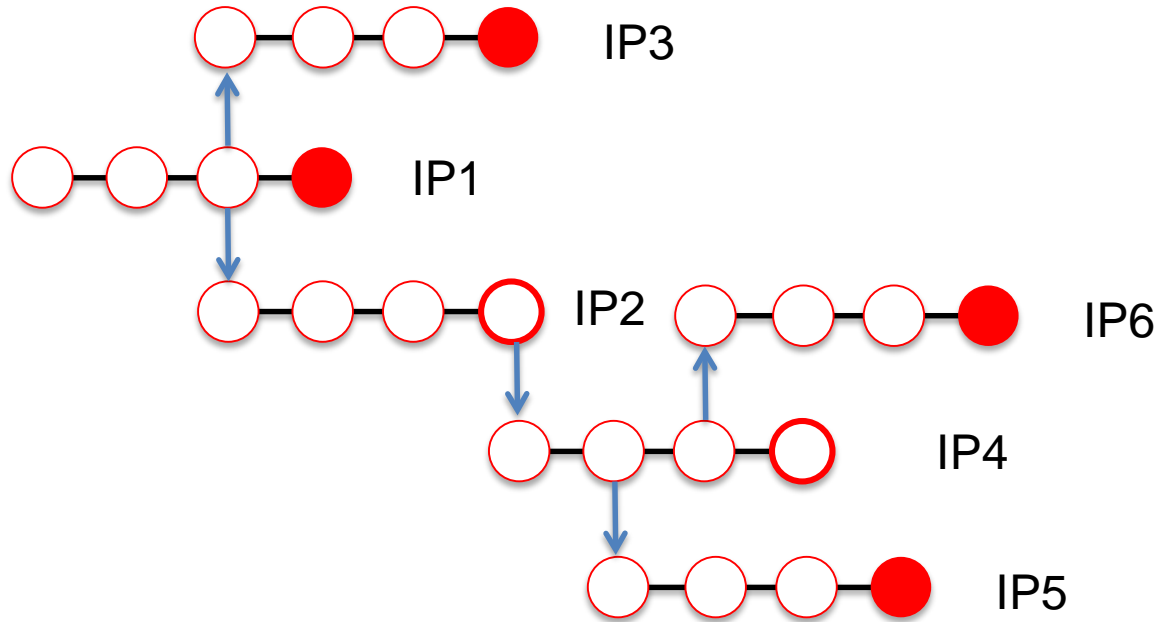
$12 / 3 = \text{Total of 4 IPs}$



$9 / 3 = 3 \text{ IPs (+1 for the index case)}$



$15 / 3 = 5$ IPs (+1 for the index case)



$15 / 3 = 5$ IPs (+1 for the index case)

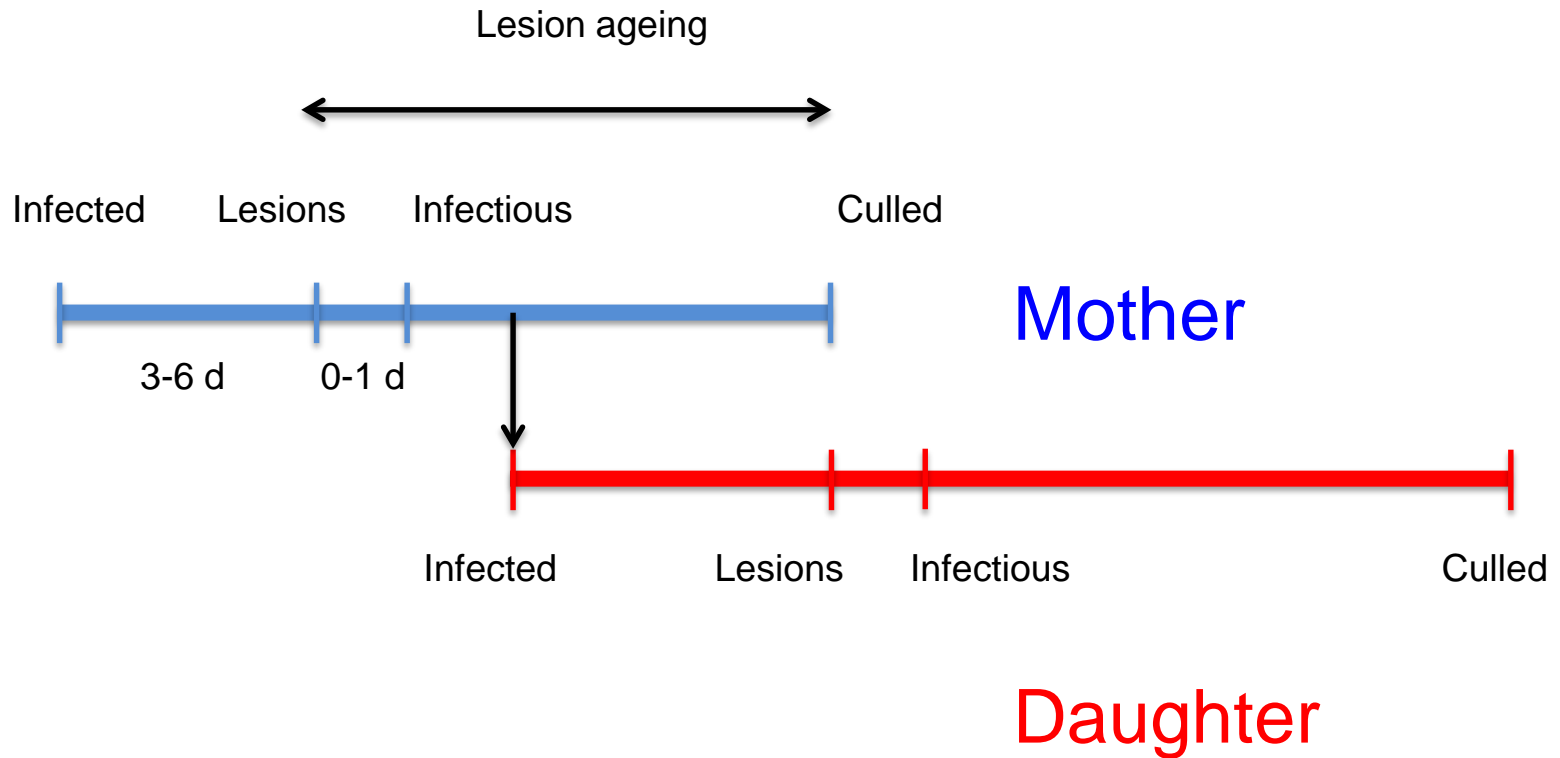
How to estimate D ?

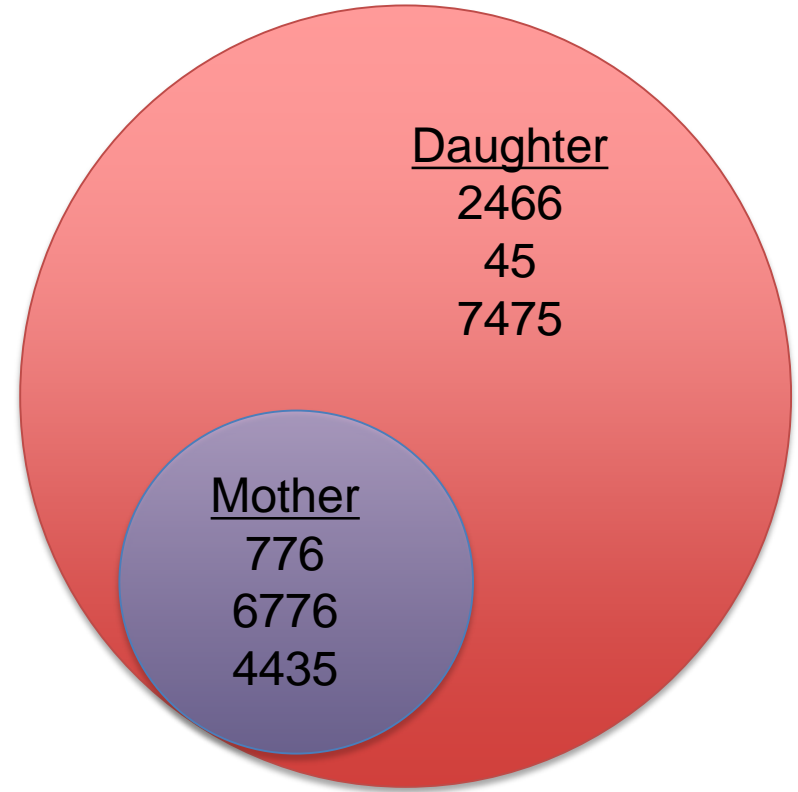
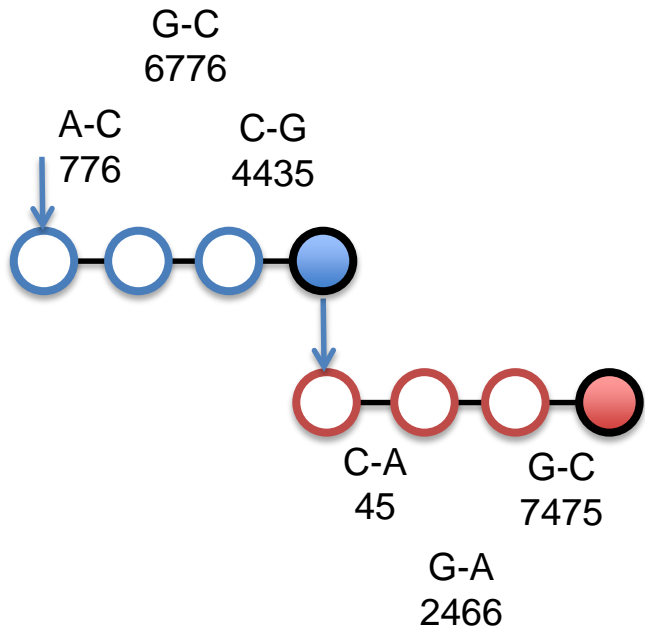
We need to identify 'mother daughter' pairs

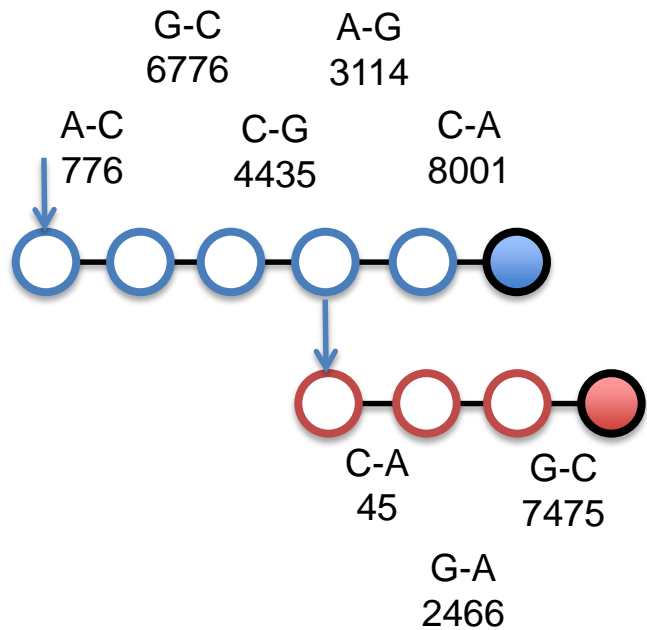
A daughter should be related to a mother according to:

- Temporal criteria
- Genetic criteria
- Perhaps spatial criteria

Temporal criteria

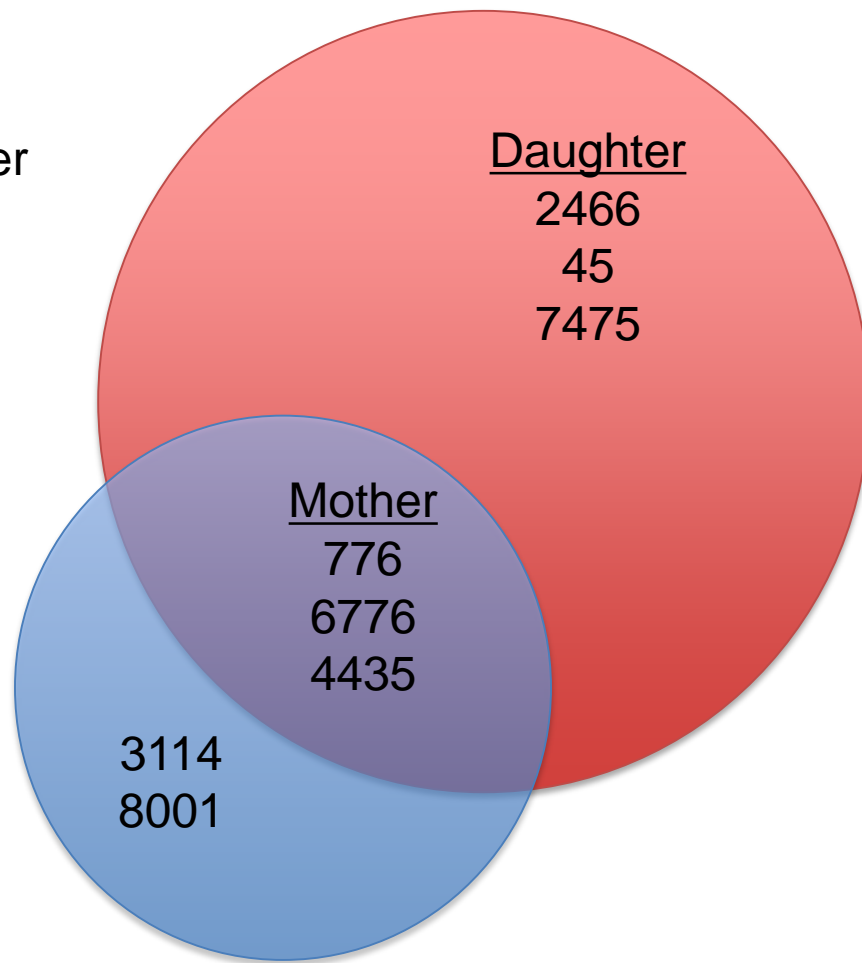


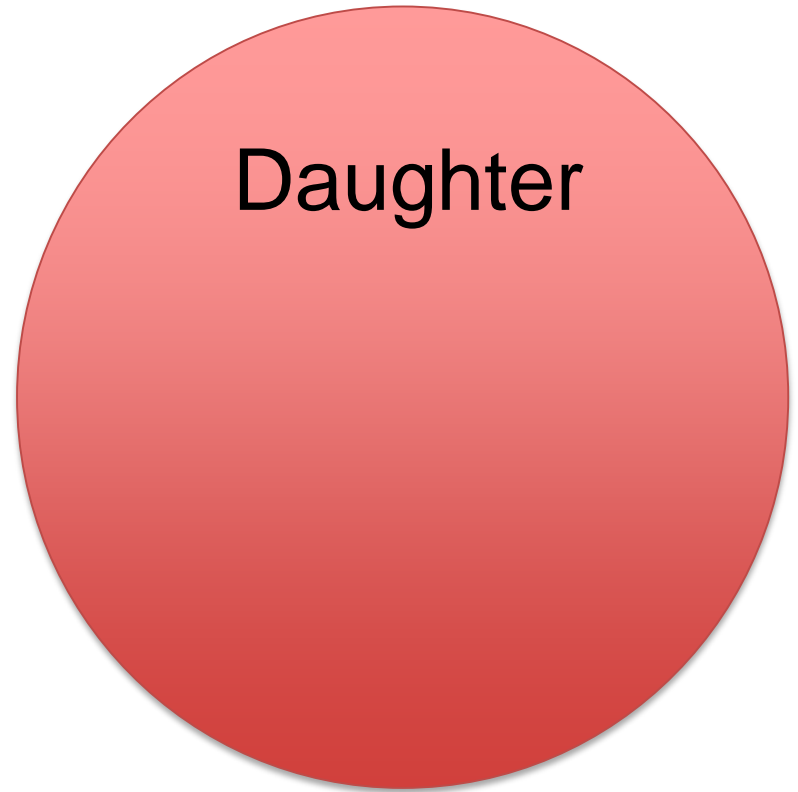
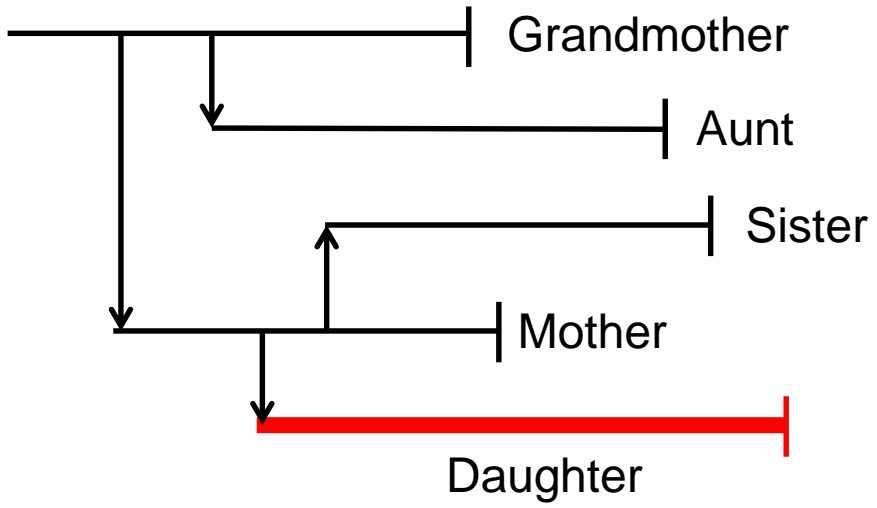


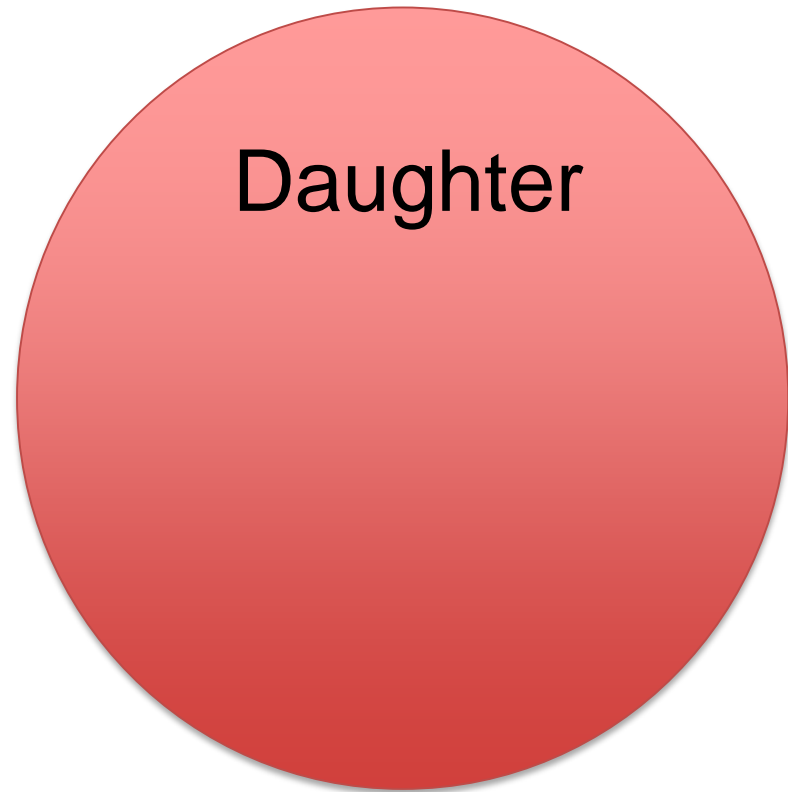
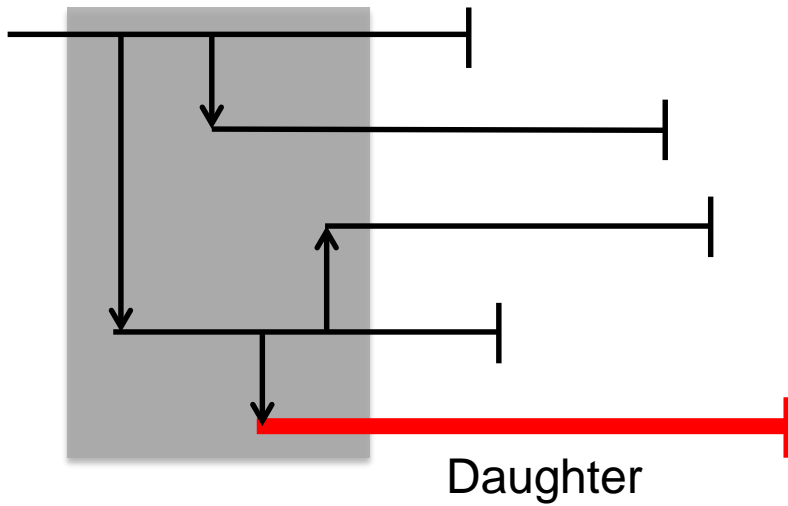


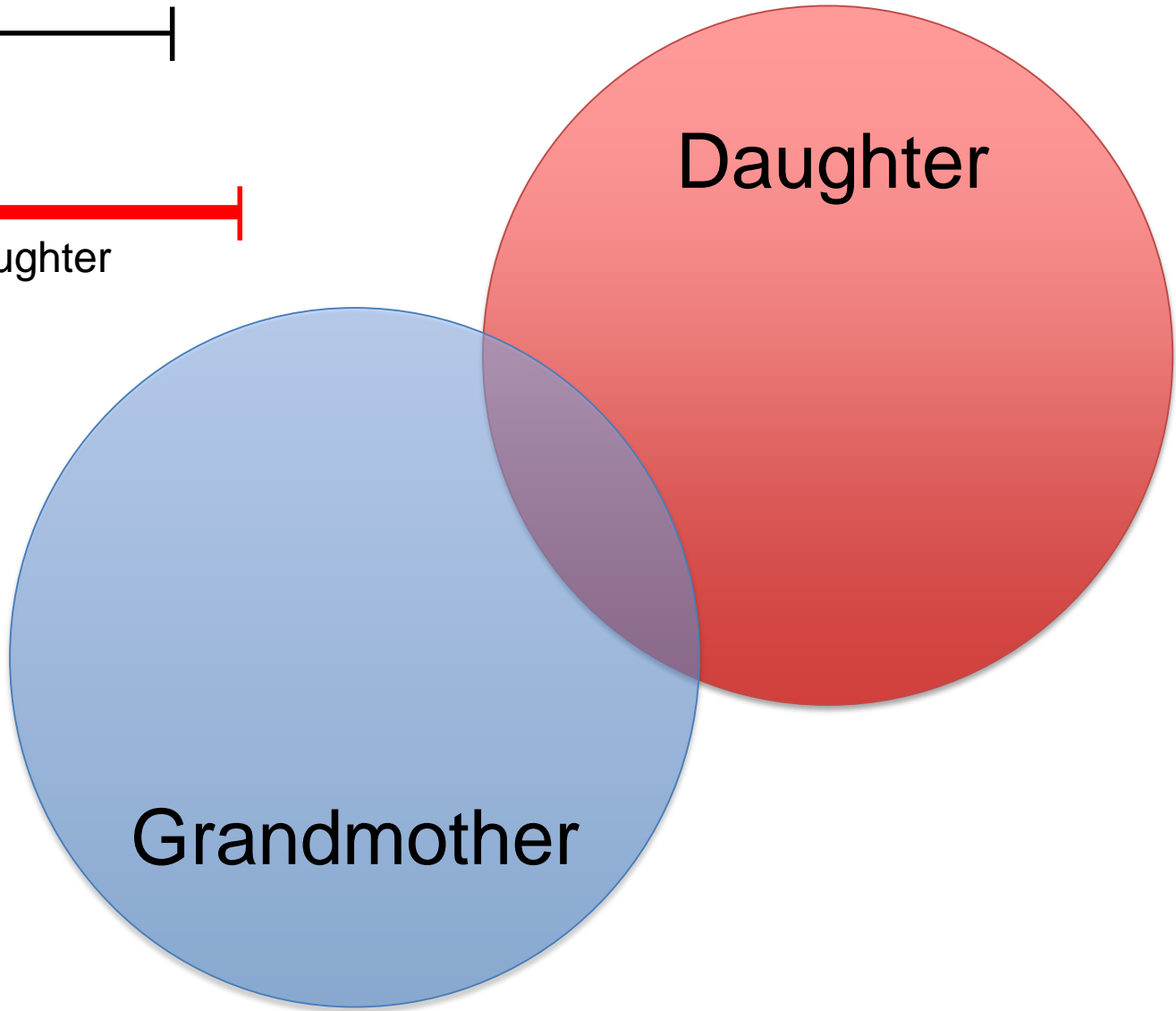
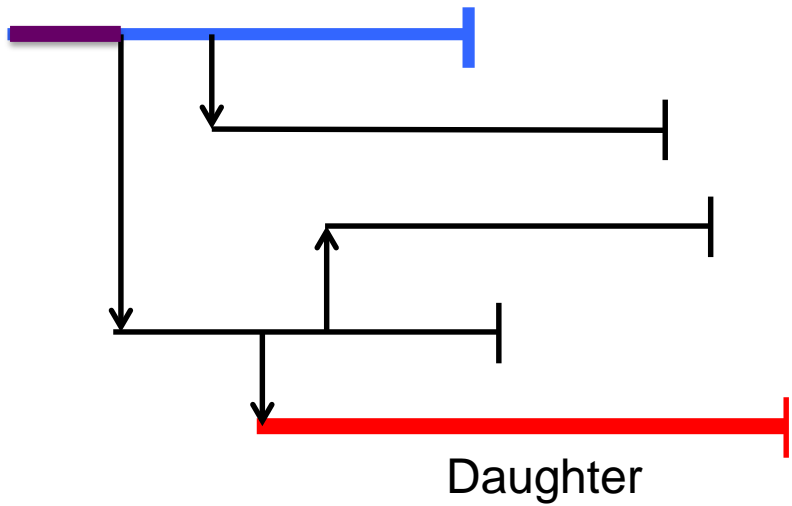
Mother

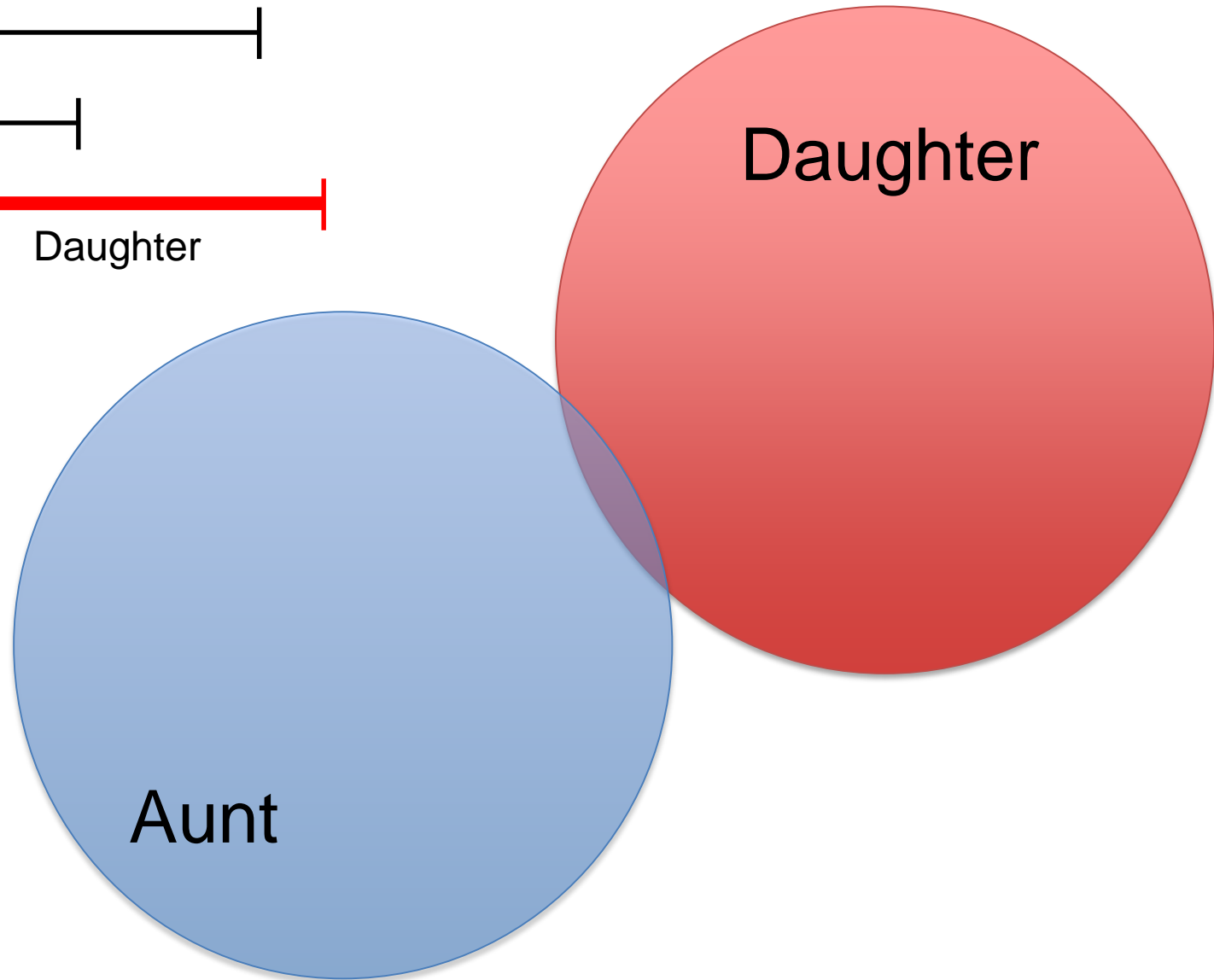
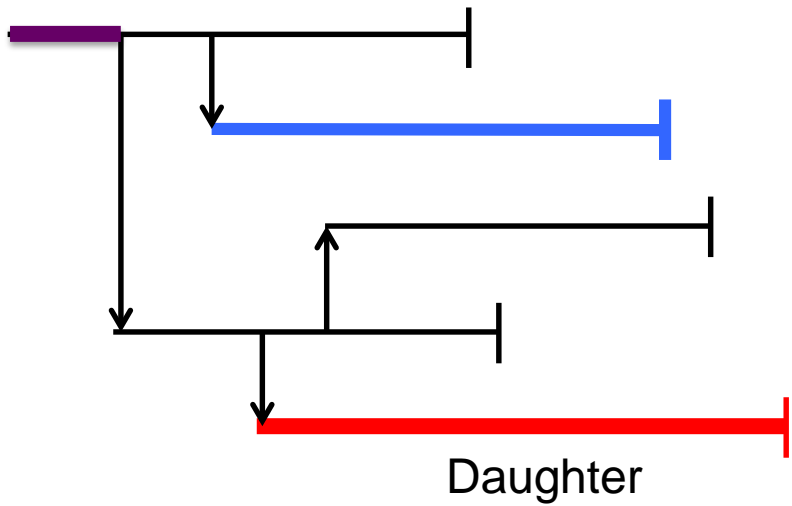
Daughter

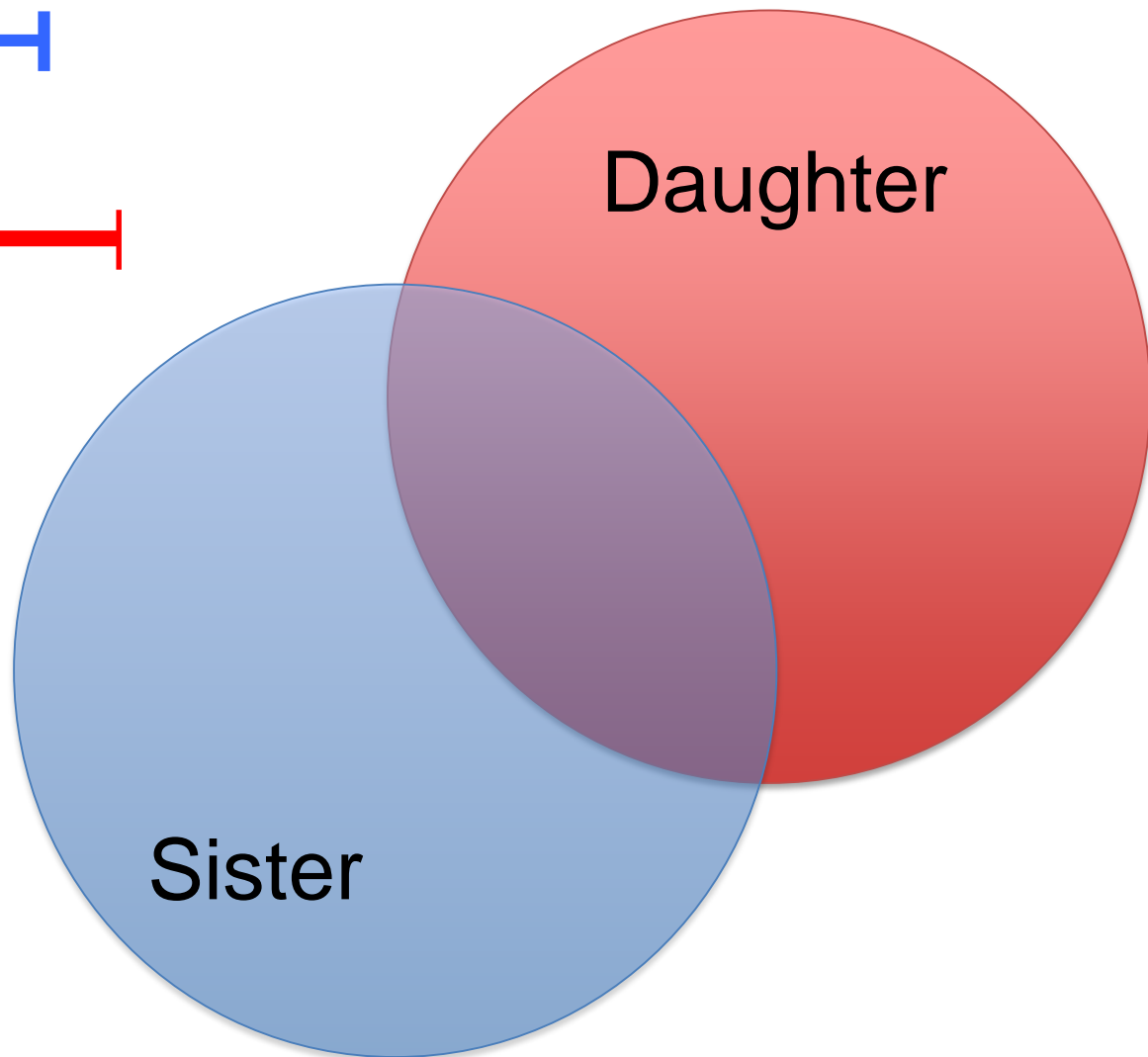
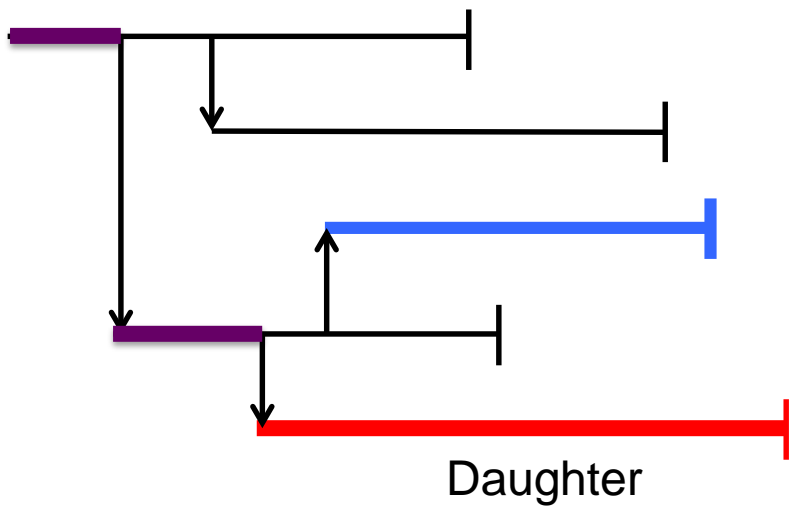


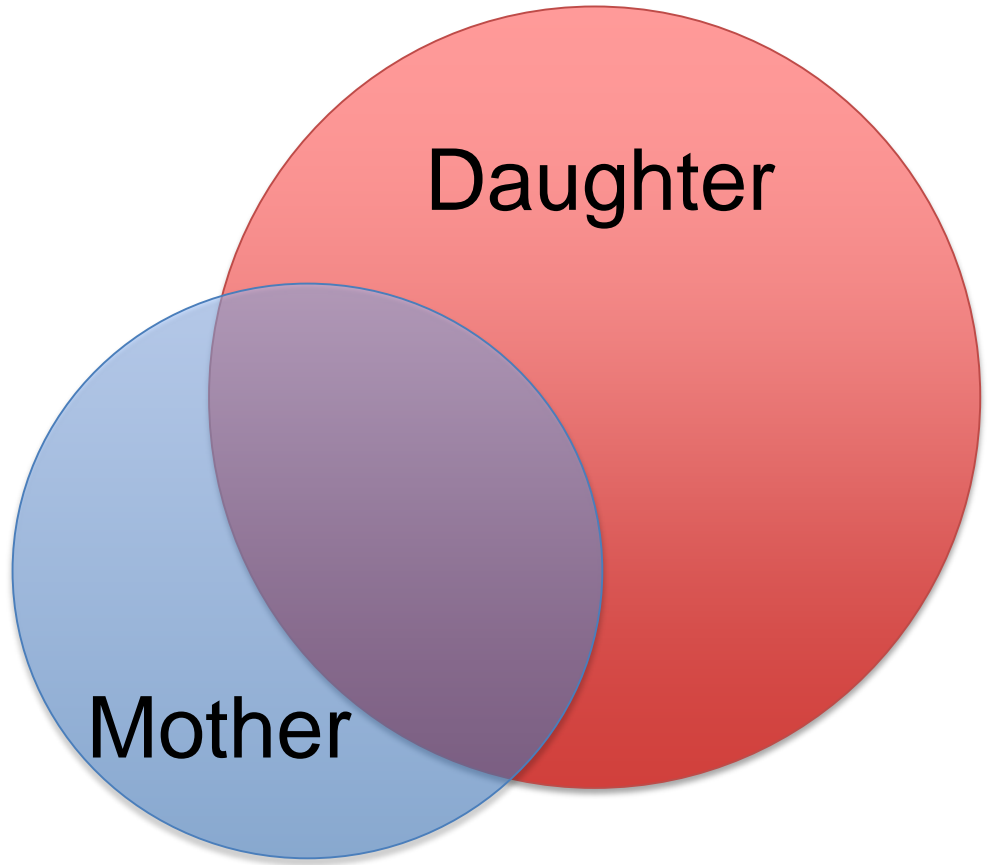
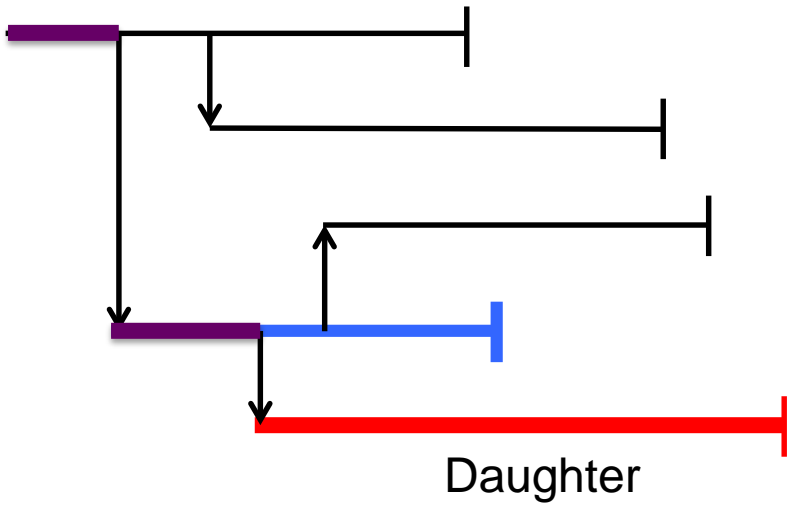








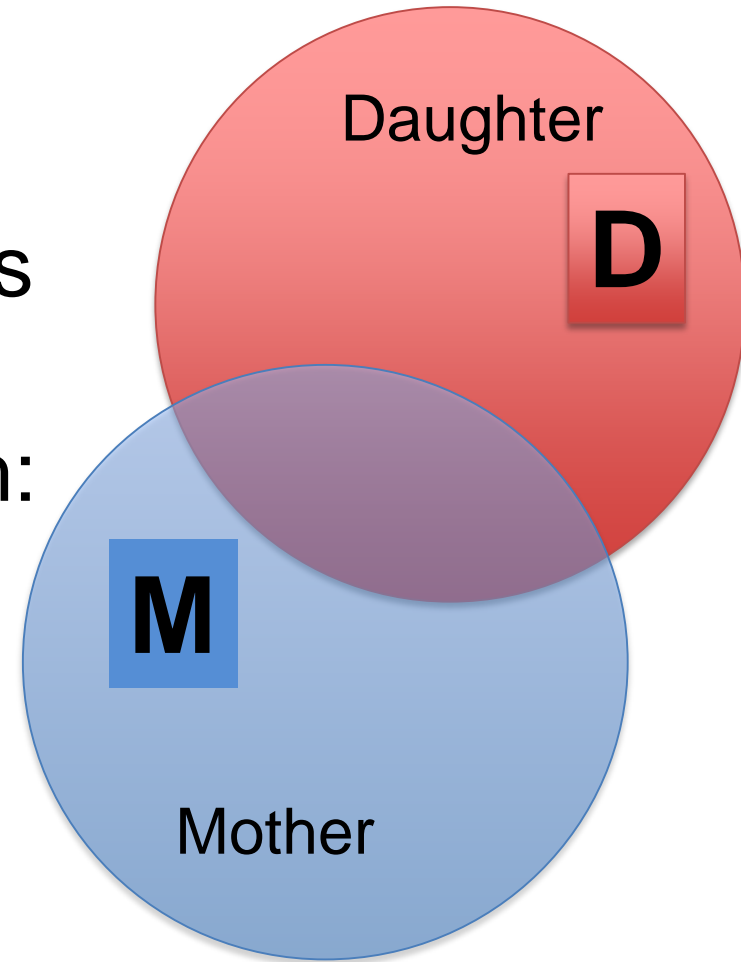




For each of the 683 daughters

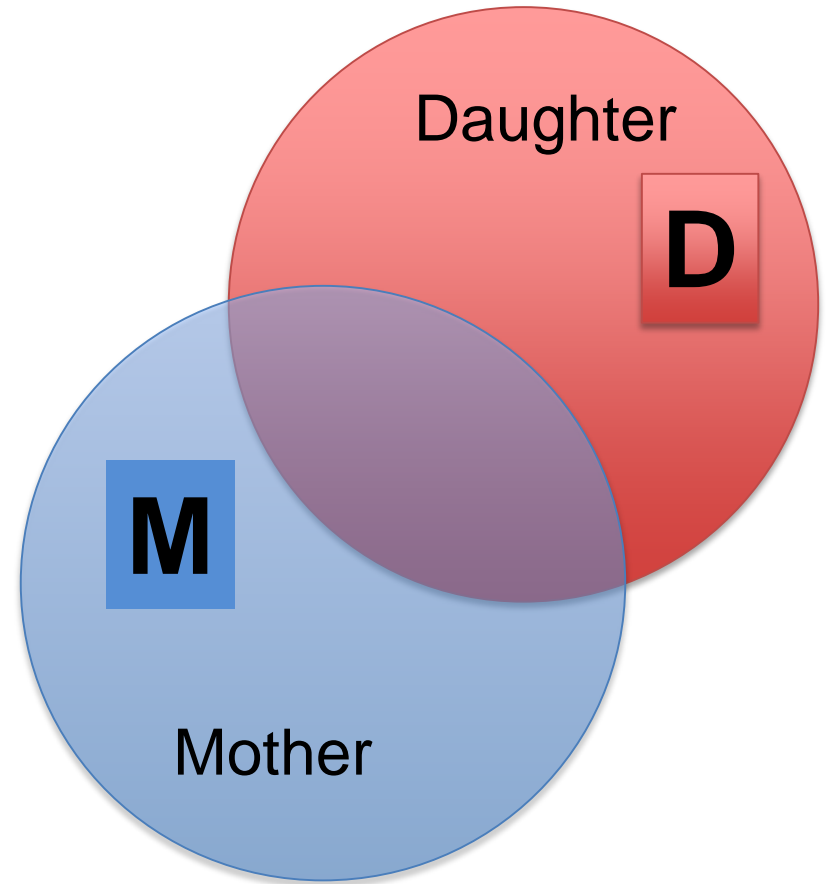
Find putative mother IP which:

- Fits timing criteria ✓
- Minimizes D ✓
- Minimizes M ✓

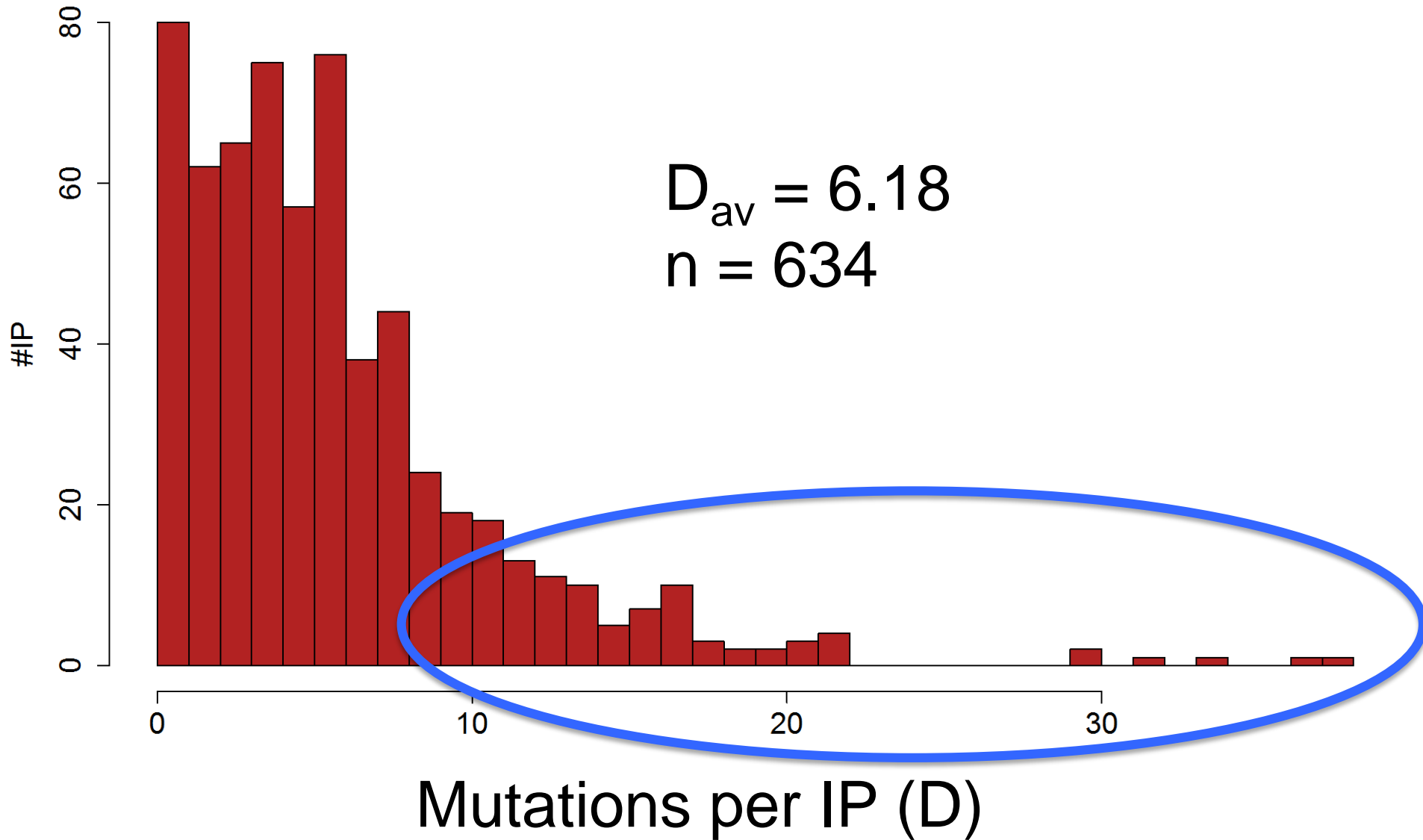


Average # mutations

on an IP = D_{av}



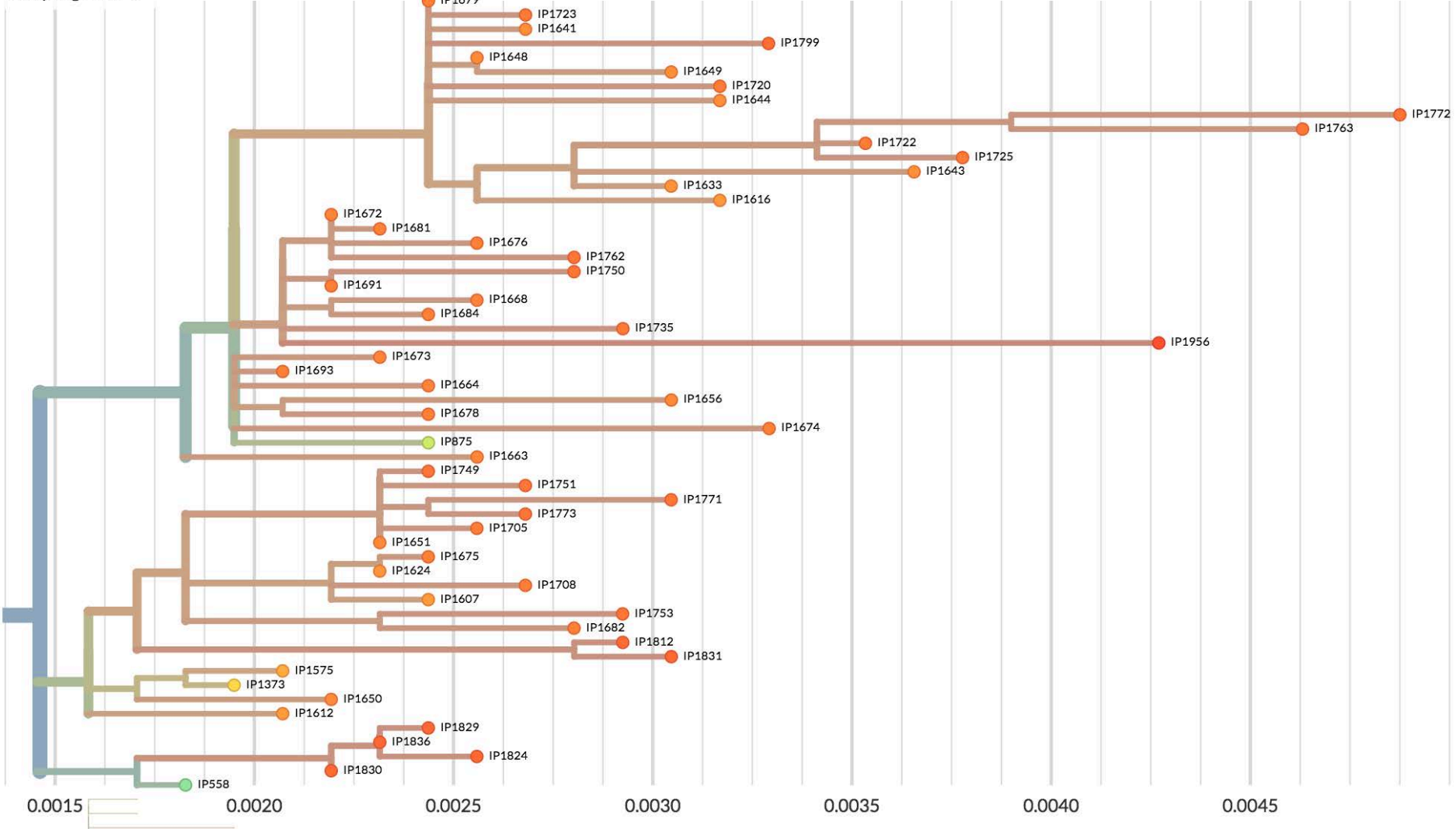
Frequency distribution of D



Phylogeny

RESET LAYOUT

Sampling date ▾



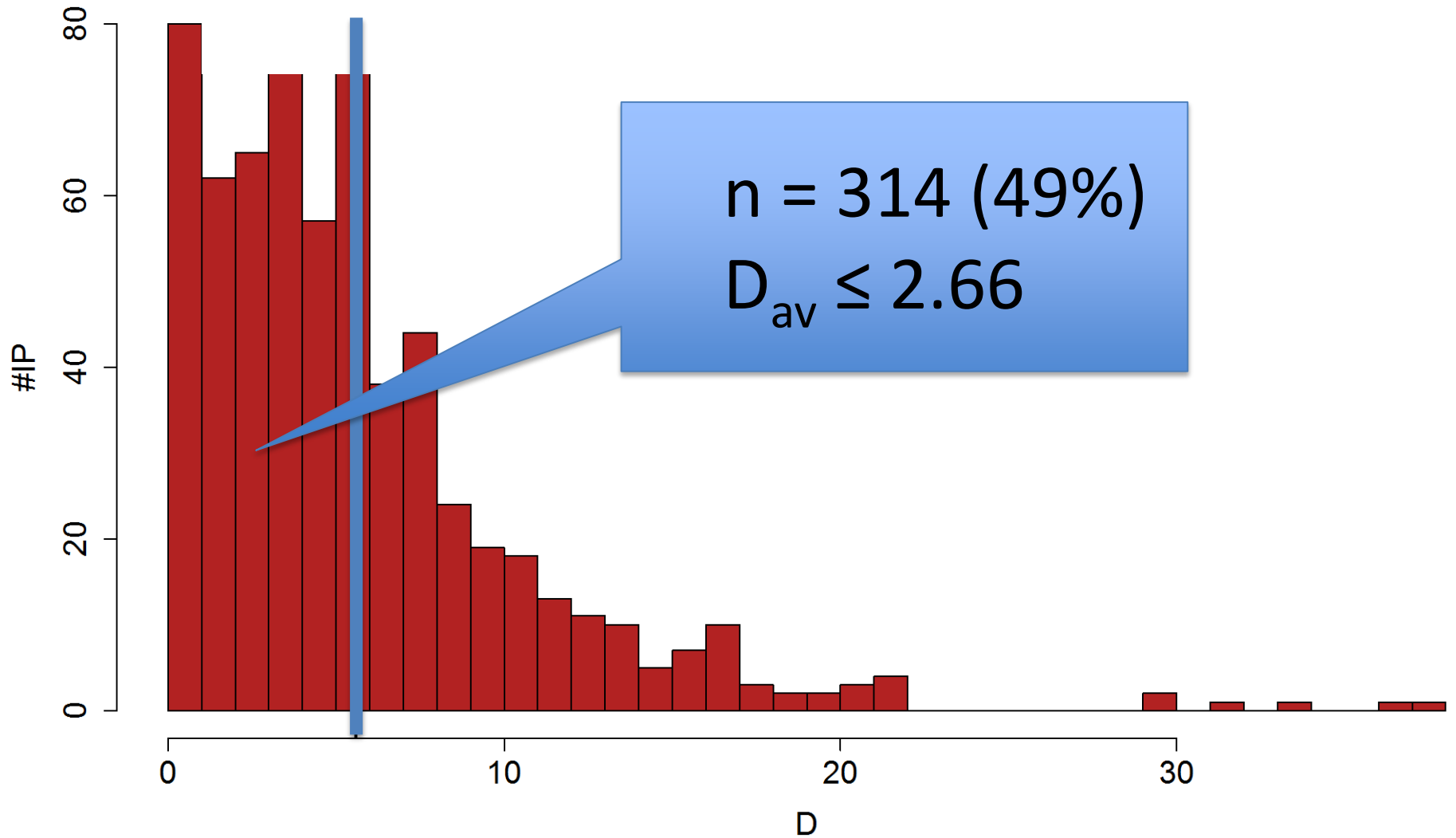
Prop IPs that left descendants (P_m) =

Sum of Internal branch lengths x $\left[M_{av} + D_{av} \right]$

Tree length x D_{av}

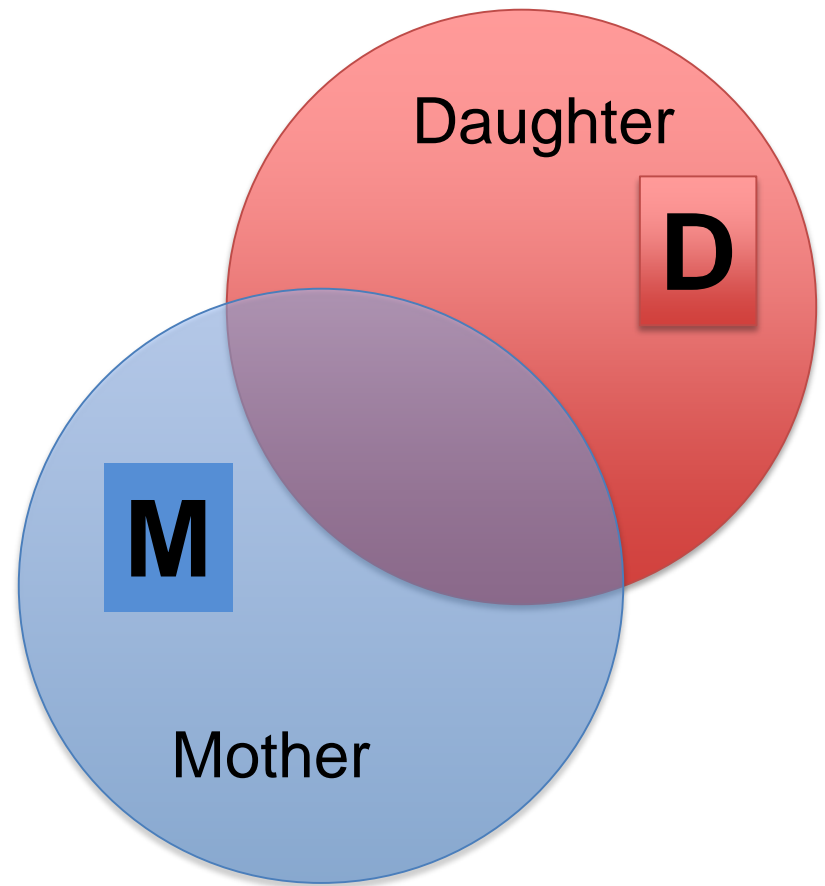
$$P_m = 0.49$$

Trimmed frequency distribution of D



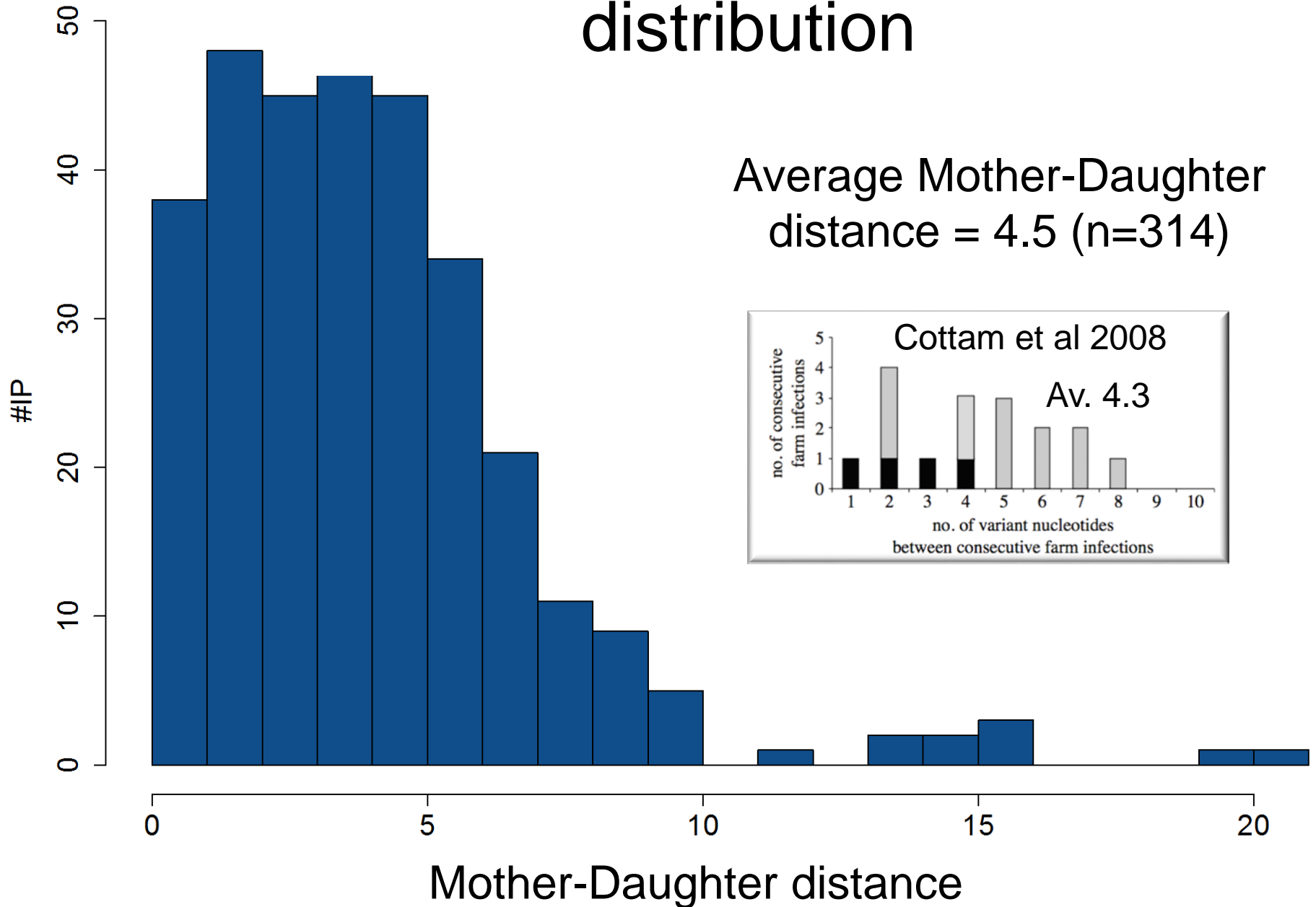
Mother to Daughter
distance

$$= \mathbf{M} + \mathbf{D}$$

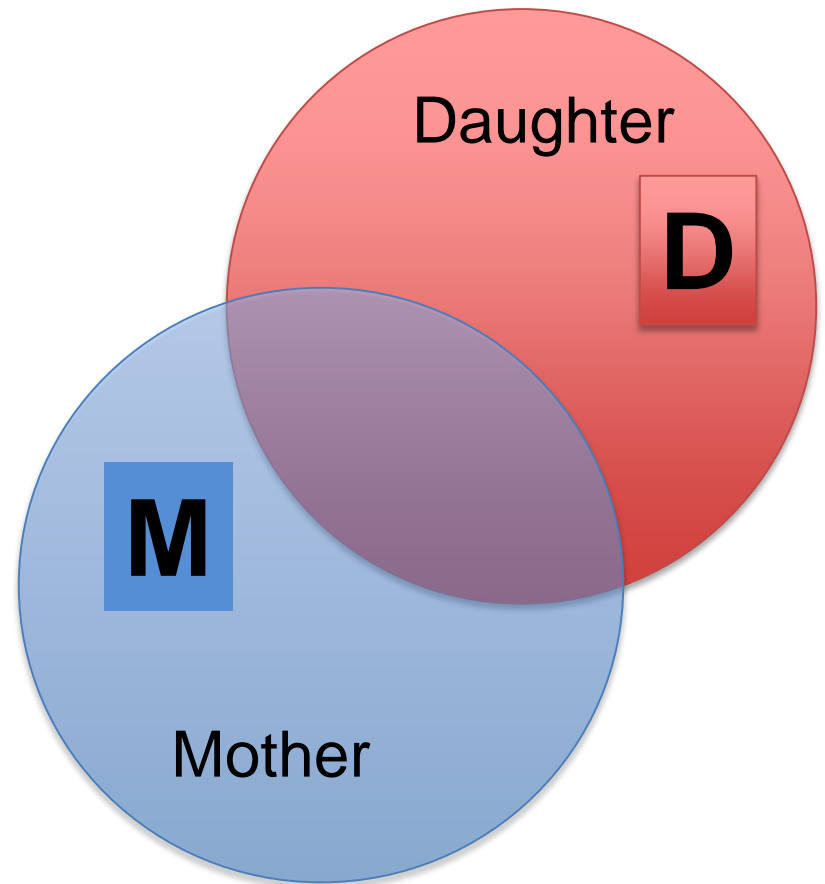


Mother-Daughter distance distribution

Average Mother-Daughter distance = 4.5 (n=314)



$$\# \text{ IPs in tree} = \text{Tree Length} / \mathbf{D}_{av}$$



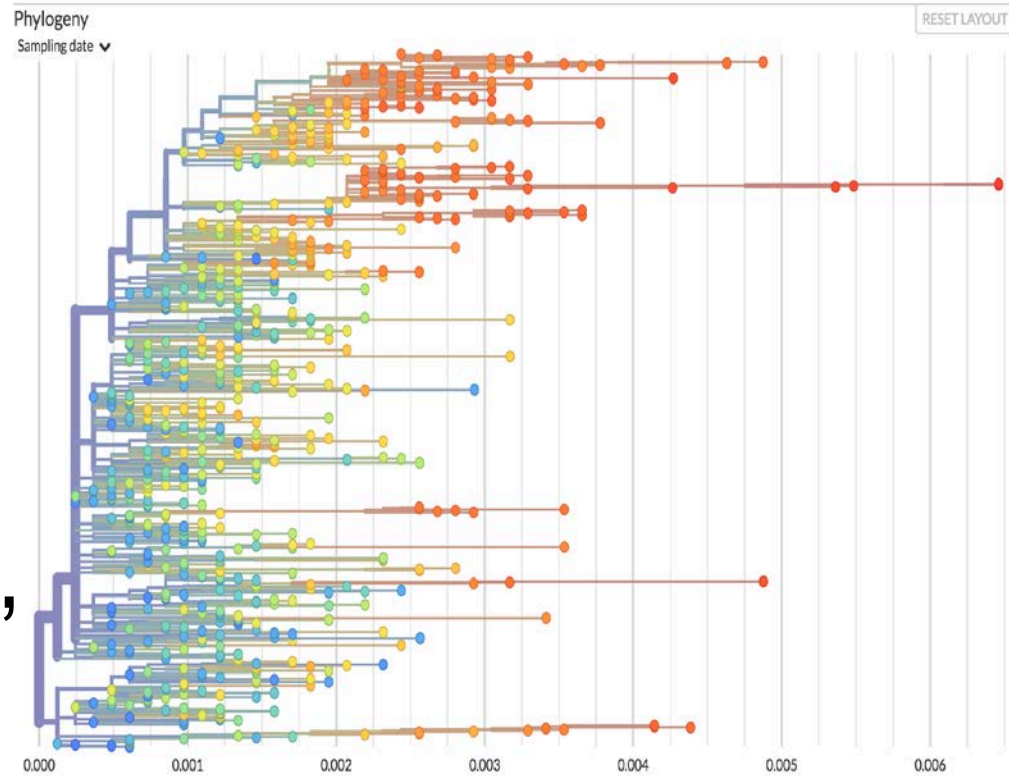
Tree length = 3058 mutations

$D_{av} \leq 2.66$

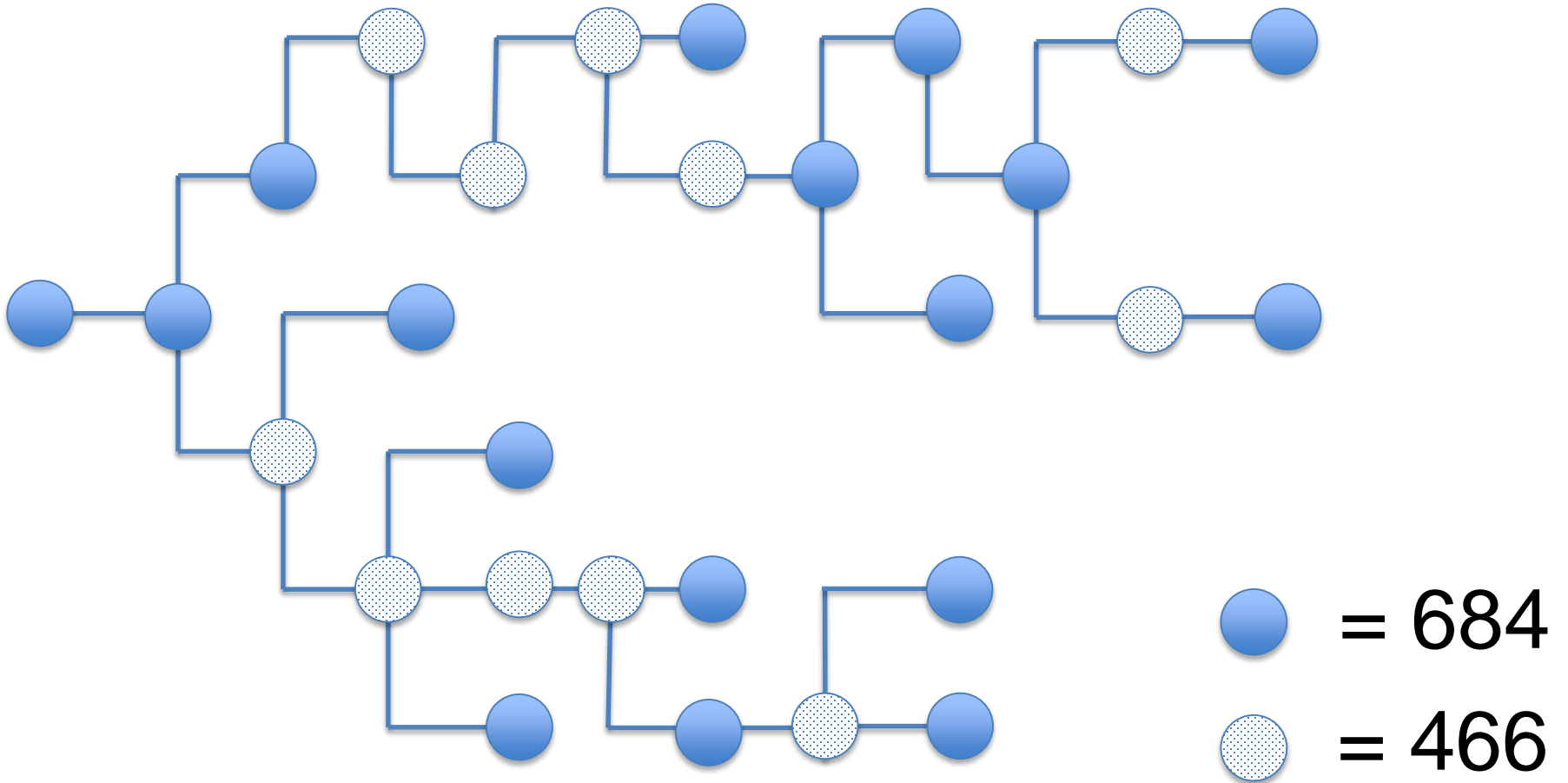
IPs 'in' tree ≥ 1150

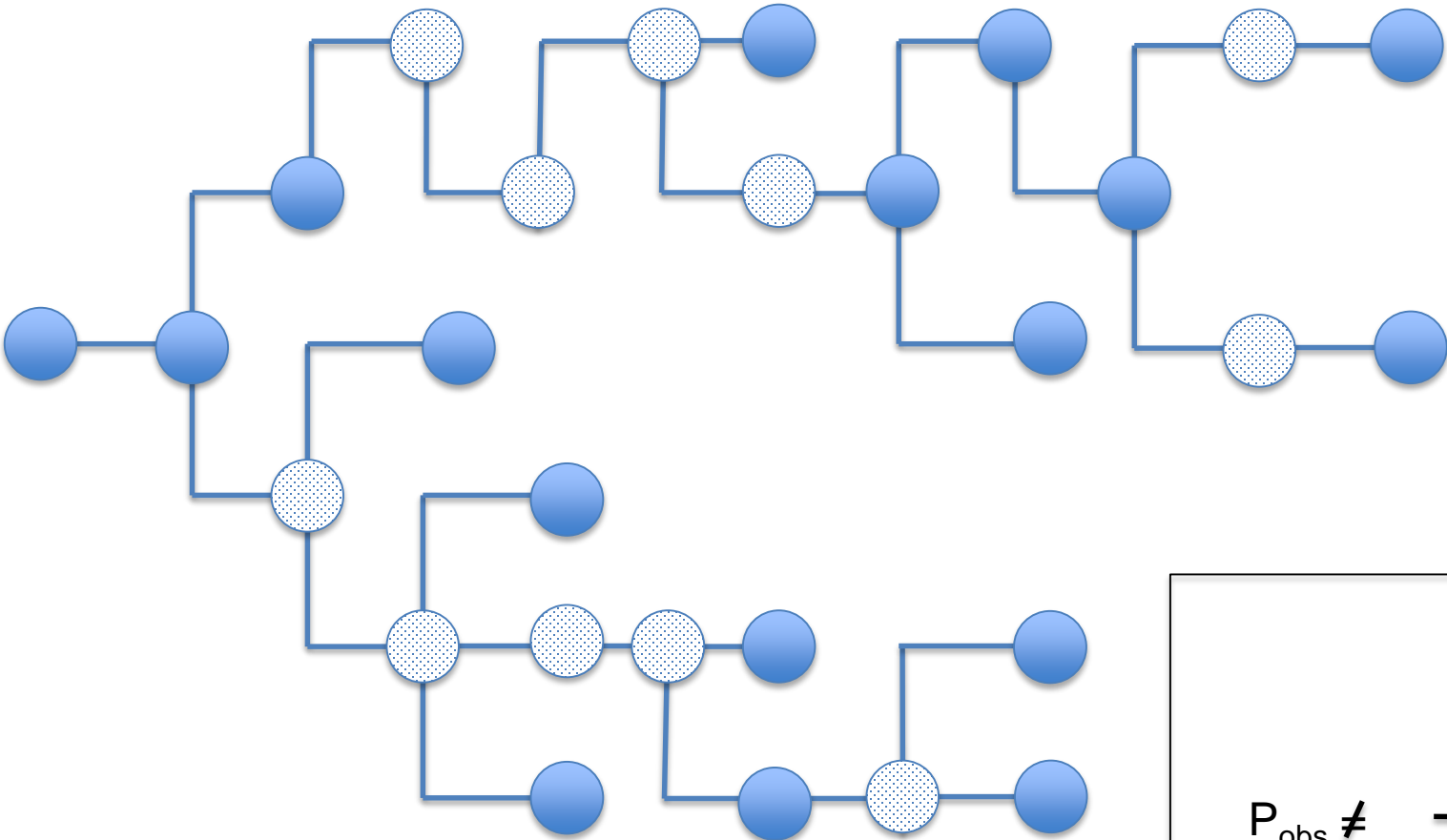
684 'known knowns'

$1150 - 684 \geq 466$ 'known unknowns'

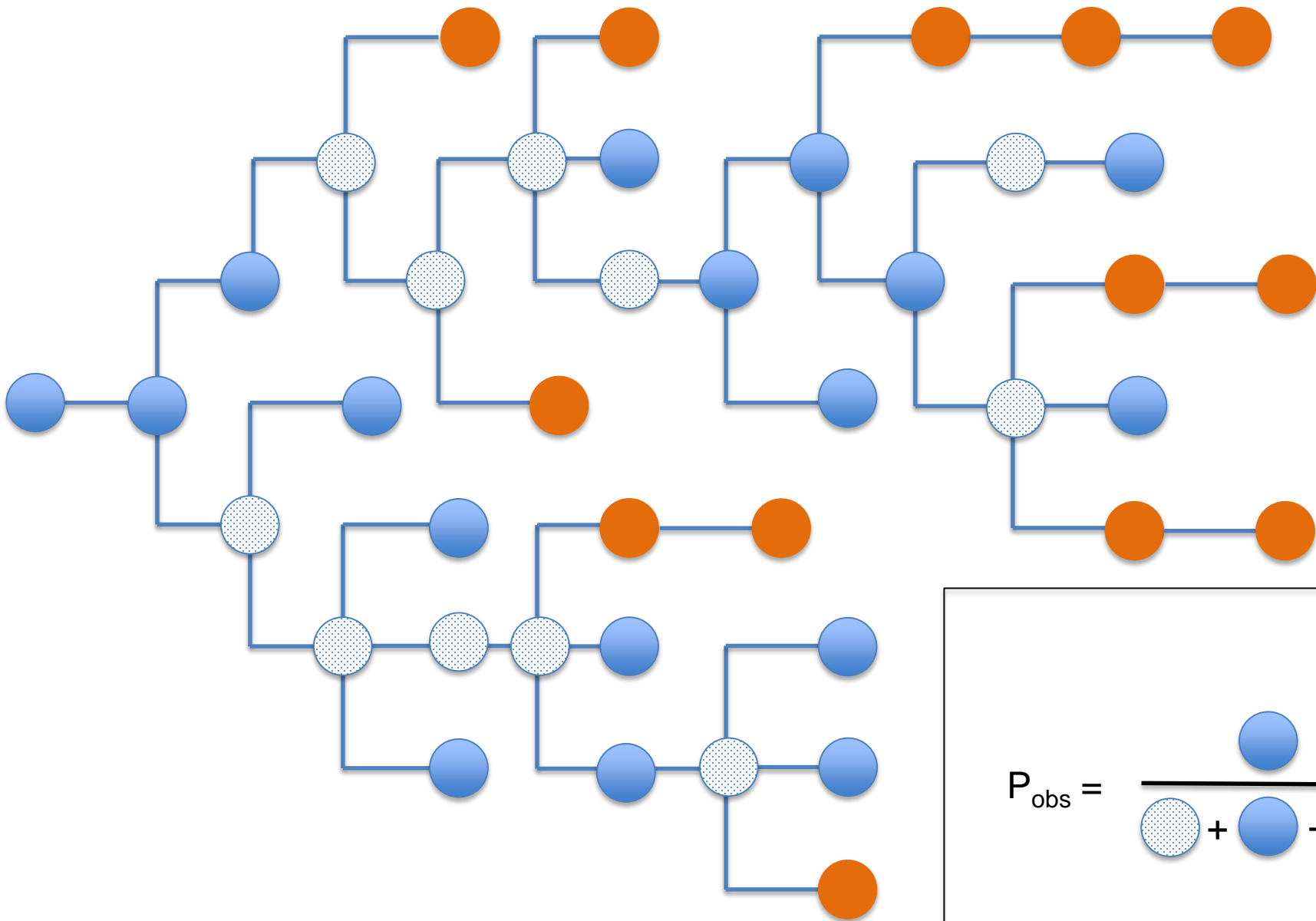


IPs 'in' tree = 1150

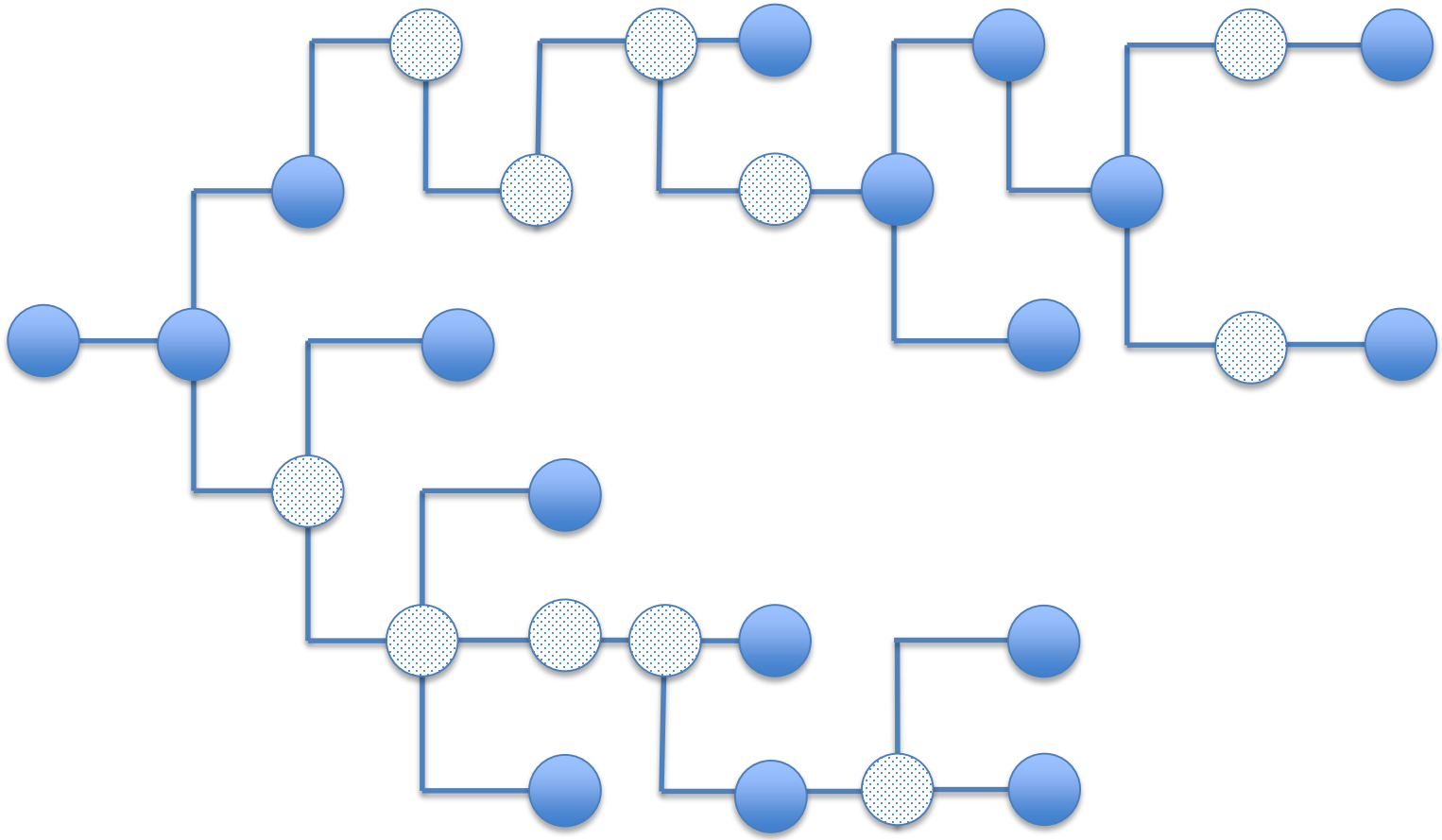




$$P_{\text{obs}} \neq \frac{\text{blue node}}{\text{dotted node} + \text{blue node}}$$



$$P_{\text{obs}} = \frac{\text{blue}}{\text{dotted} + \text{blue} + \text{orange}}$$

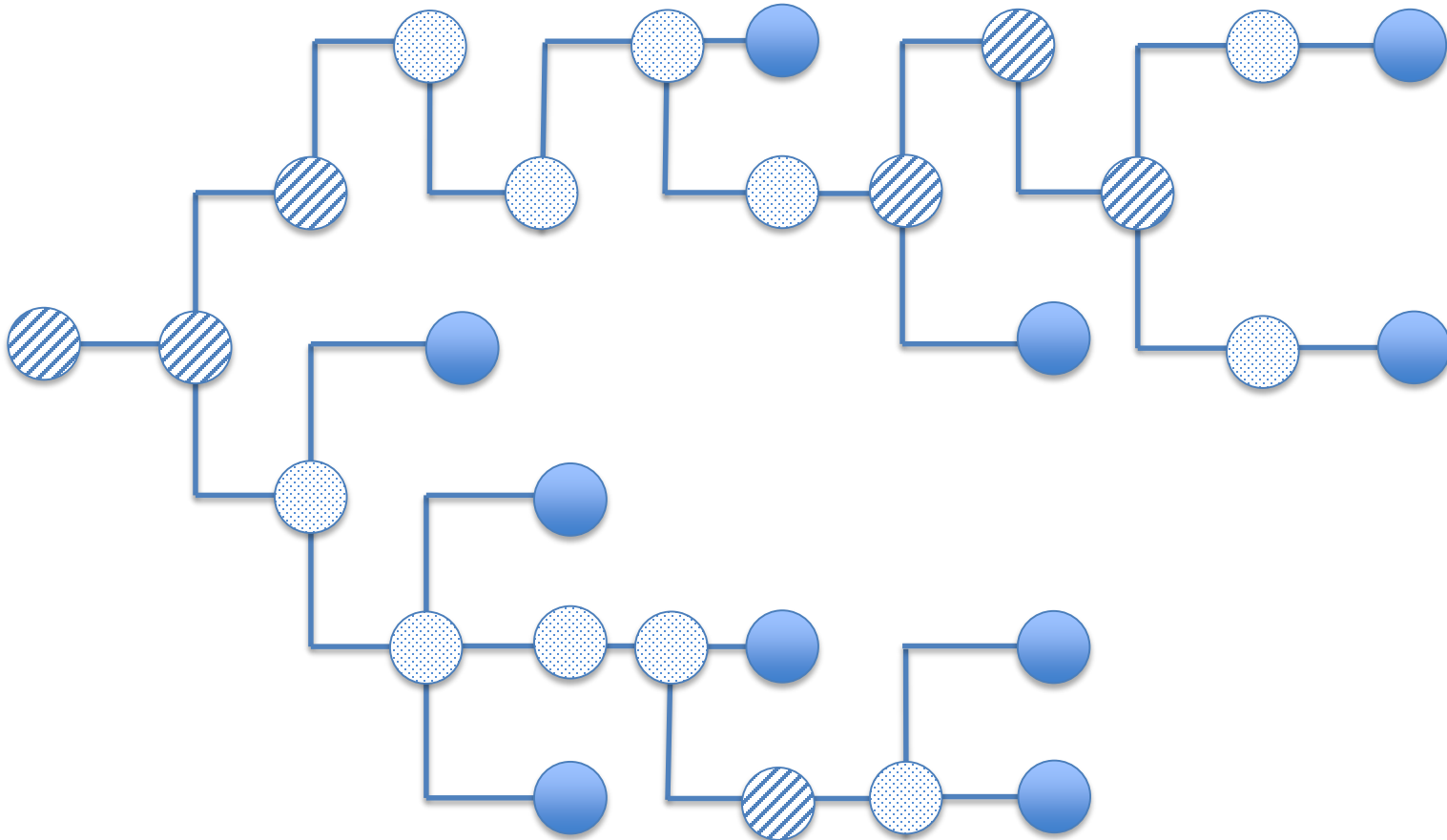




An observed IP that left observed descendants



An unobserved IP that left observed descendants

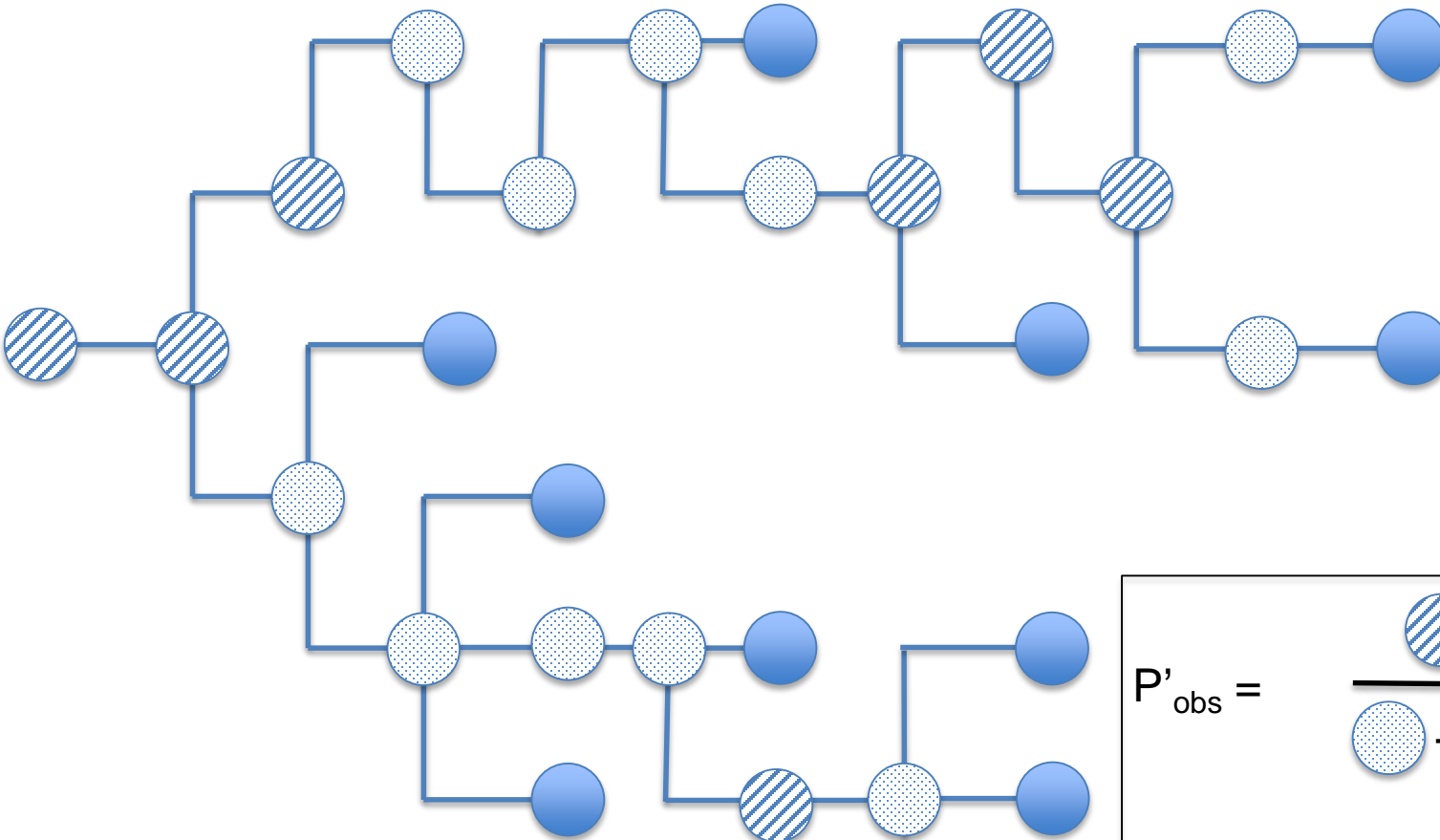




An observed IP that left observed descendants



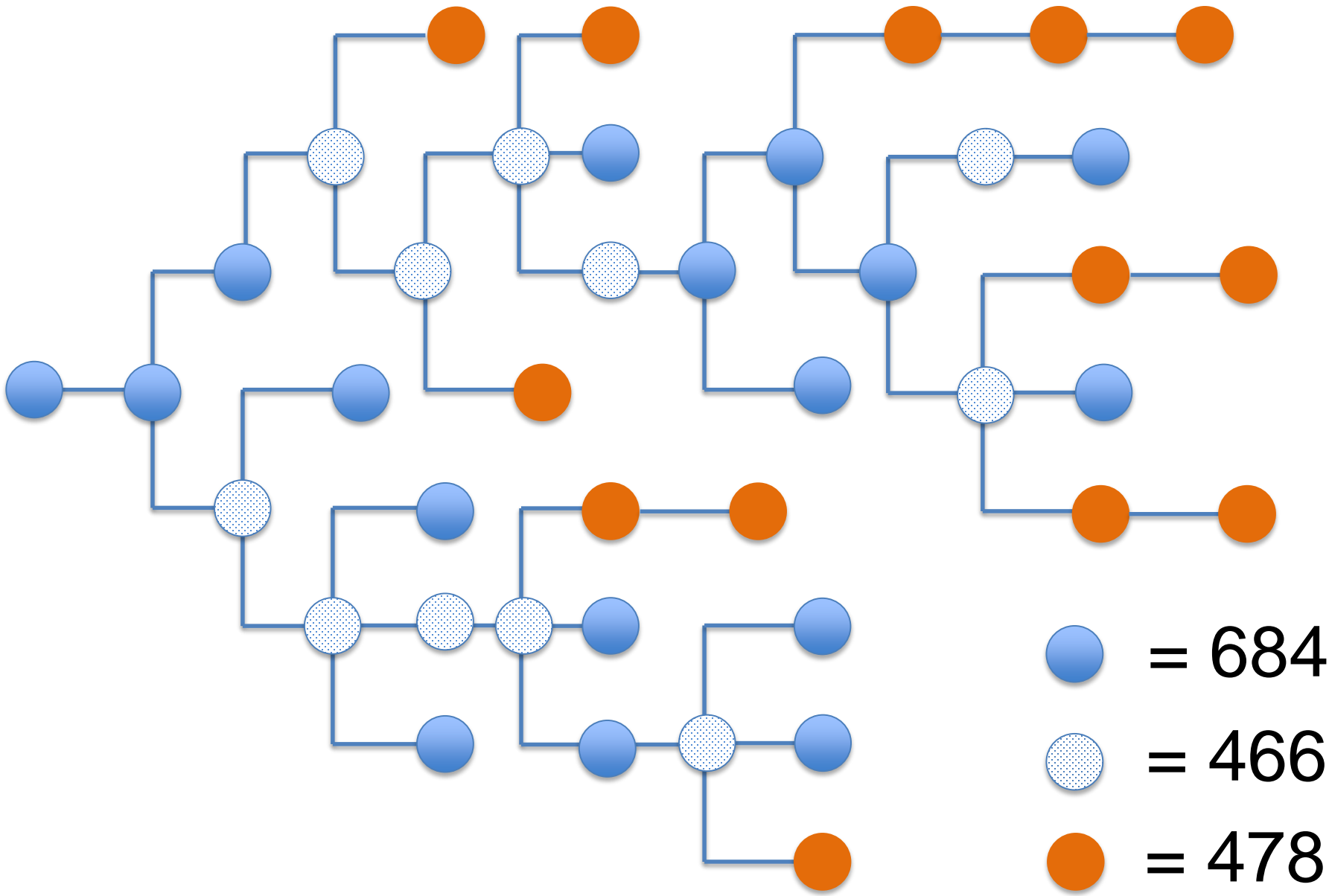
An unobserved IP that left observed descendants



$$P'_{\text{obs}} = \frac{\text{hatched circle}}{\text{dotted circle} + \text{hatched circle}}$$
$$P'_{\text{obs}} = \frac{0.49 \times 684}{0.49 \times 684 + 466} = 0.42$$

So the 684 cases we observed constitute 42% of the outbreak

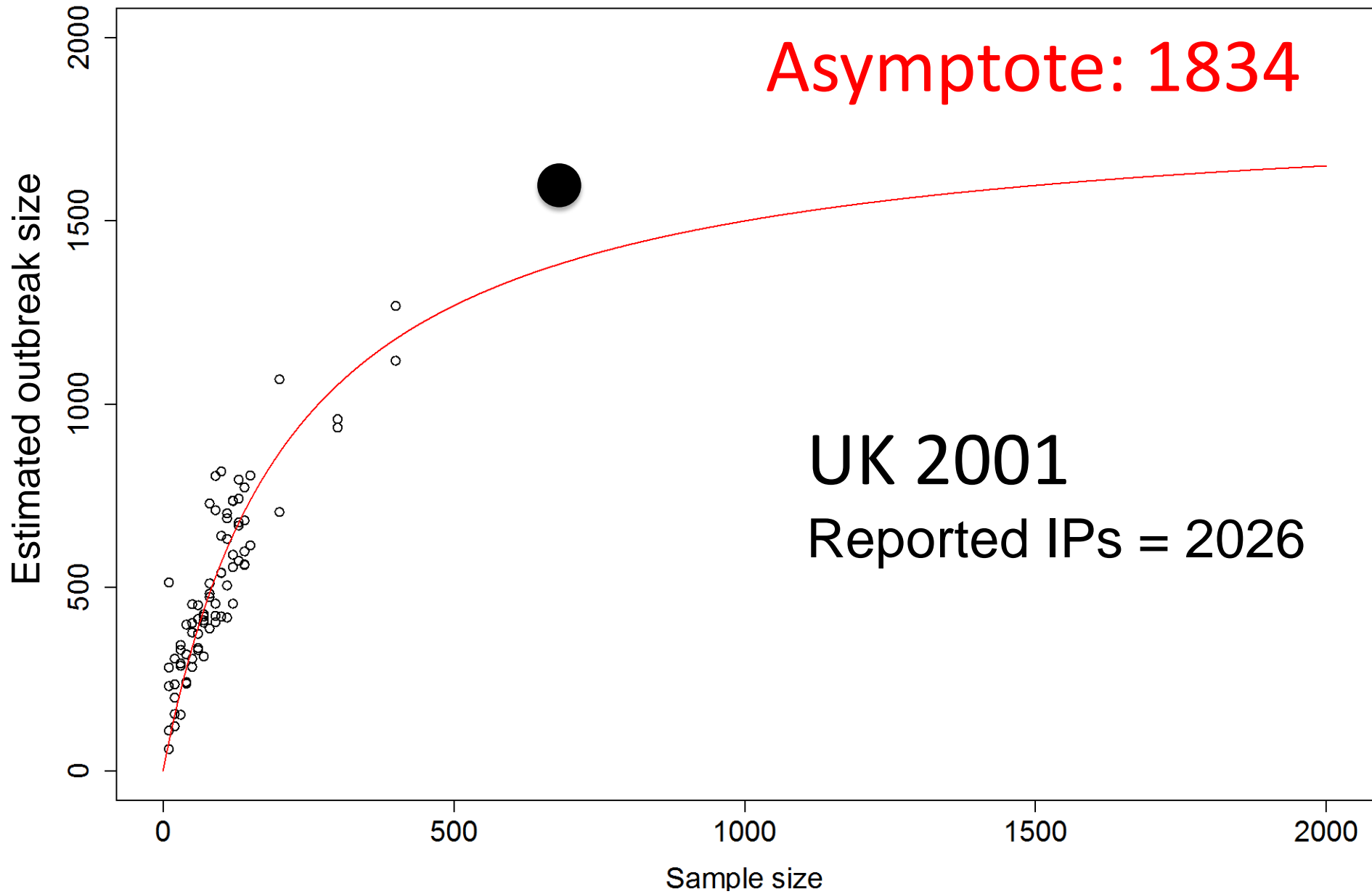
Thus we estimate the total outbreak size was ~ 1628

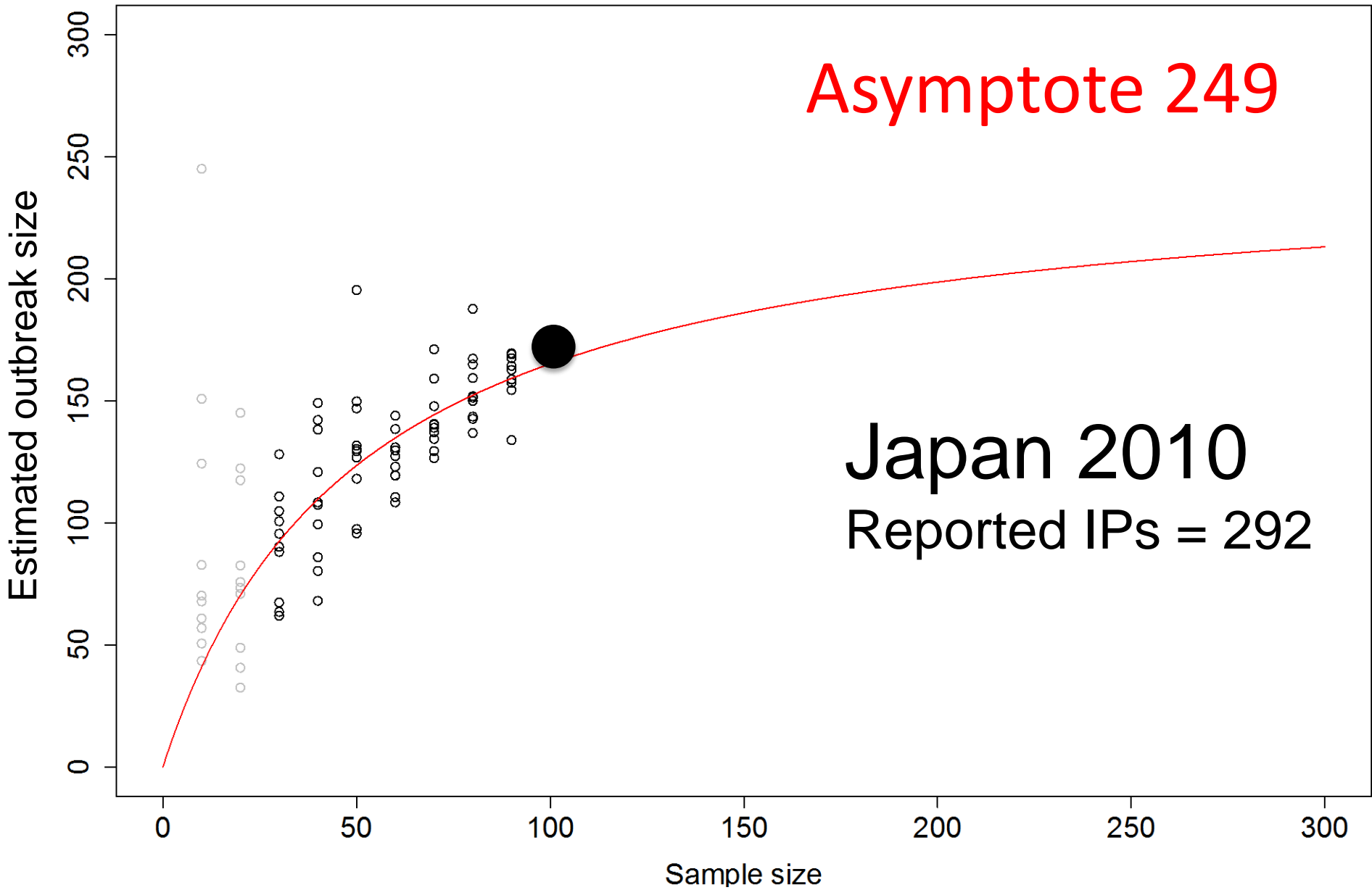


Outcomes

- We can estimate the number of mutations arising on IPs
- The distribution of ‘mother-daughter’ distances
- The proportion of IPs that left sampled descendants
- The minimum number of known unknowns ‘inside the tree’
- The unknown unknowns ‘outside of the tree’

But how sensitive is this estimate to
the sample size?
(the proportion of known knowns)





Asymptote 249

Japan 2010
Reported IPs = 292

Future Directions

- Understand more about the rarefaction dynamics
- How does tree length depend on sampling intensity?
- Sources of bias, & working with time-sliced data
- More parametric approaches based on branching structures

Mutations arise fairly uniformly over the genome

