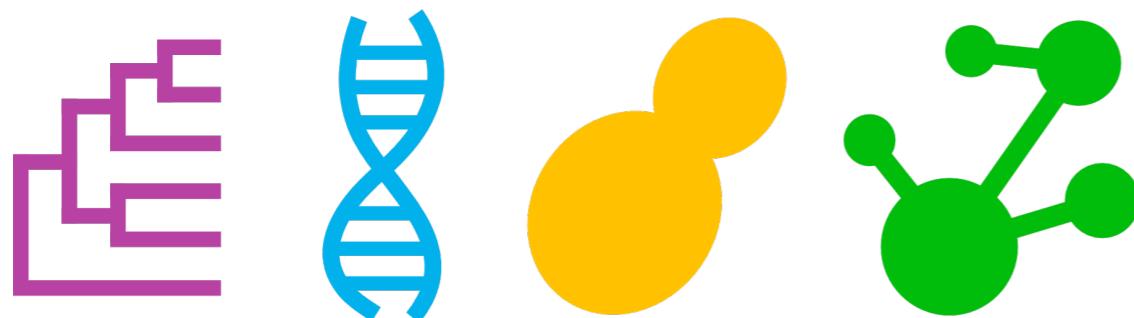


Internode certainty and related measures



Jacob L. Steenwyk



jlsteenwyk.github.io



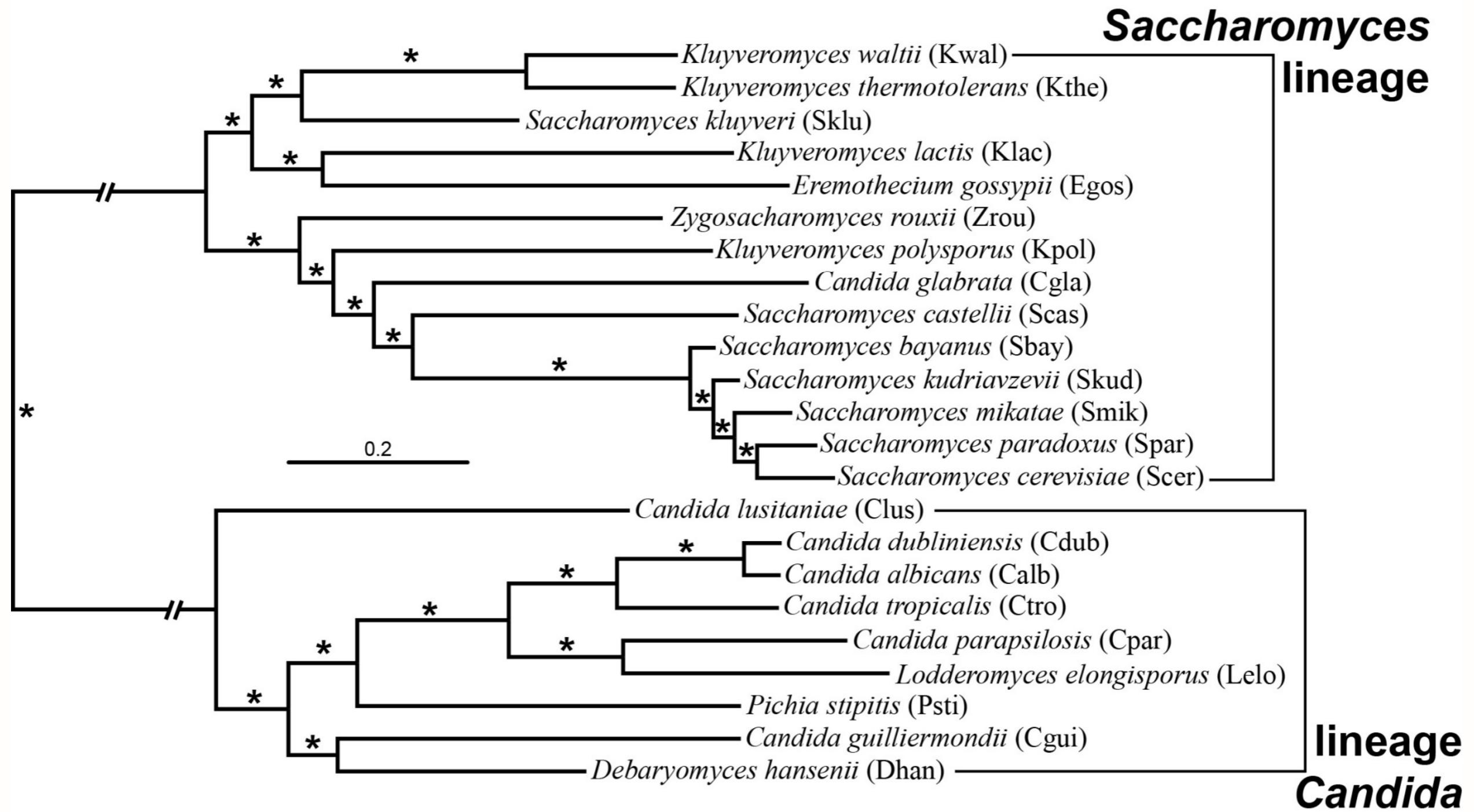
@JLSteenwyk

A refresher....

A refresher....

**The next few slides
are from Antonis Rokas**

Concatenation Yields an Absolutely Supported Phylogeny



Bootstrap Support is Misleading When Used in Large Datasets

53% 47%

A sequence alignment of four taxa (taxonA, taxonB, taxonC, taxonD) showing their DNA sequences. The sequences are: taxonA: AAAAAAAAATTTTTTTT; taxonB: AAAAAAAAACCCCCCCC; taxonC: GGGGGGGGGGTTTTTTTT; taxonD: GGGGGGGGGGCCCCCCCCC. Above the sequences, two brackets indicate bootstrap support: a bracket spanning the first seven positions is labeled '53%', and a bracket spanning the last three positions is labeled '47%'.

taxonA	AAAAAAAATTTTTTTT
taxonB	AAAAAAAACCCCCCCC
taxonC	GGGGGGGGGGTTTTTTTT
taxonD	GGGGGGGGGGGCCCCCCCCC

Bootstrap Support is Misleading When Used in Large Datasets

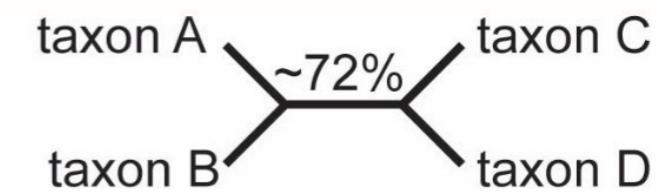


Bootstrap Support is Misleading When Used in Large Datasets

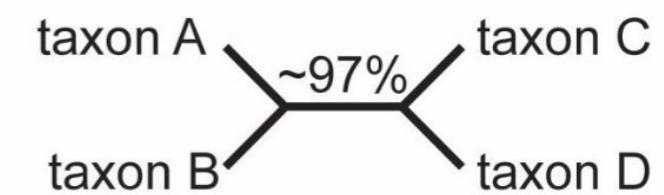
taxonA AAAAAAAAATTTTTTTT
taxonB AAAAAAAAACCCCCCCC
taxonC GGGGGGGGGGTTTTTTTT
taxonD GGGGGGGGGGCCCCCCCCC

53% 47%

100 characters



1,000 characters



Bootstrap Support is Misleading When Used in Large Datasets

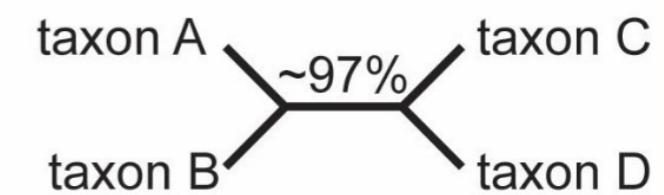
taxonA AAAAAAAAATTTTTTTT
taxonB AAAAAAAAACCCCCCCC
taxonC GGGGGGGGGGTTTTTTTT
taxonD GGGGGGGGGGCCCCCCCCC

53% 47%

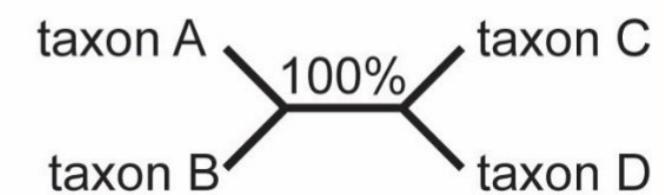
100 characters



1,000 characters



10,000 characters

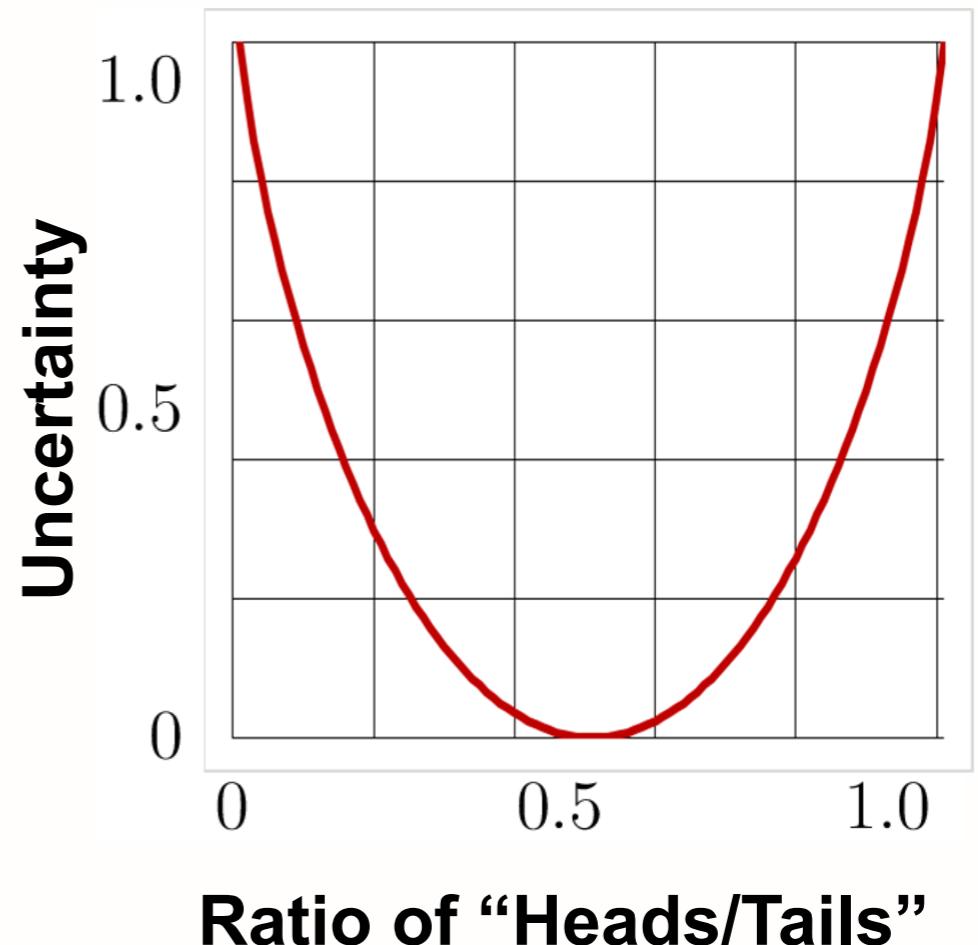


Quantifying Incongruence

Internode Certainty (IC): a measure of the support for a given internode by considering its frequency in a given set of trees jointly with that of the most prevalent conflicting internode in the same set of trees

Tree Certainty (TC): the sum of IC across all internodes

IC and TC are implemented in the latest versions of RAxML



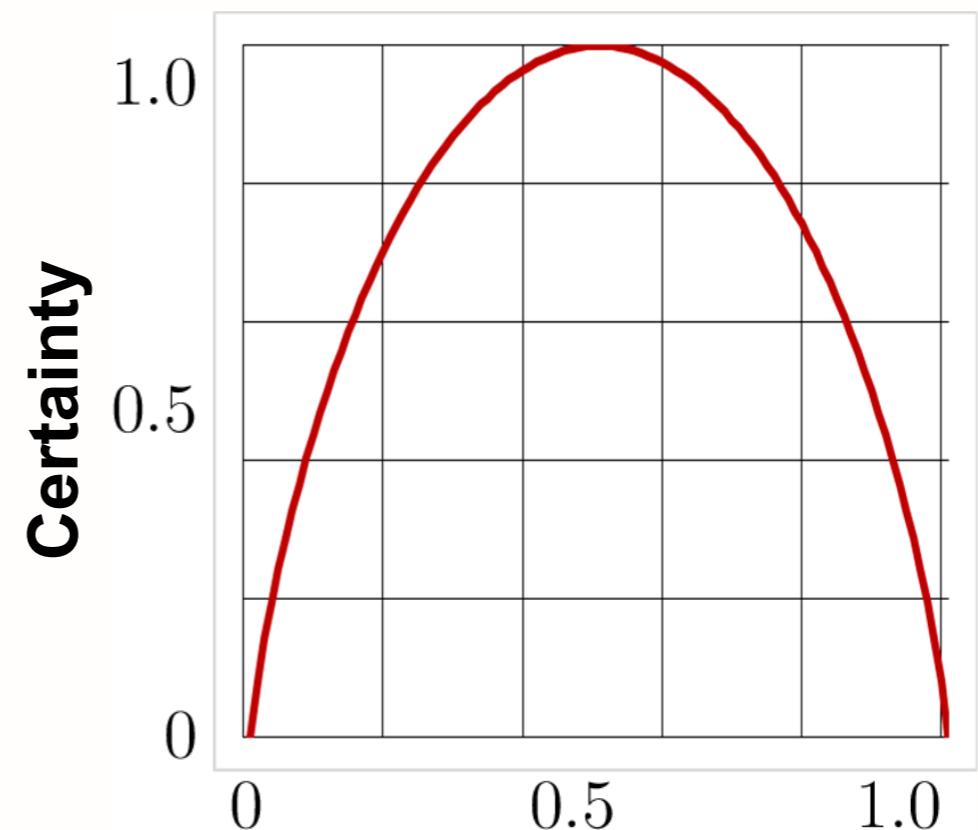
Saldanha et al. (2014) *Mol. Biol. Evol.*; Kober et al. (2016) *Mol. Biol. Evol.*;
Zhou et al. (2018) *bioRxiv*

Quantifying Incongruence

Internode Certainty (IC): a measure of the support for a given internode by considering its frequency in a given set of trees jointly with that of the most prevalent conflicting internode in the same set of trees

Tree Certainty (TC): the sum of IC across all internodes

IC and TC are implemented in the latest versions of RAxML



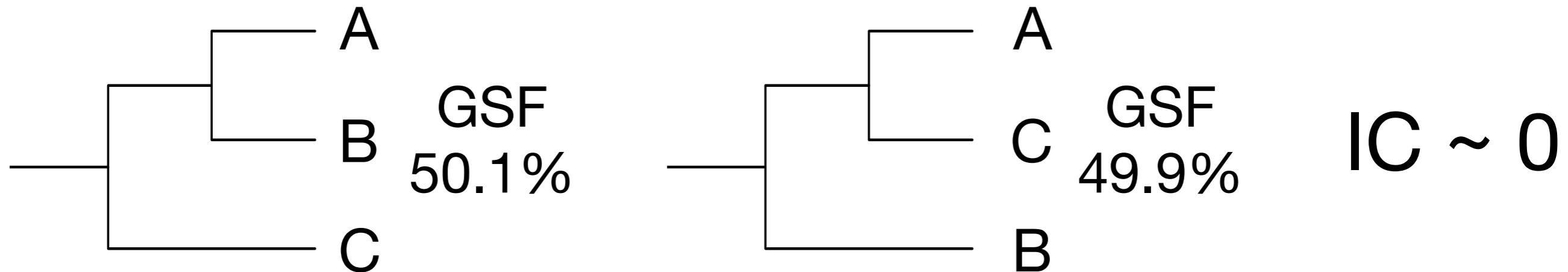
Ratio of Support for Two Conflicting Internodes



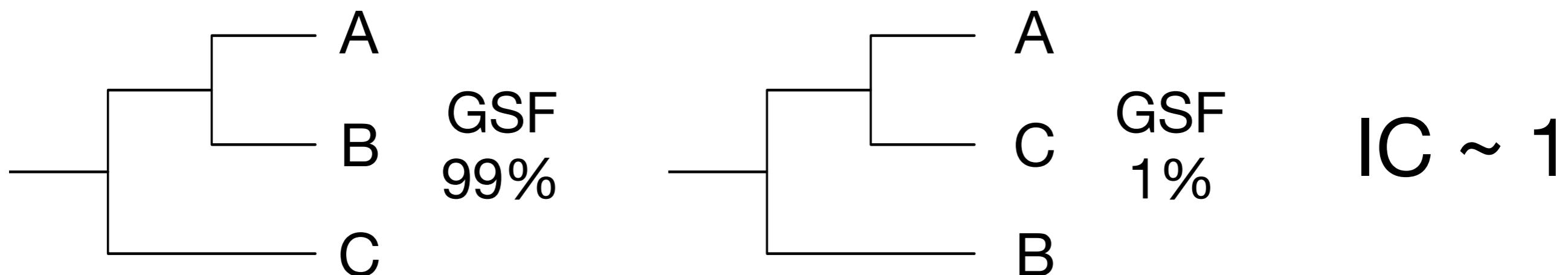
Schliebes et al. (2014) *Mol. Biol. Evol.*; Koberst et al. (2016) *Mol. Biol. Evol.*;
Zhou et al. (2018) *bioRxiv*

Internode certainty (in other words...)

Case 1: High conflict



Case 2: Low conflict



Developments of internode certainty

- The original implementation was originally developed for phylogenies with complete taxon representation
Salichos and Rokas (2013) Nature
Salichos et al. (2014) Mol. Biol. Evol.

Developments of internode certainty

- The original implementation was originally developed for phylogenies with complete taxon representation
Salichos and Rokas (2013) Nature
Salichos et al. (2014) Mol. Biol. Evol.
- Corrections for partial taxon representation has been implemented in current versions of RAxML.
Kobert et al. (2016) Mol. Biol. Evol.

Developments of internode certainty

- The original implementation was originally developed for phylogenies with complete taxon representation
Salichos and Rokas (2013) Nature
Salichos et al. (2014) Mol. Biol. Evol.
- Corrections for partial taxon representation has been implemented in current versions of RAxML.
Kobert et al. (2016) Mol. Biol. Evol.
- Quartet based IC measures, QuartetScores, are more accurate with partial gene trees but is currently under review
Zhou et al. (2018) bioRxiv

Notes on implementation

- Verbose usage of RAxML's calculations of IC provides detailed information about conflicting bipartitions

RAxML_verboseSplits.suffix

Notes on implementation

- Verbose usage of RAxML's calculations of IC provides detailed information about conflicting bipartitions

RAxML_verboseSplits.suffix

```
1. Uncinocarpus_reesii
2. Coccidioides_posadasii
3. Penicilliopsis_zonata
4. Xeromyces_bisporus
5. Monascus_ruber
6. Penicillium_camemberti
7. Penicillium_digitatum
8. Penicillium_roqueforti
9. Aspergillus_fumigatus
10. Aspergillus_niger
11. Aspergillus_oryzae
```

partition:

```
---** ----- - 1189/92.385392/0.850774
--*-* ***** * 26/2.020202/0.850774
```



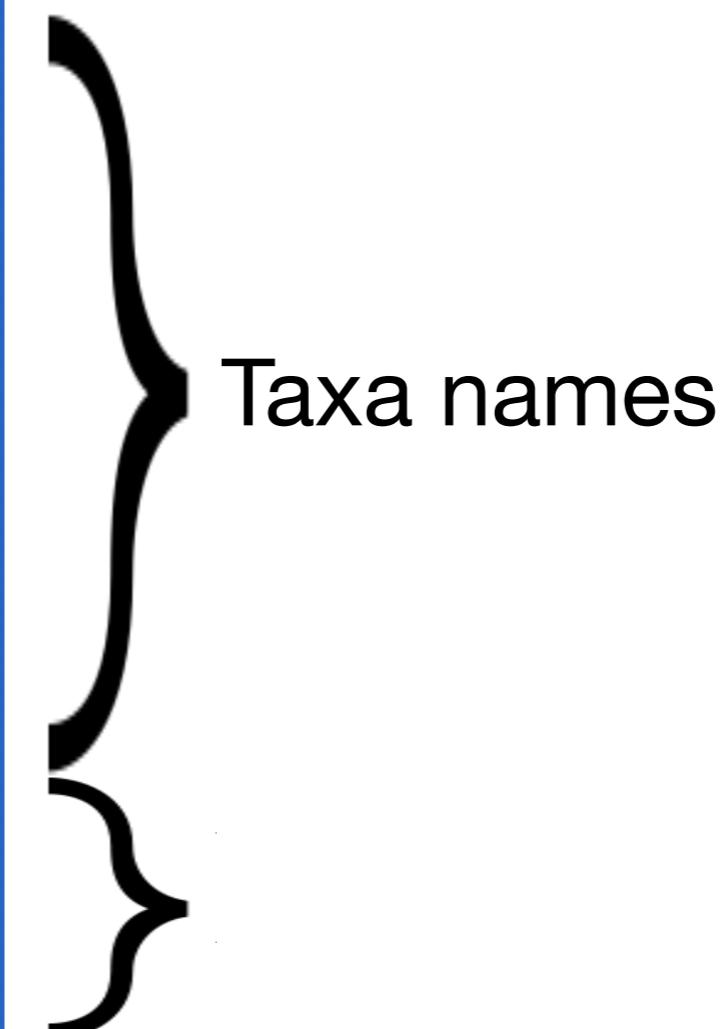
Notes on implementation

- Verbose usage of RAxML's calculations of IC provides detailed information about conflicting bipartitions
RAxML_verboseSplits.suffix

```
1. Uncinocarpus_reesii
2. Coccidioides_posadasii
3. Penicilliopsis_zonata
4. Xeromyces_bisporus
5. Monascus_ruber
6. Penicillium_camemberti
7. Penicillium_digitatum
8. Penicillium_roqueforti
9. Aspergillus_fumigatus
10. Aspergillus_niger
11. Aspergillus_oryzae
```

partition:

```
---** ----- - 1189/92.385392/0.850774
--*-* ***** * 26/2.020202/0.850774
```



Taxa names

Notes on implementation

- Verbose usage of RAxML's calculations of IC provides detailed information about conflicting bipartitions

RAxML_verboseSplits.suffix

```
1. Uncinocarpus_reesii
2. Coccidioides_posadasii
3. Penicilliopsis_zonata
4. Xeromyces_bisporus
5. Monascus_ruber
6. Penicillium_camemberti
7. Penicillium_digitatum
8. Penicillium_roqueforti
9. Aspergillus_fumigatus
10. Aspergillus_niger
11. Aspergillus_oryzae
```

partition:

```
---** ----- - 1189/92.385392/0.850774
---** ***** * 26/2.020202/0.850774
```



Taxa names

Partition information
xx/yy/zz

xx = Trees supporting ref.

yy = gene support freq.

zz = Internode certainty

Notes on implementation

- Verbose usage of RAxML's calculations of IC provides detailed information about conflicting bipartitions

RAxML_verboseSplits.suffix

```
1. Uncinocarpus_reesii
2. Coccidioides_posadasii
3. Penicilliopsis_zonata
4. Xeromyces_bisporus
5. Monascus_ruber
6. Penicillium_camemberti
7. Penicillium_digitatum
8. Penicillium_roqueforti
9. Aspergillus_fumigatus
10. Aspergillus_niger
11. Aspergillus_oryzae
```

partition:

```
---** ----- - 1189/92.385392/0.850774
---** ***** * 26/2.020202/0.850774
```

Verbose can only be used with trees that have full taxon representation

Taxa names

Partition information
xx/yy/zz
xx = Trees supporting ref.
yy = gene support freq.
zz = Internode certainty

Notes on implementation

- Exact bipartition topology for a given bipartition can be examined among files with the following syntax

RAxML_verboseIC.suffix.0 ... RAxML_verboseIC.suffix.N-1

1. *Uncinocarpus_reesii*
2. *Coccidioides_posadasii*
3. *Penicilliopsis_zonata*
4. *Xeromyces_bisporus*
5. *Monascus_ruber*
6. *Penicillium_camemberti*
7. *Penicillium_digitatum*
8. *Penicillium_roqueforti*
9. *Aspergillus_fumigatus*
10. *Aspergillus_niger*
11. *Aspergillus_oryzae*

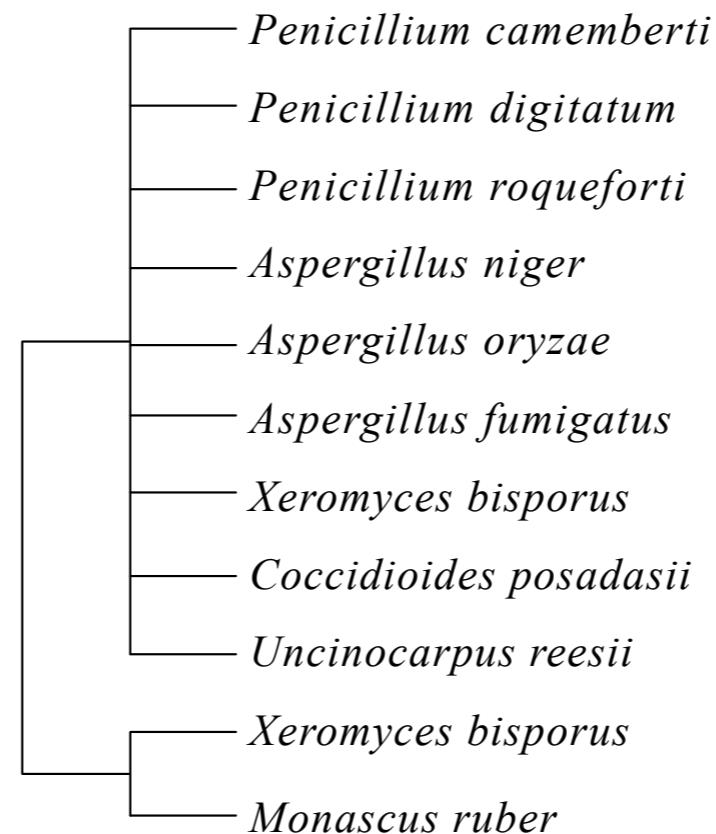
Notes on implementation

- Exact bipartition topology for a given bipartition can be examined among files with the following syntax

RAxML_verboseIC.suffix.0 ... RAxML_verboseIC.suffix.N-1

1. *Uncinocarpus_reesii*
2. *Coccidioides_posadasii*
3. *Penicilliopsis_zonata*
4. *Xeromyces_bisporus*
5. *Monascus_ruber*
6. *Penicillium_camemberti*
7. *Penicillium_digitatum*
8. *Penicillium_roqueforti*
9. *Aspergillus_fumigatus*
10. *Aspergillus_niger*
11. *Aspergillus_oryzae*

Topology 1



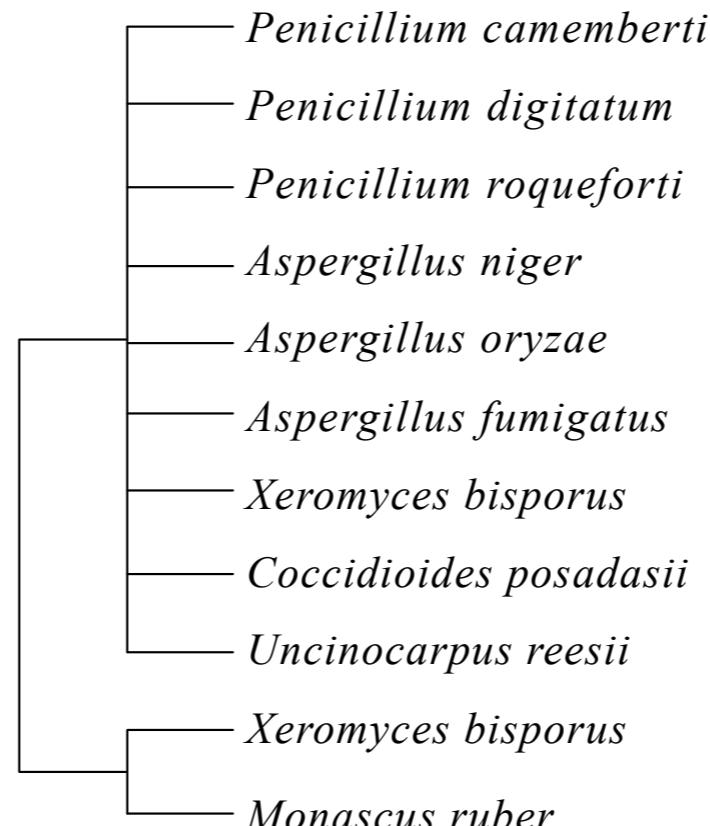
Notes on implementation

- Exact bipartition topology for a given bipartition can be examined among files with the following syntax

RAxML_verboseIC.suffix.0 ... RAxML_verboseIC.suffix.N-1

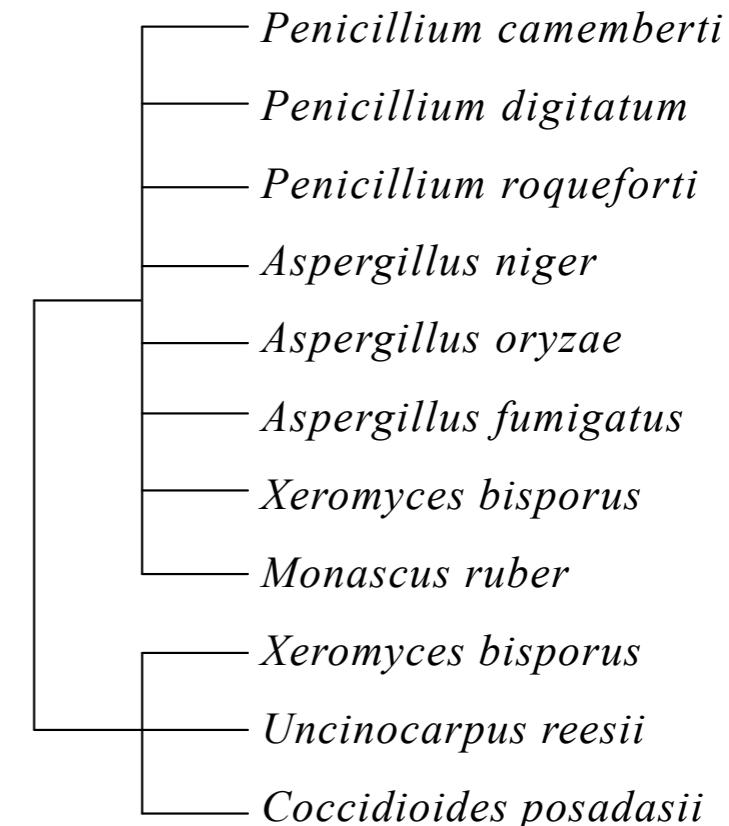
1. *Uncinocarpus_reesii*
2. *Coccidioides_posadasii*
3. *Penicilliopsis_zonata*
4. *Xeromyces_bisporus*
5. *Monascus_ruber*
6. *Penicillium_camemberti*
7. *Penicillium_digitatum*
8. *Penicillium_roqueforti*
9. *Aspergillus_fumigatus*
10. *Aspergillus_niger*
11. *Aspergillus_oryzae*

Topology 1



-----* ----- -
1189/92.385392/0.850774

Topology 2

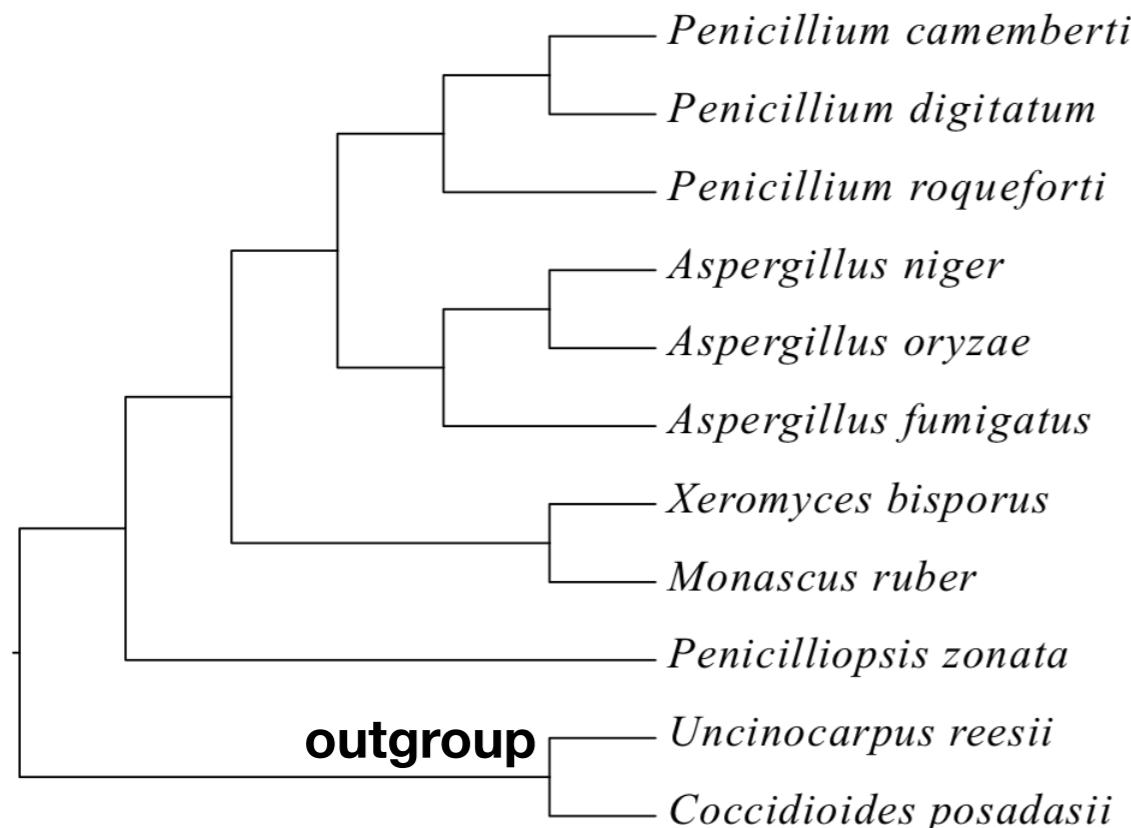


----* **** *
26/2.020202/0.850774

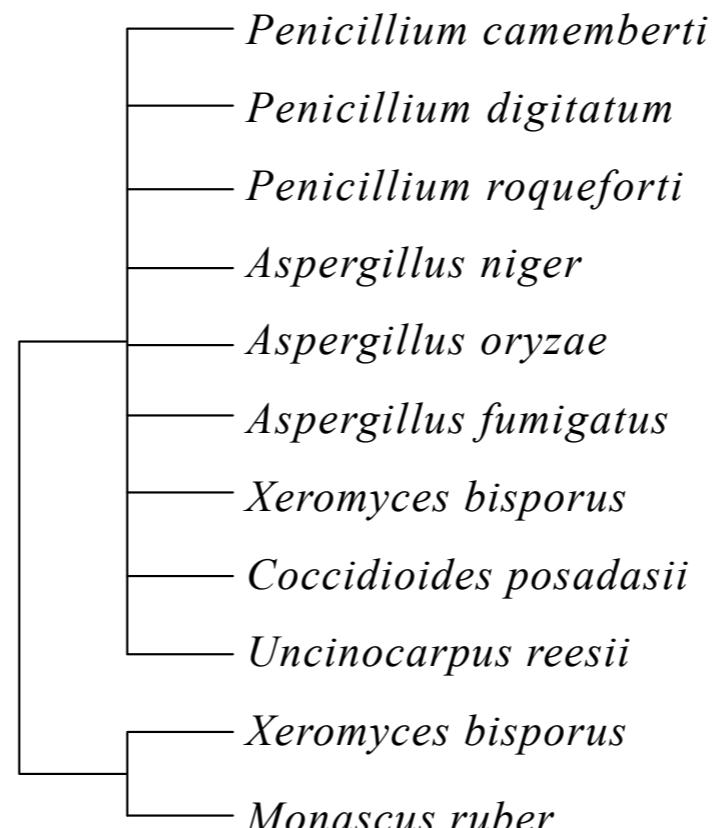
Notes on implementation

- Exact bipartition topology for a given bipartition can be examined among files with the following syntax
RAxML_verboseIC.suffix.0 ... RAxML_verboseIC.suffix.N-1

Concat topology

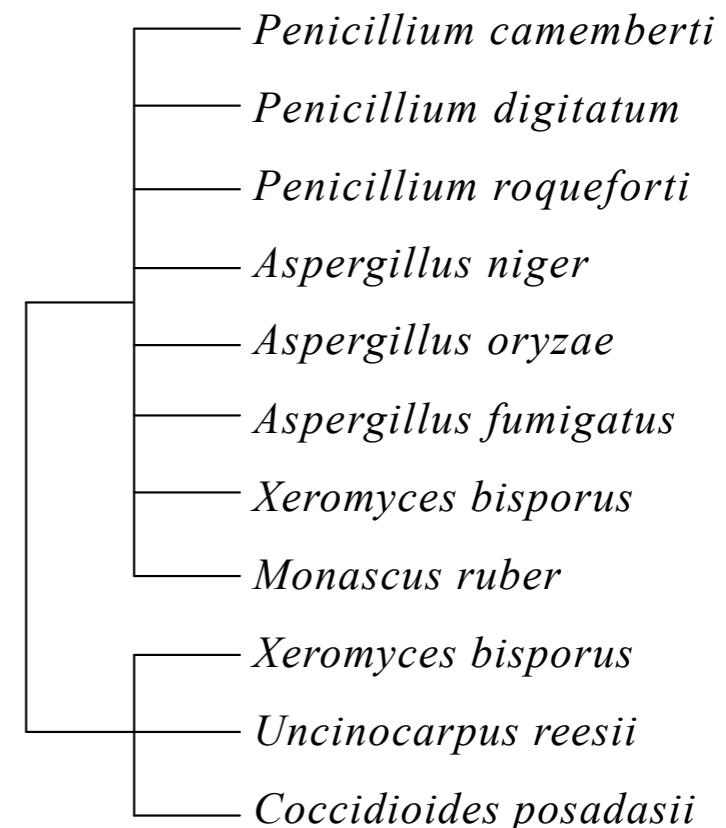


Topology 1



1189/92.385392/0.850774

Topology 2



26/2.020202/0.850774

Other definitions...

IC-AII

- computed by taking into account all conflicting bipartitions with that have $\geq 5\%$ support and not only the most supported conflict

Other definitions...

IC-All

- computed by taking into account all conflicting bipartitions with that have $\geq 5\%$ support and not only the most supported conflict

TC-All

- The sum of IC-All values

Other definitions...

IC-All

- computed by taking into account all conflicting bipartitions with that have $\geq 5\%$ support and not only the most supported conflict

TC-All

- The sum of IC-All values

Relative tree certainty

- A value from 0 (no certainty) to 1 (high certainty)

Concordance factors

Concordance factors

- proportion of the genome for which a given clade is true

Baum (2007) Taxon

Concordance factors

Concordance factors

- proportion of the genome for which a given clade is true

Baum (2007) Taxon

Gene or site concordance factors

- proportion of genes or sites for which a given clade is true
- more precisely, percentage of decisive gene trees (or sites) with a given branch

Minh (2018) bioRxiv

Concordance factors

Concordance factors

- proportion of the genome for which a given clade is true

Baum (2007) Taxon

Gene or site concordance factors

- proportion of genes or sites for which a given clade is true
- more precisely, percentage of decisive gene trees (or sites) with a given branch

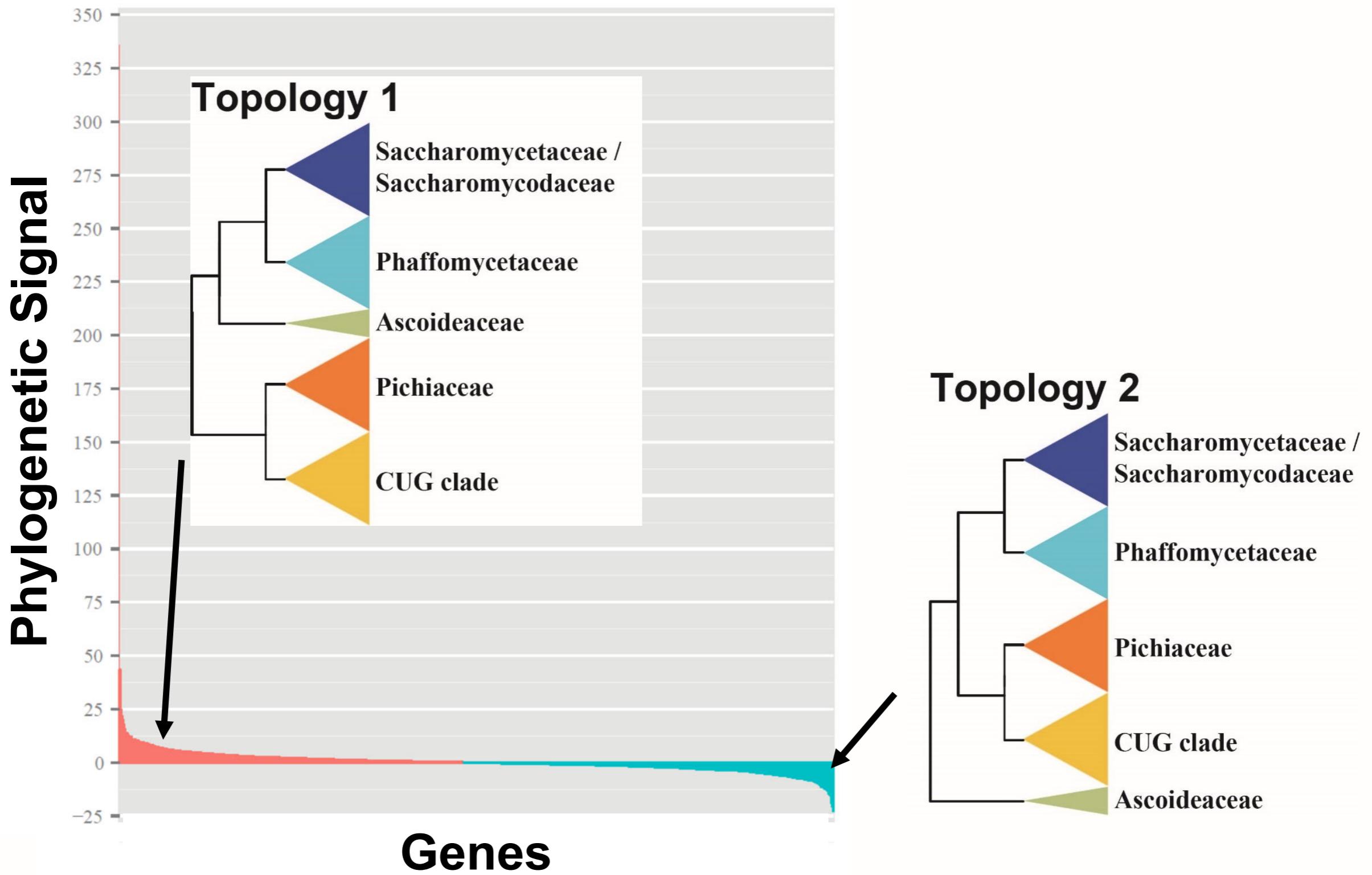
Minh (2018) bioRxiv

**Great additional analysis when
bootstrapping becomes unreliable!!**

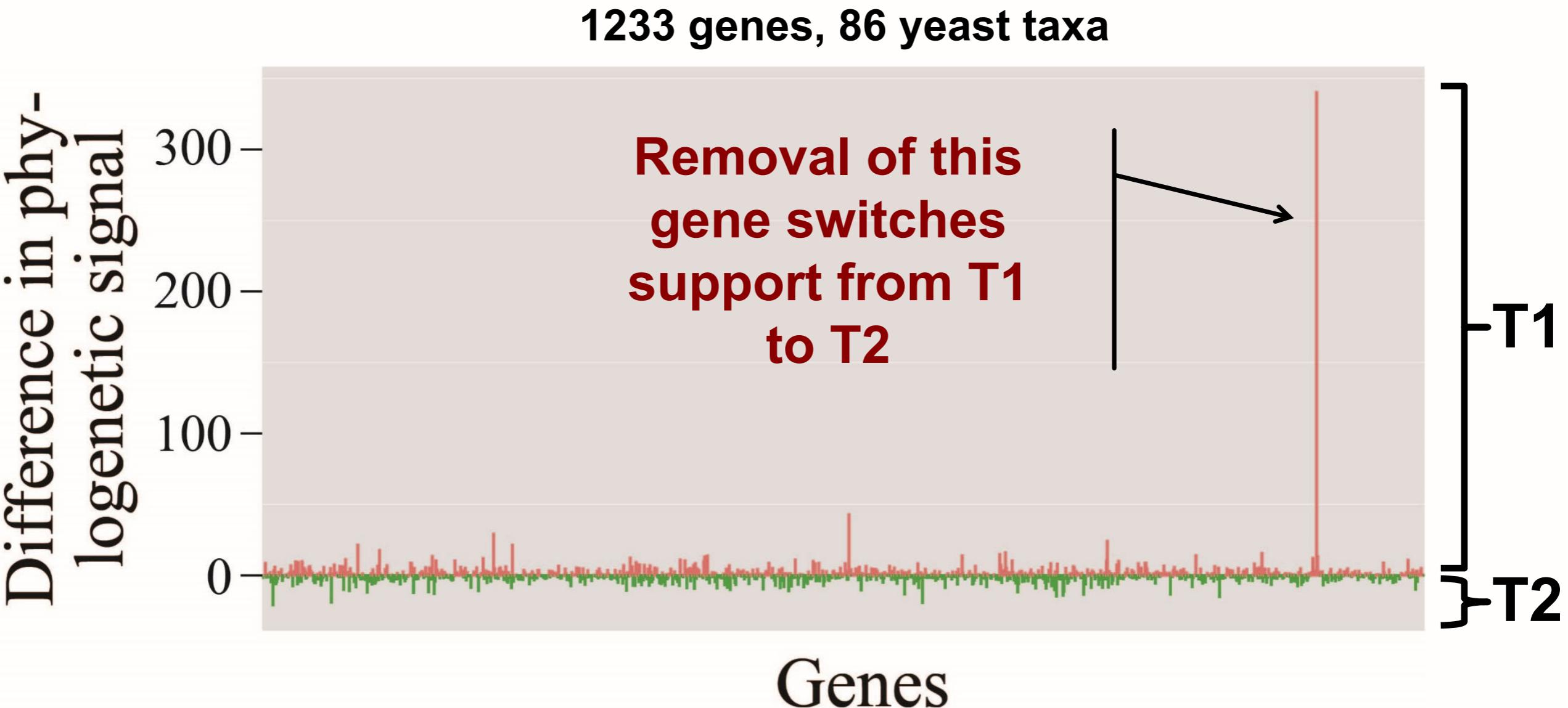
A refresher....

**The next few slides
are from Antonis Rokas**

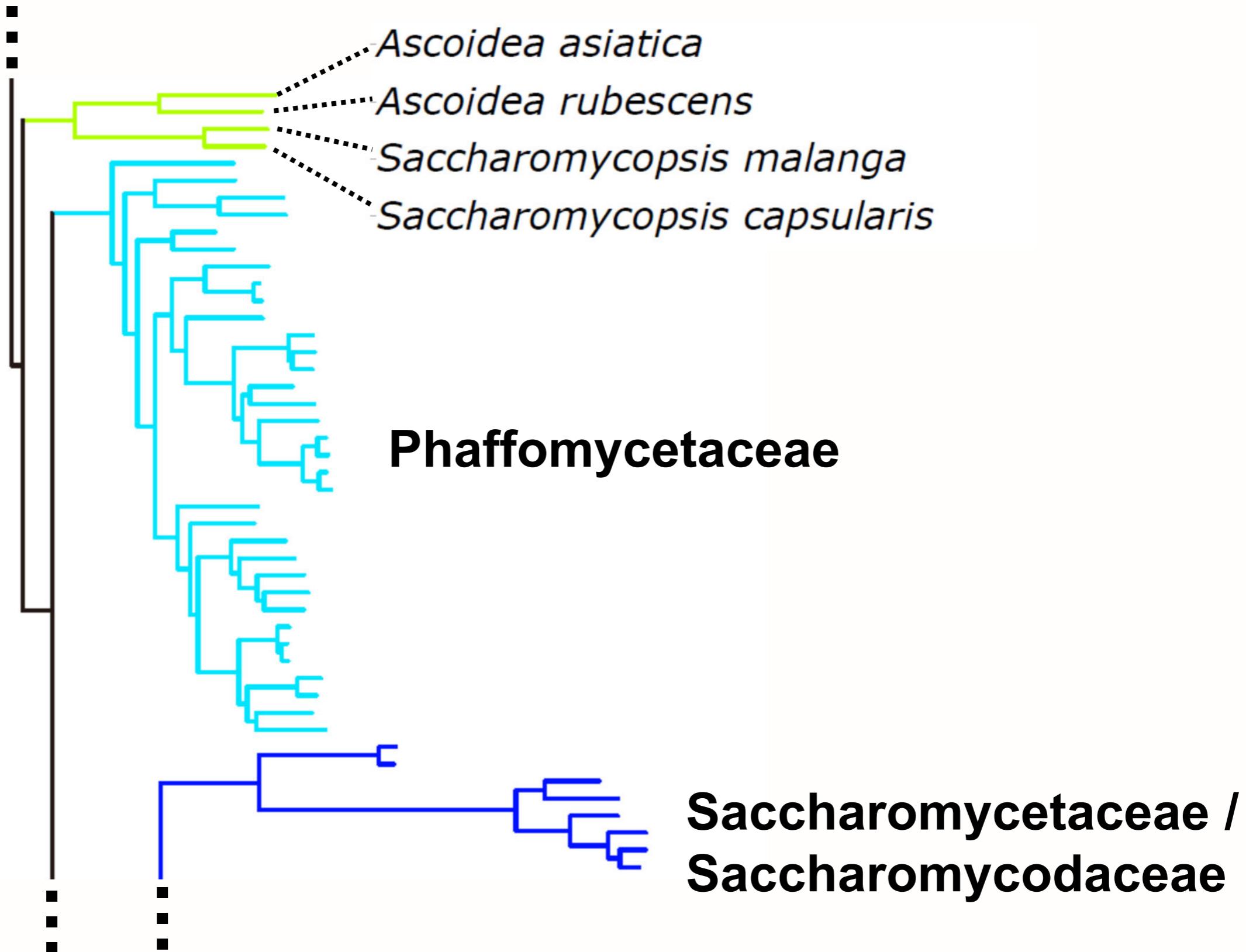
...But in Others It Stems from One or Two Genes



Phylogenetic Signal per Gene for the Two Hypotheses

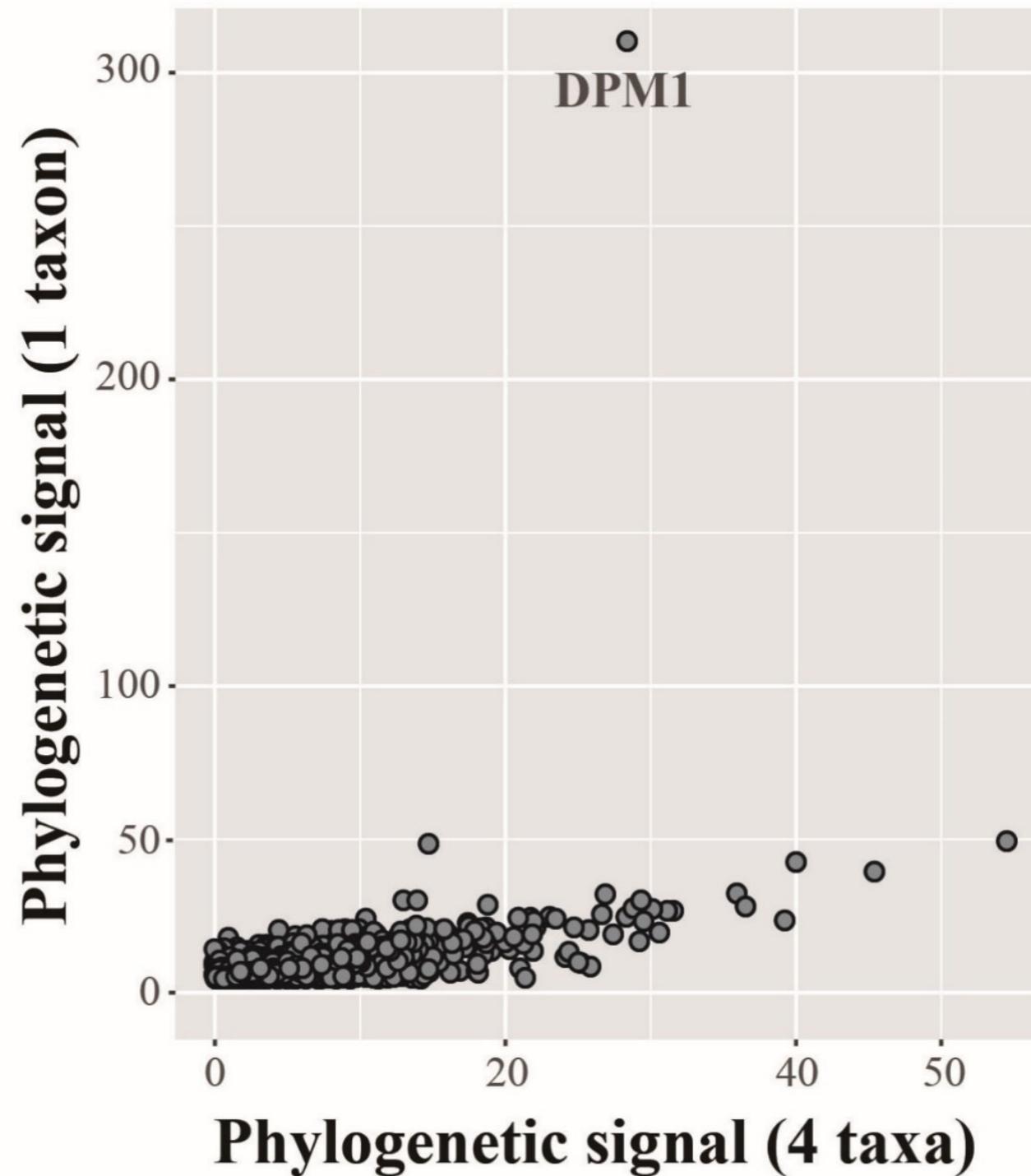


Sampling of 3 Additional Taxa “Breaks” the Long Branch

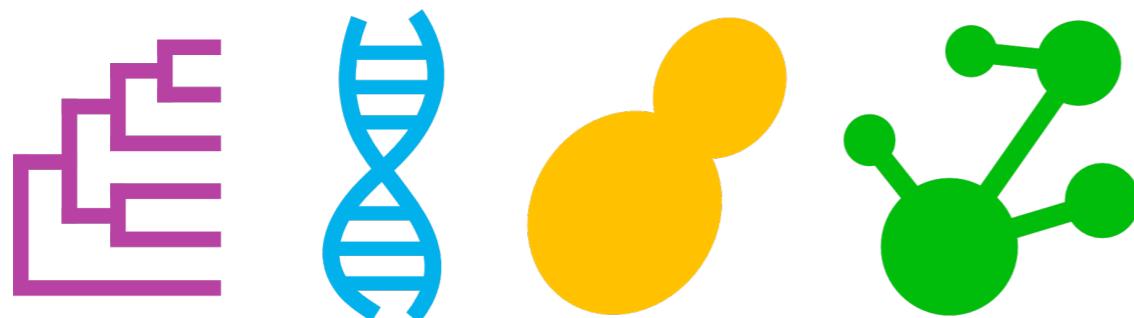


Sampling of 3 Additional Taxa Decreases Gene's Signal

2,408 genes, 329 – 332 yeast taxa



Internode certainty and related measures



Jacob L. Steenwyk



jlsteenwyk.github.io



@JLSteenwyk