Phylogenetic Support

Statistical Testing of Trees

Finlay Maguire

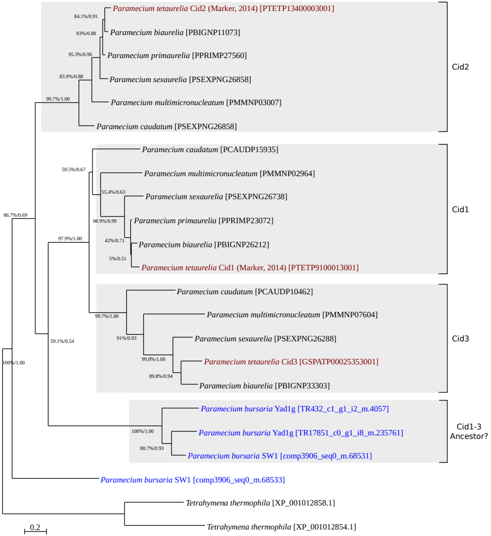
March 27, 2018

FCS, Dalhousie

Table of contents

Introduction

Phylogenies are hypotheses



Cid

Hypothesis testing

- Does another model of sequence evolution fit the data better?
- How well supported are individual branches in a tree?
- Does another tree explain the data better?

Sources of Error

- Bad data

Sources of Error

- Bad data
- Sampling error

Sources of Error

- Bad data
- Sampling error
- Misleading evolutionary events

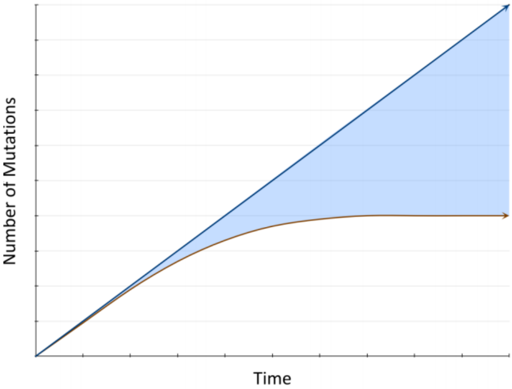
Sources of Error

- Bad data
- Sampling error
- Misleading evolutionary events
- Misspecified models

Sources of Error

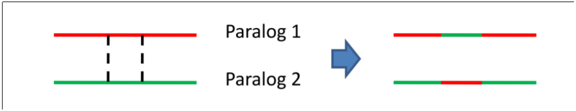
- Bad data
- Sampling error
- Misleading evolutionary events
- Misspecified models
- Inappropriate inference

Saturation

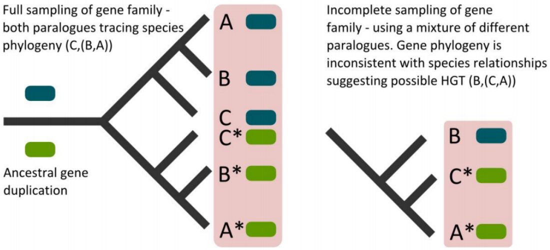


[Leonard, 2010]

Misleading Signal: Recombination

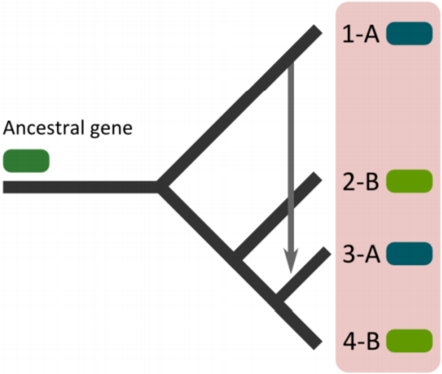


Misleading Signal: Hidden Paralogy/Incomplete Sampling



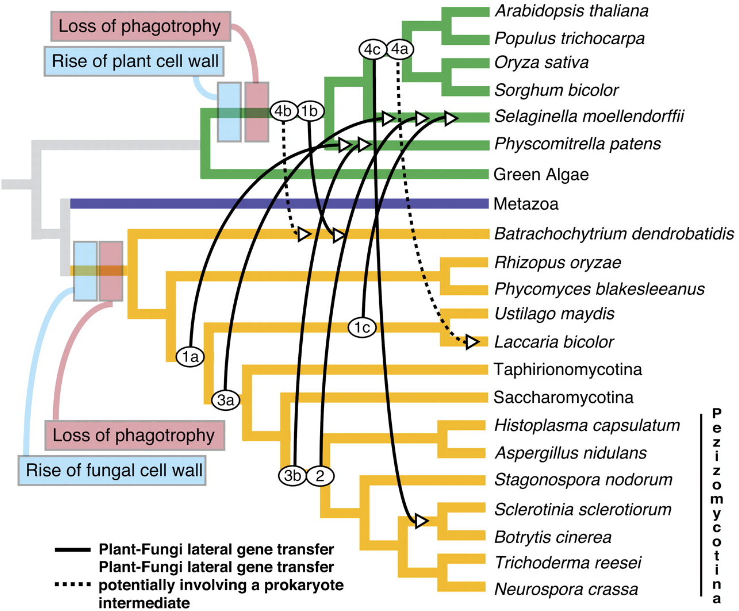
[Leonard, 2010]

Misleading Signal: Horizontal Gene Transfer



[Leonard, 2010]

Misleading Signal: Horizontal Gene Transfer

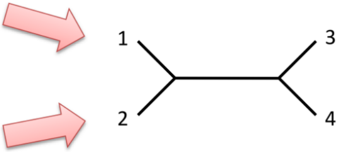


Tree not always correct paradigm

Ask for a tree get a tree.

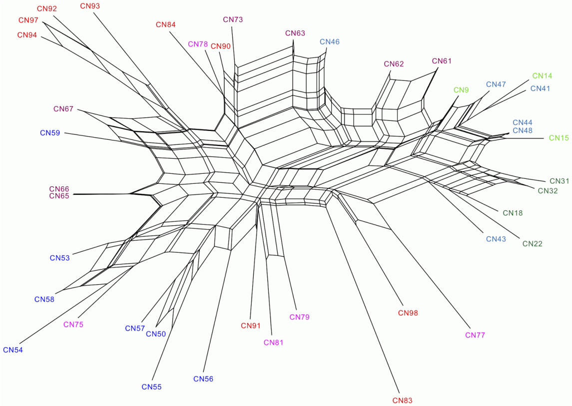
```
1  ACCGAGCAA
2  ACCGAGCAA
3  ACCGAGCAA
4  ACCGAGCAA

1  ACCGAATGA
2  ACCGAGCAG
3  GTTAGGCAG
4  GTTAGATGA
```



Tree not always correct paradigm

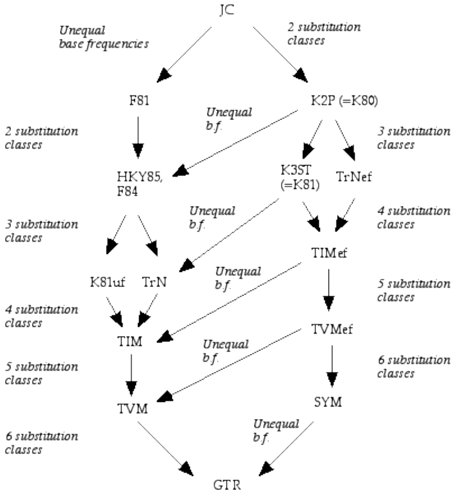
Ask for a tree get a tree.



Reanalysis of [Marwick, 2012] from `http://phylonetworks.blogspot.ca/2013/02/`

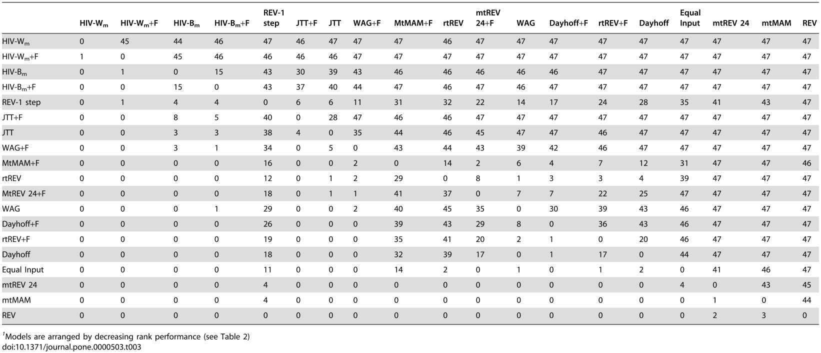
Evolutionary Model Testing

Sequence Evolution Models



`http://carrot.mcb.uconn.edu/~olgazh/bioinf2010/class24.html`

Sequence Evolution Models



	HIV-W$_m$	HIV-W$_m$+F	HIV-B$_m$	HIV-B$_m$+F	REV-1 step	JTT+F	JTT	WAG+F	MtMAM+F	rtREV	mtREV 24+F	WAG	Dayhoff+F	rtREV+F	Dayhoff	Equal Input	mtREV 24	mtMAM	REV
HIV-W$_m$	0	45	44	46	47	46	47	47	47	46	47	47	47	47	47	47	47	47	47
HIV-W$_m$+F	1	0	45	46	46	46	46	47	47	47	47	47	47	47	47	47	47	47	47
HIV-B$_m$	0	1	0	15	43	30	39	43	46	46	46	46	46	47	47	47	47	47	47
HIV-B$_m$+F	0	0	15	0	43	37	40	44	47	46	47	46	47	47	47	47	47	47	47
REV-1 step	0	1	4	4	0	6	6	11	31	32	22	14	17	24	28	35	41	43	47
JTT+F	0	0	8	5	40	0	28	47	46	46	47	47	47	47	47	47	47	47	47
JTT	0	0	3	3	38	4	0	35	44	46	45	47	47	46	47	47	47	47	47
WAG+F	0	0	3	1	34	0	5	0	43	44	43	39	42	46	47	47	47	47	47
MtMAM+F	0	0	0	0	16	0	0	2	0	14	2	6	4	7	12	31	47	47	46
rtREV	0	0	0	0	12	0	1	2	29	0	8	1	3	3	4	39	47	47	47
MtREV 24+F	0	0	0	0	18	0	1	1	41	37	0	7	7	22	25	47	47	47	47
WAG	0	0	0	1	29	0	0	2	40	45	35	0	30	39	43	46	47	47	47
Dayhoff+F	0	0	0	0	26	0	0	0	39	43	29	8	0	36	43	46	47	47	47
rtREV+F	0	0	0	0	19	0	0	0	35	41	20	2	1	0	20	46	47	47	47
Dayhoff	0	0	0	0	18	0	0	0	32	39	17	0	1	17	0	44	47	47	47
Equal Input	0	0	0	0	11	0	0	0	14	2	0	1	0	1	2	0	41	46	47
mtREV 24	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	4	0	43	45
mtMAM	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	1	0	44
REV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0

[1] Models are arranged by decreasing rank performance (see Table 2)

doi:10.1371/journal.pone.0000503.t003

[Nickle et al., 2007]

What happens if the wrong model is specified?

- Increased Inaccuracy (wrong tree more often)

What happens if the wrong model is specified?

- Increased Inaccuracy (wrong tree more often)
- Inconsistency (adding more data converges to wrong tree)

What happens if the wrong model is specified?

- Increased Inaccuracy (wrong tree more often)
- Inconsistency (adding more data converges to wrong tree)
- Wrong branch lengths (important for certain analyses)

What happens if the wrong model is specified?

- Increased Inaccuracy (wrong tree more often)
- Inconsistency (adding more data converges to wrong tree)
- Wrong branch lengths (important for certain analyses)
- Wrong tree support values

How do we select a model?

- $L(\tau, \theta) = P(X|\tau, \theta)$

How do we select a model?

- $L(\tau, \theta) = P(X|\tau, \theta)$
- With ML inference we are finding the maximum-likelihood estimate of τ and θ

How do we select a model?

- $L(\tau, \theta) = P(X|\tau, \theta)$
- With ML inference we are finding the maximum-likelihood estimate of τ and θ
- i.e. $\hat{\tau}, \hat{\theta} = \text{argmax}_{\tau,\theta} L(\tau, \theta)$

How do we select a model?

- $L(\tau, \theta) = P(X|\tau, \theta)$
- With ML inference we are finding the maximum-likelihood estimate of τ and θ
- i.e. $\hat{\tau}, \hat{\theta} = \operatorname{argmax}_{\tau,\theta} L(\tau, \theta)$
- Therefore, to compare two models we can use a likelihood ratio test (LRT δ)

How do we select a model?

- $L(\tau, \theta) = P(X|\tau, \theta)$
- With ML inference we are finding the maximum-likelihood estimate of τ and θ
- i.e. $\hat{\tau}, \hat{\theta} = \text{argmax}_{\tau,\theta} L(\tau, \theta)$
- Therefore, to compare two models we can use a likelihood ratio test (LRT δ)
- $\delta = 2(ln(L_1) - ln(L_0))$

How do we select a model?

- $L(\tau, \theta) = P(X|\tau, \theta)$
- With ML inference we are finding the maximum-likelihood estimate of τ and θ
- i.e. $\hat{\tau}, \hat{\theta} = \text{argmax}_{\tau,\theta} L(\tau, \theta)$
- Therefore, to compare two models we can use a likelihood ratio test (LRT δ)
- $\delta = 2(ln(L_1) - ln(L_0))$
- Limitations: nested models (i.e. hLRT), order matters, no regularisation

Information Criterion

- Akaike Information Criterion (AIC), penalising number of parameters:

Information Criterion

- Akaike Information Criterion (AIC), penalising number of parameters:
- $AIC = -2ln(L) + 2K$

Information Criterion

- Akaike Information Criterion (AIC), penalising number of parameters:
- $AIC = -2ln(L) + 2K$
- However, this penalises all high K models even if sample size is large too.

Information Criterion

- Akaike Information Criterion (**AIC**), penalising number of parameters:
- $AIC = -2ln(L) + 2K$
- However, this penalises all high K models even if sample size is large too.
- Corrected Akaike Information Criterion (**AICc**)

Information Criterion

- Akaike Information Criterion (AIC), penalising number of parameters:
- $AIC = -2ln(L) + 2K$
- However, this penalises all high K models even if sample size is large too.
- Corrected Akaike Information Criterion (AICc)
- $AICc = AIC + \frac{2K(K+1)}{n-K-1}$

Information Criterion

- Akaike Information Criterion (**AIC**), penalising number of parameters:
- $AIC = -2ln(L) + 2K$
- However, this penalises all high K models even if sample size is large too.
- Corrected Akaike Information Criterion (**AICc**)
- $AICc = AIC + \frac{2K(K+1)}{n-K-1}$
- Alternatively, there is the Bayesian Information Criterion (**BIC**):

Information Criterion

- Akaike Information Criterion (**AIC**), penalising number of parameters:
- $AIC = -2ln(L) + 2K$
- However, this penalises all high K models even if sample size is large too.
- Corrected Akaike Information Criterion (**AICc**)
- $AICc = AIC + \frac{2K(K+1)}{n-K-1}$
- Alternatively, there is the Bayesian Information Criterion (**BIC**):
- $BIC = -2ln(L) + Kln(n)$

Information Criterion

- Akaike Information Criterion (AIC), penalising number of parameters:
- $AIC = -2ln(L) + 2K$
- However, this penalises all high K models even if sample size is large too.
- Corrected Akaike Information Criterion (AICc)
- $AICc = AIC + \frac{2K(K+1)}{n-K-1}$
- Alternatively, there is the Bayesian Information Criterion (BIC):
- $BIC = -2ln(L) + Kln(n)$
- Decision Theory (DT) risk minimisation approach.

Limitations

- What if everything fits poorly?

Limitations

- What if everything fits poorly?
- Information criterion test relative goodness of fit instead of absolute

Limitations

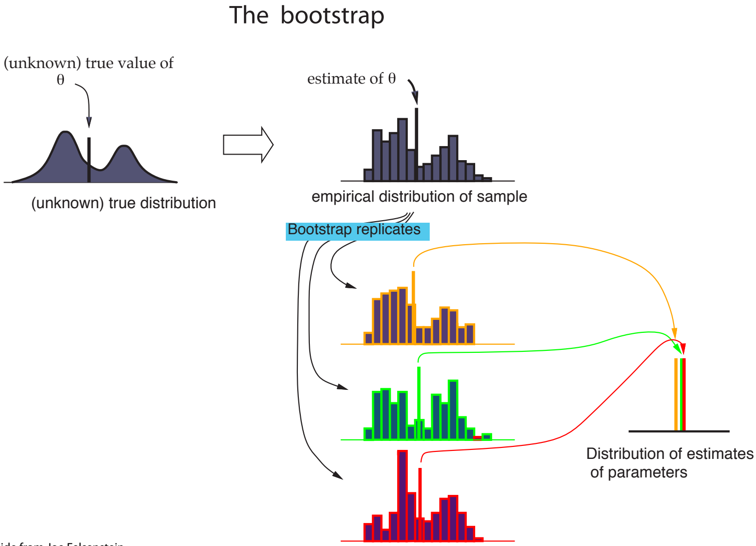
- What if everything fits poorly?
- Information criterion test relative goodness of fit instead of absolute
- Parametric Bootstrapping/Posterior Predictive Simulation

Limitations

- What if everything fits poorly?
- Information criterion test relative goodness of fit instead of absolute
- Parametric Bootstrapping/Posterior Predictive Simulation
- If the model is reasonable then data simulated under should resemble the empirical data

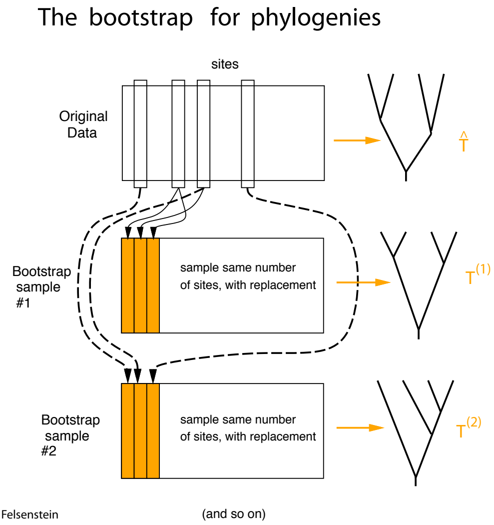
Branch Support Testing

Bootstrapping in General



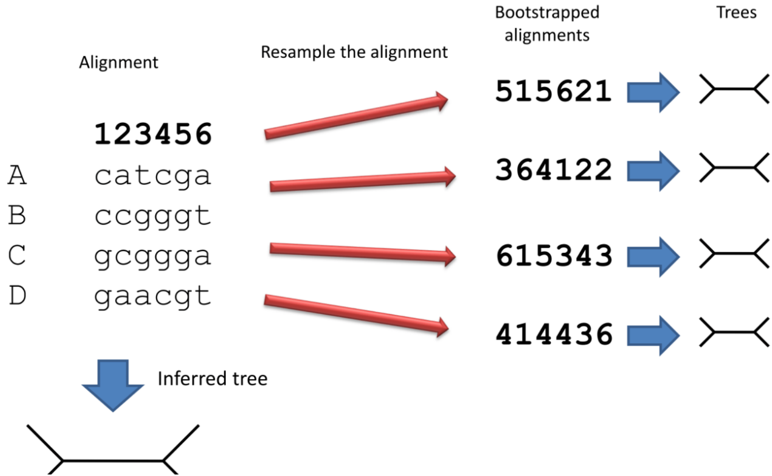
Slide from Joe Felsenstein

Bootstrapping Phylogenies



Slide from Joe Felsenstein

Bootstrapping Phylogenies



Alignment

	123456
A	catcga
B	ccgggt
C	gcggga
D	gaacgt

Resample the alignment

Bootstrapped alignments

Trees

515621

364122

615343

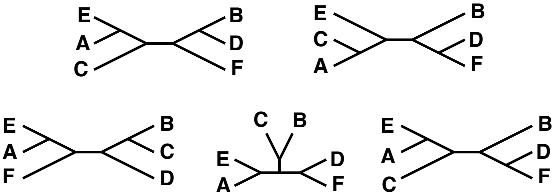
414436

Inferred tree

Bootstrapping Phylogenies

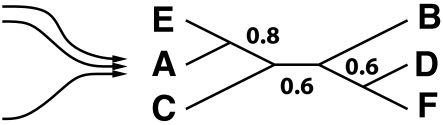
The majority-rule consensus tree

Trees:



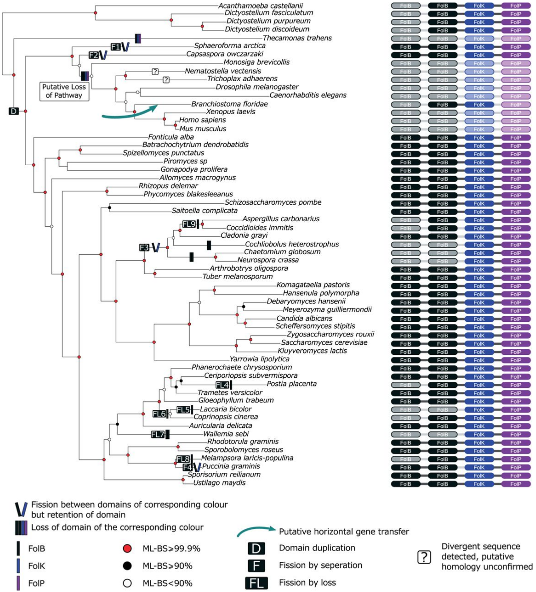
How many times each partition of species is found:

AE I BCDF	4
ACE I BDF	3
ACEF I BD	1
AC I BDEF	1
AEF I BCD	1
ADEF I BC	2
ABCE I DF	3



Slide from Joe Felsenstein

Combining the results



What is the bootstrap doing?

- Randomly reweighing the sites in an alignments

What is the bootstrap doing?

- Randomly reweighing the sites in an alignments
- Probability of a site being excluded $1 - \frac{1}{n}n$

What is the bootstrap doing?

- Randomly reweighing the sites in an alignments
- Probability of a site being excluded $1 - \frac{1}{n}n$
- Asymptotically approximately 0.36

What is the bootstrap doing?

- Randomly reweighing the sites in an alignments
- Probability of a site being excluded $1 - \frac{1}{n}n$
- Asymptotically approximately 0.36
- Goal to simulate an infinite population (number of alignment columns)

Limitations

- Typically underestimates the true probabilities

Limitations

- Typically underestimates the true probabilities
- i.e biased but conservative

Limitations

- Typically underestimates the true probabilities
- i.e biased but conservative
- Computationally demanding

Limitations

- Typically underestimates the true probabilities
- i.e biased but conservative
- Computationally demanding
- Assumes independence of sites

Limitations

- Typically underestimates the true probabilities
- i.e biased but conservative
- Computationally demanding
- Assumes independence of sites
- Relies on good input data

Limitations

- Typically underestimates the true probabilities
- i.e biased but conservative
- Computationally demanding
- Assumes independence of sites
- Relies on good input data
- Only answers to what extent does input data support a given part of the tree

Parametric Bootstraps

- Simulate data sets of this size assuming the estimate of the tree is the truth
- Key for many more sophisticated tests.
- Can be used to generate *p*-values, but non-trivial

Alternative Approaches

- Resampling estimated log-likelihoods (RELL)

Alternative Approaches

- Resampling estimated log-likelihoods (RELL)
- Instead of re-doing the full ML inference just re-sample the site $ln(L)$ values and sum

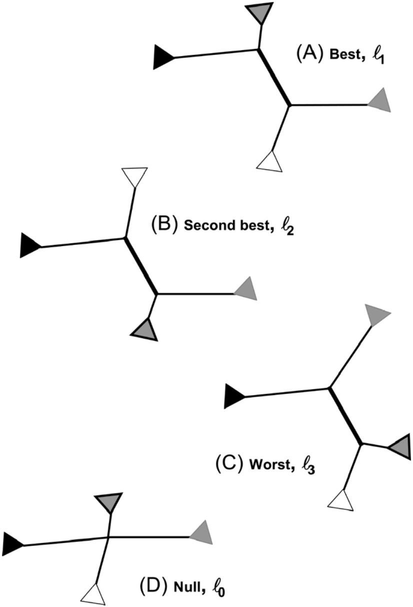
Alternative Approaches

- Resampling estimated log-likelihoods (RELL)
- Instead of re-doing the full ML inference just re-sample the site $ln(L)$ values and sum
- Rapid Bootstraps (RBS)

Alternative Approaches

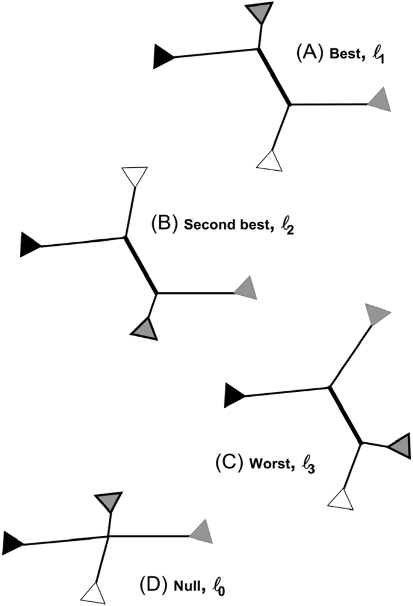
- Resampling estimated log-likelihoods (RELL)
- Instead of re-doing the full ML inference just re-sample the site $ln(L)$ values and sum
- Rapid Bootstraps (RBS)
- Ultrafast Bootstraps (UFBoot)

Likelihood Tests



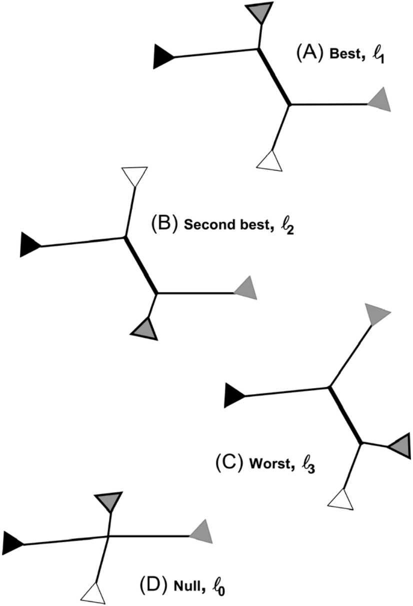
- Comparing the 3 nearest NNIs to a given branch:

Likelihood Tests



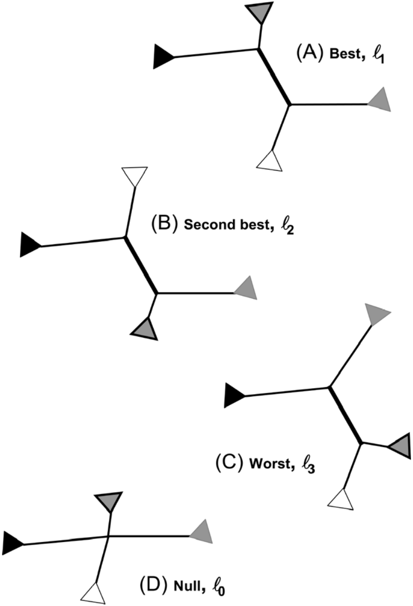
- Comparing the 3 nearest NNIs to a given branch:
- Parametric **aLRT**: χ^2 of δ for branch vs. closest NNIs

Likelihood Tests



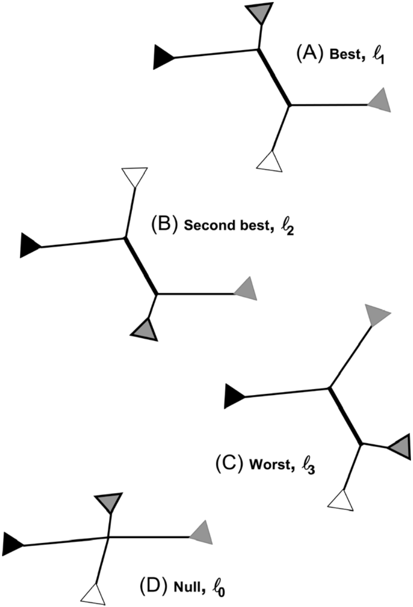
- Comparing the 3 nearest NNIs to a given branch:
- Parametric aLRT: χ^2 of δ for branch vs. closest NNIs
- Non-parametric SH-aLRT based on RELL

Likelihood Tests



- Comparing the 3 nearest NNIs to a given branch:
- Parametric aLRT: χ^2 of δ for branch vs. closest NNIs
- Non-parametric SH-aLRT based on RELL
- aBayes:

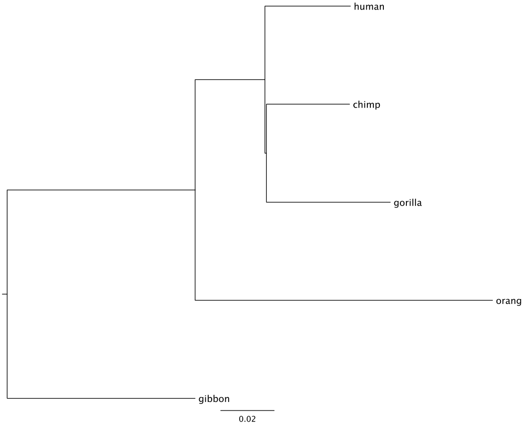
Likelihood Tests



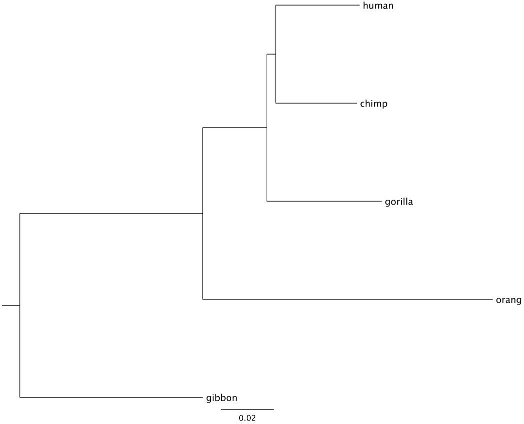
- Comparing the 3 nearest NNIs to a given branch:
- Parametric **aLRT**: χ^2 of δ for branch vs. closest NNIs
- Non-parametric **SH-aLRT** based on RELL
- **aBayes**:
- $P(T_c \mid X) = \frac{P(X|T_c)P(T_c)}{\sum_i^2=0P(X||T_i)P(T_i)}$ with flat prior $P(T_0) = P(T_1) = P(T_2)$

Comparing Trees

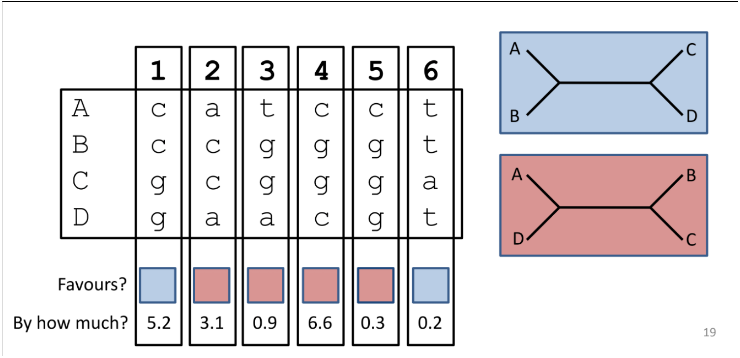
How to compare competing hypotheses?



How to compare competing hypotheses?

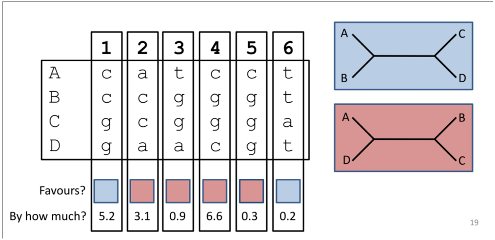


Simplistic Comparison



	1	2	3	4	5	6
A	c	a	t	c	c	t
B	c	c	g	g	g	t
C	g	c	g	g	g	a
D	g	a	a	c	g	t
Favours?	🟦	🟥	🟥	🟥	🟥	🟦
By how much?	5.2	3.1	0.9	6.6	0.3	0.2

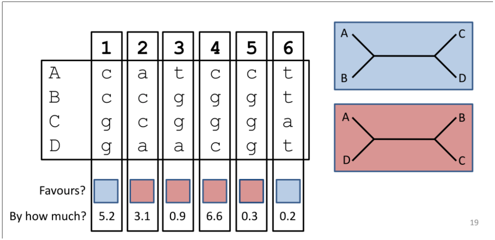
Qualitative Comparison



	1	2	3	4	5	6
A	c	a	t	c	c	t
B	c	c	g	g	g	t
C	g	c	g	g	g	a
D	g	a	a	c	g	t
Favours?	🟦	🟥	🟥	🟥	🟥	🟦
By how much?	5.2	3.1	0.9	6.6	0.3	0.2

- 4 sites favour the red tree, 2 favour the blue

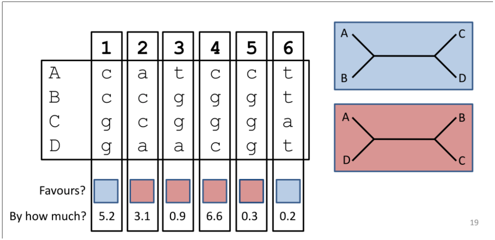
Qualitative Comparison



	1	2	3	4	5	6
A	c	a	t	c	c	t
B	c	c	g	g	g	t
C	g	c	g	g	g	a
D	g	a	a	c	g	t
Favours?	🟦	🟥	🟥	🟥	🟥	🟦
By how much?	5.2	3.1	0.9	6.6	0.3	0.2

- 4 sites favour the red tree, 2 favour the blue
- $\binom{n}{k} p^k (1-p)^{n-k}$

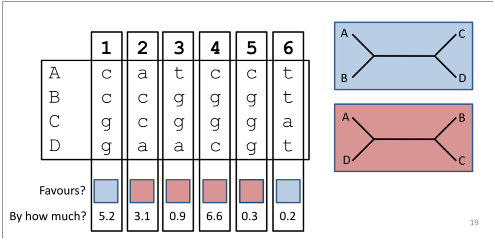
Qualitative Comparison



	1	2	3	4	5	6
A	c	a	t	c	c	t
B	c	c	g	g	g	t
C	g	c	g	g	g	a
D	g	a	a	c	g	t
Favours?	🟦	🟥	🟥	🟥	🟥	🟦
By how much?	5.2	3.1	0.9	6.6	0.3	0.2

- 4 sites favour the red tree, 2 favour the blue
- $\binom{n}{k}p^k(1-p)^{n-k}$
- 4 out of 6 $p = 0.6875$

Qualitative Comparison

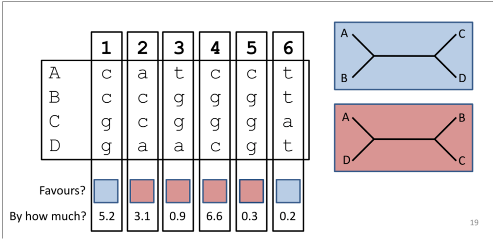


	1	2	3	4	5	6
A	c	a	t	c	c	t
B	c	c	g	g	g	t
C	g	c	g	g	g	a
D	g	a	a	c	g	t
Favours?	🟦	🟥	🟥	🟥	🟥	🟦
By how much?	5.2	3.1	0.9	6.6	0.3	0.2

19

- 4 sites favour the red tree, 2 favour the blue
- $\binom{n}{k} p^k (1-p)^{n-k}$
- 4 out of 6 $p = 0.6875$
- 40 out of 60 $p = 0.0124$

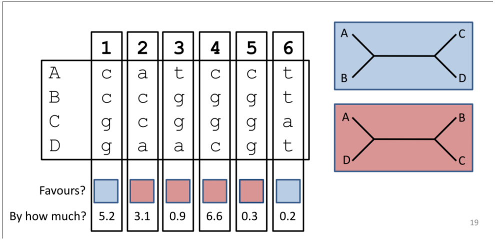
Qualitative Comparison



	1	2	3	4	5	6
A	c	a	t	c	c	t
B	c	c	g	g	g	t
C	g	c	g	g	g	a
D	g	a	a	c	g	t
Favours?	🟦	🟥	🟥	🟥	🟥	🟦
By how much?	5.2	3.1	0.9	6.6	0.3	0.2

- 4 sites favour the red tree, 2 favour the blue
- $\binom{n}{k} p^k (1-p)^{n-k}$
- 4 out of 6 $p = 0.6875$
- 40 out of 60 $p = 0.0124$
- 400 out of 600 $p = 2.3 * 10^{-16}$

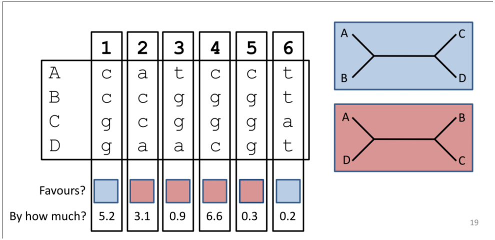
Quantiative Comparison



	1	2	3	4	5	6
A	c	a	t	c	c	t
B	c	c	g	g	g	t
C	g	c	g	g	g	a
D	g	a	a	c	g	t
Favours?	🟦	🟥	🟥	🟥	🟥	🟦
By how much?	5.2	3.1	0.9	6.6	0.3	0.2

- $\mu = (-5.2 + 3.1 + 0.9 + 6.6 + 0.3 - 0.2)/6 = 0.916$

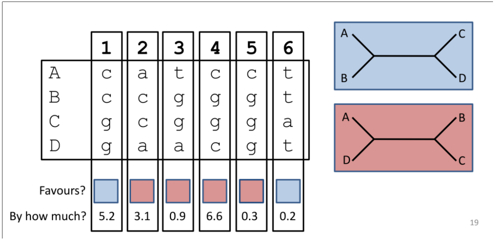
Quantiative Comparison



	1	2	3	4	5	6
A	c	a	t	c	c	t
B	c	c	g	g	g	t
C	g	c	g	g	g	a
D	g	a	a	c	g	t
Favours?	🟦	🟥	🟥	🟥	🟥	🟦
By how much?	5.2	3.1	0.9	6.6	0.3	0.2

- $\mu = (-5.2 + 3.1 + 0.9 + 6.6 + 0.3 - 0.2)/6 = 0.916$
- $\sigma^2 = 15.22$

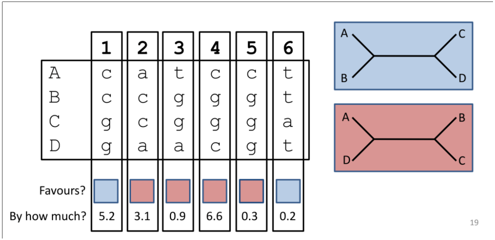
Quantiative Comparison



	1	2	3	4	5	6
A	c	a	t	c	c	t
B	c	c	g	g	g	t
C	g	c	g	g	g	a
D	g	a	a	c	g	t
Favours?	🟦	🟥	🟥	🟥	🟥	🟦
By how much?	5.2	3.1	0.9	6.6	0.3	0.2

- $\mu = (-5.2 + 3.1 + 0.9 + 6.6 + 0.3 - 0.2)/6 = 0.916$
- $\sigma^2 = 15.22$
- $t = \frac{\mu}{\sigma^2} * \sqrt{N} = 0.148$

Quantiative Comparison



	1	2	3	4	5	6
A	c	a	t	c	c	t
B	c	c	g	g	g	t
C	g	c	g	g	g	a
D	g	a	a	c	g	t
Favours?	🟦	🟥	🟥	🟥	🟥	🟦
By how much?	5.2	3.1	0.9	6.6	0.3	0.2

- $\mu = (-5.2 + 3.1 + 0.9 + 6.6 + 0.3 - 0.2)/6 = 0.916$
- $\sigma^2 = 15.22$
- $t = \frac{\mu}{\sigma^2} * \sqrt{N} = 0.148$
- therefore: $p = 0.888$ under $5 d.f.$

More robust approaches

- Null: if no sampling error (infinite data) T_1 and T_2 would explain the data equally well.

More robust approaches

- Null: if no sampling error (infinite data) T_1 and T_2 would explain the data equally well.
- $\delta(T_1, T_2 \mid X) = 2\left[\ln L(T_1 \mid X) - \ln L(T_2 \mid X)\right]$

More robust approaches

- Null: if no sampling error (infinite data) T_1 and T_2 would explain the data equally well.
- $\delta(T_1, T_2 \mid X) = 2\left[\ln L(T_1 \mid X) - \ln L(T_2 \mid X)\right]$
- Expectation under null $\mathbb{E}\left[\delta(T_1, T_2 \mid X)\right] = 0$

More robust approaches

- Null: if no sampling error (infinite data) T_1 and T_2 would explain the data equally well.
- $\delta(T_1, T_2 \mid X) = 2\left[\ln L(T_1 \mid X) - \ln L(T_2 \mid X)\right]$
- Expectation under null $\mathbb{E}\left[\delta(T_1, T_2 \mid X)\right] = 0$
- Why can't we just use χ^2 to get a critical value for δ?

More robust approaches

- Null: if no sampling error (infinite data) T_1 and T_2 would explain the data equally well.
- $\delta(T_1, T_2 \mid X) = 2\left[\ln L(T_1 \mid X) - \ln L(T_2 \mid X)\right]$
- Expectation under null $\mathbb{E}\left[\delta(T_1, T_2 \mid X)\right] = 0$
- Why can't we just use χ^2 to get a critical value for δ?
- Tree space is difficult.

Estimating variance of the null

- Many avenues:
- Non-parametric bootstrapping
- Parametric bootstrapping
- Related approaches.

Kishino-Hasegawa Test

- First, winning sites test

Kishino-Hasegawa Test

- First, winning sites test
- $H_0 : [\ln L(T_1 \mid X) - \ln L(T_2 \mid X)] = \mathbb{E}\left[\delta(T_1, T_2)\right] = 0$

Kishino-Hasegawa Test

- First, winning sites test
- $H_0 : [\ln L(T_1 \mid X) - \ln L(T_2 \mid X)] = \mathbb{E}\left[\delta(T_1, T_2)\right] = 0$
- $H_a : \mathbb{E}\left[\delta(T_1, T_2)\right] \neq 0$

Kishino-Hasegawa Test

- First, winning sites test
- $H_0 : [\ln L(T_1 \mid X) - \ln L(T_2 \mid X)] = \mathbb{E}\left[\delta(T_1, T_2)\right] = 0$
- $H_a : \mathbb{E}\left[\delta(T_1, T_2)\right] \neq 0$
- Non-parametric Bootstrap to estimate Null variance

Kishino-Hasegawa Test

- First, winning sites test
- $H_0 : [\ln L(T_1 \mid X) - \ln L(T_2 \mid X)] = \mathbb{E}\left[\delta(T_1, T_2)\right] = 0$
- $H_a : \mathbb{E}\left[\delta(T_1, T_2)\right] \neq 0$
- Non-parametric Bootstrap to estimate Null variance
- Test $\mathbb{E}\left[\delta(T_1, T_2)\right]$ two-tail t-test

Kishino-Hasegawa Test

- First, winning sites test
- $H_0 : [\ln L(T_1 \mid X) - \ln L(T_2 \mid X)] = \mathbb{E}\left[\delta(T_1, T_2)\right] = 0$
- $H_a : \mathbb{E}\left[\delta(T_1, T_2)\right] \neq 0$
- Non-parametric Bootstrap to estimate Null variance
- Test $\mathbb{E}\left[\delta(T_1, T_2)\right]$ two-tail t-test
- Due to centring assumption can’t be used for optimal tree i.e. selection bias

Kishino-Hasegawa Test

- First, winning sites test
- $H_0 : [\ln L(T_1 \mid X) - \ln L(T_2 \mid X)] = \mathbb{E}\left[\delta(T_1, T_2)\right] = 0$
- $H_a : \mathbb{E}\left[\delta(T_1, T_2)\right] \neq 0$
- Non-parametric Bootstrap to estimate Null variance
- Test $\mathbb{E}\left[\delta(T_1, T_2)\right]$ two-tail t-test
- Due to centring assumption can't be used for optimal tree i.e. selection bias
- Can't handle multiple comparisons.

Alternative tests

- Shimodaira-Hasegawa Test

Alternative tests

- Shimodaira-Hasegawa Test
- Compares candidate tree sets

Alternative tests

- Shimodaira-Hasegawa Test
- Compares candidate tree sets
- $H_0 =$ all topologies equally good

Alternative tests

- Shimodaira-Hasegawa Test
- Compares candidate tree sets
- $H_0 =$ all topologies equally good
- Very conservative when the number of candidate trees is large

Alternative tests

- **Shimodaira-Hasegawa Test**
- Compares candidate tree sets
- $H_0 =$ all topologies equally good
- Very conservative when the number of candidate trees is large
- Can be corrected with weighted SH-test overcomes.

Alternative tests

- **Shimodaira-Hasegawa Test**
- Compares candidate tree sets
- $H_0 =$ all topologies equally good
- Very conservative when the number of candidate trees is large
- Can be corrected with weighted SH-test overcomes.
- **Approximately Unbiased Test**

Alternative tests

- **Shimodaira-Hasegawa Test**
- Compares candidate tree sets
- $H_0 =$ all topologies equally good
- Very conservative when the number of candidate trees is large
- Can be corrected with weighted SH-test overcomes.
- **Approximately Unbiased Test**
- Achieves weighted by varying bootstrap size for each tree.

Alternative tests

- Shimodaira-Hasegawa Test
- Compares candidate tree sets
- $H_0 =$ all topologies equally good
- Very conservative when the number of candidate trees is large
- Can be corrected with weighted SH-test overcomes.
- Approximately Unbiased Test
- Achieves weighted by varying bootstrap size for each tree.
- Better for larger comparisons, can have issues with P-space curvature.

Alternative tests

- **Shimodaira-Hasegawa Test**
- Compares candidate tree sets
- $H_0 =$ all topologies equally good
- Very conservative when the number of candidate trees is large
- Can be corrected with weighted SH-test overcomes.
- **Approximately Unbiased Test**
- Achieves weighted by varying bootstrap size for each tree.
- Better for larger comparisons, can have issues with P-space curvature.
- **Swofford–Olsen–Waddell–Hillis** same idea but uses parametric bootstraps instead.

Alternative tests

- **Shimodaira-Hasegawa Test**
- Compares candidate tree sets
- $H_0 =$ all topologies equally good
- Very conservative when the number of candidate trees is large
- Can be corrected with weighted SH-test overcomes.
- **Approximately Unbiased Test**
- Achieves weighted by varying bootstrap size for each tree.
- Better for larger comparisons, can have issues with P-space curvature.
- **Swofford–Olsen–Waddell–Hillis** same idea but uses parametric bootstraps instead.
- Sensitive to model misspecification.

Conclusion

Summary

- Tree space makes for some interesting problems that takes away some standard statistical tricks.

Summary

- Tree space makes for some interesting problems that takes away some standard statistical tricks.
- Model selection typically relies on multiple metrics

Summary

- Tree space makes for some interesting problems that takes away some standard statistical tricks.
- Model selection typically relies on multiple metrics
- Bootstrapping is a slow, biased but conservative way to estimate the support for a given branch in your tree.

Summary

- Tree space makes for some interesting problems that takes away some standard statistical tricks.
- Model selection typically relies on multiple metrics
- Bootstrapping is a slow, biased but conservative way to estimate the support for a given branch in your tree.
- Likelihood Testing is powerful but must be used with care.

Summary

- Tree space makes for some interesting problems that takes away some standard statistical tricks.
- Model selection typically relies on multiple metrics
- Bootstrapping is a slow, biased but conservative way to estimate the support for a given branch in your tree.
- Likelihood Testing is powerful but must be used with care.
- Comparing trees directly is non-trivial due to tree-space.

Questions?

References i

Leonard, G. (2010).
Development of fusion and duplication finder blast (fdfblast): a systematic tool to detect differentially distributed gene fusions and resolve trifurcations in the tree of life.

Marwick, B. (2012).
A cladistic evaluation of ancient thai bronze buddha images: six tests for a phylogenetic signal in the griswold collection.
Connecting empires, pages 159–176.

Nickle, D. C., Heath, L., Jensen, M. A., Gilbert, P. B., Mullins, J. I., and Pond, S. L. K. (2007).
Hiv-specific probabilistic models of protein evolution.
PLoS One, 2(6):e503.

References ii

Richards, T. A., Soanes, D. M., Foster, P. G., Leonard, G., Thornton, C. R., and Talbot, N. J. (2009).
Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi.
The Plant Cell, 21(7):1897–1911.