PCP
PLANT & CELL PHYSIOLOGY

Linking plant genes to arthropod community dynamics: current progress and future challenges

Matthew A. Barbour* and Cintia Beatriz Pérez-López

Département de biologie, Faculté des Sciences, Université de Sherbrooke, 2500 boul. de l'Université, Sherbrooke J1K 2R1, Canada

*Corresponding author: E-mail, matthew.barbour@usherbrooke.ca

Received 30 April 2024; Accepted 31 January 2025

Abstract

Plant genetic variation can play a key role in shaping ecological communities. Prior work investigated the effects of coarse-grain variation among plant genotypes on their diverse arthropod communities. Several recent studies, however, have leveraged the boom of genomic resources to study how genome-wide plant variation influences associated communities. These studies have demonstrated that the effects of plant genomic variation are not just detectable but can be important drivers of arthropod communities in natural ecosystems. Field common gardens and lab-based mesocosm experiments are also revealing candidate genes that have large effects on arthropod communities. While we highlight these exciting results, we also discuss key challenges to address in future research. We argue that a major hurdle lies in the integration of genomic tools with hierarchical models of species communities (HMSCs). HMSCs are generative models that provide the opportunity to not only better understand the processes underlying community change but to also predict community dynamics. We also advocate for future research to apply models of genomic prediction to explore the genetic architecture of arthropod community phenotypes. We hypothesize that this genetic architecture will follow an exponential distribution, where a few genes of large effect, but also many genes of small effect, contribute to variation in arthropod communities. The next generation of studies linking plant genes to community dynamics will require interdisciplinary collaborations to build truly predictive models of plant genetic and arthropod community change.

Keywords: community genetics; eco-evolutionary dynamics; community resilience; genome-wide association study; plant–insect interactions

Introduction

A grand challenge we face today lies in predicting biodiversity change—from genes to communities of interacting species. Plants comprise the majority of biomass on Earth (Bar-On et al. 2018) and form the base of diverse terrestrial food webs. Genetic variation provides the raw material for adaptive evolution (Reed and Frankham 2003) and thus the persistence of plant populations in the face of environmental change. As such, plant genetic variation fosters the sustainability of basal resources, which in turn permits the assembly of multi-trophic interactions. Moreover, genetic differences in the phenotypes of individual plants can alter ecological interactions with a diverse array of organisms (Whitham et al. 2012). Thus, a major pattern that has emerged in the past three decades is that plant genetic variation can substantially influence the diversity, composition, and multi-trophic structure of their associated communities (Maddox and Root 1987, Fritz and Price 1988, Crutsinger et al. 2006, Whitham et al. 2006, Keith et al. 2010, Barbour et al. 2016). Taken together, there is clear evidence that plant genetic variation can function as a critical lever that maintains biodiversity in terrestrial food webs.

Here, we review current progress in our understanding and ability to predict linkages between plant genes and community dynamics. We focus on studies in the past 6 years that have employed novel empirical tools to explore the impact of individual plant genes on arthropod communities. We focus on studies of arthropods because they represent the vast majority of described terrestrial biodiversity (Grimaldi and Engel 2005). These studies are showing the way forward, but we highlight challenges that need to be surmounted in the next generation of studies to move toward truly predictive models that link genetic change to community dynamics. We delve little into the application of post-genome-wide association study (GWAS) analyses as the available tools have been reviewed elsewhere (Mortezaei

Plant And Cell Physiology, 2025, 66(4), 506–513
doi:https://doi.org/10.1093/pcp/pcaf015
Advance Access publication on 1 February 2025

and Tavallaei 2021, Zhu and Zhou 2021), and we do not feel it is the current limiting factor in linking plant genes to community dynamics. Instead, we focus on approaches for better understanding and predicting variation in "community phenotypes" resulting from genetic changes.

Current Progress

Kagiya et al. (2018) were among the first to use genomic tools to study how plant genetic variation in natural ecosystems influences arthropod communities. They used restriction-site-associated DNA sequencing to characterize the genetic distance among 85 mature alder trees in a 25 000-ha mixedwood forest in Japan. For each of these alder trees, they surveyed the community of arthropods over a 2-year period. They found that genetic distance was consistently associated with arthropod community dissimilarity in both years. In other words, alders with more similar genomes hosted more similar arthropod communities. Importantly, this result held even after accounting for potential confounding factors such as spatial distance, surrounding tree community composition, and temporal differences between arthropod sampling. This work provided strong evidence for the genetic similarity rule (Bangert et al. 2006) and the role of plant genetic variation outside of a controlled common garden setting.

Studies in both the lab and field have used mutants or gene-silencing techniques to identify the causal role of plant architecture and defense signaling pathways on arthropod communities (Kareiva and Sahakian 1990, Kessler et al. 2004, Kallenbach et al. 2012, Northfield et al. 2012, Schuman et al. 2012, Adam et al. 2018); however, the community-level impact of natural variation at specific genes is often unclear (Skovmand et al. 2018). Sato et al. (2019) demonstrated the causal effect of natural variation at a single gene on herbivore communities in the field. They planted 17 phenotypically diverse accessions of the genetic model plant *Arabidopsis thaliana* in two common gardens. Importantly, two of the accessions (Col-0 and L*er*-1) were paired with their respective glabrous mutants (*gl1-1* and *gl1-2*) that had independent loss-of-function mutations at the transcription factor gene *GLABRA1*(*GL1*) that controls leaf trichome development. These two accession–mutant pairs permit a strong test of the causal effect of *GLABRA1* because of the different genetic backgrounds of each accession. They found that damage from leaf-chewing herbivores, and their abundance, was lower on plants with higher trichome density, but also on the two natural accessions compared to their glabrous mutants. While the effect of *GLABRA1* on herbivore diversity and community composition was unclear, this effect could emerge in a different environmental context (e.g. where the abundance of leaf chewers was naturally higher).

Recent large-scale GWAS have provided compelling evidence for the role of variation at individual genes in shaping insect community diversity and composition. For example, Barker et al. (2019) conducted a GWAS with a population of 445 trembling aspen (*Populus tremuloides*) genotypes in a common garden to identify the genes governing variation in plant traits and insect community composition. They identified several candidate genes associated with variation in the abundance or occurrence of several insect species. Notably, they found that variation at an early nodulin-like (ENODL) transmembrane protein (*Potra001060g09097*, *A. thaliana* homolog AT1G79800) was associated with variation in insect community composition. Specifically, they found that aspen genotypes with the C allele of this ENODL gene generally hosted more insect species, but were dominated by a leaf galler, aphids, and aphid-tending ants. This was arguably the first study to link allelic variation at a plant gene to change in insect community composition.

Similarly, Simon et al. (2024) employed a large-scale GWAS with hundreds of black cottonwood (*Populus trichocarpa*) genotypes from across their native range planted in two different common gardens. Similar to Barker et al. (2019), they found evidence for a number of candidate genes associated with variation in common leaf-galling and leaf-mining herbivores as well as the diversity and composition of the insect community. A notable feature of their work is that they integrated their GWAS of insect community phenotypes with multiple layers of publicly available -omics data [e.g. metabolites, gene coexpression, gene methylation, and single-nucleotide polymorphism (SNP) correlation] to reduce the risk of false positives and thus hone in on a more reliable set of candidate genes. This analysis treated each data layer as a line of evidence (LOE) (Weighill et al. 2018), such that candidate genes that were more highly connected received higher LOE scores, lending more confidence to the interpretation of their biological function.

The above work highlights the role of an individual host plant's genes in shaping multi-trophic interactions, but what about the genes of neighboring plants? Sato et al. (2021) recently developed a novel quantitative tool called NeighborGWAS, which is able to quantify the effect of genome-wide variation of neighboring plants on focal individuals. This approach has shown that the genomic identity of neighbors leaves a clear, but polygenic, signal on herbivory and even the number of species found on focal individuals (Sato et al. 2024). While the genomic identity of the focal individual appeared to be of primary importance, the genetic effect of neighboring plants was far from negligible. Indeed, Sato et al. (2024) combined NeighborGWAS and genomic prediction tools to identify pairs of genotypes where the neighboring plants were predicted to positively affect focal individuals by reducing herbivory. They confirmed their prediction in a subsequent year (i.e. different environmental conditions), proving how their method can be used to plant genetic mixtures that modify the strength of plant–herbivore interactions. Moreover, this analytical tool can also identify the relevant spatial scale at which to quantify neighbor effects (i.e. nearest neighbors, nearest + next closest neighbors). Importantly, this method can be readily applied to most common garden experiments, providing a valuable tool to fully explore the indirect effects of plant genomic variation.

Finally, Barbour et al. (2022) provided concrete evidence for a "keystone gene" that controlled the coexistence of insect species in a simple experimental food web. Using an array of mesocosms in climate chambers, they took advantage of three different mutant lines of *A. thaliana* in a common genetic background (Col-0 accession) that mimic natural functional and loss-of-function mutations in genes that control the biosynthesis of aliphatic glucosinolates in *Arabidopsis* and many other Brassicaceae (Katz et al. 2021). They found that allelic variation at *AOP2*, a gene that controls the side-chain modification of aliphatic glucosinolates, controlled the persistence of an aphid-parasitoid food chain. Importantly, this "keystone gene" effect was robust to experimental warming of 3°C.

Future Challenges

The last 6 years have ushered in exciting results about the potential linkages between plant genes and their associated communities; however, we still lack the ability to predict how genetic change in plant populations will affect the dynamics of their associated communities. In the following sections, we highlight what we think are the major challenges for predictive modeling of gene-to-community linkages and give advice on how to surmount these challenges in future work.

Hierarchical modeling of species communities

All of the reviewed field studies sought to test for a genetic effect on an aggregate index of an ecological community, such as species diversity (e.g. richness) or distance-based metric of composition (e.g. nonmetric multidimensional scaling axis from the ordination of Bray–Curtis dissimilarities). These aggregate indices may be useful for testing for evidence of gene-to-community linkages but can also result in misleading conclusions. For example, samples from individual trees usually contain <20 species, less than a hundred individuals total, and many zeros (Wimp et al. 2005, Barbour et al. 2015, Barker et al. 2018). This situation is exacerbated for smaller plants (Johnson and Agrawal 2005, Barbour et al. 2019, Sato et al. 2019). In these scenarios, metrics of species diversity will be primarily determined by the number of individuals sampled (Gotelli and Colwell 2001) and dissimilarity indices can exhibit erratic behavior (Beck et al. 2013, Hardersen and La Porta 2023), making interpretation difficult. Moreover, current distance-based metrics (e.g. Bray–Curtis dissimilarity) do not account for the mean–variance relationship that defines species count data, which can result in misleading conclusions (Warton et al. 2012). Another limitation of this approach is that it does not allow an exploration of the mechanisms underlying community change or predictive modeling (Warton et al. 2015, Ovaskainen et al. 2017). This is because aggregate indices are ultimately determined by changes in the abundance of multiple species that are often interacting, a fact that is ignored if an index of species diversity is modeled directly. It would be much more informative to directly model the changes in multiple species simultaneously, with the appropriate statistical distribution, to derive a prediction of community change.

Fortunately, we have entered a statistical era of community ecology where data on the abundances of many potentially interacting species are relatively straightforward to model and can incorporate a rich variety of biological information (e.g. functional traits and phylogeny) (Warton et al. 2015, Ovaskainen et al. 2017). These hierarchical models of species communities (HMSCs) provide a number of advantages over traditional approaches (Ovaskainen et al. 2017). For example, sparse data from rare species do not have to be discarded and can be naturally modeled (Ovaskainen and Soininen 2011). Appropriate mean–variance relationships for species abundances can be modeled and assumptions can be checked with diagnostic tools. Information on species traits (e.g. herbivore feeding guild) or phylogenetic relationships can be readily incorporated to provide mechanistic insight to drivers of community assembly. This modeling approach is generative, meaning that it generates predictions at the species, community, or trait levels, while naturally propagating uncertainty to the level of the prediction. Importantly, HMSCs can be applied to many different study designs (e.g. hierarchical, spatial, or temporal) and many types of data (e.g. presence–absence, counts, or continuous measurements). In fact, the large-scale common garden studies previously mentioned (Barker et al. 2019, Sato et al. 2024, Simon et al. 2024) already collect the type of community data required for integrating genomic information into HMSCs (i.e. species absolute/relative abundance or presence/absence data on genotyped plant individuals). Finally, HMSCs can account for species interactions in explaining and predicting changes in species distributions.

Another advantage of HMSCs is that they build from the mixed modeling framework that forms the basis of GWAS (Zhou and Stephens 2012). Thus, common statistical approaches for GWAS can be naturally incorporated into HMSC, although this has yet to be done in practice. For example, the genomic relationship matrix (GRM) among individual plants can be modeled as a random effect to provide an estimate of narrow-sense heritability based on SNPs (h^2_{SNP}) of species and community phenotypes (Stanton-Geddes et al. 2013, Speed et al. 2017). If one is willing to assume that only many genes of small effect underlie the phenotype of interest (i.e. an infinitesimal model, Barton et al. 2017), then this same model can be used for genomic prediction of genotypes that have not been phenotyped. Note however that the above method is mainly applicable for polygenic traits, but will systematically underestimate the heritability of phenotypes controlled primarily by large-effect genes (i.e. monogenic or oligogenic architecture) (Golan et al. 2014, Speed et al. 2017).

An exciting extension of HMSC is to incorporate information on changes in species abundances over time. This is a potentially powerful way to infer the interaction matrix *A* (Novak et al. 2016), which represents the direct interactions within and between species in an ecological community (matrix *A*, Fig. 1). The interaction matrix can be inferred by modeling how the log-abundance of a species at a previous time step (i.e. sampling date or year) affects its own log-abundance and the log-abundance of other species in the next time step (Ives et al.

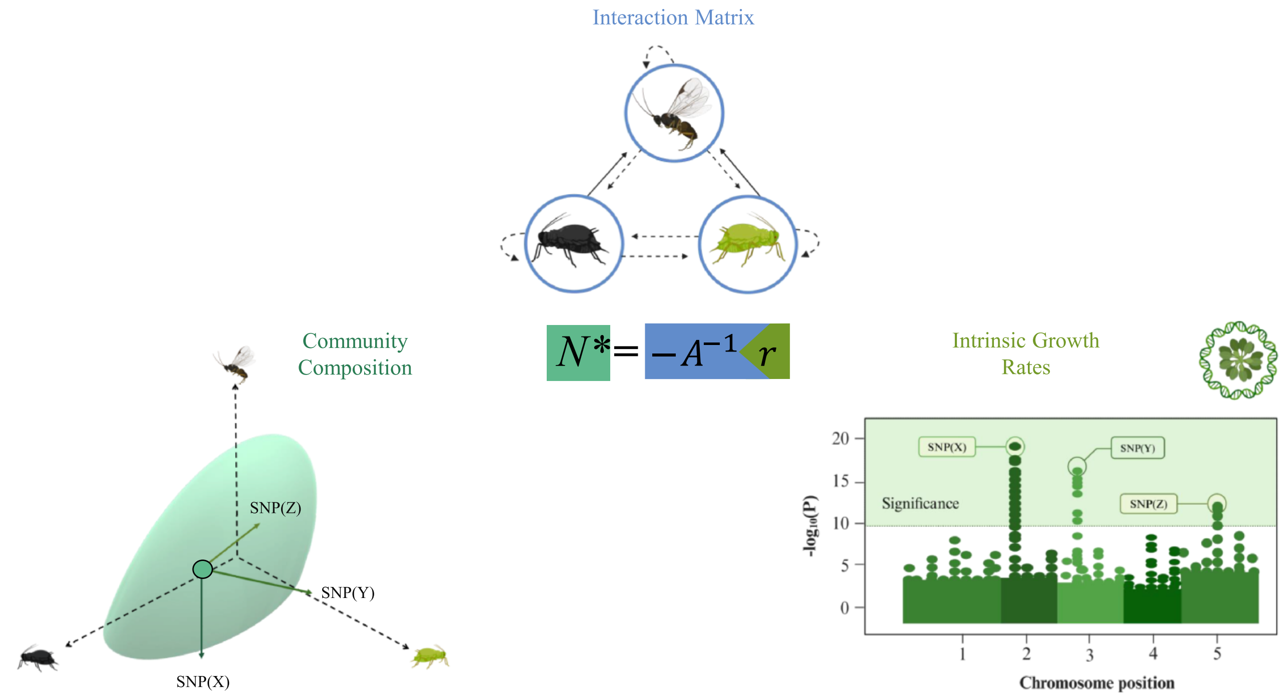

Figure 1. The interaction matrix (A) is arguably the most important theoretical abstraction in community ecology. This matrix captures the direction (positive or negative) and strength of ecological interactions within and between species in a community. The structure of the interaction matrix gives important insight to community dynamics, such as community resilience to pulse perturbations ($-\lambda_{\max}(A)$) and the net direct and indirect effects ($-A^{-1}$) of press perturbations on species equilibrium abundances (N^*), but by itself is incomplete. Another crucial piece of information lies in the vector of intrinsic growth rates r of species (i.e. their population growth rates at low density) in the community. Here, we conceptualize how plant genetic change (SNP X, Y, and Z) results in a press perturbation to herbivore intrinsic growth rates, which alters the composition of a multi-trophic community of two herbivores (aphids) and a parasitoid wasp.

2003). This multi-species Gompertz model also provides an estimate of species' intrinsic growth rates (i.e. population growth rate at low density, Fig. 1) (Ives et al. 2003). This estimated interaction matrix has a direct connection to theoretical metrics of community stability, such as community resilience to pulse perturbations (Ives et al. 2003) and the net change in community composition resulting from press perturbations (Barbour et al. 2022) (Fig. 1). An important advantage of this "inverse" approach to estimating the interaction matrix, rather than from direct observations of ecological interactions, is that interaction strengths are directly linked to fitness, which is rarely (if ever) the case for traditional ways of building species-interaction networks, which limits the ability of traditional networks to make meaningful predictions. Statistical modeling of the interaction matrix is challenging, especially in natural environments, but can be aided by incorporating natural history observations. For example, direct interactions between species that are implausible (e.g. an aphid parasitoid attacking a caterpillar) or thought to be negligible can be constrained to zero before model fitting (Ives et al. 2003).

Estimating individual interaction matrices for a diverse association panel of plants is likely impossible in practice. However, it is certainly possible to conduct repeated surveys on a manageable number of plant individuals to quantify the interaction matrix and species' intrinsic growth rates in a given location (e.g. common garden). With this interaction matrix, one can calculate the net-effects matrix ($-A^{-1}$), which represents the net direct and indirect effects of species on each other (Fig. 1), from which it is possible to identify putative keystone species or guilds in the arthropod community. This can be done by summing the direct and indirect effects of species on each other and ranking species based on their net effects (described in Berg et al. 2011). Alternative approaches for identifying keystones are an active topic of research, especially for microbial community time series (Berry and Widder 2014, Fisher et al. 2014, Amit and Bashan 2023, Wang et al. 2024). With this information in hand, one could do a targeted GWAS of plant genetic effects on the keystone species of the community. The GWAS will give insight to how plant genetic change alters the intrinsic growth rate of the "keystone species," which in turn gives insight to how community composition will change (Fig. 1). This is similar to the approach used by Barbour et al. (2022) to identify a keystone gene that controlled the coexistence of a simple experimental food web. Although the experiment of Barbour et al. (2022) was conducted in a controlled setting, this approach could be done in the field. A conceptually similar approach has been done to link plant genomic variation to microbial "hub" species in the field (Agler et al. 2016, Brachi et al. 2022).

One drawback of HMSC is that it is not currently optimized for the challenges imposed by other aspects of GWA analyses.

For example, individual SNPs can be added as a fixed effect to an HMSC model with the GRM as a random effect to estimate individual SNP effects (i.e. single-SNP GWA). Although HMSC is generally computationally efficient (Ovaskainen et al. 2017), conducting single-SNP GWAS for thousands, let alone millions, of SNPs would likely impose too much of a computational burden. This computational burden could be circumvented, however, by applying single-SNP GWAS to aggregate community indices (e.g. diversity metric or position on a community ordination axis) to first identify candidate genes, followed by an HMSC model of the identified SNPs to facilitate interpretation and prediction. This would also provide a useful check as to the robustness of inference from aggregate community indices, given that HMSC can appropriately model mean–variance relationships in the underlying data. A similar approach could be done with Bayesian alphabet models, where the focus is instead on estimating the distribution of SNP effect sizes (Gianola 2013). Alternatively, multivariate linear mixed models that are optimized for GWA analyses (Zhou and Stephens 2014) could be used, where each species is modeled as a phenotypic trait. This is essentially an HMSC model as it directly models the underlying species; however, the assumptions of a linear model are unlikely to hold with sparse count data. In this scenario, it would be strategic to log-transform the data, an approach that is valid for testing for statistical significance (but not prediction) of species count data that have a Poisson or negative binomial distribution (Ives and Freckleton 2015).

Genomic prediction of community change

The value added by genomic data will ultimately lie in whether it is able to predict community change more effectively, and efficiently, than trait-based approaches. Previous analyses suggested this could be the case, as the inclusion of a wide variety of plant functional traits (whole-plant architecture, phenology, leaf primary and secondary chemistry, leaf economic traits, etc.) routinely failed to fully explain genotypic variation in arthropod communities (Wimp et al. 2007, Barbour et al. 2015, Barker et al. 2019). We think that implementing GWAS approaches that attempt to quantify the distribution of SNP effect sizes (e.g. Bayesian alphabet models), rather than only search for candidate genes, should be of equal priority for future research. This is because genomic prediction tools leverage all of the available genomic information. The study by Sato et al. (2024) is encouraging, as it demonstrated the power of genomic prediction to reduce herbivory by planting certain mixtures of plant genotypes.

Genomic prediction has revolutionized plant and animal breeding by producing gains in key crop and livestock traits, shortening the time required for artificial selection, and avoiding regulatory constraints (Meuwissen et al. 2001, Hickey et al. 2017). From a fundamental perspective, genomic prediction tools would give insight to the genetic architecture of arthropod community phenotypes—an open question in the field of eco-evolutionary dynamics (Rudman et al. 2018)—which could also give insight to how plants coevolve in the face of a diverse arthropod community. From an applied perspective, genomic prediction could be used to foster biodiversity in breeding programs of foundation tree species used in forest plantations and even assisted migration. For example, genomic prediction has been used in white spruce to identify tree genotypes that are drought-resilient (Depardieu et al. 2020) as well as those with increased resistance to spruce budworm (most destructive insect defoliator in North America), plant growth, and wood quality (Beaulieu et al. 2020, Lenz et al. 2020), which could increase the resilience and productivity of white spruce stands under climate change, and, in turn, associated forest biodiversity.

Genomic prediction is not without its limitations though. As its name implies, these statistical models put a primer on "prediction" rather than trying to identify the underlying causal structure of genetic variation. In particular, these models rely on patterns of linkage disequilibrium with causal genes; therefore, if the pattern of linkage disequilibrium changes due to evolution, then the predictive power of these models will fail (Yabe et al. 2018). Thus, it is still important to identify the causal genes to overcome these weaknesses and to be aware of changing patterns of linkage disequilibrium, especially if the population of interest is evolving rapidly through artificial (e.g. crop breeding) or natural selection. Nevertheless, we feel that integrating models of genomic prediction with community data have the potential to give insight to the distribution of effect sizes of specific genes by leveraging all the available genomic variation.

But what exactly do we expect the distribution of plant gene effects on arthropod communities to look like? We hypothesize that the genetic architecture of arthropod community phenotypes will have an exponential distribution. This aligns with Fisher's geometric model of adaptation (Orr 1998, 2005) and implies that the arthropod community is influenced by a few genes of relatively large phenotypic effect, but also many of relatively small effect. We expect an exponential distribution of effect sizes because of the likely mix of different genetic architectures in traits that influence plant–herbivore interactions. For example, plant growth and leaf-economic traits are generally determined by many genes of small effect as these quantitative traits are coordinated by many biosynthetic and developmental pathways (Vasseur et al. 2018). In contrast, plant defense traits (e.g. trichomes and specialized plant secondary metabolites) frequently exhibit both quantitative and qualitative variation (Kliebenstein 2017). Qualitative variation (e.g. presence/absence of specialized metabolites) often results from structural genomic variation, such as a natural loss-of-function mutations (Xu and Guo 2020), or through a gene duplication event followed by the evolution of a new gene function (i.e. neofunctionalization; Kliebenstein 2009, 2017, Moore et al. 2019, Katz et al. 2021). This qualitative variation is often an indicator of the presence of a large-effect gene (or structural variant that is effectively inherited as a single gene) for a specific phenotype and often modulates the strength of species interactions (Kliebenstein 2017, Sato et al. 2019), and thus it is an excellent candidate for a keystone gene effect on community dynamics (Barbour et al. 2022). Moreover, genes coding for qualitatively different chemotypes may have pleiotropic effects on other traits, such as plant growth (e.g. growth/defense trade-off), which could magnify the individual gene's effect on multi-trophic interactions (Barbour et al. 2022). Still, given that both plant growth and defense traits are important in shaping arthropod community composition (Barbour et al. 2015, Barker et al. 2019), we hypothesize that an exponential distribution of gene effects (i.e. few large and many small) underlie variation in arthropod community composition. Interestingly, an exponential model aligns well with the escape-and-radiate hypothesis of plant–insect coevolution (Ehrlich and Raven 1964, Wheat et al. 2007), where the evolution of a novel plant defense (i.e. a relatively large effect mutation) would allow the plant to enter a new adaptive zone early in its diversification. However, with subsequent herbivore adaptation and increased number of associated species, coevolution is likely to become more diffuse (Wise and Rausher 2013) and thus rely on fine-tuned mutations of relatively small effect that are balanced against other environmental selective pressures.

Conclusions

Ovaskainen et al. (2017) noted that their statistical framework for HMSC was well suited for asking "if and how the amount and type of genetic variation influences variation in species occurrence, either among species, or in space, or through time," but lamented that "for merging perspectives from genes, individuals, and communities, the largest challenge is currently in the lack of appropriate data." We have entered an era where people are doing large-scale GWAS in the field and collecting the type of data that will permit predictive modeling of gene to community dynamics. While there are certainly challenges to this integration, we have provided guidance on how to bridge this gap. Significant advances in studies of eco-evolutionary dynamics will require interdisciplinary collaborations, as the effective integration of multi-omic datasets and HMSCs requires domain expertise in research fields that usually do not talk to each other. We suspect that the research groups that are able to effectively integrate biostructure (McCann 2007) at the genomic and community level will give fundamental insight to how biodiversity is maintained and how we can preserve it into the future.

Author contributions

Following the CRediT model, M.A.B. contributed to the conceptualization, funding acquisition, project administration, supervision, visualization, writing—original draft, and writing—review and editing; C.B.P.-L. contributed to the conceptualization, visualization, and writing—review and editing.

Conflict of interest. None declared.

Funding

Funding for this work was provided by a Start-up Grant from the Université de Sherbrooke to M.A.B., an NSERC Discovery Grant to M.A.B., and an FRQNT *Rélevé Professorale* Grant to M.A.B.

Data availability

No new datasets were generated or analyzed in this study.

References

Adam, N., Kallenbach, M., Meldau, S., Veit, D., van Dam, N.M., Baldwin, I.T., et al. (2018) Functional variation in a key defense gene structures herbivore communities and alters plant performance. *PLoS One* 13: e0197221.

Agler, M.T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S.-T., Weigel, D., et al. (2016) Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biol*. 14: e1002352.

Amit, G. and Bashan, A. (2023) Top-down identification of keystone taxa in the microbiome. *Nat. Commun*. 14: 3951.

Bangert, R.K., Turek, R.J., Rehill, B., Wimp, G.M., Schweitzer, J.A., Allan, G.J., et al. (2006) A genetic similarity rule determines arthropod community structure. *Mol. Ecol*. 15: 1379–1391.

Barbour, M.A., Erlandson, S., Peay, K., Locke, B., Jules, E.S., Crutsinger, G.M., et al. (2019) Trait plasticity is more important than genetic variation in determining species richness of associated communities. *J. Ecol*. 107: 350–360.

Barbour, M.A., Fortuna, M.A., Bascompte, J., Nicholson, J.R., Julkunen-Tiitto, R., Jules, E.S., et al. (2016) Genetic specificity of a plant–insect food web: implications for linking genetic variation to network complexity. *Proc. Natl. Acad. Sci. USA* 113: 2128–2133.

Barbour, M.A., Kliebenstein, D.J. and Bascompte, J. (2022) A keystone gene underlies the persistence of an experimental food web. *Science* 376: 70–73.

Barbour, M.A., Rodriguez-Cabal, M.A., Wu, E.T., Julkunen-Tiitto, R., Ritland, C.E., Miscampbell, A.E., et al. (2015) Multiple plant traits shape the genetic basis of herbivore community assembly. *Funct. Ecol*. 29: 995–1006.

Barker, H.L., Holeski, L.M., Lindroth, R.L. and Schädler, M. (2018) Genotypic variation in plant traits shapes herbivorous insect and ant communities on a foundation tree species. *PLoS One* 13: e0200954.

Barker, H.L., Riehl, J.F., Bernhardsson, C., Rubert-Nason, K.F., Holeski, L.M., Ingvarsson, P.K., et al. (2019) Linking plant genes to insect communities: identifying the genetic bases of plant traits and community composition. *Mol. Ecol*. 28: 4404–4421.

Bar-On, Y.M., Phillips, R. and Milo, R. (2018) The biomass distribution on Earth. *Proc. Natl. Acad. Sci. USA* 115: 6506–6511.

Barton, N.H., Etheridge, A.M. and Véber, A. (2017) The infinitesimal model: definition, derivation, and implications. *Theor. Popul. Biol*. 118: 50–73.

Beaulieu, J., Nadeau, S., Ding, C., Celedon, J.M., Azaiez, A., Ritland, C., et al. (2020) Genomic selection for resistance to spruce budworm in white spruce and relationships with growth and wood quality traits. *Evol. Appl*. 13: 2704–2722.

Beck, J., Holloway, J.D., Schwanghart, W. and Orme, D. (2013) Undersampling and the measurement of beta diversity. *Methods Ecol. Evol/Br Ecol. Soc*. 4: 370–382.

Berg, S., Christianou, M., Jonsson, T. and Ebenman, B. (2011) Using sensitivity analysis to identify keystone species and keystone links in size-based food webs. *Oikos* 120: 510–519.

Berry, D. and Widder, S. (2014) Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol*. 5: 219.

Brachi, B., Filiault, D., Whitehurst, H., Darme, P., Le Gars, P., Le Mentec, M., et al. (2022) Plant genetic effects on microbial hubs impact host fitness in repeated field trials. *Proc. Natl. Acad. Sci. USA* 119: e2201285119.

Crutsinger, G.M., Collins, M.D., Fordyce, J.A., Gompert, Z., Nice, C.C., Sanders, N.J., et al. (2006) Plant genotypic diversity predicts community structure and governs an ecosystem process. *Science* 313: 966–968.

Depardieu, C., Girardin, M.P., Nadeau, S., Lenz, P., Bousquet, J., Isabel, N., et al. (2020) Adaptive genetic variation to drought in a widely distributed conifer suggests a potential for increasing forest resilience in a drying climate. *New Phytol*. 227: 427–439.

Ehrlich, P.R. and Raven, P.H. (1964) Butterflies and plants: a study in coevolution. *Evolution* 18: 586–608.

Fisher, C.K., Mehta, P. and White, B.A. (2014) Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS One* 9: e102451.

Fritz, R.S. and Price, P.W. (1988) Genetic variation among plants and insect community structure: willows and sawflies. *Ecology* 69: 845–856.

Gianola, D. (2013) Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* 194: 573–596.

Golan, D., Lander, E.S. and Rosset, S. (2014) Measuring missing heritability: inferring the contribution of common variants. *Proc. Natl. Acad. Sci. USA* 111: E5272–81.

Gotelli, N.J. and Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett*. 4: 379–391.

Grimaldi, D.A. and Engel, M.S. (2005) Evolution of the Insects. Vol.144, pp. 1035–36. Cambridge University Press, Cambridge, New York, Melbourne.

Hardersen, S. and La Porta, G. (2023) Never underestimate biodiversity: how undersampling affects Bray–Curtis similarity estimates and a possible countermeasure. *Eur. Zool. J*. 90: 660–672.

Hickey, J.M., Chiurugwi, T., Mackay, I. and Powell, W. (2017) Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet*. 49: 1297–1303.

Ives, A.R., Dennis, B., Cottingham, K.L. and Carpenter, S.R. (2003) Estimating community stability and ecological interactions from time-series data. *Ecol. Monogr*. 73: 301–330.

Ives, A.R. and Freckleton, R. (2015) For testing the significance of regression coefficients, go ahead and log-transform count data. *Meth. Ecol. Evol*. 6: 828–835.

Johnson, M.T.J. and Agrawal, A.A. (2005) Plant genotype and environment interact to shape a diverse arthropod community on evening primrose (*Oenothera biennis*). *Ecology* 86: 874–885.

Kagiya, S., Yasugi, M., Kudoh, H., Nagano, A.J., Utsumi, S., et al. (2018) Does genomic variation in a foundation species predict arthropod community structure in a riparian forest? *Mol. Ecol*. 27: 1284–1295.

Kallenbach, M., Bonaventure, G., Gilardoni, P.A., Wissgott, A., Baldwin, I.T., et al. (2012) *Empoasca* leafhoppers attack wild tobacco plants in a jasmonate-dependent manner and identify jasmonate mutants in natural populations. *Proc. Natl. Acad. Sci. USA* 109: E1548–57.

Kareiva, P. and Sahakian, R. (1990) Tritrophic effects of a simple architectural mutation in pea plants. *Nature* 345: 433–434.

Katz, E., Li, -J.-J., Jaegle, B., Ashkenazy, H., Abrahams, S.R., Bagaza, C., et al. (2021) Genetic variation, environment and demography intersect to shape *Arabidopsis* defense metabolite variation across Europe. *eLife* 10: e67784.

Keith, A.R., Bailey, J.K. and Whitham, T.G. (2010) A genetic basis to community repeatability and stability. *Ecology* 91: 3398–3406.

Kessler, A., Halitschke, R. and Baldwin, I.T. (2004) Silencing the jasmonate cascade: induced plant defenses and insect populations. *Science* 305: 665–668.

Kliebenstein, D. (2009) Advancing genetic theory and application by metabolic quantitative trait loci analysis. *Plant Cell* 21: 1637–1646.

Kliebenstein, D.J. (2017) Quantitative genetics and genomics of plant resistance to insects. Edited by Roberts, J.A. *In* Annual Plant Reviews Online. John Wiley & Sons, Chichester.

Lenz, P.R.N., Nadeau, S., Mottet, M.-J., Perron, M., Isabel, N., Beaulieu, J., et al. (2020) Multi-trait genomic selection for weevil resistance, growth, and wood quality in Norway spruce. *Evol. Appl*. 13: 76–94.

Maddox, G.D. and Root, R.B. (1987) Resistance to 16 diverse species of herbivorous insects within a population of goldenrod, *Solidago altissima*: genetic variation and heritability. *Oecologia* 72: 8–14.

McCann, K. (2007) Protecting biostructure. *Nature* 446: 29.

Meuwissen, T.H., Hayes, B.J. and Goddard, M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.

Moore, B.M., Wang, P., Fan, P., Leong, B., Schenck, C.A., Lloyd, J.P., et al. (2019) Robust predictions of specialized metabolism genes through machine learning. *Proc. Natl. Acad. Sci. USA* 116: 2344–2353.

Mortezaei, Z. and Tavallaei, M. (2021) Recent innovations and in-depth aspects of post-genome wide association study (post-GWAS) to understand the genetic basis of complex phenotypes. *Heredity* 127: 485–497.

Northfield, T.D., Snyder, W.E., Snyder, G.B. and Eigenbrode, S.D. (2012) A simple plant mutation abets a predator-diversity cascade. *Ecology* 93: 411–420.

Novak, M., Yeakel, J.D., Noble, A.E., Doak, D.F., Emmerson, M., Estes, J.A., et al. (2016) Characterizing species interactions to understand press perturbations: what is the community matrix?. *Annu. Rev. Ecol. Evol. Syst*. 47: 409–432.

Orr, H.A. (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52: 935–949.

Orr, H.A. (2005) The genetic theory of adaptation: a brief history. *Nat. Rev. Genet*. 6: 119–127.

Ovaskainen, O. and Soininen, J. (2011) Making more out of sparse data: hierarchical modeling of species communities. *Ecology* 92: 289–295.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., et al. (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett*. 20: 561–576.

Reed, D.H. and Frankham, R. (2003) Correlation between fitness and genetic diversity. *Conserv. Biol*. 17: 230–237.

Rudman, S.M., Barbour, M.A., Csilléry, K., Gienapp, P., Guillaume, F., Hairston Jr, N.G., et al. (2018) What genomic data can reveal about eco-evolutionary dynamics. *Nat. Ecol. Evol*. 2: 9–15.

Sato, Y., Shimizu-Inatsugi, R., Takeda, K., Schmid, B., Nagano, A.J., Shimizu, K.K., et al. (2024) Reducing herbivory in mixed planting by genomic prediction of neighbor effects in the field. *Nat. Commun*. 15: 8467.

Sato, Y., Shimizu-Inatsugi, R., Yamazaki, M., Shimizu, K.K. and Nagano, A.J. (2019) Plant trichomes and a single gene GLABRA1 contribute to insect community composition on field-grown *Arabidopsis thaliana*. *BMC Plant Biol*. 19: 163.

Sato, Y., Yamamoto, E., Shimizu, K.K. and Nagano, A.J. (2021) Neighbor GWAS: incorporating neighbor genotypic identity into genome-wide association studies of field herbivory. *Heredity* 126: 597–614.

Schuman, M.C., Barthel, K. and Baldwin, I.T. (2012) Herbivory-induced volatiles function as defenses increasing fitness of the native plant *Nicotiana attenuata* in nature. *eLife* 1: e00007.

Simon, S.J., Furches, A., Chhetri, H., Evans, L., Abeyratne, C.R., Jones, P., et al. (2024) Genetic underpinnings of arthropod community distributions in *Populus trichocarpa*. *New Phytol*. 242: 1307–1323.

Skovmand, L.H., Xu, C.C.Y., Servedio, M.R., Nosil, P., Barrett, R.D.H., Hendry, A.P., et al. (2018) Keystone genes. *Trends Ecol. Evol*. 33: 689–700.

Speed, D., Cai, N., Johnson, M.R., Nejentsev, S. and Balding, D.J., (2017) Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* 49: 986–992.

Stanton-Geddes, J., Yoder, J.B., Briskine, R., Young, N.D. and Tiffin, P. (2013) Estimating heritability using genomic data. *Methods Ecol. Evol./Br. Ecol. Soc*. 4: 1151–1158.

Vasseur, F., Exposito-Alonso, M., Ayala-Garay, O.J., Wang, G., Enquist, B.J., Vile, D., et al. (2018) Adaptive diversification of growth allometry in the plant *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 115: 3416–3421.

Wang, X.-W., Sun, Z., Jia, H., Michel-Mata, S., Angulo, M.T., Dai, L., et al. (2024) Identifying keystone species in microbial communities using deep learning. *Nat. Ecol. Evol*. 8: 22–31.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C., et al. (2015) So many variables: joint modeling in community ecology. *Trends Ecol. Evol*. 30: 766–779.

Warton, D.I., Wright, S.T. and Wang, Y. (2012) Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol./Br. Ecol. Soc*. 3: 89–101.

Weighill, D., Jones, P., Shah, M., Ranjan, P., Muchero, W., Schmutz, J., et al. (2018) Pleiotropic and epistatic network-based discovery: integrated networks for target gene discovery. *Front. Energy Res*. 6: 30.

Wheat, C.W., Vogel, H., Wittstock, U., Braby, M.F., Underwood, D., Mitchell-Olds, T., et al. (2007) The genetic basis of a plant–insect coevolutionary key innovation. *Proc. Natl. Acad. Sci*. 104: 20427–20431.

Whitham, T.G., Bailey, J.K., Schweitzer, J.A., Shuster, S.M., Bangert, R.K., LeRoy, C.J., et al. (2006) A framework for community and ecosystem genetics: from genes to ecosystems. *Nat. Rev. Genet*. 7: 510–523.

Whitham, T.G., Gehring, C.A., Lamit, L.J., Wojtowicz, T., Evans, L.M., Keith, A.R., et al. (2012) Community specificity: life and afterlife effects of genes. *Trends Plant Sci*. 17: 271–281.

Wimp, G.M., Martinsen G.D., Floate K.D., Bangert R.K. and Whitham T.G. (2005) Plant genetic determinants of arthropod community structure and diversity. *Evolution* 59: 61–69.

Wimp, G.M., Wooley, S., Bangert, R.K., Young, W.P., Martinsen, G.D., Keim, P., et al. (2007) Plant genetics predicts intra-annual variation in phytochemistry and arthropod community structure. *Mol. Ecol*. 16: 5057–5069.

Wise, M.J. and Rausher, M.D. (2013) Evolution of resistance to a multiple-herbivore community: genetic correlations, diffuse coevolution, and constraints on the plant's response to selection. *Evolution* 67: 1767–1779.

Xu, Y.-C. and Guo, Y.-L. (2020) Less is more, natural loss-of-function mutation is a strategy for adaptation. *Plant Commun*. 1: 100103.

Yabe, S., Hara, T., Ueno, M., Enoki, H., Kimura, T., Nishimura, S., et al. (2018) Potential of genomic selection in mass selection breeding of an allogamous crop: an empirical study to increase yield of common buckwheat. *Front. Plant Sci*. 9: 276.

Zhou, X. and Stephens, M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet*. 44: 821–824.

Zhou, X. and Stephens, M. (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* 11: 407–409.

Zhu, H. and Zhou, X. (2021) Transcriptome-wide association studies: a view from Mendelian randomization. *Quant. Biol. (Beijing, China)* 9: 107–121.