

Deep learning for phylogenetic inference

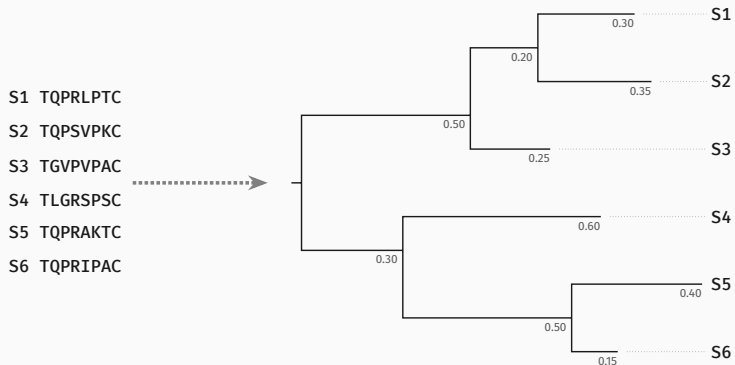


Luc Blassel, Nicolas Lartillot, Bastien Boussau, Laurent Jacob

AI4omics - Oct. 2nd 2026

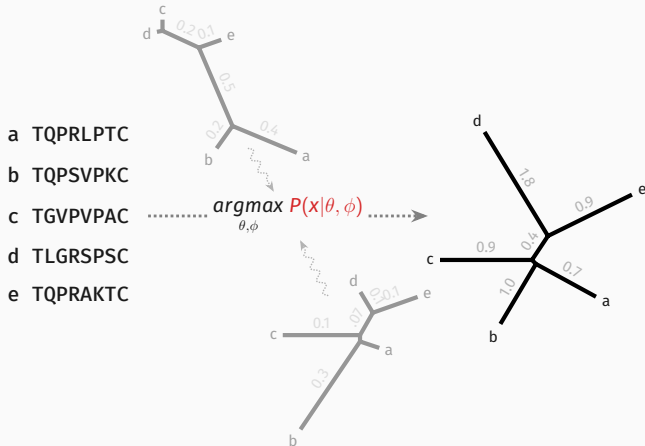


Context - What is phylogenetic inference



Goal: describe **evolutionary-history** of MSA

Context - **likelihood**-based phylogenetics (Max. Likelihood)



x : MSA, $\theta = (\tau, \ell)$: Phylogenetic tree, ϕ : Evolution model

Context - **likelihood**-based phylogenetics (Bayesian)

a TQPRLPTC

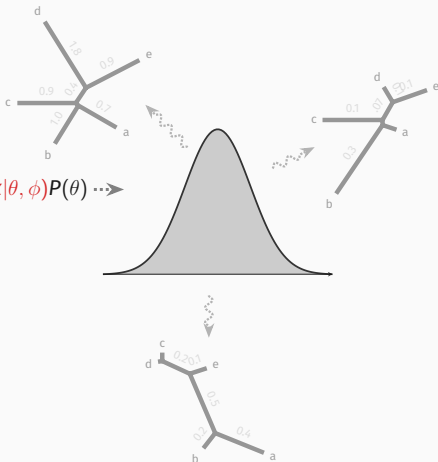
b TQPSVPKC

c TGVPVPAC

d TLGRSPSC

e TQPRAKTC

$$\dots\dots P(\theta|x, \phi) \propto P(x|\theta, \phi)P(\theta) \dots\dots \rightarrow$$



x : MSA, $\theta = (\tau, \ell)$: Phylogenetic tree, ϕ : Evolution model

Context - Dependence on likelihoods

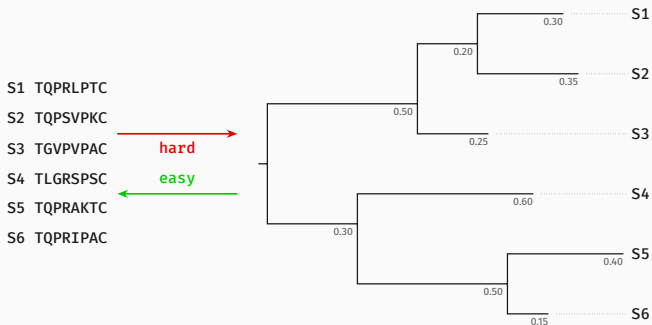
model-based inference allows for **principled assumptions**.

It is **good, but:**

- As models **grow in complexity** so does $P(x|\theta)$
- **Computing** $P(x|\theta)$ can be **slow**
- In some cases there is **no analytical form** of $P(x|\theta)$
- If we're **lucky** there is a workable **approximation**

The end **result** is that using the **likelihood** in our inference process **limits** the types of **models** we can use

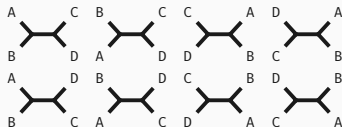
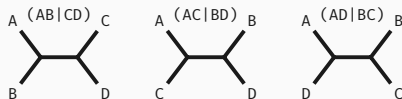
Context - SBI



- We can simulate many (tree, MSA) pairs
- **Learn** the mapping back **from MSA to tree**
- “Can we **infer**?” becomes “Can we **simulate**?”

Existing Deep-learning methods

Context - Quartet classifying



adapted from Tang et al. 2024

- Unrooted **quartet** has **3 unique** topologies
- Train a model to **classify** 4-taxa into a **topology**
- Reconstruct **tree from quartets**

Scaling problem:

$$\text{Quartets}(n) = \binom{n}{4} = \mathcal{O}(n^4)$$

Suvorov et al. 2019; Zou et al. 2020; Tang et al. 2024; Zaharias et al. 2022

Context - Machine Learning in phylogenetic workflows

Replace steps in established methods.

Azouri et al.^{1,2}: Guide tree-search: predict rearrangements¹, learn a search policy with RL².

Ada. RAxML³: Predict dataset difficulty to adapt effort ³

EBG⁴: Predict branch supports from tree and MSA⁴

DEPP/APPLES^{5,6}: Place gene sequence in reference species tree^{5,6}

Problem?

These methods **don't** produce a **tree**.

¹Azouri, Abadi, et al. 2021,²Azouri, Granit, et al. 2024,³Togkousidis et al. 2023,

⁴Wiegert et al. 2024,⁵Balaban et al. 2020,⁶Jiang et al. 2022,

Context - Learning posterior samplers

VBPI-GNN¹: VI over tree topologies and branch lengths

ARTree(Former)^{2,3}: generates topology through repeated leaf insertion

PhyloGFN⁴: Learns a tree sampler through subtree merges

GeoPhy⁵: learns, continuous embeddings $\Leftrightarrow$ trees

Problem?

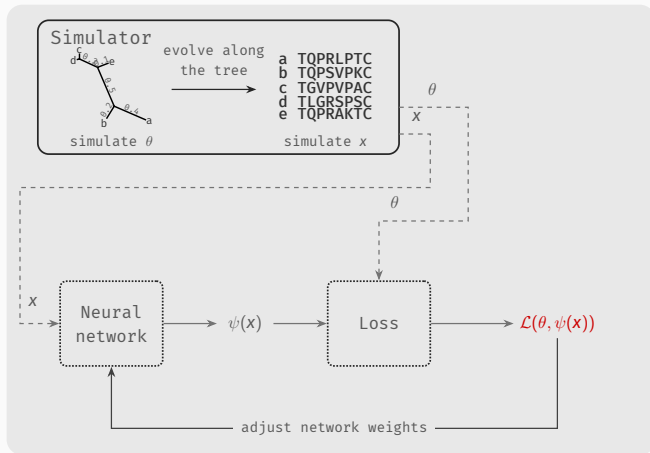
1. These methods **evaluate likelihood** during training.
2. They are also often **not amortized** accross MSAs.

¹Zhang and IV 2019,²Xie and Zhang 2023; Xie, Mao, et al. 2025,³Zhang and IV 2019,

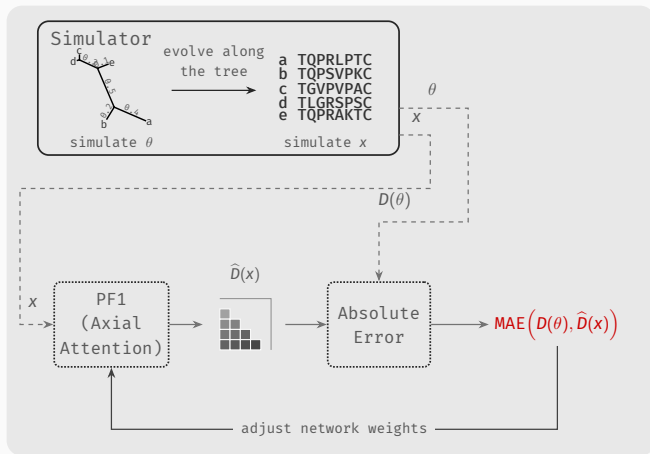
⁴M. Y. Zhou et al. 2024,⁵Mimori and Hamada 2023,

Our approach to deep phylogenetic inference

Method - learning setting



Method - Using distance as a proxy



Inference: use FastME¹ on PF1² matrix.

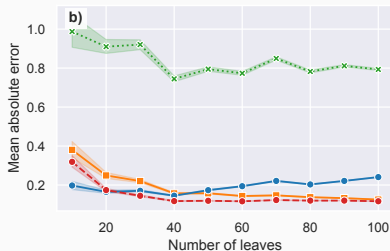
¹Lefort et al. 2015, ²Nesterenko et al. 2025

Paper →

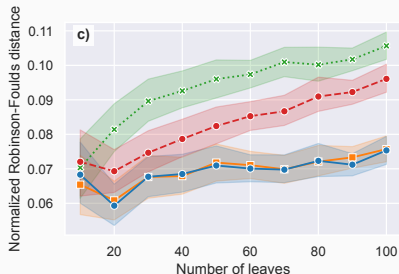


Method - Phyloformer 1 results

✓ Distances: good



✗ Topology: less good



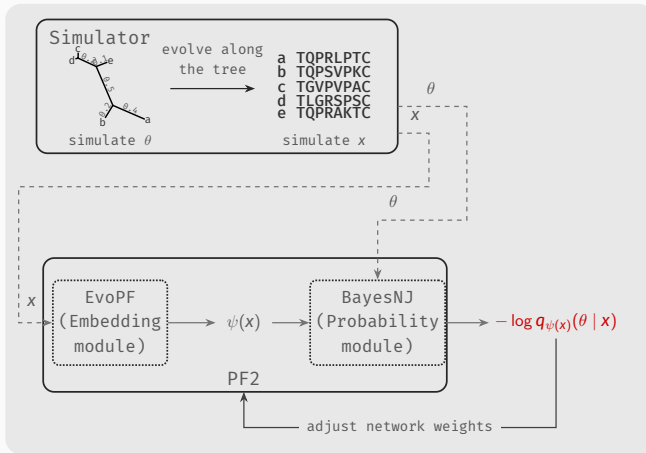
We need to do **end-to-end** phylogenetic **inference**

Method - Neural Posterior Estimation (NPE)

- Given a **probabilistic model** $p(x|\theta)$ with some prior $p(\theta)$
- We want to **estimate the posterior**: $p(\theta|x)$
- We build $q_{\psi}(\theta|x)$ a **family** of distributions **parametrized** by ψ (our NN)
- We find $q_{\psi^*} = \underset{\psi}{\operatorname{argmin}} \mathbb{E}_{p(x)}[KL(p(\theta|x)||q_{\psi}(\theta|x))]$
- In practice we **maximize** $\mathbb{E}_{p(x,\theta)}[\log q_{\psi(x)}(\theta|x)]$ by **sampling** from $p(x, \theta)$
- The two are **formally equivalent**¹

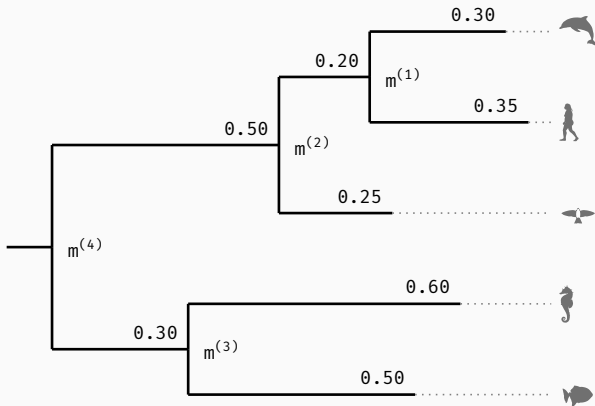
¹Radev et al. 2020 x : MSA, $\theta = (\tau, \ell)$: Phylogenetic tree, $\psi(x)$: NN applied to x

Method - How does this look like in PF2 ?



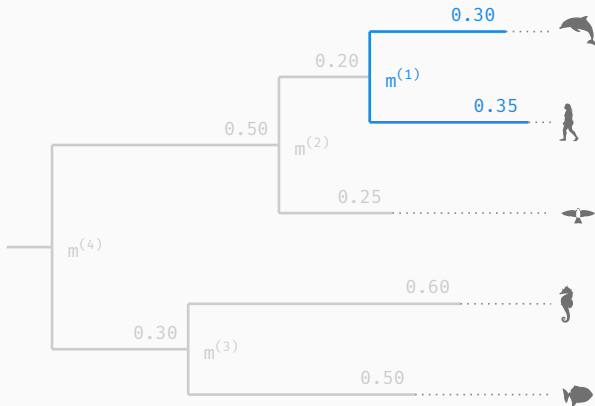
How to design BayesNJ so we can also **sample** trees?

Method - How can we design a distribution family



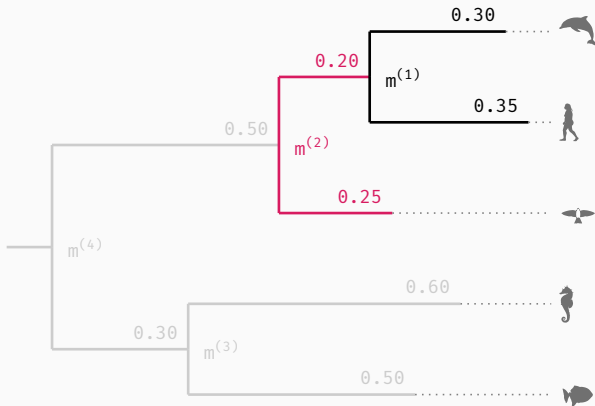
$$q_{PF2}(\theta|MSA) = ?$$

Method - How can we design a distribution family



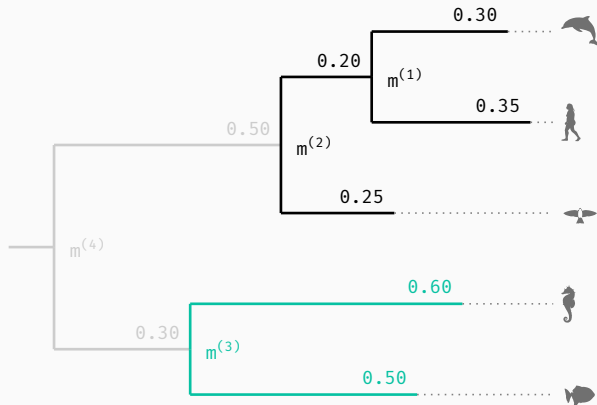
$$q_{PF2}(\theta|MSA) = q_{PF2}(m^{(1)})$$

Method - How can we design a distribution family



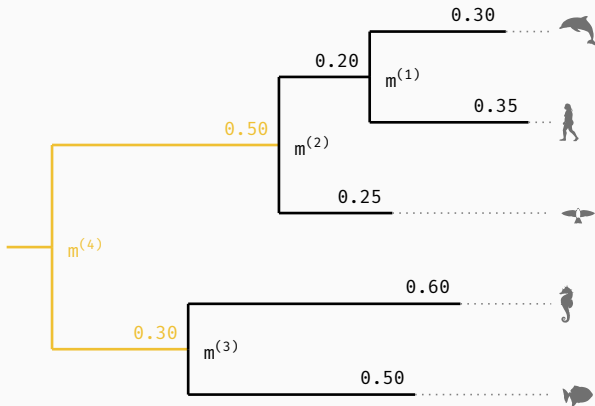
$$q_{PF2}(\theta|MSA) = q_{PF2}(m^{(1)}) \times q_{PF2}(m^{(2)}|m^{(1)})$$

Method - How can we design a distribution family



$$q_{PF2}(\theta|MSA) = q_{PF2}(m^{(1)}) \times q_{PF2}(m^{(2)}|m^{(1)}) \times q_{PF2}(m^{(3)}|m^{(\leq 2)})$$

Method - How can we design a distribution family



$$q_{PF2}(\theta|MSA) = q_{PF2}(m^{(1)}) \times q_{PF2}(m^{(2)}|m^{(1)}) \times q_{PF2}(m^{(3)}|m^{(\leq 2)}) \times q_{PF2}(m^{(4)}|m^{(\leq 3)})$$

Method - How do we sample from this?



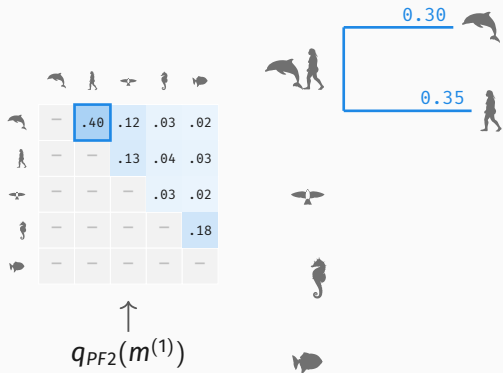
1. Start with **un-merged** taxa

Method - How do we sample from this?



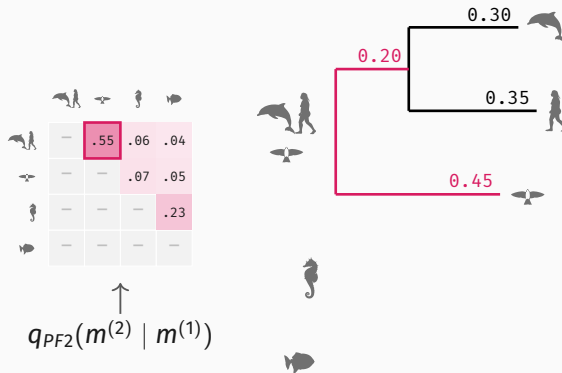
2. **Sample** a pair to **merge**

Method - How do we sample from this?



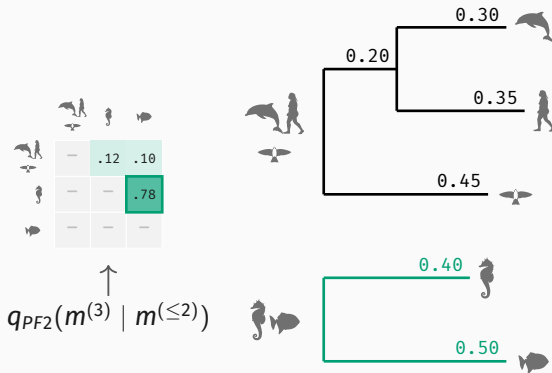
3. Sample its branch lengths

Method - How do we sample from this?



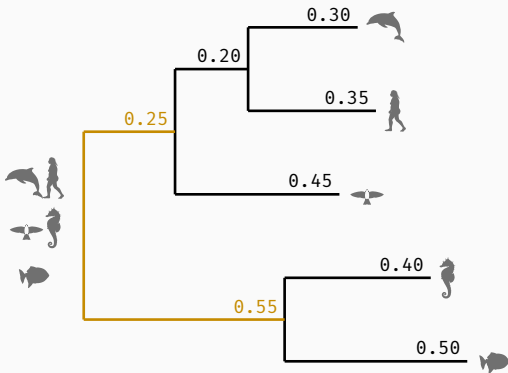
4. Sample next **merge & lengths**

Method - How do we sample from this?



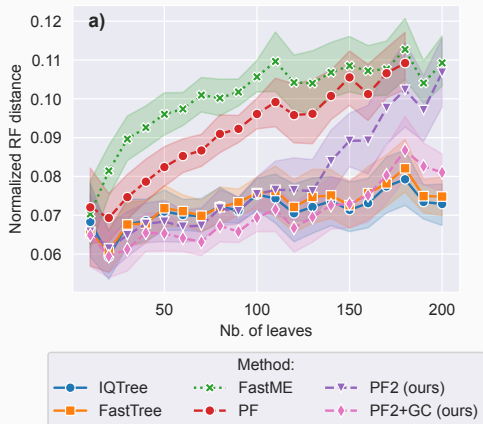
5. Sample next **merge & lengths**

Method - How do we sample from this?



6. Sample final branch lengths

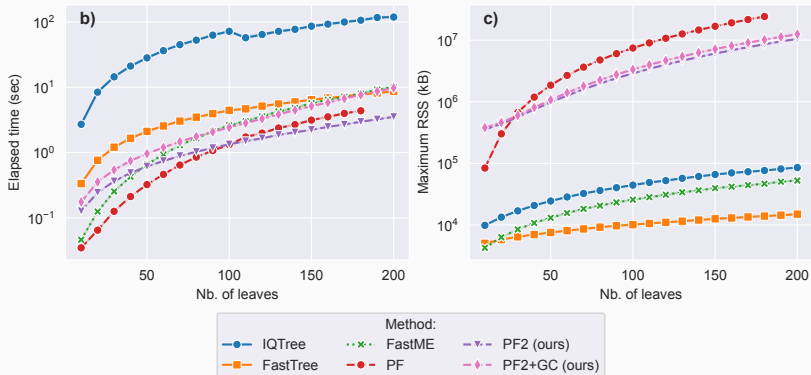
Results - Simple likelihood results (LG+GC)



LG: Le and Gascuel 2008

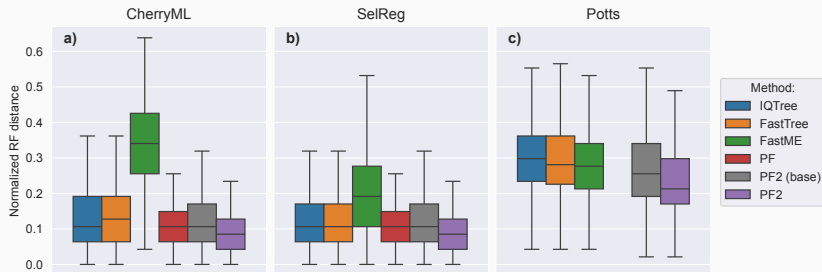
- RF = **topological distance** between trees
- Much **better** performance **than PF1**
- Greedy-MAP trees **match ML** methods
- **Posterior summary** trees **outperform** IQTree.

Results - Computational details



- **Very fast** inference because of **GPUs**¹.
- Comes at a **memory cost**.

Results - Intractable likelihood results



CherryML¹ models pairs of
co-evolving sites

SelReg² Adds **selection**
modes

Potts³ Fit on PF00072, models
structural constraints.

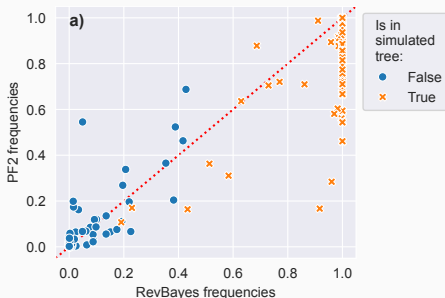
N. Sauvage & P.

Barrat-Charlaix

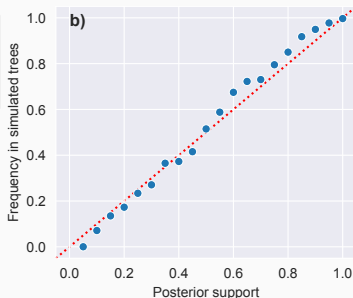
¹Prillo et al. 2023, ²Duchemin et al. 2023, ³Pagnani and Barrat-Charlaix 2025

Results - How are the estimated posteriors?

Comparison to MCMC

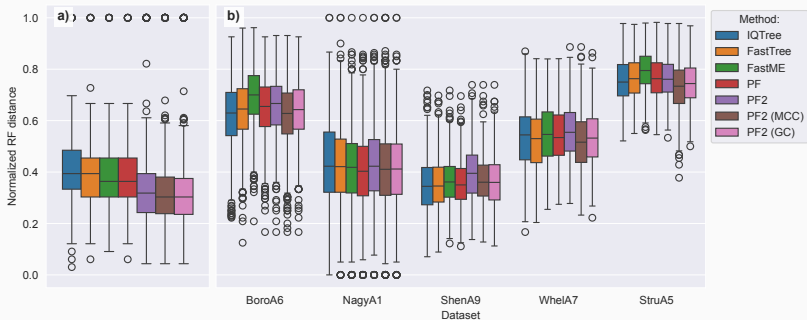


Simulation-based Calibration



RevBayes: Höhna et al. 2016

Results - Empirical results



Compare **Gene-trees** to **species-tree** build on concatenate MSA using IQTree.

a) Szöllősi et al. 2013, b) X. Zhou et al. 2018

Take-home message

- **Deep learning** can be a **powerful** tool for phylogenetic **inference**.
- **SBI** enables inference under **arbitrarily complex models**
- PF2 is a **likelihood-free**, phylogenetic **posterior** estimator, **amortized** accross MSAs.
- PF2 combines **accurate** posterior estimation with **fast inference**.

Preprint →



- **Optimize** execution **speed** and **memory**-usage under **current architecture**.
- **Explore** branch-length **parametrizations**.
- **Explore** more fundamental **architectural changes** for better scaling.
- **Train** versions of model under **more complex** and **useful** evolutionary **models**.

Thank you all!



P. Barrat-Charlaix



B. Boussau



V. Garot



L. Jacob



N. Lartillot



A. Leroy



**L.
Nesterenko**



N. Sauvage



P. Veber



Special thanks to Jean-Zay for all the GPUs!

References

- Azouri, D., S. Abadi, et al. (2021). ***Harnessing machine learning to guide phylogenetic-tree search algorithms.*** In: *Nature communications* 12.1, p. 1983.
- Azouri, D., O. Granit, et al. (2024). ***The tree reconstruction game: phylogenetic reconstruction using reinforcement learning.*** In: *Molecular Biology and Evolution* 41.6.
- Balaban, M. et al. (2020). ***APPLES: Scalable Distance-Based Phylogenetic Placement with or without Alignments.*** In: *Systematic Biology* 69.3, pp. 566–578.
- Duchemin, L. et al. (2023). ***Evaluation of methods to detect shifts in directional selection at the genome scale.*** In: *Molecular Biology and Evolution* 40.2, msac247.
- Höhna, S. et al. (2016). ***RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language.*** In: *Systematic Biology* 65.4, pp. 726–736.
- Jiang, Y. et al. (2022). ***DEPP: Deep Learning Enables Extending Species Trees using Single Genes.*** In: *Systematic Biology*, syac031.
- Le, S. Q. and O. Gascuel (2008). ***An Improved General Amino Acid Replacement Matrix.*** In: *Molecular Biology and Evolution* 25.7, pp. 1307–1320.

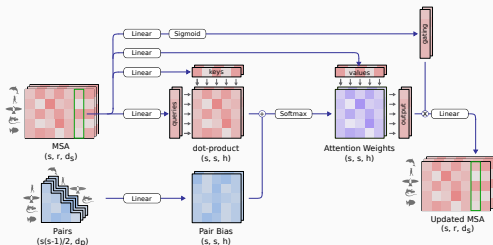
- Lefort, V. et al. (2015). **FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program.** In: *Molecular biology and evolution* 32.10, pp. 2798–2800.
- Mimori, T. and M. Hamada (2023). **GeoPhy: Differentiable Phylogenetic Inference via Geometric Gradients of Tree Topologies.** In: *Advances in Neural Information Processing Systems*. Vol. 36. Curran Associates, Inc., pp. 36591–36619.
- Nesterenko, L. et al. (2025). **Phyloformer: Fast, Accurate, and Versatile Phylogenetic Reconstruction with Deep Neural Networks.** In: *Molecular Biology and Evolution* 42.4, msaf051.
- Pagnani, A. and P. Barrat-Charlaix (2025). **Generative continuous time model reveals epistatic signatures in protein evolution.**
- Prillo, S. et al. (2023). **CherryML: scalable maximum likelihood estimation of phylogenetic models.** In: *Nature Methods* 20.8, pp. 1232–1236.
- Radev, S. T. et al. (2020). **BayesFlow: Learning complex stochastic models with invertible neural networks.** In: *IEEE transactions on neural networks and learning systems* 33.4, pp. 1452–1466.
- Suvorov, A. et al. (2019). **Accurate Inference of Tree Topologies from Multiple Sequence Alignments Using Deep Learning.** In: *Systematic Biology* 69.2, pp. 221–233.

- Szöllősi, G. J. et al. (2013). **Data from: Efficient exploration of the space of reconciled gene trees.** Version 1. Artwork Size: 45688314 bytes Pages: 45688314 bytes.
- Tang, X. et al. (2024). **Novel symmetry-preserving neural network model for phylogenetic inference.** In: *Bioinformatics Advances* 4.1, vbae022.
- Togkousidis, A. et al. (2023). **Adaptive RAXML-NG: Accelerating Phylogenetic Inference under Maximum Likelihood Using Dataset Difficulty.** In: *Molecular Biology and Evolution* 40.10, msad227.
- Wiegert, J. et al. (2024). **Predicting Phylogenetic Bootstrap Values via Machine Learning.** In: *Molecular Biology and Evolution* 41.10, msae215.
- Xie, T., Y. Mao, et al. (2025). **ARTreeFormer: A Faster Attention-based Autoregressive Model for Phylogenetic Inference.**
- Xie, T. and C. Zhang (2023). **ARTree: A Deep Autoregressive Model for Phylogenetic Inference.** In: *Thirty-seventh Conference on Neural Information Processing Systems*.
- Zaharias, P. et al. (2022). **Re-evaluating Deep Neural Networks for Phylogeny Estimation: The issue of taxon sampling.** In: *Journal of Computational Biology*.
- Zhang, C. and F. A. M. IV (2019). **Variational Bayesian Phylogenetic Inference.** In: *International Conference on Learning Representations*.

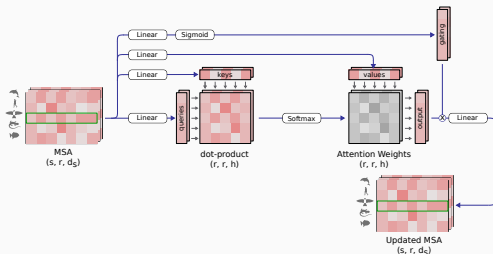
- Zhou, M. Y. et al. (2024). **PhyloGFN: Phylogenetic inference with generative flow networks**. In: *The Twelfth International Conference on Learning Representations*.
- Zhou, X. et al. (2018). **Evaluating Fast Maximum Likelihood-Based Phylogenetic Programs Using Empirical Phylogenomic Data Sets**. In: *Molecular Biology and Evolution* 35.2, pp. 486–503.
- Zou, Z. et al. (2020). **Deep residual neural networks resolve quartet molecular phylogenies**. In: *Molecular biology and evolution* 37.5, pp. 1495–1507.

Supplementary slides

Sup. Methods - EvoPF, the MSA stack

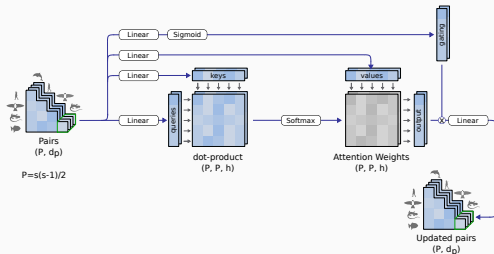


**Column-wise attention
with pair-bias**

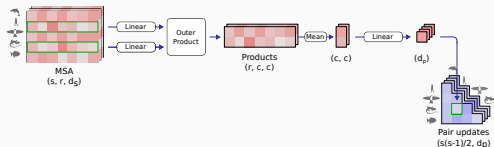


Row-wise attention

Sup. Methods - EvoPF, the pair stack



Pair attention



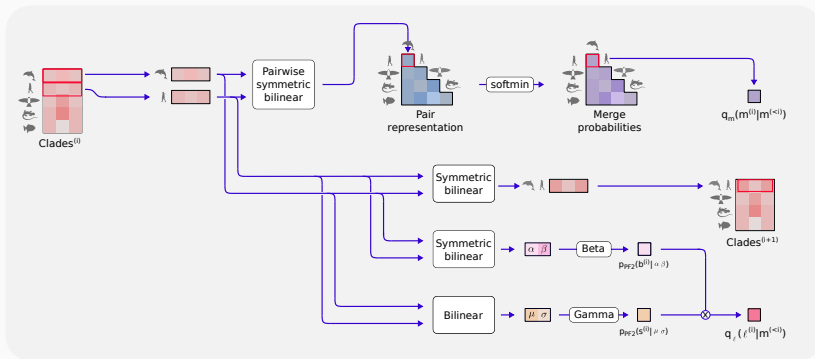
Outer product mean

Sup. Methods - The BayesNJ Module

$$m^{(i)} = \{ (l_1, l_2), (l_1, l_3) \}$$

$$s^{(i)} = l_1 + l_3$$

$$b^{(i)} = s / l_2$$



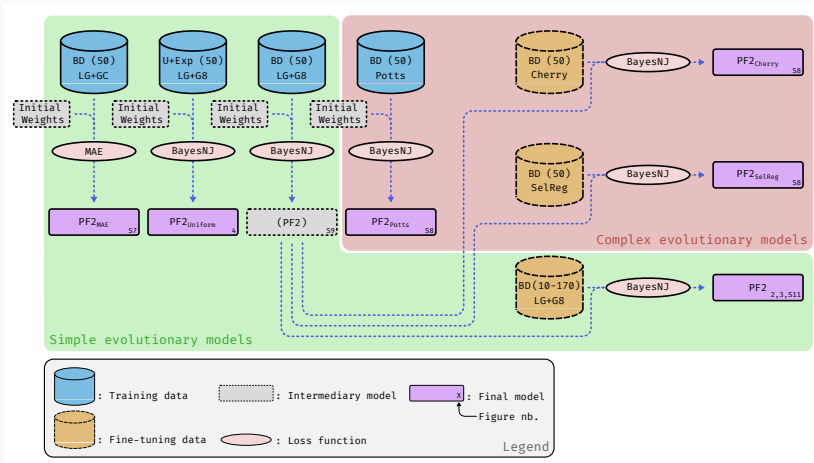
Sup. Methods - Ensuring the merge order is unique

Ensuring a **unique order** on merges ensures that we **define a distribution**. It also keeps **training** and **sampling** comparable ¹

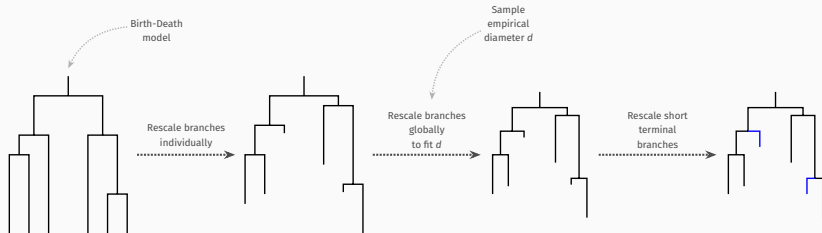
- On a given tree τ always **merge** the **shortest** available **cherry**
- When **sampling**, add **constraints**:
 1. Start with a $N \times N$ constraints matrix $M_{ij} = 0$
 2. At iteration k sample merge $m^{(k)} = (i, j)$ and cherry length $s^{(k)} = M_{ij} + X$
 3. **Update constraints** for cherries **available** when sampling $m^{(k)}$: $M'_{ij} = \max(M_{ij}, s^{(k)})$ $M'_{ui} = 0$
- During evaluation compute $p_{PF2}(s^{(k)} - M_{ij} | m^{(\leq k)})$

¹ Which is not the same if we use the NJ merge order

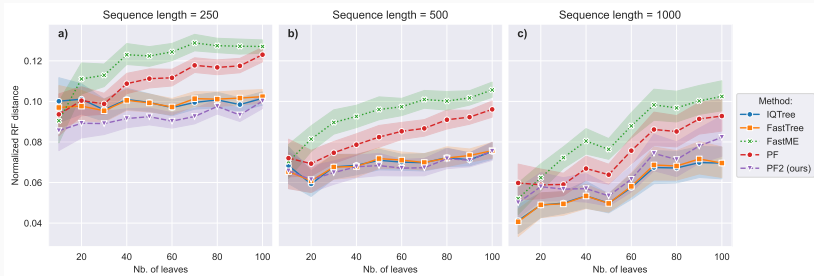
Sup. Methods - Training runs



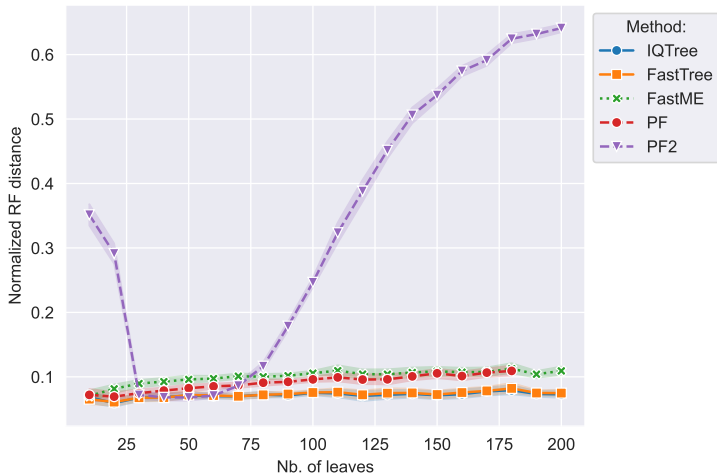
Sup. Methods - Tree simulation procedure



Sup. Results - Sensitivity to MSA length



Sup. Results - Why do we fine-tune on nb. of tips?



Sup. Results - Memory scaling

