doi:10.1111/evo.12209

GENETIC DIVERGENCE ALONG THE SPECIATION CONTINUUM: THE TRANSITION FROM HOST RACE TO SPECIES IN *RHAGOLETIS* (DIPTERA: TEPHRITIDAE)

Thomas H. Q. Powell,[1,2] Glen R. Hood,[1] Mason O. Murphy,[1] Jeffrey S. Heilveil,[3,4] Stewart H. Berlocher,[3] Patrik Nosil,[5] and Jeffrey L. Feder[1,6]

[1]*Department of Biological Sciences, University of Notre Dame, 290C Galvin Life Sciences Building, Notre Dame, Indiana 46556*

[2]*E-mail: thqpowell@gmail.com*

[3]*Department of Entomology, University of Illinois at Urbana-Champaign, 320 Morrill Hall, Urbana, Illinois 61801*

[4]*Current address: Department of Biology, 115 Science I, SUNY College at Oneonta, Oneonta, New York 13820*

[5]*Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, United Kingdom*

[6]*Environmental Change Initiative and Advanced Diagnostics and Therapeutics, University of Notre Dame, Notre Dame, Indiana 46556*

Received August 27, 2012
Accepted June 28, 2013
Data Archived: Dryad doi:10.5061/dryad.mt2hk

Studies of related populations varying in their degrees of reproductive isolation can provide insights into speciation. Here, the transition from partially isolated host races to more fully separated sibling species is investigated by comparing patterns of genetic differentiation between recently evolved (~150 generations) apple and ancestral hawthorn-infesting populations of *Rhagoletis pomonella* to their sister taxon, the undescribed flowering dogwood fly attacking *Cornus florida*. No fixed or diagnostic private alleles differentiating the three populations were found at any of 23 microsatellites and 10 allozymes scored. Nevertheless, allele frequency differences were sufficient across loci for flowering dogwood fly populations from multiple localities to form a diagnosable genotypic cluster distinct from apple and hawthorn flies, indicative of species status. Genome-wide patterns of differentiation were correlated between the host races and species pair comparisons along the majority of chromosomes, suggesting that similar disruptive selection pressures affect most loci. However, differentiation was more pronounced, with some additional regions showing elevated divergence, for the species pair comparison. Our results imply that *Rhagoletis* sibling species such as the flowering dogwood fly represent host races writ large, with the transition to species status primarily resulting from increased divergence of the same regions separating apple and hawthorn flies.

KEY WORDS: **Divergence hitchhiking, genomic divergence, genome hitchhiking, latitudinal cline, speciation with gene flow.**

> *"...these...forms may still be only well-marked varieties;...but we have only to suppose that the steps in the process of modification to be more numerous or greater in amount, to convert these ... forms into well-defined species."*
> Darwin (1859), On the Origins of Species, p. 120

Much progress has been made in understanding speciation (Coyne and Orr 2004). Nevertheless, fundamental questions remain, especially when speciation occurs without geographic barriers to gene flow (Bolnick and Fitzpatrick 2007). Although insights have come into the mechanisms underlying reproductive isolation, it is still unclear when during the divergence process populations may be considered species, as opposed to varieties or races (Hey and Pihno 2012). Although advances have been made in resolving the genetic basis for reproductive isolation,

© 2013 The Author(s). *Evolution* © 2013 The Society for the Study of Evolution.

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

the importance of standing variation versus new mutations and of genome structure in facilitating speciation remain to be resolved (Nosil and Feder 2012; Feder et al. 2013).

Analyses of genome-wide patterns of differentiation between taxa represent a powerful approach to addressing these questions. By comparing related populations at different stages of divergence, inferences can be made concerning how genetic differentiation accumulates as speciation proceeds. Although these populations do not represent a single evolutionary progression, it is reasonable to assume that most taxa will pass through these representative stages as they diverge. Many examinations of the "speciation continuum" have focused on metrics of overall genetic divergence rather than the genome-wide distribution of differentiation (Ayala et al. 1974; Avise and Smith 1977; Zimmerman et al. 1978; Nosil and Sandoval 2008; Seehausen 2008; Berner et al. 2009; Hendry et al. 2009; Peccoud et al. 2009; Merrill et al. 2011). Here, we investigate patterns of differentiation in the genome of *Rhagoletis pomonella* (Walsh) (Diptera: Tephritidae) flies as they transition from partially reproductively isolated host races to more fully isolated sibling species. Specifically, we compare patterns of microsatellite and allozyme differentiation between apple-infesting (*Malus domestica* Borkh.) and hawthorn-infesting (*Crataegus* spp. L.) host races of *R. pomonella* and the flowering dogwood fly (dogwood fly hereafter), the sister taxon of *R. pomonella* infesting the fruit of flowering dogwood (*Cornus florida* L.). Apple, hawthorn, and dogwood flies are closely related members of the *R. pomonella* species complex (Berlocher 1999, 2000), a model for speciation with gene flow for phytophagous insects (Funk et al. 2002). The known natural history and biology of these flies (see below) make them ideal for examining the nature of species, the genetic source of variation fueling differentiation, and the role of genome structure in speciation.

WHAT IS A SPECIES?

What constitutes a species has long been a topic of debate (Mayden 1997; Hey 2001; De Queiroz 2007; Hausdorf 2011). If genetic differentiation accumulates continuously during speciation with gene flow, then is there a threshold between ecological races and species and how do existing species concepts apply to this question? Here, we attempt to resolve the status of apple and hawthorn versus dogwood flies based on patterns of genomic differentiation in reference to two species concepts: (1) the biological species concept (BSC) that defines species based on reproductive isolation (Mayr 1942, 1963, 2001; Coyne and Orr 2004); and (2) genic species concepts that consider reproductive isolation to be a characteristic of specific loci (Mallet 1995; Feder 1998; Wu 2001). Originally, the BSC considered reproductive isolation to be a characteristic of the genome due to extensive epistatic coadaptation among loci precluding gene flow genome-wide (Mayr 1969). This view eventually softened such that Mayr (2001) later consider species to represent reproductive communities that were not necessarily separated by complete barriers to gene flow across the genome. This version of the BSC therefore converged to some degree with more recent genic species concepts. In practice, both the BSC and genic species concepts are usually implemented by determining whether populations form recognizable genotypic clusters in sympatry (Mallet 2008). Here, we examine whether such clusters can be distinguished using allele frequency differences, as opposed to private or fixed variants, and discuss how clustering patterns may relate to host race versus species status in *Rhagoletis*.

WHAT IS THE GENETIC SOURCE OF VARIATION FUELING SPECIATION?

Whether standing variation or novel mutations are the primary source of variation fueling divergence can have important consequences for the tempo of speciation with gene flow. Standing variation may allow populations to rapidly colonize and adapt to novel habitats, spawning new incipient species (Barrett and Schluter 2008). In contrast, if de novo mutations are required, then ecological divergence may stall until the necessary mutations arise and establish in populations. When populations are small and survival in a new environment requires adaptation along multiple dimensions, the delay can limit diversification, especially when a degree of hard selection is acting. When multiple copies of a preadapted allele exist in the ancestral population, however, there is a higher probability that the allele will establish in the novel habitat and not be lost stochastically compared to a new mutation. Although theory predicts an important role for standing variation in initiating speciation with gene flow (Hermisson and Pennings 2005), empirical data on the issue are generally lacking. In addition, it is not clear why populations may possess and maintain stores of standing variation adaptive to novel environments (Schluter and Conte 2009). Here, we address the question whether derived populations of the *R. pomonella* infesting apple and dogwood are largely defined by shifts in the frequencies of preexisting alleles in the ancestral hawthorn population or by novel mutations.

WHAT ROLE DOES GENOME STRUCTURE PLAY IN SPECIATION WITH GENE FLOW?

Genome structure may play an important role in speciation with gene flow by mitigating the antagonism between selection and recombination (Felsenstein 1981). One process by which this could happen has been termed "divergence hitchhiking" (DH) and involves local reductions in effective migration rates (m_e) between populations in gene regions physically linked to loci under divergent selection (Via and West 2008; Via 2009, 2012). Recent theory has suggested, however, that the conditions for DH to be

effective in aiding the maintenance of existing differentiation and the establishment of new mutations may be generally restricted to short recombination distances (e.g., $\leq$1 cM) surrounding sites under strong selection (e.g., $s = 0.5$) (Feder and Nosil 2010; Feder et al. 2012a). These findings are consistent with classic work on panmictic populations implying that hitchhiking effects are essentially nonexistent at genetic distances $r > s/10$, and that most strong hitchhiking occurs for $r \sim s/100$ (Maynard-Smith and Haigh 1974; Wiehe and Stephan 1993).

Another process termed "genome hitchhiking" (GH) may also lessen the antagonism between selection and recombination and facilitate speciation (Feder et al. 2012a). In the case of GH, average m_e is reduced globally across the genome by the total of divergent selection acting on all loci throughout the genome, rather than only locally by individual genes. As a result, differentiation is not restricted to isolated islands and divergence can accumulate across the genome (Michel et al. 2010). It is important to note that heterogeneous patterns of differentiation are still expected under GH due to variation in selection strength, recombination rates, and local DH effects; however, patterns of divergence will become more homogenous as speciation proceeds and reductions in m_e become greater and more uniform across the genome.

DH and GH make different and potentially testable predictions concerning patterns of divergence during speciation (Feder et al. 2012a, 2013). These predictions, however, represent different ends of a spectrum, and it is likely that both processes contribute to speciation. The key question is their relative importance. A recent model of speciation with gene flow has proposed that populations may go through four sequential stages in which direct selection on loci, DH, and GH assume differing relative importance in facilitating divergence (Feder et al. 2012b, 2013). Significant allele frequency differences have been observed between the apple and hawthorn-infesting host races of *R. pomonella* for a number of markers across the genome (Michel et al. 2010), implying that GH may be important, even in very recent cases of divergence ($\sim$150 generations). Here, we investigate whether genomic differentiation is widespread and accentuated in the flowering dogwood fly, the sister taxon to apple and hawthorn flies, in a manner consistent with GH facilitating the transition from host races to sibling species.

THE R. POMONELLA SIBLING SPECIES COMPLEX

The *R. pomonella* species complex comprises a number of populations spanning varying stages of speciation from partially reproductively isolated host races to reciprocally monophyletic species (Bush 1969; Berlocher et al. 1993; Berlocher 2000; Xie et al. 2008). The outstanding feature of the complex is the contrast displayed between the species' overlapping geographic distributions (most are broadly sympatric in eastern North America) and close morphological similarity compared to their distinct host associations (each species is a specialist on a different, nonoverlapping set of plants). This led Bush (1966) to propose that the *R. pomonella* complex radiated via a series of sympatric host shifts. The recent shift of *R. pomonella* from its native host hawthorn to introduced apple is a classic example of sympatric host race formation. Although originally reported in the Hudson Valley region of New York in the 1860s (Walsh 1867), the derived apple race is now present throughout the northeastern and midwestern United States (Bush 1966), where it co-occurs and experiences gene flow from local, native hawthorn-infesting populations. Thus, the apple race cannot be characterized as having either a single derivation versus multiple origins, as it has a complex history likely affected by elements of both processes. Other hypothesized cases of sympatric speciation include the closely related sibling species *R. mendax* (host: blueberries), *R. zephyria* (host: snowberries), and the dogwood fly. Mitochondrial sequence data suggest that together with *R. pomonella*, these three taxa form a clade of in-group species rooted by a hawthorn-infesting population of *R. pomonella* nr. from the Altiplano region of Mexico (Xie et al. 2007, 2008; Fig. 1). Nuclear sequence data imply that over the last 1.5 Ma, several cycles of allopatry and secondary contact have occurred between the Mexican and U.S. hawthorn fly populations. One result of this introgression was the establishment of adaptive latitudinal clines for diapause-related traits in the United States (Feder et al. 2003a). This standing life-history variation may have subsequently played a role in the formation of the in-group *R. pomonella* taxa by facilitating host shifts onto phenologically distinct novel hosts (Feder et al. 2005).

Here, we concentrate on the undescribed dogwood fly that allozyme studies suggested is the sister taxon to *R. pomonella* (Berlocher 1999, 2000). The current range of the dogwood fly is contained entirely within that of hawthorn-infesting populations of *R. pomonella* in the eastern United States (Fig. S1). The dogwood fly's host, *C. florida,* grows commonly throughout much of eastern North American, and genetic diversity within *C. florida* shows a continuous pattern of variation with little evidence for previous isolation (Hadziabdic et al. 2010), implying that the dogwood fly has not been spatially subdivided in the past. The dogwood fly is morphologically indistinguishable from *R. pomonella* (Bush 1966), and shows little evidence for postzygotic or nonhost related prezygotic reproductive isolation (Smith 1986). Rather, the dogwood fly is ecologically isolated from *R. pomonella* by the same types of host-specific mating (Linn et al. 2005) and diapause (Dambroski and Feder 2007) adaptations separating the apple and hawthorn fly races (Feder et al. 1994; Linn et al. 2003). The dogwood fly therefore represents a logical focus for investigation into the nature of species and genomic patterns of differentiation, as all evidence implies that

Table 1. Host plant origin, location (latitude [N] and longitude [W] in degrees), data source (provided in the footnotes), year sampled, and number of individuals genotyped (*n*) for both microsatellite (μ*sat*) and allozyme (*allo*) genotyping.

Host plant	Location	#	Lat., Long.	Source μ*sat*	Year μ*sat*	*n* μ*sat*	Source *allo*	Year *allo*	*n* *allo*	Year μ*sat*
Flowering dogwood	Cassopolis, Cass Co., MI	3	42.0, 85.97	New	2006	36				2006
	Knollwood, St. Joseph Co., IN	4	41.75, 86.22	New	2010	40				2010
	I-57 rest stop, Union Co., IL	6	37.47, 89.13	New	2005	57	a	1989	29	2005
	Reelfoot Lake, Lake Co., TN	7	36.45, 89.3	New	2010	40				2010
	Kisatchie , Natchitoches Pr., LA	8	31.48, 93.1	New	2007	47				2007
	SFA For., Nacogdoches Co., TX	9	31.31, 94.46	New	2007	39	a	1989	35	2007
	Fairfield, Wayne Co., IL		38.43, 88.36				a	1981	76	
	Carbondale, Jackson Co., IL		37.74, 89.22				a	1989	35	
	W. Patman Lake, Bowie Co., TX		32.33, 89.3				a	1989	31	
	Bogalusa, Washington Pr., LA		30.78, 89.85				a	1989	31	
	LSU, E. Baton Rouge Pr., LA		30.40, 91.15				a	1989	35	
Hawthorn	Grant, Newaygo Co., MI	1	43.35, 85.9	b	2002	48	d	1987	414	2002
	Fennville, Allegan Co., MI	2	42.6, 86.15	b	2002	96	e	1987	38	2002
	Dowagiac, Cass Co., MI	3	41.88, 86.23	b	2002	49	e	1987	32	2002
	Urbana, Champaign Co., IL	5	40.08, 88.23	b	2002	46	e	1986	60	2002
	New Madrid Co., MO	7	36.53, 89.43	c	2008	41	f	1994	15	2008
	SFA For., Nacogdoches Co., TX	9	31.31, 94.46	New	2006	94	f	1985	60	2006
	Jackson Farm, Polk Co., TX	10	30.51, 95.00	New	2006	93	f	1989	34	2006
Apple	Grant, Newaygo Co., MI	1	43.35, 85.9	b	2002	47	d	1987	672	2002
	Fennville, Allegan Co., MI	2	42.6, 86.15	b	2002	96	e	1987	30	2002
	Dowagiac, Cass Co., MI	3	41.88, 86.23	b	2002	47	e	1987	31	2002
	Urbana, Champaign Co., IL	5	40.08, 88.23	b	2002	48	e	1986	60	2002

a = Berlocher (1999); b = Michel et al. (2010); c = Cha et al. (2012); d = Feder et al. (1990); e = Feder and Bush (1989); f = Berlcoher and McPheron (1996).
IN = Indiana; IL = Illinois; LA = Louisiana; MI = Michigan; MO = Missouri; TN = Tennessee; TX = Texas.
[#]Site designations shown in Figs. S1 and 4 maps (microsatellite populations only).

it represents the next stage of divergence beyond the apple and hawthorn host races.

Materials and Methods

SAMPLING OF FLIES

Flies were collected in the field as larvae in infested flowering dogwood, hawthorn, and apple fruit from 17 different sites in the eastern United States (Table 1; Fig. S1). These encompassed the range of latitudinal overlap between the apple and hawthorn host races of *R. pomonella* and the dogwood fly. Infested fruit were transported to the laboratory where larvae were reared to adulthood using standard *Rhagoletis* husbandry techniques (Neilson and McAllan 1965). Upon eclosion, adult flies were stored at −80°C.

MICROSATELLITES AND ALLOZYMES

Twenty-three microsatellite loci were scored for a total of 967 flies representing seven hawthorn- and four apple-infesting populations of *R. pomonella*, and six dogwood fly populations (Table 1; Fig. S1). The microsatellite markers were developed by Velez et al. (2006) and map to five of the six *R. pomonella* chromosomes (the sixth dot chromosome currently lacks markers). New microsatellite data for the dogwood fly and two southern hawthorn populations were combined with previously published data for apple and northern hawthorn populations from Michel et al. (2010) and Cha et al. (2012). See Supporting Information for genotyping methods and details concerning data compilation between studies. Previously published allozyme data were combined with the microsatellite data for some analyses (see Table 1 for data sources and Supporting Information for details). A majority of the collecting sites scored for allozymes represent the same locations scored for microsatellites (Table 1), although the collection years differed. These two classes of markers were chosen to allow direct comparisons to the previously published results of Michel et al. (2010). In addition, the high mutation rate of microsatellite markers may be particularly useful in studies of very recent or on-going divergence, as single nucleotide polymorphism (SNP) differences may not have had sufficient time to accumulate.

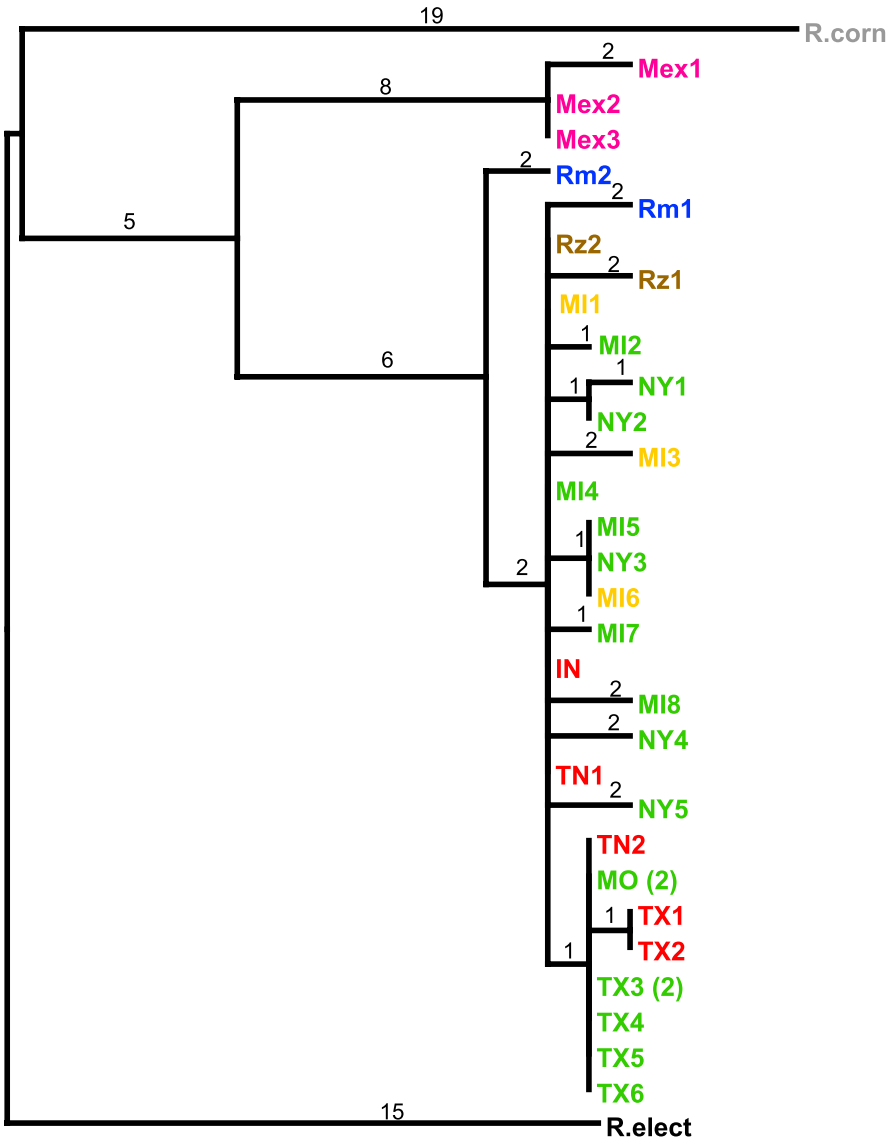

Figure 1. Mitochondrial DNA haplotype tree for the *R. pomonella* species complex, including the apple (yellow) and hawthorn host races (green), the undescribed flowering dogwood fly (red), *R. zephyira* (brown), *R. mendax* (blue), and *R. pomonella* nr. the Mexican highland hawthorn fly (pink), with the outgroup taxa *R. cornivora* (gray) and *R. electromorph* (black). Names of *R. pomonella* race and flowering dogwood fly sequences represent the state in the United States in which a fly was collected (IN = Indiana, MI = Michigan, MO = Missouri, NY = New York, TN = Tennessee, and TX = Texas). Branch labels indicate number of mutational steps. Details are given in Table S2.

mtDNA

A maximum parsimony mtDNA gene tree was constructed to quantify genetic differentiation within the *R. pomonella* species complex based on a 545 bp fragment of the cytochrome oxidase subunit II (COII) gene using PHYLIP 3.69 (Felsenstein 2005). Data for the mtDNA network came from both newly generated and previously published sequences (Smith and Bush 1997; Feder et al. 2003a). See Supporting Information for details on samples, sequencing, and alignment.

ESTIMATING MIGRATION BETWEEN HAWTHORN AND DOGWOOD FLIES

Gene flow was estimated between hawthorn and dogwood-infesting fly populations based on the identification of putative migrants and hybrids from microsatellite data using the program STRUCTURE 2.3.3 (Pritchard et al. 2000). We concentrated on these two taxa because (1) hawthorn flies are the likely ancestral population of dogwood flies (Berlocher 2000), (2) hawthorn and dogwood flies overlap at least partially in their eclosion times and

seasonal distributions as adults in the field, whereas apple and dogwood flies generally do not (Berlocher 1999), strongly limiting the potential for direct gene flow in the latter case, and (3) while previous mark recapture studies have estimated the gross migration rate (~4–6% per generation) between apple and hawthorn flies (Feder et al. 1994), observed frequency differences between the races are insufficient to reliably identify putative migrants between apple and hawthorn flies. Our strategy involved using STRUCTURE to infer the number of differentiated fly populations existing on hawthorn and dogwood host plants. These results were then used in the program as a priori population assignments to identify putative parental migrant, F1 hybrid, and later generation backcross genotypes within populations. An estimate of interhost gene flow was calculated as the frequency of individuals possessing a non-natal parental genotype plus half the frequency of individuals classified as F1 hybrids (highest posterior probability of having non-natal ancestry one generation back). For full details of the analysis see the Supporting Information. It is important to note that STRUCTURE was used in a narrow context to detect migrants and hybrids between well-differentiated clusters. Moreover, analyses were confined to populations from the same geographic region to minimizing the confounding effects of latitudinal clines. The complex clinal nature of genetic variation within the *R. pomonella* races (Feder and Bush 1989; Feder et al. 2005; Michel et al. 2010) strongly limits the usefulness of an overall STRUCTURE analysis including all populations. However, for completeness, such an analysis is presented in the Supporting Information. Discussion of the clustering of populations in the Results and Discussion sections do not refer to results from STRUCTURE, but rather to the genetic distance network described below.

ANALYSIS OF GENETIC DIFFERENTIATION

To simplify analysis of the microsatellite data, we collapsed the variation present at each locus down to two major allele classes, as previously described in Michel et al. (2010) for the apple and hawthorn host races. Our work on *R. pomonella* has indicated that major allele classes exist for nuclear genes defined by patterns of linkage disequilibrium among loci and latitudinal variation due, in part, to the presence of inversion polymorphisms and adaptive latitudinal clines for life-history traits generated by past introgression between U.S. and Mexican populations of hawthorn flies (Feder et al. 2003a, 2005). Major allele classes were identified by testing either exhaustive searches or 1,000,000 random combinations of alleles at a locus to identify the combination of variants that maximized latitudinal variation and linkage disequilibrium to flanking markers in the genome. See Michel et al. (2010) for analytical details. Resulting allele pools are described in Table S3.

To graphically depict the population structure of apple, hawthorn, and dogwood flies, a neighbor-joining network based on Nei's genetic distance D (Nei 1972) was constructed for the microsatellites using PowerMarker (Liu and Muse 2005) and PHYLIP 3.69 (Felsenstein 2005). The major results did not differ between networks constructed using either raw microsatellite allele frequencies or those pooled into the two major classes (Fig. S2), so we present the pooled network here. Bootstrap support for the network was determined using 10,000 replicates across loci. In addition, individual microsatellite and allozyme loci were analyzed for host-associated differentiation and latitudinal variation by (1) Fisher exact tests, (2) F-statistics, (3) linear regression, and (4) generalized linear modeling (GLM). For the Fisher exact tests and F-statistics, the analyses were performed for pairs of populations infesting different host plants that co-occurred with one another at collecting sites or were in close geographic proximate (<14 km) (see Table 1 for site designations; we refer to these as "local sites" hereafter). The GLM analysis was performed in R, with allele frequencies modeled as quasibinomial variables, latitude as a continuous factor, and host plant as a discrete factor. Host, latitude, and host × latitude interaction effects were tested for statistical significance by F-tests. To assess whether patterns of genome differentiation were concordant between hawthorn/dogwood flies compared to hawthorn/apple flies, we performed regression analyses between locus-specific F_{ST} values, calculated as the means across the individual paired population F_{ST} values for each of the local sites within a host comparison (Table 1).

SIMULATION MODELING

Based on the results of the analyses described above, we conducted computer simulations to investigate the range of m_e and selection coefficients that could generate the patterns of differentiation observed among apple, hawthorn, and dogwood flies. We concentrated our analysis on a subset of five markers each on chromosomes 1 and 2 that captured the range and major topological features of genomic divergence seen among these flies. The simulations were based on the two-deme model described by Feder et al. (2012a). Full modeling details are provided in the Supporting Information.

Results

MITOCHONDRIAL SEQUENCE VARIATION

The single, most parsimonious gene tree for mtDNA haplotypes is shown in Figure 1 and displayed no homoplasy. There was no evidence for host-related mtDNA differentiation among apple, hawthorn, and dogwood flies; no taxon formed a distinct matrilineage (Fig. 1).

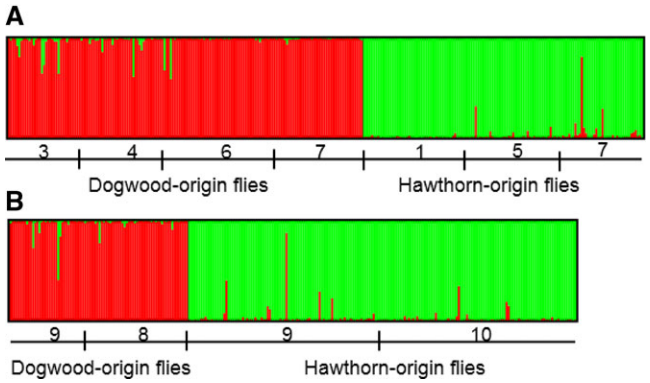

Figure 2. STRUCTURE output for $K = 2$ analyses of (A) northern hawthorn race and northern dogwood flies and (B) southern hawthorn race and dogwood flies with a priori population designations established to detect migrants and hybrids. Each vertical bar represents a sampled individual's membership coefficient to green: hawthorn race or red: dogwood fly populations. Tick marks on horizontal bar below plots represent sampled populations designated by numbers corresponding to Table 1.

There was, however, evidence for geographic variation. A southern haplotype differing by one derived substitution was present in all seven hawthorn flies sequenced from Texas and Missouri, as well as a majority of dogwood flies (3/4) sampled from southern states. The southern haplotype was absent from three apple flies, 10 hawthorn flies, and one dogwood fly sequenced from the northern states of Michigan, Indiana, and New York. These results indicated that the dogwood fly does not represent a long-standing cryptic species and confirmed its sister taxon status to the apple and hawthorn races of *R. pomonella*. Moreover, the lack of a distinct mtDNA clade defining these host-associated populations and the shared northern and southern haplotypes imply that lineage sorting is not complete and/or gene flow is ongoing among apple, hawthorn, and dogwood flies.

INTERHOST GENE FLOW

STRUCTURE revealed evidence for ongoing migration and gene flow between hawthorn and dogwood fly populations. Initial STRUCTURE runs supported $K = 3$ subpopulations representing northern hawthorn, southern hawthorn, and dogwood flies (see Supporting Information for details). We therefore preformed separate runs of northern and southern hawthorn versus dogwood populations in the same region to test for gene flow (Fig. 2). Besides conforming to the results of the initial "blind" structure runs, this approach helped mitigate the confounding effects of clinal variation in the system. Classification of flies as parental migrants and F1 hybrids was based on their higher posterior probability for being direct migrants or having migrant ancestry one generation back, respectively, than being pure members of their natal host cluster. In the combined northern hawthorn fly population ($n = 135$), one individual possessed a genotype consistent with it being a dogwood fly and two had genotypes consistent with them being F1 hybrids (Fig. 2A), resulting in a gene flow estimate of 1.48% per generation from the dogwood into the hawthorn population. In the northern dogwood fly population ($n = 171$), no individual had a hawthorn fly genotype and four flies had genotypes consistent with them being F1 hybrids, resulting in a gene flow estimate of 1.17% (Fig. 2A). Similarly, in the South, gene flow from the hawthorn into the dogwood fly population ($n = 39$) was estimated to be 1.28% per generation (one F1 hybrid) and 0.54% in the reverse direction (one F1 hybrid; $n = 93$) (Fig. 2B). The mean estimated gene flow rate between hawthorn and dogwood flies in the North and South was therefore 1.12% $\pm$ 0.203 SE ($n = 4$). Putative F2 backcross genotypes were also detected (two in northern dogwood and one in southern hawthorn populations), implying that F1 hybrids were not completely sterile and later generation hybrids not completely inviable. Based on the ratio of the number of F2 backcross progeny (3) to F1 hybrids (8), the relative fitness of F1 hybrids would be 37.5% that of resident parental genotypes. Although this estimate must be viewed cautiously due to small sample size, it does provide a rough gauge of hybrid fitness.

THE TAXONOMIC STATUS OF DOGWOOD FLIES

The neighbor-joining network for microsatellites supported the species-level distinction between dogwood flies and *R. pomonella* (Fig. 3). We emphasize that this figure represents a network reflecting current genetic distances among populations. It should not be interpreted as analogous to a phylogram indicating systematic relationships. The six dogwood fly populations surveyed across 1400 km formed a discrete cluster with 96.4% bootstrap support from the 11 apple and hawthorn fly populations analyzed (Fig. 3). Dogwood flies showed no fixed genetic differences and possessed no private allele distinguishing them from *R. pomonella* (see Table S1). Rare variants private to one or a few sites were found, but none was present across all sites in either taxon and none had a frequency greater than 3% at any site. However, dogwood flies displayed significant host-related frequency differences for 19 of 23 microsatellites and four of 10 allozymes across all dogwood and hawthorn collecting sites (Table 2; Figs. S3–S7). Fisher's exact tests for local allelic differentiation between dogwood and hawthorn sites revealed statistical differentiation for 22 of 23 microsatellites for at least one of the paired sites (Figs. S3–S7). Thus, allele frequency differences alone were sufficient to recognize dogwood flies as a diagnosable genetic cluster from *R. pomonella*, representing a sibling species.

Dogwood flies also showed evidence of significant latitudinal clines for 10 microsatellites (Table 2; Figs. S3–S7) and two allozymes (*Acon-2* and *Had*) (see Berlocher 1999). As a result, a

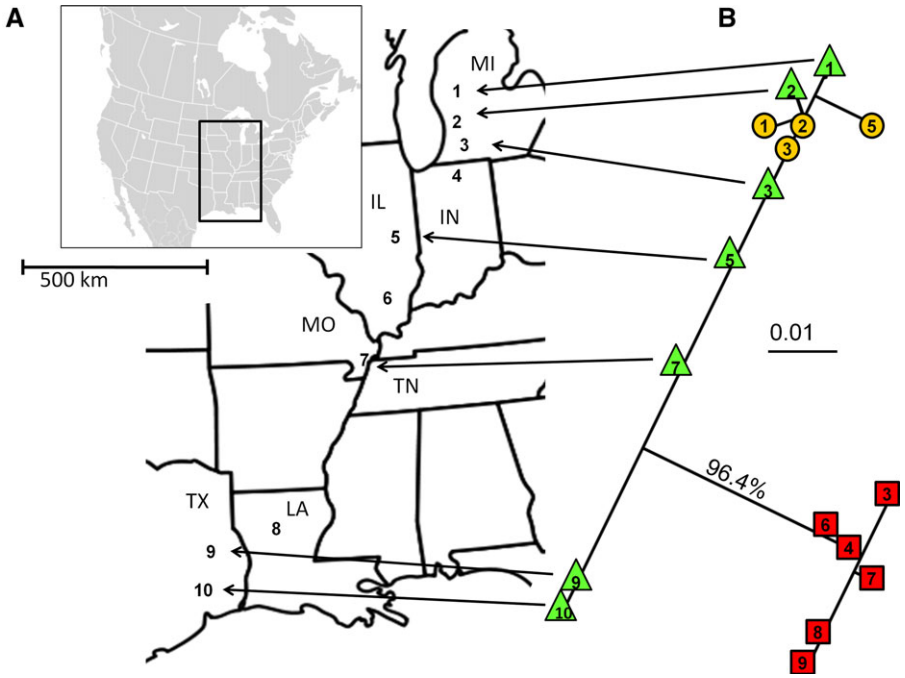

Figure 3. Microsatellite genetic distance network for apple (yellow circle) and hawthorn (green triangle) host races of *R. pomonella* and the dogwood fly (red square). (A) Map of collecting sites showing the geographic ranges of the taxa in the eastern United States. See Table 1 for full description of sites. (B) Neighbor-joining tree based on Nei's *D* (1972) estimated from microsatellite allele frequencies pooled into two classes for each of the 23 loci scored in the study. Bootstrap support comes from 10,000 replicates across loci. Note that the apple and hawthorn fly populations, while diverged from each other at local sites, do not form discrete genetic clusters across their respective ranges, whereas *R. pomonella* and the dogwood fly do, which may represent a distinction between host races and sibling species.

pattern of isolation by distance was apparent from north to south among flowering dogwood fly populations (Fig. 3). However, the latitudinal clines were generally of lesser magnitude for dogwood than hawthorn flies. Given that dogwood flies likely arose from a sympatric host shift from hawthorn, it would appear that like the apple race, dogwood flies were extracted from standing clinal variation present in the ancestral hawthorn population. However, we cannot completely rule out that the microsatellites and allozymes are linked to derived alleles at nearby loci unique to dogwood flies or the lack of high-frequency private variants is due to convergent mutation (homoplasy).

PATTERNS OF GENOME-WIDE DIFFERENTIATION

To visualize patterns of genetic divergence across the genome, we plotted log-transformed F_{ST} values for loci in relation to their relative map positions along chromosomes (Fig. 4). Due to the presence of extensive inversion polymorphism in *R. pomonella* (Feder et al. 2003b; Michel et al. 2010), there is no universal linear order of genes along chromosomes (Roethele et al. 1997; Feder et al. 2003b). Nevertheless, an evolutionary linkage map of relative gene order can be constructed based on mean recombination rates between loci averaged across crosses (Michel et al. 2010). Genetic divergence was generally higher between dogwood and hawthorn flies than it was between the host races (mean $F_{ST} = 0.067 \pm 0.012$ SE vs 0.016 ± 0.003 SE, respectively; $t = 11.27$, $P < 0.001$, paired *t*-test, 32 df; Fig. 4). However, F_{ST} values were still highly heterogeneous across the genome for both comparisons, ranging from 0.274 to 0.007 and from 0.068 to 0.0002 for the species and host race comparisons, respectively (Fig. 4). The heterogeneity was due, in part, to inverted regions of the genome displaying elevated divergence. Mean F_{ST} values were about three times higher on average for markers in putative inverted versus collinear gene regions for both apple/hawthorn ($F_{1, 31} = 4.98$, $P = 0.033$; F_{ST} collinear regions $= 0.005 \pm 0.001$ SE, $n = 13$; F_{ST} inverted regions $= 0.017 \pm 0.004$ SE, $n = 20$) and dogwood/hawthorn fly comparisons ($F_{1, 31} = 6.68$, $P = 0.0147$; F_{ST} collinear $= 0.031 \pm 0.007$ SE, $n = 13$; F_{ST} inverted $= 0.090 \pm 0.018$ SE, $n = 20$). It is possible that some of the higher differentiation in the species-level comparison is driven by more of the apple/hawthorn paired sites being strictly sympatric compared to the dogwood/hawthorn sites. However, the tightly sympatric dogwood/hawthorn site 9 does not show diminished differentiation and the high level of dogwood/hawthorn divergence was consistent across sites (Table S4). In addition, although not all flies sampled in the study were from strictly sympatric sites, it is likely that the alternate hawthorn or dogwood host was

Table 2. Results for generalized linear models of allele frequency as a quasibinomial variable as a function of host plant, latitude, and the interaction between host plant and latitude for 23 microsatellite and 10 allozyme loci. Allele frequencies for dogwood flies and apple race *R. pomonella* flies were compared separately to the hawthorn race of *R. pomonella*. Microsatellite loci have the prefix p. Allozyme designations are given in the Supporting Information Methods.

		Chromosome 1									
Dogwood versus Hawthorn		p3	p4	p37	p71	p75	Aat-2	Ak	Dia-2	Idh	Pgm
	df	12	12	12	12	8	13	13	13	13	13
Host	*F*	16.09	2.47	129.7	51.01	19.14	631.9	0.52	267.2	0.13	0.28
	P	**	–	****	****	**	****	–	****	–	–
Latitude	*F*	13.54	8.81	80.34	91.89	105.8	421.0	0.44	170.6	0.93	0.08
	P	**	*	****	****	***	****	–	****	–	–
Host:Lat	*F*	125.4	14.02	0.59	3.18	2.16	1.42	0.10	2.10	0.01	0.33
	P	****	**	–	–	–	–	–	–	–	–
Apple versus Hawthorn											
	df	10	10	10	10	10	10	10	10	10	10
Host	*F*	0.66	0.50	4.20	16.10	13.60	4.83	0	2.42	0.26	1.85
	P	–	–	–	**	**	–	–	–	–	–
Latitude	*F*	104.6	20.17	58.70	95.76	129.9	114.5	1.75	104.2	1.69	0.16
	P	****	**	***	****	****	****	–	****	–	–
Host:Lat	*F*	2.62	0.034	0.11	0.23	0.13	8.55	1.69	3.15	0.11	1.98
	P	–	–	–	–	–	*	–	–	–	–

		Chromosome 2							Chromosome 3		
Dogwood versus Hawthorn		p46	p70	p73	p54	Acon	Mpi	Me	p7	p16	p23
	df	12	11	11	8	13	13	5	12	12	12
Host	*F*	17.13	55.11	12.79	58.35	19.47	0.81	2.73	46.71	56.35	250.5
	P	**	****	**	***	**	–	–	****	****	****
Latitude	*F*	1.78	120.6	39.59	11.50	30.30	95.34	26.0	89.66	30.80	72.01
	P	–	****	***	*	***	****	*	****	***	****
Host:Lat	*F*	62.36	0.31	7.23	100.9	5.60	2.95	8.64	0.04	0.01	24.50
	P	****	–	*	***	*	–	–	–	–	**
Apple versus Hawthorn											
	df	10	8	8	9	10	10	7	10	10	10
Host	*F*	25.68	61.24	0.10	30.92	10.52	11.77	23.9	20.98	13.01	42.77
	P	**	***	–	**	*	**	**	**	**	***
Latitude	*F*	17.50	139.3	16.13	36.68	0.20	94.81	18.6	101.1	25.30	56.08
	P	**	****	*	***	–	****	*	****	**	***
Host:Lat	*F*	18.10	6.21	3.41	87.14	0.03	0.06	10.0	19.37	1.77	0.43
	P	**	–	–	****	–	–	*	**	–	–

		Chromosome 3			Chromosome 4					
Dogwood versus Hawthorn		p66	p80	Had	p11	p25	p29	p50	p60	Pgi
	df	12	12	13	12	12	12	12	12	13
Host	*F*	2.26	75.94	0.81	23.66	28.86	4.04	33.85	12.60	8.74
	P	–	****	–	**	***	–	***	**	*
Latitude	*F*	6.49	84.94	95.34	18.89	10.79	9.14	0.43	1.31	5.97
	P	*	****	****	**	**	*	–	–	*
Host:Lat	*F*	0.46	0.99	2.95	0.59	26.97	96.81	2.84	0.03	0.39
	P	–	–	–	–	***	****	–	–	–
Apple versus Hawthorn										
	df	10	10	10	10	9	10	10	10	10
Host	*F*	2.13	29.24	11.77	2.25	0.020	42.04	2.67	0.93	0.17
	P	–	**	**	–	–	***	–	–	–
Latitude	*F*	5.80	69.20	94.81	13.74	22.12	76.50	1.36	0.95	5.10
	P	*	****	****	**	**	****	–	–	–
Host:Lat	*F*	0.04	8.26	0.06	8.35	1.98	49.44	0.60	1.84	4.10
	P	–	*	–	*	–	***	–	–	–

(Continued)

Table 2. Continued.

		Chromosome 5			
Dogwood versus hawthorn		p9	p18	p27	p5
	df	12	12	12	8
Host	*F*	2.59	4.68	0.54	15.84
	P	–	–	–	*
Latitude	*F*	0.29	12.50	20.95	34.98
	P	–	**	**	**
Host:Lat	*F*	0.76	0.25	11.86	11.0
	P	–	–	**	*
Apple versus hawthorn					
	df	10	10	10	10
Host	*F*	1.65	14.08	0.27	13.38
	P	–	**	–	**
Latitude	*F*	0.00	16.61	1.93	64.45
	P	–	**	–	****
Host:Lat	*F*	0.02	4.95	0.63	0.037
	P	–	–	–	–

Significance of factors was determined by *F*-test (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$) .

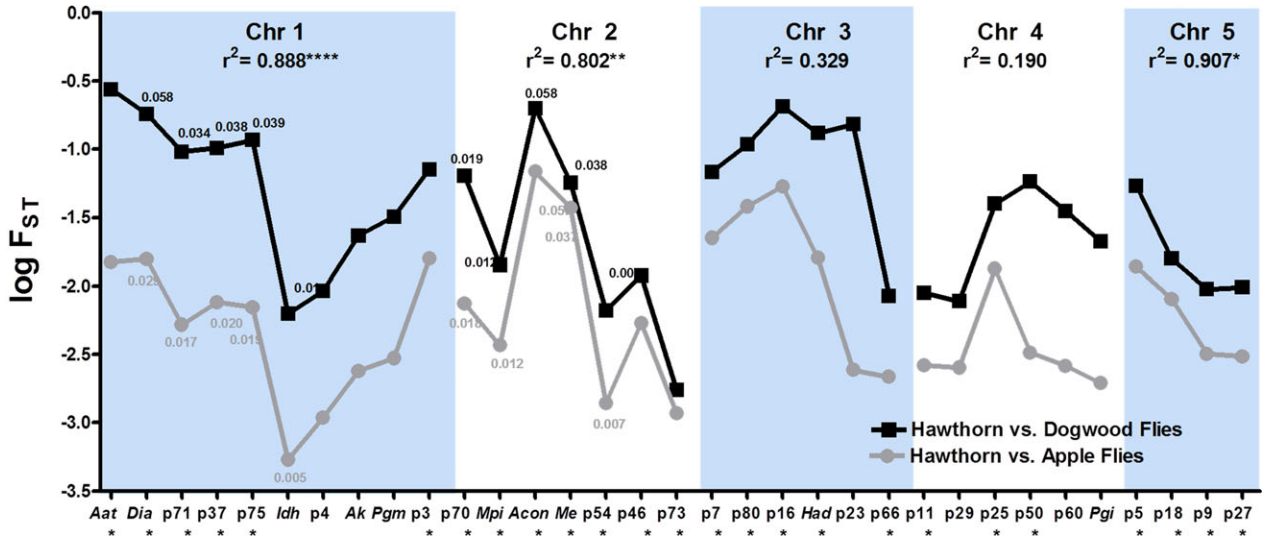

Figure 4. Mean F_{ST} values calculated between apple and hawthorn host races of *R. pomonella* (gray circles and lines) and between hawthorn and dogwood flies (black squares and lines) across local sites plotted against the relative order of loci in the genome, as determined from a series of test crosses in the apple race (Feder et al. 2003b; Michel et al. 2010). Also given are the variance explained (r^2 value) by linear regressions of F_{ST} values for apple/hawthorn versus hawthorn/dogwood flies for loci on chromosomes 1–5 (*$P \geq 0.05$, **$P \geq 0.01$, ****$P \geq 0.0001$). Numbers associated with markers *Dia* through *Idh* on chromosome 1 and p70 through p54 on chromosome 2 represent the *s* values estimated from computer simulations needed to generate the observed levels of differentiation between apple and hawthorn flies (gray numbers) and hawthorn and dogwood flies (black numbers). Asterisks below loci designated markers that displayed a significant relationship with diapause in either selection or eclosion time experiments (see Michel et al. 2010 for details).

in close proximity somewhere nearby at all nonsympatric sites sampled.

The overall pattern of genetic divergence between *R. pomonella* and the dogwood fly was significantly related to that between apple and hawthorn flies ($r^2 = 0.550$, $P = 0.0008$, 32 df). The relationship was not due to variation in allele richness or heterozygosity among loci ($r^2 < 0.05$, $P > 0.24$, 32 df for all tests using both unpooled and pooled allele frequencies; see Supporting Information for details). Moreover, the allele pooling method used here results in F_{ST} calculated from biallelic loci, eliminating some of the described inconsistencies of F_{ST} involving unequal variation (Charlesworth 1998), high diversity (Jost 2008), and constrained homozygosity (Jakobsson et al. 2013). The correspondence of F_{ST} values was significant when markers were considered separately for inverted ($r^2 = 0.478$, $P < 0.00081$, 19 df) and collinear ($r^2 = 0.624$, $P = 0.00133$, 12 df) regions.

Thus, while elevated F_{ST} associated with reduced recombination in rearrangements contributed to some of the concordance between the host races and dogwood fly, inversions alone could not account for the relationship. These results imply that divergent selection pressures acting on loci may vary in a similar manner across the genome. To better quantify the relative contributions of inversions and divergent selection to the pattern, we performed a multiple regression analysis. As a predictor for the strength of divergent selection acting either directly or indirectly on loci in the host races, we used the residuals generated from regressing apple/hawthorn fly F_{ST} values against the inverted/collinear status of loci to control for the effect of rearrangements (note this approach is similar to the chromosome center bias adjustment used by Roesti et al. 2012). We also included as a categorical variable whether a locus was located in an inverted or collinear region of the genome. Both the divergent selection and inversion terms were significant in the multiple regression, with divergent selection explaining 44.9% of the total variance of F_{ST} between dogwood and hawthorn flies ($P < 0.0001$) and inversions 10.2% ($P = 0.0144$).

With respect to individual chromosomes, F_{ST} values for loci on linkage groups 1, 2, and 5 significantly covaried between the host races and dogwood fly (Fig. 4). Markers on chromosome 1 generally displayed the greatest increases in F_{ST} for dogwood flies compared to chromosomes 2 and 5. Within chromosome 1, the allozymes *Aat-2* and *Dia-2* showed the largest increases in F_{ST}. Levels of differentiation did not significantly covary for loci on chromosomes 3 and 4 (Fig. 4). This was principally due to a subset of markers on chromosome 3 (*Had* and P23) and chromosome 4 (P50, P60, and *Pgi*) showing comparatively elevated divergence (higher F_{ST} values) for the dogwood fly than the host races. When these loci were removed from regressions, the remaining loci showed significant relationships for both chromosome 3 ($r^2 = 0.883$, $P < 0.018$, 4 df) and chromosome 4 ($r^2 = 0.997$, $P < 0.037$, 2 df).

Thus, the pattern of divergence for the dogwood fly may involve selection pressures across the genome that are generally congruent to those differentiating the apple and hawthorn races, with the exception of two gene regions on chromosomes 3 and 4, and perhaps *Aat-2* and *Dia-2* on chromosome 1, where selection may be relatively more pronounced for dogwood flies. Indeed, eliminating *Aat-2*, *Dia-2*, *Had*, P23, P50, P60, and *Pgi* increased the correspondence of F_{ST} values between dogwood flies and the host races ($r^2 = 0.677$, $P < 0.0001$, 25 df), with divergent selection explaining 61.7% of the variance ($P < 0.0001$) and inversions 6.6% ($P = 0.038$) in the corresponding multiple regression analysis.

SIMULATION MODELING

Based on a gross migration rate m of 4% per generation between apple and hawthorn flies derived from mark-release-recapture studies (Feder et al. 1994), Figure 4 gives the estimated selection coefficients s needed to produce the observed F_{ST} values for markers along chromosomes 1 and 2 between the host races, assuming that divergent selection is acting directly on these loci. Overall, we estimate that divergent selection reduces the effective gene flow rate (m_e) between hawthorn and apple fly populations to ~2.7% per generation genome-wide for unlinked, neutral genes.

Without a direct estimate of gross migration between *R. pomonella* and dogwood fly populations, we calculated a conservative proxy from our putative parental migrant/F1 identification analysis. The 1.12% gene flow estimate derived from this analysis was based on the genotypes of adult flies reared from infested fruit. Thus, this estimate encompassed the gross migration rate of adult flies and selection on the immature stages of their offspring. If we consider the proportion of putative F2 backcross to F1 hybrid flies ($3/8 = 0.375$) as providing a rough approximation of the relative fitness of the offspring of migrant flies, then this suggests a gross migration rate (m) of ~3% per generation between dogwood and hawthorn fly populations. Figure 4 gives the estimated selection coefficients s needed to produce the observed F_{ST} values for markers along chromosomes 1 and 2 between dogwood and hawthorn fly populations given a 3% gross migration rate. For loci on chromosome 1, the s values for dogwood flies needed to be about twice those between the host races to generate the observed level of differentiation, whereas for chromosome 2 the s values only needed to be about 1% greater. The simulations predicted an effective gene flow rate (m_e) of 1.35% between hawthorn and dogwood flies per generation for unlinked, neutral genes. This value is comparable to the 1.12% estimate from the STRUCTURE analysis and exactly half the simulation estimate of 2.7% between the host races.

Discussion

The results from the current study address three important issues concerning speciation with gene flow: (1) the nature of species, (2) the source of genetic variation fueling adaptive divergence, and (3) the role that genome structure plays in facilitating divergence.

THE NATURE OF SPECIES

Our findings suggest a criterion for differentiating host races from sibling species in cases of de novo speciation with gene flow related to both the BSC and genic species concepts. Specifically, because dogwood fly populations form a distinct genetic cluster, they may be considered to represent a species. The implication is that effective gene flow for a majority of the genome is higher among geographically dispersed populations of dogwood flies than it is between locally co-occurring demes of these flies and *R. pomonella*. This range-wide clustering of populations differs from Mallet's original genotypic clustering concept (1995) and

that proposed by Feder (1998). Both of these concepts are based on the clustering of individuals at the local level. Under the previous clustering concepts, it may be possible to interpret the distribution of *R. pomonella* genotypes at a single hawthorn/apple site (Feder 1998) as being indicative of species-level differences. However, as shown in Figure 3, on-going gene flow and clinal variation prevent the range-wide clustering of apple flies, despite the strong potential for intrahost gene flow. Thus, situations with strong local clustering may lead to the diagnosis of many site-specific species pairs, but geographically distant populations adapted to similar habitats are not together evolving independently relative to the populations associated with the alternative habitat. Range-wide population clustering reflects a shift in the relative strength of interhabitat to intrahabitat migration, and may therefore be more indicative of emerging evolutionary independence between differentially adapted forms. The range in question here must necessarily be restricted to populations in contact. Species status is inherently comparative between populations, and the potential for gene flow among allopatric demes cannot be meaningfully assessed (Mayr 1995).

We emphasize that under this genetic clustering definition there is no sharp boundary demarcating species from races; these stages form part of a continuum of divergence. The major difference is that the magnitude of genetic differentiation for species is greater and more consistent across geography than it is for host races. Indeed, despite forming a distinct genetic cluster, dogwood flies are not distinguished by fixed allelic differences or private alleles at any locus surveyed. Rather, the dogwood fly is differentiated in a similar manner as the host races, by frequency differences of shared alleles. Thus, allele frequency differences alone may be sufficient to distinguish species when the combination of reduced migration and divergent selection between populations adapted to different niches is strong enough for genetic clusters to be recognized across the geographic distribution of demes. This pattern will be most apparent in cases of speciation in primary contact with continuous gene flow, but it will also emerge in cases of secondary contact, once migration and selection have reached equilibrium. This shift in population structure may represent an important point along the path to complete evolutionary independence at the end of the speciation continuum. When this stage is reached, reproductive isolation likely has or is transitioning to become a characteristic of the whole genome and not individual genes, similar to Mayr's original casting of the BSC (1963), except due to divergent ecological selection and not necessarily strong genome-wide epistasis.

THE SOURCES OF GENETIC VARIATION

Our results imply that standing variation may be an important source fueling formative stages of speciation with gene flow in this system. We found no fixed differences or private alleles distinguishing apple, hawthorn, and dogwood flies. All three populations differed in the frequencies of shared alleles, many of which correlate with diapause traits differentially adapting the life history of flies to correspond to phenological differences in fruiting times among host plants (Michel et al. 2010). Moreover, the majority of loci showing host-related differences also display latitudinal clines coincident with geographic variation in the fruiting times of host plants (Table 2; Figs. S3–S7). Overall, these clines are most pronounced for hawthorn flies (Fig. 3), consistent with a mixed mode of biogeographic history at the base of the complex (Feder et al. 2005; Xie et al. 2007). Given that hawthorn is the ancestral host plant both for the apple race and the dogwood fly, the results imply that these latter two populations were derived, in large part, from standing life-history variation present in latitudinal frequency clines in hawthorn flies. The dogwood fly populations clustered in a mid-latitude location relative to the major clinal pattern of hawthorns in the neighbor-joining network (Fig. 3). However, dogwood fly genotypes do not appear to be drawn solely from this portion of the range of hawthorn flies. Rather, the clinal variation present within hawthorn-infesting *R. pomonella* provided a broad palette of diapause life-history variation, from which adaptive dogwood fly genotypes could be drawn. This is evidenced by dogwood fly populations containing allele frequencies similar to both the northern (e.g., see Figs. S1D, S2A, S3D) and southern (e.g., see Figs. S3A–C, S5A,B, S6C, S7A) ends of the hawthorn clines for different loci, as well as strong host × latitude interactions (Table 2; e.g., see Figs. S3E, S4B, S6B, S7D).

Neither line of evidence (lack of private alleles or clines) is sufficient to completely rule out the possibility that dogwood flies possess novel adaptive mutations. However, the observed patterns are consistent with the co-option of standing genetic variation being a major factor in the divergence of the dogwood fly. Higher coverage sequencing studies of the dogwood fly genome may eventually reveal private and near-fixed genetic differences. In addition, host recognition and performance genes might also differ diagnostically for the dogwood fly, whereas diapause-related genes may not. The current data do not discount the potential role of novel mutations in the system. However, these data are consistent with the hypothesis that dogwood flies represent a "polygenic species," extracted by divergent selection from largely prestanding allelic variation at multiple loci throughout the genome.

GENOME STRUCTURE AND SPECIATION

Our results imply that patterns of divergence in the *R. pomonella* group more closely resemble large genomic continents than one or a few scattered islands of differentiation. Previously, Michel et al. (2010) documented that genetic differences between the apple and hawthorn host races of *R. pomonella* was more widespread

than originally expected. Host-related differentiation was still heterogeneous through the *R. pomonella* genome, influenced by the interaction of the strength of divergent selection, migration, and recombination. Structural features of the genome also contributed to differentiation; F_{ST} values were significantly higher between the host races in inverted versus collinear regions of the genome. An important take home message from these findings was that genomic regions showing relatively lower levels of differentiation do not necessarily represent the baseline neutral expectation at mutation, migration, and drift equilibrium. Rather, the observed differentiation in these regions may still reflect the action of divergent natural selection, only less strongly so.

Here, we show that the pattern of widespread divergence seen for the host races also extends to the species-level differentiation of the dogwood fly. Indeed, genome-wide divergence for the dogwood fly roughly parallels on a locus-by-locus basis that for the host races. The main differences are that (1) the magnitude of host-related divergence is higher for the dogwood fly than between apple and hawthorn flies (F_{ST} values tend to be elevated genome-wide), and (2) a few additional gene regions show a relatively greater level of divergence for the dogwood fly (Fig. 4). These findings are more consistent with general patterns predicted for GH than DH. Specifically, during the progression from races to sibling species, divergence did not appear to expand outward from isolated genomic islands of speciation, although much more fine scale data on differentiation of flanking regions adjacent to peaks is required to test this hypothesis rigorously. Instead, we see a general genome-wide elevation of divergence coupled with a few additional peaks arising that are not evident in the host races, roughly as observed for races versus species pair of *Heliconius* butterflies (Nadeau et al. 2012). Genetic differentiation was still heterogeneous for dogwood flies; structural features of the genome associated with reduced recombination (chromosomal rearrangements) displayed higher mean divergence levels for the dogwood fly than collinear regions. However, these rearrangements did not comprise isolated islands, and the relative degree to which they were elevated in the dogwood fly was similar to that seen for collinear regions. The overall pattern of divergence implies that genome-wide effective gene flow may be reduced sufficiently for differences to accumulate between host-associated populations of *R. pomonella,* even for gene regions experiencing modest intensities of divergent selection, the characteristic feature of GH. It is possible that our analysis missed some genomic regions that may constitute islands of particularly high divergence, given the modest density of our marker coverage. However, the pattern of widespread significant differentiation, far above that expected by neutral migration at many sites across the genome fits the predictions of GH, even if genomically local peaks of differentiation exist and were undetected.

The widespread differentiation between the apple and hawthorn races reported by Michel et al. (2010) appeared to be driven by multifarious selection acting throughout the genome. Fifteen independent genomic regions were found to respond to a diapause selection experiment and 13 showed significant associations with eclosion time. The similarity in the topology of differentiation for the dogwood fly (Fig. 4) is consistent with an underlying similarity in selection regime. Adult eclosion phenology is a complex, multidimensional trait in *Rhagoletis*, involving the induction and termination of diapause (Dambroski and Feder 2007), interactions among different environmental cues (Feder et al. 2010), and widespread changes in gene expression (Ragland et al. 2011). We expect the phenological adaptation of dogwood flies may share the same pathways involved in the adaptation of the apple race, and this may become clear once the molecular underpinnings of diapause regulation are more fully elucidated in *Rhagoletis*.

Comparisons involving other members of the *R. pomonella* species complex may provide further insights into the process of speciation with gene flow. Evidence suggests that interhost migration rates are lower for *R. pomonella* with the snowberry fly, *R. zephyria*, (<0.1%; Feder et al. 1999) and blueberry fly, *R. mendax* (no detectable migration; Feder et al. 1989), than for the *R. pomonella* host races and dogwood fly. However, mtDNA sequences still show minimal differentiation among these flies (Fig. 1) (Smith and Bush 1997). Future studies involving these taxa will help elucidate how genomic differentiation builds through the speciation continuum, as on-going gene flow begins to have a negligible effect on divergence.

CONCLUSIONS

Our results imply that the dogwood fly is a sibling species to *R. pomonella* that essentially represents a host race writ large. The pattern of differentiation for the dogwood fly roughly parallels that between the apple and hawthorn host races except on a greater scale, implying increased ecological selection and lower effective migration. In addition, the results suggest that speciation with gene flow in the *R. pomonella* complex may involve incipient taxa rapidly attaining a degree of genome-wide divergence and reproductive isolation following host plant shifting akin to the BSC. Attaining this state of highly reduced m_e across the genome may allow new mutations of comparatively lower effect to establish in the face of gene flow, fostering further evolutionary divergence.

It remains to be determined whether other model systems containing divergent populations with well-resolved natural histories spanning stages along the speciation continuum will display similar patterns as manifested by *R. pomonella* flies. In this regard, *R. pomonella* could be unique with respect to the large stores of standing variation facilitating rapid divergence, although findings for stickleback (Jones et al. 2012) and *Heliconius* (Nadeau

et al. 2012) suggest otherwise. In contrast, other taxa may persist longer in earlier stages of speciation with gene flow. In these cases, m_e may primarily be reduced locally in the genome around isolated and already diverged loci. Here, the evolution of further reproductive isolation may be more heavily dependent on DH, provided that most new mutations have only minor fitness effects. It also needs to be determined whether and how populations undergoing speciation with gene flow following secondary contact fit into the four stage scheme developed for de novo divergence in primary contact. Cline theory has suggested that a similar shift from genomically local to genome-wide differentiation can occur following secondary contact where reproductive isolation transitions from being primarily genic to a characteristic of the whole genome (Barton 1983). Moreover, during the initial period of allopatry, new adaptive variation could accumulate through the genome in a manner that may be less likely to evolve de novo in sympatry, resulting in an overall distribution of prestanding divergence conducive to the evolution of further differentiation following secondary contact. Regardless, advances in mass sequencing are now enabling whole genome scans of nonmodel organisms in fine detail. Combining these surveys with selection experiments and expression data on adaptive traits generating reproductive isolation will allow us to develop a richer understanding of both the pattern and process of speciation genomics.

ACKNOWLEDGMENTS

We thank Scott Egan, Kirsten Prior, and Greg Ragland for helpful insight, Dave Costello, Andrew Forbes, and Matt Michel for help with field collections, and the editors and two anonymous reviewers for their helpful comments and criticism. This work was supported by a National Science Foundation grant to JLF and SHB and the University of Notre Dame's GLOBES IGERT program, Environmental Change Initiative, and program in Advanced Diagnostics & Therapeutics.

LITERATURE CITED

Avise, J., and M. Smith. 1977. Gene frequency comparisons between sunfish (Centrarchidae) populations. Syst. Zool. 26:319–335.

Ayala, F., M. Tracey, D. Hedgecock, and R. Richmond. 1974. Genetic differentiation during speciation process in *Drosophila*. Evolution 28:576–592.

Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation. Trends Ecol. Evol. 23:38–44.

Barton, N. 1983. Multilocus clines. Evolution 37:454–471.

Berlocher, S. 1999. Host race or species? Allozyme characterization of the "flowering dogwood fly," a member of the *Rhagoletis pomonella* complex. Heredity 83:652–662.

———. 2000. Radiation and divergence in the *Rhagoletis pomonella* species group: inferences from allozymes. Evolution 54:543–557.

Berlocher, S., and B. McPheron. 1996. Population structure of *Rhagoletis pomonella*, the apple maggot fly. Heredity 77:83–99.

Berlocher, S., B. McPheron, J. Feder, and G. Bush. 1993. Genetic differentiation at allozyme loci in the *Rhagoletis pomonella* (Diptera, Tephritidae) species complex. Ann. Entomol. Soc. Am. 86:716–727.

Berner, D., A. C. Grandchamp, and A. P. Hendry. 2009. Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. Evolution 63:1740–1753.

Bolnick, D. I., and B. M. Fitzpatrick. 2007. Sympatric speciation: models and empirical evidence. Annu. Rev. Ecol. Evol. Syst. 38:459–487.

Bush, G. L. 1966. The taxonomy, cytology, and evolution of the genus *Rhagoletis* in North America (Diptera: Tephritidae). Museum of Comparative Zoology, Cambridge, MA.

———. 1969. Sympatric host race formation and speciation in frugivorous flies of genus *Rhagoletis* (Diptera, Tephritidae). Evolution 23:237–251.

Cha, D. H., T. H. Q. Powell, J. L. Feder, and C. E. Linn. 2012. Geographic variation in fruit volatiles emitted by the hawthorn *Crataegus mollis* and its consequences for host race formation in the apple maggot fly, *Rhagoletis pomonella*. Entomol. Exp. Appl. 143:254–268.

Charlesworth, B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. 15: 538–543.

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

Dambroski, H. R., and J. L. Feder. 2007. Host plant and latitude-related diapause variation in *Rhagoletis pomonella*: a test for multifaceted life history adaptation on different stages of diapause development. J. Evol. Biol. 20:2101–2112.

Darwin, C. 1859. On the origin of species by means of natural selection. John Murray, London.

De Queiroz, K. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–886.

Feder, J. L. 1998. The apple maggot fly, *Rhagoletis pomonella*: flies in the face of conventional wisdom about speciation? Pp. 130–144 *in* S. H. Berlocher and D. J. Howard, eds. Endless forms: species and speciation. Oxford Univ. Press, Oxford, U.K.

Feder, J., and G. Bush. 1989. Gene-frequency clines for host races of *Rhagoletis pomonella* in the midwestern United States. Heredity 63:245–266.

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

Feder J. L., C. A. Chilcote, and G. L. Bush. 1989. Are the apple maggot, *Rhagoletis pomonella,* and blueberry maggot, *Rhagoletis mendax*, distinct species? Implications for sympatric speciation. Entomol. Exp. Appl. 51:113–123.

———. 1990. Regional, local and microgeographic allele frequency variation between apple and hawthorn populations of *Rhagoletis pomonella* in western Michigan. Evolution 44:595–608.

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

Feder, J. L., S. M. Williams, S. H. Berlocher, B. A. McPheron, and G. L. Bush. 1999. The population genetics of the apple maggot fly, *Rhagoletis pomonella* and the snowberry maggot, *R. zephyria*: implications for models of sympatric speciation. Entomol. Exp. Appl. 90:9–24.

Feder, J. L., S. H. Berlocher, J. B. Roethele, H. Dambroski, J. J. Smith, W. L. Perry, V. Gavrilovic, K. E. Filchak, J. Rull, and M. Aluja. 2003a. Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. Proc. Natl. Acad. Sci. USA 100:10314–10319.

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

Feder, J. L., X. F. Xie, J. Rull, S. Velez, A. Forbes, B. Leung, H. Dambroski, K. E. Filchak, and M. Aluja. 2005. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. Proc. Natl. Acad. Sci. USA 102:6573–6580.

Feder, J. L., T. H. Q. Powell, K. Filchak, and B. Leung. 2010. The diapause response of *Rhagoletis pomonella* to varying environmental conditions

and its significance for geographic and host plant-related adaptation. Entomol. Exp. Appl. 136: 31–44.

Feder, J. L., S. P. Egan, and P. Nosil. 2012a. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350.

Feder, J. L., R. Gejji, S. Yeaman, and P. Nosil. 2012b. Establishment of new mutations under divergence and genome hitchhiking. Philos. Trans. R. Soc. Lond. B 367:461–474.

Feder, J. L., S. M. Flaxman, S. P. Egan, and P. Nosil. 2013. Hybridization and the buildup of genomic divergence during speciation. J. Evol. Biol. 26:261–266. doi:10.1111/jeb.12009.

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

———. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

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

Hadziabdic, D., B. M. Fitzpatrick, X. Wang, P. A. Wadl, T. A. Rinehart, B. H. Ownley, M. T. Windham, and R. N. Trigiano. 2010. Analysis of genetic diversity in flowering dogwood natural stands using microsatellites: the effects of dogwood anthracnose. Genetica 138:1047–1057.

Hausdorf, B. 2011. Progress toward a general species concept. Evolution 65:923–931.

Hendry, A. P., D. I. Bolnick, D. Berner, and C. L. Peichel. 2009. Along the speciation continuum in sticklebacks. J. Fish Biol. 75:2000–2036.

Hermisson, J., and P. S. Pennings. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335–2352.

Hey, J. 2001. The mind of the species problem. Trends Ecol. Evol. 16:326–329.

Hey, J., and C. Pinho. 2012. Population genetics and objectivity in species diagnosis. Evolution 66:1413–1429.

Jakobsson, M., M. D. Edge, and N. A. Rosenburg. 2013. The relationship between F_{ST} and the frequency of the most frequent allele. Genetics 193:515–528.

Jones, F. C., M. G. Grabherr, Y. F. Chan, F. Yingguang, P. Russel, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. Zody, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 484:55–61.

Jost L. 2008. G_{ST} and its relatives do not measure differentiation. Mol. Ecol. 17:4015–4026.

Linn, C., J. L. Feder, S. Nojima, H. R. Dambroski, S. H. Berlocher, and W. Roelofs. 2003. Fruit odor discrimination and sympatric host race formation in *Rhagoletis*. Proc. Natl. Acad. Sci. USA 100:11490–11493.

Linn, C. E., H. Dambroski, S. Nojima, J. L. Feder, S. H. Berlocher, and W. L. Roelofs. 2005. Variability in response specificity of apple, hawthorn, and flowering dogwood-infesting Rhagoletis flies to host fruit volatile blends: implications for sympatric host shifts. Entomol. Exp. Appl. 116:55–64.

Liu, K. J., and S. V. Muse. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129.

Mallet, J. 1995. A species definition for the modern synthesis. Trends Ecol. Evol. 10:294–299.

———. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Philos. Trans. R. Soc. Lond. B 363:2971–2986.

Mayden, R. L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. Pp. 381–424 *in* M. F. Claridge, H. A. Dawah, and M. R. Wilson, eds. Species the units of biodiversity. Chapman and Hall Ltd, London.

Maynard-Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23:23–35.

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

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

———. 1969. Principles of systematic biology. McGraw-Hill, New York.

———. 1995. Species, classification, and evolution. Pp. 6–12 *in* R. Arai, M. Kato, and Y. Doi, eds. Biodiversity and evolution. National Science Foundation Museum, Tokyo.

———. 2001. What evolution is. Basic Books, New York.

Merrill, R. M., Z. Gompert, L. M. Dembeck, M. R. Kronforst, W. O. McMillan, and C. Jiggins. 2011. Mate preference across the speciation continuum in a clade of mimetic butterflies. Evolution 65:1489–1500.

Michel, A. P., S. Sim, T. H. Q. Powell, M. S. Taylor, P. Nosil, and J. L. Feder. 2010. Widespread genomic divergence during sympatric speciation. Proc. Natl. Acad. Sci. USA 107:9724–9729.

Nadeau, N. J., A. Whibley, R. T. Jones, J. W. Davey, K. K. Dasmahaptra, S. W. Baxter, M. A. Quail, M. Joron, R. F. French-Constant, M. L. Blaxter, et al. 2012. Genomic islands of divergence in hybridizing Heliconious butterflies identified by large-scale targeted sequencing. Philos. Trans. R. Soc. Lond. B 367:343–353.

Nei, M. 1972. Genetic distance between populations. Am. Nat. 106:283.

Neilson, W., and J. Mcallan. 1965. Effects of mating on fecundity of apple maggot *Rhagoletis pomonella* (Walsh). Can. Entomol. 97:276.

Nosil, P., and J. L. Feder. 2012. Genomic divergence during speciation: causes and consequences. Introduction. Philos. Trans. R. Soc. Lond. B 367:332–342.

Nosil, P., and C. P. Sandoval. 2008. Ecological niche dimensionality and the evolutionary diversification of stick insects. PLoS One 3: e1907.

Peccoud, J., A. Ollivier, M. Plantegenest, and J. C. Simon. 2009. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. Proc. Natl. Acad. Sci. USA 106:7495–7500.

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

Ragland, G. J., S. P. Egan, J. L. Feder, S. H. Berlocher, and D. A. Hahn. 2011. Developmental trajectories of gene expression reveal candidates for diapause termination: a key life-history transition in the apple maggot fly Rhagoletis pomonella. J. Exp. Biol. 214:3948–3959.

Roesti, M., A. P. Hendry, W. Salzburger, and D. Berner. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. Mol. Ecol. 21:2852–2862.

Roethele, J. B., J. L. Feder, S. H. Berlocher, M. E. Kreitman, and D. A. Lashkari. 1997. Toward a molecular genetic linkage map for the apple maggot fly (Diptera: Tephritidae): Comparison of alternative strategies. Ann. Entomol. Soc. Am. 90:470–479.

Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proc .Natl. Acad. Sci. USA 106:9955–9962.

Seehausen, O. 2008. Progressive levels of trait divergence along the 'speciation transect' in the Lake Victoria cichlid fish *Pundamilia*. Pp. 155–176 *in* R. Butlin, J. Bridle, and D. Schluter, eds. Speciation and patterns of diversity. Cambridge Univ. Press, Cambridge, U.K.

Smith, D. C. 1986. Genetics and reproductive isolation in *Rhagoletis* flies. Ph.D. diss., University of Illinois at Urbana-Champaign, Urbana, IL.

Smith, J. J., and G. L. Bush. 1997. Phylogeny of the genus *Rhagoletis* (Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase II. Mol. Phylogenet. Evol. 7:33–43.

Velez, S., M. S. Taylor, M. A. F. Noor, N. F. Lobo, and J. L. Feder. 2006. Isolation and characterization of microsatellite loci from the apple

maggot fly *Rhagoletis pomonella* (Diptera: Tephritidae). Mol. Ecol. Notes 6:90–92.

Via, S. 2009. Natural selection in action during speciation. Proc. Natl. Acad. Sci. USA 106:9939–9946.

———. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos. Trans. R. Soc. Lond. B 367:451–460.

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

Walsh, B. D. 1867. The apple-worm and the apple-maggot. Am. J. Hortic. 2:338–243

Wiehe, T., and W. Stephan. 1993. Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol. Biol. Evol. 10:842–854.

Wu, C. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.

Xie, X., J. Rull, A. P. Michel, S. Velez, A. A. Forbes, N. F. Lobo, M. Aluja, and J. L. Feder. 2007. Hawthorn-infesting populations of *Rhagoletis pomonella* in Mexico and speciation mode plurality. Evolution 61:1091–1105.

Xie, X., A. P. Michel, D. Schwarz, J. Rull, S. Velez, A. A. Forbes, M. Aluja, and J. L. Feder. 2008. Radiation and divergence in the *Rhagoletis pomonella* species complex: inferences from DNA sequence data. J. Evol. Biol. 21:900–913.

Zimmerman, E., C. Kilpatrick, and B. Hart. 1978. Genetics of speciation in rodent genus *Peromyscus*. Evolution 32:565–579.

Associate Editor: D. Fairbairn

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. (Uploaded to DRYAD) Microsatellite allele frequencies, heterozygosity (*H*), and allelic richness (*A*) for the 23 loci analyzed in study for 17 field sites, representing seven hawthorn- and four apple-infesting populations of *R. pomonella*, and six flowering dogwood fly populations.

Table S2. Collection location, reference, and Genbank accession number for flowering dogwood fly (dogwood), hawthorn race *R. pomonella* (Rp_H), apple race *R. pomonella* (Rp_A), and outgroup taxa (*R. electromopha* and *R. cornivora*) sequences used in mtDNA gene tree.

Table S3. Alleles included in one of two groups generated by Monte Carlo allele pooling method for each of 23 microsatellite loci across five chromosomes (Chr.).

Table S4. Pairwise F_{ST} values between all populations pairs, calculated across all microsatellite loci (lower triangle) and results of exact tests for allelic differentiation analyzed for all loci (upper triangle).

Table S5. Mean F_{ST} values for 23 microsatellite and 10 allozyme loci calculated between apple and hawthorn host races of *R. pomonella* (*Host Races*) and between hawthorn and dogwood flies (*Species*) across local sites.

Figure S1. Map of overlapping distributions of the flowering dogwood fly and the hawthorn and apple races of *R. pomonella* in the eastern United States with collection site designations.

Figure S2. Neighbor-joining network for apple (yellow circle) and hawthorn (green triangle) host races of *R. pomonella* and the flowering dogwood fly (red square) based on Nei's *D* (1972) estimated from unpooled microsatellite allele frequencies.

Figures S3–S7. Allele frequencies across collecting sites for the apple-infesting (gray circles and lines) and hawthorn-infesting (black triangles and dashed lines) host races of *R. pomonella* and the flowering dogwood fly (black squares and line) for the 23 microsatellite loci mapping to chromosomes 1 through 5 of the genome plotted against the latitude of populations.

Figure S8. STRUCTURE output for $K = 3$ analysis of all 17 populations (Table 1) under the same parameters as the regionally segregated $K = 2$ analyses presented in Figure 2.