PhyloFormer:

Fast, accurate and versatile phylogenetic

reconstruction with deep neural networks

Luca Nesterenko*, **Luc Blassel***, Philippe Veber, Bastien Boussau[†], Laurent Jacob[†]

MCEB 2024 - June 20th, 2024



Context - Phylogenetic inference



Context - The problem with phylogenetic inference

- Phylogenies are hard!
- 2. **Super-exponential** tree space

$$N_{topo}(n) = \frac{(2n)!}{(n+1)!}$$





Distance-based

Fast but *inaccurate* Consider each pair

independently



Maximum likelihood-based

Accurate but slow Consider whole MSA



Method - Likelihood-free inference, motivation



Method - Amortized likelihood-free inference



Method - Phyloformer overview



Method - Similarity with structure prediction



¹Jumper et al. [2021]

Method - Why self-attention²?



- Represents elements in a set as a weighted sum of all elements (including itself)
- Parametrized by learnable weights
- Yields a context-aware and learnable representation
- Applies to sets **regardless** of cardinality

²Vaswani et al. [2017]

Method - Ensuring invariance & equivariance



Self-Attention is already permutation equivariant!

Site-wise average ensures invariance w.r.t. sites.

Results - Under LG+GC model, PF performs on par with ML



Results - What about a more complex model?



idapted from 10.1038/a41598-019-55047-4

- We simulate 250 pairs of adjacent co-evolving sites
- We use a 400 × 400 substitution matrix to describe residue co-evolution, from CherryML^a
- Most ML methods would consider sites independent

^aPrillo et al. [2023]

Results - Under a co-evolution model, PF performs the best



Results - Inference speed



- **Phyloformer** is the **fastest** method^{*a*}
- Phyloformer is even **faster than FastME** on its own
- Inference speed is independent from model complexity

^ausing a GPU

Take-Home

- On the standard LG model, Phyloformer performs on par with ML methods
- 2. It is easily adaptable to more complex models
- 3. Once trained, it is the **fastest method**^{*a*}

^aProvided you have a GPU...

Take-Home

- On the standard LG model, Phyloformer performs on par with ML methods
- 2. It is **easily adaptable** to more **complex models**
- 3. Once trained, it is the **fastest method**^{*a*}

Perspectives

- **Train** on many **complex** evolution **models** (*e.g.* Potts)
- Improve scalability by moving away from pair representation
- End-to-end MSA to tree training pipeline with differentiable tree-building

^aProvided you have a GPU...

Thanks to:

- Luca Nesterenko
- Laurent Jacob
- Bastien Boussau
- Philippe Veber
- Martin Ruffel
- Dexiong Chen
- Johanna Trost

Special thanks to Jean-Zay for all the GPUs!





anr







Additional Methods - Data Encoding



Additional Methods - Network Architecture



Additional Methods - Tree simulation



Additional Methods - Axial self-attention ³

Row attention



Column attention



Update sites in a pair by looking at all other sites in the same **pair** Update sites in a pair by looking at the same site in all other pairs

³Ho et al. [2019], Rao et al. [2021]

L. Blassel - MCEB 2024

Scaled dot-product ^a

$$Z = \textit{softmax}\left(\frac{\textit{QKT}}{\sqrt{d_k}}\right) \textit{V}$$

 $ilde{\phi}(x) = \left\{ egin{array}{cc} x+1, & ext{if } x>0 \ \exp\{(x)\} & ext{if } x\leq 0, \end{array}
ight.$

Linear Kernel Attention^{*a*}

$$\mathbf{z}_{i} = \frac{\tilde{\phi}(\mathbf{q}_{i})^{\top} \sum_{j=1}^{M} \tilde{\phi}(\mathbf{k}_{j}) \mathbf{v}_{j}}{\tilde{\phi}(\mathbf{q}_{i})^{\top} \sum_{h=1}^{M} \tilde{\phi}(\mathbf{k}_{h})}$$

Rank-1 Attention

$$z'_{i} = \frac{\tilde{\phi}(q_{i})}{M^{-1}\sum_{g=1}^{M}\tilde{\phi}(q_{g})} \cdot \frac{\sum_{j=1}^{M}\tilde{\phi}(k_{j})v_{j}}{\sum_{h=1}^{M}\tilde{\phi}(k_{h})}$$

^aVaswani et al. [2017]

^aKatharopoulos et al. [2020]

Additional Results - Linear attention impact





Additional Results - Indel Model



Additional Results - SelReg Model



Network on test dataset:	PF _{Cherry} on Cherry	PF on Cherry	PF _{Cherry} on LG	PF on LG
a = co-evolution attentions	0.256	0.255	0.120	0.135
b = other attentions	0.098	0.115	0.121	0.136
Ratio a/b	4.424	3.408	0.999	0.995
auto-attentions	0.579	0.535	0.542	0.523

• **PF** Already assigns **high-attention**

values to

- **co-evolving** site pairs
- PF_{cherry} likely exploits this signal for better performance
- PF does this without needing positional encoding

Additional Results - PF performs model-based inference



Additional Results - Fine tuning with MRE loss



Additional Results - Mean relative error



Additional Results - Memory requirements



Additional Results - Phyloformer outputs likely trees



Network Name	Starting Network	Batch Size	Dataset Size	Model of evolution	Effective number of Steps/Epochs	GPUs used	Target learning rate	Target schedule steps	Selected checkpoint step	Loss Function
PF _{Base}	Initialized network	4	170k	LG+GC	145.18k/20.5	6×A100	10 ⁻³	213.2k	144k	MAE
PF	PF _{Base}	4	224k	LG+GC	40.3k/4.32	6×A100	10 ⁻⁴	66k	40,3k	MRE
PFIndel	PF _{Base}	1	55k	LG+GC+indels	240k/17.45	4×V100	10 ⁻³	240k	136.5k	MAE
PF _{Cherry}	PF _{Base}	4	1M	Cherry	30k/0.72	6×A100	10 ⁻³	66k	18k	MAE
PF _{SelReg}	PF _{Base}	4	1M	SelReg	66k/1.58	6×A100	10 ⁻³	66k	66k	MAE

- J. Ho, N. Kalchbrenner, D. Weissenborn, and T. Salimans. Axial attention in multidimensional transformers. *CoRR*, abs/1912.12180, 2019. URL http://arxiv.org/abs/1912.12180.
- J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873): 583–589, 2021.
- A. Katharopoulos, A. Vyas, N. Pappas, and F. Fleuret. Transformers are rnns: Fast autoregressive transformers with linear attention, 2020.
- S. Prillo, Y. Deng, P. Boyeau, X. Li, P.-Y. Chen, and Y. S. Song. Cherryml: scalable maximum likelihood estimation of phylogenetic models. *Nature methods*, 20(8):1232–1236, 2023.

- R. M. Rao, J. Liu, R. Verkuil, J. Meier, J. Canny, P. Abbeel, T. Sercu, and A. Rives. Msa transformer. In M. Meila and T. Zhang, editors, *Proceedings of the* 38th International Conference on Machine Learning, volume 139 of Proceedings of Machine Learning Research, pages 8844–8856. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/rao21a.html.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez,
 Ł. Kaiser, and I. Polosukhin. Attention is all you need. Advances in neural information processing systems, 30, 2017.