The phylogenetic posterior distribution

Frederick “Erick” Matsen
Fred Hutchinson Cancer Center

https://matsen.group/

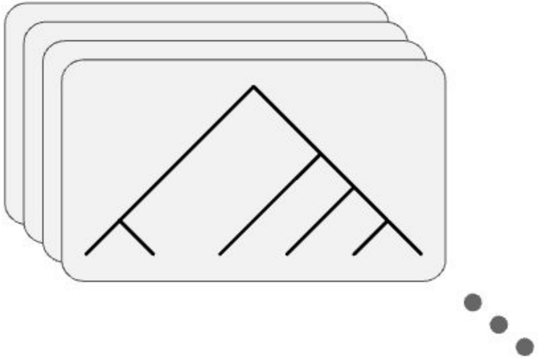
@matsen.bsky.social

https://bit.ly/icerm2024matsen

- Mathieu Fourment, Nhan Trong Ly, Minh Bui 🇦🇺
- Christiaan Swanepoel (Alexei Drummond’s lab) 🇳🇿
- Cheng Zhang 🇨🇳
- Xiang Ji & Marc Suchard 🇺🇸
- My group 🇺🇸 : Seong-Hwan Jun, Michael Karcher, Andy Magee, Hassan Nasif, Chris Jennings-Shafer, Dave Rich, Lena Colliene, Mary Barker, Will Dumm
- Chris Whidden 🇨🇦

$$P(\mathbf{z} \mid D) = \frac{P(D \mid \mathbf{z})\, P(\mathbf{z})}{\sum_{\mathbf{z}'} P(D \mid \mathbf{z}')\, P(\mathbf{z}')}$$

- $\mathbf{z}$: tree structure and branch lengths
- D: sequence data
- $P(\mathbf{z})$: prior
- $P(D \mid \mathbf{z})$: phylogenetic likelihood

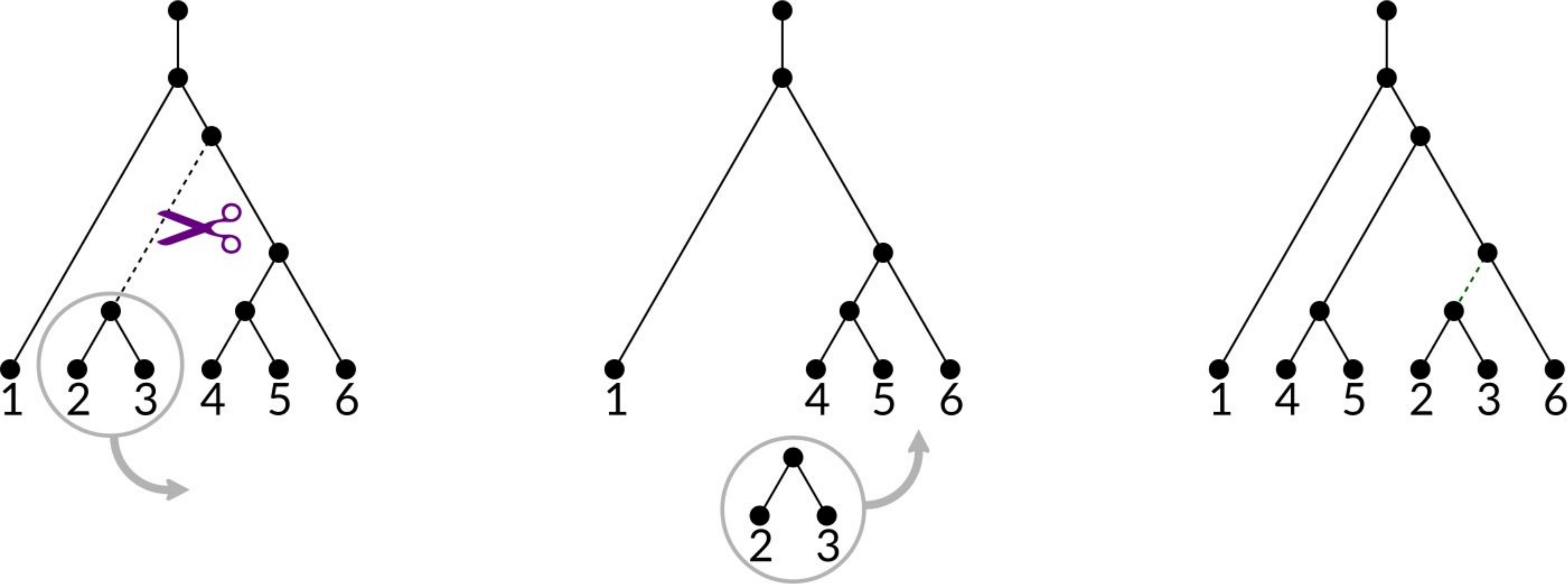


Estimating Effective Population Size and Mutation Rate From Sequence Data Using Metropolis-Hastings Sampling

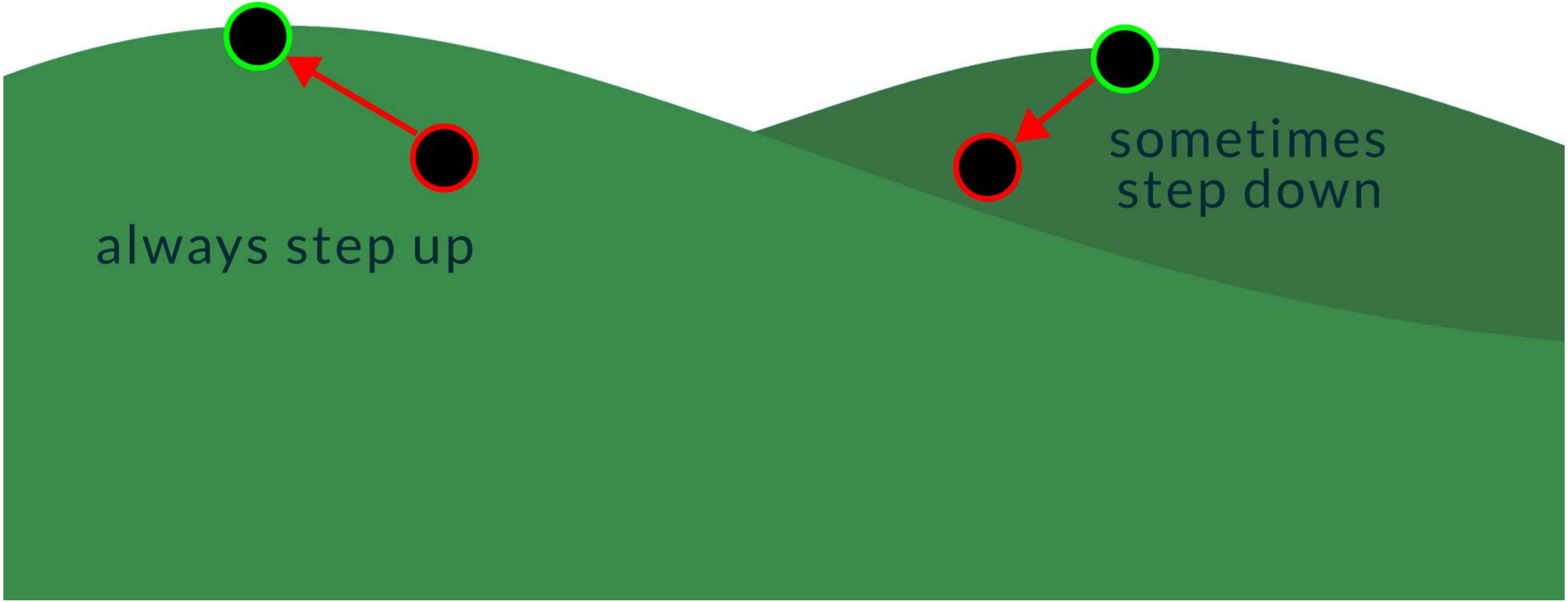
Mary K. Kuhner, Jon Yamato and Joseph Felsenstein

Department of Genetics, University of Washington, Seattle, Washington 98195-7360

Manuscript received April 5, 1994



Markov Chain Monte Carlo (MCMC)

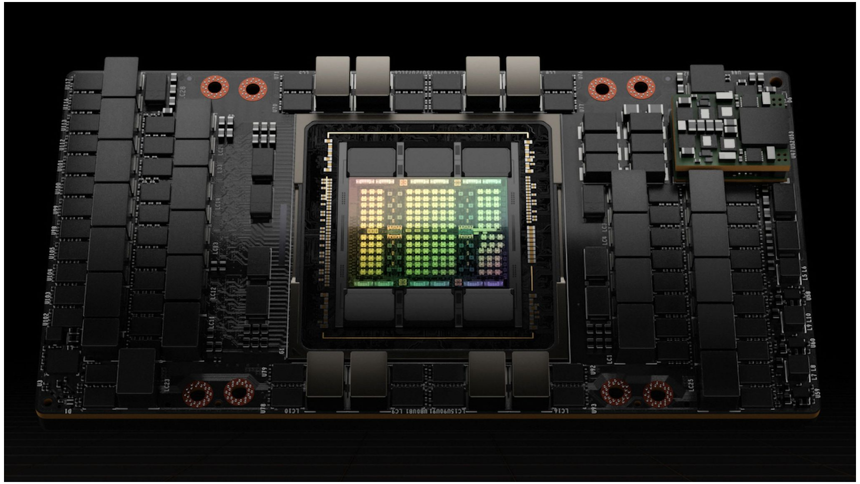


What is

1994

2024

NVIDIA

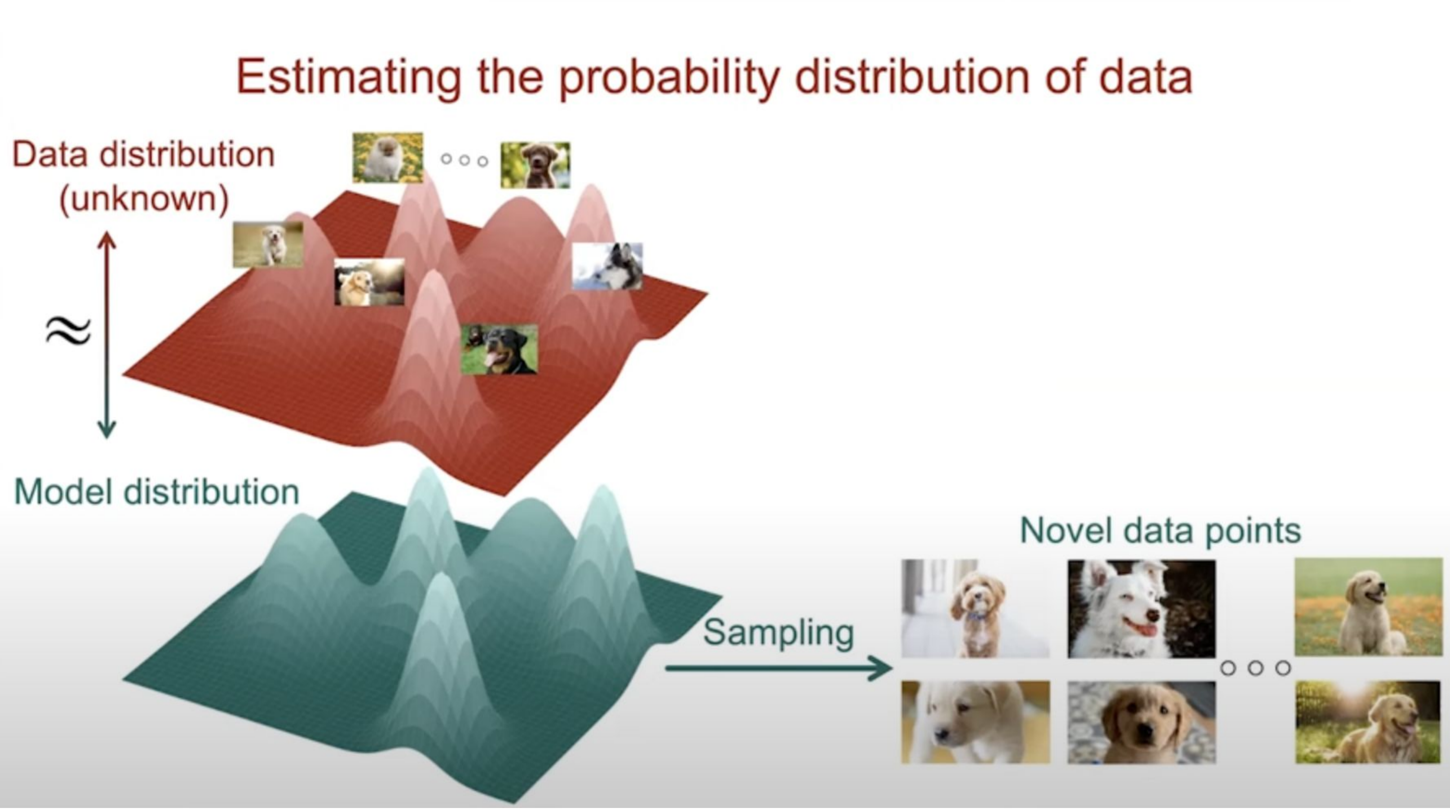


Bayesian Phylogenetics

MCMC

MCMC

What could be



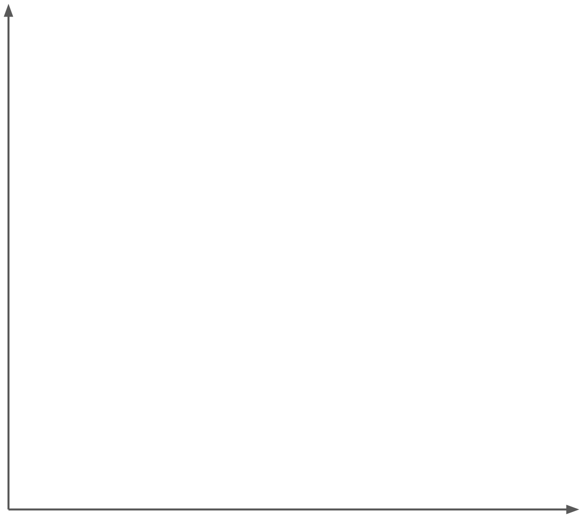
Estimating the probability distribution of data

Yang Song, https://www.youtube.com/watch?v=wMmqCMwuM2Q

- What is the shape of the posterior?
- Can we model the posterior?
- Can we infer the posterior using a method that doesn’t require MCMC?
- Does deep learning help?

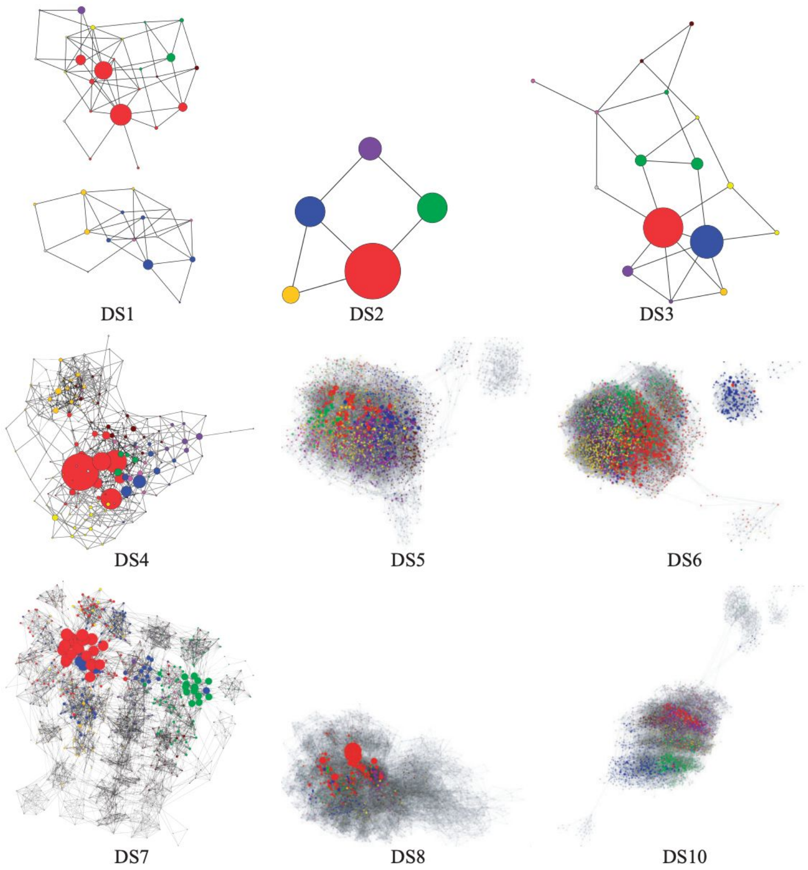
- Topological posterior
- Continuous parameters
- Joint models
- Next-gen proposals

- **What is the shape of the posterior?**
- Can we model the posterior?
- Can we infer the posterior using a method that doesn’t require MCMC?
- Does deep learning help?



- **Topological posterior**
- Continuous parameters
- Joint models
- Next-gen proposals

Phylogenetic posterior distributions have complex shapes!

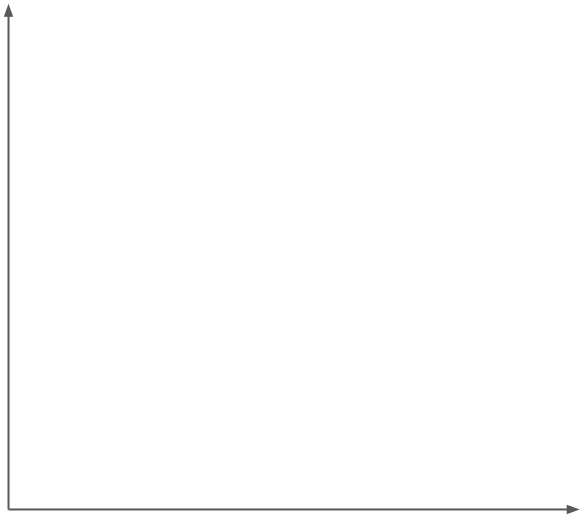


unrooted trees

http://dx.doi.org/10.1093/sysbio/syv006

- What is the shape of the posterior?
- **Can we model the posterior?**
- Can we infer the posterior using a method that doesn't require MCMC?
- Does deep learning help?

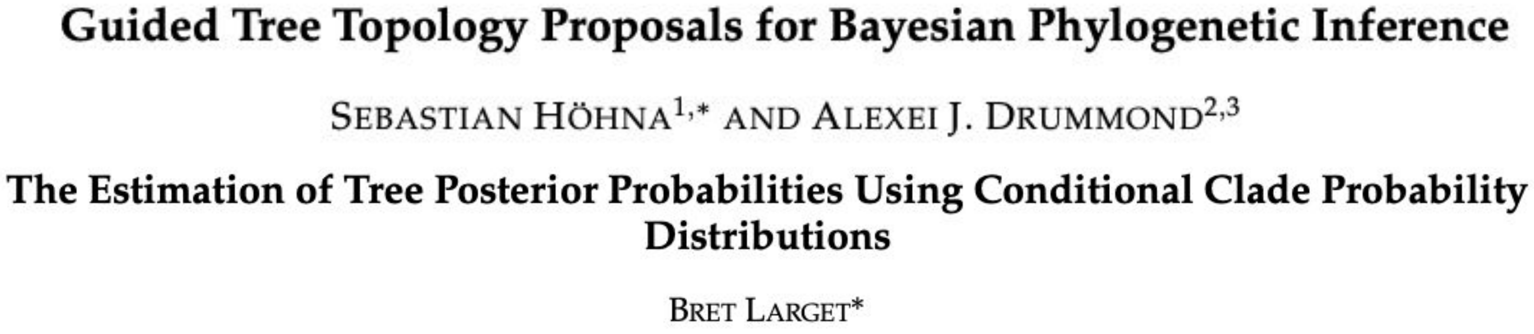
I mean a probability distribution that I can sample from directly



- **Topological posterior**

ameters

sals



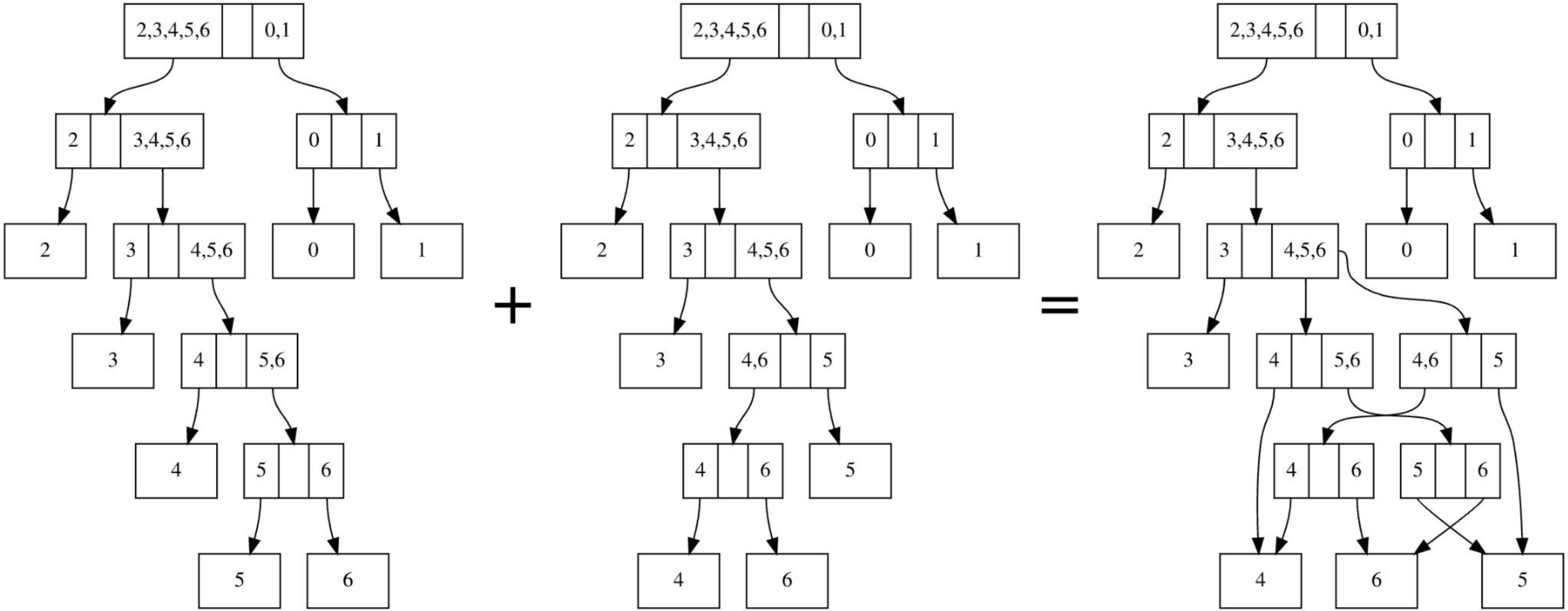
Guided Tree Topology Proposals for Bayesian Phylogenetic Inference

SEBASTIAN HÖHNA[1,*] AND ALEXEI J. DRUMMOND[2,3]

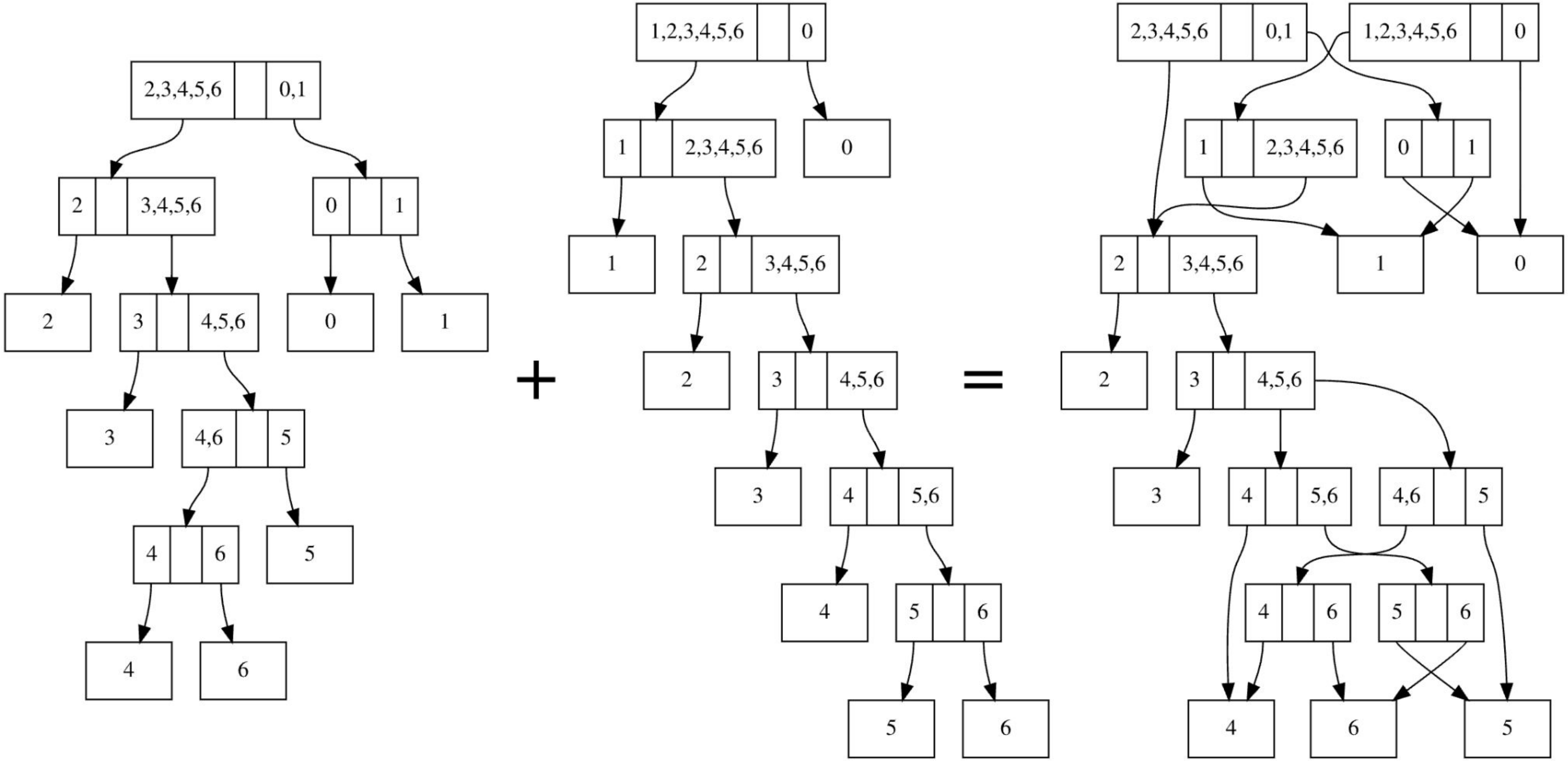
The Estimation of Tree Posterior Probabilities Using Conditional Clade Probability Distributions

BRET LARGET*

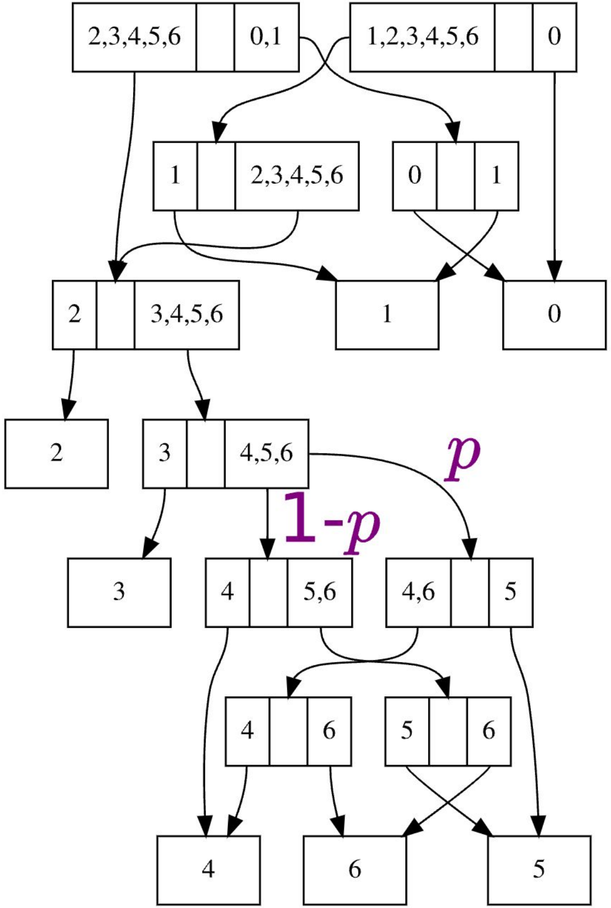
Subsplit Bayes Network (SBN) or subsplit DAG (sDAG)



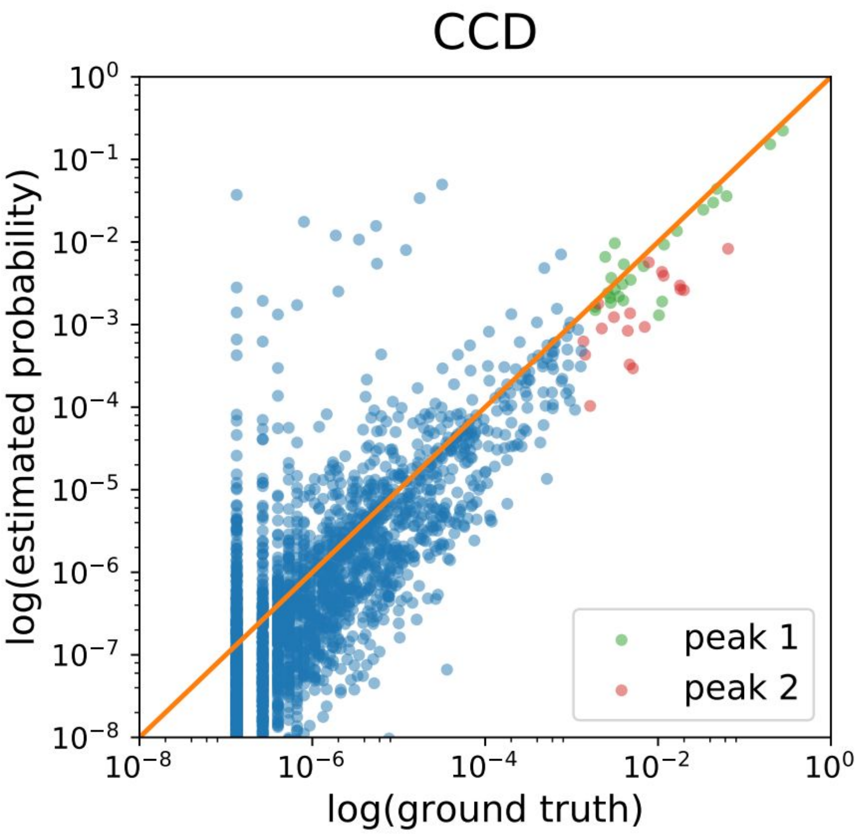
Subsplit Bayes Network (SBN) or subsplit DAG (sDAG)

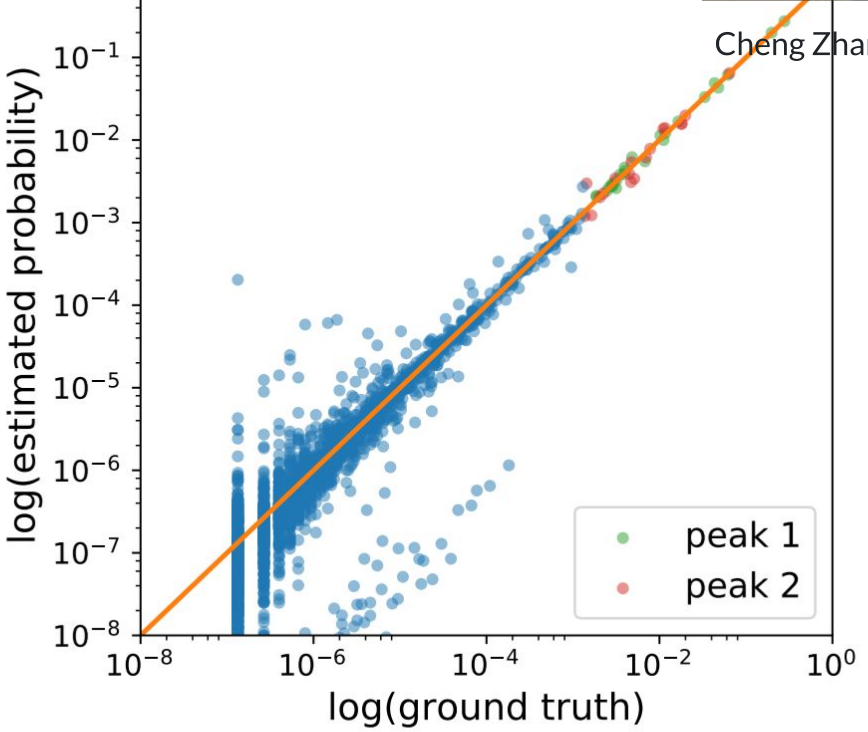


Attach probabilities to edges of DAG



SBN / sDAG can handle multi-modal posteriors



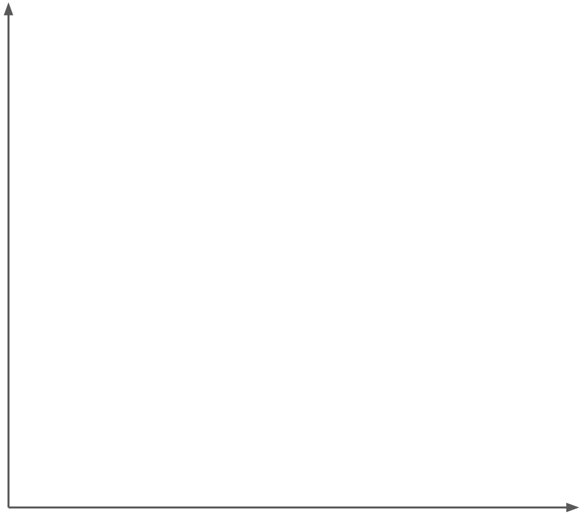


Cheng Zhang (张成)

unrooted trees

https://arxiv.org/abs/1805.07834

- What is the shape of the posterior?
- Can we model the posterior?
- **Can we infer the posterior using a method that doesn’t require MCMC?**
- Does deep learning help?

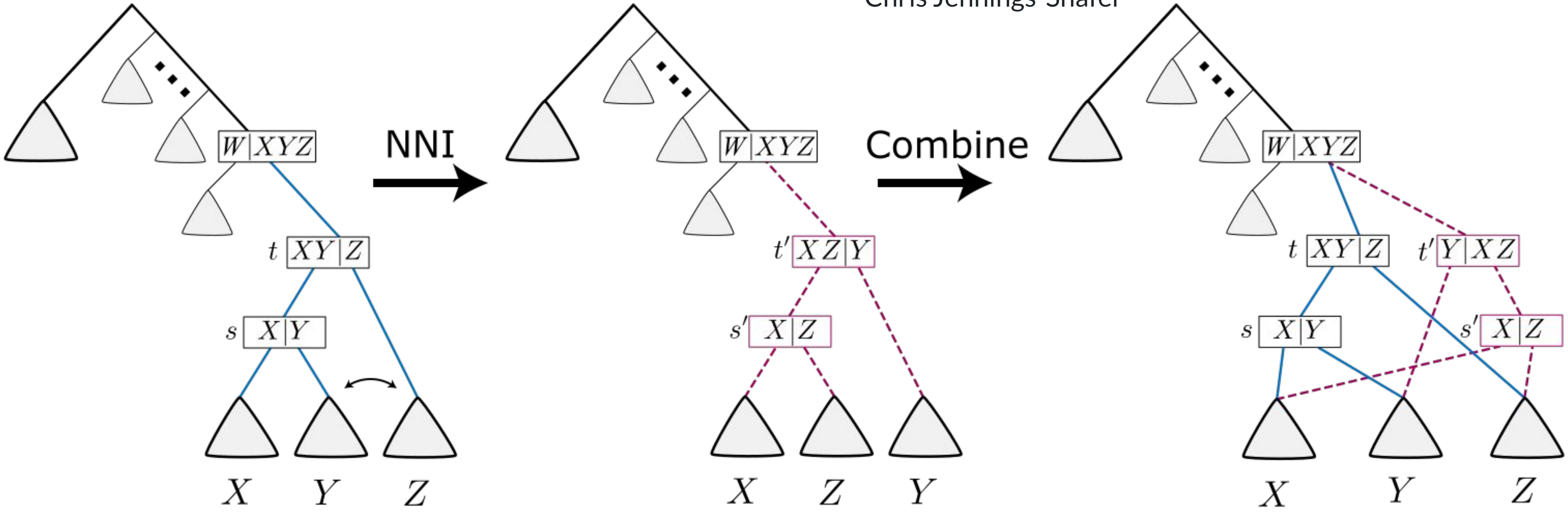


- **Topological posterior**
- Continuous parameters
- Joint models
- Next-gen proposals

Testing out an NNI on the sDAG

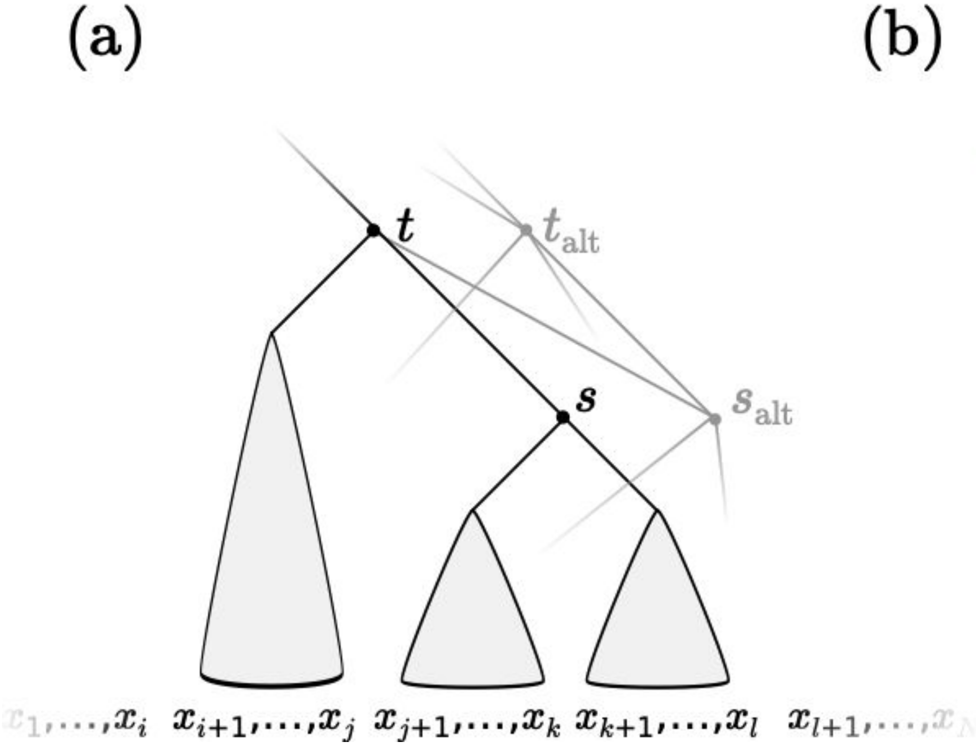
https://arxiv.org/abs/2411.09074

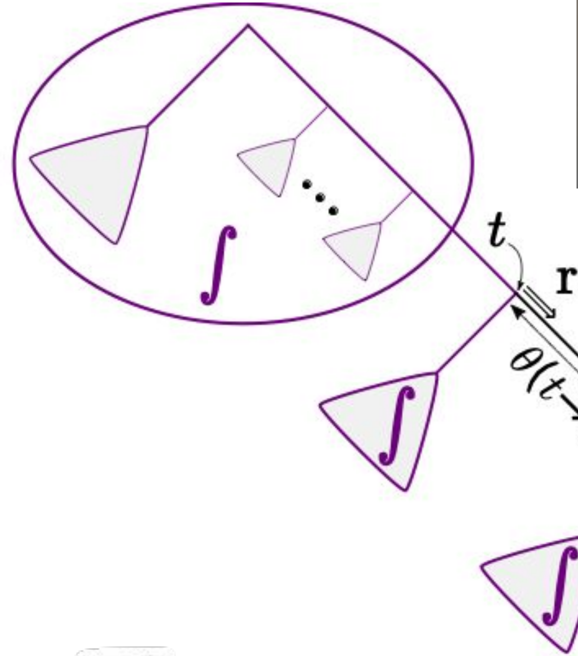
Chris Jennings-Shafer

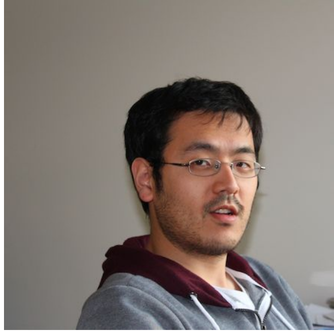


Q: How to decide if adding this structure is a good idea? There isn't one tree.

Generalized pruning





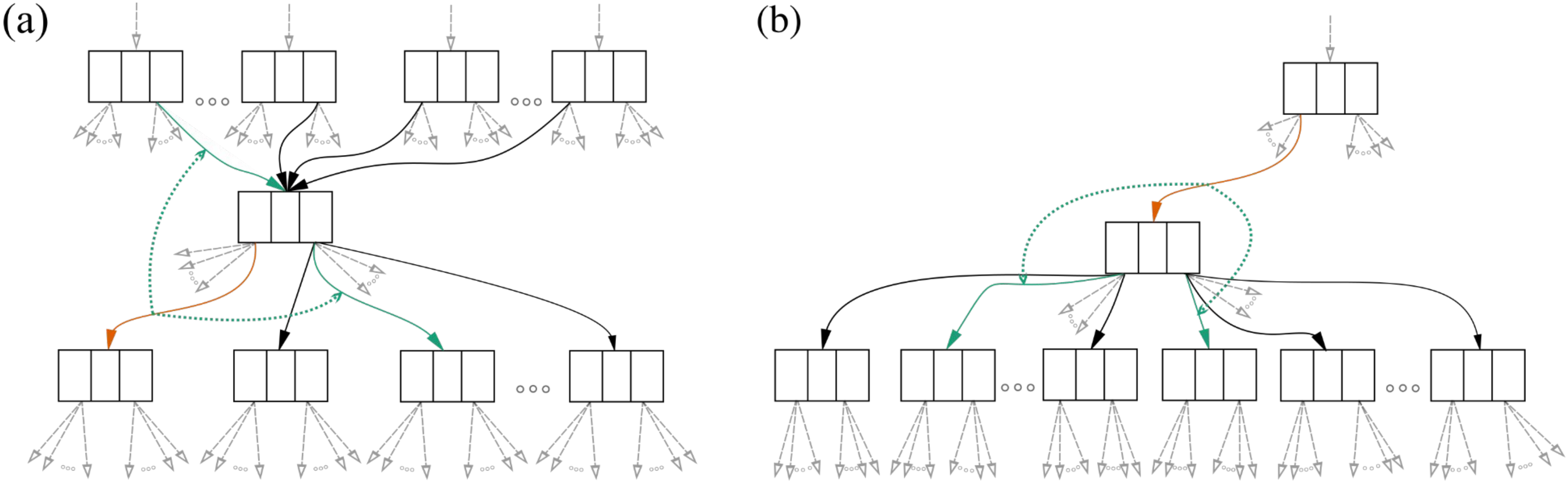


Seong-Hwan Jun

$$\prod_{k=1}^{K} \sum_{\tau \in \mathcal{D}} p_{\theta}(Y_k \mid \tau)\, p(\tau)$$

http://dx.doi.org/10.1186/s13015-023-00235-1

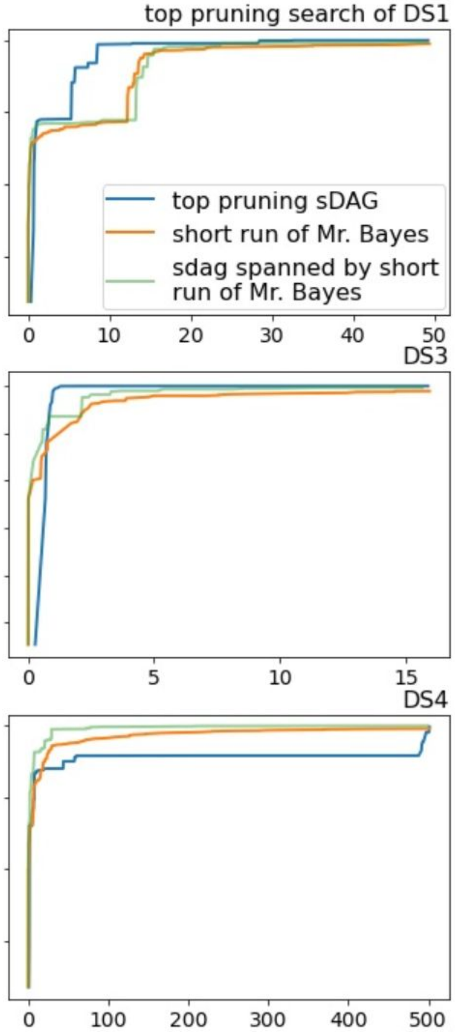
Top pruning



Associates a single “top” tree with every DAG edge. Branch lengths on edges. Likelihoods ✅

Can we infer the sDAG directly via systematic exploration?

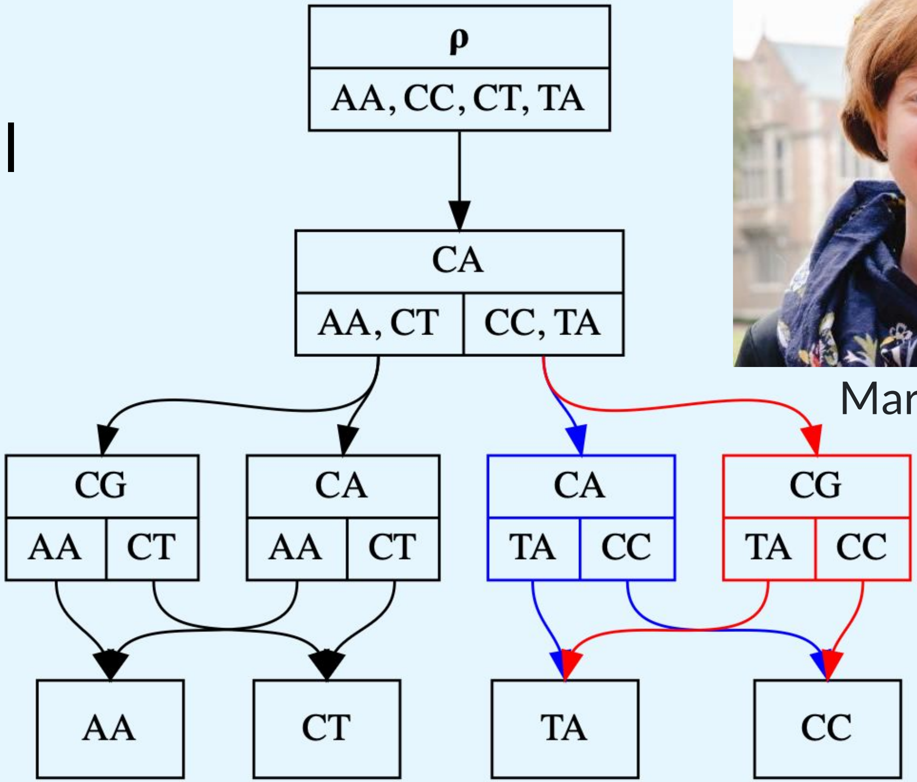
Yes, but running MrBayes & putting into an sDAG is just as good.



Problem: we can't evaluate the quality of edge in isolation.

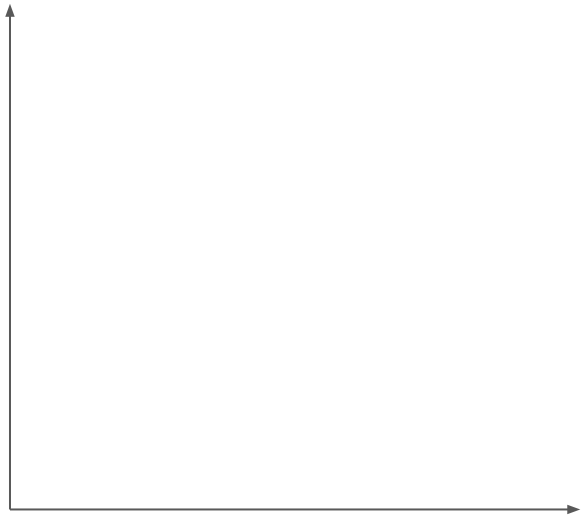
Solution: add internal node sequence information

→ "history DAG" or hDAG

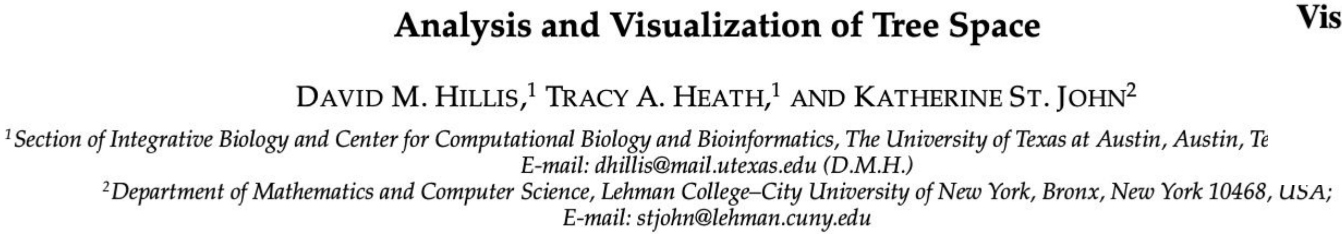



Mary Barker

- What is the shape of the posterior?
- **Can we model the posterior?**
- Can we infer the posterior using a method that doesn’t require MCMC?
- **Does deep learning help?**

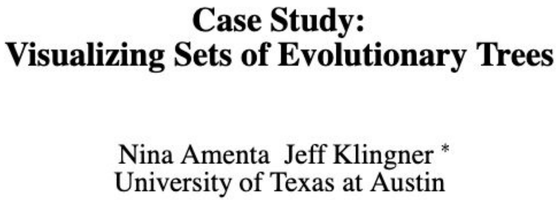


- **Topological posterior**
- Continuous parameters
- Joint models
- Next-gen proposals



Analysis and Visualization of Tree Space

David M. Hillis,[1] Tracy A. Heath,[1] and Katherine St. John[2]
[1] *Section of Integrative Biology and Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, Te*
E-mail: dhillis@mail.utexas.edu (D.M.H.)
[2] *Department of Mathematics and Computer Science, Lehman College–City University of New York, Bronx, New York 10468, USA;*
E-mail: stjohn@lehman.cuny.edu



Case Study:
Visualizing Sets of Evolutionary Trees

Nina Amenta Jeff Klingner *
University of Texas at Austin

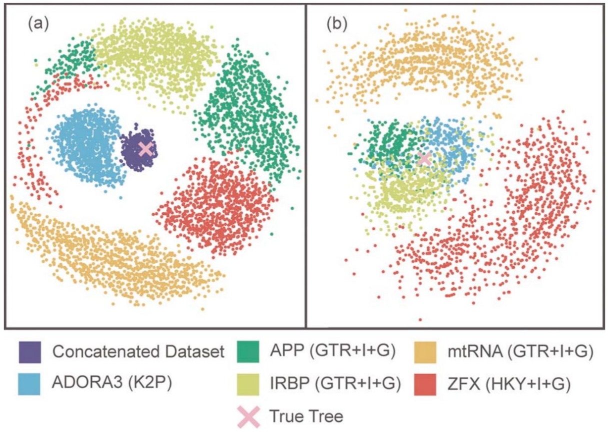
Set of trees

tree distance

Pairwise distance matrix

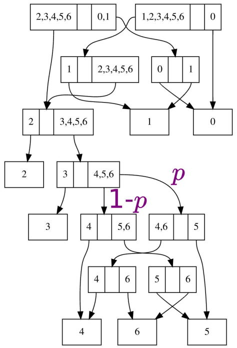
multidim scaling

(MDS)

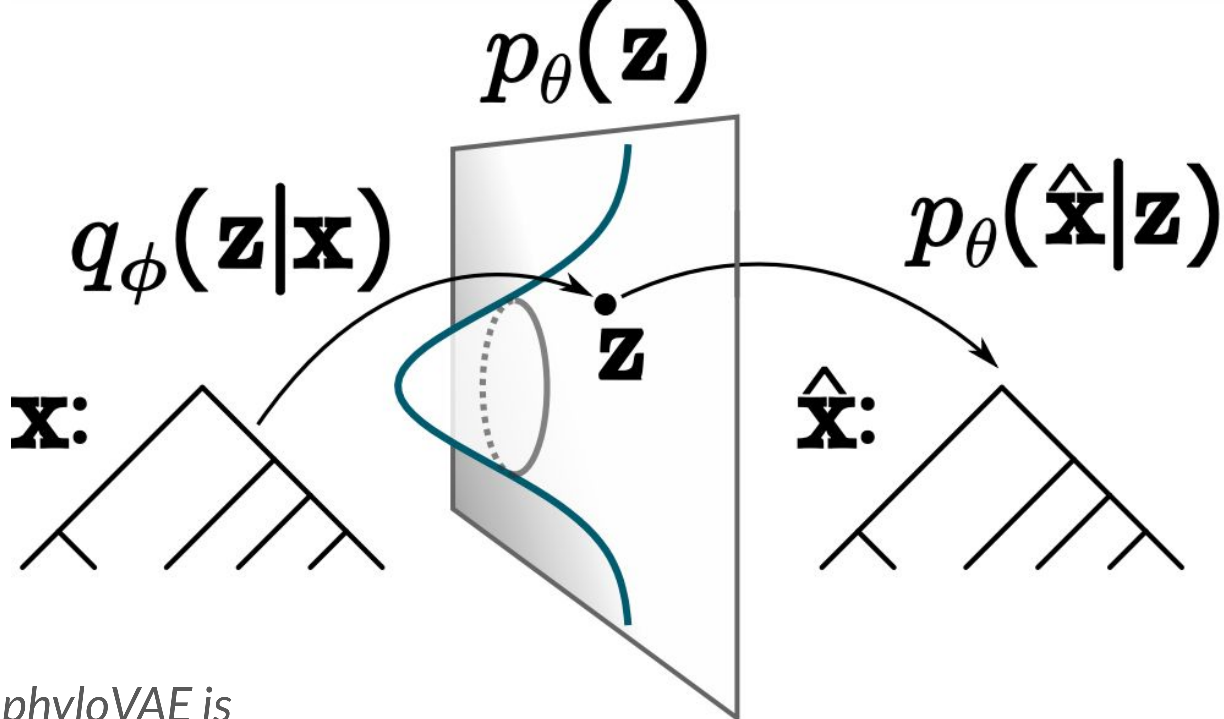


Set of trees

Probabilistic model of set of trees

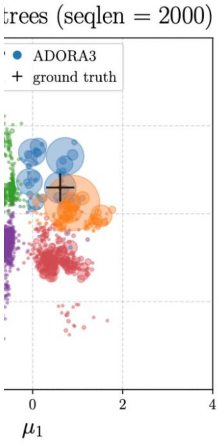


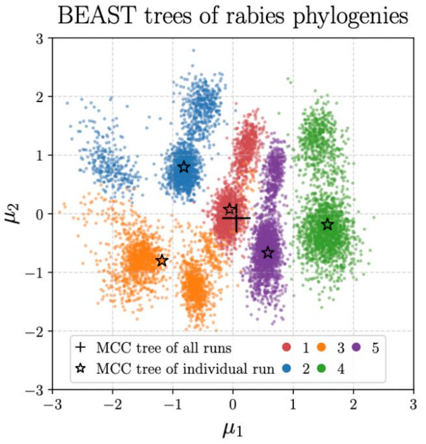
LET'S DO BOTH



Tianyu Xie

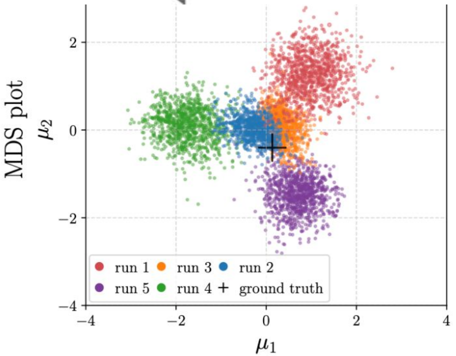
Cheng Zhang

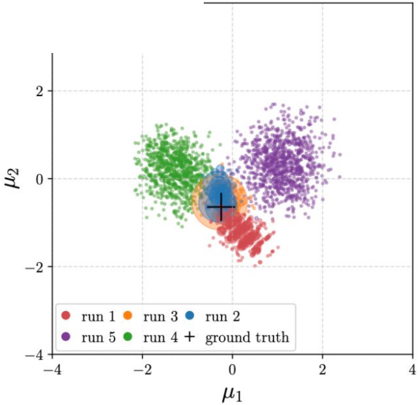


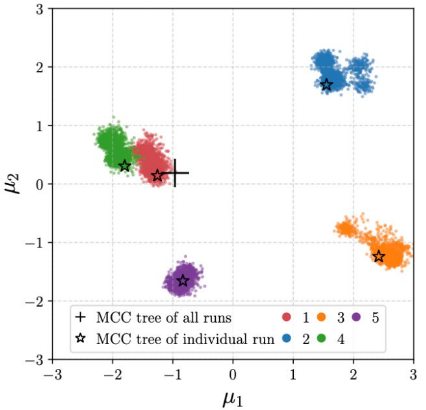


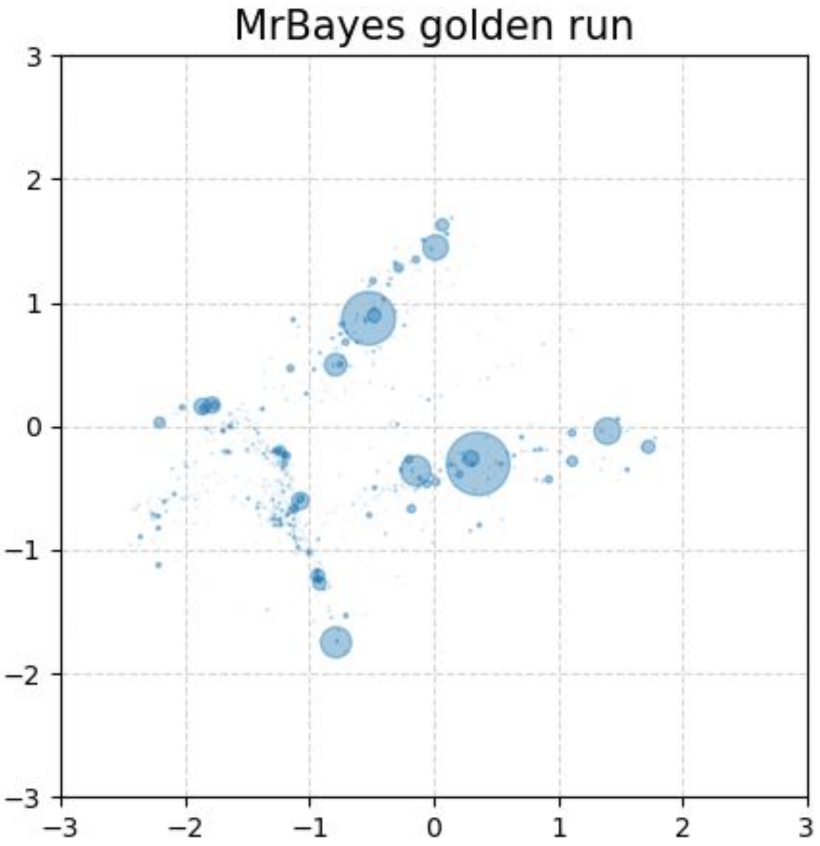
phyloVAE is simultaneously a tree visualization and a probabilistic model for trees

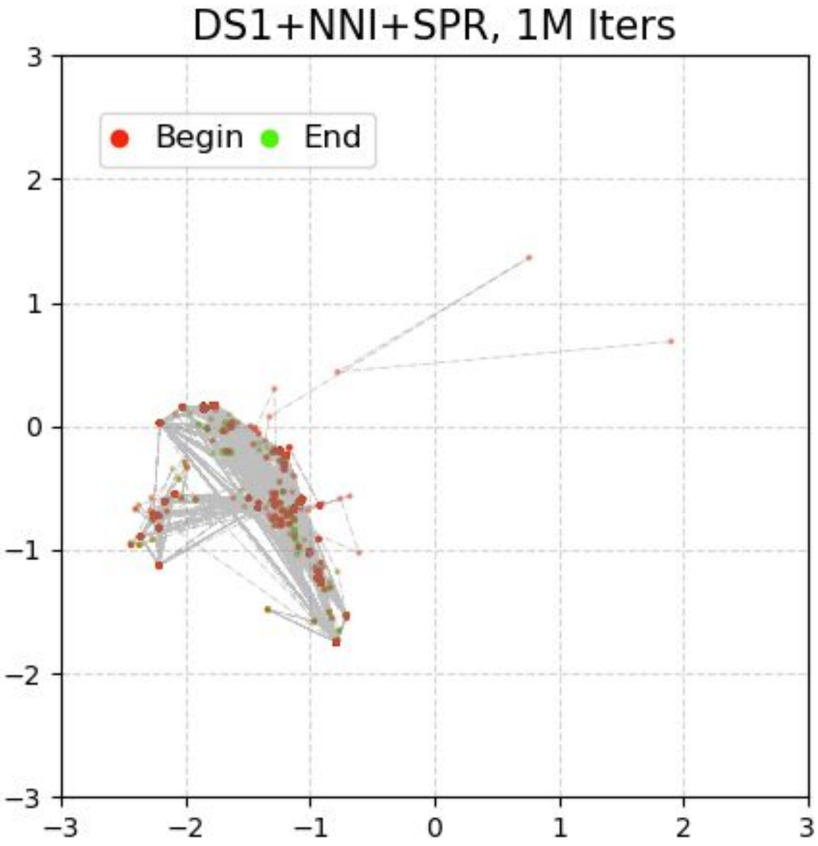
[paper (secret link)]

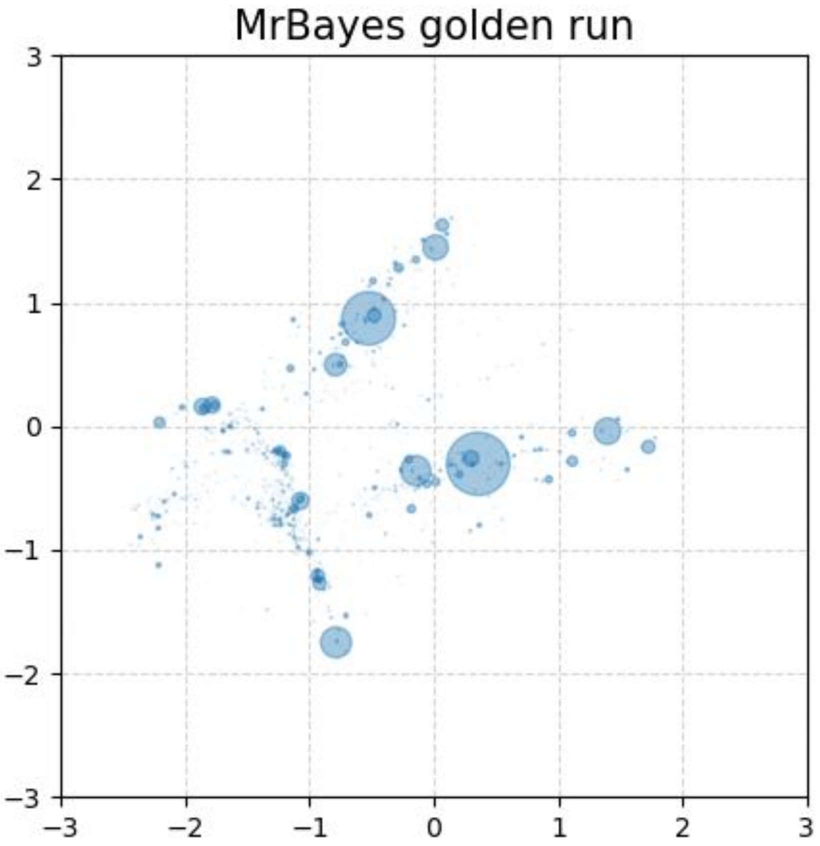


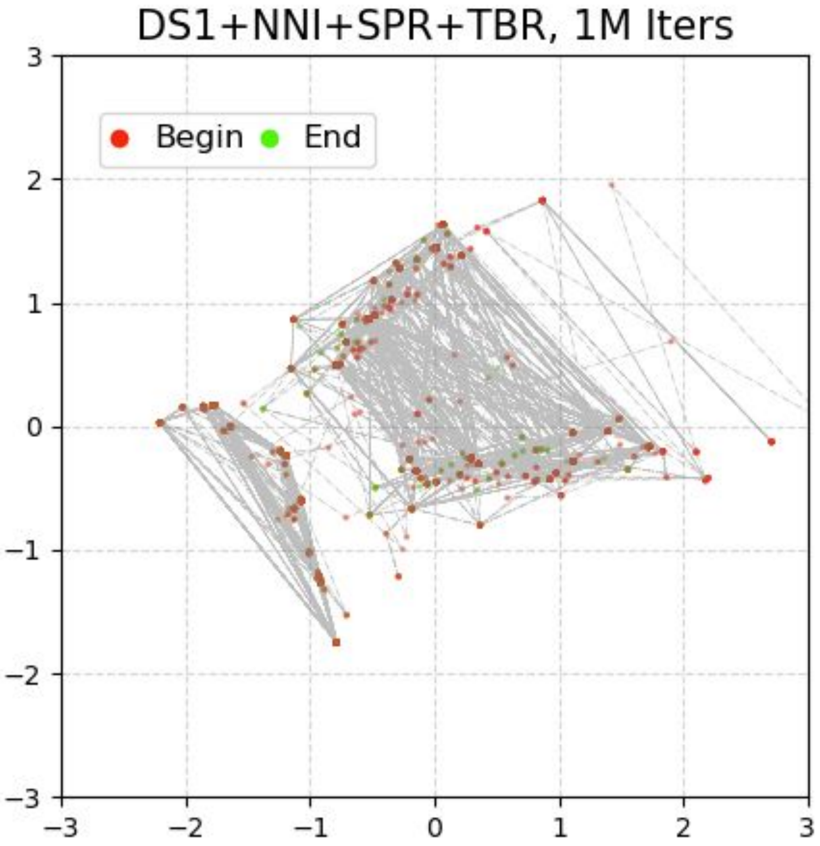




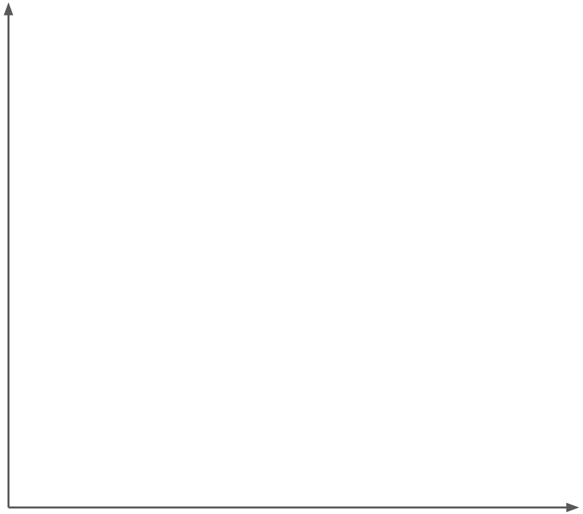

MrBayes golden run
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3


DS1+NNI+SPR, 1M Iters
● Begin ● End
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3


MrBayes golden run
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

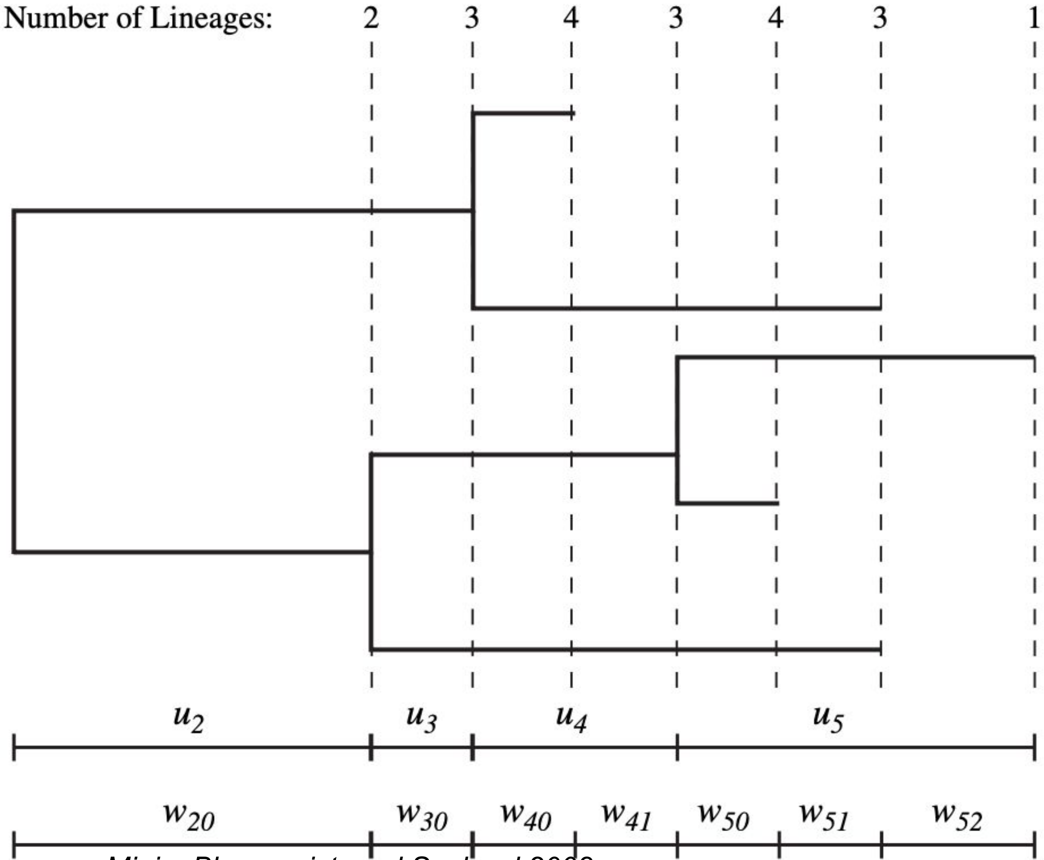

DS1+NNI+SPR+TBR, 1M Iters
Begin
End
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

- What is the shape of the posterior?
- Can we model the posterior?
- **Can we infer the posterior using a method that doesn’t require MCMC?**
- Does deep learning help?

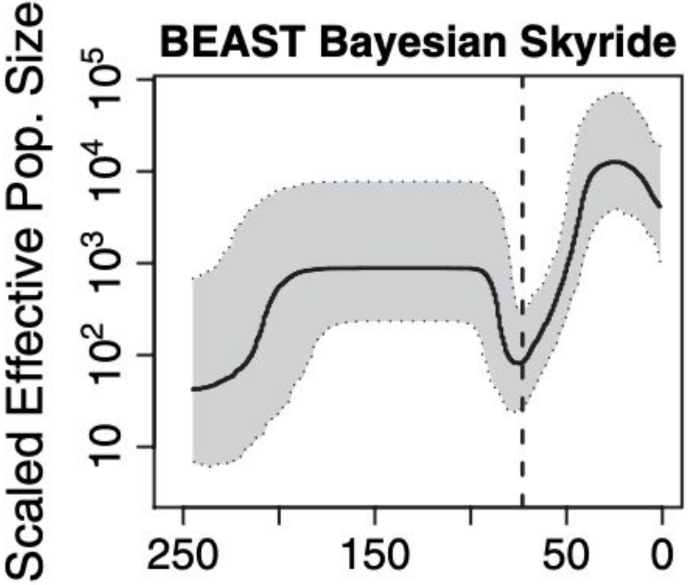


- Topological posterior
- **Continuous parameters**
- Joint models
- Next-gen proposals

Fixed tree, infer distribution of continuous parameters



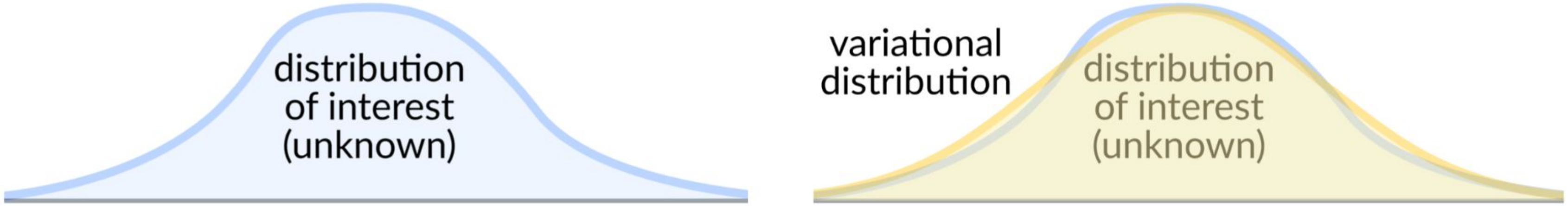
Minin, Bloomquist, and Suchard 2008



Variational inference for continuous parameters

Variational inference for continuous parameters

We can fit the variational distribution without knowing the full distribution of interest



We only need to evaluate

$$p(Y, \theta) = p(Y \mid \theta) p(\theta)$$

and (usually) $\nabla p(Y, \theta)$ wrt θ using gradient-ascent optimization

Similar to Rosenbluth-Teller-Hastings (MH) trick

Marc Suchard

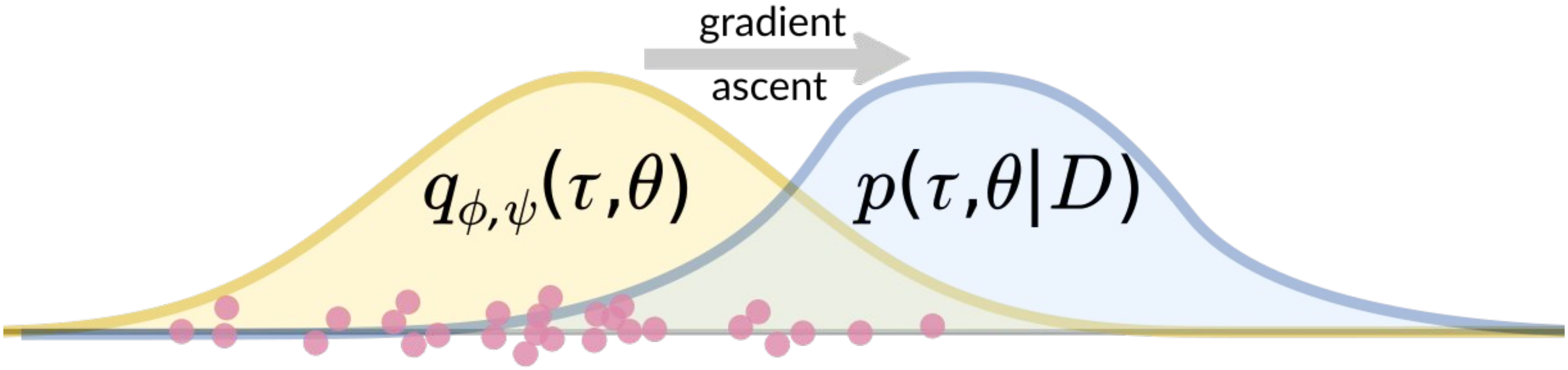
VB seeks to minimize a dissimilarity measure $\mu(\cdot \,||\, \cdot)$ between the posterior $p(\theta \mid Y)$ and a variational distribution $q(\theta; \phi) \in \mathcal{Q}$ from a family indexed by ϕ:

$$\hat{q}(\theta) = \underset{q(\theta;\phi)}{\operatorname{argmin}} \, \mu\left(q(\theta;\phi) \,||\, p(\theta \mid Y)\right)$$

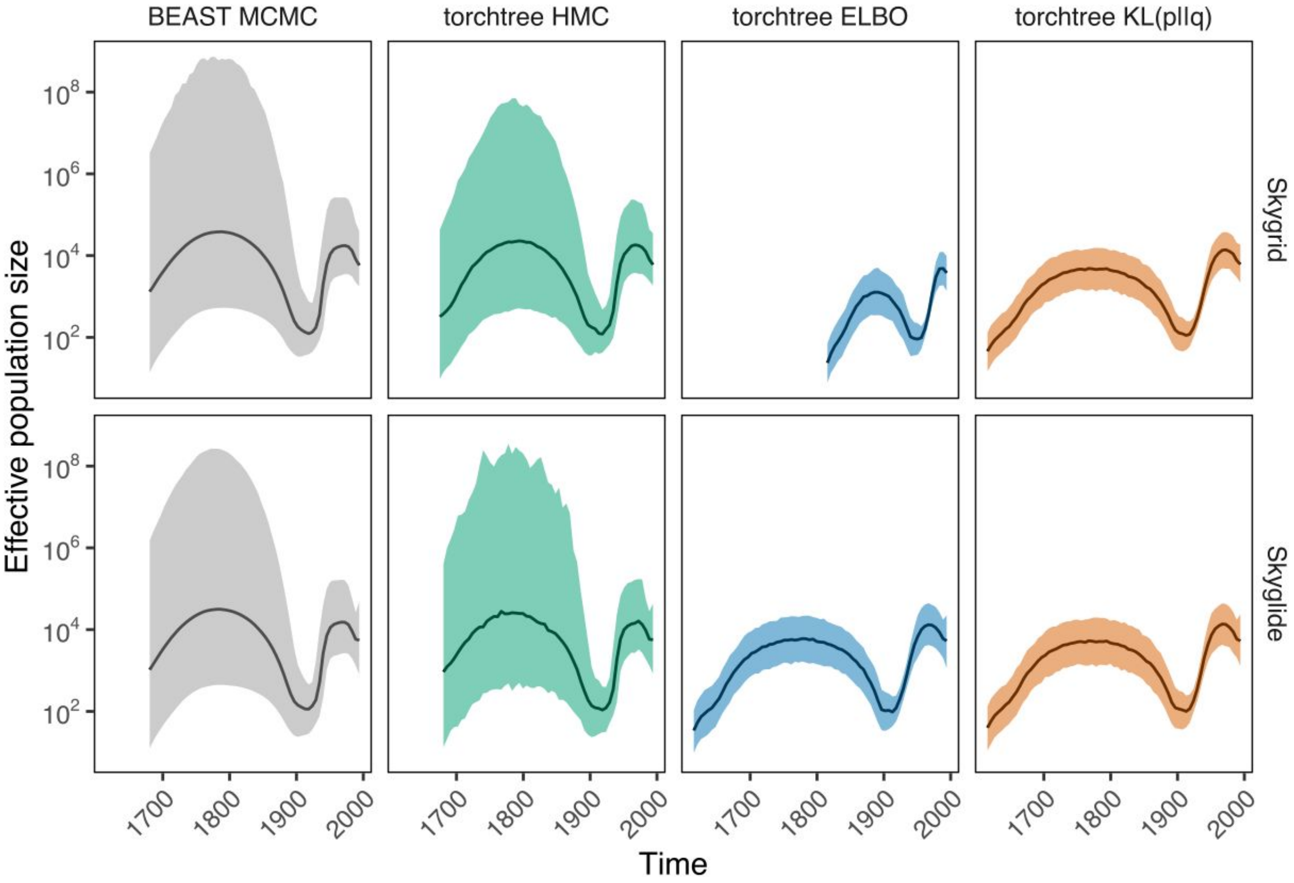
Typical distances

backward Kullback-Leibler (KL) divergence $\mathbb{E}\left[\log q(\theta;\phi) - \log p(\theta \mid Y)\right]$ wrt $q(\theta;\phi)$	called ELBO, convenient, under-dispersed ← wrong expectation
forward KL divergence $\mathbb{E}\left[\log p(\theta \mid Y) - \log q(\theta;\phi)\right]$ wrt $p(\theta \mid Y)$	intent of KL, expectation wrt "ground truth", requires (self-normalizing) importance sampling

Variational inference for continuous parameters



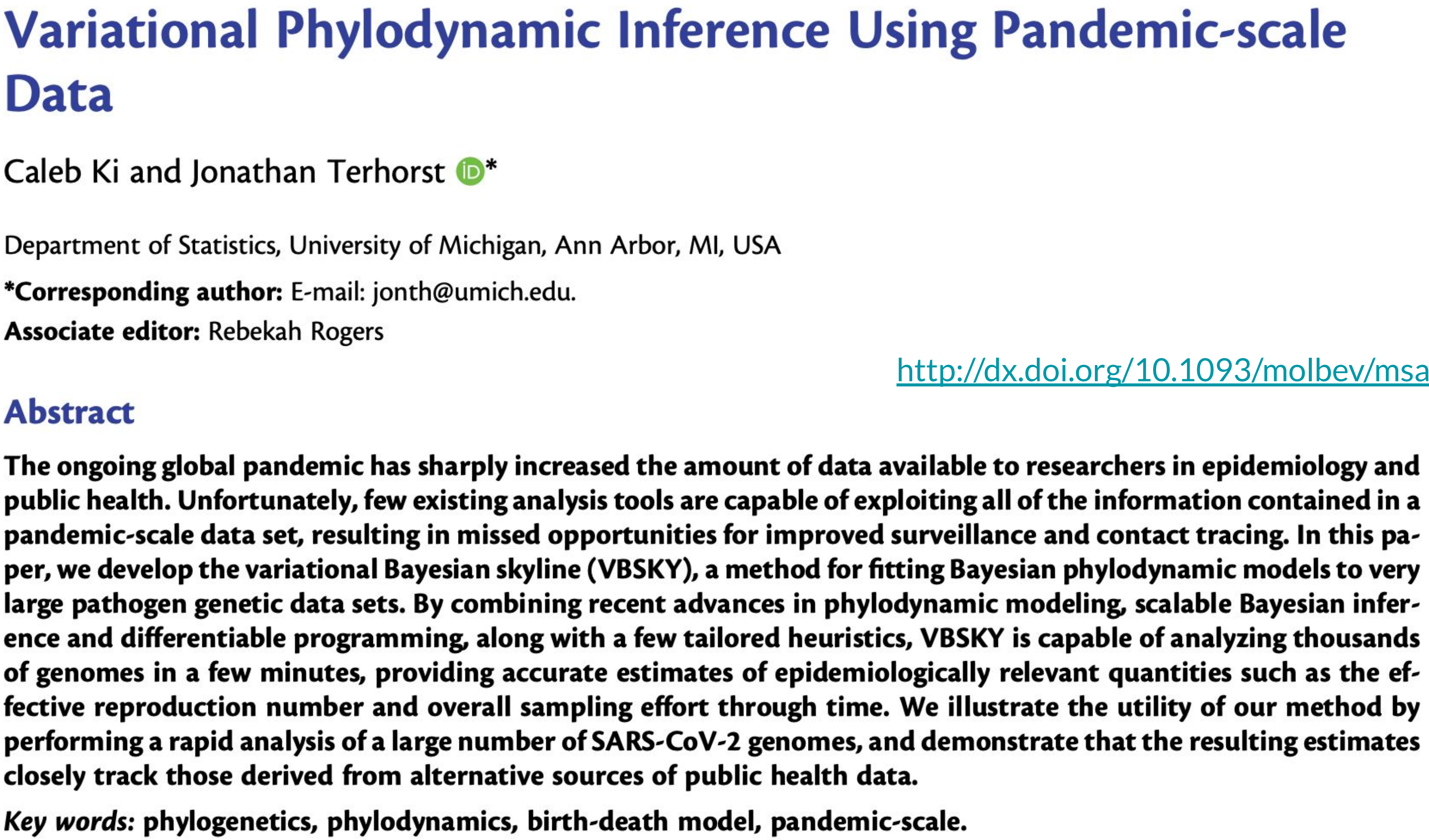
Variational inference works, with caveats



Mathieu Fourment

phylostan (2019): use Stan to formulate models

torchtree (2024): use PyTorch to formulate models



Variational Phylodynamic Inference Using Pandemic-scale Data

Caleb Ki and Jonathan Terhorst*

Department of Statistics, University of Michigan, Ann Arbor, MI, USA

***Corresponding author:** E-mail: jonth@umich.edu.

Associate editor: Rebekah Rogers

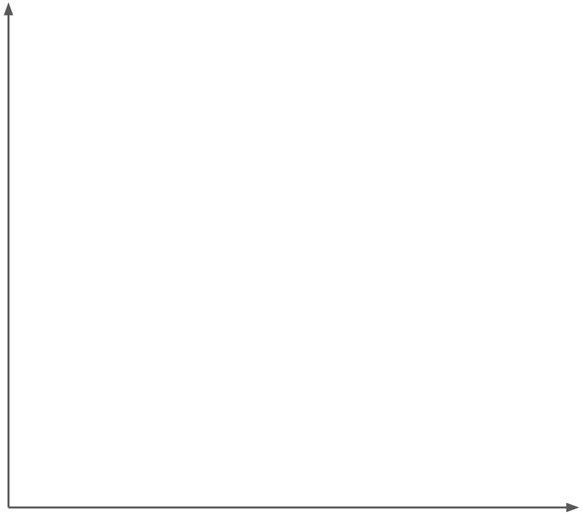
http://dx.doi.org/10.1093/molbev/msac154

Abstract

The ongoing global pandemic has sharply increased the amount of data available to researchers in epidemiology and public health. Unfortunately, few existing analysis tools are capable of exploiting all of the information contained in a pandemic-scale data set, resulting in missed opportunities for improved surveillance and contact tracing. In this paper, we develop the variational Bayesian skyline (VBSKY), a method for fitting Bayesian phylodynamic models to very large pathogen genetic data sets. By combining recent advances in phylodynamic modeling, scalable Bayesian inference and differentiable programming, along with a few tailored heuristics, VBSKY is capable of analyzing thousands of genomes in a few minutes, providing accurate estimates of epidemiologically relevant quantities such as the effective reproduction number and overall sampling effort through time. We illustrate the utility of our method by performing a rapid analysis of a large number of SARS-CoV-2 genomes, and demonstrate that the resulting estimates closely track those derived from alternative sources of public health data.

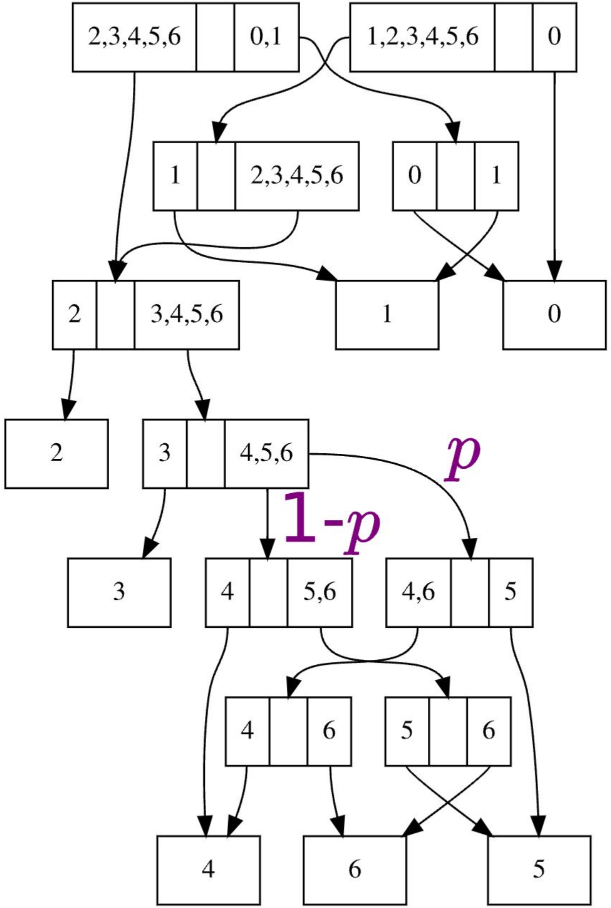
Key words: **phylogenetics, phylodynamics, birth-death model, pandemic-scale.**

- What is the shape of the posterior?
- Can we model the posterior?
- **Can we infer the posterior using a method that doesn’t require MCMC?**
- Does deep learning help?



- Topological posterior
- Continuous parameters
- **Joint models**
- Next-gen proposals

Joint models on topology and continuous parameters



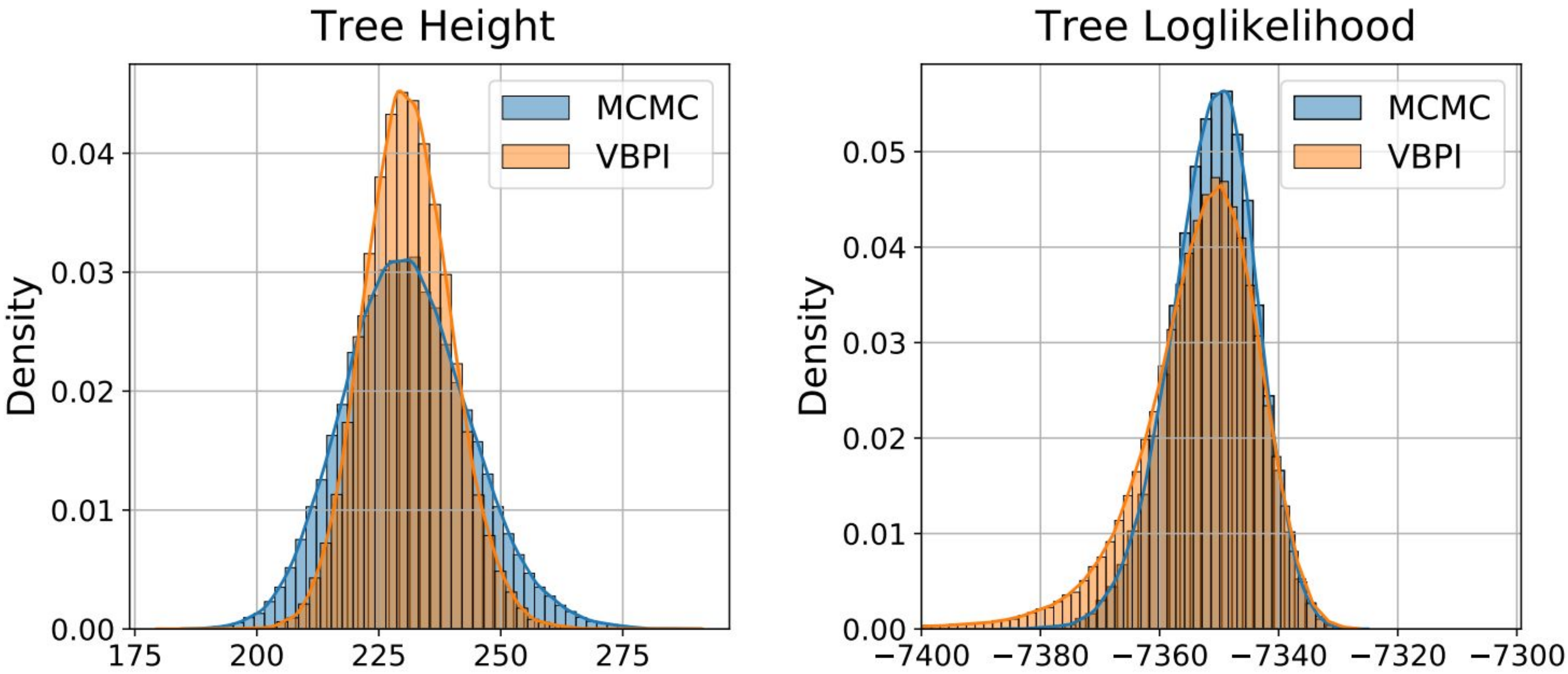
DAG edge probabilities

variational distributions

(on branch lengths, *etc*)

tree topology DAG

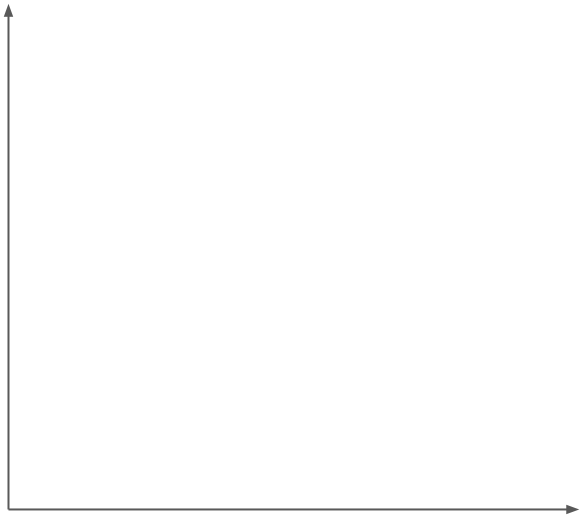
Joint topological posterior / model works, with caveats



https://www.jmlr.org/papers/v25/22-0348.html

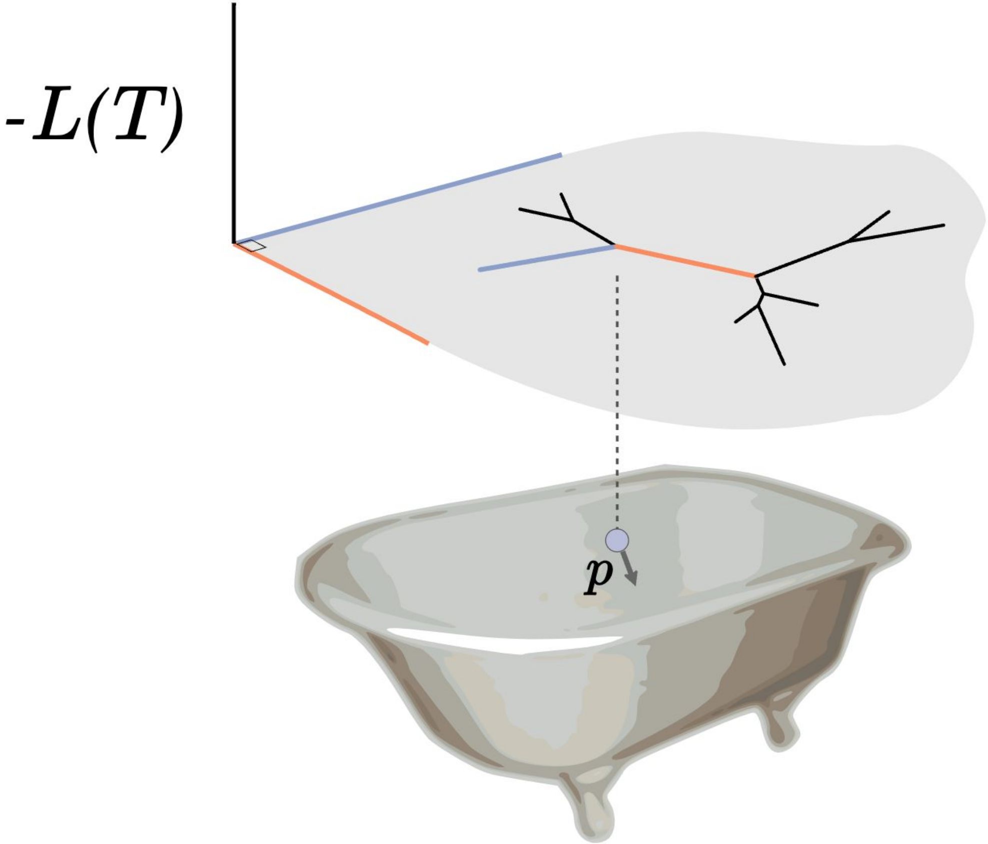
- It is possible, at least for data sets of modest size, to model the shape of the phylogenetic posterior distribution
- One can turn this into an inferential method, with caveats
- We haven’t been able to get it to beat MCMC (modeling the full posterior is a challenge, gradients are expensive, and optimization is fiddly)

- What is the shape of the posterior?
- Can we model the posterior?
- Can we infer the posterior using a method that doesn’t require MCMC?
- Does deep learning help?

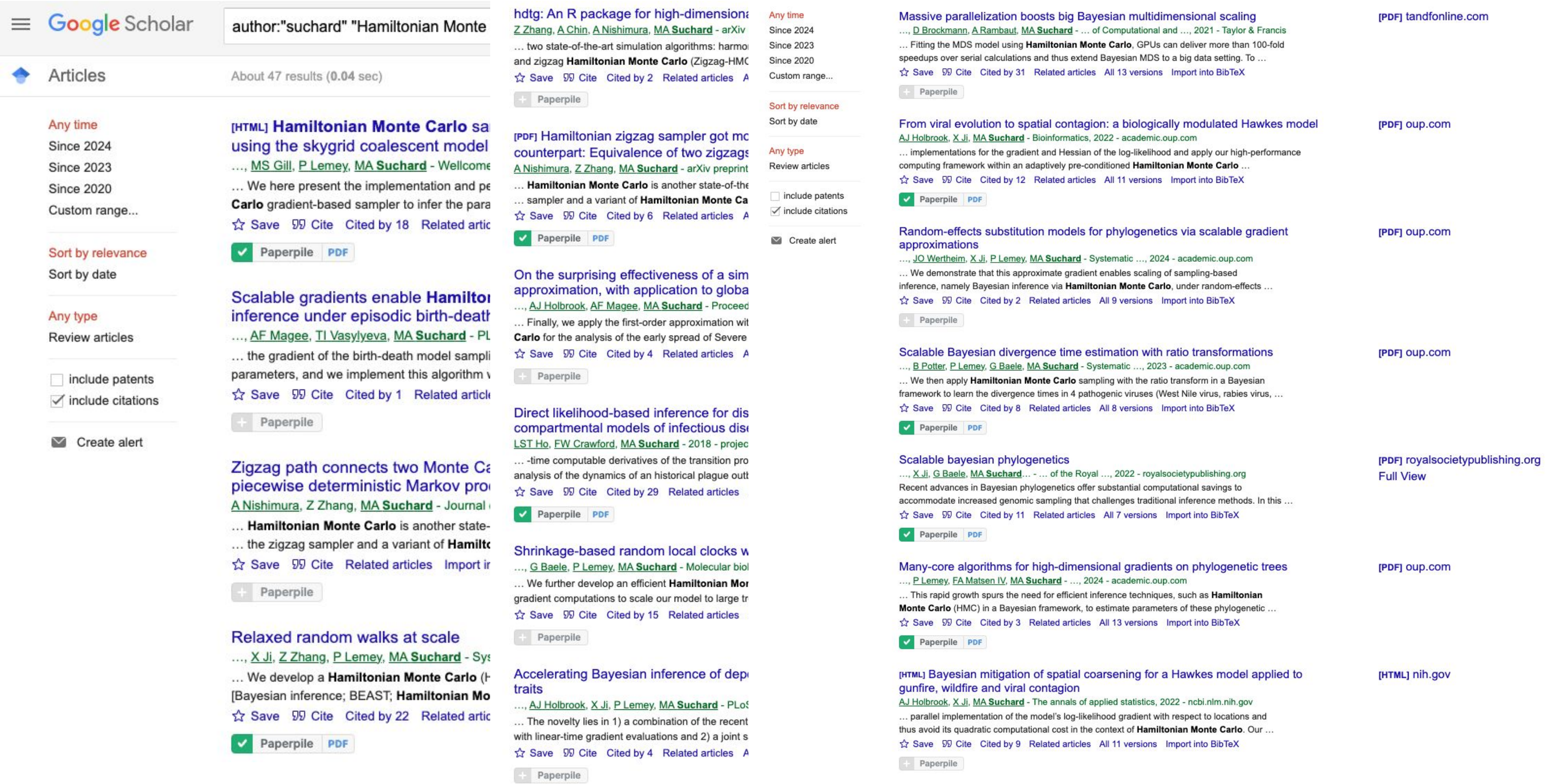


- Topological posterior
- Continuous parameters
- Joint models
- **Next-gen proposals**

Hamiltonian Monte Carlo



This is very effective for continuous parameters



About 47 results (**0.04** sec)

[HTML] **Hamiltonian Monte Carlo** sa
using the skygrid coalescent model
…, MS Gill, P Lemey, MA **Suchard** - Wellcome
… We here present the implementation and pe
Carlo gradient-based sampler to infer the para
☆ Save 99 Cite Cited by 18 Related artic
✓ Paperpile PDF

Scalable gradients enable **Hamilto**
inference under episodic birth-deatl
…, AF Magee, TI Vasylyeva, MA **Suchard** - PL
… the gradient of the birth-death model sampli
parameters, and we implement this algorithm v
☆ Save 99 Cite Cited by 1 Related article
\+ Paperpile

Zigzag path connects two Monte Ca
piecewise deterministic Markov proc
A Nishimura, Z Zhang, MA **Suchard** - Journal
… **Hamiltonian Monte Carlo** is another state-
… the zigzag sampler and a variant of **Hamilto**
☆ Save 99 Cite Related articles Import ir
\+ Paperpile

Relaxed random walks at scale
…, X Ji, Z Zhang, P Lemey, MA **Suchard** - Sys
… We develop a **Hamiltonian Monte Carlo** (H
[Bayesian inference; BEAST; **Hamiltonian Mo**
☆ Save 99 Cite Cited by 22 Related artic
✓ Paperpile PDF

hdtg: An R package for high-dimensiona
Z Zhang, A Chin, A Nishimura, MA **Suchard** - arXiv
… two state-of-the-art simulation algorithms: harmo
and zigzag **Hamiltonian Monte Carlo** (Zigzag-HMC
☆ Save 99 Cite Cited by 2 Related articles A
\+ Paperpile

[PDF] Hamiltonian zigzag sampler got mo
counterpart: Equivalence of two zigzags
A Nishimura, Z Zhang, MA **Suchard** - arXiv preprint
… **Hamiltonian Monte Carlo** is another state-of-the
… sampler and a variant of **Hamiltonian Monte Ca**
☆ Save 99 Cite Cited by 6 Related articles A
✓ Paperpile PDF

On the surprising effectiveness of a sim
approximation, with application to globa
…, AJ Holbrook, AF Magee, MA **Suchard** - Proceed
… Finally, we apply the first-order approximation wit
Carlo for the analysis of the early spread of Severe
☆ Save 99 Cite Cited by 4 Related articles A
\+ Paperpile

Direct likelihood-based inference for dis
compartmental models of infectious dis
LST Ho, FW Crawford, MA **Suchard** - 2018 - projec
… -time computable derivatives of the transition pro
analysis of the dynamics of an historical plague outl
☆ Save 99 Cite Cited by 29 Related articles
✓ Paperpile PDF

Shrinkage-based random local clocks w
…, G Baele, P Lemey, MA **Suchard** - Molecular biol
… We further develop an efficient **Hamiltonian Mor**
gradient computations to scale our model to large tr
☆ Save 99 Cite Cited by 15 Related articles
\+ Paperpile

Accelerating Bayesian inference of dep
traits
…, AJ Holbrook, X Ji, P Lemey, MA **Suchard** - PLoS
… The novelty lies in 1) a combination of the recent
with linear-time gradient evaluations and 2) a joint s
☆ Save 99 Cite Cited by 4 Related articles A
\+ Paperpile

Massive parallelization boosts big Bayesian multidimensional scaling — [PDF] tandfonline.com
…, D Brockmann, A Rambaut, MA **Suchard** - … of Computational and …, 2021 - Taylor & Francis
… Fitting the MDS model using **Hamiltonian Monte Carlo**, GPUs can deliver more than 100-fold speedups over serial calculations and thus extend Bayesian MDS to a big data setting. To …
☆ Save 99 Cite Cited by 31 Related articles All 13 versions Import into BibTeX
\+ Paperpile

From viral evolution to spatial contagion: a biologically modulated Hawkes model — [PDF] oup.com
AJ Holbrook, X Ji, MA **Suchard** - Bioinformatics, 2022 - academic.oup.com
… implementations for the gradient and Hessian of the log-likelihood and apply our high-performance computing framework within an adaptively pre-conditioned **Hamiltonian Monte Carlo** …
☆ Save 99 Cite Cited by 12 Related articles All 11 versions Import into BibTeX
✓ Paperpile PDF

Random-effects substitution models for phylogenetics via scalable gradient approximations — [PDF] oup.com
…, JO Wertheim, X Ji, P Lemey, MA **Suchard** - Systematic …, 2024 - academic.oup.com
… We demonstrate that this approximate gradient enables scaling of sampling-based inference, namely Bayesian inference via **Hamiltonian Monte Carlo**, under random-effects …
☆ Save 99 Cite Cited by 2 Related articles All 9 versions Import into BibTeX
\+ Paperpile

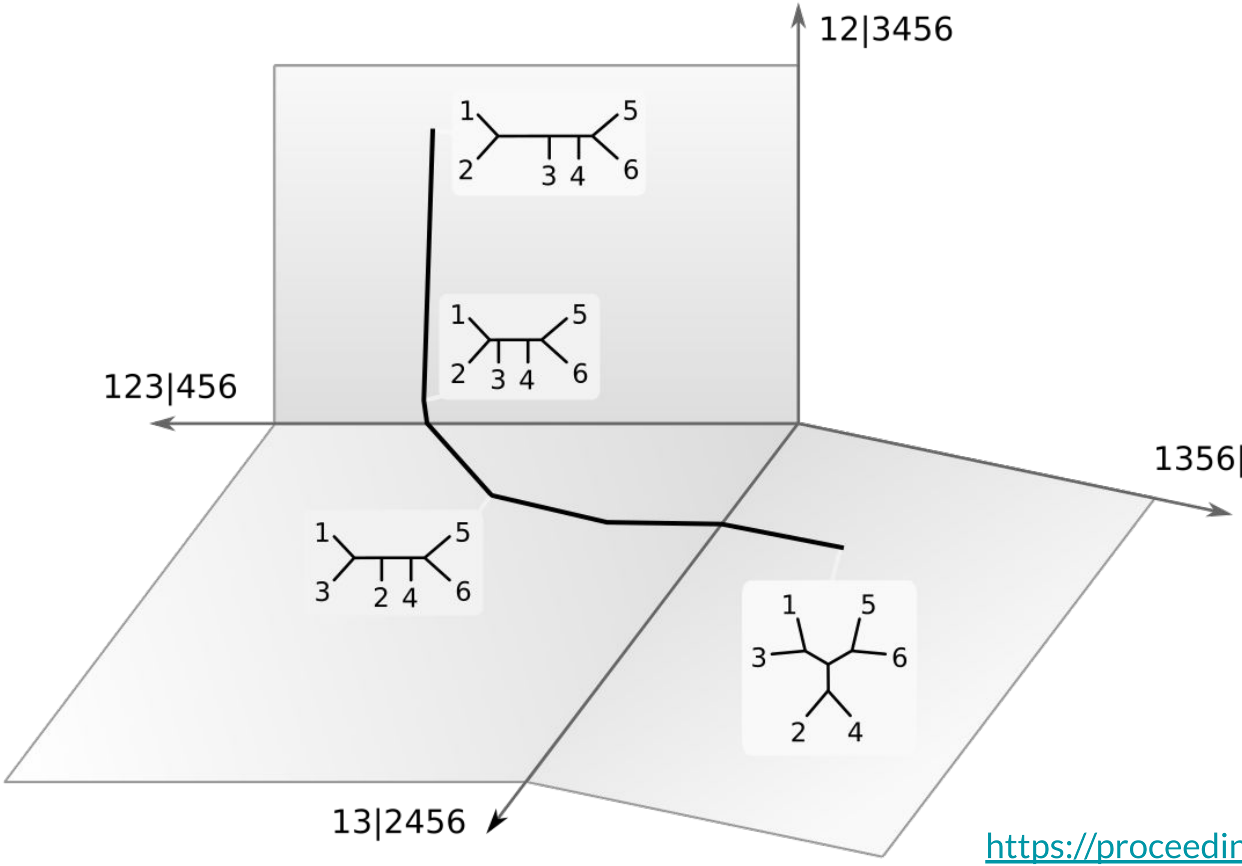
Scalable Bayesian divergence time estimation with ratio transformations — [PDF] oup.com
…, B Potter, P Lemey, G Baele, MA **Suchard** - Systematic …, 2023 - academic.oup.com
… We then apply **Hamiltonian Monte Carlo** sampling with the ratio transform in a Bayesian framework to learn the divergence times in 4 pathogenic viruses (West Nile virus, rabies virus, …
☆ Save 99 Cite Cited by 8 Related articles All 8 versions Import into BibTeX
✓ Paperpile PDF

Scalable bayesian phylogenetics — [PDF] royalsocietypublishing.org Full View
…, X Ji, G Baele, MA **Suchard**… - … of the Royal …, 2022 - royalsocietypublishing.org
Recent advances in Bayesian phylogenetics offer substantial computational savings to accommodate increased genomic sampling that challenges traditional inference methods. In this …
☆ Save 99 Cite Cited by 11 Related articles All 7 versions Import into BibTeX
✓ Paperpile PDF

Many-core algorithms for high-dimensional gradients on phylogenetic trees — [PDF] oup.com
…, P Lemey, FA Matsen IV, MA **Suchard** - …, 2024 - academic.oup.com
… This rapid growth spurs the need for efficient inference techniques, such as **Hamiltonian Monte Carlo** (HMC) in a Bayesian framework, to estimate parameters of these phylogenetic …
☆ Save 99 Cite Cited by 3 Related articles All 13 versions Import into BibTeX
✓ Paperpile PDF

[HTML] Bayesian mitigation of spatial coarsening for a Hawkes model applied to gunfire, wildfire and viral contagion — [HTML] nih.gov
AJ Holbrook, X Ji, MA **Suchard** - The annals of applied statistics, 2022 - ncbi.nlm.nih.gov
… parallel implementation of the model's log-likelihood gradient with respect to locations and thus avoid its quadratic computational cost in the context of **Hamiltonian Monte Carlo**. Our …
☆ Save 99 Cite Cited by 9 Related articles All 11 versions Import into BibTeX
\+ Paperpile

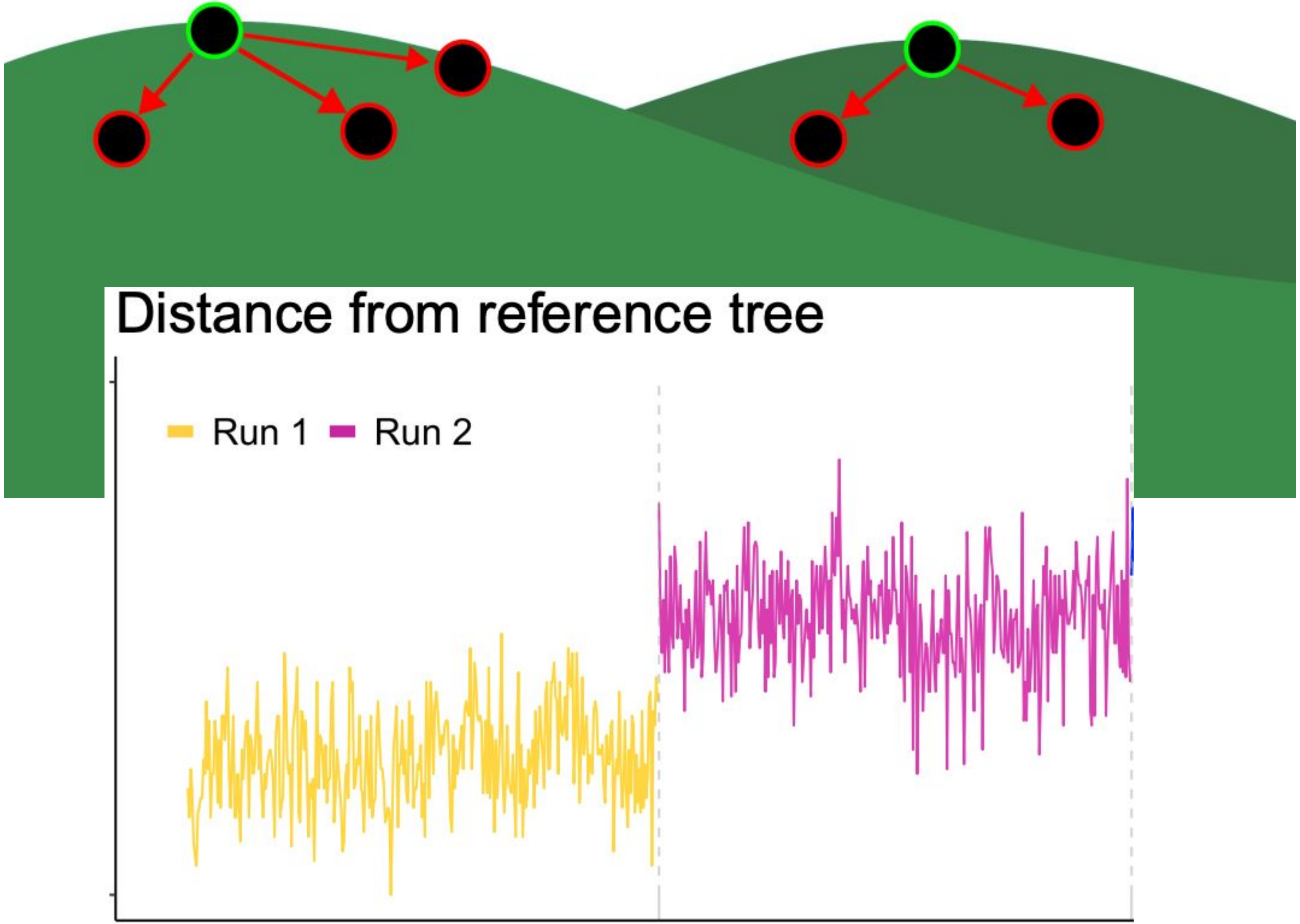
We kind-of got HMC to work across topologies



Vu Dinh

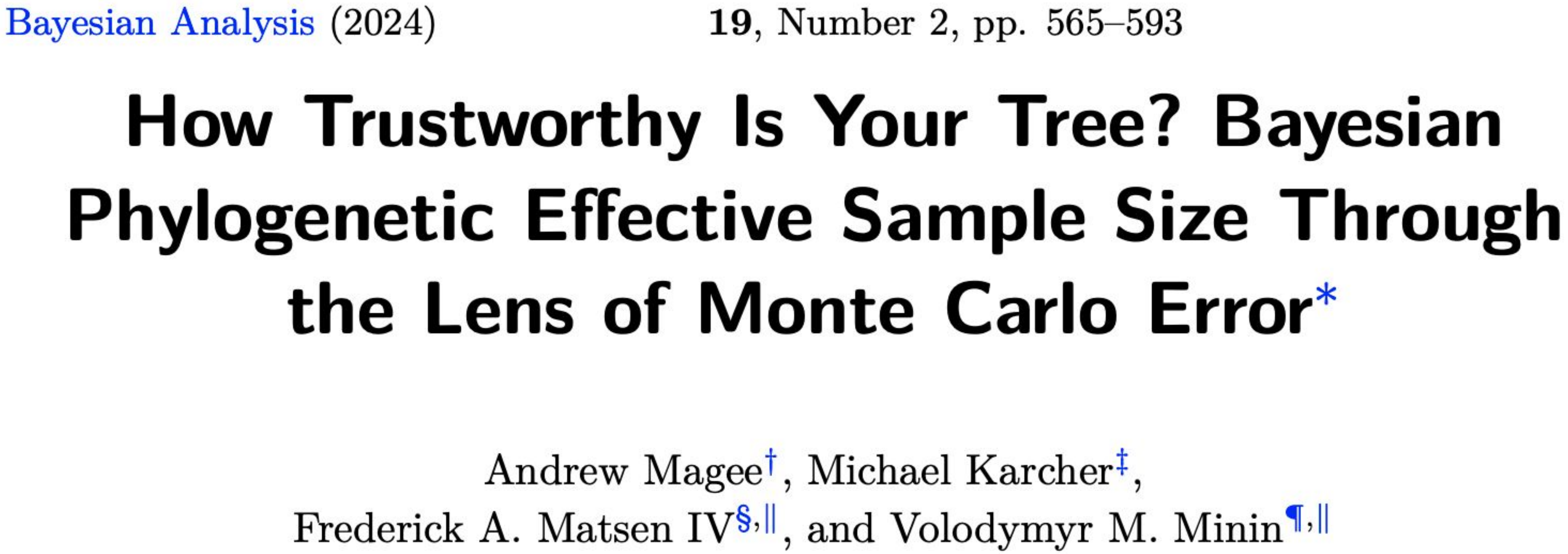
https://proceedings.mlr.press/v70/dinh17a/dinh17a.pdf

Rather than top-down trying to design a better proposal, let’s do *research* on how MCMC is exploring the space.



Distance from reference tree

What does it even mean to explore tree space well?



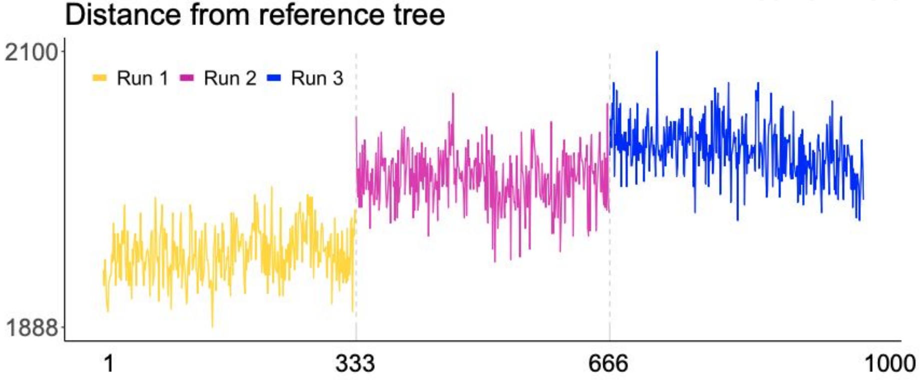
Bayesian Analysis (2024) **19**, Number 2, pp. 565–593

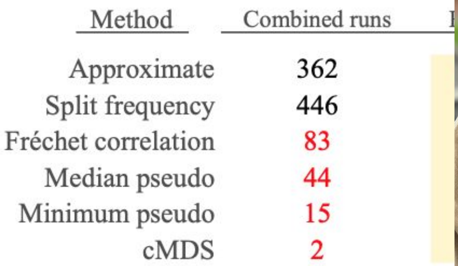
How Trustworthy Is Your Tree? Bayesian Phylogenetic Effective Sample Size Through the Lens of Monte Carlo Error[*]

Andrew Magee[†], Michael Karcher[‡],
Frederick A. Matsen IV[§,‖], and Volodymyr M. Minin[¶,‖]

Mixing over topologies can be a problem!

Metric: Robinson–Foulds

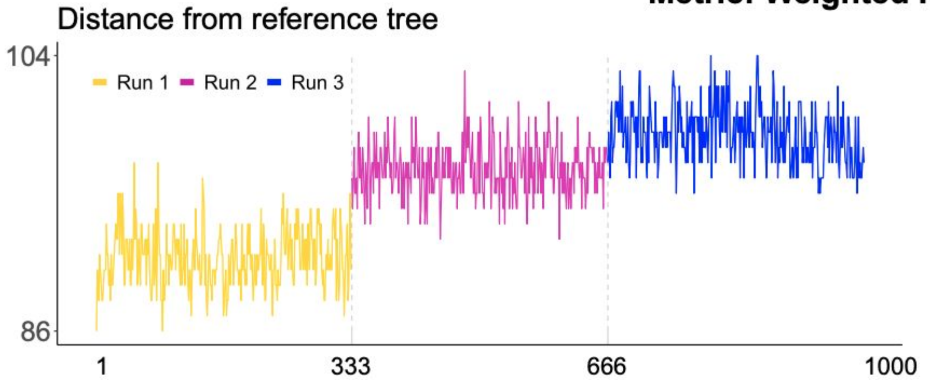


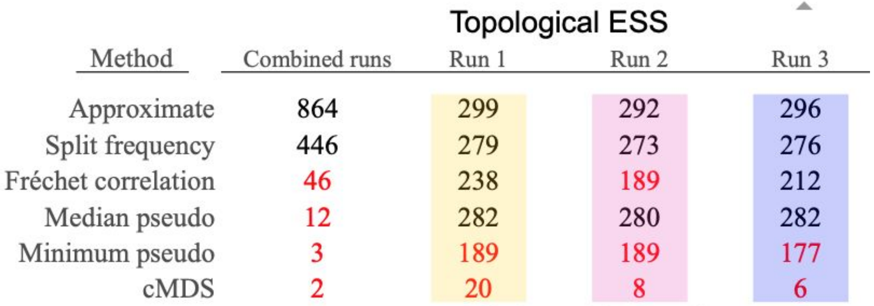


Method	Combined runs
Approximate	362
Split frequency	446
Fréchet correlation	83
Median pseudo	44
Minimum pseudo	15
cMDS	2

Marius Brusselmans

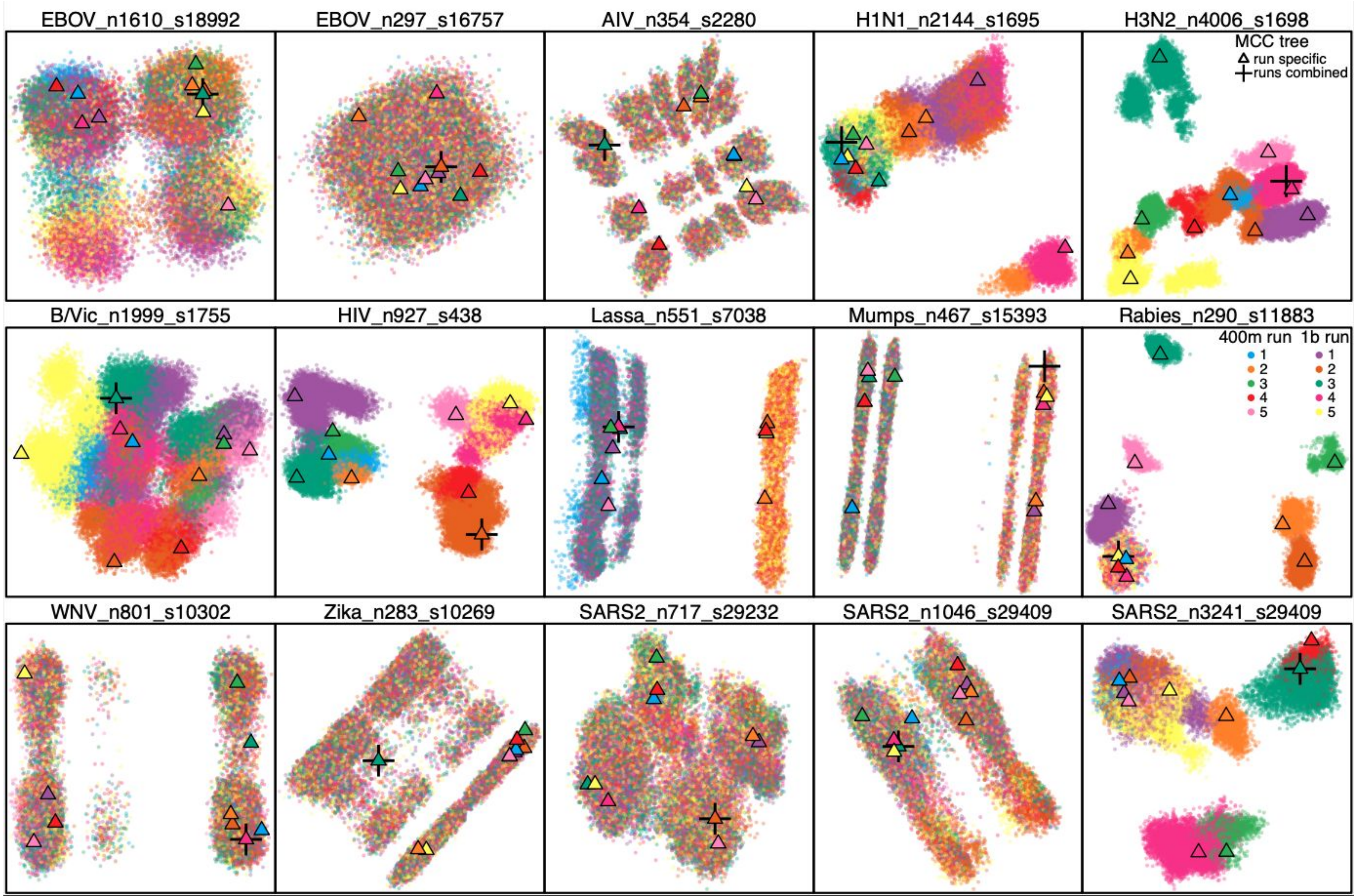
Metric: Weighted Robinson–Foulds



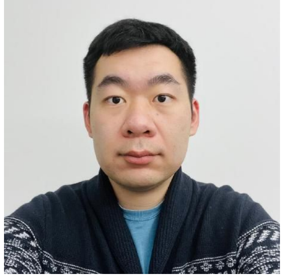


		Topological ESS		
Method	Combined runs	Run 1	Run 2	Run 3
Approximate	864	299	292	296
Split frequency	446	279	273	276
Fréchet correlation	46	238	189	212
Median pseudo	12	282	280	282
Minimum pseudo	3	189	189	177
cMDS	2	20	8	6

https://doi.org/10.1093/ve/veae081

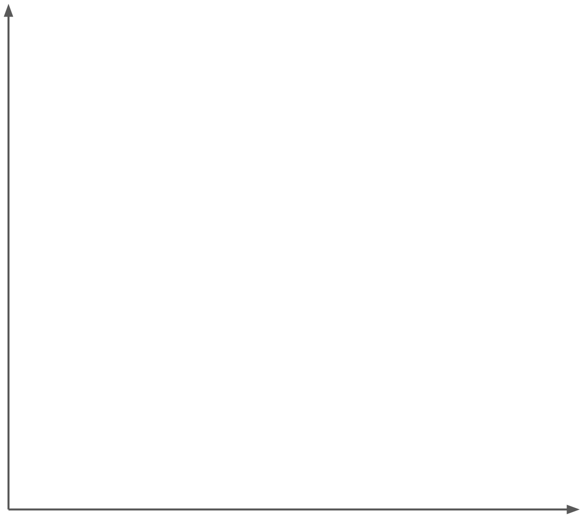


Topological lack of convergence is common in viral data sets.



Jiansi Gao

- What is the shape of the posterior?
- Can we model the posterior?
- Can we infer the posterior using a method that doesn’t require MCMC?
- **Does deep learning help?**



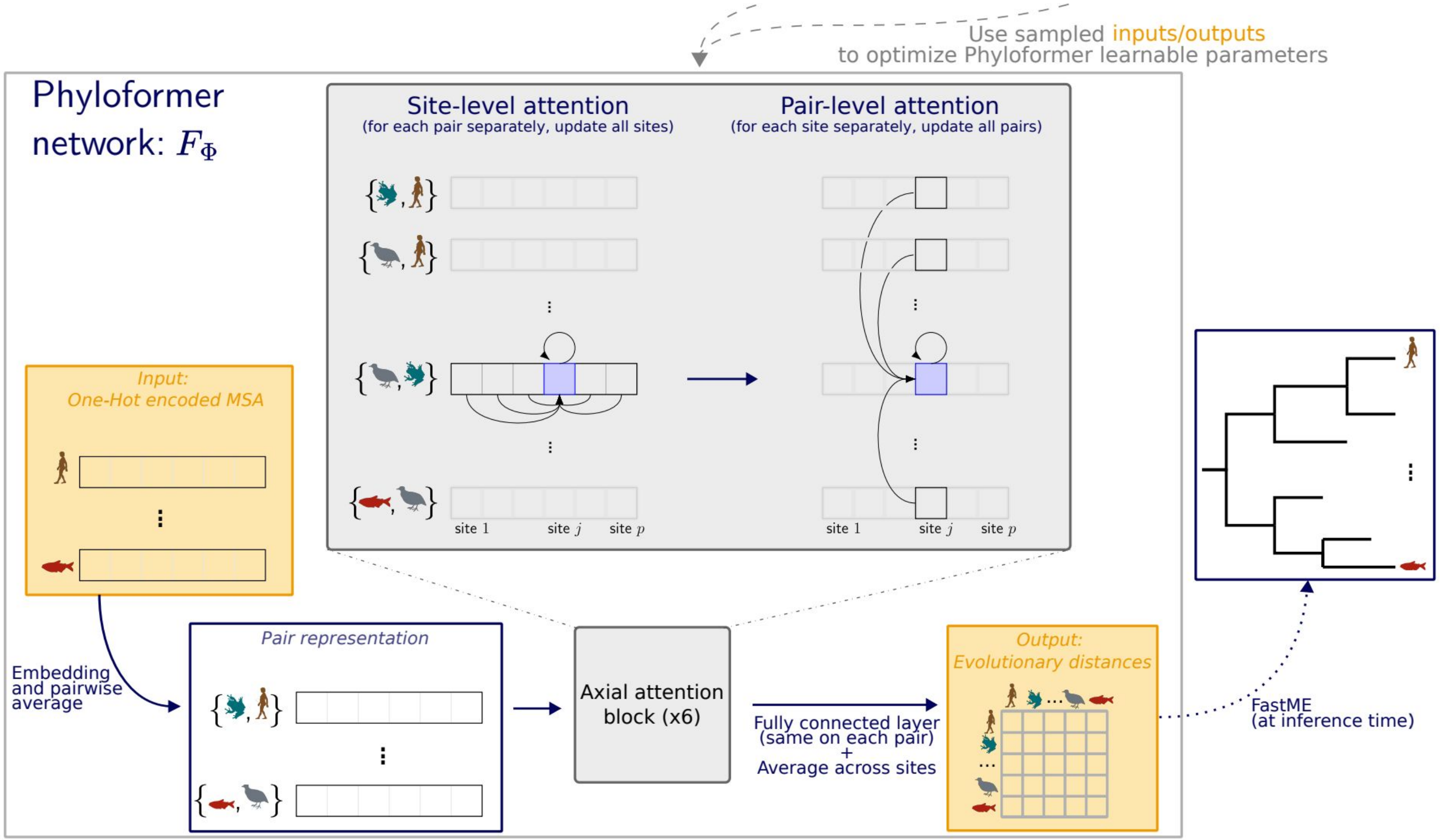
- Topological posterior
- Continuous parameters
- Joint models
- **Next-gen proposals**

The simple dream of DL for phylogenetic inference

One-hot sequence data

Fully formed tree

Phyloformer (Nesterenko et al, 2024)



IMHO, two ideas make phylogenetic inference possible.

1. Iterative tree improvement
2. Felsenstein / Fitch

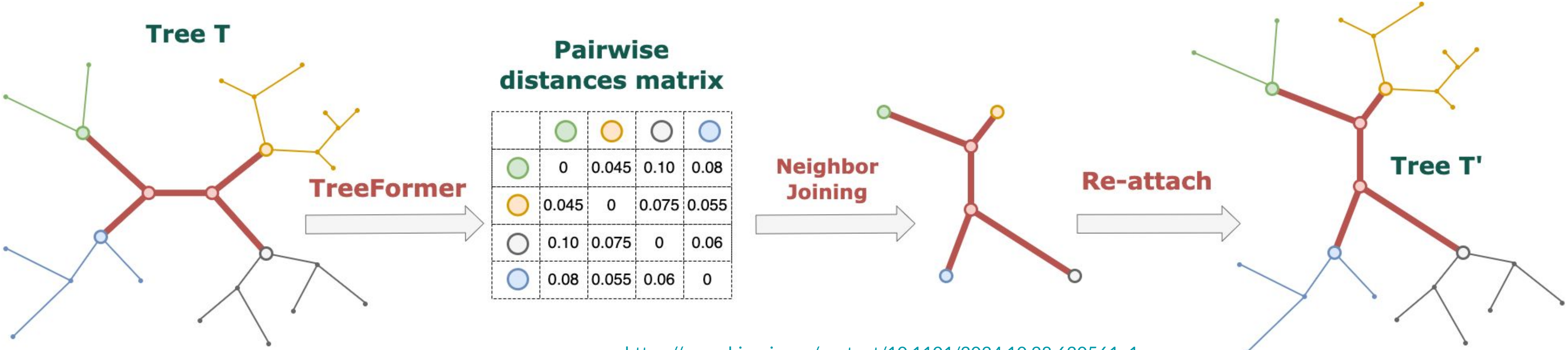
Let’s do those for DL!

A start:

We can use partial likelihood vectors instead of one-hot encoded vectors as input for a phylogenetic neural network.

Nhan Trong Ly

Minh Bui



	🟢	🟠	⚪	🔵
🟢	0	0.045	0.10	0.08
🟠	0.045	0	0.075	0.055
⚪	0.10	0.075	0	0.06
🔵	0.08	0.055	0.06	0

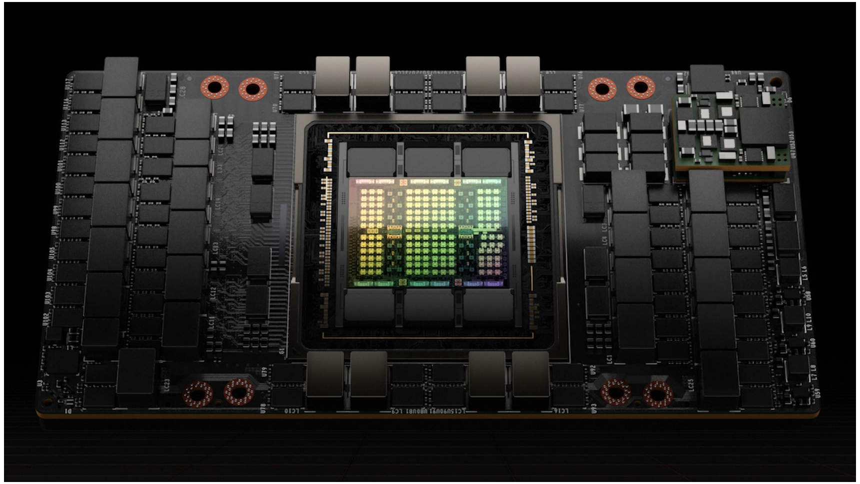
https://www.biorxiv.org/content/10.1101/2024.10.28.620561v1

What is

1994

2024

NVIDIA

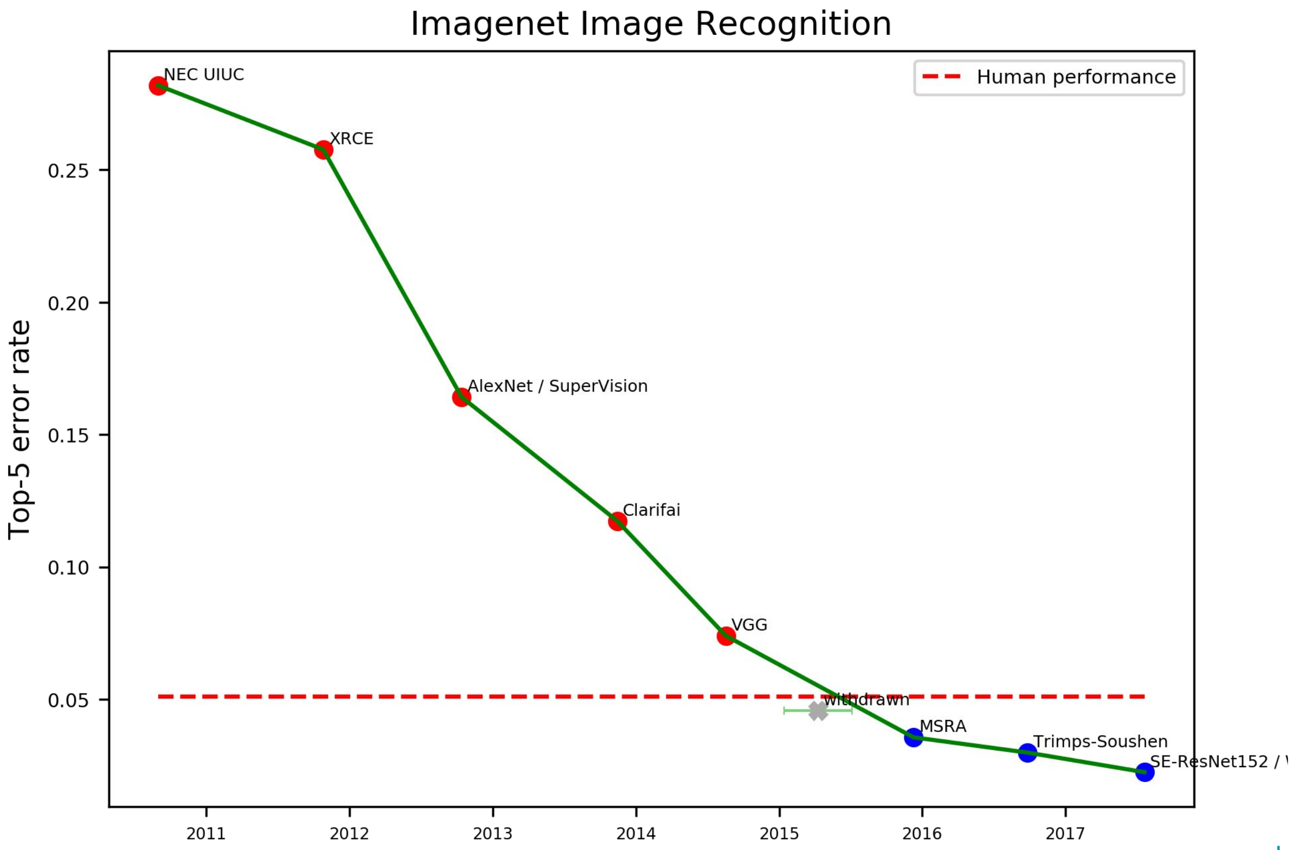


Bayesian Phylogenetics

MCMC

MCMC

What could be...



Predict:
1 *pencil box*
2 *diaper*
3 *bib*
4 *purse*
5 *running shoe*

Ground Truth:
sleeping bag

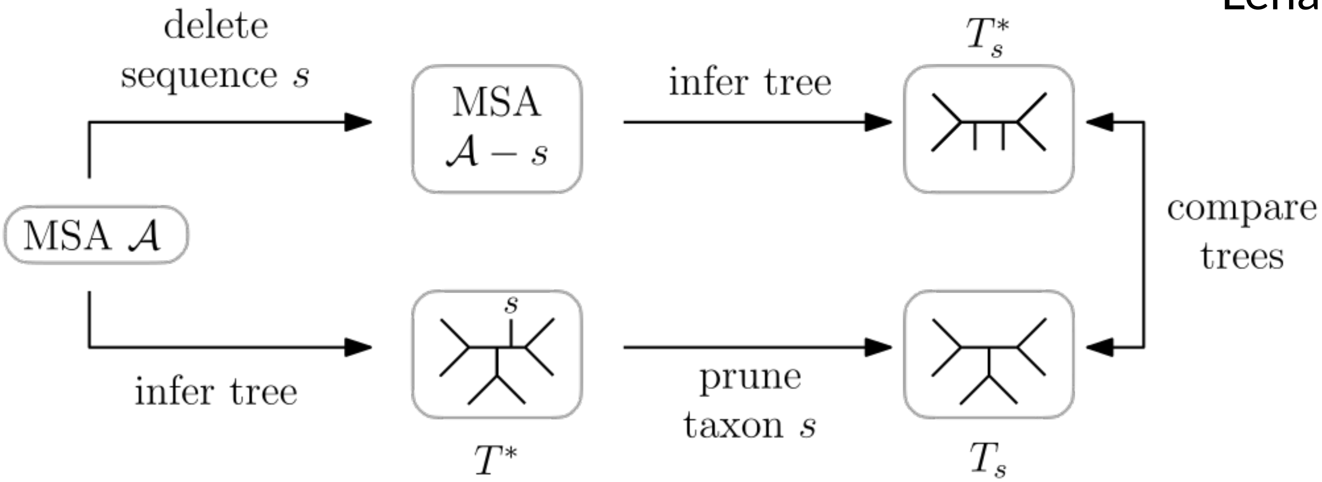
Label May Wrong (Maybe it is really a sleeping bag for Hello Kitty? lol)

https://www.eff.org/ai/metrics

Check out Lena's poster about online tree stability

"a missing paper" -- Alexis Stamatakis

Lena Collienne

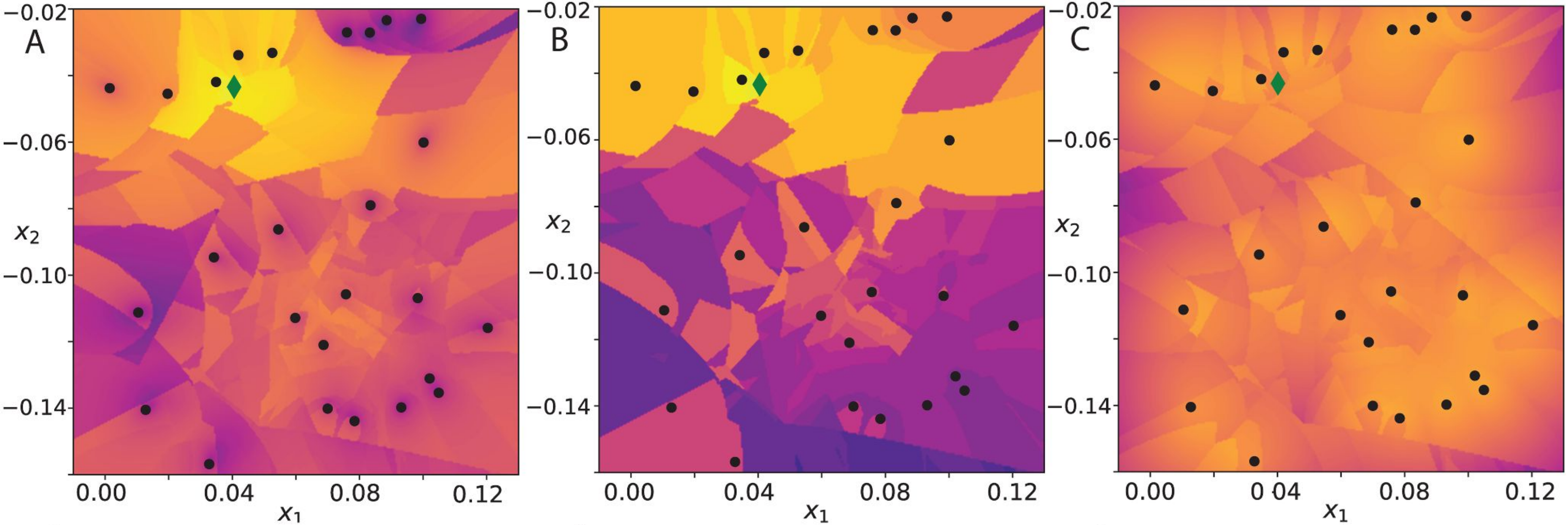


https://dx.doi.org/10.1093/sysbio/syae059

Thank you!

- Mathieu Fourment, Nhan Trong Ly, Minh Bui 🇦🇺
- Xiang Ji & Marc Suchard 🇺🇸
- Christiaan Swanepoel (Alexei Drummond’s lab) 🇳🇿
- Cheng Zhang 🇨🇳
- My group 🇺🇸 : Seong-Hwan Jun, Michael Karcher, Andy Magee, Hassan Nasif, Chris Jennings-Shafer, Dave Rich, Lena Colliene, Mary Barker, Will Dumm
- Chris Whidden 🇨🇦

NIH and HHMI



moving one node (green diamond) through embedding space. a) joint probability p(T, Y), b) symmetric difference from best tree topology and c,) total length.

https://doi.org/10.1371/journal.pcbi.1011084