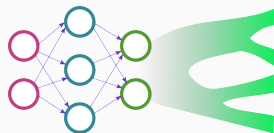


Phyloformer 2: Likelihood-free inference of phylogenetic tree posterior distributions

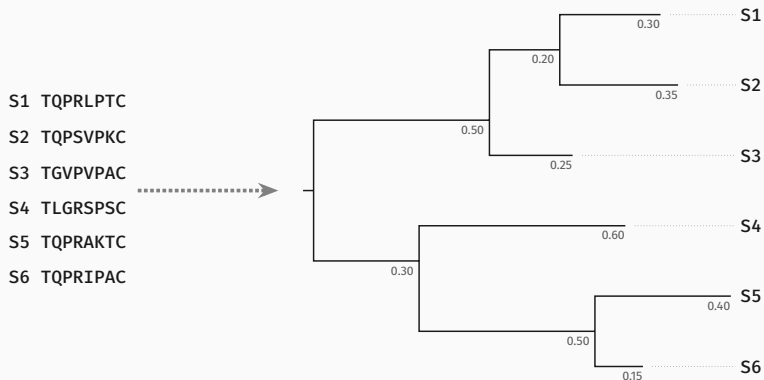


L. Blassel, N. Sauvage, P. Barrat-Charlaix, B. Boussau, N. Lartillot & L. Jacob

SMBE 2026 - June 27, 2026

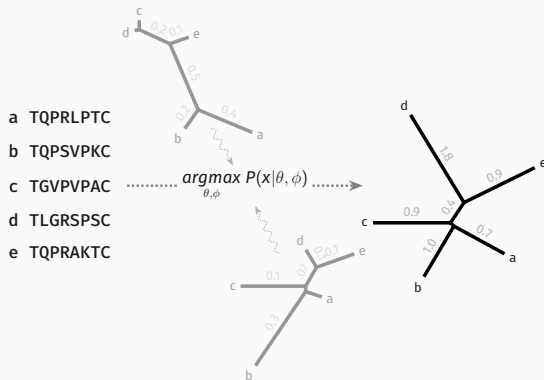


Context - Phylogenetic inference



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

Context - Likelihood-based tree reconstruction



Pros

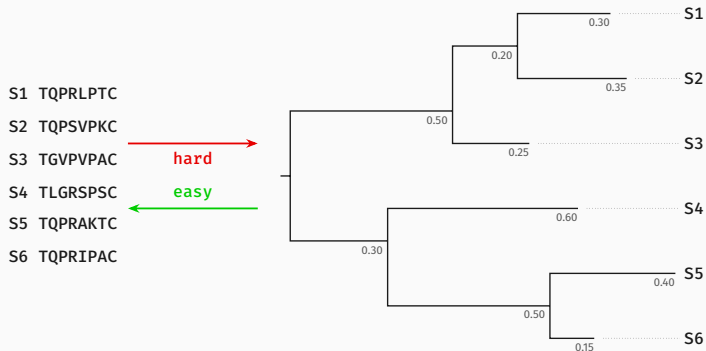
- Very **accurate**
- Principled **statistical model**

Cons

- **expensive** and **slow**
- $P(x|\theta, \phi)$ must be **computable**

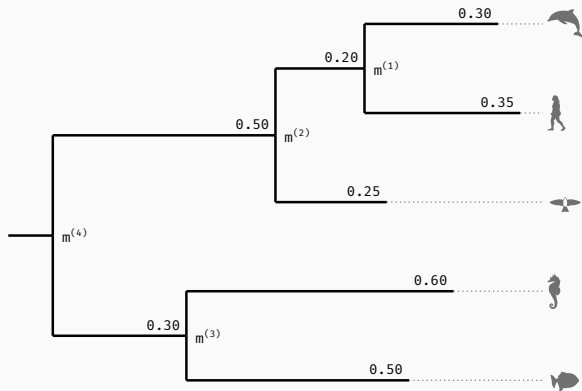
x : MSA, $\theta = (\tau, \ell)$: Phylogenetic tree, ϕ : Evolution model Felsenstein 1993; Kleinman et al. 2010

Context - Simulation-based/Likelihood-free inference



- We can simulate many (tree, MSA) pairs
- Can we **learn** the mapping **from MSA to tree**?

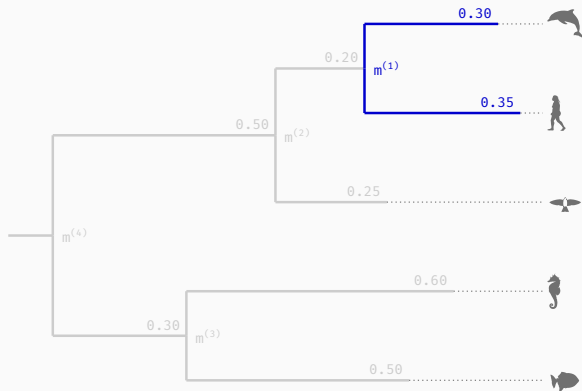
Method - Estimating tree posterior probability



- **Tree probability is product of conditional merge probabilities**

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

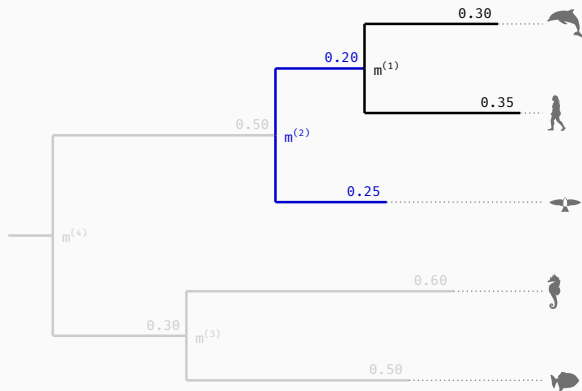
Method - Estimating tree posterior probability



- **Tree probability is product of conditional merge probabilities**

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

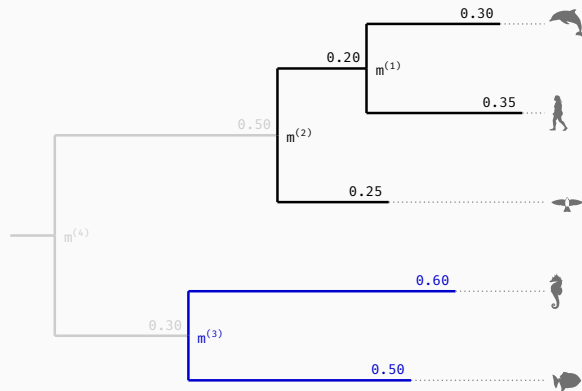
Method - Estimating tree posterior probability



- **Tree probability is product of conditional merge probabilities**

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

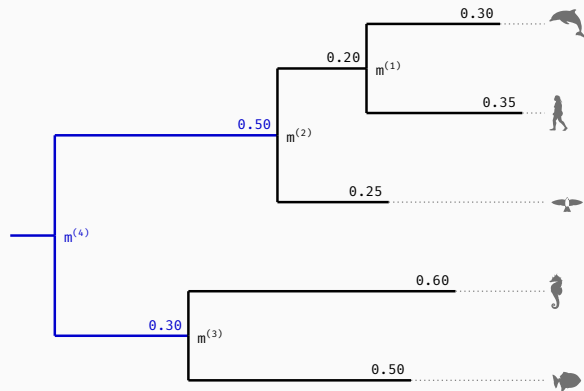
Method - Estimating tree posterior probability



- **Tree probability is product of conditional merge probabilities**

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

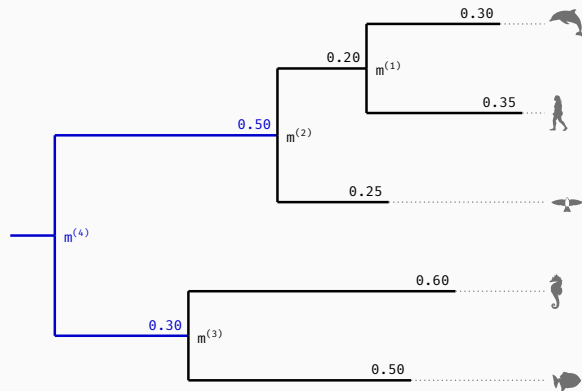
Method - Estimating tree posterior probability



- **Tree probability is product of conditional merge probabilities**

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

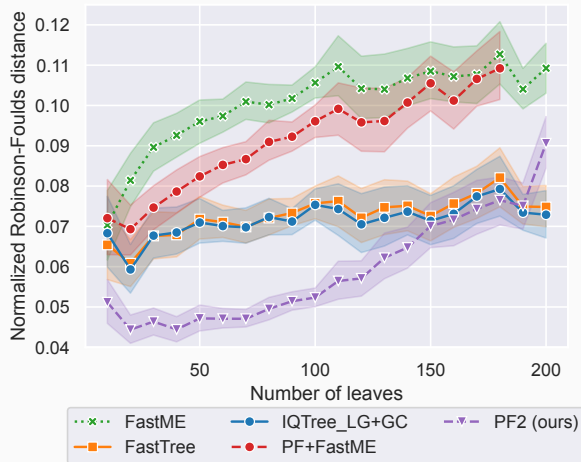
Method - Estimating tree posterior probability



- **Tree** probability is **product** of conditional **merge probabilities**
- **Sample** from the learned posterior by **sampling merges**

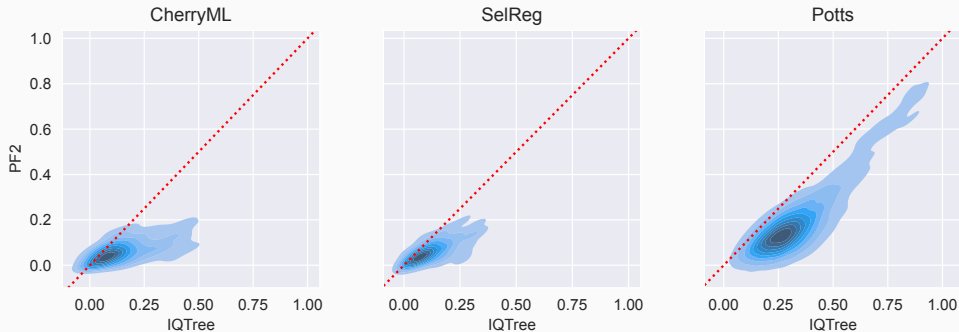
$$q_{PF2}(\theta|MSA) = q_{PF2}(m^{(1)}) \times q_{PF2}(m^{(2)}|m^{(1)}) \times q_{PF2}(m^{(3)}|m^{(\leq 2)}) \times q_{PF2}(m^{(4)}|m^{(\leq 3)})$$

Results - Topological performance



- **Very Accurate** point estimates under **tractable** LG+GC models
- Allows for **sampling** from the posterior $2.10^4 \times$ **faster than MCMC**

Results - Performance on Intracatable-Likelihood Models



Allows for **inference** under **intracatable-likelihood** models:
co-evolution, selection, heterogeneity, ...

Read the preprint



Thanks you for listening



Slides: lucblassel.com/SMBE-2026.pdf

References

Felsenstein, J. (1993). **PHYLP (phylogeny inference package), version 3.5 c.** Joseph Felsenstein.

Kleinman, C. L. et al. (2010). **Statistical Potentials for Improved Structurally Constrained Evolutionary Models.**

In: *Molecular Biology and Evolution* 27.7, pp. 1546–1560.

Supplementary Material

Sup. Methods - EvoPF architecture

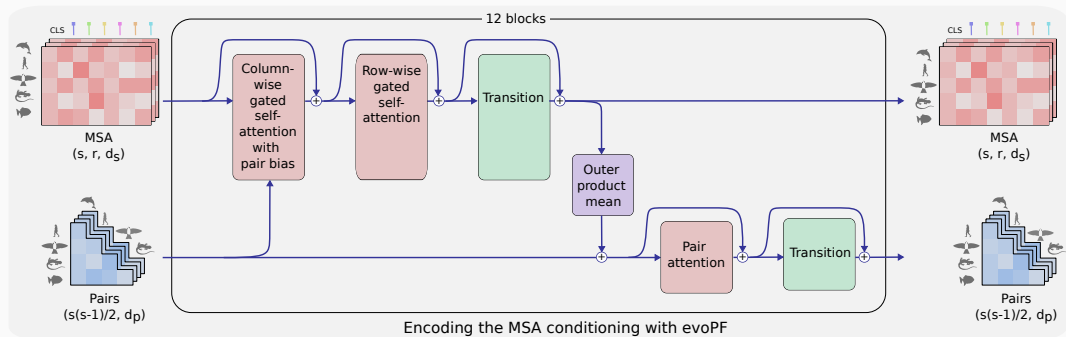
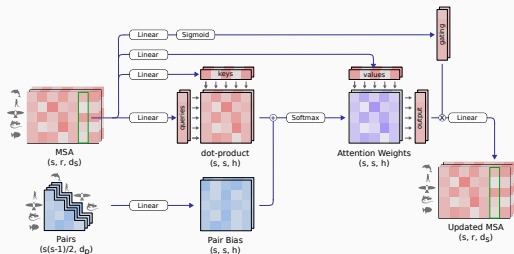
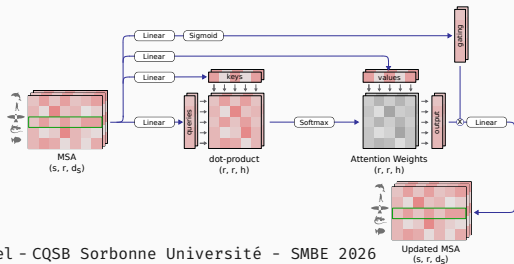


Figure inspired by Jumper et al. 2021

Sup. Methods - EvoPF, the MSA stack

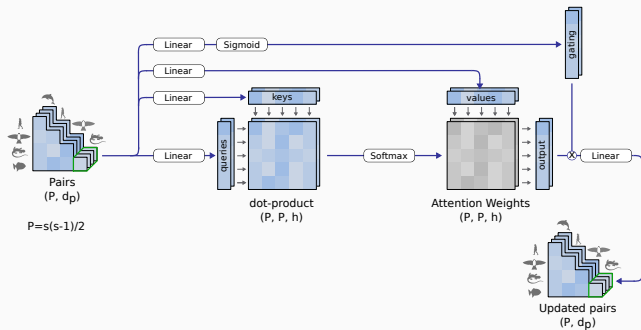


**Column-wise attention
with pair-bias**

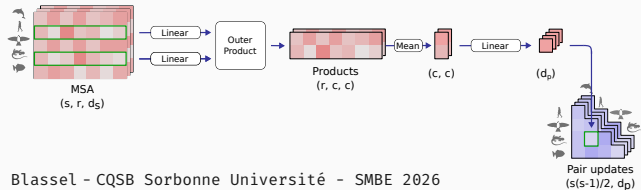


Row-wise attention

Sup. Methods - EvoPF, the pair stack



Pair attention



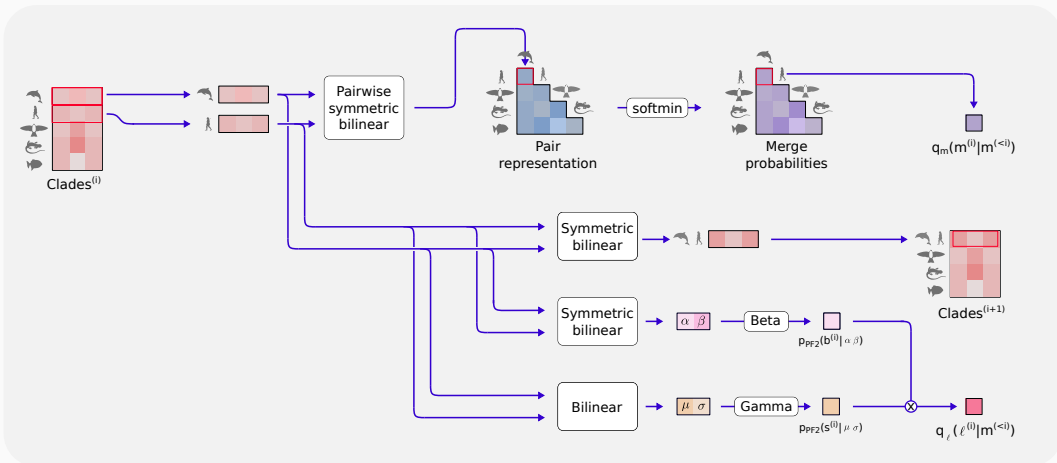
Outer product mean

Sup. Methods - The BayesNJ Module

$$m^{(i)} = \{ (l_k, l_{\ell}) , (l_k, l_{\ell}) \}$$

$$s^{(i)} = l_k + l_{\ell}$$

$$b^{(i)} = s / l_{\ell}$$



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 $m^{(k)}$:
$$M'_{ij} = \max(M_{ij}, s^{(k)}) \quad M'_{ui} = 0$$
- 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

- **Tree** is an **ordered set** of merges: $\theta : \{m^{(1)}, \dots, m^{(N-1)}\}$
- We **factorize** $q_{\psi(x)}(\theta|x)$ as the product of successive merge probabilities:

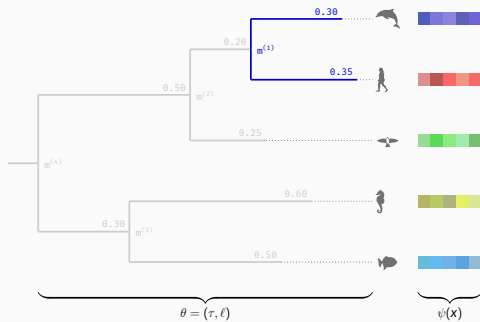
$$q_{\psi(x)}(\theta|x) = \prod_{k=1}^{N-1} q_m(m^{(k)}|m^{(<k)}) q_\ell(\ell^{(k)}|m^{(\leq k)})$$

- **Merge** probabilities have **2 components**:

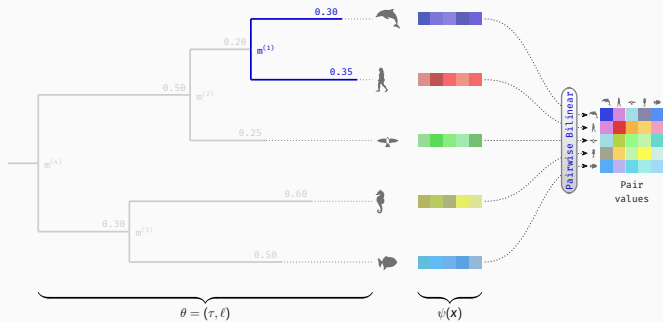
topological: $q_m(m^{(k)}|m^{(<k)})$

branch-length: $q_\ell(\ell^{(k)}|m^{(\leq k)})$

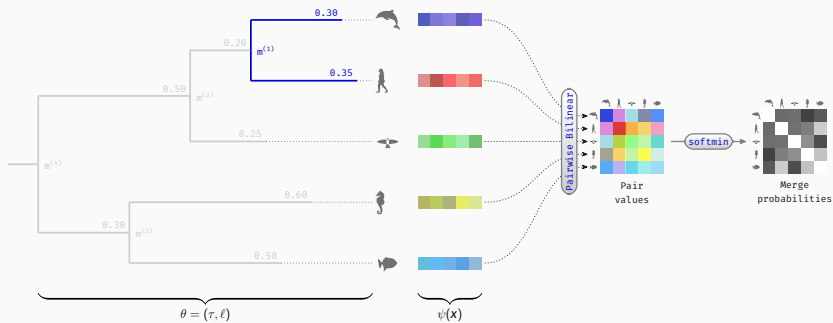
Sup. Methods - Approximating the posterior with BayesNJ (2)



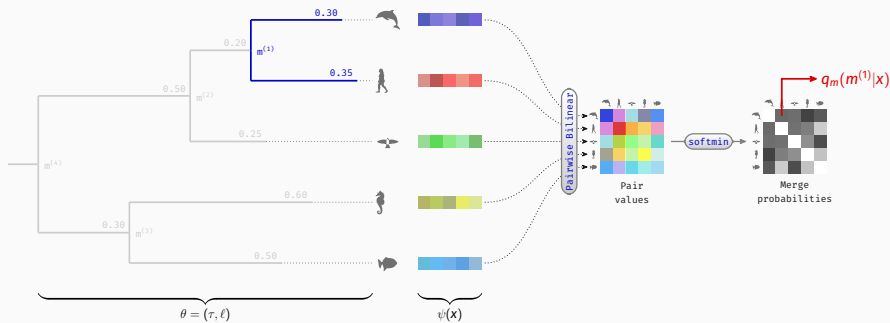
Sup. Methods - Approximating the posterior with BayesNJ (2)



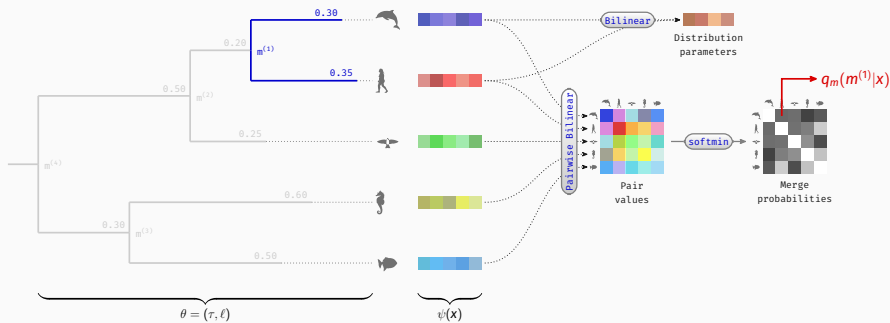
Sup. Methods - Approximating the posterior with BayesNJ (2)



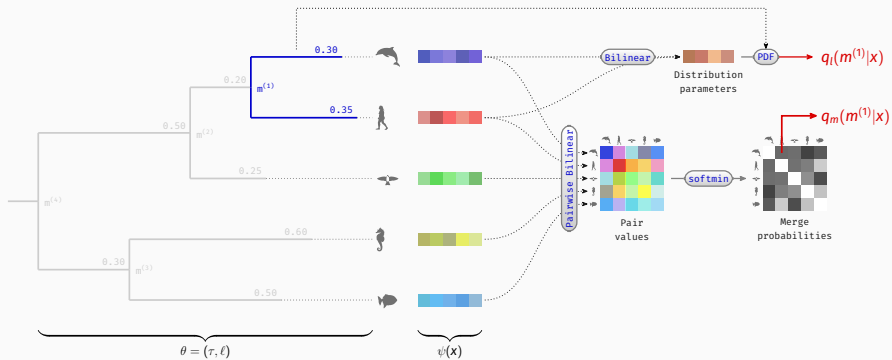
Sup. Methods - Approximating the posterior with BayesNJ (2)



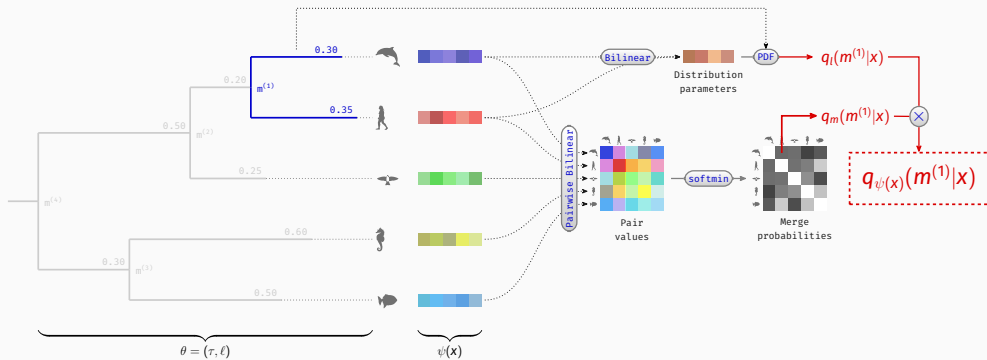
Sup. Methods - Approximating the posterior with BayesNJ (2)



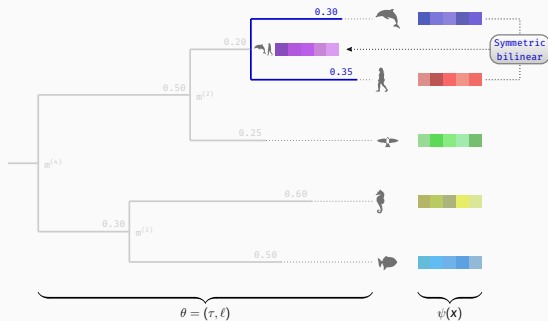
Sup. Methods - Approximating the posterior with BayesNJ (2)



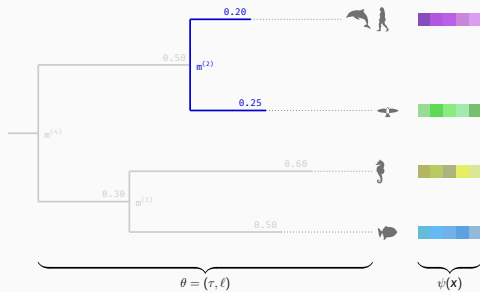
Sup. Methods - Approximating the posterior with BayesNJ (2)



Sup. Methods - Approximating the posterior with BayesNJ (2)



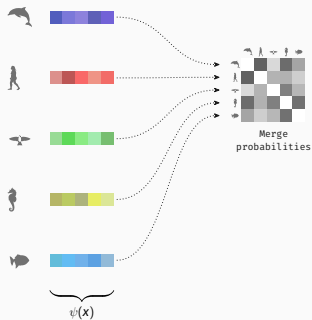
Sup. Methods - Approximating the posterior with BayesNJ (2)



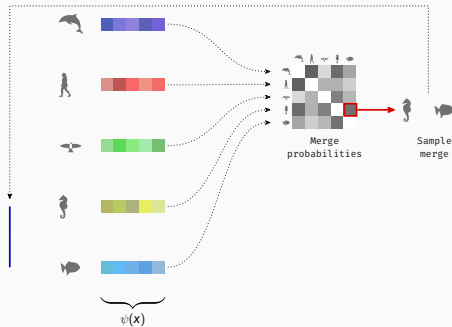
Sup. Methods - **Sampling** from the posterior with BayesNJ



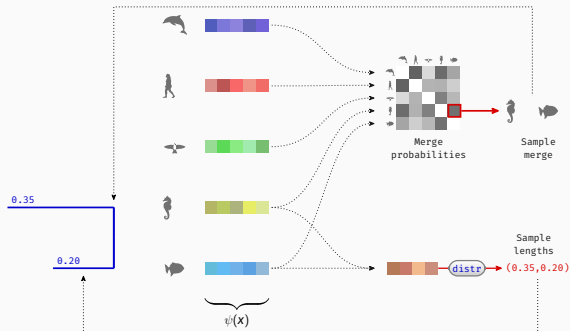
Sup. Methods - **Sampling** from the posterior with BayesNJ



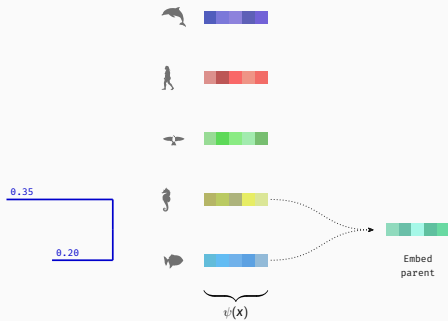
Sup. Methods - **Sampling** from the posterior with BayesNJ



Sup. Methods - **Sampling** from the posterior with BayesNJ



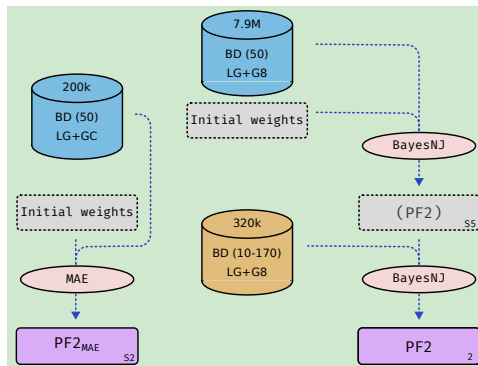
Sup. Methods - **Sampling** from the posterior with BayesNJ



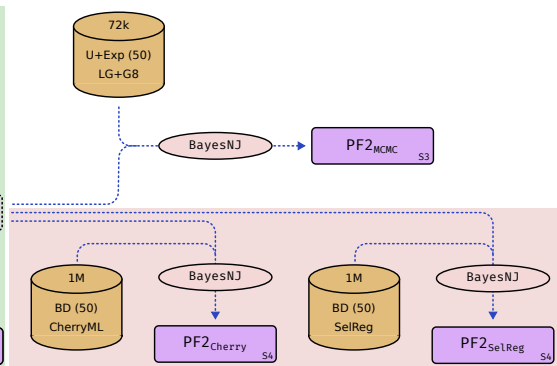
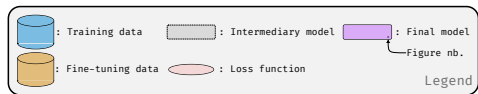
Sup. Methods - **Sampling** from the posterior with BayesNJ



Sup. Methods - Training runs

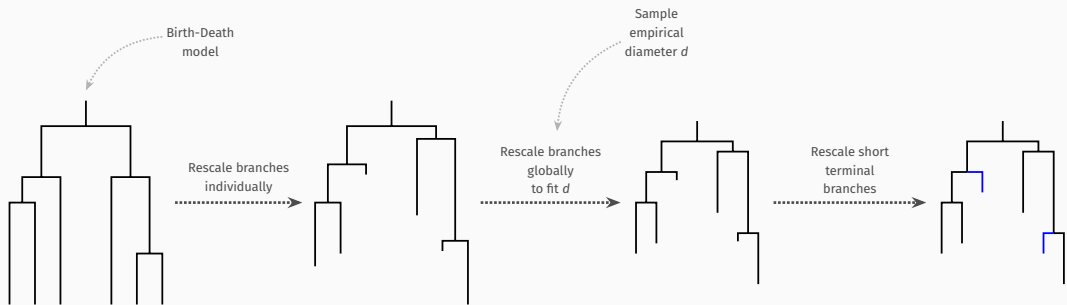


Simple evolutionary models

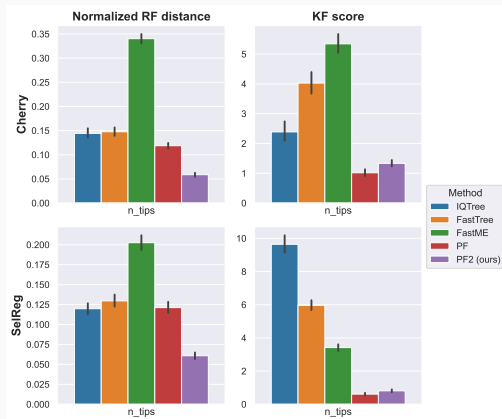


Complex evolutionary models

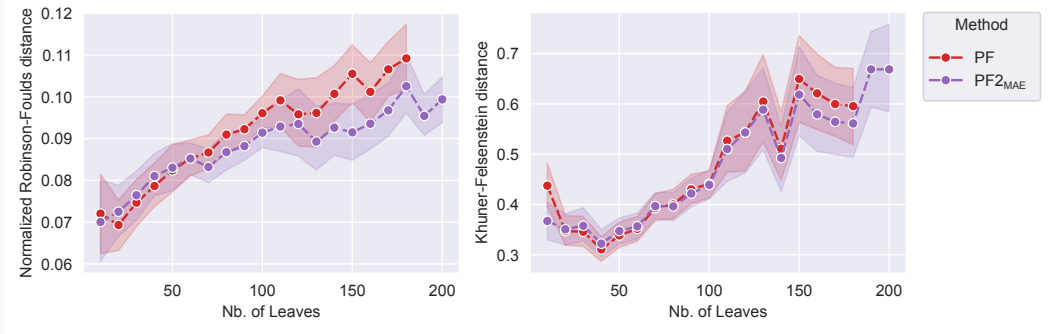
Sup. Methods - Tree simulation procedure



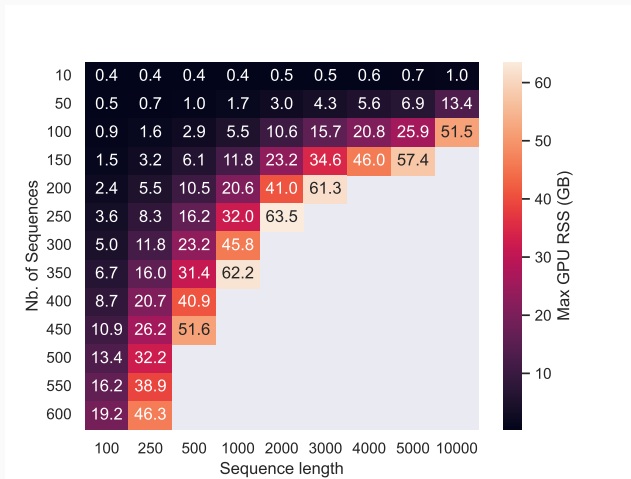
Sup. Results - PF2 performs well on complex models



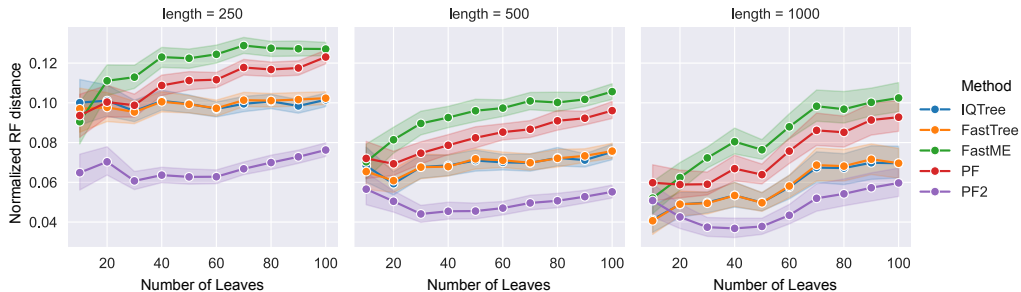
Sup. Results - Ablation study



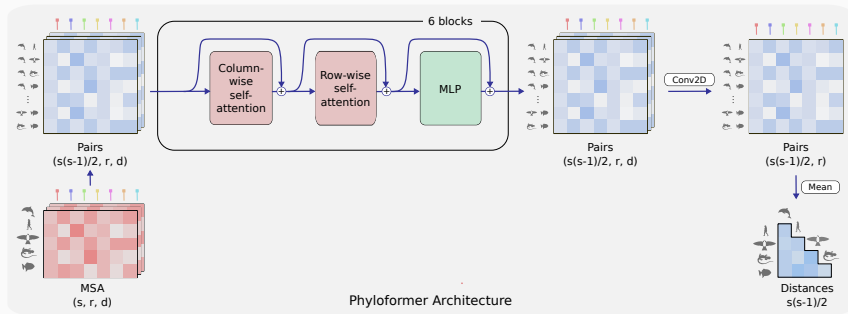
Sup. Results - Memory scaling



Sup. Results - Sensitivity to MSA length



Sup. Context - PF1 architecture



- Input an **MSA**, get a **Distance matrix**
- Feed Distance matrix to **FastME** to get **tree**

Nesterenko et al. [2025](#); Lefort et al. [2015](#)



Sup. Context - PF1 performance

