CellPress
REVIEWS

Opinion

Keystone Genes

Lotte H. Skovmand,[1] Charles C.Y. Xu,[1] Maria R. Servedio,[2] Patrik Nosil,[3] Rowan D.H. Barrett,[1] and Andrew P. Hendry[1],*

The keystone species concept is used in ecology to describe individual species with disproportionately large effects on their communities. We extend this idea to the level of genes with disproportionately large effects on ecological processes. Such 'keystone genes' (KGs) would underlie traits involved in species interactions or causing critical biotic and/or abiotic changes that influence emergent community and ecosystem properties. We propose a general framework for how KGs could be identified, while keeping KGs under the umbrella of 'ecologically important genes' (EIGs) that also include categories such as 'foundation genes', 'ecosystem engineering genes', and more. Although likely rare, KGs and other EIGs could dominate certain ecological processes; thus, their discovery and study are relevant for understanding eco-evolutionary dynamics.

Keystone Species, Keystone Genes

In his landmark 1966 paper 'Food web complexity and species diversity,' Robert T. Paine demonstrated that a community of rocky intertidal marine organisms could be dramatically altered by removing a single predator species [1]. This work empirically showed that certain species can have much larger ecological effects than others present in the community. Specifically, the effects of experimentally removing *Pisaster* starfish rippled through the trophic levels, influencing everything from barnacles to mussels to algae, leading to the local disappearance of almost half of the original biodiversity [1]. Three years later, Paine coined the famous term 'keystone species' to describe how a single species high in trophic status, such as *Pisaster*, can disproportionately affect overall species density and distribution in the community [2]. This idea proved influential, spawning countless studies and solidifying the importance of individual species on structuring communities and local ecosystems.

Following the origin of the keystone species concept, the 'keystone' label has been widely applied to organisms whose removal has community-level influences. These organisms include prey such as hares, plants on which pollinators and seed dispersers depend, mutualists, and modifier species that affect habitat features [3]. The resulting ambiguity surrounding the term prompted arguments against its use, particularly in the context of policy and species priorities for conservation purposes [3]. In direct response, Power *et al.* [4] proposed a formal definition of a keystone species as 'one whose impact on its community or ecosystem is large, and disproportionately large relative to its abundance.' Following this clarification, other related terms for ecologically important species were introduced, such as 'ecosystem engineers' that modulate resource availability to other species by physically changing the biotic or abiotic environment [5] and 'foundation species' that occupy low trophic levels, creating the physical and ecological conditions necessary for the survival of other species [6].

Just as particular species can have large ecological effects, so too might particular genes. This idea has been suggested in previous discussions within community and ecosystem genetics

Highlights

Certain genes (and alleles) might be considered EIGs if their expressed phenotype has measurable ecological effects on communities and/or ecosystems.

We provide a conceptual framework through which KGs could have large and disproportionate effects (relative to all genes or alleles in the environment) on community structure and ecosystem processes.

KGs are likely to be rare because few genes have large (and disproportionate) phenotypic effects, and those that do have such effects might not be ecologically relevant.

EIGs (such as KGs) could accentuate evolutionary processes and could therefore be important for ecological and evolutionary dynamics.

[1]Redpath Museum & Department of Biology, McGill University, Montreal, Quebec, Canada
[2]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[3]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

*Correspondence:
andrew.hendry@mcgill.ca
(A.P. Hendry).

that invoked concepts such as 'extended phenotypes', 'interspecific indirect genetic effects', and 'genes-to-ecosystems' effects [7–12]. Indeed, Okuda *et al.* [13] used the term 'keystone gene' (KG) in referring to genes of large effect underlying polymorphic phenotypes in predatory fish that influence plankton communities and ecosystem properties within aquatic systems. In addition, Whitham *et al.* [7] described how 'genes of large phenotypic effect that function through foundation species in communities of large effect, and ecosystems of large effect, will have disproportionate influences on the biosphere' [7]. Here, we explore these ideas in more detail by defining potential 'ecologically important genes' (EIGs) through the lens of potential KGs. That is, KGs could be one important class of EIGs (Table 1), wherein particular genes have ecological effects disproportionate to their 'abundance' in nature (i.e., the number of copies of that gene in the environment). Overall, we presume KGs to be relatively rare given that few single genes have large effects on traits [14–17] and that relatively few single traits are expected to have individually large effects on ecological processes. At the conjunction of these two somewhat rare situations is the expectation that very few individual genes (relative to all genes in the genome and in the environment) will have measurable ecological effects [10,18,19]. This rarity of KGs does not, however, diminish the value of their study, just as has been the case for keystone species. While we here emphasize genes with large ecological effects, we also note the importance of studying genes with small-to-modest effects that collectively contribute to phenotypes with large ecological effects [18,19].

A Conceptual Framework and Theoretical Definition

Whether or not a particular gene is an EIG is determined by the intersection of two questions: first, does that gene contribute strongly to phenotypes (broadly, including a behavior or the production of a secreted substance) that are present in the environment; and second, do those particular phenotypes have a large effect on the community or ecosystem (including social interactions with other species)? Note that we here focus on ecological effects at the level of communities and ecosystems (Figure 1), while stressing that the reality can be more complicated than a direct gene-to-environment relationship. For example, multilevel selection, trans-

Table 1. Potential Conceptual Definitions of Various Types of Ecologically Important Genes[a]

Name	Conceptual gene definition in analogy with species definition	Refs
Ecologically core gene	A gene commonly found across multiple species within a given region	[69]
Ecologically dominant gene	A gene responsible for phenotypes that competitively exclude other species by collecting a disproportionate share of resources and contributing most to productivity	[70]
Ecosystem engineering gene	A gene responsible for phenotypes that cause physical state changes and availability of resources to other species	[71]
Ecologically structural gene	A gene responsible for phenotypes that provide physical structures of environments, produce variability, provide resources, and create habitats for other species	[72]
Foundation gene	A locally abundant and regionally common gene underlying phenotypes that define structure of a community by creating locally stable conditions for other species, and by modulating and stabilizing ecosystems	[6,73]
Keystone gene	A gene with large and disproportionate ecological effects relative to its abundance	[1,2,4]

[a]Ecologically important genes include subclasses of genes that parallel the original definitions (of species) previously proposed in the literature. These subclasses can overlap with one another and merge under a general operational criterion that concerns any gene with measurable ecological effects.
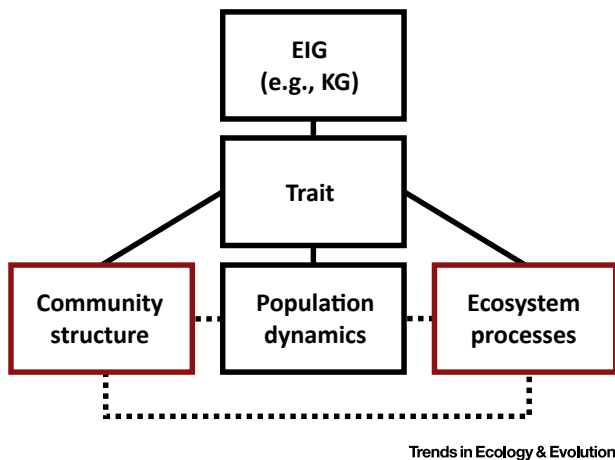
Trends in Ecology & Evolution

Figure 1. Visual Depiction of Ecologically Important Genes. A gene might be considered an ecologically important gene (EIG) if trait-mediated interactions directly (solid black lines) influence community and/or ecosystem processes (as indicated in red). Such genes might also have indirect effects (dashed lines) on communities and ecosystems, such as by influencing population dynamics of the focal species.

generational feedbacks, or any interactions among different populations (of different species) could drive various ecosystem processes [20].

Considered even more broadly, we recognize that EIGs could be recast to include genes that influence population dynamics of ecologically important organisms, such as 'housekeeping genes' or genes necessary for survival. However, we do not here consider such genes for two reasons. First, our initial motivation for this paper came from the concept of KGs, where, by analogy from keystone species, the interest is in the per capita effects of the focal organism on communities and ecosystems. Following from this motivation, for an EIG to be a KG, that phenotype, and hence that gene, would have to have ecological effects that are disproportionate to its abundance in the environment: that is, it cannot be ecologically important simply because it is abundant, such as being in a very abundant species. (Note that by 'important', we imply that such genes have detectable effects on communities and ecosystems, as we will discuss later in the context of 'operational criteria' for identifying such genes.) Second, including such genes would encompass any genes that keep important species alive, or just abundant, and so the KG concept (and the EIG concept in general) would lose much of its exclusivity and therefore utility.

As previously presented in Whitham *et al.* [7], we use the production of tannin (a category of plant secondary compounds that is expressed mainly by a single locus [21]) and its effect on decomposition rate as an empirical example. For starters, we conceptualize the contribution of a gene to the total 'amount' of a particular phenotype in the environment (amount is appropriate for tannin but the term can be substituted with other terms for different classes of phenotypes, such as frequency for a behavior, date for phenology, or other relevant measures). To do so, we first denote τ_{ijk} as the 'total phenotypic effect' (Box 1) of the contribution to phenotype i (e.g., tannin) made by an allele j in species k (i.e., all individuals of a given species in the location of interest). The total phenotypic effect thus captures all individuals in species k in that community and/or ecosystem, and each species k producing phenotype i is indexed separately. Our concept of total phenotypic effect is thus analogous to Fisher's 'average effect' [22], under the original definition of this concept in which Fisher imagined the 'substitution . . . of genes of one kind by genes of another' [23,24]. Hence, just as issues arise around measuring and calculating Fisher's average effect [23,24], similar issues would arise around measuring, and even defining, the total phenotypic effect, as we describe in Box

Box 1. Keystone Genes and the Total Phenotypic Effect

Within a species, the total phenotypic effect of an allele or a gene, τ_{ijk}, could be defined conceptually in either of two ways, which could be subtly different in some cases (see [22,24] for elaboration of these points for the concept of average effect):

(i) The component of the genotypic value G, where G is the portion of the measured phenotype of an individual that is due to genotype as opposed to the environment, summed across individuals in a species, that is caused by the presence of a single allele of interest at a locus for, in our example, tannin production. This measure would be taken given the current distribution of that allele in diploid genotypes of that species (or haploid, or polyploid, as appropriate). The total phenotypic effect will therefore vary based on allele frequency and the population size of that species. If we were instead interested in tannin production across multiple alleles at a locus that contributes to the production of tannins, we could simply group those alleles to get the total effect of that suite of alleles, or, if we grouped over all alleles at that locus, of the gene itself. Note that multiple loci in a given species could be involved in tannin production, but τ_{ijk} is measured separately at each locus; τ_{ijk} for an allele or set of alleles at a given locus and does not depend on whether other loci also contribute to the production of tannins.

(ii) The component of G, summed across individuals in that species, that would be accounted for if we experimentally removed allele j and replaced it with another allele in all individuals in that species. This definition is directly analogous to Fisher's [22] original definition of average effect, which relied on the concept of the 'substitution' of one gene for another. This substitution could occur by replacement with, for example, the remaining alleles at that locus in whatever proportions they currently were in. We could make this substitution if, say, only allele j contributed to the production of tannins, or, alternatively, if we were only interested in the effect of that particular allele on tannin production. Again, if we were interested in a gene where every allele contributed to tannin production, we could imagine we were substituting in an allele that made no contribution to tannin production across all current alleles to obtain the total phenotypic effect of that gene.

Note, importantly, that under both definitions, the total phenotypic effect depends on both the frequency of the focal allele and the rest of the genotypic composition of the species. Because the definition is taken in the current genomic context, no assumptions are made about additivity or interactions across loci. The definitions would also include indirect genetic effects (IGEs), such as social interactions within and between species, because these would potentially be affected or altered, for example, if the allele or set of alleles of interest were absent.

Considering all of the alleles (or genes) across all species that can influence tannin production, we might find a distribution of effect sizes where alleles or genes of large phenotypic effect are relatively rare (see Figure I). In the case of genes, this representation of distributions would be analogous to other previous frequency-generated distributions of QTL effects based on Orr's model that few mutations of large effect and many of smaller effect contribute to the process of adaptation [52–54].
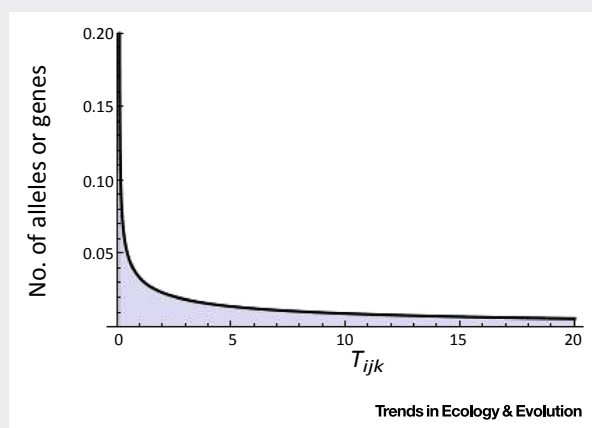

Trends in Ecology & Evolution

Figure I. Frequency Distribution of the Number of Genes (or Alleles) and Their Contribution to the Total Phenotypic Effect. Some alleles (or sets of alleles, including whole genes) j in some particular species k make a particularly large contribution to the amount of, for example, tannin to the environment, whereas most contribute very little.

1. For an allele (or gene) to be an EIG (including a KG), it should have a relatively large τ_{ijk} in comparison to other alleles (or genes) expressed in that population (Box 1). Of course, determining τ_{ijk} for real phenotypes in nature will not be an easy task in many cases; yet, we can see instances where it would be possible. In the case of tannins in a given area, for instance, it seems possible (although time-consuming and potentially expensive) to estimate the amount of tannin per unit biomass of a given genotype of each species and then the biomass of each of those genotypes.

In parallel with the above concept of the total phenotypic effect of an allele or gene, we can also consider the 'total ecological effect', e_i, of a particular phenotype, i, on the community or ecosystem (Box 2). Because different phenotypes can interact in different ways to produce a

---

Box 2. Determining Whether a Phenotype Has a Disproportionate Effect on the Community or Ecosystem

Each point on Figure I shows the hypothetical relationship between the amount of a phenotype i (or, if the phenotype is a behavior, the frequency with which it occurs) in the environment and the magnitude of ecological change due to the experimental removal of that phenotype. For example, many biotic compounds affect decomposition rate [55–57] and are present in the environment at different concentrations (note that this conceptualization assumes a standardized unit of measurement). One point on this graph would represent tannins, where T_i for that point would be the concentration of tannins and the corresponding e_i would be the effect on decomposition rate of removing tannins from the environment (for tannins, the relationship between tannin concentration and e_i would be negative, but it is the absolute value of e_i that is plotted here; see Figure I). Other points would represent the effects of other phenotypes on decomposition rate.

Because we define KGs as having both a pronounced effect and a disproportionally large effect, our second focus in identifying KGs are those points that have a particularly high absolute value of e_i and are positive outliers, meaning that the phenotype is especially crucial to ecosystem function compared to its amount in the environment, T_i. Note that in cases where a gene pleiotropically affects more than one trait that together have a high e_i (see main text), a KG could potentially be identified by summing the e_i across multiple T_i.
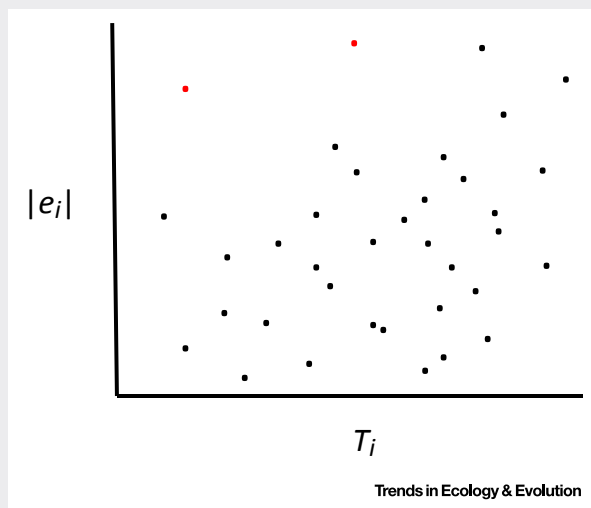

Trends in Ecology & Evolution

Figure I. Relationship between the Total Amount of a Phenotype (T_i) and the Total Magnitude (Absolute Value) of the Ecosystem Effect of That Phenotype (e_i). In the case where only one gene produces phenotype i per species, or when the alleles or genes producing phenotype i combine additively within a species, we could think of the total amount of phenotype i in the environment as being summed across all genes (alleles) and species as $T_i = \Sigma_k \Sigma_j \tau_{ijk}$. As an example, one can imagine taking the total contribution to tannin production by each allele j in each species k in a given location and then summing those amounts across all alleles and populations in that community and/or ecosystem. In other cases, such as when genes within a particular species interact to affect tannin production, the relationship between T_i and τ_{ijk} would be much more complicated. The points on the figure represent a hypothetical distribution of phenotypes and ecosystem effects. Two potential KGs (having both a high absolute value of e_i and being positive outliers) are marked in red.

particular ecological effect, a relationship between a particular phenotype and the ecosystem process of interest might not be simple. Tannins, for example, could interact with other substances to influence decomposition rate in the soil [7]. Regardless, the total ecosystem effect can be conceptually defined, analogous to the total phenotypic effect mentioned above, as the effect on the ecosystem process (e.g., for the tannin example, decomposition rate) if the phenotype i were experimentally removed from the environment. The result of many such experiments, for different phenotypes i, could be a relationship similar to the one discussed in Box 2: an EIG must have at least one allele that produces a phenotype with a high (or outlier) magnitude of e_i relative to the total amount of that phenotype in the environment. Again, this determination would be difficult, expensive, and time-consuming in real systems, but not impossible. For instance, one could experimentally remove individuals of particular phenotypes (and genotypes) and replace them with other phenotypes (and genotypes).

Finally, for an EIG to be also a KG, that is, a gene that contains at least one allele that codes for a phenotype that has a pronounced and disproportionately large effect on the environment, we want to focus on genes (i) that have alleles or sets of alleles with a high total phenotypic effect (far to the right in Figure I in Box 1, on the far right), (ii) whose phenotype determined by those alleles has a large total ecosystem effect (see Figure I in Box 2, at the top), and (iii) for which the effect is disproportionately large relative to other phenotypes and thus other genes (see Figure I in Box 2, outliers). Note that the first two requirements are sufficient for an EIG, and that the third requirement must be fulfilled for an EIG to be a KG. By partitioning the definition of a KG into these three components, we avoid scenarios where genes that produce large quantities of ecologically irrelevant phenotypes are considered KGs simply because they are expressed in a large-bodied or very common species. Although such a gene could well have a high total phenotypic effect (see Figure I in Box 1, far to the right), it would not necessarily be an outlier (see Figure I in Box 2) (i.e., the phenotype of interest might not disproportionately influence the ecosystem compared to its total amount in the environment). This third component, that the effect is disproportionate, would be the most difficult, and perhaps often impossible, to establish empirically because it requires not only determining the first two components for a given gene but also for (hypothetically all) other genes in the environment.

An Operational Criterion

In a variety of both model and non-model organisms, certain genes (alleles or sets of alleles) clearly have larger ecological effects than others [7,9,24]. How can such ecological effects be quantified and what would the threshold between EIGs and non-EIGs, and between KGs and non-KGs, look like? Given the above-noted difficulty of formally determining the three conceptual components described above, it seems appropriate to also consider an operational criterion for EIGs that would be more tractable in empirical situations. Given that all genes are rare, relative to the total number of genes in the environment, one would expect that, overall, any one gene would have a small ecological influence. Thus, any genes whose alleles have measurable effects on community structure and/or ecosystem function are likely to be EIGs. Such genes also could be KGs, with one obvious manifestation being EIGs in keystone species, which could be one operational way of identifying KGs (without excluding the possibility that some KGs could occur in non-keystone species). In addition to KGs, EIGs could be – in analogy with concepts at the species level – core genes, structural genes, foundation genes, dominant genes, and ecosystem engineering genes (Table 1). And, of course, the specific ecologically important genes (and alleles) could vary within a species depending on geographical context (i.e., different populations) and environmental influences (e.g., temperature, precipitation, elevation, predation, parasitism, etc.). That is, EIGs could be population specific and are likely to show genotype-by-environment interactions.

Ecologically important genes might differ from ecologically important species in that the same genes (or alleles) can potentially occur in multiple species and have similar pronounced ecological effects. Hence, a gene might not have a large ecological effect within a single species, but the cumulative effects of that gene across multiple species could be large. For example, a gene within the photosynthetic pathway of an algal species might have a large ecological effect because it is also present in all photosynthetic organisms that together drive the oxygen cycle.

Studies of EIGs can focus on ecologically important species, such as cottonwoods (*Populus* spp.), that are classic ecological models for gene–environment interactions in part because of their tannin production [7,25]. For instance, phenotypic variation in tannin production directly influences both communities of endophytic fungi and macroinvertebrates [26,27] and also ecosystem functions, including up to 63% of the variation in net nitrogen cycling and 97% of the variation in leaf decomposition [27,28]. Moreover, selection clearly acts on this gene because beavers (*Castor canadensis*) preferentially cut down trees with lower levels of condensed tannins, which can triple the relative abundance of trees with high tannin levels relative to those with low tannin levels [29]. Thus, we here have a particularly plausible EIG candidate that is involved in contemporary evolution (of *Populus* spp. in response to beavers) shaping ecosystem function. In Table 2, we review some other study systems with genes (or alleles) that have pronounced ecological effects and are therefore candidate EIGs and, in some cases, candidate KGs.

We predict that real EIGs might be found by targeting traits with relatively simple genetic architecture. Such traits are often polymorphic and involve a few discrete states (as opposed to more continuous trait distributions), which could enable researchers to more easily determine heritability patterns. An example is the locus responsible for color-pattern polymorphism in *Timema cristinae* (see Figure I in Box 3). In this system, maladaptation in color-pattern attracts avian predators, leading to decreases in the abundance of *Timema* and also in the abundance

Table 2. Examples of Candidate Ecologically Important Genes.

Organism	Gene(s)	Accession no.	Trait(s)	Community effect(s)	Ecosystem effect(s)	Refs
Populus spp.	Multiple candidates	XM_002302608.2, FJ573151.1	Tannin content	Animal communities, pollinator abundance	Decomposition rate	[74]
Mimulus spp.	QTL identified	Multiple (see Refs)	Petal color	Pollinator abundance		[75]
Hydrilla verticillata	*pds*	AY639658	Herbicide resistance	Fish and phytoplankton communities	CO_2 levels	[42]
Pinus contorta	Multiple candidates	Multiple (see Refs)	Serotiny, retention of seeds	Animal foragers (squirrels), forest structure	Forest structure, fire	[76]
Melitaea cinxia	*Pgi*	Multiple (see Refs)	Flight performance	Pollinator seed dispersal		[77,78]
Geospiza spp.	*HMGA2*	PRJNA263, PRJNA301892	Beak size and shape	Seed communities		[30]
Timema cristinae	*Mel-stripe*	PRJNA417530	Color polymorphism	Arthropod community richness and density	Herbivory	[60,61,79]
Rutilus rutilus	*cyp19a1a*	AB190291	Estrogen production	Sex ratios of fish populations		[80]
Zea mays L.	*Cry1Ab*		Pest resistance	Insect populations, aquatic predators		[40]
Metrosideros polymorpha	Multiple candidates	Multiple (see Refs)	Tissue quality and growth	Insect populations	Nutrient recycling	[81]

Box 3. Potential KG Found in *Timema cristinae*

Timema stick insects exhibit colors and color-patterns that camouflage them on their host plants. *Timema cristinae* is found mainly on two different hosts: *Adenostoma fasciculatum*, which has small needle-like leaves, and *Ceanothus spinosus*, which has larger ovate-shaped leaves [58]. Accordingly, *T. cristinae* exhibits two main color-pattern morphs: a green-striped morph bearing a white dorsal stripe is cryptic on the leaves of *A. fasciculatum* and a green-unstriped morph is cryptic on the leaves of *C. spinosus*. These color-pattern differences are inherited as a major locus and map to one largely non-recombining region of a single chromosome (i.e., each morph is a chromosomal variant) [38,58].

Although each morph tends to be more common on the host on which it is cryptic, departures from perfect local adaptation are common due to gene flow between stick-insect populations on different host species [59]. A measurable ecological effect of variation in local adaptation (i.e., variation in the frequency of the locally cryptic morph) occurs (see Figure I in Box 3). Specifically, birds forage preferentially on bushes where the frequency of the non-cryptic morph is high, consequently reducing the population density of *T. cristinae*. In turn, birds predate or scare away other arthropods, reducing species richness of the community by ~1/2 as 'collateral damage'. This direct effect then cascades to reduce non-*Timema* herbivory [60,61]. Thus, the locus controlling color-pattern is a candidate for a KG, although further work is required to determine which genetic regions within the non-recombining locus casually affect color-pattern.
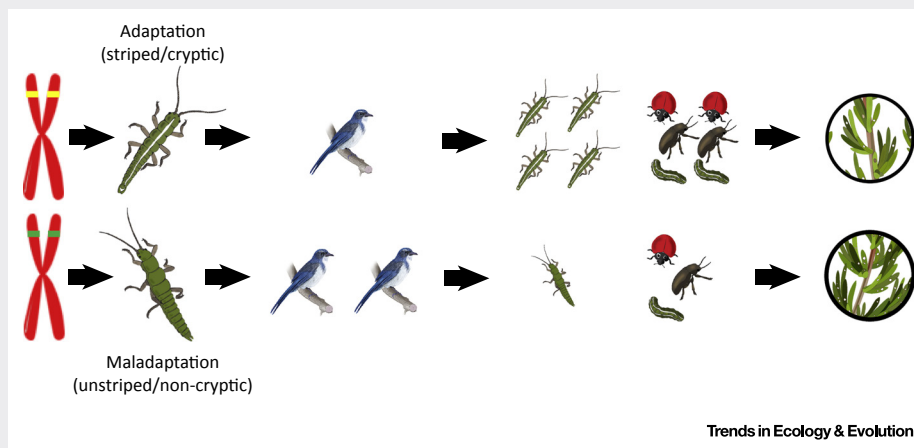

Figure I. Illustration of *Timema* Color-Pattern Morphs and Their Ecological Effects
For a Figure360 author presentation of Figure I in Box 3, see the figure legend at https://doi.org/10.1016/j.tree.2018.07.002
Timema cristinae has two color-pattern morphs (striped and unstriped), with the striped morph being rare and non-cryptic on *Ceanothus spinosus* but common and cryptic on *Adenostoma fasciculatum*. Avian predators are attracted to patches where the frequency of maladapted morphs (in the case illustrated here, the unstriped morph on *A. fasciculatum*) is high, thereby reducing the local abundance of *Timema* as well as the abundance and species richness of co-habiting arthropod communities. The effects of bird predation cascade to affect herbivory on host plants. Illustrations accredited to Rosa Marin.

and richness of other insect species in the community. A second putative example is *HMGA2*, one candidate gene in a genomic region that strongly influences beak size in Darwin's finches [30,31]. Beak size distributions interact with seed trait distributions to strongly influence the abundance and species composition of seeds in the soil [32]. Moreover, *HMGA2* allele frequencies shift during droughts, reflecting selection on beak size [30], providing another possible example of contemporary eco-evolutionary dynamics mediated by a candidate KG. Such ecological effects could then cascade across other biotic and abiotic levels. For example, effects of tannin genotypes in *Populus* spp. on herbivore diet and plant population size can influence microbial communities and other abiotic components found in soil [33], effects of *Timema* genotypes on arthropod communities cascade to influence herbivory on plants (Figure I in Box 3), and effects of Darwin's finch genotypes on plant communities could influence soil properties.

Pleiotropy and 'Supergenes'

EIGs might produce their ecological effects via pleiotropy where a single gene affects multiple traits that together have a large ecological effect. Examples might include 'hub genes' in gene regulatory networks or *trans*-acting regulatory genes. A potential case in point might be the *tb1* locus region in domestic corn types encoding a transcriptional regulator responsible for eight morphological traits including tiller height, basal branch length, and ear phenotypes [34]. Alternatively, genetic linkage can cause multiple traits to be passed on and evolve together. Indeed, some clearly important quantitative trait loci (QTL) contain hundreds or even thousands of genes [35], sometimes in an inversion or other area of reduced recombination [35,36]. These linked genes or supergenes [37] could act in much the same way as EIGs either because they contain one or more EIGs or because they have a substantial ecological effect when expressed together. For example, the major locus responsible for the color morph trait in *T. cristinae* is likely a supergene [38]. Although these supergenes are technically not EIGs (further investigation is required to pinpoint the casual gene or genes), they are expected to act functionally, be inherited, and evolve in similar ways as EIGs. For certain evolutionary questions or practical applications, the distinction between a true EIG and an 'EIG-by-linkage' will not matter.

EIGs in Applied Ecology

Identifying EIGs could be relevant for some applied cases such as transgenic organisms used in agricultural studies and resistant (and/or invasive) species found in nature. In fact, genetically modified organisms (GMOs) could be excellent candidates for the study of EIGs in the sense that alteration of one or more genes produces a large phenotypic effect, whether it be resistance or yield, that might have large ecological effects on surrounding communities. Such EIGs could be particularly important concerning introgression of transgenes between GMOs and wild crops where it has been found that transgenes in wild-type hosts tend to persist and could have negative effects on the environment [39]. Transgenic crops with pollen byproducts also can have negative ecological effects on their surrounding community. One potential case is the transgenic version of *Bacillus thuringiensis* (Bt) corn (*Zea mays*) expressing the toxin Cry1Ab for pest resistance [40]. As cornfields are often positioned near stream channels, pollen byproducts can be transferred to the water. Here, they are either buried in the sediment, consumed by aquatic invertebrates, or transported downstream and introduced to new fields. Transgenic and non-transgenic corn byproducts sampled in sediments of 12 headwater streams showed no difference in decomposition rates [40]. However, local insects (butterflies, beetles, flies, and mosquitoes) were negatively affected by Bt corn, resulting in higher mortality rates [40]. Stress on these insect consumers could then cause cascading ecological effects in riparian systems. Viewing such artificially introduced genes as potential EIGs could aid in examining ecological effects from transgenic products in agricultural practices and in detecting alternative stable states and tipping points (Box 4). In addition, transgenic tools such as CRISPR-CAS9 could enable manipulation of specific alleles in EIGs that could be used to induce large-scale ecological changes via genetic mutations [41].

Another applied case study is the phytoene desaturase gene (*pds*) in estwaite waterweed (*Hydrilla verticillata*) that causes resistance to the chemical herbicide fluridone. Fluridone is used to treat lakes invaded by *H. verticillata* and acts through inhibition of phytoene desaturase, a key enzyme in carotenoid biosynthesis. Samples collected from several lakes in Florida found three independent somatic mutations at the arginine 304 codon of *pds* responsible for phytoene desaturase production in *H. verticillata* populations that were resistant to fluridone treatment [42]. Invasion by *H. verticillata* causes dramatic ecosystem

---

Box 4. EIG Influences on Ecological Concepts

EIGs could be relevant to a wide array of important concepts in ecology including feedbacks, tipping points, hysteresis, and alternative stable states.

Feedbacks (e.g., evolution–ecology, plant–soil, predator–prey, host–parasite) are circuits where 'a change in aspects of traits, population sizes or environments . . . are either casually linked in a closed circular loop or . . . become linked as a result of the change' [62]. When feedbacks are influenced by EIGs (through their expressed trait), we expect them to be stronger because selection (evo-to-eco) and ecological effects (eco-to-evo) become concentrated on relatively few genomic regions, with few or no alternative options. That is, they are likely to be 'closed circular loops'. Feedbacks are key drivers of transitions in ecological dynamics such as shifts from generalization to specialization, rapid adaptation, and tipping points [62,63].

Tipping points are the 'critical level(s) of an external condition (*e.g.* nutrient inflow to a lake) where a system shifts to an alternative state' [64]. Alternatively, tipping points have been defined as unstable equilibrium states seen in classical representations of recovery landscapes [65]. Allele frequency changes in EIGs could reveal genetic tipping points that indicate when systems are close to shifting between stable equilibria, such as changes in a critical level of an external condition (e.g., level of tannin in an environment) that could shift local ecology or ecosystem functioning (e.g., low community abundance or decomposition rate).

Hysteresis is a characteristic of tipping points where new alternative states can differ in trajectories from a previous state, possibly making the new state irreversible [66]: 'as a parameter is changed from one value to another, the position of the equilibrium point changes, tracing a particular trajectory across the landscape' [67]. Studying hysteresis can explain how perturbations can shift equilibriums for communities and ecosystems, and their recovery afterwards, depending on their initial and new state [67]. Hysteresis could be enhanced by EIGs owing to more rapid evolution during changes in states, known as a phase shift, which also could deplete the variation needed for reverse evolution.

Alternative stable states occur when ecosystems (and/or communities) have more than one equilibrium [63] and can happen as a result of tipping points. Changes in dynamic variables such as species abundance and spatial structure, can cause one stable state, or equilibrium, to shift to a new stable state with the right amount of perturbation [67]. Combinations of gene flow, selection, and dominance can generate alternative stable states [68], which could be influenced by (or manipulated through) EIGs.

---

changes and has been shown to displace native plant communities, reduce fish weight and size, and change CO_2 levels in lakes [43]. Although *pds* is a strong putative example of a KG, it has yet to be demonstrated that lakes containing fluridone-resistant *H. verticillata* have different diversities and abundances of other aquatic flora and fauna. Of course, some of these effects might fall into the category of genes influencing presence and abundance of the evolving species, and therefore ultimately might not qualify as KGs in the sense we have emphasized in the present paper.

Implications for Key Ecological Phenomena

In eco-evolutionary dynamics, the focus has been largely on traits rather than on genes, because the former, but not the latter, directly interact with the environment and because most traits are influenced by many genes [14–16,44]. However, situations characterized by EIGs could force a re-examination of ideas about how evolution interacts with key ecological phenomena and could thereby introduce new applications in ecology (see Outstanding Questions). In particular, evolutionary dynamics of traits (and therefore their potential ecological effects) can differ when underpinned by single genes of large effect as opposed to many genes of only small-to-modest effect [45–47]. For example, EIGs might accentuate ecological events, given that evolution is expected to be more rapid in the case of single genes (once allelic variation is present) [46,48]. Such effects are pertinent for concepts such as ecosystem services including climate regulation, air purification, crop pollination, etc. [49], which are altered by contemporary evolution of key biodiversity components [50,51]. Cascading effects

in ecosystem services could also lead to tipping points in which systems show alternative stable states (Box 4).

Concluding Remarks

This opinion article discusses how single genes can have pronounced ecological effects through changes in community structure and/or ecosystem processes. Several partial examples were compiled to illustrate potential EIGs, and a general framework was provided for identifying KGs in future studies. We encourage further research to fill gaps within the genotype–phenotype–ecology pathway. As far as we know, no study has conclusively documented the existence of an EIG or KG (including location, allelic variation, number of inversions, linkage, etc.). Further work on EIGs and KGs will provide valuable insight regarding eco-evolutionary dynamics and general ecological topics such as feedbacks and tipping points. We hope that this opinion article provokes thought and inspires future research directions using the theoretical framework of EIGs and KGs, parallel to the concept of keystone species within community ecology.

Outstanding Questions

How common or rare are EIGs, such as with keystone genes?

Are KGs more commonly found in certain taxa, such as foundation, keystone, or highly abundant species?

How might EIGs be a driving force for eco-evolutionary dynamics?

What are some useful applications of EIGs with the technology that is currently available?

References

1. Paine, R.T. (1966) Food web complexity and species diversity. *Am. Nat.* 100, 65–75
2. Paine, R.T. (1969) A note on trophic complexity and community stability. *Am. Nat.* 103, 91–93
3. Mills, L.S. and Doak, D.F. (1993) The keystone-species concept in ecology and conservation. *Bioscience* 43, 219–224
4. Power, M.E. *et al.* (1996) Challenges in the quest for keystones. *Bioscience* 46, 609–620
5. Jones, C.G. *et al.* (1994) Organisms as ecosystem engineers. *Oikos* 69, 373–386
6. Ellison, A.M. *et al.* (2005) Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. *Front. Ecol. Environ.* 3, 479–486
7. Whitham, T.G. *et al.* (2006) A framework for community and ecosystem genetics: from genes to ecosystems. *Nat. Rev. Genet.* 7, 510–523
8. Mitton, J.B. (2003) The union of ecology and evolution: extended phenotypes and community genetics. *Bioscience* 53, 208–209
9. Whitham, T.G. *et al.* (2003) Community and ecosystem genetics: a consequence of the extended phenotype. *Ecology* 84, 559–573
10. Bailey, J.K. *et al.* (2011) From genes to ecosystems: a synthesis of the effects of plant genetic factors across levels of organization. *Popul. Ecol.* 53, 47–52
11. Hersch-Green, E.I. *et al.* (2011) Community genetics: what have we accomplished and where should we be going? *Philos. Trans. R. Soc. B Biol. Sci.* 366, 1453–1460
12. Wymore, A.S. *et al.* (2011) Genes to ecosystems: exploring the frontiers of ecology with one of the smallest biological units. *New Phytol.* 191, 19–36
13. Okuda, N. *et al.* (2014) Predator diversity changes the world: from gene to ecosystem. In *Biodiversity in Aquatic Systems and Environments* (Okuda, N., ed.), pp. 21–49, Springer
14. Fisher, R.A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52, 399–433
15. Lewontin, R.C. (ed.) (1974) *The Genetic Basis of Evolutionary Change*, Columbia University Press
16. Rockman, M.V. (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* 66, 1–17
17. Le Bourlot, V. *et al.* (2014) Interference versus exploitative competition in the regulation of size-structured populations. *Am. Nat.* 184, 609–623
18. Crutsinger, G.M. *et al.* (2014) Testing a "genes-to-ecosystems" approach to understanding aquatic-terrestrial linkages. *Mol. Ecol.* 23, 5888–5903
19. Hendry, A.P. (ed.) (2016) *Eco-evolutionary Dynamics*, Princeton University Press
20. Genung, M.A. *et al.* (2011) Genetic variation and community change – selection, evolution, and feedbacks. *Funct. Ecol.* 25, 408–419
21. Driebe, E.M. and Whitham, T.G. (2000) Cottonwood hybridization affects tannin and nitrogen content of leaf litter and alters decomposition. *Oecologia* 123, 99–107
22. Fisher, R.A. (ed.) (1930) *The Genetical Theory of Natural Selection*, Clarendon Press
23. Falconer, D.S. (1985) A note on Fisher's "average effect" and "average excess". *Genet. Res.* 46, 337
24. Lee, J.J. and Chow, C.C. (2013) The causal meaning of Fisher's average effect. *Genet. Res.* 95, 89–109
25. Busby, P.E. *et al.* (2015) Genetics-based interactions among plants, pathogens, and herbivores define arthropod community structure. *Ecology* 96, 1974–1984
26. Bailey, J.K. *et al.* (2005) Host plant genetics affect hidden ecological players: links among *Populus*, condensed tannins, and fungal endophyte infection. *Can. J. Bot.* 83, 356–361
27. LeRoy, C.J. *et al.* (2006) Plant genes link forests and streams. *Ecology* 87, 255–261
28. Schweitzer, J.A. *et al.* (2004) Genetically based trait in a dominant tree affects ecosystem processes. *Ecol. Lett.* 7, 127–134
29. Bailey, J.K. *et al.* (2004) Beavers as molecular geneticists: a genetic basis to the foraging of an ecosystem engineer. *Ecology* 85, 603–608
30. Lamichhaney, S. *et al.* (2016) A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* 352, 470–474
31. Chaves, J.A. *et al.* (2016) Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Mol. Ecol.* 25, 5282–5295
32. Grant, P.R. (ed.) (1986) *Ecology and Evolution of Darwin's Finches*, Princeton University Press
33. Schweitzer, J.A. *et al.* (2008) Plant-soil microorganism interactions: heritable relationship between plant genotype and associated soil microorganisms. *Ecology* 89, 773–781
34. Clark, R.M. *et al.* (2006) A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* 38, 594–597
35. Doerge, R.W. (2002) Multifactorial genetics mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.* 3, 43–52
36. Kirkpatrick, M. and Barton, N. (2006) Chromosome inversions, local adaptation and speciation. *Genetics* 173, 419–434

37. Schwander, T. *et al.* (2014) Supergenes and complex pheno-types. *Curr. Biol.* 24, 288–294

38. Lindtke, D. *et al.* (2017) Long-term balancing selection on chro-mosomal variants associated with crypsis in a stick insect. *Mol. Ecol.* 26, 6189–6205

39. Stewart, C.N. *et al.* (2003) Genetic modification: transgene intro-gression from genetically modified crops to their wild relatives. *Nat. Rev. Genet.* 4, 806–817

40. Rosi-Marshall, E.J. *et al.* (2007) Toxins in transgenic crop byprod-ucts may affect headwater stream ecosystems. *Proc. Natl. Acad. Sci. U. S. A.* 104, 16204–16208

41. Kratochwil, C.F. and Meyer, A. (2015) Closing the genotype–phenotype gap: emerging technologies for evolutionary genetics in ecological model vertebrate systems. *BioEssays* 37, 213–226

42. Michel, A. *et al.* (2004) Somatic mutation-mediated evolution of herbicide resistance in the nonindigenous invasive plant hydrilla (*Hydrilla verticillata*). *Mol. Ecol.* 13, 3229–3237

43. Langeland, K.A. (1996) *Hydrilla verticillata* (L.F.) Royle (Hydro-charitaceae), "the perfect aquatic weed". *Castanea* 61, 293–304

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

45. Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature* 217, 624–626

46. Orr, H.A. (2000) Adaptation and the cost of complexity. *Evolution* 54, 13–20

47. Yeaman, S. (2015) Local adaptation by alleles of small effect. *Am. Nat.* 186, 74–89

48. Barrett, R.D. and Schluter, D. (2008) Adaptation from standing genetic variation. *Trends Ecol. Evol.* 23, 38–44

49. Kremen, C. (2005) Managing ecosystem services: what do we need to know about their ecology? *Ecol. Lett.* 8, 468–479

50. Faith, D.P. *et al.* (2010) Evosystem services: an evolutionary perspective on the links between biodiversity and human well-being. *Curr. Opin. Environ. Sustain.* 2, 1–9

51. Rudman, S.M. *et al.* (2017) Evosystem services: rapid evolution and the provision of ecosystem services. *Trends Ecol. Evol.* 32, 403–415

52. Peichel, C.L. and Marques, D.A. (2017) The genetic and molec-ular architechture of phenotypic diversity in sticklebacks. *Philos. Trans. R. Soc. B Biol. Sci.* 3072, 20150486

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

54. Rogers, S.M. *et al.* (2012) Genetic signature of adaptive peak shift in threespine stickleback. *Evolution* 66, 2439–2450

55. Basaraba, J. and Starkey, R. (1966) Effect of plant tannins on decomposition of organic substances. *Soil Sci.* 101, 17–23

56. Wardle, D.A. *et al.* (2004) Ecological linkages between above-ground and belowground biota. *Science* 304, 1629–1633

57. García-Palacios, P. *et al.* (2013) Climate and litter quality differ-ently modulate the effects of soil fauna on litter decomposition across biomes. *Ecol. Lett.* 16, 1045–1053

58. Comeault, A.A. *et al.* (2015) Selection on a genetic polymorphism counteracts ecological speciation in a stick insect. *Curr. Biol.* 25, 1975–1981

59. Sandoval, C.P. (1994) The effects of the relative geographic scales of gene flow and selection on morph frequencies in the walking-stick *Timema cristinae*. *Evolution* 48, 1866–1879

60. Farkas, T.E. *et al.* (2013) Evolution of camouflage drives rapid ecological change in an insect community. *Curr. Biol.* 23, 1835–1843

61. Farkas, T.E. *et al.* (2016) Observational evidence that maladaptive gene flow reduces patch occupancy in a wild insect metapop-ulation. *Evolution* 70, 2879–2888

62. Crespi, B.J. (2004) Vicious circles: positive feedback in major evolutionary and ecological transitions. *Trends Ecol. Evol.* 19, 627–633

63. Scheffer, M. *et al.* (1993) Alternative equilibria in shallow lakes. *Trends Ecol. Evol.* 8, 275–279

64. Scheffer, M. *et al.* (2009) Early-warning signals for critical tran-sitions. *Nature* 461, 53–59

65. Hodgson, D. *et al.* (2015) What do you mean, "resilient"? *Trends Ecol. Evol.* 503–506

66. Brook, B.W. *et al.* (2013) Does the terrestrial biosphere have planetary tipping points? *Trends Ecol. Evol.* 28, 396–401

67. Beisner, B. *et al.* (2003) Alternative stable states in ecology. *Front. Ecol. Environ.* 1, 376–382

68. Flaxman, S.M. *et al.* (2014) Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Mol. Ecol.* 23, 4074–4088

69. Hanski, I. (1982) Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos* 38, 210–221

70. Grime, J. (1984) Dominant and subordinate components of plant communities: implications for succession, stability, and diversity. In *Colonization Succession and Stability* (Gray, A. and Crawley, M., eds), Blackwell Publishing

71. Jones, C. *et al.* (1994) Organisms as ecosystem engineers. *Oikos* 69, 373–386

72. Huston, M.A. (ed.) (1994) *Biological Diversity: The Coexistence of Species on Changing Landscapes*, Cambridge University Press

73. Dayton, P.K. (1972) Toward an understanding of community resilience and the potential effects of enrichments to the bethos at McMurdo Sound, Antarctica. In *Proceedings of the Colloquium on Conservation Problems in Antarctica* (Parker, B.C., ed.), Allen Press

74. James, A.M. *et al.* (2017) Poplar MYB115 and MYB134 tran-scription factors regulate proanthocyanidin synthesis and struc-ture. *Plant Physiol.* 174, 154–171

75. Fishman, L. *et al.* (2013) Chromosomal rearrangements and the genetics of reproductive barriers in *Mimulus* (monkey flowers). *Evolution* 67, 2547–2560

76. Budde, K.B. *et al.* (2014) In situ genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*). *New Phytol.* 201, 230–241

77. Niitepõld, K. *et al.* (2009) Flight metabolic rate and Pgi genotype influence butterfly dispersal rate in the field. *Ecology* 90, 2223–2232

78. Orsini, L. *et al.* (2009) Fitness differences associated with Pgi SNP genotypes in the Glanville fritillary butterfly (*Melitaea cinxia*). *J. Evol. Biol.* 22, 367–375

79. Nosil, P. *et al.* (2018) Natural selection and the predictability of evolution in *Timema* stick insects. *Science* 359, 765–770

80. Lange, A. *et al.* (2008) Altered sexual development in roach (*Rutilus rutilus*) exposed to environmental concentrations of the pharmaceutical 17α-ethinylestradiol and associated expression dynamics of aromatases and estrogen receptors. *Toxicol. Sci.* 106, 113–123

81. Treseder, K.K. and Vitousek, P.M. (2001) Potential ecosystem-level effects of genetic variation among populations of *Metrosi-deros polymorpha* from a soil fertility gradient in Hawaii. *Oecologia* 126, 266–275