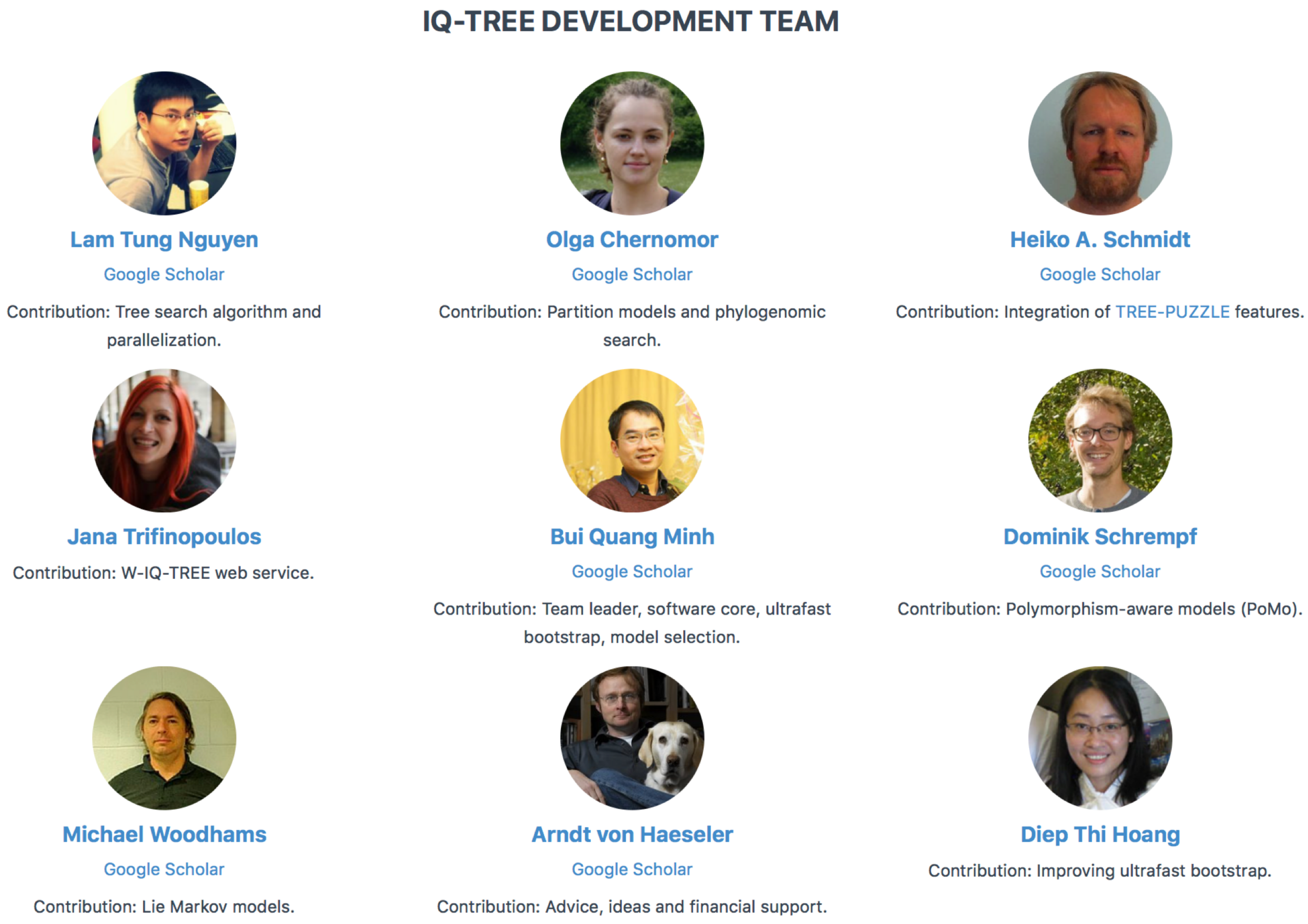
Why IQ-TREE?

Bui Quang Minh
Australian National University

Workshop on Molecular Evolution
Woods Hole, 24 July 2018



IQ-TREE DEVELOPMENT TEAM

Lam Tung Nguyen

Google Scholar

Contribution: Tree search algorithm and parallelization.

Olga Chernomor

Google Scholar

Contribution: Partition models and phylogenomic search.

Heiko A. Schmidt

Google Scholar

Contribution: Integration of TREE-PUZZLE features.

Jana Trifinopoulos

Contribution: W-IQ-TREE web service.

Bui Quang Minh

Google Scholar

Contribution: Team leader, software core, ultrafast bootstrap, model selection.

Dominik Schrempf

Google Scholar

Contribution: Polymorphism-aware models (PoMo).

Michael Woodhams

Google Scholar

Contribution: Lie Markov models.

Arndt von Haeseler

Google Scholar

Contribution: Advice, ideas and financial support.

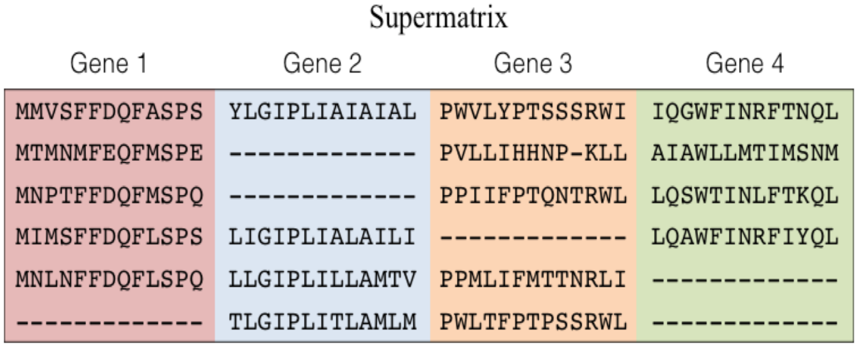
Diep Thi Hoang

Contribution: Improving ultrafast bootstrap.

Thanks to plenty of users for feedback and bug reports!

Typical phylogenetic analysis under maximum likelihood

Multiple sequence alignment

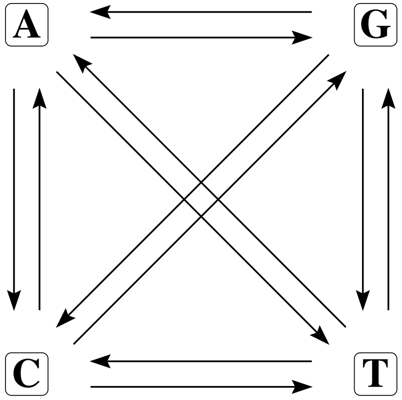


Supermatrix

Gene 1	Gene 2	Gene 3	Gene 4
MMVSFFDQFASPS	YLGIPLIAIAIAL	PWVLYPTSSSRWI	IQGWFINRFTNQL
MTMNMFEQFMSPE	-------------	PVLLIHHNP-KLL	AIAWLLMTIMSNM
MNPTFFDQFMSPQ	-------------	PPIIFPTQNTRWL	LQSWTINLFTKQL
MIMSFFDQFLSPS	LIGIPLIALAILI	-------------	LQAWFINRFIYQL
MNLNFFDQFLSPQ	LLGIPLILLAMTV	PPMLIFMTTNRLI	-------------
-------------	TLGIPLITLAMLM	PWLTFPTPSSRWL	-------------

Model selection

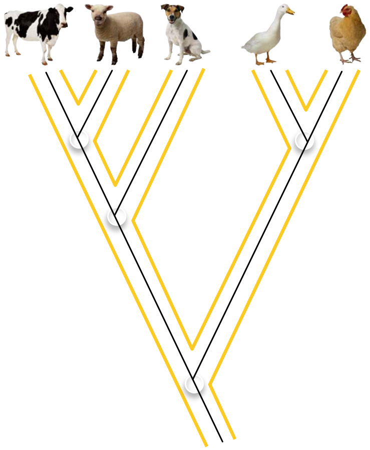
Substitution model



Tree reconstruction

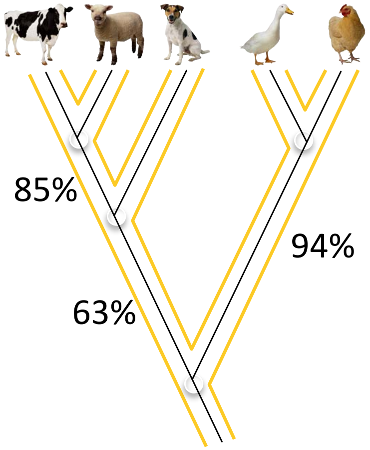
Challenges

- Flood of sequence data
- Need for realistic models of sequence evolution



Phylogenetic tree

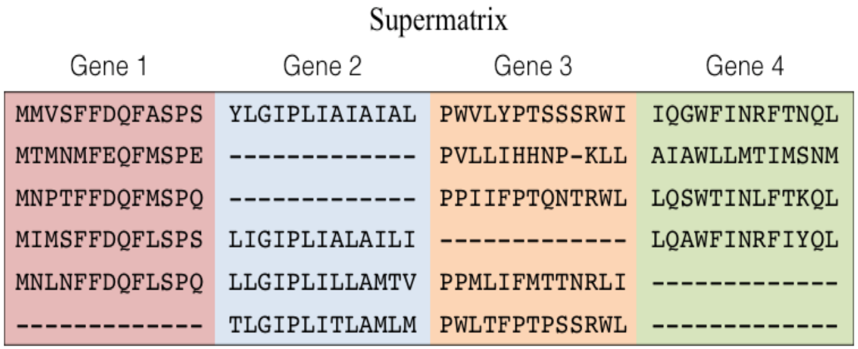
Branch supports



Tree with branch supports

Typical phylogenetic analysis under maximum likelihood

Multiple sequence alignment



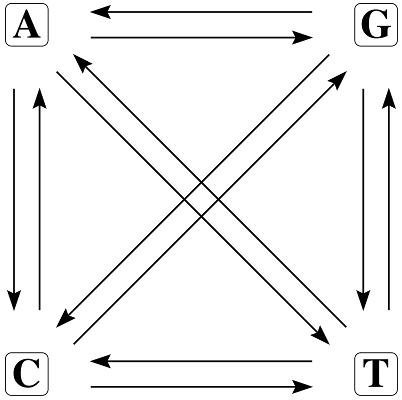
Supermatrix

Gene 1	Gene 2	Gene 3	Gene 4
MMVSFFDQFASPS	YLGIPLIAIAIAL	PWVLYPTSSSRWI	IQGWFINRFTNQL
MTMNMFEQFMSPE	-------------	PVLLIHHNP-KLL	AIAWLLMTIMSNM
MNPTFFDQFMSPQ	-------------	PPIIFPTQNTRWL	LQSWTINLFTKQL
MIMSFFDQFLSPS	LIGIPLIALAILI	-------------	LQAWFINRFIYQL
MNLNFFDQFLSPQ	LLGIPLILLAMTV	PPMLIFMTTNRLI	-------------
-------------	TLGIPLITLAMLM	PWLTFPTPSSRWL	-------------

Model selection

ModelFinder

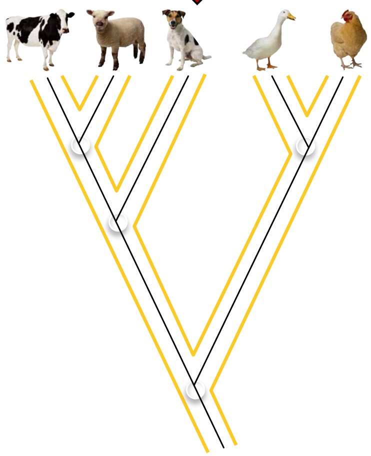
Substitution model



http://www.iqtree.org

IQ-TREE

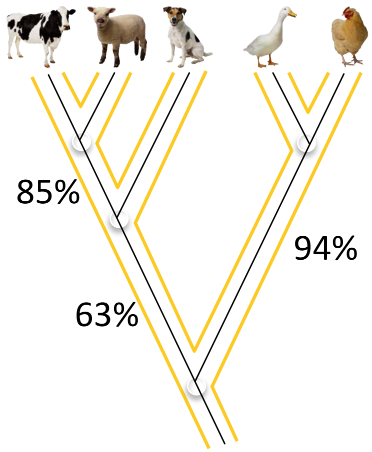
Tree reconstruction



Phylogenetic tree

Ultrafast bootstrap

Branch supports



Tree with branch supports

Advanced models of sequence evolution

Substitution models

- DNA
- Protein
- Codon
- Binary and Morphological

Partition models

- Edge-linked (equal or proportional)
- Edge-unlinked

Polymorphism-aware models (PoMo)

- Combining phylogenetic (mutations) with population genetics (drift) models.

Rate heterogeneity across sites

- Invariant sites
- Gamma model
- FreeRate model (available PhyML, RAxML-NG, IQ-TREE, PAML, BEAST2, RevBayes)

Mixture models

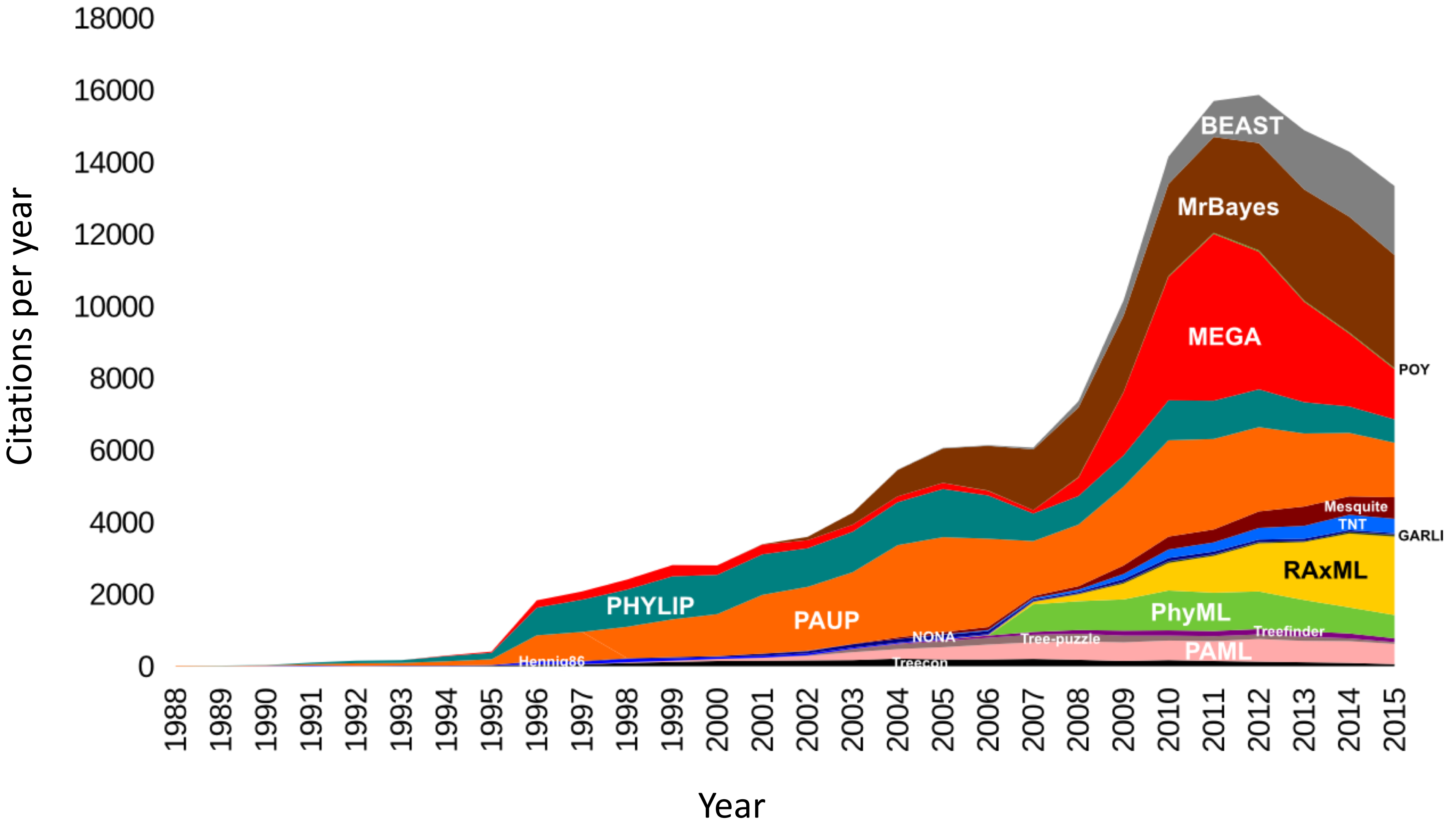
- Mixture of rate matrices (LG4M, LG4X)
- Mixture of state frequencies (C10-C60)
- Mixed branch lengths (GHOST – General Heterotachy on Single Trees)
- User-defined mixtures

Site-specific models

- Posterior mean site frequency (PMSF)

All models automatically selected with ModelFinder.
A lot faster than jModelTest and ProtTest!

Tree reconstruction: Which tree best explains the data?



Source: phylobotanist.blogspot.com

IQ-TREE: A new tree search algorithm

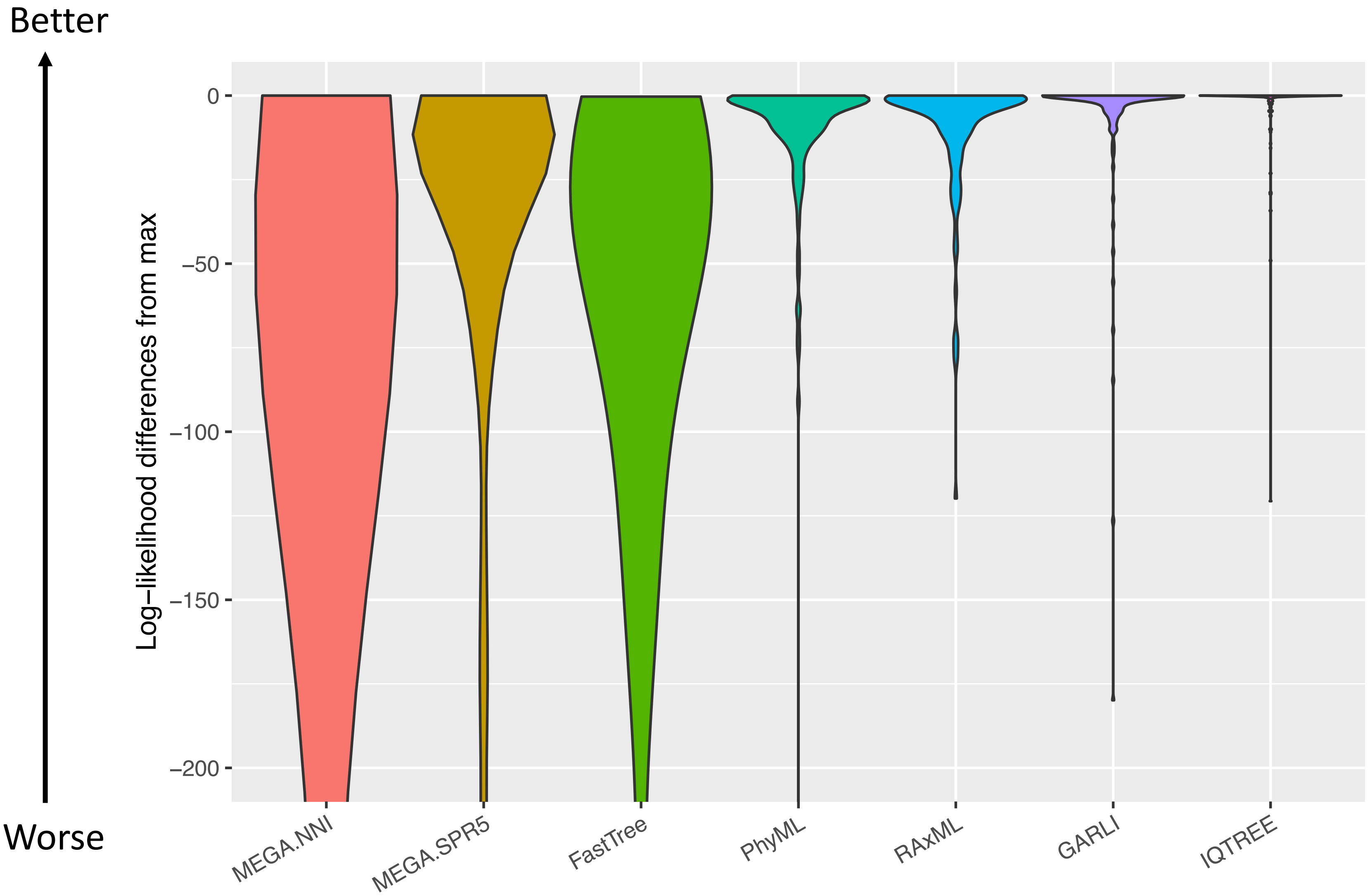
L.T. Nguyen, H. Schmidt, A. von Haeseler

Maximum parsimony

Nearest neighbor interchange (NNI)

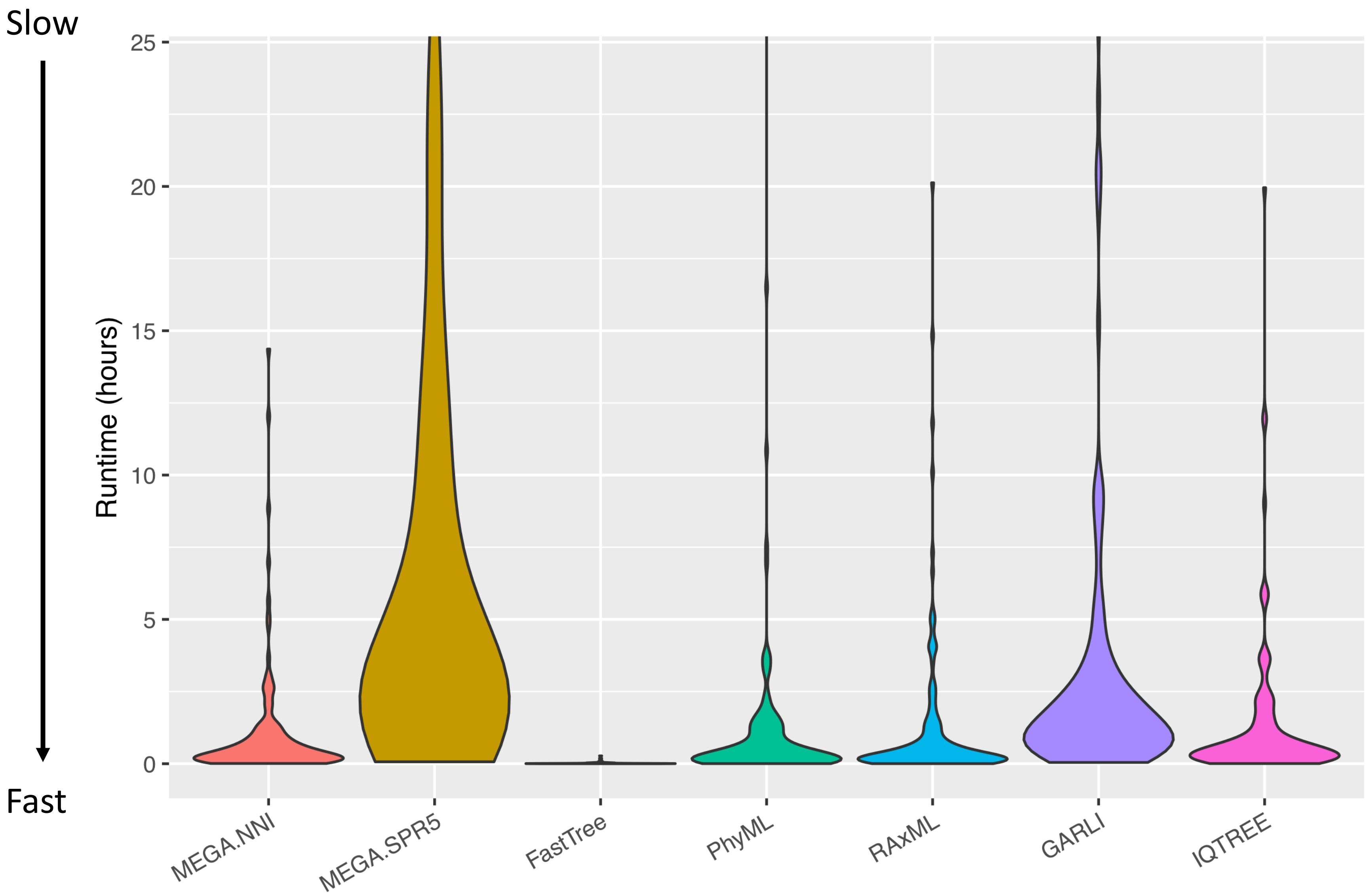
Random NNIs

Benchmark: Log-likelihood maximisation

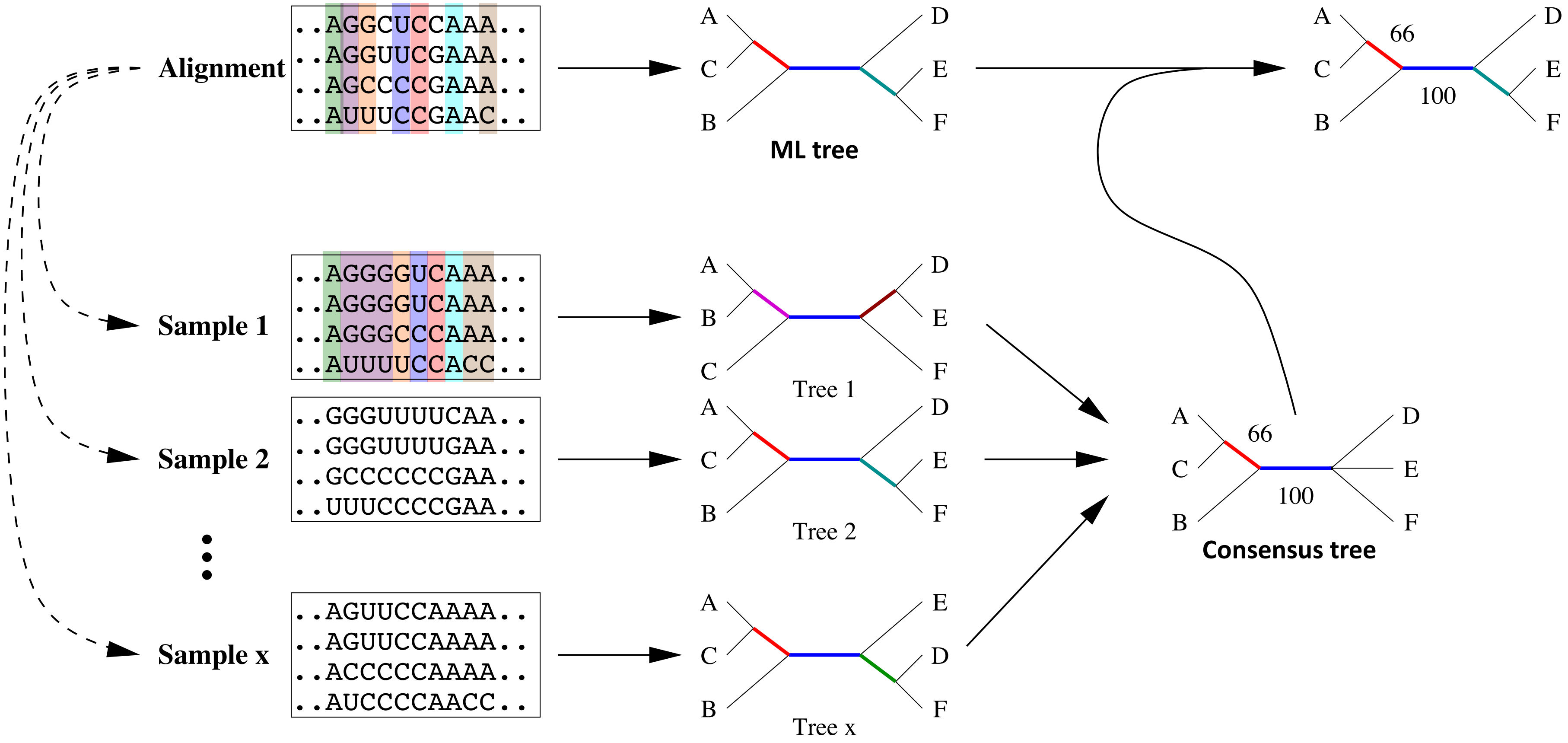


Independently confirmed by Zhou, Rokas et al. (2018)

Runtimes


Slow
Fast
Runtime (hours)
25
20
15
10
5
0
MEGA.NNI
MEGA.SPR5
FastTree
PhyML
RAxML
GARLI
IQTREE

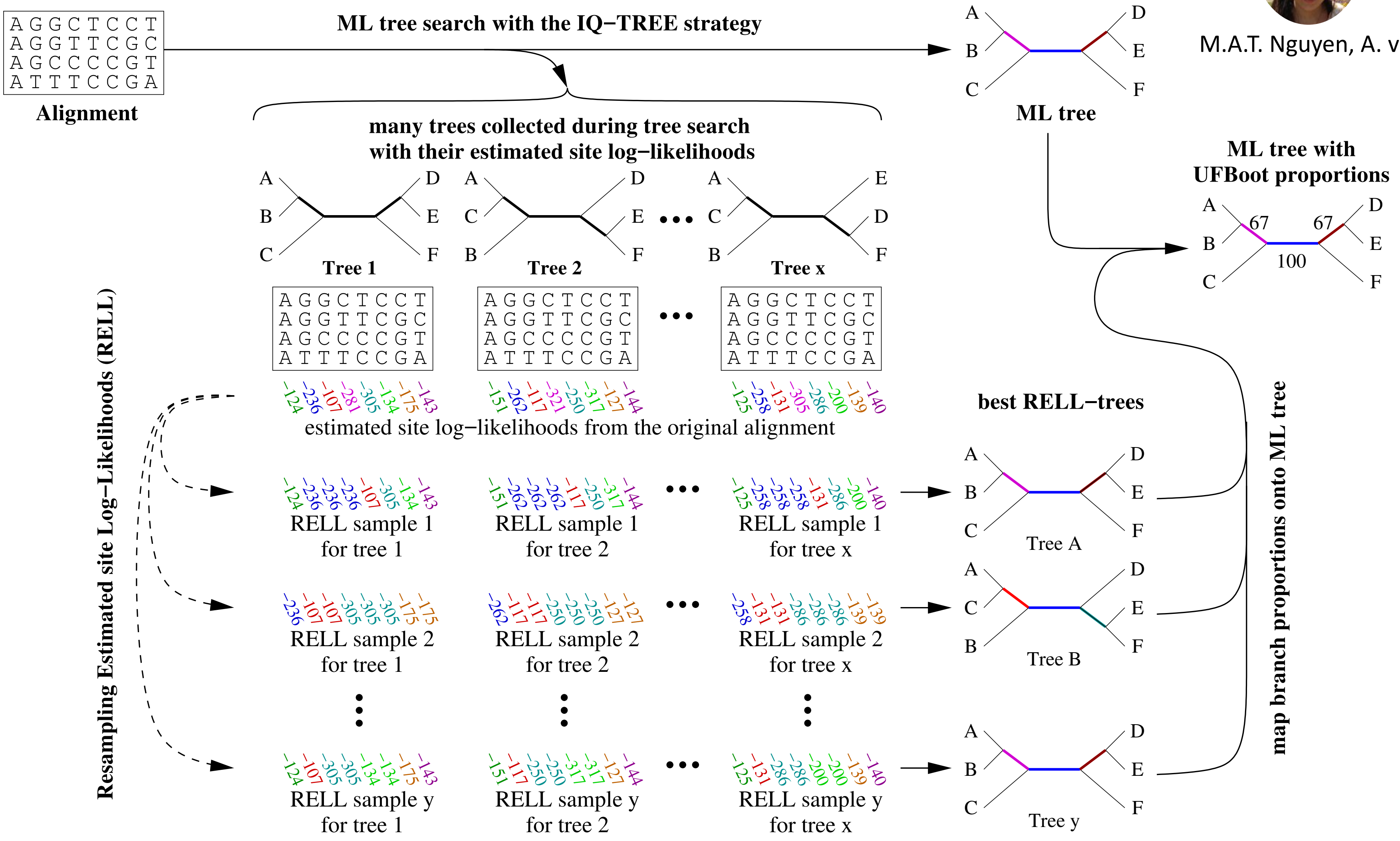
Branch supports: How reliable are branches of the tree?



Bootstrap analysis is extremely time-consuming!

UFBoot: Ultrafast bootstrap approximation

M.A.T. Nguyen, A. von Haeseler



Other features

Bootstrap resampling

- Resampling partitions
- Resampling partitions and sites

Tree topology tests

- Shimodaira-Haegawa (SH) test
- Expected Likelihood Weight (ELW)

Single branch tests

- SH-aLRT (Shimodaira-Hasegaw-like approximate likelihood ratio test)
- aBayes

Post-analysis

- Ancestral sequence reconstruction
- Inferring site-specific rates

IQ-TREE forum: https://groups.google.com/d/forum/iqtree
for questions, complaints, etc.

Exercises

Single model on a DNA data set (obtained from Phylogenetic Handbook)

- First running example
- Choosing the right substitution model
- Assessing branch supports (UFBoot, SH-aLRT)
- Utilizing multi-core CPUs

Partition model on a 248-gene Turtle DNA data set (kindly provided by J. Brown)

- Partitioned analysis for multi-gene alignments
- Choosing the right partitioning scheme *(*optional)*
- Bootstrap resampling partitions
- Tree tests
- Identifying most influential genes *(*optional)*

Mixture model on a 10-gene Microsporidia protein data set (kindly provided by L. Eme)

- Protein mixture model analysis

http://www.iqtree.org/doc/molevol