*Integrative and Comparative Biology*

*Integrative and Comparative Biology*, volume 0, pp. 1–11
https://doi.org/10.1093/icb/icaf037

*Society for Integrative and Comparative Biology*

PERSPECTIVE

# Integrating Intermediate Traits in Phylogenetic Genotype-to-Phenotype Studies

Nathan L. Clark [*], Chris Todd Hittinger [†], Hongmei Li-Byarlay [‡], Antonis Rokas [§], Timothy B. Sackton [¶] and Robert L. Unckless [||,1]

[*]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA; [†]Laboratory of Genetics, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, WI 53706, USA; [‡]Agricultural Research and Development Program, Department of Agricultural and Life Sciences, Central State University, Wilberforce, OH 45384, USA; [§]Department of Biological Sciences and Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN 37235, USA; [¶]Informatics Group, Harvard University, Cambridge, MA 02138, USA; [||]Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66045, USA

[1]E-mail: unckless@ku.edu

**Synopsis** A major goal of research in evolution and genetics is linking genotype to phenotype. This work could be direct, such as determining the genetic basis of a phenotype by leveraging genetic variation or divergence in a developmental, physiological, or behavioral trait. The work could also involve studying the evolutionary phenomena (e.g., reproductive isolation, adaptation, sexual dimorphism, behavior) that reveal an indirect link between genotype and a trait of interest. When the phenotype diverges across evolutionarily distinct lineages, this genotype-to-phenotype problem can be addressed using phylogenetic genotype-to-phenotype (PhyloG2P) mapping, which uses genetic signatures and convergent phenotypes on a phylogeny to infer the genetic bases of traits. The PhyloG2P approach has proven powerful in revealing key genetic changes associated with diverse traits, including the mammalian transition to marine environments and transitions between major mechanisms of photosynthesis. However, there are several intermediate traits layered in between genotype and the phenotype of interest, including but not limited to transcriptional profiles, chromatin states, protein abundances, structures, modifications, metabolites, and physiological parameters. Each intermediate trait is interesting and informative in its own right, but synthesis across data types has great promise for providing a deep, integrated, and predictive understanding of how genotypes drive phenotypic differences and convergence. We argue that an expanded PhyloG2P framework (the PhyloG2P matrix) that explicitly considers intermediate traits, and imputes those that are prohibitive to obtain, will allow a better mechanistic understanding of any trait of interest. This approach provides a proxy for functional validation and mechanistic understanding in organisms where laboratory manipulation is impractical.

## Introduction to the genotype-phenotype map

A central goal of evolutionary genetics is to understand the genetic differences that underlie phenotypic variation across individuals and species. Understanding this genotype-to-phenotype (G2P) map can help answer questions about the genetic architecture of traits and adaptation, the role of particular genetic loci in adaptation, and facilitate the prediction of phenotypic trait values. Indeed, many significant advances have come from quantitative genetics and association mapping (Tibbs Cortes et al. 2021; Uffelmann et al. 2021; Tan et al. 2023), methods that rely on statistical associations between genetic and phenotypic variation across individuals of a population. These approaches have been deeply informative for uncovering G2P maps, especially with very large sample sizes (Bycroft et al. 2018). Genomic selection in many agricultural systems (Crossa et al. 2017; Hayes et al. 2024), as well as the potential to predict genetic changes leading to drug resistance in pathogens and tumors (Luth et al. 2024), already demonstrate the potential power of a deep understanding of G2P links in model systems. Uncovering the G2P map for diverse traits would

Advance Access publication May 15, 2025

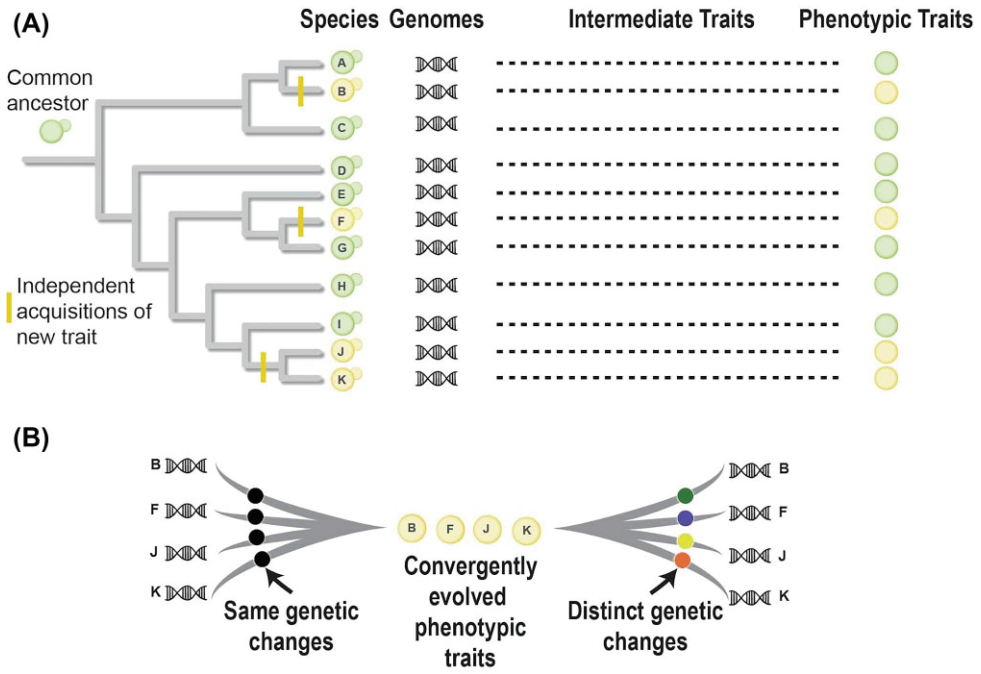



**Fig. 1** The PhyloG2P framework: traversing the genotype to phenotype map through convergence. (A) Convergent evolution is very common in the tree of life (Conway Morris 2007). Many traits, including but not limited to coloration, eyes, the ability to grow in specific substrates, and oxygen-transport systems, have evolved convergently. The PhyloG2P approach takes advantage of the repeated evolution of a given trait to identify molecular markers that are associated with the trait. The approach has been used to associate trait variation with a wide variety of molecular markers (e.g., variation in the presence or absence of genes, variation in substitution patterns in protein-coding or non-coding genomic regions, and variation in transcriptional profiles). However, the association is typically between one type of molecular marker and the trait of interest, and variation for other intermediate traits is typically not examined. Intermediate traits are marked with ellipses, and examples are given in Fig. 2. (B) Convergently evolved phenotypic traits can originate via the repeated and independent occurrence of the same genetic changes (left) or via the independent occurrence of distinct genetic changes that may affect the same or similar intermediate traits (right). Fig. modified with permission from (Gonçalves et al. 2024).

provide enormous benefits to our understanding of the biological world, allowing predictions of genetic and phenotypic change in response to climate change, and informing challenges in medicine and agriculture.

However, while DNA sequence changes are ultimately the causal factor behind most phenotypic differences, the translation from genotype to phenotype proceeds via a large number of molecular intermediates, including changes in gene expression, protein function, biochemical activity, and pathway activation. Thus, the standard G2P map can function like a black box, where even highly accurate predictions can occur in the absence of causal understanding. In this perspective, we highlight the power of phylogenetic methods in disentangling the G2P map. We then discuss how joint analysis of DNA sequences, phenotypes, and the molecular intermediates, especially when combined with new machine learning methods, can facilitate extending G2P inference across the tree of life.

## Phylogenetically bridging from genotype to phenotype: The PhyloG2P Concept

Some of life's most spectacular phenotypic changes are evolutionary adaptations that involve drastic phenotypic change over millions of years. To understand the genotypic changes underlying these diverse phenotypes requires extending the idea of statistical associations between genotype and phenotype from a population to a phylogeny, known as PhyloG2P (Smith et al. 2020, Fig. 1). From a researcher's point of view, perhaps the most useful adaptations to apply the PhyloG2P approach are those in which unrelated species evolved a convergent phenotypic change after independent exposure to similar selective pressures. Convergent evolution provides a unique opportunity to study G2P relationships, because the existence of evolutionary replicates raises the possibility that genetic changes important to trait evolution can be identified because they are shared among the convergent lineages, winnowing the seemingly limitless potential pool of mutations to those selected to produce the shared phenotype. The PhyloG2P approach has proven a powerful framework for understanding species diversity, adaptation, convergent evolution, and G2P maps (Box 1). Many studies of both candidate genes, such as cardenolide resistance in insects (Zhen et al. 2012), and genome-wide analysis, such as marine adaptation in mammals (Chikina et al. 2016) or high altitude adaptation in alpine plants (Zhang et al.

2023), have revealed convergent genetic changes potentially associated with phenotypic evolution. Convergent evolution is also evident at very deep timescales, such as the independent primary acquisitions of photosynthesis in red/green algae and Rhizaria (Howe and Nisbet 2023; Johnson et al. 2023).

Despite these successes, the nature of the problem poses certain challenges. First, not all convergent phenotypic change proceeds via similar genetic mechanisms (Fig. 1B; e.g., the independent evolution of antifreeze proteins (Rives et al. 2024)); furthermore, even in the face of largely convergent genetic mechanisms, certain aspects of the phenotypic responses of individual lineages will be idiosyncratic and/or plastic. Adaptation can proceed via different genetic mechanisms to produce a similar phenotypic change, meaning that PhyloG2P approaches will not be able to detect convergent genetic signatures.

Second, the number of genetic changes that separate species is vast, and with a huge fraction resulting from fixation of neutral mutations, making it difficult to pinpoint functionally relevant signals against the phylogenetic background (Rockman 2012). Finally, in some cases, clear signals of repeated evolutionary change can help pinpoint a particular genomic locus, but do not provide direct answers regarding the molecular mechanisms at work.

Here, we propose an expanded framework of the standard PhyloG2P approach that seeks to incorporate information about intermediate molecular traits that propagate a genetic change to a phenotypic change (Fig. 2). We argue that incorporating additional levels of mechanistic information between DNA sequence (the genotype, "G") and an organismal trait of interest (the phenotype, "P"), such as transcriptional variation, protein activities and modifications, metabolomics, or physiological responses, can help resolve some of the challenges discussed above. Consideration of intermediate molecular phenotypes can allow the determination of the levels at which convergence actually occurred, as well as help isolate sequence changes with plausible molecular routes to affect phenotype, and provide clues to the mechanism of action of sequence change. Advances over the past decade in high-throughput studies, and the revolution in genomics, have provided a vision of what such complete datasets could look like in a single species or in distantly related model species and make now an ideal time to revisit and extend the PhyloG2P framework. The ultimate goal is to find associations across the matrix from genotype through intermediate traits to phenotype. This allows mechanistic convergence to be detected either at the genotype level or any of the intermediate traits allowing for a functional understanding of the trait.

Below, we first describe the data that can be harnessed from these intermediate layers (Part 1) and why they will be highly informative (Part 2). We next discuss strategies that will enable us to work toward assembling and interrogating the PhyloG2P of life, including considerations of new methods in machine learning that allow for the prediction of missing intermediate phenotypes where such data cannot be reasonably collected experimentally (Part 3).

Our focus in this perspective is on eukaryotic organisms. While several studies use similar approaches for prokaryotic taxa (Sauer and Wang 2019; Konno and Iwasaki 2023; Ramoneda et al. 2023; Ramoneda et al. 2024), the pervasive horizontal gene transfer, reduced recombination and other characteristics of prokaryotes make careful consideration of prokaryotic PhyloG2P beyond the scope of this effort.

**Box 1: Current approaches to linking genotype to phenotype across the phylogeny**

Several studies have attempted to perform genotype to phenotype mapping across phylogenies. These can be grouped into three general groups of approaches:

*Studies that focus on candidate genes:* These studies often take a phylogenetic approach to studying variation in proteins known to be associated with a specific trait in model systems. Examples include using variation in hemoglobin or hypoxia inducible factors (HIFs) for high altitude adaptation (Projecto-Garcia et al. 2013), lysozymes in host defense and folivore digestion (Messier and Stewart 1997), flowering traits related to pollinator shifts (Wessinger et al. 2023), and RNAses (Zhang et al. 2002).

*Studies that rely on association:* Several studies have successfully used genomic variation to find the genetic basis of phenotypic traits. These include PhyloGWAS for finding the genetic basis of adaptive traits including red fruit color in tomatoes (Pease et al. 2016), PhyloACC for studying the evolution of flightlessness (Sackton et al. 2019), and gene evolution rate correlation metrics for studying the mammalian transition to a marine environment (Chikina et al. 2016).

*Broader PhyloG2P:* Several methods incorporate the phylogenetic data with other evolutionary data to link genotype to phenotype. These include methods that utilize convergent amino acid changes, correlation in branch lengths, and repeated loss or gain of gene copy number (Fukushima and Pollock 2023; Macdonald et al. 2025).

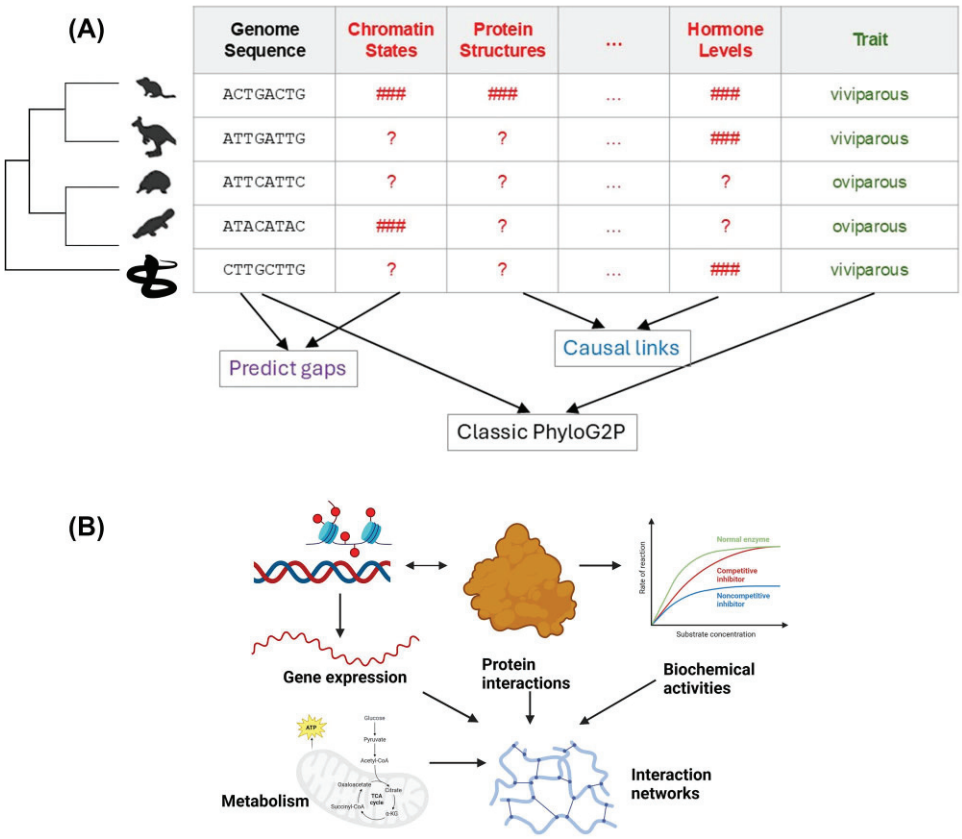



**Fig. 2** Integrating Intermediate Traits into PhyloG2P. (A) Conceptual framework illustrating how genomic sequences are linked to reproductive traits (viviparous vs. oviparous) through various intermediate traits, including chromatin states, protein structures, hormone levels, and other molecular or physiological data, which are the components of the expanded PhyloG2P matrix. Classic PhyloG2P only considers the first and last columns, but by expanding the matrix to include intermediate traits, we can predict gaps and identify causal links via imputation. ###, measured values; ?, unknown or missing data. (B) Examples of diverse intermediate data types spanning the genotype to phenotype across different organisms, highlighting the complexity of trait evolution and the role of multi-layered biological interactions in shaping phenotypic outcomes.

## Part 1: What are the intermediate data types?

We argue that incorporating additional molecular information across species (intermediate data) can extend traditional PhyloG2P analyses in important ways. Intermediate data types (Fig. 2) generally refer to the molecular and biological data that fall between raw genome sequencing data and organismal phenotypes (Hawkins et al. 2010). At the DNA level, information on the organization of 3-dimensional genomes from HiC, chromatin immunoprecipitation sequencing (CHIP-seq) data on DNA and protein interactions, and epigenetic modification of DNA can all provide a better understanding of genomic architecture and regulation (Rakyan et al. 2011; Li-Byarlay et al. 2013; Valencia and Kadoch 2019; Tu et al. 2020).

The transcription process is complex with multiple regulatory levels involving messenger RNAs, transcription factors, RNA modifications, splicing, non-coding RNAs, and other related factors (Li-Byarlay et al. 2020; Poliseno et al. 2024). At the RNA level, more data from transcriptomics (including single-cell sequencing and various forms of RNA sequencing), Assay for Transposase Accessible Chromatin Sequencing (ATAC-seq), and other new sequencing tools reveal the dynamics of the molecular process of transcription and link genotype to phenotype.

Proteomic data provide information on translational regulation, protein structure (empirical or predicted by AlphaFold), biochemical activity, protein docking $\Delta\Delta G$, protein-protein interaction networks, surface charge, and other related factors (Zhao et al. 2024). In addition to proteins, data from cellular and physiological levels, such as metabolomics, lipidomics, and hormonal levels can also be important to understanding the metabolism of molecules (Li et al. 2010).

While this is not intended as an exhaustive list, we highlight these examples to illustrate the complexity of molecular traits that are both influenced by genotype and, in turn, serve as the molecular building blocks of organismal phenotypes. We emphasize here that these

traits have a long history of being studied in a phylogenetic context (Brawand et al. 2011; Rohlfs et al. 2014; Villar et al. 2015; Hoencamp et al. 2021), and indeed as we discuss in the next section, a number of powerful examples illustrate the deep understanding that arises from combining the expertise of organismal and evolutionary biologists, molecular biologists, and physiologists to study trait evolution from a variety of perspectives. What we argue here is that as both experimental approaches and machine learning techniques advance, we have the opportunity to systematically leverage these intermediate data for complex phylogenetic modeling of genotype and phenotype that includes intermediate traits as covariates, such as through hierarchical models (Hopkins and St. John 2021; Powell et al. 2022).

## Part 2: Examples that show the power of integrating across data types

In many studies of evolutionary adaptations, elucidating intermediate functional changes arising from putatively adaptive genetic changes is a critical step in demonstrating the functional importance of an identified change.

For example, the integration of evolutionary biology and biochemistry has elucidated the role of hemoglobin and high altitude adaptation (Storz and Moriyama 2008; Simonson et al. 2010; Projecto-Garcia et al. 2013; Tufts et al. 2015; Storz 2021). Signore et al. (2019) used comparative genomics and physiological and biochemical measurements to study how Tibetan mastiffs evolved adaptations to high altitudes. Phylogenetic analysis and ancestral state reconstruction revealed that adaptation was the result of an initial ectopic gene conversion event in Tibetan wolves, followed by adaptive introgression into mastiffs. The key is that these evolutionary signatures were followed up by tests of intermediate phenotypes: oxygen saturation and the Bohr effect using purified hemoglobin protein. Opsins are another molecule for which our understanding has benefited mightily from the combined study of phylogenetic analysis with intermediate biochemical phenotypes. In this system, actual measurements of light absorption by opsins, or predictions made by trained models, elevated the findings of these studies beyond associating sequence evolution with phenotypic evolution (Hagen et al. 2023).

Intermediate phenotypes can be a powerful filter to highlight genetic changes most likely to be associated with an organismal trait. In an elegant recent study, Moreno et al. (2024)examined the convergent evolution of gliding membranes in marsupials. Screening for accelerated evolution of putative regulatory regions revealed more than a thousand candidate genes potentially enriched for glider-accelerated regulatory sequence. However, intersecting with functional data—in this case, RNA-seq expression data from the critical tissue—narrowed the list and allowed the authors to identify regulatory evolution around *Emx2* as a key functional modulator of gliding membrane development. Another illustrative example is the case of the *cortex* protein-coding gene in butterflies, long thought to be the mechanism of adaptive evolution of butterfly melanism. However, two recent studies that examined long noncoding RNA elements (an often neglected intermediate trait), have shown that melanism is not due to protein-coding variation, but is rather due to a long noncoding RNA inside the *cortex* locus (Fandino et al. 2024; Livraghi et al. 2024). Other classic examples include the evolution of pelvic reduction in sticklebacks (Shapiro et al. 2004), wing spots in *Drosophila* (Werner et al. 2010), photosynthesis in ciliates (Johnson et al. 2023) and digit loss in ungulates (Cooper et al. 2014).

While deciphering the genotype to phenotype map for behavior is likely to be particularly complex, in large part due to the complexity of the phenotype itself, we argue this may make the intermediate data types we advocate for particularly powerful. An example that demonstrates the utility and complexity of the problem is Wirthlin et al. (2024) study on the ability to produce learning vocal output in mammals, i.e., vocal learning. Previous studies presented evidence of convergence between vocal learning species at anatomical and gene expression levels, two intermediate traits (Jarvis 2004; Pfenning et al. 2014). With those insights, Wirthlin identified a new neuroanatomic region associated with vocal learning in bats, and experimentally determined its regions of open chromatin. Within those regions they inferred a set of regulatory regions associated with vocal learning by examining which showed sequence-level change associated with the vocal learning trait over four separate lineages. Thus, combining data across intermediate levels was crucial to identify a tractable set of regions to characterize, because the pool of potential regulatory regions would have otherwise been prohibitively large.

Traditional lab models (*Drosophila*, *Caenorhabditis*, yeasts, *Arabidopsis*) may be the place to start with these more integrated approaches since they are easily reared, inexpensive, and already have genetic tools and data for a wide range of species. Harrison et al. (2024) demonstrated the power of the approach in yeasts, which led to the identification of an alternative galactose utilization pathway. They employed genomic, metabolomic, and environmental data from more than 1000 yeast species grown in more than 100 conditions to train a machine learning algorithm to predict growth on different carbon sources. Importantly, while genomic

and metabolomic data both were able to train effective models on their own, the combination of genomic and metabolomic data provided the highest level of accuracy.

While this progress in traditional models is promising, the real goal is to develop these approaches for *any* species so that the above factors are less limiting. For example, the Zoonomia project aims to study the molecular basis of traits in mammals (Christmas et al. 2023; Kaplow et al. 2023), and incorporating intermediate traits into PhyloG2P analyses would accelerate discovery in this ambitious project. In the next section, we explore some possibilities for how the community could link genotype to phenotype in species that do not lend themselves to lab manipulation.

## Part 3: How do we complete the PhyloG2P matrix across the tree of life?

A complete matrix, in which intermediate phenotypes are annotated across diverse species (Fig. 2A), has the potential to greatly facilitate the linkage of sequence changes across a phylogeny to a phenotype of interest. In particular, isolating intermediate phenotypes that show patterns of molecular convergence congruent with the phenotype of interest can isolate biologically meaningful changes from evolutionary noise. A complete matrix can also potentially enable the identification of causal changes at intermediate levels, such as protein structure, transcript levels, or pathway flux, and highlight which levels of organization show signatures of convergence for a given trait.

However, in practice, a complete matrix of relevant intermediate molecular phenotypes across a diverse range of species with relevant convergent phenotypes is rarely, if ever, possible. If we consider the proposed matrix described by rows representing species and columns representing intermediate traits (Fig. 2), it may be possible to fill in nearly every column for some selected rows (e.g., for model groups that are amenable to laboratory studies, such as *Drosophila*, *Saccharomyces*, *Caenorhabditis*, *Arabidopsis*, or *Mus*), but many columns (traits) are inaccessible to wide swaths of the tree of life. For example, molecular intermediates that require laboratory manipulations (e.g., knockout experiments, time course transcriptional data, developmental traits) or require targeted reagents (ChIP-seq where antibodies to transcription factors are required) may never be feasible to generate (at least in a cost-effective manner) for the vast majority of species. While improvements in laboratory techniques can increase the accessibility of certain intermediate phenotypes (e.g., Box 2 on ATAC-seq), it is unlikely that experimental improvements alone will allow the rapid quantification of diverse intermediate phenotypes across the entire tree of life. Indeed, many species are endangered or otherwise nearly impossible to study across their life cycle.

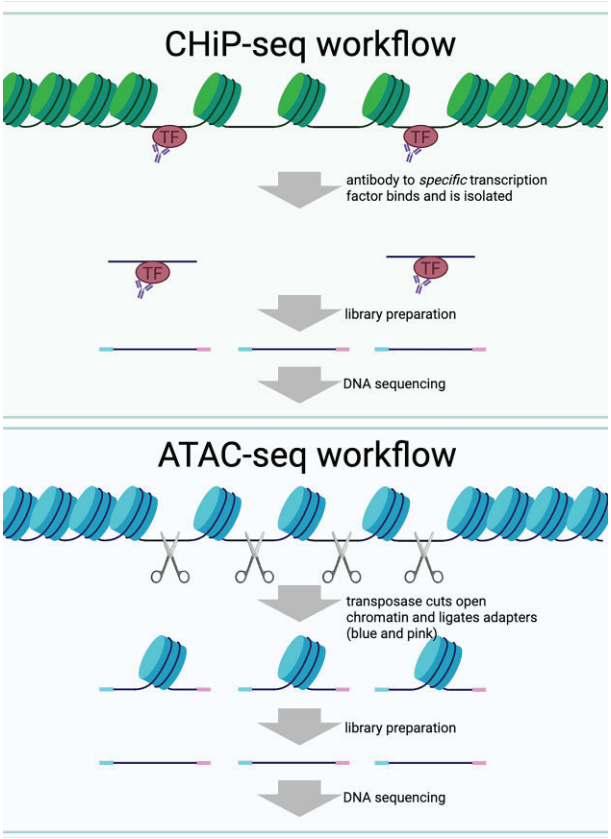
An alternative approach that is increasingly powerful is to rely on imputation methods to infer gaps in the proposed PhyloG2P matrix. Traditionally, imputation refers to predicting missing data from correlations across available data; for example, a common technique in population genetics is to impute missing genotype calls from patterns of linkage across individuals (Li et al. 2009; Das et al. 2016). However, this style of imputation is fundamentally limited as it can only infer linkages that already exist in the data structure (e.g., by using linkage disequilibrium to impute missing SNPs in a genotype matrix).

More recently, advances in machine learning have led to a flurry of prediction efforts (Song et al. 2020), where the goal is to predict a missing data type from an observed data type, often relying on training data from model organisms. These approaches have the potential to allow for prediction in a way that goes far beyond statistical imputation by learning an underlying generative function (Benegas et al. 2025). Several recent methods have been developed to predict intermediate phenotypes, including gene expression (Avsec et al. 2021; Lal et al. 2024), 3D genome organization (Brand et al. 2024; Gilbertson et al. 2024), and regions of open chromatin/putative enhancers (Kaplow et al. 2022; Kaplow et al. 2023), typically using input data that consists solely of DNA sequence. However, these approaches typically [although not always, e.g., TACIT (Huynh et al. 2024)] rely on single-species data. Approaches that predict missing data from existing data types but also leverage phylogenetic and cross-species information are likely to be much more accurate and powerful than single-species approaches alone. This kind of imputation-adjacent approach has the potential, therefore, to be a powerful way to fill gaps in the PhyloG2P matrix that are not accessible experimentally. By relying on intermediate phenotypes from accessible model organism data, we envision leveraging any data available (usually sequence) across the tree of life.

A particularly powerful demonstration of the impact of these imputation methods comes from AlphaFold and related methods (Jumper et al. 2021; Abramson et al. 2024) that provide an accurate way to predict protein structure, and, potentially other features such as post-transcriptional modifications (Shrestha et al. 2024), from primary sequence alone. The potential implications of this advance for understanding adaptation and the G2P map are just coming into focus. However, the development of protein language models (e.g., Pro-

**Box 2: ATAC-seq as an example of a lab technique revolution**

Detecting regions of the genome that function as regulatory elements is a longstanding, and difficult, problem in biology (Macdonald and Long 2005; The ENCODE Project Consortium 2007; Kellis et al. 2014). One successful molecular approach is to identify where specific transcription factors actually bind to DNA using variations of Chromatin Immunoprecipitation Sequencing (ChIP-seq) (Park 2009). In a typical ChIP-seq experiment, fragmented and fixed DNA is incubated with an antibody specific to a target transcription factor, and then the DNA fragments recovered are sequenced. While powerful, this approach is quite limited, in particular in that it requires an antibody targeted to the specific transcription factor of interest or introduction of a transgene that tags the transcription factor with a common epitope. While antibodies are commercially available for some species and some are cross-reactive, in many other cases, this limitation may be a costly and significant challenge to overcome (e.g., creating new antibodies by synthesizing peptides that represent epitopes of the protein of interest, immunizing animals, etc.). This becomes more cost-prohibitive as the number of proteins of interest increases. Transgenic techniques to add an epitope are similarly unavailable in most species. An alternative approach, called Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq), relies on the fact that regions of the genome where transcription factors bind must be accessible to protein (Grandi et al. 2022). In other words, they must be in regions of open chromatin, which means that they are also accessible to cutting by transposases. By measuring this cut rate approximated by sequence coverage, it is possible to detect regions of open chromatin in any system for which it is possible to harvest cells from target tissues. Recent advances, such as Omni-ATAC, provide flexibility in working with flash-frozen or preserved cells, raising the possibility of conducting ATAC-seq experiments even on field-collected samples (Corces et al. 2017). However, even with more flexible lab protocols, some fundamental requirements are hard to avoid (e.g., the ability to collect cells—live or flash frozen—from target tissues, development stages, etc.).




Gen, ESM3 (Madani et al. 2023; Hayes et al. 2025)) demonstrates the potential for functional convergence across a wide range of sequence space, by generating artificial proteins that share functions with known proteins despite very low sequence identity (Barrio-Hernandez et al. 2023). These protein language models also raise the possibility of being able to predict functional properties of proteins, not just structure, from se-

quence alone (e.g., inputting a primary sequence and outputting Vmax), perhaps with relatively low amounts of functional training data (Bhatnagar et al. 2025).

Nonetheless, it is important to raise a cautionary point that some forms of missing data prediction have the potential to be circular or not perform well outside the training context (Sasse et al. 2023). That is, the expectation of convergence (e.g., a similar G2P map in different species) allows imputation of missing data (e.g., expression levels under conditions that cannot be measured in the target species), but this necessary shortcut may lead to false imputation of similarities where none exist. This is analogous to a familiar problem for biologists—overfitting. A potential way to move forward under these conditions is to implement cycles of prediction, testing, and assessment (predict-test-learn cycles), instead of linear prediction and testing paths. Borrowing from the design-test-learn framework deployed by engineers, the crucial assessment step depends on acquiring novel experimental data to test the predictions made, such as for one or more new species or where the initial predictions were incorrect (Harrison et al. 2024). Relatedly, these predict-test-learn cycles can inform which data are most informative for inferring specific intermediate phenotypes across a clade, allowing researchers to prioritize the data for future collection. Thus, we envision each predict-test-learn cycle as filling in and correcting the PhyloG2P matrix, as well as driving G2P research and discovery forward.

## Conclusions

We have proposed a matrix approach to PhyloG2P studies that incorporates several intermediate data types and imputes relevant missing data: the PhyloG2P matrix. The approach starts with a standard PhyloG2P question, determining the genotypic basis of a phenotype, but then adds the likely relevant intermediate data types (e.g., gene expression, chromatin accessibility, proteomics, metabolomics) to better link genotype to phenotype. Machine learning approaches are crucial to both imputing missing data and predicting phenotypes from the data matrix. Machine learning approaches may also inform *what* data types might be most informative. It is also important to note that the basis of these analyses is still rooted in phylogenetic comparative methods (such as Phylogenetic Generalized Least Squares [PGLS]) (Revell 2010; Symonds and Blomberg 2014). The overall goal is to identify genetic changes underlying phenotypes that have such a preponderance of evidence in the form of intermediate data that functional validation *may not* be necessary. However, we acknowledge that prediction will be poor at first and therefore advocate for predict-test-learn cycles to refine the approach; many predictive models in machine learning can improve rapidly with relatively few training cases [e.g., (Bhatnagar et al. 2025)]. Additionally, it is important to note that advancements in intermediate data types (like those discussed for ATAC-seq) will continue to come for the foreseeable future. These data types will make the PhyloG2P matrix approach even more powerful and accessible for a wide range of species.

We note several caveats to the approach. First, while data collection costs continue to decline, we are advocating for massive amounts of data which will still be cost prohibitive in many cases. Second, we acknowledge that for some traits (particularly commercially important traits associated with crops and livestock), the ability to perform genetic selection might be sufficient and a detailed understanding of those traits superfluous. However, we argue that an understanding of the genetic basis of *any* trait does no harm to the commercial concerns and does move science forward. Third, many traits will covary meaning that the causal phenotype may be difficult to pin down. In this case, covarying traits could also be included in the matrix, but this may not help determine causality. Fourth, investigation of episodes of convergent evolution across vast temporal and physical scales means that study designs will greatly differ and will need to be fitted to the traits of interest. Finally, we have purposefully neglected methodological details. This is partly because it would be well beyond the scope of this perspective and partly because they are yet to exist. Working through the problem with model systems like yeast may be the best way to move forward and develop the requisite methods through relatively rapid learning cycles.

## Author contributions

N.L.C., C.T.H., H.L., A.R., T.B.S., and R.L.U. conceived, planned, wrote, and revised the paper.

## Acknowledgments

We thank two reviewers for helpful comments and the National Science Foundation LIFE (Leveraging Innovations from Evolution) Workshop for inspiring the work.

## Funding

Research in the Hittinger Lab is supported by the National Science Foundation (DEB-2110403), USDA National Institute of Food and Agriculture (Hatch Project 7005101), in part by the Department of Energy Great Lakes Bioenergy Research Center (DOE BER Office of Science DE–SC0018409), and an H.I. Romnes Faculty

Fellowship (Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation). Research in the Rokas lab is supported by the National Science Foundation (DEB-2110404) and the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R01-AI153356). Research in the Unckless Lab is supported by the National Science Foundation (MCB-2047052, DEB-2330095, and IOS-2421689). Research in the Li-Byarlay Lab is supported by the Evans-Allen fund project award no. NI241445XXXXG004, award 2021-38821-34576 from the US Department of Agriculture's National Institute of Food and Agriculture, and National Science Foundation award 1900793. Research in the Clark lab is supported by grants from the National Institutes of Health (R01-HG009299 and R01-EY030546). Research in the Sackton lab is supported by the National Science Foundation (DEB-1754397) and the National Institutes of Health (R01-HG011485).

## Conflict of interest

A.R. is a scientific consultant for LifeMine Therapeutics, Inc. All other authors declare that there are no conflicts of interest. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and should not be construed to represent any official USDA or U.S. Government determination or policy.

## Data availability

There is no data associated with this manuscript.

## References

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630:493–500.

Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, Assael Y, Jumper J, Kohli P, Kelley DR. 2021. Effective gene expression prediction from sequence by integrating long-range interactions. Nat Methods 18:1196–203.

Barrio-Hernandez I, Yeo J, Jänes J, Mirdita M, Gilchrist CLM, Wein T, Varadi M, Velankar S, Beltrao P, Steinegger M. 2023. Clustering predicted structures at the scale of the known protein universe. Nature 622:637–45.

Benegas G, Ye C, Albors C, Li JC, Song YS. 2025. Genomic language models: opportunities and challenges. Trends Genet 41:286–302.

Bhatnagar A, Jain S, Beazer J, Curran SC, Hoffnagle AM, Ching K, Martyn M, Nayfach S, Ruffolo JA, Madani A. 2025. Scaling Unlocks Broader Generation and Deeper Functional Understanding of Proteins. bioRxiv https://doi.org/10.1101/2025.04.15.649055.

Brand CM, Kuang S, Gilbertson EN, McArthur E, Pollard KS, Webster TH, Capra JA. 2024. Sequence-based machine learning reveals 3D genome differences between Bonobos and Chimpanzees. Genome Biol Evol 16:evae210.

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M et al. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343–8.

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 562:203–9.

Chikina M, Robinson JD, Clark NL. 2016. Hundreds of genes experienced convergent shifts in selective pressure in marine mammals. Mol Biol Evol 33:2182–92.

Christmas MJ, Kaplow IM, Genereux DP, Dong MX, Hughes GM, Li X, Sullivan PF, Hindle AG, Andrews G, Armstrong JC et al. 2023. Evolutionary constraint and innovation across hundreds of placental mammals. Science 380:eabn3943.

Conway Morris S. 2007. Life's Solution: Inevitable Humans in a Lonely Universe. Cambridge: Cambridge University Press.

Cooper KL, Sears KE, Uygur A, Maier J, Baczkowski KS, Brosnahan M, Antczak D, Skidmore JA, Tabin CJ. 2014. Patterning and post-patterning modes of evolutionary digit loss in mammals. Nature 511:41–5.

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B et al. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14:959–62.

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, De Los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y et al. 2017. Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci 22:961–75.

Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M et al. 2016. Next-generation genotype imputation service and methods. Nat Genet 48:1284–7.

Fandino RA, Brady NK, Chatterjee M, McDonald JMC, Livraghi L, van der Burg KRL, Mazo-Vargas A, Markenscoff-Papadimitriou E, Reed RD. 2024. The *ivory* lncRNA regulates seasonal color patterns in buckeye butterflies. Proc Natl Acad Sci USA 121:e2403426121.

Fukushima K, Pollock DD. 2023. Detecting macroevolutionary genotype–phenotype associations using error-corrected rates of protein convergence. Nat Ecol Evol 7:155–70.

Gilbertson EN, Brand CM, McArthur E, Rinker DC, Kuang S, Pollard KS, Capra JA. 2024. Machine learning reveals the diversity of Human 3D chromatin contact patterns. Mol Biol Evol 41:msae209.

Gonçalves C, Harrison M-C, Steenwyk JL, Opulente DA, LaBella AL, Wolters JF, Zhou X, Shen X-X, Groenewald M, Hittinger CT et al. 2024. Diverse signatures of convergent evolution in cactus-associated yeasts. PLoS Biol 22:e3002832.

Grandi FC, Modi H, Kampman L, Corces MR. 2022. Chromatin accessibility profiling by ATAC-seq. Nat Protoc 17:1518–52.

Hagen JFD, Roberts NS, Johnston RJ. 2023. The evolutionary history and spectral tuning of vertebrate visual opsins. Dev Biol 493:40–66.

Harrison M-C, Ubbelohde EJ, LaBella AL, Opulente DA, Wolters JF, Zhou X, Shen X-X, Groenewald M, Hittinger CT, Rokas A. 2024. Machine learning enables identification of an alternative

yeast galactose utilization pathway. Proc Natl Acad Sci USA 121:e2315314121.
Hawkins RD, Hon GC, Ren B. 2010. Next-generation genomics: an integrative approach. Nat Rev Genet 11:476–86.
Hayes BJ, Mahony TJ, Villiers K, Warburton C, Kemper KE, Dinglasan E, Robinson H, Powell O, Voss-Fels K, Godwin ID et al. 2024. Potential approaches to create ultimate genotypes in crops and livestock. Nat Genet 56:2310–7.
Hayes T, Rao R, Akin H, Sofroniew NJ, Oktay D, Lin Z, Verkuil R, Tran VQ, Deaton J, Wiggert M et al. 2025. Simulating 500 million years of evolution with a language model. Science 387:850–8.
Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, Van Schaik T, Sedeño Cacciatore Á, Contessoto VG, Van Heesbeen RGHP, Van Den Broek B et al. 2021. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. Science 372:984–9.
Hopkins MJ, St. John K. 2021. Incorporating hierarchical characters into phylogenetic analysis.Wright A, editor. Syst. Biol. 70:1163–80.
Howe CJ, Nisbet RER. 2023. Evolution: the great photosynthesis heist. Curr Biol 33:R185–7.
Huynh KLA, Tyc KM, Matuck BF, Easter QT, Pratapa A, Kumar NV, Pérez P, Kulchar R, Pranzatelli T, De Souza D et al. 2024. Spatial Deconvolution of Cell Types and Cell States at Scale Utilizing TACIT.bioRxiv https://doi.org/10.1101/2024.05.31.596861.
Jarvis ED. 2004. Learned birdsong and the neurobiology of Human language. Ann NY Acad Sci 1016:749–77.
Johnson MD, Moeller HV, Paight C, Kellogg RM, McIlvin MR, Saito MA, 2023. Lasek-Nesselquist E. 2023. Functional control and metabolic integration of stolen organelles in a photosynthetic ciliate. Curr Biol 33:973–980.e5.
Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583–9.
Kaplow IM, Lawler AJ, Schäffer DE, Srinivasan C, Sestili HH, Wirthlin ME, Phan BN, Prasad K, Brown AR,Zhang X et al. 2023. Relating enhancer genetic variation across mammals to complex phenotypes using machine learning. Science 380:eabm7993.
Kaplow IM, Schäffer DE, Wirthlin ME, Lawler AJ, Brown AR, Kleyman M, Pfenning AR. 2022. Inferring mammalian tissue-specific regulatory conservation by predicting tissue-specific differences in open chromatin. BMC Genomics [Electronic Resource] 23:291.
Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J et al. 2014. Defining functional DNA elements in the human genome. Proc Natl Acad Sci USA 111:6131–8.
Konno N, Iwasaki W. 2023. Machine learning enables prediction of metabolic system evolution in bacteria. Sci Adv 9:eadc9130.
Lal A, Karollus A, Gunsalus L, Garfield D, Nair S, Tseng AM, Gordon MG, Blischak JD, Van De Geijn B, Bhangale T et al. 2024. Decoding Sequence Determinants of Gene Expression in Diverse Cellular and Disease States.bioRxiv https://doi.org/10.1101/2024.10.09.617507.
Li H-M, Sun L, Mittapalli O, Muir WM, Xie J, Wu J, Schemerhorn BJ, Jannasch A, Chen JY, Zhang F et al. 2010. Bowman-Birk inhibitor affects pathways associated with energy metabolism in *Drosophila melanogaster*. Insect Mol Biol 19:303–13.
Li Y, Willer C, Sanna S, Abecasis G. 2009. Genotype imputation. Annu Rev Genom Hum Genet 10:387–406.
Li-Byarlay H, Boncristiani H, Howell G, Herman J, Clark L, Strand MK, Tarpy D, Rueppell O. 2020. Transcriptomic and epigenomic dynamics of honey bees in response to lethal viral infection. Front Genet 11:566320.
Li-Byarlay H, Li Y, Stroud H, Feng S, Newman TC, Kaneda M, Hou KK, Worley KC, Elsik CG, Wickline SA et al. 2013. RNA interference knockdown of *DNA methyl-transferase 3* affects gene alternative splicing in the honey bee. Proc Natl Acad Sci USA 110:12750–5.
Livraghi L, Hanly JJ, Evans E, Wright CJ, Loh LS, Mazo-Vargas A, Kamrava K, Carter A, van der Heijden ESM, Reed RD et al. 2024. A long noncoding RNA at the *cortex* locus controls adaptive coloration in butterflies. Proc Natl Acad Sci US. 121:e2403326121.
Luth MR, Godinez-Macias KP, Chen D, Okombo J, Thathy V, Cheng X, Daggupati S, Davies H, Dhingra SK, Economy JM et al. 2024. Systematic in vitro evolution in *Plasmodium falciparum* reveals key determinants of drug resistance. Science 386:eadk9893.
Macdonald AR, James ME, Mitchell JD, Holland BR. 2025. From Trees to Traits: a Review of Advances in PhyloG2P Methods and Future Directions. arXiv https://doi.org/10.48550/arXiv.2501.07043
Macdonald SJ, Long AD. 2005. Prospects for identifying functional variation across the genome. Proc Natl Acad Sci USA 102:6614–21.
Madani A, Krause B, Greene ER, Subramanian S, Mohr BP, Holton JM, Olmos JL, Xiong C, Sun ZZ, Socher R et al. 2023. Large language models generate functional protein sequences across diverse families. Nat Biotechnol 41: 1099–106.
Messier W, Stewart C-B. 1997. Episodic adaptive evolution of primate lysozymes. Nature 385:151–4.
Moreno, J.A., Dudchenko, O., Feigin, C.Y., Mereby S.A., Chen Z., Ramos R., Almet A.A., Sen H., Brack B.J., Johnson M.R. et al. 2024. Emx2 underlies the development and evolution of marsupial gliding membranes. Nature 629:127–35.
Park PJ. 2009. ChIP–seq: advantages and challenges of a maturing technology. Nat Rev Genet 10:669–80.
Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation.. PLoS Biol 14:e1002379.
Pfenning AR, Hara E, Whitney O, Rivas MV, Wang R, Roulhac PL, Howard JT, Wirthlin M, Lovell PV, Ganapathy G et al. 2014. Convergent transcriptional specializations in the brains of humans and song-learning birds. Science 346:1256846.
Poliseno L, Lanza M, Pandolfi PP. 2024. Coding, or non-coding, that is the question. Cell Res 34:609–29.
Powell OM, Barbier F, Voss-Fels KP, Beveridge C, Cooper M. 2022. Investigations into the emergent properties of gene-to-phenotype networks across cycles of selection: a case study of shoot branching in plants. Silico Plants 4:diac006.
Projecto-Garcia J, Natarajan C, Moriyama H, Weber RE, Fago A, Cheviron ZA, Dudley R, McGuire JA, Witt CC, Storz JF. 2013. Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. Proc Natl Acad Sci USA 110:20669–74.
Rakyan VK, Down TA, Balding DJ, Beck S. 2011. Epigenome-wide association studies for common human diseases. Nat Rev Genet 12:529–41.

Ramoneda J, Hoffert M, Stallard-Olivera E, Casamayor EO, Fierer N. 2024. Leveraging genomic information to predict environmental preferences of bacteria. ISME J 18:wrae195.
Ramoneda J, Stallard-Olivera E, Hoffert M, Winfrey CC, Stadler M, Niño-García JP, Fierer N. 2023. Building a genome-based understanding of bacterial pH preferences. Sci Adv 9:eadf8998.
Revell LJ. 2010. Phylogenetic signal and linear regression on species data: *p hylogenetic regression*. Methods Ecol Evol 1:319–29.
Rives N, Lamba V, Cheng CHC, Zhuang X. 2024. Diverse origins of near-identical antifreeze proteins in unrelated fish lineages provide insights into evolutionary mechanisms of new gene birth and protein sequence convergence. Mol Biol Evol 41:msae182.
Rockman MV. 2012. The Qtn program and The alleles that matter for evolution: all that's gold does not glitter. Evolution 66:1–17.
Rohlfs RV, Harrigan P, Nielsen R. 2014. Modeling gene expression evolution with an extended Ornstein–Uhlenbeck process accounting for within-species variation. Mol Biol Evol 31:201–11.
Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M et al. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. Science 364:74–8.
Sasse A, Ng B, Spiro AE, Tasaki S, Bennett DA, Gaiteri C, De Jager PL, Chikina M, Mostafavi S. 2023. Benchmarking of Deep Neural Networks for Predicting Personal Gene Expression from DNA Sequence Highlights Shortcomings.bioRxivhttps://doi.org/10.1101/2023.03.16.532969.
Sauer DB, Wang D-N. 2019. Predicting the optimal growth temperatures of prokaryotes using only genome derived features. Bioinformatics 35:3224–31.
Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D, Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428:717–23.
Shrestha P, Kandel J, Tayara H, Chong KT. 2024. Post-translational modification prediction via prompt-based fine-tuning of a GPT-2 model. Nat Commun 15:6699.
Signore AV, Yang Y-Z, Yang Q-Y, Qin G, Moriyama H, Ge R-L, Storz JF. 2019. Adaptive changes in hemoglobin function in high-altitude Tibetan canids were derived via gene conversion and introgression. Mol Biol Evol 36:2227–37.
Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB et al. 2010. Genetic evidence for high-altitude adaptation in Tibet. Science 329:72–5.
Smith SD, Pennell MW, Dunn CW, Edwards SV. 2020. Phylogenetics is the New genetics (for Most of Biodiversity). Trends Ecol Evol 35:415–25.
Song M, Greenbaum J, Luttrell J, Zhou W, Wu C, Shen H, Gong P, Zhang C, Deng H-W. 2020. A review of integrative imputation for Multi-omics datasets. Front Genet 11:570255.
Storz JF, Moriyama H. 2008. Mechanisms of hemoglobin adaptation to high altitude hypoxia. High Alt Med Biol 9:148–57.
Storz JF. 2021. High-altitude adaptation: mechanistic insights from integrated genomics and physiology. Mol Biol Evol 38:2677–91.
Symonds MRE, Blomberg SP. 2014. A primer on phylogenetic generalised least squares. In: Garamszegi LZ, editor. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Berlin, Heidelberg: Springer Berlin Heidelberg. p.105–30. Available from: http://link.springer.com/10.1007/978-3-662-43550-2_5.
Tan X, He Z, Fahey AG, Zhao G, Liu R, Wen J. 2023. Research progress and applications of genome-wide association study in farm animals. Ani Res One Health 1:56–77.
The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447:799–816.
Tibbs Cortes L, Zhang Z, Yu J. 2021. Status and prospects of genome-wide association studies in plants. Plant Genome 14:e20077.
Tu X, Mejía-Guerra MK, Valdes Franco JA, Tzeng D, Chu P-Y, Shen W, Wei Y, Dai X, Li P, Buckler ES et al. 2020. Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. Nat Commun 11:5089.
Tufts DM, Natarajan C, Revsbech IG, Projecto-Garcia J, Hoffmann FG, Weber RE, Fago A, Moriyama H, Storz JF. 2015. Epistasis constrains mutational pathways of hemoglobin adaptation in high-altitude pikas. Mol Biol Evol 32:287–98.
Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D. 2021. Genome-wide association studies. Nat Rev Methods Primers 1:59.
Valencia AM, Kadoch C. 2019. Chromatin regulatory mechanisms and therapeutic opportunities in cancer. Nat Cell Biol 21:152–61.
Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ et al. 2015. Enhancer evolution across 20 mammalian species. Cell 160:554–66.
Werner T, Koshikawa S, Williams TM, Carroll SB. 2010. Generation of a novel wing colour pattern by the Wingless morphogen. Nature 464:1143–8.
Wessinger CA, Katzer AM, Hime PM, Rausher MD, Kelly JK, Hileman LC. 2023. A few essential genetic loci distinguish penstemon species with flowers adapted to pollination by bees or hummingbirds. PLoS Biol 21:e3002294.
Wirthlin ME, Schmid TA, Elie JE, Zhang X, Kowalczyk A, Redlich R, Shvareva VA, Rakuljic A, Ji MB, Bhat NS et al. 2024. Vocal learning–associated convergent evolution in mammalian proteins and regulatory elements. Science 383:eabn3263.
Zhang J, Zhang Y, Rosenberg HF. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat Genet 30:411–5.
Zhang X, Kuang T, Dong W, Qian Z, Zhang H, Landis JB, Feng T, Li L, Sun Y, Huang J et al. 2023. Genomic convergence underlying high-altitude adaptation in alpine plants. JIPB 65:1620–35.
Zhao N, Wu T, Wang W, Zhang L, Gong X. 2024. Review and comparative analysis of methods and advancements in predicting protein complex structure. Interdiscip Sci Comput Life Sci 16:261–88.
Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012. Parallel molecular evolution in an herbivore community. Science 337:1634–7.