PERSPECTIVE

Extending Genomics to Natural Communities and Ecosystems

Thomas G. Whitham,[1,2] Stephen P. DiFazio,[3] Jennifer A. Schweitzer,[4] Stephen M. Shuster,[1,2] Gery J. Allan,[1,2] Joseph K. Bailey,[4] Scott A. Woolbright[1,2]

An important step in the integration of ecology and genomics is the progression from molecular studies of relatively simple model systems to complex field systems. The recent availability of sequenced genomes from key plants is leading to a new understanding of the molecular drivers of community composition and ecosystem processes. As genome sequences accumulate for species that form intimate associations in nature, a detailed view may emerge as to how these associations cause changes among species at the nucleotide level. This advance could dramatically alter views about the structure and evolution of communities and ecosystems.

The emerging field of community and ecosystem genetics has shown that genetic variation in one species can influence the composition of associated communities (the interacting species that live in a particular area) and the functioning of ecosystems (the biotic community and its abiotic environment) (*1*, *2*). To date, this field has focused mainly on the genetics of foundation plant species—those that structure their ecosystems by creating locally stable conditions and providing specific resources for diverse organisms (*3*). When a foundation species' genotype influences the relative fitness of other species, it constitutes an indirect genetic interaction (*4*), and when these interactions change species composition and abundance among individual tree genotypes, they cause community and ecosystem phenotypes to become statistically distinct from one another (*2*). Figure 1 shows some of the community and ecosystem phenotypes that have been quantified for the foundation tree species, *Populus angustifolia*. These include canopy arthropod communities, soil microbial communities, soil nutrient pools, and interactions between an insect herbivore and its avian predators. Common garden studies show that these phenotypes are all heritable (*4*–*6*) in the broad sense (i.e., additive, dominant, and epistatic genetic effects influence these phenotypes). For example, genes control the susceptibility to the leaf-galling aphid, *Pemphigus betae*, in *P. angustifolia*. The presence or absence of this insect is determined by susceptible or resistant tree genotypes. This interaction predictably affects other trophic levels and alters the composition of a diverse community of fungi, insects, spiders, and avian predators (*2*, *5*). This work suggests that indirect genetic interactions among relatively few species (e.g., a tree, herbivores, mutualists, and/or pathogens) and their environment may structure the composition and abundance of a much larger community of organisms and define ecosystem processes such as nutrient cycling. However, much work needs to be undertaken to understand the genomic components underlying these phenotypes, such as additive genes and epigenetic effects. Here, we examine how a genomics perspective may aid our understanding of communities and ecosystems by focusing on the indirect genetic interactions that arise among a few foundation species rather than attempting to

[1]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA. [2]Merriam-Powell Center for Environmental Research, Northern Arizona University, Flagstaff, AZ 86011, USA. [3]Department of Biology, West Virginia University, Morgantown, WV 26506–6057, USA. [4]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA.

*To whom correspondence should be addressed. E-mail: Thomas.Whitham@nau.edu

understand the interactions among all species in an ecosystem.

Although much progress has been made in obtaining genome sequences for of a few agricultural (e.g., rice) or model plant species (e.g., *Arabidopsis*), extending this knowledge to complex communities and ecosystems in natural systems represents a major frontier. Ecologically important traits with known effects on communities and ecosystems have been mapped in *Populus*, including sex, dormancy, disease resistance, leaf chemistry, and biomass (Fig. 2) (*7–10*). Many of these same traits can expand our understanding of community and ecosystem phenotypes among related species in other ecosystems; for example, markers from *Populus trichocarpa* x *P. deltoides*

interspecific comparisons. However, a focus on the shared and unique aspects driving ecological interactions among many different species with a community genetics perspective justifies genomic investigation of nonmodel foundation species (e.g., *Zostera* and *Solidago* spp.) (*14, 15*) that support diverse communities but lack the necessary mapping resources or genomics tools.

Community heritability (H^2_C), the tendency of related individuals within a species to support similar communities of organisms and ecosystem processes, provides a bridge linking ecological interactions and genomics (*2, 4*). For example, regardless of where a replicate clone of an individual tree genotype was planted in a common garden, it accumulated a similar arthropod com-

and soil microbial community phenotypes as well as ecosystem phenotypes (e.g., rates of litter decay and soil nitrogen mineralization processes) covary with plant leaf chemistry, especially levels of condensed tannins and phenolic glycosides (*16–18*). Given the importance of condensed tannins and lignins in energy flow and nutrient cycling, now that their genes are being identified (*10*), we can extend our understanding of the functional drivers of ecosystem processes in natural forests (*16, 19*).

A genomics perspective can further add to our understanding of community structure and ecosystem processes by elucidating the specific genes, alleles, and epigenetic mechanisms that underlie community heritability and ecosystem phenotypes (Fig. 1) associated with mapped traits in foundation species (Fig. 2). When interspecific interactions have fitness consequences for one or more species, selection occurs within a community context (*4*). Foundation species imposing selection on other community members are expected to cause genetic differentiation among communities (*20*). This process provides the simplest explanation for significant heritability in community phenotypes (*2, 4*). It also suggests that when allelic or genomic variation within foundation species with phenotypic effects exhibit population genetic structure, interacting species will also show population level structure to a degree determined by the concordance of their life cycles (*20*). For example, the aphid, *Pemphigus betae*, has evolved two life cycles. In a natural hybrid zone of *Populus angustifolia* x *P. fremontii*, where trees are highly susceptible, aphids alternate between their primary tree hosts and their herbaceous secondary hosts, *Rumex* and *Chenopodium* spp. In contrast, in the adjacent zone of pure *P. angustifolia*, trees are more resistant and aphid survival is lower. In these habitats, aphids live year-round on the roots of their secondary host; this is a simpler life cycle. Reciprocal transfer experiments show that these life cycles are maintained in the field (*21*), which supports the hypothesis that aphids and their tree hosts genetically covary. The theoretical prediction (*20*) that differential population genetic structure arising from these associations should be identifiable by statistical covariance between neutral markers in each species remains to be tested.

Because interspecific interactions and their fitness effects are underlain by the heritable traits of an individual, species within the community undergo continuous evolutionary change (*22*). For example, the mutualism between ectomycorrhizal fungi and trees commonly affects other species, both directly and indirectly. However, little is known about the molecular mechanisms regulating the interactions between the host plant and these fungi. Joint analysis of whole genome sequences and transcriptional profiles of the *Laccaria bicolor* fungus and its *P. trichocarpa* host (*23*) shows that the fungal genome encodes 28 pre-
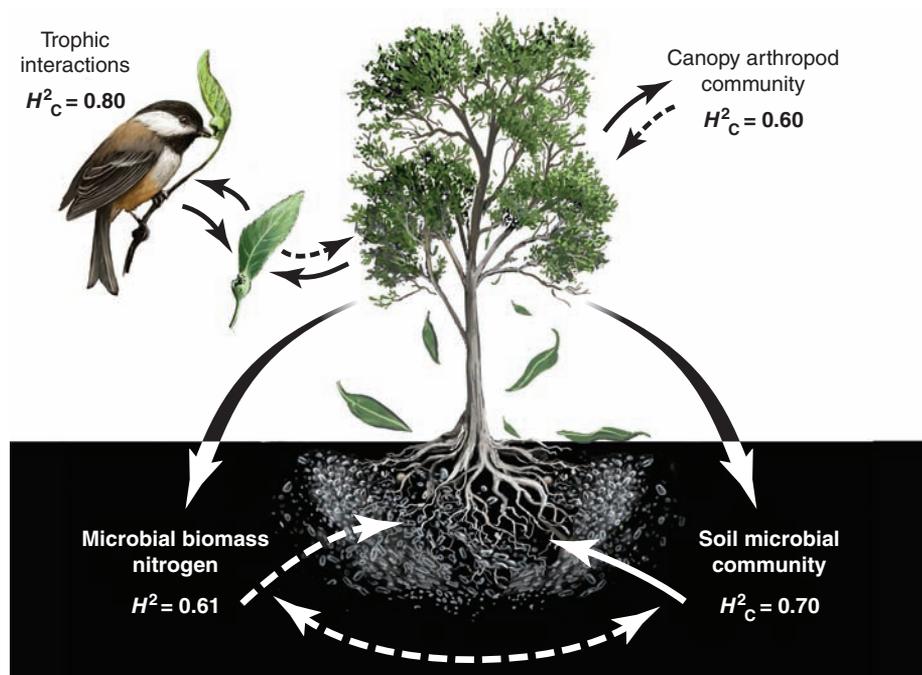

Fig. 1. Community and ecosystem phenotypes of individual tree genotypes of *Populus angustifolia* show broad-sense heritability (H^2). Significant heritability of the canopy arthropod community, soil microbial community, trophic interactions between birds and insects, and soil nutrient pools demonstrate how trait variation in a foundation tree can structure communities and ecosystem processes (H^2_C for community traits; H^2 for single ecosystem trait) (*4–6*). Because soil microorganisms mediate many ecosystem processes, including litter decomposition and rates of nutrient mineralization, the formation of these communities may feed back to affect plant fitness. Solid lines indicate known interactions; dashed lines indicate possible interactions. Quantitative genetic patterns such as these argue that genomic approaches will enhance our understanding of how interacting community members influence ecosystem processes.

genetic maps (*10*) have been transferred to other *Populus* and *Salix* species (i.e., aspen and willow) (*11, 12*), thereby bridging genetic resources between these species (Fig. 2). Similar links between *Eucalyptus globulus* and a community of herbivorous arthropods and mammals have been established with the identification of quantitative trait loci (QTLs) involved in plant defensive chemicals (*13*). It is important to note that this line of inquiry is in its infancy, especially as it pertains to

munity that was significantly different from other replicated tree genotypes. Overall, tree genotype accounted for about 60% of the variation in the arthropod community (i.e., $H^2_C = 0.60$) (Fig. 1). Because these analyses confirm a genetic basis for community phenotypes, they link community-level traits with plant genetic maps and genome sequences (Fig. 2), that is, they reveal genomic locations where genes underlying species interactions may reside. For example, foliar arthropod
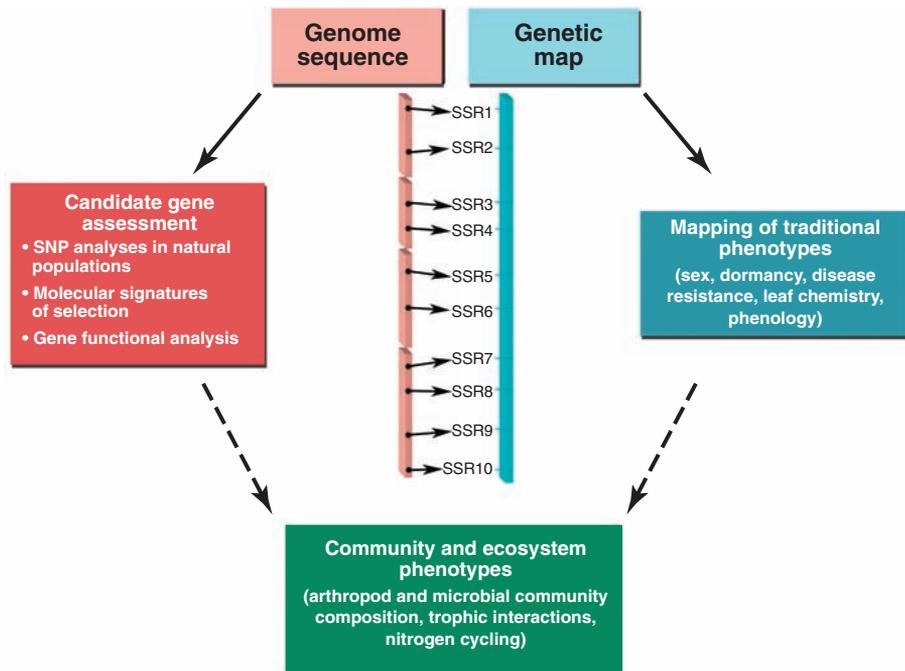
Plant Genomes

Fig. 2. Traditional phenotypes (i.e., those expressed in the individual or population) can be used as a bridge for uncovering the molecular determinants of community and ecosystem phenotypes. First, genomic regions controlling traditional phenotypes are identified with QTL analysis and/or genetic association studies. Candidate genes from these intervals are identified from the genome sequence and anchored to genetic maps with sequence-tagged markers like simple sequence repeats (SSRs). Second, these candidate genes are assessed for their involvement in traditional phenotypes with high-resolution association mapping in natural populations, searching for signatures of selection such as selective sweeps, enhanced local linkage disequilibrium, and/or rates of synonymous and nonsynonymous polymorphisms. Third, genes with evidence for associations and recent selection are then subjected to functional analysis, including transgenic overexpression and knockouts and transcriptome analysis. Fourth, genes responsible for traditional phenotypes can be linked to community and ecosystem phenotypes through studies in common gardens and natural populations. SNP, single-nucleotide polymorphism.

viously undescribed, small secreted proteins that are highly induced upon establishment of the mycorrhizal mutualism. Similarly, the *P. trichocarpa* genome contains homologs of proteins important in establishing ectomycorrhizal mutualisms in other species (*24*). Identified candidate genes controlling the establishment and maintenance of this relationship allow for detailed studies of genetic covariation among species (*4, 20*).

A genomics approach to ecosystem processes is especially important because these processes represent the combined effects of interactions among multiple species, environmental variation, and complex feedback mechanisms that can be difficult to partition or understand in an evolutionary context. Understanding the molecular basis (and expression) of plant and subsequent community traits may simplify our understanding of ecological interactions that affect ecosystem processes. In turn, these processes may feed back to have fitness consequences for the individual expressing those traits. For example, the condensed tannins that affect trophic interactions, soil microbial communities, leaf litter decomposition, and soil nitrogen mineralization rates (*6, 16*) may also feed back to affect the fitness of the host

plant itself. The heritability of specific community and ecosystem phenotypes (Fig. 1) suggests that a genomics approach may link the covarying

genetic interactions that have resulted in these quantitative patterns.

As the above studies emphasize, quantifying the genetic covariance among species is key to understanding the evolution of communities and ecosystems (*20*). Thus, the information gained by studying multiple interacting genomes will inform larger questions than understanding any one genome alone. With the sequencing of the *Populus trichocarpa* genome (*10*) and with *Eucalyptus* in progress, we are now in a position to understand how genes in these foundation tree species, directly and indirectly, affect the evolution of their associated forest communities and mediate ecosystem processes. Multiple genomes and genomic tools are becoming available for diverse organisms that interact with *Populus*, including the mycorrhizal fungus, *Laccaria bicolor* (*24*), with other mycorrhizal, pathogen, and endophytic bacteria in progress (Table 1). This integrative approach is being used in the study of microbial symbiotic communities in animals. Here, genome sequencing has revealed that these microbial communities undergo evolutionarily convergent reductions in genome size, resulting in obligate associations driven by complementary metabolic pathways (*23*).

Although integrating interactions among foundation species at the genomic level is no small task, it is far less imposing than understanding the interactions of all species in an ecosystem. A focus on genome-based interactions among foundation species provides scientists with a realistic approach for unraveling the ecology and evolution of complex ecosystems. Without such an overview, ecosystems may seem intricate beyond description, except as random or stochastic assemblages of communities. A merger of genomics with community and ecosystem ecology may ultimately provide an understanding of how

Table 1. Links for ongoing and completed genome sequencing projects for *Populus* and other organisms that were specifically selected for sequencing due in part to their ecological associations with *Populus*.

Species	URL for Sequencing Project
Populus trichocarpa	www.ornl.gov/sci/ipgc
Laccaria bicolor	http://mycor.nancy.inra.fr/IMGC/LaccariaGenome/about_laccaria.html
Glomus intraradices	http://repositories.cdlib.org/cgi/viewcontent.cgi?article=3173&context=lbnl
Paxillus involutus	www.jgi.doe.gov/sequencing/why/CSP2008/pinvolutus.html
Melampsora larici-populina	www.jgi.doe.gov/sequencing/why/CSP2006/poplarrust.html
Microbial community in decaying *Populus* biomass	www.jgi.doe.gov/sequencing/lspssseqplans2007.html
Pseudomonas putida w619	http://genome.jgi-psf.org/finished_microbes/psepw/psepw.home.html
Serratia proteamaculans 568	http://genome.jgi-psf.org/finished_microbes/serpr/serpr.home.html
Stenotrophomonas maltophilia R551-3	http://genome.jgi-psf.org/draft_microbes/stema/stema.home.html

complex natural ecosystems are formed and maintained through evolutionary time.

References and Notes

1. J. Antonovics, in *Plant Resistance to Herbivores and Pathogens*, R. S. Fritz, E. L. Simms, Eds. (Univ. of Chicago Press, Chicago, 1992), pp. 426–449.
2. T. G. Whitham *et al.*, *Nat. Rev. Gen.* **7**, 510 (2006).
3. A. M. Ellison *et al.*, *Front. Ecol. Environ.* **3**, 479 (2005).
4. S. M. Shuster, E. V. Lonsdorf, G. M. Wimp, J. K. Bailey, T. G. Whitham, *Evol. Int. J. Org. Evol.* **60**, 991 (2006).
5. J. K. Bailey, S. C. Wooley, R. L. Lindroth, T. G. Whitham, *Ecol. Lett.* **9**, 78 (2006).
6. J. A. Schweitzer *et al.*, *Ecology* **89**, 773 (2008).
7. V. Jorge, A. Dowkiw, P. Faivre-Rampant, C. Bastien, *New Phytol.* **167**, 113 (2005).
8. A. M. Rae, P. J. Tricker, S. M. Bunn, G. Taylor, *New Phytol.* **175**, 59 (2007).
9. T. M. Yin *et al.*, *Genome Res.* **18**, 422 (2008).
10. G. A. Tuskan *et al.*, *Science* **313**, 1596 (2006).
11. S. J. Hanley, M. D. Mallott, A. Karp, *Tree Genet. Genom.* **3**, 35 (2006).
12. S. A. Woolbright *et al.*, *Heredity* **100**, 59 (2008).
13. J. S. Freeman *et al.*, *New Phytol.* 10.1111/j.1469-8137.2008.02417.x (2008).
14. A. R. Hughes, J. J. Stachowicz, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8998 (2004).
15. G. M. Crutsinger *et al.*, *Science* **313**, 966 (2006).
16. J. A. Schweitzer *et al.*, *Ecol. Lett.* **7**, 127 (2004).
17. B. Rehill *et al.*, *Biol. Syst. Ecol.* **33**, 125 (2005).
18. R. K. Bangert *et al.*, *Heredity* **100**, 121 (2008).
19. S. A. Harding *et al.*, *Tree Physiol.* **25**, 1475 (2005).
20. M. J. Wade, *Nat. Rev. Gen.* **8**, 185 (2007).
21. N. A. Moran, T. G. Whitham, *Evol. Int. J. Org. Evol.* **42**, 717 (1988).
22. J. N. Thompson, *The Geographic Mosaic of Coevolution* (Univ. of Chicago Press, Chicago, 2005).
23. A. Moya, J. Pereto, R. Gil, A. Latorre, *Nat. Rev. Gen.* **9**, 218 (2008).
24. F. Martin *et al.*, *Nature* **452**, 88 (2008).
25. Land for common gardens provided by the Ogden Nature Center; funding provided by NSF DEB-0425908, NSF DEB-0743437, and Reclamation.

PERSPECTIVE

From Genotype to Phenotype: Systems Biology Meets Natural Variation

Philip N. Benfey[1,2]* **and Thomas Mitchell-Olds**[1]

The promise that came with genome sequencing was that we would soon know what genes do, particularly genes involved in human diseases and those of importance to agriculture. We now have the full genomic sequence of human, chimpanzee, mouse, chicken, dog, worm, fly, rice, and cress, as well as those for a wide variety of other species, and yet we still have a lot of trouble figuring out what genes do. Mapping genes to their function is called the "genotype-to-phenotype problem," where phenotype is whatever is changed in the organism when a gene's function is altered.

Substantial progress in identifying gene function has been made. Studying the effects of modifying individual genes in model organisms such as *Drosophila*, *Caenorhabditis*, and *Arabidopsis* has allowed several thousand genes to be associated with phenotypes. Through similarities in the encoded protein sequence, we have also managed to identify the general function of many genes, classifying them as enzymes, receptors, transcription factors, and so forth. Another informative approach has been to compare genes descended from the same ancestor across many different organisms. In bacteria, this comparative genomics approach has been used to map genes shared among organisms that have similar phenotypes, resulting in the assignment of putative function to these genes (*1*). And yet we still do not know the function of a large number of the genes in either plants or animals, and we still cannot predict with any accuracy what the effect will be of modifying the activity of an uncharacterized gene, even when it has been assigned to a functional class. (Indeed, natural selection may act on effects,

which are too subtle to be identified by experimental manipulations; hence, it may be impossible to determine the function of some genes.) Equally daunting is starting with a phenotypic variant and trying to predict what genes are likely to be involved. The problem is complicated by the fact that most phenotypes of medical or agricultural interest are "complex," which means that more than one gene, in addition to environmental factors, contributes to expression of the phenotype. Not that single-gene traits are necessarily uninteresting for medicine or agriculture, but these were easier for geneticists to decipher. Now we are left with multigene traits that are harder to work out.

The difficulty in mapping genotype to phenotype can be traced to several causes, including inadequate description of phenotypes, too little data on genotypes, and the underlying complexity of the networks that regulate cellular functions. Recent technical advances for acquiring genome-wide data hold promise for improvements in genotyping and phenotyping. It is particularly exciting to contemplate the application of these advances to the myriad of interesting phenotypes found in nature. This natural variation is generated by additive and epistatic effects of alleles across multiple genes, resulting in many individuals with phenotypes near the population mean, and a minority showing ex-

treme phenotypes. Some combinations result in enhanced traits, whereas other combinations are deleterious to fitness in specific environments. Phenotypic alterations are usually in matters of amount, rather than in the presence or absence of a trait. The field of statistical genetics has developed sophisticated tools to map such quantitative traits to regions of chromosomes. The chromosomal regions are known as quantitative trait loci (QTLs) and are described in terms of the percentage of the variation of a trait that can be attributed to each region.

What has been generally missing is the context in which to place these percentages associated with QTLs. What does it mean, at the cellular or molecular level, that a particular allelic polymorphism has a large or small effect on a trait? This is where the complexity of the underlying cellular networks comes into play. Until recently, most molecular processes occurring within cells were described in terms of linear pathways. A signal received by a cell would be transmitted by a linear series of molecular interactions, ultimately resulting in a response such as a change in gene expression. The field of systems biology is expanding this view, replacing the linear pathways with interconnected networks. These networks frequently look like the "hub-and-spoke" configurations of airline routes. When viewed from the perspective of a network in which there are preferred and alternative routes, the magnitudes associated with quantitative trait loci take on new meaning. Because of the hub-and-spoke organization of the major airline routes, a snowstorm in Chicago can result in disruption of 35% of transcontinental air traffic, whereas a snowstorm in Des Moines might only cause a 2% change.

This analogy illustrates another way in which systems biology is changing the way we think about biological processes. The relative importance of the different cities is a function of the dynamics of transcontinental air traffic, not of the cities' intrinsic size or location. A city that is central for one airline's network is frequently peripheral for another airline's network. Although the dynamics of metabolic networks have been

[1]Department of Biology, Duke University, Durham, NC 27708, USA. [2]Institute for Genome Sciences and Policy—Center for Systems Biology, Duke University, Durham, NC 27708, USA.

*To whom correspondence should be addressed. E-mail: philip.benfey@duke.edu