

Reproductive isolation grows on trees

Leonie C. Moyle¹ and Bret A. Payseur²

¹ Department of Biology, Indiana University, Bloomington, IN 47405, USA

² Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA

Despite rapid advances in the complementary fields of speciation genetics and molecular phylogenetics, little attention has been devoted to understanding how genetic changes associated with reproductive isolation accumulate across clades of species. We argue that comparative analyses of reproductive isolation QTL (quantitative trait loci) can generate unique insights into the underlying mechanisms of speciation. By revealing the timing and evolutionary progression of genetic changes underlying reproductive isolation, these approaches allow the direct evaluation of conceptual and theoretical predictions, such as the ‘snowball effect’, within a statistically robust hypothesis-testing framework. We outline the potential and challenges of this synthetic framework, with the aim of stimulating empirical and theoretical progress towards understanding the mechanisms that underlie the origin of species.

Two complementary approaches to understanding speciation

Speciation is of special interest to evolutionary biologists. Although the nature of species has been debated for at least 150 years [1,2], most contemporary researchers accept that one important component of speciation is the evolution of barriers to interbreeding between diverging lineages, because these barriers enable the preservation of evolutionarily independent units. Two complementary fields provide essential information about this process. The reconstruction of phylogenetic history uncovers the relationships and timing of divergence among species, whereas the genetic dissection of reproductive barriers reveals the genetic changes involved in speciation. The last three decades have seen dramatic advances in both fields. From relatively simple models of trait change, sophisticated phylogenetic frameworks for understanding the evolution of DNA sequences and quantitative traits are now well developed, as are the computational tools required for broad-scale application of these approaches [3–6]. The revival of interest in the evolution of traits contributing to reproductive isolation has also been remarkable [2]. Much of this enthusiasm has been stimulated by the application of quantitative trait locus (QTL) mapping and associated molecular and functional approaches, to dissecting the genetic basis of isolating traits in numerous study systems, including both animals and plants (Box 1). In many of these systems, empirical data on reproductive isolation loci are available or emerging for multiple species pairs within the same clade [7–10], and some studies have begun to reconstruct the

evolutionary history of individual ‘speciation loci’ (e.g. Refs [11–14]).

Despite this progress, there are surprisingly few comparative studies of the genetic basis of reproductive isolation, as measured via the predominant approach to speciation genetics, QTL mapping. Here we argue that comparative studies of reproductive isolation QTL can provide unique mechanistic insights that cannot be gained from other strategies, including comparative studies of reproductive isolation phenotypes. We also identify two primary obstacles to comparative studies of isolation QTL: empirical constraints on the generation of appropriate data and a lack of powerful analytical methods. Throughout, we focus specifically on the value of comparative QTL mapping studies of reproductive isolation because QTL mapping is the most common strategy for understanding the genetic architecture of species barriers, and the most feasible approach for generating the required comparative data, especially in non-model systems. Accordingly, when we refer to ‘speciation genetics’ approaches we are referring to QTL mapping studies, although we recognize that the genes that contribute to species barriers can be identified in other ways (Box 1).

Unique biological inferences from comparative analyses of reproductive isolation QTL

Comparative approaches can provide unique insights into speciation. This is amply demonstrated by studies that have examined associations between either rates of speciation or reproductive isolation phenotypes and biological variables predicted to underlie cladogenesis (Box 2). Incorporating phylogenetic history into speciation studies improves biological realism. Modeling the effects of phylogenetic relatedness when examining patterns of divergence recognizes that lineages generally begin as reproductively integrated populations and then progressively accumulate barriers to successful interbreeding (although new species can also, more rarely, arise via hybridization [15]). Comparative studies can also manage the biological and statistical dependencies that arise when examining numerous species pairs that have individual species or sections of their phylogenetic history in common (Box 2). For example, isolating barriers appearing between two taxa early in their evolutionary divergence are likely to be present in all species subsequently descended from each of these progenitor lineages. Most importantly, by providing a statistically robust framework in which more than one species or species pair (for reproductive isolation studies) can be analyzed, comparative speciation studies can evaluate broader evolutionary patterns within and between groups (Box 2). Although detailed studies of

Corresponding author: Moyle, L.C. (lmoyle@indiana.edu).

Box 1. The genetics of reproductive isolation

An explosion of research into the genetic basis of species barriers has occurred in the last 30 years [2]. Quantitative trait locus (QTL) mapping, the most widespread of these approaches, involves crossing two genotypes (often different species) to generate recombinant hybrid populations. In these populations, taxon-specific molecular markers are used to locate chromosomal regions (QTL) that are statistically associated with the expression of reduced fitness at one or more reproductive stage(s) in hybrid individuals (see Box 3 Figure 1). For example, post-zygotic isolation QTL are chromosomal locations that are significantly associated with the expression of sterility or inviability phenotypes. The goal of these QTL studies is to describe the number of loci involved, the sizes of their phenotypic effects, their modes of gene action, their locations in the genome and (ultimately) their underlying molecular bases.

QTL mapping can identify the genetic basis of a variety of isolating barriers. This approach has been applied to trait differentiation that might contribute to prezygotic isolation between species [39], in some cases establishing a direct mechanistic link between specific QTL differences and prezygotic isolating barriers [63]. QTL underlying hybrid intermediacy that could contribute to extrinsic (environment-specific) postzygotic reproductive isolation have also been identified in several systems [44,64]. Perhaps most commonly, studies have successfully mapped QTL underlying intrinsic hybrid sterility or inviability, including between animal species from *Drosophila*

[35,36,38], mosquito [65], and mouse [40–42], and plant species from *Solanum* [22,66], *Helianthus* [67], *Oryza* [10,68] and *Mimulus* [8,69], among other organisms. Individually, each of these studies provides a snapshot of the ‘genetic architecture’ underlying barriers to interbreeding between two lineages. Other complementary approaches, including deletion mapping [37] and hybrid rescue mutations [17] have also been used to uncover the genetic basis of species differences that contribute to reproductive isolation.

It is important to recognize that most QTL identified in mapping studies are chromosomal regions with statistical effects on a phenotype, rather than specific genes. As such, one QTL could contain multiple genetic factors that contribute to trait variation. By contrast, in several cases the precise genetic changes underlying hybrid inviability and sterility have been identified [11,16–21]. Several such ‘speciation genes’ have been analyzed for their clade-wide patterns of molecular evolution, to determine when and where the relevant molecular changes took place [12,13,16,70]. Ultimately these cases provide the greatest resolution about the mechanisms of divergence during speciation. However, because QTL mapping is currently the predominant approach to speciation genetics, and is likely to remain so for all but the most well studied systems into the immediate future, we focus exclusively on this approach when discussing comparative approaches to speciation genetics.

individual species pairs are essential for understanding divergence processes, an important goal of speciation studies is to identify generalities in the mechanisms responsible for the formation of new species.

Although comparative approaches have provided insight into the pattern and process of speciation, this framework has yet to address several fundamental questions about the underlying mechanisms of reproductive isolation, including its genetic basis and detailed evolution through time. By contrast, by identifying the number and phenotypic effects of loci responsible for reproductive isolation, studies of the genetics of speciation between individual species pairs provide this necessary mechanistic information, but only for limited snapshots of divergence time. These single case studies are unable to reconstruct

the order and timing of causal genetic changes, or to evaluate general patterns or predictions about the accumulation of isolation over an entire clade. Achieving these goals requires combining the perspective of comparative approaches with the mechanistic insight of genetic studies, to develop comparative genetics of speciation. Ideally, this new approach would focus on the individual genes involved in species barriers. However, the identification of these genes remains a significant challenge, even though the list of genes involved in species barriers continues to grow [11,16–21] (Box 1). As long as QTL mapping remains the ‘state of the art’ for most empirical systems, comparative studies will focus on this approach. Table 1 summarizes a suite of problems we propose would be best addressed with comparative analyses of reproductive isolation QTL.

Box 2. Phylogenetic comparative approaches to speciation

Closely related species often share biological features solely because of their recent common ancestry. Accordingly, hypothesis tests that include multiple closely related species must take into account the potential statistical and biological dependencies that can arise from shared history. In order to understand the evolutionary dynamics of character change [6,61], comparative approaches aim to differentiate evolutionary trends or patterns caused by common underlying mechanism(s) from those that simply result from this shared phylogenetic history

Phylogenetic comparative approaches have contributed substantially to our understanding of speciation. By reconstructing the phylogenetic history of large clades, heterogeneity in the relative rates of speciation and extinction can be measured [71–73]. The comparative framework also allows the identification of associations between measures of diversification (including speciation rates and reproductive isolation phenotypes) and biological variables thought to underlie cladogenesis (reviewed in Ref. [2]). For example, comparative studies have suggested correlations between elevated rates of speciation and the frequency or intensity of sexual selection [74], the switch to phytophagy in insects [75], or differences in the parasite species richness among host species [76].

Comparative studies have also specifically examined the accumulation of reproductive isolation phenotypes by evaluating the relationship between phylogenetic distance (as a proxy for time since

divergence) and one or more metrics of pairwise reproductive isolation, using multiple pairs of species in a clade (e.g. Refs [77–80], reviewed in Ref. [2]). In general, these studies reveal an increase in the strength of reproductive isolation phenotypes over increasing evolutionary divergence as well as, in some cases, differences in the rates of accumulation of isolation phenotypes that act at different reproductive stages [2]. These studies have also incorporated or adapted phylogenetic comparative methods to take into account the statistical and biological dependencies that come from examining multiple species pairs that share one or more individual taxa in common. For example, a modified Phylogenetic Independent Contrasts method [3] uses nested averaging of reproductive isolation to reduce all pairwise comparisons across a node to a single comparison [77,81].

Appropriately corrected comparative methods have also been used to examine more directed questions about the evolution of isolation phenotypes. For example, Noor [82] used a phylogenetically based comparative method to evaluate evidence for ‘reinforcement’ by comparing the strength of prezygotic isolation between triads of *Drosophila* species, each of which includes one sympatric and one allopatric species pair. In this way, rather than posing a statistical problem of non-independence, the shared evolutionary history of species is used as a tool to provide insight into the process of speciation.

Table 1. Hypotheses about the evolution of reproductive isolation that can be tested with comparative isolation mapping

Prediction	Testable null hypothesis
Postzygotic isolation grows faster than linearly with time ('snowball effect' [8])	Number of QTL contributing to isolation is linearly proportional to branch length
Prezygotic isolation evolves before postzygotic isolation (e.g. due to rapid evolution of traits involved in sexual selection [87] or sexual conflict [88])	No difference in where prezygotic and postzygotic isolation QTL occur on the phylogeny
Prezygotic isolation evolves by reinforcement	Clades that have species with historical range overlaps accumulate isolation QTL at the same rate as clades without range overlaps
Isolation is driven by natural or sexual selection	Number of isolation QTL is linearly proportional to branch length; Isolation QTL are uncorrelated with phylogenetic shifts in putative agents of selection
Isolation is catalyzed by or causes shifts in organismal biology (e.g. outcrossing rate, recombination rate)	Isolation QTL are randomly distributed with respect to trait shifts on the phylogeny

Below, as illustrations, we expand on two specific applications of this approach to salient issues in speciation genetics.

The evolutionary order and timing of reproductive isolation mutations

Reproductive isolation QTL that differ between species pairs must have evolved along non-shared branches of a phylogeny (Box 3 Figure I). Conversely, where shared QTL among more than one species cross are due to the same underlying genetic change, this change must have occurred along an evolutionary branch shared by all species involved in the comparison. Thus, simply identifying homologous versus lineage-specific QTL in a phylogenetic context can broadly delimit the timing of the origin of new 'isolating' alleles at these loci [22]. Precisely pinpointing the evolutionary timing of shared alleles is likely to be more complex, although additional biological information could help to resolve this challenge (Box 3).

The ability to delimit the evolutionary timing of QTL allows hypothesis testing about the relative importance of specific genetic changes during different stages of divergence. For example, there is considerable debate about which isolating barriers are first associated with species differentiation [2]. Examining the accumulation of different isolation phenotypes over increasing evolutionary distances can go some way to addressing this question (Box 2). However, by ordering the evolution of isolating QTL in time, comparative analyses can implicate which loci and corresponding barriers evolved 'early' versus 'late' with respect to an evolutionary split between two lineages. Such studies also enable other mechanistic observations, including whether the relative size (magnitude of phenotypic effect) of these traits differs between early versus late accumulating QTL. Although there are well developed expectations about the distribution of effect sizes of loci underlying adaptive change [23,24], comparable predictions are not yet available for reproductive isolation loci. Conceivably, the effect sizes of isolation loci might be expected to increase over evolutionary time because the potential for incompatibility will grow as genomes and developmental systems become increasingly divergent.

Similar comparisons of relative timing or changes in effect size could be conducted between different classes of reproductive barrier or different ecological conditions. Testable predictions include how and when different kinds of isolating barrier are expected to develop, as well as the influence of external (e.g. geographical), biological (e.g.

mating system, population structure and genetic architecture of reproductive isolation), or mechanistic (e.g. genetic drift versus natural selection) factors on this rate of accumulation (Table 1). For example, major ecological shifts causing rapid evolutionary change might be responsible for the accelerated accumulation of isolation at one or more stages [25]. This could be tested by evaluating whether evolutionary branches associated with ecological shifts also show enriched accumulation of isolation QTL, in groups with lineages that occur in several distinct ecological conditions.

Evaluating theoretical predictions: The 'snowball effect' for postzygotic reproductive isolation

Much of the available mathematical theory of speciation is population genetic, focusing on the accumulation of mutations that underlie reproductive isolation, rather than on the isolation phenotypes themselves [26]. An especially useful theoretical framework has been the Dobzhansky-Muller model, which describes the evolution of loci underlying postzygotic reproductive isolation (hybrid inviability and sterility) [27–29]. Specifically, this model proposes that hybrid dysfunction is due to negative epistatic interactions between alleles that have independently arisen and been fixed in geographically separate populations. Because incompatibilities are due to disrupted interactions between two or more loci, this model predicts that the number of hybrid incompatibilities ['Dobzhansky-Muller Incompatibilities' (DMIs)] that isolate two lineages should increase faster than linearly with divergence time – the 'snowball effect' [27]. Similar logic predicts that the expected number of DMIs between two species increases exponentially from the root to the tips of a phylogeny [22].

Comparative QTL analyses provide the data necessary to test such predictions. For example, in Box 3 Figure I the snowball effect between species 1 and 3 can be tested because the rooted tree topology (1, (2,3)) allows each isolation QTL acting between these two species to be assigned to branches above and below the most recent common ancestor of species 2 and 3. Although there is strong empirical support for the mechanisms postulated to produce the snowball effect – including the existence of DMIs – the effect itself has never been tested. Statistically and biologically robust comparative analyses of isolation loci are essential for evaluating theoretical expectations such as the snowball hypothesis. These analyses could also provide a framework for evaluating more directed

Box 3. Reconstructing the sequence of origin of reproductive isolation QTL by comparative mapping

Given a known set of phylogenetic relationships among a group of species, and high-quality QTL mapping data from multiple species pairs within this group (Figure 1a), information about the evolutionary order and timing of specific QTL can be inferred (Figure 1b). For illustration we show only a three species phylogeny, although larger trees can provide more informative and robust comparative analyses. Among the three depicted species in the Figure, isolation QTL unique to species pair (1,2) or (1,3) must have evolved along nonshared evolutionary branches between species 1 and 2 (branch c), or species 1 and 3 (branch b), or species 1 and 3 (branch a). (Note that this reconstruction says nothing about the allelic changes that underlie isolation between species 2 and 3.) Reconstructing the timing of origin of unique QTL is therefore relatively straightforward in the scenario where isolation alleles, once gained between two lineages, permanently contribute to barriers between all descendent species of these lineages. This assumption might be more or less reasonable, depending on the type of barrier under consideration. For example, intrinsic postzygotic isolating barriers are generally considered 'irreversible' [2]. By contrast, for potentially more labile traits such as those contributing to ecologically mediated isolation or differentiation in reproductive timing, transitions between 'isolating' and 'nonisolating' states might be more frequent. As additional information about the lability of different isolating traits accumulates, more complex phylogenetic models that incorporate reversible changes can be constructed.

In contrast to delimiting the timing of unique QTL, at least two challenges face the phylogenetic reconstruction of QTL apparently shared among more than one species pair. The first is largely empirical: shared loci are identified as co-localizing to the same genomic region; however, co-localized QTL need not be underpinned by exactly the same (i.e. evolutionarily homologous) mutation. (By contrast, in cases where the precise DNA change underlying species barriers is known, it is more straightforward to establish homology via sequence comparisons). This challenge can be met empirically. For example, if co-localized QTL from different species crosses are homologous, they are not expected to complement each other (i.e. restore species interfertility) when genetically combined. Therefore,

even without knowing the underlying DNA change, standard genetic approaches can be used to evaluate hypotheses about underlying homology.

The second (arguably larger) challenge is analytical. For any single allelic change that is shared (homologous) among more than one species pair, there are always at least two evolutionary scenarios consistent with its evolutionary origin, and therefore at least two possible phylogenetic reconstructions of the underlying mutation. For example, in Figure 1, shared (homologous) QTL between sp1-sp2 and sp1-sp3 hybrids could have evolved along branch a or d. In some cases, resolution might be relatively straightforward. For example, in each QTL mapping population, the allele that is consistently associated with the expression of incompatibility could suggest which species has the derived allele, and therefore delimit the evolutionary change to one of the two alternative branches. Nonetheless, resolution is likely to be more complex for some reproductive isolation traits. The genetic changes that typically underlie postzygotic hybrid incompatibility (DMIs), for example, are necessarily epistatic; they are the product of disrupted interactions caused by allelic changes at two or more loci. Therefore, for these traits an additional problem to be considered is which of the two (or more) genetic interactors is being mapped at a given QTL. This complexity is amplified by the possibility that DMIs could be due either to interactions between derived alleles in the two diverging lineages or to derived-ancestral interactions (i.e. interactions between loci that have both undergone substitutions in one of the lineages [27]), although theory indicates that the former interactions should occur more commonly [27]. In principle, reconstructions are possible for multiple partners when epistatic combinations of loci are identified through mapping, but will be more complex than reconstructions for loci underlying non-epistatic loci that can contribute to other forms of reproductive isolation.

This challenge of deciding which of the two (or more) possible evolutionary scenarios is better explained by shared QTL data is one that could benefit from further theoretical attention. The addition of more species will also probably aid in resolving the phylogenetic timing of shared genetic changes.

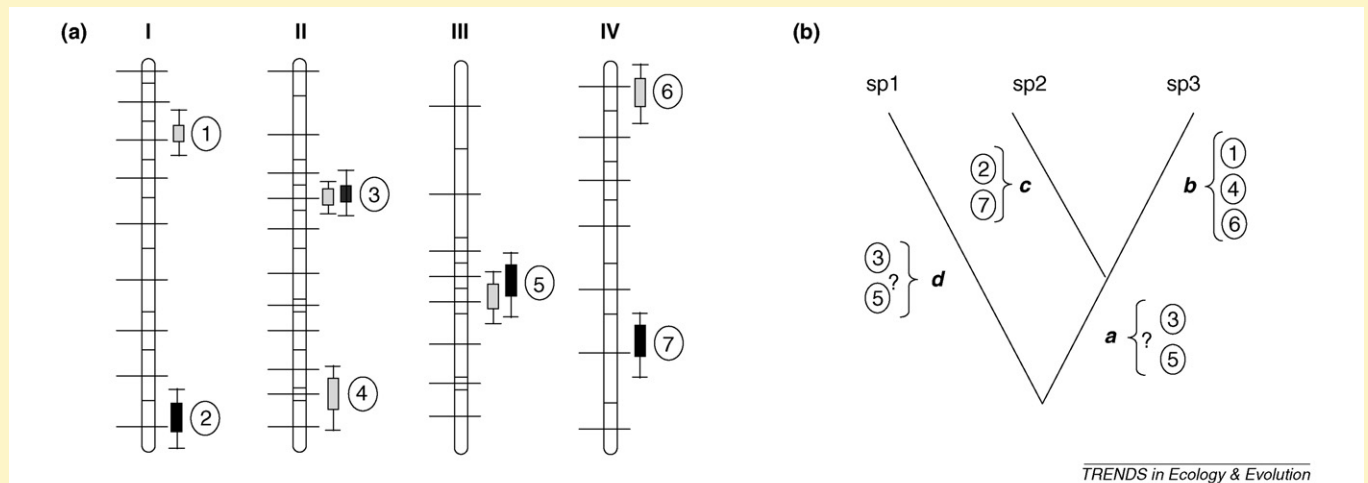


Figure 1. Evolutionary accumulation of reproductive isolation loci. (a) Mapping data showing the genomic location of isolation QTL acting between species pairs (1,2) and (1,3). QTL are labeled with circled numbers. Roman numerals indicate chromosomes 1 to 4. (b) A rooted phylogeny describing relationships among the three depicted species, and the proposed evolutionary timing of changes contributing to reproductive isolation based on QTL that are unique to either species pairs (1,2) or (1,3) or shared (homologous) between (1,2) and (1,3). Circled numbers correspond to QTL labeled in (a). Question-marks indicate ambiguous evolutionary reconstructions for the adjacent QTL. Letters a, b, c, d label the four evolutionary branches on this three species phylogeny.

hypotheses about subsets of species pairs within a clade, such as whether the number of shared hybrid incompatibility loci between different groups of species pairs is greater (or less) than would be expected based on the known phylogenetic relationships between species. Devi-

ations from expected values could occur if a particular evolutionary branch (or branches) had experienced an acceleration or deceleration of DMI accumulation. Such tests also could be applied to data on QTL for other kinds of isolating barriers (Table 1). Finally, existing comparative

Opinion

methods could be modified to test for a direct association between the rate of accumulation of DMIs and the number of speciation events (nodes) within a phylogeny [30].

Challenges to comparative isolation mapping

Given the accepted value of comparative analyses in evolutionary biology, and the rapid proliferation of QTL underlying reproductive isolation in the literature, why are there not more comparative analyses of isolation QTL? Limitations on these studies are partly explained by the logistical and practical challenges of gathering appropriate data from multiple taxon pairs within a clade. QTL mapping is expensive and time-consuming, and has often proceeded on an *ad hoc* basis depending upon the interests of particular researchers. Nonetheless, these factors do not completely explain the lack of comparative QTL analyses of reproductive isolation, for there are many comparative QTL analyses of other kinds of quantitative genetic traits (e.g. Refs [31–33]). Rather, the primary obstacles appear to be twofold: first, historically there have been unique experimental constraints on QTL studies of reproductive isolation, which have limited the availability of appropriate comparative data; and second, there are some crucial analytical methods and tools that must be developed to address problems specific to reproductive isolation data. Fortunately, neither of these challenges is insurmountable.

Historical experimental limitations on comparative isolation QTL data

Genetic analyses of reproduction isolation have unique experimental constraints. In particular, the most appropriate ‘evolutionary window’ in which to examine the genetics of species barriers is more circumscribed than for other kinds of phenotypes: the taxonomic units examined must be sufficiently diverged to express reproductive isolation (e.g. sub-species and above) but still inter-fertile enough to allow genetic analyses [2]. This problem is amplified in the context of comparative analyses, which require multiple taxon pairs in the same clade.

Nonetheless, the recent explosion of QTL studies of reproductive isolation suggests that this experimental challenge can be met by concentrating on experimentally appropriate biological systems. For example, the expansion of isolation mapping studies to include subspecific taxon pairs has added valuable chronological breadth to our picture of how and when isolation QTL accumulate (e.g. Refs [34,35]). In fact, some of the most successful models for the genetics of speciation satisfy the criteria necessary for comparative analyses of reproductive isolation QTL. *Drosophila*, for example, already has the potential to integrate multiple different mapping studies that examine reproductive isolation at different taxonomic levels or at different stages (e.g. Refs [35–39]). The ecologically diverse plant group *Mimulus* has already been the focus of detailed QTL analyses of barriers to interbreeding among multiple groups at different taxonomic levels [8,34]. Similarly, agricultural and medical model systems including tomato, rice and mouse contain many diverse and inter-fertile taxa, have sufficient molecular genetic tools to enable the development of multiple mapping popu-

lations, and feature studies of isolation phenotypes among one or more taxon pairs [9,10,22,40–42]. Other ecological and evolutionary model systems, such as the plant group *Aquilegia* (columbines) [43] and the animal group *Gasterosteus* (sticklebacks) (e.g. Refs [44–46]), are particularly well suited to studying the genetic basis of prezygotic and extrinsic postzygotic isolation. In addition to these developing systems, comparative mapping might be particularly suited for clades in which the capacity for hybridization is lost relatively slowly, such as birds and some plant groups [47–49], where there is the potential to examine changes spanning longer timescales. Overall, the field is poised for a flood of data that could be used in comparative analyses of isolation QTL.

Analytical and methodological challenges

Genetic mapping data for reproductive isolation also introduce analytical complications that do not apply to other kinds of QTLs. In particular, although we wish to infer the evolution of isolation and associated QTL along specific lineages, reproductive isolation is fundamentally a joint property of two lineages, and the data are collected from pairs of species [2,26]. The best description of such ‘relational’ phenotypes in a comparative analysis is unclear. This problem does not arise for traits whose evolutionary change can be straightforwardly assigned to a single evolutionary lineage or branch. In addition, reproductive barriers might be highly specific to species pairs, precluding inferences about ‘hidden data’ (i.e. tip states that are unknown due to missing data). For example, postzygotic isolation (hybrid inviability and sterility) frequently involves epistasis between different loci in two lineages (DMIs: see above); the hybrid sterility effect of any given substitution is therefore dependent on the particular genetic background in which it is found. Accordingly, knowing the location and number of loci that contribute to sterility interactions between two species (e.g. species 1 and 2 in Box 3 Figure I) might not provide information about the loci that contribute to sterility between each of these lineages and a third, closely related, species (e.g. species 3 in Box 3 Figure I). When isolation QTL analyses are only conducted for a subset of the taxon pairs within a clade, statistical methods will need to address this isolation-specific ‘hidden data’ issue.

In other ways, existing comparative methods provide a useful starting framework. Rates of evolution of QTL number or individual QTL effects could be measured using generalized least squares [50], standardized contrasts [3,51], likelihood [52] or other methods [53]. The presence or absence of particular QTL, the total number of QTL, or the QTL effect size at ancestral nodes could also be inferred using methods developed for discrete [54] or continuous [55] traits. But some modifications will be needed to better model the evolution of reproductive isolation. For example, in contrast to the standard Brownian motion model, which assumes equal probabilities of trait increase and decrease along branches, reproductive isolation typically increases over time.

In addition to these isolation-specific problems, as in any comparative analysis, appropriate methods will also need to accommodate several sources of statistical error,

including uncertainty in the inferred locations and effects of QTL (owing to such factors as the sample sizes in the crosses, the fidelity of the genetic map, and assumptions about the relationship between genotype and phenotype in the mapping approach), and uncertainty in the estimates of the phylogenetic topology and branch lengths. QTL mapping [56] and phylogenetic and comparative methods [57–59] that could account for such statistical uncertainties are available. In the case of species that are only recently diverged and/or only partially reproductively isolated from one another, sources of biological variation could also complicate inference. For example, genome-wide phylogenetic incongruence, which can arise from incomplete lineage sorting and/or introgression via hybridization, could be especially problematic among closely related taxa. These factors could influence both confidence in the reconstruction of relationships and the interpretation of QTL data from mapping populations that are differentially affected by, for example, historical introgression between the species of interest. Although these are problems for all comparative analyses of QTLs, they might be exacerbated in systems used to study reproductive isolation.

Resolutions and future directions

Successfully integrating genetic mapping of reproductive isolation with phylogenetic inference will require the creative attention of a diverse community of researchers, including statistical geneticists, systematists, theoretical population geneticists, and experimentalists studying the genetics of speciation. Existing frameworks can provide some direction and guidance. For example, although much of the available theory that quantitatively describes the evolution of reproductive isolation [26,60] focuses on the evolution of isolating barriers between individual species pairs, many existing models could be expanded to explicitly incorporate the accumulation of barriers among multiple species within a clade. Additionally, as outlined above, existing comparative methods already provide sophisticated analytical approaches for incorporating phylogenetic information into models of trait evolution [3,61]. Similar approaches based on alternative stochastic models, such as Levy processes (which could include ‘phenotypic jump’ components in addition to continuous drift and diffusion) or Brownian motions with reflecting boundaries, could be used to more realistically model the unique characteristics of the evolution of reproductive isolation, including its tendency to increase over time. Similarly, methods used for the phylogenetic depiction of reticulation, hybridization, or recombination [62] could be adapted to represent relational reproductive isolation traits.

From a statistical perspective, solutions are also required for the practicalities of explicitly designing experiments for comparative isolation mapping, and for handling, analyzing, and interpreting the experimental results. For the latter, QTL mapping approaches that test hypotheses about shared versus unshared QTL (Box 4) are already available. These strategies could be modified to include phylogenetic information. More broadly, a useful goal would be the joint analysis of mapping data and phylogenetic history. Ideally, such an approach would

Box 4. Comparing QTL in Multiple Mapping Populations

Evolutionary change in the genetic architecture of a trait can be examined by measuring the same phenotype in mapping populations generated by crossing different parental strains. In particular, evidence for a QTL in a specific genome interval can be statistically compared among different crosses [83–85].

Li *et al.* [84] provided a straightforward algorithm for jointly analyzing mapping data from multiple inbred-line crosses. Mapping is first conducted separately in each cross to obtain an initial indication of the genetic patterns in the data. Individual phenotypes and marker genotypes are then combined across mapping populations, recoding the parental marker alleles in a binary fashion ([84] suggest recoding based on phenotypic patterns). Finally, the combined set of phenotypes and genotypes are analyzed. By conducting separate analyses that treat the indicator variable *cross* as either an additive covariate or an interactive (*cross* × QTL) covariate, shared and cross-specific QTL (respectively) are identified.

The increased sample size that comes from combining multiple mapping populations brings both higher statistical power for detecting QTL and higher mapping resolution when QTL are shared among crosses [83,86]. Combined cross analysis also enables the joint consideration of mapping populations generated in different laboratories and/or with different genetic structures (e.g. F_2 , RIL, NIL). Most important, divergence in QTL patterns among crosses can be tested statistically, rather than “eye-balling” the overlap of LOD-score peaks from different mapping populations. Hypotheses of particular interest to studies of the evolution of reproductive isolation, including changes in the epistatic partners that underlie Dobzhansky-Muller incompatibilities, can also be evaluated using this framework.

We see advantages to directly incorporating phylogenetic information into methods such as combined cross analysis (which currently ignore the evolutionary relationships among the parental strains). For example, phylogenetics provides a natural approach for recoding parental marker alleles. This approach would be particularly useful for “transgressive” traits such as reproductive isolation, where the lack of divergence between the parental strains (both are fertile) provides little intuition for recoding based on the phenotype. Furthermore, the probability that QTL are cross-specific should be affected by the phylogeny of the crossed strains, and this dependence could be incorporated into the mapping algorithm by adding parameters that reflect phylogenetic relationships.

simultaneously account for uncertainty in phylogenetic reconstruction, QTL parameters, and the relationships between them.

Conclusions

In their landmark book, *Speciation*, Coyne and Orr describe the third (contemporary) phase of speciation research as involving five themes: genetic analysis, ‘mathematical theory, an emphasis on ecology, molecular analysis, and the use of comparative studies’ (p. 5, Ref. [2]). Surely an additional theme should be the fusion of two or more of these traditionally separate approaches. Comparative studies of reproductive isolation loci can combine the power of QTL mapping with the broad scope of the comparative perspective to generate unique insights into the mechanistic basis of reproductive isolation. These studies enable the testing of mathematical theory and the evaluation of ecological and evolutionary mechanisms of speciation, within a statistically and biologically robust framework. In doing so, they will help to build a more complete view of the evolution of reproductive isolation and thereby the mechanisms and causes of speciation.

Acknowledgments

Thanks to M. Hahn and four anonymous reviewers for thoughtful and constructive comments on earlier versions of the manuscript, to C. Ané for useful discussions on phylogenetic comparative methods, and to M. Turelli and members of the Moyle and Payseur labs for contributions to our thinking on speciation. This research was supported by NSF grant DEB 0918000 to BAP, and NSF grant DEB 0849157 to LCM.

References

- 1 Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*, John Murray
- 2 Coyne, J.A. and Orr, H.A. (2004) *Speciation*, Sinauer Assoc., Inc
- 3 Felsenstein, J. (1985) Phylogenies and the comparative method. *Am. Nat.* 125, 1–15
- 4 Swofford, D.L. *et al.* (1996) Phylogenetic inference. In *Molecular Systematics* (Hillis, D.H. *et al.*, eds), pp. 407–514, Sinauer
- 5 Huelsenbeck, J.P. and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755
- 6 Felsenstein, J. (2004) *Inferring Phylogenies*, Sinauer
- 7 Orr, H.A. *et al.* (2007) Speciation in *Drosophila*: From phenotypes to molecules. *J. Hered.* 98, 103–110
- 8 Wu, C.A. *et al.* (2008) *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100, 220–230
- 9 Moyle, L.C. (2008) Ecological and evolutionary genomics in the wild tomatoes (*Solanum* Sect. *Lycopersicon*). *Evolution* 62, 299–3013
- 10 Koide, Y. *et al.* (2008) Genetics of speciation in rice. *Biotechnol. Agri. For.* 247–259
- 11 Presgraves, D.C. *et al.* (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* 423, 715–719
- 12 Ting, C.T. *et al.* (2004) Gene duplication and speciation in *Drosophila*: Evidence from the Odysseus locus. *Proc. Nat. Acad. Sci. U.S.A.* 101, 12232–12235
- 13 Maheshwari, S. *et al.* (2008) Recurrent positive selection of the *Drosophila* hybrid incompatibility gene Hmr. *Mol. Biol. Evol.* 25, 2421–2430
- 14 Cattani, M.V. and Presgraves, D.C. (2009) Genetics and lineage-specific evolution of a lethal hybrid incompatibility between *Drosophila mauritiana* and its sibling species. *Genetics* 181, 1545–1555
- 15 Arnold, M.L. (1997) *Natural Hybridization and Evolution*, Oxford University Press
- 16 Ting, C.T. *et al.* (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282, 1501–1504
- 17 Barbash, D.A. *et al.* (2003) A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proc. Nat. Acad. Sci. U.S.A.* 100, 5302–5307
- 18 Brideau, N.J. *et al.* (2006) Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* 314, 1292–1295
- 19 Phadnis, N. and Orr, H.A. (2009) A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* 323, 376–379
- 20 Mihola, O. *et al.* (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323, 373–375
- 21 Tang, S.W. and Presgraves, D.C. (2009) Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science* 323, 779–782
- 22 Moyle, L.C. and Nakazato, T. (2008) Comparative genetics of hybrid incompatibility: Sterility in two *Solanum* species crosses. *Genetics* 179, 1437–1453
- 23 Orr, H.A. (1998) The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* 52, 935–949
- 24 Orr, H.A. (2003) The distribution of fitness effects among beneficial mutations. *Genetics* 163, 1519–1526
- 25 Schluter, D. (2000) *The Ecology of Adaptive Radiation*, Oxford University Press
- 26 Turelli, M. *et al.* (2001) Theory and speciation. *Trends Ecol. Evol.* 16, 330–343
- 27 Orr, H.A. (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* 139, 1805–1813
- 28 Turelli, M. and Orr, H.A. (2000) Dominance, epistasis and the genetics of postzygotic isolation. *Genetics* 154, 1663–1679
- 29 Turelli, M. and Moyle, L.C. (2007) Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* 176, 1059–1088
- 30 Pagel, M. *et al.* (2006) Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* 314, 119–121
- 31 Causse, M. *et al.* (2007) Both additivity and epistasis control the genetic variation for fruit quality traits in tomato. *Theor. Appl. Genet.* 115, 429–442
- 32 Larson, S.R. and Mayland, H.F. (2007) Comparative mapping of fiber, protein, and mineral content QTLs in two interspecific *Leymus* wildrye full-sib families. *Mol. Breed.* 20, 331–347
- 33 Lou, P. *et al.* (2007) Quantitative trait loci for flowering time and morphological traits in multiple populations of *Brassica rapa*. *J. Exp. Bot.* 58, 4005–4016
- 34 Hall, M.C. *et al.* (2006) Pleiotropic quantitative trait loci contribute to population divergence in traits associated with life-history variation in *Mimulus guttatus*. *Genetics* 172, 1829–1844
- 35 Reed, L.K. *et al.* (2008) Genetic architecture of hybrid male sterility in *Drosophila*: analysis of intraspecific variation for interspecific isolation. *PLoS ONE* 3, e3076
- 36 Tao, Y. *et al.* (2003) Genetic dissection of hybrid Incompatibilities between *Drosophila simulans* and *D. mauritiana*. II. Mapping hybrid male sterility loci on the third chromosome. *Genetics* 164, 1399–1418
- 37 Presgraves, D.C. (2003) A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. *Genetics* 163, 955–972
- 38 Chang, A.S. and Noor, M.A.F. (2007) The genetics of hybrid male sterility between the allopatric species pair *Drosophila persimilis* and *D. pseudoobscura bogotana*: dominant sterility alleles in collinear autosomal regions. *Genetics* 176, 343–349
- 39 Moehring, A.J. *et al.* (2006) The genetic basis of prezygotic reproductive isolation between *Drosophila santomea* and *D. yakuba* due to mating preference. *Genetics* 173, 215–223
- 40 Storchova, R. *et al.* (2004) Genetic analysis of X-linked hybrid sterility in the house mouse. *Mamm. Genome* 15, 515–524
- 41 Oka, A. *et al.* (2007) Disruption of genetic interaction between two autosomal regions and the X chromosome causes reproductive isolation between mouse strains derived from different subspecies. *Genetics* 175, 185–197
- 42 Good, J.M. *et al.* (2008) A complex genetic basis to X-linked hybrid male sterility between two species of house mice. *Genetics* 179, 2213–2228
- 43 Hodges, S.A. and Kramer, E.M. (2007) Columbines. *Curr. Biol.* 17, R992–R994
- 44 Peichel, C.L. *et al.* (2001) The genetic architecture of divergence between threespine stickleback species. *Nature* 414, 901–905
- 45 Kitano, J. *et al.* (2008) Divergence of male courtship displays between sympatric forms of anadromous threespine stickleback. *Behaviour* 145, 443–461
- 46 Barrett, R.D.H. *et al.* (2008) Natural selection on a major armor gene in threespine stickleback. *Science* 322, 255–257
- 47 Price, T.D. (2002) Domesticated birds as a model for the genetics of speciation by sexual selection. *Genetica* 116, 311–327
- 48 Price, T.D. and Bouvier, M.M. (2002) The evolution of F1 postzygotic incompatibilities in birds. *Evolution* 56, 2083–2089
- 49 Levin, D.A. (2000) *The Origin, Expansion, and Demise of Plant Species*, Oxford University Press
- 50 Martins, E.P. (1994) Estimating the rate of phenotypic evolution from comparative data. *Am. Nat.* 144, 193–209
- 51 Garland, T. (1992) Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am. Nat.* 140, 509–519
- 52 O'Meara, B.C. *et al.* (2006) Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60, 922–933
- 53 Lynch, M. (1991) Methods for the analysis of comparative data in evolutionary biology. *Evolution* 45, 1065–1080
- 54 Pagel, M. (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *System. Biol.* 48, 612–622
- 55 Schluter, D. *et al.* (1997) Likelihood of ancestor states in adaptive radiation. *Evolution* 51, 1699–1711
- 56 Lynch, M. and Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits*, Sinauer
- 57 Pagel, M. *et al.* (2004) Bayesian estimation of ancestral character states on phylogenies. *System. Biol.* 53, 673–684

- 58 Martins, E.P. (1996) Conducting phylogenetic comparative studies when the phylogeny is not known. *Evolution* 50, 12–22
- 59 Carstens, B.C. and Knowles, L.L. (2007) Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: An example from *Melanoplus* grasshoppers. *System. Biol.* 56, 400–411
- 60 Gavrilets, S. (2004) *Fitness Landscapes and the Origin of Species*, Princeton Univ. Press
- 61 Harvey, P.H. and Pagel, M.D. (1991) *The Comparative Method in Evolutionary Biology*, Oxford Univ. Press
- 62 Huson, D.H. and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267
- 63 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
- 64 Lexer, C. *et al.* (2005) Genetics of species differences in the wild annual sunflowers, *Helianthus annuus* and *H-petiolaris*. *Genetics* 169, 2225–2239
- 65 Slotman, M. *et al.* (2004) The genetics of inviability and male sterility in hybrids between *Anopheles gambiae* and *Anopheles arabiensis*. *Genetics* 167, 275–287
- 66 Moyle, L.C. and Graham, E.B. (2005) Genetics of hybrid incompatibility between *Lycopersicon esculentum* and *L. hirsutum*. *Genetics* 169, 355–373
- 67 Rieseberg, L.H. *et al.* (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152, 713–727
- 68 Matsubara, K. *et al.* (2007) Identification and linkage mapping of complementary recessive genes causing hybrid breakdown in an intraspecific rice cross. *Theor. Appl. Genet.* 115, 179–186
- 69 Sweigart, A.L. *et al.* (2006) A simple genetic incompatibility causes hybrid male sterility in *Mimulus*. *Genetics* 172, 2465–2479
- 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 Barraclough, T.G. *et al.* (1998) Revealing the factors that promote speciation. *Philos. Trans. Roy. Soc. B* 353, 241–249
- 72 Nee, S. (2001) Inferring speciation rates from phylogenies. *Evolution* 55, 661–668
- 73 Pybus, O.G. and Harvey, P.H. (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. Roy. Soc. Lond. B* 267, 2267–2272
- 74 Arnqvist, G. *et al.* (2000) Sexual conflict promotes speciation in insects. *Proc. Nat. Acad. Sci. U.S.A.* 97, 10460–10464
- 75 Mitter, C. *et al.* (1988) The phylogenetic study of adaptive zones - Has phytophagy promoted insect diversification. *Am. Nat.* 132, 107–128
- 76 Nunn, C.L. *et al.* (2004) Parasites and the evolutionary diversification of primate clades. *Am. Nat.* 164, S90–S103
- 77 Coyne, J.A. and Orr, H.A. (1989) Patterns of speciation in *Drosophila*. *Evolution* 43, 362–381
- 78 Sasa, M.M. *et al.* (1998) Patterns of postzygotic isolation in frogs. *Evolution* 52, 1811–2182
- 79 Mendelson, T.C. (2003) Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (*Percidae: Etheostoma*). *Evolution* 57, 317–327
- 80 Moyle, L.C. *et al.* (2004) Patterns of reproductive isolation in three angiosperm genera. *Evolution* 58, 1195–1208
- 81 Fitzpatrick, B.M. (2002) Molecular correlates of reproductive isolation. *Evolution* 56, 191–198
- 82 Noor, M.A.F. (1997) How often does sympatry affect sexual isolation in *Drosophila*? *Am. Nat.* 149, 1156–1163
- 83 Walling, G.A. *et al.* (2000) Combined analyses of data from quantitative trait loci mapping studies: Chromosome 4 effects on porcine growth and fatness. *Genetics* 155, 1369–1378
- 84 Li, R.H. *et al.* (2005) Combining data from multiple inbred line crosses improves the power and resolution of quantitative trait loci mapping. *Genetics* 169, 1699–1709
- 85 Guo, B. *et al.* (2006) Pooled analysis of data from multiple quantitative trait locus mapping populations. *Theor. Appl. Genet.* 113, 39–48
- 86 Lander, E. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Gene.* 11, 241–247
- 87 Snook, R.R. *et al.* (2009) Interactions between the sexes: new perspectives on sexual selection and reproductive isolation. *Evol. Ecol.* 23, 71–91
- 88 Gavrilets, S. and Hayashi, T.I. (2005) Speciation and sexual conflict. *Evol. Ecol.* 19, 167–198