Heredity

npg

ORIGINAL ARTICLE

Quantitative genetics of feeding behavior in two ecological races of the pea aphid, *Acyrthosiphon pisum*

MC Caillaud[1] and S Via[2]

Much of the diversity of herbivorous insects stems from the adaptive divergence of populations onto different host plants. This often involves the evolution of specialized patterns of host acceptance that in turn lead to assortative mating for insects that mate exclusively on their hosts. Here, we explore the genetic architecture of feeding behavior in a herbivorous insect that has become a model for the study of incipient speciation, the pea aphid (*Acyrthosiphon pisum*). We use crosses between individuals specialized to either alfalfa or red clover in order to perform both a biometrical analysis and a quantitative trait locus (QTL) analysis of key feeding behaviors. For each character in each environment, Castle–Wright's estimator for the number of effective factors segregating ranged from 0.11 to 2.54. Similarly, between 0 and 3 QTLs were detected. In one case, a single QTL explained over 50% of the variance in the F2, suggesting that at least one gene (or a complex of tightly linked genes) has a major effect on feeding behavior in the pea aphid. However, the identified QTL explain only 23–73% of the genetic variance for these characters thus additional genes of minor effect are also involved. We found a variety of modes of gene action, including several cases of non-additive gene action. Our results suggest that feeding behavior in pea aphids is neither simple nor highly polygenic. The oligogenetic basis of variation in feeding behavior may facilitate host shifts, providing one explanation for the frequent divergence and speciation of herbivorous insects.
Heredity (2012) **108**, 211–218; doi:10.1038/hdy.2011.24; published online 7 September 2011

INTRODUCTION

Behaviors that affect mate choice have an important role in the process of speciation by determining the chances of meeting between potential sexual partners or the recognition of these partners (Slater and Halliday, 1994; Schluter, 1998). These behaviors have been shown to evolve at a relatively rapid rate, and to frequently be the first phenotypes to vary among recently diverged lineages (Foster and Endler, 1999). For organisms that mate within their resource environment, habitat choice behavior is a form of mate choice, because it leads to *de facto* assortative mating (Bush, 1994; Schluter, 1998). The genetic underpinnings of habitat choice behavior may crucially affect the likelihood of speciation driven by selection in different habitats (Gavrilets, 2004; Via, 2009). Thus, the genetic architecture of habitat choice within species is an important aspect of speciation research.

How adaptation to different ecological environments could reduce gene flow has been examined in great detail in host-plant races of phytophagous insects (Itami *et al.*, 1997; Feder, 1998; Funk, 1998; Via, 1999; Nylin *et al.*, 2004; reviewed in Matsubayashi *et al.*, 2010). Research on plant-feeding insects has actually contributed greatly to the development of a general framework for studying a type of speciation known as 'ecological speciation' (Schluter and Conte, 2009; Via, 2009). What do we know about the genetic architecture of habitat choice, or in this case host-plant preference, in host races of phytophagous insects? First, host preference is usually determined by a few (1–5) loci (reviewed in Matsubayashi *et al.*, 2010). Those loci were not characterized further, except for two genes coding for odorant binding proteins in *Drosophila melanogaster* (Matsuo *et al.*, 2007).

Second, preference genes are often located on autosomes, except in two butterfly systems where Z-linked inheritance was reported. Third, dominance of preference for one host over another was found in about half the reviewed studies. Here, we report on the genetic architecture of a suite of behavioral characters that determine host-plant preference in the pea aphid (*Acyrthosiphon pisum* Harris), a phytophagous insect that has become a model system for the study of ecological speciation (Via, 2009; Peccoud and Simon, 2010). Unlike previous studies, we combine biometrical analyses and quantitative trait locus (QTL) mapping analyses. Also unique to this study is the analysis of different steps of the aphid behavior that eventually lead to the acceptance or rejection of the plant as host.

Pea aphids can be found on numerous legume species. However, this broad host range at the species level does not reflect generalized host use at the individual level—instead, reciprocal transplant experiments have revealed genetically differentiated host races in nearly all cases in which individuals collected from different hosts have been tested on alternative plants (Via, 1991; Simon *et al.*, 2003; Ferrari *et al.*, 2008), and these divergent populations span the continuum between populations and species (Peccoud *et al.*, 2009). A large part of this genetic divergence in host use by pea aphids appears to be caused by variation in habitat choice (Via, 1999; Via *et al.*, 2000). Habitat choice is a complex trait as it involves several sequential steps and multiple behavioral components. This trait was dissected by Caillaud and Via (2000). Winged pea aphids from alfalfa and red clover in New York State (USA) land without discrimination on alfalfa and clover. But if a specialist lands on the alternative host, the aphid rapidly abandons the

[1]Department of Biology, Ithaca College, Ithaca, NY, USA and [2]Department of Entomology, University of Maryland, College Park, MD, USA
Correspondence: Dr MC Caillaud, Department of Biology, Ithaca College, 165 CNS, Ithaca, NY 14850, USA.
E-mail: mcaillaud@ithaca.edu

plant without taking the time to find the phloem, which is the actual food source (Caillaud and Via, 2000). This rapid assessment of plant type involves tasting cells in the leaf or stem tissue nearly as soon as the feeding stylets are inserted into the plant. The decision to accept or reject depends on the recognition of stimulants specific to each host plant, not on deterrents or toxins found in alternate hosts (Del Campo et al., 2003). If very specialized pea aphids are prevented from leaving the alternate host, they will often remain unwilling to even search for the phloem, or ingest phloem sap at all, ultimately starving to death (Caillaud and Via, 2000). We know that pea aphids can experience severely reduced fitness on the alternate host, as shown in 'sham migrations' of winged aphids, which strongly favors the evolution of accurate and rapid habitat detection and choice (Via et al., 2000). The unwillingness to feed from the alternate host is therefore an important component of habitat choice in the pea aphid, and it is particularly important with respect to population divergence and speciation because it decreases random mating between sympatric populations on different hosts.

Here, the unwillingness to feed on non-hosts was dissected into simpler components for genetic analysis. In the honeybee, dissection of the character 'dance communication' showed that this complex behavior was regulated by subsets consisting of simple genic systems (Rinderer and Beaman, 1995). This is also possible for the character of interest here because feeding behavior in aphids involves not only the phloem sap, which is the final food source, but also other plant tissues, which contain plant allelochemicals that may affect plant acceptance or rejection. Different chemical signals, different sensory systems and different subsets of aphid genes, may control feeding. Although most of the feeding behavior involves the movements of the feeding stylets within the plant tissues, we were able to observe these using an electronic monitor, the direct current electrical penetration graph technique (Tjallingii, 1988, described in Caillaud and Via, 2000). Using both a modified version of line-cross analysis (Kearsey and Pooni, 1996; Lynch and Walsh, 1998), and a linkage map (as in Hawthorne and Via, 2001), we analyzed the mode of gene action and estimated the number of effective genetic factors for each of these traits.

MATERIALS AND METHODS
Clones and crosses
We made controlled crosses to analyze the genetic architecture of short-term feeding behavior in the pea aphid, exploiting the fact that pea aphids are cyclically parthenogenetic. After the sexually produced fertilized eggs hatch, progeny reproduce parthenogenetically. This allows replicated phenotypic measurements of each hybrid genotype to be made, increasing the experimental power of the analyses. In purely sexual species, replication of F2 progeny can only be obtained through making recombinant inbred lines, in which inbreeding depression or residual genetic variation may bias analyses.

A reciprocal single-pair cross between two genotypes specialized on different hosts was used for these analyses. The two genotypes were collected in 1989 in Tompkins county (NY, USA) from an alfalfa field (genotype 'A1') and a clover field (genotype 'C1'). These two genotypes were chosen for these experiments because field experiments of demography on both hosts revealed that they typify the ecological specialization of a set of field-collected clones (Caillaud and Via, 2000). F1 hybrids were obtained from reciprocal crosses between these two parental genotypes, and two F1 were then crossed to produce the F2 generation. The two parental clones, the two F1 and the F2 were then used in a line cross analysis (Kearsey and Pooni, 1996; Lynch and Walsh, 1998).

The genotypes crossed to produce either F1 or F2 generations were induced to form sexual morphs through exposure to a declining photoperiod and a cold temperature. Note that each clone produces sexual females, all with the same genotype, and multiple XO males that differ in which chromosome X they bear. Ten replicate mating dishes of two males and three females for each direction of the cross were established so that the F1 and F2 progeny would have, on average, equal numbers of the two possible male-transmitted X chromosomes. All fertilized eggs produced over life of the females were harvested, surface sterilized and placed in an incubator under daily temperature cycles: $4\,°C$ during a 12-h day and $0\,°C$ during a 12-h night. After about 100 days of this winter treatment, eggs were removed from the incubator and the hatchling progeny were reared in Petri-dishes containing both alfalfa and clover foliage. These progeny reproduce clonally under long-day photoperiod, and thus a parthenogenetic lineage was established for each F1 or F2 clone. Each parthenogenetic lineage was then maintained individually in a container containing two 7 cm square pots planted with both clover and alfalfa.

Phenotypic characterization
We have previously identified the specific characters that explain differences in feeding behavior between the parental genotypes A1 and C1 (Caillaud and Via, 2000). Here, we evaluate four of these characters on both host plants in A1, C1, 2 F1 hybrids and 102 F2 hybrids. The characters are: the amount of time spent on a plant, but without penetrating it in search of food (NON-PEN, duration of waveform NP in direct current electrical penetration graph (see Caillaud and Via, 2000 for pictures of the various waveforms seen in feeding monitor trials)), the total time spent searching for feeding sites (SEARCH, duration of waveform C), the time before an aphid started injecting saliva in the phloem vessels to prepare the ingestion phase (START PREPARATION, time to waveform E1) and the amount of time spent ingesting sieve sap (INGESTION, duration of waveform E2) (Table 1). Typically, aphids deposited on host plants (C1 on clover for instance), search actively for the feeding sites (long SEARCH) and spend at least a third of their time on the plant ingesting nutrients (long INGESTION) (Caillaud and Via, 2000). In contrast, aphids deposited on non-host plants (C1 on alfalfa for instance) spend most of their time 'sitting' on the plant (long NON-PEN), and do not ingest nutrients (short INGESTION).

Aphid stylet activities during plant penetration were recorded for 390 min using the direct current electrical penetration graph technique as described in Caillaud and Via (2000). The aphid and the plant were included in an electrical circuit. The penetration of the aphid stylets in the plant modified the characteristics (voltage) of the electrical signal recorded, which provided

Table 1 Biological significance of four characters typifying behavioral divergence between parents A1 and C1

Behavioral characters	Biological significance	Trend when suitable for feeding	Trend plant is rejected
NON-PEN	Time spent non-penetrating the plant tissues (the aphid is either immobile on the plant or walking)	Short	Long
SEARCH	Time spent searching actively for the nutritional tissues	Long	Short
START PREPARATION	Time to the first attempt to inject saliva in the phloem vessels to prepare ingestion	Short	Long
INGESTION	Time spent ingesting phloem sap from the phloem vessels (mostly sugars and aminoacids)	Long	Short

A plant is suitable for feeding if 'NON-PEN' and 'START PREPARATION' are short while 'SEARCH' and 'INGESTION' are long.

reliable information about the behavior (ingestion, salivation) and the stylet tip position (phloem, xylem, and so on) of the aphid during plant penetration. The electric characterization and biological significance of these waveforms has been carefully calibrated (Tjallingii, 1988). This method permitted us to follow the activity of aphid mouthparts inside the plant tissues as they penetrated the epidermis, reached the phloem vessels, injected saliva in these vessels to prepare the ingestion step and eventually ingested phloem sap.

In a given day, four genotypes were tested on each of the two host plants. Aphid positions were randomized within each day to avoid any effect of their location in the Faraday cage. We performed 9–14 replicates per host for parental genotypes, 3–5 for 2 F1 genotypes and 2–5 for 121 F2 genotypes (total of 792 observations).

All experiments involved wingless adult aphids, starved for 3 h before experiments in order to increase their reactivity to the plant and to standardize the pre-experimental physiological conditions. Although winged aphids are responsible for the majority of host plant acceptance–rejection in nature, it is quite difficult to attach a gold wire to their thorax without interfering with their mobility. The feeding behavior of wingless individuals on the host plants of this study is not significantly different from the one of winged individuals (Flory and Caillaud, unpublished (Supplementary Figure 1). The recordings of the individuals that died (3%), or escaped (9% because of breakage of the gold wire) before the end of the recording period, were discarded. Experiments were performed at the same time every day to remove any effect of diurnal rhythm on the results. To even further reduce the effect of environmental variance, all aphids were set up by the same person and analysis of the recordings was done blind.

Temperature was held at $21 \pm 1\,°C$ and aphids were under continuous and homogeneous artificial illumination coming from above (fluorescent tubes, 1500 lux). Plants were grown from seeds in 2.5″ pots. Plants were maintained in a growth chamber at $20\,°C$ and 16:8 light:dark, and were 5–6 weeks old at the time of the experiments.

The behavioral data were examined for normality (Proc Univariate; SAS Institute Inc., 1990). START PREPARATION and INGESTION were log $(x+1)$ transformed before analysis. Line-cross and QTL mapping analyses were performed on least square means for parental lines and F1 hybrid lines, and best linear unbiased predictors for F2 hybrid lines (PROC MIXED, SAS Institute Inc., 1990). We were unable to examine the behavior of all 121 F2 hybrids on the same day, and observations were grouped into 'blocks' in our experimental design. The behavior of parental genotypes A1 and C1 on both alfalfa and clover was tested in each block (two replicates per plant). We compared the feeding behavior of A1 and C1 across blocks using PROC MIXED (SAS Institute Inc., 1990) where 'plant' and 'parent' were fixed effects while 'block' and all interactions including 'block' were random effects. A 'block' effect for a given behavioral character would suggest that part of the variation between the different clones studied is due to the fact that they were tested on different days. We will mention 'block' effects only when they were significant.

Line-cross analysis
To determine which behavioral characters showed significant genetic variation among the F2 hybrids, an analysis of each character on each plant was performed (PROC MIXED, SAS Institute Inc., 1990). 'F2 hybrid' was considered as a random factor because the F2 individuals studied are only a sample of the possible F2s that could have been obtained, while 'parent' and 'F1 hybrid' were considered as fixed because they were specifically chosen. No analysis of gene action or estimate of number of loci was attempted on traits for which there was no significant segregation variance among F2.

We tested the generation means for goodness of fit to genetic models incorporating additive or additive-dominant effects using a Joint scaling test as described in Kearsey and Pooni (1996). This test has been shown to be applicable to non-homozygous lines (as A1 and C1 are) in the absence of mating between close relatives (Lynch and Walsh, 1998). This method uses the means and variances of behavioral characters for three generations (parents, F1 and F2 hybrids) to derive estimates of the composite (that is, net) additive and dominance effects on the phenotypic difference between the parents for each plant. Here, we implemented the weighted least square method as described in Lynch and Walsh (1998). The object is to explain the variation between the observed generation means with as simple a model as possible. We started with the simple model involving only net additive effects ($Wt_i.y_i=wt_i(m+a.x1_i)$),

tested its significance (χ^2_a), then gradually added higher-order composite effects to the model until no further significant improvement in the model fits occurred. We added a composite dominance effect first ($Wt_i.y_i=wt_i(m+a.x1_i+d.x2_i)$) ($\chi^2_{aXd}$). If the additive-dominance model was rejected, we proceeded on the analysis of a model containing an epistatic component (an additive by additive composite effect: $Wt_i.y_i=wt_i(m+a^{\star}x1_i+aa^{\star}x3_{i...})$)($\chi^2_{aXa}$). For the latter model, and because we have only four generations, we had to drop the composite dominance effect from the model. This was possible only when dominance effects did not improve significantly the fit of the model, which we tested by calculating the likelihood ratio test statistic: $\wedge=\chi^2_a-\chi^2_{aXd}$ (Lynch and Walsh, 1998). The weighted regression analysis and the χ^2 test were implemented using PROC REG (SAS Institute Inc., 1990).

The minimum number of segregating factors involved in genetic divergence in behavior between alfalfa and clover specialists (n_e) was estimated using the method first developed by Castle (1921), then modified by Lande (1981) for use with non-homozygous populations. We use here the method suggested by Cockerham (1986) that corrects for sampling variances in the estimates of parental populations (equation 9.27 in Lynch and Walsh, 1998). The sampling variances of the parental means and the segregational variance were estimated using a weighted least squares procedure comparable to the one used to estimate the composite effects of the generation (PROC REG; SAS Institute Inc., 1990). After computing n_e using the Cockerham equation (1986), we substituted this estimate in an expression suggested by Zeng (1992) that takes into account possible linkage (c) and unequality of allelic effects (C_α). C_α is the squared coefficient of variation of effects and c is the average recombination frequency between random pairs of loci throughout the genome. Estimates of c using this approximation were shown to not greatly differ from the more refined estimates that can be obtained with a genetic map (Lynch and Walsh, 1998). C_α is equal to 0 when all loci have equal effects. Nor C_α neither c are known for the pea aphid. A downwardly biased estimate of c is given by: $c=(M-1)/2M$, where M is the haploid number of chromosomes.

QTL mapping analysis
We used the same cross and the same linkage map of pea aphids that Hawthorne and Via (2001) used to map QTL affecting fecundity and acceptance (measured by looking at the location of winged individuals 70 h after release in a cage) on alfalfa and clover. This linkage map was made mostly from dominant AFLP markers, requiring a separate map for each parental genome. Linkage groups Ia–IVa pertain to the parental genotype specialized on alfalfa (A1). Linkage groups Ic–IVc pertain to the parental genotype specialized on clover (C1) (Hawthorne and Via, 2001). Seven co-dominant sequence-tagged AFLP markers allowed alignment between the two maps.

The genotype of 102 F2 progeny was assessed using 116 AFLP markers as described in Hawthorne and Via (2001). Primer information for all these primers is provided in Supplementary Tables 2 and 3. The F2 phenotype was assessed using the EPG technique as described earlier. QTL affecting feeding behavior were identified using composite interval mapping (Zeng, 1994) and model 6 of QTL Cartographer (Basten et al., 1996). In our analysis, the size of the 'conditioning window' used around the test interval was 15cM. The significance level of the likelihood ratio for each analysis was determined by permutations (Churchill and Doerge, 1994; Doerge and Churchill, 1996). One thousand permutations of the aphid phenotypes with respect to genotypes were performed for each chromosome and each trait separately because there is evidence of segregation distortion in the pea aphid, particularly for linkage group I, which inflates the permutation maximum likelihood ratio statistics (Doerge, personal communication). When the permutation maximum likelihood ratio statistics < 50 times was significant at $\alpha=0.1$ under the null hypothesis, the QTL was considered as suggestive. Estimates of the additive and dominance effects of each behavioral QTL (a, d), as well as the proportion of the variance explained by the QTL conditioned on the background markers (r^2), were given by QTL cartographer.

RESULTS
Descriptive statistics
No significant variation was found between EPG recordings performed on either of the two parental genotypes (A1 and C1) between

the beginning and the end of the experiment 18 months later. This suggests that significant variation among F2 tested at different times during this period can be reasonably attributed to genetic variation and not to environmental variation. The descriptive statistics of the F2 generation are presented in the Supplementary Table 1. Genetic variation in the F2 generation was found for all characters. Variance between the F2 represented between 9.3% (SEARCH on clover) and 32.7% (INGESTION on clover) of the total phenotypic variance.

Line-cross analysis

If alfalfa and clover specialists have diverged primarily in genes with additive effects, then hybrid means for all generations should fall along the dotted lines joining the observed parental means in Figure 1.

The extent to which the hybrid means are displaced from this line is proportional to the degree of dominance. In the presence of epistasis, the displacement for the F1 hybrids and the F2 hybrids is comparable. Hybrid means fall along the line for three characters: SEARCH in the alfalfa habitat (Figure 1b), START PREPARATION in the clover habitat (Figure 1g) and INGESTION in the clover habitat (Figure 1h), suggesting that these characters are almost completely additive. A joint scaling test reveals that the best model for explaining line means for these characters is indeed the simple additive model (Table 2). For all other characters, F1 hybrid means are displaced from the line. The additive-dominant model fits best the hybrid means of two characters: START PREPARATION in the alfalfa habitat (Figure 1c) and NON-PEN in the clover habitat (Figure 1e). In contrast, the
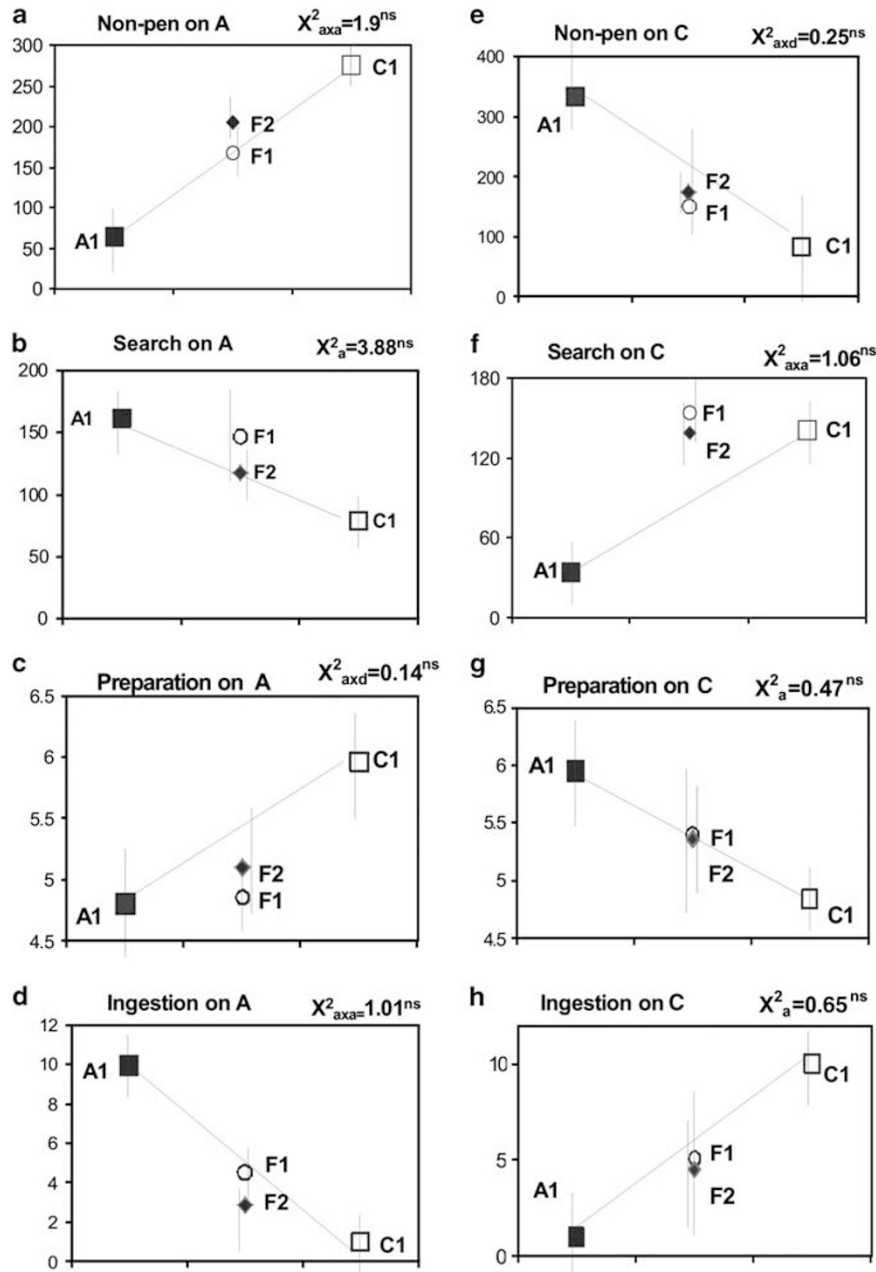
Figure 1 Observed character means and standard errors for the four behavioral traits measured on alfalfa (**a–d**) and clover (**e–h**), in four lines (parentals A1 and C1, hybrids F1 and F2). *A priori* expectations of additive gene effects are represented by the lines drawn between parental means A1 and C1.

Table 2 Maximum likelihood estimates of four components of the generation means (m, a, d and aXa) and test of the significance of models incorporating these components (χ^2 test)

Behavioral characters	Alfalfa					Clover				
	m	a	d	aXa	χ^2	m	a	d	aXa	χ^2
NON-PEN										
y=m + xa	198.7	−123.9			6.69	178.5	120.1			11.3
y=m + xa + xd	200.1	−114.9	15.66		5.14	**174.3**	**124.5**	**−32.2**		**0.25**[NS]
y=m + xa + xaa	**202.7**	**−105.7**		**−32.8**	**1.9**[NS]	173	124.8		35.3	0.7
SEARCHING										
y=m + xa	**119.7**	**41.1**			**3.88**[NS]	133.1	−64.4			38.3
y=m + xa + xd	119.5	43.3	15.2		1.82	136.7	−55.3	42.1		3.91
y=m + xa + xaa	119.7	40.9		0.35	3.88	**139.3**	**−53.6**		**−51.7**	**1.06**[NS]
START PREPARATION										
y=m + xa	5.15	−0.53			9.26	**5.33**	**0.56**			**0.47**[NS]
y=m + xa + xd	**5.1**	**−0.56**	**−0.27**		**0.14**[NS]	5.32	0.54	−0.05		0.1
y=m + xa + xaa	5.08	−0.56		0.32	1.2	5.31	0.53		0.07	0
INGESTION										
y=m + xa	3.12	4.29			5.58	**4.55**	**−4.87**			**0.65**[NS]
y=m + xa + xd	3.1	4.62	−1.01		3.98	4.51	−4.92	−0.38		0.31
y=m + xa + xaa	**3.03**	**4.9**		**1.99**	**1.01**[NS]	4.49	−4.95		0.56	0.09

Abbreviation: NS, nonsignificant.
Bold entries correspond to the model that best fits the data.

additive and additive by additive epistatic gene action explain best genetic divergence between A1 and C1 for NON-PEN in the alfalfa habitat (Figure 1a, Table 2), SEARCH in the clover habitat (Figure 1f, Table 2) and INGESTION in the alfalfa habitat (Figure 1d, Table 2).

Biometrical estimates of gene number using equation the Castle (1921) estimator, modified as suggested by Cockerham (1986), varied little between characters and ranged from 0.11 to 2.54 (Table 3). An estimate of c is 3/8 for pea aphids. As for C_α, its estimation is more elusive. With a C_α equal to 1, Otto and Jones (2000) found values of n_e very close to the real number of simulated underlying loci (n) when $n \leqslant 20$. When $n=100$, n_e underestimated n and only 35% of the loci involved were detected. In the pea aphid, using equation 9.27 of Lynch and Walsh (1998) with $c=3/8$ and $C_\alpha=1$, we find n_e between 0.43 (SEARCH on alfalfa) and 9.5 (INGESTION on clover) (Table 3).

QTL mapping analysis

On alfalfa, no QTL was found for START PREPARATION on alfalfa, suggesting the involvement of many genes of small effect. For all other behavioral characters, we found 1–3 QTL spread over all four linkage groups (Table 4). The proportion of variance explained by each QTL varied from 7.3% to 52.1%. For each trait, and in each plant environment, all detected QTLs explained between 23.3 and 73.8% of the genetic variance for that trait in that environment (Table 4).

Most of the QTL have positive additive effects on characters expressed on the 'native host' and negative additive effects on characters expressed on the alternate host (Table 4). That is, for instance, QTL for SEARCH in A (that is, alfalfa) on Ia (the alfalfa homolog for linkage group I) increases the time spent actively looking for nutritional tissues on alfalfa (thus increases feeding), while the QTL for SEARCH in A on IIIc (the clover homolog for linkage group III) decreases that time (thus decreasing feeding). The three exceptions were found for behavioral characters in the clover environment. The QTL for NON-PEN on clover mapped onto IVc has an effect opposite

Table 3 Estimates of gene number (n_e) for the four behavioral characters, in each parental habitat

Behavioral parameters	Alfalfa			Clover		
	n_e	Gene effect[a]	n(QTL)[b]	n_e	Gene effect[a]	n(QTL)[b]
NON-PEN	1.6	aXa	3	2.02	aXd	2
SEARCH	0.77	a	3	1.005	aXa	2
START PREPARATION	0.14	aXd	0	0.11	a	3
INGESTION	0.56	aXa	2	2.54	a	1

[a]From Table 2.
[b]From Table 4.

to the predicted effect because its directionality is positive thus decreasing feeding on clover. Similarly, QTLs for START PREPARA-TION and INGESTION on clover mapped onto IIIc have a direction-ality that decreases feeding on clover instead of increasing it.

In some cases, QTLs for several behavioral characters colocalized. For example, on linkage group Ia, between chromosomal positions 28 and 31cM, we found a QTL that decreases NON-PEN on alfalfa while increasing SEARCH and INGESTION on that plant (Table 4). Similarly, on linkage group IIIa, between positions 31 and 33cM, a QTL decreases NON-PEN on alfalfa while increasing SEARCH on alfalfa. Nevertheless, unique QTLs were detected for each character. For instance, SEARCH and NON-PEN on alfalfa are influenced by other QTLs than the common ones mentioned above, located on linkage groups IIIc and Ic, respectively (Table 4).

DISCUSSION

One of our main goals was to distinguish between two alternative genetic models for the evolution of adaptation. We asked: can

Table 4 QTL position and effect on the phenotype. START PREPARATION and INGESTION were transformed before analysis

Behavioral characters	Plant	Peak LR			Effect		% Var explained	Effect on behavior
		Linkage group	cM	LR	a	d		
NON-PEN	A	Ia	30.8	18.9	−43.3	−23.1	**52.1**	+ on A
		IIIa	32.7	13.1	−6.84	39.2	14.4	+ on A
		Ic[a]	108.6	10.6	20.2	5.6	7.3	− on A
	C	IIc	36.1	12.9	−7.1	−4.21	8.8	+ on C
		IVc	4.3	16.6	11.1	3.05	14.5	− on C
SEARCH	A	Ia	30.8	13.4	21.8	−2.44	14.9	+ on A
		IIIa	31.1	11.2	6.1	−23.7	11.1	+ on A
		IIIc	11.48	12.5	−21.7	8.29	10.7	− on A
	C	IIa	48.2	10.9	−6.4	−1.34	13.4	− on C
		IIIa[a]	11.5	10.1	−15.3	17.3	10.8	− on C
START PREPARATION	A	None	—	—	—	—	—	—
	C	IIa	30.2	11.7	0.041	0.003	11.1	− on C
		IIc[a]	47.7	10.8	−0.04	−0.013	7.9	+ on C
		IIIc[a]	76.8	12.9	0.05	0.09	26.4	− on C
INGESTION	A	Ia	28.8	14.7	0.17	−0.06	16	+ on A
		Ic[a]	84.1	15.7	−0.17	0.18	30.9	− on A
	C	IIIc[a]	74.8	14.6	−0.8	−1.57	35.7	− on C

Abbreviations: A, alfalfa; C, clover; LR, likelihood ratio; QTL, quantitative trait locus.
The bold entry corresponds to the QTL that explains the highest percentage of the variance found in the F2 population.
Linkage group names and location are from Hawthorne and Via (2001).
[a]For suggestive QTLs ($P=0.1$).

variation between populations in ecologically important traits be explained by allelic differentiation at a few major genes, or are key differences between populations because of many genes, each with a relatively small phenotypic effect? If divergence between A1 and C1 is caused by one-to-few major segregating factors with large effects on the phenotype, then host shifts could occur relatively rapidly through allelic change at only a few loci, providing a reasonable mechanism for the rapid diversification of herbivorous insects. Also, if divergence between A1 and C1 is caused by a one-few major segregating factors, then a search for molecular markers tightly linked to these loci is likely to be successful, and the eventual molecular characterization of these loci is likely to be possible. The biometrical estimates of effective genetic factors are strikingly low (Table 3). Similarly, QTL mapping analysis detected 0 to 3 QTL per character in each plant habitat (Table 4). One of these QTL maps on linkage group Ia and explains 52.1% of the genetic variance between specialized parents A1 and C1. It is tempting to speculate that a major gene influencing feeding behavior segregates in our F2 generation. True et al. (1997) have proposed a quantitative definition of a major gene effect as one for which the distributions of alternative homozygotes show little overlap such that the probability of misclassifying them is <0.05. When both homozygous classes are normally distributed with equal variance, the probability of misclassification is 0.05 when the means are 3.28 s.d. apart. For the character NON-PEN on alfalfa, the parental means are 5.61 s.d. apart, so a major gene would have an effect that explains 58.4% of the parental difference. The QTL for NON-PEN on alfalfa mapped onto Ia has an effect estimate of 52.1%. This may very well be a major gene. However, an inherent bias in QTL analyses comes from

the fact that the same data are used to detect QTL and to determine their effect sizes (Beavis, 1998). As a consequence, effect sizes of detected QTLs are usually overestimated. For instance, in a simulation study involving 20 underlying QTLs and 500 F2's, Otto and Jones (2000) found that the 9.65 detected QTLs appeared to explain 100% of the parental difference, on average, although less than half of the underlying QTLs were detected. In the case of the pea aphid, this means that the QTL on linkage group Ia that explains 52.1% of the parental difference may actually have a smaller effect on the phenotype. Several studies have suggested the presence of genes with major effects on behaviors that may be involved in feeding, such as olfactory avoidance and foraging behavior in larvae of Drosophila melanogaster (reviewed in Anholt and MacKay, 2004) and adults of Apis mellifera (Hunt et al., 1995). However, few have used quantitative genetic methods to estimate gene numbers, and the QTL analysis of feeding behavior in pea aphids remains, to our knowledge, the only mapping study of this key behavioral trait in herbivorous insects.

The 1–3 QTLs detected for a given behavioral character, in a given plant environment explain 23–73% of the parental difference, thus showing that other QTL having a minor effect on feeding behavior have not been detected. Our QTL mapping analysis involved a limited number of recombinant genotypes ($n=102$). Although we increased the experimental power of our genetic analysis by replicating phenotypic measurements for each F2 genotype, on each plant, we only increased it by 1.73. Limitations in the size of the segregating generation in which the QTL mapping analysis is performed is known to lead to underestimates of the total number of loci involved (Beavis, 1998). With a small F2 population, QTLs that explain large

fractions of the phenotypic difference between parental lines can be detected but loci of more subtle effect can be overlooked. In other words, feeding behavior in the pea aphid appears to have an oligogenic basis.

Thus, neither the single gene, nor the infinitesimal model of genetic adaptation, explains feeding behavior in pea aphids. An oligogenic basis of phenotypic divergence has also been reported in several studies of host-plant-associated traits. In the first one, Jones (1998) used a chromosomal assay to show that differences in performance between *D. seychellia* and *D. simulans* on the toxic *Morinda* fruit was influenced by five small regions of large phenotypic effect while large regions of the chromosomes have no effect on this trait. He then used these results to suggest that adaptation to *Morinda* was neither monogenic nor highly polygenic. In the second one, Sezer and Butlin (1998b) found that a small effective number of loci (n_e) underlied differences in oviposition preference between two sympatric host races of the brown planthopper *N. lugens*. Craig *et al.* (2001) used Mendelian genetics to show that differences in host preference between two sympatric host races of the fly *Eurosta solidaginis* was inherited at a limited number of loci. Last, Dambrosky *et al.*, 2005 used segregation patterns in F2 and backcross hybrids to show that only a modest number of allelic differences at a few loci may underlie host fruit odor discrimination in host races of *Rhagoletis pomonella*. In the pea aphid, we reach a similar conclusion but using QTL mapping. In addition, we provide suggestive evidence that a major gene influencing feeding behavior segregates in our mapping population. This type of genetic architecture may facilitate the rapid and repeated evolution of host range in herbivorous insects. Allelic diversification in just a few genes of large effect could allow initial establishment on the new host, followed by the spread of novel alleles at other minor loci that improve the 'quality' of this establishment.

Interestingly, the putative major QTL is on the X chromosome. Several other studies have reported a strong sex-linked component of host-associated behavioral traits. In two closely related species of the swallowtail butterfly (*Papilio* spp.), host preference was shown to be X-linked (Thompson, 1988). In two geographical populations of another butterfly, *Polygonia c-album*, Janz (1998) found that population differences in oviposition preference was also X-linked. Traits encoded on the X chromosome are believed to evolve at a faster rate, under natural selection, than autosomal loci, if mutations are fully or partially recessive because these alleles are shielded from selection by the heterozygous condition and exposed to selection when in the hemizygous state. In other words, X-linkage may function as a means to keep important combinations of traits intact in the face of recombination and gene flow. More studies on a variety of herbivorous insects are needed to examine the particular role that genomic regions on the X chromosome could have in the evolution of insect–plant interactions. In the case of aphids, the location of this putative major QTL on the X chromosome may facilitate future attempts to use positional cloning because males of pea aphids are haploid and a dense linkage map of this chromosome can be obtained relatively easily.

Moreover, this putative major QTL affecting NON-PEN on Alfalfa is located on the same chromosomal segment, between AFLP markers C6–675 and E8–330, as two other characters measured previously by Hawthorne and Via (2001). These characters are 'Fecundity on Alfalfa' and 'Acceptance of Alfalfa'. 'Fecundity' was measured as the number of progeny during the first 9 days of adult life. 'Acceptance' was measured as the percentage of winged aphids that were on a plant with offspring 70 h after being released in a cage containing both clover and alfalfa plants (Hawthorne and Via, 2001). An analysis of genetic correlations between all these characters and another mapping study using a dense linkage map would help determine whether this colocalization is due to pleiotropy of a master locus, or to the presence of tightly linked loci affecting each character independently.

What about QTL affecting other behavioral steps besides NON-PENETRATION? Monitoring aphid stylets as they penetrate the plant tissues gives us the opportunity to identify unique genes that would otherwise be invisible if the behavior was analyzed with less detail. Our study suggests that they are indeed different loci influencing different parts of the exploration of the plant. The putative major gene located between AFLP markers C6–675 and E8-330 on the X chromosome, influences NON-PENETRATION, SEARCH and INGESTION on Alfalfa. But this locus does not influence START PREPARATION that is a step at which the stylet tips are inside the phloem vessels and intense salivation occurs in preparation for the switch to phloem ingestion (Table 4). This suggests that hybrid acceptance of alfalfa and clover could involve two events, one related to plant chemistry before the phloem vessels are reached, and another related to plant chemistry once the stylet tips are inside the phloem vessels.

Our other major goal was to investigate the mode of gene action on traits associated with host choice in pea aphids. The biometrical analysis revealed that a simple model including only additive effects explains the behavioral divergence between alfalfa and clover specialists for three of the eight characters (four traits on each of two hosts). A model including additive plus dominance effects were sufficient for two traits, while additive plus epistatic effects were required to explain divergence in the final three traits (Figure 1 and Table 2). This diversity of gene action is consistent with previous applications of line-cross analysis that suggest non-additive gene action is almost always involved in the case of differentiation of divergent lines (Hard *et al.*, 1992; Fenster *et al.*, 1997; Hatfield, 1997; Lynch and Walsh, 1998; Sezer and Butlin, 1998a). Non-additive genetic effects on behaviors associated with host-plant were also detected in two other phytophagous insects, the seed-feeding beetle *Callosobruchus maculates* (Fox *et al.*, 2004), and the soapberry bug, *Jadera haematoloma* (Carroll *et al.*, 2001).

One striking result shown by the QTL mapping analysis, is that all but 3 of the 16 estimated QTL alleles alter the phenotype toward the form of specialization predicted by the parental phenotypes (Table 4). For most quantitative traits, there is a mixture of plus and minus alleles in each parental line (Tanksley, 1993). For example, 36% of the QTLs detected in a cross between two tomato species had effects opposite to those predicted by the parental phenotypes (DeVicente and Tanksley, 1993). In *D. mauritiana* and *simulans*, QTL analysis of male-specific bristle number traits also revealed a mixture of plus and minus alleles (True *et al.*, 1997). In contrast, a rare situation was described in another cross between *D. mauritiana* and *simulans*: all but 1 out of the 19 additive effect estimates showed the same positive sign (Zeng *et al.*, 2000). In this context, Zeng *et al.* (2000) hypothesized that a strong preponderance of allelic effects in only a single direction reflected a history of unusually strong directional (or in our case, divergent) selection. The consistent directionality of the allelic effects on key traits in divergent pea aphid host races seen here and in Hawthorne and Via (2001), supports this hypothesis. We know that divergent selection on various components of host use in pea aphids is very strong (Via *et al.*, 2000), as required for genetic divergence between populations that utilize different environments with no physical barriers to gene flow (Via and West, 2008; Via, 2009). Thus, it seems likely that future analyses in other systems will reveal that consistent directionality of allelic effects on ecologically important traits is a general feature of the genetic architecture of taxa that have diverged under strong selection in the face of gene flow.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Anholt RRH, Mackay TFC (2004). Quantitative genetic analyses of complex behaviours in *Drosophila*. *Nat Rev Gen* **5**: 838–849.

Basten CJ, Weir BS, Zeng ZB (1996). *QTL Cartographer (Computer Software)*. Program in Statistical Genetics, North Carolina State Univ. Raleigh, NC.

Beavis WD (1998). QTL analyses: power, precision, and accuracy. In: AH Patterson (ed), *Molecular Dissection of Complex Traits*. CRC Press: Boca Raton, FL. pp 145–162.

Bush GL (1994). Sympatric speciation in animals: new wine in old bottles. *Trends Ecol Evol* **9**: 285–288.

Caillaud CM, Via S (2000). Specialized feeding behaviour influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *Am Nat* **156**: 606–621.

Carroll SP, Dingle H, Famula TR, Fox CW (2001). Genetic architecture of adaptive differentiation in evolving host races of the soapberry bug, *Jadera haematoloma*. *Genetica* **112**: 257–272.

Castle WE (1921). An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. *Science* **54**: 223.

Churchill GA, Doerge RW (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.

Cockerham CC (1986). Modifications in estimating the number of genes for a quantitative character. *Genetics* **137**: 191–199.

Craig TP, Horner JD, Itami JK (2001). Genetics, experience, and host-plant preference in *Eurosta solidaginis*: implications for host shifts and speciation. *Evolution* **55**: 773–782.

Dambroski HR, Linn Jr C, Berlocher SH, Forbes AA, Roelofs W, Feder JF (2005). The genetic basis for fruit odor discrimination in Rhagoletis flies and its significance for sympatric host shifts. *Evolution* **59**: 1953–1964.

Del Campo M, Via S, Caillaud MC (2003). Recognition of host specific stimulants mediates host specificity and assortative mating in two host races of the pea aphid. *Ecol Entomol* **28/4**: 405–412.

DeVicente MC, Tanksley SD (1993). QTL analysis of transgressive segregation in an inyterspecific tomato cross. *Genetics* **134**: 585–596.

Doerge RW, Churchill GA (1996). Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**: 285–294.

Feder JL (1998). The Apple Maggot Fly, *Rhagoletis pomonella*. Flies in the face of conventional wisdom about speciation. In: Schluter D, Howard DJ, Berlocher SB (eds). *Endless Forms: Species and Speciation*. Oxford University Press: New York, NY, pp 130–144.

Fenster CB, Galloway LF, Chao L (1997). Epistasis and its consequences for the evolution of natural populations. *Trends Ecol Evol* **12**: 282–286.

Ferrari J, Via S, Godfray HC (2008). Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. *Evolution* **62**: 2508–2524.

Foster SA, Endler JA (1999). *Geographic Variation in Behavior: Perspectives on Evolutionary Mechanisms*. Oxford University Press: New York, NY.

Fox CW, Stillwell RC, Amarillo-S AR, Czesak M, Messina FJ (2004). Genetic architecture of population differences in oviposition behaviour of the seed beetle *Callosobruchus maculatus*. *J Evol Biol* **17**: 1141–1151.

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

Gavrilets S (2004). *Fitness Landscapes and the Origin of Species*. Princeton University Press: Princeton, NJ.

Hard JJ, Bradshaw WE, Holzapfel CM (1992). Epistasis and the genetic divergence of photoperiodism between populations of the pitcher-plant mosquito, *Wyeomyia smithii*. *Genetics* **131**: 389–396.

Hatfield T (1997). Genetic divergence in adaptive characters between sympatric species of stickleback. *Am Nat* **149**: 1009–1029.

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

Hunt GJ, Page RE, Fondrk MK, Dullum CJ (1995). Major quantitative trait loci affecting honey bee foraging behaviour. *Genetics* **141**: 1537–1545.

Itami JK, Craig TP, Horner JD (1997). Factors affecting gene flow between the host races of *Eurosta solidaginis*. In: Mopper S, Strauss SY (eds). *Genetic Structure and Local Adaptation in Natural Insect Populations Effects of Ecology, Life History and Behaviour*. Chapman and Hall: New York, NY. pp 375–407.

Janz N (1998). Sex-linked inheritance of host-plant specialization in a polyphagous butterfly. *Proc R Soc Lond Ser B* **265**: 1–4.

Jones CD (1998). The genetic basis of *Drosophila sechellia*'s resistance to a host plant toxin. *Genetics* **149**: 1899–1908.

Kearsey MJ, Pooni HS (1996). *The Genetic Analysis of Quantitative Traits*. Chapman and Hall: London (UK). p 381.

Lande R (1981). The minimum number of genes contributing to quantitative genetic variation between and within populations. *Genetics* **99**: 541–553.

Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates: Sunderland, MA. p 803.

Matsubayashi KW, Ohshima I, Nosil P. (2010). Ecological speciation in phytophagous insects. *Entomologia experimentalis et applicata* **134**: 1–27.

Matsuo T, Sugaya S, Yasukawa J, Aigaki T, Fuyama Y (2007). Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. *PLoS Biol* **5**: e118.

Nylin S, Nygren GH, Winding JJ, Janz N, Bergstrom A (2004). Genetics of host-plant preference in the comma butterfly *Polygonia c-album* (Nymphalidae), and evolutionary implications. *Biol J Linnean Soc* **84**: 755–765.

Otto SP, Jones CD (2000). Detecting the undetected: estimating the total number of loci underlying a quantitative trait. *Genetics* **156**: 2093–2107.

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. *Proc Natl Acad Sci USA* **106**: 7495–7500.

Peccoud J, Simon JC (2010). The pea aphid complex as a model of ecological speciation. *Ecol Entomol* **35**: 119–130.

Rinderer TE, Beaman LD (1995). Genic control of honey bee dance language dialect. *Theor Appl Genet* **91**: 727–732.

SAS Institute Inc. (1990). *SAS/StAT User's Guide, Version 6*, 4th edn, vol 1, SAS Institute Inc.: Cary, NC. p 943.

Schluter D (1998). Ecological causes of speciation. In: Schluter D, Howard DJ, Berlocher SB (eds). *Endless Forms: Species and Speciation*. Oxford University Press: New York, NY, pp 114–129.

Schluter D, Conte GL (2009). Genetics and ecological speciation. *Proc Natl Acad Sci USA* **106**: 9955–9962.

Sezer M, Butlin RK (1998a). The genetic basis of host-plant adaptation in the brown planthopper (*Nilaparvata lugens*). *Heredity* **80**: 499–504.

Sezer M, Butlin RK (1998b). The genetic basis of oviposition preference between sympatric host races of the brown planthopper (*Nilaparvata lugens*). *Proc R Soc Lond Ser B* **265**: 2399–2404.

Simon JC, Carre S, Boutin M, Prunier-Leterme N, Sabater-Munoz B, Latorre A, Bournoville R (2003). Host-based divergence in populations of the pea aphid: Insights from nuclear markers and the prevalence of facultative symbionts. *Proc R Soc Lond Ser B* **270**: 1703–1712.

Slater PJB, Halliday TR (1994). *Behaviour and Evolution*. Cambridge University Press: Cambridge (UK).

Tanksley SD (1993). Mapping polygenes. *Annu Rev Genet* **27**: 205–233.

Thompson JN (1988). Evolutionary genetics of oviposition preference in swallowtail butterflies. *Evolution* **42**: 1223–1234.

Tjallingii WF (1988). Electrical recording of stylet penetration activities. In: Minks AK, Harrewijn P (eds). *Aphids, Their Biology, Natural Enemies and Control*, vol. 2 B. Elsevier: Amsterdam (The Netherlands). pp 95–108.

True JR, Liu J, Stam LF, Zeng ZB, Laurie CC (1997). Quantitative genetic analysis of divergence in male secondary sexual traits between *Drosophila simulans* and *Drosophila mauritiana*. *Evolution* **51**: 816–832.

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

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

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

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

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

Zeng ZB (1992). Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method. *Genetics* **131**: 987–1001.

Zeng ZB (1994). Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.

Zeng ZB, Liu J, Stam LF, Kao CH, Mercer JH, Laurie CC (2000). Genetic architecture of morphological shape difference between two *Drosophila* species. *Genetics* **154**: 299–310.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)