Graph neural networks: A unifying predictive model architecture for evolutionary applications

The transition from explanatory to predictive models in evolutionary biology is a significant and challenging task. We propose that graph representations and graph neural networks may play a crucial role in this transition.

Published Jun 18, 2025

Arcadia Science

DOI: 10.57844/arcadia-e7kq-frwh

Purpose

Neural networks are increasingly used in evolutionary biology research. Despite this burgeoning interest, most work uses just a few model architectures. This bias matters: the alignment of data structure, task, and architecture influences predictive and explanatory outcomes.

We propose that graph neural networks (GNNs), a comparatively underutilized architecture, are uniquely well-suited for evolutionary applications. We detail how GNNs leverage relational structures embedded in evolutionary data where other architectures can't. We review example applications and discuss promising avenues where GNNs could advance evolutionary research. Our goal is to highlight the value of GNNs and encourage other evolutionary biologists to leverage the full extent of their utility.

- All associated **code** and **data** are available in this GitHub repository.

Explanatory vs. predictive models in evolution

Evolutionary biologists are driven to answer fundamental questions about *how* the world works. What led to the adaptive radiation of Darwin's finches [1]? What facilitated the repeated speciation and parallel ecological divergence between limnetic and benthic freshwater threespine sticklebacks [2]? Does epistasis increase or decrease phenotypic diversity [3]? Evolution is, historically speaking, the domain of *explanatory* rather than *predictive* models.

For example, when studying macroevolution, it's common to interpret real data by fitting idealized models of evolution (e.g., Brownian motion (BM) or Ornstein-Uhlenbeck (OU) [4]) to them. Doing so has helped advance our understanding of a number of phenomena, such as resolving how species diversification along ecological gradients can underlie adaptive radiations (e.g., Anolis lizards [5]). However, the features driving these models' explanatory power also restrict their predictive utility.

Though providing valuable biological insight, explanatory model design inherently limits their ability to predict unobserved or future outcomes. This mismatch between model intention and application isn't a shortcoming *per se* — these models were never intended to enable accurate prediction. It does mean, however, that when explanatory models are applied to predictive tasks, they rely on overly simplistic assumptions that maintain interpretability yet harm predictive capabilities. This issue isn't unique to evolutionary biology (for discussions, see [6] & [7]). For instance, phylogenetic imputation methods use explanatory models like BM or OU to predict missing trait values, constrained by assumptions such as constant rates of trait evolution across lineages and through time [8]. Dedicated predictive modeling frameworks tailored to evolutionary biology are needed.

Accordingly, evolutionary biologists have increasingly turned to machine learning frameworks more amenable to predictive tasks, particularly neural networks (NNs) (Figure 1, A–B) [9][10][11]. By leveraging multiple interconnected layers of artificial neurons, NNs can learn complex, non-intuitive relationships within data [12][13]. Despite challenges to interpretability, NNs' predictive capabilities make them

highly valuable statistical tools, especially given the intricate and subtle patterns often present in biological data.

Convolutional neural networks (CNNs: Figure 1, C–D) have become the dominant architecture used in evolutionary biology. CNNs specialize in grid-structured data, such as images and sequences, leveraging spatial autocorrelation through convolutional kernels. Somewhat famously, CNNs have been shown to be “unreasonably effective” for population genetics inference, matching or exceeding existing explanatory models [14].

However, only some biological data are structured appropriately for CNNs, and restructuring comes with trade-offs. For example, genetic data are often converted into 2D "images" despite biologically irrelevant structuring in one input dimension, potentially limiting predictive accuracy and efficiency. Data preprocessing such as this can have an outsized impact on CNN performance [15]. While 1D CNNs offer a more natural and appropriate fit for linear genomic data — and have been successfully applied across a range of population genetic tasks — both 1D and 2D CNNs require input to conform to a regular grid. This requirement restricts possible applications since biological systems are often better represented as irregular non-Euclidean relational structures. Thus, although effective in some cases, the widespread use of CNNs may reflect convenience and historical precedent as much as innate architectural suitability.

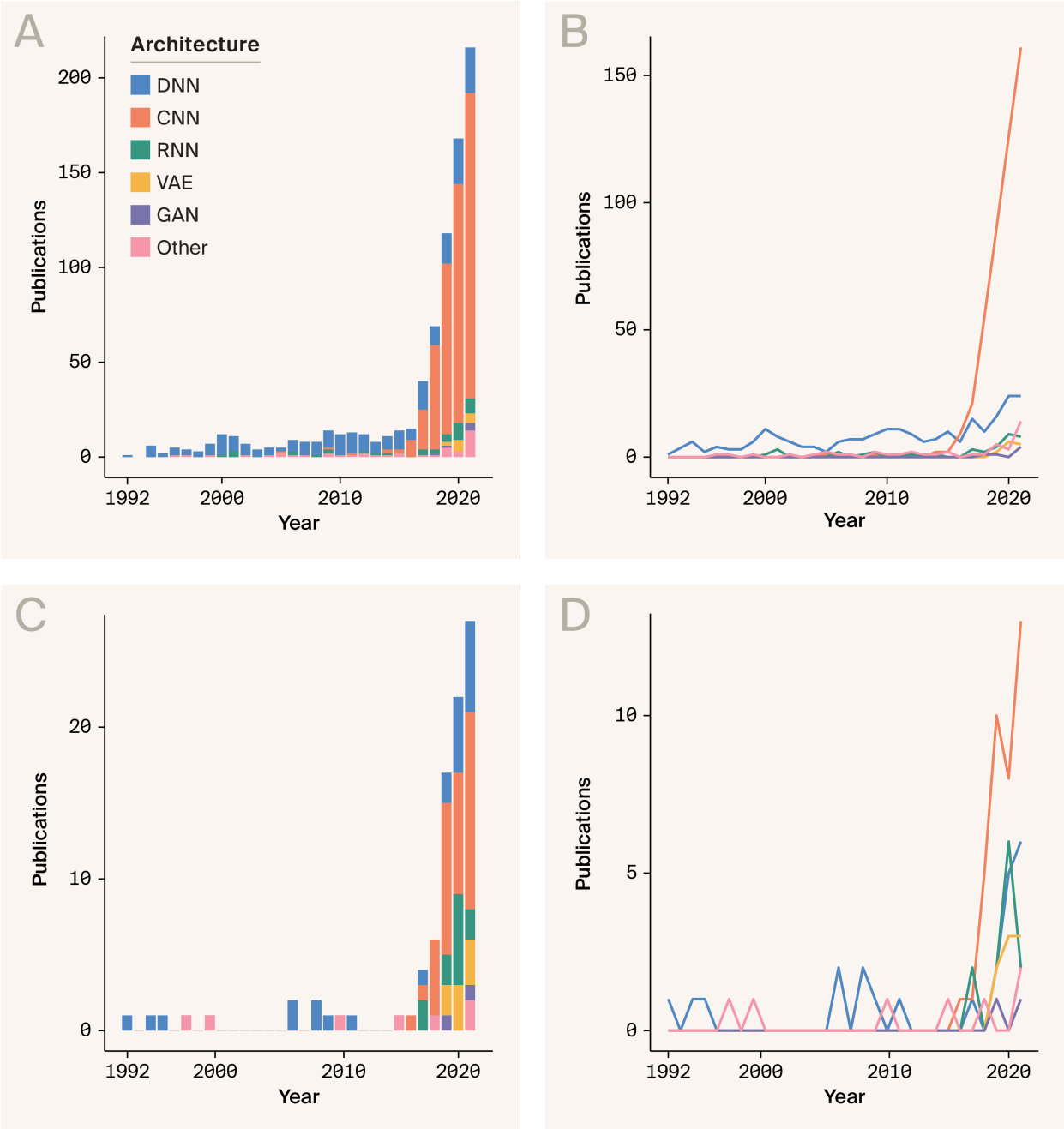

Figure 1. **Trends in the use of neural networks (NNs) in ecology and evolution (data from [9]) through 2021**.

(A–B) Count of publications using each architecture type, considering all data types.

(C–D) Count of publications using each architecture type, considering only studies using molecular data.

In all panels, any publication that used more than one architecture type is counted once per architecture. **DNN**: deep neural network, **CNN**: convolutional neural network, **RNN**: recurrent neural network, **VAE**: variational auto-encoder, **GAN**: generative adversarial network.

NOTE: The trends shown here are meant to be exemplars — we have not extended this literature review to the present day.

So, is there a model architecture better suited for evolutionary data? This is an important question. Model architectures often act like Bayesian priors, each with

unique inductive biases. Architectures can impose constraints on what models expect to see and, ultimately, what and how they learn. Effective alignment can simplify the learning task and improve predictive performance, particularly in small datasets common in biology. CNNs have succeeded in population genetic applications because genetic autocorrelation is amenable to convolution. But is there an alternative architecture better suited to the relational structures that evolution produces?

Evolution: It's graphs on graphs

We think the answer may be graphs. From phylogenies (bifurcating graphs) to ancestral recombination graphs (ARGs) to interaction networks and genotypic fitness landscapes, a vast swath of biology can be meaningfully represented as graphs. Moreover, graphs may provide the key to spanning from microevolution to macroevolution by drawing connections *between* biological scales. This puts on the table the possibility of a universal evolutionary representation, from proteins to genes and species, even ecological communities, each represented as hierarchically nested graphs (Figure 2).

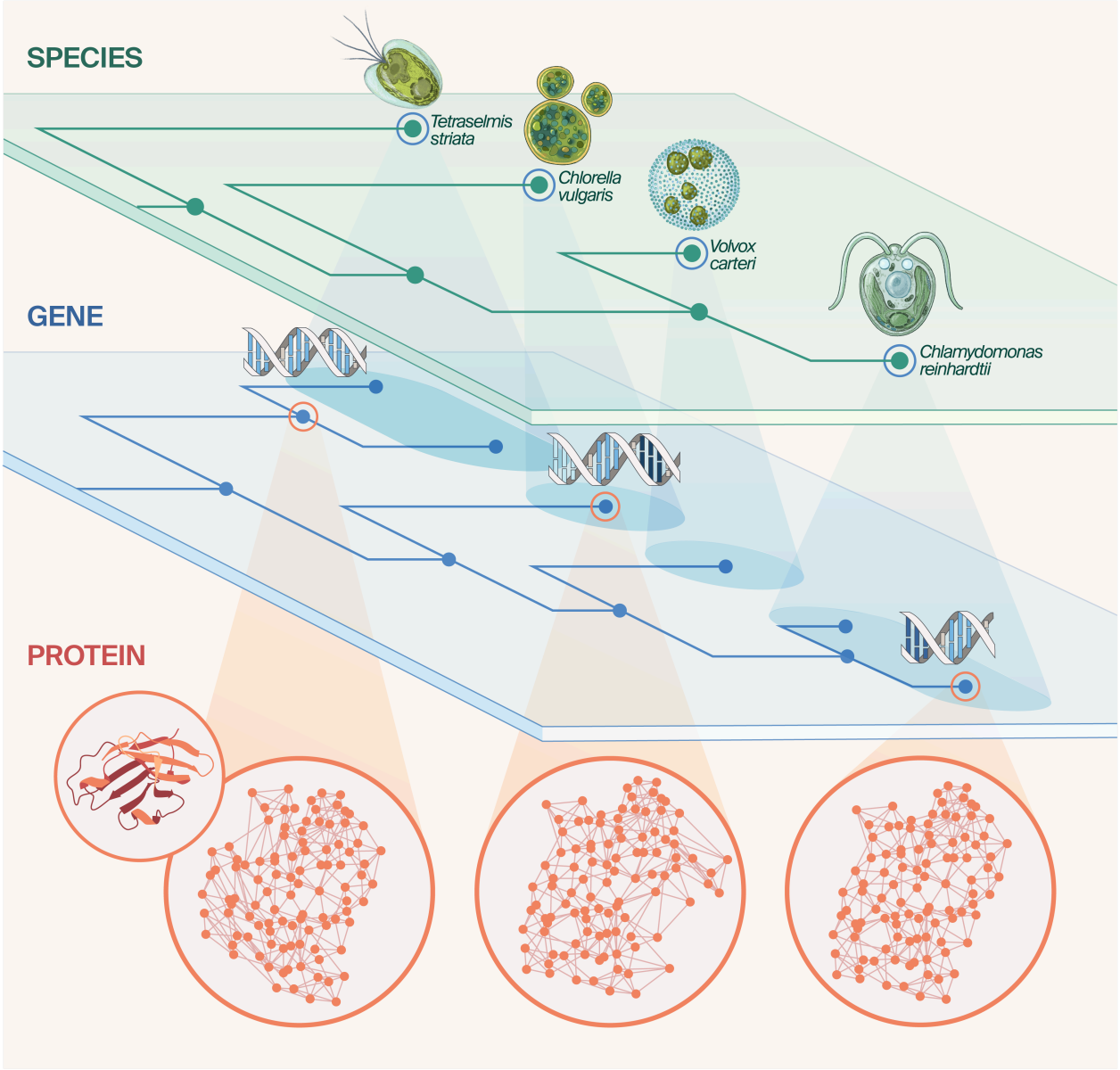

Figure 2. **Graphs are a unifying biological data structure across scales, from macroevolution to microevolution**.

(Top) Species trees (phylogenies) are fully bifurcating graphs that represent the relationships among extant species (terminal nodes) and their common ancestors (internal nodes) through descent with modification.

(Middle) A gene family tree — structured similarly — depicts the relationships among homologous gene copies possessed by the same species as in the species tree.

(Bottom) Proteins encoded by each homologous gene copy (and their common ancestors) in this gene family can be meaningfully and richly represented as protein residue graphs, where nodes correspond to amino acids, and edges correspond to interacting or spatially proximal residues, capturing detailed structural and physicochemical information.

Why is this the case? Because evolution through descent with modification induces a graph-like relational structure in biological data. We often represent these relationships as phylogenetic trees wherein each species or gene

corresponds to a node interconnected through edges representing common ancestry. Ultimately, a phylogeny is inherently a regular, fully bifurcating graph. Similarly, genetic structures such as ARGs explicitly capture the complex histories of genomic segments across populations and recombination events [16]. Furthermore, ecological networks depicting species interactions like predation, mutualism, competition, or gene regulatory networks depicting complex genetic pathways are also naturally expressed as graphs. This ubiquity underscores graph representations' inherent suitability and explanatory power for evolutionary and ecological questions.

Given the inherent suitability of graph structures to address questions in ecology and evolution, we're thus prompted to ask: Is there a predictive model architecture capable not only of handling such non-Euclidean, graph-structured data but also managing — and even exploiting — the complex nested hierarchical structures induced by evolutionary processes? After all, it's previously been shown that CNN architectures aligned to image data markedly outperform non-convolutional NNs [17][18], and architectures specialized for non-Euclidean data lead to improved outcomes by inherently respecting the data's geometry [19]. Could leveraging graph-based approaches thus bridge explanatory and predictive paradigms, harnessing the inherent relational structure of evolutionary data to improve both biological understanding and predictive accuracy?

Introducing graph neural networks

Yes! The solution we propose lies in graph neural networks (GNNs: [20]). Graph neural networks are exactly what they sound like — a neural network architecture specifically designed to process and learn from graph-structured data comprising nodes (individual entities or observations) and edges (Box 1). GNNs can be used for a variety of prediction tasks: node regression/classification (e.g., variant effect prediction), edge prediction (e.g., phylogenetic inference), and graph regression/classification (e.g., gene-regulatory network functional classification). Given that graph-structured data is abundant in biology, the potential of GNNs is vast.

Why might GNNs work so well? For one, all GNNs use message passing to aggregate information from neighboring nodes along edges, thus allowing the model to learn complex *local* relationships in the data in a manner explicitly informed by graph structure [21]. In effect, this assumes that nodes that are closer to and more connected to one another in a graph are more similar to each other. Why might we care about this as evolutionary biologists?

Because this message-passing mechanism functionally leverages something that's both the bane and boon of any evolutionary comparative study — evolutionary non-independence. Descent with modification renders biological samples statistically non-independent "evolutionary pseudoreplicates," as demonstrated compellingly in Felsenstein's seminal 1985 publication "Phylogenies and the Comparative Method" [22]. Thankfully, there now exists a wealth of statistical methods based on explanatory models that explicitly use the inferred phylogeny to account for evolutionary non-independence [4]. Just as accounting for evolutionary non-independence is essential to the adequacy and performance of explanatory models, so will it be for predictive models. In fact, we're likely to push these models even further by explicitly making the model aware of that evolutionary non-independence by baking it into the model architecture and data representation. GNNs provide us with the key to do so.

GNNs are also exceptionally flexible. For example, message-passing can incorporate convolution (as in graph convolutional networks; GCNs [23]) or attention mechanisms (as in graph attention networks; GATs [24]) to more fully learn complex relationships present in the data at both local and global scales. Furthermore, many common neural architectures are special cases of GNNs: CNNs are a special case of GCNs on regular grids, RNNs are a special case of GNNs on sequential chain graphs, and transformers [25] are a special case of GATs with fully connected attention graphs.

Thus, while CNNs have been undeniably useful — particularly in population genetic contexts with spatially or sequentially structured (i.e., Euclidean) genomic data — GNNs offer even broader flexibility. This flexibility is reassuring. Biological data frequently exhibit relational complexity beyond simple adjacency or grid-like structures. GNNs inherently accommodate these complexities, making them highly versatile tools. In the following section, we discuss several applications that

have been particularly fruitful and propose a couple of promising future applications.

GNNs in practice

Population genetic inference

As discussed previously, an early application of neural networks to evolution was the use of CNNs for population genetics. How do GNNs stack up here? Recent work [26] has found that a GCN matches and often exceeds CNN performance on population genetic tasks, particularly at identifying genomic regions under selective sweeps. Notably, the GCN achieves this performance with nearly two orders of magnitude fewer parameters than the CNN (~200 thousand parameters compared to ~21 million). This disconnect between model size and performance supports our earlier suspicion: that using an architecture aligned with the data structure indeed helps to learn more, and from less.

What about GNNs makes them suited for these tasks? The data used here — tree sequences — are highly efficient representations of genomic data that capture the changing evolutionary relationships among samples while walking along the genome [27]. These tree sequences approximate ARGs, complex graph structures capturing recombination and coalescent histories [16]. The message-passing framework inherent to GNNs allows for adaptive weighting of neighbors, enabling them to selectively integrate relevant local signals such as lineage-specific demographic events or recombination hotspots that are otherwise obscured by fixed receptive fields in CNNs. Indeed, recent studies have applied GNNs directly to ARGs, proving helpful in estimating demographic histories and identifying regions subject to selection under complex population scenarios [28].

Thus, using evolutionarily meaningful, graph-structured data, GNNs can infer everything from demographic history to the genomic landscape of natural selection and introgression/horizontal gene flow. While CNNs remain useful for specific structured genomic data tasks, the flexibility and general applicability of GNNs position them as a potentially superior choice across a broader range of population genetic and evolutionary biology problems. Despite these initially promising demonstrations, we emphasize that we have only begun to scratch the surface of GNN's potential for population genetic problems.

Diversification dynamics

GNNs may also be helpful for the inference of diversification dynamics using phylogenetic trees (e.g., for understanding speciation/extinction or mapping pathogen transmission dynamics [29][30][31]). Historically, this work has disproportionately relied upon birth–death (BD) and coalescent models. Both model types are highly interpretable. For instance, BD models employ just two primary parameters: birth (λ), corresponding either to speciation or transmission events, and death (μ), corresponding either to extinction or loss of infected individuals, respectively.

The interpretability of these models has had immense practical value. During the COVID-19 pandemic, BD and coalescent models applied to SARS-CoV-2 phylogenies provided early and critical insights into the epidemiology of this novel infectious disease, directly informing public health decisions [32][33][34][35]. Beyond COVID-19, the application of these models has long been a critical component of coordinated responses to infectious disease outbreaks [36]. For example, they've historically been instrumental in identifying emerging seasonal influenza strains around which vaccines are developed and assessing vaccine efficacy (e.g., [37]).

So, how can GNNs propel the field forward? Phylodynamics is a field where many explanatory models have been useful for prediction tasks almost by coincidence. We can move beyond this, however. For instance, GNNs could explicitly leverage the temporal structure of pathogen phylogenies to simultaneously model shifts in transmission dynamics and predict the emergence of epidemiologically important variants, something traditionally challenging for simpler models.

Initial applications of GNNs to phylodynamic problems have demonstrated substantial promise, notably in classifying transmission clusters [38][39]. However, there are several immediate areas where GNNs could be refined for this application, such as comprehensive epidemiological parameter estimation. Interestingly, a comparative study evaluating macroevolutionary diversification parameter estimates (speciation and extinction) noted that other neural network architectures often outperformed GNNs [40]. However, these GNNs lacked features that improve performance, such as skip connections or attention-based graph convolutional layers. Thus, given the inherent flexibility of GNN implementations,

a more comprehensive exploration of possibilities will be of interest here (as elsewhere).

Phylogenetic imputation and ancestral state reconstruction

Finally, GNNs may be uniquely well-suited to common tasks in comparative biology, such as trait imputation and ancestral state reconstructions. For example, ancestral state reconstruction is one of the most common use cases of phylogenetic comparative methods in evolutionary studies. Writ large, this includes the inference of everything from geographic ranges [41] and quantitative or discrete phenotypes [42] to even protein sequences [43] of the common ancestors of extant species.

Many of these tasks are built on a common methodological approach we stereotype here (for a review, see [44]). First, an explanatory model of how a trait has evolved is fit to a reconstructed phylogeny and trait data for a set of species. The fitted model is then used to probabilistically reconstruct trait values at the internal (ancestral reconstruction) or terminal (phylogenetic imputation) nodes, returning the most likely values based on the model parameters. Although intuitive, this approach can lead to biased or incorrect trait estimates, as commonly used models make unrealistic assumptions, such as constant evolutionary rates through time and shared rates across species.

GNNs, on the other hand, have the potential to model more realistic evolutionary scenarios. For instance, using a combination of graph convolution and graph attention, GNNs may be capable of flexibly and accurately modeling the underlying heterogeneity of evolutionary rates. Additionally, if modeling the evolution of multiple traits, GNNs may be able to capture additional complexity and nuance in patterns of correlated trait evolution that are typically out of reach of standard models. Last, mechanisms like jumping knowledge [45] may help GNNs to flexibly integrate information from both local and global phylogenetic neighborhoods to model and learn where saltational jumps in trait evolution occur. Fortunately, sophisticated simulation tools are readily available, enabling researchers to create realistic evolutionary scenarios for effective GNN training (e.g., [46][47][48][49]). Thus, while simulation quality remains essential, GNNs are an optimally structured architecture to handle these predictive tasks efficiently and accurately.

Tip of the iceberg

We have only begun to scratch the surface of the potential utility of GNNs for application to questions and subjects in evolutionary biology. Entire publications could be written about each. From the potential of GNNs to directly infer phylogenetic trees themselves (e.g., [50]) from genetic sequence data to predicting protein–protein interactions (e.g., [51]) and facilitating the inference of orthology at deep evolutionary time scales, the number and diversity of prospective use cases are vast. Excitingly, in many cases, we're beginning to see this exploration unfold, though we emphasize that it's just that — only the beginning. Ultimately, the creativity of implementation and thoughtful application, more than innate architectural limitations, will likely determine the success of GNNs in evolutionary biology.

Although outside of the scope of this pub, we encourage readers to familiarize themselves more with the technical details of how GNNs are implemented and how different individual architectural components may play key roles in their success and performance for any given application [20]. For instance, just as we've seen with the rampant success of the transformer architecture [25] in the context of large language models, it seems incredibly likely that GNN architectures that incorporate some form of attention mechanism will be vitally important to capture the complexity inherent to biological data. Furthermore, we emphasize that models needn't rely on a single architectural type. For instance, one recent study successfully combined protein language models with GNNs to enable the prediction of essential genes in metazoans [52].

In many cases, the primary utility of GNNs may be in bridging *across* architectures — explicitly building in the hierarchical relationships induced by evolution through descent with modification (e.g., Figure 2). Building sophisticated, complex hierarchical models such as these spanning biological scales is undoubtedly challenging, but GNNs present an explicit means by which to do so (e.g., [53][54]). Still, the value gained from more completely building in the evolutionary structure we know to exist in our models may be transformative. Ultimately, the boundary to GNN success in evolutionary biology lies primarily in our creativity and ingenuity in leveraging this powerful architecture.

Methods

We downloaded the supplementary table from Borowiec et al., 2022 [9] (found here) and converted it to a tab-separated text file. We loaded these data into R (v4.3.3), processed them, and visualized outputs with the following packages: readr (v2.1.5), dplyr (v1.1.4), tidyr (v1.3.1), stringr (v1.5.1), ggplot2 (v3.5.1) [55], reshape2 [56] (v1.4.4), cowplot (v1.1.3), and arcadiathemeR (v0.1.0) [57]. We excluded publications for which the entry for Architecture (i.e., the NN architecture used in the study) was “NA,” as these corresponded to review articles, as well as studies for which the architecture was "unknown." We counted each type once when multiple architecture types were used in a single study. For example, if a study used both a convolutional neural network and a recurrent neural network, we incremented the count for both architectures by one for that year.

We used ChatGPT to help write code and provide suggestions to restructure writing.

Contributors (A–Z)

- **Audrey Bell**: Visualization
- **Austin H. Patton**: Conceptualization, Software, Visualization, Writing
- **Ryan York**: Supervision, Validation

References

1. Grant PR, Grant BR. (2024). From microcosm to macrocosm: adaptive radiation of Darwin’s finches. https://doi.org/10.1093/evolinnean/kzae006
2. Schluter D, McPhail JD. (1992). Ecological Character Displacement and Speciation in Sticklebacks. https://doi.org/10.1086/285404
3. Baier F, Gauye F, Perez-Carrasco R, Payne JL, Schaerli Y. (2023). Environment-dependent epistasis increases phenotypic diversity in gene regulatory networks. https://doi.org/10.1126/sciadv.adf1773

4. Harmon L. (2019). Phylogenetic Comparative Methods: Learning From Trees. https://doi.org/10.32942/osf.io/e3xnr

5. Losos JB. (2009). Lizards in an Evolutionary TreeEcology and Adaptive Radiation of Anoles. https://doi.org/10.1525/california/9780520255913.001.0001

6. Shmueli G. (2010). To Explain or to Predict?. https://doi.org/10.1214/10-sts330

7. Sainani KL. (2014). Explanatory Versus Predictive Modeling. https://doi.org/10.1016/j.pmrj.2014.08.941

8. Molina-Venegas R. (2023). How to get the most out of phylogenetic imputation without abusing it. https://doi.org/10.1111/2041-210x.14198

9. Borowiec ML, Dikow RB, Frandsen PB, McKeeken A, Valentini G, White AE. (2022). Deep learning as a tool for ecology and evolution. https://doi.org/10.1111/2041-210x.13901

10. Schrider DR, Kern AD. (2018). Supervised Machine Learning for Population Genetics: A New Paradigm. https://doi.org/10.1016/j.tig.2017.12.005

11. Christin S, Hervet É, Lecomte N. (2019). Applications for deep learning in ecology. https://doi.org/10.1111/2041-210x.13256

12. McCulloch WS, Pitts W. (1943). A logical calculus of the ideas immanent in nervous activity. https://doi.org/10.1007/bf02478259

13. Rosenblatt F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain.. https://doi.org/10.1037/h0042519

14. Flagel L, Brandvain Y, Schrider DR. (2018). The Unreasonable Effectiveness of Convolutional Neural Networks in Population Genetic Inference. https://doi.org/10.1093/molbev/msy224

15. Cecil RM, Sugden LA. (2023). On convolutional neural networks for selection inference: Revealing the effect of preprocessing on model learning and the capacity to discover novel patterns. https://doi.org/10.1371/journal.pcbi.1010979

16. Rasmussen MD, Hubisz MJ, Gronau I, Siepel A. (2014). Genome-Wide Inference of Ancestral Recombination Graphs. https://doi.org/10.1371/journal.pgen.1004342

17. Lecun Y, Bottou L, Bengio Y, Haffner P. (1998). Gradient-based learning applied to document recognition. https://doi.org/10.1109/5.726791

18. Krizhevsky A, Sutskever I, Hinton GE. (2017). ImageNet classification with deep convolutional neural networks. https://doi.org/10.1145/3065386

19. Bronstein MM, Bruna J, LeCun Y, Szlam A, Vandergheynst P. (2017). Geometric Deep Learning: Going beyond Euclidean data. https://doi.org/10.1109/msp.2017.2693418

20. Corso G, Stark H, Jegelka S, Jaakkola T, Barzilay R. (2024). Graph neural networks. https://doi.org/10.1038/s43586-024-00294-7

21. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. (2017). Neural Message Passing for Quantum Chemistry. https://doi.org/10.48550/arxiv.1704.01212

22. Felsenstein J. (1985). Phylogenies and the Comparative Method. https://doi.org/10.1086/284325

23. Scarselli F, Gori M, Ah Chung Tsoi , Hagenbuchner M, Monfardini G. (2009). The Graph Neural Network Model. https://doi.org/10.1109/tnn.2008.2005605

24. Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. (2017). Graph Attention Networks. https://doi.org/10.48550/arxiv.1710.10903

25. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. (2017). Attention Is All You Need. https://doi.org/10.48550/arxiv.1706.03762

26. Whitehouse LS, Ray DD, Schrider DR. (2024). Tree Sequences as a General-Purpose Tool for Population Genetic Inference. https://doi.org/10.1093/molbev/msae223

27. Kelleher J, Wong Y, Wohns AW, Fadil C, Albers PK, McVean G. (2019). Inferring whole-genome histories in large population datasets. https://doi.org/10.1038/s41588-019-0483-y

28. Korfmann K, Sellinger TPP, Freund F, Fumagalli M, Tellier A. (2024). Simultaneous Inference of Past Demography and Selection from the Ancestral Recombination Graph under the Beta Coalescent. https://doi.org/10.24072/pcjournal.397

29. Morlon H. (2014). Phylogenetic approaches for studying diversification. https://doi.org/10.1111/ele.12251

30. Featherstone LA, Zhang JM, Vaughan TG, Duchene S. (2022). Epidemiological inference from pathogen genomes: A review of phylodynamic models and applications. https://doi.org/10.1093/ve/veac045

31. Morlon H, Andréoletti J, Barido-Sottani J, Lambert S, Perez-Lamarque B, Quintero I, Senderov V, Veron P. (2024). Phylogenetic Insights into Diversification. https://doi.org/10.1146/annurev-ecolsys-102722-020508

32. Fountain-Jones NM, Appaw RC, Carver S, Didelot X, Volz E, Charleston M. (2020). Emerging phylogenetic structure of the SARS-CoV-2 pandemic. https://doi.org/10.1093/ve/veaa082

33. Lai A, Bergna A, Acciarri C, Galli M, Zehender G. (2020). Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. https://doi.org/10.1002/jmv.25723

34. Vaughan TG, Sciré J, Nadeau SA, Stadler T. (2020). Estimates of outbreak-specific SARS-CoV-2 epidemiological parameters from genomic data. https://doi.org/10.1101/2020.09.12.20193284

35. Attwood SW, Hill SC, Aanensen DM, Connor TR, Pybus OG. (2022). Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. https://doi.org/10.1038/s41576-022-00483-8

36. Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, Prosperi MC. (2017). Phylodynamic applications in 21st century global infectious disease research. https://doi.org/10.1186/s41256-017-0034-y

37. Castro LA, Bedford T, Ancel Meyers L. (2020). Early prediction of antigenic transitions for influenza A/H3N2. https://doi.org/10.1371/journal.pcbi.1007683

38. Sun C, Fang R, Salemi M, Prosperi M, Rife Magalis B. (2024). DeepDynaForecast: Phylogenetic-informed graph deep learning for epidemic transmission dynamic prediction. https://doi.org/10.1371/journal.pcbi.1011351

39. Sun C, Li Y, Marini S, Riva A, Wu DO, Fang R, Salemi M, Rife Magalis B. (2024). Phylogenetic-informed graph deep learning to classify dynamic transmission clusters in infectious disease epidemics. https://doi.org/10.1093/bioadv/vbae158

40. Lajaaiti I, Lambert S, Voznica J, Morlon H, Hartig F. (2023). A Comparison of Deep Learning Architectures for Inferring Parameters of Diversification Models from Extant Phylogenies. https://doi.org/10.1101/2023.03.03.530992

41. Ree RH, Moore BR, Webb CO, Donoghue MJ. (2005). A LIKELIHOOD FRAMEWORK FOR INFERRING THE EVOLUTION OF GEOGRAPHIC RANGE ON PHYLOGENETIC TREES. https://doi.org/10.1111/j.0014-3820.2005.tb00940.x

42. Schluter D, Price T, Mooers AØ, Ludwig D. (1997). LIKELIHOOD OF ANCESTOR STATES IN ADAPTIVE RADIATION. https://doi.org/10.1111/j.1558-5646.1997.tb05095.x

43. Merkl R, Sterner R. (2016). Ancestral protein reconstruction: techniques and applications. https://doi.org/10.1515/hsz-2015-0158

44. Revell L. (2024). Ancestral state reconstruction of phenotypic characters. https://doi.org/10.32942/x2h61q

45. Xu K, Li C, Tian Y, Sonobe T, Kawarabayashi K-i, Jegelka S. (2018). Representation Learning on Graphs with Jumping Knowledge Networks. https://doi.org/10.48550/arxiv.1806.03536

46. Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon LJ. (2014). geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. https://doi.org/10.1093/bioinformatics/btu181

47. Morlon H, Lewitus E, Condamine FL, Manceau M, Clavel J, Drury J. (2016). RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. https://doi.org/10.1111/2041-210x.12526

48. Guillerme T. (2024). treats: A modular R package for simulating trees and traits. https://doi.org/10.1111/2041-210x.14306

49. Roa Lozano J, Duncan M, McKenna DD, Castoe TA, DeGiorgio M, Adams R. (2024). TraitTrainR: accelerating large-scale simulation under models of continuous trait evolution. https://doi.org/10.1093/bioadv/vbae196

50. Zhang C. (2023). Learnable Topological Features for Phylogenetic Inference via Graph Neural Networks. https://doi.org/10.48550/arxiv.2302.08840

51. Xu M, Qian P, Zhao Z, Zeng Z, Chen J, Liu W, Yang X. (2024). Graph Neural Networks for Protein-Protein Interactions -- A Short Survey. https://doi.org/10.48550/arxiv.2404.10450

52. Ma J, Song J, Young ND, Chang BCH, Korhonen PK, Campos TL, Liu H, Gasser RB. (2023). 'Bingo'—a large language model- and graph neural network-based workflow for the prediction of essential genes from protein data. https://doi.org/10.1093/bib/bbad472

53. Rampasek L, Wolf G. (2021). Hierarchical Graph Neural Nets can Capture Long-Range Interactions. https://doi.org/10.1109/mlsp52302.2021.9596069

54. Sobolevsky S. (2021). Hierarchical Graph Neural Networks. https://doi.org/10.48550/arxiv.2105.03388

55. Wickham H. (2016). ggplot2. https://doi.org/10.1007/978-3-319-24277-4

56. Wickham H. (2007). Reshaping Data with the**reshape**Package. https://doi.org/10.18637/jss.v021.i12

57. https://github.com/arcadia-science/arcadiathemer