doi:10.1111/j.1558-5646.2008.00352.x

SELECTION AND GENOMIC DIFFERENTIATION DURING ECOLOGICAL SPECIATION: ISOLATING THE CONTRIBUTIONS OF HOST ASSOCIATION VIA A COMPARATIVE GENOME SCAN OF *NEOCHLAMISUS BEBBIANAE* LEAF BEETLES

Scott P. Egan,[1,2] Patrik Nosil,[3,4] and Daniel J. Funk[1,5]

[1]*Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235*

[2]*E-mail: scott.p.egan@vanderbilt.edu*

[3]*Zoology Department and Centre for Biodiversity Research, University of British Columbia, Vancouver, B.C. V6T 1Z4, Canada*

[4]*E-mail: pnosil@zoology.ubc.ca*

[5]*E-mail: daniel.j.funk@vanderbilt.edu*

Received September 17, 2007
Accepted January 23, 2008

This study uses a comparative genome scan to evaluate the contributions of host plant related divergent selection to genetic differentiation and ecological speciation in maple- and willow-associated populations of *Neochlamisus bebbianae* leaf beetles. For each of 15 pairwise population comparisons, we identified "outlier loci" whose strong differentiation putatively reflects divergent selection. Of 447 AFLP loci, 15% were outliers across multiple population comparisons, and low linkage disequilibrium indicated that these outliers derived from multiple regions of the genome. Outliers were further classified as "host-specific" if repeatedly observed in "different-host" population comparisons but never in "same-host" comparisons. Outliers exhibiting the opposite pattern were analogously classified as "host-independent." Host-specific outliers represented 5% of all loci and were more frequent than host-independent outliers, thus revealing a large role for host-adaptation in population genomic differentiation. Evidence that host-related selection can promote divergence despite gene flow was provided by population trees. These were structured by host-association when datasets included host-specific outliers, but not when based on neutral loci, which united sympatric populations. Lastly, three host-specific outliers were highly differentiated in all nine different-host comparisons. Because host-adaptation promotes reproductive isolation in these beetles, these loci provide promising candidate gene regions for future molecular studies of ecological speciation.

KEY WORDS: **AFLPs, divergent natural selection, herbivorous insects, population divergence, reproductive isolation, speciation genetics.**

Ever since Darwin (1859), the study of natural selection has been a primary focus of evolutionary biology (Fisher 1930; Mayr 1942; Endler 1986; Kingsolver et al. 2001). Investigations of selection include attempts to demonstrate its occurrence, identify the various roles it plays, and evaluate its relative importance as a cause of biological differentiation. Two areas of current interest are its capacity to explain molecular genetic variation and to cause speciation.

Debates about the contributions of selection to the fate of genetic variation heated up with the formulation of the neutral theory (Kimura 1968; King and Jukes 1969). Although abundant evidence now supports a role for genetic drift in explaining much molecular genetic polymorphism and divergence (Kimura 1986; Ohta 1992, 2002), recent technical and analytical advances have increasingly helped identify how selection also shapes genomic evolution. QTL approaches can identify gene regions associated with phenotypes under selection (Hawthorne and Via 2001; Noor et al. 2001; Rieseberg 2001; Feder et al. 2003; Lexer et al. 2004; Burke et al. 2005), whereas candidate gene approaches can characterize selection at a yet finer genetic scale using molecular evolutionary techniques (Bradshaw and Schemske 2003; Lexer et al. 2003; Nachman et al. 2003; Colosimo et al. 2005; Hoekstra et al. 2006; Joron et al. 2006; for review, see Nielsen 2005; Vasemagi and Prinner 2005; Noor and Feder 2006).

Although these techniques allow high resolution investigations of particular loci, a complementary, "genome scan," approach provides a coarser evaluation of each of hundreds of loci distributed throughout the genome, allowing different questions to be asked (Black et al. 2001; Storz 2005; Stinchcombe and Hoekstra 2008). Such data may be used in combination with simulations to identify loci that exhibit greater differentiation than expected under neutrality (Beaumont and Nichols 1996; Beaumont and Balding 2004; Beaumont 2005). The elevated differentiation of these "outlier loci" is interpreted as the consequence of divergent selection, whereas nonoutlier loci are interpreted as neutrally evolving. By identifying which of hundreds of loci are selected versus neutral, genome scans offer the opportunity to evaluate the contributions of selection to genetic differentiation across diverse regions of the genome. The genome scan approach has begun to attract more practitioners, with a modest, yet recently accumulating number of studies demonstrating its utility (Wilding et al. 2001; Emelianov et al. 2003; Campbell and Bernatchez 2004; Scotti-Saintagne et al. 2004; Acheré et al. 2005; Vasemagi et al. 2005; Bonin et al. 2006; Murray and Hare 2006; Savolainen et al. 2006; Yatabe et al. 2007; Nosil et al. 2008).

The other area in which the study of selection has generated recent interest concerns "ecological speciation" (Schluter 2000, 2001; Funk et al. 2002; Rundle and Nosil 2005). Divergent sources of natural selection promote the divergent ecological adaptation of populations to different environments and ecological speciation theory posits that the resulting genetic differentiation at selected loci may incidentally promote the evolution of reproductive isolation as a byproduct. This could occur if the selected loci have pleiotropic effects on reproductive isolation or are linked to other loci that influence reproductive barriers. Although this idea was implicit in the writings of thinkers from the evolutionary synthesis (Muller 1942; Mayr 1947, 1963), only more recently has this hypothesis become an important, empirically tested, aspect of evolutionary study (e.g., Feder et al. 1998; Funk 1998; Via 1999; Via et al. 2000; Rundle et al. 2000; Jiggins et al. 2001; Bradshaw and Schemske 2003; Nosil 2007).

Selection's effects on speciation have been empirically studied via "mechanistic" investigations that focus on experimentally identifying proximate ecological causes of reproductive barriers in particular species (e.g., Wood and Keese 1990; Craig et al. 1993; Feder et al. 1994; Boughman 2001; Jiggins et al. 2001; Rundle 2002; Nosil 2004; Nosil and Crespi 2006) and, less frequently, by comparative approaches that seek to isolate and evaluate the general contributions of selection across varied populations or taxa. The latter studies have, for example, compared levels of reproductive isolation between populations as a function of the degree to which they are subject to divergent selection pressures. One such approach contrasts comparisons of population pairs using different habitats (and thus subject to habitat-specific divergent selection) versus comparisons of population pairs using the same habitat (and thus not differing in habitat-associated selection) (Schluter and Nagel 1995; Funk 1996, 1998; Rundle et al. 2000; Rundle 2002; Nosil et al. 2002; Vines and Schluter 2006; Langerhans et al. 2007). Such studies offer tests of ecological speciation theory, which predicts that, other things being equal, the former sort of population comparisons should tend to exhibit more reproductive isolation than the latter sort, given their greater opportunity for genetically based adaptive divergence (Schluter and Nagel 1995; Funk 1996, 1998). The controlled nature of such comparative approaches and their application across multiple pairs of populations allows rigorous and informative explorations of the evolutionary consistency with which selection contributes to speciation (Funk et al. 2002, 2006; Funk and Nosil 2008). In these respects, comparative studies of ecological speciation are to mechanistic investigations of reproductive barriers as genome scans are to QTL and candidate gene approaches. In both comparisons that comprise this analogy, the first of the two compared approaches provides inferences about the generality of selection's influence. This is accomplished by making repeated comparisons (across suites of populations in the case of ecological speciation and across loci in genome scans) that evaluate deviations from a null expectation of no effect (provided by ecologically similar habitats and neutral loci, respectively). The present study combines the advantages of both population-level and locus-level comparative approaches in an attempt to investigate the roles of divergent selection in shaping genome-wide patterns of genetic differentiation during ecological speciation.

Specifically, this study represents an ecologically comparative genome scan of 447 AFLP loci for 15 pairwise comparisons of populations of the leaf beetle *Neochlamisus bebbianae*. Each population in this study is specifically associated with, and adapted to, one of two tree species (maple versus willow) and these host-associated populations appear to be in the process of

ecological speciation (Funk 1998; Funk et al. 2002; Funk and Nosil 2008) (see Methods for further biological details). Our approach uses contrasts between "different-host comparisons" of populations, each of which is associated with a different host plant, versus "same-host comparisons" of pairs of populations associated with the same host plant. From the locus perspective, we analogously contrast loci observed to be outliers in multiple different-host comparisons but no same-host comparisons versus those observed to be outliers in multiple same-host comparisons but no different-host comparisons. These two kinds of outliers represent host-specific and host-independent sources of selection, respectively. These comparisons are supplemented by analyses of genetic structure and linkage disequilibrium.

The fundamental questions investigated by this study are: (1) To what degree is differentiation in the genomic regions assayed by this study caused by divergent selection? This central issue in the selectionist-neutralist debate has been evaluated by a modest number of genome scans. Yet answering it requires such studies on many taxa. (2) What proportion of this genomic differentiation is attributable to a specific source of selection? Very little is known about the contribution of specific ecological factors to overall evolutionary genetic differentiation. Addressing this requires comparative genomic approaches (see also AFLP genome scans by Wilding et al. 2001 [shore level]; Emelianov et al. 2003 [host plant]; Campbell and Bernatchez 2004 [trophic niche]; Bonin et al. 2006 [altitude]; Nosil et al. 2008 [host plant]). (3) To what degree is overall divergent selection promoted by a particular ecological factor, as opposed to alternative factors? Addressing this represents the unique and perhaps most important contribution of the present study and requires the controlled comparison of host-specific versus host-independent selective differentiation. Here the issue is the degree to which overall divergent selection is specifically host related. The importance of host adaptation in the phenotypic evolution of insect herbivores is well recognized (e.g., Bernays and Chapman 1994; Winkler and Mitter 2008). However, our study is the first to quantify its relative importance in a genomic context. (4) Can specific outlier loci be identified that represent special targets of strongly divergent host-related selection? This represents a quest for candidate "ecological speciation gene regions" for future studies that take advantage of the powerful and complementary molecular approaches described above.

Methods

STUDY ANIMALS, STUDY POPULATIONS, AND NATURAL HISTORY

Neochlamisus bebbianae (Brown) (Coleoptera: Chrysomelidae) is a univoltine, eastern North American leaf beetle that uses specific tree species from six taxonomically disparate genera as host plants, on which all life activities—from oviposition through adult feeding and mating—are conducted (Karren 1972; Funk 1998). Populations associated with different host plants are partially differentiated in host preference and performance traits, and exhibit partial reproductive isolation as a consequence of divergent host adaptation (Funk 1998, 1999; Funk et al. 2002; Egan and Funk 2006; Funk and Nosil 2008). These beetles thus offer informative systems for investigating ecological speciation. The present study evaluates *N. bebbianae* populations associated with red maple (*Acer rubrum*:P Aceraceae) and Bebb's willow (*Salix bebbiana*: Salicaceae). Populations of these "maple beetles" and "willow beetles" are morphologically homogeneous (Karren 1972) yet exhibit host-associated ecological divergence and premating reproductive barriers (Funk 1998; Funk et al. 2002; Egan and Funk 2006) that include behavioral isolation, habitat isolation, and immigrant inviability (cf., Nosil et al. 2005). Recent evidence demonstrates that F1 crosses yield hybrids yet also some cryptic and ecologically dependent postmating isolation (Egan et al., unpubl. ms.). Significant host preference variation exists among maple and willow beetle populations associated with the same host. Mitochondrial DNA variation among our study populations is $< 2\%$ and polyphyletic among and within host-associated populations, suggestive of ongoing or evolutionarily recent gene flow, or incomplete lineage sorting (Funk 1999, unpubl. data). Replicated comparisons of sympatric versus allopatric mtDNA differentiation for different-host comparisons consistently reveal less differentiation in sympatry (Funk et al., unpubl. data), a pattern most parsimoniously explained by gene flow (Grant et al. 2005). In sum, past evidence suggests partial but incomplete evolutionary differentiation among study populations, irrespective of host association.

This study treats beetles from five geographically scattered localities, all 162–529 km apart (Fig. 1, Table 2). Maple beetles and willow beetles were each collected from three localities, one of which (Caledonia Co., Vermont) was the same for both, offering a sympatric comparison. This site represented the region of the northeast—much of upstate New York and New England, and parts of southern Quebec and Ontario—where maple and willow beetle populations commonly co-occur and their host plants intermingle in the same microhabitats (Funk 1999, unpubl. data). The remaining populations were collected from localities just beyond this region of general overlap, at sites in which the alternative host form is rare or absent (D. J. Funk, unpubl. data): Oswego Co., New York; Suffolk Co., New York; Wyoming Co., Pennsylvania; and near East Angus, Quebec. This sampling design was chosen to increase the chance of evaluating adaptively divergent populations, while also including a sympatric different-host pair in which host-related selection and gene flow would interact to a larger degree. Numbers of individuals analyzed were: $M_{LI} = 30$, $M_{PA} = 21$, $M_{VT} = 25$, $W_{OS} = 30$, $W_{QU} = 29$, $W_{VT} = 30$ (See Fig. 1 for population abbreviations used throughout this article). In all cases, insects from each population were collected from

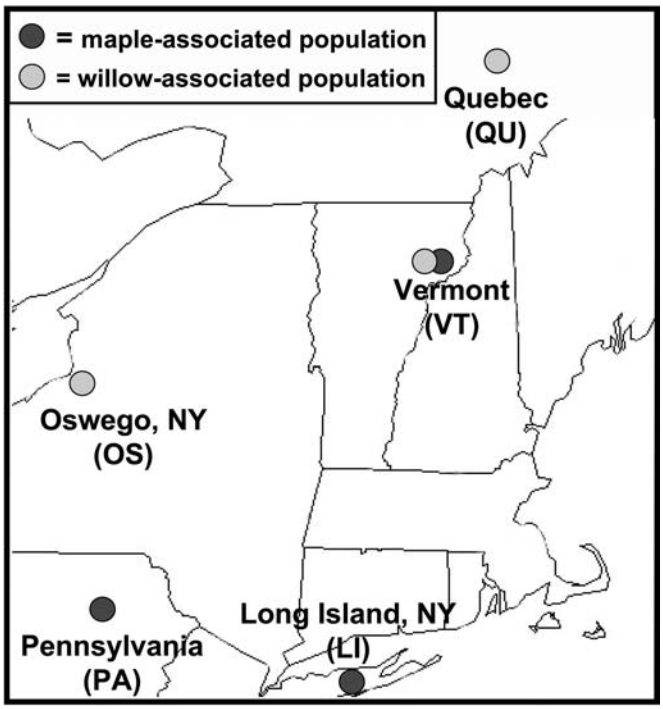

Figure 1. Map of *Neochlamisus bebbianae* leaf beetle study populations. Maple and willow populations from Vermont are sympatric. Throughout the article, beetle population names are based on the first letter of the host plant (M or W) plus a subscripted locality abbreviation (provided in parentheses here). See text for further details.

many different plant individuals. Numbers of populations and individuals sampled are typical of related studies (e.g., Wilding et al. 2001; Campbell and Bernatchez 2004; Bonin et al. 2006; Nosil et al. 2008). These beetles fly between individual host plants in field and laboratory (D. J. Funk, pers. obs.), but no quantitative data are available on their dispersal capacity.

DNA EXTRACTION AND AFLP PROTOCOLS

Whole genomic DNA was extracted from larvae and adults following DNAeasy Animal Extraction Kit (Qiagen, Valencia, CA) protocols. *N. bebbianae* includes some asexual individuals. These were identified and removed prior to AFLP analysis using a restriction enzyme that identifies their diagnostic mitochondrial lineage (Egan et al., unpubl. data). Contamination by host material was unlikely to be problematic given: (1) our use of animal-specific rather than plant-specific DNA extraction and AFLP kits, (2) the likely degradation of DNA in the gut, and (3) the low likelihood that such DNA would outcompete higher-concentration beetle DNA during PCR. Nonetheless, we PCR-screened all beetle extractions with universal plant primers for both cpDNA (primers C and F; Taberlet et al. 1991) and mtDNA (primers nad5/4 and nad5/5.; Dumolin-Lapeague et al. 1997). Among > 300 screened products, only one yielded a band, and this individual was removed from the study. In contrast, maple and willow DNA extracted with plant-specific protocols, and used as positive controls, consistently amplified with both primer pairs. Given our lack of amplification from beetle DNAs using universal primers for high-copy plant genomes, it seems highly unlikely that low-copy nuclear regions were significantly amplified from hosts with our anonymous AFLP primers.

Table 1. Primers used for AFLP analysis, and number of repeatable loci for each primer combination (in parentheses) that were used for analyses.

Primer	(Sequence 5′ – 3′)
Preselective	
Eco+C	GACTGCGTACCAATTCC
Mse+C	GATGAGTCCTGAGTAAC
Selective	
Eco+CTC	GACTGCGTACCAATTCCTC
Eco+CAG	GACTGCGTACCAATTCCAG
Mse+CTG	GATGAGTCCTGAGTAACTG
Mse+CGA	GATGAGTCCTGAGTAACGA
Mse+CAA	GATGAGTCCTGAGTAACAA
Mse+CCT	GATGAGTCCTGAGTAACCT
Primer Combinations	
D (41)	Eco+CTC/Mse+CGA
E (37)	Eco+CTC/Mse+CAA
F (102)	Eco+CTC/Mse+CCT
G (28)	Eco+CTC/Mse+CTG
H (48)	Eco+CAG/Mse+CTG
I (85)	Eco+CAG/Mse+CGA
J (61)	Eco+CAG/Mse+CAA
K (45)	Eco+CAG/Mse+CCT

To generate AFLP data following Vos et al. (1995), we used AFLP Core Reagent Kits (Invitrogen, Carlsbad, CA). After the preselective amplification, eight selective amplification primer combinations were used to generate PCR products that were purified using Sephadex (GE Healthcare, Piscataway, NJ) (Table 1). Products were run on 6% polyacrylamide gels in the Department of Biological Sciences at Vanderbilt University using an MJ Base Station Automated Sequencer (MJ Research, Waltham, MA). Fragment analysis was completed using Cartographer software associated with the Base Station. Loci between 100 and 500 bp were identified manually and locus ranges were set within Cartographer. Presence versus absence of peaks for these anonymous loci was first called automatically within Cartographer according to a noise threshold set by the program. All loci for all individuals were then visually inspected and manually adjusted (hand-called) by SPE. While handcalling, samples that had a weak or noisy signal were noted and rerun. All scoring was done blind to population of origin.

REPEATABILITY ESTIMATES

To ensure high reliability of analyzed AFLP loci, every individual was genotyped twice for all primer pairs. Each gel included

individuals from both hosts and multiple study populations. Replicate samples of the same individual were run on different gels. The repeatability of each individual locus was estimated as one minus the ratio of the total number of differences for that locus to the total number of individuals genotyped (following Bonin et al. 2004; Pompanon et al. 2005). Individual loci that were less than 90% repeatable were excluded, leaving 447 loci for our analyses. These remaining loci were used to estimate the average "per locus genotyping error rate," as the ratio of the total number of differences across all loci to the total number of comparisons across loci (Bonin et al. 2004; Pompanon et al. 2005). This value was 5.5%.

EMPIRICAL DISTRIBUTION OF F_{ST} AND OUTLIER DETECTION USING SIMULATIONS

Simulations can be used to model the differentiation of neutral loci and thereby identify loci from an empirical dataset that are more highly differentiated than expected under neutrality. Such loci are termed "outlier loci" that are interpreted as the putative subjects of divergent selection, either directly, or via close physical linkage to selected loci. To identify outlier loci, we adopted the approach of Beaumont and Nichols (1996) and Beaumont and Balding (2004) as implemented in the program Dfdist, which is suitable for the analysis of AFLP data and has been frequently used for this purpose (e.g., Bonin et al. 2006; Savolainen et al. 2006). Such divergence-based methods have considerable power to detect loci subject to divergent selection (reviewed by Beaumont 2005; Beaumont and Balding 2004).

This approach involved three steps, each of which was applied separately to each of the 15 possible pairwise population comparisons provided by our six study populations. First, the empirical distribution of F_{ST} values among loci was generated using a Bayesian method in which where allele frequencies are estimated from the proportion of recessive genotypes in the sample (Zhivotovsky 1999). Second, a distribution of simulated F_{ST} values targeted to match the empirical distribution was obtained. To do so, a hierarchical Bayesian approach was used to compute F_{ST} conditional on heterozygosity in a subdivided population under an island model (Wright 1943). Note, however, that the method is robust to the application of alternative models (finite island, infinite island, colonization, stepping stone) and thus that results are not dependent on adherence to the assumptions of an island model (Beaumont and Nichols 1996). We thereby generated 50,000 simulated loci with a mean F_{ST} similar to the "trimmed mean" F_{ST}. This trimmed mean was calculated by removing the highest 30% and lowest 30% of F_{ST} values observed in the empirical dataset. Our use of trimmed means and the 30% threshold is suggested by Beaumont (2005) and employed in other AFLP genome scans (e.g., Bonin et al. 2006). It is an approach used to better estimate the average "neutral" F_{ST} value used in the simulations by removing those loci most likely to be influenced by divergent or stabilizing selection (Beaumont and Balding 2004). This provides a baseline F_{ST} against which potential outlier loci can be statistically evaluated. Third, the empirical and simulated distributions were compared to identify outliers. From the simulated loci, an outlier threshold was calculated at both the upper 95th and 99th quantiles of F_{ST} using the default smoothing parameter of 0.04. Loci above these thresholds were considered outliers.

Because of our particular interest in host-associated divergence, we used two additional approaches to corroborate the host-specific outlier status of loci. Both approaches involved combining individuals from different populations using the same host prior to analysis into one of two pooled "host populations." First, we adopted a "global" Dfdist approach that identified outlier loci based on a Dfdist analysis of the two pooled host populations. Because this analysis pooled populations with different patterns of neutral differentiation, it should be cautiously interpreted. Nonetheless, it is useful for identifying highly supported host-specific outliers (Bonin et al. 2006). Second, we separately conducted a type of logistic regression for each earlier-identified host-specific outlier, assigning individual genotype (AFLP band present vs. absent) as the dependent variable and associated host as the independent variable. A likelihood-ratio test then determined whether host significantly predicted genotype. The regression approach has also been advocated for outlier verification because it is independent of a trimming threshold (Bonin et al. 2007).

CATEGORIZING OUTLIERS

After outlier detection, loci were categorized according to the types of comparisons in which they were outliers (see Table 3). Loci that were not outliers in any comparison were considered neutral. The replication of outlier behavior across population pairs was used to further categorize loci. Loci that were outliers in only one pairwise comparison were deemed "nonrepeated outliers;" those observed in more than one comparison were denoted "repeated outliers." Both categories were further subdivided according to whether outliers were observed only in different-host (DH) pairwise population comparisons (DH-specific outliers), or only in same-host (SH) pairs (SH-specific). We specifically noted DH-specific outliers observed in two or three (the maximum possible) statistically independent comparisons, that is, comparisons that did not share a population in common. We also identified "mixed outliers," that is, those that appear in both DH and SH comparisons. Mixed outliers are difficult to interpret as they may have arisen under host-independent selection or under host-related selection followed by the transfer of alleles to populations using the alternative host via gene flow. Due to this ambiguity we focused on DH-specific and SH-specific outliers for certain interpretations. Further, various analyses treat datasets representing four "classes" of loci: (1) "Total Loci" = all analyzed AFLP loci,

(2) "Neutral Loci" = loci that were not outliers in any comparison, (3) "DH-Only Outliers" = loci that were outliers in multiple different-host comparisons only, and (4) "Other Outliers" = loci that were outliers in multiple comparisons, at least some of which were same-host comparisons (see Tables 3 and 4).

ADDITIONAL ISSUES FOR OUTLIER INFERENCE

Four further issues should be considered when evaluating outliers. First, the large number of loci screened in a genome scan raises the possibility of type I error due to multiple comparisons. For example, within any single pairwise population comparison the number of loci expected to be outliers by random chance (type I error) alone = 0.05 × no. of loci. However, our inferences further minimize such problems by focusing on loci detected as outliers in multiple pairs of populations, and especially on DH-specific outliers. This yields increased confidence because the compound probability of a locus being erroneously identified as an outlier decreases markedly as a function of the number of comparisons in which its is observed and the consistency with which it is observed in the same type of comparison. In the extreme, our study recovered loci that were detected as outliers in all nine DH-comparisons and none of the SH-comparisons, a pattern that would be extraordinarily unlikely to reflect type I error.

Second, among-locus variation in levels of genetic differentiation could reflect variation in mutation rates rather than selection (Balloux and Lugon-Moulin 2002; Hedrick 2005; Noor and Feder 2006). However, this factor is very unlikely to explain our findings. This is partly because mutation rate is not predicted to strongly affect patterns of genetic differentiation when gene flow occurs between populations (Beaumont and Nichols 1996; Balloux and Lugon-Moulin 2002; Hedrick 2005), as is likely in *N. bebbianae* given the evidence for incomplete reproductive isolation and mitochondrial differentiation described earlier, and results from the present study. This argument derives from theory showing that in an island model (Wright 1931), F_{ST} decreases as a function of N ($m + \mu$), where N = local population size, m = migration rate, and μ is mutation rate (Hartl and Clark 1997; Balloux and Lugon-Moulin 2002). Thus, when migration exceeds mutation, as is likely even with very little migration, mutation is expected to contribute little to patterns of genetic differentiation. Furthermore, the mutation rate hypothesis cannot account for loci that are highly differentiated only in DH-comparisons, nor for the greater frequency of DH-specific than SH-specific outliers (see Results). In this respect, our SH-comparisons act as ecological controls for the potential contributions of (presumably ecologically independent) mutation rate variation. Similar arguments apply to other potentially confounding factors, such as sex linkage, whereby reduced effective population size of sex-linked loci could increase genetic divergence by drift, resulting in high F_{ST} of sex-linked loci that are not affected by selection. As for mutation rate variation, this process is not expected to result in host-specific outliers.

Third, the ability to detect and quantify loci under divergent selection depends both on the strength and the history of selection. Because selection varies continuously, the likelihood of identifying a locus as an outlier depends on the strength of divergent selection. Although simulations have shown that Dfdist is very powerful at detecting divergent selection (Beaumont and Nichols 1996; Beaumont and Balding 2004), weakly selected loci are less likely to be observed as outliers than strongly selected loci, due to the increased statistical power necessary to identify significant deviations from neutrality when differentiation is weaker. A history of selective sweeps can similarly cause an underestimation of the proportion of loci that have historically been subject to divergent selection because such sweeps will erase past signatures of selection on linked chromosomal regions. Finally, different mutations/loci may be involved in the divergent adaptation of different pairs of populations, reducing the number of repeated outliers that are detected relative to the actual frequency of parallel selective divergence. None of these factors call into question the status of loci observed to be outliers. However, they suggest that numbers of inferred outliers offer conservative estimates of selection's contributions to genomic differentiation.

Fourth, the independence of outlier loci warrants consideration, and this issue can be addressed by analyzing linkage disequilibrium. Such analyses indicated low linkage disequilibrium in our data (see Results), arguing for a dispersed as opposed to clustered distribution of observed outliers in the genome. However, because AFLP genome scans provide no information on the specific regions of the genome that have been assayed, it is possible that sampling outlier "hot" or "cold" spots could have biased our findings and that our estimates of linkage disequilibrium were affected by shared polymorphisms and independent segregation. Various authors have suggested using multiple primer pairs as a means of ameliorating such biases (e.g., Campbell and Bernatchez 2004; Rogers et al. 2007) and we hope that our use of eight different primer pairs has thus reduced any bias and increased the genomic coverage of our study. Moreover, our sampling of several hundred AFLP loci (gene regions) represents coverage as good as or greater than that presented by most other "genome scans" (Nosil et al., unpubl. ms.).

LINKAGE DISEQUILIBRIUM ANALYSES

To draw inferences about the genomic distribution of outlier loci, we quantified linkage disequilibrium within each population for the Neutral Loci, Other Outlier, and DH-Only Outlier classes. We estimated linkage disequilibrium using two methods, which examine pairs of loci and overall multilocus disequilibrium,

respectively. First, we used the program DIS (Dasmahapatra et al. 2002) to estimate pairwise linkage disequilibria between loci for dominant data (R = the gametic correlation coefficient correcting for variable allele frequencies) using a maximum-likelihood equivalent to that of Hill (1974). This program is limited to 40 loci, so we analyzed the 23 DH-Specific Outlier loci and—to allow for comparable analyses from other classes—also randomly selected 23 Neutral Loci and 23 Other Outlier loci for evaluation. We then used paired t-tests, with population as the unit of replication, to examine whether classes of loci differed significantly in levels of linkage disequilibrium. Second, we used LIAN 3.1 (Haubold and Hudson 2000), which allowed the evaluation of all loci within each class. LIAN tests for independent assortment by computing the number of loci at which pairs of individuals differ. From the distribution of these mismatch values, a variance is calculated. This is compared to the variance expected for linkage equilibrium, allowing computation of a standardized index that represents multilocus association (I_A^S). Paired t-tests determined whether classes of loci differed in these association values across populations.

EVALUATING GENETIC STRUCTURE

We used three approaches to evaluate genetic differentiation among study populations for each class of loci. First, overall genetic differentiation (F_{ST}) was calculated for each population comparison in AFLP-SURV version 1.0 (Vekemans et al. 2002). Second, we constructed population trees. To do so, we created 1000 bootstrapped Nei's genetic distance matrices, generated by AFLP-SURV, that were used to construct a neighbor-joining 50% majority rule bootstrap consensus tree with the programs NEIGHBOR and CONSENSE within PHYLIP 3.6 (Felsenstein 2004). Third, we used STRUCTURE 2.1 (Pritchard et al. 2000), assuming $K = 2$ genetic clusters to evaluate whether the assignment proportions of individuals to the two clusters tended to reflect their host association. STRUCTURE is conventionally used to determine the number of genetic clusters that best fits the data. However, it has many additional uses (Pritchard et al. 2000) including the one adopted here and, for example, its recent application by Gompert et al. (2006) to study hybrid speciation with AFLP data. Our analyses involved Bayesian assignment analysis (Pritchard et al. 2000) and implemented a burn-in of 50,000 generations and a Markov Chain of 500,000 generations using the admixture model. Differences in assignment patterns across classes were visually interpreted. Paired t-tests were also used to assess whether classes of loci differed in degree of assignment to the two clusters. These analyses evaluated mean individual assignment proportions from the six study populations.

Results

OUTLIER DETECTION

Patterns of outlier detection and distribution are summarized in Tables 2–5. See the "Categorizing Outliers" section above for definitions of terms used here and throughout the rest of the article. With respect to the 95th (99th) quantiles, a total of 132(67) outliers were detected, including some from each population comparison (Fig. 2). These included 65(42) nonrepeated and 67(25) repeated outliers. A total of 47(31) nonrepeated and 23(11) repeated DH-specific outliers were observed. The M_{VT} vs. W_{QU} comparison exhibited the highest numbers of total (33) and repeated DH-specific outliers (13) at the 95th quantile, whereas the M_{LI} vs. W_{QU} comparison yielded the highest values (11, 8, respectively) at the 99th quantile. Notably, 16(5) repeated outliers were specifically associated with comparisons involving M_{LI}, whereas no other population was associated with more than 6(5) such outliers. The additional global Dfdist and logistic regression analyses corroborated the existence of DH-specific outliers. The global analyses identified eight outliers at the 95th quantile (loci 113, 127, 131, 172, 219, 265, 295, 347; see Appendix). All of these had been documented in DH-comparisons by the initial Dfdist analyses, and six represented DH-specific outliers that had been observed in multiple independent comparisons. The regression analyses found that host significantly predicted genotype for 22/23 DH-specific outliers (all but locus 437). Because the global and regression analyses were performed primarily to corroborate the Dfdist results, they are not further discussed. By contrast, as outlier patterns are a major focus of this study, they are primarily evaluated and further analyzed in the Discussion to avoid redundancy here.

LINKAGE DISEQUILIBRIUM

DIS estimates of linkage disequilibrium were consistently low ($R < 0.10$) and did not significantly vary between classes of loci across populations (DH-Only Outliers vs. Neutral Loci: $t_5 = 0.66$, $P = 0.73$; DH-Only Outliers vs. Other Outliers: $t_5 = -1.44$, $P = 0.10$; Neutral Loci vs. Other Outliers: $t_5 = -1.37$, $P = 0.11$; Fig. 3). All of the LIAN analyses detected weak but significant (all $P < 0.05$) departures from multilocus linkage equilibrium. However, paired t-tests detected no differences in disequilibrium among classes of loci (DH-Only Outliers vs. Neutral Loci: $t_5 = 1.56$, $P = 0.18$; DH-Only Outliers vs. Other Outliers: $t_5 = 0.79$, $P = 0.46$; Neutral Loci vs. Other Outliers: $t_5 = -0.64$, $P = 0.55$), corroborating the DIS analyses.

ASPECTS OF GENETIC STRUCTURE

Patterns of overall F_{ST} across population comparisons and classes of loci are summarized in Table 2. The highest value among all Total Loci estimates was 0.106.

Table 2. Aspects of differentiation between all pairs of study populations, and with respect to the four focal classes of AFLP loci.

			F_{ST}[2]				Outliers			
	Geographic distance (km)	No. of loci[1]	Total Loci	Neutral Loci	DH-only outliers	Other outliers	Total no.	As percentage of loci 95%(99%)	No. of DH-Only or SH-Only[3]	As percentage of loci 95%(99%)
Different-host										
M_{LI} vs. W_{OS}	368	384	0.1060	0.0734	0.1826	0.2958	23 (7)	6.0% (1.8%)	6 (4)	1.6% (1.0%)
M_{LI} vs. W_{QU}	529	408	0.0853	0.0669	0.2796	0.1848	25 (11)	6.1% (2.7%)	9 (8)	2.2% (2.0%)
M_{LI} vs. W_{VT}	439	359	0.0937	0.0719	0.2613	0.2093	26 (6)	7.2% (1.7%)	8 (5)	2.2% (1.4%)
M_{VT} vs. W_{OS}	376	408	0.0762	0.0553	0.0682	0.2354	29 (8)	7.1% (2.0%)	5 (5)	1.2% (1.2%)
M_{VT} vs. W_{QU}	171	409	0.0455	0.0312	0.1434	0.0947	33 (6)	8.1% (1.5%)	13 (3)	3.2% (0.7%)
M_{VT} vs. W_{VT}	0	396	0.0442	0.0354	0.1154	0.1035	16 (6)	4.0% (1.5%)	8 (6)	2.0% (1.5%)
M_{PA} vs. W_{OS}	238	393	0.0363	0.0324	0.1022	0.0639	17 (10)	4.3% (2.5%)	5 (5)	1.3% (1.3%)
M_{PA} vs. W_{QU}	485	391	0.0339	0.0299	0.1887	0.0609	22 (11)	5.6% (2.8%)	10 (7)	2.5% (1.8%)
M_{PA} vs. W_{VT}	449	370	0.0403	0.0367	0.1499	0.0865	23 (10)	6.2% (2.7%)	11 (7)	3.0% (1.9%)
Same-host										
M_{LI} vs. M_{PA}	162	387	0.0901	0.0832	0.1279	0.2692	19 (9)	4.9% (2.3%)	0 (0)	0% (0%)
M_{PA} vs. M_{VT}	449	376	0.0212	0.0168	0.0031	0.1089	24 (10)	6.4% (2.7%)	2 (1)	0.5% (0.3%)
M_{VT} vs. M_{LI}	439	386	0.0757	0.0552	0.0893	0.2249	20 (10)	5.2% (2.6%)	1 (1)	0.3% (0.3%)
W_{OS} vs. W_{QU}	325	412	0.0318	0.0226	0.0438	0.1142	19 (3)	4.6% (0.07%)	1 (1)	0.2% (0.2%)
W_{QU} vs. W_{VT}	171	396	0.008	0.0001	0.0171	0.0190	14 (4)	3.5% (1.0%)	4 (3)	1.0% (0.8%)
W_{VT} vs. W_{OS}	376	380	0.0457	0.0153	0.0380	0.0380	15 (4)	3.9% (1.0%)	3 (3)	0.8% (0.8%)

[1]Number of polymorphic loci for each population comparison, i.e., data used in Dfdist analyses of that comparison.

[2]For each class of loci, genetic differentiation (F_{ST}) between populations was calculated using 1000 bootstraps in AFLP-SURV version 1.0 (Vekemans et al. 2002).

[3]Numbers of repeated DH-specific outliers in DH-comparisons and of repeated SH-specific outliers in SH-comparisons.

Phylogenetic analyses of population-level differentiation for each class of loci yielded three different topologies (Fig. 4). Total Loci and DH-Only Outliers yielded strong bootstrap support for identical topologies that were consistent with the monophyletic grouping of populations by host plant. By contrast, the Neutral Loci and Other Outlier trees supported a lack of host-associated monophyly. Only the Neutral Loci dataset united the sympatric Vermont populations of maple and willow beetles.

The consistency with which STRUCTURE assigned individuals from each study population to each of the $K = 2$ clusters

Table 3. Categorization of detected outlier loci and possible causes for their observation.

Outlier distribution	[1]Type of comparison	[2]No. of outlier loci 95% (99%)	Possible inferences
Observed	Different host	47 (31)	host-associated selection
in a single	Same host	18 (11)	host-independent selection
comparison			host-associated selection
("nonrepeated")			involving different loci in different populations
Observed	Different host only	23 (11)	host-associated selection
in multiple	Same host only	5 (2)	host-independent selection
comparisons	Different host &	39 (12)	host-independent selection
("repeated")	same host ("mixed")		local effects (population-specific loci)
			host-associated selection
			involving different loci in different populations

[1]Based on the two possible types of pairwise population comparison: those between populations associated with different host plants and those between populations associated with the same host plant.

[2]Number of outlier loci exceeding the 95th or 99th quantiles generated from simulated neutral loci using Dfdist.

Table 4. Proportions of genomic differentiation attributable to outlier type.[1]

95th or 99th quantile[2]	Outlier distribution	Percentage of loci that are outliers	Percentage of loci that are DH-specific outliers	Percentage of outliers that are DH-specific[3]
95th	all	29.5	16.7	53.0
95th	nonrepeated	14.5	10.5	72.3
95th	repeated	15.0	5.1	34.3
99th	all	15.0	9.4	31.8
99th	nonrepeated	9.4	6.9	73.8
99th	repeated	5.6	2.5	44.0

[1]See Table 3 for data from which tabled proportions are calculated.
[2]Threshold from Dfdist analyses that must be exceeded for locus to be considered an outlier.
[3]Percentages are lower for calculations for the "all" and "repeated" outliers because of the inclusion of "mixed" outliers in these calculations. Given the ambiguous causes of mixed outliers (see text), these values might thus be viewed as conservative estimates.

imposed by our analysis is visually documented in Fig. 5. The DH-Only Outliers most consistently separated groups of populations into alternative clusters. This was specifically the case for maple versus willow populations. This contrasts with the lack of strong differentiation among populations at neutral loci. Analyses of the admixture proportions determined by STRUCTURE's proportional assignment of individuals from different populations to different clusters supported these qualitative observations. Most importantly, DH-Only Outliers exhibited more host-associated differentiation than did Neutral Loci (paired-t 5 $= -2.9$, $P = 0.04$) or Total Loci (paired-t 5 $= -3.4$, $P = 0.02$), whereas the Other Outliers showed more population differentiation than Neutral Loci (paired-t 5 $= -3.2$, $P = 0.02$) but not Total Loci (paired-t 5 $= -1.17$, $P = 0.29$). DH-Only Outliers and Other Outliers did not differ in strength of assignment (paired-t 5 $= -0.36$, $P = 0.74$). (See Table 5 for mean assignment proportions.) Another notable pattern was the strong differentiation of the M_{LI} population from all others with respect to the Other Outliers, an observation consistent with the previously noted high frequency of outliers specific to comparisons involving this population. This population resides on an island and in a habitat (pine barrens) differing from that of the other study populations. These factors may have contributed to greater degrees of spatial and associated genetic isolation as well as local host-independent selection in this population, phenomena that would explain our results as well as certain patterns of phenotypic differentiation in this population (D. J. Funk, pers. obs.).

Table 5. Summary of mean individual assignment proportions to each of two clusters by the program STRUCTURE 2.1, across predefined study populations.

AFLP Dataset	M_{LI}	M_{PA}	M_{VT}	W_{OS}	W_{QU}	W_{VT}
Total loci						
Cluster 1	0.771	0.412	0.501	0.235	0.298	0.245
Cluster 2	0.229	0.588	0.499	0.765	0.702	0.755
Neutral loci						
Cluster 1	0.678	0.608	0.647	0.436	0.614	0.503
Cluster 2	0.322	0.392	0.353	0.564	0.386	0.497
DH-specific outliers						
Cluster 1	0.809	0.666	0.626	0.219	0.082	0.108
Cluster 2	0.191	0.334	0.374	0.781	0.918	0.892
Other outliers						
Cluster 1	0.075	0.819	0.696	0.838	0.600	0.637
Cluster 2	0.925	0.181	0.304	0.162	0.400	0.363

Discussion

Evaluating how natural selection explains genetic variation has been a central theme of population genetics for decades. Broad-scale insights into this issue are now being provided by "genome scans" that can survey hundreds of gene regions scattered across the genome and allow the identification of those that are differentiating under divergent selection. The use of AFLPs for such scans has provided access to such population genomic insights in even nonmodel organisms. Such studies are just beginning to accumulate and many more will be required before general patterns can emerge and be evaluated. Like other genome scans, the present study evaluates the frequency of gene regions under selection. However, a more specific goal was to quantify the contributions of particular ecological sources of selection to genetic differentiation: those associated with an insect's host plants. Doing so required comparative analyses that isolated these contributions by evaluating multiple populations associated with different host plants using appropriately controlled experimental designs. Specifically, we evaluated the role of "host-related selection" among populations of *N. bebbianae* leaf beetles associated with either maple or willow host plants. Our design allowed the absolute contribution of host-related selection to be estimated, but also the degree to which it accounts for overall divergent selection in these insects, as compared to host-independent factors. Thus it allows an initial genomic evaluation of the general evolutionary importance of host adaptation. Because these beetles appear to exist somewhere along the "ecological speciation continuum" (see Methods), we also discuss the potential implications of our findings for understanding this phenomenon. Our results compellingly document various contributions of selection and indicate a principal role for host-related selection in ongoing diversification within *N. bebbianaae*.

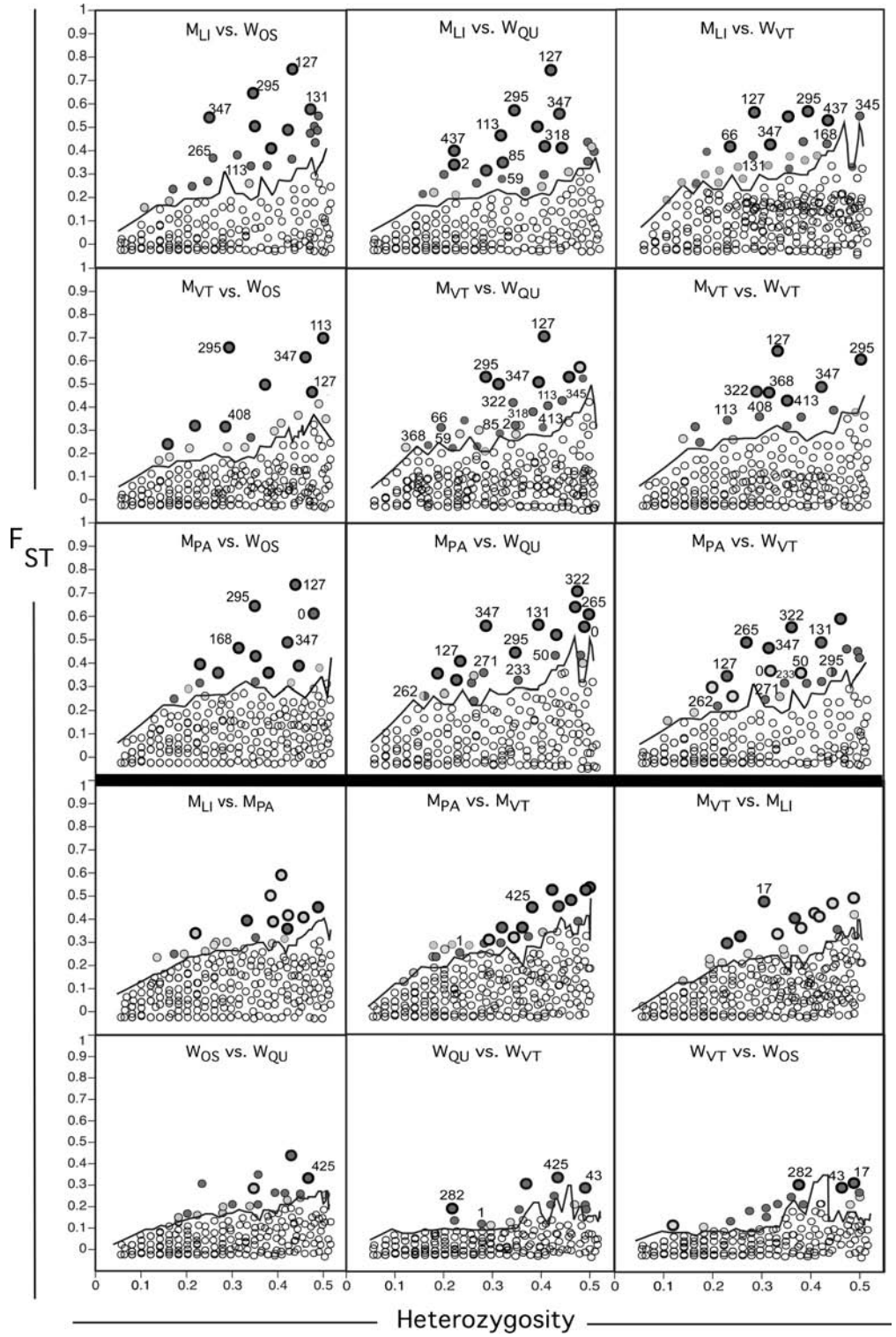

Figure 2. Results of Dfdist analyses. Each of the 15 plots illustrates the empirical distribution of F_{ST} values for individual AFLP loci (circles) in relation to a simulated 95th quantile for neutrally evolving loci, for a given population comparison. The shaded loci with F_{ST} values exceeding this quantile are "outlier" loci putatively evolving under divergent selection. "Repeated" outliers that are observed in multiple population comparisons are dark gray. Outliers that additionally exceed the 99th quantile are outlined in bold. Comparisons between populations of different host forms (DH-comparisons) are above the thick black line and those between populations of the same host form (SH-comparisons) are below it. Repeated outliers that are restricted to different-host comparisons or to same-host comparisons are numbered (see also Table A1). Circles that are half light-gray and half dark-gray indicate two separate loci that have identical values but belong to different categories (described above).

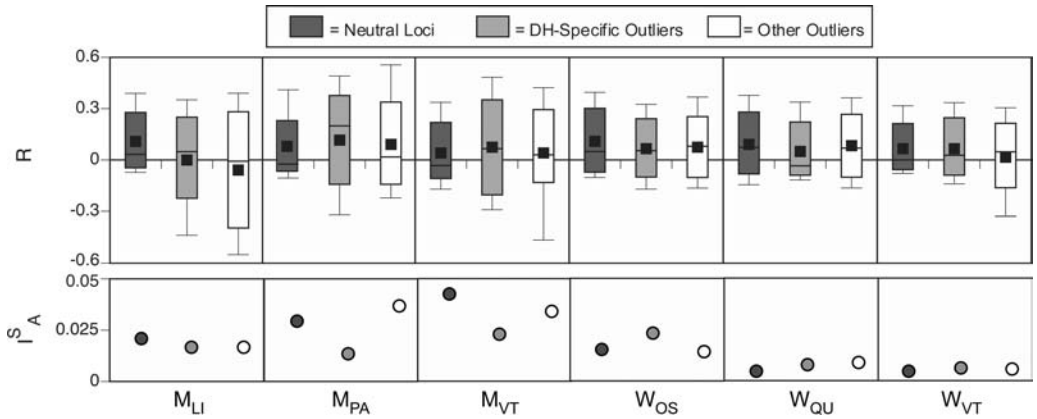

Figure 3. Box plots providing estimates of linkage disequilibrium for three of the focal classes of AFLP loci. The top panel shows box plots of pairwise linkage disequilibrium values (*R*), calculated using DIS. Black square, mean; bottom, middle, and top lines of box, 25th, 50th, and 75th quantiles, respectively; stems, 10th and 90th quantiles. Bottom panel depicts multilocus linkage disequilibrium values (I_A^S) calculated using LIAN 3.0.

COMPARATIVE INSIGHTS ON SELECTION'S CONTRIBUTIONS TO GENOMIC DIFFERENTIATION

The focal interpretations of this study derive from estimates of the number and proportion of loci affected by divergent selection. However, evaluating these values depends on the criteria employed to determine whether a given locus is considered an outlier. Genome scans studies are still in their infancy and the existing literature illustrates that accepted standards for these criteria have not emerged, leading to challenges when comparing results across different studies. Many such studies (e.g., Campbell and Bernatchez 2004; Scotti-Saintagne et al. 2004; Savolainen et al. 2006) use the 95th quantile of simulated F_{ST} values as the threshold for outlier acceptance (analogous to the convention of accepting $P < 0.05$ as statistically significant), whereas certain studies have more conservatively adopted the 99th quantile (Wilding et al. 2001; Bonin et al. 2006; Murray and Hare 2006). As compared to nonrepeated outliers, repeated outliers offer additional confidence that divergent selection is responsible for outlier status (Campbell and Bernatchez 2004; Bonin et al. 2006; Nosil et al. 2008) because of the decreased likelihood that the F_{ST} of a locus would repeatedly exceed its neutral expectation by chance. Combining the use of different quantiles and outlier replication as criteria yields multiple thresholds for identifying outliers. Depending on which of these is used, the estimated contributions of selection to genomic differentiation vary. We report estimates based on all these criteria (Table 4) to illustrate this and to facilitate the identification of general trends.

The contributions of selection can be also quantified at two different levels, another source of heterogeneity in the literature. We report our findings at both these levels and thus will discuss both "summary" and "pairwise" estimates. First, our "summary" estimates of selection are based on the sum of outlier loci observed in any of the 15 pairwise population comparisons in our analysis (Table 4). These estimates evaluate the proportion of total assayed loci that contribute to divergent selection in at least some population comparisons, that is, at least some of the time. Second, our "pairwise" estimates treat only those loci that are observed to be outliers in a single specific pairwise population comparison (Table 2). These estimates allow the evaluation of selection's contributions to genetic differentiation per se between a specific pair of populations. Below, we discuss our summary and pairwise insights on the first three questions raised in the Introduction. For the purposes of this discussion our pairwise estimates are all derived from 95th quantile data. (See Tables 2–4 for additional information.)

First, the distinction between putatively selected (i.e., outlier) versus putatively neutral loci provided by a Dfdist-based genome scan provides a first approximation of the proportion of total observed genetic differentiation that reflects overall divergent selection (i.e., of all sources). Depending on the stringency of the outlier threshold adopted, our summary calculation of this value varied considerably, from 5.6% to 29.5%. The two threshold criteria of intermediate stringency (repeated 95th quantile loci and all 99th quantile loci) both yielded an estimate of 15%. By contrast, our pairwise estimates varied from 3.5% to 8.1% of the assayed portions of the genome that have been affected by selection.

Second, evaluating outliers according to their distribution across ecological categories of population comparison (different-host versus same-host) allowed us to more specifically evaluate the proportion of total assayed genomic differentiation that reflects host-related selection. This calculation is possible because repeated DH-specific outliers provide de facto comparative evidence for the effects of this phenomenon. Summary estimates indicated that 2.5–16.7% of the assayed genome appears to be influenced by host-related selection, whereas pairwise estimates varied

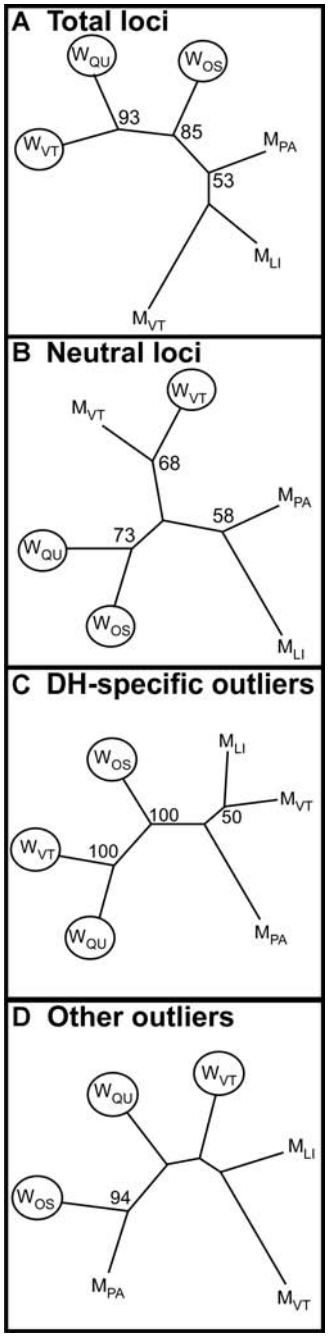

Figure 4. Fifty percent majority rule neighbor-joining population trees for each of the four focal classes of AFLP loci, based on Nei's genetic distances between study populations. Bootstrap values greater than 50% are indicated at nodes. Willow-associated beetle populations are circled to emphasize host-associated patterns.

between 1.2% and 3.2% among individual DH-comparisons. It is worth noting that these pairwise estimates are conservative because if divergent host-associated selection does act on a particular locus, but does so only in a single population comparison, it will not be included when estimating these values. Yet the contributions of unique genes/mutations to divergent host-related adaptation of particular populations pairs can yield such nonrepeated DH-outliers. The likelihood that many nonrepeated DH-outliers do indeed reflect host-related selection is supported by the observation that a high proportion of total outliers are associated with DH-comparisons in nonrepeated as well as repeated categories (Table 3).

Third, the proportion of outliers that represent DH-specific outliers provides an estimate of the proportion of total divergent selection that is specifically host-related. Our summary estimates thereby indicated that 31.8%–73.8% of selection on these beetles is associated with adaptation to the host plant. These values vary considerably in part because we decided to include mixed outliers in certain calculations, despite the ambiguity of their host-independent versus host-specific origin (see discussion in Methods) (Table 4). For our pairwise estimates, we calculated the proportion of outliers in each DH-comparison that were repeated and DH-specific and similarly calculated the proportion of outliers in each SH-comparison that were repeated and SH-specific. We then compared DH versus SH values using a Mantel test, based on 10,000 randomizations using the program MANTEL (Manly 1997). This analysis demonstrated that (DH-specific) loci diverging under host-specific selection were significantly more frequent than (SH-specific) loci most parsimoniously interpreted as diverging due to host-independent factors ($r = -0.753$, Mantel's $t = -2.0698$, $P = 0.0192$).

Although the literature on AFLP-based genome scans is still scant, we compared our findings with existing examples (Wilding et al. 2001; Emelianov et al. 2003; Campbell and Bernatchez 2004; Scotti-Saintagne et al. 2004; Acheré et al. 2005; Bonin et al. 2006; Jump et al. 2006; Murray and Hare 2006; Savolainen et al. 2006; Nosil et al. 2008). To ensure commensurability we compared results based on the same outlier criteria and type of estimate (summary vs. pairwise) reported for a given study. First, these comparisons revealed the proportion of genomic differentiation attributable to selection in our study to be consistently higher than in previous studies. Thus, selection appears to be playing a comparatively large role in the genomic differentiation of these beetles. Intriguingly a rather high estimate was also obtained in another published study of insect herbivores (Emelianov et al. 2003). However, the use of different techniques for quantifying locus differentiation in that study limits the informativeness of this comparison. Second, our proportions of assayed genomic differentiation attributable to host-related selection were likewise higher than the estimated contributions of particular selective factors deriving from the few comparable "ecological" genome scans (Wilding et al. 2001; Campbell and Bernatchez 2004; Bonin et al. 2006; Nosil et al. 2008), with the exception of one summary estimate of 5% at the 99th quantile that exceeded the 2.5% from our study at this quantile.

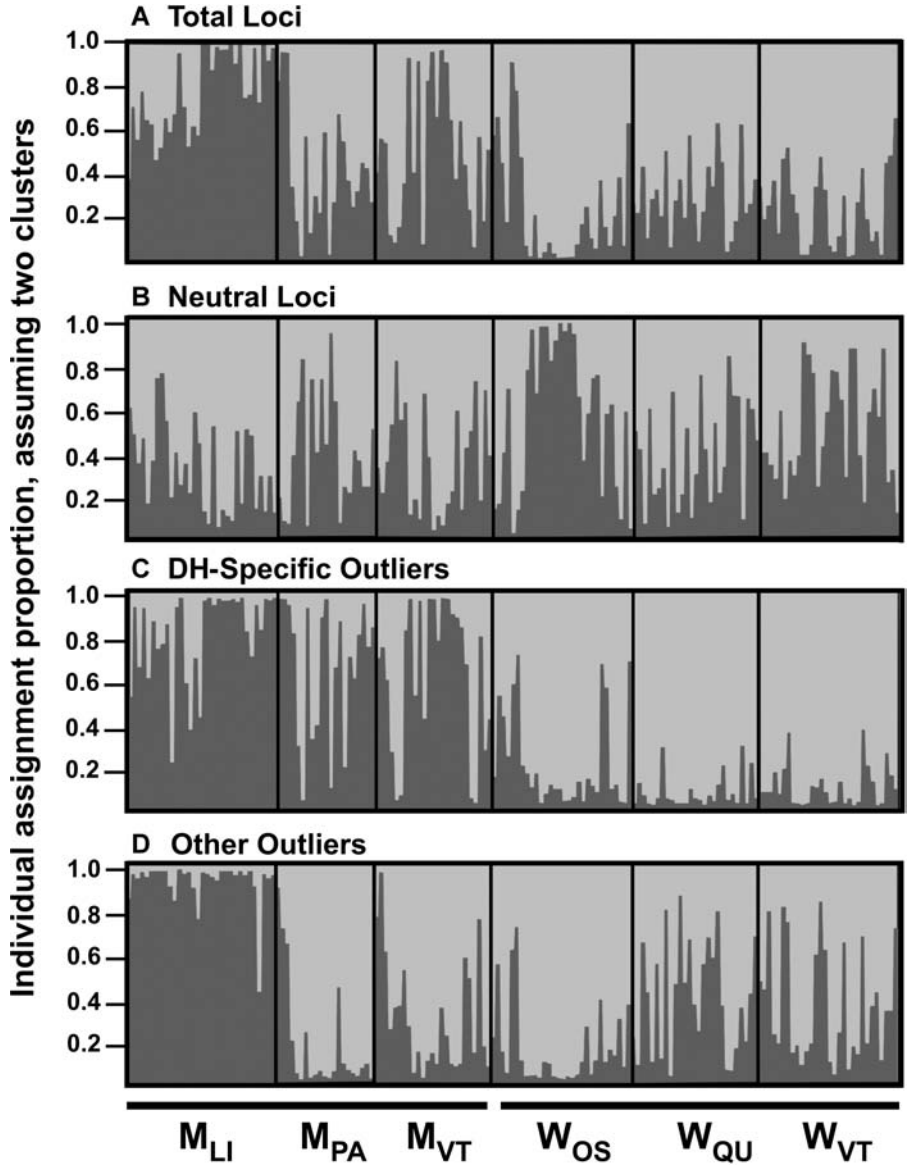

Figure 5. Bar plots showing Bayesian assignment proportions for $K = 2$ clusters provided by STRUCTURE 2.1 for the four focal classes of AFLP loci and each study population. Each vertical bar corresponds to one individual. The proportion of each bar that is light or dark gray represents an individual's assignment proportion to clusters one versus two. See text for details.

Third, and most intriguing, are our findings on the proportion of selection inferred to be due to host-related factors. One might expect adaptive population divergence to typically reflect a variety of disparate selection factors. However, the proportion of total divergent selection across the genome that can be attributable to particular ecological sources have not previously been evaluated. Here, both summary and pairwise estimates of these proportions strikingly indicated that more selected loci in these beetles are diverging under host-related selection than under all host-independent sources (e.g., factors associated with geographic variation in climate, generalist predators, etc.) combined. The results of this initial genome-scan-based evaluation of this issue thus provide quantitative support for the long-held supposition that host-related factors play a principal role in the adaptive differentiation of specialist insect herbivores (e.g., Bernays and Chapman 1994; Winkler and Mitter 2008).

CANDIDATE LOCI

Further evidence on the biological significance of host-related selection in these leaf beetle populations derives from a locus-specific consideration of the 23 repeated 95th quantile DH-specific outliers. At least some of these loci were observed to be outliers in every DH-comparison (mean = 8.2 ± 0.8 of these 23 outlier loci per comparison; range = 5–13), illustrating the

population-level generality of particular host-related selection pressures. Nine of the 23 were outliers in at least two statistically independent population comparisons, and three of these were outliers in three independent comparisons, the maximum possible. Moreover, each of these three loci proved to be an outlier in all nine DH-comparisons, in which they cumulatively exceeded the 99th quantile in 26/27 instances. These loci generally exhibited the highest or among the highest F_{ST} values in each population comparison (mean F_{ST} across DH-comparisons: locus 127 $= 0.58 \pm 0.05$; locus 295 $= 0.55 \pm 0.11$; locus 347 $= 0.51 \pm 0.06$). These results apparently provide compelling locus-specific examples of uniformly strong and geographically consistent host-related selection pressures. These loci thus provide prospective "candidate gene regions" for future investigations into the molecular biology and populations genetics of traits under host-related selection.

LINKAGE DISEQUILIBRIUM

Studies interpreting the proportion of loci found to be outliers to reflect the proportion of the genome affected by selection implicitly assume that these outliers are randomly distributed across the genome. The truth of this depends on the degree of physical linkage among outlier loci. In the extreme, all outliers could derive from one linkage group (e.g., an inversion) and thus reflect selection on one region (Noor et al. 2001; Feder et al. 2003). However, the low linkage disequilibrium observed here (mean $R <$ 0.15 for all classes of loci; Fig. 3), and its nonsignificant difference between DH-Only Outliers and Neutral Loci, both indicate that this study's outliers represent many genomically dispersed regions that are unlikely to have adversely affected our results. By comparison, values of R in hybrid zone studies often range up to 0.60 even for unlinked loci (Szymura and Barton 1986; Mallet et al. 1990; Dasmahapatra et al. 2002). Further, Beaumont and Nichols (1996) suggest that effectively estimating F_{ST} with their method requires > 20 independent loci whereas we evaluated 447. Moreover, even given nonindependence, outliers in different-host comparisons would have to have been disproportionately detected versus those in same-host comparisons to affect our critical inferences on the relative frequency of these outlier classes, an unlikely scenario. In sum, the design and findings of this study support its conclusions in the context of the issue of linkage disequilibrium.

PATTERNS OF POPULATION DIFFERENTIATION AND THE BALANCE OF SELECTION AND GENE FLOW

This study also advances understanding of general patterns of evolutionary differentiation among the study populations, neutral as well as selected. Specifically, several forms of evidence support past findings indicating that these are incompletely differentiated conspecific populations that exhibit a degree of evolutionary independence (divergence) despite some interaction (via evolutionarily recent or ongoing gene flow): (1) Population-level F_{ST} values for neutral AFLP loci in our 15 comparisons reach but never exceed moderate levels (maximum $= 0.0832$; Table 2) that are typical of those observed among conspecific populations of other herbivore species at similar spatial scales. For example, across many species of herbivores with populations separated by 50–500 km, mean F_{ST} was reported to be 0.07 ± 0.01 by Peterson and Denno (1998). (2) The STRUCTURE analyses revealed some individuals in each population with a majority of their assayed genome assigned to cluster 1 and other individuals largely assigned to cluster 2 (Fig. 5). This suggests a recent history of genetic exchange between maple and willow populations. (3) A lack of alternatively fixed alleles for any locus in any population comparison also argues for recent gene flow (Fig. 2). (4) Population trees contradicted host-associated monophyly while providing insights on with the balance of selection and gene flow (see also Wilding et al. 2001; Campbell and Bernatchez 2004; Bonin et al. 2006; Nosil et al. 2008): (a) Neither Neutral Loci nor Other Outlier trees were consistent with the monophyletic grouping of populations by host. (b) Datasets including DH-specific loci (Total Loci and DH-Only Outlier classes) did support host-associated monophyly. (c) Bootstrap support for host-associated monophyly was stronger for the DH-Only Outlier tree even though the Total Loci dataset was more than an order of magnitude larger. (d) The Neutral Loci tree was the only one to unite the sympatric maple and willow beetle populations. In sum, these patterns (1–4) suggest that neutral gene flow has occurred between maple and willow populations, yet host-related selection is sufficiently strong to promote consistent enough differentiation for various outlier loci to affect tree structure. The sympatric Vermont maple and willow populations are particularly informative. The union of these populations only in the Neutral Loci tree (Fig. 4) combined with the observation of eight DH-specific outliers in this comparison (Table 2) suggests that divergent host-related selection must be strong enough to yield adaptive differentiation in the face of recent or ongoing gene flow.

RELEVANCE FOR ECOLOGICAL SPECIATION

Past work demonstrates partial host-associated ecological divergence and reproductive isolation among populations of *N. bebbianae* leaf beetles associated with different host plants, presenting them as an informative study system for the investigation of ecological speciation. Divergent host adaptation appears to have directly resulted in the evolution of inherently ecological barriers such as habitat isolation and immigrant inviability, whereas the indirect effects of divergent host adaptation are apparent in the consistent sexual isolation—a noninherently ecological barrier—exhibited between populations associated with different hosts, even in the absence of host foliage (e.g., Funk 1998; Funk et al.

2002; Funk and Nosil 2008). A positive association between population divergence in host use traits and degree of sexual isolation (Funk 1998; Funk and Nosil 2008) provides more direct evidence that host-related selection is driving the evolution of reproductive isolation, pushing these populations closer to speciation. All these patterns apply to maple- and willow-associated *N. bebbianae* populations.

This evidence that divergent host adaptation is driving speciation in our focal populations allows our present study to also be viewed as an evaluation of the genomic basis of ecological speciation. On this view, our identification of dozens of loci specifically associated with divergent host-related selection is consistent with the hypothesis that ecological speciation has a polygenic basis in this species. Indeed, as any genome scan can sample only a small fraction of the genome, as well as for additional reasons described earlier (see Methods), the absolute numbers of selected loci reported here must considerably underestimate the actual number of gene regions responding to selection. Speciation models differ considerably in their predictions about the number of loci involved in the evolution of reproductive isolation and on the required strength of divergent selection (Maynard Smith 1966; Felsenstein 1981; Dieckmann and Doebeli 1999; Kondrashov and Kondrashov 1999; Berlocher and Feder 2002; Coyne and Orr 2004; Gavrilets 2004). These are especially relevant parameters for evaluating the plausibility of speciation models involving initially nonallopatric populations. Maple- and willow-associated populations of *N. bebbianae* commonly occur sympatrically and syntopically and the sympatric Vermont populations seem to be adaptively differentiating in the face of gene flow. Additionally, our three "candidate gene" outliers raise important questions about the strength of divergent host-related selection and—via comparison with less consistently repeated and less highly differentiated outliers—on the uneven degree to which various selected loci may promote ongoing ecological speciation. These data may thus form the basis for future studies that evaluate alternative speciation models in this system.

CONCLUSIONS AND FUTURE PROSPECTS

Inferences from the present study were made possible by a comparative experimental design that integrates controlled population comparisons with a genome scan approach to evaluate several outstanding questions about the contributions of natural selection to genomic differentiation and ecological speciation. This study identifies dozens of genomically dispersed loci (= gene regions) that contribute to these processes in maple- and willow-associated populations of *N. bebbianae* leaf beetles. Compared to prior work, this study demonstrates high contributions of selection generally, and of host-related selective factors in particular, to assayed genomic differentiation. Most importantly, it suggests that host-related selection is largely responsible for overall adaptive divergence in this system, with host-independent sources of selection playing a smaller role. Future work will further explore the balance between selection and gene flow in this system and obtain additional data for the comparison of alternative speciation models. An obvious and exciting future direction is the molecular characterization of the linkage groups associated with our three candidate "ecological speciation gene regions." Such work might detect specific genes playing major roles in adaptive divergence and speciation, identify the function of those genes, and allow the molecular evolutionary evaluation of the selective and demographic factors explaining the history of diversification in this species.

ACKNOWLEDGMENTS

We thank M. Chapman, D. Wills, C. Nice, Z. Gompert, J. Burke, and J. Ellis for technical advice and assistance, J. Galindo for considerable input on conceptual and analytical issues, A. Brown and N. Spiegel for contributions to data collection, D. McCauley for discussion, and D. McCauley, C. Benkman, D. Pfennig, and four anonymous reviewers for comments on earlier versions of the manuscript. SPE was supported by Vanderbilt University, and DJF was funded by Vanderbilt University and the National Science Foundation (DEB 0221262) during the conduct of this study. PN was funded by a post-doctoral fellowship from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

LITERATURE CITED

Achere, V., J. M. Favre, G. Besnard, and S. Jeandroz. 2005. Genomic organization of molecular differentiation in Norway spruce (*Picea abies*). Mol. Ecol. 14:3191–3201.

Balloux, F., and N. Lugon-Moulin. 2002. The estimation of population differentiation with microsatellite markers. Mol. Ecol. 11:155–165.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. Lond. B 263:1619–1626.

Beaumont, M. A. 2005. Adaptation and speciation: what can F_{ST} tell us? Trends Ecol. Evol. 20:435–440.

Beaumont, M. A., and D. J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. Mol. Ecol. 13:969–980.

Berlocher, S. H., and J. L. Feder. 2002. Sympatric speciation in phytophagous insects: moving beyond controversy? Annu. Rev. Entomol. 47:773–815.

Bernays, E. A., and R. F. Chapman. 1994. Host-plant selection by phytophagous insects. Chapman & Hall, New York.

Black, W. C., C. F. Baer, M. F. Antolin, and N. M. DuTeau. 2001. Population genomics: genome-wide sampling of insect populations. Annu. Rev. Entomol. 46:441–469.

Bonin, A., E. Bellemain, P. B. Eidesen, F. Pompanon, C. Brochmann, and P. Taberlet. 2004. How to track and assess genotyping errors in population genetics studies. Mol. Ecol. 13:3261–3273.

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

Bonin, A., D. Ehrich, and S. Manel. 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. Mol. Ecol. 16:3737–3758.

Boughman, J. W. 2001. Divergent sexual selection enhances reproductive isolation in sticklebacks. Nature 411:944–948.

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

Burke, J. M., S. J. Knapp, and L. H. Rieseberg. 2005. Genetic consequences of selection during the evolution of cultivated sunflower. Genetics 171:1933–1940.

Campbell, D., and L. Bernatchez. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. Mol. Biol. Evol. 21:945–956.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 307:1928–1933.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, MA.

Craig, T. P., J. K. Itami, W. G. Abrahamson, and J. D. Horner. 1993. Behavioral evidence for host-race formation in *Eurosta solidaginis*. Evolution 47:1696–1710.

Darwin, C. 1859. On the origin of species by means of natural selection or the preservation of favored races in the struggle for life. J. Murray, London.

Dasmahapatra, K. K., M. J. Blum, A. Aiello, S. Hackwell, N. Davies, E. P. Bermingham, and J. Mallet. 2002. Inferences from a rapidly moving hybrid zone. Evolution 56:741–753.

Dieckmann, U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400:354–357.

Dumolin-Lapegue, S., M.-H. Pemonge, and R. J. Petit. 1997. An enlarged set of consensus primers for the study of organelle DNA in plants. Mol. Ecol. 6:393–397.

Egan, S. P., and D. J. Funk. 2006. Individual advantages to ecological specialization: insights on cognitive constraints from three conspecific taxa. Proc. R. Soc. Lond. B. 273:843848.

Emelianov, I., F. Marec, and J. Mallet. 2003. Genomic evidence for divergence with gene flow in host races of the larch budmoth. Proc. R. Soc. Lond. B 271:97–105.

Endler, J. A. 1986. Natural selection in the wild. Princeton Univ. Press, Princeton, NJ.

Feder, J. L., S. B. Opp, B. Wlazlo, K. Reynolds, W. Go, and S. Spisak. 1994. Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. Proc. Natl. Acad. Sci. USA 91:7990–7994.

Feder, J. L., S. H. Berlocher, and S. B. Opp. 1998. Sympatric host-race formation and speciation in *Rhagoletis* (Diptera: Tephritidae): a tale of two species for Charles D. Pp. 408–441 *in* S. Mopper, and S. Strauss, eds. Genetic structure and local adaptation in natural insect populations: effects of ecology, life history, and behavior. Chapman & Hall, New York.

Feder, J. L., F. B. Roethele, K. Filchak, J. Niedbalski, and J. Romero-Severson. 2003. Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. Genetics 163:939–953.

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

———. 2004. PHYLIP. Distributed by the author, Department of Genome Sciences, Univ. of Washington, Seattle.

Fisher, R. A. 1930. The genetical theory of natural selection. Clarendon Press, Oxford.

Funk, D. J. 1996. The evolution of reproductive isolation in *Neochlamisus* leaf beetles: a role for selection. PhD Dissertation., State Univ. of New York, Stony Brook, NY.

———. 1998. Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. Evolution 52:1744–1759.

———. 1999. Molecular systematics of cytochrome oxidase I and 16S from *Neochlamisus* leaf beetles and the importance of sampling. Mol. Biol. Evol. 16:67–82.

Funk, D. J., and P. Nosil. 2008. Comparative analyses and ecological speciation in herbivorous insects. Pp. 117–135 *in* K. Tilmon, ed. Ecology and evolution of herbivorous insects. University of California Press, Berkeley, CA.

Funk, D. J., K. E. Filchak, and J. L. Feder. 2002. Herbivorous insects: model systems for the comparative study of speciation ecology. Genetica 116:251–267.

Funk, D. J., P. Nosil, and W. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proc. Natl. Acad. Sci. USA 103:3209–3213.

Gavrilets, S. 2004. Fitness landscapes and the origin of species. Princeton Univ. Press, Princeton, NJ.

Gompert, Z., J. A. Fordyce, M. L. Forister, A. M. Shapiro, and C. C. Nice. 2006. Homoploid hybrid speciation in an extreme habitat. Science 314:1923–1925.

Grant, P. R., B. R. Grant, and K. Petren. 2005. Hybridization in the Recent Past. Am. Nat. 166:56–67.

Hartl, D. L., and A. G. Clark. 1997. Principles of population genetics. Sinauer Associates, Sunderland, MA.

Haubold, B., and Hudson, R. R. 2000. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Bioinformatics 16:847–848.

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

Hedrick, P. W. 2005. A standardized genetic differentiation measure. Evolution 59:1633–1638.

Hill, W. G. 1974. Estimation of linkage disequilibrium in randomly mating populations. Heredity 33:229–239.

Hoekstra, H. E., R. J. Hirschmann, R. A. Bundey, P. A. Insel, and J. P. Crossland. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313:101–104.

Jiggins, C. D., R. E. Naisbit, R. L. Coe, and J. M. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411:302–305.

Joron, M., R. Papa, M. Beltran, N. Chamberlain, J. Mavarez, S. Baxter, M. Abanto, E. Bermingham, S. J. Humphray, J. Rogers, et al. 2006. A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. PLOS Biol. 4:1831–1840.

Jump, A. S., J.M Hunt, J.A Martinez-Izquirdo, and J. Penuelas. 2006. Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. Mol. Ecol. 15:3469–3480.

Karren, J. B. 1972. A revision of the subfamily Chlamisinae of America north of Mexico (Coleoptera: Chrysomelidae). Univ. Kansas Sci. Bull. 49:875–988.

Kimura, M. 1968. Evolutionary rate at the molecular level. Nature 217:624–626.

———. 1986. DNA and the neutral theory. Philos. Trans. R. Soc. Lond. B 312:343–354.

King, J. L., and T. H. Jukes. 1969. Non-Darwinian evolution. Science 164:788–798.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gilbert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. Am. Nat. 157:245–261.

Kondrashov, A. S., and F. A. Kondrashov. 1999. Interactions among quantitative traits in the course of sympatric speciation. Nature 400:351–354.

Langerhans, R. B., M. E. Gifford, and E. O. Joseph. 2007. Ecological speciation in *Gambusia* fishes. Evolution 61:2056–2074.

Lexer, C., M. E. Welch, J. L. Durphy, and L. H. Rieseberg. 2003. Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. Mol. Ecol. 12:1255–1235.

Lexer, C., Z. Lai, and L. H. Rieseberg. 2004. Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implication for the origin of *Helianthus paradoxus*, a diploid hybrid species. New Phytol. 161:225–233.

Mallet, J., N. Barton, G. Lamas, C. Jose Santisteban, M., Muedas, and H. Eeley. 1990. Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. Genetics 124:921–936.

Manly, B. F. J. 1997. Randomization, bootstrap and Monte Carlo methods in biology. Chapman and Hall, London.

Maynard Smith, J. 1966. Sympatric speciation. Am. Nat. 100:637–650.

Mayr, E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York.

———. 1947. Ecological factors in speciation. Evolution 1:263–288.

———. 1963. Animal species and evolution. Harvard Univ. Press, Cambridge.

Muller, H. J. 1942. Isolating mechanisms, evolution and temperature. Biol. Symp. 6:71–125.

Murray, M. C., and M. P. Hare. 2006. A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. Mol. Ecol. 15:4229–4242.

Nachman, M. W., H. E. Hoekstra, and S. L. D'Agostino. 2003. The genetic basis of adaptive melanism in pocket mice. Proc. Natl. Acad. Sci. USA 100:5268–5273.

Nielsen, R. 2005. Molecular signatures of natural selection. Annu. Rev. Genet. 39:197–218.

Noor, M. A. F., and J. L. Feder. 2006. Speciation genetics: evolving approaches. Nat. Rev. Genet. 7:851–861.

Noor, M. A. F., K. L. Grams, L. A. Bertucci, and J. Reiland. 2001. Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. USA 98:12084–12088.

Nosil, P. 2004. Reproductive isolation caused by visual predation on migrants between divergent environments. Proc. R. Soc. Lond. B 271:1521–1528.

———. 2007. Divergent host-plant adaptation and reproductive isolation between ecotypes of *Timema cristinae* walking-sticks. Am. Nat. 169:151–162.

Nosil, P., and B. J. Crespi. 2006. Experimental evidence that predation promotes divergence during adaptive radiation. Proc. Natl. Acad. Sci. USA 103:9090–9095.

Nosil, P., B. J. Crespi, and C. P. Sandoval. 2002. Host-plant adaptation drives the parallel evolution of reproductive isolation. Nature 417:440–443.

Nosil, P., T. H. Vines, and D. J. Funk. 2005. Perspective : Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705–719.

Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: 'isolation by adaptation' and multiple roles for divergent selection. Evolution 62:316–336.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. 23:263–286.

———. 2002. Near-neutrality in evolution of genes and gene regulation. Proc. Natl. Acad. Sci. USA 99:16134–16137.

Peterson, M. A., and R. F. Denno. 1998. Life-history strategies and the genetic structure of phytophagous insect populations. Pp. 263–322 *in* S. Mopper, and S. Strauss, eds. Genetic structure and local adaptation in natural insect populations: effects of ecology, life history, and behavior. Chapman & Hall, New York.

Pompanon, F., A. Bonin, E. Bellemain, and P. Taberlet. 2005. Genotyping errors: causes, consequences and solutions. Nat. Rev. Genet. 6:847–859.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. Trends Ecol. Evol. 16:351–358.

Rogers, S. M., N. Isabel, and L. Bernatchez. 2007. Linkage maps of the dwarf and normal lake whitefish (*Coregonus clupeaformis*) species complex and their hybrids reveal the genetic architecture of population divergence. Genetics 175:375–398.

Rundle, H. D. 2002. A test of ecologically dependent post-mating isolation between sympatric sticklebacks. Evolution 56:322–329.

Rundle, H., and P. Nosil. 2005. Ecological speciation. Ecol. Letters 8:336–352.

Rundle, H. D., L. Nagel, J. W. Boughman, and D. Schluter. 2000. Natural selection and parallel speciation in sympatric sticklebacks. Science 287:306–308.

Savolainen, V., M. C. Anstett, C. Lexer, I. Hutton, J. J. Clarkson, M. V. Norup, M. P. Powell, D. Springate, N. Salamin, and W. J. Baker. 2006. Sympatric speciation in palms on an oceanic island. Nature 441:210–213.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford Univ. Press, Oxford.

———. 2001. Ecology and the origin of species. Trends Ecol. Evol. 16:372–380.

Schluter, D., and L. M. Nagel. 1995. Parallel speciation by natural selection. Am. Nat. 146:292–301.

Scotti-Saintagne, C., S. Mariette, I. Porth, P. G. Goicoechea, T. Barreneche, K. Bodenes, K. Burg, and A. Kremer. 2004. Genome scanning for interspecific differentiation between two closely related oak species (*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.). Genetics 168:1615–1626.

Stinchcombe, J. R., and H. E. Hoekstra. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100:158–170.

Storz, J. F. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. 14:671–688.

Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *Bombina variegata*, near Cracow in Southern Poland. Evolution 40:1141–1159.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17:1105–1109.

Vasemagi, A., and C. R. Primmer. 2005. Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. Mol. Ecol. 14:3623–3642.

Vasemagi, A., J. Nilsson, and C. R. Primmer. 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.). Mol. Biol. Evol. 22:1067–1076.

Vekemans, X., T. Beauwens, M. Lemaire, and I. Roldan-Ruiz. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Mol. Ecol. 11:139–151.

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

Via, S., A. C. Bouck, and S. Skillman. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. Evolution 54:1626–1637.

Vines, T. H., and D. Schluter. 2006. Strong assortative mating between allopatric sticklebacks as a by-product of adaptation to different environments. Proc. R. Soc. Lond. B 273:911–916.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Vandelee, M. Hornes, A. Fritjers, J. Pot, J. Peleman, M. Kuiper, et al. 1995. AFLP—a new technique for DNA-fingerprinting. Nucl. Acids Res. 23:4407–4414.

Wilding, C. S., R. K. Butlin, and J. Grahame. 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. J. Evol. Biol. 14:611–619.

Winkler, I. S., and C. Mitter. 2008. The phylogenetic dimension of insect/plant assemblages: a review of recent evidence. Pp. 240–263 *in* K. Tilmon, ed. Ecology and evolution of herbivorous insects. University of California Press, Berkeley, CA.

Wood, T. K., and M. C. Keese. 1990. Host-plant-induced assortative mating in *Enchenopa* treehoppers. Evolution 44:619–628.

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

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

Wright, S. 1943. Isolation by distance. Genetics 28:114–138.

Zhivotovsky, L. A. 1999. Estimating population structure in diploids with multilocus dominant DNA markers. Mol. Ecol. 8:907–913.

Associate Editor: D. Pfennig

Appendix. Distribution of individual repeated outlier loci across population comparisons

	Population comparisons															
	Different host									Same host						
Repeated loci[1]	M_{PA} vs. W_{OS}	M_{PA} vs. W_{QU}	M_{PA} vs. W_{VT}	M_{LI} vs. W_{OS}	M_{LI} vs. W_{QU}	M_{LI} vs. W_{VT}	M_{VT} vs. W_{OS}	M_{VT} vs. W_{QU}	M_{VT} vs. W_{VT}	M_{LI} vs. M_{PA}	M_{PA} vs. M_{VT}	M_{VT} vs. M_{LI}	W_{OS} vs. W_{QU}	W_{QU} vs. W_{VT}	W_{VT} vs. W_{OS}
127^2,295^2,347^2	x	x	x	x	x	x	x	x	x						
265^2		x	x	x											
322^2		x	x					x	x						
66^2, 345^2						x		x							
113^2				x	x		x	x	x						
131^2		x	x	x		x									
50, 233, 271, 262		x	x												
0	x	x	x												
408							x		x						
413, 368								x	x						
437					x	x									
168	x					x									
2, 59, 85, 318					x			x							
282,43														x	x
425											x		x	x	
1											x			x	
17												x			x
13		x	x											x	
16			x											x	
11	x	x	x											x	x
285	x														x
150, 79		x	x											x	x
108	x		x											x	x
48,199	x	x												x	x
445									x				x		
364								x	x						x
435				x	x								x	x	
206					x						x		x		
243					x								x		
249				x	x								x		

Continued

Appendix. Continued

	Population comparisons															
	Different host									Same host						
Repeated loci[1]	M_{PA} vs. W_{OS}	M_{PA} vs. W_{QU}	M_{PA} vs. W_{VT}	M_{LI} vs. W_{OS}	M_{LI} vs. W_{QU}	M_{LI} vs. W_{VT}	M_{VT} vs. W_{OS}	M_{VT} vs. W_{QU}	M_{VT} vs. W_{VT}	M_{LI} vs. M_{PA}	M_{PA} vs. M_{VT}	M_{VT} vs. M_{LI}	W_{OS} vs. W_{QU}	W_{QU} vs. W_{VT}	W_{VT} vs. W_{OS}
20							x				x				
239,187,181	x											x			
433					x					x					
399						x			x	x					
160			x							x					
74,29				x	x						x				
73			x								x		x	x	
366				x		x							x		
219	x	x			x			x		x					
32					x		x					x			
405								x	x			x			
421	x		x	x		x				x	x		x		
6	x	x	x										x		x
376		x	x				x				x	x		x	
12						x								x	x
197	x										x		x		
196	x		x					x					x		x
184	x											x	x		x
183	x								x		x		x		x
44					x	x		x			x				
68					x	x	x				x		x	x	

[1] Numbers refer to individual loci and were determined by starting with the locus from the first primer pair (=D) with the smallest fragment length and counting through the 447 total loci as a function of increasing fragment length for each primer pair and then through successive primer pairs. That is, locus 447 represents the longest fragment scored for primer pair K.

[2] Outlier detected in multiple (two or three) statistically independent different-host comparisons, but not in any same-host comparisons.