The genetic mosaic suggests a new role for hitchhiking in ecological speciation

SARA VIA and JOAN WEST

Department of Biology, Department of Entomology, and Program in Behaviour, Ecology, Evolution and Systematics, University of Maryland, College Park, MD 20742, USA

Abstract

Early in ecological speciation, the genomically localized effects of divergent selection cause heterogeneity among loci in divergence between incipient species. We call this pattern of genomic variability in divergence the 'genetic mosaic of speciation'. Previous studies have used F_{ST} outliers as a way to identify divergently selected genomic regions, but the nature of the relationship between outlier loci and quantitative trait loci (QTL) involved in reproductive isolation has not yet been quantified. Here, we show that F_{ST} outliers between a pair of incipient species are significantly clustered around QTL for traits that cause ecologically based reproductive isolation. Around these key QTL, extensive 'divergence hitchhiking' occurs because reduced inter-race mating and negative selection decrease the opportunity for recombination between chromosomes bearing different locally adapted QTL alleles. Divergence hitchhiking is likely to greatly increase the opportunity for speciation in populations that are sympatric, regardless of whether initial divergence was sympatric or allopatric. Early in ecological speciation, analyses of population structure, gene flow or phylogeography based on different random or arbitrarily chosen neutral markers should be expected to conflict — only markers in divergently selected genomic regions will reveal the evolutionary history of adaptive divergence and ecologically based reproductive isolation. Species retain mosaic genomes for a very long time, and gene exchange in hybrid zones can vary dramatically among loci. However, in hybridizing species, the genomic regions that affect ecologically based reproductive isolation are difficult to distinguish from regions that have diverged for other reasons.

Keywords: ecological speciation, F_{ST} outlier, genetics of speciation, host races, sympatric hitchhiking

Received 16 April 2008; revision received 9 May 2008; accepted 23 May 2008

Introduction

Over 60 years ago, Dobzhansky and Mayr determined the course of speciation research by strongly asserting that geographical isolation is required for speciation, that reproductive isolation is the essence of speciation, and that postzygotic genetic incompatibilities are the primary cause of reproductive isolation (e.g. Dobzhansky 1937; Mayr 1942). Their ideas defined a retrospective approach to speciation research that focuses on analysis of reproductive isolation between 'good species'.

This retrospective approach has dominated speciation research since the 1940s (Otte & Endler 1989; Howard & Berlocher 1998; Coyne & Orr 2004). Despite its nearly

Correspondence: Sara Via, Fax: 301-314-6262; E-mail: svia@umd.edu

universal acceptance, however, the retrospective study of speciation has an overwhelming limitation. By the time diverging lineages are accepted as good species, the genetic changes that initially caused reproductive isolation cannot be distinguished from species differences that evolved much later (e.g. Templeton 1981). In fact, by the time a typical retrospective analysis begins, so much time has passed that genetic incompatibilities may far outnumber the earlier genetic changes responsible for the initial barriers to gene flow between sister species. This makes it easy to conclude that the scores of genetic incompatibilities that contribute to hybrid sterility and inviability (Coyne & Orr 2004) are the cause rather than the consequence of speciation.

Why wait so long to study the evolution of reproductive isolation? In ecological speciation, reproductive isolation evolves as a result of divergent selection on resource or

habitat use (review in Rundle & Nosil 2005). Extant species thought to have resulted from this process are reproductively isolated primarily by habitat choice, mate choice, selection against migrants or extrinsic selection against hybrids (Schluter 2001; Via 2001), although postzygotic genetic incompatibilities can be found in some cases (Ramsey *et al.* 2003). At some point during divergence, such species must have been divergent and partially reproductively isolated populations, suggesting that contemporary host races or ecotypes are good models for the study of early phases of ecological speciation. This approach could clarify the identity of early causes of reproductive isolation, reveal the kinds of genetic changes that produce them, and show whether ecologically based barriers to gene flow frequently evolve before the appearance of intrinsic genetic incompatibilities. Despite the common concern that not all divergent populations will go on to become good species (Coyne & Orr 2004. p. 60), any generalities that can be discovered about the causes of reproductive isolation in partially reproductively isolated contemporary populations are likely to provide important insights into ecological speciation.

'Divergence-with-gene-flow' speciation and porous species boundaries

There is now solid evidence that divergent selection can maintain incomplete ecologically based reproductive isolation in the presence of gene flow (Bush 1994; examples in Rice & Hostert 1993; Howard & Berlocher 1998; Schluter 2001; Via 2001; Rundle & Nosil 2005). This rapidly growing body of work on 'divergence-with-gene-flow' speciation (Rice & Hostert 1993) reveals that complete allopatry is unnecessary during speciation, that the evolution of reproductive isolation is not the fragile process assumed by Mayr, and that the conditions under which isolation can evolve and be maintained are neither as restrictive nor as infrequent as previously thought (Rice & Hostert 1993; Schluter 2001; Via 2001; Mallet 2005a; Lexer *et al.* 2006).

Recent analyses of natural populations suggest that adaptive genetic divergence in just a few key traits (usually associated with resource use, mate choice, or pollination) can lead to both pronounced phenotypic differentiation and reduced interpopulation mating and reproduction (Schemske 2000; Via *et al.* 2000; Via 2001; Bradshaw & Schemske 2003; Mallet 2005a; Rundle & Nosil 2005). Because divergent selection causes genetic change only in loci affecting the target phenotypic traits, taxa may become almost completely reproductively isolated while remaining genetically extremely similar at genomic regions that are unaffected by divergent selection. We refer to the genomic mix of divergent and non-divergent genomic regions within incipient species as the 'genetic mosaic of speciation'.

Taxa can retain mosaic genomes and show genomic heterogeneity in divergence and gene flow long after they are accorded species status (e.g. Rieseberg *et al.* 1999; Machado *et al.* 2002; Machado & Hey 2003; Payseur *et al.* 2004; Mallet 2005b; Vasemagi *et al.* 2005; Bonin *et al.* 2006; Harr 2006; Mallet *et al.* 2007; Yatabe *et al.* 2007). Foreshadowing what is now called the 'porous species boundary', Barton's classic analyses of hybrid zones showed that neutral or mutually beneficial alleles can be freely exchanged between species, while alleles that are disadvantageous in the genetic background or ecological situation of the other species are prevented from introgressing (reviews in Barton & Gale 1993; Barton 2000). However, in interspecific hybrids, the nonintrogressing part of the genetic mosaic includes not only divergently selected genes, but also chromosomal rearrangements (e.g. Rieseberg *et al.* 1999), and incompatibilities that have accumulated within each species in response to uniform selection, balancing selection or drift (e.g. Machado *et al.* 2002). Unfortunately, retrospective analyses of speciation usually begin after the portion of the genetic mosaic caused by divergent selection has been assimilated into this more complex pattern of species divergence, making it very difficult to distinguish either the causes of particular genetic differences between hybridizing species or their role (if any) in early forms of reproductive isolation.

The nature of the genetic mosaic of speciation thus changes in important ways during the millions of years required for the evolution of complete and irreversible reproductive isolation between new species (Coyne & Orr 2004; Ch. 12). During this long process, genomic variation within each species slowly comes into phylogenetic concordance through independent responses to selection and genetic drift within the new species (Avise 2000; Ch. 6). In ecological speciation, virtually complete ecologically based reproductive isolation may result from genetic change in just a few genomic regions (Schemske 2000; Rundle & Nosil 2005), and yet, these small changes crucially affect the rest of the speciation process by establishing the branching pattern with which variability across the rest of the genome will eventually come into concordance (Avise 2000; Ch. 6).

Here, we focus on exploring the genetic mosaic early in ecological speciation, when genetic differences between incipient species are dominated by responses to divergent selection on habitat or resource use. At this early stage, the genetic regions that cause the first forms of ecologically based reproductive isolation under divergent selection can be detected by their excessive divergence relative to the rest of the genome. Later, the genomic regions affected by divergent selection will be much more difficult to distinguish from regions that diverged in other ways.

Detecting the genetic mosaic in early ecological speciation

The idea that key genomic regions under selection can be identified with marker-specific estimates of genetic divergence, such as Wright's F_{ST}, dates back to Cavalli-Sforza

(1966). Motivated by the desire to obtain evidence about the frequency of selected markers in the context of the selectionist/neutralist debate, Lewontin & Krakauer (1973) provided a test using F_{ST} heterogeneity among markers: F_{ST} values significantly larger than other markers suggest divergent selection, while F_{ST} values significantly smaller than others suggest uniform selection. During the 1980s, this approach was often used to identify and exclude F_{ST} outliers from analyses requiring marker neutrality (e.g. Slatkin 1987).

The detection of F_{ST} outliers has recently enjoyed a renaissance, aided by the development of new methods that utilize coalescent simulation to account for stochastic variability in F_{ST} among loci (Beaumont & Nichols 1996; Beaumont & Balding 2004). However, the current goal of outlier analysis differs from that of the 1980s. Now, researchers usually discard unselected markers, and use F_{ST} outliers to probe the genetics of adaptation or speciation (reviewed in Luikart *et al*. 2003; Beaumont 2005; Storz 2005). Recent outlier analyses have revealed considerable variation in F_{ST} across the genomes of incipient species under divergent selection (Wilding *et al*. 2001; Campbell & Bernatchez 2004; Emelianov *et al*. 2004; Rogers & Bernatchez 2007; Nosil *et al*. 2008). Although the universal conclusion of these studies is that the outliers result from marker linkage to quantitative trait loci (QTL) for ecologically important traits under divergent selection, only Rogers & Bernatchez (2005) have shown a statistically significant association between the location of outliers and these key QTL.

The probable association between F_{ST} outliers and QTL for divergently selected traits raises a number of crucial questions about genetic change during speciation. How large are the genomic regions around key QTL that are affected by divergent selection? How close does a marker have to be to a divergently selected QTL in order to become a significant outlier? Given the conventional wisdom that hitchhiking regions are short-lived except where recombination is limited (Begun & Aquadro 1992; Turner *et al*. 2005), tight outlier–QTL linkage seems like an appropriate null hypothesis.

However, this hypothesis conflicts with the relatively large number of outliers identified in previous studies of divergent ecological races that are thought to be incipient species (5–18% of randomly chosen markers, e.g. Wilding *et al*. 2001; Emelianov *et al*. 2004; Nosil *et al*. 2008). If hitchhiking is restricted to a 1–2 kb region (Ting *et al*. 2000; Turner *et al*. 2005) around just a handful of QTL for a few traits, how can so many markers each be tightly enough linked to a QTL to become outliers? The hypothesis that linkage is tight between outliers and QTL also conflicts with reports of large regions of genomic differentiation between subspecies of mice [2–12 centimorgans (cM); Harr 2006, Payseur *et al*. 2004] and whitefish morphs (16.5 cM; calculated from data shown in Rogers & Bernatchez 2007).

Additional estimates of the size of chromosome blocks affected by divergent selection during early ecological speciation are required to resolve this apparent conflict between the conventional wisdom about hitchhiking and observed genomic divergence between incipient species. Here, we quantify the spatial distribution of outlier-QTL linkages using markers of known map distance from QTL for traits known to cause reproductive isolation between two insect host races. We examine the persistence of these outlier-QTL linkages by repeating the outlier and distance analysis in two other locations, and we evaluate whether the genetic mosaic pattern could be the cause of phylogenetic conflict in analyses of incipient and very young species (e.g. Shaw 2002; Beltran *et al*. 2002; Machado *et al*. 2002; Machado & Hey 2003; Mallarino *et al*. 2004). Finally, we present a mechanism, 'divergence hitchhiking', by which large genomic regions around divergently selected QTL are protected from recombination during ecological speciation.

Materials and methods

The system

We used two partially reproductively isolated host races of the pea aphid (*Acyrthosiphon pisum pisum*) as a model for taxa during early ecological speciation. Reciprocal transplants in the USA and Europe (Via 1991, 1999; Sandstrom 1996; Simon *et al*. 2003; Ferrari *et al*. 2007) have clearly established that pea aphid populations found sympatrically on alfalfa and red clover are ecologically specialized and genetically differentiated for host use. These races are highly reproductively isolated by a combination of habitat choice, selection against migrants and ecologically based selection against hybrids (Via *et al*. 2000). No significant intrinsic hybrid inviability or sterility has yet been detected (S. Via, unpublished data).

Outlier analysis

In our outlier analysis, we used 40 AFLP markers and five sequence-tagged codominant markers from a QTL map of early fecundity on and behavioural acceptance of each host plant species (Hawthorne & Via 2001; Via & Hawthorne 2001). These are the key traits involved in reproductive isolation, and we know from previous ecological work that they are under strong divergent selection (Via 1991, 1999; Via *et al*. 2000). This QTL map thus provides a unique opportunity to test the inference that markers with elevated genetic divergence are linked to genomic regions that affect ecologically based reproductive isolation, and to estimate the overall spatial distribution of F_{ST} outliers relative to key QTL.

The specialized pea aphid genotypes used in the QTL mapping cross were collected from fields in a dairy farming

area outside of Ithaca, New York, in 1991. In 2002, we returned to alfalfa and red clover fields on the same farms and collected an additional 100 pea aphids from each host. DNA was extracted, and each field-collected aphid was genotyped for the mapped markers as in Hawthorne & Via (2001).

Previous analyses in pea aphid populations (Via 1999) revealed that Hardy–Weinberg equilibrium within the races is typical, and we assume this here. F_{ST} for each marker and its outlier status was determined using coalescent simulation in the programs FDIST2 and DFDIST (Beaumont & Nichols 1996; Beaumont & Balding 2004). Our data are too sparse to produce a profile of F_{ST} values around each QTL; hence, we measured the distance of each marker to the nearest QTL for either habitat choice or performance on one of the hosts and then pooled the data across markers and QTL. Therefore, the resulting distribution represents an overall picture of marker-QTL distance across the entire genome. We used logistic regression to test for an association between the outlier status of a given marker and its map distance to the nearest QTL. The significance of the regression was tested by randomization. For each of 1000 trials, the observed F_{ST} values were randomized across the observed set of distances (to preserve the sampling distribution of map distances), and the logistic regression coefficient of F_{ST} on map distance was estimated. The observed logistic regression coefficient was then compared to the distribution of simulated regression coefficients to obtain a probability level for the relationship between outlier status and proximity to QTL. For comparison, we also performed a linear regression of F_{ST} on distance from the nearest QTL, and again used randomization to determine the significance level.

To test the persistence of outlier-QTL linkages, we repeated the F_{ST} outlier analyses on samples from two additional North American locations (Maryland and Iowa). Given the ecological specialization seen in pea aphid populations in Europe (Sandstrom 1996; Simon *et al.* 2003; Ferrari *et al.* 2007), the host races are likely to have diverged long ago in Europe or the Middle East. After their introduction to North America in the mid-1800s, pea aphids that colonized different regions of North America have been more or less geographically isolated for at least 100–150 years (one sexual generation/year). On this basis, we reasoned that if outlier-QTL linkages are not protected from inter-race recombination, we would find few outliers, and that they would differ among collections from different locations in the USA. For these analyses, we used aphids collected in 2002 in Iowa (48 aphids from one alfalfa field, and 47 aphids from one clover field) and Maryland (94 aphids from three alfalfa fields, and 84 aphids from four clover fields). In both of these locations, the two pea aphid races are sympatric and known to be host-specialized (Via 1991 and unpublished data). These aphids were genotyped for 31 of the mapped ALFP markers employed in the New York outlier analysis, but not for the codominant markers.

Neighbour-joining analyses

We explored whether the genetic mosaic pattern would affect the estimated genetic similarity of pea aphid populations from alfalfa and red clover in New York, Iowa and Maryland by contrasting the results of neighbour-joining analyses made with different sets of neutral markers. For one analysis, we used five AFLP markers that are relatively close to QTL and were F_{ST} outliers in all three locations. For the other analysis, we used the five AFLP markers located farthest from any QTL, none of which were outliers in any location. These marker groups were chosen to provide a clear contrast between results obtainable by separating markers into those that are demonstrably affected by divergent selection and those that are not. The analyses were done in PHYLIP (Felsenstein 2004), and bootstrap support was estimated for each branch. Only bootstrap values greater than 80% are considered significant.

Results

Because the F_{ST} values for AFLP markers estimated in DFDIST were extremely highly correlated with those estimated in FDIST2 ($r = 0.999$), we used the FDIST2 values so that the codominant markers could be included in the same outlier analysis. We found that 27% of our markers are F_{ST} outliers between the alfalfa and red clover host races of pea aphids in New York (10 of the 40 mapped AFLP markers and two of the five mapped codominant markers). This is a higher frequency of outliers than found for unmapped markers (5–17% e.g. Wilding *et al.* 2001; Emelianov *et al.* 2004; Scotti-Saintagne *et al.* 2004; Oetjen & Reusch 2007; Nosil *et al.* 2008), which we attribute to the fact that our markers are probably more differentiated on average than if they had been chosen at random, because they can only be mapped if they differ between the mapping parents.

To test whether the high F_{ST} values of the outlier markers could be the result of either a particularly high mutation rate or some demographic artefact (e.g. Beaumont 2005), we used the same data set to estimate F_{ST} values between fields of the same crop, where free gene flow is expected. In these analyses (Fig. 1B,C), only one outlier was found among the 40 AFLP markers tested for aphids from alfalfa, and there were no F_{ST} outliers among the aphids from different red clover fields. This result supports our interpretation that the outlier F_{ST} values between the host races (Fig. 1A) are the result of linkage to QTL under divergent selection.

The logistic regression revealed a significant relationship between outlier status and map distance to the nearest QTL ($P < 0.05$, Fig. 1A). As a check, we also ran a linear regression of F_{ST} value on distance to QTL ($P < 0.02$, predicted values not shown). Our results thus confirm previous claims that F_{ST} outliers have diverged due to their linkage to QTL
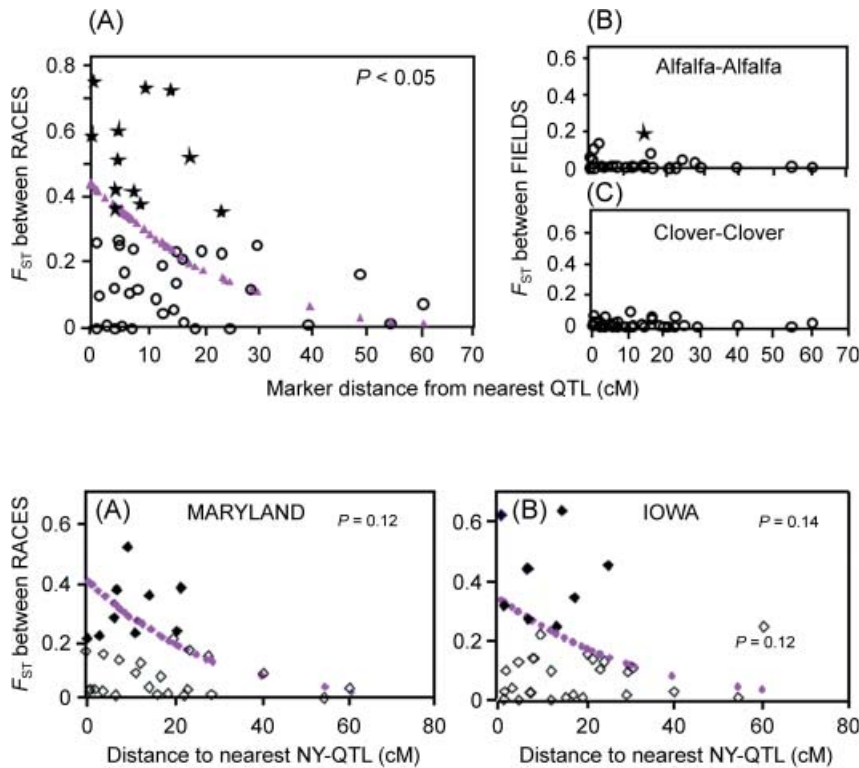
Fig. 1 Relationship between F_{ST} and marker distance to the nearest QTL for traits involved in ecological speciation ($n = 45$). Black stars mark outlier values of F_{ST}, blue circles show non-outlier F_{ST} values, and pink squares show the predicted value for each marker in a logistic regression of outlier status on distance to the nearest QTL. P-values for the regression coefficients were determined by randomization of the observed F_{ST} values across the observed marker distances. (A) F_{ST} calculated between the alfalfa and clover races In New York. (B, C) F_{ST} calculated among fields of each crop type using the same data.

Fig. 2 F_{ST} between alfalfa and clover populations in Maryland, USA (A) and Iowa, USA (B), plotted against marker distance to the nearest QTL on the map made using New York (USA) populations. Symbols are as in Fig. 1, except blue diamonds denote non-outlier F_{ST} values.

under divergent selection (Wilding *et al.* 2001; Emelianov *et al.* 2004; Storz 2005; Rogers & Bernatchez 2005).

In our study, the average distance of the F_{ST} outliers from the nearest QTL is 10.6 cM (Fig. 1A). However, we found that many markers more closely linked to QTL are not outliers (Fig. 1A). We suggest that these are ancestral polymorphisms established before population subdivision occurred under selection. These markers contain no information about population divergence, and should not be used when estimating the size of genomic regions affected by divergent selection. As we discuss later, the prevalence of ancestral polymorphisms until very late in the speciation process seriously complicates empirical estimation of the size of chromosomal blocks affected by divergence hitchhiking.
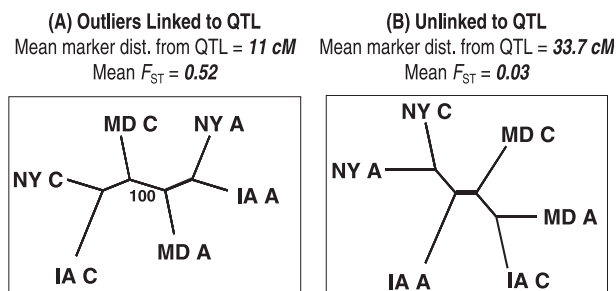
Two of the most distant outliers in the New York population (at 18.5 cM and 25 cM, Fig. 1A), both map outside the 95% CI for the same QTL in a genomic region where we had relatively few markers. Because these two markers could be linked instead to a QTL that we could not resolve on our map (Hawthorne & Via 2001), the inference that markers as far as 20 cM from QTL are affected by divergent selection must be tentative. However, given that Hawthorne & Via (2001) mapped the 3–5 QTL of largest effect for each trait, it seems unlikely that linkage to undiscovered QTL could explain all nine outliers that are > 5 cM from mapped QTL (Fig. 1A). Thus, we suggest twice the distance of the average outlier to the nearest QTL (21.2 cM) as a conservative estimate of the size of the hitchhiking regions around QTL

in these partially reproductively isolated populations of pea aphids in New York.

In the collections from Maryland and Iowa, we genotyped eight of the 12 AFLP markers that were F_{ST} outliers in New York. Of these, seven were also F_{ST} outliers in Maryland (Fig. 2A), and six of the eight were outliers in Iowa (Fig. 2B). Three markers that were not significant F_{ST} outliers in New York were significant in Maryland and/or Iowa. The map distributions of F_{ST} outliers in the Iowa and Maryland populations relative to QTL mapped in the New York population (Fig. 2A,B) are very similar to that seen in Fig. 1A. The logistic regressions of F_{ST} outlier status on marker distance to QTL for both populations are suggestive, given the small sample size ($P = 0.12$ for Maryland and $P = 0.14$ for Iowa). Thus, although F_{ST} outlier-QTL linkages are not completely invariant over time and space, our results show that individual F_{ST} outliers with reliable and persistent linkages to key QTL can readily be identified.

Using the genetic mosaic to study the early history of ecological speciation

We have shown that F_{ST} outliers are significantly associated with key QTL, providing evidence that they reflect the effects of divergent selection on the QTL. This, and their repeatability across populations, suggests that analyses of population structure or phylogenetic relationships based on outlier markers should reveal patterns that could be

(A) Outliers Linked to QTL
Mean marker dist. from QTL = *11 cM*
Mean F_{ST} = *0.52*

(B) Unlinked to QTL
Mean marker dist. from QTL = *33.7 cM*
Mean F_{ST} = *0.03*

Fig. 3 Neighbour-joining trees for pea aphids sampled from alfalfa and red clover in New York (NY A, and NY C), Maryland (MD A and MD C) and Iowa (IA A and IA C). (A) Neighbour-joining tree made using five AFLP markers that are linked to a QTL and have outlier F_{ST} values in all three locations. (B) Neighbour–joining tree for the same populations made using five markers that are far from a QTL and have non-outlier F_{ST} values in all populations. Only bootstrap values greater than 75% are shown.

obscured by ongoing gene exchange at genomic regions unaffected by divergent selection (Wilding *et al*. 2001; Emelianov *et al*. 2004; Dopman *et al*. 2005).

We tested this hypothesis with neighbour-joining analyses of the relationships among pea aphids collected from three regions of the USA (Iowa, New York and Maryland). One neighbour-joining tree was made using data from five mapped AFLP markers that were close to QTL and outliers in all three locations (average distance from QTL = 11 cM; average F_{ST} = 0.52). This analysis reveals a well-supported branching pattern in which populations cluster by host plant affiliation, not by geographical location (Fig. 3A). Despite the sympatry of the two host races in each location, the genomic regions containing F_{ST} outliers are clearly resisting introgression. The 100% bootstrap value between the races suggests that the underlying gene trees of the five individual outliers are highly concordant, as expected for markers affected by the same pattern of divergent selection (Avise 2000).

In contrast, the neighbour-joining tree made with markers known to be unlinked to divergently selected QTL (average distance from QTL = 33.7 cM; average F_{ST} = 0.03) shows no significant pattern of genetic similarity among either geographical or host-associated populations (Fig. 3B). The absence of any clear signal in this analysis suggests that the underlying gene trees for the five markers distant from divergently selected QTL are highly variable and unreliable for use in the analysis of population structure or phylogeography.

Discussion

During ecological speciation, significant among-locus heterogeneity in divergence arises because divergent selection affects some genomic regions and not others. We

call this transient pattern the genetic mosaic of speciation. Early in this phase of ecological speciation, patterns of change at genomic regions directly involved in reproductive isolation can be studied directly by identifying excessively divergent chromosomal regions. In contrast, in mature species, most polymorphic loci will have come into phylogenetic concordance through independent responses to drift, uniform selection or balancing selection (Avise 2000; Ch. 6). By this time, the genomic regions that first caused reproductive isolation under divergent selection are no longer distinctive, having become assimilated into a complex pattern of genetic divergence between species.

Although outlier analysis and the observation of genomic heterogeneity in divergence and gene flow are not new ways to detect selected markers, this is the first time that F_{ST} outliers estimate the size of genomic regions affected by divergent selection. By combining outlier analysis with a QTL map of traits known to cause reproductive isolation between the pea aphid host races, we showed that markers with an average of 10.6 cM from divergently selected QTL carry the footprint of divergent selection as F_{ST} outliers. The surprisingly large size of these genomic regions could not have been revealed by outlier analysis on unmapped markers.

Divergence hitchhiking during ecological speciation

Although the persistent distribution of outliers around divergently selected QTL during ecological speciation is similar in broad outline to hitchhiking within a panmictic population after a selective sweep, it results from a very different mechanism. To emphasize this important difference and minimize confusion, we call the protection from interrace recombination around key QTL during early speciation 'divergence hitchhiking'.

In our analysis, nine of 12 outliers were > 5 cM away from the nearest QTL, with a mean outlier-QTL distance of 10.6 cM (Fig. 1A). This is a much larger hitchhiking region than expected within a panmictic population after a selective sweep, where the region of linkage disequilibrium around a selected gene rapidly shrinks except in regions of reduced recombination (e.g. Begun & Aquadro 1992; Charlesworth 1998). Similarly small hitchhiking regions are thought to occur during speciation (Ting *et al*. 2000; Turner *et al*. 2005), leading to the conventional wisdom that hitchhiking regions in natural populations are small, unless recombination is reduced.

Curiously, recent reports of large genomic regions that are divergent between incipient species appear to have been ignored by students of speciation. In subspecies of house mouse, regions of divergence extend for 11–31 Mb (Harr 2006), and the average distance between markers in significant linkage disequilibrium was 12.2 cM (Payseur *et al*. 2004). Emelianov *et al*. (2004) found that F_{ST} outliers
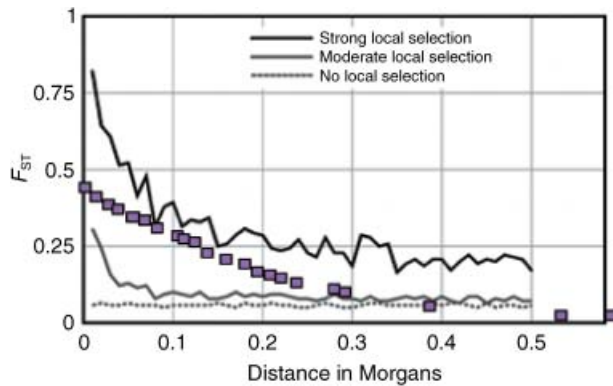
Fig. 4 Predicted values (pink squares) from the logistic regression of data shown in Fig. 1 (A), overlaid on simulation results from Charlesworth *et al.* (1997), using their figure as redrawn by Storz (2005). Strong local selection is $s_{(QTL)} = 0.5$, weak local selection is $s_{(QTL)} = 0.1$.

between two host races of moth clustered on just a few linkage groups, on which 50–90% of markers were significant F_{ST} outliers, suggesting large genomic regions of divergence. The mean distance of outlier F_{ST} to QTL for traits that distinguish dwarf and normal whitefish populations is 16.5 cM (calculated from data in Rogers & Bernatchez 2005), and here, the average outlier-QTL distance was 10.6 cM. All of these far exceed the small hitchhiking regions expected given the conventional wisdom.

These large regions of divergence hitchhiking do not result from selective sweeps, which occurs within populations. Instead, we suggest that they result from a mechanism at the very crux of speciation itself — population subdivision and reduced interbreeding in divergent populations. As strong local (divergent) selection on resource or habitat use minimizes the frequency of migrants and hybrids, and favours habitat or mate choice, populations become subdivided. As successful interbreeding is increasingly limited, the opportunity for recombination between chromosomes from the two specialized races is reduced, causing the *effective* recombination rate around divergently selected QTL to be much lower than the nominal rate based on map distance (which can be calculated in experimental crosses because recombination *per se* is not altered). Although it has been known for a decade that population subdivision can cause a persistent hitchhiking region on either side of a selected locus that is of a size (in Morgans) on the order of the selection coefficient (Charlesworth *et al.* 1997; Fig. 4; Barton 2000; review in Storz 2005), the importance of this result for our understanding of speciation has not yet been fully appreciated.

To check the correspondence of our results with the Charlesworth *et al.* (1997) simulations, we overlaid the predicted values from our logistic regression on their simulation

results (Fig. 4). Like F_{ST}, the probability that a marker will be an outlier varies between 0 and 1, but it is a more appropriate metric for the extent of the hitchhiking region than mean F_{ST}, given the number of our markers whose F_{ST} reflect ancestral polymorphism rather than divergent selection. Our results clearly fall within the simulated range for hitchhiking around a moderately ($s = 0.1$) to strongly ($s = 0.5$) divergently (locally) selected allele (Fig. 4). Could divergent selection on our QTL possibly be this strong?

If *n* QTL affect a given trait, then the average selection coefficient on each QTL allele (s_{QTL}) can be expressed as $s_{QTL} = s_{phenotypic}/2n$, where $s_{phenotypic}$ is selection on a given trait, and the factor of two accounts for diploidy (Rieseberg & Burke 2001). We can use independent estimates of phenotypic selection on the mapped traits from reciprocal transplants on the New York population and an estimate of the number of QTL from the Hawthorne & Via (2001) map to calculate s_{QTL} as follows:

When aphid genotypes from alfalfa are tested on clover, their reduction in fitness is 70%, while the reduction of fitness for clover genotypes tested on alfalfa is 95% (S. Via, unpublished data), giving an average selection coefficient on early fecundity, $s_{phenotypic(fecundity)} = 0.83$. If four major QTL affect performance on each host, the average selection on each QTL allele is $s_{QTL(fecundity)} = 0.83/(4*2) = 0.10$. For habitat acceptance, the average $s_{phenotype(habitat\ choice)} = 0.95$ (calculated from data in Via *et al.* 2000), giving $s_{QTL(habitat\ choice)} = 0.95/8 = 0.12$. These calculations, based on independent estimates of divergent selection on the mapped traits, confirm the good fit of our data to the Charlesworth *et al.* (1997) model. They also suggest that persistent effects of local selection should be felt on the order of 10 cM on either side of our mapped QTL, closely matching the observed mean outlier distance to QTL in our study (10.6 cM, Fig. 1A).

It is possible that some of these regions of divergence hitchhiking in our system may contain several QTL for the same trait. Although the presence of multiple genes rather than one gene in a divergent region is a crucial difference from a functional standpoint, it is unlikely to affect the population genetic interpretation or consequences of divergence hitchhiking. Regions of divergence hitchhiking of the same size could result from either a single QTL of very large effect, or from several individual QTL with smaller individual effects. In the latter case, the individual regions of divergence hitchhiking would be smaller, reflecting smaller selection coefficients. However, if the QTL are close enough that they cannot be distinguished on a map, the individual regions of divergence hitchhiking are likely to overlap, creating a region of about the same size as that for a single QTL of the same total effect.

In sum, our results provide strong evidence that population subdivision, an intrinsic element of the speciation process, can lead to large regions around divergently selected

QTL that are protected from recombination between genomic regions containing different locally adapted QTL alleles. This is consistent with the conventional wisdom that restricted recombination is required for extensive hitchhiking, because recombination that could separate locally adapted QTL and their flanking markers is indeed reduced to a rate far below the expected level based on map distance. Although recombination within the divergent populations occurs as expected, it has little population genetic effect because individuals within populations are likely to bear the same (or ancestrally polymorphic) marker alleles flanking the locally selected QTL.

Detecting divergence hitchhiking is only possible early in speciation

F_{ST} outliers resulting from divergent selection can only be identified relatively early in the genetic mosaic phase of ecological speciation, before other causes of divergence begin to accumulate. Moreover, as the overall level of genetic divergence increases, detecting outliers caused by any mechanism becomes increasingly difficult. Thus, regions of divergence hitchhiking will generally be missed in retrospective analyses of reproductive isolation using highly reproductively isolated taxa.

Within incipient species, the extent of divergence hitchhiking may be seriously underestimated due to pervasive ancestral polymorphism. For example, with a dense map or a genome sequence, markers or sequence adjacent to outliers can be examined for signs of divergent selection. However, ancestrally polymorphic loci contain no information about divergent selection; thus, a low F_{ST} marker close to an outlier will not necessarily represent the edge of the divergence hitchhiking region. Before making that inference, it is necessary to verify that a low F_{ST} marker is actually affected by gene flow and is not an ancestral polymorphism. The use of autocorrelation analysis of marker divergence to detect regions affected by divergent selection is similarly affected by ancestral polymorphism and should be avoided as a method for detecting the size of a given region of divergence hitchhiking. The final problem in detecting regions of divergence hitchhiking stems from the considerable stochastic variability in F_{ST} values of individual markers that results from independent realizations of the coalescent process. This is why outlier identification requires coalescent simulation (e.g. Beaumont 2005). Although outliers can be statistically distinguished from non-outliers, it is not appropriate to interpret differences between the nominal F_{ST} values of adjacent markers when attempting to estimate the extent of divergence hitchhiking.

In this type of analysis, only the markers identified as F_{ST} outliers provide information about divergent selection on nearby QTL that is not confounded by either ancestral polymorphism or the stochastic variance of nominal F_{ST}

values. Given a large outlier analysis and a dense map, a curve-fitting or spline technique might be useful for approximating the shape of regions of divergence hitchhiking around individual QTL. Ultimately, however, we will need to move beyond the simple identification of F_{ST} outliers to determine the size of individual regions of divergence hitchhiking. It is likely that detailed coalescent analyses of multiple closely linked markers will be required to distinguish between low-divergence markers that are ancestrally polymorphic and those that are subject to contemporary gene flow. Accurately delineating individual regions of divergence hitchhiking is an important issue that warrants further research.

Population subdivision and divergence hitchhiking in sympatry and allopatry

During ecological speciation, population subdivision begins as soon as divergent selection against migrants and hybrids reduces the fraction of matings that occur between individuals bearing alternate QTL alleles. In allopatric populations under divergent selection on ecologically important traits, the genetic mosaic will still occur, although it may be difficult to detect if divergent selection is weak. Divergence hitchhiking does not occur in allopatry because there is no interpopulation mating and thus no potential for reduction of effective recombination around divergently selected QTL. Moreover, divergence hitchhiking is unlikely to be seen in outlier analyses performed on taxa that are well into the speciation process, such as species that form hybrid zones after secondary contact. Such taxa are likely to have accumulated considerable allelic divergence in allopatry through independent responses to uniform selection, balancing selection, or drift. Not only do these genetic differences elevate the overall degree of divergence, making it more difficult to detect F_{ST} outliers, but genetic divergence accumulated in allopatry will often be unrelated to the map locations of QTL under divergent selection.

However, if taxa that diverge in allopatry become sympatric relatively early in speciation, ecologically based selection against migrants and hybrids will immediately produce divergence hitchhiking, protecting large regions around key QTL from recombination. At the same time, gene flow in genomic regions unaffected by divergent selection will homogenize neutral or beneficial inter-race polymorphisms that may have accumulated while the populations were allopatric, and allelic substitutions that cause genetic incompatibilities in hybrids may be purged if their selective benefit within populations does not compensate for the fitness disadvantages in hybrids. The instant appearance of divergence hitchhiking upon secondary contact and the potential elimination of some genetic incompatibilities that accumulated in allopatry substantially reduce the degree

to which the geographical location of initial divergence is relevant to the outcome of speciation for divergent taxa that become sympatric relatively early in the process.

Through its association with reduced inter-race recombination, divergence hitchhiking poses an additional challenge to the conventional view of sympatric speciation. Classic models of sympatric speciation without linkage or pleiotropy (e.g. Felsenstein 1981; review Via 2001) conclude that free gene flow and recombination foil the evolution of assortative mating, severely reducing the potential for speciation in sympatry. A similar expectation holds in models of secondary contact, where it is assumed that free recombination will hasten the collapse of divergence (e.g. Liou & Price 1994).

The existence of large regions of divergence hitchhiking around locally selected QTL caused by reduced inter-race recombination eliminates this key constraint on sympatric speciation. Our data suggest that on average in pea aphids from alfalfa and red clover in New York, a 21.2 cM region around a QTL affecting resource use is protected from inter-race recombination. If this result is general, then even rather loosely linked genes for habitat choice or mate choice could readily be held in long-term linkage disequilibrium between races without the need for tight physical linkage or pleiotropy (Hawthorne & Via 2001). Recognition of divergence hitchhiking may thus neutralize one of the main ongoing challenges to sympatric speciation.

By protecting loosely linked genes from inter-race recombination, divergence hitchhiking has essentially the same effect as a chromosomal inversion. However, unlike an inversion, divergence hitchhiking does not affect within-race linkage disequilibrium, and it can only be seen during the genetic mosaic phase of early ecological speciation. These protected regions are likely to be invisible in a typical retrospective analysis, because by the time good species are recognized, much of the genome will have diverged by other means and become phylogenetically concordant with the branching pattern initiated by divergent selection (Avise 2000; Ch. 6). By this time, the divergence hitchhiking regions around key QTL, originally visible as a cluster of F_{ST} outliers, will have been assimilated into the overall pattern of divergence.

Consequences of the genetic mosaic for phylogenetic analyses of taxa close to the species boundary

Determining the phylogenetic history of very closely related taxa is always problematic, due to stochastic variation in gene trees and incomplete lineage sorting (Maddison 1997; Pollard *et al.* 2006). During the genetic mosaic phase of ecological speciation, the localized effects of divergent selection cause additional variance in gene genealogies. However, rather than compounding the problems of phylogenetic analysis below the species level, the genetic mosaic provides a solution: using F_{ST} outliers, a phylogenetically important signal can be extracted from the noise caused by coalescent stochasticity and incomplete lineage sorting. Studies, including this one, that have contrasted phylogenetic analyses of incipient species with and without F_{ST} outliers, show that outliers encode a picture of genetic relatedness paralleling that of the quantitative traits under divergent selection, while analyses without outliers reveal little of significance (Wilding *et al.* 2001; Emelianov *et al.* 2004).

Recognition that diverging taxa move through the genetic mosaic phase is crucial for those interested in phylogenetic and phylogeographical analyses below the species level. Given the very different evolutionary histories of markers affected by divergent selection and those that are uninvolved in divergence, it should be of no surprise that phylogenetic analyses of very closely related taxa made with randomly or arbitrarily chosen markers conflict with one another (e.g. Shaw 2002; Beltran *et al.* 2002; Machado & Hey 2003; Mallarino *et al.* 2004; Dopman *et al.* 2005). They may even contradict known patterns of phenotypic divergence, as did two analyses using mtDNA in pea aphids, which failed to reflect the pronounced quantitative genetic divergence of the host-associated races (Boulding 1998; Birkle & Douglas 1999), despite its earlier confirmation by reciprocal transplants (e.g. Via 1989, 1991).

Neutral markers are the gold standard for higher level phylogenetics because they minimize homoplasy. This is fine for anciently diverged taxa in which there is no gene flow and ancestral polymorphisms have generally been resolved. However, for taxa in the genetic mosaic phase of divergence under selection, only markers associated with the selected traits will give an accurate picture of the history of adaptation and the branching patterns among lineages. It is likely to be many years after the initial lineage split under selection before the majority of the genome reaches phylogenetic concordance with the branching pattern established early in the speciation process by the divergently selected QTL (Avise 2000; Ch. 6). Although markers that are demonstrably unaffected by divergent selection will be very useful for demographic analyses, using F_{ST} outliers linked to key QTL (in the absence of the selected genes themselves) is the only practical way to avoid conflicting stories about adaptive divergence. We caution that simple proximity to QTL in the absence of outlier information is not a good criterion for choosing markers for phylogenetic analyses of incipient species because ancestral polymorphisms may become trapped in the hitchhiking regions around QTL when populations initially diverge.

Conclusions

Exploiting the simplicity of the genetic mosaic early in ecological speciation by divergent selection can reveal the genetic changes that cause the earliest reproductive isolation without the confounding effects of genetic differences that

accumulate between species over time. Given the ability to identify individual genomic regions that contribute to adaptive divergence and reproductive isolation, we can probe the genomic basis of parallel evolution, ask whether taxa evolve collectively during divergence (e.g. Rieseberg & Burke 2000; Morjan & Rieseberg 2004), and estimate the relative order in which traits (and even individual QTL) associated with ecological speciation have evolved in particular cases. We can also study the history of adaptation and visualize new branches on the species tree in very closely related taxa without the phylogenetic conflict that has plagued previous analyses.

We have much to learn about the nature of increasing genetic divergence along the long timeline of ecological speciation, from the establishment of the genetic mosaic by divergent selection, through the long period in which the species boundary is porous, and finally to the stage at which successful interbreeding is completely blocked. An integrated picture of the various mechanisms involved in genetic divergence as speciation proceeds in different ecological situations would be a worthy goal for future speciation research.

Acknowledgements

References

Avise JC (2000) *Phylogeography: The History and Formation of Species.* Harvard University Press, Cambridge, Massachusetts.

Barton NH (2000) Genetic hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **355**, 1553–1562.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 13–45. Oxford University Press, Oxford, UK.

Beaumont MA (2005) Adaptation and speciation: what can F_{ST} tell us? *Trends in Ecology & Evolution*, **20**, 435–440.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.

Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*, **356**, 519–520.

Beltran M, Jiggins CD, Bull V *et al.* (2002) Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Molecular Biology and Evolution*, **19**, 2176–2190.

Birkle L, Douglas AE (1999) Low genetic diversity among pea aphid (*Acyrthosiphon pisum*) biotypes of different plant affiliation. *Heredity*, **82**, 605–612.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773–783.

Boulding EG (1998) Molecular evidence against phylogenetically distinct host races of the pea aphid (*Acyrthosiphon pisum*). *Genome*, **41**, 769–775.

Bradshaw HD Jr, Schemske DW (2003) Allele substitution at a flower color locus produces a pollinator shift in two monkeyflower species (*Mimulus*). *Nature*, **426**, 176–178.

Bush GL (1994) Sympatric speciation in animals: new wine in old bottles. *Trends in Ecology & Evolution*, **9**, 285–288.

Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to asses the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**, 945–956.

Cavalli-Sforza LL (1966) Population structure and human evolution. *Royal Society of London*, B, **164**, 362–369.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155–174.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Dobzhansky T (1937) *Genetics and the Origins of Species*. Columbia University Press, New York.

Dopman EB, Perez L, Bogdanowicz SM, Harrison RG (2005) Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proceedings of the National Academy of Sciences, USA*, **102**, 14706–14711.

Emelianov I, Marec F, Mallet J (2004) Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 97–105.

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

Felsenstein J (2004) PHYLIP. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, Washington.

Ferrari J, Godfray CHJ, Faulconbridge AS, Prior K, Via S (2007) Population differentiation and genetic variation in host choice among pea aphids from eight host plant genera. *Evolution*, **60**, 1574–1584.

Harr B (2006) Genomic islands of differentiation between house mouse subspecies. *Genome Research*, **16**, 730–737.

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

Howard DJ, Berlocher SH, eds. (1998) *Endless Forms: Species and Speciation*. Oxford University Press, New York.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Lexer C, Kremer A, Petit RJ (2006) Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology*, **15**, 2007–2012.

Liou LW, Price TD (1994) Speciation by reinforcement of premating isolation. *Evolution*, **48**, 1451–1459.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Machado CA, Hey J (2003) The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 1193–1202.

Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution*, **19**, 472–488.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Mallarino R, Bermingham E, Willmott K, Whinnett A, Jiggins CD (2004) Molecular systematics of the butterfly genus *Ithomia* (Lepidoptera: Ithomiinae): a composite phylogenetic hypothesis based on seven genes. *Molecular Phylogenetics and Evolution*, **34**, 625–644.

Mallet J (2005a) Speciation in the 21st century. *Heredity*, **95**, 105–109.

Mallet J (2005b) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**, 229–237.

Mallet J, Beltran M, Neukirchen W, Linares M (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology*, **7**, 28.

Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York.

Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, **13**, 1341–1356.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: 'Isolation by adaptation' and multiple roles for divergent selection. *Evolution*, 62, 316–336.

Oetjen K, Reusch TB (2007) Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Molecular Ecology*, **16**, 5156–5167.

Otte D, Endler JA, eds. (1989) *Speciation and its Consequences*. Sinauer Associates, Sunderland, Massachusetts.

Payseur BA, Krenz JG, Nachman M (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, **58**, 2064–2078.

Pollard DA, Iyer VN, Moses AM, Eisen MB (2006) Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *Public Library of Science, Genetics*, **2**, 1634–1647.

Ramsey JR, Bradshaw HD Jr, Schemske DW (2003) Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution*, **57**, 1520–1534.

Rice WR, Hostert EE (1993) Laboratory experiments on speciation: what have we learned in 40 years? *Evolution*, **47**, 1637–1653.

Rieseberg LH, Burke JH (2001) The biological reality of species: gene flow, selection and collective evolution. *Taxon*, **50**, 47–67.

Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of barrier to gene flow between two sunflower species. *Genetics*, **152**, 713–727.

Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, **14**, 351–361.

Rogers SM, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus sp. Salmonidae*) species pairs. *Molecular Biology and Evolution*, **24**, 1423–1428.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.

Sandstrom J (1996) Temporal change in host adaptation in the pea aphid, *Acyrthosiphon pisum*. *Ecological Entomology*, **21**, 56–62.

Schemske DW (2000) Understanding the origin of species. *Evolution*, **54**, 1069–1073.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, **16**, 372–380.

Scotti-Saintagne C, Mariete S, Porth I *et al.* (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Leibl]. *Genetics*, **168**, 1615–1626.

Shaw KL (2002) Conflict between nuclear and mictochondrial DNA phylogenies of a recent species radiation: what mitochondrial DNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences, USA*, **99**, 16122–16129.

Simon JC, Carre S, Boutin M *et al.* (2003) Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 1703–1712.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Templeton AR (1981) Mechanisms of speciation- a population genetic approach. *Annual Review of Ecology and Systematics*, **12**, 23–48.

Ting C-T, Tsaur S-C, Wu C-I (2000) The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proceedings of the National Academy of Sciences, USA*, **97**, 5313–5316.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *Public Library of Science, Biology*, **9**, 1572–1578.

Vasemagi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology and Evolution*, **22**, 1067–1076.

Via S (1989) Field estimation of variation in host plant use between local populations of pea aphids from two crops. *Ecological Entomology*, **14**, 357–364.

Via S (1991) The genetic structure of host plant adaptation in a spatial patchwork: demographic variability among reciprocally transplanted pea aphid clones. *Evolution*, **45**, 827–852.

Via S (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution*, **53**, 1446–1457.

Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, **16**, 381–390.

Via S, Hawthorne DJ (2001) Genetic architecture of ecological specialization and incipient speciation in divergent races of pea aphids. *The American Naturalist*, **159**, S47–S60.

Via S, Bouck AC, Skillman S (2000) Reproductive isolation between sympatric races of pea aphids. II. Selection against migrants and hybrids in the parental environment. *Evolution*, **54**, 1626–1637.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange

between parapatric morphs of *Littorina saxatillis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**, 611–619.

Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*, **175**, 1883–1893.

S. Via's work focuses on the ecological genetics of adaptation to different environments and the genetics of ecological speciation as a result of divergent selection on resource use. Joan West is interested in evolutionary genetics of insect-plant interactions.