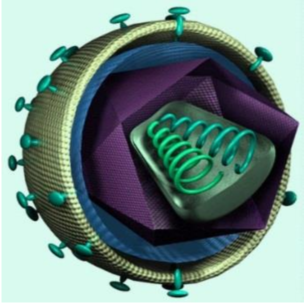


Bayesian methods for detecting epistatic interactions and compensatory mutations in Influenza A viruses

Dr Samantha Lycett, Mojca Zelnikar,
Prof Andrew Leigh Brown,
Prof Andrew Rambaut

THE UNIVERSITY of EDINBURGH

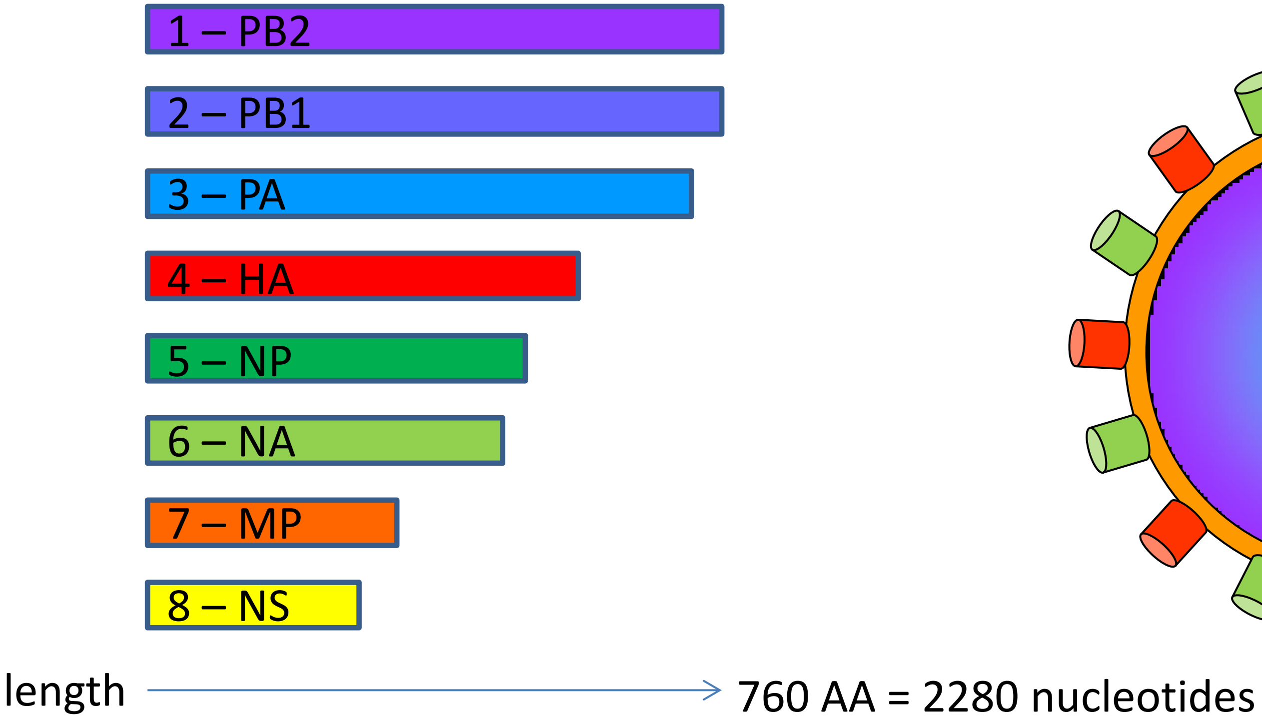
samantha.lycett@ed.ac.uk

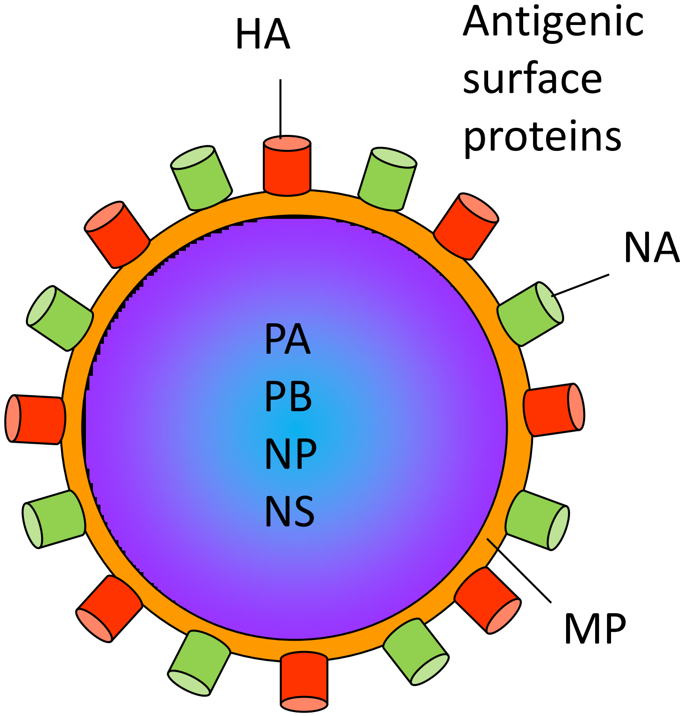
Contents

- Introduction
  - Interactions between sites in influenza
- Methods for quantifying interactions:
  - Correlation of mapped changes
  - Joint models

Influenza Virus

- 8 RNA Segments coding for 10-11 Proteins

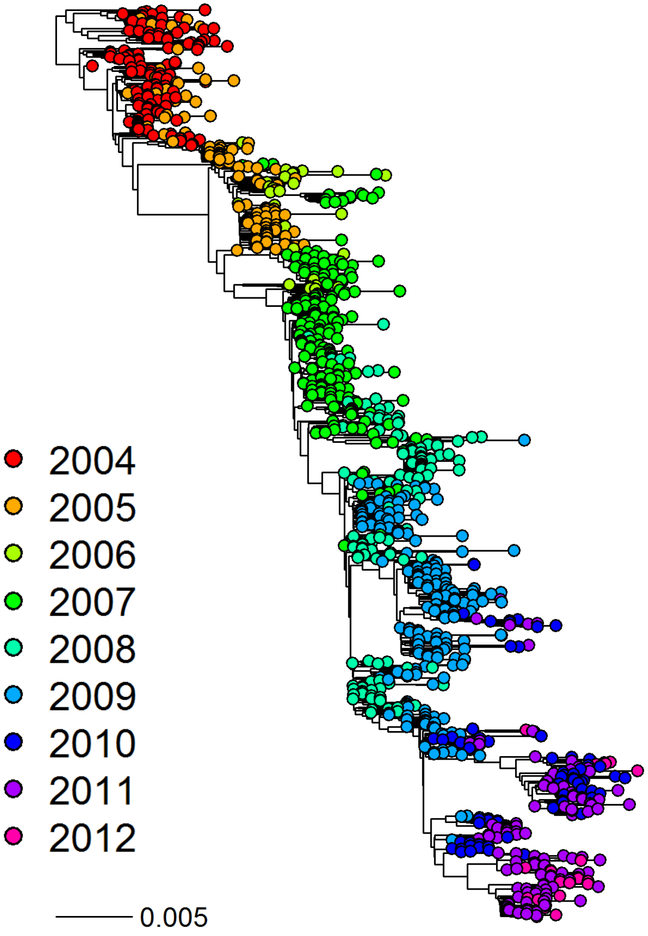




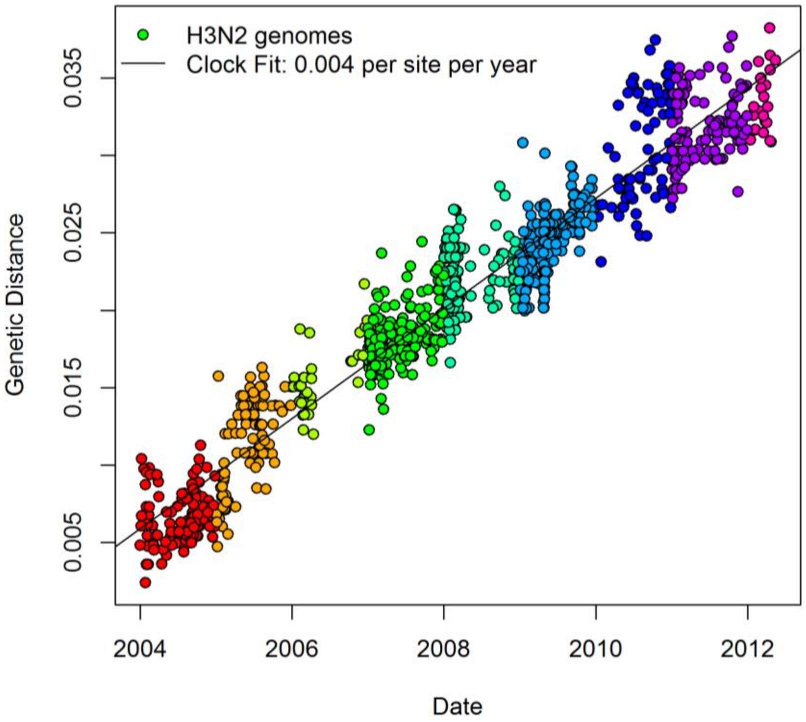
- Nucleotide mutation rate about 1 – 5 x 10-3 per site per year
- Genome is 13Kb : expect 10 – 70 mutations per genome per year

Influenza Evolution is Clock Like in Humans

Tree of H3N2 – HA sequences 2004-2012



Genetic Distance from Root

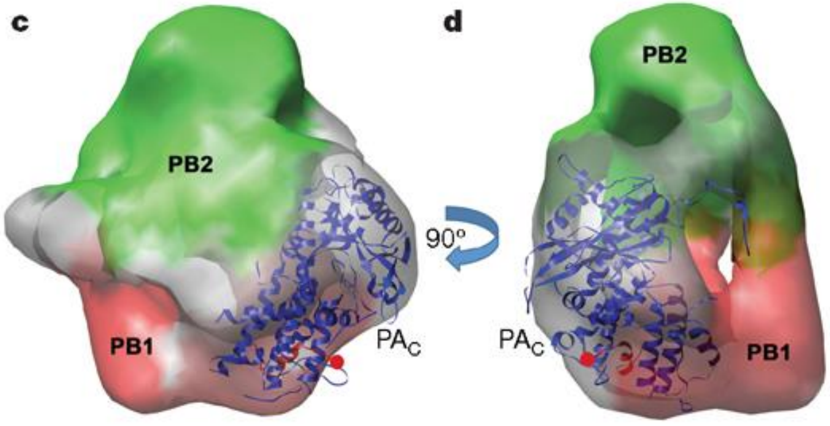


- Human immunity drives antigenic novelty in influenza

Types of Interactions

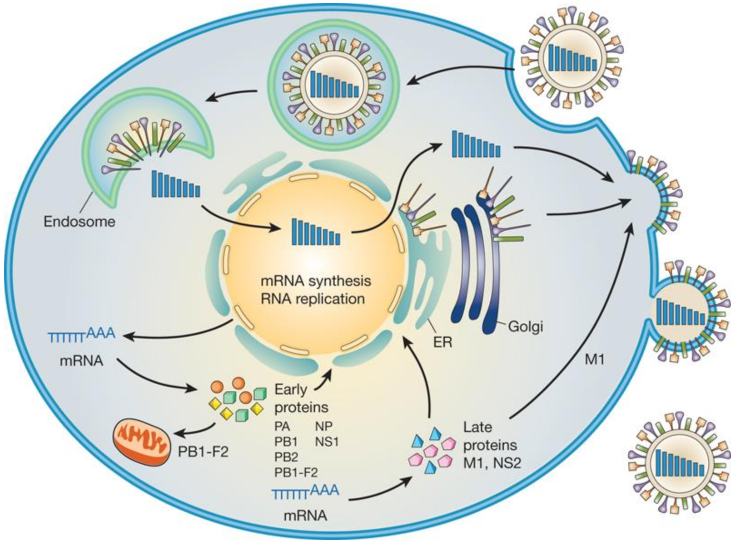
- Structural
  - Between Polymerases

XJ He *et al.* *Nature* **000**, 1-4 (2008)
doi:10.1038/nature07120



- Functional balance
  - Between HA & NA

G Neumann *et al.* *Nature* **459**, 931-939 (2009)
doi:10.1038/nature08157



Epistatic expectations and evidence

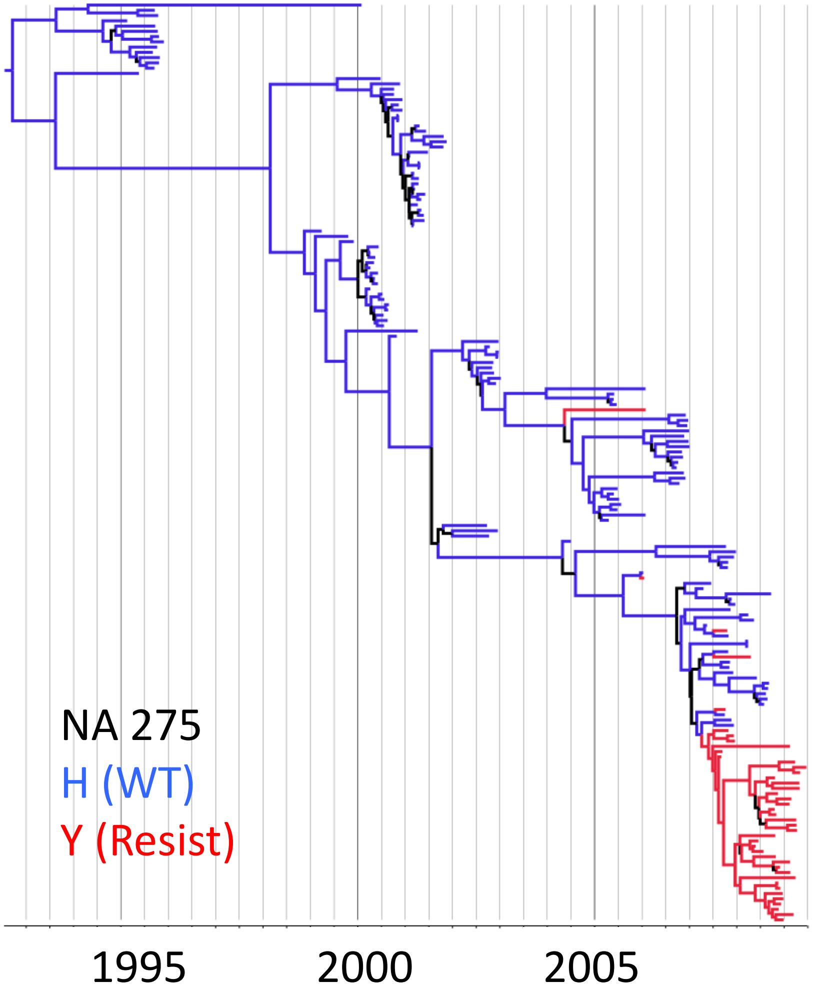
- Sites or genomic features are correlated
  - Fitness enhancing, so expect higher prevalence of favourable combinations
  - Expect unfavourable combinations to die out
  - Expect selection for compensatory mutations
- Evidence:
  - Fitness competition assays
  - Prevalence of reassortments
  - Statistical signatures

Motivating Data - Background

- Neuraminidase inhibitor drugs in use since 1999
  - An experimental study showed that drug resistant NA in H1N1 was less fit than wild type
- But rapid increase in drug resistance in H1N1
  - particularly oseltamivir during 2007 & 2008
  - in the absence of direct selective pressure
- Possible explanations:
  - “Permissive” mutations in NA, which compensate for the drug resistant mutation in the NA active site
  - “Hitch-hiking” with antigenically favourable HA, or other mutations in HA which restore HA-NA balance

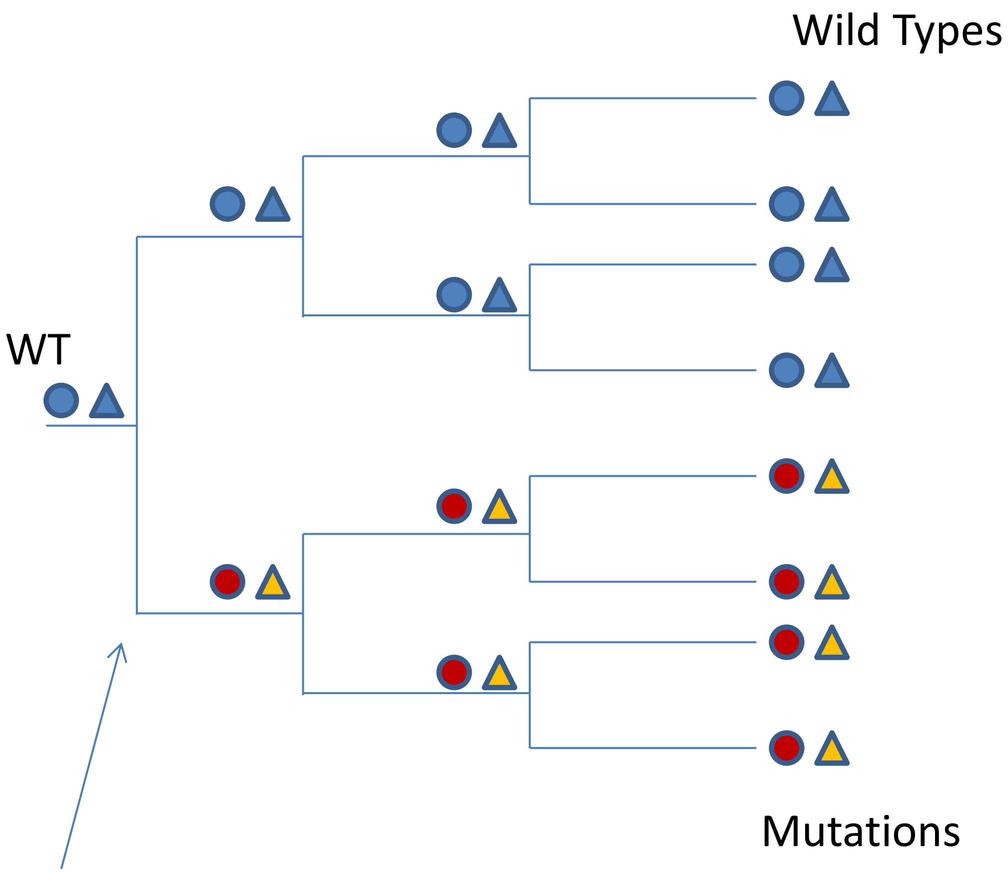
H1N1 Seasonal Data

- Data set = complete genome sequences with day, month, year
- Subsample to 3 per country per year => 175 from 1995 – 2009
- Concatenate HA and NA sequences, and test for recombination/reassortment with HyPhy (GARD and Single Break Point)
- No reassortment detected between HA and NA (on this set)

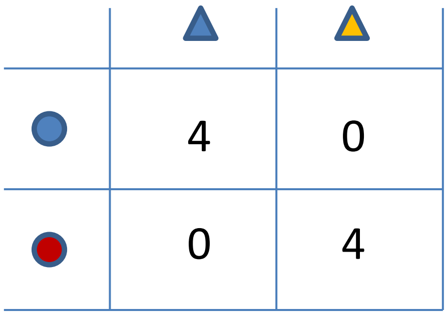


Detect Epistatic Interactions ?

False Signatures – Founder Effect



Correlation of sequences



	▲	△
🔵	4	0
🔴	0	4

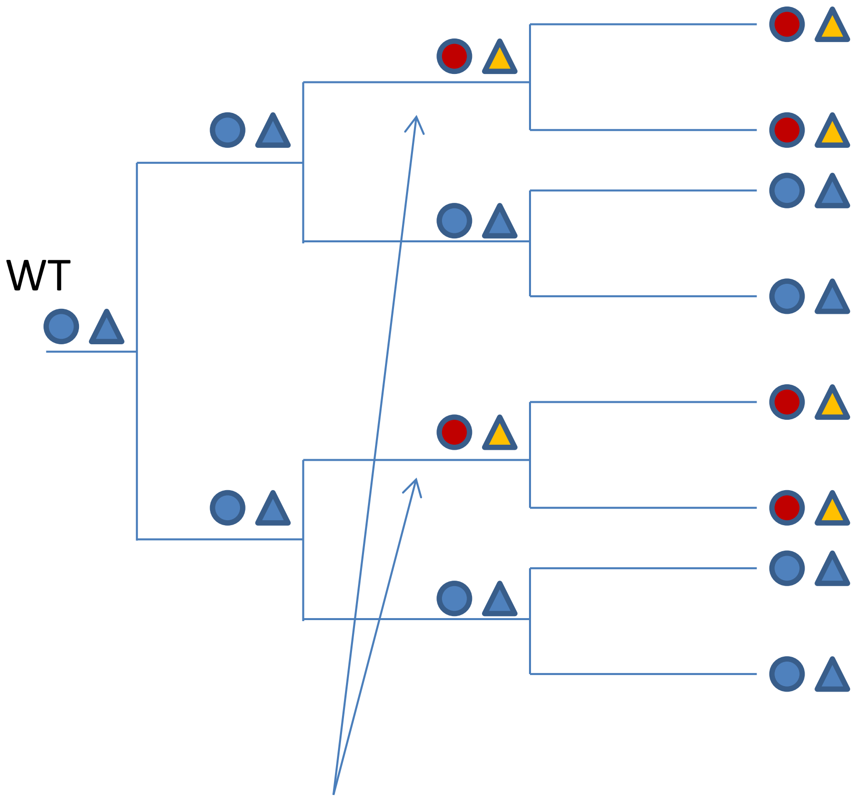
Looks like:

- wild types co-occurr
- mutations co-occurr

Just one founder event

Is the correlation really significant ?

Signatures – Co-evolution



The same double mutation has happened more than once !

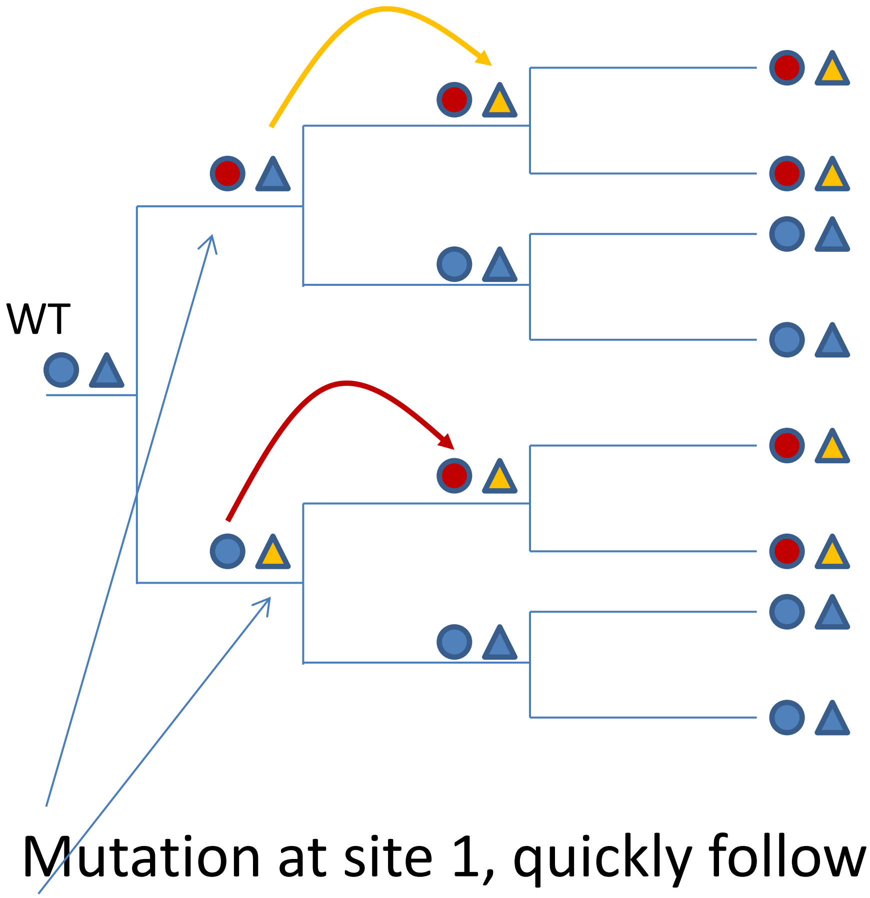
Correlation of sequences

	▲	△
🔵	4	0
🔴	0	4

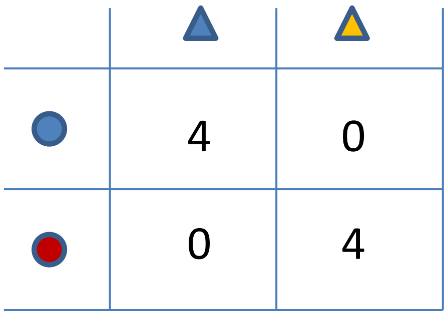
Method 1 – Mapping

Signatures – Co-evolution

(Especially if short branches and dense sampling)



Correlation of sequences



	▲ (blue)	▲ (yellow)
● (blue)	4	0
● (red)	0	4

Method 2 – Modelling

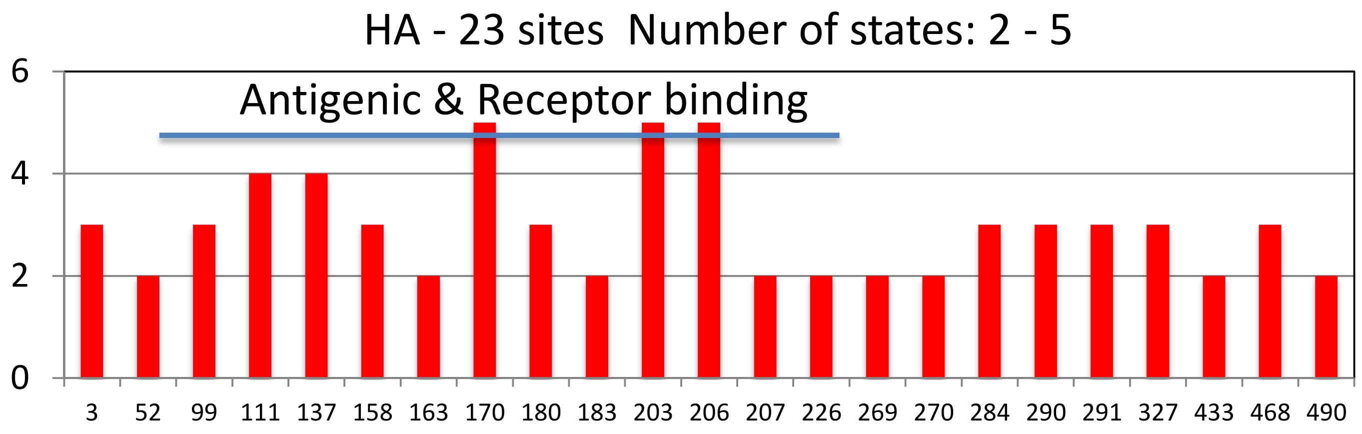
Mutation at site 1, quickly followed by site 2
Mutation at site 2, quickly followed by site 1

Methods 1 - Mapping

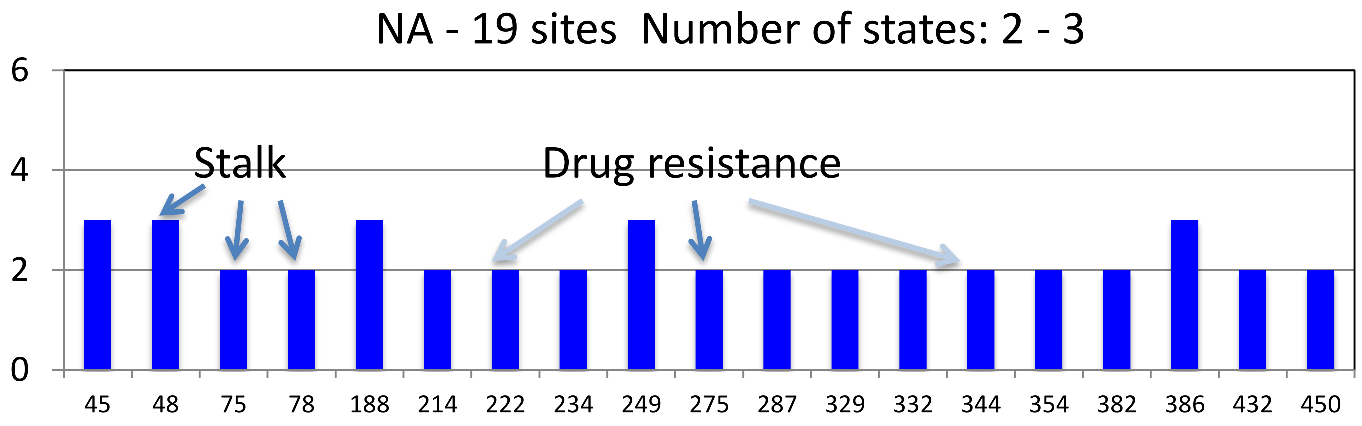
- Construct tree and reconstruct ancestral states at internal nodes
    - Tree from nucleotide sequences
    - Asymmetric discrete state amino acid model
- For each branch, compare node to ancestral node:
    - 0 = no change from ancestor
    - 1 = mutation
- Calculate correlation of pattern of changes between sites (fisher test on: 00, 01, 10, 11)
- Repeat over 1000 trees with 10 sampled reconstructions per tree
    - Trees and mappings performed in BEAST

Initial Site Selection

HA

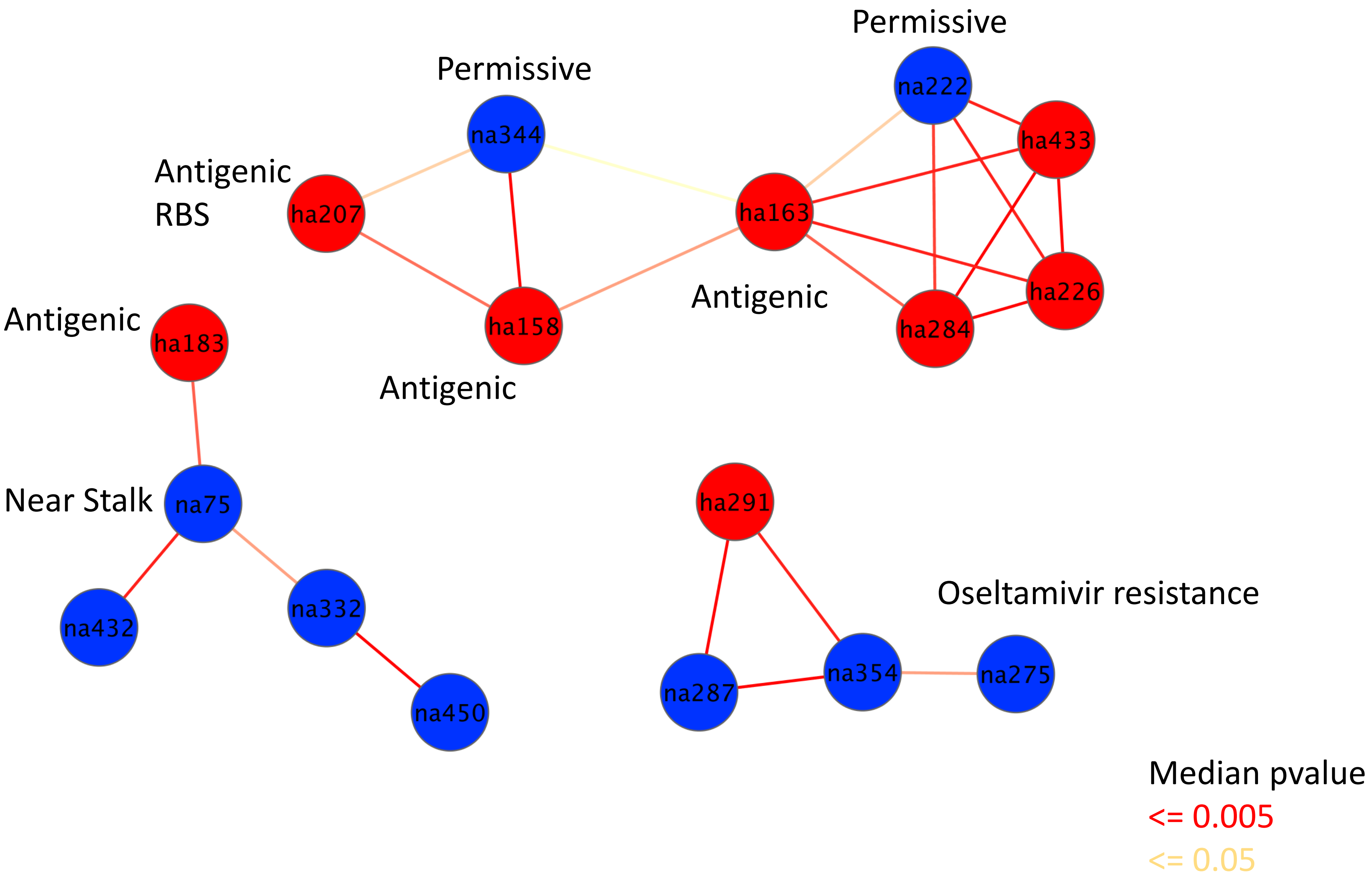

HA - 23 sites Number of states: 2 - 5

NA


NA - 19 sites Number of states: 2 - 3

Variable sites with minor frequency of 10%

HA-NA Interactions by Mapping



But…

- Not multiple testing corrected
- Links are apparent associations, but could be indirect
  - Could have used this data for structural inference of a Bayesian Graphical Model (HyPhy)
- But have low numbers of co-mutations on a branch
  - Averaged over trees and reconstructions:
  - HA-158 – NA-344 : 1.85 co-mutations (median=2)
  - HA-206 – HA-207 : 1.80 co-mutations (median=2)

Method 2

Joint Discrete Trait Models

Detecting Coevolving sites

- Consider 2 sites: site 1 with A or B; site 2 with a or b
- Does the state of site 1 affect site 2 ?
  - Independent: rate of a <-> b (site 2) does not depend upon state at site 1 (A or B)
  - Dependent: need to consider rates of joint sites e.g. rate Aa – Ab, Ba – Bb etc

Independent Rates

Site 1	A	B
A		1
B	2	

Site 2	a	b
a		1
b	2	

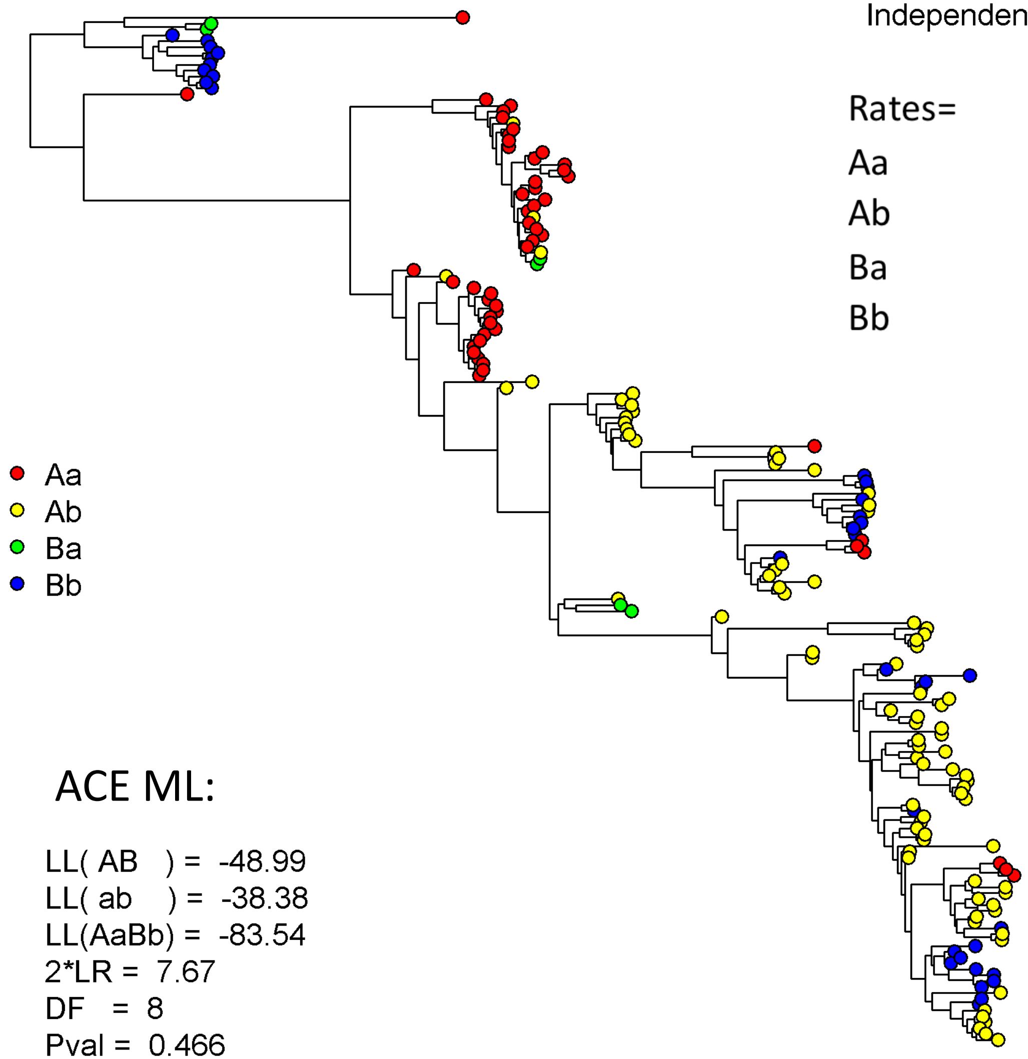
Dependent Rates

Site 1+2	Aa	Ab	Ba	Bb
Aa		1	2	3
Ab	7		4	5
Ba	8	10		6
Bb	9	11	12	

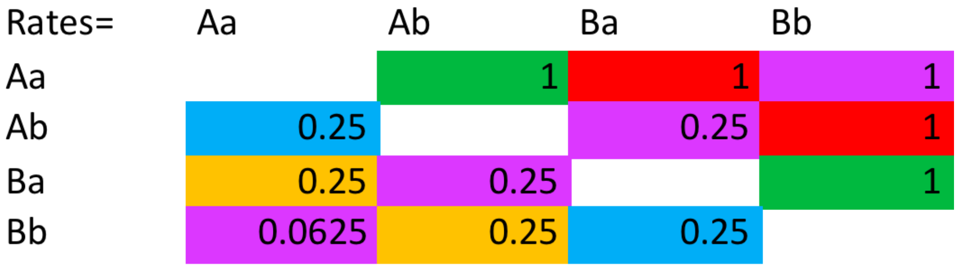
Detecting Coevolving sites

- Fit discrete asymmetric models to tip data
  - Model 1: Site 1 – e.g. A – B (2 states, 2 rates)
  - Model 2: Site 2 – e.g. a – b (2 states, 2 rates)
  - Model 3: Joint Site 1 & 2 e.g. Aa – Bb (4 states, 12 rates)
- Calculate likelihood ratio between dependent and independent models:
  - LR = LL(Model 3) – (LL(Model 1) + LL(Model 2))
- The 2*LR will be chi-square distributed with degrees of freedom = difference in parameters
  - DF = 12 – (2 + 2) = 8
  - P-value = 1 – prob(chi.square(2*LR, DF))

Example Independent Traits



Independent 20



Rates=	Aa	Ab	Ba	Bb
Aa		1	1	1
Ab	0.25		0.25	1
Ba	0.25	0.25		1
Bb	0.0625	0.25	0.25	

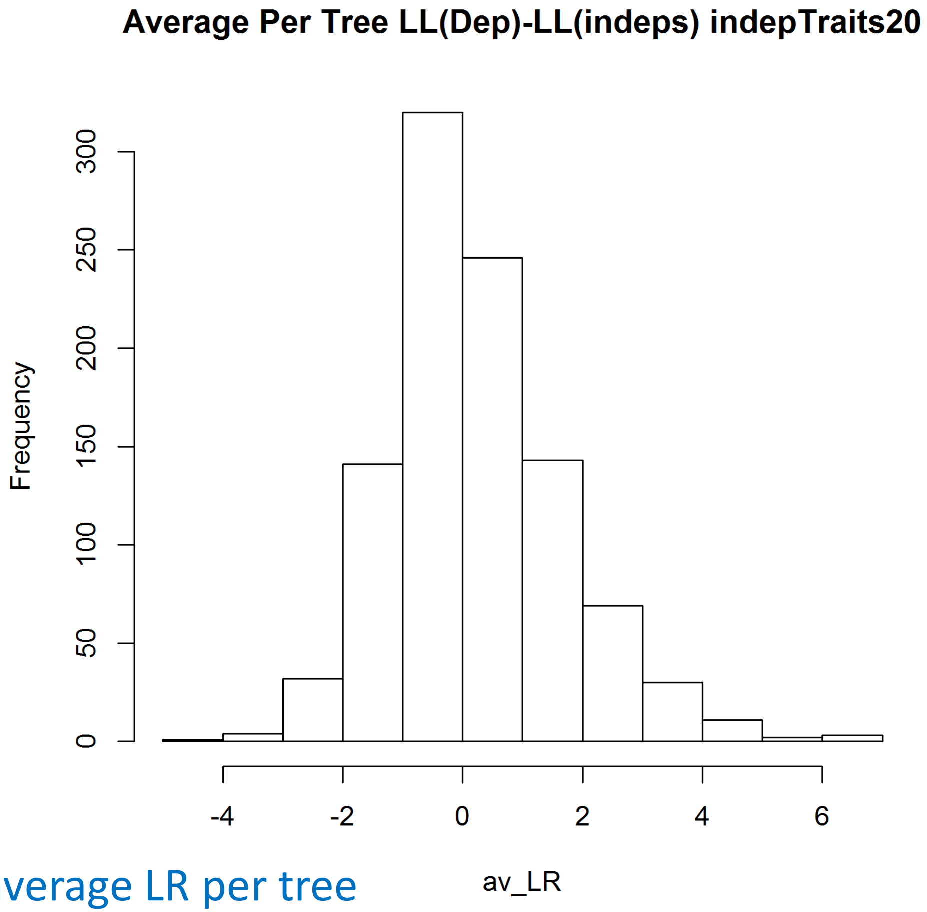
Rates
a -> b = 1
b -> a = 0.25
A -> B = 1
B -> A = 0.25

ACE ML:

LL(AB) = -48.99
LL(ab) = -38.38
LL(AaBb) = -83.54
2*LR = 7.67
DF = 8
Pval = 0.466

Generated Independent Traits
LR test using ACE =>
dependent model not favoured
(LR not significant)

Example Independent Traits: BEAST Results

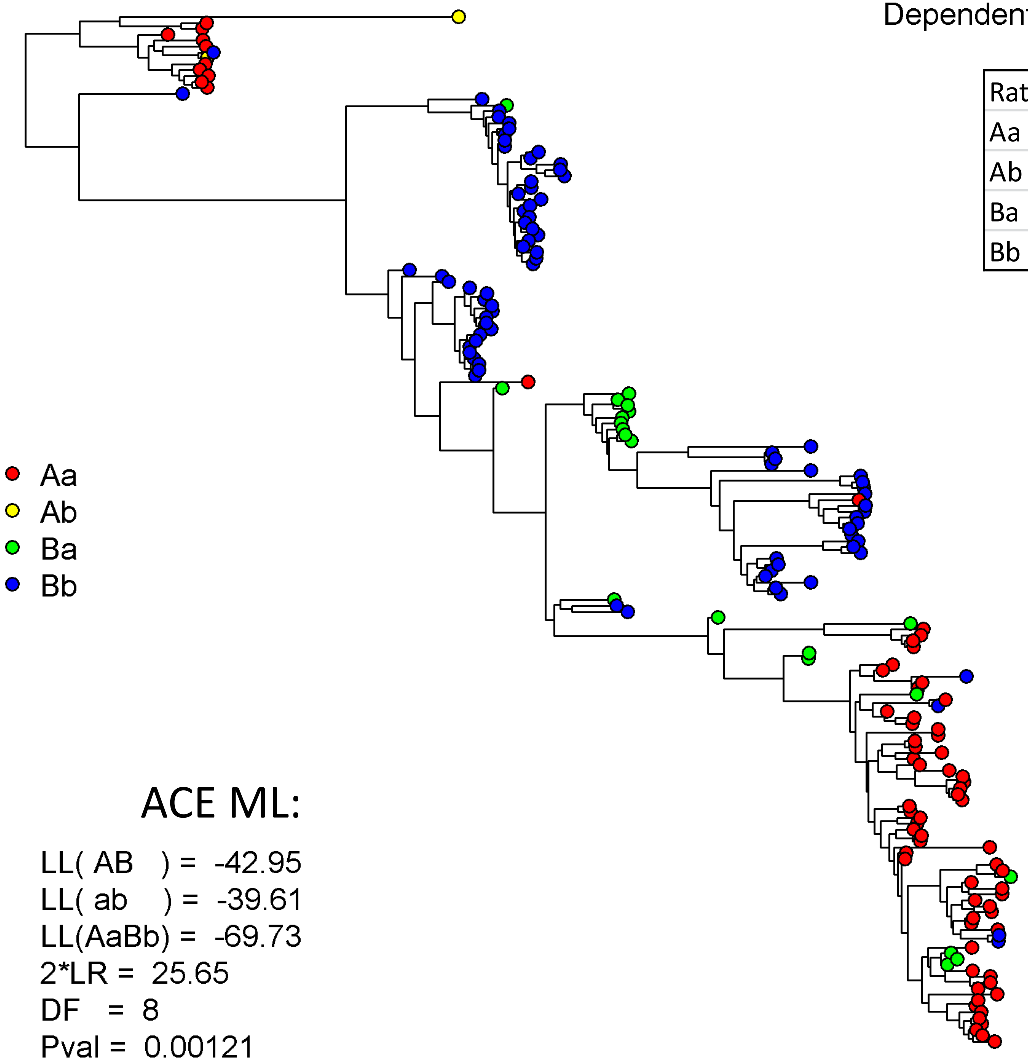


Independent traits are independent by LR Test (SUCCESS!)

Distribution of average LR per tree
Mean LR = 0.2   95% HPD = -2.3 – 3.2
Mean p   = 0.95 95% HPD = 0.67 – 0.99 (not significant)

[For BEAST: Used 1000 empirical trees so multiple comparisons per tree]

Example Dependent Traits



Dependent 39

Rates=	Aa	Ab	Ba	Bb
Aa		0.1	1	1
Ab	1		0.01	1
Ba	0.25	0.01		1
Bb	0.25	0.25	0.1	

ACE ML:

LL(AB) = -42.95
LL(ab) = -39.61
LL(AaBb) = -69.73
2*LR = 25.65
DF = 8
Pval = 0.00121

Generated Dependent Traits
LR test using ACE =>
dependent model favoured
(LR significant)

Example Dependent Traits

Rate Aa -> Ab = slow

slow rate for a-b when in state A (dont want to be in state Ab)

Rate Ba -> Bb = fast

fast rate for a-b when in state B (want to be in state Bb)

Rate Ab -> Aa = fast

fast rate for b-a when in state A (want to be in state Aa)

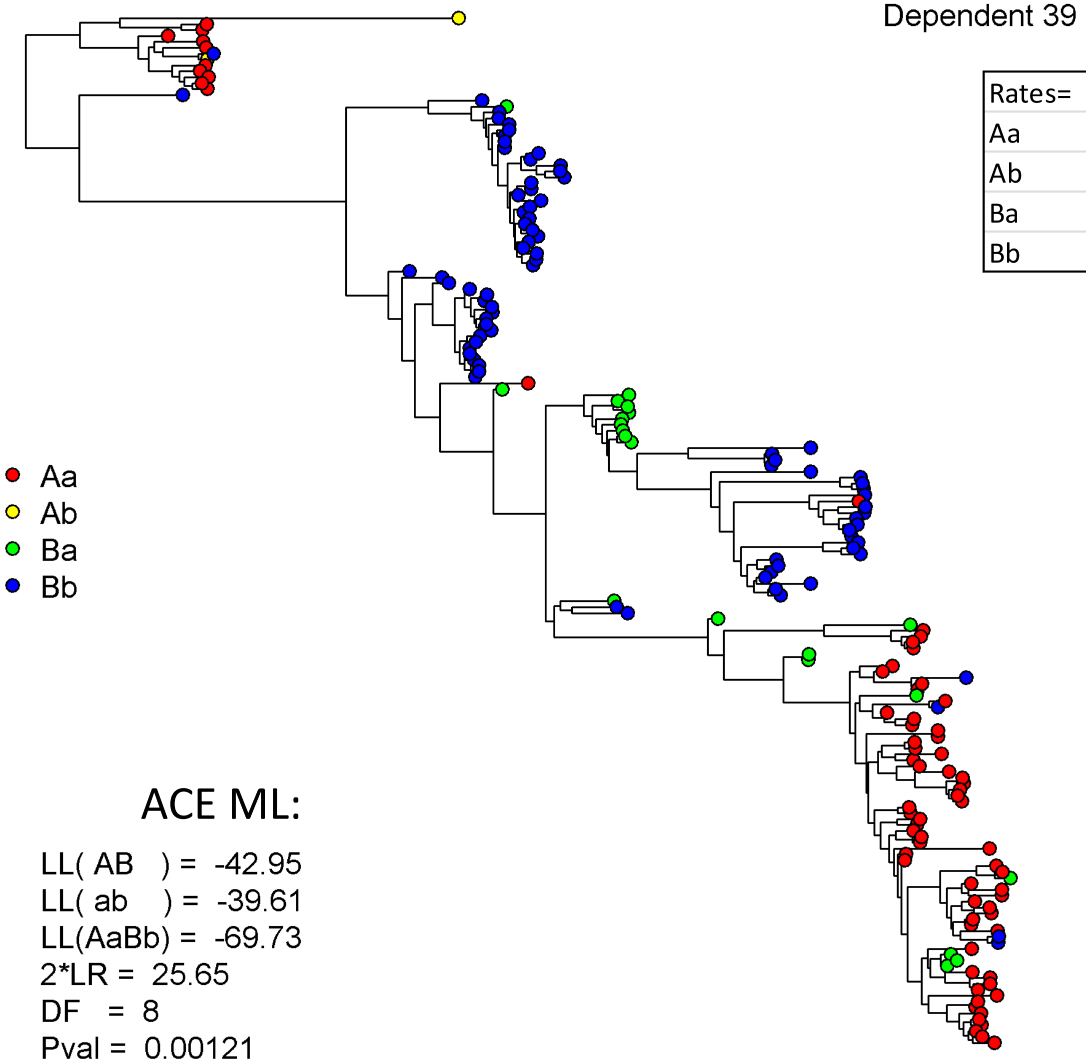
Rate Bb -> Ba = slow

slow rate for b-a when in state B (dont want to be in state Ba)

Rate Ab -> Ba = slow * slow very slow

Rate Ba -> Ab = slow * slow very slow

Example Dependent Traits


Dependent 39

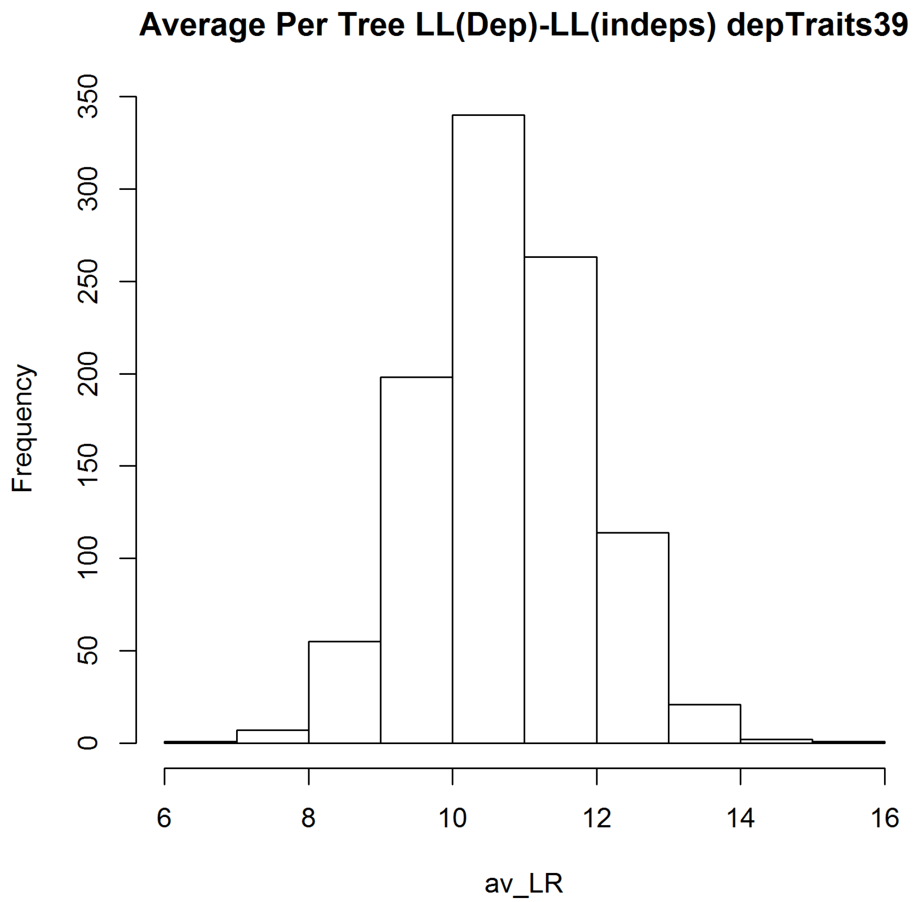
Rates=	Aa	Ab	Ba	Bb
Aa		0.1	1	1
Ab	1		0.01	1
Ba	0.25	0.01		1
Bb	0.25	0.25	0.1	

ACE ML:

LL(AB) = -42.95
LL(ab) = -39.61
LL(AaBb) = -69.73
2*LR = 25.65
DF = 8
Pval = 0.00121

Generated Dependent Traits
LR test using ACE =>
dependent model favoured
(LR significant)

Example Dependent Traits: BEAST Results



Dependent traits are
Dependent by LR Test
(SUCCESS!)

Distribution of average LR per tree
Mean LR = 10.7   95% HPD = 8.6 – 13.0
Mean p   = 0.008 95% HPD = 0.0002 – 0.02 (significant)

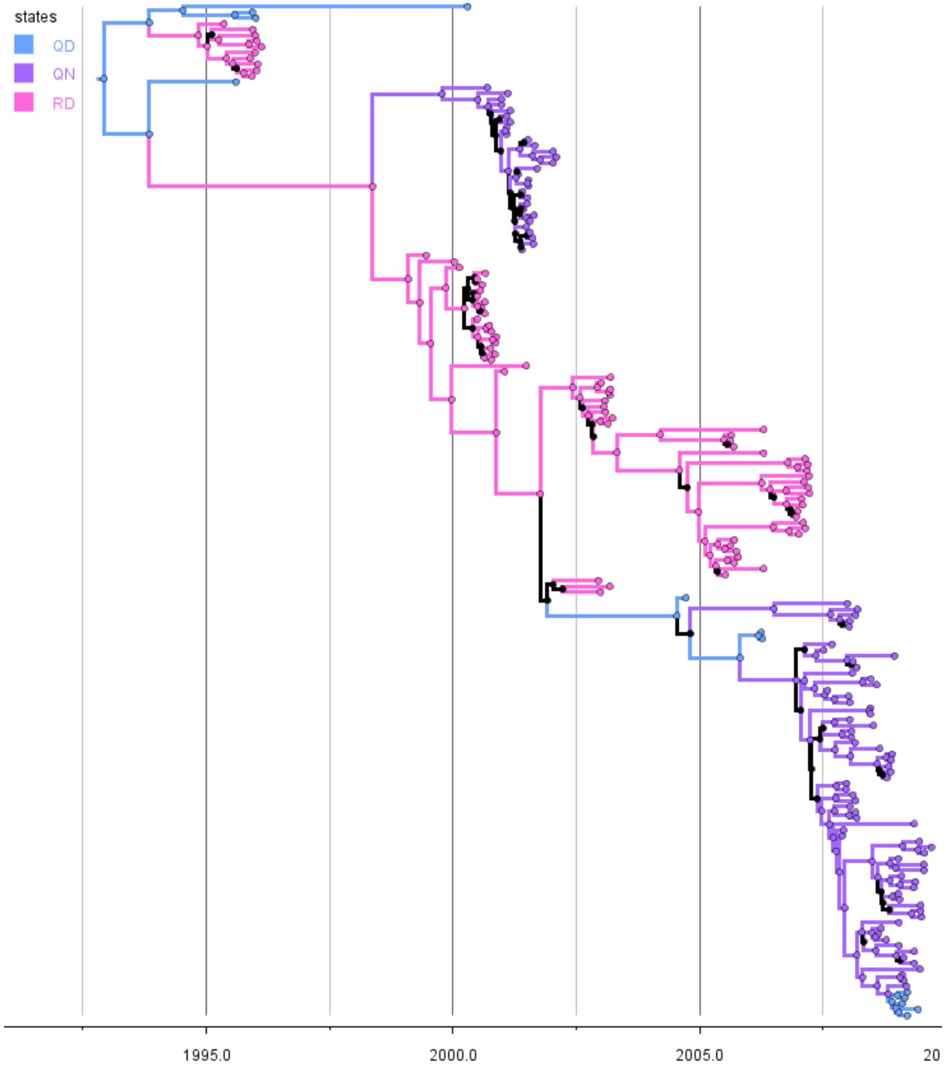
[For BEAST: Used 1000 empirical trees so multiple comparisons per tree]

Interactions by Modelling using BEAST

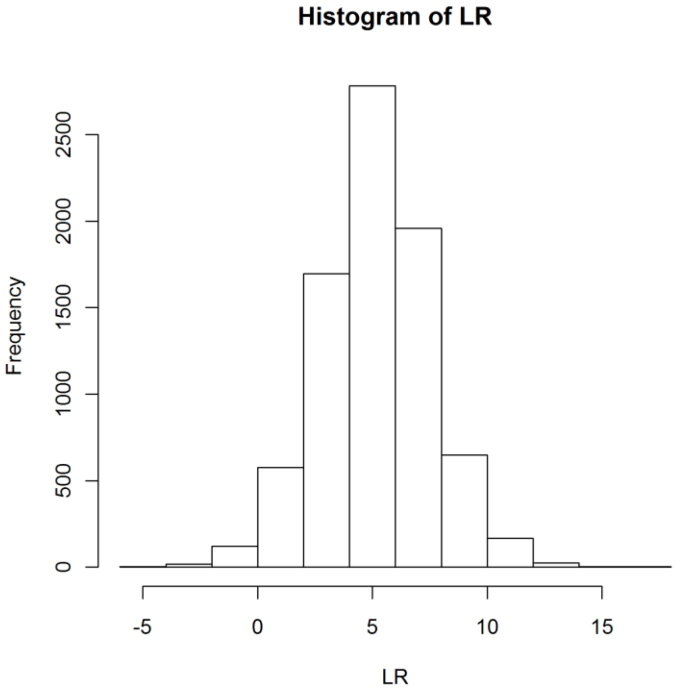
- Calculate dependent and independent scores for selected sites over posterior distribution of trees and models
- Median p-values

Model	HA-203	HA-206	HA-207	NA-222	NA-275	NA-344
HA-203	-	1.000	0.337	0.665	0.239	1.000
HA-206	1.000	-	0.513	1.000	0.993	0.998
HA-207	0.337	0.513	-	0.083	1.000	0.707
NA-222	0.665	1.000	0.083	-	1.000	0.006
NA-275	0.239	0.993	1.000	1.000	-	1.000
NA-344	1.000	0.998	0.707	0.006	1.000	-

NA 222 + NA 344 – ‘Permissive Mutations’

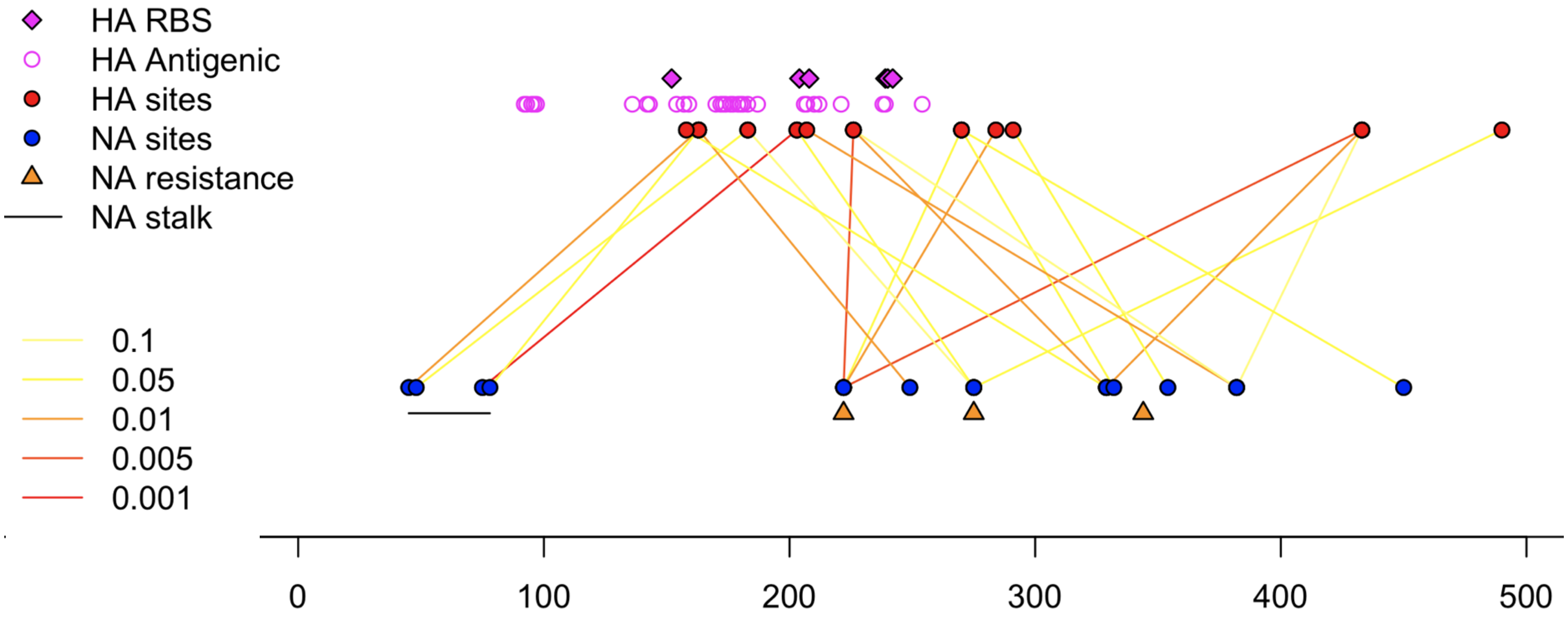


NA 222 + NA 344



NA 222 + NA 344	
LR Mean	= 5.15
DF	= 2
Median pvalue	= 0.0056

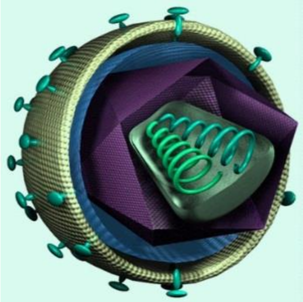
Interactions Between HA and NA by Modelling



Conclusions

- Detect epistatic interactions
  - Mapping works but is best for multiple independent changes
  - Joint state modelling is preferable for Influenza example, because want to detect changes on different branches
- Found interesting looking interactions with HA and NA

Acknowledgements



- Andrew J. Leigh Brown
- Mojca Zelnikar
- Lu Lu
- Emma Hodcroft
- Manon Ragonnet-Cronin
- Melissa Ward

- Andrew Rambaut
- Trevor Bedford
- Jess Hedge
- Gytis Dudas
- Matthew Hall

University of Edinburgh

Supported by wellcome trust

samantha.lycett@ed.ac.uk