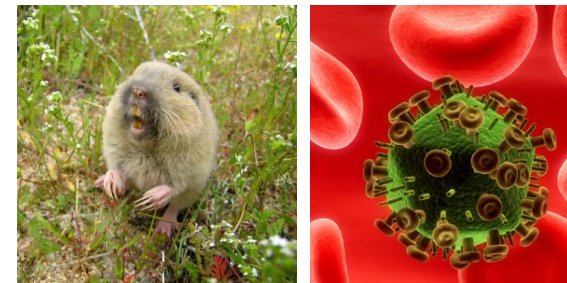
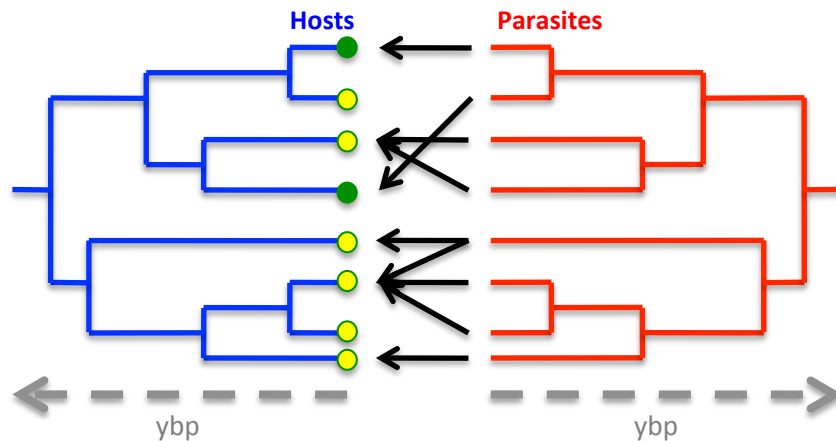


# Inferring patterns of parasite host-shifting



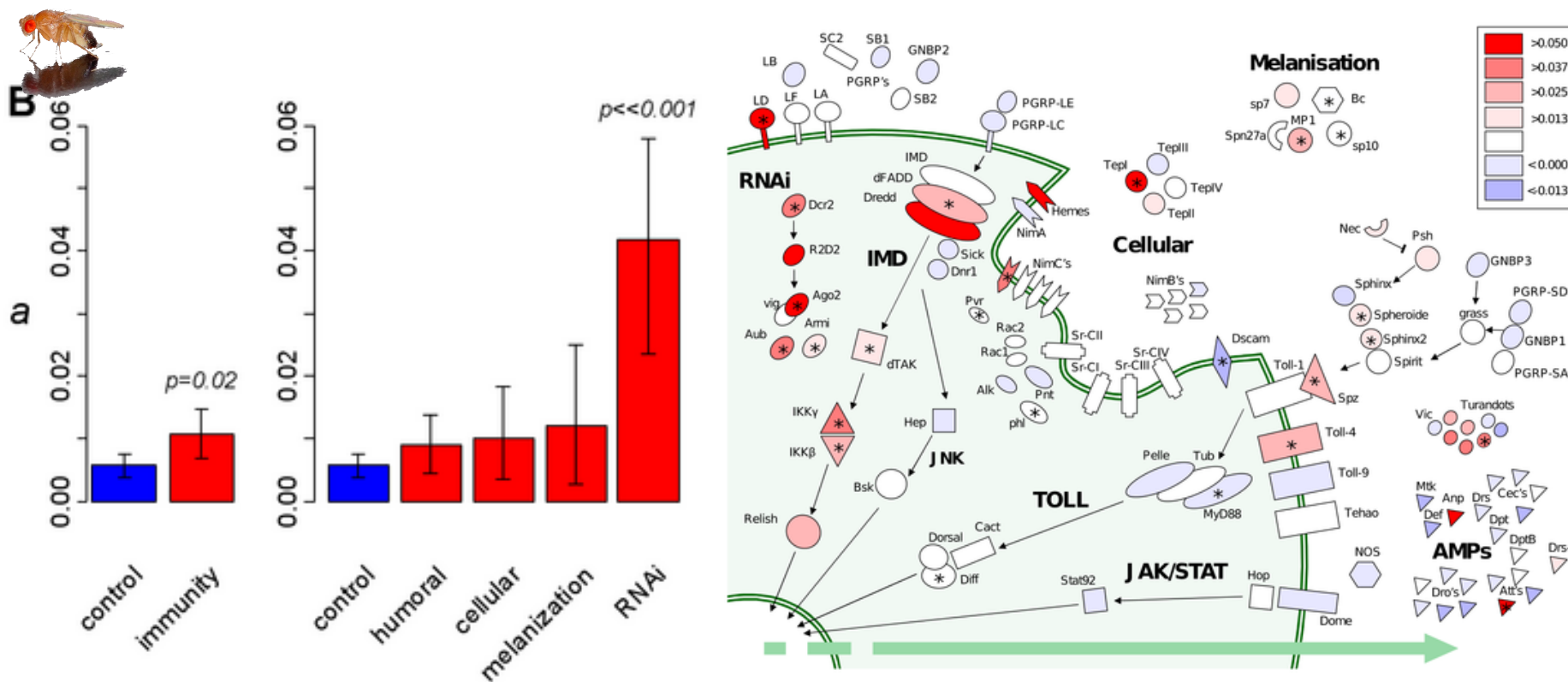
**John Welch**

[j.j.welch@gen.cam.ac.uk](mailto:j.j.welch@gen.cam.ac.uk)



# Host-parasite coevolution

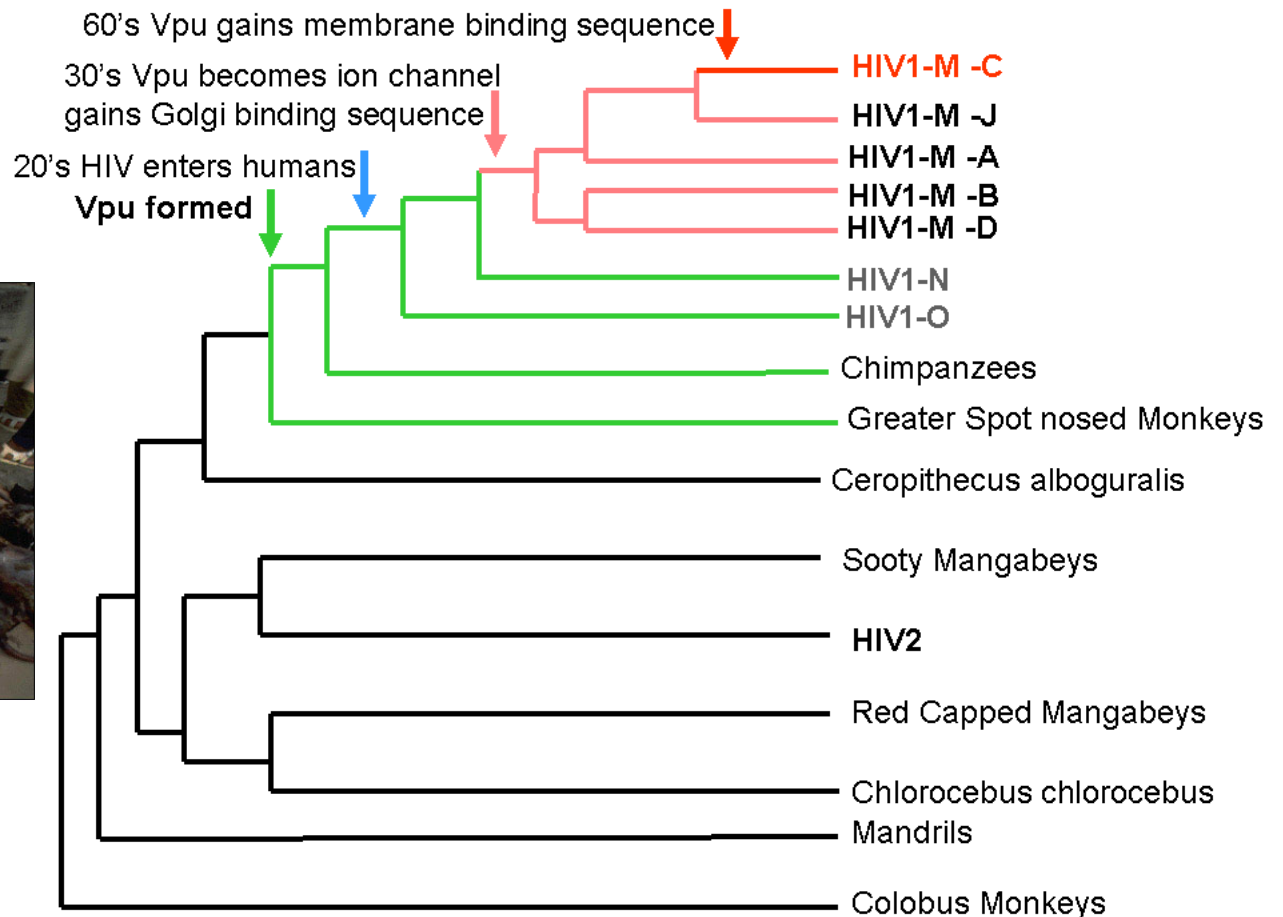
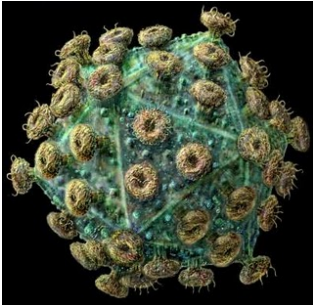
Parasites and pathogens implicated in much molecular evolution



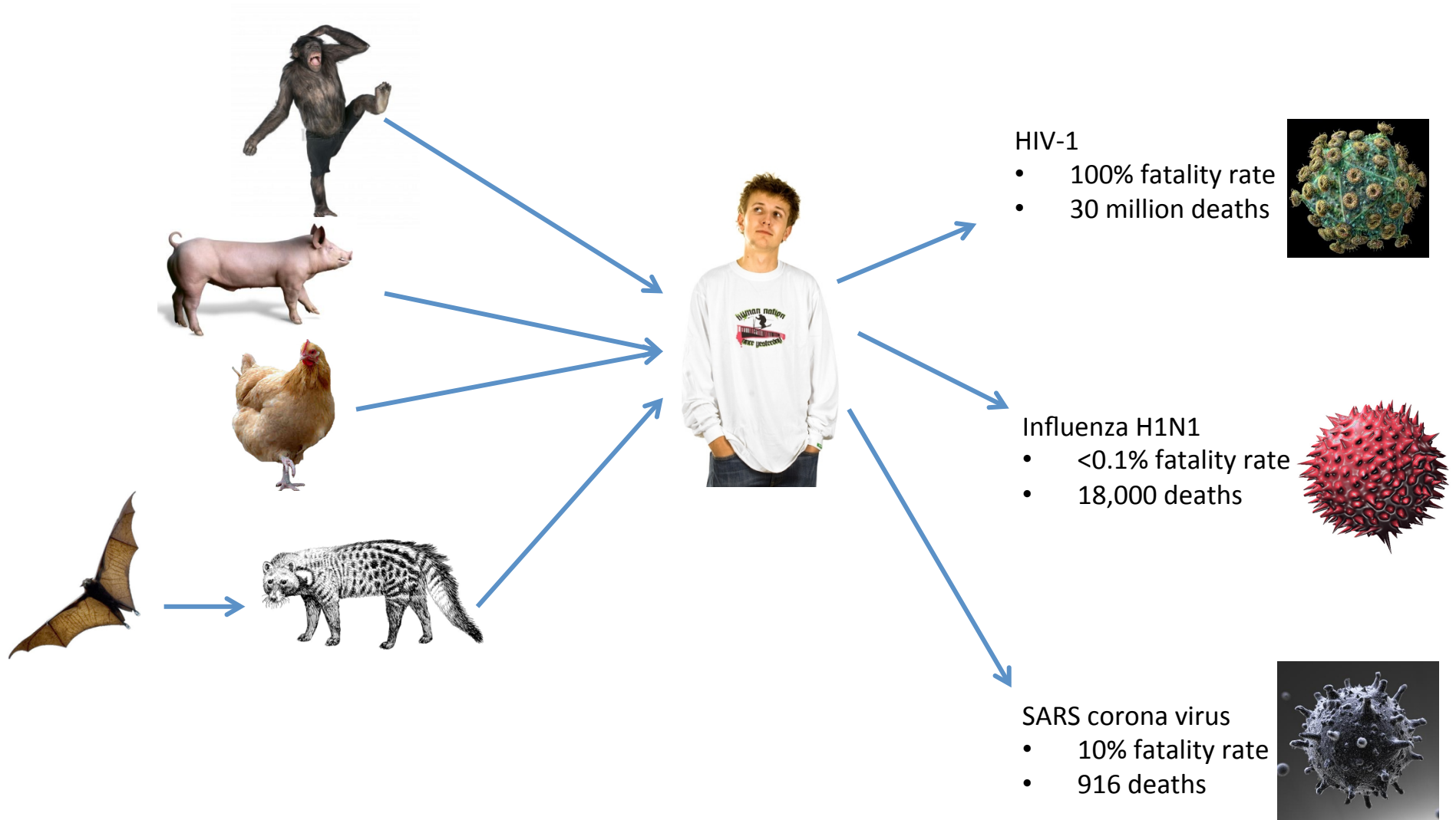
Obbard, Welch, Kim & F. Jiggins (PLoS Genetics 2009)

# Host shifts and emerging diseases

## Origins and spread of the HIV pandemic



# Host shifts and emerging diseases

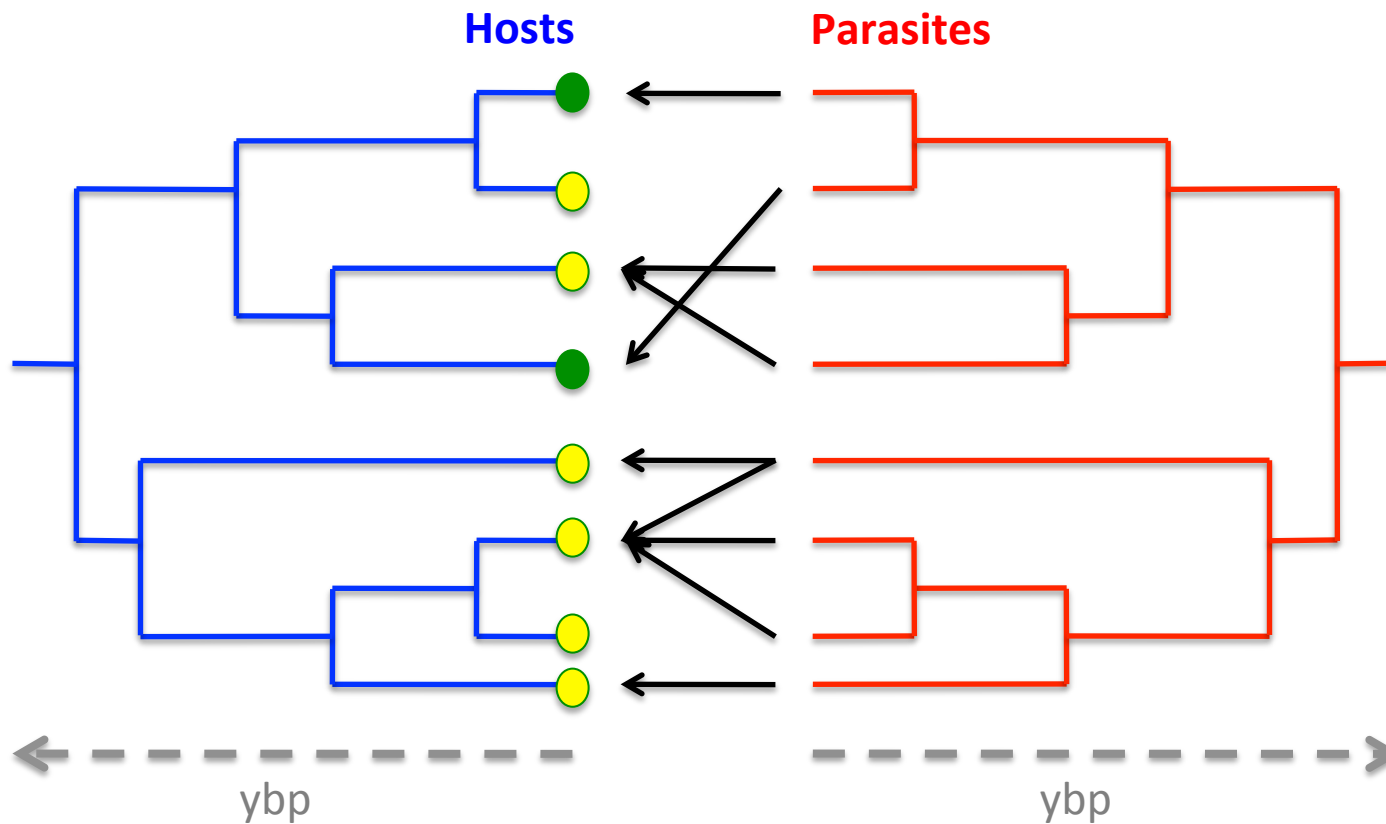


# The questions

**What are the genetic/ecological predictors of host-shift success?**

**What is the history of parasite-host associations in a given group?**

# The basic framework

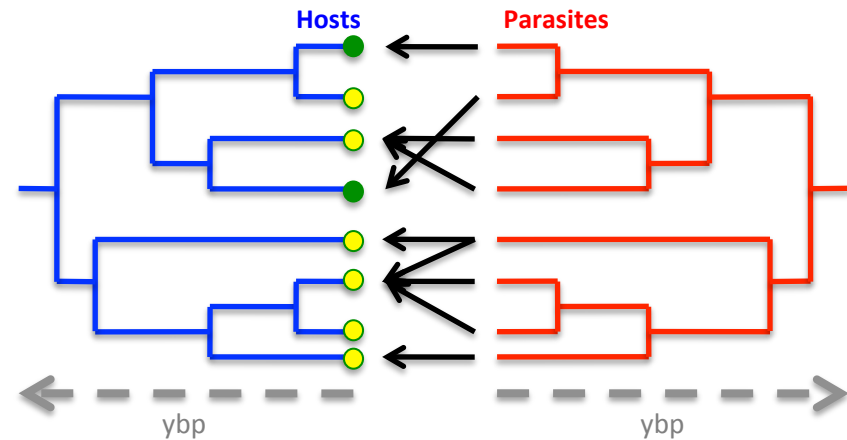


- Dated phylogeny of the **host** taxa
- Dated phylogeny of the **parasite** taxa
- Contemporary associations of **hosts** and **parasites**
- Other ecological variable(s)

# Outline

What are the genetic/ecological predictors of host-shift success?

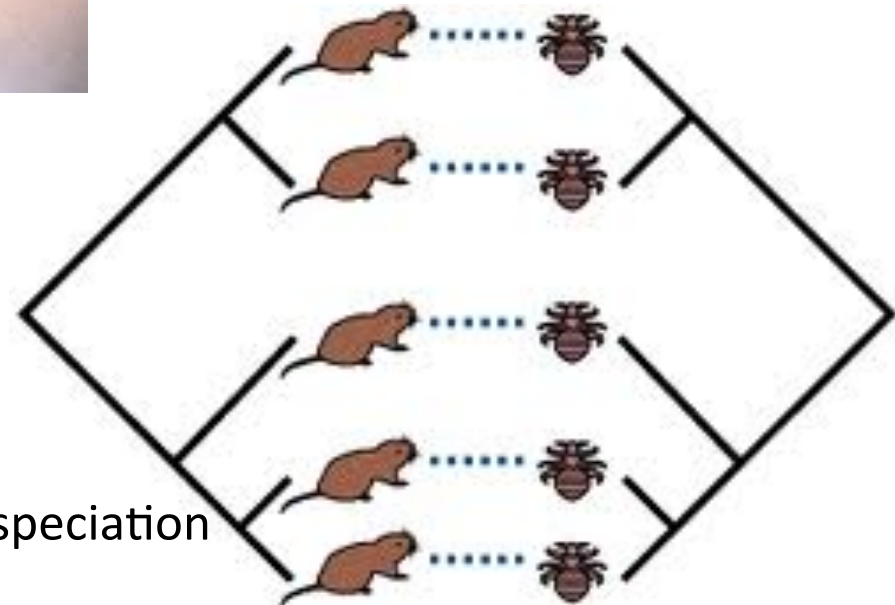
What is the history of parasite-host associations in a given group?



1. Tests for complete congruence
2. Tests for topological similarity
3. Cophylogeny reconstruction
4. Ancestral host state reconstruction
5. Incidence dynamics
6. Tree as predictor variable

# 1. Tests for complete congruence

Pocket gophers and chewing lice



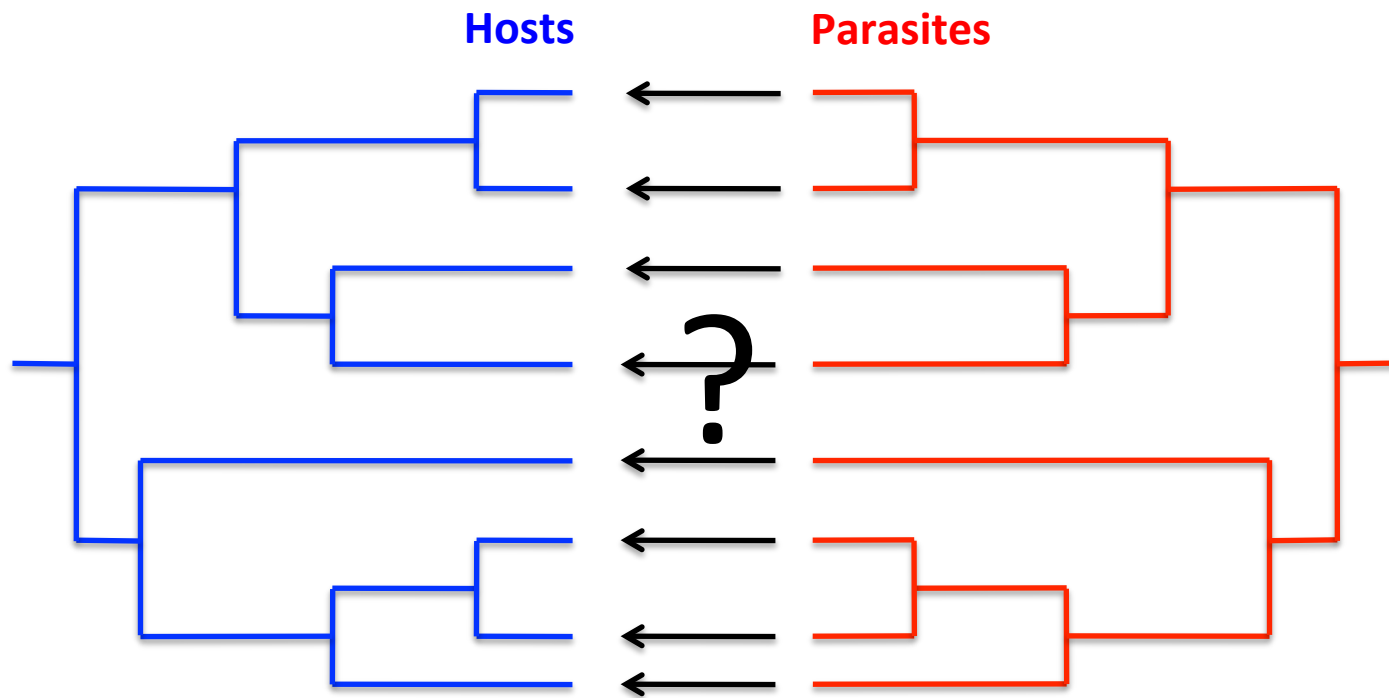
Suggestive of vertical transmission and cospeciation

(Fahrenholz 1913)



# 1. Tests for complete congruence

Are topologies the same?



# 1. Tests for complete congruence

- **Does a single topology provide a significantly worse explanation of data?**

*Incongruence Length Difference test*

Test statistic for Maximum Parsimony trees:

$$\delta = \text{Length}_{\text{Concat}} - (\text{Length}_{\text{Host}} + \text{Length}_{\text{Parasite}})$$

Null distribution by random assignment of sites to “host” or “parasite” genomes.

- Susceptible to process heterogeneity (e.g., different rates)
- Localised versions of the test available

(Mickey & Farris 1981; Hafner et al. 1994; Farris et al. 1994; Dolphin et al. 2000; Thornton & DeSalle 2000; Darlu & Lecointre 2002; Planet 2006)

# 1. Tests for complete congruence

- **Does a single topology provide a significantly worse explanation of data?**

*Incongruence Length Difference test*

Test statistic for Maximum Parsimony trees:

$$\delta = \text{Length}_{\text{Concat}} - (\text{Length}_{\text{Host}} + \text{Length}_{\text{Parasite}})$$

Null distribution by random assignment of sites to “host” or “parasite” genomes.

*Likelihood ratio test*

$$\delta = \ln L_{\text{Concat}} - (\ln L_{\text{Host}} + \ln L_{\text{Parasite}})$$

Parametric bootstrap to obtain the null distribution

*Bayesian approach*

One and two-topology models compared by estimating marginal log likelihoods from the harmonic means of the log likelihood in the MCMC run.

# 1. Tests for complete congruence

- Does a single topology provide a significantly worse explanation of data?
- What is the probability that the two trees have different topologies?  
Use an MCMC posterior sample of trees

$$p = 1 - \sum_t \Pr_H(t) \Pr_P(t)$$

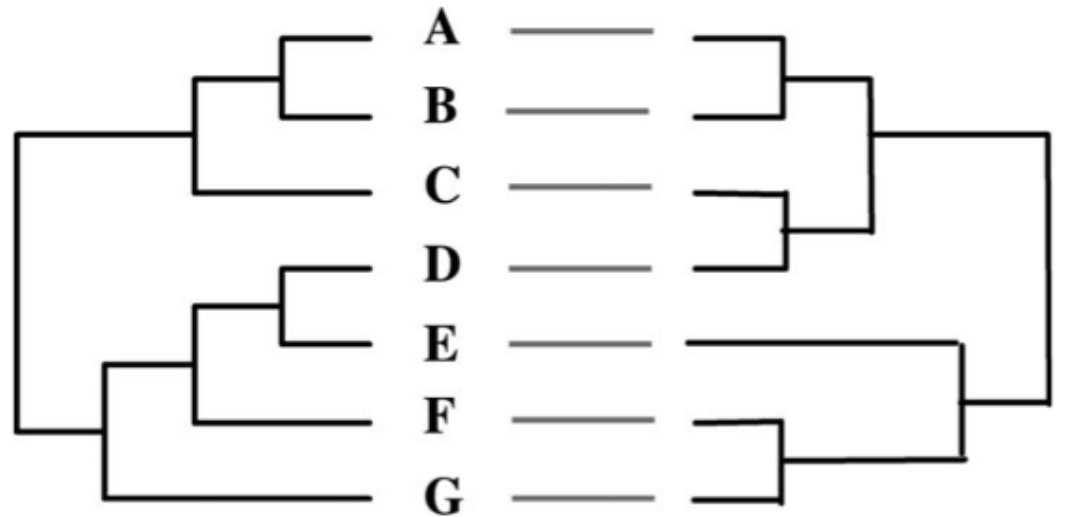
- Does ML parasite tree differ significantly from the ML host tree?
  - Non-parametric: *SH test*  
(are all of a set of topologies are equally good explanations of the data?)
  - Parametric bootstrap: *SOWH test*  
(simulate parasite-like data under the host topology)

## 2. Tests for topological similarity

### Tanglegrams



Pocket gophers



Chewing lice



## 2. Tests for topological similarity

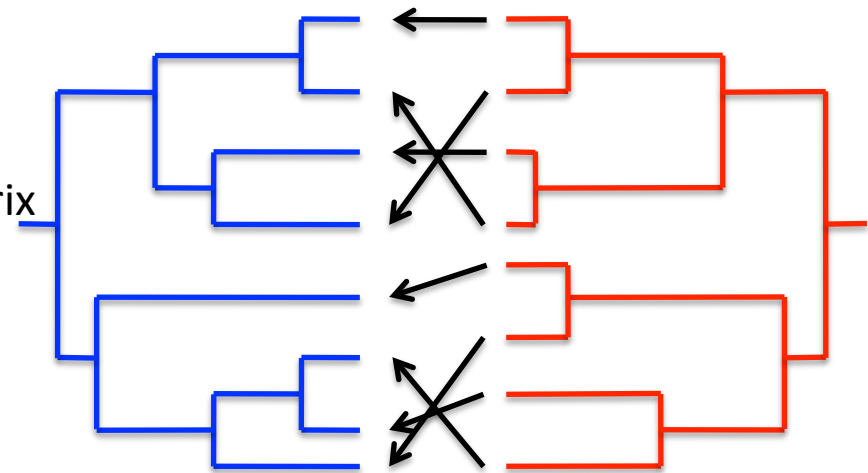
Are topologies more similar than chance?

### One-to-one host-parasite relationships

#### *Mantel Test*

Calculate Pearson's correlation coefficient for entries of distance matrix implied by the phylogenies.

Permute tree labels to obtain null distribution.



$$\text{Cor}(h_{ij}, p_{ij})$$

## 2. Tests for topological similarity

Are topologies more similar than chance?

### Many-to-many host-parasite relationships

#### *ParaFit*

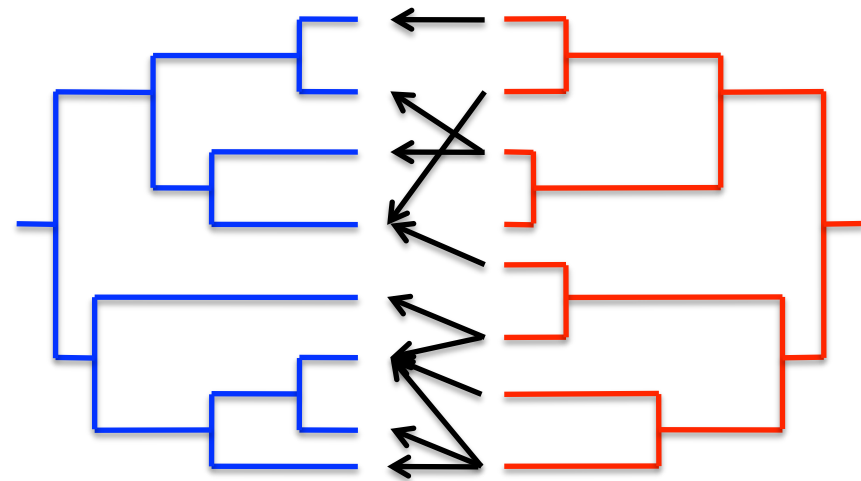
Transform host and parasite distances into matrices of principal coordinates

Test statistic is  $\text{Tr}(D'D) = \text{sum of squares of all entries in:}$

Rows in  $A$  permuted (hosts randomly associated with parasites)

$$D = C A^T B$$

Can test for individual links



$A$  binary association matrix

$C$  host tree with pcs on rows

$B$  parasite tree with pcs in columns

## 2. Tests for topological similarity

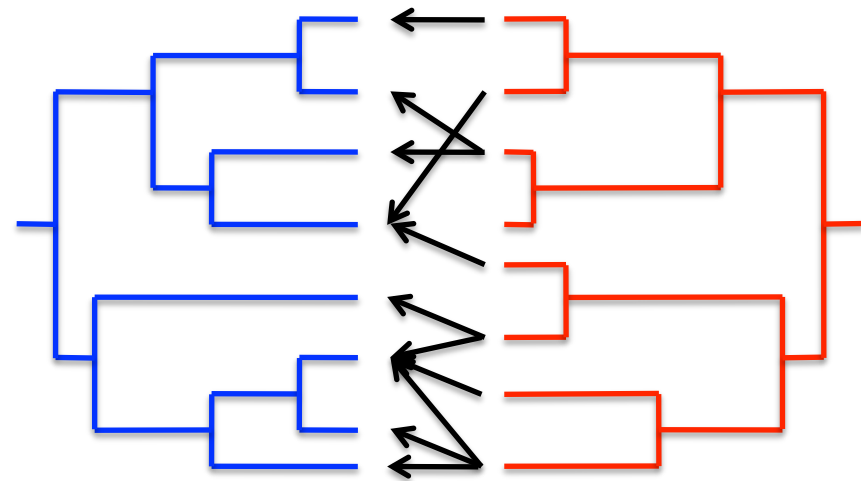
Are topologies more similar than chance?

### Many-to-many host-parasite relationships

Correlation coefficient between host and parasite distances for each pair of host-parasite links

Permute host and parasite labels, retaining associations

More power than *ParaFit*, unbiased unlike other possible extensions of the Mantel test



$$\text{Cor}(h_{ij}, p_{ij})$$

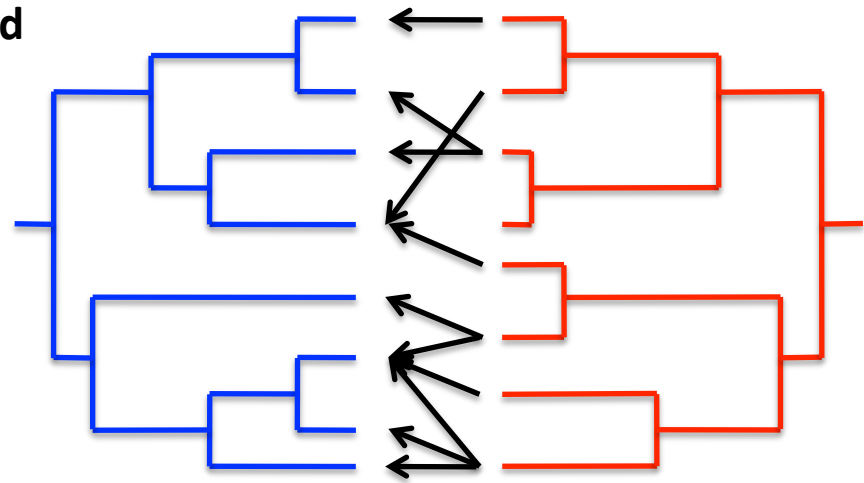


## 2. Tests for topological similarity

Are topologies more similar than chance?

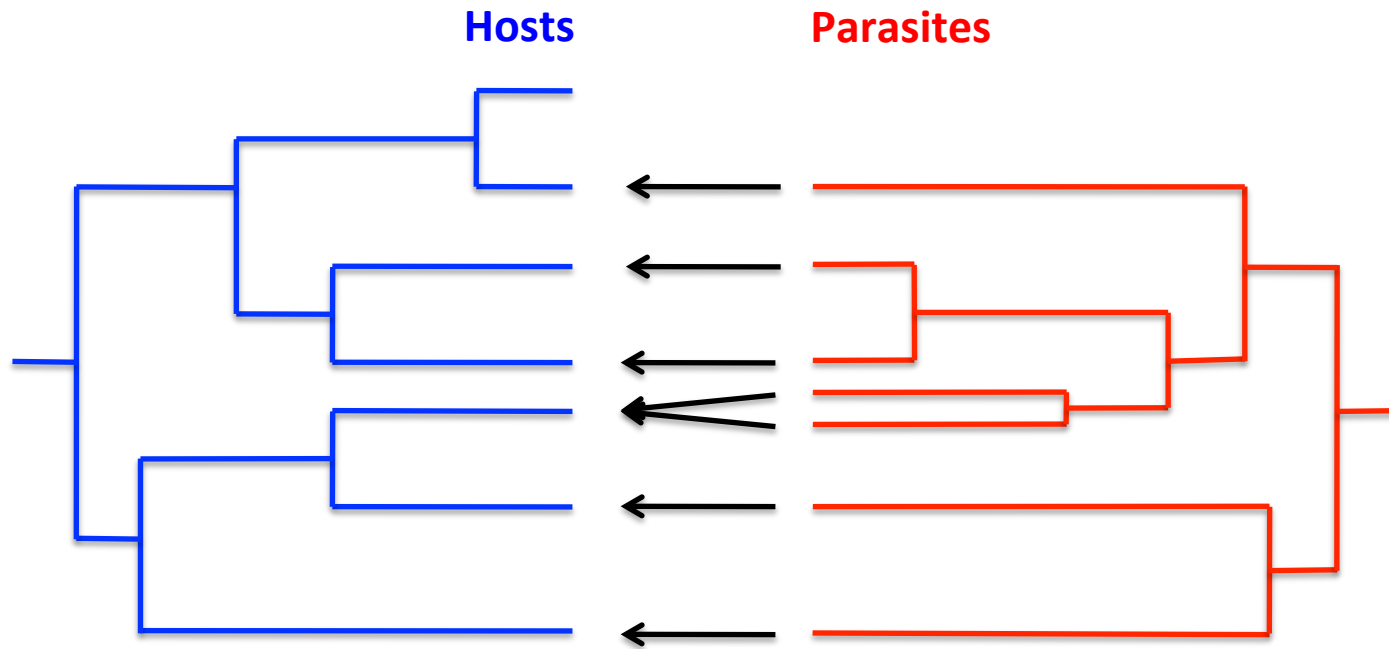
There are any other “consensus indices” and “tree distance metrics”

Often with known null distributions



# 3. Cophylogeny reconstruction

How do the trees relate to each other?

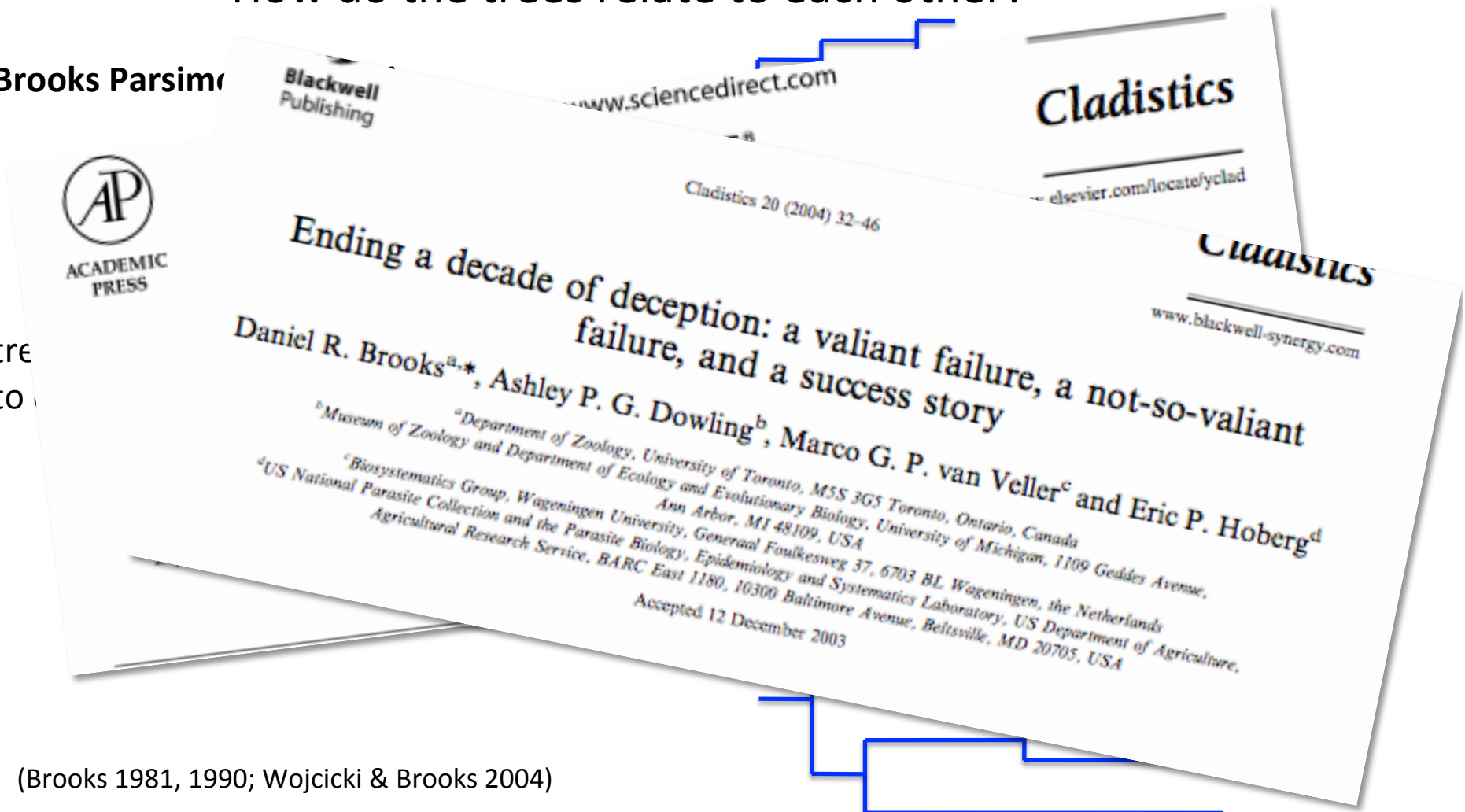


# 3. Cophylogeny reconstruction

How do the trees relate to each other?

Brooks Parsim

tre  
to



(Brooks 1981, 1990; Wojcicki & Brooks 2004)

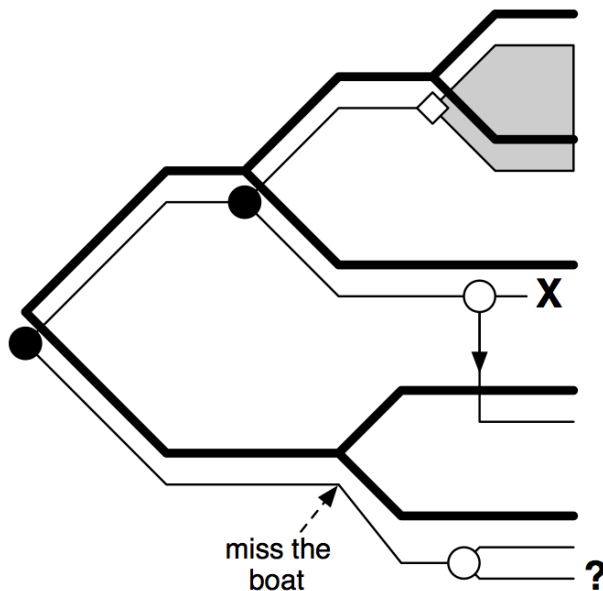
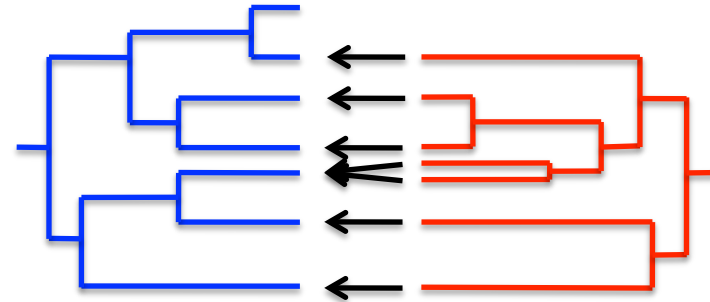
# 3. Cophylogeny reconstruction

How do the trees relate to each other?

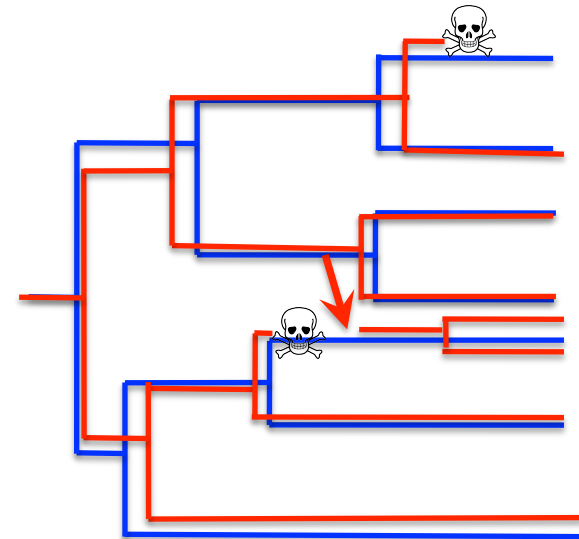
## Events-based methods

Assume trees are known with certainty

Map **parasite** tree onto **host** tree by posing different types of event



- codivergence
  - duplication
  - ↘ host switch
  - X extinction
  - ? sampling failure
  - ↘ miss the boat
  - ◇ failure to diverge
- loss



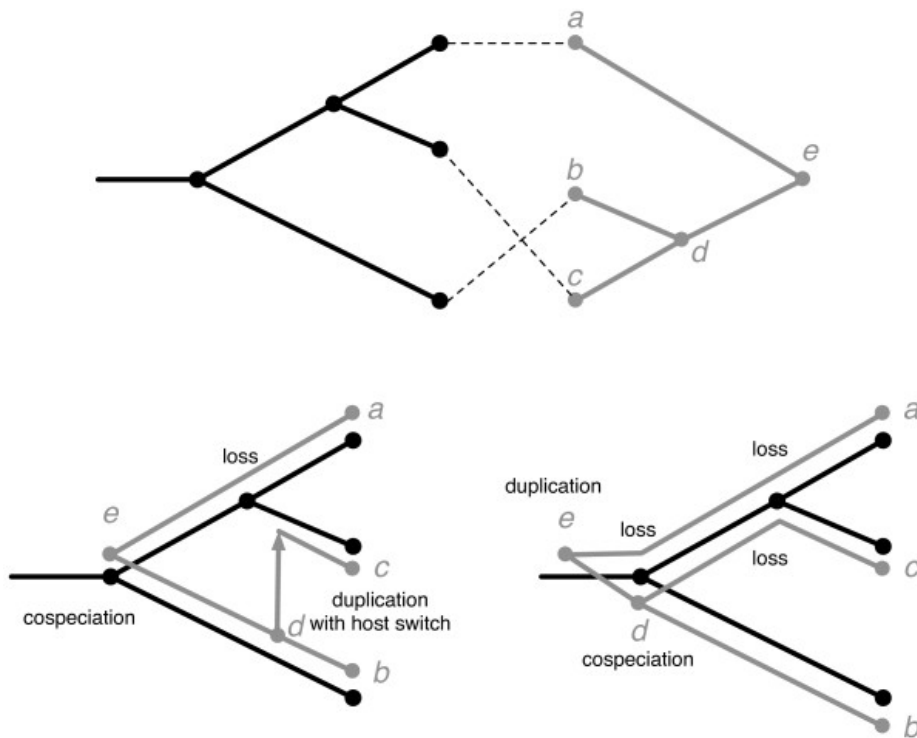
# 3. Cophylogeny reconstruction

How do the trees relate to each other?

## Events-based methods

Possible events mapped using parsimony, or explicit models

weighted



Change the cost table  
Set the cost values...

Cospeciation cost:	0
Duplication cost:	2
Sorting cost:	2
Host Switch cost:	3
Extinction cost:	2

#Costs	Minimize
Nothing	Maximize
Nothing	Minimize
Nothing	Minimize
Nothing	Minimize
Nothing	Minimize

Ok Cancel

# 3. Cophylogeny reconstruction

How do the trees relate to each other?

## Events-based methods

*Treefitter*: Parsimony based.

*TreeMap 2.02*: using *Jungles*, finds optimal solutions.

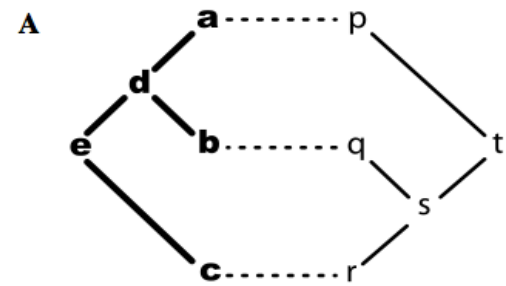
Exponential running time in the worst case

*Tarzan*: very fast. Uses divergence times, but can give inconsistent solutions

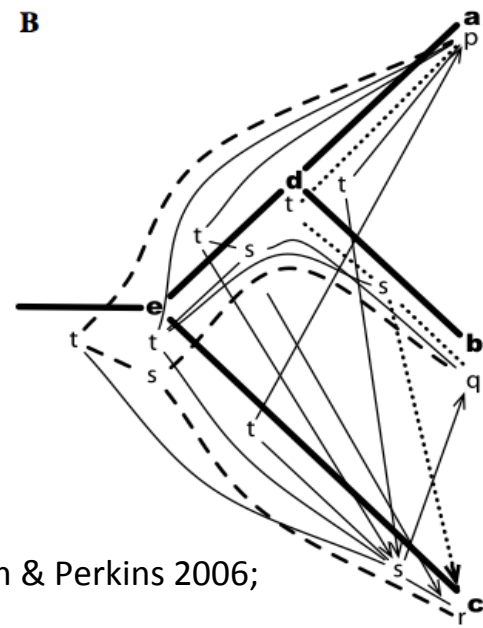
*Jane*: Genetic algorithm + dynamic programming etc. etc.



optimal



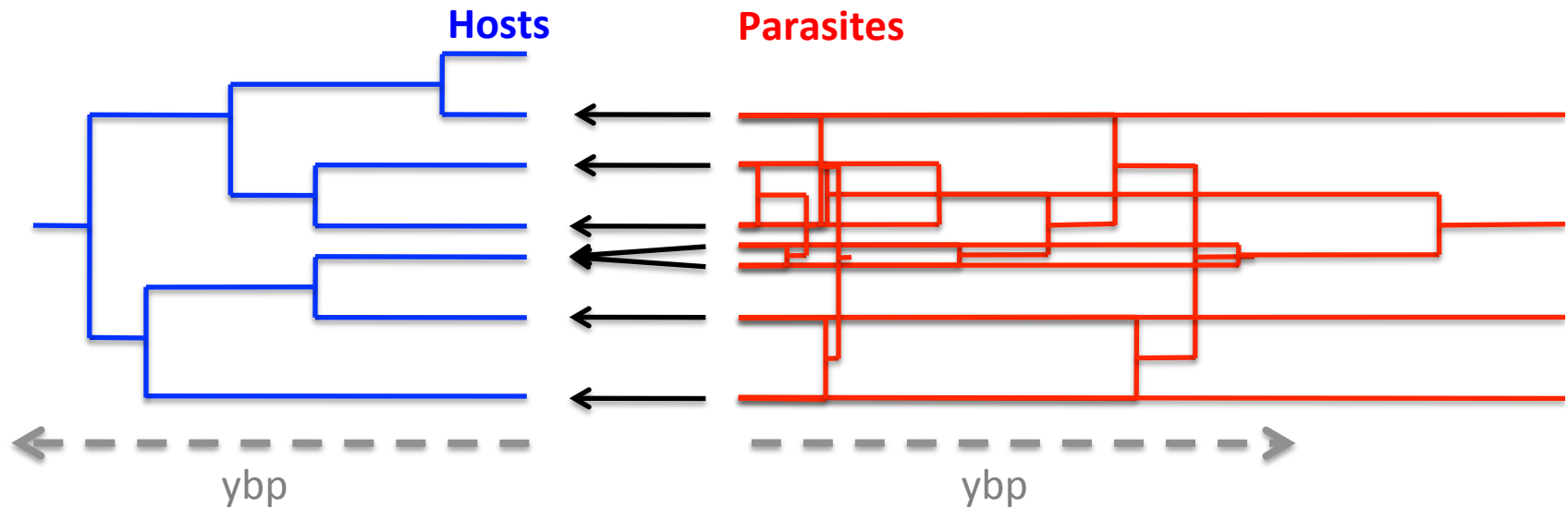
0,12]



(Charleston 1998; Ronquist 2002; Merkle & Middendorf 2005; Charleston & Perkins 2006; Huelsenbeck et al. 1997, 2000; Conow et al. 2010)

# 3. Cophylogeny reconstruction

Co-speciation or switching to related hosts?

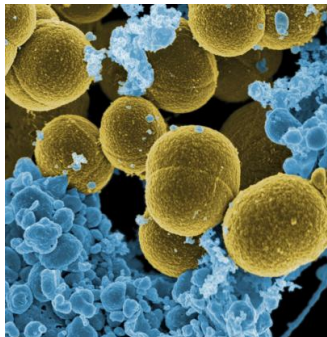


(Charleston & Robertson 2002; Merkle & Middendorf 2005; Shah et al. 2010)

# 4. Ancestral host state reconstruction

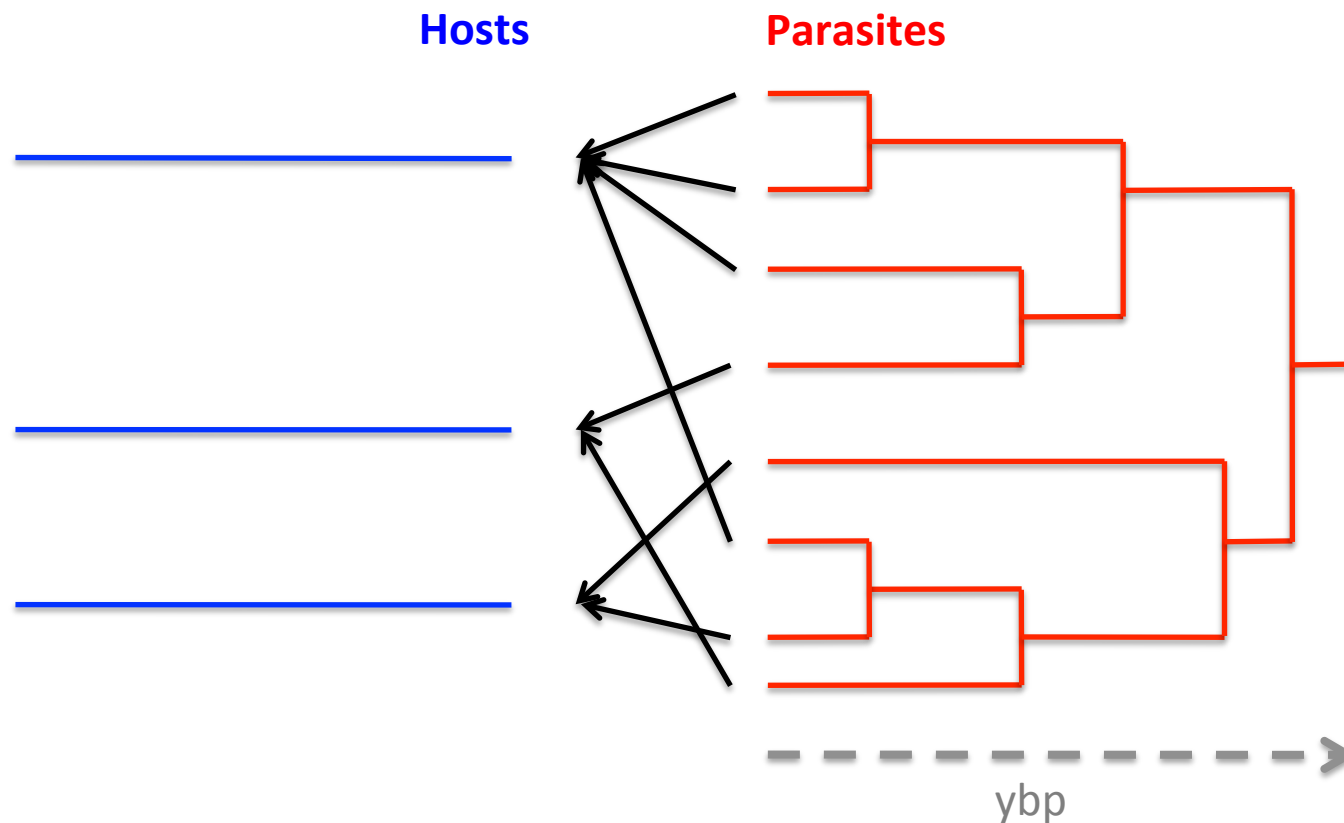
## *Staphylococcus aureus*

- Pathogenic MRSA outbreak in hospitals
- Nasal-carriage in humans
- Categorized into multiple Strain Types
- Many STs found in non-human hosts
- Molecular evidence for human to animal jumps
  - Inactivation of genes involved in human pathogenicity
  - Acquisition of MGE from other animal strains





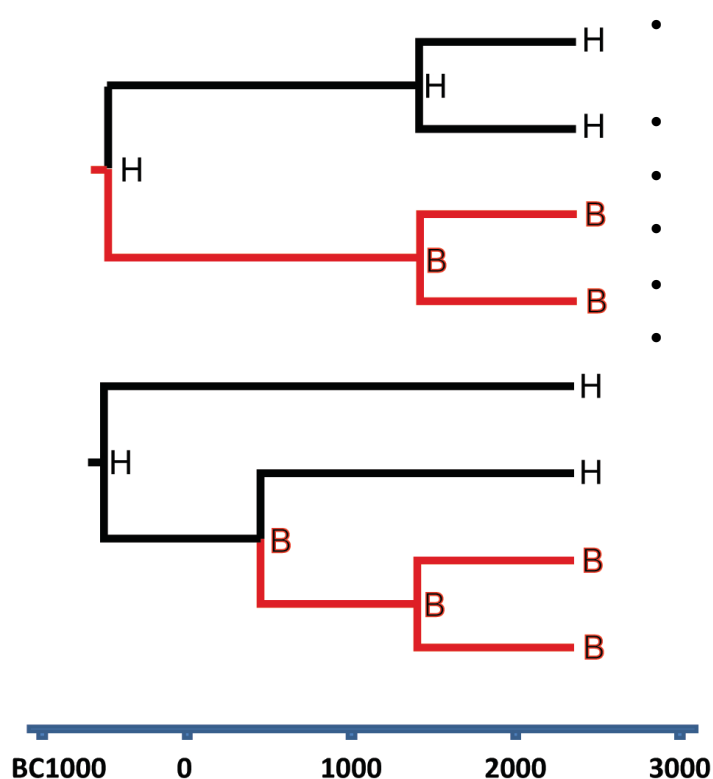
# 4. Ancestral host state reconstruction









- Parasite radiation rapid relative to that of hosts
- Each host associated with multiple parasites
- (host phylogeny ignored/unknown)

# 4. Ancestral host state reconstruction

## Bayesian ancestral host state reconstruction



- Host-switching modeled as a continuous-time Markov Chain with **non-reversible** instantaneous rate matrix
- Bayesian stochastic variable selection possible for large matrices
- Can fix certain node states *a priori*.
- Uses Markov Jumps Methods to estimate number of host jumps
- Implemented in BEAST using BEAGLE
- Estimates parasite topology, variable substitution rates, divergence dates, ancestral host states and number of host switches in a single simultaneous analysis.

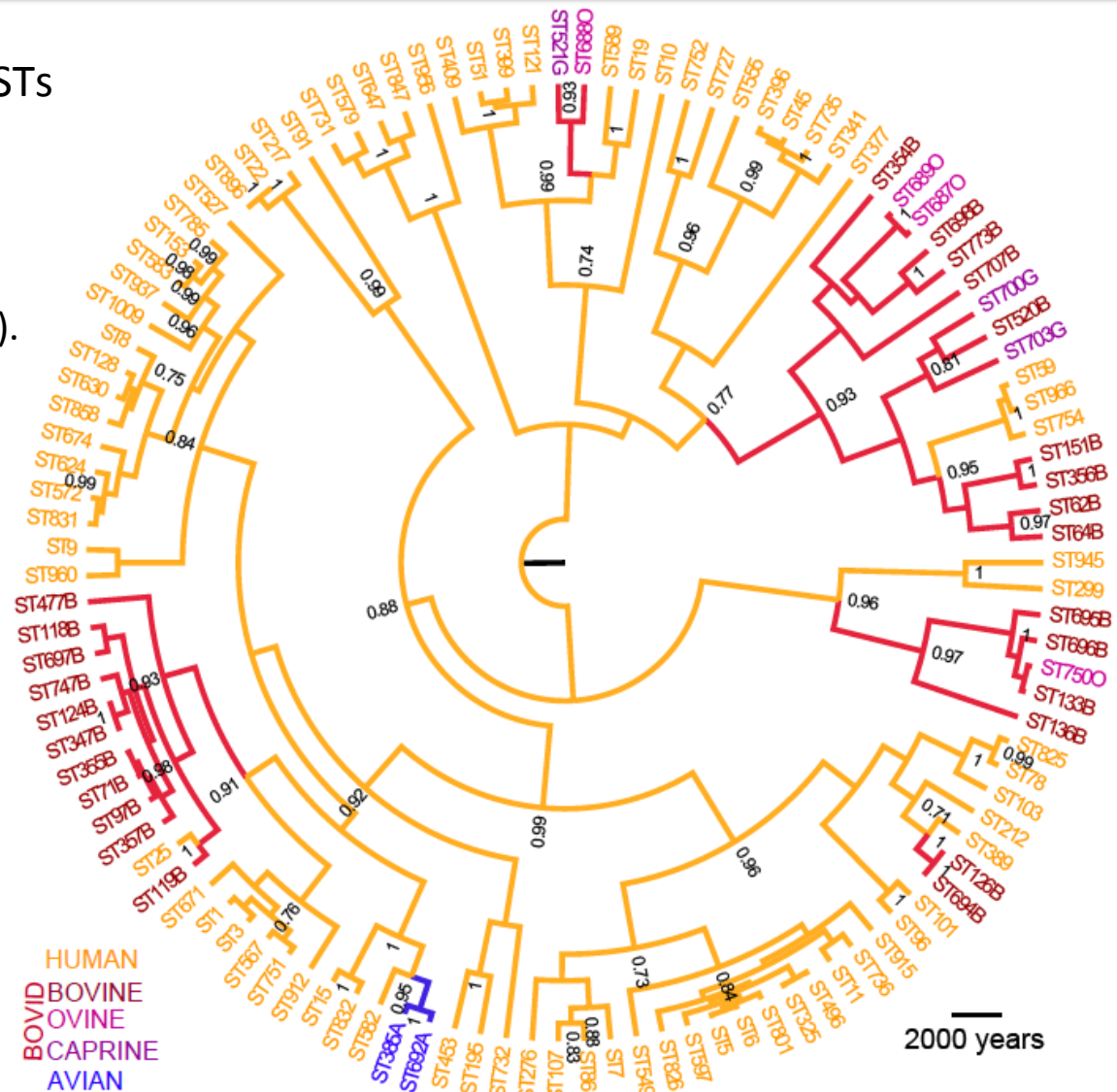
Host State	A 	B 	C 
A 	--	$q_{AB}$	$q_{AC}$
B 	$q_{BA}$	--	$q_{BC}$
C 	$q_{CA}$	$q_{CB}$	--

(Lemey *et al.* 2009; Edwards *et al.* 2011; Weinert *et al.* *In review*)

# 4. Ancestral host state reconstruction

MLST data from 115 *S. aureus* STs

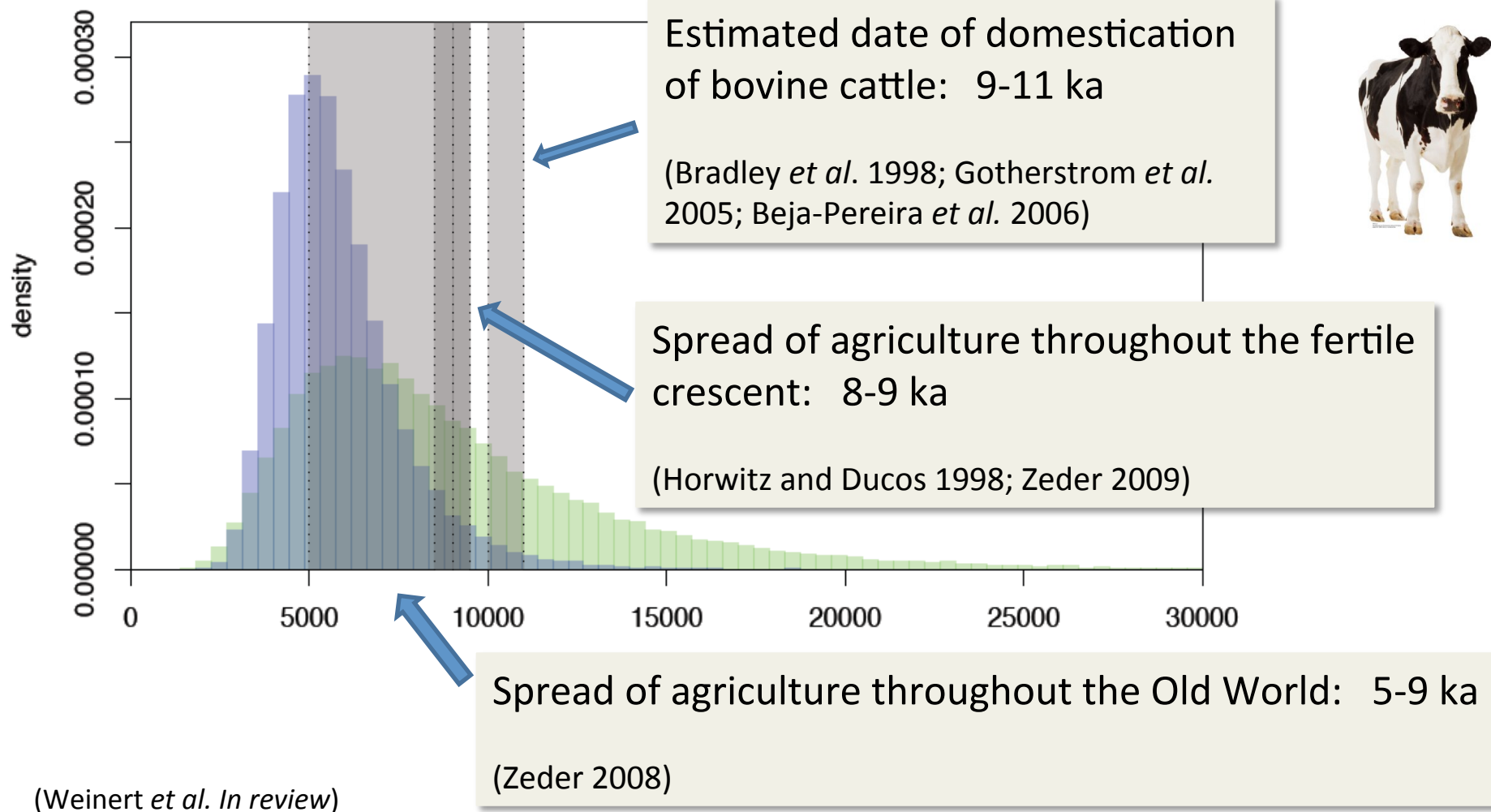
Human to bovid host jumps:  
Median 5.34  
(95% Bayesian CI: 4.95–6.89).



(Weinert *et al.* In review)

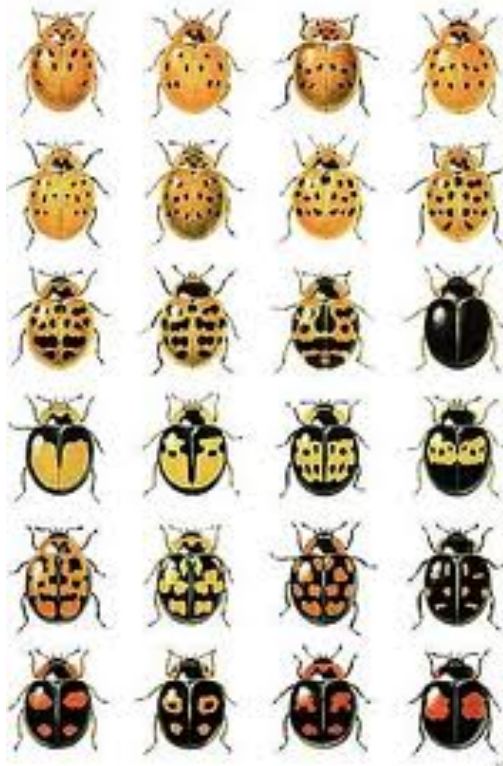
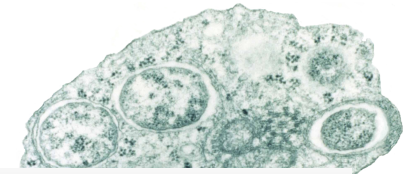
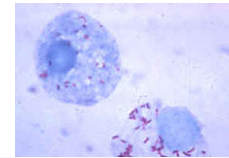
# 4. Ancestral host state reconstruction

Dating the first human-to-bovine host jump



# 5. Incidence dynamics

Parasitic intracellular endosymbionts of arthropods



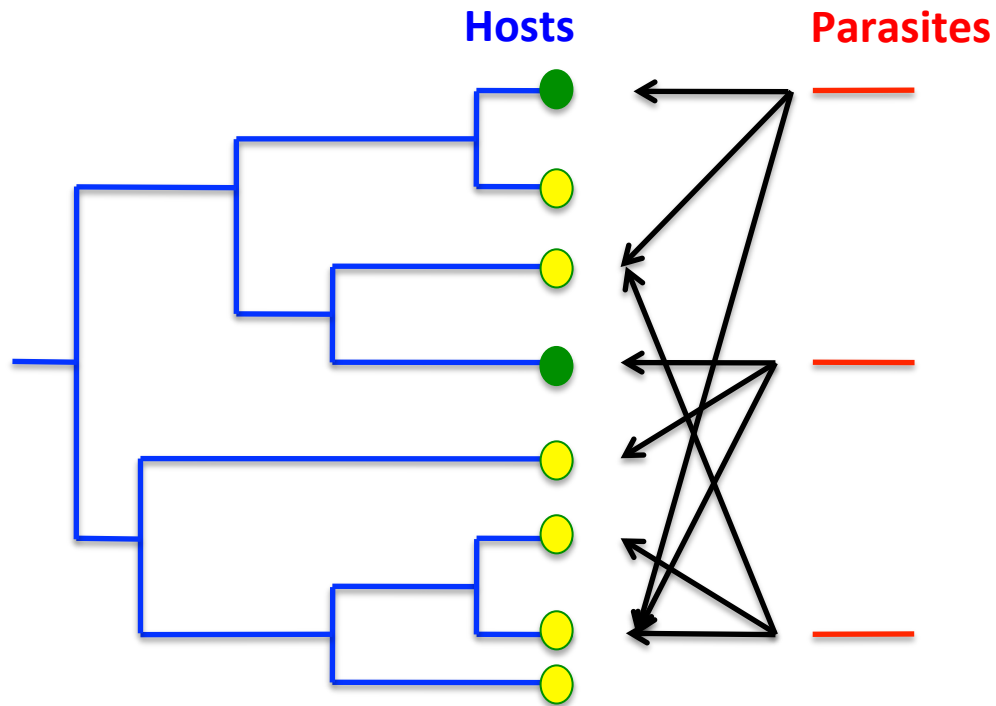
A	B	C	D
Species	Wolbachia	Spiroplasma	Rickettsia
<i>Adalia bipunctata</i>	0	1	1
<i>Adalia decempunctata</i>	0	0	1
<i>Anatis ocellata</i>	0	0	0
<i>Anisosticta novemdecimpunctata</i>	0	1	0
<i>Aphidecta oblitterata</i>	0	0	0
<i>Calvia quattuordecimguttata</i>	1	0	1
<i>Chilocorus bipustulatus</i>	1	1	0
<i>Coccidula rufa</i>	1	0	1
<i>Coccinella hieroglyphica</i>	0	0	0
<i>Coccinella septempunctata</i>	1	0	0
<i>Exochomus quadripustulatus</i>	0	0	0
<i>Halyzia sedecimguttata</i>	1	0	1
<i>Harmonia quadripunctata</i>	0	0	0
<i>Myrha octodecimguttata</i>	0	0	0
<i>Myzia oblongoguttata</i>	0	0	0
<i>Propylea quattuordecimpunctata</i>	0	0	0
<i>Rhyzobius (Rhizobius) litura</i>	1	0	1
<i>Scymnus frontalis</i>	0	0	1
<i>Subcoccinella vigintiquatuor punctata</i>	0	0	1
<i>Tytthaspis sedecimpunctata</i>	0	0	0

millions of years ago

ent

(Engelstaedter & Hurst 2011; Weinert et al. 2009; Welch et al. *In prep.*)

# 5. Incidence dynamics



- Each parasite associated with multiple hosts
- Parasite phylogeny ignored/unknown
- Host shifts are rapid compared to host radiation

# 5. Incidence dynamics

A model of parasite incidence dynamics:

In infinitesimal time interval  $dt$ , an infected host clear the infection with probability  $u dt$

and switches from infected host  $i$  to uninfected host  $j$  with probability

$$v_{ij} dt = \alpha e^{-\beta d_{ij}} dt$$

- If parasite not globally extinct, the population at any given time is in one of  $2^n - 1$  possible infection states.
- The probability that the population is in each of these states, reaches a **quasi stationary distribution** (an eigenvector of the  $2^n - 1 \times 2^n - 1$  transition matrix).
- Can therefore obtain ML estimates of  $u/\alpha$  and  $\beta$  from the data (for  $n < 17$  hosts)
- Work in progress....

(Engelstaedter & Hurst 2009; Welch et al. *In prep.*)

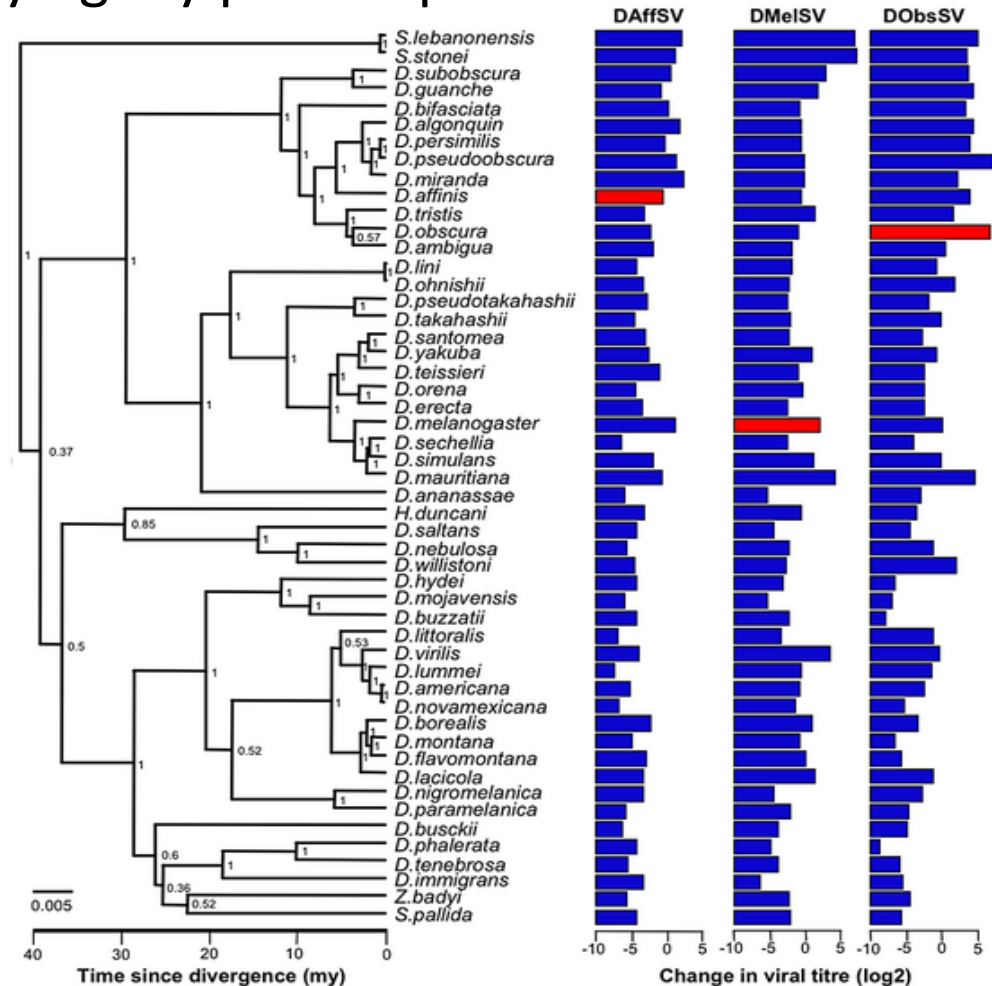
# 6. Tree as predictor variable

To what extent does host phylogeny predict parasite success?



3 viruses injected into 51 different species of Drosophilidae, including their natural hosts.

Viral titre measured after 15 days



(Longdon et al. 2011; Hadfield 2010; Gilmour et al. 2002)



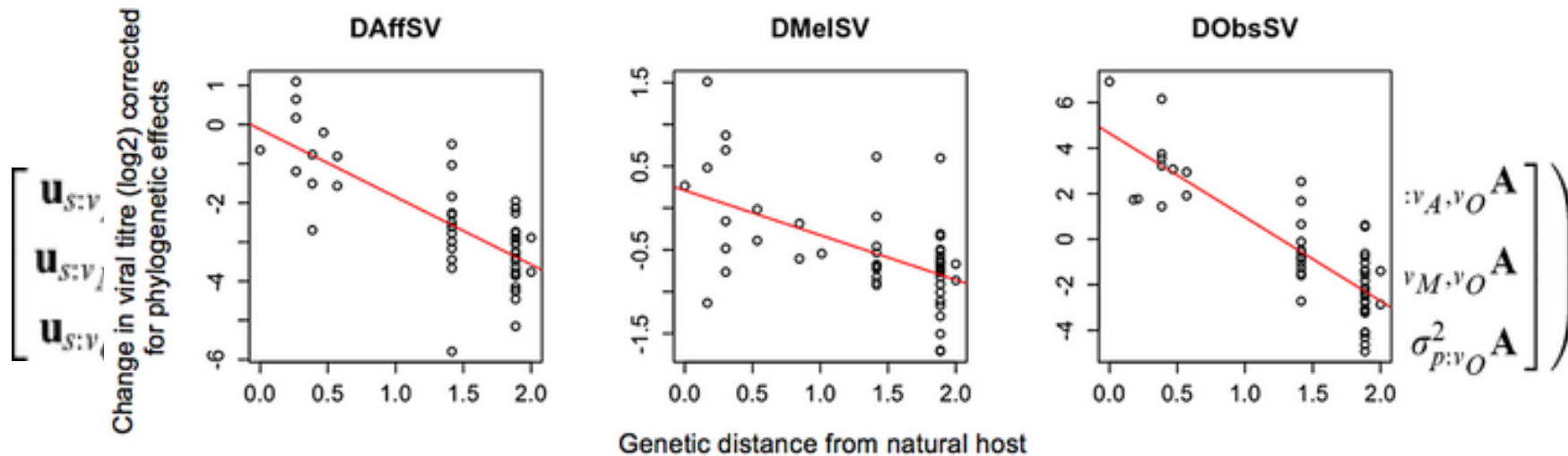
# 6. Tree as predictor variable

To what extent does host phylogeny predict parasite success?



$$y_{vhi} = \beta_v + d_{vh}\gamma_v + u_{p:vh} + u_{s:vh} + e_{vhi}$$

- $y_{vhi}$  viral titre for  $i$ th replicate with host  $h$  and virus  $v$
- $\beta_v$  intercept for virus  $v$
- $d_{vh}$  genetic distance of host  $h$  from natural host of virus  $v$
- $u_{p:vh}$  (random effect) phylogeny
- $u_{s:vh}$  (random effect) excluding phylogeny



(Longdon et al. 2011; Hadfield 2010; Gilmour et al. 2002)