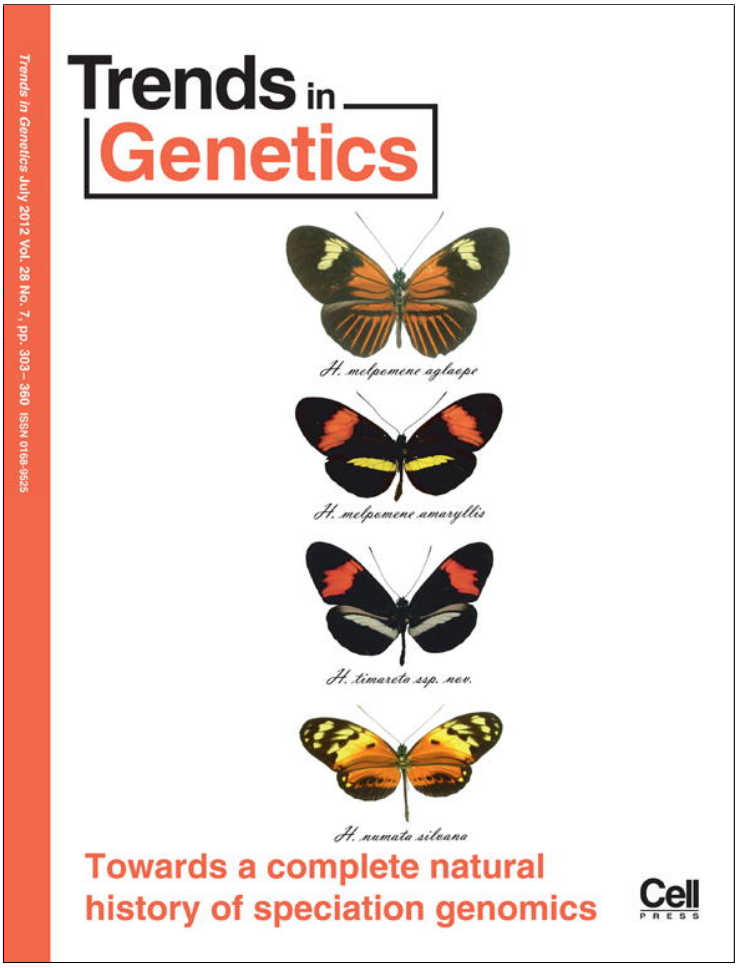

Trends in Genetics July 2012 Vol. 28 No. 7, pp. 303–360 ISSN 0168-9525
Trends in Genetics
H. melpomene aglaope
H. melpomene amaryllis
H. timareta ssp. nov.
H. numata silvana
Towards a complete natural history of speciation genomics
Cell PRESS

Review

The genomics of speciation-with-gene-flow

Jeffrey L. Feder[1], Scott P. Egan[1,2] and Patrik Nosil[3,4]

[1] Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA
[2] Advanced Diagnostics and Therapeutics, University of Notre Dame, Notre Dame, IN 46556, USA
[3] Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, CO 80309, USA
[4] Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

The emerging field of speciation genomics is advancing our understanding of the evolution of reproductive isolation from the individual gene to a whole-genome perspective. In this new view it is important to understand the conditions under which 'divergence hitchhiking' associated with the physical linkage of gene regions, versus 'genome hitchhiking' associated with reductions in genome-wide rates of gene flow caused by selection, can enhance speciation-with-gene-flow. We describe here a theory predicting four phases of speciation, defined by changes in the relative effectiveness of divergence and genome hitchhiking, and review empirical data in light of the theory. We outline future directions, emphasizing the need to couple next-generation sequencing with selection, transplant, functional genomics, and mapping studies. This will permit a natural history of speciation genomics that will help to elucidate the factors responsible for population divergence and the roles that genome structure and different forms of hitchhiking play in facilitating the genesis of new biodiversity.

Speciation genomics: perspective and key issues

Next-generation sequencing (NGS) is imparting a fresh genomics perspective to an old and evolving question, speciation. Substantial progress has been made in the past decades on resolving the importance of different factors and traits causing speciation [1–6]. In addition, individual 'speciation genes' contributing to reproductive isolation have been identified [2,7–12]. However, we lack a thorough understanding of (i) how these speciation genes are organized in the genomes of diverging populations (i.e. genome structure, the extent to which loci are physically linked on chromosomes and reside within structural features restricting recombination, such as inversions), and (ii) the significance of genome structure for speciation. In this regard, a crucial issue concerns the relative importance of different types of genetic hitchhiking in facilitating speciation when gene flow occurs between populations. Advances in NGS are enabling researchers to address these questions through comparative studies of genome-wide patterns of differentiation between populations at varying stages along the speciation continuum from partially isolated races to fully isolated taxa. Although these populations do not form a direct evolutionary progression, and in particular instances may represent taxa in selection/migration equilibrium, inferences can nevertheless be made concerning the roles that genome structure and associated hitchhiking processes play in speciation for populations that are increasingly isolated.

Here, we review results from recent theoretical [13–17] and empirical studies (e.g. [18–25]) in the emerging field of speciation genomics to aid the transition from individual gene to whole-genome understanding of speciation-with-gene-flow. We focus on the issue of the relative importance of 'divergence hitchhiking' (DH) and 'genome hitchhiking' (GH) for facilitating speciation-with-gene-flow [13,14,26] (Glossary). We first examine theory concerning the efficacy

Corresponding author: Feder, J.L. (feder.2@nd.edu).

Keywords: speciation genomics; divergence hitchhiking; genome hitchhiking; gene flow; genome scan; reproductive isolation.

Glossary

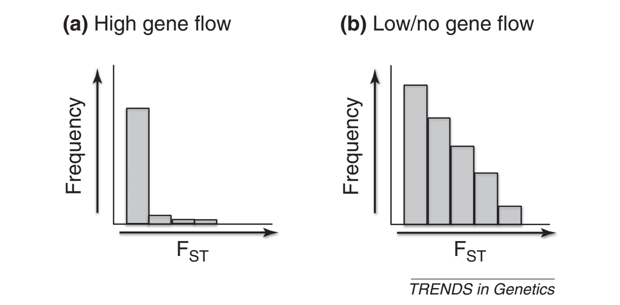
Divergence hitchhiking (DH): a process in which divergent selection on a locus can reduce the effective migration rate for physically linked gene regions and, thus, increase divergence in the surrounding region.
Divergent selection: selection that acts in contrasting directions between two populations, usually with respect to ecological differences between their environments (e.g. large body size confers high survival in one environment and low survival in the other). Divergent selection generates 'extrinsic' reproductive isolation when migrants between environments do not survive well and when their hybrid offspring do not fare as well as resident genotypes.
Fitness epistasis: synergistic effects between loci exceeding the effects of individual loci acting independently of one another that impact upon the fitness of an organism.
Fixation index (F_{ST}): a measure of genetic diversity describing the relative degree of allele frequency differences between populations.
Genome hitchhiking (GH): process in which divergent selection reduces the average effective migration rate globally across the genome fostering increased divergence genome-wide.
Genetic hitchhiking: the change in frequency of an allele in a population due to it being carried along at a higher (or lower) frequency with other gene(s) under selection.
Genome scan: a survey of numerous genetic markers distributed throughout the genome for differentiation between populations.
Genomic island of divergence (speciation): a region of the genome of any size, but usually considered to be relatively small and isolated from other such regions, whose divergence exceeds neutral background expectations in the absence of divergent selection.
Outlier locus: a gene marker showing divergence statistically departing from (usually above) background null or neutral expectations. Outlier loci are often interpreted as being affected by divergent selection and/or causing reproductive isolation and are associated with genomic islands of speciation.
Linkage disequilibrium (LD): non-random associations of alleles at two or more loci. Note that physical linkage facilitates, but does not guarantee, LD, and unlinked markers can sometimes be in LD (e.g. due to admixture in hybrid zones).
Multiplicative fitness interactions: the cumulative effects of individual loci on the overall fitness of an organism determined by the product of individual locus effects.

0168-9525/$ – see front matter © 2012 Elsevier Ltd. All rights reserved. doi:10.1016/j.tig.2012.03.009

Box 1. Genomic divergence depending on gene flow

Genomic divergence in the face of gene flow could be very different from that during allopatric divergence without gene flow. Strictly allopatric divergence, be it via selection or drift, proceeds unfettered by the homogenizing effects of gene flow [31]. Thus, the extent of genetic linkage and recombination among genes relative to the strength of selection is not a major constraint on divergence in allopatry. By contrast, physical linkage relationships and recombination rates among genes, together with levels of gene flow and the strength of selection, are crucial considerations with respect to speciation-with-gene-flow (see main text for details).

In terms of expected empirical patterns, it has been argued that speciation-with-gene-flow will be characterized by divergence in only a few regions that harbor genes under strong divergent selection that causes reproductive isolation [28,29], whereas the rest of the neutral or more weakly-selected genome is homogenized by gene flow. This is predicted to generate an 'L-shaped' frequency distribution of genetic differentiation across loci in the genome (i.e. most loci have low F_{ST} values) (Figure Ia). By contrast, allopatric speciation might be characterized by divergence across much more of the genome (Figure Ib). This can lead to a different distribution of genomic differentiation than is observed with gene flow; specifically a distribution characterized by less skew, more density in the center, and a less pronounced tail of extreme values. Explicit comparisons of the distribution of genomic differentiation for populations varying in their degree of gene flow are lacking. Moreover, even if documented, patterns such as those described above should be interpreted with caution. For example, recent divergence might preclude strong differences for many regions in allopatry and selection on many loci might generate widespread divergence even with gene flow [13,14]. Further theoretical and empirical work on this issue is required.



TRENDS in Genetics

Figure I. Hypothetical distributions of genetic differentiation (F_{ST}) among loci expected for divergence in the face for relatively recently separated taxa experiencing conditions of **(a)** high versus **(b)** low or no gene flow.

of DH and GH for population divergence, framing these processes in the context of a new four-phase model for speciation-with-gene-flow. We show that the theory makes testable predictions about patterns of genomic divergence associated with the four-phase model of speciation-with-gene-flow defined by differences in the relative effectiveness of DH versus GH through time. We next assess currently available data to examine whether patterns of genomic differentiation qualitatively concur with these theoretical predictions. We then discuss difficulties that arise when attempting to infer evolutionary process solely from patterns in genome scans. We conclude by outlining types of future studies that could aid our understanding of a natural history of speciation genomics, resolving the role that genome structure and different forms of genetic hitchhiking play in creating biodiversity.

Geographic modes of speciation

Genome structure is most relevant for speciation-with-gene-flow, where population divergence is generally driven by divergent selection associated with different habitats or environments (Box 1). When gene flow accompanies speciation, an antagonism exists between divergent selection that builds up favorable combinations of locally adapted genes, and migration and recombination that break them down and homogenize populations [27]. Hence, features of genome structure that reduce recombination between populations (e.g. inversions, translocations or centromeres) can enhance the effectiveness of divergent selection by creating and maintaining linkage disequilibrium (LD) and be important for speciation-with-gene-flow. By contrast, there is no antagonism between selection and recombination during speciation between allopatric populations because geographic barriers to migration preclude gene flow; hence, linkage is not as crucial for allopatric divergence [2,24]. The extent to which the physical linkage of genes is needed for speciation-with-gene-flow therefore forms the crux of current debate [28,29].

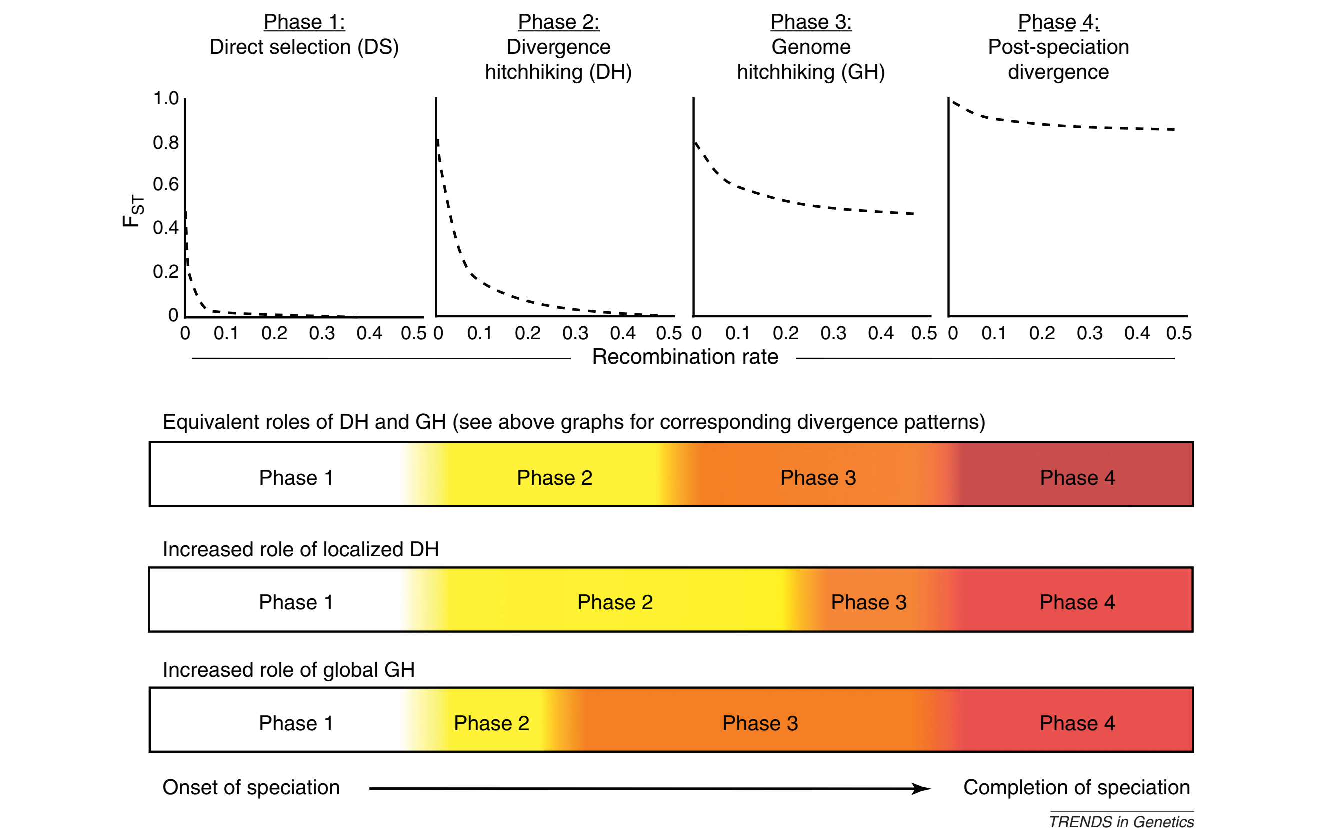
Types of genetic hitchhiking

Divergent selection promotes the genetic differentiation of the target sequences it directly acts upon (i.e. 'direct selection', DS). In addition to these direct effects, genetic hitchhiking caused by divergent selection could potentially facilitate the evolution of increased genomic divergence in two ways. The first way is DH [19,26,30]. When a locus is under divergent selection, not only are the alleles at that locus restricted from moving between populations, but so are nearby linked regions, even if such regions are neutral [31,32]. The reason is that, following migration, selection on the target locus can also eliminate alleles at nearby genes before they have had a chance to recombine and introgress into the resident gene pool. The effective migration rate (m_e) is therefore reduced locally in the genome around selected genes compared to the gross migration rate (m) between populations. When, $m_e < m$, neutral and adaptive divergence can differentially accumulate and generate elevated peaks of F_{ST} in the region around a selected gene compared to the remainder of the genome (Box 2). This is because drift and selection do not have to overcome as great a homogenizing effect from gene flow for linked compared to unlinked loci. Consequently linkage, particularly of new mutations of modest selective advantage to already strongly selected loci, may promote speciation-with-gene-flow.

The second way that the effects of divergent selection can be accentuated is through GH [13,15]. GH occurs when the average rate of m_e is reduced globally across the genome by divergent selection. Instead of the loss of a migrating allele being dictated only by selection against a nearby locus, its fate is tied to the elimination of all maladapted genes in the genome. By reducing the average m_e genome-wide, F_{ST} can become elevated even for unlinked neutral regions compared to expectations based on m (Box 2). A large degree of heterogeneity in m_e will still probably exist through the genome, however, because DS will affect some regions and not others, and DH arising

Box 2. Four phases of speciation-with-gene-flow

An important theoretical result is the prediction of four different phases of speciation-with-gene-flow (Figure I). During the first, initial phase when gene flow is still high, loci directly subject to strong divergent selection relative to migration will tend to diverge independently from other genes (DS is of prime importance). After this, a second, intermediate stage is reached when DH rises in relevance, and new mutations tightly-linked to the few, already diverged genes are now able to differentiate owing to the locally reduced m_e surrounding the selected sites. DH will be most significant when selection on the new mutation is weak relative to migration ($s << 0.5m$). Thus, instead of being crucial, DH may supply a fortuitous push towards speciation when new mutations of weak effect occur close to an already diverged locus or loci. When multiple loci are under divergent selection, the third stage of speciation ensues. Here, the situation changes and m_e begins to become globally reduced across the genome instead of being only locally reduced in a small window around individually diverged genes. GH now facilitates genome-wide divergence, and mutations with modest to weak effects are established across the genome (but heterogeneity among regions is still expected due to variation in selection strengths, recombination rates, and other parameters). This GH-dominated phase 3 can theoretically occur with as few as three loci under strong divergent selection ($s = 0.5$) [13,14]. Previous theory and data concerning hybrid zones following secondary contact have provided important insights concerning this transition from independent clines at selected loci (phases 1 and 2) to a genome-wide barrier to gene flow (phases 3 and 4) as a result of extensive disequilibrium and selection among many loci distributed around the genome [32,50–53]. Eventually, a fourth stage is attained where alleles ecologically favored in one habitat and neutral in the other, as well as universally favored variants, do not introgress readily. At this point, diagnostically fixed differences can accumulate between populations. Key research questions for speciation genomics therefore concern whether these phases, the conditions promoting them, and their relative importance and length, can be identified in nature (Figure I).



TRENDS in Genetics

Figure I. The four potential phases of speciation-with-gene-flow involving differences in the relative importance of DS, DH, and GH. Plots depict the general expected relationship of divergence (F_{ST}) for a neutral site at varying recombination rates ranging from $r = 0$ cM (completely linked) to $r = 0.5$ (unlinked) to a divergently selected locus, as speciation proceeds through the four phases. Color panels denote three different scenarios representing varying contributions of DH and GH to the accumulation of genetic divergence as speciation-with-gene-flow proceeds. Note that these phases are not absolute, and the barriers between them are somewhat diffuse, but they provide a framework for interpreting genomic speciation (see text for details).

from linkage of sites to a gene under selection will reduce m_e below the level caused solely by GH. In contrast to DH, GH can facilitate the establishment of new adaptive mutations of modest to minor effect across the genome, instead of their spread being restricted to regions containing already strongly diverged loci. Note that new mutations under strong selection with $s >> m$ generally do not need DH or GH to become established and will accrue genome-wide during speciation-with-gene-flow [13]. A key question then is the extent to which DH and GH make speciation-with-gene-flow more likely compared to only strong divergent selection acting directly on individual target genes (i.e. DS with no hitchhiking).

Genomic islands and continents of speciation

The roles that DS, DH, and GH play in speciation have been described using an oceanic island metaphor conceptualizing patterns of differentiation observed in genome scans. The metaphor arose from a seminal study between different forms of mosquitoes [33] which implied that gene

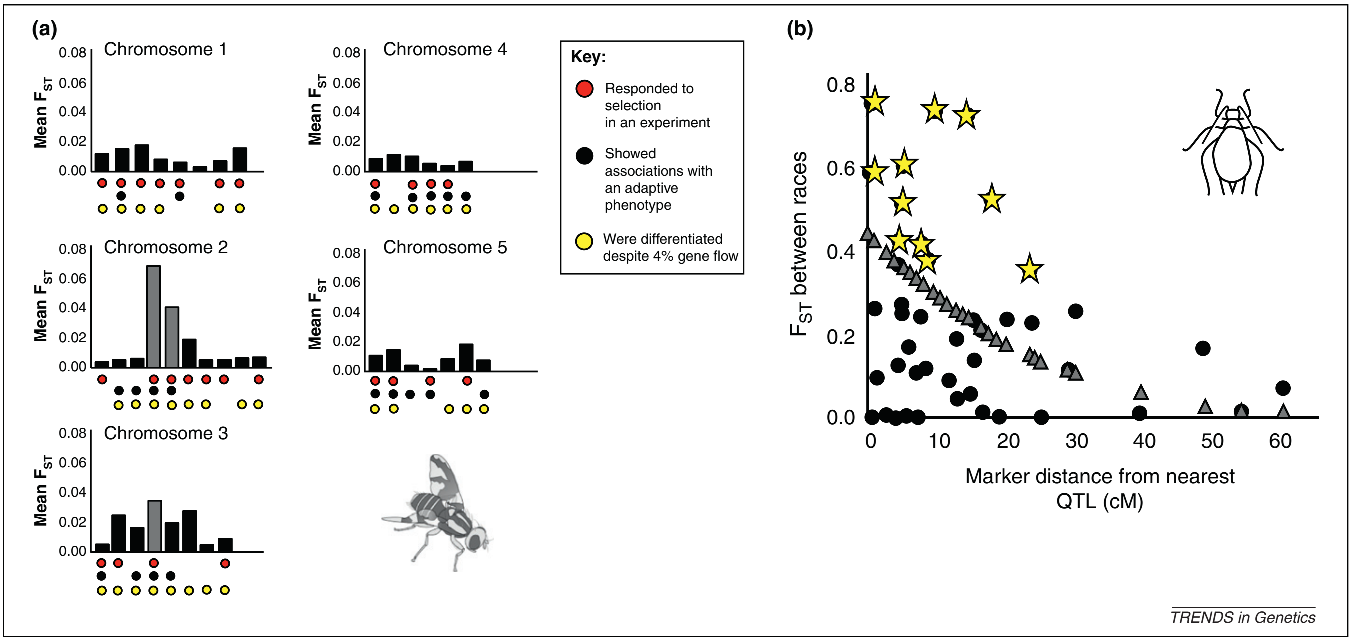


Figure 1. Empirical examples of potential GH and DH in nature. **(a)** Mean F_{ST} for loci on chromosomes 1–5 between hawthorn- and apple-infesting host races of *Rhagoletis pomonella* flies. Gray bars represent loci that were statistical outliers in a genome scan. Dots below each bar denote if a locus displayed evidence for divergent selection in the form of significant host-related differentiation between hawthorn and apple flies despite high levels of gene flow in nature (yellow), associations with an adaptive diapause phenotype (black) or responses to selection in a manipulative overwintering experiment (red). The key observation is that despite detecting only three statistical outliers in a genome scan, the diapause and selection experiments and more detailed genetic analysis revealed evidence for genome-wide divergence, consistent with multifarious selection and GH. Modified from [39] with permission of the National Academy of Sciences. **(b)** Genetic differentiation (F_{ST}) between clover- and alfalfa-associated races of pea aphids (*Acyrthosiphon pisum*) for putatively neutral markers as a function of recombination distance in cM from known adaptive QTL. Triangles show the predicted values from a logistic regression of outlier status on distance to the nearest QTL based on the model of [40]. Statistical outliers with high F_{ST} (designated by yellow stars) at considerable map distances from QTLs were argued to arise from DH. Modified from [19] and reprinted with permission of Blackwell publishing.

flow was sufficient to homogenize most of the *Anopheles gambiae* genome to a low baseline level of neutral differentiation ([34–36] for contrasting views). However, divergent selection acting on a few regions was strong enough to create 'genomic islands of divergence' that contained F_{ST} 'outlier' loci rising above the neutral sea level. It has been proposed that DH could then allow for the sequential build-up of divergence around such isolated islands because increasing numbers of linked sites under selection in a region will synergistically reduce m_e locally below the sum total of each site individually. Thus, genomic islands might increase in both height and width until large genomic regions become differentiated [19,26].

At the other end of the spectrum, when speciation-with-gene flow occurs, it may often be driven by multifarious divergent selection acting on many loci and affecting many different traits through the genome [37–39]. This could generate a genome-wide reduction in m_e favorable for GH, resulting in a rapid general uplifting of large 'continents' of divergence above neutral sea level. An example may be the hawthorn-infesting and recently derived sympatric apple-infesting host races of the fly *Rhagoletis pomonella*. By coupling analysis of microsatellites with selection experiments for diapause life-history traits that ecologically isolate the races, it was shown [39] that the majority of the genome of *R. pomonella*, and not only a few outlier loci, was differentiated and affected by divergent selection (Figure 1a). However, it is important to note that genomic continents can still have highly variable 'topography': baseline levels of divergence can be elevated above that expected in the absence of selection but only moderately so, and thus only the regions more directly affected by selection or undergoing low recombination will be exceptionally differentiated, as seen in *R. pomonella* [39].

Theory of speciation genomics-with-gene-flow

General implications

Although the verbal arguments for DH and GH are appealing, assessment of the conceptual metaphors will require a formal theory of speciation genomics. The foundation for such a theory was developed by (i) expanding a single-locus model of local adaptation with gene flow [40] to any number of genes under divergent selection and a wider set of parameter values [14], and (ii) conducting computer simulations examining the probability that new mutations become established under varying conditions [13]. The main finding was that DH around a selected locus can generate large regions of neutral differentiation and increase the establishment of new mutations under some, but relatively limited, conditions. For a single locus, regions of differentiation did not generally extend more than 1–2 centiMorgans (cM) along a chromosome from a selected site, and often less so. The exception for extending DH farther along a chromosome was when selection was very strong (s = 0.5) on the target locus, effective population size was relatively small (n_e = 1000), and migration rate low (m = 0.001). With multiple unlinked loci under selection, regions of differentiation can be larger, but m_e becomes globally reduced such that genome-wide divergence accumulates due to GH. The theory therefore raises questions about the general efficacy of DH for enhancing speciation-with-gene-flow relative to GH. A final consideration is that divergently selected mutations will often arise in non-favored habitats and genetic backgrounds. Indeed, in a two-deme model with equal population sizes, this will occur half the time. In these cases, loose rather than tight linkage of the mutation to a diverged locus is more conducive to its establishment because looser linkage helps the mutation recombine out of non-favorable genetic backgrounds [13].

Caveats of theory

The implications of the theory should be tempered, however, by several caveats. If the vast majority of new mutations early in speciation have minor effects on fitness, then linkage and DH can be more important. It has also been shown that under prolonged periods of stabilizing selection in patches with different optima, tightly linked genes of major effect can replace multiple genes of minor effect [17]. In addition, chromosomal rearrangements that capture locally favorable combinations of genes could foster the evolution of increased divergence in inverted regions due to their recombination-suppressing effects ([41,42], but see [43]). The past biogeography of taxa and standing genetic variation in ancestral populations could also pre-package adapted alleles into concentrated islands conducive for DH. Indeed, episodes of allopatry and secondary contact favor selection for recombination modifiers (e.g. inversions) that keep beneficial genotypic combinations inherited together as 'supergenes' [42,44–48].

Particular types of epistasis involving genes affecting habitat choice, assortative mating, and mimicry could also increase the importance of DH for speciation. For example, a new mutation resulting in greater preference for a particular habitat will be favored only in individuals that also possess genes for high performance (survivorship) in that habitat [49]. These types of 'two-allele' systems [27] in which fitness is a function of genetic background can exacerbate the selection–recombination antagonism, placing an increased premium on tighter linkage than the standard multiplicative fitness interactions modeled in [21] and [14]. By contrast, for 'one-allele' systems where a single allele causes individuals to select the environment they survive best in, linkage is not an issue [24].

Phases of speciation-with-gene-flow

An important insight from the theory is the prediction of four phases of speciation-with-gene-flow, and this generates testable hypotheses about how patterns of genomic divergence may change as speciation proceeds (Box 2). We stress that these four phases do not form discrete boundaries, but instead reflect differences in the relative importance of DS, DH, and GH during speciation. In phase 1, speciation starts with a key role for DS in establishing divergence of a few loci generating ecological specialization. Populations may then enter phase 2, where DH can aid in the sequential build-up of differentiation for linked sites surrounding the initially diverged genomic islands. In phase 3, the cumulative strength of selection is sufficient for m_e to be significantly reduced globally and GH elevates differentiation genome-wide [32,50–53]. Finally, in stage 4, m_e is reduced so strongly and genome-wide that populations are essentially allopatric, and the distribution of F_{ST} across the genome becomes a flat, high-elevation plateau.

A key research goal currently is to determine from NGS data how prevalent these different phases of speciation are (Figure I, Box 2). For example, if DH is essential for speciation, then discrete islands of differentiation (phase 2) should often be observed in genome scans between recently diverged taxa, with multiple quantitative trait loci (QTL) for ecological adaptation mapping to a few regions of the genome. Alternatively, if GH drives crucial formative stages of speciation-with-gene-flow, then transitions between phases 1 and 3 can occur rapidly, and differentiated loci will be spread throughout the genome rather than being clustered. Finally, if both DH and GH are important, then comparisons of related taxa along the speciation continuum should reveal populations transitioning from phase 2 to 3 as divergence proceeds.

Empirical data

Divergence hitchhiking

Several observations are consistent with a role for DH in genomic divergence and speciation [26,30]. In alfalfa and clover host-races of pea aphids, major QTL affecting performance and preference for the native host have been detected [49]. Significant outlier loci have been detected from these QTLs for markers up to 10–20 cM away (Figure 1b) [19]. Similar extended regions of divergence have been reported for dwarf and normal ecotypes of freshwater whitefish [54,55]. A recent study of stickleback fish also reported substantial short- and long-distance LD along chromosomes in both freshwater and oceanic populations [21], a pattern that could facilitate large regions of differentiation via DH. These results appear to contradict the predictions that DH should generally be confined to a narrow recombination window around a selected site, suggesting that caveats concerning the theory such as strong fitness epistasis may apply.

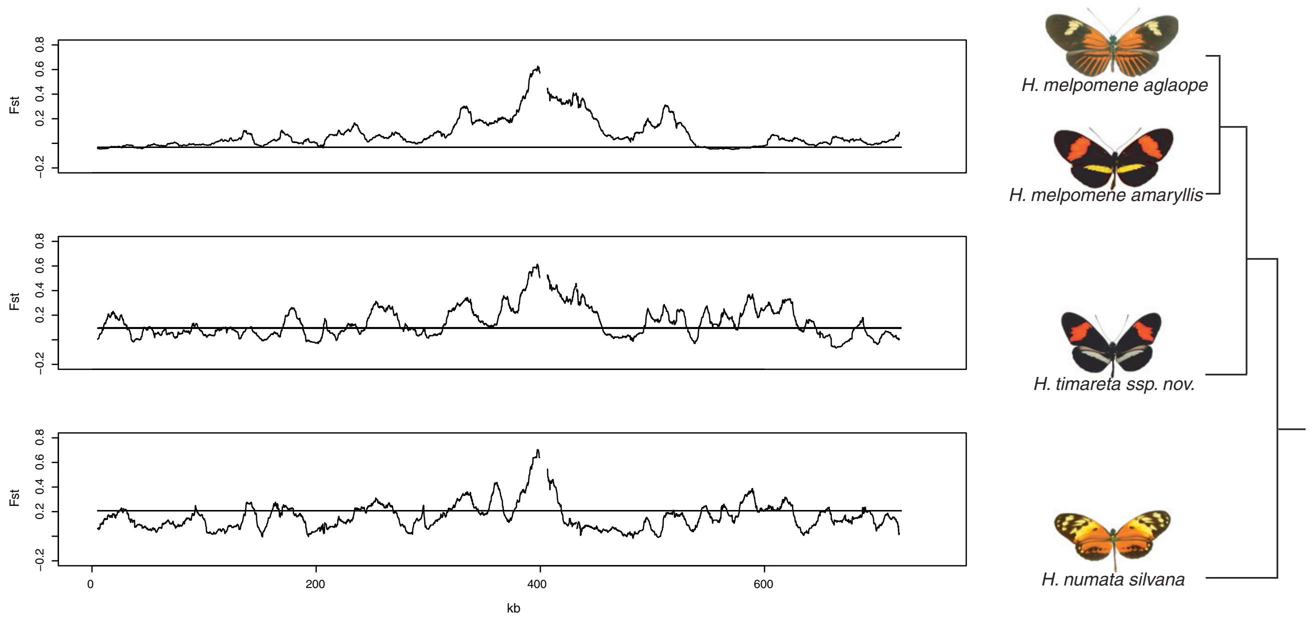
However, several empirical observations, as well as issues associated with the studies discussed above, argue against a crucial role for DH. Many studies have reported individual regions of genomic divergence to be small, including examples from mosquitoes [56,57], fruit flies [58], fish [22], snails [59], and plants [20,53]. These findings are consistent with theory implying that when DH occurs, its effects will generally be limited to linked sites in close recombination proximity to selected loci [14]. Moreover, outliers displaying pronounced differentiation are often scattered at many locations across the genome [20,23,24,48,56,60], instead of being clumped within a few islands (but see [61]). Finally, the interpretation of large regions of divergence (e.g. in pea aphids) is difficult; instead of representing DH away from a single QTL, such blocks might represent regions containing multiple QTL, some of which are undetected. This issue is not necessarily solved by increased statistical power because identifying and mapping all possible phenotypes under selection is difficult, necessarily leading to undetected QTL.

In summary, there are observations for and against a role for DH in genomic divergence. Further work is needed to determine if departures from theory are (i) real and represent a problem with model assumptions or are special cases associated with the theory (e.g. epistasis or inversions), or (ii) problems with inadequate marker sampling

Box 3. *Heliconius* butterflies: a case-study of the 'four-phase model'

A recent study of genomic divergence in hybridizing *Heliconius* butterflies provides an opportunity to evaluate the four-phase model. *Heliconius* butterflies represent a radiation of species in which wing color-pattern divergence related to Müllerian mimicry has been implicated in speciation. Several loci that control wing-pattern phenotypes have been mapped and two identified through sequencing. A recent study [18] used targeted NGS capture methods to survey patterns of divergence across these regions in divergent geographic races and species of *Heliconius* and then compared divergence in color-pattern regions to unlinked BAC clones (i.e. putative neutrally evolving regions). Three different points in the speciation continuum were examined. In order of increasing divergence, these comparisons were between commonly hybridizing '*aglaope*' and '*amaryllis*' races of *H. melpomene*, between *H. melpomene* and *H. timareta*, which are likely to hybridize relatively frequently, and between the more distantly related species *H. melpomene* and *H. numata*, which hybridize occasionally in the wild [77] (Figure I).

The researchers found major peaks of elevated population differentiation in the color-pattern regions. A few islands of divergence were detected between the *aglaope* and *amaryllis* races of *H. melpomene* races and low baseline sea-levels of divergence made them highly evident. The races might thus be in phase 1 or early phase 2, where the few most strongly selected regions have diverged but most of the genome is homogenized by gene flow. The more closely related species pair *H. melpomene* and *H. timareta* exhibited a greater number of genomic islands than the races and somewhat elevated baseline differentiation. Thus, this species pair is probably at a more advanced stage of divergence in phase 2 or early phase 3, where DH could facilitate the differentiation of more weakly selected regions, and some GH also takes place, to elevate baseline differentiation. Finally, the distantly related pair *H. melpomene* and *H. numata* exhibited high levels of baseline differentiation, which often obscured or erased genomic islands. This pattern is indicative of widespread divergence via GH, as occurs in phase 3. Further analysis of additional regions of the *Heliconius* genome and studies of genomic divergence across the speciation continuum for other taxa are highly warranted.



TRENDS in Genetics

Figure I. Genetic differentiation (F_{ST}) between pairwise comparisons of *Heliconius* butterflies at different stages of speciation across a divergently selected color-pattern region of the genome. Specifically, the *HmBD* region associated with Müllerian mimicry (jagged lines) versus the 95% upper confidence level for neutral divergence based on three unlinked and putatively non-selected regions sequenced from BAC clones (solid, flat lines). The three different levels of divergence correspond to comparisons of different races within species (top panel: *H. melpomene amaryllis* versus *H. m. aglaope*), between closely related species (middle panel: *H. timareta* versus *H. m. aglaope*), and between more distantly related species (bottom panel: *H. numata* versus *H. m. aglaope*). The results imply that as speciation has proceeded, genomic islands under selection from the start of the process have remained highly differentiated whereas other regions, including unlinked putatively neutral ones, have steadily increased in divergence, potentially due to GH effects. Modified from [18] with permission of the Royal Society of London.

in genome scans and QTL detection. The issue of marker density will be addressed by application of NGS technology. However, resolving the other issues will require much more extensive data and may challenge our ability to quantify the role of DH in speciation except in general terms.

Genome hitchhiking

Several observations also support and argue against GH. Perhaps the most widespread evidence for GH stems from studies testing for associations between the degree of ecological or adaptive divergence between populations (a proxy for the strength of divergent selection) and the level of genetic differentiation between populations. This pattern, termed 'isolation-by-adaptation' (IBA), is analogous to isolation-by-distance (IBD), but the gene-flow reductions generating IBA occur via increased adaptive divergence rather than via greater geographic distance [62]. GH can generate IBA, even for neutral loci unlinked to those under divergent selection. A recent review of 22 studies of IBA at putatively neutral loci [48] found that 15 (68%) showed evidence for IBA, independently from IBD. This survey was not a formal meta-analysis and new examples have since emerged (e.g. [63]), and the results therefore should not be over-interpreted. However, at the very least IBA appears to be fairly common in nature. Classic examples concern morphological divergence positively correlated with neutral molecular distance in a rainforest passerine bird [64], lake whitefish [54,65], and an island lizard [66,67]. Several other observations are also consistent with GH – such as significantly elevated F_{ST} levels across the genome [22,60,68,69] and associations between ecology and neutral genetic differentiation in mosaic hybrid zones [70–72].

In addition, the aforementioned study of *R. pomonella* [39], whole-genome sequencing in mosquitoes [56], genome scans in *Arabidopsis* [23,68], and a recent restriction site associated DNA marker (RAD-tag) genome scan of lake and stream eco-types of stickleback fish [73], have revealed widespread genomic differentiation across the genome. Such patterns are consistent with a rapid transition to phase 3, implying that when speciation-with-gene-flow occurs it may often involve multifarious selection rapidly enabling GH and genome-wide divergence.

However, GH is not always expected in nature because high gene-flow can readily swamp differentiation of neutrally evolving or weakly selected genomic regions [26,74]. Indeed, a lack of IBA has been reported in many studies testing for it [48], and similar 'negative' results may often go unpublished. Likewise, several population genomic studies report low F_{ST} across the genome, despite some regions of exceptional differentiation [48], suggesting that selection operating on only a small number of regions has led to their differentiation, but has had little effect on unlinked regions. However, purely observational outlier results from genome-scan studies may be biased to miss selection acting more weakly on differentiated regions (Figure 1a) [39], and this problem is compounded when genomic coverage is poor. In addition, it can take time for neutral differentiation to accumulate to expected levels even when selection acts directly on linked loci in the region. In sum, there is convincing evidence for GH in some but not all cases, and particular studies imply that it can occur early in the speciation process.

Empirical evidence regarding the 'four-phase model'

How well do empirical data fit the four-phase model and the conditions predicted to underlie the phases? These questions cannot be answered readily at this time because very few studies have examined how genomic divergence varies across the speciation continuum. We can see possible glimpses of these phases, however, by comparing results from taxa lying at different points in speciation. For example, some taxa that appear to have diverged only to the point of moderate reproductive isolation exhibit divergence in only a few regions [48], whereas more diverged species-pairs exhibit widespread differentiation across much of the genome [26,30]. However, there is much variability among these trends. For example, the same pattern of multiple and relatively small islands of divergence has been reported in two systems (mosquitoes and butterflies) that appear to vary widely in how far speciation has proceeded [18,75].

The results highlight problems with trying to make comparisons between study systems that span drastically different taxa and in relying on pattern alone to try to understand mechanism when a balance among multiple processes is involved. Although making comparisons between disparate systems is a starting point, it is akin to comparing apples and oranges – not only can the biology be different, but studies often involve different experimental approaches, knowledge of natural history, and type and number of molecular markers. What is required now are standardized and detailed analyses of genomic divergence between closely related taxa (e.g. population pairs within species that vary strongly in their degree of reproductive isolation or different ecotype and species pairs within a single genus) that span the speciation continuum and that have well-characterized natural and biogeographic histories. Such work is increasing at the phenotypic level [38,76–81], but has yet to be applied fully at the genomic level. Nonetheless, one case-study for *Heliconius* butterflies examining patterns of genomic divergence at different points in the speciation continuum [18,82] (Box 3) appears to be consistent with DS, DH, and GH increasing in their relative importance as speciation proceeds.

Concluding remarks

A framework for the field of speciation genomics is taking shape, including clarification of the outstanding issues that must be resolved. From a theoretical standpoint, more work is needed to determine more clearly the consequences that epistasis, inversions and other physical aspects of the genome, standing genetic variation, and variable spatial/geographic scenarios have on the standard two-deme, multiplicative fitness model. In addition, new theory is needed to transform current predictions concerning genetic divergence into more dynamic recreations of how genomic differentiation unfolds through time during speciation to (i) delimit better the size and distribution of islands and continents of divergence, and (ii) assess more fully the prevalence and duration of the potential four different phases. This will allow for more quantitative predictions of patterns of genomic divergence generated by DS, DH, and GH and lay the foundation for more powerful statistical tests that can distinguish between them.

Although theoretical predictions may become reasonably precise, discerning process from empirical patterns of NGS divergence may remain difficult due to overlapping expectations generated by different combinations of processes (e.g. strong selection and low recombination can both facilitate genomic divergence). Accurate empirical tests of predictions may thus require large amounts of information on mutation rates, numbers and distribution of selected sites and their *s* values, migration rate, recombination rate, and the past history of populations [25]. Nevertheless, broad generalities may still be gleamed from meta-analyses of related groups of taxa at varying stages in the speciation continuum. In this regard, although all specific loci may not be known, combining knowledge of the natural history of a system and, in particular, the key factors and traits generating divergent selection, and the locations of genomic regions containing genes contributing to reproductive isolation, will allow general trends to be ascertained.

The field of genomics is yet to move into a truly experimental phase, where factors such as selection and gene flow are manipulated (e.g. in reciprocal transplant experiments) to test the processes driving and constraining genomic divergence. Some cutting-edge 'experimental genomic' studies of this type have now been conducted in the laboratory using microbes or flies [83–86], but these do not address reproductive isolation or speciation in natural populations. Such experiments in natural systems might then be combined with patterns from genome scans to allow hypothesis-testing concerning the role of genome structure in speciation. Eventually, through molecular evolution, functional genomics, and gene transformation and knockout studies, the individual genes that mattered for different points in the speciation process may also be resolved, at least for a subset of model systems, to form a complete natural history of speciation genomics.

Acknowledgments

We thank T. Arcella, S. Berlocher, L. Bernatchez, A. Comeault, T. Farkas, G. Hood, A. Martin, M. Muschick, T. Powell, G. Ragland, S. Sim, G. St. Jean, the evolution group at the University of Colorado, Boulder, and two anonymous reviewers for discussion and useful comments on the manuscript. We also thank Guy L. Bush for his profound influence on our thinking about speciation-with-gene-flow. The work was supported, in part, from grants to J.L.F. from the National Science Foundation and the US Department of Agriculture, and to P.N. from the European Research Council.

References

1 Schluter, D. (2001) Ecology and the origin of species. *Trends Ecol. Evol.* 16, 372–380

2 Coyne, J.A. and Orr, H.A. (2004) *Speciation*, Sinauer Associates

3 Rundle, H.D. and Nosil, P. (2005) Ecological speciation. *Ecol. Lett.* 8, 336–352

4 Feder, J.L. *et al.* (2005) Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6573–6580

5 Funk, D.J. (1998) Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. *Evolution* 52, 1744–1759

6 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

7 Wu, C. (2001) The genic view of the process of speciation. *J. Evol. Biol.* 14, 851–865

8 Wu, C.I. and Ting, C.T. (2004) Genes and speciation. *Nat. Rev. Genet.* 5, 114–122

9 Presgraves, D.C. (2007) Speciation genetics: epistasis, conflict and the origin of species. *Curr. Biol.* 17, R125–R127

10 Rieseberg, L.H. and Blackman, B.K. (2010) Speciation genes in plants. *Ann. Bot.* 106, 439–455

11 Nosil, P. and Schluter, D. (2011) The genes underlying the process of speciation. *Trends Ecol. Evol.* 26, 160–167

12 Orr, H.A. (2005) The genetic basis of reproductive isolation: Insights from *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6522–6526

13 Feder, J.L. *et al.* (2012) Establishment of new mutations under divergence and genome hitchhiking. *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 461–474

14 Feder, J.L. and Nosil, P. (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* 64, 1729–1747

15 Nosil, P. and Feder, J.L. (2012) Genomic divergence during speciation: causes and consequences. *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 332–342

16 Yeaman, S. and Otto, S.P. (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* 65, 2123–2129

17 Yeaman, S. and Whitlock, M.C. (2011) The genetic architecture of adaptation under migration–selection balance. *Evolution* 65, 1897–1911

18 Nadeau, N.J. *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 343–353

19 Via, S. and West, J. (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.* 17, 4334–4345

20 Strasburg, J.L. *et al.* (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 364–373

21 Hohenlohe, P.A. *et al.* (2012) Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 395–408

22 Hohenlohe, P.A. *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* 6, e1000862

23 Fournier-Level, A. *et al.* (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* 333, 86–89

24 Hancock, A.M. *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 333, 83–86

25 Bernatchez, L. *et al.* (2010) On the origin of species: insights from the ecological genomics of lake whitefish. *Philos. Trans. R. Soc. B: Biol. Sci.* 365, 1783–1800

26 Via, S. (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 451–460

27 Felsenstein, J. (1981) Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* 35, 124–138

28 Via, S. (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* 16, 381–390

29 Savolainen, V. *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* 441, 210–213

30 Via, S. (2009) Natural selection in action during speciation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9939–9946

31 Gavrilets, S. (2004) *Fitness Landscapes and the Origin of Species*, Princeton University Press

32 Barton, N. and Bengtsson, B.O. (1986) The barrier to genetic exchange between hybridizing populations. *Heredity* 57, 357–376

33 Turner, T.L. *et al.* (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 3, 1572–1578

34 White, B.J. *et al.* (2010) Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Mol. Ecol.* 19, 925–939

35 Turner, T.L. and Hahn, M.W. (2010) Genomic islands of speciation or genomic islands and speciation? *Mol. Ecol.* 19, 848–850

36 Noor, M.A.F. and Bennett, S.M. (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* 103, 439–444

37 Rice, W.R. and Hostert, E.E. (1993) Laboratory experiments on speciation – what have we learned in 40 years? *Evolution* 47, 1637–1653

38 Nosil, P. *et al.* (2009) Ecological explanations for (incomplete) speciation. *Trends Ecol. Evol.* 24, 145–156

39 Michel, A.P. *et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9724–9729

40 Charlesworth, B. *et al.* (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* 70, 155–174

41 Kirkpatrick, M. and Barton, N. (2006) Chromosome inversions, local adaptation and speciation. *Genetics* 173, 419–434

42 Feder, J.L. *et al.* (2011) Adaptive chromosomal divergence driven by mixed geographic mode of evolution. *Evolution* 65, 2157–2170

43 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

44 Kimura, M. (1956) A model of a genetic system which leads to closer linkage by natural selection. *Evolution* 10, 278–287

45 Butlin, R.K. (2005) Recombination and speciation. *Mol. Ecol.* 14, 2621–2635

46 Charlesworth, D. and Charlesworth, B. (1975) Theoretical genetics of Batesian mimicry. 2. Evolution of supergenes. *J. Theor. Biol.* 55, 305–324

47 Kouyos, R.D. *et al.* (2006) Effect of varying epistasis on the evolution of recombination. *Genetics* 173, 589–597

48 Nosil, P. *et al.* (2009) Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* 18, 375–402

49 Hawthorne, D.J. and Via, S. (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* 412, 904–907

50 Barton, N.H. (1979) Dynamics of hybrid zones. *Heredity* 43, 341–359

51 Barton, N.H. and Hewitt, G.M. (1985) Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* 16, 113–148

52 Barton, N.H. and Hewitt, G.M. (1989) Adaptation, speciation and hybrid zones. *Nature* 341, 497–503

53 Vines, T.H. *et al.* (2003) The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. *Evolution* 57, 1876–1888

54 Rogers, S.M. and Bernatchez, L. (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonas* sp. Salmonidae) species pairs. *Mol. Biol. Evol.* 24, 1423–1438

55 Renaut, S. *et al.* (2011) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 354–363

56 Lawniczak, M.K.N. *et al.* (2010) Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* 330, 512–514

57 Turner, T.L. and Hahn, M.W. (2007) Locus- and population-specific selection and differentiation between incipient species of *Anopheles gambiae*. *Mol. Biol. Evol.* 24, 2132–2138

58 Turner, T.L. *et al.* (2008) Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics* 179, 455–473

59 Wood, H.M. *et al.* (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Mol. Ecol.* 17, 3123–3135

60 Strasburg, J.L. *et al.* (2009) Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. *Mol. Biol. Evol.* 26, 1341–1355

61 Emelianov, I. *et al.* (2004) Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proc. R. Soc. Lond. Ser. B: Biol. Sci.* 271, 97–105

62 Nosil, P. *et al.* (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: 'isolation by adaptation' and multiple roles for divergent selection. *Evolution* 62, 316–336

63 Vonlanthen, P. *et al.* (2009) Divergence along a steep ecological gradient in lake whitefish (*Coregonus* sp.). *J. Evol. Biol.* 22, 498–514

64 Smith, T.B. *et al.* (1997) A role for ecotones in generating rainforest biodiversity. *Science* 276, 1855–1857

65 Lu, G. and Bernatchez, L. (1999) Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. *Evolution* 53, 1491–1505

66 Thorpe, R.S. and Richard, M. (2001) Evidence that ultraviolet markings are associated with patterns of molecular gene flow. *Proc. Natl. Acad. Sci. U.S.A.* 98, 3929–3934

67 Ogden, R. and Thorpe, R.S. (2002) Molecular evidence for ecological speciation in tropical habitats. *Proc. Natl. Acad. Sci. U.S.A.* 99, 13612–13615

68 Turner, T.L. *et al.* (2010) Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nat. Genet.* 42, 260–263

69 Wilding, C.S. *et al.* (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J. Evol. Biol.* 14, 611–619

70 Nosil, P. *et al.* (2005) Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* 59, 705–719

71 MacCallum, C.J. *et al.* (1998) Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution* 52, 227–239

72 Rand, D.M. and Harrison, R.G. (1989) Ecological genetics of a mosaic hybrid zone – mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution* 43, 432–449

73 Roesti, M. *et al.* (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Mol. Ecol.* DOI: 10.1111/j.1365-294X.2012.05509.x

74 Thibert-Plante, X. and Hendry, A.P. (2010) When can ecological speciation be detected with neutral loci? *Mol. Ecol.* 19, 2301–2314

75 Hahn, M.W. *et al.* (2011) No evidence for biased co-transmission of speciation islands in *Anopheles gambiae*. *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 374–384

76 Peccoud, J. *et al.* (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7495–7500

77 Mallet, J. *et al.* (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evol. Biol.* 7, 28

78 Seehausen, O. *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature* 455, 620–623

79 Nosil, P. and Sandoval, C.P. (2008) Ecological niche dimensionality and the evolutionary diversification of stick insects. *PLoS ONE* 3, e1907

80 Berner, D. *et al.* (2009) Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. *Evolution* 63, 1740–1753

81 Hendry, A.P. *et al.* (2009) Along the speciation continuum in sticklebacks. *J. Fish Biol.* 75, 2000–2036

82 Jiggins, C.D. (2008) Ecological speciation in mimetic butterflies. *Bioscience* 58, 541–548

83 Paterson, S. *et al.* (2010) Antagonistic coevolution accelerates molecular evolution. *Nature* 464, 275–278

84 Araya, C.L. *et al.* (2010) Whole-genome sequencing of a laboratory-evolved yeast strain. *BMC Genomics* 11, 88

85 Barrick, J.E. *et al.* (2009) Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* 461, 1243–1274

86 Burke, M.K. *et al.* (2010) Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* 467, 587–590