MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/mec.12021

Localizing F_{ST} outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow

SARA VIA,*† GINA CONTE,*‡ CASEY MASON-FOLEY* and KELLY MILLS*

*Department of Biology, University of Maryland, College Park, MD, 21042, USA, †Department of Entomology, University of Maryland, College Park, MD 21042, USA

Abstract

Populations that maintain phenotypic divergence in sympatry typically show a mosaic pattern of genomic divergence, requiring a corresponding mosaic of genomic isolation (reduced gene flow). However, mechanisms that could produce the genomic isolation required for divergence-with-gene-flow have barely been explored, apart from the traditional localized effects of selection and reduced recombination near centromeres or inversions. By localizing F_{ST} outliers from a genome scan of wild pea aphid host races on a Quantitative Trait Locus (QTL) map of key traits, we test the hypothesis that between-population recombination and gene exchange are reduced over large 'divergence hitchhiking' (DH) regions. As expected under divergence hitchhiking, our map confirms that QTL and divergent markers cluster together in multiple large genomic regions. Under divergence hitchhiking, the nonoutlier markers within these regions should show signs of reduced gene exchange relative to nonoutlier markers in genomic regions where ongoing gene flow is expected. We use this predicted difference among nonoutliers to perform a critical test of divergence hitchhiking. Results show that nonoutlier markers within clusters of F_{ST} outliers and QTL resolve the genetic population structure of the two host races nearly as well as the outliers themselves, while nonoutliers outside DH regions reveal no population structure, as expected if they experience more gene flow. These results provide clear evidence for divergence hitchhiking, a mechanism that may dramatically facilitate the process of speciation-with-gene-flow. They also show the power of integrating genome scans with genetic analyses of the phenotypic traits involved in local adaptation and population divergence.

Keywords: adaptation, contemporary evolution, ecological genetics, natural selection, population genetics—empirical, speciation

Received 19 November 2011; revision received 24 June 2012; accepted 11 July 2012

Introduction

The physical impossibility of gene flow between allopatric populations causes complete genomic isolation and allows genetic divergence to accumulate between populations even under weak selection or drift. Without physical isolation, the homogenizing effects of gene flow and between-population recombination have made the potential for speciation controversial (review in Coyne & Orr 2004). However, wild populations that utilize different resources in sympatry often maintain significant phenotypic divergence (Via 1991, 2001), suggesting that the difficulties suggested by simple models of speciation (Felsenstein 1981) may not reflect the range of mechanisms operating in nature. Typically, divergent locally adapted populations or ecotypes in sympatry are genomic mosaics of divergent and nondivergent regions, suggesting that divergence involves an underlying mosaic of genomic isolation in which gene exchange is reduced at

Correspondence: Sara Via, Fax: 301-314-9358; E-mail: svia@umd.edu
‡ Present address: Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada

© 2012 Blackwell Publishing Ltd

sites affecting local adaptation, but is ongoing elsewhere (Via 2009, 2012; Pinho & Hey 2010). Detailed genome scans are now revealing such heterogeneous genome-wide patterns of divergence (Hohenlohe *et al.* 2010; Lawniczak *et al.* 2010; Jones *et al.* 2012), but little attention has been paid to the underlying genomic mosaic of isolation beyond invoking the usual suspects: selection that overcomes migration and restricted recombination in centromeric or inverted regions of the genome.

Increasingly, genome scans are being used to identify candidate genes involved in speciation (Begun *et al.* 2007; Sapir *et al.* 2007; Yatabe *et al.* 2007; Turner *et al.* 2008; Fournier-Level *et al.* 2011). Such studies follow the logic of Lewontin & Krakauer (1973) and interpret each excessively divergent marker (i.e. 'outlier' or 'F_{ST} outlier') as tightly linked to a target of divergent selection. Accordingly, apart from divergence around centromeres or chromosomal break points, genetically divergent regions are generally thought to be small and maintained by a localized alteration in the migration/selection balance that extends just a few kb beyond the selected gene (Ting *et al.* 2000; Wu 2001; Wood *et al.* 2008; Feder & Nosil 2010). Via (2012) called this view of genomic isolation between sympatric populations 'multilocus migration/selection balance' (MM/SB) because it extends a one-locus population genetic process across the genome and depicts the pattern of genomic isolation in sympatry as an exact replica of the heterogeneous pattern of genomic divergence (Via 2009; Fig. 1b).

We test an alternative view in which gene flow is restricted over large genomic regions that are expected to include multiple peaks of genomic divergence. Theory suggests that early in the process of divergence-with-gene-flow while population sizes are small, moderately strong divergent selection on gene(s) affecting one or more key phenotypic traits can reduce the realized frequency of inter-race recombination and gene exchange over a substantial genomic region (Charlesworth *et al.* 1997; Feder & Nosil 2010). Although each divergence hitchhiking (DH) region may begin around a single divergently selected gene/QTL, subsequent divergence of additional loosely linked QTL is expected, producing clusters of divergent genes/QTL that experience less gene exchange and extend across a larger genomic region than expected around a single divergently selected gene (Via 2009). Within these regions of genomic isolation, which may ultimately extend for many Mb, loosely linked genes/QTL can diverge coordinately without disruption by between-population recombination, as if they were far more tightly linked than their map distance suggests. Because these regions of coordinated evolution only occur around divergently selected genes/QTL, Via & West (2008) called this mechanism of genomic isolation 'divergence hitchhiking' (DH) to distinguish it from the transient regions of traditional hitchhiking that occur within populations after a selective sweep.

Within divergence hitchhiking regions, reduced gene exchange permits loosely linked loci uninvolved in speciation to diverge by either genetic drift or independent directional selection. Thus, F_{ST} outliers may not always be tightly linked to a 'speciation gene' (Via 2009, 2012). Moreover, clusters of outliers and divergently selected genes/QTL caused by divergence hitchhiking are not necessarily limited to regions of structurally reduced recombination. Indeed, evidence is mounting that clusters of divergent sites outside centromeres and inversion break points are common (Hohenlohe *et al.* 2010; Lawniczak *et al.* 2010; Neafsey *et al.* 2010; Jones *et al.* 2012).

To understand the genetic mosaic of isolation, we must determine the locations of outlier markers relative to the divergently selected QTL that cause speciation. This has rarely been done (Rogers & Bernatchez 2007; Via & West 2008; Nadeau *et al.* 2012). Although QTL maps are generally too imprecise to identify individual genes (MacKay *et al.* 2009), genomic regions of a few cM that are directly involved in local adaptation and speciation can be identified by mapping phenotypic traits known to cause ecological specialization and reproductive isolation (Barrett & Hoekstra 2011). The proposed regions of divergence hitchhiking that manifest as colocalizing outliers and divergently selected QTL can then be used to test the common assumption that each F_{ST} outlier tags a gene under divergent selection (Stapley *et al.* 2010).

Here, we present a QTL map of traits known to cause assortative mating and ecologically based reproductive isolation between sympatric host-associated races of pea aphids, a model system for incipient speciation-with-gene-flow (Peccoud & Simon 2010). Using this map, we show that molecular signatures of divergent selection from a genome scan colocalize with genomic regions affecting traits that cause reproductive isolation. This integrated analysis is one of the first for which ecological, quantitative genetic and molecular data have been jointly interpreted to analyse the genomic basis of ecological speciation (LeCorre & Kremer 2012; Nosil & Feder 2012; Strasburg *et al.* 2012; reviews in Via 2012). This work extends our previous analyses (Hawthorne & Via 2001; Via & West 2008) in several important ways. First, the addition of 118 microsatellite markers yields an integrated linkage map on which QTL can be mapped more precisely than on our AFLP-based map and also increases the number of mapped markers used in the genome scans of wild pea aphid populations. In addition, this second-generation map permits evaluation of how individual genomic regions contribute to phenotypic selection against migrants and to the genetic correlations

© 2012 Blackwell Publishing Ltd

that facilitate the evolution of ecological divergence and assortative mating (Hawthorne & Via 2001).

We use this integrated outlier/QTL map to identify proposed regions of divergence hitchhiking and to test whether F_{ST} outliers and divergently selected QTL are significantly clustered. We also introduce a clear test for the existence of divergence hitchhiking: under MM/SB, each divergent marker is assumed to be tightly linked to a target of divergent selection. Because nonoutlier markers between divergent sites are considered equally likely to experience gene flow in this view, they should have similar population genetic properties. In contrast, if nonoutlier markers within DH regions are protected from gene exchange as predicted, they could have accumulated more signal of population divergence than nonoutlier markers exposed to ongoing gene flow. Thus, under divergence hitchhiking, we expect two classes of outlier markers that are distinguished by the predicted degree of gene exchange to which they are exposed (Table 1). We test this key prediction of divergence hitchhiking by comparing analyses of population structure made using nonoutlier markers from within and outside the proposed DH regions.

The divergent host races of pea aphids on alfalfa and red clover: a model system for the analysis of speciation-with-gene-flow

Pea aphids (*Acyrthosiphon pisum pisum*) were introduced to North America from Europe, probably in the mid-1800s (Eastop 1971). Although sympatric populations of pea aphids on alfalfa and red clover in Europe are genetically divergent and somewhat ecologically specialized (Ferrari *et al.* 2006, 2008; Peccoud & Simon 2010), the most specialized populations found on these crops to date are in the eastern and midwestern USA (Via 1991, 1999; Via *et al.* 2000; Hawthorne & Via 2001). In upstate New York, USA, field sampling of winged migrants early in the summer revealed that only about 10% of pea aphids colonizing new fields of alfalfa and clover originated from the other host plant (Via 1999). Given that reciprocal transplants in both Iowa (Via 1999) and New York showed intense phenotypic selection against such migrants (Via *et al.* 2000), realized migration between hosts is likely to be even < 10%. Finally, F_1 hybrids average only about 50% of the fecundity of the locally adapted parental forms on both hosts (Via *et al.* 2000), further reducing the opportunity for recombination and gene exchange between these specialized host races. The extensive field and laboratory work on host plant adaptation in pea aphids make this one of the few experimental systems in which molecular signatures of divergence can be interpreted in the light of both quantitative genetic and ecological analyses of the phenotypic traits that cause reproductive isolation between sympatric populations.

Materials and methods

Sampling pea aphids for the mapping cross and the population genetic analyses

In 1991, 10 adult pea aphids were collected at widely spaced locations within each of two alfalfa and two red clover fields within a 20 sq. mile area in Tompkins Co, NY, USA (see Via 1999 for map). Each of these 40 genotypes (henceforth 'clones') was maintained in individual clonal culture on the collection host under 'summer' conditions (16L:8D, 20 °C), using rigorous protocols to avoid contamination. The specialized parents of the mapping cross came from this sample. In 2001, we returned to this study area and collected 100 adult pea aphids from several alfalfa fields and another 100 adults from adjacent red clover fields. These aphids were immediately frozen at −80 °C for DNA extraction and were used for the genome scan of wild populations. Using 15 microsatellite markers, we verified that each aphid in this collection has a unique genotype (Via, unpublished data), which simplifies the interpretation of population genetic analyses, given the potential for clonal replication in this species.

The mapping cross

Based on reciprocal transplants using the 1991 collections (data not shown), we chose one pea aphid genotype that was highly specialized on alfalfa and another that was a clover specialist as parents for the reciprocal mapping crosses. Sexual forms of each clone were induced by manipulating photoperiod over a 3-month period. Because pea aphids are cyclically parthenogenetic, many sexual individuals (both male and female) can be obtained from a single genotype. Reciprocal matings between these alfalfa and clover specialists were set up by pairing a male of one genotype with several females of the other genotype. Fertilized eggs were surface sterilized with a 10% bleach solution, then placed into sterile dishes with autoclaved water for a 100-day incubation period with temperature alternating between 4 and 0 °C during the 12L:12D photoperiod. Egg hatch was between 75% and 85%.

Because pea aphid eggs hatch into parthenogenetic individuals, we established a clonal lineage from each F_1 progeny. After phenotype testing 35 F_1 genotypes in a reciprocal transplant (Via *et al.* 2000), we identified two F_1 genotypes with fecundities close to the midparent average on each host. We reasoned that these two clones should each contain approximately equal frac-

© 2012 Blackwell Publishing Ltd

Table 1 Marker classes defined by outlier status and genomic location

Class
Class 1: Markers shown by coalescent simulation to be significantly more divergent between populations than the average neutral marker (F_{ST} outliers)
Class 2: Nonoutlier markers interspersed among outliers within a proposed region of divergence hitchhiking
Class 3: Nonoutlier markers located at least 3 cM from the boundary of the nearest divergence hitchhiking region (defined as the outermost outlier or QTL in a cluster)

tions of the parental QTL alleles for host specialization. We then induced multiple males and females from each of these two F_1 genotypes and reciprocally crossed them as before to obtain 198 F_2 progeny. Each parthenogenetic F_2 genotype was maintained in individual clonal culture with free choice of both alfalfa and red clover and was phenotyped for both fecundity in the first 9 days of adult life and behavioural acceptance. Each trial consisted of a single individual of each clone tested on one of the host plants (mean 3.4 replicates of each F_2 genotype for each trait/plant combination, methods in Hawthorne & Via 2001). A large sample of every genotype was frozen at −80 °C for genotypic analysis.

Marker development and genotyping the F_2 and the field samples

The microsatellites used in this map were developed from a set of pea aphid EST sequences (IDs in Table S1, Supporting information; Sabatir-Munoz *et al.* 2006). Each prospective microsatellite was tested for polymorphism and phase in the mapping family by genotyping the parents and F_1 using a labelled M13 universal primer (Boutin-Ganache *et al.* 2001), resulting in 115 polymorphic microsatellites that could be reliably amplified (primers in Table S1, Supporting information). These were assembled into multiplex groups with 22 additional microsatellites obtained by screening genomic libraries (Caillaud *et al.* 2004; Caillaud unpublished data; Table S1, Supporting information). The 198 F_2 and the 200 wild-collected individuals used in the genome scan were genotyped for these 137 microsatellites. All allele calls from GeneMapper v3.7 (Applied Biosystems, Foster City, CA, USA) were manually verified.

Linkage map and QTL map construction

For linkage mapping, we used the software JoinMap4 (van Ooijen 2006). Starting with a map that included every marker, we progressively eliminated poorly fitting markers until we had an acceptable balance between coverage and the goodness of fit chi-square test for each of the four linkage groups. Only 118 microsatellites could be placed confidently on the linkage map, which also includes 66 AFLP, 11 sequence-tagged codominant markers and three allozyme markers. The QTL map was constructed with MapQTL6 (van Ooijen 2009), using the multiple QTL model (MQM) method. We chose these mapping programs because they accept the variety of segregation patterns typical of outcrossed populations. Additive and dominance effects of each QTL were estimated, but we did not pursue epistatic effects for technical reasons. We identified the sex chromosome by genotyping XO males for a selection of mapped markers and noting that LG1 had no heterozygotes (Via, unpublished data).

In MapQTL6, the additive phenotypic effect of a QTL is calculated as the expected phenotypic difference between an individual bearing two alleles at that QTL from the alfalfa specialist parent (henceforth 'alfalfa alleles') and an individual with two QTL alleles from the clover specialist parent (henceforth 'clover alleles'). Calculations are based on an F_1 cross coded as $ab \times cd$ (where a, c are the alleles inherited from the alfalfa specialist, and b, d are inherited from the clover specialist (MapQTL6 manual, p. 27; van Ooijen 2009). In this coding scheme, a positive QTL effect means that the phenotypic value is larger in individuals with two alfalfa alleles than in those with two clover alleles, while a negative effect means the reverse.

Estimating QTL for traits expressed in different environments

Instead of estimating the genotype × environment interaction (GEI) for each trait, we chose to consider the expression of a given trait in each of the two host plant environments as a separate 'character state' (Yamada 1962). This approach acknowledges that partially nonoverlapping sets of genes/QTL may underlie the phenotype produced by a given genotype in different environments, and it allows the evolutionary impact of GEI to be interpreted quantitatively using the cross-environment genetic correlations (Via & Lande 1985). We mapped each character state separately, that is, fecundity on alfalfa or fecundity on clover and behavioural acceptance of alfalfa or acceptance of clover. This immediately distinguishes genomic regions exhibiting antagonistic pleiotropy (QTL for both character states colocalize and have opposite directionality) from regions that contain a QTL for only one character state (i.e. show conditional neutrality) with-

© 2012 Blackwell Publishing Ltd

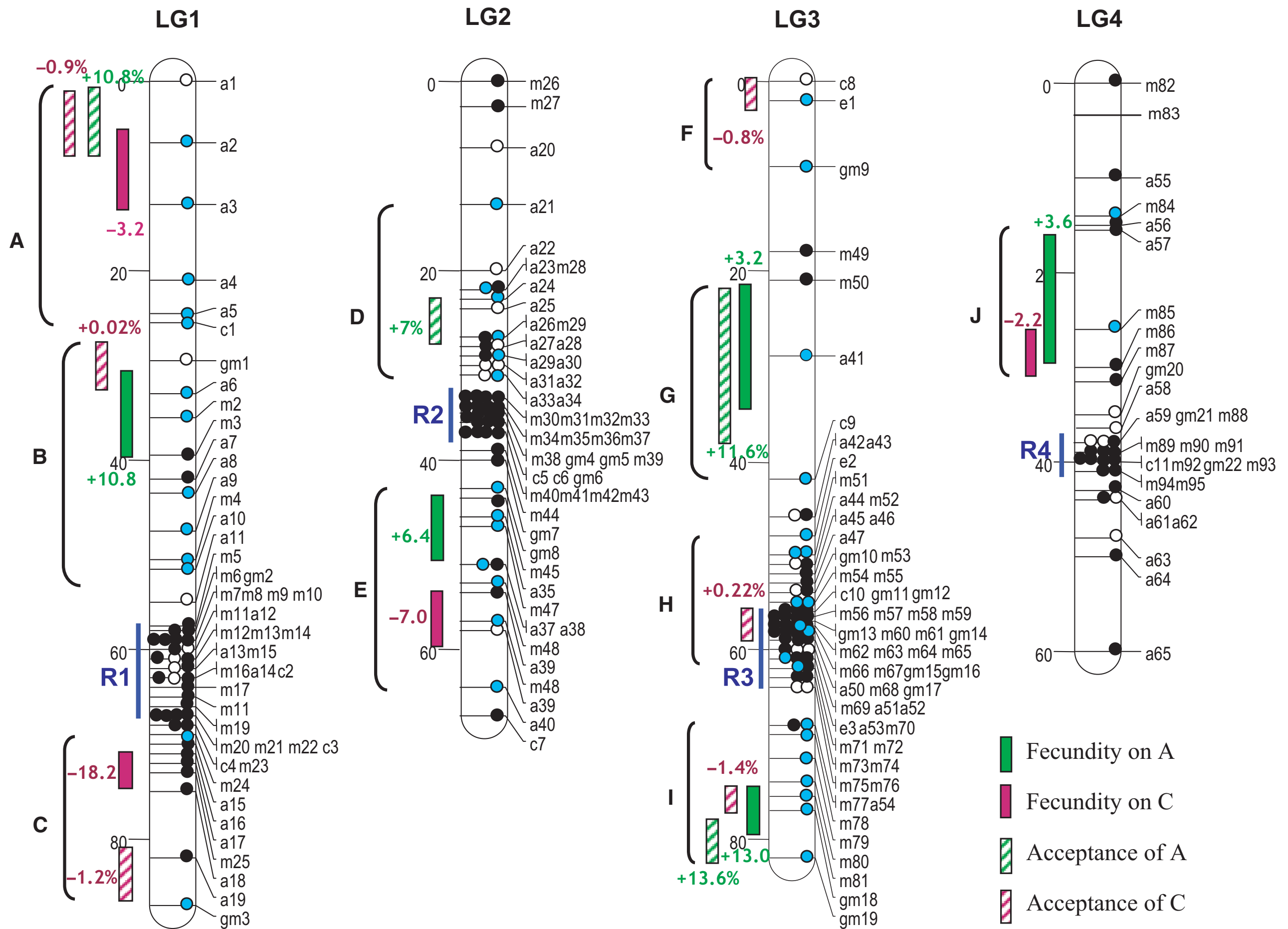

Fig. 1 Second-generation QTL map of differential performance and host acceptance by pea aphid host races specialized on either alfalfa or red clover. Each trait was analysed as a separate character state in each environment (i.e. fecundity on alfalfa and fecundity on clover, colours indicated on key). Bars show the 2-LOD interval for each QTL. Marker location was rounded to the nearest cM to make the marker names legible. Numbers next to each QTL are the effect size with reference to the alfalfa specialist and are in units of number of offspring for the QTL affecting fecundity and percentage change in host acceptance for the QTL affecting host choice. Results of the genome scan between the wild aphid populations on alfalfa and red clover are shown as dots within the bars for each linkage group (black = nonoutlier, blue = F_{ST} outlier at $P < 0.01$, open = not tested). Brackets with letters A–J mark proposed regions of divergence hitchhiking; regions R1–R4 are the clumps of microsatellites evaluated in the text.

out the additional statistical testing required by the GEI approach (Anderson *et al.* 2012).

F_{ST} estimation and outlier analysis

F_{ST} (the standardized allele frequency difference between populations) can be highly variable among markers (Holsinger & Weir 2009). Not only are F_{ST} values affected by divergent or balancing selection (Beaumont 2005), they also vary due to the stochasticity of the coalescent process that acts independently at each locus. Given the appreciable variation among loci in nominal F_{ST} values caused by the randomness inherent in the coalescent process, it is not meaningful to test individual F_{ST} values for differences either from zero or from one another (Rosenberg & Nordborg 2002). Instead, a formal outlier analysis should be used to identify markers ('F_{ST} outliers') that are significantly more (or less) divergent than expected given the noise from the coalescent process and the average genome-wide level of divergence (Holsinger & Weir 2009). For this reason, we analyse F_{ST} as a discrete variable (outlier or not), rather than attempting to use differences in nominal F_{ST} values to classify markers as suggested by Nosil *et al.* (2009).

For the genome scan, we genotyped the collection of 200 aphids from wild populations made in 2001 for each polymorphic microsatellite and added genotypes of the same clones for the 40 AFLPs used in Via & West (2008) and 22 AFLP that were not analysed previously. We did a joint outlier analysis of these markers in FDIST2 (Beaumont & Nichols 1996), which

© 2012 Blackwell Publishing Ltd

employs coalescent simulations to obtain an empirical distribution against which to test individual F_{ST} values. We used an iterative approach in which each round of 30 000 simulations involved testing markers relative to the computed average F_{ST}, then removing markers significant at $P < 0.01$ and recalculating the average F_{ST} for the next round. This approach is conservative relative to using the median F_{ST} as a baseline, because the mean F_{ST} values in the first two rounds (0.153, 0.112) were both higher than the median F_{ST} (0.099), raising the threshold for outlier detection. To further minimize false positives, only markers significant at the $P < 0.01$ level after two rounds were classified as outliers. After calculating the distance from each mapped marker to the nearest QTL at the highest LOD score, we performed a logistic regression of outlier status on marker distance from the nearest QTL to test whether outliers cluster around QTL. For significance testing, we randomized the set of observed F_{ST} values over the observed marker distances 10 000 times using a program written in SAS Stat 9.2 (SAS Institute, Cary, NC, USA) and used the distribution of regression coefficients for the randomized data to determine a *P*-value for the observed regression coefficient.

Analysis of population structure using markers classified by outlier status and map location

Using the QTL map, we assigned mapped markers to one of three classes based on outlier status and location within or outside proposed DH regions (Table 1). Under divergence hitchhiking, markers in these three classes are predicted to be influenced by different evolutionary forces: Class 1 markers (F_{ST} outliers) are affected by divergent selection either directly or as divergence hitchhikers that diverge after gene exchange is reduced, Class 2 markers are nonoutliers that are protected from gene exchange by their location within DH regions, and Class 3 markers are nonoutliers that are assumed to be subject to ongoing gene flow (Via 2009, 2012). We used these marker classes to formulate a clear test for the existence of divergence hitchhiking. Under MM/SB, each outlier is thought to be maintained individually by selection on a tightly linked gene, so most nonoutlier markers will be sites of ongoing gene flow; no distinction is made between nonoutlier markers in different genomic locations. Under divergence hitchhiking, however, Class 2 nonoutlier markers are thought to be protected from gene exchange and could therefore have begun the long process of attaining genealogical concordance with the outliers. As such, the Class 2 nonoutliers may carry a greater signal of divergence than expected for Class 3 nonoutlier markers, despite the similarity of their mean F_{ST} values in comparison with the average F_{ST} of outlier markers.

To compare the population structure revealed by the different marker classes, we analysed four replicate data sets for each class of mapped markers in STRUCTURE 2.3.3 (Pritchard *et al.* 2000). Each data set consists of genotypes for the 200 aphids collected from wild populations on alfalfa or clover in 2001 at five unlinked markers (> 10 cM apart) to avoid correlations among linked markers in admixed populations. Although linkage can facilitate the detection of population structure in some cases, Falush *et al.* (2003) caution that the linkage model in STRUCTURE should not be used to correct for linkage disequilibrium between tightly linked markers, which affects about 22% of Class 2 and 3 markers (Via, unpublished data). Because many of the Class 3 markers map to dense clusters (Fig. 1), we could identify only five unlinked markers at a time, and so all data sets were limited to five markers (see Table S2, Supporting information for information on markers used in each analysis).

For each of the 12 data sets (3 marker classes × 4 replicates), we performed five replicate STRUCTURE runs (Pritchard *et al.* 2000). We assumed two genetic clusters ($k = 2$) to correspond to the two host races that we already know are genetically differentiated. We used the admixture model with 15 000 replicates in the burnin period and 30 000 after burnin, assuming correlated allele frequencies. Other parameters were set to the defaults. From the five runs on each data set, we used the estimated distribution of ancestries (*Q* values) and the summary statistics from the run with the greatest log-likelihood (LnP[D]). We then used analysis of variance (implemented in PROC GLM, SAS 9.2; SAS Institute) to test for differences among marker classes in F_{ST}, alpha (an estimate of cluster 'distinctness', where $\alpha = 0$ corresponds to no admixture; Pritchard *et al.* 2000) and the per cent of pea aphids with the majority of their genotype correctly assigned to the cluster corresponding to their 'home' host. These tests quantify the relative ability of each marker class to resolve the genetic structure of these ecologically divergent and partially reproductively isolated host races. Although the residuals of these analyses may not be strictly normally distributed, nonparametric Kruskal–Wallis tests produced virtually the same results.

Results

Second-generation linkage and QTL map of traits causing ecologically based reproductive isolation between pea aphids on two hosts

The linkage map (Fig. 1) has a total map length of about 300 cM and consists of the sex chromosome

© 2012 Blackwell Publishing Ltd

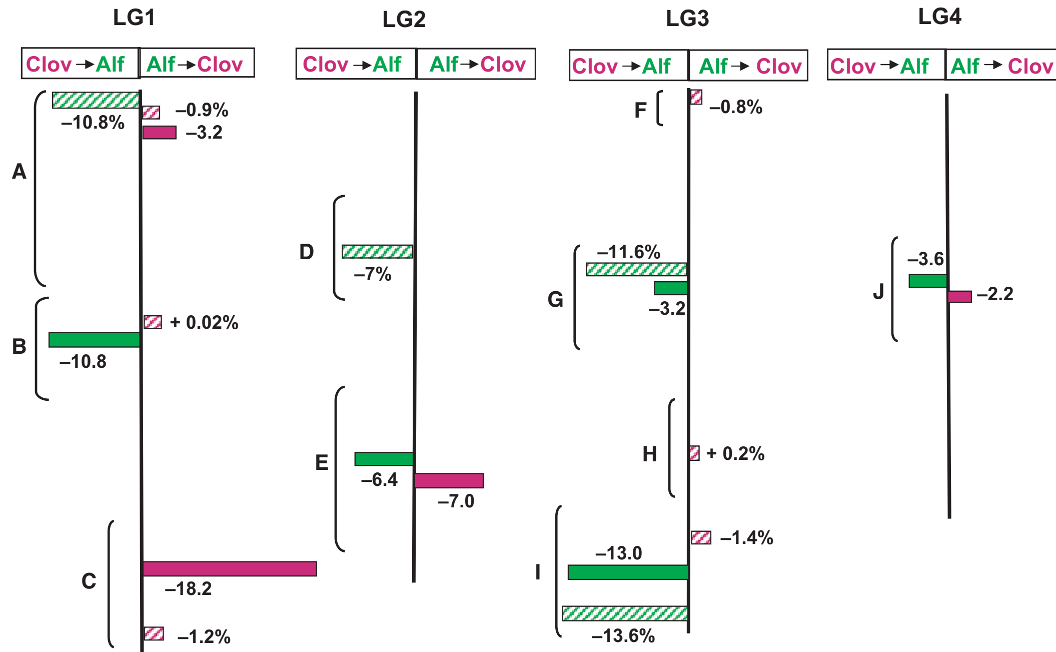

Fig. 2 Estimated contribution of each proposed divergence hitchhiking region to selection against between-host migrants, plotted on a cartoon of each linkage group. Bars extending to the left are the expected phenotypic values for migrants from red clover to alfalfa, while right-facing bars are the expected phenotypic values for migrants from alfalfa to red clover. Negative values mean that migrants have lower fecundity or acceptance than residents. QTL colours for each character state and the interpretation of the lettered brackets are as in Fig. 1.

(LG1) and three autosomes. This map is a clear advance from our previous work in which the dominant AFLP markers required a separate map for each specialized parent (Hawthorne & Via 2001). Early fecundity and host acceptance of aphids on each host were mapped as two separate 'character states' (solid symbols for fecundity on each host and hatched symbols for behavioural acceptance), and character states expressed on the same host are shown in the same colour. For each of the four character states, there are QTL on every linkage group. Although the 2-LOD support intervals are broad, it is clear that QTL are not scattered randomly across the genome: 84% (16/19) of the identified QTL colocalize with at least one other QTL (Fig. 1). Moreover, each QTL is itself likely to be cluster of several genes, as commonly found in fine mapping studies (Frary *et al.* 2000; Turner *et al.* 2008; MacKay *et al.* 2009; Counterman *et al.* 2010; Nadeau *et al.* 2012).

Each group of QTL is associated with a cluster of F_{ST} outliers from the genome scan of the wild populations (blue dots on Fig. 1). Because F_{ST} outliers can only be maintained in genomic regions where gene exchange is reduced, the colocalization of F_{ST} outliers and QTL suggests large underlying regions of reduced gene flow around divergently selected QTL. These regions range from approximately 9–25 cM (noted as A–J on Fig. 1, although clusters A and B and H and I may each be a single large cluster). The logistic regression of outlier status on marker distance from the nearest QTL was highly significant ($P < 0.0019$ from the randomization test). In addition, outlier markers tend to be closer to a QTL than nonoutliers (mean distance 4.84 and 7.87 cM respectively, $P < 0.0025$). Both of these results support our hypothesis that outliers cluster around divergently selected QTL.

The additive phenotypic effect of each QTL is estimated as the expected phenotypic difference between an individual with both QTL alleles from the alfalfa race (i.e. 'alfalfa alleles') and individuals with both alleles at that QTL from the red clover race ('clover alleles'). Regardless of host, all QTL for fecundity increase the early fecundity in residents relative to migrants from the other host (Figs 1 and 2). Similarly, seven of nine QTL for host acceptance increase the affinity of residents for their home plant while causing migrants from the other host to be more likely to reject it. To show how different genomic regions contribute to selection against migrants, we plotted the expected phenotypic values of migrants (i.e. individuals with two specialized QTL alleles from the alternate host) on a cartoon of the map (Fig. 2). On this plot, a negative value means that migrants have lower expected fecun-

© 2012 Blackwell Publishing Ltd

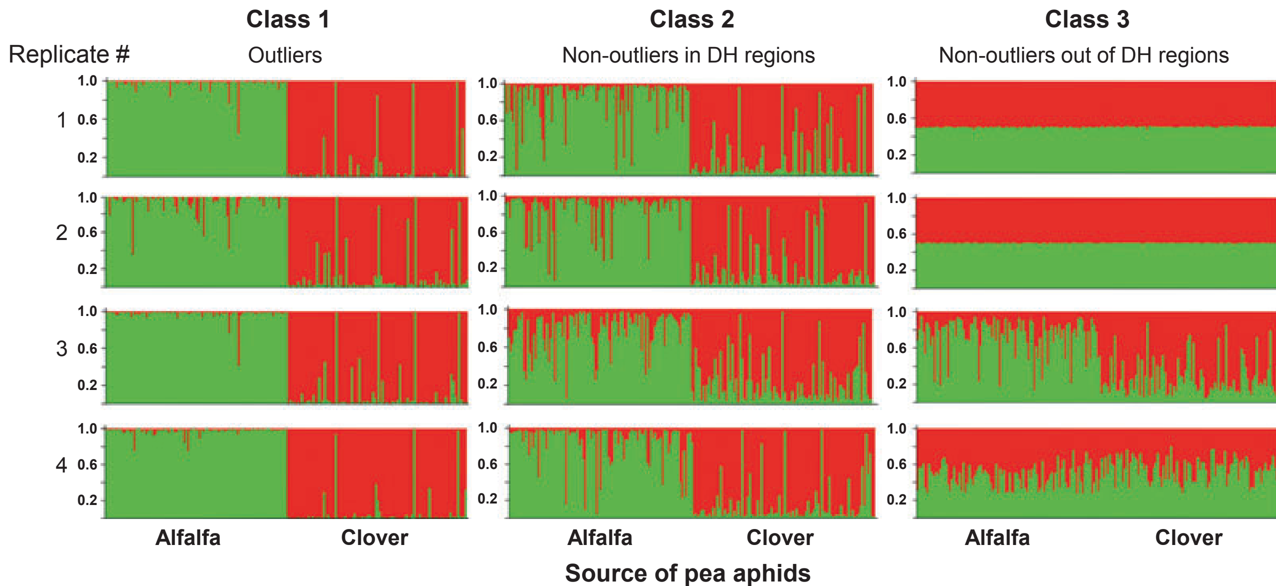

Fig. 3 Replicated STRUCTURE analyses using three classes of markers defined by outlier status and map location (Table 1). Each column shows the results of analyses of four replicate data sets that each consists of five unlinked markers chosen from the QTL map (Fig. 1). Within the column for each marker class, the replicate analyses are shown in no particular order. Individual bars within each analysis are the proportional contribution of each of the two genomic clusters to the genotypes of 100 pea aphids collected from alfalfa (on left of each panel) and 100 genotypes collected from red clover (on right). Additional information about the data sets and the individual analyses in Table S2 (Supporting information).

dity or host acceptance than the residents due to locally adapted QTL alleles at that genomic location.

The mapped clusters of QTL affect various combinations of the four character states: some groups of QTL have antagonistic effects on fecundity or host acceptance in the two environments (clusters A, E, I and J), while others affect traits on only one of the two hosts and add variance to those character states without affecting the cross-environment covariance (clusters B, D, G for alfalfa; cluster C for clover). For example, an individual with both alfalfa alleles at the QTL in cluster C (Figs 1 and 2) that migrates to red clover would be expected to accept the host just slightly less than the clover specialists (−1.2%), but would have on average 18.2 fewer offspring than an individual with both clover alleles from QTL in cluster C. Because there are no QTL expressed in alfalfa in cluster C, this genomic region is expected to have no fitness effects on migrants from clover to alfalfa. In contrast, QTL in cluster E affect fecundity antagonistically on the two hosts: a migrant to clover with two alfalfa alleles at the QTL in cluster E could expect to have seven fewer progeny than a clover specialist, while migrants to alfalfa with two clover alleles at the QTL in cluster E can expect 6.4 fewer offspring than the specialized alfalfa residents. Thus, the negative genetic correlations across environments seen among the F_2 (Hawthorne & Via 2001) are composites of antagonistic effects from QTL within the same DH region and uncorrelated variation from DH regions with QTL that affect the phenotype in only one environment.

The summed QTL effects on each character state (from Figs 1 and 2) reveal that individuals with both alleles from the alfalfa specialist at all mapped QTL are expected to accept alfalfa 43% more and to have 37 more offspring on alfalfa than will individuals that have inherited all QTL alleles from the clover specialist. In contrast, F_2 with all QTL alleles from the specialized clover parent are expected to have 30.6 more offspring on red clover than are migrants bearing alfalfa alleles, although residents are expected to accept clover as a host plant only 4.5% more than will the migrants from alfalfa. These summed QTL effects explain approximately half of the phenotypic difference for each character state observed between the parents in reciprocal transplants (Via *et al.* 2000).

We identified four unusually dense clusters of microsatellites (noted as blue lines on Fig. 1), three of which harbour no F_{ST} outliers or QTL. To test whether these might be regions of reduced recombination, we used the reference pea aphid genome sequence to identify the scaffolds containing our mapped microsatellites (NCBI, BioProject Accession #PRJNA13657, International Aphid Genomics Consortium 2010). We then used the scaffold lengths to calculate the estimated Mb/cM for 1 cM regions either within or outside one

© 2012 Blackwell Publishing Ltd

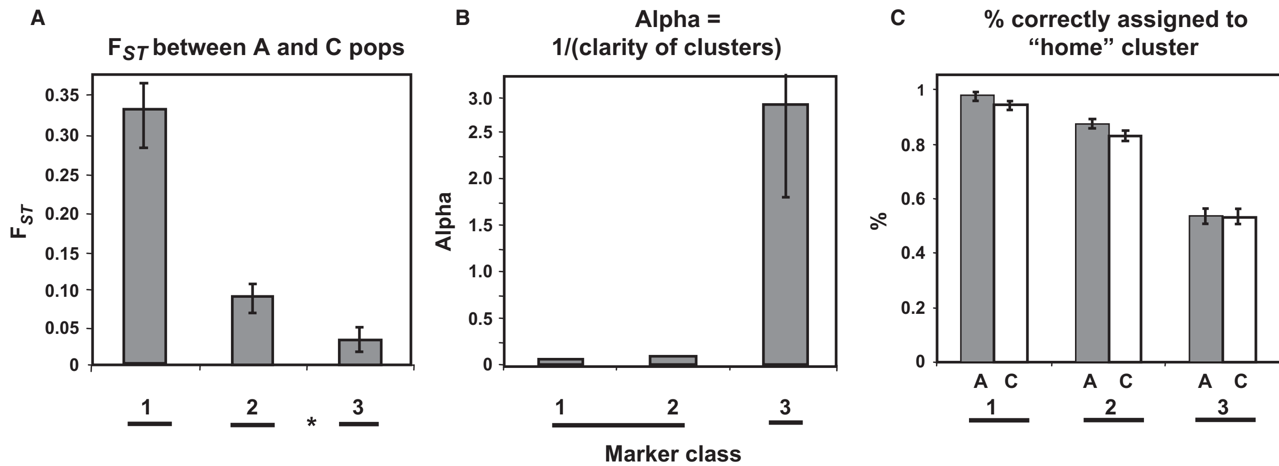

Fig. 4 Comparison of summary statistics from STRUCTURE for the three marker classes. Bars beneath the class numbers connect marker classes that are not significantly different at $P < 0.01$. (a) Differences between marker classes in mean F_{ST} [*the average F_{ST} for Classes 2 and 3 was significantly different only in the entire set of mapped markers ($P < 0.006$), not in this subset of markers]. (b) Differences between marker classes in alpha (an estimate of cluster clarity, where alpha = 0 represents no admixture). (c) Differences between marker classes in the percentage of each individual's genotype correctly classified to the genetic cluster associated with the crop from which it was collected.

of the clumps. Results show means of 4.69 Mb/cM within clumps and 0.85 Mb/cM outside the clumps, $P < 0.0008$ (Table S3, Supporting information). These preliminary results suggest that the dense clusters of microsatellites in Fig. 1 correspond to regions of reduced recombination. They also imply that the proposed DH regions on LG1–LG3 are farther apart on the genome than they appear on the linkage map and that the average Class 3 marker is farther from the nearest QTL than the map data suggest.

The population structure of pea aphid host races in the wild as revealed by different classes of markers

The replicated STRUCTURE analyses show clear differences among marker classes in resolution of the known population structure between the specialized host races (Fig. 3). As expected (Rosenberg 2002; Via & West 2008), the outliers (Class 1 markers) provide the best resolution of adaptive genetic differentiation between the two host races (Fig. 3, left column) and have a higher mean F_{ST} (Fig. 4a) than the other two marker classes. In addition, the analyses using outliers have a lower mean alpha value (where $\alpha = 0$ indicates no admixture; Fig. 4b) and more accurately assigned each genotype to the genetic cluster associated with its collection host (mean 93.7%, Fig. 4c) than did analyses using Class 3 markers.

As expected under divergence hitchhiking, Class 2 and Class 3 markers differ in the resolution of population structure much more than expected based on their mean F_{ST} values (Fig. 4a). Class 2 markers reveal the population structure of the specialized host races almost as clearly as the outliers (Fig. 3, compare middle and left columns), despite a significantly lower mean F_{ST} (Fig. 4a). Correspondingly, the alpha values of analyses using outliers and Class 2 markers were also the same (Fig. 4b). In contrast, the Class 3 markers reveal little population structure (Fig. 3 right column), and their mean alpha was significantly higher than seen in the Class 2 analyses (Fig. 4b). Finally, Class 2 markers classified individuals to the genetic cluster representing their home host plant almost as well as the outliers (82.5%), while the mean accuracy of genotypic classification by the Class 3 markers was barely better than random (57%, Fig. 4c). The higher resolution of population structure by the Class 2 nonoutlier markers supports our hypothesis that they are protected from gene exchange relative to the Class 3 nonoutliers and provides strong evidence for the existence of divergence hitchhiking in this system.

Discussion

The pea aphid host races on alfalfa and red clover in the eastern United States maintain pronounced ecological divergence in sympatry and are considered a model system for the study of incipient ecological speciation-with-gene-flow (Peccoud & Simon 2010). We explored the genomic patterns of divergence and isolation between these incipient species by coupling a QTL map of fecundity and host plant acceptance with a genome scan of F_{ST} in the same wild populations from which the mapping cross originated. F_{ST} outliers are

© 2012 Blackwell Publishing Ltd

markers that are significantly more divergent than the genome-wide average, given the noise inherent in the coalescent process (Beaumont & Nichols 1996). By plotting the genomic locations of these F_{ST} outliers on the QTL map, we could associate molecular signatures of divergent selection in the wild with divergently selected QTL for traits known to produce ecologically based reproductive isolation. Results show colocalizing clusters of F_{ST} outliers and QTL affecting fecundity and host choice on all four linkage groups (Figs 1 and 2). After putting mapped markers into three classes based on outlier status and genomic location (Table 1), we compared the ability of each marker class to resolve the genetic structure of these divergent populations. Results of these analyses support a key prediction from divergence hitchhiking that nonoutlier markers within DH regions experience less gene exchange than do nonoutliers outside such regions (Figs 3 and 4). These results provide the clearest evidence to date for the existence of divergence hitchhiking during speciation-with-gene-flow.

Integrating a genome scan with a QTL map to determine the genomic distribution of divergence

Despite the blazing rate of advance in molecular genetics, a significant gulf remains between population genomic analyses of selection and the study of adaptive phenotypic divergence. Bottom-up population genomic analyses can readily identify markers bearing the signature of divergent selection, but the relationship between such markers and adaptive phenotypic divergence is difficult to determine, particularly for quantitative traits (Dalziel *et al.* 2009; Nielsen 2009; Houle *et al.* 2010; Pritchard & DiRienzo 2010; Storz & Wheat 2010; Barrett & Hoekstra 2011; LeCorre & Kremer 2012). On the other hand, standard top-down quantitative genetics provide estimates of phenotypic plasticity and the genetic (co)variances among key traits, but say nothing about the nature or genomic location of these polygenes (MacKay *et al.* 2009). A QTL map bridges this chasm by providing a genomic template on which to associate markers bearing the signature of divergent selection and QTL for divergently selected phenotypic traits that cause reproductive isolation. Using this approach, we identified genomic regions at which divergently selected QTL and F_{ST} outliers are clustered. Because the allele frequency differences that cause outliers can only be maintained in regions of reduced gene exchange, we propose that each cluster of QTL and F_{ST} outliers reflects an underlying region of divergence hitchhiking, across which the realized magnitude of gene exchange between these incipient species is reduced. Because the only information we have about genomic isolation is the presence of an outlier, we arbitrarily used the outermost F_{ST} outlier in a cluster to mark the boundaries of each DH region.

These clusters of outliers and divergent QTL cannot be explained by structural reductions in recombination, because there is no evidence of chromosomal inversions in pea aphids and pea aphid chromosomes are holocentric. Moreover, we found no outliers or QTL in three of four regions that we identified as likely sites of reduced recombination, contradicting the common assumption that divergence is concentrated in regions of reduced recombination (Turner *et al.* 2008; Neafsey *et al.* 2010; Jones *et al.* 2012; Nosil & Feder 2012). Our findings are not unusual: recent genome-wide analyses in other systems provide compelling evidence for regions of genomic divergence that are not associated with centromeres or chromosomal break points (*Anopheles:* Lawniczak *et al.* 2010; Neafsey *et al.* 2010; Weetman *et al.* 2012; sticklebacks: Jones *et al.* 2012). In these examples, sites of genomic divergence appear to be clustered, although adequate statistical tests for clustering have not been performed (Via 2012). We expect many additional examples of genomic divergence outside such regions to be discovered as more fine-scale genome scans become available. To our knowledge, divergence hitchhiking provides the only clear mechanism for the production and maintenance of clusters of genomic divergence outside regions of structurally reduced recombination.

Testing the existence of divergence hitchhiking using nonoutlier markers in different genomic locations

Under divergence hitchhiking, outliers are predicted to cluster in genomic regions where strong divergent selection on one or more QTL has produced a barrier to gene flow by reducing the realized rate of between-population recombination. Within these regions, some outliers may be linked to divergently selected genes, but other loosely linked markers might diverge by lineage sorting or within-population selective sweeps in the increasingly independent incipient species. Given protection from gene exchange within a DH region, genes/markers that diverge for any of these reasons will be maintained as F_{ST} outliers. Thus, under divergence hitchhiking, outliers will not necessarily tag divergently selected genes. This contrasts with the typical expectation that each outlier is tightly linked to a target gene where divergence is maintained by an independent alteration in the migration/selection balance (Wood *et al.* 2008; Stapley *et al.* 2010). When extended across the entire genome, this ‘multilocus migration-selection balance’ [MM-SB] predicts many small regions of genomic divergence punctuating a genome that is otherwise exposed to gene flow. In this view, the unit of genomic isolation is thought to be as small as one gene (Ting

© 2012 Blackwell Publishing Ltd

et al. 2000; Wu 2001; Yatabe *et al.* 2007), and just two classes of loci are expected during speciation (Wu 2001; Beaumont 2005): those where gene flow is reduced by divergent selection, and those that are uninvolved in speciation and experience ongoing gene flow. Although Nosil *et al.* (2009) propose three marker classes based on nominal F_{ST} values, this classification is inappropriate due to the unreliability of individual F_{ST} estimates (Holsinger & Weir 2009).

The three classes of markers identified under divergence hitchhiking (Table 1) include two classes of nonoutliers: those within DH regions that are predicted to be at least partially protected from between-race recombination and gene exchange (Class 2) and those outside DH regions that are expected to be subject to ongoing gene flow (Class 3). This clear prediction can be tested with empirical data: if divergence hitchhiking explains observed clusters of divergence, Class 2 nonoutlier markers within clusters should show a signature of reduced gene exchange relative to Class 3 nonoutliers. In contrast, if patterns of gene flow around F_{ST} outliers are better explained by MM/SB, Class 2 and Class 3 markers should yield equivalent results in population genetic analyses.

We performed this test using mapped microsatellite markers classified by outlier status and genomic location (Table 1; Table S2, Supporting information). The goal of the STRUCTURE analysis was to compare the relative abilities of nonoutlier markers in Class 2 and Class 3 to resolve the genetic structure already identified both by reciprocal transplants of the two host races (Via 1991; Via *et al.* 2000) and by the STRUCTURE analyses using the outliers (Fig. 3, left column). The results are clear and could not have been predicted by the relatively low F_{ST} values of both nonoutlier classes. Although the mean F_{ST} of Class 2 markers is much lower than that of the Class 1 outliers ($F_{ST_Class1} = 0.32$, $F_{ST_Class2} = 0.096$, $P < 0.01$, Fig. 4a), the Class 2 markers resolve the adaptive population structure of the host races almost as clearly as do the F_{ST} outliers (Fig. 3, centre and left columns, Fig. 4b, c). In contrast, the population structure of the divergent races was poorly resolved by Class 3 markers (Fig. 3, right column): both the fraction of an individual's genotype assigned to each cluster in the Class 3 analyses (Fig. 3, right column) and the high mean value of alpha (Fig. 4b) suggest that Class 3 markers are affected by gene flow. Although the mean F_{ST} of Class 2 and Class 3 markers differ only slightly for the subset of markers used in these analyses (Fig. 4a), when the entire set of mapped markers is considered, the small difference in F_{ST} between Class 2 and Class 3 markers is highly significant (mean $F_{ST_Class2} = 0.098$, mean $F_{ST_Class3} = 0.063$, $P < 0.006$).

These results suggest that divergence hitchhiking has provided the Class 2 nonoutliers with enough protection from gene flow for genetic drift and independent selective sweeps to initiate movement towards genealogical concordance with the branching pattern first defined by the outliers (Rosenberg 2002; Via 2009). Our results clearly contradict suggestions that the gene is the unit of isolation (Yatabe *et al.* 2007), and that the boundary of a region protected from gene flow under migration/selection balance should be drawn at the nearest nonoutlier marker (Ting *et al.* 2000; Wood *et al.* 2008).

This empirical test for the existence of divergence hitchhiking is best applied when outliers can be mapped relative to key QTL, but data from whole-genome scans could be used if sites of genomic divergence appear to be clustered when plotted on a genome assembly (Hohenlohe *et al.* 2010; Jones *et al.* 2012). With such data, nonoutlier markers within clusters could be assigned to Class 2, while nonoutliers in other genomic locations would be Class 3 markers. Such empirical tests using data from wild populations are a far more reliable way to evaluate the probability of divergence hitchhiking than the application of simplistic genetic models (Feder & Nosil 2010), which depend crucially on the values of key demographic parameters. Because almost nothing is known about the demographic conditions that typify speciation-with-gene-flow in wild populations, these models cannot presently reveal the potential for divergence hitchhiking to occur in the wild.

Divergence hitchhiking, genotype x environment interaction, and the genetic correlations that cause ecological specialization and assortative mating

In quantitative traits, differential gene expression across environments can cause a given genotype to exhibit different phenotypes in different environments (phenotypic plasticity); genetic variation in this plastic response is traditionally estimated in animal and plant breeding as genotype x environment interaction (GEI) (MacKay *et al.* 2009). Although the evolutionary effects of GEI cannot be easily quantified (Via & Lande 1985), most QTL analyses of evolution in multiple environments continue to estimate it (Edwards & Weinig 2011; Pelgas *et al.* 2011; Anderson *et al.* 2012). To understand phenotypic evolution in variable environments, it is far more useful to consider a character expressed in two environments as a pair of genetically correlated 'character states' (Yamada 1962). Any genetic correlation between these character states less than + 1 corresponds to significant GEI (Via 1987). Using the character state approach in QTL mapping, the potential for the evolution of assortative mating can be readily interpreted in the context of direct and correlated responses to selection on environment-specific character states. Similarly,

© 2012 Blackwell Publishing Ltd

evolutionary constraints (antagonistic pleiotropy or tight linkage of genes with antagonistic effects across environments) can be easily identified when QTL for two environmentally associated character states colocalize and have opposite directionality.

Compared to the attention lavished on linkage disequilibrium in population genetic models of speciation (Felsenstein 1981), the effects of genetic correlations among quantitative traits on the evolution of assortative mating have barely been considered in the speciation literature. Nevertheless, these effects are clear (Hawthorne & Via 2001): positive genetic correlations within environments between traits involved in local adaptation and mate choice accelerate the evolution of assortative mating, while negative cross-environment genetic correlations speed population divergence and increase ecologically based reproductive isolation. Divergence hitchhiking allows loosely linked QTL to contribute to genetic correlations as if they were much more tightly linked than their relative map distances would suggest. This foils the destructive effects of recombination and facilitates the evolution of assortative mating, mitigating the major criticism of speciation-with-gene-flow (Felsenstein 1981; Smadja *et al.* 2008).

By separately mapping each environmentally dependent character state, it becomes apparent that the facilitating pattern of genetic correlations seen in pea aphid populations (Via 1999; Hawthorne & Via 2001) is caused by mixture of independent variation from singleton QTL and covariance from pleiotropic, tightly linked and loosely linked QTL within the same DH regions (Figs 1 and 2). We identified 10 genomic regions in which divergently selected QTL colocalize (Fig. 1) and three regions that harbour only a single QTL. In all multi-QTL groups except cluster B (Fig. 1), the directionality of individual QTL effects causes positive genetic correlations between fecundity and host acceptance on the same host (clusters A, C, G, I), and/or negative genetic correlations between character states on different hosts (clusters A, E, J). In half of these clusters, overlap between the map locations of QTL is large enough to suggest that they represent either single genes with pleiotropic effects or tightly linked genes affecting different character states (clusters A, B, G, I, J). In clusters C, E and I, however, the 2-LOD support intervals for some or all of the QTL do not overlap, suggesting that these QTL are too far apart to evolve coordinately given nominal rates of recombination based on map distance. However, because between-race recombination is likely to be just a fraction of the rate expected from map distances, these loosely linked QTL within a DH region contribute to genetic correlations among divergently selected traits just as they would if tightly linked physically.

Divergently selected QTL that are loosely linked may often contribute to genetic correlations. In a recent QTL analysis of domesticated chicken, genetic correlations among domestication traits were described as including covariance attributable to QTL as far away as 15 cM from a core of pleiotropic and tightly linked QTL (Wright *et al.* 2010). These authors suggest no mechanism for how such loosely linked QTL could contribute to genetic correlations without disruption by recombination. We suggest that ancient domestication could have been very similar to speciation-with-gene-flow; strong selection for domestication was probably accompanied by gene flow with the wild ancestor. If so, then regions of divergence hitchhiking would be expected around domestication genes, and the loosely linked clusters of QTL expected in these regions are likely to persist in domesticated animals and plants. This may be a general pattern, given that clustered QTL have also been seen in many recent studies of wild populations under divergent selection (Heliconius: Kronforst *et al.* 2006; lake whitefish: Rogers & Bernatchez 2007; sticklebacks: Albert *et al.* 2008; white spruce: Pelgas *et al.*, 2011).

Genomic isolation versus genomic divergence during speciation-with-gene-flow

The absence of a geographical barrier to gene exchange causes a heterogeneous pattern of genomic divergence during speciation-with-gene-flow. In this study, we have addressed whether the genomic pattern of isolation exactly mirrors this mosaic pattern of divergence, as expected under the view we call MM/SB (Via 2009; Fig. 1a). Our results do not support this view. Instead, we present evidence that large genomic regions around divergently selected genes/QTL are protected from gene exchange between locally adapted populations. These regions of divergence hitchhiking, which may span many megabases, will be manifest in genome scans as clusters of divergent sites and visible on QTL maps as groups of colocalizing outliers and divergently selected QTL. Although such clusters of divergence could be associated with inversions or centromeric regions, reduced recombination is not required; divergence hitchhiking can occur anywhere in the genome.

Until recently, the automatic reduction in between-population recombination that occurs due to the disruption of random mating between locally adapted populations (Charlesworth *et al.* 1997) has been ignored in analyses of speciation, although it is well known in the hybrid zone literature (Barton & Bengtsson 1986). The mechanisms that produce divergence hitchhiking are surprisingly simple: as local adaptation evolves, the frequency of interbreeding between sympatric

© 2012 Blackwell Publishing Ltd

populations declines due to selection against increasingly specialized migrants and their F_1 hybrids (Via *et al.* 2000), as well as to habitat choice (Via 1999) or mate choice if present. When fewer F_1 than expected under random mating are produced, the realized frequency of between-population recombination drops accordingly. In this way, local adaptation monotonically reduces the *opportunity* for between-race recombination across the entire genome and reduces effective migration rates. At any given time, the magnitude of this genome-wide effect [called 'genome hitchhiking' by Feder *et al.* (2012)] is a function of allelic divergence at all QTL that affect local adaptation (Via 2012).

Overlaid on this genome-wide reduction in realized between-population recombination, divergent selection against QTL recombinants with the 'wrong' locally adapted QTL alleles reduces the *effectiveness* of recombination and introgression at that genomic region. This occurs because alleles at neutral loci that flank a recombinant QTL allele (now deleterious in the alternate genetic background) can be eliminated by selection before they can 'escape' by recombining away from the QTL into the local genetic background. This effect is disproportionately increased if several loosely linked genes are under divergent selection (Charlesworth *et al.* 1997). The persistent reduction in the ability of flanking regions to recombine away from a divergently selected gene explains why DH regions remain much larger than traditional hitchhiking regions after a selective sweep within populations.

It is unfortunate that genomic divergence and genomic isolation are sometimes equated (Feder & Nosil 2010). Genomic isolation cannot be expected to immediately result in genomic divergence because it takes an extremely long time after gene flow is extinguished for loci not under divergent selection to come into concordance with the branches defined by the outliers, which will eventually be recognized as the species tree. The time to complete genealogical concordance is likely to be measured in millions of generations (Avise 2000; Wu 2001); it may never be reached in species that continue to hybridize (Sambatti *et al.* 2012). In the recently diverged taxa most useful for genetic analysis of speciation-with-gene-flow, we expect that most genomic divergence will still be concentrated in regions of divergence hitchhiking because these regions harbour the genes under divergent selection. The highly heterogeneous and apparently clustered divergence in recent genome scans (Begun *et al.* 2007; Turner *et al.* 2008; Counterman *et al.* 2010; Lawniczak *et al.* 2010; Neafsey *et al.* 2010; Hohenlohe *et al.* 2010; Jones *et al.* 2012) support this hypothesis. In sum, there is little empirical support for Feder & Nosil's (2010, p. 1742) prediction that the entire genome will rapidly diverge once multiple loci are involved in speciation-with-gene-flow.

To our knowledge, divergence hitchhiking is the first mechanism to be proposed that provides both the genomic isolation required for speciation-with-gene-flow and a predictable pattern of divergence for genes and markers in different genomic regions. We eagerly await data from coupled genome scans and QTL maps in additional recently diverged systems, and we encourage other researchers to test for the existence of divergence hitchhiking by comparing the signatures of gene flow at the three marker classes we define.

Acknowledgements

Dave Hawthorne provided the AFLP markers and Amy Bouck; Marina Caillaud, Stasia Skillman Greenwalt, Christy Olson and Carrie Wu provided expert assistance with the crosses and F_2 phenotyping. Further thanks to Amy Bouck for the list of EST-based microsatellites and to Marina Caillaud for non-EST microsatellites. Support was provided by NSF DEB0528288 and USDA-NIFA Competitive Grant 2006-35607-16707 to SV.

References

Albert SYK, Sawaya S, Vines TH *et al.* (2008) The genetics of adaptive shape shift in sticklebacks: pleiotropy and effect size. *Evolution*, **62**, 76–85.

Anderson JT, Lee C-R, Rushworth CA, Colautti RI, Mitchell-Olds T (2012) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology*, doi: 10.1111/j.1365-294X.2012.05522.x.

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

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, **12**, 767–780.

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

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

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

Begun DJ, Holloway AK, Stevens K *et al.* (2007) Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology*, **5**, e310.

Boutin-Ganache I, Raposo M, Raymond R, Deschepper CF (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *BioTechniques*, **31**, 25–28.

Caillaud CMC, Mondor-Genson G, Levine-Wilkinson S *et al.* (2004) Microsatellite DNA markers for the pea aphid: *Acyrthosiphon pisum* (Harris, 1776) [Homoptera, Sternorrhynca]. *Molecular Ecology Notes*, **4**, 446–448.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and

© 2012 Blackwell Publishing Ltd

background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155–174.

Counterman BA, Araujo-Perex F, Hines HM *et al.* (2010) Genomic hotspots for adaptation: the population genetics of Mullerian mimicry in *Heliconius erato*. *PLoS Genetics*, **6**, e10000796.

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

Dalziel AC, Rogers SM, Schulte PM (2009) Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. *Molecular Ecology*, **18**, 4997–5017.

Eastop VF (1971) Keys for the identification of *Acyrthosiphon* (Hemiptera: Aphididae). *Bulletin of the British Museum (Natural History) Entomology*, **26**, 1–115.

Edwards CE, Weinig C (2011) The quantitative-genetic and QTL architecture of trait integration and modularity in Brassica rapa across simulated seasonal settings. *Heredity*, **106**, 661–677.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

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

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 303–306.

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

Ferrari J, Godfray HCJ, Faulconbridge AS, Prior K, Via S (2006) Population differentiation and genetic variation in host choice among pea aphids from eight host genera. *Evolution*, **60**, 1574–1584.

Ferrari J, Via S, Godfray HCJ (2008) Population differentiation and genetic variation in performance of pea aphids on plants from eight host genera. *Evolution*, **62**, 2508–2523.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86–89.

Frary A, Nesbitt TC, Frary A *et al.* (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, **289**, 85–88.

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

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics*, **10**, 639–650.

Houle D, Govindaraju DR, Omholt S (2010) Phenomics: the next challenge. *Nature Reviews Genetics*, **11**, 855–866.

International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biology*, **8**, e1000313.

Jones FC, Grabherr MG, Chan YF *et al*. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Kronforst MR, Young LG, Kaplan DD, McNeelty C, O″Neill RJ, Gilbert LE (2006) Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 6575–6580.

Lawniczak MKN, Emrich SJ, Holloway AK *et al.* (2010) Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, **330**, 512–514.

LeCorre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**, 1548–1566.

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

MacKay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, **10**, 565–577.

Nadeau , NJ , Whibley A, Jones RT *et al*. (2012) Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Philisophical Transactions of the Royal Society B*, **367**, 343–353.

Neafsey DE, Lawniczak MKN, Park DJ, Redmand SN, Coulibaly MB, Traore SF (2010) SNP genotyping defines complex gene flow boundaries among African malaria vector mosquitoes. *Science*, **330**, 514–517.

Nielsen R (2009) Adaptationism—30 years after Gould and Lewontin. *Evolution*, **63**, 2487–2490.

Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **367**, 332–342.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

van Ooijen JW (2006) JoinMap4: Software for the Calculation of Genetic Linkage Maps in Experimental Populations. Kyazma BV, Wageningen, The Netherlands.

van Ooijen JW (2009) MapQTL6: Software for the Mapping of Quantitative Trait Loci in Experimental Populations of Diploid Species. Kyazma BV, Wageningen, The Netherlands.

Peccoud J, Simon J-C (2010) The pea aphid complex as a model of ecological speciation. *Ecological Entomology*, **35**, 119–130.

Pelgas B, Bousquet J, Meirmans PG, Ritland K, Isabel N (2011) QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. *BMC Genomics*, **12**, 145.

Pinho C, Hey J (2010) Divergence with gene flow: models and data. *Annual Review of Ecology and Systematics*, **41**, 215–230.

Pritchard JK, DiRienzo A (2010) Adaptation—not by sweeps alone. *Nature Reviews Genetics*, **11**, 665–667.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–949.

Rogers SM, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural populations of lake whitefish (Coregonus sp. Salmonidae). *Molecular Biology and Evolution*, **24**, 1423–1438.

Rosenberg NA (2002) The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology*, **61**, 225–257.

Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, **3**, 380–390.

© 2012 Blackwell Publishing Ltd

Sabatir-Munoz B, Legeai F, Rispe C *et al*. (2006) Large-scale gene discovery in the pea aphid *Acyrthosiphon pisum* (Hemiptera). *Genome Biology*, **7**, R21.

Sambatti JBM, Strasburg JL, Ortiz-Barrientos D, Baack EJ, Rieseberg LH (2012) Reconciling extremely strong barriers with hich levels of gene exchange in annual sunflowers. *Evolution*, **66**, 1459–1473.

Sapir Y, Moody ML, Brouillette LC, Donovan LA, Rieseberg LH (2007) Patterns of genetic diversity and candidate genes for ecological divergence in a homoploid hybrid sunflower, *Heliantuhus anomalus*. *Molecular Ecology*, **16**, 5017–5029.

Smadja C, Galindo J, Butlin R (2008) Hitching a lift on the road to speciation. *Molecular Ecology*, **17**, 4177–4180.

Stapley J, Reger J, Geulner PGD *et al*. (2010) Adaptation genomics: the next generation. *Trends in Ecology and Evolution*, **25**, 705–712.

Storz JF, Wheat CW (2010) Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution*, **64**, 2489–2509.

Strasburg JL, Sherman NA, Wright KM, Moyle LC, Willis JH, Rieseberg LH (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **367**, 364–373.

Ting CT, Tsaur SC, 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 of the United States of America*, **97**, 5313–5316.

Turner TL, Levine MT, Eckert ML, Begun DJ (2008) Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics*, **179**, 455–473.

Via S (1987) Genetic constraints on the evolution of phenotypic plasticity. In: *Evolutionary Constraints in Ecology* (ed. Loeschke V), pp. 4. Springer-Verlag, New York.

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 and Evolution*, **16**, 381–390.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9939–9946.

Via S (2012) Divergence hitchhiking and the spread of genomic isolation during speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **367**, 451–460.

Via S, Lande R (1985) Genotype–environment interaction and the evolution of phenotypic plasticity. *Evolution*, **39**, 505–522.

Via S, West JA (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

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.

Weetman D, Wilding CS, Steen K, Pinto J, Donnelly MJ (2012) Gene flow-depoendent genomic divergence between *Anopheles gambiae* M and S forms. *Molecular Biology and Evolution*, **29**, 279–291.

Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123–3135.

Wright D, Rubin C-J, Martinez Barrio A *et al*. (2010) The genetic architecture of domestication in the chicken: effects of pleiotropy and linkage. *Molecular Ecology*, **19**, 5140–5156.

Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Yamada Y (1962) Genotype x environment interaction and the genetic correlation of the same trait under different environments. *Japanese Journal of Genetics*, **37**, 498–509.

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.V. and G.C. are interested in the genetics of ecological speciation-with-gene-flow. C.M.-F. and K.M. were Honors undergraduates in the S.V. lab and now teach middle-school and high-school biology, respectively.

Data accessibility

DNA sequences of scaffolds used in Table S3 (Supporting information): NCBI BioProject Accession PRJNA13657.

Joinmap genotype file, Joinmap traits file, Marker IDs, FDIST2 output file, STRUCTURE input files: DRYAD entry doi:10.5061/dryad.9cf75.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Table S1 Details for mapped markers.

Table S2 Markers used in STRUCTURE analyses, with parameter estimates for each replicate.

Table S3 Map intervals used to calculate Mb/cM.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2012 Blackwell Publishing Ltd