MOLECULAR ECOLOGY

Molecular Ecology (2013) doi: 10.1111/mec.12423

Host-associated genomic differentiation in congeneric butterflies: now you see it, now you do not

ALEXANDER S. MIKHEYEV,* CAROLYN S. MCBRIDE,† ULRICH G. MUELLER,‡ CAMILLE PARMESAN,‡§ MELANIE R. SMEE,¶ CONSTANTI STEFANESCU,‖** BRIAN WEE†† and MICHAEL C. SINGER‡‡‡

*Okinawa Institute of Science and Technology, Onna-son, Okinawa 904-0495, Japan, †Laboratory of Neurogenetics and Behavior, The Rockefeller University, New York, NY, USA, ‡Integrative Biology, University of Texas, Austin, TX 78712, USA, §Marine Institute, Level 3 Marine Bldg., Plymouth University, Drakes Circus, Plymouth Devon PL4 8AA, UK, ¶Department of Biology, University of York, York YO10 5DD, UK, ‖Catalan Butterfly Monitoring Scheme, Museu de Ciències Naturals, Granollers 08402, Spain, **Global Ecology Unit, CREAF, Cerdanyola del Vallès 08193, Spain, ††NEON, Inc. c/o Smithsonian Institution, 1100 Jefferson Drive SW, Suite 3123, MRC 705, Washington DC 20560-0001, USA, ‡‡School of Biomedical and Biological Sciences, Plymouth University, Drake Circus, Plymouth PL4 8AA, UK*

Abstract

Ecotypic variation among populations may become associated with widespread genomic differentiation, but theory predicts that this should happen only under particular conditions of gene flow, selection and population size. In closely related species, we might expect the strength of host-associated genomic differentiation (HAD) to be correlated with the degree of phenotypic differentiation in host-adaptive traits. Using microsatellite and Amplified Fragment Length Polymorphism (AFLP) markers, and controlling for isolation by distance between populations, we sought HAD in two congeneric species of butterflies with different degrees of host plant specialization. Prior work on *Euphydryas editha* had shown strong interpopulation differentiation in host-adapted traits, resulting in incipient reproductive isolation among host-associated ecotypes. We show here that *Euphydryas aurinia* had much weaker host-associated phenotypic differentiation. Contrary to our expectations, we detected HAD in *Euphydryas aurinia*, but not in *E. editha*. Even within an *E. aurinia* population that fed on both hosts, we found weak but significant sympatric HAD that persisted in samples taken 9 years apart. The finding of significantly stronger HAD in the system with less phenotypic differentiation may seem paradoxical. Our findings can be explained by multiple factors, ranging from differences in dispersal or effective population size, to spatial variation in genomic or phenotypic traits and to structure induced by past histories of host-adapted populations. Other infrequently measured factors, such as differences in recombination rates, may also play a role. Our result adds to recent work as a further caution against assumptions of simple relationships between genomic and adaptive phenotypic differentiation.

Keywords: *Castilleja*, *Collinsia*, *Lonicera*, Nymphalidae, *Pedicularis*, *Succisa*

Received 28 February 2013; revision received 13 June 2013; accepted 13 June 2013

Introduction

Spatial variation in habitat quality generates heterogeneous patterns of natural selection acting on locally beneficial traits. Population divergence in these traits should result even in the face of moderate homogenizing gene flow (Slatkin 1985; Yeaman & Otto 2011). Such divergence is frequently observed (Jain & Bradshaw 1966; Lenormand *et al.* 1999; Hoekstra *et al.* 2004; Kawecki & Ebert 2004; McCracken *et al.* 2009; Nosil 2009; Egan *et al.* 2012a; Roesti *et al.* 2012). From the moment that populations first diverge in a specific ecological adaptation, they are systematically differentiated

Correspondence: Alexander S. Mikheyev,
Fax: +81 (0) 98-966-1063; E-mail: sasha@homologo.us

at loci controlling traits directly involved in that adaptation. In practice, population-level differentiation may remain restricted to such loci and be undetectable at neutral sites. For example, Hoekstra *et al.* (2004) found no genomic differentiation in a spatial mosaic of pocket mice that had evolved different coat colours on different substrates. Likewise, Feldman *et al.* (2010) found no association between neutral markers and tetrodotoxin resistance that allowed garter snakes to feed on toxic newts (see also McCracken *et al.* 2009; Egan *et al.* 2012b).

Alternatively, genetic correlates of ecological adaptation may spread more widely through the genome, a process that meta-analysis indicates is quite prevalent in nature (Shafer & Wolf 2013). The nature and extent of this spread, and its relationship between landscape and geography, can provide information about the history of trait acquisition and about progress towards possible speciation (Emelianov *et al.* 2004; Rogers & Bernatchez 2007; Egan *et al.* 2008; Via & West 2008; Chamberlain *et al.* 2009; McCracken *et al.* 2009 Peccoud *et al.* 2009; Counterman *et al.* 2010; DeLeon *et al.* 2010; Roesti *et al.* 2012). Lowry (2010), in a mini-review, opined that these studies usher in an era of 'landscape evolutionary genomics' in which histories of the evolution of local adaptation will be elucidated in a spatially explicit manner. To attempt this, we need to know what processes have the potential to cause ecologically associated differentiation at loci that are not directly involved in ecological adaptation. At least three such processes have been described.

First, when gene flow occurs between populations under contrasting selection regimes, natural selection can reduce effective recombination around divergently selected loci and generate a genomic signature of adaptation with divergence between habitats centred on the genomic regions involved in local adaptation (Begun & Aquadro 1992; Barton 2000; Emelianov *et al.* 2004; Via & West 2008; Counterman *et al.* 2010; Feder & Nosil 2010). Second, resource-associated differentiation can develop when divergent adaptations generate reproductive isolation and thereby facilitate genetic drift (Nosil *et al.* 2008). Thibert-Plante & Hendry (2010) predict that this occurs when gene flow between populations using the same resource is high enough for the effects of drift to be negligible, while gene flow between populations using different resources is reduced to the point where drift becomes important. Third, a large literature, from Slatkin (1975) and Barton & Hewitt (1985) to Bierne *et al.* (2011), predicts that barriers to gene flow caused by endogenous incompatibilities will tend to drift across landscapes until they coincide in space with exogenous barriers caused by ecological factors. By this process, Bierne *et al.* (2011) predict that 'coupling of endogenous and exogenous factors can occur easily in spatially subdivided populations even if the loci concerned are unlinked'. They argue that genotype–environment associations can be developed by this coupling process even if the principal factor reducing gene flow is endogenous and not causally related to environmental variables.

Under these diverse influences, the genome can become a mosaic of regions differing in the degree to which they have been influenced by specific ecological adaptations (Emelianov *et al.* 2004; Nosil *et al.* 2009; Via 2009; Nosil & Feder 2012; Roesti *et al.* 2012). Herbivorous insects have figured conspicuously in efforts to document the genomic signatures of ecological adaptation stemming from the use of different host plant species as hosts. Despite the predictions that resource-associated differentiation should develop only under rather restricted conditions, host-associated genomic differentiation – hereafter abbreviated as 'HAD' – has been frequently reported in herbivorous insects (Stireman *et al.* 2012 & references therein). Why HAD is so prevalent in insects, despite apparently restrictive conditions for its evolution, is not fully understood.

We investigated HAD within two congeneric nymphalid butterflies that differ in the degree of population differentiation in host-adaptive traits. Because stronger phenotypic differentiation implies stronger divergent selection and/or stronger barriers to gene flow, all else being equal, we expected that HAD would be more developed in the species with the stronger phenotypic differentiation. Instead, we found the opposite pattern, suggesting that factors other than selection on host-adaptive traits play a role in shaping the extent of HAD.

Materials and methods

Study system

Our two study species are *Euphydryas aurinia* and *E. editha*, in the subfamily Melitaeinae, a group of which the population biology has been extensively studied (Hanski 1999, 2011; Wahlberg *et al.* 2002; Ehrlich & Hanski 2004). Published information on host specialization of *Euphydya*s indicates strong phenotypic differentiation among *E. editha* populations using different hosts (Singer 1971; Rausher 1982; Singer & Parmesan 1993; Singer *et al.* 1994; Singer & McBride 2010) and relatively weak differentiation in *E. aurinia* (Mazel 1986). We present a detailed comparison of host-associated phenotypic differentiation in the two species. We begin by summarizing previously published data on three host-associated traits of *E. editha*: adult oviposition preference, larval performance and clutch size. This information

was obtained by comparing populations using either of two different hosts, *Pedicularis semibarbata* (Orobanchaceae) or *Collinsia torreyi* (Plantaginaceae). We complement this published information by obtaining parallel data on the same three traits from *E. aurinia*, using populations feeding on either of two hosts: *Succisa pratensis* (Dipsacaceae) or *Lonicera sp.* (*L. implexa* or *L. etrusca*, Caprifoliaceae). For *E. aurinia*, we have prior data only for a single trait, host preference, in a single population (Singer *et al.* 2002), and with this exception, the *E. aurinia* data presented here are new. For the purpose of our study, we should ideally also have a detailed comparison of dispersal in the two species. Dispersal has been extensively studied in both species, but we lack a comparison that applies the same techniques in the same manner across the same spatial scale. Nonetheless, the cumulative published evidence (Ford & Ford 1930; Ehrlich 1965; Gilbert & Singer 1973; Harrison 1989; Zimmerman *et al.* 2011) suggests that the two species should not differ dramatically in their patterns of dispersal (see Appendix S1, Supporting information for more details).

Diet recording and population sampling

Adult butterflies are typically more specialized in their oviposition choices than are larvae searching for food (Wiklund 1975). We therefore recorded the diet of a *Euphydryas* population as the proportion of eggs laid on each host at that site.

For our sampling of *E. aurinia*, we used populations in two habitat types: those feeding on *Lonicera* in open Mediterranean woodland and those feeding on *Succisa* in damp meadows (Fig. 1b). We found a single unique, biphagous site, La Barroca, where both habitat types were juxtaposed, the insects flew freely between them and used both hosts. At all other sites, the butterflies had only a single host available and were monophagous. Our study sites were all at low-to-moderate elevation, between sea level and 1200 m. We ignored populations living at high elevation (>1800 m) that feed on *Gentiana* (Gentianaceae) or *Succisa*. The reasons for this were partly logistical and partly our desire to investigate the effects of isolation stemming from host use and not from phenological isolation unrelated to host adaptation. The Catalan Butterfly Monitoring Scheme (http://www.catalanbms.org/) indicates that *E. aurinia* populations at high elevation, which flew in July, were phenologically isolated from those in our study, which flew in April/May. The monitoring scheme records a bimodal distribution of the insects across altitudes, with records from elevations below 1200 m and above 1800 m, indicating absence or scarcity of sites at 1200–1800 m that might, if they were to exist, act as efficient corridors for gene flow between insects at low and high elevation.

Because *E. editha* exists in populations distributed more evenly than *E. aurinia* across elevations from sea level to >3000 m, gene flow among sites at different altitudes should be less constrained and we did not restrict our sampling to particular elevations. We sampled populations feeding principally on *Pedicularis* (Orobanchaceae), *Castilleja* (Orobanchaceae) or *Collinsia* (Plantaginaceae) (Fig. 1a). We ignored populations feeding principally on *Penstemon*, *Veronica* or *Antirrhinum* (Plantaginaceae; mapped in Singer & Wee 2005) because we know of only a small number of populations using these hosts: two populations using *Penstemon*, one using *Veronica* and one using *Antirrhinum*. We also ignored populations whose principal host is *Plantago*, because, although we were able to sample five such populations, four of them occurred clumped in space and were also genetically very close, making it difficult to separate effects of host from effects of geographical position.

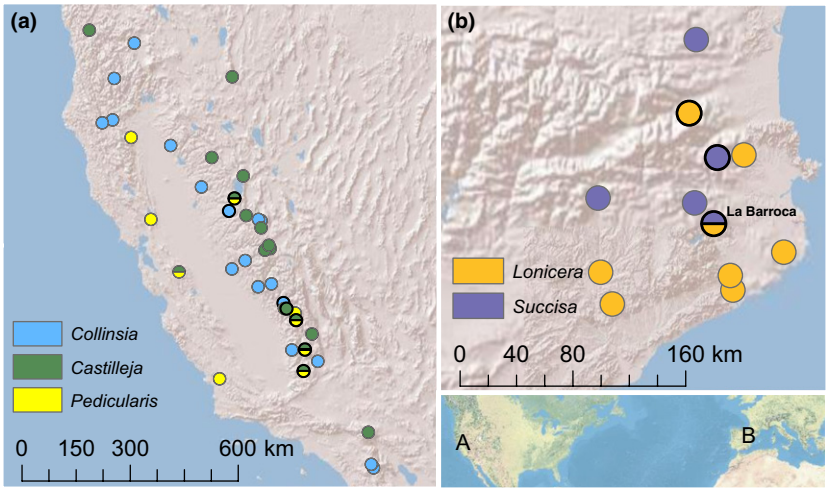


Fig. 1 Geographical distribution of populations used in this study and their host use. *Euphydryas editha* (a), *Euphydryas aurinia* (b). The subset of *E. editha* populations outlined with bold circles has a previously characterized suite of host-specific adaptations (Singer & McBride 2010). *E. aurinia* populations used for phenological comparison in Fig. 3 are also outlined in bold. La Barroca is the biphagous *E. aurinia* population.

We treated individual flying adults captured on the same day as independent samples from their populations. Groups of larvae living in separate webs were likewise considered to be independent. To maintain this independence of samples, we used only one egg or larva from any given group.

Although we controlled for isolation by distance (IBD) when analysing for HAD, we nonetheless strove to sample in a way that minimized the spatial clumping of same-host populations as much as possible. This was more difficult in *E. aurinia*, in which the *Lonicera*-feeding ecotype reaches its northern limit on the north slopes of the Pyrenees and *Succisa*-feeding ecotype reaches its known southern limit about 60 km further south, leaving a relatively narrow latitudinal belt within which populations on the two hosts both occur. We restricted our sampling to sites within and close to this latitudinal band. As a result, although the mean position of *Lonicera*-feeding populations is further south than that of *Succisa* feeders, samples from the two hosts were sufficiently interspersed that they did not form spatially separate groups (Fig. 1b). In contrast, *E. editha* populations on different hosts are naturally interdigitated all across California and southern Oregon (Singer & Wee 2005), so it was relatively easy to extend our sampling of insects on each of the three host genera across most of the study area without introducing strong geographical bias (Fig. 1a).

Phenotypic differentiation associated with host use

Populations used for studies of phenotypic differentiation are listed in Table 1 and displayed graphically in Fig. 1b.

Oviposition preference. The strength of preference was measured by recording host acceptances and rejections in a series of staged encounters with each host in alternation, using a technique developed for Melitaeine butterflies (Singer *et al.* 1992). Host acceptance was recorded when the insect extruded its ovipositor and probed the underside of a leaf (see videos supplemental to McBride & Singer 2010). The experimenter prevented oviposition by manually removing the insect from the plant as soon as acceptance was observed.

In the absence of actual oviposition, the range of hosts that would be accepted expands over time, as motivation to lay eggs increases. As a consequence, when two hosts of differing acceptability are offered, the more preferred host is the first to be accepted, while the first acceptance of the less preferred host comes later, if at all. Butterflies that accepted the foreign host at different times relative to the native host were assigned to the following scores: −2 (first accepted foreign host one or more days prior to first accepting native host), −1 (accepted foreign host earlier than native host but on the same day), 0 (accepted foreign host at same time as native host), 1 (accepted foreign host later than native host but on same day) and 2 (first accepted foreign host 1 or more days after first accepting native host). Positive scores thus indicate preferences for the native host, while negative scores indicate preference for the foreign host.

Clutch size. To investigate possible differentiation of clutch size among host-affiliated populations of *E. aurinia*, we made field observations of natural clutches at four *Lonicera*-feeding sites and two *Succisa*-feeding sites.

Table 1 Locations, samples sizes and host use of *Euphydryas aurinia* populations used in our study. The populations in grey have been used for phenological comparisons (Fig. 3)

Population		Sample size		Host	Location	
Long name	Code	AFLP	Microsatellites		Latitude	Longitude
Can Jorda	CJ	9	11	Succisa	42.1443	2.503
Col Estelales	COL	2	5	Lonicera	41.6635	1.97906
Coustouges	COU	9	14	Succisa	42.3584	2.64733
Darnius	DAR	12	14	Lonicera	42.3704	2.81783
El Guix	ELG	4	6	Lonicera	41.8159	1.90441
Col del Forn	FORN	10	14	Lonicera	42.5681	2.46861
La Barroca	LALO	9	10	Lonicera	42.0457	2.62543
La Barroca	LASU	9	12	Succisa	42.0457	2.62543
La Barroca	LALO2010	0	14	Lonicera	42.0457	2.62543
La Barroca	LASU2010	0	15	Succisa	42.0457	2.62543
La Nou de Bergeuda	HIS	2	2	Succisa	42.167	1.88627
Mas Calc	MC	14	16	Lonicera	41.9112	3.07234
Col de la Redoulade	RED	3	9	Succisa	42.9122	2.5145
Tordera	TOR	5	7	Lonicera	41.7308	2.7475
Sequia de Sils	SILS	0	6	Lonicera	41.80038	2.73104

When clutches were too large to count in the field, we took photographs, from which we could later make egg counts.

Larval performance. To compare larval performance on the two hosts, we split families of neonate *E. aurinia* larvae into two equal-sized groups and reared them on cut leaves of each host in captivity, recording survival to diapause.

Phenological separation. To investigate possible phenological separation of *E. aurinia* populations on different hosts, we visited two monophagous populations that used different hosts and were separated by <50 km. We made the visits on consecutive days and recorded developmental stages of insects found in the field on each host, categorizing insects as larvae in first, second or third instar. We also recorded simultaneous developmental stages of insects on the two hosts at the only site in our study, La Barroca, where both hosts were used. Sites used for phenological comparisons are highlighted by black outlines in Fig. 1b.

Analysis of molecular data

We used two sources of molecular data. One data set comprised nuclear DNA AFLPs genotyped by Wee on samples gathered between 1999 and 2003 (2004). That analysis used four primer combinations for *E. aurinia* and three primer combinations for *E. editha*, producing 318 polymorphic loci in *E.* aurinia and 547 polymorphic loci in *E. editha*. The other data sets were from EST-derived microsatellites identified in two previous studies, which developed the molecular markers, and genotyped the populations used in the present study, but did not conduct analysis of HAD (Mikheyev *et al.* 2010; Smee *et al.* 2013). These two studies developed 9 loci for *E. aurinia* (3.4 ± 1.2 alleles/locus) and 10 loci for *E. editha* (5.3 ± 2.3 alleles/locus) and used the same individuals that were previously genotyped for AFLPs, but also added additional samples gathered between 2008 and 2010. In consequence, the present study had a larger sample of populations in the microsatellite data set than in the AFLP data set. For *E. aurinia*, we analysed microsatellite data on 157 individuals from 12 populations and AFLP data on 88 individuals from 11 populations. For *E. editha*, we had microsatellite data on 445 individuals from 40 populations and AFLP data on 246 individuals from 23 populations. Tables 1 and 2 show the number of individuals genotyped for each type of marker in each population.

The AFLP and microsatellite data sets provide technically independent estimates of population differentiation. To generate these estimates, we used Arlequin 3.5 (Excoffier 2010) to obtain θ_{st} statistics from AFLPs and F_{st} statistics from microsatellites. θ_{st} is an estimator of F_{st} based upon the proportion of allele frequencies shared between individuals in different subpopulations (Weir & Cockerham 1984). As these two statistics are analogous and strongly correlated in our data (Mikheyev *et al.* 2010; Smee *et al.* 2013), we used averages of the θ_{st} and F_{st} matrices where both values were available. Because AFLP data were from an earlier and less comprehensive survey, they were missing for several populations, forcing us to rely on microsatellite data only in those cases. However, we also verified that our principal conclusions hold for AFLP and microsatellite data sets analysed separately. R scripts and output from analysis under different data partitions can be found on DataDryad (doi:10.5061/dryad.1v0tb).

Isolation-by-distance analysis

Genetic matrices were linearized using Slatkin's method (Slatkin 1995), and the geographical distances were log-transformed. With these transformations, the slope of the regression of genetic distance on geographical distance is inversely proportional to the product of effective population size and the variance of the offspring position relative to parental position ($1/(4N*\pi*\sigma)$) (Rousset 1997). Isolation by distance and the regression slope were evaluated with Mantel tests using 1000 bootstrap replicates of all points, computed in software IBD (Bohonak 2002).

Host-associated differentiation analysis

In the presence of significant IBD, the partial Mantel test remains the most commonly used tool for analysing HAD while controlling for effects of spatial structure (Smouse *et al.* 1986). However, there has been debate concerning its validity (Legendre 2000; Raufaste & Rousset 2001; Rousset & Waller 2002; Castellano & Balletto 2002). As our data were analysed using Euclidean distances between geographical coordinates, the partial Mantel test should be appropriate for our data (Legendre & Fortin 2010). We nonetheless also conducted an alternative test of HAD called a distance-based redundancy analysis (dbRDA) (Legendre & Anderson 1999). In the dbRDA, we examined the effect of host use variables on the genetic distance response matrix while controlling for IBD using principal coordinates of neighbour matrices vectors computed from the distance matrix (Borcard & Legendre 2002).

In addition to omnibus tests using Mantel and dbRDA, we specifically looked at population differentiation at the biphagous *E. aurinia* site (where both host plants were used) by examining the significance of θ_{st}

Table 2 Locations, samples sizes and host use of *Euphydryas editha* populations used in our study. The populations in grey have previously documented adaptive host-associated suites of behaviour (McBride & Singer 2010)

Population		Sample size		Host			Location	
Long name	Code	AFLP	Microsatellites				Latitude	Longitude
Agua Fria	AFL	10	6	Collinsea			37.5	−120.17
Ash canyon	AC		15	Collinsea			41.83	−122.6
Big Meadow	BM	14	13		Castilleja	Pedicularis	35.88	−118.35
Bircham Flat Road	BF	10	20	Collinsea			38.45	−119.445
Colony Meadows	CM	12	13			Pedicularis	36.623	−118.6
Crawford Creek	CC		15	Collinsea			41.167	−123.1
Del Puerto Canyon	DP	13	13		Castilleja	Pedicularis	37.44	−121.49
Dubakella Mtn	DK		15	Collinsea			40.384	−123.141
Ebbets Pass	EP	10	15		Castilleja		38.547	−119.817
Franklin Point	FP		6			Pedicularis	40.052	−122.685
Gold Lake	AU	11	12		Castilleja		39.671	−120.668
Indian Flat	IF	10	7	Collinsea			37.664	−119.841
Knoxville Road	KN		14			Pedicularis	38.47	−122.19
Leek Springs	LK	11	10	Collinsea			38.637	−120.244
Mill Canyon	MIL		10	Collinsea			38.473	−119.51
Mount Dana	MD	9	10		Castilleja		37.9	−119.22
Mud Creek	MUD	6	7	Collinsea			39.9	−121.7
Obrien Bog	OB		14		Castilleja		42.07	−123.72
Piute	PI	9	11		Castilleja	Pedicularis	35.456	−118.384
Powerhouse Road	PH		8	Collinsea			37.143	−119.52
Pozo	PZ		24			Pedicularis	35.3	−120.48
Pratts	PR		17	Collinsea			33.465	−116.64
Road to Emma Lake	REL		6		Castilleja		38.307	−119.448
Rowell Meadows	RO	10	12		Castilleja		36.717	−118.812
Ruth Reservoir	RR		15	Collinsea			40.33	−123.4
Tahoe Meadows	TAH	4	5		Castilleja		39.313	−119.89
Tamarack Ridge B	TR	16	17	Collinsea			37.207	−119.185
Timber Gap/Mineral King	MK	12	12		Castilleja	Pedicularis	36.48	−118.57
Tuolumne Meadows	TM	11	12		Castilleja		37.87	−119.35
Walker Pass	WK		9	Collinsea			35.649	−118.035
Warner Mt-Horseshoe flat	WM		5		Castilleja		41.199	−120.162
Wild Horse Springs	WH	9	9		Castilleja		34.203	−116.769
Yucca Point	YP	10	9	Collinsea			36.824	−118.888
Glen Alpine/Mount Tallac	MT	3	4		Castilleja	Pedicularis	38.88	−120.11
Barbara Trail	BT		4	Collinsea			33.54	−116.7
California Hot Springs	CHS		4	Collinsea			35.883	−118.678
Iowa Hill	IOH	4	4	Collinsea			39.1	−120.93
Monache Meadows	MM		4		Castilleja		36.201	−118.18
Rabbit Meadow	RM	19	21		Castilleja	Pedicularis	36.711	−118.8655
Saddlebag/Gardisky	SRGL	23	18		Castilleja		37.9685	−119.255

and F_{st} values at this site, using 1000 permutations of individuals between the two subpopulations using Arlequin 3.5 (Excoffier 2010).

Analyses were conducted in R using the ecodist package (Goslee 2007) for Mantel tests and multidimensional scaling or the vegan package (Dixon 2009) for dbRDA. The primary analyses focused on monophagous populations. For *E. editha*, we also examined genetic differentiation between monophagous populations feeding on *Collinsia* and monophagous or biphagous populations feeding on any combination of *Castilleja* and *Pedicularis*. The latter contrast was justified by the observation that of the 18 populations that used *Collinsia,* all but one were monophagous, while most populations that used *Pedicularis* also used *Castilleja* (fig. 5 in Singer & McBride 2010).

Analysis of HAD in a subset of E. editha populations with known adaptive suites

Given the surprising lack of HAD among populations of the more highly specialized species, *E. editha*, we

conducted two further searches for HAD in this species only, focusing on a subset of six populations that show clear phenotypic differentiation (Fig. 2; Singer & McBride 2010). These are the populations that use either *C. torreyi* or *P. semibarbata* (bold circles in Fig. 1a; see also fig. 2 in Singer & McBride 2010). Table 2 has names and locations of these and other populations. Due to the large number of markers available, these analyses considered AFLP data only.

First, we used an AMOVA to estimate the proportion of total genotypic variation in the sample attributable to variation among individuals within populations, to variation among populations nested within host and to host affiliation itself. The AMOVA was implemented in Arlequin 3.5 (Excoffier 2010). For this set of populations, we also conducted an IBD analysis using the same methodology as described above.

Our second approach was to look for an association between host adaptation and narrow regions of genomic differentiation (outlier loci) by conducting a comparative genome scan in the same manner as Egan *et al.* (2008) and Nosil *et al.* (2008) using the program Dfdist (Beaumont & Balding 2004). The goal of this method is to detect regions under divergent selection by identifying F_{st} values that are higher than expected given locus-specific heterozygosity and the mean level of differentiation for a given population pair. If adaptation to alternative host plants involves strong divergent selection on limited regions of the genome, then we might expect different-host comparisons to produce more outliers than same-host comparisons, as observed by Egan *et al.* (2008) for *Neochlamissus* leaf beetles. We adopted this approach so that our results would be directly comparable to those from other insect populations that are adapting to alternative host plants. Briefly, we performed the following three steps for each individual population pair. (1) We generated the empirical distribution of F_{st} across all AFLP loci. This is implemented in Dfdist using a Bayesian method in which allele frequencies are estimated from the proportion of recessive genotypes in the sample under the assumption of Hardy–Weinberg equilibrium (Zhivotovsky 1999). (2) We simulated F_{st} for 50,000 'neutral' loci conditional upon heterozygosity. The target neutral F_{st} value used in these simulations was estimated by trimming the highest 30% and lowest 30% of the empirical loci and then averaging F_{st} across the remaining intermediate 40% of empirical loci. (3) We compared the empirical distribution to the simulated 'neutral' distribution in order to identify outlier loci likely to be under divergent selection between populations. The outlier threshold was calculated at the upper 95% quantile using a smoothing parameter of 0.04.

After completing these three steps, we compared the number and identity of outlier loci for pairs of populations that are adapted to the same host to those for pairs of populations that are adapted to different hosts.

Results

E. aurinia showed less phenotypic differentiation than E. editha

Phenotypic differentiation in all three host-adaptive traits – oviposition preference, clutch size and larval performance – was less developed among *Euphydryas aurinia* populations feeding on *Lonicera* and *Succisa* than among the subset of *E. editha* populations feeding on *Pedicularis semibarbata* and *Collinsia torreyi* characterized by Singer & McBride 2010 (Fig. 2). *E. editha* females showed moderate-to-strong preferences for their hosts of origin. While *E. aurinia* from *Succisa* also preferred their native host, the preferences of *E. aurinia* from *Lonicera* showed no such adaptation, with most insects preferring *Succisa* (Fig. 2a). Likewise, while *E. editha* populations showed striking differentiation in clutch size according to host of origin, *E. aurinia* populations did not differ (Fig. 2b). Finally, while *E. editha* larvae from *C. torreyi* are unable to develop on *P. semibarbata*, *E. aurinia* larvae of both origins survived moderately on both hosts (Fig. 2c).

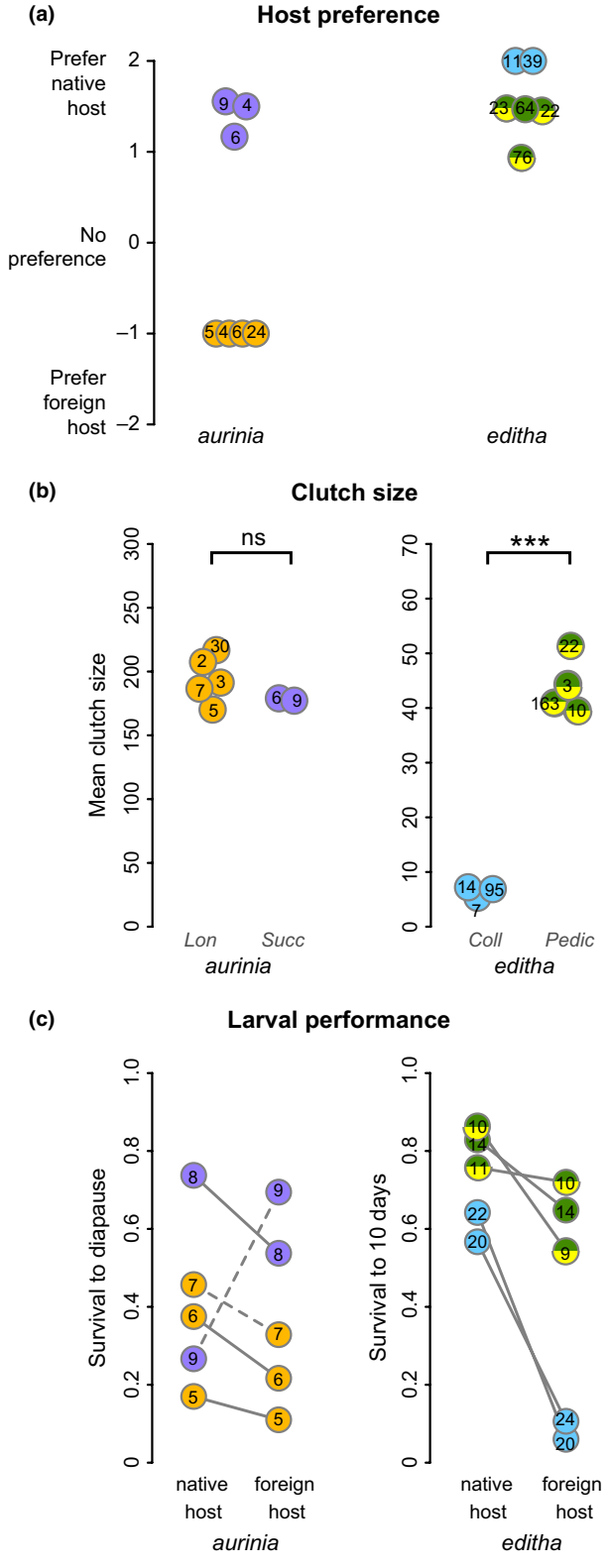
With respect to phenology, we observed substantial phenological overlap between two populations of *E. aurinia* using *Lonicera* and *Succisa* at similar latitude and elevation; there was also overlap between *E. aurinia* using the two plants at the site where both were hosts (Fig. 3). Similar overlap has previously been shown between *E. editha* feeding on *P. semibarbata* and *C. torreyi* (Singer & McBride 2012).

Both species showed significant IBD

Results of the regression of F_{st} on geographical distance are shown in Fig. 4 (for *E. aurinia*: $r = 0.48$, $P = 0.0060$; for *E. editha*: $r = 0.39$, $P = 0.0070$). The slope of the reduced major regression axis was 1.1 (99% CI: 0.70, 1.58) for *E. aurinia* and 0.30 (99% CI: 0.26, 0.34) for *E. editha*. Expressed as the product of effective population size and parent–offspring displacement, these values are 0.072 (*E. editha*) and 0.26 (*E. aurinia*), suggesting that *E. editha* has either greater effective population sizes, which allow for greater gene flow, or greater dispersal tendency (Rousset 1997).

E. aurinia, but not E. editha, showed significant HAD

Contrasting patterns of HAD in the two species can be seen graphically in the NMDS plot in Fig. 5. Monophagous populations of *E. aurinia* showed a strong pattern of HAD according to the both partial Mantel test



(*r* = 0.77, *P* = 0.001) and dbRDA (*P* = 0.02). Even within the biphagous population, the larvae collected from different hosts remained slightly, although significantly,

Fig. 2 Ecological specialization in the two butterfly species. Circles represent population means and are colour-coded by host as in Fig. 1. Numbers inside circles indicate the number of individuals/families sampled per population. *Euphydryas aurinia* data are new, with the exception of host preference data from a single *Lonicera*-feeding population published by Singer *et al.* 2002;. *E. editha* data from a subset of well-characterized populations using *Collinsia torreyi* or *Pedicularis semibarbata* (Fig. 1 bold circles) are shown for comparison and were previously published by Singer & McBride 2010. (a) Host preference: *E. aurinia* populations native to *Succisa* preferred their native host (mean score = 1.4), but those native to *Lonicera* did not (mean score = −1.0). All *E. editha* populations prefer their native host. (b) Clutch size: *E. aurinia* populations native to different hosts did not differ in clutch size (mean = 194 on *Lonicera* and 178 on *Succisa*; *t*-test *P* = 0.29), while *E. editha* populations native to different hosts differed strongly. (c) Larval performance: solid lines connect means for individual populations. Dotted lines connect means for *Succisa*- or *Lonicera*-feeding *E. aurinia* larvae collected from the biphagous population at La Barroca. *E. aurinia* larvae were diverse, with some evidence of increased survival on the native host. *E. editha* populations have consistently, and in some cases dramatically, higher survival on their native host.

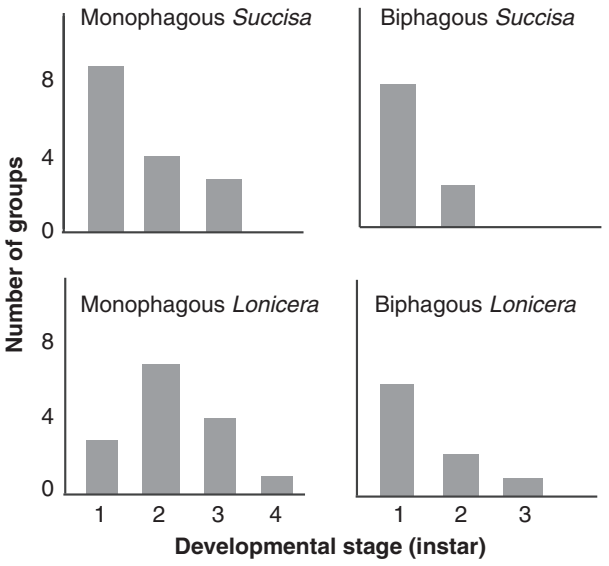


Fig. 3 Phenological overlap between two monophagous populations of *Euphydryas aurinia* on different hosts and between larvae using different hosts at a biphagous site. Two monophagous populations of *E. aurinia*, sampled 1 day apart, showed substantial phenological overlap despite using different hosts (left panels). The same was true of insects sampled simultaneously from the two hosts at the biphagous site (right hand panels). Data are the number of gregarious larval groups in each developmental stage. Equivalent data for two *E. editha* populations using different hosts are given by Singer & McBride (2012). The populations used in the phenological comparison are marked by black outlines in Fig. 1B.

genetically differentiated over the course of 9 years (2001: F_{st} = 0. 11, P = 0.0049; 2010: F_{st} = 0. 031, P = 0.045). This effect was not significant using the AFLP markers alone (2001: θ_{st} = 0.040, P = 0.10), possibly due to the lower sample size of that data set. In contrast, the 22 monophagous *E. editha* populations showed no HAD by either method (Mantel: r = 0.090, P = 0.10; dbRDA: P = 0.60). There was also no HAD detected in the full 40-population *E. editha* data set where *Castilleja*-feeding and *Pedicularis*-feeding populations were pooled (Mantel r = 0, P = 0.45; dbRDA: P = 0.71).

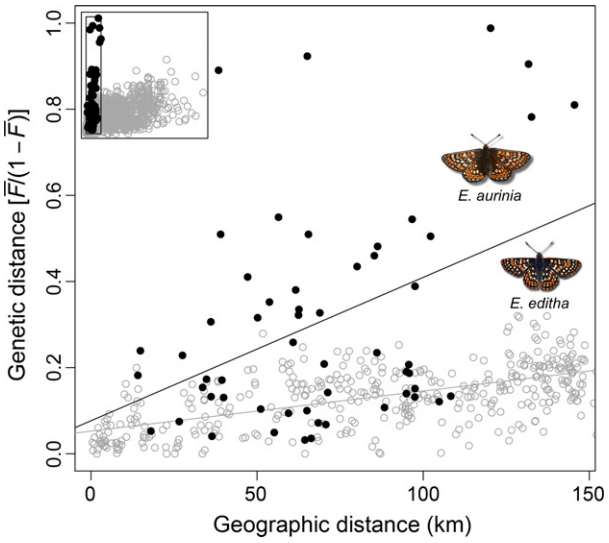


Fig. 4 Isolation by distance in *Euphydryas aurinia* and *E. editha*. Both species showed strong isolation by distance, although the effects were much stronger for *E. aurinia* than for *E. editha*. The main panel shows *E. aurinia* populations and the *E. editha* populations in the same distance range. The inset shows all of the data for both species. *E. aurinia*: closed circles, *E. editha*: open circles.

Further analyses of the subset of six *E. editha* populations in which adaptive phenotypic divergence was previously documented (Singer & McBride 2010) also failed to detect HAD. Both the Mantel test and dbRDA were not significant (Mantel r = –0.043, P = 0.78; dbRDA: P = 0.27). The AMOVA estimated the fraction of total genetic variation attributable to host plant affiliation, differentiation among populations that use the same host and genetic variation within populations. Host plant affiliation accounted for an insignificant fraction of variation (2.9%, P = 0.2). A substantial proportion of the variation was assigned to population differences within host plant categories (11.8%, P < 0.0001). Most of the variation, however, segregated within individual populations (85.3%, P < 0.0001). The Dfdist outlier analysis showed a weak trend in the opposite direction from that expected according to HAD. Same-host comparisons yielded slightly more outliers than different-host comparisons, in terms of both the absolute number of outliers and the mean number of outliers per population pair (Table 3).

Finally, we also assessed the relative degree of HAD in the two species using the partial Mantel correlation coefficient. The 95% confidence intervals of the partial Mantel correlation coefficients of *E. aurinia* and *E. editha* did not overlap under any partitioning of the data,

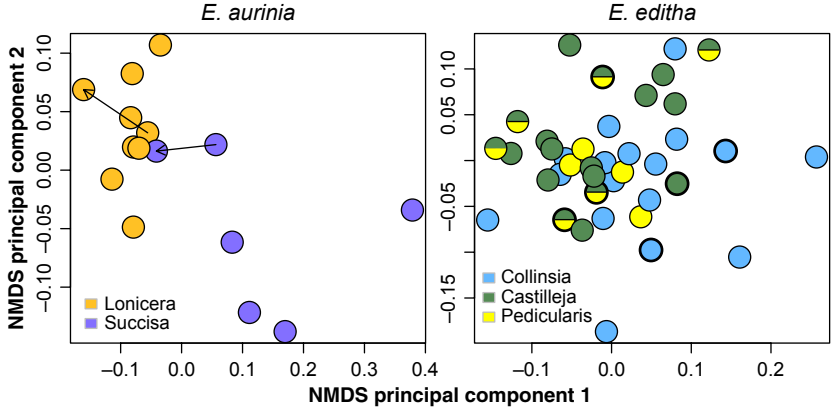


Fig. 5 Nonmetric multidimensional scaling (NMSD) plot of *Euphydryas aurinia* and *E. editha* genetic distances and host use. Using the same types of markers, the two species show contrasting patterns of host-associated genomic differentiation (HAD) – strong in *E. aurinia* and nonexistent in *E. editha*. Butterflies collected from different hosts in the biphagous population of *E. aurinia* showed significant genetic differentiation in two samples, collected 9 years apart. Circles represent populations and are colour-coded by host as in Fig. 1. *Succisa*- and *Lonicera*-feeding individuals from the sole biphagous *E. aurinia* population are shown separately. This biphagous population was sampled in 2001 and again in 2010, as indicated by connecting arrows pointing forward in time. The subset of *E. editha* populations feeding on *Collinsia torreyi* or *Pedicularis semibarbata* and showing a well-defined suite of divergent host adaptations are highlighted by bold outlined circles.

Table 3 Numbers of AFLP loci identified as outliers in genome scans for the 6-population *Euphydryas editha* data set. Outlier loci were categorized as occurring only in comparison with *same*-host populations, only in comparison with *different*-host populations or in comparison with *both* types. Same-host and different-host only outliers were further divided into those that occurred in a single comparison and those that occurred in multiple (repeat) comparisons. *N* = number of comparisons of each type conducted. The table shows slightly more outliers in the same-host comparison, which is a trend in the opposite direction than would be expected under HAD

		Number of outliers			Number of outliers per comparison		
Type	*N*	Single	Repeat	Total	Single	Repeat	Total
Same host	7	29	6	35	4.1	0.9	5
Different host	8	15	12	27	1.9	1.5	3.4
Both hosts	15		29	29		1.9	1.9

which we interpret as statistical evidence that *E. aurinia* had significantly greater HAD than *E. editha* (see DataDryad doi:10.5061/dryad.1v0tb).

Discussion

History of HAD

Thirty years ago, there was a general assumption that host adaptations of herbivorous insects should be detectable from 'neutral' markers, reflecting the view that allozyme correlates of host association in insects such as *Rhagoletis* stemmed from isolation between host-associated forms brought about by host-associated mating (Feder *et al.* 1994). Further investigation revealed additional layers of biological complexity, such as allozyme loci that were either directly involved in host adaptation or hitch-hiking with loci under selection (Feder *et al.* 1997). The most recent work shows that HAD exists in several large areas of the *Rhagoletis* genome, but not in others (Michel *et al.* 2010), and suggests that long-distance migration of chromosomal inversions has assisted adaptation to the novel host, apple.

We might expect the prevalence of HAD to be exaggerated by publication bias, because observations of significant HAD seem more likely to be published than those failing to find it. In this context, there is a role for unbiased studies in which a set of insects is examined without prior knowledge of HAD, and all results are reported. In doing this, Stireman *et al.* (2005) found HAD in about half of the herbivorous insect species attacking Goldenrod (*Solidago*), suggesting that the apparent widespread nature of HAD is not an artefact of publication bias.

Host-associated genomic differentiation (HAD) is extremely variable among herbivorous insects. In some cases it is undetectable, while in others it is sufficiently strong and pervasive to cause reclassification of 'host races' as 'species' (Dorchin *et al.* 2009). It may be restricted to small 'genomic islands' or may spread to affect half of the genome (Via 2009). The existence of such diversity in the strength and extent of HAD raises the possibility that comparison of insects showing different degrees of HAD and different strengths of local host adaptation might be a useful tool for dissecting the stages in evolution of host adaptation and clarifying their relationships between population differentiation and speciation.

Contrasting patterns of host-associated phenotypic differentiation in E. aurinia and E. editha

Prior work has shown that *Euphydryas editha* populations feeding on *C. torrey*i and *P. semibarbata* have diverged in a complex adaptive suite involving not only the obvious traits, host preference and larval performance (Singer 1971; Rausher 1982; Singer & McBride 2010), but also a diversity of additional behavioural and life history traits, including foraging height, egg laying height, partitioning of reproductive output among many small clutches or few large ones, and preference for host phenology (Singer & McBride 2010). These *E. editha* populations were more host-specialized than our study populations of *E. aurinia* in all three traits that we compared: oviposition preference, larval performance and clutch size (Fig. 2).

Contrasting patterns of HAD in E. aurinia and E. editha

Overall patterns of HAD in both species are visualized in Fig. 5. We partitioned our data sets in a variety of ways and consistently found significant HAD in all analyses of *E. aurinia*, but in no analysis of *E. editha*. After adjustment for isolation by distance, the 95% confidence limits for HAD within the two species were nonoverlapping, indicating that the difference in HAD was itself statistically significant (see DataDryad doi:10.5061/dryad.1v0tb).

Previously Descimon *et al.* (2001) reported strong IBD in *E. aurinia* using allozyme markers. We find that monophagous populations of *E. aurinia* showed both significant isolation by distance and a strong genomic

signature of host affiliation after controlling for IBD. This conclusion held whether we used dbRDA or partial Mantel tests. As might be expected by those who criticize partial Mantel tests as being insufficiently conservative (Rousset 2002), the statistical significance of HAD was higher in the Mantel analysis than the dbRDA. The biphagous *E. aurinia* population also showed HAD, but to a much lesser degree than among monophagous sites.

AMOVA analysis of *E. editha* supported the conclusion that populations were differentiated, but gave no indication that this differentiation was broadly associated with host use. Analysis with Dfdist provided no evidence for host-associated selection/differentiation restricted to narrow portions of the genome. None of our analyses of *E. editha* revealed HAD, not within the set of six populations using *Pedicularis semibarbata* or *Collinsia torreyi* and showing known host-adaptive suites, not within the set of 22 monophagous populations using *Castilleja, Collinsia* or *Pedicularis*, nor in the set of 40 populations in which we used just two host-association categories by combining *Castilleja* and *Pedicularis* and comparing populations using either or both of these hosts with those using *Collinsia*. The 40-population analysis gave a HAD estimate of zero after allowance for IBD, despite considerable statistical power from its large sample size.

Why was HAD so strong in E. aurinia?

One possible source of HAD in *E. aurinia* is historical. The geographical region where nearby populations feed on either *Succisa* or *Lonicera* is confined to a relatively narrow (60 km) latitudinal band (Fig. 1b). This distribution suggests a recent contact, of which HAD might be a residual symptom. However, if this is the case, we are puzzled by the persistence of sympatric HAD across 9 years at our biphagous study site. It is possible that frequent immigration to this site has occurred from nearby populations monophagous on *Lonicera*, but unlikely that *Succisa*-feeding insects at the site have been reinforced by immigration from populations monophagous on *Succisa*, because such populations are very rare in lowland Catalonia and none are known nearby. If they reduce offspring fertility, nonrecombining genomic rearrangements, which were ancestrally fixed and are correlated with host, could be one reason for the sympatric persistence of HAD. Alternatively, sympatric HAD may somehow be generated directly by the use of different hosts, rather than by phylogeographical features of the system or by patterns of immigration to the biphagous site.

A possible facilitator of HAD in *E. aurinia* might be strong host fidelity, such as could be generated by host-associated mating (Caillaud & Via 2000; Emelianov *et al.* 2001; Berlocher & Feder 2002). *Euphydryas* larvae typically leave their host plants to pupate, and most matings occur near the eclosion site of the female (Singer & Thomas 1992). Therefore, mating pairs are not normally found on host plants, but scattered through the habitats. In the case of *E. aurinia*, the distribution of territorial males at one of our study sites, Darnius (Catalunya), gave no indication that mate search was host-associated, because males were just as likely to choose patches with hosts as patches devoid of hosts (Singer & Wee 2005). Further, captive insects did not require the presence of hosts in order to mate and showed no barriers to fertile matings between insects from different hosts. Three laboratory matings between insects from a monophagous *Succisa*-feeding site and a monophagous *Lonicera*-feeding site all produced fertile eggs and larvae that survived well to diapause, but we did not attempt to raise them to adults. The same initial success was observed in three laboratory crosses between *E. aurinia* from the two hosts at the biphagous site, La Barroca. These matings between insects from different hosts followed encounters staged by the experimenters and do not exclude the possibility of mating barriers in the field. However, the difference between the strength of allopatric HAD and sympatric HAD indicates that HAD is reduced by some influence of sympatry, an influence for which the most likely candidate is successful mating between insects originating from different hosts.

A trait that affected both host adaptation and mating patterns simultaneously could be another explanation of our result (Kronforst *et al.* 2006; Chamberlain *et al.* 2009). The list of known traits that directly tie mating to ecological adaptation has been increasing (Servedio *et al.* 2011) and such a trait could as yet remain undetected among our *E. aurinia* populations. Pending any such discovery, we have no strong candidate for the generator of HAD that we found in this species.

What does the genetic evidence suggest?

IBD regression slopes can provide insight into dynamics of gene flow. Rousset (1997) has shown that, when appropriately transformed, the slope is inversely proportional to the product of effective population size and parent–offspring displacement, which is one measure of migration rate. Hence our data on IBD cannot estimate both effective population sizes and migration rates, but only the relationship between them. On the assumption that our study species do not differ in spatial variance of selection and given that mark–recapture data suggest that species do not markedly differ in their dispersal ability (see Appendix S1, Supporting information for an expanded discussion), we can infer that *E. editha* has greater effective population sizes than *E. aurinia* (Rousset 1997). Within populations, greater effective population sizes should strengthen the effects of selec-

tion relative to drift and increase the likelihood that host-associated adaptations will develop in *E. editha* rather than in *E. aurinia*. At the same time, higher levels of gene flow between populations should prevent the emergence of HAD among populations. In view of the fact that modelling suggests very restrictive parameters under which HAD might develop (Feder & Nosil 2010; Thibert-Plante & Hendry 2010), these features of the two butterflies may account for much of the difference.

However, in view of the ongoing genetic differentiation at the single biphagous *E. aurinia* population in our study, we should consider alternative explanations for the difference between our study species. Any difference between them in recombination rate could be a factor, because, if one species had a lower recombination rate, this would increase the chances that large sections of the genome would be linked to loci under selection (Kirkpatrick & Barton 2006; Thibert-Plante & Hendry 2010). We should also warm to the suggestion by Bierne *et al.* (2011) that endogenous barriers to gene flow may be important; for example, cytoplasmic incompatibility factors can be associated with host specificity and *Wolbachia* can cause genetic differentiation between sympatric populations of insect hosts (Werren 1997; Werren *et al*. 2008). *Wolbachia* is indeed present in the *E. aurinia* populations at a high level, but there the distribution of multilocus genotypes is not associated with host use (M. R. Smee, Y. Pauchet, P. Wilkinson, C. R. Bulman, M. C. Singer, B. Wee, R. H. ffrench-Constant, M. M. Y. Tin, A. S. Mikheyev & D. J. Hodgson, unpublished).

Our findings run counter to our expectation that our study species should differ in the strength of HAD in the same direction as their difference in degree of host adaptation. Factors other than the traits directly associated with host specialization seem to play a larger role. Future comparative investigations of recombination rates or endogenous gene flow barriers could shed light on the mechanism responsible for the differences we have observed. We hope that the contrast we draw, showing that host-associated genetic differentiation is stronger in the species with less host-associated phenotypic differentiation, will help understand how landscape-level patterns can be used to illuminate evolution of species' interactions.

Acknowledgements

We are most grateful to Jordi Artola, Agnès Batlle, Rafel Carbonell, Henri Descimon, Juexin Jiang, Mark Kirkpatrick, Michael Lockwood, James Mallet, Toni Mariné, Joan Strassman, Tanya Vo and Chris Wheat. Funding was provided by the authors, by the San Diego Foundation and by NSF DEB-0215436 to MCS and Ulrich Mueller and by the Okinawa Institute of Science and Technology. CSM was supported by NSF DEB-0815145 to Michael Turelli during the early stages of this work and by the Howard Hughes Medical Institute.

References

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

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Reviews of Ecology, Evolution and Systematics*, **16**, 113–148.

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

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

Berlocher SH, Feder JK (2002) Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Reviews of Entomology*, **47**, 773–815.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044–2072.

Bohonak AJ (2002) IBD (Isolation by Distance): a program for analyses of isolation by distance. *Journal of Heredity*, **93**, 153–154.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Caillaud MC, Via S (2000) Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *American Naturalist*, **156**, 606–621.

Castellano S, Balletto E (2002) Is the partial Mantel test inadequate? *Evolution; International Journal of Organic Evolution*, **56**, 1871–1873.

Chamberlain NL, Hill RI, Kapan DD *et al*. (2009) Polymorphic butterfly reveals the missing link in ecological speciation. *Science*, **326**, 847–850.

Counterman BA, Araujo-Perez F, Hines HM *et al*. (2010) Genomic hotspots for adaptation: the population genetics of Mullerian mimicry in *Heliconius erato*. *PLoS Genetics*, **6**, e1000796.

DeLeon LF, Bermingham E, Podos J, Hendry AP (2010) Divergence with gene flow as facilitated by ecological differences: within-island variation in Darwin's finches. *Philosophical Transactions of the Royal Society of London*, **365**, 1041–1052.

Descimon H, Zimmermann M, Cosson E, Barascud B, Neve G (2001) Diversité génétique, variation géographique et flux géniques chez quelques lépidoptères rhopalocères français. *Genetics Selection Evolution*, **33**, 223–249.

Dixon P (2009) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, **14**, 927–930.

Dorchin N, Scott ER, Clarkin CE, Luongo MP, Jordan S, Abrahamson WG (2009) Behavioural, ecological and genetic evidence confirm the occurrence of host-associated differentiation in goldenrod gall-midges. *Journal of Evolutionary Biology*, **22**, 729–739. doi:10.1111/j.1420-9101.2009.01696.x.

Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution*, **62**, 1162–1181.

Egan SP, Hood GR, Ott JR (2012a) Testing the role of habitat isolation among ecologically divergent gall wasp populations. *International Journal of Ecology*, **2012**. Article 809897, doi:10.1155/2012/809897

Egan SP, Hood GR, Feder JL, Ott JR (2012b) Divergent host plant use promotes reproductive isolation among cynipid gall wasp populations. *Biology Letters*, **8**, 605–608.

Ehrlich PR (1965) The population biology of the butterfly *Euphydryas editha* II. The structure of the Jasper Ridge colony. *Evolution*, **19**, 327–336.

Ehrlich PR, Hanski I (eds.) (2004) *On the Wings of Checkerspots: A Model System for Population Biology*. Oxford University Press, Oxford.

Emelianov I, Dres M, Baltensweiler W, Mallet J, W, Mallet J (2001) Host-induced assortative mating in host races of the larch budmoth. *Evolution*, **55**, 2002–2010.

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

Excoffier L (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

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

Feder JL, Opp SB, Wlazlo B *et al.* (1994) Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proceedings of the National Academy of Sciences USA*, **91**, 7990–7994.

Feder JL, Roethele JB, Wlazlo B *et al.* (1997) Selective maintenance of allozyme differences among sympatric races of the apple maggot fly. *Proceedings of the National Academy of Sciences USA*, **94**, 11417–11421.

Feldman CR, Brodie ED, Brodie ED III, Pfrender ME (2010) Genetic architecture of a feeding adaptation: Garter snake (*Thamnophis*) resistance to tetrodotoxin bearing prey. *Proceedings of the Royal Society of London series B Biological Sciences*, **277**, 3317–3325.

Ford HD, Ford EB (1930) Fluctuation in numbers and its effect on variation in *Melitaea aurinia* (Rottembourg, 1775) (Lepidoptera: Nymphalidae). *Transactions of the Royal Entomological Society of London*, **78**, 345–351.

Gilbert LE, Singer MC (1973) Dispersal and gene flow in a butterfly species. *American Naturalist*, **107**, 58–72.

Goslee S (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1–19.

Hanski I (1999) *Metapopulation Ecology*. Oxford University Press, Oxford, UK.

Hanski I (2011) Eco-evolutionary spatial dynamics in the Glanville Fritillary butterfly. *Proceedings of the National Academy of Sciences USA*, **108**, 14397–14404.

Harrison S (1989) Long-distance dispersal and colonization in the bay checkerspot butterfly, *Euphydryas editha bayensis*. *Ecology*, **70**, 1236–1243.

Hoekstra HE, Drumm KE, Nachman MW (2004) Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*, **58**, 1329–1341.

Jain SK, Bradshaw AD (1966) Evolutionary divergence among adjacent plant populations. I. The evidence and its theoretical analysis. *Heredity*, **21**, 407–441.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics*, **173**, 419–434. doi:10.1534/genetics.105.047985.

Kronforst MR, Young LG, Kapan DD *et al.* (2006) Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proceedings of the National Academy of Sciences USA*, **103**, 6575–6580.

Legendre P (2000) Comparison of permutation methods for the partial correlation and partial Mantel tests. *Journal of Statistical Computation and Simulation*, **67**, 37–73.

Legendre P, Anderson M (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 1–24.

Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.

Lenormand T, Bourguet D, Guillemaud T, Raymond M (1999) Tracking the evolution of insecticide resistance in the mosquito Culex pipiens. *Nature*, **400**, 861–864.

Lowry DB (2010) Landscape evolutionary genomics. *Biology Letters*, **6**, 502–504.

Mazel R (1986) Structure et evolution du peuplement *d'Euphydryas aurinia* (Rottembourg) Lepidoptera dans le sud-ouest européen. *Vie Milieu*, **36**, 205–225.

McBride CS, Singer MC (2010) Field studies reveal strong postmating isolation between ecologically divergent butterfly populations. *PLoS Biology*, **8**, e10005289. doi 10.1371/journal.pbio.1000529

McCracken KG, Bulgarella M, Johnson KP *et al.* (2009) Gene flow in the face of countervailing selection; adaptation to high-altitude hypoxia in the ßA hemoglobin subunit of yellow-billed pintails in the Andes. *Molecular Biology and Evolution*, **26**, 815–827.

Michel AP, Sim S, Powell THQ *et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences USA*, **107**, 9724–9729.

Mikheyev AS, Vo T, Wee B *et al.* (2010) Rapid microsatellite isolation from a butterfly by de novo transcriptome sequencing: performance and a comparison with AFLP-derived distances. *PLoS ONE*, **5**, e11212.

Nosil P (2009) Adaptive population divergence in cryptic color-pattern following a reduction in gene flow. *Evolution*, **63**, 1902–1912.

Nosil P, Feder J (2012) Widespread yet heterogeneous genomic divergence. *Molecular Ecology*, **21**, 2829–2832.

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

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

Peccoud J, Ollivier A, Plantegenest M, Simon JC (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences of the USA*, **106**, 7495–7500. doi:10.1073/pnas. 0811117106.

Raufaste N, Rousset F (2001) Are partial Mantel tests adequate? *Evolution*, **55**, 1703–1705.

Rausher MD (1982) Population differentiation in *Euphydryas editha* butterflies: larval adaptation to different hosts. *Evolution*, **36**, 581–590.

Roesti M, Hendry AP, Salzburger W *et al.* (2012) Genome divergence during evolutionary diversification as revealed in lake-stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.

Rogers SM, Bernatchez I (2007) The genetic architecture of ecological speciation and the association with signatures of

selection in natural lake whitefish (*Coregonas* sp.) species pairs. *Molecular Biology and Evolution*, **24**, 1423–1438.
Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.
Rousset F (2002) Partial Mantel tests: reply to Castellano and Balletto. *Evolution*, **56**, 1874–1875.
Rousset F, Waller D (2002) Partial mantel tests: reply to Castellano and Balletto. *Evolution*, **56**, 1874–1875.
Servedio MR, van Doorn GS, Kopp M, Frame AM, Nosil P (2011) Magic traits in speciation: magic but not rare? *Trends in Ecology & Evolution*, **26**, 389–397.
Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, **16**, 940–950.
Singer MC (1971) Evolution of food-plant preference in the butterfly *Euphydryas editha*. *Evolution*, **25**, 383–389.
Singer MC, McBride CS (2010) Multi-trait, host-associated divergence among sets of butterfly populations: implications for reproductive isolation and ecological speciation. *Evolution*, **64**, 921–933.
Singer MC, McBride CS (2012) Geographic mosaics of species association: a definition and an example driven by plant-insect phenological synchrony. *Ecology*, **93**, 2658–2673.
Singer MC, Parmesan C (1993) Sources of variation in patterns of plant-insect association. *Nature*, **361**, 251–253.
Singer MC, Thomas CD (1992) The difficulty of distinguishing between resource use and resource preference: an example from hilltopping behavior in butterflies. *American Naturalist*, **140**, 654–664.
Singer MC, Wee B (2005) Spatial pattern in checkerspot butterfly-host plant association at local, metapopulation and regional scales. *Annales Zoologi Fennici*, **42**, 347–361.
Singer MC, Vasco DA, Parmesan C, Thomas CD, Ng D (1992) Distinguishing between preference and motivation in food choice: an example from insect oviposition. *Animal Behaviour*, **44**, 463–471.
Singer MC, Thomas CD, Billington HL, Parmesan C (1994) Correlates of speed of evolution of host preference in a set of twelve populations of the butterfly *Euphydryas editha*. *EcoScience*, **1**, 107–114.
Singer MC, Stefanescu C, Pen I (2002) When random sampling does not work: standard design falsely indicates maladaptive host preferences in a butterfly. *Ecology Letters*, **5**, 1–6.
Slatkin M (1975) Gene flow and selection in a two-locus system. *Genetics*, **81**, 787–802.
Slatkin M (1985) Gene flow in natural populations. *Annual Reviews of Ecology, Evolution and Systematics*, **16**, 393–430.
Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.
Smee MR, Pauchet Y, Wilkinson P *et al*. (2013) Microsatellites for the Marsh Fritillary Butterfly: *De Novo* Transcriptome Sequencing, and a Comparison with Amplified Fragment Length Polymorphism (AFLP) markers. *PLoS ONE*, **8**, e54721.
Smouse P, Long J, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.
Stireman JO, Nason JD, Heard SB (2005) Host-associated genetic differentiation in phytophagous insects: general phenomenon or isolated exceptions? Evidence from a goldenrod-insect community. *Evolution*, **59**, 2573–2587.
Stireman JO, Devlin H, Abbott P (2012) Rampant host- and defensive phenotype-associated differentiation in a goldenrod gall midge. *Journal of Evolutionary Biology*, **25**, 1991–2004.
Thibert-Plante X, Hendry AP (2010) When can ecological speciation be detected with neutral loci? *Molecular Ecology*, **19**, 2301–2314.
Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences USA*, **106**, 9939–9946.
Via S, West J (2008) The genetic mosaic suggests a new role for hitch-hiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.
Wahlberg N, Klemetti T, Selonen V, Hanski I (2002) Metapopulation structure and movements in five species of checkerspot butterflies. *Oecologia*, **130**, 33–43.
Wee P-S (2004) *Effects of Geographic Distance, Landscape Features And Host Association on Genetic Differentiation of Checkerspot Butterflies*. Ph.D. dissertation. University of Texas, Austin, TX.
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **6**, 1358–1370.
Werren JH (1997) *Wolbachia*: master manipulators of invertebrate biology. *Annual Review Of Entomology*, **42**, 587–609.
Werren JH, Baldo L, Clark ME (2008) Wolbachia: master manipulators of invertebrate biology. *Nature reviews Microbiology*, **6**, 741–751.
Wiklund C (1975) The evolutionary relationship between adult oviposition preferences and larval host range in *Papilio machaon* L. *Oecologia*, **18**, 185–197.
Yeaman S, Otto SP (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution*, **65**, 2123–2129.
Zhivotovsky LA (1999) Estimating the population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, **8**, 907–913.
Zimmerman K, Zdenec F, Jiskra P *et al*. (2011) Mark-recapture on large spatial scale reveals long distance dispersal in the Marsh Fritillary. *Ecological Entomology*, **36**, 499–510.

Data accessibility

Sample locations, host use, raw AFLP data, distance matrixes, R scripts and their output: DRYAD entry doi:10.5061/dryad.1v0tb. Raw microsatellite data for *E. aurinia*: DRYAD entry doi:10.5061/dryad.309t3. Raw microsatellite data for *E. editha*: DRYAD entry doi:10.5061/dryad.1540.

A.S.M., C.S.M., U.G.M., C.P. and M.C.S. designed the study. M.R.S., C.S., B.W., C.P. and M.C.S. contributed genetic data for the analysis, or conducted field work. A.S.M. and C.S.M. performed the analysis. A.S.M., C.S.M. and M.C.S. wrote the manuscript.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Population dynamics and dispersal in melitaeine butterflies.