

Predicting the occurrence of host-associated differentiation in parasitic arthropods: a quantitative literature review

Kyle Harrison^{1,2*} , Aaron M. Tarone², Thomas DeWitt³ & Raul F. Medina² 

¹USDA-ARS, Agroecosystem Management Research, 1600 Campus Mall, Lincoln, NE 68503, USA, ²Department of Entomology, Texas A&M University, TAMU 2475, College Station, TX 77843, USA, and ³Department of Wildlife & Fisheries Sciences, Texas A&M University, 2258 TAMU, College Station, TX 77843, USA

Accepted: 13 September 2021

Key words: Arthropoda, ecological speciation, host-associated differentiation, HAD, literature review, parasite–host case study, reproductive isolation, isolating reproductive barrier, diversification, immigrant inviability, gallmaking, discriminant function analysis

Abstract

Parasite populations associated with different host species can encounter a variety of isolating reproductive barriers, leading to each population independently accumulating genome-wide genetic differences due to their host associations. This phenomenon is called host-associated differentiation (HAD) and has been proposed as an indicator of early diversification among parasitic arthropods. Although many parasite–host case study systems have been tested for the genetic signature of HAD (e.g., $F_{ST} \geq 0.15$ between sympatric, host-associated populations in the absence of allopatry), it is unknown which isolating reproductive barriers best predict the general occurrence of HAD. HAD development has been attributed to biological and ecological factors that either directly generate reproductive isolation between parasites living on different hosts, such as ‘immigrant inviability’ (i.e., lower fitness of immigrants in non-native environments), or that promote the accumulation of host-specific genetic adaptations, such as the gallmaking feeding mode. In fact, some of these factors are shared across multiple case studies, suggesting that the occurrence of HAD is generalizable and can be predicted based on the incidence of significant biological and ecological factors. By means of a discriminant function analysis (DFA), this research assessed 108 arthropod parasite–host case studies for ecological and biological factors significantly correlated with the occurrence of HAD and whether these factors could be used to distinguish the presence of HAD from its absence. The DFA demonstrated that case studies that developed HAD could be distinguished from case studies that did not develop HAD. The results of the DFA were corroborated by a ‘non-iterative partial least squares’ (NIPALS) discriminant model and a nominal logistic regression. Case studies with HAD could be robustly separated from case studies without HAD based on the incidence of these predictive factors: immigrant inviability, gallmaking, endophagy, recent range invasions of either hosts or parasites, differential host phenology, and differential parasite morphology. These results were used in an infinite random forest analysis to generate a hierarchy of conditional probabilities that separated HAD presence from absence. The results provide researchers with a tool for reliably predicting which untested parasite–host system would likely develop HAD. Immigrant inviability, gallmaking, and their combination were strongly correlated with the presence of HAD, which indicated parasite–host systems with these traits were highly likely to develop HAD. Contrary to expectation, endophagous feeding was negatively correlated with HAD presence, which indicated phytophagous endophagous feeders (excepting gallmakers) were highly unlikely to develop HAD. Furthermore, parasitoids were shown to be just as likely to develop HAD as not. Unfortunately, potentially significant predictive factors (e.g., allochrony) were excluded from analysis because too few case studies have been specifically

*Correspondence: Kyle Harrison, USDA-ARS, Agroecosystem Management Research, 1600 Campus Mall, Lincoln, NE 68503, USA. E-mail: kyle.harrison@usda.gov

tested for these factors. Furthermore, this analysis was biased by the lack of ‘negative’ publication results and the overrepresentation of research laboratories that primarily study HAD. Future research should accumulate novel HAD case studies that specifically test for allochrony, differential microbial associations, and morphological differentiation.

Introduction

Parasitic Arthropoda are one of the most diverse groups of multicellular life on Earth. Of the estimated 1 million described arthropod species, over 60% are involved in some form of parasitism with a plant or animal host or hosts (Price, 1980; Poulin, 1997; Dobson et al., 2008). In fact, the majority of Coleoptera, Hemiptera, and Hymenoptera are phytophagous parasites of plant hosts (Mitter et al., 1988; Wiegmann et al., 1993; Farrell, 1998). Furthermore, modern estimates of total arthropod species predict that the number of extant parasitic arthropod species are a magnitude greater than the number of currently described species, especially among obligate parasite clades (e.g., Braconidae, Ichneumonidae, and Acariformes) (Rodriguez et al., 2013). Therefore, the actual proportion of parasitic arthropods among all arthropods may be greater than the previously estimated 60%. The apparent overabundance of parasitic arthropods has prompted the scientific search for the origins of their biodiversity since Linnaeus (Fisher, 1988).

Forbes et al. (2009) succinctly described the relationship between parasitism and species diversity when they postulated ‘a major cause for biodiversity may be biodiversity itself’ (Ehrlich & Raven, 1964; Janz et al., 2006). Ecological speciation, or population divergence in response to distinct environments, has been proposed as the primary mechanism contributing to parasitic arthropod diversity (Futuyma & Moreno, 1988; Schluter, 2001; Via, 2001; Rundle & Nosil, 2005; Mallet et al., 2009; Matsubayashi et al., 2010). This is because parasite populations can experience divergent selection pressures when exposed to different host species or biologically distinct host lineages. The physiological and chemical differences between these hosts represent divergent ‘environments’, creating or promoting reproductive isolation between host-associated parasite populations (Price, 1980; Funk, 1998). In the presence of an isolating reproductive barrier, host-mediated evolution will drive the independent accumulation of host-specific adaptations and genome-wide genetic differences through selection and genetic drift (Bush, 1975; Jaenike, 1990; Berlocher & Feder, 2002). If host-associated, disruptive selection acts on traits linked to performance or reproduction, then reinforcement for host association can result in relatively quick genetic differentiation (i.e.,

developing in a few centuries; Bush, 1969; Butlin, 1987; Howard, 1993; Groman & Pellmyr, 2000; Althoff et al., 2001; Servedio, 2001; Servedio & Noor, 2003; Rundle & Nosil, 2005). This process, called host-associated differentiation (HAD), can either cause or be caused by reproductive isolation (Forbes et al., 2017). Therefore, HAD constitutes an early stage of ecological speciation, creating opportunities for host race evolution and total reproductive incompatibility (Bush, 1994; Pritchard et al., 2000; Via & Hawthorne, 2002; Linn et al., 2004; Futuyma, 2008; Peccoud et al., 2009; Michel et al., 2010; Powell et al., 2013). For this reason, HAD has been proposed as an indicator of early diversification among parasitic arthropods (Funk, 1998, 2010; Stireman et al., 2005). To test for HAD, researchers have compared genome-wide genetic markers between sympatric parasite populations sampled from different host species or host lineages. If a tested population demonstrated significant, host-associated genome-wide differentiation (e.g., $F_{ST} > 0.15$ or a cluster analysis depicting multiple genetic groups), then the genetic signature of HAD was ‘present’. As allopatric differentiation is a more likely alternative hypothesis to HAD, case studies that tested for HAD simultaneously ruled out allopatry as the origin for differentiation by sampling from multiple geographic localities for the same parasites and hosts.

It is unknown how common HAD is among parasitic arthropods and which parasite–host systems are prone to the development of HAD. However, specific biological or ecological factors may account for the occurrence of HAD in specific case study systems. In these case studies, the genetic signature of HAD was tested between host-associated parasite populations that possessed biological or ecological factors that promote either reproductive isolation or the independent accumulation of genetic differences. This research assessed the available case studies for the commonly shared biological or ecological factors (listed below) and their correlation with HAD.

Allochronic isolation

Parasites depend on their hosts to survive, meaning they must be in the same physical space and time as their hosts’ resources (e.g., fruits and seeds coming into season). This obligate dependency imposes selection pressures on parasites to match their hosts’ phenology.

If host-associated parasite populations adapt to different host species or lineages with divergent phenologies, then parasites can experience disruptive selection on traits relevant to seasonal timing. This can generate temporal, or ‘allochronic’, pre-mating reproductive isolation (Alexander & Bigelow, 1960; Groman & Pellmyr, 2000; Rajimann & Menken, 2000; Cooley et al., 2001; Abbot & Withgott, 2004; Fudickar et al., 2016). For example, differences in fruit availability between apple and hawthorn trees impose differential selection pressures on adult emergence and eclosion times in apple maggot flies (*Rhagoletis pomonella* Walsh), generating host-associated populations with genome-wide divergence (Bush, 1969; Feder et al., 1988; Feder & Filchak, 1999; Rajimann & Menken, 2000). However, allochronic isolation has only been directly demonstrated in a handful of parasite species. Therefore, this research evaluated case studies for distinct host phenologies and their correlation with HAD.

Feeding mode

Endophagous feeders are thought to experience stronger disruptive selection compared to exophagous feeders or ectoparasites (Cornell et al., 1998; Stiling & Rossi, 1998; Stireman et al., 2005). For example, endophagous feeders are subjected to prolonged and intense exposure to plant chemical defenses and have fewer opportunities to escape host defenses compared to exophagous feeders (Waring et al., 1990; Abbot, 2001; Abrahamson et al., 2001). This is especially true for the specialized biology of parasitoid or gallmaking insects, which have evolved gene-by-gene mechanisms that allow them to overcome host defenses and hijack their hosts’ physiology to create suitable developmental environments for their offspring (Abrahamson et al., 1994, 2001; Brown et al., 1995; Stireman et al., 2008; Dickey & Medina, 2012). The selection pressures inherent to gallmaking promote pre- and post-mating reproductive isolation between gallmakers associated with phenologically distinct hosts. This is because migrant gallmakers and their hybrid offspring should be unlikely to possess the ‘lock-and-key’ mechanisms required for gall formation (Craig et al., 1994; Itami et al., 1998). Similarly, parasitoids initiate behavioral and physiological changes in their arthropod hosts to overcome their immunological response to internal invasion (Strand & Pech, 1995). Like gallmakers, these conditions may promote the development of pre- and post-mating isolation between parasitoids associated with immunologically distinct hosts (Faeth et al., 1981; Schwarz et al., 2005; Loxdale, 2009, 2010). This research evaluated the trophic positions and feeding niches (e.g., herbivore, gallmaker, or parasitoid) as well as their concealment (i.e., endophagous vs. exophagous).

Host fidelity

Host-associated parasitic arthropods can experience host fidelity, or the tendency of insects to remain on and return to their developmental environment (Bush, 1969; Feder et al., 1994; Wood et al., 1999; Craig et al., 2001; Stelinski & Liburd, 2005). For example, host fidelity has been demonstrated in host-associated populations of pea aphid populations (*Acyrtosiphon pisum* Harris): pea aphids from clover prefer to feed on clover even when alfalfa was abundantly available (Via, 1999; Via et al., 2000). Similarly, the apple maggot fly consists of two host-associated populations, one preferring to oviposit and mate on hawthorn trees and another population preferring apple trees (Feder et al., 1994; Dambroski et al., 2005). Importantly, the host fidelity of apple maggot flies was mediated by host-specific volatiles that drew them to their natal host (Frey et al., 1998). This reliance on host volatile cues demonstrated that experience can pre-condition parasites to prefer the host environment in which they developed, selecting for traits that intrinsically link a population to their host (Cunningham et al., 2001). For example, cuticular hydrocarbons (CHC) are adaptive traits of most terrestrial arthropods that are directly influenced by their host environment (Otte et al., 2018). Despite this plasticity, differences in CHC can determine host and mate recognition, potentially creating pre-mating reproductive isolation between host-associated populations with different CHC profiles (Etges et al., 2009). When parasite populations use different host species or lineages, selection for natal host preference can result in divergent selection in response to differential host cues (e.g., plant volatile or oviposition substrate preference) (Abrahamson et al., 1989; McCall et al., 1993; Frey et al., 1998; Pureswaran et al., 2004). Furthermore, selection for host preference can be exacerbated in parasites that exclusively choose mates on their hosts (Futuyma & Moreno, 1988; Jaenike, 1990; Egan & Funk, 2006). In this research, the presence of host fidelity was evaluated in different parasite–host systems based on previous tests for host-associated volatile preference, oviposition preference, and host-mediated mate-finding behavior.

Phylogeny

The evolution of parasitism is biased within phylogenetic clades. The 10 largest parasitic clades contain 90% of all described parasite species. Specific clades, especially flies and mites, have multiple evolutionary origins for parasitism (Weinstein & Kuris, 2016). These results suggest that some parasitic arthropods are better suited for specializing in host resources compared to others. Therefore, some parasitic arthropods might be better suited to host specialization and the development of HAD. This research assessed the correlation between parasite clade (i.e., order

and family) and HAD occurrence. In addition, some plants are less susceptible to herbivory based on their anatomy and life-history parameters (Carmona et al., 2011). Since these traits may be linked to phylogeny, some plant clades may be more vulnerable to arthropod herbivory and host-resource specialization. This study also tested the relationship between host clade (including non-plants) and HAD occurrence.

Selection against migrants

Pre-mating isolation can evolve when migrants move from a natal, locally adapted host niche to a novel host niche occupied by locally adapted parasite populations. As migrants have lower fitness than locally adapted populations, host-to-host migration is unlikely to result in successful colonization or hybridization (Funk, 1998; Via et al., 2000; Nosil et al., 2005; Dickey & Medina, 2011a). Furthermore, hybrids of migrants and locally adapted parasites will likely retain the maladapted traits of their migrant parents, meaning post-mating isolation can also evolve between immigrant and natal parasites (Wade & Johnson, 1994; Wu & Palopoli, 1994; Naisbit et al., 2001; Schluter, 2001; Rundle, 2002). The lowered fitness of immigrants in novel environments, or ‘immigrant inviability’, is often due to antagonistically pleiotropic adaptations inherent to host-adapted populations. Immigrant inviability can also work in concert with other biological and ecological factors (e.g., habitat preference and allochrony), potentially reinforcing host associations (Matsubayashi et al., 2010). Although it would seem apparent that HAD and immigrant inviability are intrinsically linked, case study examples exist in which the genetic signature of HAD is present in the absence of immigrant inviability (e.g., specialist *Myzus persicae* Sulzer populations evolving from generalists; Margaritopoulos et al., 2007) and vice versa (e.g., *Aphelinus albipodus* Hayat & Fatima associated with soybean and bird cherry-oat aphids; Yokomi & Tang, 1995). Therefore, the correlation between immigrant inviability and HAD occurrence was evaluated in this study.

Morphological/sexual selection

Divergent mating cues or habitat preference can reduce the probability that individual parasites can successfully reproduce with conspecifics occurring on different hosts. For example, host-mediated mate choice has been documented between populations of goldenrod gall flies (*Eurosta solidaginis* Fitch) associated with different goldenrod species (Craig et al., 1993, 1997). Furthermore, sexual isolation has been observed between populations of walking sticks (*Timema cristinae* Vickery) associated with two host-plant species (*Ceanothus spinosus* Nutt and

Adenostoma fasciculatum Hook & Arn) (Nosil et al., 2002). In these cases, morphological differences arose between host-associated populations in response to divergent ecological selection pressures (e.g., body color and morph), yet some of these differences were not linked to performance, suggesting that the relationship between sexual isolation and morphological differentiation can be linked (Roy et al., 2013). Sexual isolation can arise due to distinct morphological traits linked to host plant use or vice versa, but the origins of sexual isolation are usually left unexplored due to the experimental limitations involved in proving causation (Jiggins et al., 2000; Dobson et al., 2002; Fordyce & Nice, 2003; Nosil, 2007; Egan et al., 2008; Fordyce, 2010). Therefore, specific kinds of morphological differentiation (e.g., distinct leg, penis, or wing cell shape) were chosen as potential indicators of sexual selection among case study systems (Katayama et al., 2014; Kelly, 2014). Other kinds of morphological differences (e.g., distinct mouthpart or ovipositor lengths) were considered host-associated adaptations that evolved in response to distinct host environments (Pappers et al., 2002). This research evaluated both sexual and non-sexual morphological differentiation and their correlations with HAD.

Recent invasions

Parasites may use a newly encountered host species or lineage if the novel host is biologically similar to their natal host (Berlocher & Feder, 2002; Drès & Mallet, 2002). The closer the novel host’s characteristics to their natal host’s, the higher the probability that the parasite will accept and survive on the novel host. However, when a parasite species encounters opportunities to expand its range (e.g., an invasive insect pest arrives on a new continent), host shifts can take place in spite of the fitness costs associated with parasitizing a maladaptive host (Murphy, 2004; Schwarz et al., 2005; Agosta, 2006; Craig et al., 2011). As host–parasite interactions select for host-specific adaptations in parasites (see ‘Habitat fidelity’ above and ‘Selection against migrants’ below), any introduction of a novel host-parasite interaction potentially promotes the development of HAD (Kawata, 2002). Furthermore, the absence of competition in vacant niches or the overabundance of a niche (e.g., monocultures) may accelerate the accumulation of these adaptations and potentially offset the fitness costs associated with occurring on a maladaptive host (Barman et al., 2012). Geological time also influences HAD occurrence. Although HAD has been shown to occur in older parasite–host systems, it is currently believed that these systems were more likely to transition from host-associated populations into separate species (Rabosky, 2009), meaning invasions that occurred hundreds of thousands to millions of years ago could have

developed HAD. Likewise, invasions that occurred within the last few years may yet develop HAD. This study evaluated case studies for recently observed geographic range expansions by either parasites or hosts. Furthermore, this study evaluated systems for anthropogenic management of hosts or parasites.

Reproductive mode

Parthenogenetic and cyclically parthenogenetic parasites (e.g., aphids, thrips, and mites) evolve reproductive isolation more easily than sexual insects because asexual reproduction facilitates the accumulation of host-specific adaptations without trait loss due to recombination. Thus, adaptive alleles can rapidly increase in frequency and become fixed within asexual populations, promoting genome-wide genetic differentiation in host-associated populations (Hartl, 1972; Lynch & Gabriel, 1983; Lynch, 1984; King, 1993; Neiman & Linksvayer, 2006; Loxdale, 2008; Dickey & Medina, 2010). In addition, asexually reproducing arthropods possess short generation times, decreasing the geological time required for host-specific adaptations to become apparent to researchers (Martin & Palumbi, 1993; Li et al., 1996). Multivoltine parasites with short generation times would experience several more rounds of host-mediated selection pressure per season compared to univoltine parasites. For example, the rapid radiation of Aphidoidea has been attributed to low recombination rates and quick generation turnover (see table 3 of Dickey & Medina, 2010). This study evaluated parasites for their described reproductive modes as well as its annual generation turnover.

The biological and ecological factors described above have been shown to be directly involved in HAD development in specific parasite–host systems (e.g., allochry in apple maggot flies, habitat fidelity in pea aphids, or immigrant inviability in *Timema* walking sticks). However, it is unknown which of these biological and ecological factors best predict HAD occurrence. If ecological and biological factors predict the occurrence of HAD, then the incidence of those factors should be highly correlated with the presence or absence of HAD. This hypothesis is falsified because several examples exist in which HAD is absent in parasite–host systems even though they involve multiple factors thought to promote or cause the development of HAD (Jaenike & Selander, 1980; Clements et al., 2000; Baer et al., 2004; Gómez-Díaz et al., 2007; Althoff, 2008; Lozier et al., 2009; Dickey & Medina, 2010, 2011b; Kohlen et al., 2011; Simonato et al., 2012). For example, immigrant inviability can develop independently of HAD and vice versa (see ‘Selection against migrants’ above). As another example, host-specific volatile preference and allochry have been shown to cause reproductive isolation and were both

present in the cranberry fruitworm, *Acrobasis vaccinii* Riley, yet HAD has not developed between blueberry- and cranberry-associated populations of this endophagous insect (Medina et al., 2014). In contrast, HAD has developed in the blueberry gall midge, *Dasineura oxycoccana* Johnson, associated with the same hosts, suggesting that another factor (e.g., gallmaking) may account for the development of HAD in this ecosystem (Cook et al., 2011). Similarly, six Aphidoidea species have populations associated with pecan and water hickory trees throughout central Texas and all six species were expected to exhibit some degree of HAD due to the hosts’ differential phenologies and the parasites’ parthenogenesis. In addition, three of the species were gallmakers, a factor believed to promote HAD. However, only three of these parasite species exhibited the genetic signature of HAD and one of the species that did not develop HAD was a gallmaker (*Phylloxera texana* Stoetzel) (Medina et al., 2017). These apparent inconsistencies in the theory of HAD development necessitate a quantitative comparison of potentially predictive factors and their relationship with HAD.

Predicting HAD occurrence has implications for evolutionary biology, conservation biology, agriculture, and disease management (Bernal & Medina, 2018). First, as discussed above, HAD may serve as an indicator of ongoing diversification among parasitic arthropods. Second, HAD can create rare but ecologically significant genotypes living in sympatry with morphologically similar conspecifics (e.g., parasitic wasps, each specializing on different host-associated aphid populations). Ignoring this phenomenon could result in inadvertent biodiversity loss. Third, host-associated populations of pestiferous insects can differ in host-associated adaptations relevant to pest control (Burban et al., 1992; Medina, 2012). Using the above example, parasitic wasps specializing in a specific, host-associated aphid population would provide better biological control than a closely related wasp specializing in another aphid population. Lastly, HAD has important implications for disease ecology. Specifically, parasites would be more competent vectors of pathogens without host-associated populations (Hastie et al., 1999; McCoy et al., 2013; Esteve-Gassent et al., 2016). To predict which parasite–host systems are prone to the development of HAD, a quantitative literature review was used to compare the occurrence of HAD against the incidence of factors hypothesized to generate HAD in specific case study systems.

Materials and methods

Data sources and searches

This study was conducted to identify and quantify the potential sources of reproductive isolation resulting in

genetic differentiation among parasite–host systems and test their correlation with HAD occurrence. This study analyzed a data matrix comprised of biological and ecological factors shared among parasite–host case studies tested for HAD. These case studies included descriptions of the parasite and host life histories as well as descriptions of reproductive isolation between host-associated populations. Specifically, all the factors described in the ‘Introduction’ section were considered for each parasite–host case study system: allochronic isolation, feeding mode, host fidelity, phylogeny, selection against migrants, morphological/sexual selection, recent invasions, and reproductive mode. A single investigator (K Harrison) searched for abstracts that described case studies of parasite–host systems tested for genetic population structuring. These searches were conducted using Google Scholar, Web of Science, and JSTOR internet databases. Key words and phrases used in the searches are provided in Table S1. Case studies were included in the database if: (1) genome-wide genetic markers (e.g., microsatellites, RFLP’s, SNP’s) were used to test for genetic structuring among multiple parasite populations, (2) the system included at least one sympatric location for host-associated parasite populations, and (3) the experimental design specifically characterized genetic structuring by host species or lineage associations and collection locations, ruling out allopatry as the origin of genetic differentiation. The literature cited within each case study was also searched for examples of previously characterized parasite–host systems that preceded the case study. Case studies included in the database were placed into one of two categories: case studies that unambiguously demonstrated genetic population structuring by host species or lineage (HAD present) and case studies that unambiguously demonstrated population genetic structuring did not exist, or that demonstrated that any present population genetic structuring was not due to host associations (HAD absent). Assignment of case studies to a HAD category was codified with a 1 or 0, creating an independent variable data matrix comprised of 108 case studies: 72 HAD present and 36 HAD absent.

In addition to assigning case studies to HAD category, the scientific literature cited was searched for potentially predictive ecological and biological factors within each parasite–host system. Descriptions of predictive factors were searched for all parasites and hosts involved in the case study (Table S1). In total, 129 journal articles were used to describe the biology and ecology of each parasite and host involved in the evaluated case studies. Parasite–host descriptions were used to construct a data matrix of potentially predictive factors based on the following criteria: (1) the authors described or tested for the presence of a predictive factor that could generate an isolating

reproductive barrier between parasite populations, or (2) provide citations of authors who tested these factors in previous research. A list of the parasite–host case study systems and their descriptions is provided in columns 2–20 of Table S2. The most common factors (i.e., represented by >10 case studies) were included in downstream analysis (Table 1). Like HAD presence and absence, predictive factors were codified with a 1 (present) or 0 (absent), creating a data matrix comprised of 108 case studies and the most common predictive factors (see Table 1), or 2160 (= 108 × 20) total cells. Factor categorization was handled independently, meaning some factor combinations (e.g., ‘endophagous’) could be linked. Several potentially predictive factors have yet to be tested in many parasite–host systems, meaning only 1763 out of 2160 (81.6%) matrix cells were available for analysis. To avoid the exclusion of entire case studies or entire predictive factors from analysis, a multiple imputation method was used to simulate missing cell values where applicable (Horton & Lipsitz, 2001; Schmitt et al., 2015). The singular value decomposition method of multiple imputation was chosen because it works well with large datasets (Prasad et al., 2006) and

Table 1 List of biological and ecological factors commonly shared by arthropod parasite–host case studies. These factors were represented by ≥10 case studies and were included in the discriminant function analysis (DFA). Uncommon factors (e.g., Dipteran parasites, ectoparasites, frugivores, detritivores, long generation times, etc.) were implicitly included in the DFA as factors by contrast

Biological or ecological factor	No. studies
Asterales host	18
Asexual reproduction	35
Distinct host phenologies	72
Endophagous	52
Gallmaker	16
Hemipteran parasite	32
Host used to find mate	93
Hymenopteran parasite	19
Immigrant inviability	52
Lepidopteran parasite	20
Oviposition preference	42
Parasite feeds on leaves	17
Parasite feeds on vascular tissues	28
Parasitoid	16
Phytophagous	77
Recent invasions	48
Rosales host	12
Short generation times (≥4 generations/year)	65
Anthropogenic management	57
Volatile preference	34

could be performed in JMP v.12 (SAS Institute, Cary, NC, USA). Unfortunately, several potentially significant predictive factors (allochryony, infection by differential microbial species, morphological differentiation, and differential oviposition preference) were comprised of >2/3 missing values, meaning these factors could not be appropriately imputed without major error. Therefore, these factors were excluded from downstream analysis.

Statistical analysis

Statistical analysis was performed using JMP v.12. Data were analyzed using three complementary procedures to define basic relationships between HAD occurrence and potentially predictive factors, check for correspondence across analytical methods, and test for variable interactions. A discriminant function analysis (DFA) tested the correlation of each factor with HAD as well as each factor's relative contribution to the separation of 'HAD present' systems from 'HAD absent'. The DFA calculated the percentage of parasite–host system cases correctly classified, which provided an intuitive metric for model success. The a priori probability of group membership (i.e., 'HAD present' or 'HAD absent') was assumed to be proportional to the number of case studies assigned to each group in the initial pool of case studies, meaning group membership was biased toward HAD presence. As the classification variable has two levels, case studies could only be plotted against canonical axis 1, meaning canonical weights for each covariate (and their associated biplot rays) relate to canonical axis 1 only. The model also assumed homogeneous, symmetrical variances (i.e., the model was linear and non-quadratic).

To check the correspondence of the DFA, a nominal logistic regression was also used to identify predictive factors significantly correlated with HAD occurrence. Model validities of the DFA and nominal logistic regressions were independently tested with a 'non-iterative partial least squares' (NIPALS) discriminant model using the 'leave-one-out cross-validation' (LOOCV) procedure, which is a type of cross-validation approach in which individual observations are validated with an algorithm trained by all other observations. In this way, a principal component analysis (PCA) could be performed on this dataset, which has missing values. Variable Importance in Projection (VIP) scores estimated the importance of each variable in the projection used in the NIPALS.

Lastly, complex interactions between predictive factors and HAD occurrence were tested using regression partitions, or an infinite random forest decision tree. The decision tree with the best logworth support was plotted as a hierarchy of predictive factors based on the factor-based, conditional probabilities that best separate HAD presence

from absence (Cutler et al., 2007; Welling & Kurihara, 2009; Criminisi et al., 2012). The K-means clustering method was used to determine the number of statistically distinct data clusters (K) among all HAD case studies based on the same predictive factors used above (Groman & Pellmyr, 2000).

Results

Canonical scores obtained from the DFA demonstrated that 'HAD present' parasite–host systems could be distinctly separated from 'HAD absent' systems based on a few, significant predictive factors (Figure 1). The DFA showed that nine out of 20 predictive factors made meaningful, whole-model contributions to HAD case study classification and were at least partially correlated with the X-axis of Figure 1: asexual reproduction, distinct host phenologies, endophagy, gallmaking, immigrant inviability, recent invasions, host used to find mate(s), short generations times (four or more generations per year), and anthropogenic management (Wilks' $\Lambda = 0.514$, $d.f._{num} = 35$, $d.f._{denom} = 71$, $P < 0.0001$; $R^2 = 0.3125$, $d.f. = 20$, $P < 0.0001$). The DFA also showed a significant interaction effect between exophagous feeding and the incidence of distinct host phenologies, which was correlated with 'HAD present' case studies. Based on these factors, case study systems could be robustly sorted into their respective 'HAD present' (88% success rate) or 'HAD absent' (78% success rate) classification. The nominal logistic regression corroborated the results of the DFA, showing four shared predictive factors related to HAD occurrence: immigrant inviability and endophagy ($\alpha = 0.05$) as well as gallmaking and recent invasions ($\alpha = 0.10$; Table 2). Furthermore, the nominal logistic regression had a significant whole-model fit ($\chi^2 = 46.21$, $d.f._{num} = 35$, $d.f._{denom} = 71$, $P < 0.0001$), meaning these four factors could account for most of the variance in HAD occurrence. These same four factors each possessed the largest biplot rays in Figure 1, meaning they contributed most to the canonical scores used to separate HAD presence from absence. Consistent with these observations, the NIPALS analysis independently recovered the same significant predictive factors as the DFA and nominal logistic regression; however, the large number of variables included in the analysis lead to the loss of ordination in the NIPALS. These factors cumulatively accounted for 60.9% of variance along the x-axis and 38.5% along the y-axis which, together, separate HAD 'presence' from 'absence' (Table 3). Importantly, immigrant inviability was very highly correlated ($r = 0.996$) with the presence of HAD in case studies and accounted for 31.2% of the y-axis variance on its own. VIP scores showed that 'immigrant inviability' is a sufficiently high-score variable to achieve a score

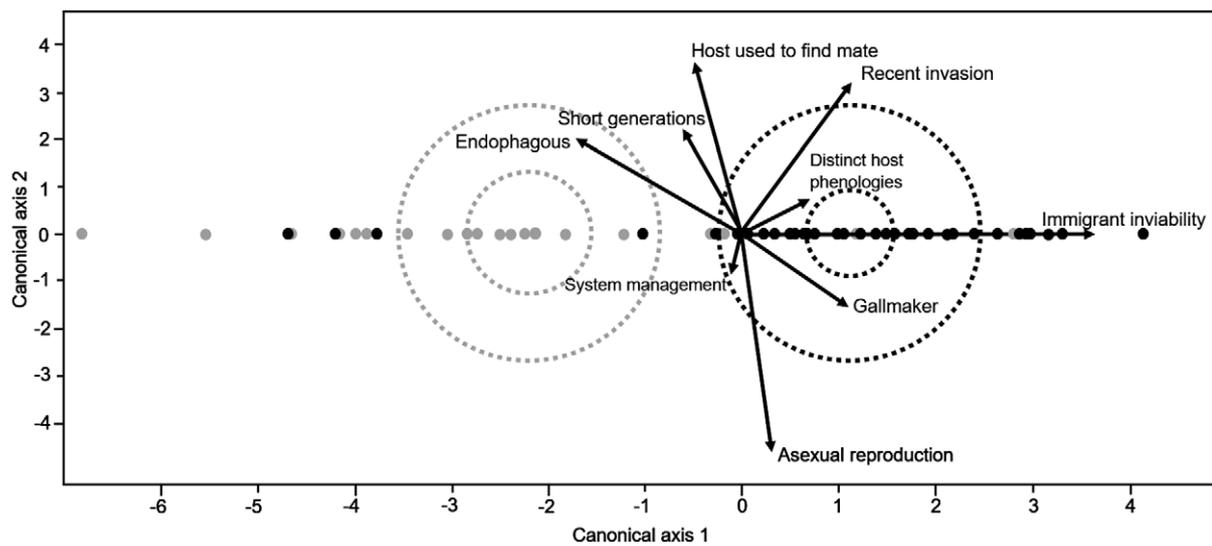


Figure 1 Discriminant scores for parasite–host case studies based on significant predictive factors. ‘Host-associated differentiation (HAD) present’ case studies are represented by black dots and ‘HAD absent’ cases by gray dots. Dotted lines represent the 95% multivariate centroid confidence ellipsis (inner circle) and the 50% normal ellipsis (outer circle). Biplot rays represent the relative contribution of a given predictive factor to canonical axis 1, which separated HAD presence from absence. The Y-axis was not involved in the separation of ‘HAD present’ from ‘HAD absent’ case studies; vertical positions of the biplot rays were included for reading purposes.

Table 2 Nominal logistic regression of predictive factors correlated with host-associated differentiation (HAD) occurrence. d.f._{num} and d.f._{denom} stand for degrees of freedom for the numerator and denominator of the χ^2 ratio, respectively

Model effect	SE	d.f. _{num}	d.f. _{denom}	χ^2	Prob> χ^2	Correlation coefficient
Whole model	1.0026	35	71	46.21	<0.0001	1.0
Asexual reproduction	0.6329	1	107	0.35	0.55	−0.374
Distinct host phenologies	0.5645	1	107	0.60	0.44	−0.264
Endophagy	0.6707	1	107	6.47	0.011	−0.332
Gallmakers	0.9374	1	107	2.84	0.092	0.447
Immigrant inviability	0.8514	1	107	25.49	<0.0001	0.946
Recent invasions	0.6536	1	107	2.79	0.095	0.308
Host used to find mate	0.7740	1	107	0.75	0.39	−0.329
Short generation times ¹	0.6145	1	107	1.15	0.28	−0.322
Anthropogenic management	0.6745	1	107	0.10	0.76	−0.374

¹More than four generations per year.

threshold of >0.8 on its own, whereas ‘endophagy’, ‘gall-makers’, ‘recent invasion’, and ‘distinct host phenologies’ must be paired with at least one other variable. The LOOCV likewise supported these results, showing the same four factors could account for most of the model’s predictive power ($R^2 = 0.3698$, d.f. = 20, $P < 0.0001$). Lastly, the infinite random forest algorithm generated a hierarchy of factor-based conditional probabilities that robustly separated HAD presence from absence ($R^2 = 0.431$, d.f. = 20, $P < 0.0001$). Unfortunately, the overrepresentation of ‘HAD

present’ case studies available likely meant that correlations between predictive factors and the presence of HAD were unintentionally given more weight than correlations between factors and HAD absence.

Based on these results, a hierarchical schematic was used to show the ecological and biological conditions most likely to involve HAD as well as the conditions least likely to involve HAD (Figure 2). Logworth support for branch splits was significant ($\alpha = 0.05$) for all branches [1.38–11.19 where $\logworth = -\log_{10}(P\text{-value})$]. Importantly,

Table 3 Non-iterative partial least squares (NIPALS) table for all parasite–host case studies (n = 108). The X and Y variables represent the directions of maximum variation through the dataset. Variable importance in partial least squares projection (VIP) scores showed that ‘immigrant inviability’ is a sufficiently high-score variable to achieve VIP threshold >0.8 on its own, whereas ‘endophagy’, ‘gallmakers’, ‘recent invasion’, and ‘distinct host phenologies’ must be paired with at least one other variable

No. factors	Factors included	% of total variance for cumulative		No. VIP >0.8
		X	Y	
1	Immigrant inviability	15.71	31.24	1
2	Endophagy	30.04	37.00	2
3	Gallmakers	46.01	38.19	2
4	Recent invasion	58.18	38.45	2
5	Distinct host phenologies	60.97	38.50	2

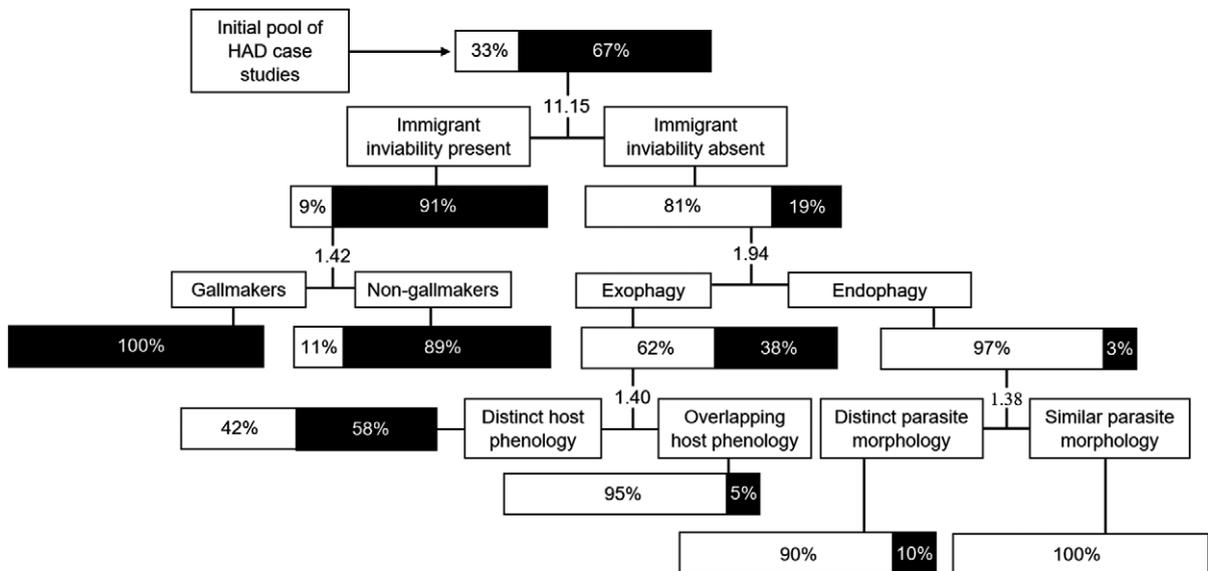


Figure 2 A hierarchy of conditional probabilities based on the incidence of biological and ecological factors separating ‘host-associated differentiation (HAD) present’ parasite–host from ‘HAD absent’ case studies. Each level of the dichotomous hierarchy represents a predictive factor (i.e., immigrant inviability, gallmaking, endophagy, host phenology, and recent invasions). Branches split the presence of a given factor from its absence. Values superimposed over each branch represent the logworth support for a given split in the hierarchy. The black and white bars represent the division of case studies (%) expected to be ‘HAD present’ (black) or ‘HAD absent’ (white).

the factors shown to be significantly correlated with HAD occurrence were all represented in the hierarchy. This hierarchy can be used to predict the occurrence of HAD in untested case studies based on the incidence of significant factors. Immigrant inviability was the highest-level predictor because it was the largest contributor to separating ‘HAD present’ systems and ‘HAD absent’: almost all case studies involving immigrant inviability were ‘HAD present’ (91%), whereas case studies without immigrant inviability were ‘HAD absent’ (81%). Among the case studies

involving immigrant inviability, gallmaking was the next level predictor of HAD occurrence: 100% of case studies involving both immigrant inviability and gallmaking were ‘HAD present’ and almost all non-gallmakers involving immigrant inviability were likewise ‘HAD present’ (89%). In case studies without immigrant inviability, feeding mode (i.e., endophagy vs. exophagy) was the next best predictor of HAD occurrence. Contrary to expectation, ‘exophagous feeders without immigrant inviability’ were more likely to be ‘HAD present’ (38%) than ‘endophagous

feeders without immigrant inviability' (3%). Therefore, the next level under 'endophagous feeders without immigrant inviability' (i.e., recent invasions) could only weakly separate HAD presence from absence. In contrast, 'exophagous feeders without immigrant inviability' were more likely to be 'HAD present' when distinct host phenologies were also present (58%) compared to case studies with overlapping host phenologies (5%).

Notably, several predictive factors expected to be highly correlated with the presence of HAD were not. Specifically, oviposition preference, parasitoid feeding mode, and volatile preference made no meaningful model contributions to separating HAD occurrence from its absence, suggesting that these factors were not correlated with HAD presence or absence. Similarly, the phylogenetic identities of hosts (e.g., Asterales) and parasites (e.g., Hymenoptera) made no significant model contributions, suggesting that none of the phylogenetic clades tested were correlated with HAD presence or absence.

Discussion

The initial hypothesis was supported: at least four biological and ecological factors were highly correlated with the presence or absence of HAD and could be used as predictors of HAD occurrence. Analysis results agreed that immigrant inviability, gallmaking, endophagy, and recent invasions were correlated with the occurrence of HAD among the tested parasite–host case studies. The DFA also revealed that specific combinations of predictive factors (e.g., exophagous feeders experiencing differential host phenologies) were also significantly correlated with HAD occurrence. The infinite random forest model successfully separated 'HAD present' case studies from 'HAD absent' based on the incidence of significant predictive factors. This model was used to construct a hierarchy of conditional probabilities describing the likelihood that a given parasite–host system would be 'HAD present' or 'HAD absent'.

Four findings from this study are consistent with previously published research. First, immigrant inviability was highly correlated with the presence of HAD and had a comparatively large contribution to the DFA and other models separating 'HAD presence' from 'HAD absence'. Immigrant inviability and HAD development may not be intrinsically linked, but it is apparent that immigrant inviability is the best predictor of HAD occurrence. Immigrant inviability was also the first major predictor for the absence of HAD in the DFA. In addition, immigrant inviability was one of the most valuable factors contributing to the infinite random forest model and, thus, the hierarchy of conditional probabilities. For example, immigrant

inviability and gallmaking were independently shown to be highly correlated with the presence of HAD but also appeared to have a compounding effect on HAD development (Dickey & Medina, 2011a). These results support the hypothesis that selection against migrants living on maladaptive hosts is a significant contributor to reproductive isolation in parasites (Funk, 1998; Via et al., 2000; Nosil et al., 2005; Nosil, 2007; Dickey & Medina, 2011a). Furthermore, these results suggest that selection is the major driving force behind HAD development.

Second, gallmaking was highly correlated with the presence of HAD and made significant whole-model contributions to the separation of HAD presence from absence. These results support the hypothesis that gallmakers experience relatively strong, host-associated selection pressures that generate reproductive isolation (Abrahamson et al., 1994, 2001; Craig et al., 1994; Brown et al., 1995; Itami et al., 1998; Stireman et al., 2008; Dickey & Medina, 2012). This study, however, also showed that gallmaking is a special case among phytophagous, endophagous feeders. Contrary to expectation, other forms of endophagy (i.e., fruit-feeders, leaf-miners, and stem-feeders) were significantly associated with the absence of HAD whereas most exophagous feeders included in this study developed HAD. The negative relationship between plant endophagy and HAD occurrence suggested that, with the exception of gallmakers, most endophagous arthropod herbivores would be unlikely to develop HAD and that exophagous, leaf-feeding arthropods are more likely to develop HAD than previously thought (Stireman et al., 2006; Dickey & Medina, 2010, 2012). Notably, parasitoid feeding mode made no significant model contributions to the separation of HAD presence from absence, which suggested that parasitoid endophagous feeders were just as likely to develop HAD as not.

Third, distinct host phenologies were correlated with the presence of HAD. This result, in addition to the results concerning immigrant inviability and gallmaking, supported the hypotheses that allochrony and habitat isolation promote disruptive selection and the evolution of host fidelity (Futuyma & Moreno, 1988; Jaenike, 1990; Feder et al., 1994; Dambroski et al., 2005; Egan & Funk, 2006). Distinct host phenologies were shown to be significant in the DFA and the NIPALS analysis, but not the nominal logistic regression. This result suggested that distinct host phenologies were only partially correlated with HAD occurrence. Interestingly, phenological differences were correlated with the presence of HAD whereas differences in host-specific volatiles were not significantly correlated with HAD presence or absence. However, this result is not surprising because differential responses to host phenologies have been shown to impose isolating

reproductive barriers, independent of differential parasite responses to different olfactory host cues (Onstad et al., 1986; Stamp & Bowers, 1990; Hunter, 1992; Feder et al., 1993; Abrahamson et al., 1994; Yukawa, 2000; Mopper, 2005; Brooks & Hoberg, 2007).

Fourth, the incidence of recent range invasions, by either parasites or hosts, was correlated with the presence of HAD. This result suggested HAD was more likely to develop when parasites were exposed to a novel niche or given an opportunity to expand their host range. Therefore, parasite range invasions promote HAD development despite the fitness costs associated with living on a maladaptive host. The survival of these early immigrants may be due to the fitness benefits of parasitizing a highly abundant host with few competitors or predators (Sakai et al., 2001). Based on this observation, the prevalence of HAD will likely increase among newly established parasite–host systems (i.e., invasive species expanding their territories). This contrasts with older systems in which parasites and hosts were both native, which were not correlated with either HAD presence or absence. However, it is possible that older systems involving distinct species living on similar hosts may have, at one point, been genetically distinct, host-associated populations. Furthermore, systems with biological and ecological factors favoring the development of HAD, yet lacking the genetic signature of HAD (e.g., *A. vaccinii* living on cranberry and blueberry), may still develop into genetically distinct, host-associated populations over time. Importantly, almost every one of the recent invasions evaluated in this study was due to anthropogenic interference. This observation has implications for the expected increase in invasive species range expansions following global climate changes (Gandon & Michalakis, 2002).

Contrary to expectation, some ecological and biological factors made no whole-model contributions to the separation of HAD case studies despite being hypothesized as major contributors to HAD in their own systems: oviposition preference, parasitoid feeding mode, parthenogenetic reproductive mode, and phylogeny were not correlated with HAD. This result suggested that Hemiptera, Hymenoptera, Lepidoptera, parasitoids, and parthenogens were all just as likely to develop HAD as not. Therefore, HAD could not be predicted by low recombination rates or the rapid accumulation of alleles in host-associated populations through short generation turnover. Interestingly, the likelihood of HAD developing in a parasite–host system is independent of evolutionary rate-of-change. In addition, parasitoid feeding mode did not appear to be influenced by the same host-associated selection pressures that appeared to drive HAD development in gallmakers (Craig et al., 1994). Lastly, the results of this study indicate that

physiological and biological differences between the major arthropod orders tested did not predispose any parasite clade to the development of HAD.

Although the findings of this research largely support previously published research, there are limitations to this study. Specifically, the information threshold required to test the significance of potentially predictive factors (i.e., representation by at least 10 case study systems) could not be met. Therefore, they were excluded from analysis. For example, several case studies have supported the hypothesis that allochrony can generate reproductive isolation between host-associated populations (Akimoto, 1990; Feder et al., 1993; Feder & Filchak, 1999; Mopper, 2005; Ueno et al., 2006; Santos et al., 2007; Schöfl et al., 2009). However, relatively few case studies have been specifically tested for allochronic isolation. Furthermore, gallmaking may have an undiscovered, significant interaction with allochrony. Specifically, the gallmaking aphid *Kaltenbachella japonica* Matsumura consists of eight genetically distinct, host-associated populations which occur in sympatry at various times throughout the growing season, corresponding to budburst in their natal host-plant species (Komatsu & Akimoto, 1995; Abbot, 2001; Leonardo & Muiru, 2003). To appropriately test the relative importance of allochrony for predicting HAD occurrence, future research needs to accumulate more examples of allochrony specifically tested in systems where HAD can be characterized. As another example, the characterization of microbial communities associated with different arthropod populations is a relatively recent scientific focus due to recent methodological breakthroughs in metagenomic analyses, meaning there are few case studies testing the relationship between microbiota and HAD occurrence to date. Despite the lack of case study examples, recent research has shown that microbial symbionts of insect herbivores play significant roles in host-shift survival and HAD development by granting some phytophagous parasites the ability to use different host-plant species (Simon et al., 2003; Tsuchida et al., 2004; Medina et al., 2011; Brady & White, 2013). This relationship between microbial community species composition, host use, and HAD has already been found in some systems (Hopkins & Freckleton, 2002; Medina et al., 2011). For example, the occurrence of two bacterial species (i.e., *Pantoea agglomerans* Gavini and *Serratia marcescens* Bizio) in the pecan leaf phylloxera, *Phylloxera notabilis* Pergande, depends on the phylloxera's host associations, but the reason for this association is unknown (Hosokawa et al., 2007). Like with allochrony, more case study systems need to be tested for differential infection by microbes. Lastly, geometric morphometric analysis has been performed in relatively few parasite species, especially among parasite populations simultaneously

tested for the genetic signature of HAD. Therefore, the relative contribution of morphological or sexual selection for generating HAD remains untested.

Due to the several knowledge gaps described above, multiple imputation was required to simulate missing data for several case studies. Although this is a valid approach for working with incomplete datasets, there is always the possibility of unintentionally introducing bias (Pedersen et al., 2017). Furthermore, each analysis presented in this research was unintentionally biased against ‘HAD absent’ case studies. This outcome highlights the importance of reporting ‘negative’ results in the scientific literature. Case study examples where a phenomenon was expected to be observed but failed to materialize are necessary for quantitative literature reviews, such as this one, to appropriately identify correlates. Future research should place more emphasis on reporting negative results. Lastly, the paucity of case studies tested for HAD meant that specific authors (i.e., RF Medina and JO Stireman) were overrepresented in the dataset. This result was unsurprising because these laboratories specialized in HAD development. It was possible that these laboratories were biased towards discovering the presence of HAD, but these same laboratories produced the largest number of ‘HAD absent’ case studies as well. Furthermore, and more importantly, the techniques used for assessing sympatric genetic differentiation require that (1) allopatry be tested as an alternative hypothesis and (2) numerical evidence must be provided before any hypothesized genetic differentiation can be accepted by the peer-review. Therefore, the designation of ‘HAD present’ is largely unambiguous.

Conclusions

This study shows that the presence of HAD in parasite–host case studies was correlated with the incidence of significant biological and ecological factors. These factors could be used to successfully predict the occurrence of HAD among parasite–host systems. Specifically, case studies involving immigrant inviability and gallmaking were highly likely to develop HAD, whereas most case studies without immigrant inviability did not develop HAD. In addition, case studies that developed HAD in the absence of immigrant inviability did so if exposed to differential host phenologies or recent invasion events. Otherwise, the development of HAD appeared to be idiosyncratic. Therefore, immigrant inviability and gallmaking were the best predictors for HAD, suggesting selection is the major driving force behind HAD development. The authors advocate for the expansion of the HAD case study library and encourage future researchers to specifically test for key predictive factors that could

not be included in this study (i.e., allochryony, differential microbe infection, and morphological differentiation). The authors also advocate for the publication of negative results. Future research into the evolution of HAD and parasite diversity will refine the results of this study and provide novel insights into the factors involved in the development of HAD.

Acknowledgments

We extend our thanks to Dr. Cecilia Tamborindeguy for providing useful critiques and reviews. Partial financial support for this research was provided by USDA Hatch funding (TEX09185, RFM). Software product names are provided as a reference to the reader and do not represent an endorsement of these products by the USDA or Texas A&M University.

Author contributions

Kyle Harrison: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (equal); Methodology (equal); Visualization (lead); Writing-original draft (lead); Writing-review & editing (lead). Aaron M Tarone: Formal analysis (equal); Investigation (supporting); Methodology (supporting); Project administration (supporting); Supervision (supporting); Validation (equal); Writing-original draft (supporting); Writing-review & editing (supporting). Thomas DeWitt: Formal analysis (equal); Methodology (equal); Supervision (supporting); Validation (equal); Writing-original draft (supporting); Writing-review & editing (supporting). Raul F Medina: Conceptualization (lead); Formal analysis (equal); Funding acquisition (lead); Investigation (equal); Methodology (supporting); Project administration (lead); Resources (equal); Supervision (lead); Validation (equal); Writing-original draft (equal); Writing-review & editing (equal).

Data Availability Statement

Supplementary Table S2 provides the raw data and their supporting literature, which were used to construct the data matrix analyzed in this study.

References

- Abbot DK (2001) Evolutionary Genetics of Gall-Forming Aphids: Population and Behavioral Processes. PhD Dissertation, University of Arizona, Tucson, AZ, USA.
- Abbot P & Withgott JH (2004) Phylogenetic and molecular evidence for allochronic speciation in gall-forming aphids (*Pemphigus*). *Evolution* 58: 539–553.

- Abrahamson WG, Brown JM, Roth SK, Sumerford DV, Horner JD et al. (1994) Gallmaker speciation: an assessment of the roles of host-plant characters and phenology, gallmaker competition, and natural enemies. *The Ecology and Evolution of Gall-Forming Insects* (ed. by PW Price, WJ Mattson & YN Baranchikov), pp. 208–222. USDA – Forest Service, St. Paul, MN, USA.
- Abrahamson WG, Eubanks MD, Blair CP & Whipple AV (2001) Gall flies,inquilines, and goldenrods: a model for host-race formation and sympatric speciation. *American Zoologist* 41: 928–938.
- Abrahamson WG, McCrea KD & Anderson SS (1989) Host preference and recognition by the goldenrod ball gallmaker *Eurosta solidaginis* (Diptera: Tephritidae). *American Midland Naturalist* 121: 322–330.
- Agosta SJ (2006) On ecological fitting, plant–insect associations, herbivore host shifts, and host plant selection. *Oikos* 114: 556–565.
- Akimoto S (1990) Local adaptation and host race formation of a gall-forming aphid in relation to environmental heterogeneity. *Oecologia* 83: 162–170.
- Alexander RD & Bigelow RS (1960) Allochronic speciation in field crickets, and a new species, *Acheta veletis*. *Evolution* 14: 334–346.
- Althoff DM (2008) A test of host-associated differentiation across the ‘parasite continuum’ in the tri-trophic interaction among yuccas, bogus yucca moths, and parasitoids. *Molecular Ecology* 17: 3917–3927.
- Althoff DM, Groman JD, Segraves KA & Pellmyr O (2001) Phylogeographic structure in the bogus yucca moth *Prodoxus quinquepunctellus* (Prodoxidae): comparisons with coexisting pollinator yucca moths. *Molecular Phylogenetics and Evolution* 21: 117–127.
- Baer C, Tripp D, Bjorksten T & Antolin M (2004) Phylogeography of a parasitoid wasp (*Diaeretiella rapae*): no evidence of host-associated lineages. *Molecular Ecology* 13: 1859–1869.
- Barman AK, Parajulee MN, Sansone C & Medina RF (2012) Host preference of cotton fleahopper, *Pseudatomoscelis seriatus* (Reuter) is not labile to geographic origin and prior experience. *Environmental Entomology* 41: 125–132.
- Berlocher SH & Feder JL (2002) Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology* 47: 773–815.
- Bernal JS & Medina RF (2018) Agriculture sows pests: how crop domestication, host shifts, and agricultural intensification can create insect pests from herbivores. *Current Opinion in Insect Science* 26: 76–81.
- Brady CM & White JA (2013) Cowpea aphid (*Aphis craccivora*) associated with different host plants has different facultative endosymbionts. *Ecological Entomology* 38: 433–437.
- Brooks DR & Hoberg EP (2007) How will global climate change affect parasite–host assemblages? *Trends in Parasitology* 23: 571–574.
- Brown JM, Abrahamson WG, Packer RA & Way PA (1995) The role of natural-enemy escape in a gallmaker host-plant shift. *Oecologia* 104: 52–60.
- Burban C, Fishpool L, Fauquet C, Fargette D & Thouvenel J (1992) Host-associated biotypes within West African populations of the whitefly *Bemisia tabaci* (Genn.), (Hom., Aleyrodidae). *Journal of Applied Entomology* 113: 416–423.
- Bush GL (1969) Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera, Tephritidae). *Evolution* 23: 237–251.
- Bush GL (1975) Modes of animal speciation. *Annual Review of Ecology and Systematics* 6: 339–364.
- Bush GL (1994) Sympatric speciation in animals: new wine in old bottles. *Trends in Ecology & Evolution* 9: 285–288.
- Butlin R (1987) Species, speciation, and reinforcement. *American Naturalist* 130: 461–464.
- Carmona D, Lajeunesse MJ & Johnson MT (2011) Plant traits that predict resistance to herbivores. *Functional Ecology* 25: 358–367.
- Clements KM, Sorenson CE, Wiegmann BM, Neese PA & Roe RM (2000) Genetic, biochemical, and behavioral uniformity among populations of *Myzus nicotianae* and *Myzus persicae*. *Entomologia Experimentalis et Applicata* 95: 269–281.
- Cook MA, Ozeroff SN, Fitzpatrick SM & Roitberg BD (2011) Host-associated differentiation in reproductive behaviour of cecidomyiid midges on cranberry and blueberry. *Entomologia Experimentalis et Applicata* 141: 8–14.
- Cooley JR, Simon C, Marshall DC, Slon K & Ehrhardt C (2001) Allochronic speciation, secondary contact, and reproductive character displacement in periodical cicadas (Hemiptera: *Magicicada* spp.): genetic, morphological, and behavioural evidence. *Molecular Ecology* 10: 661–671.
- Cornell HV, Hawkins BA & Hochberg M (1998) Towards an empirically-based theory of herbivore demography. *Ecological Entomology* 23: 340–349.
- Craig TP, Horner JD & Itami JK (1997) Hybridization studies on the host races of *Eurosta solidaginis*: implications for sympatric speciation. *Evolution* 51: 1552–1560.
- Craig TP, Horner JD & Itami JK (2001) Genetics, experience, and host-plant preference in *Eurosta solidaginis*: implications for host shifts and speciation. *Evolution* 55: 773–782.
- Craig TP, Itami JK, Abrahamson WG & Horner JD (1993) Behavioral evidence for host-race formation in *Eurosta solidaginis*. *Evolution* 47: 1696–1710.
- Craig TP, Itami JK, Horner JD & Abrahamson WG (1994) Host shifts and speciation in gall-forming insects. *The Ecology and Evolution of Gall-Forming Insects* (ed. by PW Price, WJ Mattson & YN Baranchikov), pp. 194–207. USDA – Forest Service, St. Paul, MN, USA.
- Craig TP, Itami JK, Ohgushi T, Ando Y & Utsumi S (2011) Bridges and barriers to host shifts resulting from host plant genotypic variation. *Journal of Plant Interactions* 6: 141–145.
- Criminisi A, Shotton J & Konukoglu E (2012) Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends in Computer Graphics and Vision* 7: 81–227.

- Cunningham J, West S & Zalucki M (2001) Host selection in phytophagous insects: a new explanation for learning in adults. *Oikos* 95: 537–543.
- Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT et al. (2007) Random forests for classification in ecology. *Ecology* 88: 2783–2792.
- Dambroski HR, Linn C Jr, Berlocher SH, Forbes AA, Roelofs W & Feder JL (2005) The genetic basis for fruit odor discrimination in *Rhagoletis* flies and its significance for sympatric host shifts. *Evolution* 59: 1953–1964.
- Dickey A & Medina R (2010) Testing host-associated differentiation in a quasi-endophage and a parthenogen on native trees. *Journal of Evolutionary Biology* 23: 945–956.
- Dickey AM & Medina RF (2011a) Immigrant inviability in yellow pecan aphid. *Ecological Entomology* 36: 526–531.
- Dickey AM & Medina RF (2011b) Lack of sequential radiation in a parasitoid of a host-associated aphid. *Entomologia Experimentalis et Applicata* 139: 154–160.
- Dickey AM & Medina RF (2012) Host-associated genetic differentiation in pecan leaf phylloxera. *Entomologia Experimentalis et Applicata* 143: 127–137.
- Dobson A, Lafferty KD, Kuris AM, Hechinger RF & Jetz W (2008) Homage to Linnaeus: how many parasites? How many hosts? *Proceedings of the National Academy of Sciences of the USA* 105: 11482–11489.
- Dobson SL, Fox CW & Jiggins FM (2002) The effect of *Wolbachia*-induced cytoplasmic incompatibility on host population size in natural and manipulated systems. *Proceedings of the Royal Society B* 269: 437–445.
- Drès M & Mallet J (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society B* 357: 471–492.
- Egan SP & Funk DJ (2006) Individual advantages to ecological specialization: insights on cognitive constraints from three conspecific taxa. *Proceedings of the Royal Society B* 273: 843–848.
- Egan SP, Nosil P & Funk DJ (2008) Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution* 62: 1162–1181.
- Ehrlich PR & Raven PH (1964) Butterflies and plants: a study in coevolution. *Evolution* 18: 586–608.
- Esteve-Gassent MD, Castro-Arellano I, Fera-Arroyo TP, Patino R, Li AY et al. (2016) Translating ecology, physiology, biochemistry, and population genetics research to meet the challenge of tick and tick-borne diseases in North America. *Archives of Insect Biochemistry and Physiology* 92: 38–64.
- Etges WJ, De Oliveira CC, Ritchie MG & Noor MA (2009) Genetics of incipient speciation in *Drosophila mojavensis* II. Host plants and mating status influence cuticular hydrocarbon QTL expression and G×E interactions. *Evolution* 63: 1712–1730.
- Faeth SH, Mopper S & Simberloff D (1981) Abundances and diversity of leaf-mining insects on three oak host species: effects of host-plant phenology and nitrogen content of leaves. *Oikos* 37: 238–251.
- Farrell BD (1998) 'Inordinate fondness' explained: why are there so many beetles? *Science* 281: 555–559.
- Feder JL, Chilcote CA & Bush GL (1988) Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature* 336: 61–64.
- Feder JL & Filchak KE (1999) It's about time: the evidence for host plant-mediated selection in the apple maggot fly, *Rhagoletis pomonella*, and its implications for fitness trade-offs in phytophagous insects. *Entomologia Experimentalis et Applicata* 91: 211–225.
- Feder JL, Hunt TA & Bush L (1993) The effects of climate, host plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella*. *Entomologia Experimentalis et Applicata* 69: 117–135.
- Feder JL, Opp SB, Wlazlo B, Reynolds K, Go W & Spisak S (1994) Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proceedings of the National Academy of Sciences of the USA* 91: 7990–7994.
- Fisher RC (1988) An inordinate fondness for beetles. *Biological Journal of the Linnean Society* 35: 313–319.
- Forbes AA, Devine SN, Hippee AC, Tvedte ES, Ward AKG et al. (2017) Revisiting the particular role of host shifts in initiating insect speciation. *Evolution* 71: 1126–1137.
- Forbes AA, Powell TH, Stelinski LL, Smith JJ & Feder JL (2009) Sequential sympatric speciation across trophic levels. *Science* 323: 776–779.
- Fordyce JA (2010) Host shifts and evolutionary radiations of butterflies. *Proceedings of the Royal Society B* 277: 3735–3743.
- Fordyce JA & Nice CC (2003) Variation in butterfly egg adhesion: adaptation to local host plant senescence characteristics? *Ecology Letters* 6: 23–27.
- Frey JE, Feder JL, Palma J & Bush GL (1998) Differences in the electroantennal responses of apple- and hawthorn-infesting races of *Rhagoletis pomonella* to host fruit volatile compounds. *Chemoecology* 8: 175–186.
- Fudickar AM, Greives TJ, Atwell JW, Stricker CA & Ketterson ED (2016) Reproductive allochrony in seasonally sympatric populations maintained by differential response to photoperiod: implications for population divergence and response to climate change. *American Naturalist* 187: 436–446.
- Funk DJ (1998) Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. *Evolution* 52: 1744–1759.
- Funk DJ (2010) Does strong selection promote host specialisation and ecological speciation in insect herbivores? Evidence from *Neochlamisus* leaf beetles. *Ecological Entomology* 35: 41–53.
- Futuyma DJ (2008) Sympatric speciation: norm or exception? *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects* (ed. by KJ Tilmon), pp. 136–148. University of California Press, Berkeley, CA, USA.
- Futuyma DJ & Moreno G (1988) The evolution of ecological specialization. *Annual Review of Ecology and Systematics* 19: 207–233.
- Gandon S & Michalakis Y (2002) Local adaptation, evolutionary potential and host-parasite coevolution: interactions between

- migration, mutation, population size and generation time. *Journal of Evolutionary Biology* 15: 451–462.
- Gómez-Díaz E, González-Solís J, Peinado M & Page RD (2007) Lack of host-dependent genetic structure in ectoparasites of *Calonectris* shearwaters. *Molecular Ecology* 16: 5204–5215.
- Groman J & Pellmyr O (2000) Rapid evolution and specialization following host colonization in a yucca moth. *Journal of Evolutionary Biology* 13: 223–236.
- Hartl D (1972) A fundamental theorem of natural selection for sex linkage or arrhenotoky. *American Naturalist* 106: 516–524.
- Hastie T, Tibshirani R, Sherlock G, Eisen M, Brown P & Botstein D (1999) Imputing Missing Data for Gene Expression Arrays. Technical Report, Statistics Department, Stanford University, Stanford, CA, USA.
- Hopkins G & Freckleton RP (2002) Declines in the numbers of amateur and professional taxonomists: implications for conservation. *Animal Conservation* 5: 245–249.
- Horton NJ & Lipsitz SR (2001) Multiple imputation in practice: comparison of software packages for regression models with missing variables. *American Statistician* 55: 244–254.
- Hosokawa T, Kikuchi Y, Shimada M & Fukatsu T (2007) Obligate symbiont involved in pest status of host insect. *Proceedings of the Royal Society B* 274: 1979–1984.
- Howard DJ (1993) Reinforcement: origin, dynamics, and fate of an evolutionary hypothesis. *Hybrid Zones and the Evolutionary Process* (ed. by RG Harrison), pp. 46–69. Oxford University Press, New York, NY, USA.
- Hunter MD (1992) A variable insect–plant interaction: the relationship between tree budburst phenology and population levels of insect herbivores among trees. *Ecological Entomology* 17: 91–95.
- Itami JK, Craig TP & Horner JD (1998) Factors affecting gene flow between the host races of *Eurosta solidaginis*. *Genetic Structure and Local Adaptation in Natural Insect Populations: Effects of Ecology, Life History, and Behavior* (ed. by S Mopper & SY Straus), pp. 375–407. Chapman & Hall, New York, NY, USA.
- Jaenike J (1990) Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics* 21: 243–273.
- Jaenike J & Selander RK (1980) On the question of host races in the fall webworm, *Hyphantria cunea*. *Entomologia Experimentalis et Applicata* 27: 31–37.
- Janz N, Nylin S & Wahlberg N (2006) Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evolutionary Biology* 6: 4.
- Jiggins FM, Hurst GD & Majerus ME (2000) Sex-ratio-distorting *Wolbachia* causes sex-role reversal in its butterfly host. *Proceedings of the Royal Society B* 267: 69–73.
- Katayama N, Abbott JK, Kjærandsen J, Takahashi Y & Svensson EI (2014) Sexual selection on wing interference patterns in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* 111: 15144–15148.
- Kawata M (2002) Invasion of vacant niches and subsequent sympatric speciation. *Proceedings of the Royal Society B* 269: 55–63.
- Kelly CD (2014) Sexual selection, phenotypic variation, and allometry in genitalic and non-genitalic traits in the sexually size-dimorphic stick insect *Micrarchus hystricleus*. *Biological Journal of the Linnean Society* 113: 471–484.
- King CE (1993) Random genetic drift during cyclical ameiotic parthenogenesis. *Hydrobiologia* 255: 205–212.
- Kohnen A, Wissemann V & Brandl R (2011) No host-associated differentiation in the gall wasp *Diplolepis rosae* (Hymenoptera: Cynipidae) on three dog rose species. *Biological Journal of the Linnean Society* 102: 369–377.
- Komatsu T & Akimoto S (1995) Genetic differentiation as a result of adaptation to the phenologies of individual host trees in the galling aphid *Kaltenbachiella japonica*. *Ecological Entomology* 20: 33–42.
- Leonardo TE & Muir GT (2003) Facultative symbionts are associated with host plant specialization in pea aphid populations. *Proceedings of the Royal Society B* 270: S209–S212.
- Li W-H, Ellsworth DL, Krushkal J, Chang BH-J & Hewett-Emmett D (1996) Rates of nucleotide substitution in primates and rodents and the generation–time effect hypothesis. *Molecular Phylogenetics and Evolution* 5: 182–187.
- Linn CE, Dambroski HR, Feder JL, Berlocher SH, Nojima S & Roelofs WL (2004) Postzygotic isolating factor in sympatric speciation in *Rhagoletis* flies: reduced response of hybrids to parental host-fruit odors. *Proceedings of the National Academy of Sciences of the USA* 101: 17753–17758.
- Loxdale HD (2008) The nature and reality of the aphid clone: genetic variation, adaptation and evolution. *Agricultural and Forest Entomology* 10: 81–90.
- Loxdale HD (2009) What's in a clone: the rapid evolution of aphid asexual lineages in relation to geography, host plant adaptation and resistance to pesticides. *Lost Sex: The Evolutionary Biology of Parthenogenesis* (ed. by I Schön, K Martens & P van Dijk), pp. 535–557. Springer, Dordrecht, The Netherlands.
- Loxdale HD (2010) Rapid genetic changes in natural insect populations. *Ecological Entomology* 35: 155–164.
- Lozier J, Roderick G & Mills N (2009) Molecular markers reveal strong geographic, but not host associated, genetic differentiation in *Aphidius transcaspicus*, a parasitoid of the aphid genus *Hyalopterus*. *Bulletin of Entomological Research* 99: 83–96.
- Lynch M (1984) The limits to life history evolution in *Daphnia*. *Evolution* 38: 465–482.
- Lynch M & Gabriel W (1983) Phenotypic evolution and parthenogenesis. *American Naturalist* 122: 745–764.
- Mallet J, Meyer A, Nosil P & Feder JL (2009) Space, sympatry and speciation. *Journal of Evolutionary Biology* 22: 2332–2341.
- Margaritopoulos JT, Malarky G, Tsitsipis JA & Blackman RL (2007) Microsatellite DNA and behavioural studies provide evidence of host-mediated speciation in *Myzus persicae* (Hemiptera: Aphididae). *Biological Journal of the Linnean Society* 91: 687–702.
- Martin AP & Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the USA* 90: 4087–4091.

- Matsubayashi KW, Ohshima I & Nosil P (2010) Ecological speciation in phytophagous insects. *Entomologia Experimentalis et Applicata* 134: 1–27.
- McCall PJ, Turlings TCJ, Lewis WJ & Tumlinson JH (1993) Role of plant volatiles in host location by the specialist parasitoid *Microplitis croceipes* Cresson (Braconidae: Hymenoptera). *Journal of Insect Behavior* 6: 625–639.
- McCoy KD, Léger E & Dietrich M (2013) Host specialization in ticks and transmission of tick-borne diseases: a review. *Frontiers in Cellular and Infection Microbiology* 3: 57.
- Medina RF (2012) Implications of host-associated differentiation in the control of pest species. *Insect Outbreaks Revisited* (ed. by P Barbosa, DK Letourneau & AA Agrawal), pp. 291–310. John Wiley & Sons, Chichester, UK.
- Medina R, Dickey A, Harrison K & Miller G (2017) Host-associated differentiation in a pecan and water hickory Aphidomorpha community. *Entomologia Experimentalis et Applicata* 162: 366–378.
- Medina R, Nachappa P & Tamborindéguy C (2011) Differences in bacterial diversity of host-associated populations of *Phylloxera notabilis* Pergande (Hemiptera: Phylloxeridae) in pecan and water hickory. *Journal of Evolutionary Biology* 24: 761–771.
- Medina R, Szendrei Z, Harrison K, Isaacs R, Averill A et al. (2014) Exploring host-associated differentiation in the North American native cranberry fruitworm, *Acrobasis vaccinii*, from blueberries and cranberries. *Entomologia Experimentalis et Applicata* 150: 136–148.
- Michel AP, Sim S, Powell TH, Taylor MS, Nosil P & Feder JL (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the USA* 107: 9724–9729.
- Mitter C, Farrell B & Wiegmann B (1988) The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *American Naturalist* 132: 107–128.
- Mopper S (2005) Phenology – how time creates spatial structure in endophagous insect populations. *Annales Zoologici Fennici* 42: 327–333.
- Murphy SM (2004) Enemy-free space maintains swallowtail butterfly host shift. *Proceedings of the National Academy of Sciences of the USA* 101: 18048–18052.
- Naisbit RE, Jiggins CD & Mallet J (2001) Disruptive sexual selection against hybrids contributes to speciation between *Heliconius cydno* and *Heliconius melpomene*. *Proceedings of the Royal Society B* 268: 1849–1854.
- Neiman M & Linksvayer T (2006) The conversion of variance and the evolutionary potential of restricted recombination. *Heredity* 96: 111–121.
- Nosil P (2007) Divergent host plant adaptation and reproductive isolation between ecotypes of *Timema cristinae* walking sticks. *American Naturalist* 169: 151–162.
- Nosil P, Crespi BJ & Sandoval CP (2002) Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* 417: 440–443.
- Nosil P, Vines TH & Funk DJ (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* 59: 705–719.
- Onstad DW, Reissig WH & Shoemaker CA (1986) Influence of apple cultivar, tree phenology, and leaf quality on the development and mortality of *Choristoneura rosaceana* (Lepidoptera: Tortricidae). *Canadian Entomologist* 118: 123–132.
- Otte T, Hilker M & Geiselhardt S (2018) Phenotypic plasticity of cuticular hydrocarbon profiles in insects. *Journal of Chemical Ecology* 44: 235–247.
- Pappers SM, van der Velde G, Ouborg NJ & van Groenendael JM (2002) Genetically based polymorphisms in morphology and life history associated with putative host races of the water lily leaf beetle, *Galerucella nymphalae*. *Evolution* 56: 1610–1621.
- 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. *Proceedings of the National Academy of Sciences of the USA* 106: 7495–7500.
- Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM et al. (2017) Missing data and multiple imputation in clinical epidemiological research. *Clinical Epidemiology* 9: 157.
- Poulin R (1997) Species richness of parasite assemblages: evolution and patterns. *Annual Review of Ecology and Systematics* 28: 341–358.
- Powell TH, Hood GR, Murphy MO, Heilveil JS, Berlocher SH et al. (2013) Genetic divergence along the speciation continuum: the transition from host race to species in *Rhagoletis* (Diptera: Tephritidae). *Evolution* 67: 2561–2576.
- Prasad AM, Iverson LR & Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9: 181–199.
- Price PW (1980) *Evolutionary Biology of Parasites*. Princeton University Press, Princeton, NJ, USA.
- Pritchard JK, Stephens M & Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Pureswaran DS, Gries R & Borden JH (2004) Antennal responses of four species of tree-killing bark beetles (Coleoptera: Scolytidae) to volatiles collected from beetles, and their host and non-host conifers. *Chemoecology* 14: 59–66.
- Rabosky DL (2009) Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. *Systematic Biology* 58: 629–640.
- Raijmann LEL & Menken SBJ (2000) Temporal variation in the genetic structure of host-associated populations of the small ermine moth *Yponomeuta padellus* (Lepidoptera, Yponomeutidae). *Biological Journal of the Linnean Society* 70: 555–570.
- Rodriguez JJ, Fernández-Triana JL, Smith MA, Janzen DH, Hallwachs W et al. (2013) Extrapolations from field studies and known faunas converge on dramatically increased estimates of global microgastrine parasitoid wasp species richness (Hymenoptera: Braconidae). *Insect Conservation and Diversity* 6: 530–536.
- Roy D, Seehausen O & Nosil P (2013) Sexual dimorphism dominates divergent host plant use in stick insect trophic morphology. *BMC Evolutionary Biology* 13: 135.

- Rundle HD (2002) A test of ecologically dependent postmating isolation between sympatric sticklebacks. *Evolution* 56: 322–329.
- Rundle HD & Nosil P (2005) Ecological speciation. *Ecology Letters* 8: 336–352.
- Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J et al. (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics* 32: 305–332.
- Santos H, Rousset J, Magnoux E, Paiva M-R, Branco M & Kerdelhué C (2007) Genetic isolation through time: allochronic differentiation of a phenologically atypical population of the pine processionary moth. *Proceedings of the Royal Society B* 274: 935–941.
- Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution* 16: 372–380.
- Schmitt P, Mandel J & Guedj M (2015) A comparison of six methods for missing data imputation. *Journal of Biometrics & Biostatistics* 6: 224.
- Schöfl G, Heckel DG & Groot AT (2009) Time-shifted reproductive behaviours among fall armyworm (Noctuidae: *Spodoptera frugiperda*) host strains: evidence for differing modes of inheritance. *Journal of Evolutionary Biology* 22: 1447–1459.
- Schwarz D, Matta BM, Shakir-Botteri NL & McPherson BA (2005) Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature* 436: 546–549.
- Servedio MR (2001) Beyond reinforcement: the evolution of pre-mating isolation by direct selection on preferences and post-mating, prezygotic incompatibilities. *Evolution* 55: 1909–1920.
- Servedio MR & Noor MA (2003) The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics* 34: 339–364.
- Simon J-C, Carré S, Boutin M, Prunier-Leterme N, Sabater-Muñoz B et al. (2003) Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. *Proceedings of the Royal Society B* 270: 1703–1712.
- Simonato M, Battisti A, Zovi D & Medina RF (2012) Testing for host-associated differentiation in two egg parasitoids of a forest herbivore. *Entomologia Experimentalis et Applicata* 145: 124–133.
- Stamp NE & Bowers MD (1990) Phenology of nutritional differences between new and mature leaves and its effect on caterpillar growth. *Ecological Entomology* 15: 447–454.
- Stelinski L & Liburd O (2005) Behavioral evidence for host fidelity among populations of the parasitic wasp, *Diachasma alloeum* (Muesebeck). *Naturwissenschaften* 92: 65–68.
- Stiling P & Rossi AM (1998) Deme formation in a dispersive gall-forming midge. *Genetic Structure and Local Adaptation in Natural Insect Populations: Effects of Ecology, Life History, and Behavior* (ed. by S Mopper & SY Straus), pp. 22–36. Chapman & Hall, New York, NY, USA.
- Stireman JO, Janson EM, Carr TG, Devlin H & Abbot P (2008) Evolutionary radiation of *Asteromyia carbonifera* (Diptera: Cecidomyiidae) gall morphotypes on the goldenrod *Solidago altissima* (Asteraceae). *Biological Journal of the Linnean Society* 95: 840–858.
- Stireman JO, Nason JD & Heard SB (2005) Host-associated genetic differentiation in phytophagous insects: general phenomenon or isolated exceptions? Evidence from a goldenrod-insect community. *Evolution* 59: 2573–2587.
- Stireman JO, Nason JD, Heard SB & Seehawer JM (2006) Cascading host-associated genetic differentiation in parasitoids of phytophagous insects. *Proceedings of the Royal Society B* 273: 523–530.
- Strand MR & Pech LL (1995) Immunological basis for compatibility in parasitoid-host relationships. *Annual Review of Entomology* 40: 31–56.
- Tsuchida T, Koga R & Fukatsu T (2004) Host plant specialization governed by facultative symbiont. *Science* 303: 1989.
- Ueno H, Furukawa S & Tsuchida K (2006) Difference in the time of mating activity between host-associated populations of the rice stem borer, *Chilo suppressalis* (Walker). *Entomological Science* 9: 255–259.
- Via S (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution* 53: 1446–1457.
- Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution* 16: 381–390.
- Via S, Bouck AC & Skillman S (2000) Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution* 54: 1626–1637.
- Via S & Hawthorne DJ (2002) The genetic architecture of ecological specialization: correlated gene effects on host use and habitat choice in pea aphids. *American Naturalist* 159: S76–S88.
- Wade MJ & Johnson NA (1994) Reproductive isolation between two species of flour beetles, *Tribolium castaneum* and *T. freemani*: variation within and among geographical populations of *T. castaneum*. *Heredity* 72: 155–162.
- Waring GL, Abrahamson WG & Howard DJ (1990) Genetic differentiation among host-associated populations of the gall-maker *Eurosta solidaginis* (Diptera: Tephritidae). *Evolution* 44: 1648–1655.
- Weinstein SB & Kuris AM (2016) Independent origins of parasitism in Animalia. *Biology Letters* 12: 20160324.
- Welling M & Kurihara K (2009) Bayesian *k*-means as a ‘maximization-expectation’ algorithm. *Neural Computation* 21: 1145–1172.
- Wiegmann BM, Mitter C & Farrell B (1993) Diversification of carnivorous parasitic insects: extraordinary radiation or specialized dead end? *American Naturalist* 142: 737–754.
- Wood T, Tilmon K, Shantz A, Harris C & Pesek J (1999) The role of host-plant fidelity in initiating insect race formation. *Evolutionary Ecology Research* 1: 317–332.
- Wu C-I & Palopoli MF (1994) Genetics of postmating reproductive isolation in animals. *Annual Review of Genetics* 28: 283–308.
- Yokomi R & Tang Y (1995) Host preference and suitability of two aphelinid parasitoids (Hymenoptera: Aphelinidae) for aphids (Homoptera: Aphididae) on citrus. *Journal of Economic Entomology* 88: 840–845.

Yukawa J (2000) Synchronization of galls with host plant phenology. *Population Ecology* 42: 105–113.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Search terms for discovery of arthropod parasite–host case studies tested for the genetic signature of host-associated differentiation (HAD) and the potential predictive factors described in each case study. Words in parentheses were included in search terms after searching for words outside of parentheses. The words ‘host’ and ‘parasite’ were also replaced with the specific names of the organisms involved in each case study.

Table S2. Table of parasite–host case studies tested in the quantitative literature review. Rows represent the parasite–host case studies assessed. Columns represent the biological and ecological descriptions of each parasite–host study system: (1) Scientific name of the parasitic arthropod species involved in a case study, (2)

scientific name for the hosts or host lineage, (3) presence/absence of host-associated genetic structuring in a given system, (4) order of the parasitic arthropod, (5) order(s) of the host(s), (6) parasite’s feeding mode and trophic position, (7) host resource used by the parasite, (8) endophagous vs. exophagous parasite life history, (9) parasite reproductive mode, (10) number of generations the parasite typically experiences annually, (11) whether the parasite obligately relies on a host to find mates, (12) whether host resources are available at the same time or different times, (13–15) whether parasite populations were tested for host volatile preference, oviposition preference, or immigrant inviability, (16) whether parasite populations exist at different times of the year, (17) whether parasite populations have distinct morphological features, (18) whether the parasite or hosts are recent invaders, (19) whether the hosts are anthropogenically managed, and (20) whether evidence exists of host-mediated, differential infection by microbes. Missing data are marked ‘.’