Highlight: Leveraging Phylogenomics to Uncover the Loci of Repeated Evolution

Pedro Andrade *

*Corresponding author: E-mail: pedroamandrade@gmail.com.

Accepted: August 19, 2025

Highlights editors: Laura A. Katz and Maud Tenaillon

Observations of similar phenotypes occurring in distinct lineages have profoundly shaped early views on the role of adaptation in evolution. A classic example is the suite of morphological and physiological adaptations associated with multiple independent transitions to marine life in mammals (Fig. 1). Analogous phenotypes often arise when distinct lineages are exposed to similar selective pressures, resulting in the evolution of shared adaptive traits. An interesting corollary is that the repeated evolution of certain traits can be associated with genetic changes at a relatively small number of loci, which act as "hotspots" of evolutionary change (Martin and Orgogozo 2013). This may be because these genes play a central role in metabolic or developmental pathways relevant to the trait, or because mutations at these loci tend to have limited effects on other traits (i.e. low pleiotropy), making them more likely to be retained by natural selection; these factors likely promote "gene reuse" during the evolution of a trait across lineages. With the rapid accumulation of genomic and phenotypic data, there is now an opportunity to leverage signals of replicated evolution to better understand the genetic basis of organismal adaptation—a particularly urgent goal in the context of ongoing global environmental changes.

One researcher working to tackle this question is Arlie Macdonald (Fig. 2), a PhD student at the ARC Centre of Excellence for Plant Success (Australia) under the supervision of Barbara Holland, Jonathan Mitchell, and Maddie James. They are particularly interested in exploring how fields such as agriculture and conservation could use the growing availability of phylogenetic and trait data to characterize adaptation in plants. "In the face of climate change, there is great interest in understanding the genetic underpinnings of traits that allow plants to thrive in arid environments" they explain, adding that many traits that mediate adaptation to these environments "aren't obvious targets for genome-wide association studies as they tend to vary across species rather than within species; however, shifts from wetter to more arid environments have occurred independently many times in different species". Hence, Macdonald and colleagues are exploring the potential of a new family of approaches, termed "Phylogenetic Genotype to Phenotype mapping" (PhyloG2P), which are the focus of a new review in *Genome Biology and Evolution* (Macdonald et al. 2025).

PhyloG2P methods leverage phylogenetic reconstruction (i.e. the generation of evolutionary trees) and trait data to associate genotype to phenotype across lineages, from closely related to highly divergent taxa. Although some PhyloG2P approaches can be based on traits that have evolved only once within a lineage (using genetic and phenotypic state reconstruction to correlate genetic changes to the evolution of the target phenotypes), most studies focus on traits that have arisen repeatedly across multiple lineages, taking advantage of replicated evolution to separate confounding lineage-specific genetic changes from those shared across lineages. The latter approach is the primary focus of Macdonald and colleagues' review article (Fig. 1). "The recent burst in development of PhyloG2P methods is probably a consequence of the combination of large, well-resolved phylogenies and the increasing number of detailed trait databases - together, these allow for the identification of replicated evolution," argues Macdonald, when referring to the recent development of the field, which was also facilitated by increasingly abundant nucleotide sequence data. In their review, the authors explore in detail how to query phylogenomic data to investigate trait evolution—with the important caveat that accurately defining the phenotype of interest is the keystone to any trait mapping approach, particularly when mapping traits across species, or even deeper phylogenetic scales.

Methodological advances include approaches based on modeling changes in evolutionary rate between lineages

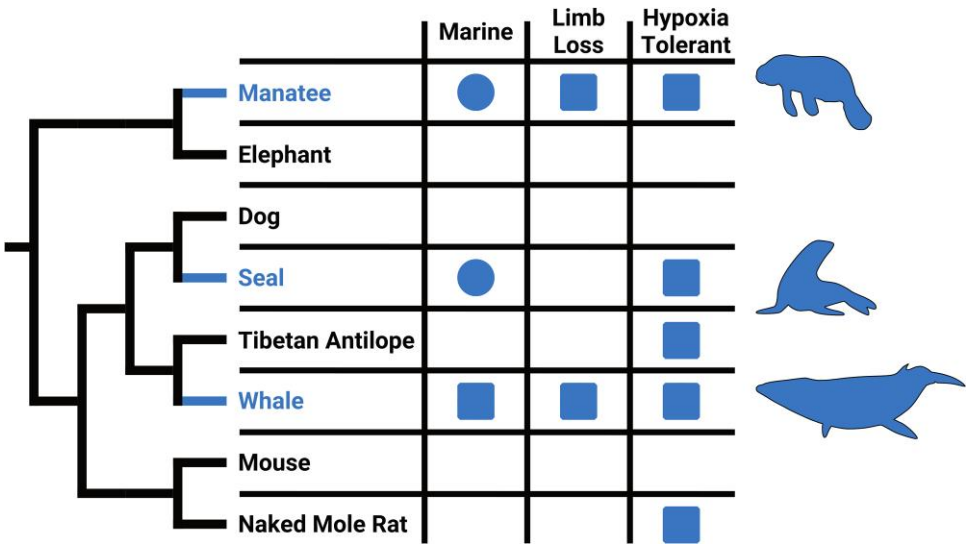

Fig. 1. Conceptual schematic representing the main approach for identifying biological targets for PhyloG2P studies; these take advantage of replicated phenotypic evolution to detect signatures of shared genotypic evolution (from Macdonald et al. 2025). In this hypothetical example, traits associated with the independent adaptation to a marine environment in three lineages of mammals. Silhouettes were retrieved from PhyloPic (phylopic.org) and used under a CC0 1.0 Universal license.

with and without the focal trait. These are becoming increasingly popular, as they detect broader changes in evolutionary conservation at a locus, without the need to identify causal mutations. Two specific bioinformatics tools, RERconverge and PhyloAcc, have received particular attention. RERconverge (Kowalczyk et al. 2019) estimates the relative evolutionary rate (RER) of each genomic locus, for every branch of the tree, normalizing estimates for genome-wide effects; it then tests for association between relative evolutionary rate and the evolution of a focal trait, comparing rates in branches where the trait is present versus branches where the trait is absent. PhyloAcc (Hu et al. 2019) works by identifying a lineage with a trait of interest, then using a Bayesian approach with three nested models to detect non-coding regions with evidence of accelerated evolution in this lineage, compared to others. These two approaches are only a small sample of the variety of methods that have been used under a PhyloG2P framework; this diversity of approaches makes Macdonald particularly fascinated with this topic, but they reiterate that improving accessibility to trait and sequence databases, as well as bioinformatic tools, is of primary importance to ensure their wider adoption by the evolutionary genomics community.

Macdonald et al. (2025) also underline the importance of considering ways to score traits other than binary "presence-or-absence". They suggest that modeling traits as continuous variables may help to capture their underlying biological complexity. While most studies published so far have relied on straightforward binary scoring of the target trait, novel methods are being developed that correlate continuous change in a trait's value and the evolutionary rate of specific lineages. One example cited by the authors of the integration of continuous trait data is Kowalczyk et al. (2020), who applied RERconverge to compare relative evolutionary rates of proteins of 61 mammal species. Using this approach, Kowalczyk et al. (2020) detected not only specific genes underlying longevity but also identified increased evolutionary constraint on genes participating in longevity-associated pathways.

Looking forward, PhyloG2P and traditional trait mapping and functional approaches may serve complementary roles for understanding trait evolution. For example, GWAS (genome-wide association studies) focuses on mapping phenotypes that segregate within interbreeding populations, making it better suited for in-depth studies of single species; PhyloG2P approaches, on the other hand, are more suitable for a broader initial detection of signatures of association across lineages. As Macdonald concludes: "it may still be challenging for PhyloG2P methods to reveal causal links between genotype and phenotype on their own, so we will still likely require follow-up transgenic or knockout experiments, coupled with QTL mapping or GWAS".

Fig. 2. GBE author spotlight: Arlie R. Macdonald. Arlie is a queer PhD student at the Australian Research Council Centre of Excellence for Plant Success. Their PhD project focuses on using phylogentic approaches to map genotype to phenotype, in an emerging field known as PhyloG2P. Arlie particularly enjoys working in the scientific environment created by the interdisciplinary Plant Success research center, which brings many fields of biology and mathematics together. As part of the center, Arlie has worked with the Researcher Development Group to organize and deliver several training workshops for their fellow early career researchers, ranging from statistical analysis to scientific presentation skills. They are also deeply passionate about science communication and using education to empower and uplift marginalized people. Outside of their scientific work, you might find Arlie on the stage of an improvised comedy show, or on one of the amazing bush walks near their home in Meanjin/Brisbane. Arlie R. Macdonald (photo: Natalie C. Procopio).

PhyloG2P could thus serve as a framework for the initial exploration of the genomic basis of traits across multiple species, leveraging evolutionary replication, trait data, and phylogenomics for a first look at the loci underlying phenotypic evolution across the tree of life.

Want to learn more? Check out these other articles on methodological approaches in evolutionary genomics recently published in *Genome Biology and Evolution*:

- Henry GA, Stinchcombe JR. Predicting fitness-related traits using gene expression and machine learning. Genome Biol Evol. 2025:17(2):evae275. https://doi.org/10.1093/gbe/evae275.
- Kim D, Park S, Steinegger M. Unicore enables scalable and accurate phylogenetic reconstruction with structural core genes. Genome Biol Evol. 2025:17(6):evaf109. https://doi.org/10.1093/gbe/evaf109.
- Xia X. On rooting and dating viral trees with a changing evolutionary rate following host-switching. Genome Biol Evol. 2025:17(7):evaf134. https://doi.org/10.1093/gbe/evaf134.

Literature Cited

Hu Z, Sackton TB, Edwards SV, Liu JS. Bayesian detection of convergent rate changes of conserved noncoding elements on phylogenetic trees. Mol Biol Evol. 2019:36:1086–1100. https://doi.org/10.1093/molbev/msz049.

Kowalczyk A, et al. RERconverge: an R package for associating evolutionary rates with convergent traits. Bioinformatics. 2019:35:4815–4817. https://doi.org/10.1093/bioinformatics/btz468.

Kowalczyk A, Partha R, Clark NL, Chikina M. Pan-mammalian analysis of molecular constraints underlying extended lifespan. Elife. 2020:11:e51089. https://doi.org/10.7554/eLife.51089.

Macdonald AR, James ME, Mitchell JD, Holland BR. From trees to traits: a review of advances in PhyloG2P methods and future directions. Genome Biol Evol. 2025:17(9):evafXXX. https://doi.org/10.1093/gbe/evafXXX.

Martin A, Orgogozo V. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution. 2013:67:1235–1250. https://doi.org/10.1111/evo.12081.