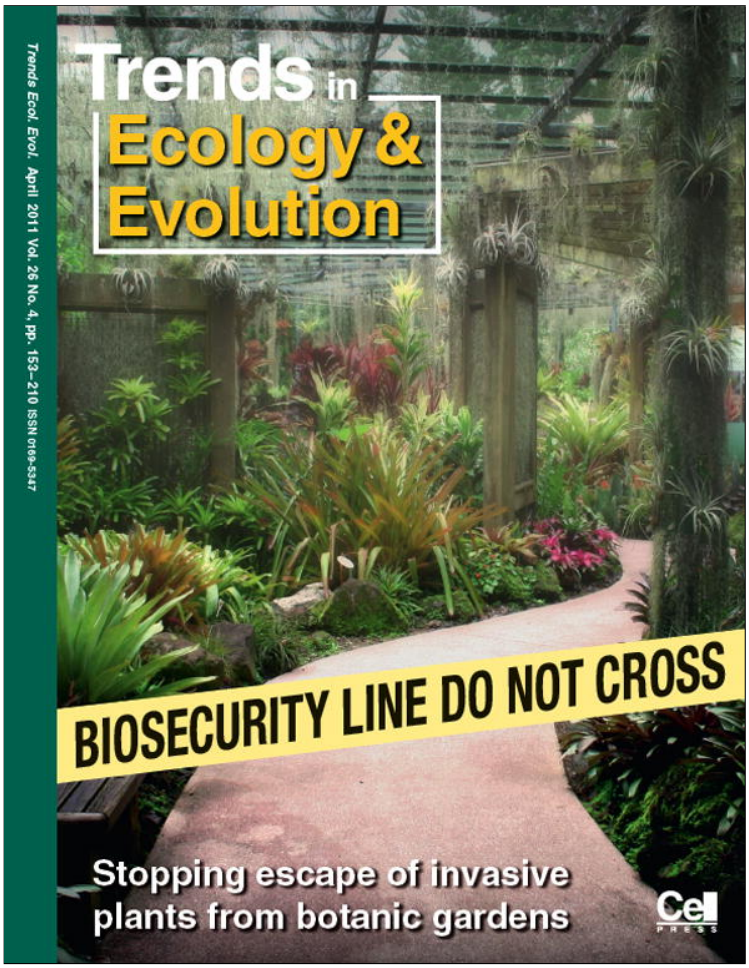

Opinion

Cell PRESS

The genes underlying the process of speciation

Patrik Nosil[1,2] and Dolph Schluter[3]

[1] Department of Ecology and Evolutionary Biology, University of Boulder, CO 80309, USA
[2] Institute for Advanced Study, Wissenschaftskolleg, Berlin, 14193, Germany
[3] Zoology Department and Biodiversity Research Centre, University of British Columbia, Vancouver BC, Canada, V6T 1Z4

The long-standing goal of finding genes causing reproductive isolation is being achieved. To better link the genetics with the process of speciation, we propose that 'speciation gene' be defined as any gene contributing to the evolution of reproductive isolation. Characterizing a speciation gene involves establishing that the gene affects a component of reproductive isolation; demonstrating that divergence at the locus occurred before completion of speciation; and quantifying the effect size of the gene (i.e. the increase in total reproductive isolation caused by its divergence). Review of a sample of candidate speciation genes found that few meet these criteria. Improved characterization of speciation genes will clarify how numerous they are, their properties and how they affect genome-wide patterns of divergence.

What are speciation genes and why study them?

Understanding the genetic basis of speciation is a long-standing goal in evolutionary biology, but many questions remain unanswered or debated [1–3]. For example, how many genes contributed to speciation? What are their effect sizes? Are the same genes involved repeatedly in independent speciation events? Did mutations involved in speciation arise *de novo* or from older standing genetic variation? How prevalent are Dobzhansky-Muller incompatibilities (see Glossary) compared with additive genetic effects? Are changes at regulatory sites or coding regions more likely to underlie speciation? Likewise, are particular classes of gene, such as transposable elements or male-expressed loci, or classes of mutation, such as gene duplications, involved more often in speciation compared with others?

The solutions to such questions are of broad interest not only to geneticists, but also to ecologists and evolutionary biologists in general, as they will help clarify the process of speciation and, hence, the origins of species diversity. For example, the rate of evolution at genes underlying speciation will help decide the relative roles of natural selection and genetic drift in species formation [1,4]. The classes of gene involved, and their phenotypic effects, will help determine the roles of adaptation to environment and genomic conflict in speciation, and whether selection is typically divergent, favoring different alleles in different populations, or parallel, amplifying chance differences [4]. Genes underlying speciation can also be used to test theoretical models that make predictions about the effect size and physical location of genes in the genome [1,5].

The genes to help answer these questions are 'speciation genes', which we define as those genes whose divergence made a significant contribution to the evolution of reproductive isolation between populations. Our definition contrasts with most previous definitions that consider a speciation gene to be any gene contributing to a contemporary component of reproductive isolation between populations or species (Table 1). Here, we outline steps to identifying a speciation gene and evaluate a sample of known candidates.

Most of the first specific genes put forth as speciation genes affected intrinsic hybrid sterility or inviability. Our second goal is to broaden coverage of the many forms of reproductive isolation, including extrinsic components, whose genetic basis remains largely unknown. We end by summarizing what has been learnt so far about the mechanisms of speciation from speciation genes. We do not

Glossary

Divergent natural selection: selection that either acts in contrasting directions between two populations, usually with reference to ecological differences between their environments (e.g. large body size confers high survival in one environment and low survival in the other), or that favors opposite extremes of a trait within a single population (i.e. disruptive selection).
Dobzhansky-Muller incompatibilities: hybrid dysfunction arising from negative interactions (epistasis) between alleles at two or more loci; an allelic substitution at a locus causes no reduction in fitness on its own genetic background, but leads to reduced fitness when placed on the alternative background.
Ecologically dependent post-mating isolation: a form of extrinsic reproductive isolation that occurs when the intermediate phenotype of hybrids leads to reduced fitness in parental environments and when intermediate environments are either lacking or highly restricted.
Effect size of a speciation gene: the positive increase in total reproductive isolation that resulted from divergence between populations at the locus.
Extrinsic reproductive isolation: forms of reproductive isolation that are dependent on the ecological setting; in some cases, these can disappear in an ecologically altered lab environment or if environmental conditions in nature change.
Genealogical discordance: differences among loci in their gene trees.
Immigrant inviability: a form of extrinsic reproductive isolation arising from natural selection against individuals that immigrate into foreign habitats, because such immigrants are maladapted to that environment.
Intrinsic reproductive isolation: reproductive isolation that is not strongly dependent on the ecological setting, the most common example being intrinsic genetic incompatibilities causing hybrid sterility or inviability, even in benign lab environments (other examples exist, such as mating signal or gametic incompatibilities that are independent of ecological setting).
Reproductive isolation: genetically based barriers to gene flow between populations or species.

Corresponding authors: Nosil, P. (patrik.nosil@colorado.edu); Schluter, D. (schluter@zoology.ubc.ca)

 0169-5347/$ – see front matter doi:10.1016/j.tree.2011.01.001 Trends in Ecology and Evolution, April 2011, Vol. 26, No. 4

Table 1. A sample of earlier definitions or usages of the term 'speciation gene'[a]

Quotation	Ref.
'by ''speciation gene'' we merely mean any gene that reduces hybrid fitness'	[78]
'speciation genes are those that contribute to reproductive isolation, often in the form of hybrid inviability, sterility, or behavioral aberration. This definition can include genes that cause isolation owing to physiological, behavioral or even ecological factors.'	[18]
'until recently, the genes that cause reproductive isolation remained black boxes. Consequently, evolutionary biologists were unable to answer several questions about the identities and characteristics of ''speciation genes'''	[79]
'this perhaps unfortunate term, which is now entrenched in the literature, refers to any locus that causes reproductive isolation, whether in F_1 or later generation hybrids, and whether the gene was amongst the first to cause isolation or not'	[2]
'speciation genes restrict gene flow between the incipient species and related taxa'	[55]
'a gene that contributed to the splitting of two lineages by reducing the amount of gene flow between them'	[6]

[a]The definition by [6] is closest to our definition discussed in this article.

provide a comprehensive review of genes affecting reproductive isolation, which can be found elsewhere [1,6,7]. Rather, using key examples, our aim is to sharpen discussion on what genes can reveal about the speciation process. Along the way, we make several assumptions (e.g. genes underlying divergent adaptations might have a role in extrinsic reproductive isolation). These are motivated by the early stage of the field and the need to start somewhere, thereby paving avenues for future work.

Criteria for speciation genes

We propose that identifying a speciation gene involves the following steps: (i) demonstrating that the gene has an effect on a component of reproductive isolation today; (ii) demonstrating that divergence at the locus occurred before speciation was complete; and (iii) quantifying the 'effect size' of the gene at the time it diverged; that is, the degree to which divergence at the locus increased total reproductive isolation.

Past definitions of speciation genes invoke criterion (i) above. By our definition, population divergence at a speciation gene increases the total amount of reproductive isolation (for simplicity, 'divergence' here refers to evolution at the locus, whether or not it led to fixation of distinct alleles; we also recognize that, in special cases, substitution of the same allele in two populations might increase reproductive isolation). It follows that this divergence must occur before reproductive isolation between populations is complete, necessitating criterion (ii). Not all genes affecting reproductive isolation today had a role in the process of speciation [1,8]. Instead, many 'post-speciation' genes will have diverged instead after the completion of speciation (i.e. after reproductive isolation is complete, or is so strong that little further evolution of it can occur), thereby duplicating the effects of genes that diverged earlier in the process. By contrast, speciation genes meet our definition even if reproductive isolation is presently incomplete. In such cases, one cannot predict whether speciation will eventually be completed, but we nevertheless learn about the process of speciation (i.e. the evolution of reproductive isolation) in the meantime. We retain the term 'speciation gene' here, albeit via a modified definition, rather than coin a new term, because 'speciation gene' is entrenched in the literature and encapsulates our main points.

It is possible that post-speciation genes are informative about the evolution of reproductive isolation generally. However, we cannot evaluate this until we clearly distinguish genes that diverged during and after speciation. There are good reasons to suspect that genes that diverge during speciation will differ from those that diverge after. For example, if there is gene flow between populations, then genes that diverge before speciation is complete might require stronger selection or more specific genetic architectures to fix than those that diverge after. Later-fixing genes also have more chances for epistatic effects on reproductive isolation [5,9]. Nevertheless, a good place to start is to find genes that underlie components of present-day reproductive isolation, considering them as candidate speciation genes until direct evidence is obtained of their contribution to the speciation process. Eventually, information will emerge on speciation genes in taxon pairs differing in their strengths of total reproductive isolation, enabling the contribution of genes to various 'stages' of the speciation process (i.e. from beginning to end) to be understood.

Our definition of a speciation gene emphasizes the magnitude of the increment of a gene to total reproductive isolation at the time of its divergence, which we call its 'effect size'. This leads to criterion (iii). The effect size of a speciation gene might be small if numerous genes are involved in a given component of reproductive isolation. For example, the number of genes underlying intrinsic post-mating isolation between 36 taxon pairs (mostly *Drosophila*) was estimated to be 18 genes on average [1], and was as large as 191 in the case of *Drosophila simulans* and *Drosophila melanogaster* [10]. Moreover, total reproductive isolation can result from the cumulative effects of multiple components of isolation [1,11–13], which can limit the effect size of a given gene still further. The concept of effect size is straightforward when applied to genes that made stand-alone increments to total reproductive isolation. However, sometimes only the later-fixing gene of a pair (or set) of interacting genes led to an immediate increase in reproductive isolation, whereas the earlier-fixing gene had little or no effect at the time of its divergence. In such cases the effect size of the later-fixing gene is assigned retroactively to both members of the pair. Effect size is difficult to measure but there is little option if we wish to distinguish the genes that mattered to speciation from those that did not (Box 1).

Candidate speciation genes

It is only with technological advances made during the past decade or so that numerous genes underlying reproductive isolation have been robustly identified and analyzed at the

Box 1. Estimating speciation effect sizes

Ideally, effect size would be estimated by how much a gene increased total reproductive isolation. Practically, the effect size of a candidate speciation gene is often estimated from its current effect on a specific component of reproductive isolation. Estimating effect size will then also require estimation of the fraction of current reproductive isolation contributed by a given component. The magnitude of reproductive isolation (RI) caused by a component is often quantified using indices ranging from 0 (no effect) to 1 (complete RI) [1,11,13]. The effect of a gene on present-day RI can be estimated crudely as: (RI contributed by a component × effect of the gene on that component). For example, *OdsH* reduces fertility of male hybrids between *Drosophila simulans* and *Drosophila mauritiana* by half (with other nearby genes required to confer full sterility) [28,80]. Hypothetically, if hybrid male sterility represented half the total reproductive isolation between the two *Drosophila* species, then the effects of *OdsH* on the total current reproductive isolation would be estimated as $0.5 \times 0.5 = 0.25$, or 25%. When a gene has effects on more than one form of reproductive isolation, such estimates should incorporate the contribution of the gene to each form affected. All such estimates will be biased if components of reproductive isolation interact, will be less accurate if simultaneous divergence occurs at many loci, and are prone to detection limits, which make large effect sizes easier to document.

Despite these issues, progress can be made. Take the case of the QTL locus *YUP* in *Mimulus* monkeyflowers, which strongly affects both flower color and pollinator isolation between *Mimulus cardinalis* and *Mimulus lewisii* [1,6,7]. Current total isolation between these species is strong (>0.99) and arises via the combined effects of multiple components [1]. Although pollinator isolation on its own is strong (=0.976), total reproductive isolation between species would remain high even in its absence, because of the effects of the other barriers. From Table 2 of Ramsey *et al.* [1], we estimated that total isolation in the absence of pollinator isolation would be 0.957 (averaging across estimates of reproductive isolation in each direction). If pollinator isolation is completely due to the result of the *YUP* locus, and if *YUP* is the most recent speciation gene to diverge, the effect size of *YUP* would be $\approx 0.99–0.957 = 0.033$, approximately 3%. By contrast, if *YUP* was the first gene to diverge (an unlikely scenario [11,13,81,82]), its effect size could be as large as 0.976.

These calculations demonstrate that estimating effect size is challenging, but that upper and lower bounds can be obtained. Future work should also consider the degree of progress towards complete fixation of alternative alleles at speciation genes, because no matter the effect of an allele on phenotype (i.e. reproductive isolation) greater genetic divergence will result in stronger reproductive isolation.

DNA-sequence level [1,14–16]. How strong is the evidence that they are speciation genes? To answer this, we compared examples from the literature against our criteria (Tables S1 and S2 in the supplementary material online). Our goal was to illustrate the application of our criteria using a sample of candidate genes, rather than to provide a comprehensive review. Much attention has been given to genes causing intrinsic hybrid sterility or inviability, which can be readily assessed in the laboratory (but see [6,17,18]). We also searched for genes underlying pre-mating sexual isolation and gametic incompatibility [19–23] and genes that contribute to extrinsic, ecologically based components of reproductive isolation [4,12,24]. We found few examples of genes underlying genetically based extrinsic isolation, a situation that we regard as unsatisfying, given that such isolation is often strong in nature [4,12,13,25]. To compensate, we tabulated cases of genes under divergent natural selection between populations and species, on the assumption that these genes contribute in some way to extrinsic isolation [4,12,24]. Examples other than those listed in Tables S1 and S2 in the supplementary material online exist, but those considered span a range of reproductive barriers and illustrate our main points, providing us with a starting point for evaluating genes and the speciation process.

Genes affecting components of reproductive isolation today

The strength of evidence that a given gene is the cause of reproductive isolation varies greatly. The strongest evidence comes from studies that have mapped reproductive isolation to a candidate gene and then used experimental methods, such as positional cloning, gene replacement or knockout, gene expression assays and transgenic manipulations, to confirm the gene involved. Most examples that initially drew attention involved genes causing intrinsic post-mating isolation in *Drosophila*, and include *Odysseus* (*OdsH*) [26–28], hybrid male rescue (*Hmr*) [29,30], lethal hybrid rescue (*Lhr*) [14], nucleoporin 96 (*Nup96*) [31] and nucleoporin 160 (*Nup160*) [16].

In other cases, reproductive isolation has been mapped to a genomic region (i.e. a quantitative trait locus, QTL) and a candidate gene within the QTL identified. In these cases, the involvement of the gene in reproductive isolation remains correlative because its effects are yet to be disentangled from those of physically linked genes. For example, the *wingless* gene resides within a QTL affecting wing color differences and sexual isolation between two species of mimetic *Heliconius* butterflies [32]. Other examples are triosephosphate isomerase (*Tpi*), a gene within a QTL contributing to temporal isolation between populations of *Ostrinia nubilalis* corn borers [33]; pentatricopeptide repeat (*PPR*) genes, which lie within QTL for cytoplasm-dependent anther sterility in *Mimulus* monkeyflower hybrids [34,35]; calcium-dependent protein kinase (*CDPK*), which lies within a QTL for immigrant inviability owing to salt intolerance in *Helianthus* sunflowers [36], and *Lysin* and *Bindin*, which correlate with gametic compatibility in spawning invertebrates [23,37,38].

In a third class of examples, the evidence is even more indirect, consisting of cases in which a specific gene has been found to affect a phenotypic trait known to be under divergent natural selection between the contrasting environments of the parent species. Such selection contributes to reproductive isolation if parental individuals have difficulties growing and surviving in the wrong environment before hybridization can occur ('immigrant inviability'; c.f. [12]), or if hybrids suffer ecologically based reductions in fitness [24,25,39]. An example of such a gene is *Ectodysplasin* (*Eda*), which is the major locus controlling differences in the number of bony lateral plates between wild marine and freshwater stickleback populations [40]. Lateral plate number is probably under divergent selection, with a high plate number being advantageous in marine environments where toothed predators are abundant [41,42], and disadvantageous in freshwater habitats where

low-plated fish have faster growth rates [43,44] and possibly better survival in the presence of invertebrate predators [45]. Similar examples include pituitary homeobox 1 (*Pitx1*) in stickleback [46,47], melanocortin-1 receptor (*Mc1r*), Agouti-signaling protein in mice [48,49], and the long-wave-sensitive opsin gene in cichlids [50]. However, in all such examples to date, the effects on extrinsic reproductive isolation have yet to be directly measured.

Finally, there are examples in which reproductive isolation has been mapped to a QTL, but a candidate gene has yet to be identified. Examples include QTLs affecting immigrant inviability in host races of *Acyrthosiphon pisum* pea aphids [51]; temporal isolation between host races of *Rhagoletis* flies [52]; intrinsic post-mating isolation between various plant taxa [22] (reviewed in [53]); and sexual isolation in *Laupala* crickets [54]. These QTL studies inform about the location of genomic regions involved in speciation, but further work identifying specific genes is required to address questions about the nature (e.g. classes of gene and number of mutations) of genetic changes involved in speciation.

Timing of divergence

How much is known about the timing of divergence of candidate speciation genes? The criterion that divergence takes place before the end of the speciation process is automatically met in examples in which reproductive isolation is not now complete. Two cases are PR domain containing 9 (*Prdm9*) and *Overdrive*, which affect hybrid sterility between subspecies in house mice (*Mus musculus*) and fruit fly (*Drosophila pseudoobscura*), respectively [15,55].

Timing of divergence is more difficult to determine for species already completely reproductively isolated. In such cases, approaches are available to help establish timing of divergence after the fact. One approach uses phylogenetic methods. For example, the gene *Nup96* interacts with *Nup160* to cause intrinsic hybrid inviability between *D. melanogaster and D. simulans* [16,31]. Because reproductive isolation arises from their interaction, both genes would need to have diverged before the evolution of complete reproductive isolation for either of them to be considered a speciation gene. Mapping nucleotide substitutions at *Nup96* onto the known phylogeny of the *D. melanogaster* species group revealed that each substitution in the *D. simulans* lineage was relatively ancient, having occurred before the split with its sister species, *Drosophila mauritiana* [31] (Figure 1). This rules out a late fixation event, but it remains uncertain whether *Nup96* diverged before *D. melanogaster* split from the common ancestor of *D. simulans*.

A second approach compares gene genealogies of candidate speciation genes to genealogies of genes not involved in the speciation process, such as unlinked neutral loci. The logic is that genes affecting reproductive isolation do not flow readily between the species, and if they cease to flow before the time of gene flow cessation at neutral loci, then their divergence must have occurred before gene flow between the species ceased [8,27,56]. The pattern predicted in such a case is discordance between the genealogies of speciation genes and other loci (Figure 2) [57,58].

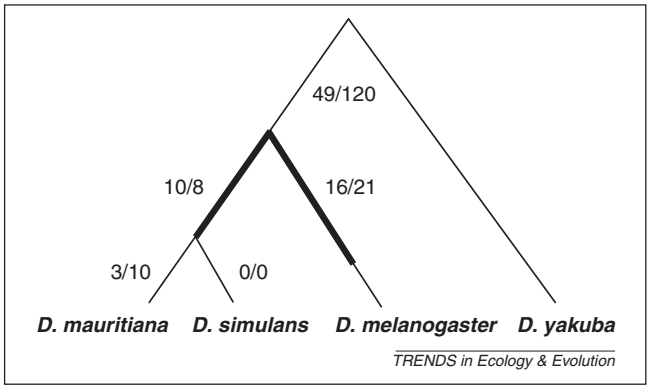

Figure 1. Inferred timing of divergence at *Nup96*, a gene that causes intrinsic post-mating isolation between *Drosophila melanogaster* and *Drosophila simulans*. The species, which currently exhibit complete reproductive isolation, split approximately 1–3 million years ago [77]. Depicted are the replacement:silent substitution ratios mapped onto the phylogeny of the *D. melanogaster* group of species (results shown are for fixed differences; an indel and an insertion are not depicted). Bold branches indicate those in which *Nup96* experienced significant positive selection. All substitutions present in *D. simulans* occurred before the split with *Drosophila mauritiana*, indicating that divergence at *Nup96* was not recent, although the exact timing of divergence in relation to the completion of speciation is unknown. Modified, with permission, from [31].

Speciation genes should 'reflect species boundaries', whereas loci not involved in speciation might, for example, show little phylogenetic resolution, or group taxa by geographic location [59,60]. An example stems from the pheromone races of *O. nubilalis*, where only genealogies for a

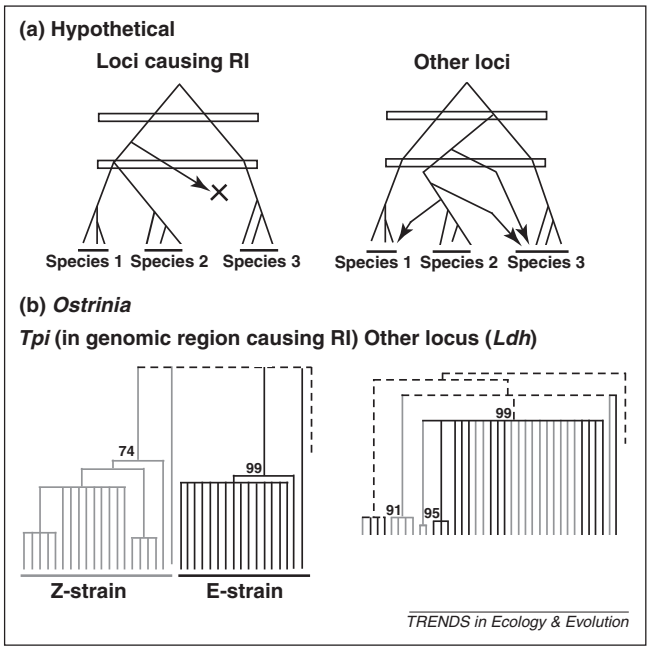

Figure 2. An example of a method to infer that divergence at a candidate speciation gene occurred before the evolution of complete reproductive isolation (following [27]). Depicted are gene trees for different types of loci (e.g. candidate speciation genes versus other loci, such as neutral genes). **(a)** Hypothetical gene flow between species (depicted by arrows) is prevented (large 'X') at speciation genes. Horizontal bars represent speciation events. **(b)** The candidate speciation gene *Tpi* resides in a QTL underlying temporal reproductive isolation between two strains of *Ostrinia nubilalis* corn borers. A phylogeny based on this gene clearly sorts the two strains: grey lines, Z-strain individuals; black lines, E-strain individuals. However, a phylogeny based on *Ldh*, which has no known effects on reproductive isolation in these same strains, fails to sort the strains. A possible explanation is that gene flow has occurred frequently at *Ldh* since the time of divergence at *Tpi*, implying that reproductive isolation was not yet complete when *Tpi* diverged. Results for three other genes were similar to those shown for *Ldh*. Numbers by the nodes represent bootstrap support. Reproduced, with permission, from [33].

gene lying within a QTL for reproductive isolation showed pheromone race exclusivity [33]. A similar pattern occurs for *OdsH* [27].

Although this approach is promising, the degree of genealogical discordance required to be confident that a gene diverged before gene flow ceased requires further study. In general, coalescent-based approaches, such as the isolation-with-migration (IM) model are now commonly used to tease apart the effects of divergence time and gene flow on neutral loci [57]. Such approaches could also be applied to speciation genes, although their efficacy is complicated by potential selection acting on such genes [57].

Effect size of a speciation gene

The amount of reproductive isolation that already exists between populations limits the degree to which the subsequent divergence at a gene can increase reproductive isolation (i.e. it limits the effect size of a gene). For example, a gene that completely sterilizes hybrids has an effect size of only 1% if, at the time of its divergence, it reduced total reproductive isolation from 99% to 100%. Estimating effect size will be especially difficult if divergence occurred long ago and many different barriers to gene flow have evolved. Nonetheless, as described in Box 1, upper and lower bounds might still be estimated. In many circumstances, the effect of a gene will be underestimated because redundant effects of other genes that diverged later will mask the contribution of the target gene at the time of its divergence. Nevertheless, an underestimate is better than no estimate at all and provides a conservative measure.

Other than zero, is there a minimum effect size for a gene to be considered a speciation gene? It is too soon to answer this question definitively, as we currently lack information on the distribution of effect sizes, and even small effects might be important. Reproductive isolation at the level of 95% will often enable ecologically divergent populations to coexist and maintain their distinctiveness, as is seen in many host races of phytophagous insects [12,61] and in sympatric stickleback [62]. A gene that raised the total to 100% reproductive isolation (i.e. effect size = 5%) would staunch all gene flow, enable populations to diverge further by stochastic processes and perhaps ensure that speciation is not reversed [63,64]. Conversely, a gene that brings total reproductive isolation from 0% to 5% will not by itself have as much impact, although it can make it slightly easier for later genes to diverge [1,17]. These examples suggest that the effect size of a speciation gene is related, but not identical, to its importance in the speciation process. A small effect size will render a gene 'less' of a speciation gene in the quantitative, but not in the qualitative, sense. If many genes with small speciation effect sizes exist, a more 'genomic' view that considers how individual genes are arrayed in the genome and how they interact might best characterize the speciation process. Such a finding would inform general understanding of speciation.

Candidate speciation genes: conclusions

Although the number of candidates is growing, only a few genes identified so far come with sufficient evidence to be called speciation genes (Tables S1 and S2 in the supplementary material online). The clearest examples are those known to affect hybrid sterility between incompletely reproductive isolated taxa [15,55]. Such genes have large speciation effect sizes, assuming that other components of reproductive isolation only manifested in the wild have not been overlooked. What evidence is usually lacking? Above all else, we lack estimates of effect size. The example of the *YUP* locus in Box 1 demonstrates how it is possible to put bounds on effect size.

Speciation genes and the mechanisms of speciation

Candidate speciation genes have already provided information about the speciation process. Particularly in animal taxa, there is evidence that genes causing reproductive isolation were influenced by natural selection. First, an increasing number of genes are being discovered whose alternative alleles affect phenotypic traits that are adapted to contrasting environments. These genes might contribute to extrinsic reproductive isolation. In some cases, the mechanism of selection affecting the gene is indicated from comparative or functional data and even selection experiments [44,48]. Second, several discovered genes underlying intrinsic post-zygotic isolation show molecular signatures of positive selection [1,2,16,17]. In one case, *Overdrive*, the mechanism of selection appears to be intragenomic conflict rather than agents of external environment ([15], but see also [65–67]), but it is too early to say whether intragenomic conflict is responsible in other examples. In most cases, the mechanisms of selection are not yet known. Recent reviews suggest that persistent directional selection on speciation genes is less pervasive in plants [6,7]. Future work will determine whether there are consistent differences between major taxa in the evolutionary dynamics of speciation genes.

Candidate speciation genes are also informative about the importance of two classes of speciation via natural selection. In one class, divergence results from divergent natural selection between contrasting environments ('ecological speciation') [4,25,68]. Under this process, different alleles at speciation genes are favored in different populations because of environmental differences in salinity, predation, water availability, disease, light spectrum, and so on (e.g. [40,44,48,50]). In other words, selection favors divergence, and reproductive isolation is its by-product. Candidate genes underlying extrinsic isolation represent genes putatively involved in ecological speciation. In the other class, divergence results despite equivalent or parallel selection pressures on different populations, because each population experiences and fixes a different sequence of mutations at speciation genes ('mutation-order speciation') [4,9,63,69]. Natural selection drives alleles to fixation, but initial divergence itself is stochastic. Genes that diverged under intragenomic conflict, such as meiotic drive [15,16,65,66,70] and cytonuclear interactions [35], provide the most compelling evidence for such mutation-order effects. Divergence occurs because the mutations that cause drive, and mutations that counter drive, are unlikely to be the same in different populations. Interestingly, many cytonuclear incompatibilities in plants

are highly polymorphic [6], perhaps because selection is frequency dependent, in which case changes at multiple loci will be required to produce substantial reproductive isolation between populations.

The results to date also show that intrinsic post-mating isolation usually results from the interaction of at least two genes [1,14]. Such so-called 'Dobzhansky-Muller incompatibilities' are evidence that, despite the low fitness of hybrid gene combinations, the path of divergence probably did not occur against the force of selection [1]. Finally, speciation can involve the divergent resolution or silencing of duplicate genes [6,71,72], but the mechanism by which this occurs, whether selection or drift, deserves further attention.

Candidate speciation genes also provide insights into the role of genetic architecture in speciation with gene flow. According to theory [5], gene flow leads to recombination that breaks down existing associations between genes under selection and genes underlying pre-mating reproductive isolation. This erodes existing pre-mating isolation between populations and stalls further divergence ('selection and recombination antagonism'). This antagonism can be eliminated or diminished in at least three major ways, all of which appear to occur in nature, but with data currently too limited to infer their relative importance. The first is when the genes under selection and those causing reproductive isolation are one and the same (i.e. pleiotropy), as might occur for genes affecting host preference in *Drosophila* [73]. The second scenario is when pre-mating isolation arises via the fixation of the same allele in both of two diverging populations (the 'one-allele mechanism'), as is probable in *Drosophila persimilis* [74]. Third, antagonism between selection and recombination might be lessened when genes underlying pre-mating isolation and those experiencing divergent selection are physically linked on a chromosome, as might occur in pea aphids [51], butterflies [32] and crickets [54]. Such linkage is facilitated when the underlying genes reside in chromosomal rearrangements, but even then theory indicates that physical linkage must be tight to resolve selection and recombination antagonism [75,76]. Collectively, these examples illustrate how speciation genes can shed light on the process of speciation.

Conclusions and future directions

Genes causing reproductive isolation have been called 'speciation genes'. Here, we outlined a modified definition and the criteria to meet it. Our survey of examples highlights the work that remains to be done to better characterize speciation genes. For example, some genes expected to lead to extrinsic isolation have been identified, but their effects on reproductive isolation have yet to be measured. For candidate speciation genes affecting all types of reproductive isolation, future studies need to consider the timing of divergence relative to the completion of speciation. Work is needed to better establish the degree of genealogical discordance required to infer timing of divergence confidently and retroactively. Finally, the challenging, yet crucial, task of determining the effect size of speciation genes must be tackled, particularly across different stages of the speciation process. Recent advances in genomic technologies facilitate the precise study of larger numbers of genomic regions [3]. Thus, future studies of speciation genes will further inform the number, effect size, genomic distribution and types of gene involved in speciation. Such data will unravel how individual speciation genes are embedded within the collective genome and are likely to reveal the mechanisms driving and constraining speciation.

Acknowledgments

J. Galindo, S. Pavey, J. Mallet, H. Collin, J. Feder and A. Meyer provided useful discussion concerning the genetics of speciation. J. Feder, R. Barrett, S. Rogers, L. Rieseberg, and four anonymous reviewers provided comments on an earlier draft of the article. During the writing of this article, P.N. was funded by the Institute for Advanced Study Berlin (Wissenschaftskolleg) and D.S. by the Natural Sciences and Engineering Council of Canada.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tree.2011.01.001.

References

1 Coyne, J.A. and Orr, H.A. (2004) *Speciation*, Sinauer Associates
2 Orr, H.A. (2005) The genetic basis of reproductive isolation: insights from *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6522–6526
3 Noor, M.A.F. and Feder, J.L. (2006) Speciation genetics: evolving approaches. *Nat. Rev. Genet.* 7, 851–861
4 Schluter, D. (2009) Evidence for ecological speciation and its alternative. *Science* 323, 737–741
5 Felsenstein, J. (1981) Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* 35, 124–138
6 Rieseberg, L.H. and Blackman, B.K. (2010) Speciation genes in plants. *Ann. Bot.* 106, 439–455
7 Bomblies, K. (2010) Doomed lovers: mechanisms of Isolation and Incompatibility in plants. *Annu. Rev. Plant Biol.* 61, 109–124
8 Via, S. (2009) Natural selection in action during speciation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9939–9946
9 Gavrilets, S. (2004) *Fitness Landscapes and the Origin of Species*, Princeton University Press
10 Presgraves, D.C. (2003) A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. *Genetics* 163, 955–972
11 Ramsey, J. *et al.* (2003) Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* 57, 1520–1534
12 Nosil, P. *et al.* (2005) Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* 59, 705–719
13 Lowry, D.B. *et al.* (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philos. Trans. R. Soc. B: Biol. Sci.* 363, 3009–3021
14 Brideau, N.J. *et al.* (2006) Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* 314, 1292–1295
15 Phadnis, N. and Orr, H.A. (2009) A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* 323, 376–379
16 Tang, S.W. and Presgraves, D.C. (2009) Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science* 323, 779–782
17 Mallet, J. (2006) What does *Drosophila* genetics tell us about speciation? *Trends Ecol. Evol.* 21, 386–393
18 Wu, C.I. and Ting, C.T. (2004) Genes and speciation. *Nat. Rev. Genet.* 5, 114–122
19 Lassance, J.M. *et al.* (2010) Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. *Nature* 466, 486–489
20 Swanson, W.J. and Vacquier, V.D. (2002) Reproductive protein evolution. *Annu. Rev. Ecol. Syst.* 33, 161–179

21 Lessios, H.A. (2007) Reproductive isolation between species of sea urchins. *Bull. Mar. Sci.* 81, 191–208
22 Rieseberg, L.H. and Willis, J.H. (2007) Plant speciation. *Science* 317, 910–914
23 Palumbi, S.R. (2009) Speciation and the evolution of gamete recognition genes: pattern and process. *Heredity* 102, 66–76
24 Rundle, H.D. and Whitlock, M.C. (2001) A genetic interpretation of ecologically dependent isolation. *Evolution* 55, 198–201
25 Rundle, H.D. and Nosil, P. (2005) Ecological speciation. *Ecol. Lett.* 8, 336–352
26 Ting, C.T. *et al.* (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282, 1501–1504
27 Ting, C.T. *et al.* (2000) The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5313–5316
28 Sun, S. *et al.* (2004) The normal function of a speciation gene, *Odysseus*, and its hybrid sterility effect. *Science* 305, 81–83
29 Barbash, D.A. *et al.* (2003) A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5302–5307
30 Barbash, D.A. *et al.* (2004) Functional divergence caused by ancient positive selection of a *Drosophila* hybrid incompatibility locus. *PLoS Biol.* 2, 839–848
31 Presgraves, D.C. *et al.* (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* 423, 715–719
32 Kronforst, M.R. *et al.* (2006) Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6575–6580
33 Dopman, E.B. *et al.* (2005) Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14706–14711
34 Sweigart, A.L. *et al.* (2006) A simple genetic incompatibility causes hybrid male sterility in *Mimulus*. *Genetics* 172, 2465–2479
35 Barr, C.M. and Fishman, L. (2010) The nuclear component of a cytonuclear hybrid incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat (PPR) genes. *Genetics* 184, 455–465
36 Lexer, C. *et al.* (2004) Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *New Phytologist* 161, 225–233
37 Palumbi, S.R. (1999) All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proc. Natl. Acad. Sci. U.S.A.* 96, 12632–12637
38 Springer, S.A. and Crespi, B.J. (2007) Adaptive gamete-recognition divergence in a hybridizing *Mytilus* population. *Evolution* 61, 772–783
39 Via, S. *et al.* (2000) Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution* 54, 1626–1637
40 Colosimo, P.F. *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* 307, 1928–1933
41 Reimchen, T.E. (1995) Predator-induced cyclical changes in lateral plate frequencies of *Gasterosteus*. *Behaviour* 132, 1079–1094
42 Reimchen, T.E. (2000) Predator handling failures of lateral plate morphs in *Gasterosteus aculeatus*: functional implications for the ancestral plate condition. *Behaviour* 137, 1081–1096
43 Marchinko, K.B. and Schluter, D. (2007) Parallel evolution by correlated response: lateral plate reduction in threespine stickleback. *Evolution* 61, 1084–1090
44 Barrett, R.D.H. *et al.* (2008) Natural selection on a major armor gene in threespine stickleback. *Science* 322, 255–257
45 Marchinko, K.B. (2009) Predation's role in repeated phenotypic and genetic divergence of armor in threespine stickleback. *Evolution* 63, 127–138
46 Shapiro, M.D. *et al.* (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428, 717–723
47 Chan, Y.F. *et al.* (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a pitx1 enhancer. *Science* 327, 302–305
48 Hoekstra, H.E. *et al.* (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313, 101–104
49 Steiner, C.C. *et al.* (2007) Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol.* 5, 1880–1889
50 Seehausen, O. *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature* 455, 620–623
51 Hawthorne, D.J. and Via, S. (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* 412, 904–907
52 Feder, J.L. *et al.* (2003) Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10314–10319
53 Moyle, L.C. and Graham, E.B. (2005) Genetics of hybrid incompatibility between *Lycopersicon esculentum* and *L. hirsutum*. *Genetics* 169, 355–373
54 Shaw, K.L. and Lesnick, S.C. (2009) Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9737–9742
55 Mihola, O. *et al.* (2009) A mouse speciation gene encodes a meiotic histone h3 methyltransferase. *Science* 323, 373–375
56 Wu, C.I. (2001) The genic view of the process of speciation. *J. Evol. Biol.* 14, 851–865
57 Hey, J. (2006) Recent advances in assessing gene flow between diverging populations and species. *Curr. Opin. Genet. Dev.* 16, 592–596
58 Via, S. and West, J. (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.* 17, 4334–4345
59 Nosil, P. *et al.* (2009) Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* 18, 375–402
60 Via, S. (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* 16, 381–390
61 Drès, M. and Mallet, J. (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. Lond. Ser. B: Biol. Sci.* 357, 471–492
62 Gow, J.L. *et al.* (2006) Contrasting hybridization rates between sympatric three-spined sticklebacks highlight the fragility of reproductive barriers between evolutionarily young species. *Mol. Ecol.* 15, 739–752
63 Price, T.D. (2007) *Speciation in Birds*, Roberts and Co.
64 Seehausen, O. *et al.* (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Mol. Ecol.* 17, 30–44
65 Presgraves, D.C. (2007) Does genetic conflict drive rapid molecular evolution of nuclear transport genes in *Drosophila*? *Bioessays* 29, 386–391
66 Presgraves, D.C. (2007) Speciation genetics: epistasis, conflict and the origin of species. *Curr. Biol.* 17, R125–R127
67 Johnson, N.A. (2010) Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends Genet.* 26, 317–325
68 Funk, D.J. *et al.* (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3209–3213
69 Barton, N.H. (2001) The role of hybridization in evolution. *Mol. Ecol.* 10, 551–568
70 Presgraves, D.C. and Stephan, W. (2007) Pervasive adaptive evolution among interactors of the *Drosophila* hybrid inviability gene, Nup96. *Mol. Biol. Evol.* 24, 306–314
71 Masly, J.P. *et al.* (2006) Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science* 313, 1448–1450
72 Lynch, M. and Force, A.G. (2000) The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat.* 156, 590–605
73 Matsuo, T. *et al.* (2007) Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. *PLoS Biol.* 5, 985–996
74 Ortíz-Barrientos, D. and Noor, M.A.F. (2005) Evidence for a one-allele assortative mating locus. *Science* 310, 1467–11467
75 Feder, J.L. and Nosil, P. (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* 64, 1729–1747
76 Feder, J.L. and Nosil, P. (2009) Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution* 63, 3061–3075
77 Powell, J.R. (1997) *Progress and Prospects in Evolutionary Biology: the Drosophila Model*, Oxford University Press

78 Orr, H.A. and Presgraves, D.C. (2000) Speciation by postzygotic isolation: forces, genes and molecules. *Bioessays* 22, 1085–1094

79 Orr, H.A. *et al.* (2004) Speciation genes. *Curr. Opin. Genet. Dev.* 14, 675–679

80 Perez, D.E. and Wu, C.I. (1995) Further characterization of the *Odysseus* locus of hybrid sterility in *Drosophila* – one gene is not enough. *Genetics* 140, 201–206

81 Schemske, D.W. and Bradshaw, H.D. (1999) Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc. Natl. Acad. Sci. U.S.A.* 96, 11910–11915

82 Bradshaw, H.D. and Schemske, D.W. (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426, 176–178