

## ORIGINAL ARTICLE

# Maintaining genetic diversity and population panmixia through dispersal and not gene flow in a holocyclic heteroecious aphid species

LC Orantes<sup>1</sup>, W Zhang<sup>1</sup>, MAR Mian<sup>1,2</sup> and AP Michel<sup>1</sup>

Heteroecious holocyclic aphids exhibit both sexual and asexual reproduction and alternate among primary and secondary hosts. Most of these aphids can feed on several related hosts, and invasions to new habitats may limit the number of suitable hosts. For example, the aphid specialist *Aphis glycines* survives only on the primary host buckthorn (*Rhamnus* spp.) and the secondary host soybean (*Glycine max*) in North America where it is invasive. Owing to this specialization and sparse primary host distribution, host colonization events could be localized and involve founder effects, impacting genetic diversity, population structure and adaptation. We characterized changes in the genetic diversity and structure across time among *A. glycines* populations. Populations were sampled from secondary hosts twice in the same geographical location: once after secondary colonization (early season), and again immediately before primary host colonization (late season). We tested for evidence of founder effects and genetic isolation in early season populations, and whether or not late-season dispersal restored genetic diversity and reduced fragmentation. A total of 24 single-nucleotide polymorphisms and 6 microsatellites were used for population genetic statistics. We found significantly lower levels of genotypic diversity and more genetic isolation among early season collections, indicating secondary host colonization occurred locally and involved founder effects. Pairwise  $F_{ST}$  decreased from 0.046 to 0.017 in early and late collections, respectively, and while genetic relatedness significantly decreased with geographical distance in early season collections, no spatial structure was observed in late-season collections. Thus, late-season dispersal counteracts the secondary host colonization through homogenization and increases genetic diversity before primary host colonization.

*Heredity* (2012) **109**, 127–134; doi:10.1038/hdy.2012.21; published online 2 May 2012

**Keywords:** aphid population genetics; dispersal; *Aphis glycines*; SNPs

## INTRODUCTION

Aphids (Hemiptera: Aphididae) are one of the most diverse groups among insects. About 4700 described species exist (Blackman and Eastop, 2006; Peccoud *et al.*, 2010), some of which represent diverging populations, host races, biotypes, or other potential forms of incipient speciation (Van Emden and Harrington, 2007; Carletto *et al.*, 2009; Peccoud *et al.*, 2009). Aphids also vary in important life-history characteristics such as reproductive mode and host specificity (Moran, 1992; Blackman and Eastop, 2006). Species can be heteroecious (alternating between primary and secondary hosts) or autoecious (complete life cycle on a single host), holocyclic (undergoing sexual and asexual reproduction) or anholocyclic (asexual only), with populations of the same species exhibiting different strategies. Any adaptation that evolves during the asexual stage can quickly become common and spread, as the generation time can be as little as a few days and selection in the form of clonal amplification markedly increases favored clones (Halkett *et al.*, 2005; Vialatte *et al.*, 2005; Harrison and Mondor, 2011). Through holocycly, favored clones have a better opportunity to transfer favorable genetic variation into new combinations through sexual reproduction, increasing genetic

diversity and potentially the adaptation response. However, in obligate holocyclic species that are also heteroecious, sexual reproduction occurs on different hosts (the primary host) than asexual reproduction (secondary hosts). During spring, emigrants (winged aphids) leave the primary host in search of secondary hosts, and, in the following autumn, males and gynoparae (precursors of mating females) are generated and migrate to primary hosts. Thus, genetic variation in heteroecious aphids must pass through a potentially hazardous seasonal migration involving two independent host colonizations before sexual reproduction and therefore gene flow.

Of the host-alternating aphid species, 63% feed on more than five plant species, but usually hosts within the same family (Eastop, 1973; Hales *et al.*, 1997). The benefits of host alternation include escape from natural enemies, avoidance of intraspecific competition and independence from relying on a single-host species (Mackenzie and Dixon, 1992; Hales *et al.*, 1997). Despite these benefits, there are significant costs associated with heteroecy, such as dependence on the availability and sufficient quality of at minimum two hosts, and the potential founder effects related to primary and secondary host colonization. Colonization costs are amplified in the 6% of

<sup>1</sup>Department of Entomology, Ohio Agricultural Research and Development Center, The Ohio State University, Wooster, OH, USA and <sup>2</sup>USDA-ARS and Department of Horticulture and Crop Sciences, Ohio Agricultural Research and Development Center, The Ohio State University, Wooster, OH, USA

Correspondence: Dr AP Michel, Department of Entomology, Ohio Agricultural Research and Development Center, The Ohio State University, 1680 Madison Ave, Wooster, OH 44691, USA.

E-mail: michel.70@osu.edu

Received 3 November 2011; revised 20 March 2012; accepted 22 March 2012; published online 2 May 2012

heteroecious aphids classified as specialists that are solely dependent on one primary and one secondary host (Hales *et al.*, 1997). For example, Ward *et al.* (1998) calculated that only 0.6% of migrants successfully colonized primary hosts in the specialist, heteroecious bird cherry-oat aphid, *Rhopalosiphum padi* in the United Kingdom. If indeed migration events result in founder effects and localized colonization, then specialist heteroecious aphid populations could suffer from decreased genetic diversity and increased fragmentation and isolation, further impacting population sustainability and adaptation potential.

In this study, we implemented a population genetic approach to characterize changes in the genetic diversity and population structure in the holocyclic, heteroecious aphid specialist, *A. glycines*. This aphid is invasive in North America, having first been detected in 2000 and, since then, has successfully established across most of the North Central US and Great Lakes region and southeastern Canada (Ragsdale *et al.*, 2007, 2011). Like *R. padi*, *A. glycines* (the soybean aphid) is a significant crop pest, but feeds on soybean (*Glycine max*) as its secondary host across a wide geographical area in North America. Various buckthorn species (*Rhamnus* spp.) can serve as the primary host of *A. glycines*, but the patchily distributed common buckthorn (*Rhamnus carthartica*) is the main primary host in North America (Voegtlin *et al.*, 2004). Although it can feed on other Leguminosae hosts in its native Asian range, the suitable host range has severely contracted after its invasion to North America (Voegtlin *et al.*, 2004; Blackman and Eastop, 2006). In early spring, aphids emerge from overwintering eggs on buckthorn and develop into fundatrices, which are highly specialized and fecund female morphs that generate the secondary host colonizers (Blackman and Eastop, 2006, life cycle in Supplementary Figure 1). After two to three asexual generations on buckthorn, alate females are eventually produced that colonize soybean fields, typically following soybean emergence in late May to early June (Ragsdale *et al.*, 2004; Tilmon *et al.*, 2011). Up to 15 asexual generations can occur on soybean (Ragsdale *et al.*, 2004), and upon soybean senescence, decreasing temperature and changing photoperiod, winged males and gynoparae are produced for migration to common buckthorn.

For *A. glycines*, there are three separate movement events: (1) primary host to secondary, (2) among secondary hosts and (3) from secondary host back to primary. As soybean is widely distributed in North America, secondary host dispersal during the asexual phase is more limited by its reproductive capacity and production of asexual alates than finding suitable host plants. Population sizes can double in the field every 6–7 days (Ragsdale *et al.*, 2007), and an individual female aphid can produce >20 nymphs in a week (Michel *et al.*, 2010a). *A. glycines* can disperse quite large distances during the asexual phase, including into areas where little to no buckthorn exists (Ragsdale *et al.*, 2004, 2011; Tilmon *et al.*, 2011). Most aphids are weak fliers, so much of the dispersal among secondary hosts is wind aided (Taylor *et al.*, 1979; Dixon and Howard, 1986; Loxdale *et al.*, 1993). During both host transition events, however, environmental conditions and host proximity are possible key factors that determine the success of colonization. For example, although mostly sympatric, the primary and secondary hosts of *A. glycines* differ in their geographical distribution in more southern latitudes. Soybean is widely cultivated across the United States and southern Canada, but common buckthorn is most abundant in latitudes north of 41°N (Ragsdale *et al.*, 2004; Tilmon *et al.*, 2011). Suitable *Rhamnus* species are patchily distributed south of 41°N in North America, but can be found in dense thickets—relics of the historical use of buckthorn for hedgerows or landscaping before its serious invasiveness was realized

(Heimpel *et al.*, 2010). This difference in distributions can lead to significant mortality and founder effects during host transitions, thereby decreasing genetic diversity and increasing genetic isolation among populations. Indeed, the abundance of buckthorn near soybean fields was a key predictor of soybean infestation in ON, Canada (Bahlai *et al.*, 2010). Furthermore, aphids collected from soybean in early spring (representing founding individuals) showed less genetic diversity and more genetic differentiation than aphids collected later in the summer (Michel *et al.*, 2009a), suggesting that time becomes a major factor for genetic differentiation rather than geographical space. An additional constraint is the phenological disjunction between alate production on buckthorn and soybean emergence. In years with poor or delayed planting conditions, there may be little or no soybean available for winged migrants to colonize, and no other secondary hosts are known.

Despite this potential for founder effects during host colonization, the reproductive output of soybean aphid on secondary hosts can rapidly increase populations. With limited number of clones, selection in the form of clonal amplification favors the fittest clones with the highest reproductive output. In laboratory colonies of *A. glycines*, almost 50% of genetic diversity (measured by the number of unique clones) can be lost in as little as 10 generations (Michel *et al.*, 2010b). Unless counteracted by the immigration of new genotypes and genetic variation before migration to primary hosts, these isolated populations may continually be at risk of founder effects and decreased genetic diversity.

Characterizing changes in the genetic diversity and structure in *A. glycines* will help understand the population genetic implications of host colonization in specialist heteroecious aphids. These implications include predicting adaptation potential of soybean aphid in its invasive habitat to overcome aphid-resistant soybean varieties (Kim *et al.*, 2008; Hill *et al.*, 2010) or the possibility of insecticide resistance (should it occur). If founder effects and clonal amplification occur during and after secondary host colonization, then we would expect decreased levels of genetic diversity within populations, and a significant genetic differentiation among populations. As asexual reproduction proceeds, we would expect large-scale dispersal among soybean fields to spread genetic variation, increase genetic diversity and homogenize populations immediately before migration and gene flow on primary hosts. To address these implications, we compared population genetic characteristics using 24 single-nucleotide polymorphisms (SNPs) and 6 microsatellites from 16 field populations of *A. glycines*. In eight North American soybean fields, we collected soybean aphids from two time points (early season and late season), reflecting important phases of dispersal: immediately after soybean colonization and before primary host migration.

## MATERIALS AND METHODS

### Soybean aphid samples and DNA isolation

This study used eight sites across North America representing much of the north central soybean growing region (Figure 1). Two collections were taken in each field: an early season collection (before 6 July 2009), representing the soybean colonization population, and a late-season collection (after 30 July 2009), representing populations after many asexual generations and secondary host dispersal (Table 1). For each collection, 1 aphid-infested leaf from 50 different plants in a field was placed in an individual, sealed plastic bag, and sent overnight to the corresponding author where an individual aphid was removed from each bag, placed in a 0.2-ml microcentrifuge tube, and stored at –80 °C for later genetic analysis. Only one aphid per leaf and per plant was sampled to limit the possibility of including clonal individuals. All aphids were transported or collected under USDA/APHIS permit P526P-08-00872 to the corresponding author. DNA was extracted from each aphid with the E.Z.N.A.

Genomic DNA Isolation Kit (Omega Bio-Tek, Norcross, GA, USA) following the manufacturer's instructions, with a 100- $\mu$ l elution.

**Molecular marker genotyping**

The six microsatellites used in this study were originally developed from related species *A. fabae* and *A. gossypii* (Vanlerberghe-Masutti *et al.*, 1999; Coeur d'acier *et al.*, 2004; Gauffre and Coeur d'acier, 2006). Owing to the small size of the founding, invasive soybean aphid population in North America, all six microsatellite loci behaved as diallelic polymorphic markers, similar to the 24 SNP markers used in this study. Full details of microsatellite testing and PCR conditions are published elsewhere (Michel *et al.*, 2009a, b). Briefly, microsatellites were amplified in 20  $\mu$ l PCR reactions using fluorescently labeled forward primers. Genotyping was performed using a Beckman Coulter CEQ8800 at the Molecular Cellular and Imaging Center (MCIC, OARDC, Wooster, OH, USA) by pooling six microsatellites in a single genotyping run. Samples were diluted according to fluorescent dye per manufacturer's instructions. Individual genotypes were scored using the CEQ Fragment analysis software (Beckman Coulter, Miami, FL, USA) followed by manual inspection of allele determinations.

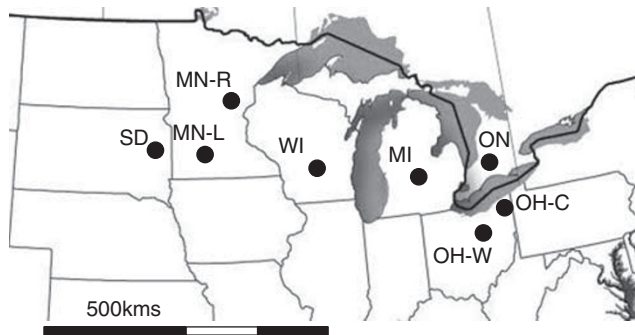
The Molecular Ecology Resources Consortium (2011) described testing and validation of 30 SNPs for *A. glycines*. For this study a total of 24 SNPs were used, owing to the poor amplification and a small minor allele frequency in 6 of the SNPs. Briefly, the standard Luminex (Austin, TX, USA) allele specific primer extension protocol was used where two sets of primers were designed to (1) restrict and amplify the genomic area containing the SNP and (2) target the SNP by creating primers with allele specific primer extension. Amplification of the genomic area was performed combining 12 forward and 12 reverse primers with PCR conditions following the instructions for the Qiagen Multiplex PCR kit (Qiagen, Valencia, CA, USA). PCR product was vortexed and centrifuged at 1000 r.p.m. for 1 min then cleaned with ExoSAP-IT (Affymetrix Corp., Santa Clara, CA, USA) according to manufacturer's instructions. For the allele specific primer extension reaction, the allele specific primer extension protocol was followed, using 4  $\mu$ l of multiplex PCR template in a 10- $\mu$ l aliquot. Data

were collected with the Luminex 200system and the alleles were detected and called using the Masterplex QT and Masterplex GT software from MiraBio (San Francisco, CA, USA).

**Polymorphism and genetic diversity statistics**

Owing to the microsatellites behaving similar to SNPs (diallelic), we combined data from both marker data sets. Allele frequencies, observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity were calculated using the Microsatellite Analyser (MSA 4.05; Dieringer and Schlötterer, 2003). We used GenAEx 6.41 (Peakall and Smouse, 2001) to calculate deviation from Hardy-Weinberg Equilibrium as measured by the inbreeding coefficient,  $F_{IS}$ . Linkage disequilibrium was calculated using FSTAT v 2.9.3.2 (Goudet, 1995, 2001). Neutrality of loci was assessed using LOSITAN to detect outlying alleles under selection to eliminate bias that could make the data shift towards balancing or positive selection (Beaumont and Nichols, 1996; Lopes *et al.*, 2008). Those loci with significant bias and evidence of selection were removed from data sets and analyses were recalculated. Frequency of null alleles for both marker data sets was estimated using ML-NullFreq (Kalinowski and Taper, 2006). To test the ability of the marker set for genetic diversity estimated, we performed resampling tests of loci and individuals using a jackknife procedure with 500 replications within the program GenClone (Arnaud-Haond and Belkhir, 2007). In addition, all statistics were also calculated removing repeated MLGs (common practices for parthenogenic aphids to limit clonal bias; Sunnucks, *et al.*, 1997). As genotypic diversities were high, removing MLGs did not significantly change results, and we report results using entire data. Polymorphism statistics  $H_o$ ,  $H_e$ ,  $F_{IS}$ , were compared among early and late populations using 10 000 random permutations implemented in FSTAT. As a control, samples were also grouped by geography (East: OH, ON and MI; West: MN, SD and WI).

For population genetic statistics we followed suggestions provided in Arnaud-Haond *et al.* (2007) and the program GenClone for use with partially clonal organisms. However, in some cases, especially with late populations, genotypic diversities were close or at maximum, that is, every individual was represented by a different clonal genotype. In these instances, we also included common population genetic statistics used in previous aphid genetic studies (Miller *et al.*, 2003; Vialatte *et al.*, 2005; Klueken *et al.*, 2011), as data mirrored randomly admixed populations atypical of clonal reproduction. The probability of two individuals that share a MLG (that is, clones) resulting from a sexual reproduction event was calculated using  $P_{sex}$  using GenClone (Arnaud-Haond *et al.*, 2007). To compare genotypic diversity, GenAEx generated a list of distinct, MLGs among populations, and calculated genotypic richness,  $R$  ( $R = (G - 1)/(N - 1)$ ), where  $G$  is number of MLGs and  $N$  is the total number of samples (Dorken and Eckert, 2001). We also calculated the Simpson's evenness statistic,  $V$ , and the Pareto distribution index,  $c$ . All three statistics were shown to be the least redundant in estimating the clonal diversity and abundance (Arnaud-Haond *et al.*, 2007). Population assignment was calculated with GenAEx, using the Paetkau assignment test (Paetkau *et al.*, 2004), where each individual of a population is frequency based assigned to the highest log population likelihood computed for each population per sample. For the late populations, the 'self' population assignment (that is, number of individuals



**Figure 1** Approximate locations of collection sites from Table 1.

**Table 1** Collection information for *A. glycines* populations

Population	Abbreviation	Early collection date	Late collection date	N <sup>a</sup>	N <sup>b</sup>	Latitude	Longitude
South Dakota	SD	13 June	30 July	32	32	44.32	-96.78
Minnesota-Lamberton	MN-L	23 June	8 September	40	32	44.24	-95.31
Minnesota-Rosemont	MN-R	3 June	26 August	48	32	44.99	-93.18
Wisconsin	WI	22 June	8 September	32	40	43.31	-89.33
Michigan	MI	10 June	14 September	32	40	42.71	-84.46
Ontario	ON	6 July	25 August	48	32	43.04	-81.26
Ohio-Wooster	OH-W	15 June	24 August	46	40	40.76	-81.90
Ohio-Cortland	OH-C	8 July	20 August	32	32	41.29	-80.65

All collections from summer of 2009.

<sup>a</sup>Sample size in early collection.

<sup>b</sup>Sample size in late collection.

assigned to their own sampled population) was regressed over the ordinal collection dates, and significance was determined by generating a correlation coefficient using the Minitab 16 Statistical Software (State College, PA, USA). To compare the overall genotypic populations parameters, we used the Wilcoxon signed-rank (WSR) test for paired early vs late population, and Mann–Whitney U test was used to compare East and West populations.

### Genetic differentiation and population structure

Matrices of  $F_{ST}$  values and the Bonferroni corrected  $P$ -values between populations were generated using MSA 4.05, calculated through 10000 random permutations.  $F_{ST}$  was compared among populations grouped by time (early and late) and geography (East and West) using FSTAT (see above). A principal component analysis was generated based on a matrix of Nei's genetic distance between population per loci (Nei, 1972, 1978) using GenAlEx 6.41. To determine the effect of geographical distance on spatial structure, we analyzed the level of global spatial autocorrelation by calculating  $r$ , the spatial autocorrelation coefficient in GenAlEx (Smouse and Peakall, 1999; Peakall *et al.*, 2003; Double *et al.*, 2005). The statistical significance is assessed through 10000 random permutations, as well as bootstrapping values of  $r$  10000 times. Values of  $r$  within 95% confidence intervals fail to reject the null hypothesis of no spatial genetic structure. Values of  $r > 0$  indicate relatedness increases with geographical distance, while values  $< 0$  indicate a decline in relatedness (and hence increase in structuring) with geographical distance (Peakall *et al.*, 2003). Spatial autocorrelation was performed at scales of 150 and 300 km and for early and late populations separately.

## RESULTS

### Neutrality, equilibrium and polymorphism

No significant null alleles ( $> 0.05$  frequency) were found in either the microsatellite or the SNP data (data not shown). After running LOSITAN to check for marker neutrality, 4 loci (SNPs 4730, 42701, 1538 and 5820) exhibited significant evidence of selection with the first 3 loci suspected of divergent selection and 5820 in balancing selection (Supplementary Figure 2). Upon closer inspection of allele frequencies, no association among geography, latitude or other potential environmental parameters responsible for divergent selection were found, and only SNP 4730 was found within a known gene (reverse transcriptase) based on similarity to the pea aphid whole-genome sequence (IAGC, 2010). For these loci under divergent selection, at least two populations in both early and late collections showed allelic fixation (major allele frequency  $> 0.95$ ), suggestive of ascertainment bias and likely impacted  $F_{ST}$  values. Nonetheless, population genetic statistics were calculated including and excluding these four SNPs, although we report analyses using only the neutral loci. In total, 20 SNPs and 6 microsatellites were used to generate results in this study. The 26 markers used were able to recover genetic diversity as indicated by resampling procedures provided in GenClone (Supplementary Figure 3).

Across all loci and populations, the frequency of the minor allele ranged from 0.06 to 0.40. Expected heterozygosity ( $H_e$ ) was lower than observed heterozygosity ( $H_o$ ) in all populations (both early and late), averaging 0.427 in early populations and 0.437 in late populations, whereas  $H_o$  averaged 0.536 and 0.589, respectively. Tests of Hardy–Weinberg Equilibrium showed the presence of excess of heterozygosity ( $F_{IS} < 0$ ) in at least one locus in all populations (Supplementary Table 1). The number of affected loci ranged from 2 in the early population from MN–L, to 10 in the late population from OH–W. However, no individual population had more than half of the loci with heterozygote excess. No significant linkage disequilibrium was found in any population. There was no significant difference in heterozygosity or  $F_{IS}$  among early and late populations as determined by FSTAT.

### Genotypic diversity and population assignment

On the secondary host, *A. glycines* propagates clonally through asexual reproduction until autumn migration to buckthorn. During this stage, comparing the number of clones, as represented by MLGs, can provide relevant information of genetic diversity and relatedness among populations. We determined the number and distribution of unique (that is, singleton MLG) vs matching (found more than once) MLGs among all samples. None of the matching MLGs exhibited evidence of a distinct sexual reproduction event based on  $P_{sex}$  values (data not shown). Overall, the level of genotypic diversity in both early season and late-season populations was high. A substantial number of distinct MLGs were found among all samples, totaling 192 MLGs of 310 individuals in the early season populations (61.94%) and, significantly higher, 258 of 288 individuals in late-season populations (89.58%, WSR = 0,  $n_{s/r} = 8$ ,  $P < 0.01$ ). There was no significant difference in the number of unique MLGs when comparing East and West populations ( $P > 0.05$ ). In early populations, the most common North American MLG had a frequency of only 3.55% (11 individuals) and was found in SD, MI, WI and OH–W. The second highest MLG was 2.90% (9 individuals) in both populations of OH and in WI. In the late populations there were no predominant MLGs. The early season population from SD had the lowest genotype diversity ( $R = 0.68$ ), having 9 matching genotypes and 21 unique genotypes (Table 2). Among the late-season populations, SD was also the population with the lowest  $R$  (0.87), having 21 unique genotypes out of 32 total individuals. For the rest of the late-season populations, the number of matching MLGs ranged from zero to four, with  $R$  reaching 1.00 in MI, OH–W and ON. When comparing each population early season vs late season we found an increase of  $R$  at each location and significantly higher overall  $R$  in late-season populations (WSR = 0,  $n_{s/r} = 8$ ,  $P < 0.01$ ). Not surprisingly, no significant difference was found between  $R$  among Eastern and Western populations when comparing within early, late and all data combined (data not shown).

To further compare distribution of genotypes among populations we performed a population assignment test. Early season populations had significantly higher self-population assignment (individuals assigned to own population) than in late-season populations (WSR = 0,  $n_{s/r} = 8$ ,  $P < 0.01$ ) with 56.1% average self-assignment in early season and 31.6% average self-assignment in late-season populations (Table 2). Early season populations showed no pattern of self-assignment by temporal or geographical distribution and largely were assigned to their own populations. However, late-season populations were strongly influenced by time of collection. Regression of self-assignment/sample size of late-season populations over time (Figure 2) showed a significant fitted negative correlation ( $R^2 = 0.90$ ,  $P = 0.01$ ) where the number of self-assigned individuals decreased as the collection time progress later into the season. No such correlation was found with the early collected samples ( $R^2 = 0.028$ , data not shown). The late-season SD population was collected 21 days before any other late population (Table 1) and showed a substantial increase in the number of self-assigned individuals, as well as the lowest  $R$  among all late-season populations.

### Dispersal and temporal effect on population structure

As *A. glycines* populations rapidly increase on secondary hosts, alates are produced and widely disperse that may impact population structure. We constructed a pairwise matrix of Nei's genetic distance and generated a principal component analysis, which explained the variation among populations by the interaction of the multiple loci. The early season populations showed stronger effects of the

**Table 2** Population assignment and genotypic diversity per population

Population	Assignment			Genotypes				V <sup>f</sup>	c <sup>g</sup>	
	Self-population <sup>a</sup>	Other population <sup>b</sup>	Self-populationP/N	No. matching <sup>c</sup>	No. unique <sup>d</sup>	Total	R <sup>e</sup>			
<i>Early</i>										
SD	24	8	0.75	9	13	22	0.68	0.71	1.8	
MN-L	24	16	0.60	10	24	34	0.85	0.82	<2.6	
MN-R	26	22	0.54	13	32	41	0.85	0.79	2.8	
WI	19	13	0.59	8	21	19	0.90	0.69	<3.4	
MI	14	18	0.44	9	22	31	0.97	0.52	<3.4	
ON	23	25	0.48	11	32	42	0.87	0.79	3.5	
OH-W	26	20	0.57	10	29	40	0.87	0.68	2.7	
OH-C	18	14	0.56	9	19	27	0.84	0.66	2.9	
Average	21.75	17.00	0.57	9.25	24.00	33.25	0.85	0.71	2.9	
<i>Late</i>										
SD	16	17	0.50	7	21	28	0.87	0.58	2.5	
MN-L	8	24	0.25	3	28	31	0.97	0	<5.0	
MN-R	11	21	0.34	3	28	31	0.97	0	<4.4	
WI	11	29	0.23	1	46	47	0.98	0	<4.3	
MI	5	35	0.13	0	40	40	1.00	∞	ND	
ON	12	20	0.38	4	28	32	1.00	∞	ND	
OH-W	16	24	0.40	2	38	40	1.00	∞	ND	
OH-C	12	20	0.38	2	29	31	0.97	0	<5.0	
Average	11.38	23.75	0.32	2.75	32.25	35.00	0.97	0.12	4.2	

Abbreviations: MI, Michigan; MN-L, Minnesota-Lamberton; MN-R, Minnesota-Rosemont; ND, not determined; ON, Ontario; OH-C, Ohio-Cortland; OH-W, Ohio-Wooster; SD, South Dakota; WI, Wisconsin.

<sup>a</sup>Individuals assigned to their own population.

<sup>b</sup>Individuals assigned to any other population but its own.

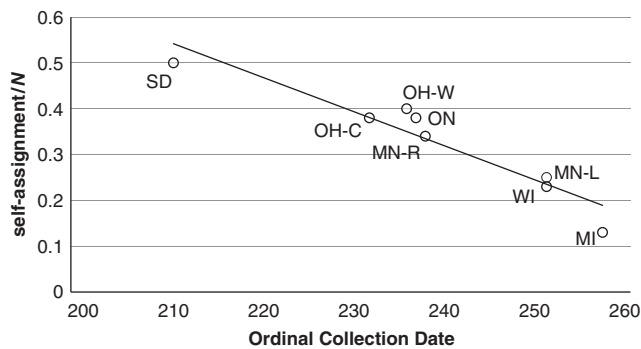
<sup>c</sup>Genotypes that were found more than once across all populations.

<sup>d</sup>Genotypes that were found only once within the population.

<sup>e</sup>R, genotypic richness where  $R = \# \text{ genotypes} - 1/N - 1$ .

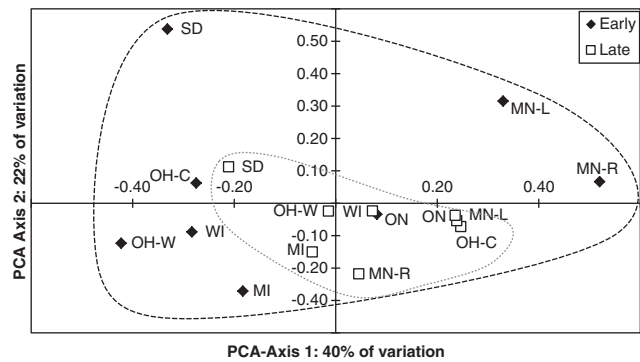
<sup>f</sup>V, Simpson's evenness statistic.

<sup>g</sup>c, Pareto distribution index, ND owing to max R; values with < represent estimates only owing to a lack of matching genotypes within a population.



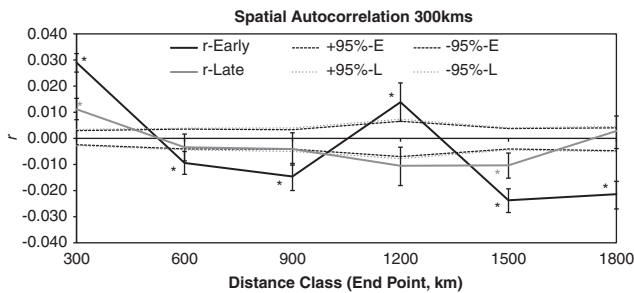
**Figure 2** Regression showing decreasing self-population assignment/N over time of collection, indicating later collected populations share more migrants.

components given by variation among loci and genetic separation among populations (Figure 3). The average pairwise separation among early collections was 0.55. These populations showed no particular clustering or effect by specific components such as geography, but rather each population was independently impacted by a particular locus or loci, reflective of clonal amplification. In the late-season populations, there was a strong clustering effect drawing all the populations closer towards the origin with a significantly smaller average pairwise separation at 0.238 ( $WSR = -14$ ,  $n_{s/r} = 28$ ,  $P < 0.001$ ). The reduction of variation among components was an



**Figure 3** Principal component analysis (PCA) showing genetic distance (*D*) among early and late population in North America. Early populations collected in June are represented by black diamonds, encircled by black dashed line; late populations collected in August and September are represented by white squares encircled by gray dotted lines. Dispersion of the populations is determined by the influence of the different components (that is, loci) such that populations that are further apart are isolated by genetic uniqueness, while populations closer together indicate shared variation.

effect of the homogenization by dispersal of aphids among populations which decreased genetic divergence among them. The late-season population of SD did not show the same trend as the other late populations, possibly owing to its sampling before widespread dispersal began. At the time of collection the late SD still had a



**Figure 4** Spatial autocorrelation among early (black line) and late (gray line) collected populations at 300 km distance classes. Dotted and dashed lines represent 95% confidence intervals of the neutral level of  $r$ , error bars represent  $r$  estimates from 10 000 bootstraps. Values labeled with \* represent evidence of a significant spatial structure.

greatest number of matching genotypes, higher number of self-assigned individuals and lowest genetic diversity (Table 2) among any late population.

Pairwise  $F_{ST}$  results concurred with the principal component analysis in that the early season populations had a significantly higher average pairwise  $F_{ST}$  value (0.046) than late collected populations (0.017,  $P < 0.001$ ). Overall, 27 out of 28 pairwise comparisons were significantly differentiated among early populations, compared with only 14 out of 28 significant comparisons in late populations (Supplementary Table 2). Differences in population subdivision between West and East collections were not significant, despite the magnitude of  $F_{ST}$  within geographical sub-populations being greater than what was observed within early season populations (early season West and East,  $F_{ST} = 0.055$  and  $0.026$ , respectively,  $P = 0.06$ ; late season West and East,  $F_{ST} = 0.011$  and  $0.025$ , respectively,  $P = 0.40$ ).

Spatial autocorrelation analyses with early and late collections showed discrepancies in the extent of significant structuring based on geographical distance at both 150 km (data not shown) and 300 km (Figure 4). For early season populations, all  $r$  values at all 6 distance classes of 300 km were significant, with 2 interceptions of the  $x$ -axis. Notably, a positive correlation was seen at distances  $< 300$  km, whereas a significant negative correlation was seen with the remaining 5 size classes (except for the 1200 km class), indicating restrictions on gene flow over larger distances. On the contrary, significant spatial structure was observed for only 2 out of 6 size classes ( $< 300$  km and 1500 km) for late populations, consistent with random dispersal.

## DISCUSSION

The North American invasion by *A. glycines* has severely limited its host availability, and now this species has become an extreme specialist, utilizing only one primary and one secondary host. During sexual reproduction, only buckthorn (*Rhamnus* spp.) is utilized, and, in North America, primary host use is mostly restricted to *R. carthartica*. The lone secondary host (where most of the asexual phase takes place) is soybean. We investigated how colonization related to the dependence on only two hosts with different distributions in North America impacted the genetic diversity, structure and potentially the adaptability of *A. glycines*.

### Heterozygosity excess

Most populations (both early and late) exhibited evidence of heterozygote excess, but each population differed in terms of which loci significantly deviated from Hardy–Weinberg Equilibrium.

Previous genetic studies on aphids have reported deviation from Hardy–Weinberg Equilibrium because of heterozygosity excess (Papura *et al.*, 2003; Vialatte *et al.*, 2005). Michel *et al.* (2009a) observed similar levels of heterozygote excess with the soybean aphid, with more deviation present in early collected populations. In this study, early season populations did not differ significantly in  $F_{IS}$  relative to late-season populations. The persistence of slight heterozygote excess for the duration of the season is likely caused by clonal amplification with subsequent dispersal. In asexual organisms, clonal amplification results when clones with the highest fitness reproduce faster and eventually outnumber less fit clones (Sunnucks *et al.*, 1997). This selection reduces the number of distinct clones, consistent with our significantly lower values of  $R$  in early populations. Our data support the hypothesis that there may be a slight advantage for heterozygous aphid lineages. As the season progresses, these genotypes produce winged alates that then disperse among soybean fields, maintaining and, in some cases, increasing heterozygosity. Future tests, currently underway, could include fitness comparisons of aphids that differ in levels of inbreeding and determine whether heterozygote advantages persist during asexual reproduction of this holocyclic aphid.

### Factors of secondary host colonization and founder effects

Previous work through ecological modeling has demonstrated the importance of the primary host by identifying local buckthorn abundance ( $< 4$  km from soybean field) as the best predictor of *A. glycines* secondary host colonization (Bahlai *et al.*, 2010). Our genetic data support the ecological modeling, revealing less genotypic diversity, higher relatedness and more population structuring among aphids in early collected populations when compared with late populations. These results are consistent with local colonization. However, most locations used to date in studies examining secondary host colonization are north of  $41^{\circ}\text{N}$  (OH–W in this study is the lone exception at  $40^{\circ}\text{N}$ ), where primary and secondary hosts are largely sympatric. Therefore, the factors leading to secondary host colonization in more southern locations with a more scattered and isolated distribution of buckthorn are unknown. The absence of large quantities of the primary host in these areas suggests that southern secondary host colonization results through dispersal by aphids coming from secondary hosts. Our data indicated a fair amount of matching MLGs in several early populations. As no reports are known of successful adult overwintering (McCornack *et al.*, 2005), and  $P_{sex}$  values did not indicate independent sexual reproduction events, these matching MLGs are likely from the same clone. This observation then suggests that long distance dispersal from primary to secondary hosts can occur.

In this study, a decrease in genotypic diversity and higher self-population assignment in early season populations implicate a founder effect during secondary host colonization. After sexual reproduction on the primary host in autumn, genetic diversity should be at its peak the following spring, when fundatrices emerge and produce spring migrants. Proximity of soybean to buckthorn is a likely factor (Bahlai *et al.*, 2010), as smaller dispersal distances likely increase colonization success. Although mortality from primary to secondary host has not been studied, Ward *et al.* (1998) estimated survival of *R. padi* during autumn migration to primary hosts at 0.6%. Using four microsatellites, Klueken *et al.* (2011) compared three potential primary plant sources and found that aphids in the nearest secondary host field were genetically similar to just a single primary plant with little or no contribution from the other two potential sources. Perhaps the largest mortality factor, especially when the

secondary host is a cultivated crop, is the phenological disjunction between alate aphids on the primary host and secondary host emergence (Ragsdale *et al.*, 2004; Michel *et al.*, 2009a). In several observations, alate flight from buckthorn had been recorded weeks before any local soybean emerged (Ragsdale *et al.*, 2004). Furthermore, by the time soybean was emerging, soybean aphid populations could not be found on buckthorn. No other secondary hosts in North America are known that might act as transitional hosts to bridge the disjunction, and reverse migration back to buckthorn in the absence of soybean has not been detected. Thus, the surviving secondary host colonization population would be lacking in genotypic diversity, as suggested by this study.

It is tempting to speculate that colonization alone may lead to the founder effect, but the interaction of selection and drift after secondary host colonization may also provide the signal of a founder effect. Although founder effects are typically caused by random mortality during colonization and the subsequent impact of genetic drift, selection in these early populations cannot be ruled out for *A. glycines*. Notably, the early season individuals in this study were of apterae, and not alates. Thus, the entire potential colonizing population was not sampled—only individuals present after two to three asexual generations within the soybean field. Many unique alate clones with high genetic diversity could still potentially colonize a secondary host field but would be difficult to detect because clonal amplification only favors those with the highest reproductive output. In Klueken *et al.* (2011), secondary host colonizers from the primary plant source had a *R* of 0.96, and the earliest aphid populations on the secondary host had a *R* of 1.0, suggesting high initial genotypic diversity during secondary host colonization. Selection could be in the form of plant quality (Noma *et al.*, 2010), insecticidal seed treatments (Magalhaes *et al.*, 2009), soybean variety or other characteristics related to common agronomic practices.

This pattern of decreased genotypic diversity in early populations may be different with holocyclic autoecious aphids—those that complete the entire life cycle on a single host but have an obligatory sexual generation. As there is no secondary host colonization phase, there is likely not a resulting decrease of genetic diversity in the spring. Assuming negligible impacts of overwintering mortality, the spring populations represent offspring from sexual reproduction, which tends to reshuffle MLGs and restore equilibrium. As the season progresses, autoecious populations are likely to be similar to heteroecious aphids in that clones can be lost owing to the clonal amplification. Indeed, for the aphid *Tuberculatus quercicola*, a decrease in genetic diversity was observed during the season within populations on individual oak trees (Yao and Akimoto, 2009). However, genetic differentiation among populations remained high with *T. quercicola*, suggesting, at least with this species, that late-season migration might not overcome the loss of diversity associated with clonal propagation. The differences between life histories between *A. glycines* and *T. quercicola* may also reflect host quality or availability in that the vast abundance of soybean in the United States provides opportunity to generate a large number of alates.

#### Dispersal, homogenization and the spread of genetic variation

Within aphids, the evolution of complex reproductive and life histories, including facultative asexual lineages, have allowed adaptation, host shifts and even speciation (Via, 1999; Peccoud *et al.*, 2009). The maintenance of specialization, or alternatively the prevention of host shift adaptation, within holocyclic heteroecious aphids depends on many factors including fitness in searching for and feeding on new hosts (Hales *et al.*, 1997; Ward *et al.*, 1998), or inherent population

genetic processes that limit gene flow and drive local adaptation. Given the level of significant genetic structuring seen in early season populations, local adaptation may lead to divergent selection among populations. Our data indicate that despite this early season structuring during secondary host colonization, late season dispersal causes panmixia across much of the North American range of *A. glycines*. Pairwise estimates of  $F_{st}$  decreased by over 60% among early season and late-season populations (0.046–0.017, respectively), and late-season populations exhibited significantly less variation in the principal component analysis (Figure 3). The reduction in genetic variation and differentiation among populations was concurrent with a shift towards more random dispersal based on spatial structuring analysis (Figure 4). Homogenization through dispersal counteracts genetic differentiation during secondary host colonization and ensures that any local adaptation that may occur through drift or selection is admixed immediately before autumn primary host colonization and sexual reproduction. Without late-season dispersal, populations may become more isolated and structured over time as host colonization events suffer from founder effects and occurs on a local scale. Our data also indicate that adaptations to insect management tactics such as aphid-resistant soybean varieties or insecticide resistance that arise locally can spread quite rapidly.

Homogenization through dispersal before sexual reproduction and gene flow may be reflected in several other specialized holocyclic and heteroecious aphids, such as *R. padi*, that rely on a common and abundant cultivated crop as a secondary host. Additional work will focus on genetic variation on primary hosts, specifically before and after sexual reproduction. Unfortunately, while *R. cathartica* is easy to find, significant *A. glycines* populations are often variable and difficult to sample. Further characterizing the interplay of genetic drift, selection, host plant distribution and dispersal can lead to a better understanding of the biological and evolutionary basis for adaptation and specialization in holocyclic heteroecious aphids and help predict the risk of adaptation and resistance to management such as insecticide resistance and evolution of aphids on *A. glycines*-resistant soybean varieties (Kim *et al.*, 2008; Hill *et al.*, 2010).

#### DATA ARCHIVING

Genotype data have been deposited at Dryad: doi:10.5061/dryad.3hq923hn.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGEMENTS

We thank the members of the Michel Laboratory (Geoff Parker, Jake Wenger) for assistance with data collection. SNP genotyping was performed at the MCIC with assistance from Jody Whittier. D Francis (OARDC) and three reviewers provided comments to improve the manuscript. Soybean samples were provided by K Tilmon, B Potter, D Ragsdale, E Cullen, C DiFonzo, C Krupke and T Baute. Funding was provided by the Ohio Soybean Council #OSC 10-2-03 and #OSC 08-2-08, USDA-ARS, Department of Entomology, an OARDC-SEEDS Graduate Research Award (LCO) and the Ohio Agricultural Research and Development Center, the Ohio State University.

Arnaud-Haond S, Belkhir K (2007). Genclone: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Mol Ecol Notes* 7: 15–17.

Arnaud-Haond S, Duarte CM, Alberto F, Serrão EA (2007). Standardizing methods to address clonality in population studies. *Mol Ecol* 16: 5115–5139.

- Bahlai CA, Sikkema S, Hallett RH, Newman J, Schaafsma AW (2010). Modeling distribution and abundance of soybean aphid in soybean fields using measurements from the surrounding landscape. *Env Entomol* **39**: 50–56.
- Beaumont MA, Nichols RA (1996). Evaluating loci for use in the genetic analysis of population structure. *Proc Roy Soc Lond* **263**: 1619–1626.
- Blackman RL, Eastop VF (2006). *Aphids of the World's Herbaceous Plants and Shrubs—an identification and information guide*. John Wiley, Inc.: New York, NY.
- Carletto J, Lombaert E, Chavigny P, Brevault T, Lapchin L, Vanlerberghe-Masutti F (2009). Ecological specialization of the aphid *Aphis gossypii* Glover on cultivated host plants. *Mol Ecol* **18**: 2198–2212.
- Coeur d'acier A, Sembène M, Audirot P, Rasplus JY (2004). Polymorphic microsatellite loci in the black aphid, *Aphis fabae* Scopoli, 1763 (Hemiptera, Aphididae). *Mol Ecol Notes* **4**: 306–308.
- Dieringer D, Schlotterer C (2003). Microsatellite analyzer (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol Ecol Notes* **3**: 167–169.
- Dixon AFG, Howard MT (1986). Dispersal in aphids: a problem in resource allocation. In Danthanarayanan W (ed.) *Insect Flight: Dispersal and Migration*. Springer: New York, pp 145–151.
- Dorken ME, Eckert CG (2001). Severely reduced sexual reproduction in northern populations of a clonal plant, *Decodon verticillatus* (Lythraceae). *J Ecol* **89**: 339–350.
- Double MC, Peakall R, Beck NR, Cockburn A (2005). Dispersal, philopatry and infidelity: dissecting local genetic structure in superb fairy-wrens (*Malurus cyaneus*). *Evolution* **59**: 625–635.
- Eastop VF (1973). Deductions from the present day host plants of aphids and related insects. In van Emden HF (ed.) *Insect/Plant Relationships*. Blackwell: Oxford, pp 157–178.
- Gaufrre B, Coeur d'acier A (2006). New polymorphic microsatellite loci, cross-species amplification and PCR multiplexing in the black aphid, *Aphis fabae* Scopoli. *Mol Ecol Notes* **6**: 440–442.
- Goudet J (1995). FSTAT (Version 1.2): a computer program to calculate F-statistics. *J Hered* **86**: 485–486.
- Goudet J (2001). FSTAT: a program to estimate and test gene diversities and fixation indices, version 2.9.3.
- Hales DF, Tomiuk J, Woehrmann K, Sunnucks P (1997). Evolutionary and genetic aspects of aphid biology: a review. *Eur J Entomol* **94**: 1–55.
- Halkett F, Plantegenest M, Prunier-Leterme N, Mieuze L, Delmott F, Simon J-C (2005). Admixed sexual and facultatively asexual aphid lineages at mating sites. *Mol Ecol* **14**: 325–336.
- Harrison JS, Mondor EB (2011). Evidence for an invasive aphid 'superclone': extremely low genetic diversity in oleander aphid (*Aphis nerii*) populations in the southern United States. *PLoS One* **6**: e17524.
- Heimpel GE, Frelich LE, Landis DA, Hopper KR, Hoelmer KA, Sezen Z *et al.* (2010). European buckthorn and Asian soybean aphid as components of an extensive invasional meltdown in North America. *Biol Invas* **12**: 2913–2931.
- Hill CB, Crull L, Herman TK, Voegtlin DJ, Hartman GL (2010). A new soybean aphid (Hemiptera: Aphididae) biotype identified. *J Econ Ent* **103**: 509–515.
- International-Aphid-Genomics-Consortium (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS: Biol* **8**: e1000313.
- Kalinowski S, Taper M (2006). Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. *Conser Genet* **7**: 991–995.
- Kim KS, Hill CB, Hartman GL, Mian RM, Diers BW (2008). Discovery of soybean aphid biotypes. *Crop Sci* **48**: 923–928.
- Klueken AM, Simon J-C, Hondelmann P, Mieuze L, Gilabert A, Poehling H-M *et al.* (2011). Are primary woody hosts 'island refuges' for host-alternating aphids and important for colonization of local cereals? *J Appl Ent* **36**: 347–360.
- Lopes AA, Lopes RJ, Beja-Pereira A, Luikart G (2008). LOSITAN: a workbench to detect molecular adaptation base on an  $F_{st}$ -outlier method. *BMC Bioinformatics* **9**: 323.
- Loxdale HD, Hardie J, Halber S, Fottit R, Kidd NAC, Carter CI (1993). The relative importance of short and long-range movement of flying aphids. *Biol Rev* **68**: 291–311.
- Mackenzie A, Dixon AFG (1992). An ecological perspective of host alternation in aphid (Homoptera:Aphidinea:Aphididae). *Entomol Gen* **16**: 265–284.
- Magalhaes LC, Hunt TE, Siegfried BD (2009). Efficacy of neonicotinoid seed treatments to reduce soybean aphid populations under field and controlled conditions in Nebraska. *J Econ Ent* **102**: 187–195.
- McCornack BP, Carrillo MA, Vennette RC, Ragsdale DW (2005). Physiological constraints on the overwintering potential of the soybean aphid (Homoptera: Aphididae). *Environ Entomol* **34**: 235–240.
- Michel AP, Mian MA, Davila-Olivas NH, Cañas LA (2010a). Detached leaf and whole plant assays for soybean aphid resistance: differential responses among resistance sources and biotypes. *J Econ Ent* **103**: 949–957.
- Michel AP, Zhang W, Jung JK, Kang ST, Mian MAR (2009a). Population genetic structure of *Aphis glycines*. *Env Entomol* **38**: 1301–1311.
- Michel AP, Zhang W, Jung JK, Kang S, Mian MAR (2009b). Cross-species amplification and polymorphism of microsatellite loci in the soybean aphid, *Aphis glycines*. *J Econ Entomol* **102**: 1389–1392.
- Michel AP, Zhang W, Mian MA (2010b). Genetic diversity and differentiation among laboratory and field populations of the soybean aphid, *Aphis glycines*. *Bull Ent Res* **100**: 727–734.
- Miller NJ, Birley AJ, Overall ADJ, Tatchell GM (2003). Population genetic structure of the lettuce root aphid, *Pemphigus bursarius* (L.), in relation to geographic distance, gene flow and host plant usage. *Heredity* **91**: 217–223.
- Molecular Ecology Resources Primer Development Consortium (2011). Permanent genetic resources added to molecular ecology resources database 1 June 2011–31 July 2011. *Mol Ecol Res* **11**: 1124–1126.
- Moran NA (1992). The evolution of aphid life cycles. *Ann Rev Entomol* **37**: 321–348.
- Nei M (1972). Genetic distance between populations. *Am Nat* **106**: 283–292.
- Nei M (1978). Estimation of heterozygosity and genetic distance from a small number of individuals. *Genetics* **76**: 379–390.
- Noma T, Gratton C, Colunga-Garcia M, Brewer MJ, Mueller EE, Wyckhuys KA *et al.* (2010). Relationship of soybean aphid (Hemiptera: Aphididae) to soybean plant nutrients, landscape structure, and natural enemies. *Env Ent* **39**: 31–41.
- Paetkau D, Slade R, Burden M, Estoup A (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Mol Ecol* **13**: 55–65.
- Papura D, Simon J-C, Halkett F, Delmotte F, Le Gallic J-F, Dedryve CA (2003). Predominance of sexual reproduction in Romanian populations of the aphid *Sitobion avenae* inferred from phenotypic and genetic structure. *Heredity* **90**: 397–404.
- Peakall R, Ruibal M, Lindenmayer DB (2003). Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, *Rattus fuscipes*. *Evolution* **57**: 1182–1195.
- Peakall R, Smouse PE (2001). GenAlEx V5: Genetic analysis in Excel; population genetic software for teaching and research. *Mol Ecol Notes* **6**: 288–295.
- Peccoud J, Ollivier A, Plantegenest M, Simon J-C (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 J-C, von Dohlen C, Coeur d'acier A, Plantegenest M, Vanlerberghe-Masutti F *et al.* (2010). Evolutionary history of aphid-plant associations and their role in aphid diversification. *CR Biol* **333**: 474–487.
- Ragsdale D, Voegtlin DJ, O'Neil RJ (2004). Soybean aphid biology in North America. *Ann Entomol Soc Am* **97**: 204–208.
- Ragsdale DW, Landis DA, Brodeur J, Heimpel GE, Desneux N (2011). Ecology and management of the soybean aphid in North America. *Ann Rev Ent* **56**: 375–399.
- Ragsdale DW, McCornack BP, Venette RC, Potter BD, MacRae IV, Hodgson EW *et al.* (2007). Economical threshold for soybean aphid (Hemiptera: Aphididae). *J Econ Ent* **100**: 1258–1267.
- Smouse PE, Peakall R (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* **82**: 561–573.
- Sunnucks P, De Barro PJ, Lushai G, Maclean N, Hales DF (1997). Genetic structure of an aphid studied using microsatellites: cyclic parthenogenesis, differentiated lineages and host specialization. *Mol Ecol* **6**: 1059–1073.
- Taylor LR, Woivod IP, Taylor RAJ (1979). The migratory ambit of the hop aphid and its significance in aphid population dynamics. *J Anim Ecol* **48**: 955–997.
- Tilmon KJ, Hodgson EW, O'Neal ME, Ragsdale DW (2011). Biology of the soybean aphid, *Aphis glycines* (Hemiptera: Aphididae) in the United States. *J Integ Pest Manage* **2**: 1–7.
- Van Emden H, Harrington R (2007). *Aphids as Crop Pests*. CABI Publishing: London.
- Vanlerberghe-Masutti F, Chavigny P, Fuller S (1999). Characterization of microsatellite loci in the aphid species *Aphis gossypii* Glover. *Mol Ecol* **8**: 693–695.
- Via S (1999). Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution* **53**: 1446–1457.
- Vialatte A, Dedryve CA, Simon J-C, Galman M, Plantegenest M (2005). Limited genetic exchanges between populations of an insect pest living on uncultivated and related cultivated host plants. *Proc R Soc B* **1567**: 1075–1082.
- Voegtlin DJ, O'Neil RJ, Graves WR (2004). Tests of suitability of overwintering hosts of *Aphis glycines*: identification of a new host association with *Rhamnus alnifolia* L'Heritier. *Ann Entomol Soc Am* **97**: 233–234.
- Ward SA, Leather SR, Pickup J, Harrington R (1998). Mortality during dispersal and the cost of host-specificity in parasites: how many aphids find hosts? *J Animal Ecol* **67**: 763–773.
- Yao I, Akimoto S-I (2009). Seasonal changes in the genetic structure of an aphid-ant mutualism as revealed using microsatellite analysis of the aphid *Tuberculatus quercicola* and the ant *Formica yessensis*. *J Insect Sci* **9**: 9.

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