

Using host-associated differentiation to track source population and dispersal distance among insect vectors of plant pathogens

Gina Marie Angelella¹  | Andy P. Michel² | Ian Kaplan¹

¹Department of Entomology, Purdue University, Lafayette, Indiana

²Department of Entomology, The Ohio State University, Wooster, Ohio

Correspondence

Gina Marie Angelella, Department of Entomology, Purdue University, Lafayette, IN.

Email: ginama@vt.edu

Present Address

Gina Marie Angelella, Department of Horticulture, Virginia Tech University, Eastern Shore Agricultural Research and Extension Center, Painter, Virginia

Funding information

USDA NIFA AFRI Predoctoral Fellowship Program, Grant/Award Number: 11588839; USDA North Central Region IPM Grants Program, Grant/Award Number: 11-34103-30723; Virginia Tech

Abstract

Small, mobile insects are notoriously challenging to track across landscapes and manage in agricultural fields. However, genetic differentiation among insect populations and host plants acquired through host-associated differentiation could be exploited to infer movement within crop systems and damage potential. Although many insects exhibit host-associated differentiation, management strategies for insect vectors of plant pathogens assume a homogenous population. Nevertheless, phenotypic changes derived from host-associated differentiation could manifest in altered behavior or physiology affecting the likelihood of vector–pathogen–plant interactions, or the subsequent efficiency of pathogen transmission. We used SNPs to assess genotypic structure and host-associated differentiation in the cowpea aphid, *Aphis craccivora* Koch (Hemiptera: Aphididae). To do so, we sampled *A. craccivora* across the Midwestern United States, from two host plants, alfalfa (*Medicago sativa*) and black locust (*Robinia pseudoacacia*)—putative source populations for winged migrants. Simultaneously, we sampled winged *A. craccivora* landing in pumpkin fields where they transmit viruses. Structure analyses supported host-associated differentiation by identifying two major genotypic groups: an alfalfa group containing a single multi-locus genotype and a locust group containing all others. Winged locust-group aphids landed at a much greater magnitude within focal fields during year 2 than year 1, while those in the alfalfa group remained fairly consistent. Spatial autocorrelation analyses indicated locust-group aphid movement was characterized by small-scale dispersal during year 2, likely originating from populations within 10 km. We also detected strong temporal differences in colonization from the two host plants. Early in the summer, most winged aphids (79.4%) derived from the locust group, whereas late in the summer more (58.3%) were from the alfalfa group. Because early crop growth stages are more susceptible to damage from aphid-vectored viruses, these data implicate locust as the more important source and illustrate how host-associated differentiation can be used to track dispersal and inform management of heterogeneous pest populations.

KEYWORDS

Aphis craccivora, *Medicago sativa*, multilocus genotype, *Robinia pseudoacacia*, single nucleotide polymorphism, *Trifolium* spp.

1 | INTRODUCTION

Host-associated differentiation (HAD) is a type of ecological isolation whereby herbivorous insects specialize on different plant species, creating genetic divergence among sympatric subpopulations (Berlocher & Feder, 2002; Scheffer & Hawthorne, 2007; Stireman, John, Nason, & Heard, 2005). This may eventually result in speciation (Abrahamson, Eubanks, Blair, & Whipple, 2001; Bush, 1969), but gene flow can prevent the complete isolation of subpopulations and thus HAD is a mechanism that generates and maintains intraspecific genetic diversity. Although HAD is a fundamental concept for understanding incipient species formation and basic evolutionary principles, it can also be applied as a tool to improve insect management in agricultural systems (Diehl & Bush, 1984; Medina, 2012). Numerous insect pests display HAD or related forms of population substructure (Antwi, Sword, & Medina, 2015; Bethenod et al., 2005; Brown, Frohlich, & Rosell, 1995; Pashley, 1986), yet their management strategies virtually always assume a homogenous population colonizing crops.

Accounting for predictable host races could prove especially useful for identifying the source populations and dispersal distances of small, mobile insects that are nearly impossible to otherwise track across agricultural landscapes. Aphids, for example, are minute even by insect standards and are known for their complex flight patterns, including the propensity to engage in long-distance migrations (Loxdale et al., 1993). Importantly, aphids exhibit a suite of life history characteristics (e.g., asexual reproduction) that make them prone to exhibit HAD (Dickey & Medina, 2010), and many populations display some sort of underlying genetic structure according to host-plant or geography (Delmotte, Leterme, Gauthier, Rispe, & Simon, 2002; Guillemaud, Mieuxet, & Simon, 2003; Loxdale & Lushai, 2007; Loxdale et al., 2011; Orantes, Wei Zhang, Mian, & Michel, 2012; Wilson, Sunnucks, Blackman, & Hales, 2002). However, few studies have used HAD-mediated population differences to infer the source and movement of key plant pathogen vectors: one recent study used microsatellite loci to identify the originating host plants for pea aphids, *Acyrtosiphon pisum*, migrating into legume crops across the north-western United States (Eigenbrode et al., 2016).

In this study, we used molecular markers to explore the roles of HAD and geographical scale across putative source populations in determining the spatial ecology of the cowpea aphid, *Aphis craccivora* Koch (Hemiptera: Aphididae), an asexually reproducing vector of plant viruses. Our prior research into the epidemiology of cucurbit crop viruses identified *A. craccivora* as the most important vector among a community of 53 aphid species alighting in fields (Angelella et al., 2015). While cucurbits are nonhosts for *A. craccivora* (i.e., they cannot survive and reproduce on these plants), winged

alates can nevertheless introduce viruses through a stylet-borne nonpersistent route of transmission. The source for these vectors is completely unknown, but *A. craccivora* is widely considered a specialist on legumes (Blackman, 2015), which likely cannot overwinter in the Midwest. Because this plant family (Fabaceae) is quite large, we focused primarily on two common legumes in the Midwestern United States that occur near agricultural areas and are frequently reported to contain *A. craccivora*: black locust (*Robinia pseudoacacia*) and alfalfa (*Medicago sativa*). Interestingly, prior research has shown that *A. craccivora* possess unique bacterial symbionts that maintain separate host-associated populations for these two host plants (Brady & White, 2013; Wagner et al., 2015). We therefore expect that *A. craccivora* originating from locust are genetically distinct from those coming from alfalfa.

Determining the origin of winged *A. craccivora* that serve as crop virus vectors is potentially important for several reasons. One, legumes can serve as reservoirs for cucurbit viruses, including black locust (Laney, Avanzato, & Tzanetakis, 2012), alfalfa and clover (Brunt et al., 2017). Thus, aphids originating from these plants may acquire different viruses before alighting in fields. Two, vector competence varies with host plant associations in other aphid species (Coutts, Hawkes, & Jones, 2006; Gray, Smith, Barbiéri, & Burd, 2002; McGrath & Harrison, 1995). The *A. craccivora* system is further complicated by the fact that migrants from locust and alfalfa possess different endosymbionts, which affect plant probing behaviors that mediate virus transmission (Angelella, Nalam, Nachappa, White, & Kaplan, 2017). Last, *A. craccivora* population dynamics and production of winged morphs likely vary between a crop, such as alfalfa, and an unmanaged tree like black locust. Alfalfa is regularly mown several times a year, which could allow populations to build up and induce dispersal during these disturbance events (Kennedy & Storer, 2000; Losey & Eubanks, 2000), generating pulses of migrating alates. In comparison, black locust seedlings—the preferred growth stage for aphids—are less disturbed and grow opportunistically in smaller, more widely distributed patches. Also, *A. craccivora* are commonly tended by mutualistic ants on locust, while this was almost never observed in alfalfa (G. Angelella, *personal observation*). This difference may be ecologically significant since ants usually suppress aphid wing formation (Müller, Williams, & Hardie, 2001).

Here, we describe the results from a 2-year field analysis aimed at identifying the source of *A. craccivora* flying into focal pumpkin fields where they transmit economically important viruses. To do so, we sampled populations of *A. craccivora* on alfalfa, locust, and other fabaceous host plants (e.g., clover) at local and regional scales surrounding pumpkin fields. We then used single nucleotide polymorphisms (SNPs) to test for evidence of HAD and, simultaneously, compared the genetic structure of *A. craccivora* populations with

migrant winged aphids flying into focal fields. Finally, we examined geographical patterns in genetic relatedness with spatial autocorrelation analyses to assess patterns in dispersal behavior, predicting that a tendency toward more frequent small-scale movements among populations would lead to positive local, rather than regional, spatial autocorrelation.

2 | METHODS

2.1 | Aphid collection

We trapped winged aphids landing in four focal pumpkin fields with pans (DiFonzo, Ragsdale, Radcliffe, Gudmestad, & Secor, 1997) in a 1:4 propylene glycol and water solution, collecting aphids once a week in 2012 (year 1) and 2013 (year 2). Twenty pan traps were set out in each field for 14 weeks after pumpkin plant emergence or transplant in fields and ending with the onset of plant senescence (6/19/2012–9/23/2012; 6/24/2013–9/23/2013). Focal fields were located in north-west Indiana in Tippecanoe, White, and Jasper Cos. (Figure 1; Supporting Information Table S1). We identified *A. craccivora* from all alates trapped weekly within each field and stored them separately by field and collection date at -80°C in undiluted ethanol.

We collected apterae by hand from potential source populations on black locust and with sweep nets in alfalfa or clover (*Trifolium* spp. L.). Collections from potential source populations occurred within the duration of focal field sampling, except for clover which was only sampled in year 1 due to the difficulty in locating clover fields. Any temporal variation between alfalfa and locust collection dates largely results from variable periods of aphid colonization,

as collection attempts were made throughout the sampling period until at least 20 individuals could be collected in a single sampling event. Source population locations were selected to represent a wide variety of distances both locally (ca. 3–50 km) and regionally (ca. 100–700 km) relative to the four focal pumpkin fields (Figure 1; Supporting Information Table S1). To include a more representative composition of aphid clones, or multilocus genotypes (MLGs), per population, we swept alfalfa or clover fields 10 times in 10 locations at least 30 m apart. Populations of *A. craccivora* on locust were found on seedlings, often concentrated on new growth on one location. Because populated seedlings were isolated, *A. craccivora* in locust populations were collected from a single plant at each location. All aphid samples were stored at -80°C in undiluted ethanol.

We extracted individual aphid DNA with DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). Samples with DNA concentration $<5\text{ ng}/\mu\text{l}$ were amplified with REPLI-g Mini kit (Qiagen). In total, we extracted DNA from 503 aphids collected in year 1, and 465 aphids collected in year 2. We additionally extracted DNA following this protocol from 20 winged *A. craccivora* trapped in pan traps at two pumpkin field sites in 2010 and 2011 for RAD library preparation and sequencing. The pumpkin sites were located within the radius of the four focal fields sampled in years 1 and 2.

2.2 | RAD library preparation, SNP identification and assessment of function

For RAD library preparation, we used $3\text{ }\mu\text{g}$ DNA from a pool of 20 winged *A. craccivora* trapped at two pumpkin field sites in 2010 and 2011. Library preparation followed protocol described by Etter and Johnson (2012) for RAD paired-end sequencing. The DNA

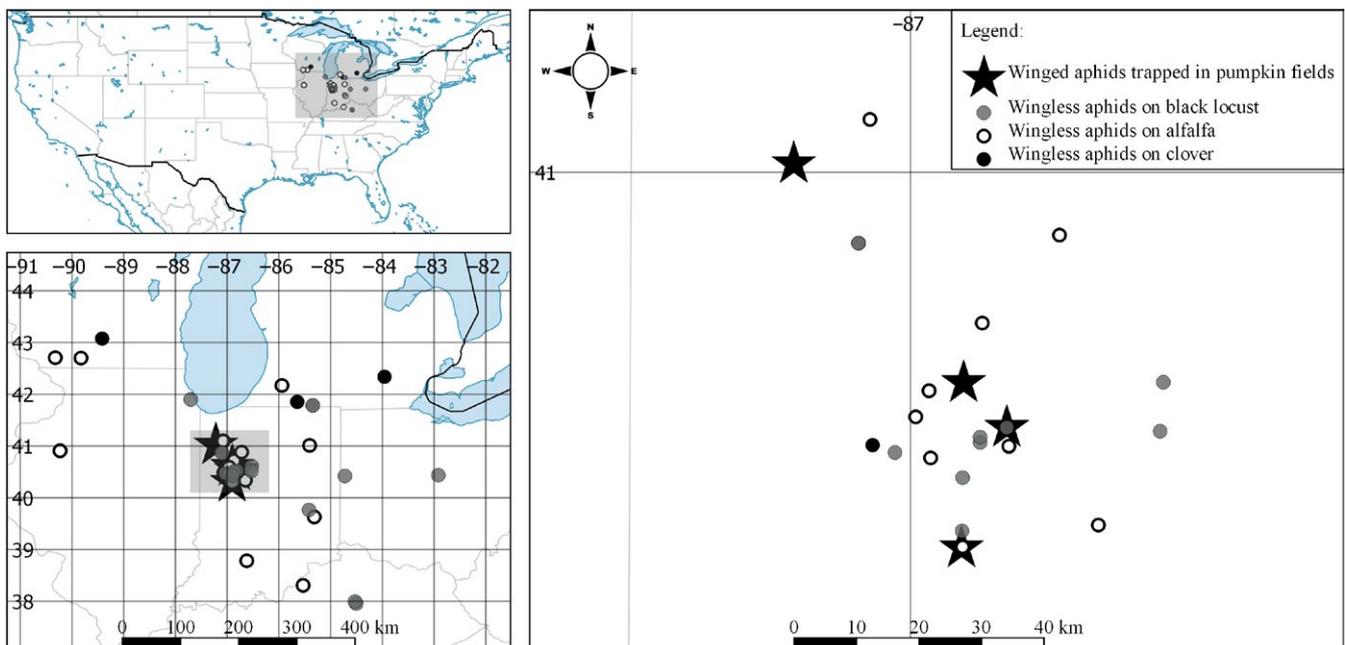


FIGURE 1 Map of *Aphis craccivora* collection sites (also see Supporting Information Table S1). Decimal degree latitudinal and longitudinal coordinates indicated

underwent a single digestion with the restriction enzyme EcoRI-HF, after which P1 adaptors incorporating six barcodes, each 12 nucleotides in length, were ligated to digested DNA. Barcodes were selected to differ as much as possible. The ligated DNA fragments were then sheared to an average 500 bp in length, and size-selected via gel electrophoresis to select samples 300–500-bp in length. Next, PE-P2 adaptors were ligated to DNA fragments. High-fidelity PCR amplified the P1 and PE-P2 adaptor-ligated DNA fragments, after a test amplification to evaluate library quality. We then sequenced the library on an Illumina MiSeq platform, which produced 6,065,758 single reads passing the filter, with a read length of 250 bp. Raw reads were exported in fastq files. We conducted library preparation and sequencing at the OARDC Molecular-Cellular Imaging Center.

The RAD tags were then used to identify de novo candidate SNPs with Stacks v. 1.19 (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013), as we lacked a reference genome for *A. craccivora*. First, we demultiplexed and cleaned raw reads with the process_radtags program. We truncated reads to 224 bp, removed those with any uncalled bases, and discarded those with low-quality scores. In all, we retained 5,347,646 high-quality reads. Next, we ran the denovo_map.pl pipeline which utilizes ustacks to align matching sequences (i.e., putative alleles), cstacks to make a catalog of the consensus loci, and sstacks to match all of the samples in the population against the catalog. We allowed a maximum of two mismatches between stacks within individuals and a minimum depth of coverage of three nucleotides for ustacks, and allowed no mismatches between stacks between individuals for cstacks.

Fifty candidate SNPs were run across all *A. craccivora* DNA samples collected in years 1 and 2, of which seven candidate SNPs failed validation (LGC Genomics LLC, Beverly, MA). We subsequently identified both aphid samples and loci with >15% missing data using GenAIEx v6.5 program (Peakall & Smouse, 2006, 2012) and discarded them from the dataset. Loci with a minor allele frequency <0.05 in one or more populations were then identified using the poppr package (Kamvar, Brooks, & Grünwald, 2015; Kamvar, Tabima, & Grünwald, 2014) and R v3.2.2 statistical software program (R Core & Team, 2015) and discarded. Additionally, we identified four nonneutral loci undergoing balancing selection, and one locus (Ac44139) undergoing positive selection using the LOSITAN program (Beaumont & Nicols, 1996; Supporting Information Figure S1). Upon examination, there was markedly greater homozygosity at locus Ac44139 in locust-origin populations. The locus Ac44139 sequence was similar to an *Ac. pisum* transposon (based on BlastX), but was not found in an *A. craccivora* transcriptome (J. White, unpublished data). Regardless, we conducted analyses using neutral loci only as those undergoing selection can disrupt gene frequency calculations such as Wright's F_{ST} (Beaumont & Nicols, 1996). Ultimately, the dataset included 817 aphid individuals and 21 loci.

Lastly, we explored the possibility of SNP effects on virus epidemiology by checking for alignment between SNP sequences and an *A. craccivora* transcriptome (J. White, unpublished data), and running a BLASTX search on those highly similar sequences (alignment

identification >98%) to identify putative functions. We then compared open reading frames within highly similar transcriptome sequences with each SNP variant with ORF-Finder (NCBI) to predict whether they would change resultant protein products.

2.3 | Genetic diversity and polymorphism

We used the Paetkau assignment test (Paetkau, Slade, Burden, & Estoup, 2004) in GenAIEx to assign samples to groups. GenAIEx also matched individual aphid samples to MLGs; however, missing data can inflate the number of MLGs by reducing matches. To correct for this, we also manually assigned samples to MLG by “best estimate,” assigning an MLG only after eliminating all other possible matches (Gitzendanner, Weekley, Germain-Aubrey, Soltis, & Soltis, 2012). We calculated pairwise codominant genetic distance between MLGs to assess the total number of differing alleles across loci. To assess genetic diversity, we followed recommendations made by Arnaud-Haond, Duarte, Alberto, and Serrão (2007) in selecting diversity parameters appropriate for clonal organisms: we used the poppr package in R statistical software program to calculate Simpson complement (D'), Simpson evenness indices (V), and the slope of the Pareto distribution (c), and calculated genotypic richness (R : $R = (G - 1)/(N - 1)$, where G = number of MLGs and N = population size). Again, following recommendations for analysis of clonal organisms by Arnaud-Haond et al. (2007), we recalculated all of the statistical analyses after removing repeat MLGs within a population to exclude the effect of clonality unless otherwise stated. Observed (H_o) and expected heterozygosity (H_e), inbreeding coefficient (F_{is}), allele frequencies, and deviation from Hardy–Weinberg equilibrium were calculated with GenAIEx. To assess the likelihood of clonality or recombination within populations sampled, indicating sexual or asexual reproduction within populations, we calculated, a \bar{r}_d modified version of the index of association (I_A) robust to data with multiple loci (Agapow & Burt, 2001), using the poppr package.

2.4 | Population genetic structure

To visualize genetic structure among and within aphid populations, we used STRUCTURE v2.3.4, which infers populations among samples and can identify migrant or admixed samples (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000). Because STRUCTURE methods assume HWE, only the datasets with all repeat MLGs removed within populations underwent the analysis. We used a burn-in length of 50,000 and 100,000 Monte Carlo Markov-Chain repetitions to test $K = 2$ –21 number of genetic groups, and each value of K was replicated 10 times. We ran admixture models with LOCPRIOR, which uses sample location data to inform the assignment of clusters in studies with fewer markers, individuals or weaker structure without increasing spurious structure detection (Hubisz, Flush, Stephens, & Pritchard, 2009; Pritchard, Stephens, & Donnelly, 2010). The probable value of K was generated using the Evanno, Regnaut, and Goudet (2005) method with STRUCTURE HARVESTER program (Earl, 2012). The graphical output was

visualized with Distruct v1.1 (Rosenberg, 2004). We calculated the proportion of individuals within each population identifying within each STRUCTURE-assigned group. To quantify genetic differentiation between STRUCTURE-assigned groups, we calculated Wright's F_{ST} (Wright, 1978).

2.5 | Spatial/Temporal patterns in migrant aphids

We assessed the relationship of codominant genotypic distance and geographical distance among sample sites with global spatial autocorrelation in GenAlEx, generating 999 bootstrapped coefficient values (r) with 999 random permutations. We included geographical distance class end points of 0, 10, 20, 50, 100, 200, 300, 400, and >400 km in analyses. The 0 km distance class was included to distinguish between aphids collected from the same population (0 km geographical distance separation) vs. from nearby populations separated by some distance within 10 km. Because it is thought most aphid dispersal occurs over short distances (Loxdale et al., 1993), incremental distance units were much closer together to identify the most influential scale within 50 km. Beyond the local scale, we increased distance classes by 100 km, ensuring multiple paired comparisons were present in each distance class. Because weather patterns and other related external factors can influence aphid population dynamics, and these patterns varied between years, years 1 and 2 were analyzed separately. Aphids were analyzed by STRUCTURE-assigned group.

Although varying sample collection dates can lead to variations in self-population assignment, with clonal diversity within populations either increasing (Michel, Zhang, Jung, Kang, & Rouf Mian, 2009; Orantes et al., 2012) or decreasing (Komazaki et al., 2011; Sunnucks, Barro, Lushai, Maclean, & Hales, 1997) over time, levels of self-population assignment in this study were very low across all populations (Supporting Information Table S2). As such, we chose not to assess frequency of self-assignment as a function of collection date. However, we assessed the temporal relationship of host-associated MLGs in focal field alatae (migrants) with a logistical regression of migrant alfalfa- or locust-group identity assigned by STRUCTURE analysis, by a full factorial cross of year and early or late seasonal

occurrence. The "early" bin contained samples collected in weeks 1–7 of the sampling period, and the "late" bin, weeks 8–14. All migrants were included in the analysis.

3 | RESULTS

3.1 | SNP transcription

Fourteen out of the 21 SNP loci were highly similar to sequences in an *A. craccivora* transcriptome search (Supporting Information Table S3A). Four of these transcriptome sequences were highly similar in a BlastX search with *Ac. pisum*, *Diuraphis noxia*, and *Myzus persicae* sequences predicting proteins as well (Supporting Information Table S3B). One locus, Ac107071, was highly similar to a sequence within the *A. craccivora* transcriptome and also with a predicted filaggrin or filaggrin-like protein in *Ac. pisum*, *D. noxia*, and *M. persicae*. An additional locus, Ac88299, was highly similar to an *A. craccivora* transcriptome sequence, which in turn matched uncharacterized proteins in *Ac. pisum*, *D. noxia*, and *M. persicae*. The SNP variants induced open reading frame changes altering an unnamed protein product within the transcriptome sequence matching locus Ac88299, but not at any other loci.

3.2 | Polymorphism and genetic diversity

Before the removal of repeat MLGs, all loci in 11 populations were not in HWE, including all alfalfa-origin populations and one locust-origin population (Supporting Information Table S4). The remaining black locust populations had almost entirely monomorphic loci. After repeat MLGs (assigned by best estimate) were removed, there was only a single locus in a focal field population, "K," not in equilibrium or monomorphic (Supporting Information Table S5). Seven repeat MLGs were identified, with memberships of 428, 256, 73, 18, 6, 3, and 3 individuals (Table 1). The remaining 31 MLGs occurred in only one individual across all populations.

Among the nonmigratory aphids, the dominant MLG was found largely in alfalfa-origin populations: 84.5% of the populations within which the largest MLG was identified were of alfalfa origin, and 98.6%

TABLE 1 Aphid multilocus genotype (MLG) assignment using best estimate method, by year and population. Letter categories indicate names of an MLG to which multiple aphids were assigned, while aphids with a unique MLG are indicated by the "Single MLG" category

Population	W	R	Q	P	O	L	D	Single MLG
Year 1 (2012)								
Migrant	0	0	0	0	0	13	0	4
Alfalfa	0	0	0	0	0	162	0	0
Locust	161	0	0	0	7	1	16	0
Clover	20	1	0	2	0	25	0	1
Clover + Alfalfa	0	0	0	0	0	22	0	0
Year 2 (2013)								
Migrant	4	4	0	0	6	8	1	18
Alfalfa	0	0	3	0	1	182	1	3
Locust	71	0	0	0	59	15	0	5

TABLE 2 Pairwise genetic distance (total number of differing alleles across loci) between each best estimate MLG and repeat MLGs. Single MLGs found in only one individual are numbered 1–31, while repeat MLGs found in at least two individuals are indicated by a letter

	L	W	D	O	P	Q	R
L:	0						
W:	21	0					
D:	16	5	0				
O:	20	1	6	0			
P:	18	11	14	10	0		
Q:	17	16	19	15	9	0	
R:	15	14	15	13	11	10	0
1:	21	22	25	23	13	8	12
2:	20	13	16	14	4	9	13
3:	18	13	16	12	12	17	7
4:	20	15	18	14	6	9	15
5:	18	15	20	14	8	1	9
6:	15	10	13	9	5	4	6
7:	17	8	11	7	1	10	8
8:	19	22	23	21	11	6	8
9:	7	12	15	11	9	4	6
10:	14	11	14	10	4	7	5
11:	17	16	17	17	5	6	8
12:	20	15	16	16	12	11	9
13:	19	20	21	17	9	14	10
14:	16	13	14	12	4	9	5
15:	22	21	22	18	6	13	13
16:	22	11	12	10	14	19	11
17:	17	4	7	3	7	10	10
18:	16	17	18	16	4	3	9
19:	20	11	16	12	12	7	9
20:	18	11	14	10	2	7	13
21:	20	19	22	20	12	7	13
22:	18	19	22	18	6	3	7
23:	21	14	19	13	7	8	16
24:	18	15	18	16	10	3	9
25:	17	12	15	13	5	8	8
26:	17	8	13	7	3	6	6
27:	15	14	15	13	7	6	12
28:	20	1	6	2	12	17	15
29:	15	10	13	11	7	8	4
30:	27	16	21	17	15	20	24
31:	16	13	16	12	4	3	9

of alfalfa-origin aphids had this MLG. The second most dominant MLG was found primarily in locust-origin populations: 90.6% of the populations within which this MLG was identified were of locust origin, and 70.3% of locust-origin aphids had this MLG. Of the remaining

repeat MLGs, two were largely locust-origin (20.0% of locust-origin aphids), one of alfalfa origin (0.9% of alfalfa-origin aphids), and two were found only in clover aphid populations. Three repeat MLGs (W, O, D) found in locust were only slightly distinct, with genetic distances of 1–6, and 28 of 31 single MLGs had total allele differences of <10 (Table 2). The dominant locust MLG (W) and dominant alfalfa MLG (L) had the greatest genetic distance between them (Table 2).

The dominant alfalfa-origin MLG exhibited excess heterozygosity, whereas a heterozygote deficit was found in the dominant locust-origin MLG, which had many monomorphic loci. The dominant alfalfa-origin MLG differentiated from all other MLGs in heterozygosity at nine loci. Locus Ac88299 was heterozygous in all aphids but the dominant alfalfa-origin MLG contained one SNP while all others expressed the alternative nucleotide. Overall, levels of genotypic diversity were low, with focal field migrant populations, a clover field (“BuckCreek”) and a black locust population (“OH”) exhibiting the highest relative diversity (Supporting Information Table S2). All multilocus disequilibrium estimates (\bar{r}_d) were significant ($p < 0.01$), indicating sexual recombination seems to be absent within *A. craccivora* across all populations (Supporting Information Table S6).

3.3 | Population genetic structure

STRUCTURE analysis results suggested a probability value of $K = 2$ distinct genetic groups (Supporting Information Figure S2). There were no differences in the optimal number of population clusters or in individuals’ group assignment comparing datasets with repeat MLGs removed after assignment by conservative vs. best estimate approach. There was very little evidence of admixture in the manually assigned dataset, but several individuals in the conservative dataset showed admixed lineages. However, due to inclusion of “unique” MLGs containing missing loci in the conservative dataset, the admixture could be a byproduct of genotyping error generated associated with missing data (Reeves, Bowker, Fettig, Tembrock, & Richards, 2016).

The STRUCTURE diagram predicts population MLG assortment into two groups, largely separated between alfalfa- and locust-origin populations (Figure 2): membership of one group contained the MLG “L,” and membership of the second group contained all other MLGs. As “L” was the dominant MLG of alfalfa-origin aphids in both years, the first STRUCTURE-assigned group will be referred to as the alfalfa group, and the second as the locust group. A separation by group between locust- and alfalfa-origin populations is clear in year 1 and somewhat mixed in year 2. The STRUCTURE diagram reflects the MLGs assigned to each group, however; examining the proportion of individuals of each group within populations illustrates the strength of separation by group between alfalfa and black locust populations in year 2 as well (Figure 3). Clover populations represented intermediate group membership, while clover + alfalfa mix populations were entirely derived from the alfalfa group (Figures 2 and 3). Migrant populations of *A. craccivora* landing in pumpkin fields derived from both groups (Figures 2 and 3). The alfalfa and locust groups had an F_{ST} of 0.156, indicating great genetic differentiation (Wright, 1978).

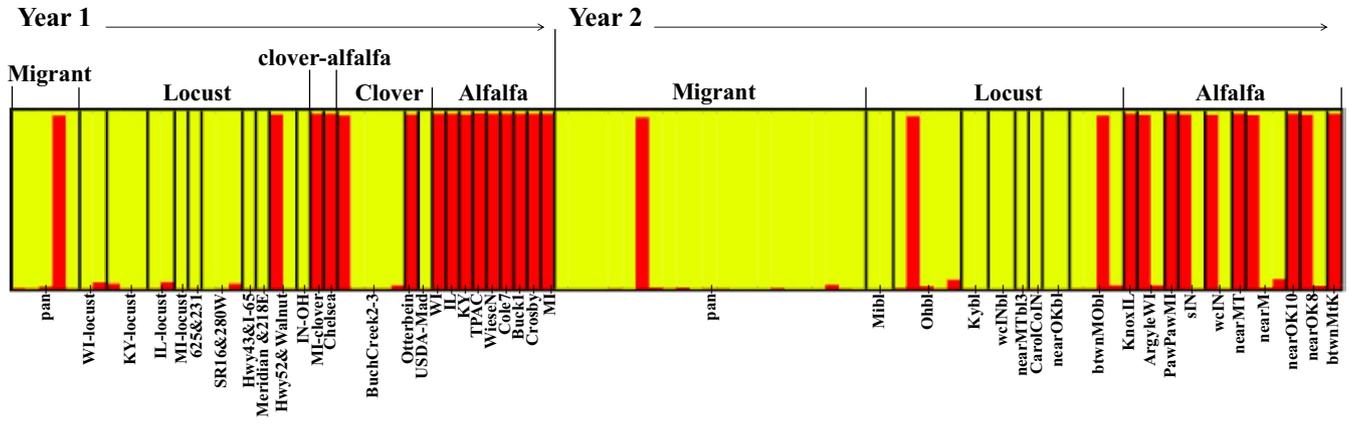


FIGURE 2 STRUCTURE diagram of $K = 2$ genetic groups, assigned to individual *Aphis craccivora*. Sampled populations are delineated and labeled below, with individual MLGs within each population represented by vertical bars of equal size, colored to indicate STRUCTURE-assigned group membership (red = alfalfa group, yellow = locust group). Actual population sample type and collection year are indicated above the graph. Analysis was run under an admixture model, with combined year 1 (2012) and year 2 (2013) datasets, excluding repeat MLGs after best estimate MLG assignment

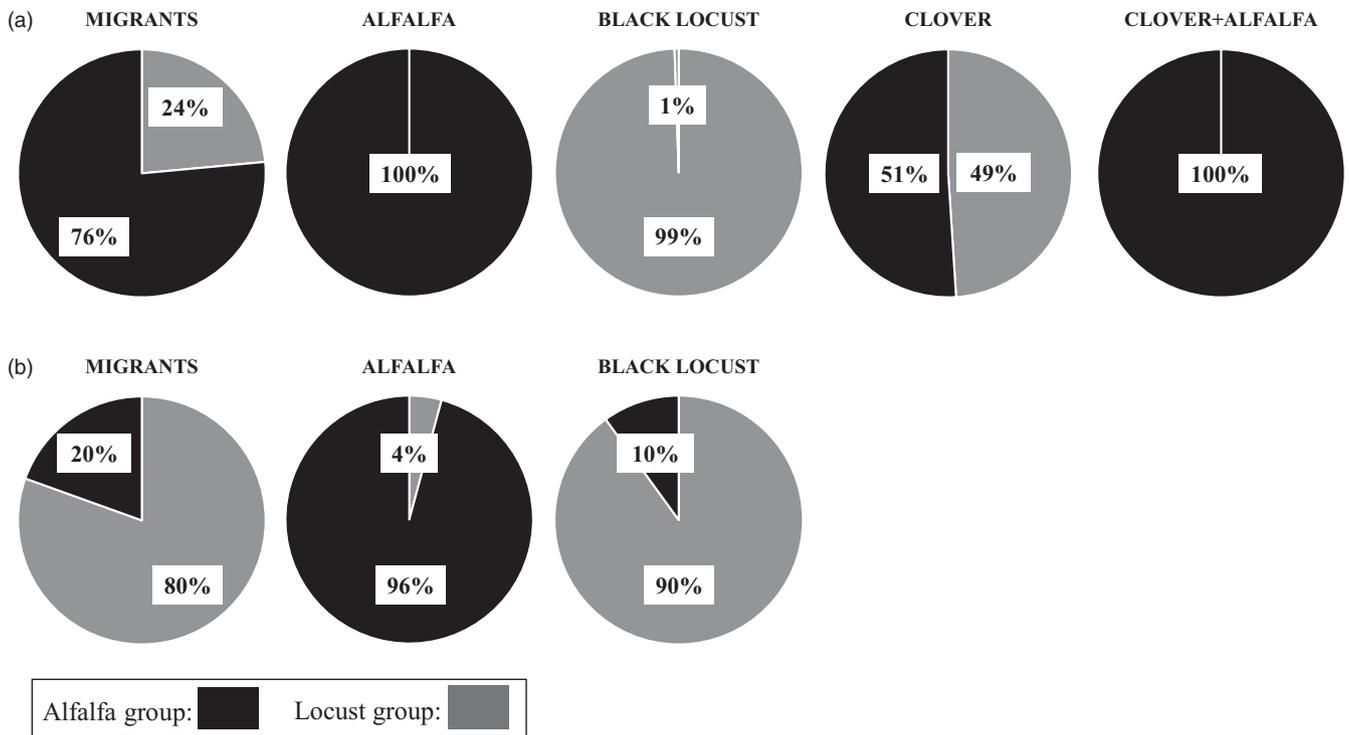


FIGURE 3 Proportion of STRUCTURE-assigned group among individual wingless *Aphis craccivora* collected by host plant (“alfalfa,” “black locust,” “clover,” “clover + alfalfa” mix) and winged *Aphis craccivora* trapped within focal pumpkin fields (“migrants”). (a) Year 1 (2012) dataset, (b) Year 2 (2013) dataset

3.4 | Spatial/Temporal patterns in migrant aphids

Only aphids assigned by STRUCTURE analysis to the locust group were analyzed, as all members of the alfalfa group were genetically identical and therefore lacked spatial genetic structure. The degree of spatial patterning within locust-group aphids varied by year. In year 1, aphids exhibited positive spatial autocorrelation within a given population at 0 km ($r = 0.16, p = 0.001$), and very weak negative

effects across populations at 10, 50, and 400 km as well ($r = -0.073, p = 0.001; r = -0.027, p = 0.003; r = -0.008, p = 0.045$, respectively; Figure 4a). In year 2, aphids exhibited positive spatial autocorrelation again within a given population at 0 km ($r = 0.37, p = 0.001$), but also across populations at 10 km, and 100 km ($r = 0.15, p = 0.001; r = 0.048, p = 0.001$, respectively), and exhibited negative spatial autocorrelation at 20, 50, and 300 km ($r = -0.14, p = 0.001; r = -0.040, p = 0.002; r = -0.093, p = 0.001$, respectively; Figure 4b).

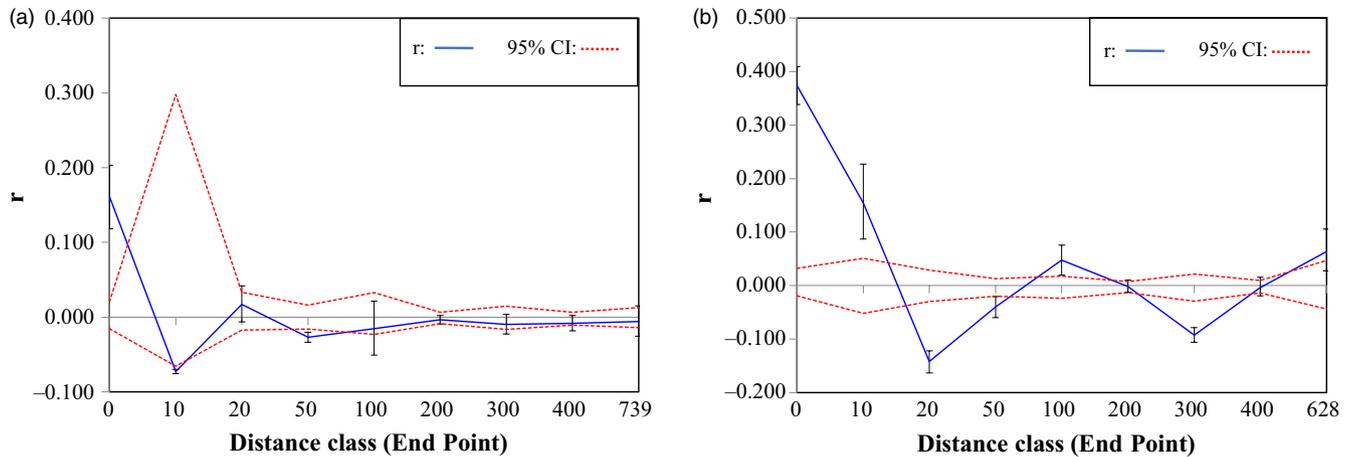


FIGURE 4 Spatial autocorrelation analyses of *Aphis craccivora* genetic relatedness by geographical distance for locust-group aphids, as assigned by STRUCTURE analysis. Error bars indicate r estimates from 999 bootstraps. Red dashed lines indicate $\pm 95\%$ confidence intervals. (a) Year 1 (2012) dataset, (b) Year 2 (2013) dataset

Results were similar following analyses of locust-group datasets excluding repeat MLGs. Aphids were marginally significant within a given population at 0 km in both years (year 1: $r = 0.21$, $p = 0.053$; year 2: $r = 0.087$, $p = 0.053$; Supporting Information Figure S3A,B). In year 2, aphids also showed positive spatial autocorrelation across populations within 10 km ($r = 0.19$, $p = 0.037$), and in year 1 populations within 300 and >400 km ($r = 0.25$, $p = 0.023$; $r = 0.23$, $p = 0.026$, respectively). Negative spatial autocorrelation occurred in year 1 populations within 400 km ($r = -0.084$, $p = 0.038$) and in year 2 populations within 20 km and 300 km ($r = -0.126$, $p = 0.007$; $r = -0.10$, $p = 0.019$, respectively).

The membership of migrant aphids landing within focal fields to either the locust-group or alfalfa group varied significantly between years and between early and late seasonal timing (year: $\chi^2 = 4.16$, $SE = 0.97$, $p = 0.041$; seasonal timing: $\chi^2 = 3.90$, $SE = 0.84$, $p = 0.048$). In total, there were 13 alfalfa group *A. craccivora* alates in year 1 migrant populations and 8 in year 2, with relatively more landing in focal fields late-season (year 1:69.2%; year 2:62.5%; Table 3). Locust-group migrants varied between years with a total of four individuals trapped in focal fields in year 1, evenly distributed between early- and late-season (50.0%). In year 2, 33 total alates landed in focal fields, with a greater percentage arriving early in the season (75.8%; Table 2).

4 | DISCUSSION

Genetic grouping of *A. craccivora* individuals clearly varied by alfalfa or locust host association according to STRUCTURE analysis, making it possible to identify putative host plant sources of winged aphid migrants landing in focal fields. The host plant origin of migrants was influenced by both year and early/late seasonal timing. Alfalfa-group migrants tended to occur later in the season with relatively less year-to-year variation in magnitude (relative ratio of early-to-late-season migrants was ca. 1:2; $n_{\text{year1}} = 13$, $n_{\text{year2}} = 8$), whereas locust-derived aphids differed in early:late proportion and magnitude

TABLE 3 Total number of winged *Aphis craccivora* trapped in lower north-western Indiana focal pumpkin fields, and number of MLGs grouped with predominately black locust (N_{LG}) or alfalfa (N_{AG}) populations early and late in the growing season (early: weeks 1–7; late: weeks 8–14); MLGs assigned by STRUCTURE analysis

	Early			Late		
	N	N_{LG}	N_{AG}	N	N_{LG}	N_{AG}
Year 1 (2012)	6	2	4	11	2	9
Year 2 (2013)	28	25	3	13	8	5

by year (1:1 early-to-late ratio in year 1, $n_{\text{year1}} = 4$; 25:8 in year 2, $n_{\text{year2}} = 33$). Thus, alfalfa may be a relatively minor contributor of migrants in focal pumpkin fields with low total numbers overall and biased toward late-season activity, and the contribution of migrants from black locust may fluctuate more by year, but years with greater migrant production could generate more early-season dispersal.

The pattern has important implications for virus epidemiology. Black locust, alfalfa, and clover are host to several pumpkin viruses. Locust is host to papaya ringspot virus type W (PRSV-W) and watermelon mosaic virus type 2 (WMV; Laney et al., 2012), and alfalfa and at least two species of clover (*T. incarnatum*, *T. pratense*) are susceptible to WMV (Brunts et al., 2017). Cucumber mosaic virus was also isolated from alfalfa and white clover (*T. repens*) in Wisconsin (Mueller, Groves, & Gratton, 2012) and is an additional aphid-vector periodically found in pumpkins in the midwestern region of the United States (Jossey & Babadoost, 2008; Walters, Kindhart, Hobbs, & Eastburn, 2003). Crops are more susceptible to reductions in yield due to virus infection in earlier stages of growth; thus, vectors interacting with crops earlier in the growing season are of far greater concern.

Although the relatively low number of migrating aphids we trapped in focal fields is not unrealistic, it does raise some concern regarding sample size. The total *A. craccivora* trapped per field was comparable to that of a previous study in area pumpkin fields

(Angelella et al., 2015), and to that of a similar study of *Ac. pisum* migrant population genetics in peas (Eigenbrode et al., 2016). However, because migrant trap catches can be quite low within crop fields, it is possible that patterns in MLG composition could be impacted by sample size and the ability to obtain a representative sample. As such, we would recommend investigating whether increasing the number and/or density of pan traps per crop field could lead to a more optimal sampling effort for migrant vectors such as *A. craccivora*. If, however, aphids are preferentially landing in pan traps relative to the surrounding crop canopy (Boiteau, 1990), it is possible that roughly the same total abundance could be trapped per field irrespective of trap density. In fact, we obtained roughly similar trap catches to that of Angelella et al. (2015) and Eigenbrode et al. (2016) despite using a greater number of pan traps per field: 20 field⁻¹ compared to 5 field⁻¹ and 3 field⁻¹, respectively. Notwithstanding, care should be taken to optimize the migrant vector sampling method.

Our results support patterns of genetic relatedness and differentiation among locust-group *A. craccivora* across populations by geographical distance and year. Alfalfa-group *A. craccivora* were genetically identical both within and among collection sites, and thus exhibited no genotypic spatial structure. Results of year 2 locust-origin aphids may indicate a connection between small-scale geographical movement and migrant alightment in focal fields. Positive genetic structure within 10 km and differentiation at 20 km suggest small-scale movement among populations was common in year 2, corresponding with a large number of locust-origin migrants in focal fields. In contrast, the differentiation of year 1 locust-origin aphids within even the small geographical scale of 10 km may suggest a lack of small-scale dispersal, which corresponds with infrequent migrant arrival in focal fields. This smaller-scale dispersal range could make repositioning crop fields to alleviate influx of migrant aphids more feasible, potentially helping growers to move away from areas prone to repeated virus infection. Locust-group aphids also showed some indication of moderate migratory dispersal capabilities, as evidenced by differentiation at moderate geographical distances of 300–400 km in both years. This would be consistent with a spring recolonization of the Midwest from southern overwintering populations requiring regional-scale movement. However, these overwintering populations would not likely be located at the relatively high latitudes within the 300–400 km range of the focal pumpkin fields (Dykstra et al., 2014).

Aphis craccivora populations on alfalfa were dominated by a single MLG characterized by excess heterozygosity. While locust populations had one dominant MLG, an additional MLG was moderately represented among individuals in year 2, and several other MLGs were also present at low frequency in both year 1 and year 2. In contrast to the dominant alfalfa-origin MLG, the dominant locust-origin MLG exhibited homozygote excess. There are multiple possibilities for these outcomes, two of which include inbreeding (Simon & Hebert, 1995), or null alleles (Foltz, 1986). These possibilities are unlikely for the *A. craccivora* in this study because we confirmed the likelihood of obligate parthenogenesis and strictly asexual reproduction, and we selectively removed all

loci suspected to contain null alleles. Another potential explanation is the Wahlund effect, which can occur when subpopulation structure leads to artificially decreased heterozygosity (Wahlund, 1928); for example, if subpopulations are separated geographically or temporally, genetic drift could subsequently occur within these subpopulations decreasing heterozygosity (Prugnolle & De Meeûs, 2010). The Wahlund effect is possible: spatial autocorrelation analysis indicated geographical isolation among locust-group populations around 20 and 300 km apart in year 2. Remaining possibilities include side-effects of life cycle, and clonal selection. There is some suggestion that asexual anholocyclic lifecycle might favor the production of homozygote excess (Simon et al., 1999), because patterns of mutation might accumulate and persist throughout the overwintering and recolonization of a single plant-host (Loxdale et al., 2011). Conversely, others have speculated that asexual populations tend to favor heterozygosity by accumulating mutations (Balloux, Lehmann, & Meeûs, 2003; Delmotte et al., 2002; Harrison John & Mondor, 2011; Lynch, 1984; Vorburger, Lancaster, & Sunnucks, 2003) while sexual populations select for homozygosity in favored alleles (Hales, Tomuik, Wöhrmann, & Sunnucks, 1997; Simon & Hebert, 1995; Simon et al., 1999).

Nevertheless, the formation of predominant “superclones” can indicate relative fitness advantage (Vorburger et al., 2003). The concept of clonal amplification (Sunnucks et al., 1997) predicts clones with greater fitness will reproduce more rapidly to eventually outnumber other clones. This concept has found support in several aphid species, including *Myzus persicae* (Vorburger 2005), *Aphis glycines* (Wenger, Monica Ramstad, Mian, & Michel, 2014), *Aphis nerii* (Harrison John & Mondor, 2011), and *Ac. pisum* (Eigenbrode et al., 2016). Interestingly, authors of one study of aphids on tansy (*Tanacetum vulgare* L.) found no evidence of dominant clones and contrasted this to *M. persicae* and other pest species of intensively cultivated monocultures (Loxdale et al., 2011). If managed monoculture crops exert a stronger selective pressure on aphid MLGs, it would follow that aphid specialists in alfalfa would experience stronger selective pressure than those on unmanaged black locust seedlings. Unmanaged locusts also likely present a more varied nutritional profile and host quality; thus, naturally occurring black locust diversity may be relatively high (Chang, Bongarten, & Hamrick, 1998), and additionally, black locusts are thought to associate with a high diversity of rhizobia symbionts leading to variations in N fixation, total N, and plant growth (Batzli, Graves, & Berkum, 1992). The frequency of unique MLGs within focal field migrants also suggests that *A. craccivora* may be arriving from other populations, or even host plants other than alfalfa or black locust. However, it is also likely that mitotic mutations during clonal reproduction have resulted in multiple MLGs descending from the same clonal lineage if they exhibit low genetic distance (differences in alleles length; Arnaud-Haond et al., 2007). Indeed, at least two repeat MLGs found in locust populations and 28/31 single MLGs were only slightly different from more numerically dominant MLG, suggesting they may belong to the same multilocus lineage (MLL). Thus, it is likely that the vast

majority of migrants derived from an alfalfa-associated or black locust-associated genotype.

Genetic grouping by locust- or alfalfa-association is compelling in light of recent evidence of distinct locust-associated and alfalfa-host-associated populations driven by facultative symbiont infection (Wagner et al., 2015). *Aphis craccivora* populations on black locust are usually infected with the facultative endosymbiont, *Arsenophonus*, while those on alfalfa and other plant hosts are not (Brady & White, 2013). Wagner et al. (2015) found that *Arsenophonus* enhances aphid performance on black locust while decreasing it on other host plants, essentially facilitating its use as a host plant. They also observed *A. craccivora* clonal lines originating on alfalfa were especially well-adapted to alfalfa: while locust-origin clonal lines infected with *Arsenophonus* significantly outperformed on locust compared to alfalfa, alfalfa-origin clonal lines transacted with *Arsenophonus* performed equally well on locust and alfalfa. It would be interesting to assess facultative endosymbiont infection relative to population genotypic structure and dispersal among host-plant populations. The endosymbionts are horizontally transferred (Oliver, Degnan, Burke, & Moran, 2010; Sandström, Russell, White, & Moran, 2001), which could facilitate a sustained relationship between clonal populations and black locust, resulting in the accumulation of adaptive mutations in an MLG. In fact, recent research showed that genetic lineage, rather than endosymbiont association or host-plant diet, drives the toxicity of black locust-origin *A. craccivora* to the predator *Harmonia axyridis*, while alfalfa-origin lineages are nontoxic and vulnerable to predation (White, McCord, Jackson, Dehnel, & Lenhart, 2017). The alfalfa "superclone," in contrast, could not be attributed to facultative endosymbiont infection.

It is possible that host race formation within locust- and alfalfa-associated populations has also generated phenotypic changes to behavior or physiology impacting virus transmission efficiency, and/or the likelihood of interacting with virus reservoirs and susceptible crop plants. Although it was evident that many of the study SNP sequences are being transcribed in *A. craccivora*, it is unclear at present whether they would result in phenotypic differentiation. But this would not be unprecedented. For example, regional populations of *Schizaphis graminum* (Rondani) within the United States varied in their host-plant associations and ability to transmit five viruses that cause barley yellow dwarf disease (Gray et al., 2002). The authors found that the poorest virus vectors were cyclically parthenogenetic, holocyclic populations, and speculated that because one might expect higher rates of genetic diversity in those populations due to sexual recombination and random mating, it could follow that vector competence is more readily retained in obligately parthenogenetic, anholocyclic *S. graminum* if it is a recessive trait. The obligately parthenogenetic nature of *A. craccivora* reproduction could similarly facilitate the retention of MLGs with enhanced vector competence. Follow-up studies could examine how relative transmission efficiencies vary among *A. craccivora* alfalfa- and locust-associated clonal types.

5 | CONCLUSION

We observed a pattern of distinct genotypic *A. craccivora* grouping by alfalfa- or black locust-origin, suggesting HAD has generated two host races. Aphids placed into the predominantly alfalfa group by STRUCTURE analysis possessed a single MLG, which was by far the dominant MLG found in aphid populations on alfalfa. Aphids placed in the locust group largely possessed one dominant MLG, but were more diverse and contained several additional MLGs. Due to the lack of diversity among alfalfa aphids, we could not detect genetic structure relative to geographical distance; however, locust aphids supported a prediction for small-scale dispersal in year 2, with some evidence of moderate migratory dispersal capabilities in both years. Due to evidence for an appreciable dispersal capability, the potential for generating large numbers of migrants, and the early-season timing of dispersal flights, our data suggest that locust-origin migrants are more important vectors in the pumpkin virus pathosystem.

ACKNOWLEDGEMENTS

We thank J. White for Kentucky specimen contributions and for searching their *Aphis craccivora* transcriptome for SNP sequence hits, and J. Anderson, C. Krupke, J. Holland, and anonymous reviewers for comments on the manuscript. Financial support was provided by the USDA North Central Region IPM Grants Program, grant 11-34103-30723, and the NIFA AFRI Predoctoral Fellowship program, grant 11588839. We would also like to thank Virginia Tech's Open Access Subvention Fund for supporting publication costs.

CONFLICT OF INTEREST

None declared.

DATA ARCHIVING STATEMENT

All raw RAD library sequences are available at NCBI SRA: BioProject accession #PRJNA503177, BioSample accession #SAMN10354042. The validated SNPs developed in this study as well as the genotypic, geographical, population and host-plant or pan collection source data have been deposited in Dryad at <https://doi.org/10.5061/dryad.3560pr3>.

ORCID

Gina Marie Angelella  <https://orcid.org/0000-0003-2874-9470>

REFERENCES

Abrahamson, W. G., Eubanks, M. D., Blair, C. P., & Whipple, A. V. (2001). Gall flies, inquilines, and goldenrods: A model for host-race formation

- and sympatric speciation. *American Zoologist*, 41(4), 928–938. <https://doi.org/10.1093/icb/41.4.928>
- Agapow, P. M., & Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, 1(1–2), 101–102. <https://doi.org/10.1046/j.1471-8278.2000.00014.x>
- Angelella, G. M., Egel, D. S., Holland, J. D., Nemacheck, J. A., Williams, C. E., & Kaplan, I. (2015). Differential life history trait associations of aphids with nonpersistent viruses in cucurbits. *Environmental Entomology*, 44(3), 562–573. <https://doi.org/10.1093/ee/nvv058>
- Angelella, G. M., Nalam, V., Nachappa, P., White, J. A., & Kaplan, I. (2017). Endosymbionts differentially alter exploratory probing behavior of a nonpersistent plant virus vector. *Microbial Ecology*, 76(2), 453–458. <https://doi.org/10.1007/s00248-017-1133-5>
- Antwi, J. B., Sword, G. A., & Medina, R. F. (2015). Host-associated differentiation in a highly polyphagous, sexually reproducing insect herbivore. *Ecology and Evolution*, 5(13), 2533–2543. <https://doi.org/10.1002/ece3.1526>
- Arnaud-Haond, S., Duarte, C. M., Alberto, F., & Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular Ecology*, 16(24), 5115–5139. <https://doi.org/10.1111/j.1365-294X.2007.03535.x>
- Balloux, F., Lehmann, L., & de Meeûs, T. (2003). The population genetics of clonal and partially clonal diploids. *Genetics*, 164(4), 1635–1644.
- Batzli, J. M. C., Graves, W. R., & Van Berkum, P. (1992). Diversity among rhizobia effective within *Robinia pseudoacacia* L. *Applied and Environmental Microbiology*, 58(7), 2137–2143.
- Beaumont, M. A., & Nicols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, 263, 1619–1626. <https://doi.org/10.1098/rspb.1996.0237>
- Berlacher, S. H., & Feder, J. L. (2002). Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annual Review of Entomology*, 47(1), 773–815. <https://doi.org/10.1146/annurev.ento.47.091201.145312>
- Bethenod, M. T., Thomas, Y., Francois Rousset, B., Frérot, L. P., Genestier, G., & Bourguet, D. (2005). Genetic isolation between two sympatric host plant races of the European corn borer, *Ostrinia nubilalis* Hübner. II: Assortative mating and host-plant preferences for oviposition. *Heredity* 94(2):264–270. <https://doi.org/10.1038/sj.hdy.6800611>
- Blackman, R. L. (2015). *Aphids on the world's plants*. London, UK: Natural History Museum. <http://www.aphidsonworldsplants.info>
- Boiteau, G. (1990). Effect of trap color and size on relative efficiency of water-pan traps for sampling alate aphids (Homoptera: Aphididae) on potato. *Journal of Economic Entomology*, 83(3), 937–942. <https://doi.org/10.1093/jee/83.3.937>
- Brady, C. M., & White, J. A. (2013). Cowpea aphid (*Aphis craccivora*) associated with different host plants has different facultative endosymbionts. *Ecological Entomology*, 38(4), 433–437.
- Brown, J. K., Frohlich, D. R., & Rosell, R. C. (1995). The sweetpotato or silverleaf whiteflies: Biotypes of *Bemisia tabaci* or a species complex? *Annual Review of Entomology*, 40(1), 511–534. <https://doi.org/10.1146/annurev.en.40.010195.002455>
- Brunt, A. A., Crabtree, K., Dallwitz, M. J., Gibbs, A. J., Watson, L., & Zurcher, E. J. (Eds.). (2017). *Plant Viruses Online: Descriptions and Lists from the VIDE Database*. Version: 16th January 1997. Retrieved from <http://biology.anu.edu.au/Groups/MES/vide/>
- Bush, G. L. (1969). Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera, Tephritidae). *Evolution*, 23(2), 237–251.
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, 1(3):171–182.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Chang, C.-S., Bongarten, B., & Hamrick, J. (1998). Genetic structure of natural populations of black locust (*Robinia pseudoacacia* L.) at Coweeta, North Carolina. *Journal of Plant Research*, 111(1):17–24.
- Coutts, B. A., Hawkes, J. R., & Jones, R. A. C. (2006). Occurrence of Beet western yellows virus and its aphid vectors in over-summering broad-leaved weeds and volunteer crop plants in the grainbelt region of south-western Australia. *Australian Journal of Agricultural Research*, 57(9), 975–982.
- Delmotte, F., Leterme, N., Gauthier, J.-P., Rispé, C., & Simon, J.-C. (2002). Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Molecular Ecology*, 11(4), 711–723. <https://doi.org/10.1046/j.1365-294X.2002.01478.x>
- Dickey, A. M., & Medina, R. F. (2010). Testing host-associated differentiation in a quasi-endophage and a parthenogen on native trees. *Journal of Evolutionary Biology*, 23(5), 945–956. <https://doi.org/10.1111/j.1420-9101.2010.01962.x>
- Diehl, S. R., & Bush, G. L. (1984). An evolutionary and applied perspective of insect biotypes. *Annual Review of Entomology*, 29(1), 471–504. <https://doi.org/10.1146/annurev.en.29.010184.002351>
- DiFonzo, C. D., Ragsdale, D. W., Radcliffe, E. B., Gudmestad, N. C., & Secor, G. A. (1997). Seasonal abundance of aphid vectors of Potato virus Y in the Red River Valley of Minnesota. *Journal of Economic Entomology*, 90(3), 824–831.
- Dykstra, H. R., Weldon, S. R., Martinez, A. J., White, J. A., Hopper, K. R., Heimpel, G. E., ... Oliver, K. M. (2014). Factors limiting the spread of the protective symbiont *Hamiltonella defensa* in *Aphis craccivora* aphids. *Applied and Environmental Microbiology*, 80(18), 5818–5827. <https://doi.org/10.1128/AEM.01775-14>
- Earl, D. A. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Eigenbrode, S. D., Davis, T. S., Adams, J. R., Husebye, D. S., Waits, L. P., & Hawthorne, D. (2016). Host-adapted aphid populations differ in their migratory patterns and capacity to colonize crops. *Journal of Applied Ecology*, 53(5), 1382–1390. <https://doi.org/10.1111/1365-2664.12693>
- Etter, P. D., & Johnson, E. (2012). RAD paired-end sequencing for local de novo assembly and SNP discovery in non-model organisms. In F. Pompanon, & A. Bonin (Eds.), *Data production and analysis in population genomics* (pp. 135–151). New York City, NY: Humana Press.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164(4), 1567–1587.
- Foltz, D. W. (1986). Null alleles as a possible cause of heterozygote deficiencies in the oyster *Crassostrea virginica* and other bivalves. *Evolution*, 40(4), 869–870.
- Gitzendanner, M. A., Weekley, C. W., Germain-Aubrey, C. C., Soltis, D. E., & Soltis, P. S. (2012). Microsatellite evidence for high clonality and limited genetic diversity in *Ziziphus celata* (Rhamnaceae), an endangered, self-incompatible shrub endemic to the Lake Wales Ridge, Florida, USA. *Conservation Genetics*, 13(1), 223–234. <https://doi.org/10.1007/s10592-011-0287-9>
- Gray, S. M., Smith, D. M., Barbierri, L., & Burd, J. (2002). Virus transmission phenotype is correlated with host adaptation among genetically diverse populations of the aphid *Schizaphis graminum*. *Phytopathology*, 92(9), 970–975.

- Guillemaud, T., Mieuze, L., & Simon, J.-C. (2003). Spatial and temporal genetic variability in French populations of the peach-potato aphid, *Myzus persicae*. *Heredity*, 91(2), 143. <https://doi.org/10.1038/sj.hdy.6800292>
- Hales, D. F., Tomuik, J., Wöhrmann, K., & Sunnucks, P. (1997). Evolutionary and genetic aspects of aphid biology: A review. *European Journal of Entomology*, 94(1), 1–55.
- Harrison John, S., & Mondor, E. B. (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(3), e17524. <https://doi.org/10.1371/journal.pone.0017524>
- Hubisz, M. J., Flush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322–1332. <https://doi.org/10.1111/j.1755-0998.2009.02591.x>
- Jossey, S., & Babadoost, M. (2008). Occurrence and distribution of pumpkin and squash viruses in Illinois. *Plant Disease*, 92(1): 61–68.
- Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, 6, 208. <https://doi.org/10.3389/fgene.2015.00208>
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281.
- Kennedy, G. G., & Storer, N. P. (2000). Life systems of polyphagous arthropod pests in temporally unstable cropping systems. *Annual Review of Entomology*, 45(1), 467–493. <https://doi.org/10.1146/annurev.ento.45.1.467>
- Komazaki, S., Toda, S., Shigehara, T., Kanazaki, S., Izawa, H., Nakada, K., & Souda, E. (2011). The genetic structure of *Aphis gossypii* populations in Japanese fruit orchards. *Entomologia Experimentalis Et Applicata*, 140(2), 171–179. <https://doi.org/10.1111/j.1570-7458.2011.01145.x>
- Laney, A. G., Avanzato, M. V., & Tzanetakis, I. E. (2012). High incidence of seed transmission of *Papaya ringspot virus* and *Watermelon mosaic virus*, two viruses newly identified in *Robinia pseudoacacia*. *European Journal of Plant Pathology*, 134(2), 227–230. <https://doi.org/10.1007/s10658-012-9985-5>
- Losey, J. E., & Eubanks, M. D. (2000). Implications of pea aphid host-plant specialization for the potential colonization of vegetables following post-harvest emigration from forage crops. *Environmental Entomology*, 29(6), 1283–1288. <https://doi.org/10.1603/0046-225X-29.6.1283>
- Loxdale, H. D., Hardie, J., Halbert, S., Footitt, R., Kidd, N. A. C., & Carter, C. I. (1993). The relative importance of short- and long-range movement of flying aphids. *Biological Reviews*, 68(2), 291–311. <https://doi.org/10.1111/j.1469-185X.1993.tb00998.x>
- Loxdale, H. D., & Lushai, G. (2007). Population genetic issues: The unfolding story using molecular markers. In R. Harrington, & H. F. van Emden (Eds.), *Aphids as crop pests* (pp. 31–67). Wallingford, Oxfordshire, UK: CABI.
- Loxdale, H. D., Schoefer, G., Wiesner, K. R., Nyabuga, F. N., Heckel, D. G., & Weisser, W. W. (2011). Stay at home aphids: Comparative spatial and seasonal metapopulation structure and dynamics of two specialist tansy aphid species studied using microsatellite markers. *Biological Journal of the Linnean Society*, 104(4), 838–865. <https://doi.org/10.1111/j.1095-8312.2011.01761.x>
- Lynch, M. (1984). Destabilizing hybridization, general-purpose genotypes and geographic parthenogenesis. *Quarterly Review of Biology*, 59(3), 257–290. <https://doi.org/10.1086/413902>
- McGrath, P. F., & Harrison, B. D. (1995). Transmission of tomato leaf curl geminiviruses by *Bemisia tabaci*: Effects of virus isolate and vector biotype. *Annals of Applied Biology*, 126(2), 307–316. <https://doi.org/10.1111/j.1744-7348.1995.tb05368.x>
- Medina, R. F. (2012). Implications of host-associated differentiation in the control of pest species. In P. Barbosa, D. K. Letourneau, & A. Agrawal (Eds.), *Insect outbreaks revisited* (pp. 291–310). Chichester, UK: John Wiley & Sons Ltd.
- Michel, A. P., Zhang, W., Jung, J. K., Kang, S.-T., & Rouf Mian, M. A. (2009). Population genetic structure of *Aphis glycines*. *Environmental Entomology*, 38(4), 11.
- Mueller, E. E., Groves, R. L., & Gratton, C. (2012). Crop and non-crop plants as potential reservoir hosts of Alfalfa mosaic virus and Cucumber mosaic virus for spread to commercial snap bean. *Plant Disease*, 96(4): 506–514. <https://doi.org/10.1094/PDIS-02-11-0089>
- Müller, C. B., Williams, I. S., & Hardie, J. (2001). The role of nutrition, crowding and interspecific interactions in the development of winged aphids. *Ecological Entomology*, 26(3), 330–340. <https://doi.org/10.1046/j.1365-2311.2001.00321.x>
- Oliver, K. M., Degnan, P. H., Burke, G. R., & Moran, N. A. (2010). Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual Review of Entomology*, 55, 247–266. <https://doi.org/10.1146/annurev-ento-112408-085305>
- Orantes, L. C., Wei Zhang, M. A., Mian, R., & Michel, A. P. (2012). Maintaining genetic diversity and population panmixia through dispersal and not gene flow in a holocyclic heteroecious aphid species. *Heredity*, 109(2), 127–134. <https://doi.org/10.1038/hdy.2012.21>
- 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. *Molecular Ecology*, 13(1), 55–65. <https://doi.org/10.1046/j.1365-294X.2004.02008.x>
- Pashley, D. P. (1986). Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae): A sibling species complex? *Annals of the Entomological Society of America*, 79(6), 898–904. <https://doi.org/10.1093/aesa/79.6.898>
- Peakall, R., & Smouse, P. E. (2006). Genalex 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6, 288–295.
- Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28(19):2537–2539.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2010). *Documentation for STRUCTURE software: Version 2.3*. Chicago, IL: University of Chicago.
- Prugnolle, F., & De Meeûs, T. (2010). Apparent high recombination rates in clonal parasitic organisms due to inappropriate sampling design. *Heredity*, 104(2), 135–140. <https://doi.org/10.1038/hdy.2009.128>
- R Core Team (2015). *R: A language and environment for statistical computing 3.2.2*. Vienna, Austria: The R Foundation for Statistical Computing.
- Reeves, P. A., Bowker, C. L., Fetting, C. E., Tembrock, L. R., & Richards, C. M. (2016). Effect of error and missing data on population structure inference using microsatellite data. [bioRxiv:080630](https://doi.org/10.1101/080630).
- Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Resources*, 4(1), 137–138.
- Sandström, J. P., Russell, J. A., White, J. P., & Moran, N. A. (2001). Independent origins and horizontal transfer of bacterial symbionts of aphids. *Molecular Ecology*, 10(1), 217–228. <https://doi.org/10.1046/j.1365-294X.2001.01189.x>
- Scheffer, S. J., & Hawthorne, D. J. (2007). Molecular evidence of host-associated genetic divergence in the holly leafminer *Phytomyza glabricola* (Diptera: Agromyzidae): Apparent discordance among marker systems. *Molecular Ecology*, 16(13), 2627–2637. <https://doi.org/10.1111/j.1365-294X.2007.03303.x>
- Simon, J.-C., Baumann, S., Sunnucks, P., Hebert, P. D. N., Pierre, J.-S., Le Gallic, J.-F., & Dedryver, C.-A. (1999). Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, 8(4), 531–545.

- Simon, J.-C., & Hebert, P. D. N. (1995). Patterns of genetic variation among Canadian populations of the bird cherry-oat aphid, *Rhopalosiphum padi* L. (Homoptera: Aphididae). *Heredity*, *74*(4), 346–353.
- Stireman, I. I. I., John, O., Nason, J. D., & Heard, S. B. (2005). Host-associated genetic differentiation in phytophagous insects: General phenomenon or isolated exceptions? *Evidence from a golden-rod-insect Community*. *Evolution*, *59*(12), 2573–2587. <https://doi.org/10.1554/05-222.1>
- Sunnucks, P., De Barro, P., Lushai, G., Maclean, N., & Hales, D. (1997). Genetic structure of an aphid studied using microsatellites: Cyclic parthenogenesis, differentiated lineages and host specialization. *Molecular Ecology*, *6*(11), 1059–1073. <https://doi.org/10.1046/j.1365-294X.1997.00280.x>
- Vorburger, C., Lancaster, M., & Sunnucks, P. (2003). Environmentally related patterns of reproductive modes in the aphid *Myzus persicae* and the predominance of two 'superclones' in Victoria, Australia. *Molecular Ecology*, *12*(12), 3493–3504. <https://doi.org/10.1046/j.1365-294X.2003.01998.x>
- Wagner, S. M., Martinez, A. J., Ruan, Y.-M., Kim, K. L., Lenhart, P. A., Dehnel, A. C., ... White, J. A. (2015). Facultative endosymbionts mediate dietary breadth in a polyphagous herbivore. *Functional Ecology*, *29*, 1402–1410. <https://doi.org/10.1111/1365-2435.12459>
- Wahlund, S. (1928). Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas*, *11*(1), 65–106. <https://doi.org/10.1111/j.1601-5223.1928.tb02483.x>
- Walters, S. A., Kindhart, J. D., Hobbs, H. A., & Eastburn, D. M. (2003). Viruses associated with cucurbit production in southern Illinois. *HortScience*, *38*(1), 65–66.
- Wenger, J., Monica Ramstad, M. A., Mian, R., & Michel, A. P. (2014). The use of refuge in host plant resistance systems for the control of virulent biotype adaptation in the soybean aphid (Hemiptera: Aphididae). *Journal of Economic Entomology*, *107*(4), 1599–1609. <https://doi.org/10.1603/EC13545>
- White, J. A., McCord, J. S., Jackson, K. A., Dehnel, A. C., & Lenhart, P. A. (2017). Differential aphid toxicity to ladybeetles is not a function of host plant or facultative bacterial symbionts. *Functional Ecology*, *31*(2), 334–339. <https://doi.org/10.1111/1365-2435.12736>
- Wilson, A. C. C., Sunnucks, P., Blackman, R. L., & Hales, D. F. (2002). Microsatellite variation in cyclically parthenogenetic populations of *Myzus persicae* in Southeastern Australia. *Heredity*, *88*(4), 258. <https://doi.org/10.1038/sj.hdy.6800037>
- Wright, S. (1978). *Evolution and the genetics of populations: a treatise in four volumes: Vol. 4: variability within and among natural populations*. Chicago, IL: University of Chicago Press.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Angelella GM, Michel AP, Kaplan I. Using host-associated differentiation to track source population and dispersal distance among insect vectors of plant pathogens. *Evol Appl*. 2018;00:1–13. <https://doi.org/10.1111/eva.12733>