Maintenance of genetic diversity through plant–herbivore interactions
Andrew D Gloss, Anna C Nelson Dittrich, Benjamin Goldman-Huertas and Noah K Whiteman

Identifying the factors governing the maintenance of genetic variation is a central challenge in evolutionary biology. New genomic data, methods and conceptual advances provide increasing evidence that balancing selection, mediated by antagonistic species interactions, maintains genome-wide functionally important genetic variation within species and natural populations. Because diverse interactions between plants and herbivorous insects dominate terrestrial communities, they provide excellent systems to address this hypothesis. Population genomic studies of Arabidopsis thaliana and its relatives suggest spatial variation in herbivory maintains adaptive genetic variation controlling defense phenotypes, both within and among populations. Conversely, inter-species variation in plant defenses promotes adaptive genetic variation in herbivores. Emerging genomic model herbivores of Arabidopsis could illuminate how genetic variation in herbivores and plants interact simultaneously.

Addresses
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA

Corresponding author: Whiteman, Noah K (whiteman@email.arizona.edu)

Why should plant–herbivore interactions maintain genetic diversity?
Host–pathogen interactions are among the most important selective forces known to maintain genetic variation in both hosts [5–7] and pathogens [8*], and SVS plays a key role in this process. For example, geographic variation in pathogen communities may be the strongest selective force maintaining non-neutral genetic variation across human populations [6]. Spatial variation in plant and herbivore populations and communities is likely to produce a similar effect. SVS may be particularly important for herbivores, as plants comprise a large fraction of an herbivore’s environment and may be more important than abiotic factors in determining herbivore fitness [9].

Under SVS, selective advantages or disadvantages of alleles at a locus differ between environments that individuals of a species occupy (Figure 1b,f; Supplementary Table 1). A simple model of populations inhabiting multiple environments, connected by varying levels of gene flow, forms the foundation of theoretical models of polymorphism maintenance within populations [1,10] and among locally adapted populations and host races [11,12]. Figure 1 illustrates an application of this model to plant–herbivore interactions: spatial variation in a plant defensive trait (e.g. [13*]) — which may arise through complex biotic interactions, abiotic interactions, or
Two simplified models for spatially varying selection (SVS) due to spatial heterogeneity in host plant characteristics. Two alleles ('A1' and 'A2') in an herbivorous insect have opposing effects on insect fitness on two host plant types ('P1' and 'P2'), which may represent plant genotypes or species that differ in a defensive trait. When a single insect population feeds on both sympatric host types (a), the two alleles can be maintained at intermediate frequency within the population (c, d). When host plant types are spatially separated so that insect gene flow between host types is low (e), allele frequencies will diverge between populations feeding on each host type, and the two allele polymorphism will be maintained at the species level (g, h) by migration-selection balance [12]. Both novel mutations and standing genetic variation can be driven by SVS to the intermediate frequencies depicted in the figure. The two models presented above are simplified extremes of situations in nature, which can fall along a continuum of high (right column) versus low (left column) host plant segregation with insect gene flow among host types.

Genetic drift — is expected to maintain genetic variation within or among insect populations.

Levene [10] first demonstrated mathematically that SVS can maintain multiple alleles at stable equilibrium frequencies in a single, randomly mating population. Subsequent studies revealed that when gene flow across environments is low, maintenance of polymorphism becomes more favorable [1,10,14]. Further, when the costs of host resistance and enemy virulence vary between environments [15], global polymorphisms in interacting host and enemy genes are even more likely; for plants, geographically variable components of the environment can alter the cost of defense [16]. A final important insight is that the maintenance of polymorphism is more favorable as environment-specific fitness advantages or disadvantages of an allele increase [10,17]. Thus, alleles maintained by SVS are likely to have large phenotypic effects and contribute a disproportionately large amount to fitness [17,18]. These predictions are consistent with the finding that traits under biotic selection are controlled by loci with larger effects than traits under abiotic selection in plants [19].

**Do herbivores maintain genetic variation in plants?**

Population genomic analyses, enabled by whole-genome resequencing of natural *Arabidopsis thaliana* (*Arabidopsis*) accessions ([20], http://www.1001genomes.org/), suggest abundant adaptive variation exists for defense-related traits. Loci underlying defense-related traits [21] are highly differentiated between populations compared to
the genome overall [22*]. The same loci showed little evidence for selective sweeps, inconsistent with an arms race model in which repetitive sweeps reduce diversity [22*]. Instead, plant enemies maintain species-wide defense polymorphisms over broad geographic scales. These polymorphisms manifest through both protein structure and gene expression; genes controlling defense traits, such as glucosinolate production, show high levels of genetic polymorphism and high variation in expression between individuals [23]. However, the extent to which anti-herbivore or anti-pathogen defense genes each contribute to these patterns in Arabidopsis is unclear.

Climate-responsive genes in Arabidopsis show elevated polymorphism [24] and predict fitness in common gardens [3,25], suggesting climatic variation maintains ecologically important genetic variation in Arabidopsis. However, heterogeneity in biotic interactions (e.g. herbivory) may also contribute to these patterns if biotic and climatic variables co-vary. In fact, allele frequencies at genes involved in defense varied with climate more often than expected by chance [25]. Observational data of herbivore distributions, integrated with common garden experiments to identify genetic variation underlying fitness trade-offs in the presence/absence of herbivores in different geographic contexts, may help link genetic variation to spatially varying herbivory.

Additional studies connecting genotype to phenotype at individual loci have also illuminated functionally important variation maintained by herbivores. In Arabidopsis, geographic variation at a locus controlling variation in glucosinolate profiles correlated with the relative long-term abundance of two specialist aphids across Europe, consistent with the direction of differential selection imposed by these species in the laboratory [13*]. Similarly, variation in glucosinolate biosynthetic genes in Boerhiera stricta, a close relative of Arabidopsis, explained geographic variation in herbivore damage and fitness in common gardens [26*]. Herbivory may also contribute to the maintenance of defense variation at finer spatial scales: amino acid polymorphisms underlying a trade-off between growth or defense against biotrophic pathogens and aphids are maintained at intermediate frequencies across populations in ACD6, a gene controlling leaf necrosis [27].

Do plants maintain genetic variation in herbivores?

The strongest evidence that plant diversity drives genetic variation in herbivorous insects exists for host races — sympatric insect populations that use different hosts and are genetically differentiated, despite gene flow among populations [11]. Antagonistic selection when feeding on different hosts, a form of SVS, is hypothesized to generate and maintain genetic divergence at loci affecting preference and performance on different host plants, reducing inter-race gene flow and creating more subtle divergence in nearby genomic regions [11]. Genome-wide scans revealed that regions with loci affecting preference for, and population growth rate on, different hosts have diverged in pea aphid (Acyrthosiphon pisum) host races [28*,29]. Meanwhile, divergent genomic regions in apple maggot races (Rhagoletis pomonella) control diapause timing [30]. These findings highlight that adaptations to host-specific defenses can maintain genetic differences between host races, but other differences between host plants (e.g. phenotype) can also be important.

At present, there is little evidence from the literature for host plant variation maintaining polymorphism within herbivores in the absence of host race formation (but see [31,32]). However, differential performance of different spider mite (Tetranychus urticae) genotypes across hosts suggests spatial mosaics of host plants can maintain significant phenotypic variation in generalist herbivores [33]. Similar patterns may occur even in relatively specialized insects as well: though geography explains patterns of relatedness among populations of the large pine weevil (Hylobius abietis), allele frequencies at a few loci of unknown function differ between individuals feeding on spruce or pine [34].

Detecting plant-driven balancing selection in herbivorous insects

Illuminating signatures of balancing selection in herbivorous insects driven by plant variation requires three phases: (1) identifying plant genes or traits affecting insect fitness, (2) identifying insect genes interacting with plant genes or traits that mediate effects on insect fitness, and (3) using population genetic tests for balancing selection with appropriate null hypotheses. The ability to rapidly generate genomic sequence data from many individuals within natural populations, and to conduct experimental evolution and common garden experiments using completely sequenced plant accessions, now makes achieving these criteria feasible at the scale of genomic analyses. Cost-effective methods relying on pooled sequencing are particularly promising (Box 1).

False positive and negative results stemming from confounding factors, such as population structure, are a major obstacle to mapping loci through genome-wide association studies [35]. Meanwhile, many population genetic tests for balancing selection (Supplementary Table 2) suffer high rates of false positives resulting from genetic drift. Further, simulations [36] reveal that when SVS favors multiple alleles within a population, partial selective sweeps during which a new mutation rises to intermediate frequency proceed extremely slowly. As a result, recombination during the sweep limits hitchhiking of neutral polymorphism to narrow windows near the selected site, and signatures of balancing selection are difficult to detect. Reduced-representation sequencing
Box 1 The utility of pooled sequencing experiments for mapping targets of balancing selection in herbivorous insects

Genome sequences can be generated easily for non-model insects. However, generating, sequencing, and maintaining inbred lines required for traditional quantitative trait locus (QTL) mapping or genome wide association (GWA) studies remains laborious and expensive. Conversely, short read, next-generation sequencing of pooled individuals, in which allele frequencies can be compared between biologically meaningful groups of individuals, offers a desirable alternative in the following contexts:

**Extreme-QTL mapping** requires the generation of large populations that exhibit segregating variation for a trait, isolation of many progeny with extreme trait values, and estimation of allele frequencies in phenotypically extreme individuals through pooled sequencing [52]. Accuracy of the approach is similar to GWA in *Drosophila melanogaster* [53], though pooled extreme-QTL approaches preclude estimation of trait heritability, epistasis, and locus effect sizes. High-throughput phenotyping of herbivore weight gain or development time on different plant mutant genotypes, ecotypes, or species offers an avenue to identify herbivore genetic variation maintained by variation in plant defenses. Experimental populations for phenotyping could be generated by crossing phenotypically divergent parents from a single population or across locally adapted populations, or derived from directly sampling wild individuals to take advantage of natural, low levels of linkage disequilibrium.

**Evolve-and-resequence approaches** [54] involve altering phenotypes of experimental populations through artificial selection or divergent growth conditions, followed by pooled resequencing of experimental populations to uncover causal genetic variants. Rarely a substitute for high and low performance insect populations on different plant types could directly uncover loci with antagonistic effects that depend on host plant characteristics.

Allelic distributions in nature have been used to infer local adaptation [55]. Alleles underlying preferential feeding or high survival on particular, sympatric plants should be at higher frequencies in insects consuming those plants, and the distributions of locally adaptive alleles across populations should be explained more by habitat characteristics (e.g. common host plant species, genotypes, or chemotypes) than population structure.

strategies such as RAD-tag sequencing [37], therefore, may not generate sequence data within small genomic regions showing signatures of balancing selection, particularly in species with low levels of linkage disequilibrium. Integrating genetic mapping studies with population analysis using whole genome sequences (e.g. re-sequencing) is therefore necessary to reveal highly informative, genome-scale patterns.

**Model systems for testing if plant-herbivore interactions maintain variation**

Studying the maintenance of variation by reciprocal plant–insect interactions requires model systems that are experimentally tractable, genomically characterized and interact in nature. While features affecting the ability to perform experiments and generate genomic data — such as genome size or the ease of rearing and manipulating the organism in the laboratory or field — are important to consider, ecological and evolutionary inferences require knowledge of the distributions and ecologies of the interacting species. Ideally, herbivore species amenable to addressing these questions would be nested evolutionarily among a set of species with genomic resources, have low linkage disequilibrium and a large effective population size, and feed naturally on model plant species with sequenced accessions and functional mutants.

Given the available genomic resources and detailed insight into the phenotypic and genetic basis of important defense traits, *Arabidopsis* and its close relatives are excellent models to study genetic variation under balancing selection. A number of herbivorous arthropod species that attack *Arabidopsis* have genomic resources in various stages of development; and these species differ in feeding mode (piercing-sucking or chewing), host breadth (specialist or generalist), and mode of reproduction (Figure 2). New genomic resources for chewing herbivores such as the genome sequence of the diamondback moth (*Plutella xylostella*) and the two-spotted spider mite (*T. urticae*) have facilitated novel insight into the evolution of herbivory and resistance to mustard defenses [38,39]. Other species are also emerging as useful models for addressing if and how plant–herbivore interactions maintain genetic variation. For example, the leaf-mining drosophilid fly *Scaptomyza flaca* specializes on *Arabidopsis* and relatives (Brassicaceae) in the wild [40] (Figure 3). *S. flaca* is closely related to the many *Drosophila* species with completely sequenced genomes and has a relatively small genome (290 Mb) [41]. Remarkably, this species recently evolved from within the microbe-feeding *Drosophila* species [42]. *S. flaca*, like other chewing herbivores, is sensitive to canonical jasmonate-dependent defenses, including glucosinolates, and exhibits variation in performance across *Arabidopsis* accessions [41,43]. Genes that are functionally important in detoxification against glucosinolates have been shown to be under positive selection [Gloss et al., in review]. Because each generation is sexual, it promises to be a good candidate for use in laboratory selection experiments. A sequenced transcriptome [43] and genome [R. Lapoint et al., unpublished] will allow this species to be leveraged in a population genomics context to complement the important lessons learned from other systems.

In addition to *Arabidopsis*, population and comparative genomic resources are rapidly accumulating for plant species of economic and ecological importance, particularly crop plants and their wild relatives [44–46] and forest trees [47]. Importantly, these species span a wide range of taxonomic diversity and vary in defensive traits and life history strategies. Emerging genomic resources for herbivores in these systems will complement those already available (e.g. [48,49]), enabling identification of common patterns underlying the maintenance of variation through plant–herbivore interactions across diverse model systems.

**Outstanding questions/future directions**

Together, population genomic analysis of natural populations coupled with genetic mapping and experimental
evolution (Box 1) can address whether local adaptation or balanced polymorphisms explain much of the adaptive nucleotide variation found in genomic datasets [50]. Outstanding questions include:

1. Is balancing selection, particularly through SVS, an important force maintaining variation in plant–herbivore systems?
2. How important are different sources of plant variation—intraspecific, interspecific, and non-genetic—for the maintenance of genetic variation in herbivores, and vice versa?
3. How does the amount of variation maintained by balancing selection differ between systems with specific versus diffuse species interactions?
4. What are the spatial and temporal scales over which plant–herbivore interactions can maintain balanced polymorphism?
5. How often do genes under balancing selection within populations diverge adaptively among populations or species, given that selective sweeps favoring new mutations erode polymorphism?
6. Do genes under balancing selection through plant–herbivore interactions provide standing variation?
co-opted for other adaptations, such as pesticide resistance?

Answers to these questions may differ between plants and arthropod herbivores, primarily because many plant defense traits can be constrained by the multitude of diverse herbivores attacking each host plant species [51]. Systems in which one, highly specialized herbivore heavily influences plant fitness are ideal for studying both sides of the interaction (plant and herbivore) simultaneously, but may be less generalizable (e.g. [9]).

**Conclusions**

Established evolutionary theory indicates that adaptive processes can facilitate the maintenance, rather than simply the erosion, of genetic variation within and among populations of plants and herbivores. Rapid progress on the development of genomic resources for model plant species with wild relatives has facilitated the illumination of the genes and alleles underlying natural trait variation, as well as how genomes are shaped by adaptive and neutral processes. *Arabidopsis* has been a key model in this regard, and the promise of >1000 completely sequenced genomes, an active research community investigating all facets of its biology, and emerging model herbivores will enable studies linking genetic variation in plants to variation within herbivore species and communities. The extent to which balancing selection *sensu lato* can account for the large amount of genetic variation present in plant and herbivorous insect populations is a general one considering that most named species of life are herbivorous insects and the plants on which they feed.

**Acknowledgments**

A.D.G. and B.G.H. were each supported by Graduate Research Fellowships from the National Science Foundation (DGE-0646147). A.C.N.D. was supported by a PERT fellowship from the National Institutes of Health (SK12GM080708-13). N.K.W. was supported by grants from the National Science Foundation (DEB-1256789), the National Geographic Society (#9097-12), and the University of Arizona (Faculty Seed Grant).

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.pbi.2013.06.002.

**References and recommended reading**

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest


9. The authors identified genes showing balancing selection across 65 fully sequenced isolates of the malaria parasite, *Plasmodium falciparum*. Similar population genomic scans, which are being embraced in model organisms and in studies of human pathogens, could help reveal the extent of balancing selection on counter-defense genes in insect herbivores.


11. Levene H: *Genetic equilibrium when more than one ecological niche is available*. *Am Nat* 1953, 87:331-333.


1307 worldwide accessions of *A. thaliana*, genotyped at 250,000 SNPs, provide a high-resolution picture of genetic variation and natural selection. Previously-identified SNPs underlying defense phenotypes show a unique evolutionary history relative to SNPs controlling flowering time, development, and mineral accumulation.


Hierarchical, population genomic analysis of pea aphid races identified genes encoding salivary proteins and olfactory receptors as candidates driving host-specific divergence in the face of gene flow.


450  Biotic interactions


