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. Hendry1,∗

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.

1Redpath Museum & Department of Biology, McGill University, Montreal, Quebec, Canada
2Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
3Department 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

<table>
<thead>
<tr>
<th>Name</th>
<th>Conceptual gene definition in analogy with species definition</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecologically core gene</td>
<td>A gene commonly found across multiple species within a given region</td>
<td>[69]</td>
</tr>
<tr>
<td>Ecologically dominant gene</td>
<td>A gene responsible for phenotypes that competitively exclude other species by collecting a disproportionate share of resources and contributing most to productivity</td>
<td>[70]</td>
</tr>
<tr>
<td>Ecosystem engineering gene</td>
<td>A gene responsible for phenotypes that cause physical state changes and availability of resources to other species</td>
<td>[71]</td>
</tr>
<tr>
<td>Ecologically structural gene</td>
<td>A gene responsible for phenotypes that provide physical structures of environments, produce variability, provide resources, and create habitats for other species</td>
<td>[72]</td>
</tr>
<tr>
<td>Foundation gene</td>
<td>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</td>
<td>[6,73]</td>
</tr>
<tr>
<td>Keystone gene</td>
<td>A gene with large and disproportionate ecological effects relative to its abundance</td>
<td>[1,2,4]</td>
</tr>
</tbody>
</table>

*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.
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 $\tau_k$ as the ‘total phenotypic effect’ (Box 1) of the contribution to phenotype (e.g., tannin) made by an allele 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 1.
Box 1. Keystone Genes and the Total Phenotypic Effect

Within a species, the total phenotypic effect of an allele or a gene, $\tau_{ik}$, 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 $\tau_{ik}$ is measured separately at each locus; $\tau_{ik}$ 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 1). 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].

Figure 1. 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) 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.
For an allele (or gene) to be an EIG (including a KG), it should have a relatively large $\tau_{ik}$ in comparison to other alleles (or genes) expressed in that population (Box 1). Of course, determining $\tau_{ik}$ 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

\[ |e_i| \]

\[ T_i \]

**Figure 1.** 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 $T_i = \sum_{i,k} \tau_{ik}$. 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 $\tau_{ik}$ 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 be 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 1 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>
<thead>
<tr>
<th>Organism</th>
<th>Gene(s)</th>
<th>Accession no.</th>
<th>Trait(s)</th>
<th>Community effect(s)</th>
<th>Ecosystem effect(s)</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Populus</em> spp.</td>
<td>Multiple candidates</td>
<td>XM_002302608.2, FJ573151.1</td>
<td>Tannin content</td>
<td>Animal communities, pollinator abundance</td>
<td>Decomposition rate</td>
<td>[74]</td>
</tr>
<tr>
<td><em>Mimulus</em> spp.</td>
<td>QTL identified</td>
<td>Multiple (see Refs)</td>
<td>Petal color</td>
<td>Pollinator abundance</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hydilla</em> verticillata</td>
<td><em>pds</em></td>
<td>AY639658</td>
<td>Herbicide resistance</td>
<td>Fish and phytoplankton communities</td>
<td></td>
<td>[42]</td>
</tr>
<tr>
<td><em>Pinus</em> contorta</td>
<td>Multiple candidates</td>
<td>Multiple (see Refs)</td>
<td>Spermaty, retention of seeds</td>
<td>Animal foragers (squirrels), forest structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Melitaea</em> cinxia</td>
<td><em>Pgi</em></td>
<td>Multiple (see Refs)</td>
<td>Flight performance</td>
<td>Pollinator seed dispersal</td>
<td></td>
<td>[77,78]</td>
</tr>
<tr>
<td><em>Geospiza</em> spp.</td>
<td><em>HMIGA2</em></td>
<td>PRJNA263, PRJNA301892</td>
<td>Beak size and shape</td>
<td>Seed communities</td>
<td></td>
<td>[30]</td>
</tr>
<tr>
<td><em>Timema</em> cristinae</td>
<td><em>Mel-striped</em></td>
<td>PRJNA417530</td>
<td>Color polymorphism</td>
<td>Arthropod community richness and density</td>
<td>Herbivory</td>
<td>[60,61,70]</td>
</tr>
<tr>
<td><em>Rutilus</em> rutilus</td>
<td><em>cyt19a7a</em></td>
<td>AB190291</td>
<td>Estrogen production</td>
<td>Sex ratios of fish populations</td>
<td></td>
<td>[80]</td>
</tr>
<tr>
<td><em>Zea</em> mays L.</td>
<td><em>Cry1Ab</em></td>
<td></td>
<td>Pest resistance</td>
<td>Insect populations, aquatic predators</td>
<td></td>
<td>[40]</td>
</tr>
<tr>
<td><em>Metrosideros</em> polymorpha</td>
<td>Multiple candidates</td>
<td>Multiple (see Refs)</td>
<td>Tissue quality and growth</td>
<td>Insect populations</td>
<td>Nutrient recycling</td>
<td>[81]</td>
</tr>
</tbody>
</table>
Trends in Ecology & Evolution

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 [68]. 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 [69]. A measurable ecological effect of variation in local adaptation (i.e., variation in the frequency of the locally cryptic morph) occurs (see Figure 1 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 1. Illustration of *Timema* Color-Pattern Morphs and Their Ecological Effects

For a Figure360 author presentation of Figure1 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 Marín.

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 1 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 estwate 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
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.

References

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?
Trends in Ecology & Evolution

64. Brook, B.W. et al. (2013) Does the terrestrial biosphere have planetary tipping points? Trends Ecol. Evol. 28, 396–401
76. Lange, A. et al. (2008) Altered sexual development inroach (Rutilus rutilus) exposed to environmental concentrations of the pharmaceutical 17a-ethinylestradiol and associated expression dynamics of aromatases and estrogen receptors. Toxicol. Sci. 108, 113–123