IQ-TREE

Efficient software for phylogenomic inference

Stable release 1.6.12 (August 15, 2019)

Download v1.6.12 for macOS

Latest release 2.2.2.6 (May 27, 2023)

Download v2.2.2.6 for macOS

All Downloads | Documentation

IQ-TREE has been developed by 12+ contributors:

From ANU:


James Barbetti


Thomas Wong


Robert Lanfear


Bui Quang Minh


Nhan Ly-Trong


Piyumal Demotte

From international:


Michael Woodhams


Olga Chernomor


Arndt von Haeseler


Dominik Schrempf


Heiko A. Schmidt


Diep Thi Hoang

Past members:

Lam Tung Nguyen

Jana Trifinopoulos

IQ-TREE enables to infer phylogenetic trees of "SARS-CoV-2" virus

For identifying new variants and key mutations for vaccine design

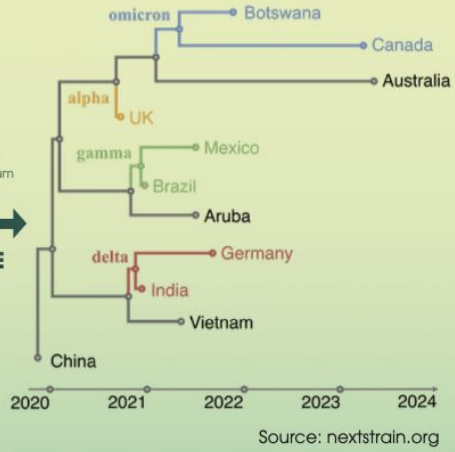
Sequencing

DNA Sequence Alignment

Canada	ATGTTTGTTTTTC....TCT....CGTTC....CATA....TAA
Botswana	ATGTTTGTTTTTC....TCT....AGTTC....CATA....TAA
Australia	ATGTTTGTTTTTC....TCT....AGTTC....CATC....TAA
UK	ATGTTTGTTTTTC....TCT....AGTTC....CACC....TAA
Mexico	ATGTTTGTTTTTC....TTT....AGTTC....TGCC....TAA
Brazil	ATGTTTGTTTTTC....TTT....AGTTC....TACC....TAA
Aruba	ATGTTTGTTTTTC....TCT....AGTTC....TACC....TAA
Germany	ATGTTTGTTTTTC....TCA....GAGTT....TACC....TAA
India	ATGTTTGTTTTTC....TCT....GAGTT....TACC....TAA
Vietnam	ATGTTTGTTTTTC....TCT....GATTT....TACC....TAA
China	ATGTTTGTTTTTC....TCT....AGTTT....TACC....TAA

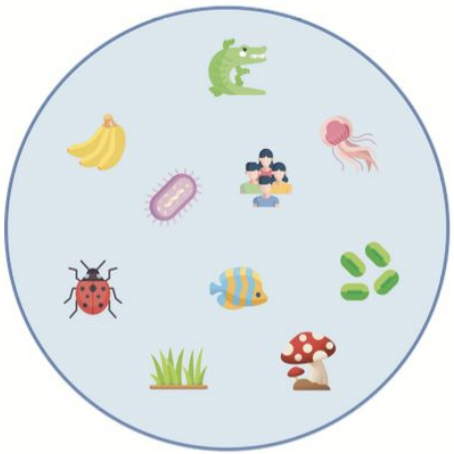
IQ-TREE
Small & Medium alignments

CMAPLE
Very large alignments



Source: nextstrain.org

IQ-TREE enables to infer the origins of life on earth

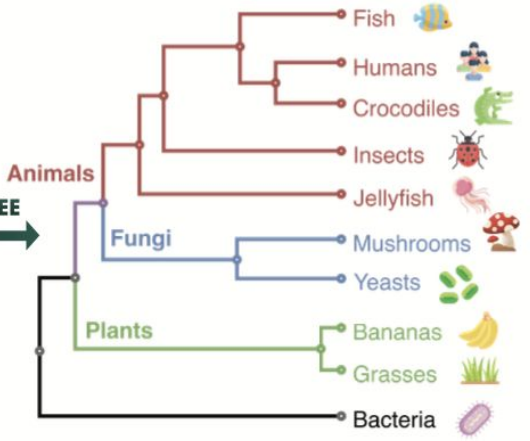


Sequencing

Protein Sequence Alignment

Humans	RMHYD....LIHTS....LARVV....FTLEYGYS....EYRS
Crocodiles	RMHYD....LIHTS....LARRV....FTLEYGYS....EYRS
Fish	RMHYD....LIHTS....LARRV....FTQEYGTS....EYRS
Insects	RMHYD....LIHTS....LARRV....FTLEYGTS....EYYS
Jellyfish	RMHYD....LIHTS....LARRV....FTLEYGTS....EYRS
Mushrooms	RMHYD....LLNRS....LARRV....FTLEYGTS....EYRS
Yeasts	RMHYD....LINRS....LARRV....FTLEYGTS....EYRS
Bananas	RMHYD....LIHTS....LASRV....FTLRYGTS....EYRS
Grasses	RMHYD....LIHTS....LASRV....FTLEYGTS....EYRS
Bacteria	RMHHD....LIHTS....LARRV....FTLEYGTS....EYRS

IQ-TREE



M

What is IQ-TREE?

IQ-TREE is a software program for phylogenetic inference, which means it is used to construct evolutionary trees that represent the relationships between different biological sequences such as DNA or protein sequences. The name "IQ-TREE" stands for "Intelligent Quartet Tree" and it is a reference to the algorithm used to infer the phylogenetic trees, which is based on the analysis of quartets of sequences.

IQ-TREE uses a number of advanced algorithms and statistical models to estimate the evolutionary history of the sequences, including models that account for rate heterogeneity among sites, among lineages, and among partitions. It also includes a number of tools for visualizing and interpreting the resulting trees.

IQ-TREE is widely used in molecular evolution and phylogenetics research, and is considered to be one of the fastest and most accurate programs available for phylogenetic inference. It is available for download as a standalone software package and also as a web server for users who prefer a graphical user interface.

Regenerate response

Typical phylogenetic analysis under maximum likelihood

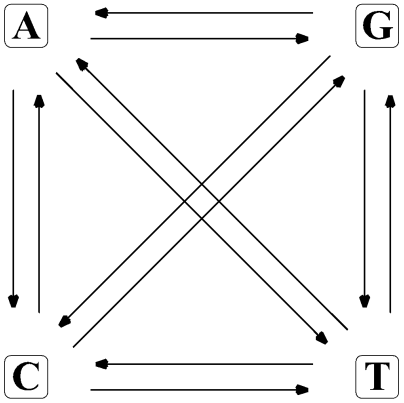
Multiple sequence alignment

```
ACGGGAT--C--C----CATTAC
ACGGGAT--C--C----CACTAC
CCGGGATAGCTTC----CATTAC
ACCCCCTATC--CACTGGATTAC
ACGACATATC--CACTGGATTCC
```

Model selection

ModelFinder (2017)

Substitution model

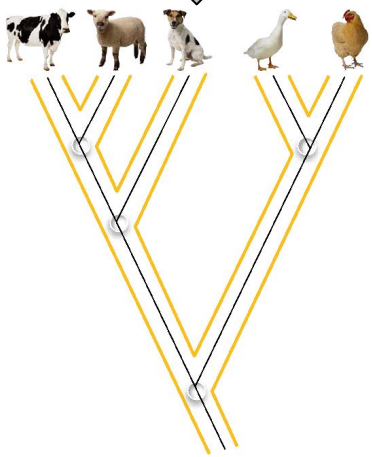


We focused on improving all three steps for large datasets!

IQ-TREE (2015, 2020)

Tree reconstruction

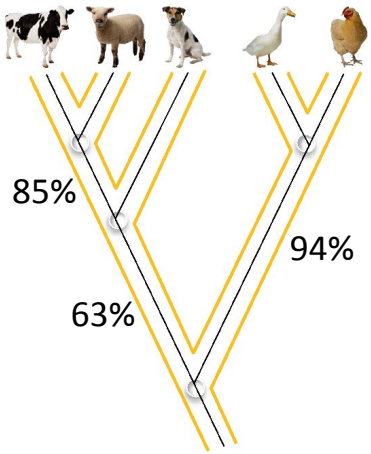
iqtree2 -s ALN_FILE -B 1000



Phylogenetic tree

Ultrafast bootstrap (2013, 2018)

Assessment of branch supports



Tree with branch supports

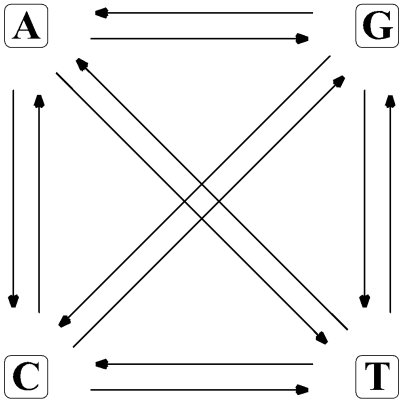
IQ-TREE tree search algorithm

Multiple sequence alignment

```
ACGGGAT--C--C----CATTAC
ACGGGAT--C--C----CACTAC
CCGGGATAGCTTC----CATTAC
ACCCCCTATC--CACTGGATTAC
ACGACATATC--CACTGGATTCC
```

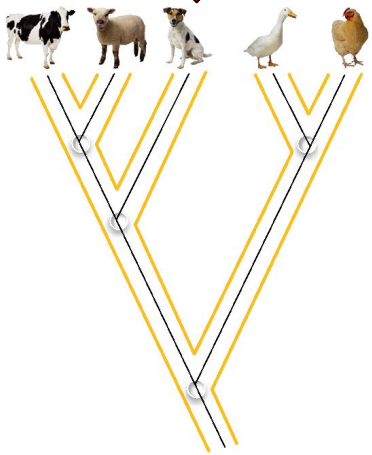
Model selection

Substitution model



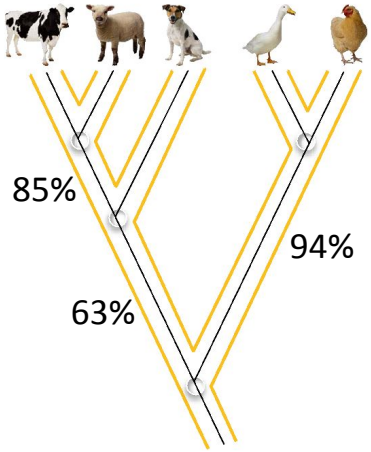
IQ-TREE (2015, 2020)

Tree reconstruction



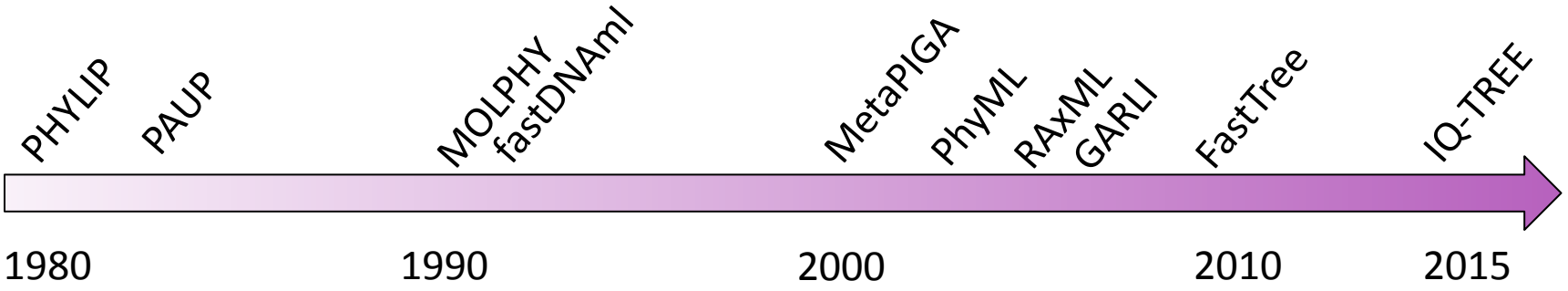
Phylogenetic tree

Assessment of branch supports

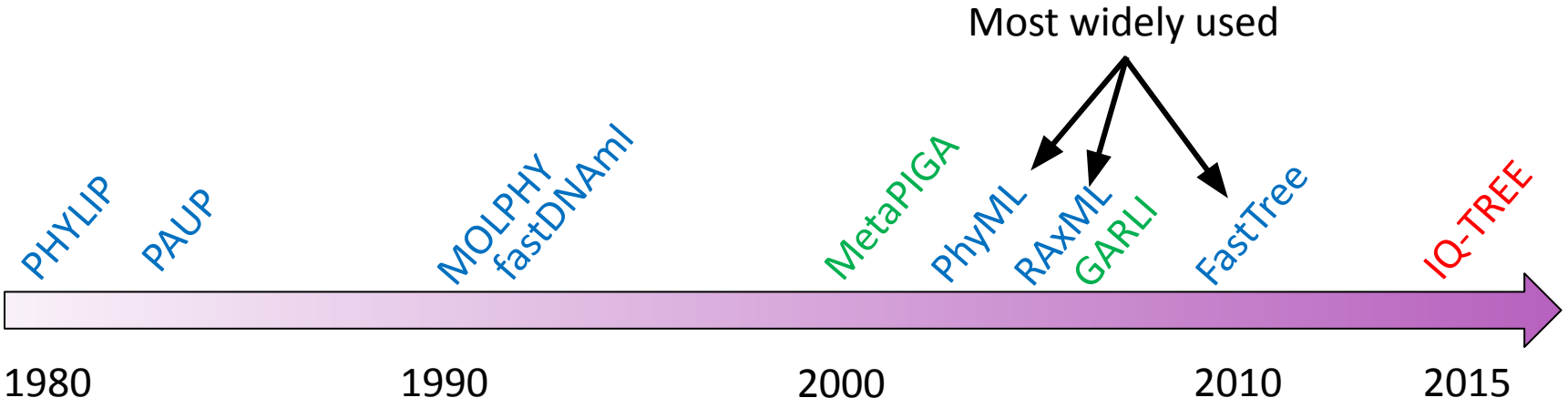


Tree with branch supports

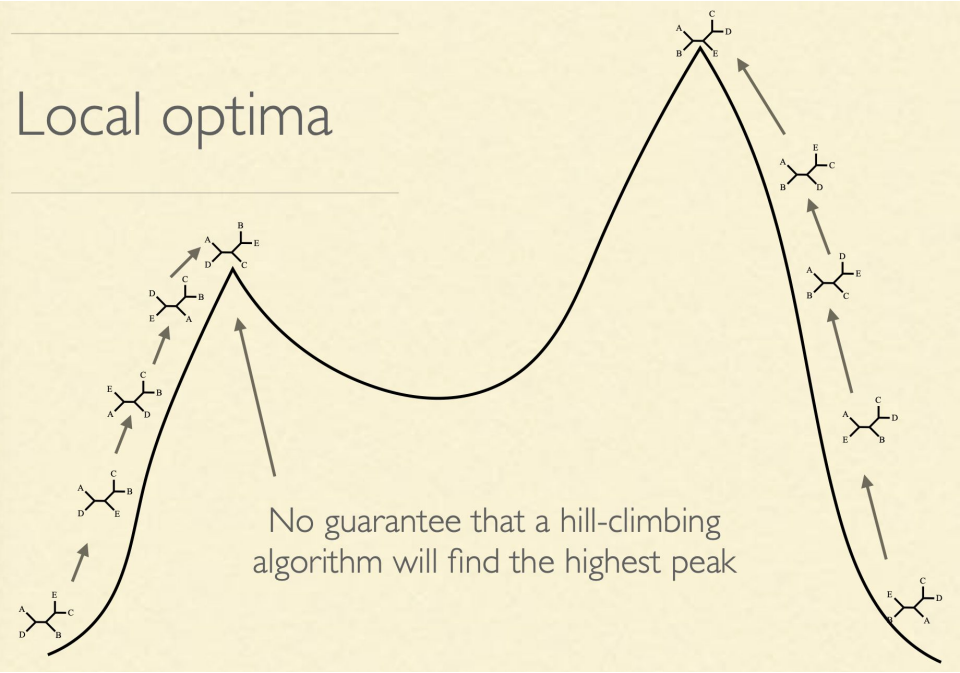
Search heuristics for finding maximum likelihood trees



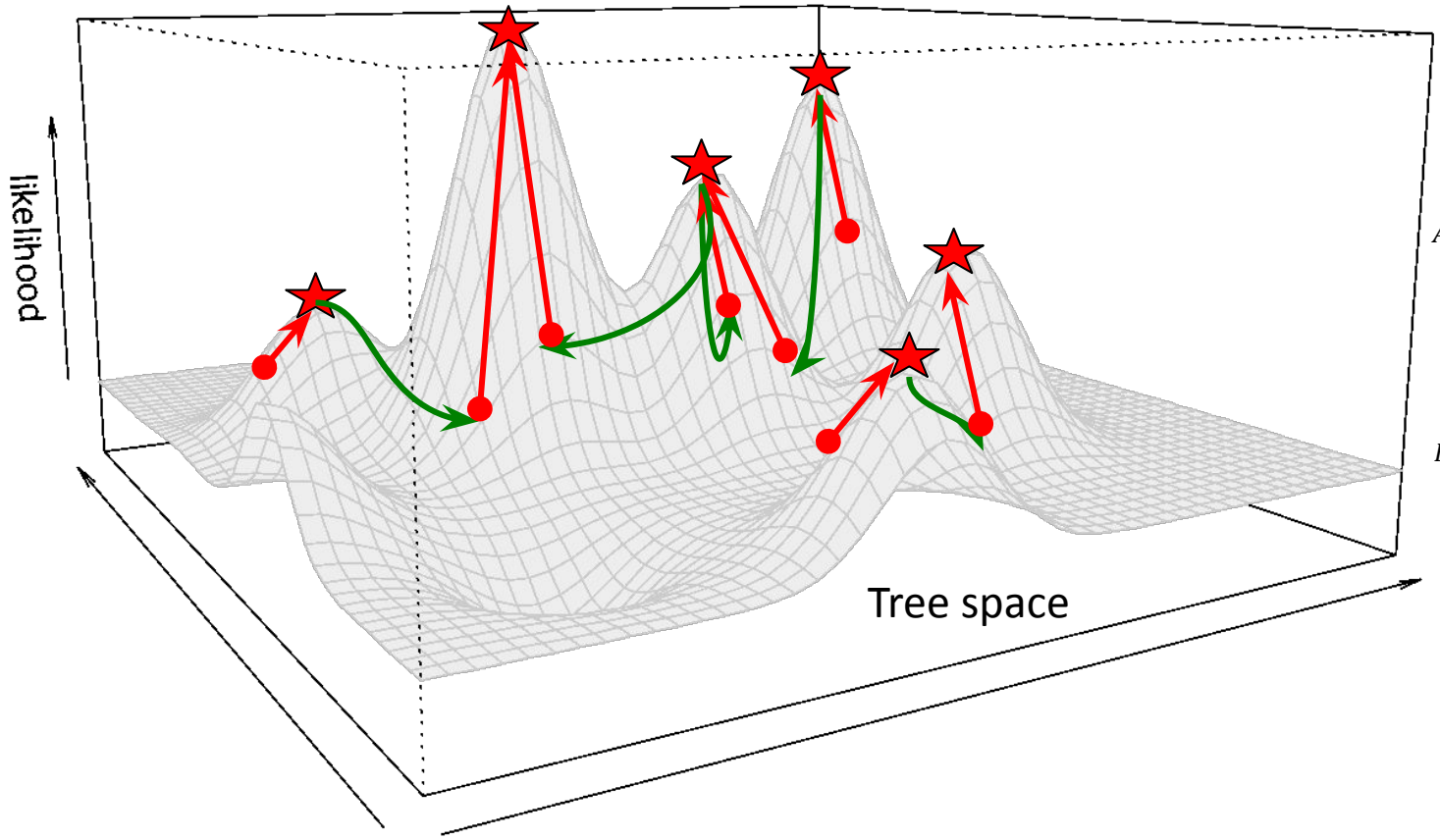
Search heuristics for finding maximum likelihood trees

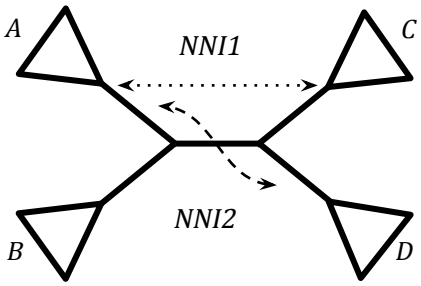


1. Hill-climbing / greedy algorithms:
   Fast but local optimum
2. Genetic algorithm:
   Slow but escaping local optima
3. IQ-TREE:
   Fast and escaping local optima



IQ-TREE: A new stochastic algorithm





Nearest neighbor interchange

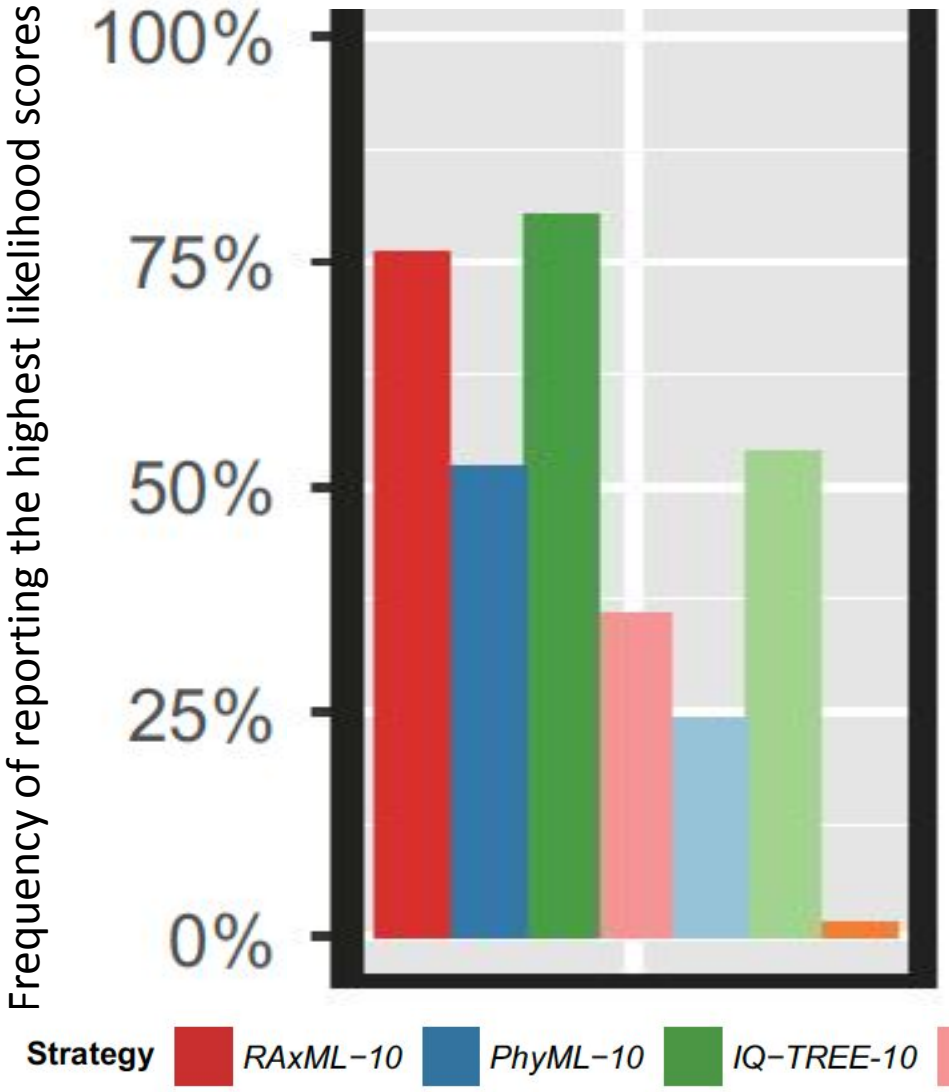
* 100 starting trees (99 parsimony, 1 NJ)
* Keeping a “population” of 20 best trees
* Stop if unsuccessful for 100 consecutive down-hill + up-hill moves

Lam-Tung Nguyen

Heiko Schmidt

Arndt von Haeseler

An independent benchmark by Zhou et al. (2018)



Runtimes:

FastTree is fastest

PhyML is slow/crashed

IQ-TREE & RAxML are equally fast

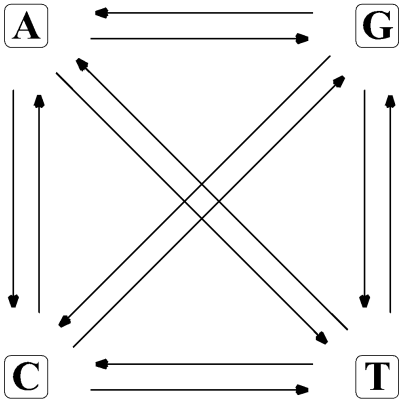
IQ-TREE tree search algorithm

Multiple sequence alignment

```
ACGGGAT--C--C----CATTAC
ACGGGAT--C--C----CACTAC
CCGGGATAGCTTC----CATTAC
ACCCCCTATC--CACTGGATTAC
ACGACATATC--CACTGGATTCC
```

Model selection

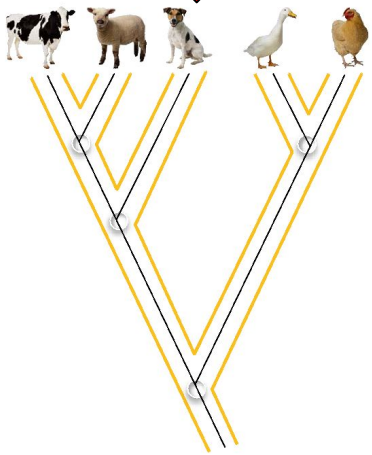
Substitution model



IQ-TREE algorithm efficiently explores tree space

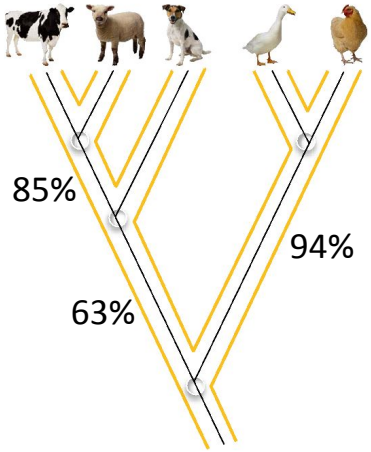
IQ-TREE (2015, 2020)

Tree reconstruction



Phylogenetic tree

Assessment of branch supports



Tree with branch supports

Genome-scale data: Concatenation methods

Supermatrix

Gene 1	Gene 2	……	Gene 1,000
CACCTGTCGT	----------	----------	TCTGGTGCAG
CAGCTGTCGT	GCTCTTTCTG	TTGAGCCTGG	TCTGGTGCAG
CAGCTGCCGT	GTTTTCTCTG	TTGAGCCTGG	TCTGGTACAG
CAGCTGCCGC	GTTCTCTCCG	----------	TCTGGTGCAA
CTCCTGCCGG	GTGCTCTCAG	----------	----------
CTCCTGCCGG	----------	CTGAGCCGGG	TCTGGTGCAG
CTCTTGCCGG	----------	CTGAGCCTTG	----------

Concatenation

Assume that all genes share the same evolutionary history

Phylogenomic Inference

All genes

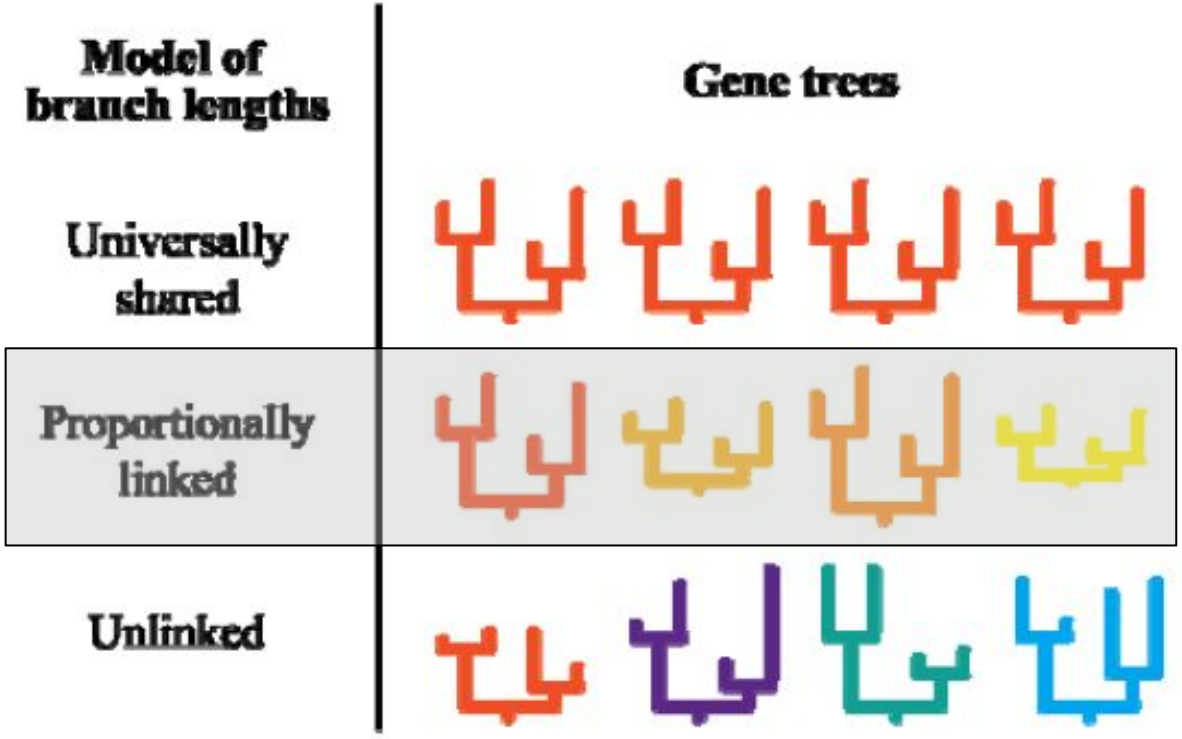
But this ignores the occurrence of different gene trees

Partition model

Supermatrix

Gene 1	Gene 2	……	Gene 1,000
CACCTGTCGT	----------	----------	TCTGGTGCAG
CAGCTGTCGT	GCTCTTTCTG	TTGAGCCTGG	TCTGGTGCAG
CAGCTGCCGT	GTTTTCTCTG	TTGAGCCTGG	TCTGGTACAG
CAGCTGCCGC	GTTCTCTCCG	----------	TCTGGTGCAA
CTCCTGCCGG	GTGCTCTCAG	----------	----------
CTCCTGCCGG	----------	CTGAGCCGGG	TCTGGTGCAG
CTCTTGCCGG	----------	CTGAGCCTTG	----------

Substitution models: JC HKY+G …… GTR+G



iqtree2 -s ALN_FILE –q PARTITION_FILE

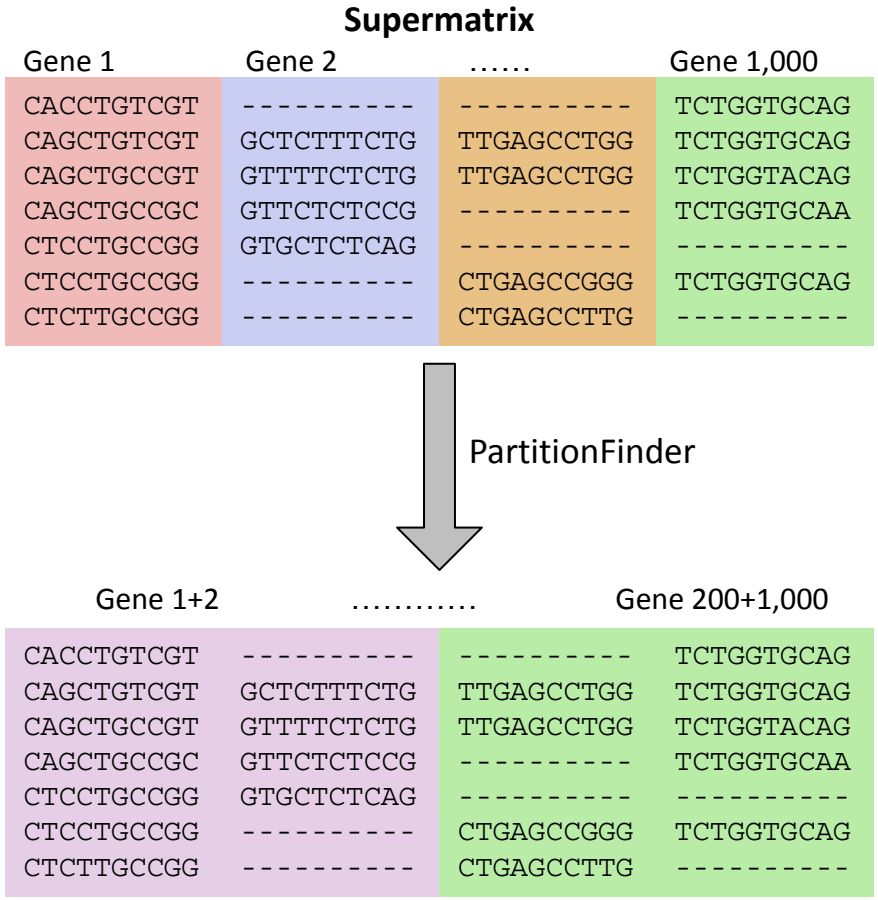
iqtree2 -s ALN_FILE –p PARTITION_FILE

iqtree2 -s ALN_FILE –Q PARTITION_FILE

Example partition file (turtle.nex)

```
#nexus
begin sets;
  charset ENSGALG00000000223.macse_DNA_gb = 1-846;
  charset ENSGALG00000001529.macse_DNA_gb = 847-1368;
  charset ENSGALG00000002002.macse_DNA_gb = 1369-2040;
  charset ENSGALG00000002514.macse_DNA_gb = 2041-2772;
  charset ENSGALG00000003337.macse_DNA_gb = 2773-3738;
  charset ENSGALG00000003700.macse_DNA_gb = 3739-4623;
  charset ENSGALG00000003702.macse_DNA_gb = 4624-6168;
  charset ENSGALG00000003907.macse_DNA_gb = 6169-6648;
  charset ENSGALG00000005820.macse_DNA_gb = 6649-7224;
  charset ENSGALG00000005834.macse_DNA_gb = 7225-7920;
  charset ENSGALG00000005902.macse_DNA_gb = 7921-8490;
  charset ENSGALG00000008338.macse_DNA_gb = 8491-9282;
  charset ENSGALG00000008517.macse_DNA_gb = 9283-9822;
  charset ENSGALG00000008916.macse_DNA_gb = 9823-10368;
  charset ENSGALG00000009085.macse_DNA_gb = 10369-11298;
  charset ENSGALG00000009879.macse_DNA_gb = 11299-11895;
  charset ENSGALG00000011323.macse_DNA_gb = 11896-12795;
  charset ENSGALG00000011434.macse_DNA_gb = 12796-13242;
  charset ENSGALG00000011917.macse_DNA_gb = 13243-14223;
  charset ENSGALG00000011966.macse_DNA_gb = 14224-14691;
  charset ENSGALG00000012244.macse_DNA_gb = 14692-15444;
  charset ENSGALG00000012379.macse_DNA_gb = 15445-15963;
  charset ENSGALG00000012568.macse_DNA_gb = 15964-16593;
  charset ENSGALG00000013227.macse_DNA_gb = 16594-17895;
  charset ENSGALG00000014038.macse_DNA_gb = 17896-18456;
  charset ENSGALG00000014648.macse_DNA_gb = 18457-18954;
  charset ENSGALG00000015326.macse_DNA_gb = 18955-19551;
  charset ENSGALG00000015397.macse_DNA_gb = 19552-20145;
  charset ENSGALG00000016241.macse_DNA_gb = 20146-20820;
end;
```

How to reduce potential model overfitting?



Substitution models: HKY ..... GTR+G

PartitionFinder algorithm (Lanfear et al. 2012):

1. Evaluate all pairs of genes.
2. Find the pair with best score.
3. If score improves, merge two genes and repeat steps 1-3.
4. Otherwise, stop.

iqtree2 … -m MFP+MERGE

Relaxed clustering algorithm (Lanfear et al. 2014):

In step 1: only examine the top k% of most "promising" pairs.

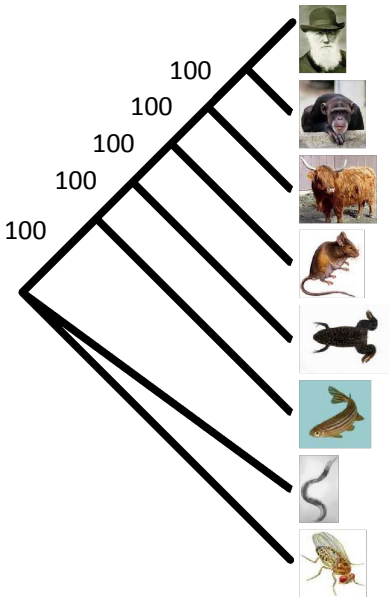
iqtree2 … -rcluster 10

Concatenation methods: Limitation

Supermatrix

Gene 1	Gene 2	.....	Gene
1,000			
CAGCTGTCGT	----------	----------	TCTGGTGCAG
CAGCTGTCGT	GCTCTTTCTG	TTGAGCCTGG	TCTGGTGCAG
CAGCTGCCGT	GTTTTCTCTG	TTGAGCCTGG	TCTGGTACAG
CAGCTGCCGC	GTTCTCTCCG	----------	TCTGGTGCAA
CTCCTGCCGG	GTGCTCTCAG	----------	----------
CTCCTGCCGG	----------	CTGAGCCGGG	TCTGGTGCAG
CTCTTGCCGG	----------	CTGAGCCTTG	----------

Phylogenomic Inference



Species tree of life

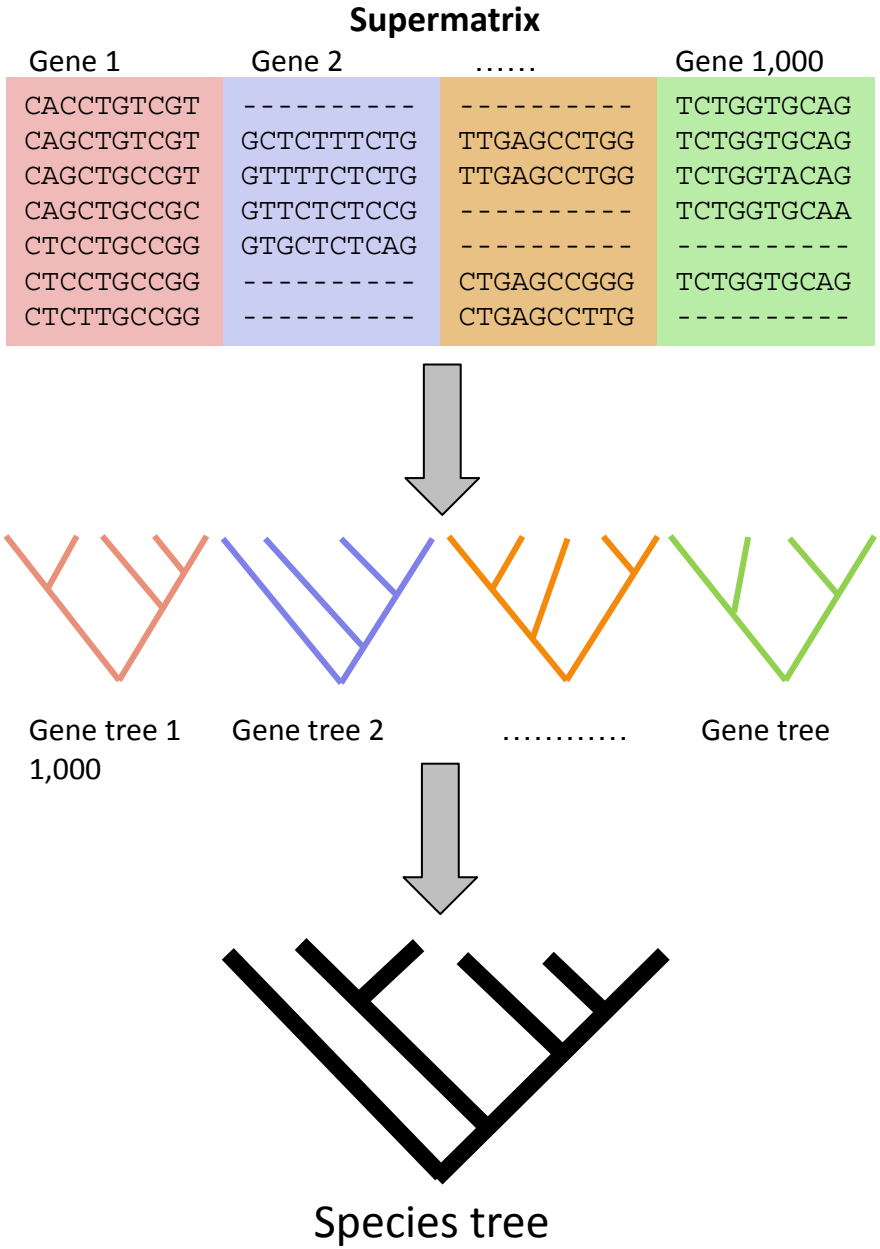
Bootstrap supports and Bayesian posteriors tend to 100% as #genes increases!

Concatenation assumes a single tree across all loci

Potential *systematic bias*

"When the method of inferring phylogenies is one with undesirable statistical properties such as inconsistency, the bootstrap does not correct for these" (Felsenstein, 1985)

Coalescent methods



Gene Concordance Factor (gCF): How often a branch in species tree is found among gene trees?
0% ≤ gCF ≤ 100%

Implementation in IQ-TREE fully accounts for missing data

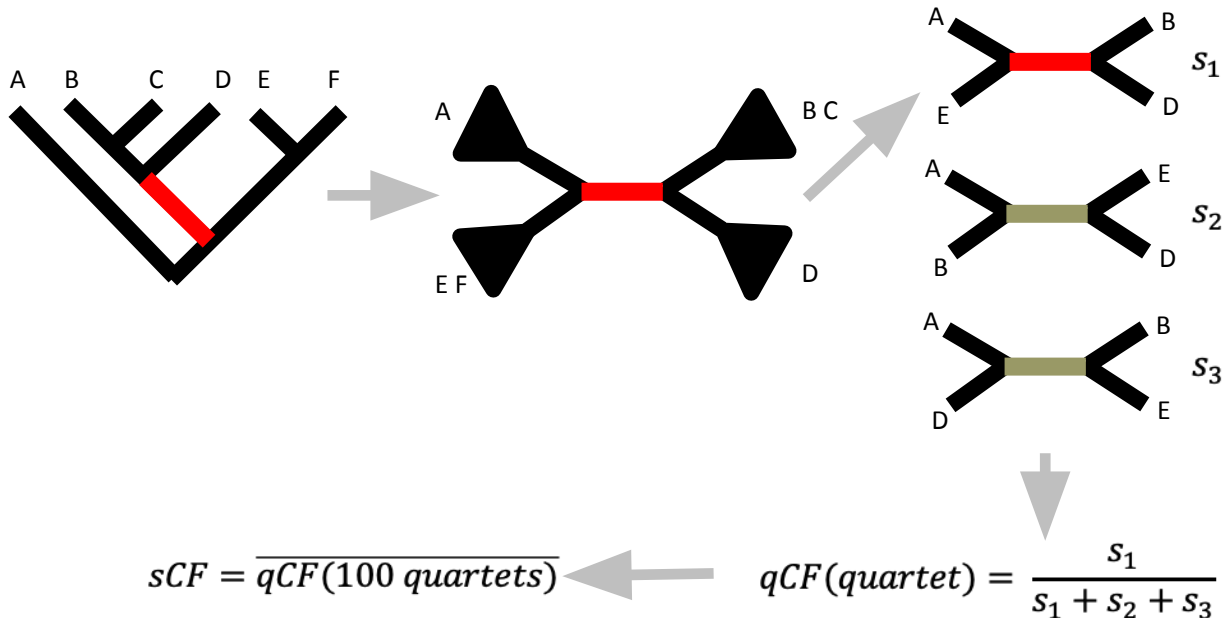
Problem: Uncertainties in gene trees!

Site Concordance Factor (sCF)

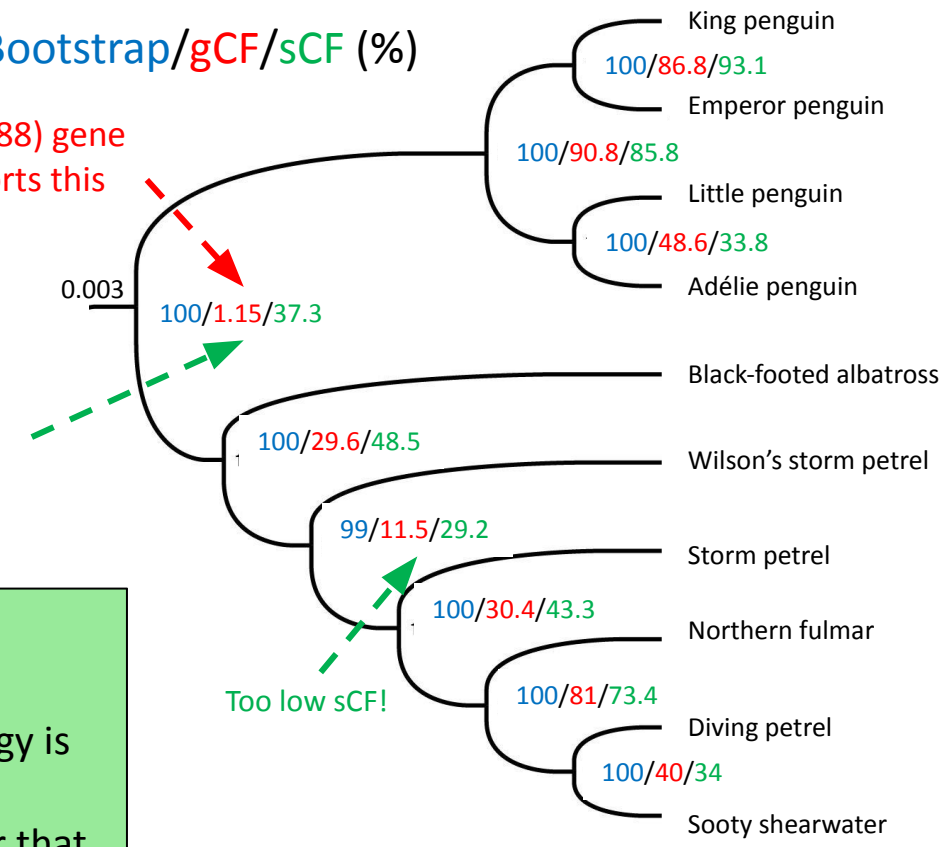
Supermatrix

Gene 1	Gene 2	……	Gene 1,000
CAGCTGTCGT	----------	----------	TCTGGTGCAG
CAGCTGTCGT	GCTCTTTCTG	TTGAGCCTGG	TCTGGTGCAG
CAGCTGCCGT	GTTTTCTCTG	TTGAGCCTGG	TCTGGTACAG
CAGCTGCCGC	GTTCTCTCCG	----------	TCTGGTGCAA
CTCCTGCCGG	GTGCTCTCAG	----------	----------
CTCCTGCCGG	----------	CTGAGCCGGG	TCTGGTGCAG
CTCTTGCCGG	----------	CTGAGCCTTG	----------

Site Concordance Factor (sCF):
How often a branch is "supported" by alignment sites?
33.3% ≦ sCF ≤ 100%



An example birds data set (Reddy et al., 2017)



- 131 sites support this branch
- 105 sites support NNI branch 1
- 114 sites support NNI branch 2

Felsenstein (1985): a difference of 20 sites favouring one topology is enough to give 100% bootstrap support for that one topology!

Penguins

Tubenoses

- gCF and sCF are useful when bootstrap supports reach 100%.
- CAUTION when gCF ~ 0% or sCF ~ 33%, even if BS ~ 100%.
- GREAT when gCF and sCF > 50%.

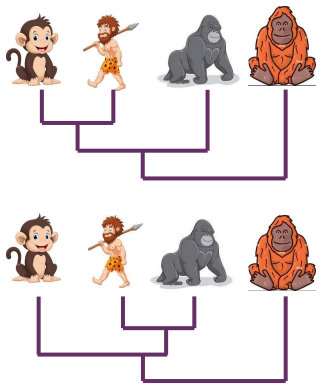
Mixture Across Sites and Trees (MAST) model

Concatenated alignment

S1:	A	A	-	T	A	A	A	T
S2:	T	A	A	C	C	T	T	T
S3:	T	A	T	A	A	G	T	T
S4:	A	C	-	A	C	A	A	A

L_1^1 L_2^1 L_3^1 L_4^1 L_5^1 L_6^1 L_7^1 L_8^1

L_1^2 L_2^2 L_3^2 L_4^2 L_5^2 L_6^2 L_7^2 L_8^2



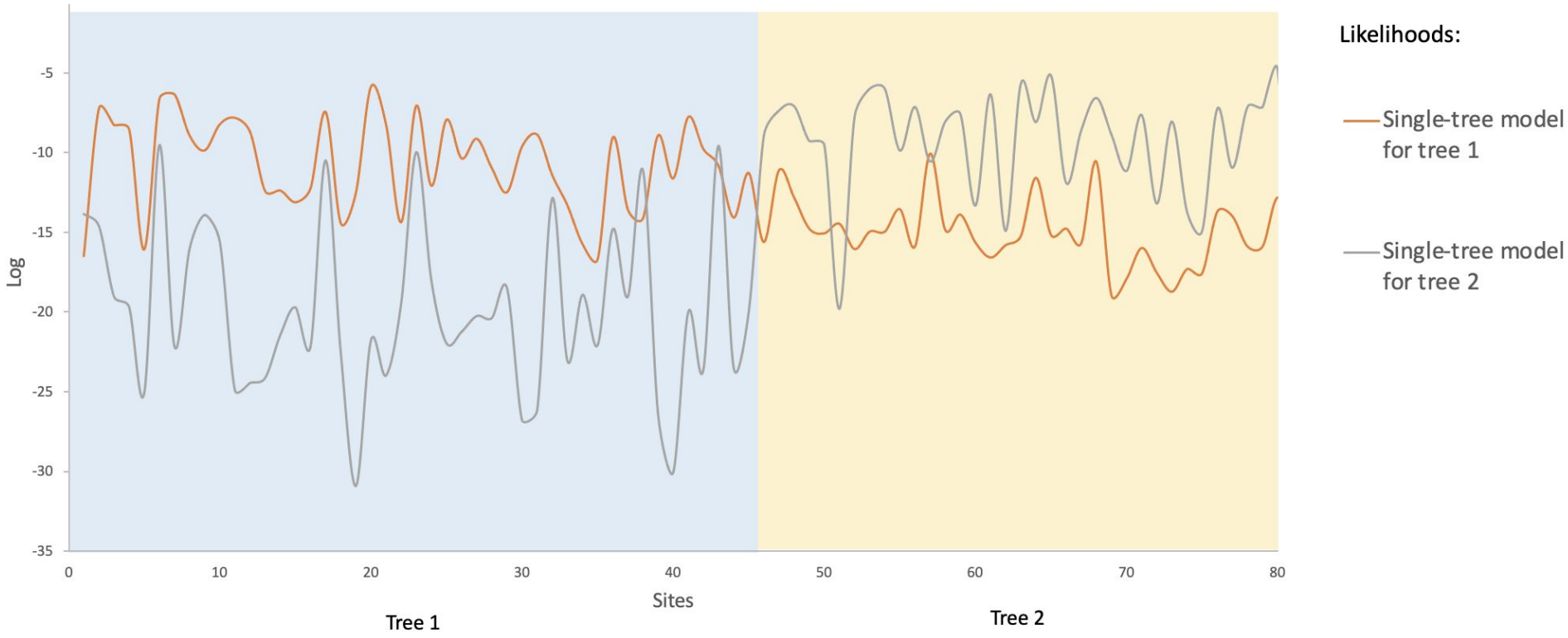
Likelihood for site *i*: $L_i = w_1 L_i^1 + w_1 L_i^2$

where w_j represents the portion of sites belonging to tree j

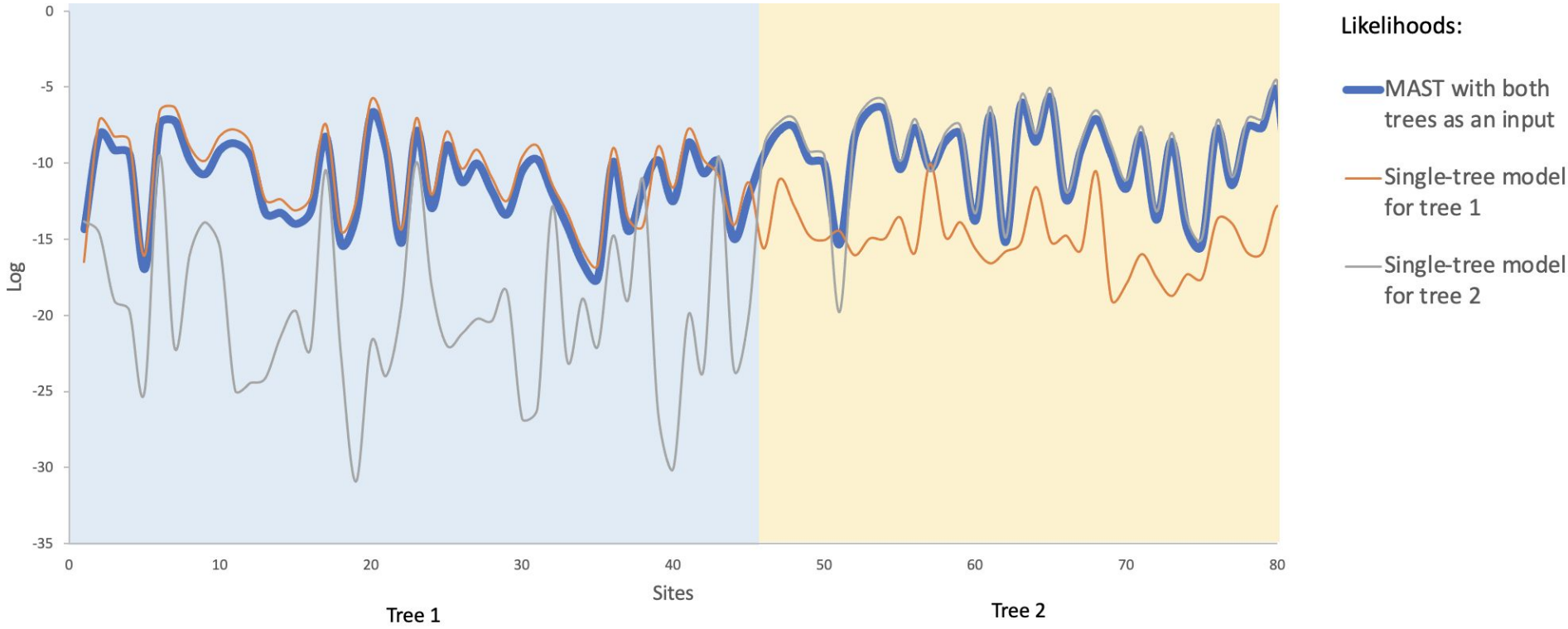
Log-likelihood of the trees: $\sum_i \log(L_i)$

iqtree2 -s ALN_FILE –te TREES_FILE –m GTR+G+T

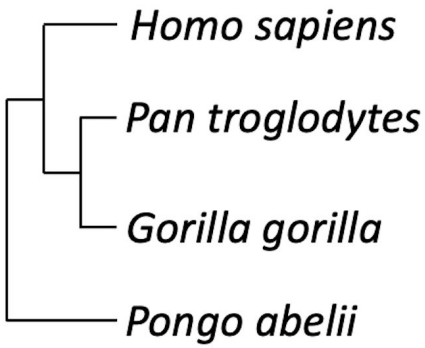
Toy example: Site log-likelihood

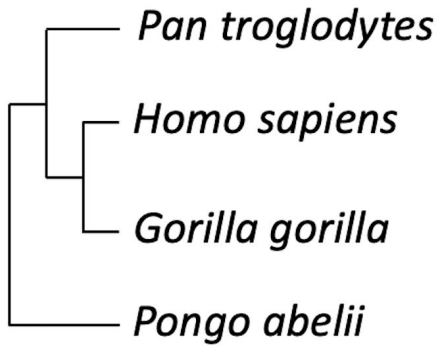


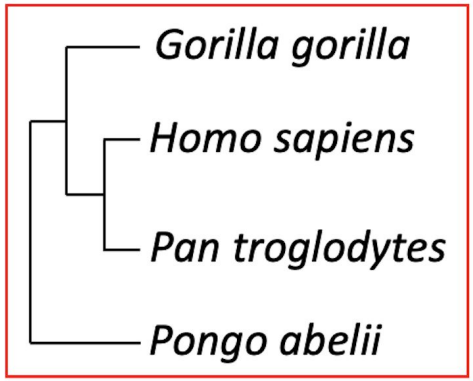
Toy example: Site log-likelihood



The classical example of Human, Chimp, Gorilla







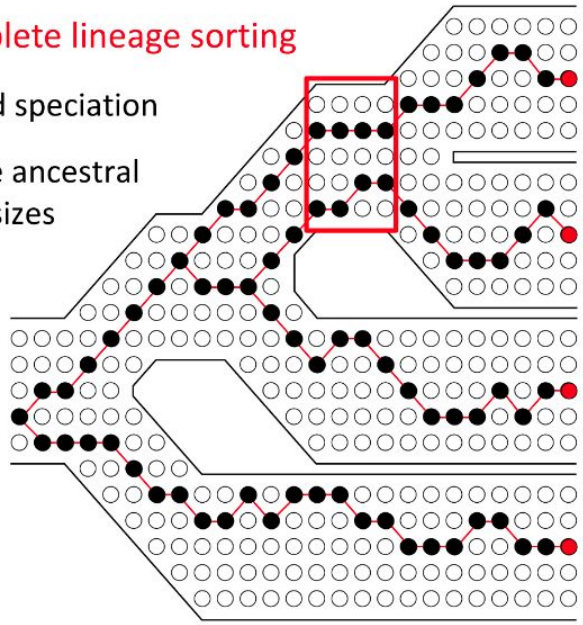
	T_{A1}	T_{A2}	T_{A3}
Gene tree frequencies:	19.8%	20.1%	60.1%
MAST model weights:	**17.9%**	**17.4%**	**64.7%**

Data: 1,595 genes; 1,618,506 bp (Vanderpool et al. 2020)

Gene trees discordance due to deep coalescence

Incomplete lineage sorting

- Rapid speciation
- Large ancestral pop sizes



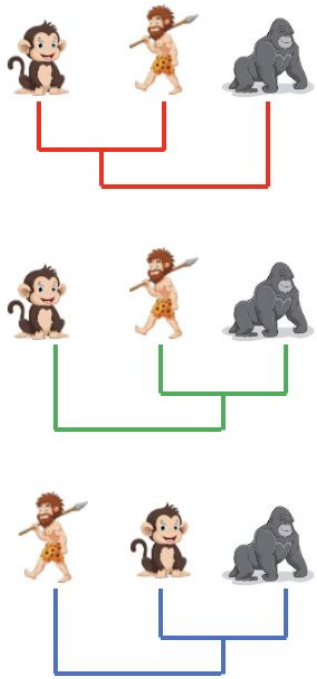
Human

Chimpanzee

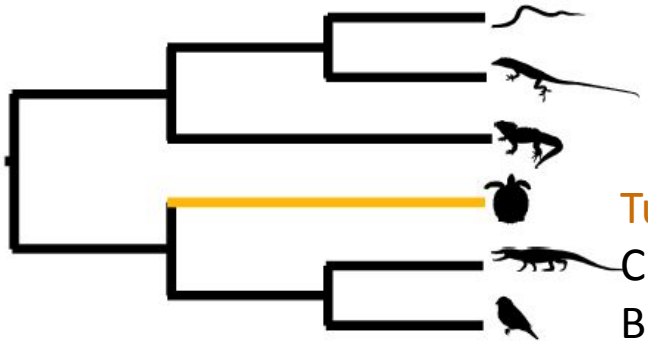
Gorilla

Orangutan

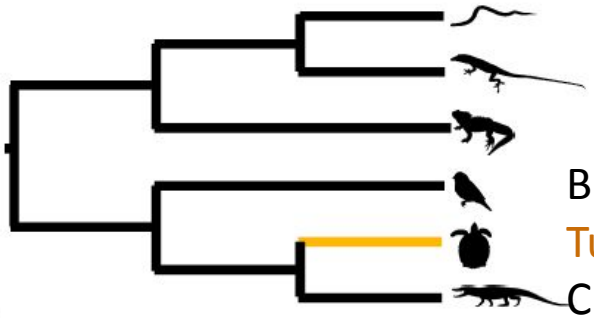
Deep coalescence
(ILS)



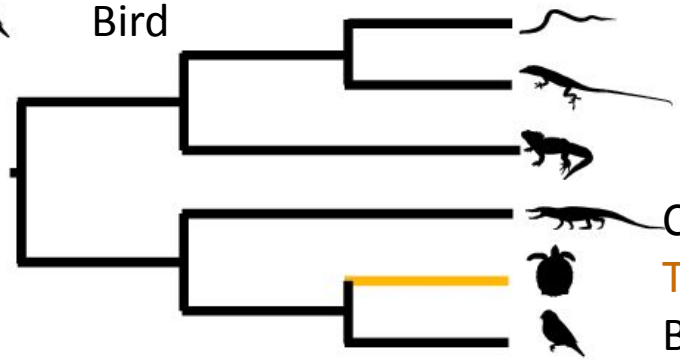
Dataset for IQ-TREE lab: Where is Turtle in the tree?



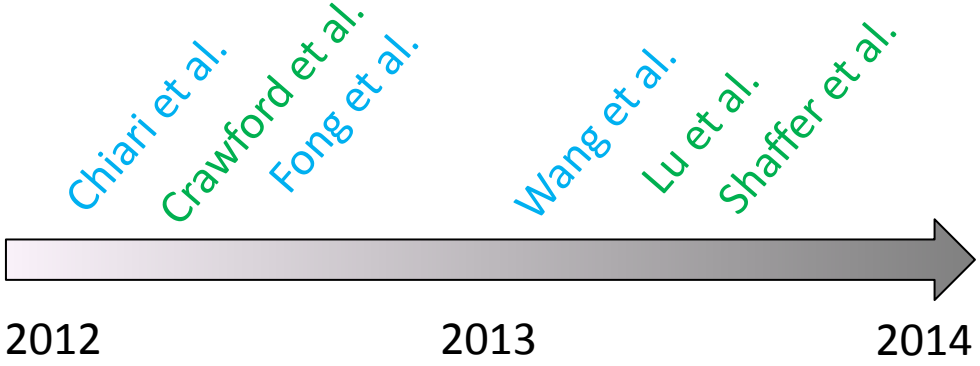
Turtle
Crocodile
Bird



Bird
Turtle
Crocodile



Crocodile
Turtle
Bird



Different studies led to different trees!

Dataset: 16 species, 29 genes, 20,820 bp
(a subset of Chiari et al. 2012)

Thanks Jeremy Brown

IQ-TREE lab

1. Input data
2. Inferring the first phylogeny
3. Applying partition model
4. Choosing the best partitioning scheme
5. Tree mixture model (**NEW**)

-- Break --

6. Identifying most influential genes
7. Removing influential genes
8. Concordance factors (advanced)

http://www.iqtree.org/workshop/sydney2023