PARALLEL AND NONPARALLEL PATTERNS OF GENETIC CO-DIFFERENTIATION: EVIDENCE FOR HOST ASSOCIATED DIFFERENTIATION AMONG TROPHIC LEVELS OF THE OAK GALL WASP SYSTEM

HONORS THESIS

Presented to the Honors College of
Texas State University
in Partial Fulfillment
of the Requirements

for Graduation in the Honors College

by

Simon Lee Beymer

San Marcos, Texas
May 2019

PARALLEL AND NONPARALLEL PATTERNS OF GENETIC CO-DIFFERENTIATION: EVIDENCE FOR HOST ASSOCIATED DIFFERENTIATION AMONG TROPHIC LEVELS OF THE OAK GALL WASP SYSTEM

by

Simon Lee Beymer

Thesis Supervisor:

_____

James R. Ott, Ph.D.
Department of Biology

Second Reader:

_____

Chris Nice, Ph.D.
Department of Biology

Approved:

_____

Heather C. Galloway, Ph.D.
Dean, Honors College

TABLE OF CONTENTS

LIST OF FIGURES

APPENDICES

Page

Abstract

Populations of phytophagous insects whose life cycles are intimately related to their host plants can express host-associated differentiation (HAD) in the form of genetic divergence and/or RI isolation across alternative host plants. In this work, we ask if HAD cascades to higher trophic levels in parasitoids of an insect herbivore that expresses HAD. *Belonocnema treatae* (Hymenoptera: Cynipidae) induces galls on three species of live oak (Quercus) across the southern USA. Both HAD and geography structure genomic variation among host associated populations of *B. treatae*. The generalist inquiline parasitoid, *Synergus* sp. (Hymenoptera: Cynipidae) develops within *B. treatae* galls. Herein, we test the hypothesis of co-genetic divergence between the inquiline and its insect host and between the inquiline and the host plant on which it develops. Genotyping-by-sequencing was conducted for 758 *Synergus* from 35 populations selected to reflect known genomic structure of *B. treatae* across its host plants and geographic range. Population genomic structure of the inquiline based on 57,664 single nucleotide polymorphisms was then investigated using Principal Component Analysis and the hierarchical Bayesian model ENTROPY to assign individuals to genetic clusters and estimate admixture proportions. ENTROPY revealed significant substructure within *Synergus* sp. corresponding to five independent lineages, three of which represent cryptic taxa. The five lineages display various degrees of overlap with respect to both host plant and geography including a longitudinal division shared with the gall former but no strong

evidence of shared HAD with their insect host or their host plant. Thus, while the herbivorous insect displays HAD its inquiline parasites do not.

I. INTRODUCTION

Understanding the processes that generate biodiversity is central to the modern synthesis of biology that links evolution, ecology, and genetics (Dobzhansky, 1951; Mayr, 1942; Nosil, 2008; Via, 2002). Ecology and geography are two of the most common drivers of diversification (Cabej 2012; Turelli *et al*., 2001) and herbivorous insects are among the most diverse organisms on the globe with their diversity stemming from adaptive radiation relative to their host (Ehrlich, & Raven, 1964; Forbes *et al*., 2017; Price, 2002; Walsh, 1861) and relative to their insect natural enemies (Bernays, & Graham, 1988; Singer, & Stireman, 2005; Vamosi, 2005). At the population level, the initial stages of differentiation of insect herbivores can be initiated by divergent selection acting on populations occupying alternative host plants in a process known as "Host Associated Differentiation" (HAD) (Bush, 1969; Walsh, 1861). When alternative host plant use is coupled with the development of reproductive isolation, divergence initiated by HAD can lead to host race formation (reviewed by Forbes *et al*., 2017). There has been a growing interest in the comparative examination of HAD among generalist and specialist insect species and the possible role that HAD can play in initiating speciation (Bush, 1969; Dres, & Mallet, 2002; Forbes *et al*., 2009; Funk *et al*., 2002; Nosil *et al*., 2002; Sword *et al*., 2005).

The degree of intimacy with the host is thought to be a contributing factor to HAD (Price *et al*., 1980; Medina, 2012). Thus, specialist insect herbivores whose feeding, development, and mating may be tightly linked to a single, or series of closely related, host plant(s) appear more likely to display HAD relative to species that exploit multiple hosts (Forbes *et al*., 2017). Thus, increasing specialization is linked to greater

likelihood of HAD (Forbes *et al*., 2009; Funk *et al*. 2002; Medina, 2012). However, generalist insect herbivores are likely sources in which to investigate cryptic HAD when such species are distributed across multiple host plants throughout their geographic range and selection favors regional specialization via local adaptation to specific host plants (Antwi *et al*., 2015; Dopman *et al*., 2002; Sword *et al*., 2005; Fox, & Morrow, 1981). For example, generalist herbivorous insect species such as cotton flea hoppers and snakeweed grasshoppers have been shown to be composed of regional or local specialist populations (Barman *et al*., 2012; Sword *et al*., 2005).

Similarly, there has been long standing interest in the patterns of co-differentiation between host and parasites. Parasitoids may exhibit HAD as they are intimately related to their host insect (Cook, & Segar 2010; Godfray, 1994; Stireman *et al*., 2006). As phytophagous insects become specialized in their local environment, they provide new habitats (new niches) for their natural enemies to exploit (Hood *et al*., 2015). Selection for niche matching generates a form of ecological speciation where the divergence of the parasite can parallel that of the host (Schluter, 2000; Stireman *et al*., 2006). Alternatively, the host plants may release chemical cues (Rosenthal, & Berenbaum, 2012; Tuomi *et al.*, 1988) that attract the natural enemies of insect herbivores as a method of self-defense (Turlings *et al*., 1990; Dicke, & Sabelis, 1987). In this way, host plants can influence the natural enemy community of herbivores, which might lead to HAD between the host plant and the parasitoid (Lill *et al*., 2002). These intimate cross-trophic level relationships form the predictions for cascading HAD across trophic levels in plant–herbivore–parasitoid systems (Althoff, 2008; Forbes *et al*., 2009;

Singer, & Stireman, 2005; Stireman *et al*., 2006). If the herbivorous insect displays HAD, then one may predict that obligate parasites of the insect might also display HAD.

Gall wasps (Cynipidea: Hymenoptera) are highly specialized insect herbivores with each species typically inducing galls on specific, ephemeral tissues on a single, or series of closely related, plant species (Askew, 1984; Quicke, 1997). Cynipid gall formers feed exclusively and complete development within galls which are composed of plant tissue but develop under the control of the insect (Rohfritsch, 1992; Stone *et al*., 2002). Gall formers can be considered endoparasites of plants as they develop surrounded by plant tissue where they confront plant defenses (Price, 2002). While galls are hypothesized to function as defensive structures (Price, & Pschorn-Walcher, 1988), developing cynipid larvae typically suffer from high levels of mortality due to a diverse community of insect natural enemies (Forbes *et al*., 2015; Hood, & Ott, 2010). Included within the insect natural enemy communities of cynipid gall formers are two guilds which function as parasites: parasitoids and inquilines. Parasitoids develop within gall formers where they confront the insect's host immune responses (Carton *et al*., 2008; Strand & Noda, 1991; Strand & Pech, 1995) while inquilines develop by feeding within galls on plant tissue that has been modified by the gall former wherein they confront plant defenses (Price *et al*., 1980).

Cynipid natural enemy communities are models for understanding community structure and evolution (Askew, 1984, Bailey *et al*., 2009; Hayward, & Stone 2005; Forbes *et al*., 2015; Stone *et al*., 2009). An outstanding question of the evolution of insect herbivores and their natural enemy communities is whether HAD within species of insect herbivores can drive parallel patterns of HAD among their associated natural

enemies and/or parasites (Althoff, 2008; Stireman *et al*., 2006). Given the intimate relationships between gall formers, their host plants, parasitoids and inquilines, this question can be asked at two levels within gall former-natural enemy systems. Herein we investigate the hypothesis of HAD of a widespread inquiline in relation to the patterns of host plant associated differentiation exhibited by its gall former host across three host plants as well as in relation to patterns of genomic structure exhibited by the gall former across its geographic range.

Study System

Belonocnema treatae (Hymenoptera: Cynipidae) induces galls on three species of American live oaks: *Quercus fusiformis*, (*Qf*), *Q. virginiana* (*Qv*), and *Q. geminata*, (*Qg*) across the southern and southeastern United States (Driscoe 2018; Schuler *et al*., 2018). Temporally distinct sexual and asexual generations alternate to complete the lifecycle with sexual organisms inducing galls on live oak leaves and asexual organisms inducing galls on roots (Lund *et al*., 1998). Leaf galls are subsequently attacked by at least 25 species of insect natural enemies (Forbes *et al*., 2015). This natural enemy community includes both parasitoids and inquilines. *Synergus spp*. (Hymenoptera: Cynipidae) are the most abundant natural enemy (Forbes *et al*., 2015; Busbee, 2018). The genus *Synergus* represents a clade within the family Cynipidae in which the ability to induce galls has been lost (Ronquist, 1994). Instead, *Synergus* oviposit into the galls induced by other gall forming species including Cynipids. Larvae then develop as inquilines that feed primarily on gall tissue and advantageously on gall occupants. *Synergus* are generalist inquilines most commonly associated with oak galls (Askew, 1984; Stone *et al*., 2002), but also with rose galls (Askew, 1984) and other hosts (Csoka *et al*., 2005). *Synergus sp*. develop

within leaf galls induced by *B. treatae* on all three species of live oaks on which *B. treatae* develops (Busbee, 2018). Thus, *Synergus* are intimately linked throughout development to both the live oak species that their larvae feed on and to *B. treatae* which induces these modified plant structures.

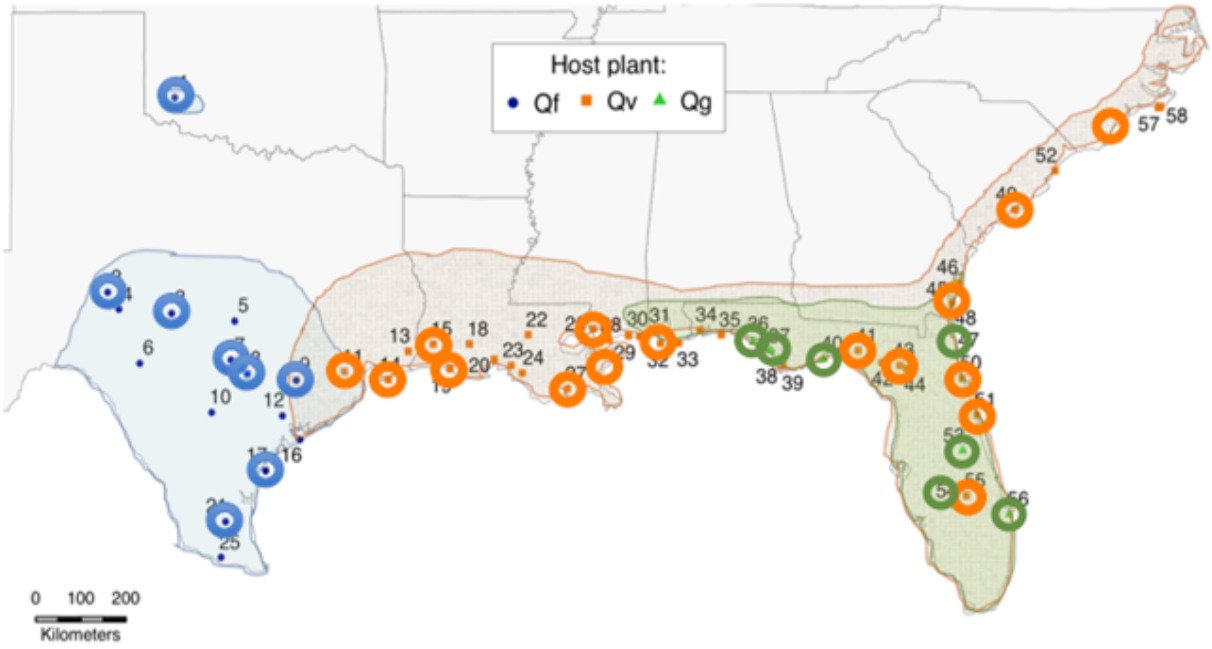

Figure 1A.Collection sites of *Synergus* relative to *B. treatae* geographic structure. 176 individuals from 8 sites were collected from *Q. fusiformis*, 412 individuals from 19 sites on *Q. virginiana*, and 180 individuals from 8 sites on *Q. geminata*. Mean = 21.9 ± 4.9 from 35 sites. Each dot represents a *B. treatae* sampling location and a circle represents a *Synergus* sampling location.

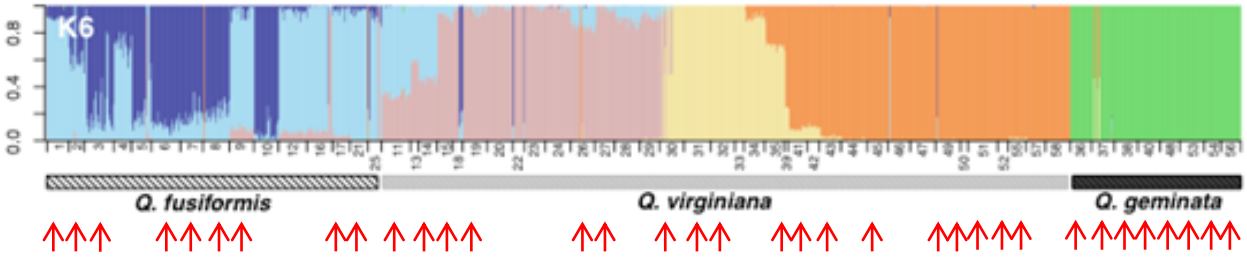

Figure 1B. Collection sites of *Synergus* relative to *B. treatae* geographic genomic structure shown via ENTROPY (k = 6). Admixture proportions (q) of 1,219 *B. treatae* sampled from 58 sites distributed across three host plants (*Qf* = 14, *Qv* = 36, *Qg* = 8). Shown are results for *k* =6. Maximum likelihood estimates of the admixture proportion (y-axis) using 6,987 loci. Individuals are ordered by host plant affiliation and the absolute

distance from the northwestern most and isolated locality (site 1, Quartz Mt., OK). The proportion of each individual's ancestry is denoted by the height of each block of color (genetic cluster). The red arrows indicate the locations of the 35 *Synergus* sampled site relative the k=6 clusters identified within *B. treatae*.

Previous research examined evidence of HAD in *B. treatae* by sampling *B. treatae* throughout the geographic range of the three live oak species throughout the United States and, thus, the known range of *B. treatae* (Driscoe *et al*., 2019 in review; Fig. 1A). Driscoe (2019) extracted DNA from 1,219 individual *B. treatae* distributed across the three host plants from 58 sites for Next Generation Sequencing. Figure 1B shows the observed genomic substructure across the geographic range of *B. treatae* as depicted by the ancestry proportions for $k = 6$, where k is the number of inferred populations, estimated by ENTROPY (Driscoe *et al*. 2019). Figure 1 depicts evidence for HAD as organisms designated by the green cluster correspond *to B. treatae* reared from galls collected exclusively from *Qg*. That these organisms collected from multiple *Qg* sites cluster together means they are genetically similar as a result of shared host plant. If HAD was the only pattern driving the genomic structure of *B. treatae* across its geographic range, and genomic differentiation was deep among host plant related lineages, then at $k = 3$ *B. treatae* would cluster corresponding to each of the three host plants. Instead, geography also plays a significant role in driving the genetic differentiation of *B. treatae* as ENTROPY shows evidence for a central longitudinal division at $k = 2$ (not shown) which corresponds to the Mississippi region (Figure 1). This longitudinal division is central to the system and is also observed across systems in the Southeastern United States (Gonzalez *et al*., 2008; Walker, & Avise, 1998). At $k = 6$ the

contribution of geographic variation and HAD to the pattern of genomic variation in *B. treatae* is clear (Fig 1B).

The pattern of genetic structure in *B. treatae* as a function of geography and host plant form the basis for predictions tested herein of parallel patterns of genetic substructure within *Synergus*. If genetic divergence among *Synergus* populations is strictly imposed by the host plants used by *B. treatae*, then we predict that ENTROPY analysis of genotyping-by-sequencing data for *Synergus* would show three clusters at $k = 3$ that correspond to the host plant from which samples were collected. Alternatively, if divergence aligns with the observed patterns of divergence exhibited by single nucleotide polymorphism (SNP) data for *B. treatae*, (i.e., the divergence in *Synergus* parallels the evolutionary history of its host insect) then we predict that the ENTROPY model on *Synergus* for $k = 6$ will cluster so as to match the lineages observed in *B. treatae*. If divergence in *Synergus* is solely the result of isolation by distance, then we predict clusters by geography or a gradual gradient along the ENTROPY plot for $k = 2$.

We also seek to use SNP data to resolve taxonomic uncertainty in *Synergus*. While inquilines are both abundant and integral to cynipid natural enemy communities, they are the least understood (Askew 1961, Stone *et al*., 2002). Taxonomic identification is difficult (Wiebes-Rijks, 1979; Acs *et al*., 2010) due in part to high between generation variation–specifically size and color (Mayr, 1872). Based on COI sequence data, Forbes *et al*. (2015) reported three "species" of *Synergus* reared from *B. treatae* leaf galls collected from *Qf* and *Qv* in central Texas. Busbee (2018) sampled *Synergus* across 74 sites distributed across the full geographic range of *Qf*, *Qv*, and *Qg* in the USA (described below) and noted three distinct morphotypes and their geographic distributions (Fig. 2).

M1 M2 M3

Figure 2. Three morphotypes of *Synergus* sp. identified by Busbee (2018) reared from the *B. treatae* leaf galls collected from *Qf*, *Qv* and *Qg* across the geographic range of each host plant.

The relationship between the taxa identified by Forbes and morphs identified by Busbee is uncertain as is the relationship between morphs and patterns of genetic divergence with respect to host plants and geography. Thus, in addition to testing for HAD within *Synergus*, the analysis of population genetic structure of *Synergus* using SNP data supported three closely related goals. First, to resolve the relationships among the "COI-based species" identified by Forbes, the "morphs" identified by Busbee, and the lineages uncovered herein based on patterns of SNP variation. Second, to understand the relationships of SNP-based lineages relative to one another. Third, uncover the geography of, and patterns of co-occurrence of, SNP based-lineages within and among host associated populations.

II. Materials & Methods

As described by Busbee (2018), mature leaf galls containing penultimate stage *B. treatae* and or developing insect natural enemies were collected during October and November of 2015 and again in 2016 from a total of 74 sites distributed across the geographic ranges of *Qf*, *Qv*, and *Qg*, which together span the southern and southeastern

United States. Leaf galls were removed from leaves and placed into collection traps stored under ambient conditions at the Texas State University Greenhouse where emergent *B. treatae* and natural enemies including *Synergus sp*. were collected daily over the course of two years. Emergent organisms were stored by site in 95% ethanol at the time of emergence. All insect natural enemies were subsequently identified (Busbee 2018). Driscoe *et al*. (in review) as summarized above subsequently investigated population genomic structure within *B. treatae* based on sampling 58 of the 74 sites that were selected to span the geographic range of *B. treatae* across the geographic ranges of its three host plants (Fig. 1A).

Herein we investigate population genomic structure within *Synergus* using genotyping-by-sequencing of 768 individuals drawn from 35 of the 58 sites studied by Driscoe *et al*. The 35 sites included 8 sites from *Qf*, 19 sites from *Qv*, and 8 sites from *Qg* (Fig. 1A). Supplemental table 1 illustrates the correspondence between sample site selection across studies and provides details of sample sizes for each site for *Synergus*. Central to testing the hypothesis of parallel genetic divergence between *B. treatae* and *Synergus*, the 35 *Synergus* sites were selectively drawn to sample within each of the $k = 6$ clusters identified by ENTROPY for *B. treatae* (Driscoe *et al*. in review 2019; Fig. 1B). Overall, this sampling design allows us to compare the patterns of genomic structure for *Synergus* in relation to A) host plant affiliation and B) the pattern evident in *B. treatae*. Prior to DNA extraction, all 768 individual *Synergus* sp. were identified to one of three morphotypes designed as (M1–M3) following Busbee *et al*. (2018) (Fig. 2).

Genomic Library Preparation

Genomic DNA was isolated from 768 individual (diploid) adult female *Synergus* (mean = 21.9 ± 4.9 per site; range = 9–36) by homogenizing whole bodies following the DNeasy Blood and Tissue Kit protocol (Qiagen Inc.) We created a reduced representation genomic library for each individual using a multiplexed genotyping-by-sequencing (GBS) approach, following the protocols of Parchman *et al*. (2012), Gompert *et al*. (2014) and Mandeville *et al*., (2015). Briefly, genomic DNA was digested using two restriction enzymes, EcoR1 and Mse1, at non-targeted sites throughout the genome. Customized Illumina adaptor sequences containing the primer sequences and unique 8-10 bp individual identifiers were ligated to DNA fragments. Fragments were then amplified in two separate rounds of PCR for each sample. Individual libraries were pooled then size selected for fragments of 250-350 bp using BluePippin quantitative electrophoresis at The University of Texas Genomic Sequencing and Analysis Facility (Austin, TX). Before and after size selection, DNA concentration and quality were verified using a BioAnalyzer prior to Illumina sequencing (University of Texas, Austin). The two pooled, size-selected libraries were each sequenced twice at the University of Texas Genomic Sequencing and Analysis Facility (Austin, TX) across four lanes on the Illumina HiSeq 2500 platform. Reads were generated as single-end 100 bp sequences. We obtained a total of 5.25×10^8 parsed reads.

Assembly and Variant Calling

Identifier and restriction site sequences were removed from sequence reads and identified to individuals using custom perl scripts (available from the authors). An artificial reference genome was then created using the clustering approach of dDocent with minor modifications following the methods of Puritz *et al*. (2014a) and Puritz *et al*. (2014b) using

the 475 million reads retained after parsing the sequence data of all individuals. For sequences to be considered for inclusion into the reference genome, we required 4 or more reads represented by ≥4 individuals with a minimum of 80% sequence homology. This resulted in an artificial reference comprised of 126,937 contigs, to which all sequence reads were assembled using BWA ver. 0.7.13 (Burrows-Wheeler Aligner, Li, & Durbin 2009). SAMtools ver. 0.1.19 and BCFtools ver. 0.1.19 (Li *et al*., 2009) were used to index, sort and merge the individual alignments, and then identify bi-allelic SNPs for which at least 50% of all individuals had reads. We used a full prior for variant calling and set the threshold probability for identifying a variant site at P = 0.05. We incorporated genotype uncertainty due to sequencing and alignment errors in downstream analysis by retaining genotype likelihoods (Li, 2011, Skotte *et al*., 2013). SNPs with more than one alternative allele were removed to eliminate potential paralogs. One SNP per fragment was chosen at random and retained for analysis to minimize linkage disequilibrium among SNPs. We then sorted variants by minor allele frequencies (MAF), retaining SNPs with a MAF > 5%. We removed low coverage individuals (N = 10 with < 1x median coverage), then repeated subsequent filtering starting at variant calling. In total, we retained 758 individuals (mean = 21.7 ± 5.0 per site, range = 9–36) and identified 57,664 loci with an average median sequence coverage of 5.9 and an average of 2.8 reads per individual. This level of sequence coverage was sufficient for downstream analyses that incorporate genotype uncertainty (Buerkle, & Gompert, 2013).

Population Genetic Structure

Population genomic structure of *Synergus* was investigated using the program ENTROPY (Gompert *et al*., 2014). This hierarchical clustering Bayesian model

incorporates uncertainty in sequencing coverage and error within loci and estimates of allele frequencies to calculate genotype probabilities based on estimated genotype likelihoods. ENTROPY requires only specification of the number of hypothesized ancestral clusters (*k*) without the need for *a priori* assumptions about an individual's assignment probability and produces estimates of genotype probabilities of individuals. We ran ENTROPY models for $k = 2–10$. For each clustering analysis, posterior estimates of genotype probabilities were obtained by running two chains of 75,000 Markov Chain Monte Carlo (MCMC) steps with a 5,000 step-burn in and thinning by retaining every 10th value. MCMC mixing and convergence were checked by estimating effective sample size (ESS) and by examining Gelman-Rubin convergence diagnostics using coda in R (Gelman, & Rubin, 1992; Plummer *et al*., 2006, R Core Team 2017). Mean assignment probabilities (q) were averaged between the two chains run for each *k* model. For a more comprehensive picture of genetic structure, we chose not to select the "best" k, (Evanno *et al*., 2005), but instead present all model solutions across the various *k* values (Gilbert *et al*., 2012; Janes *et al*., 2017; Meirmans, 2015). Mean posterior genotype probabilities were averaged across *k*'s. Principal Component Analysis (PCA) was then performed to visualize the patterns of genomic differentiation among individuals (Price *et al*., 2006) and compared to clusters identified by ENTROPY for *Synergus* and in relation to patterns previously observed between the host *B. treatae* lineage and the host plant. When deep genomic differentiation is detected between lineages as evidenced by the amount of variation explained by PC1, genomic variation among lineages can reflect not only differences in SNP allelic states, but also can arise due to the absence of SNPs and contigs due to restriction site evolution. Therefore, we

performed further analyses for two deeply diverged lineages that were each distributed across two of the three host plants to investigate the hypothesis of HAD within each lineage. We repeated assembly and variant calling for these lineages and then ran ENTROPY for $k = 2$-6 using the same methods as above to allow comparisons between global and fine scale patterns (see appendix for detailed methods). Nei's D_A was then computed to estimate pairwise genetic distances between the *Synergus* lineages detected by ENTROPY across sample sites and to quantitatively investigate how these correspond to the observed morphotypes (Takezaki, & Nei 1996). A dendrogram was then constructed in R to visualize the relationships among the lineages of *Synergus* that were detected (R Core Team 2017).

III. Results

Population Genetic Structure of *Synergus*

We first analyzed genetic structure within *Synergus* from $k = 2$–10 using the data from all 35 sites. From $k = 2$–5 distinct lineages successively emerge representing the nested levels of divergence detected by ENTROPY among lineages (Fig. 3). At $k = 5$, five clearly delineated lineages are present which show no evidence of admixture except for two individuals at site 59. At $k = 6$ and beyond, (Fig. S1), ENTROPY models started to break down as evidenced by Gelman-Rubin scores and low ESS values. Importantly, with respect to testing the hypothesis of HAD within *Synergus*, if HAD is strictly related to the host plant on which development of the inquiline proceeds we would expect to see the three clusters identified by ENTROPY at $k = 3$ to align with the three host plants.

Instead, the distribution of the three clusters appears to be independent of host plant with each cluster appearing across two or three of the host plant species (Fig. 3, row 2).

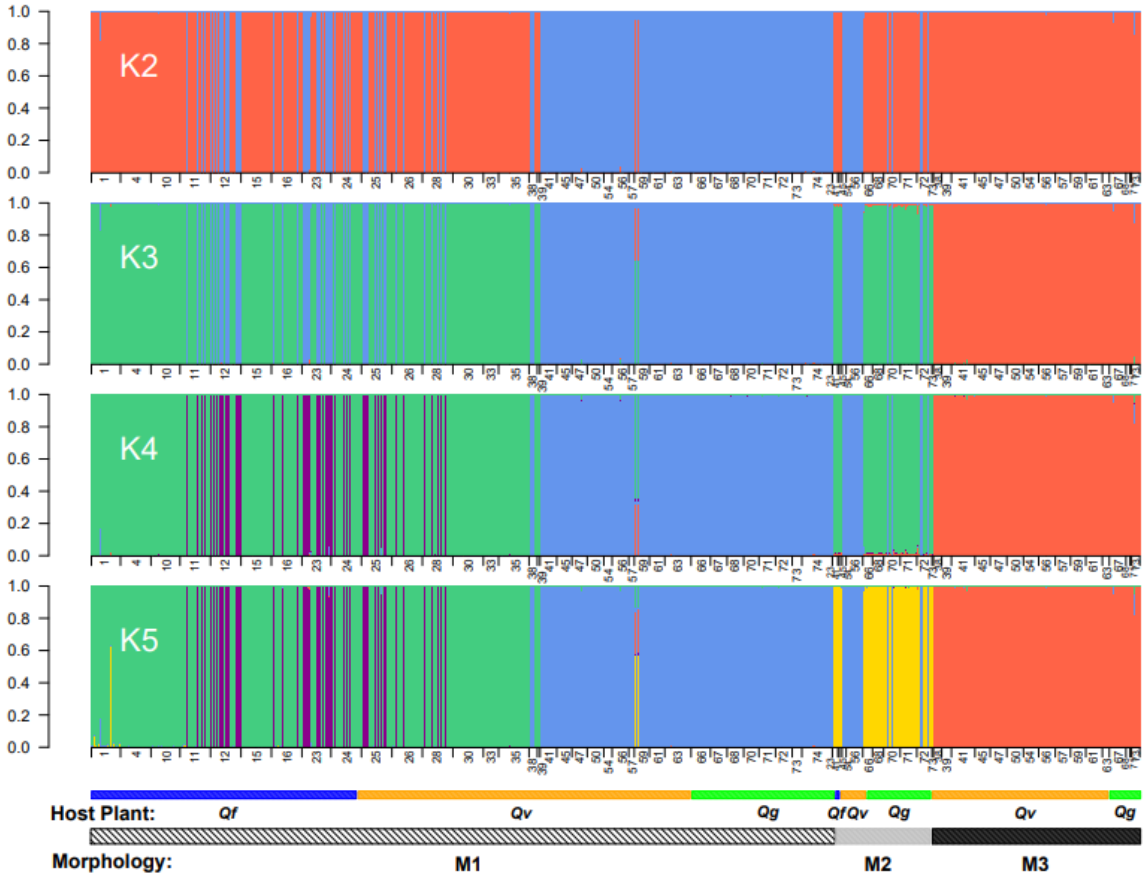

Figure 3. Ancestry proportions (q) based on 758 individuals estimated in ENTROPY (MCMC settings: 75,000 steps, 5,000 burn in and thinning every 10th step; *k* = 2– 5. Each bar corresponds to a sample individual. Colored segments within each bar depict the proportion of an individual's genome inherited from one of *k* inferred source populations (admixture proportion). Individuals are ordered first by a priori morphology, bottom-lowest bar, then host plant, then sample sites arranged from east to west. The five lineages (clusters) detected by ENTROPY are coded M1A (green); M1B (purple); M1C (Blue); M2 (Yellow) and M3 (Red).

At *k* = 3, M1 splits into the two deepest lineages with M1A clustering with M4 while M1B and M1C cluster together. What was initially identified as morph 1 (the

bicolor morph in Fig. 2) is resolved as three genetically distinct but morphologically cryptic lineages at $k = 5$ designated as M1A (green), M1B (purple), and M1C (blue in Fig 3).

Testing for HAD

Mapping the genetic data back onto the collection map shows how the lineages are distributed with respect to host plants (Figure 4). All morphotypes are present on at least two host plants. Lineages M1A and M1B are present to the east of Mississippi, while M1C and M2 are present the west of Mississippi (Fig. 4). This pattern of clustering provides evidence that lineages M1A and M2 shared a common ancestor before some sort of geographical division. M3 is only present to the east of Mississippi.

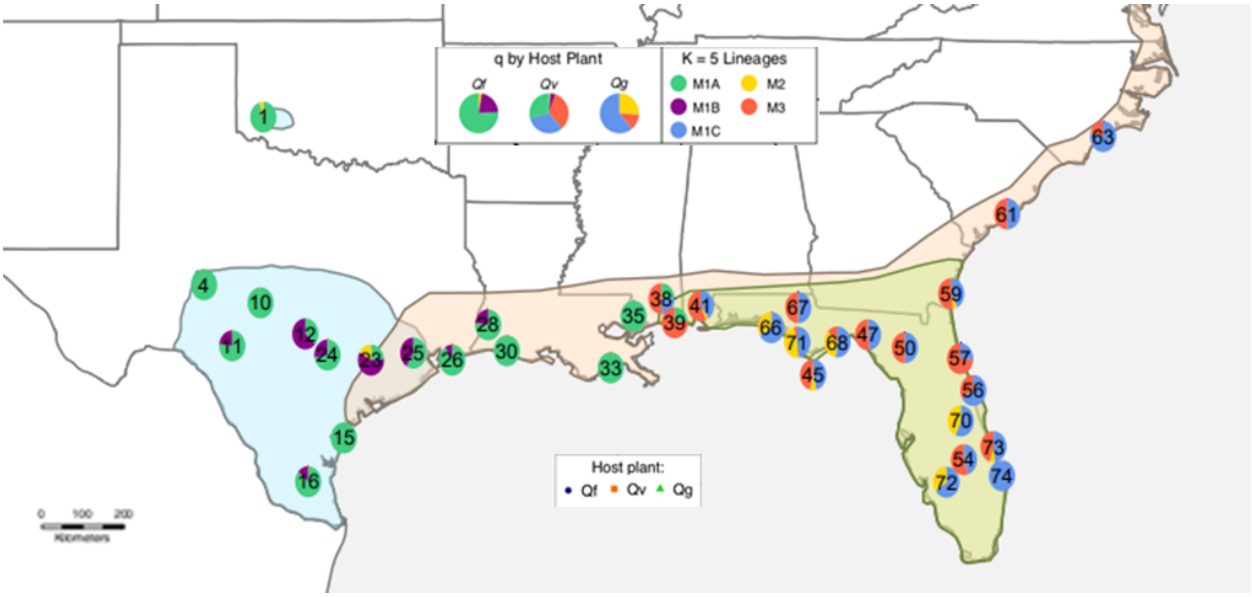

Figure 4. Proportion of individuals sampled at each of the 35 study sites assigning to each of the five lineages detected by ENTROPY. Sequenced individuals were not drawn randomly at every site but rather were drawn so as to include representatives of the a priori distinguishable morphs (M1–3) in the study design. Thus, the proportion of individuals assigning to each morph at each site does not in all cases represent the

relative frequency of each lineage per site. The map does however depict the approximate proportions and approximate geographic distribution of lineages.

To investigate HAD between the host plant and *Synergus*, we used a Principal Component Analysis (PCA) (Figure 5A). The PCA shows evidence of three deeply diverged clusters which are separated by PC1 which accounts for 68.7% of the genomic variation. PC2 accounts for 21.4% of the genomic variation. This PCA shows that there is no clear evidence of HAD. However, the M2 lineage is almost entirely restricted to *B. treatae* galls developing on *Q. geminata*, which might suggest HAD, except a small percentage of these organisms were also collected from *Q. fusiformis*. Figure 5A corroborates that two lineages that were split geographically into (1) M1A and M2, and (2) M1C and M1B—as these morphotypes are collected from different plant species from opposite sides of the longitudinal division in Mississippi.

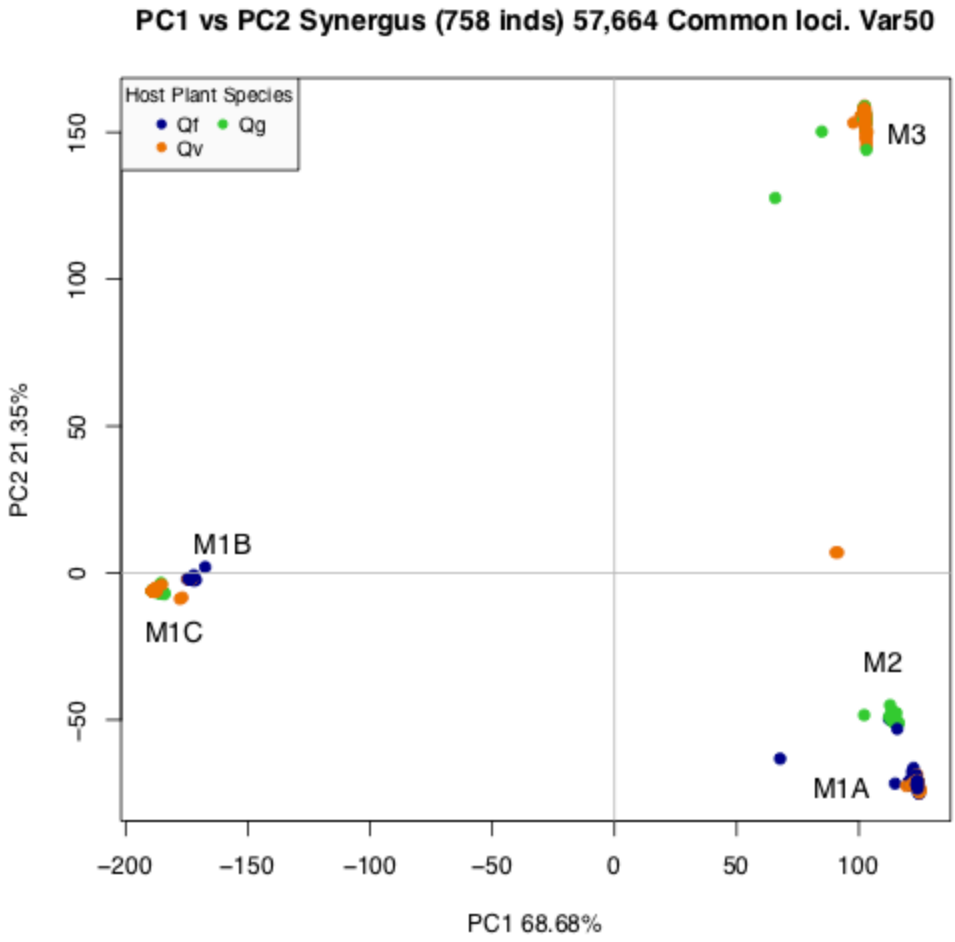

Figure 5A. Principal Component Analysis comparing the genetic structure of the observed five lineages to the host plants that they were collected from.

The ENTROPY and PCA sub analyses we performed on lineages (m1C and m3) allowed us to look at the patterns of genetic variation within each lineage to examine evidence of HAD at a finer scale than the global analysis. If HAD is present, then we predict to see two clusters for $k = 2$ where organisms assign to a cluster corresponding to one or the other alternative host plant. Instead the sub-analysis reveals that morphotype 1C is distributed across *Qv* and *Qg* (Figure S2), and at $k = 2$ the patterns of genetic similarity between small groups of organisms is independent of the host plant. If HAD

occurs between *B. treatae* and *Synergus*, the pattern of differentiation within M1C present at $k = 3$ would appear as three clusters that neatly correspond to the three clusters that were observed for *B. treatae*. Instead, $k = 3$ shows evidence of genetic similarity among nearby collection sites for these organisms.

Figure S2 and the sub-analysis of morphotype three shows a similar pattern: $k = 2$ shows no division based on host plant. What might look like evidence for HAD appears at $k = 3$ as a cluster in *Synergus* corresponds to the *Qv_gulf* lineage identified in *B. treatae*. However, this is likely due to biogeography that drives the differentiation of the host oak, *B. treatae*, and *Synergus*. The *Qv_gulf* lineage from *B. treatae* is most likely the result of a biogeographical division seen across systems in the Southeastern United States, and parallel structure in *Synergus* is the result of this divide across the trophic levels of the system. Genetic similarity is present between nearby geographical sites here as well.

As *Synergus* were collected among the genomic structure of *B. treatae*, we can ask about the patterns of co-differentiation among organisms from the same sites, plants, and geographic regions. If there was evidence of strict HAD, then we would predict genetic clusters in *Synergus* that completely correspond to genetic clusters in *B. treatae*. Instead, *Synergus* show evidence of five clusters that we compared to the structure in *B. treatae* for evidence of HAD (Fig. 5B). M2 almost entirely corresponds to the structure of *B. treatae* evident among *Q. geminata* oaks, however there are also seen on the completely other side of the map on the *Qf_Qf lineage*, which suggests no evidence for HAD between the genetic structure of *B. treatae* and *Synergus*.

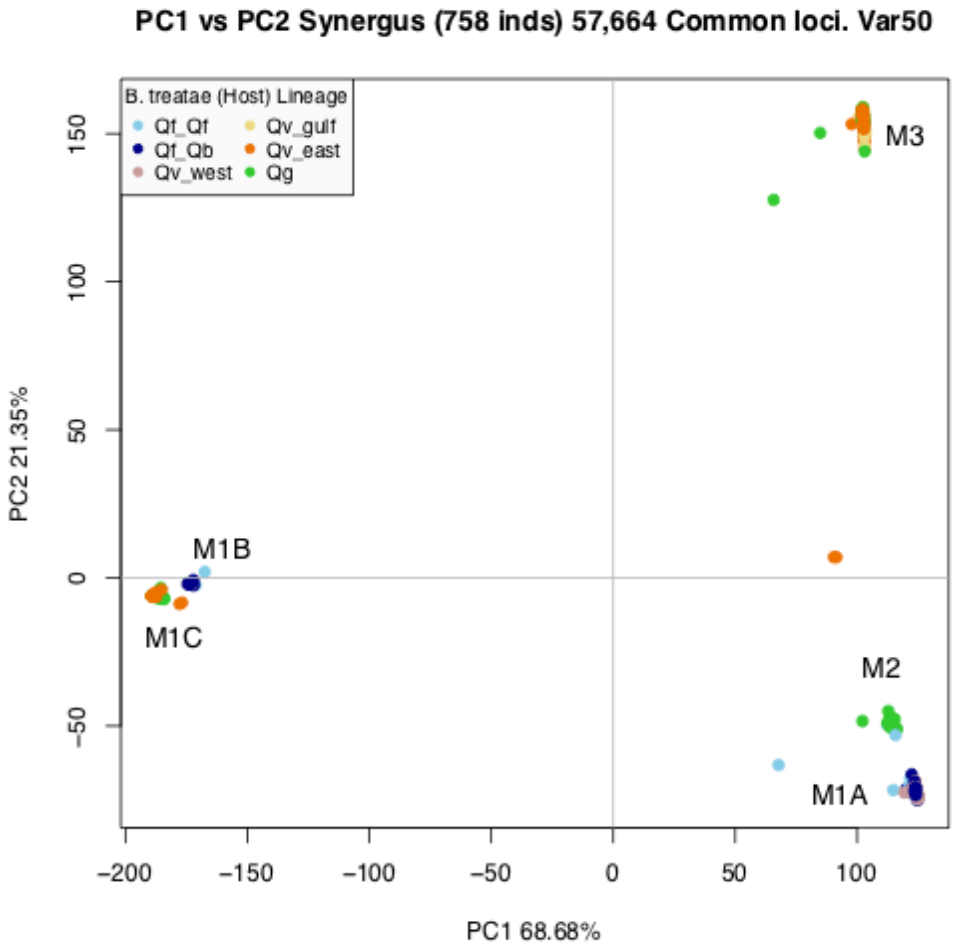

Figure 5B. Principal Component Analysis of the genetic structure of the observed five lineages of *Synergus* in comparison to the six *B. treatae* lineages from which they were sampled.

Genetic Diversity in *Synergus*

We then explored the relationships between the five lineages detected in ENTROPY using a dendrogram of Nei's D_A (Fig 6). This figure shows that M1A and M2 are more genetically similar while M3 forms an outgroup. M1B and M1C form a second genetic pair of lineages that are more similar. This is interesting as M1A, M1B, and M1C are most phenotypically similar yet most genetically dissimilar (Fig. 6).

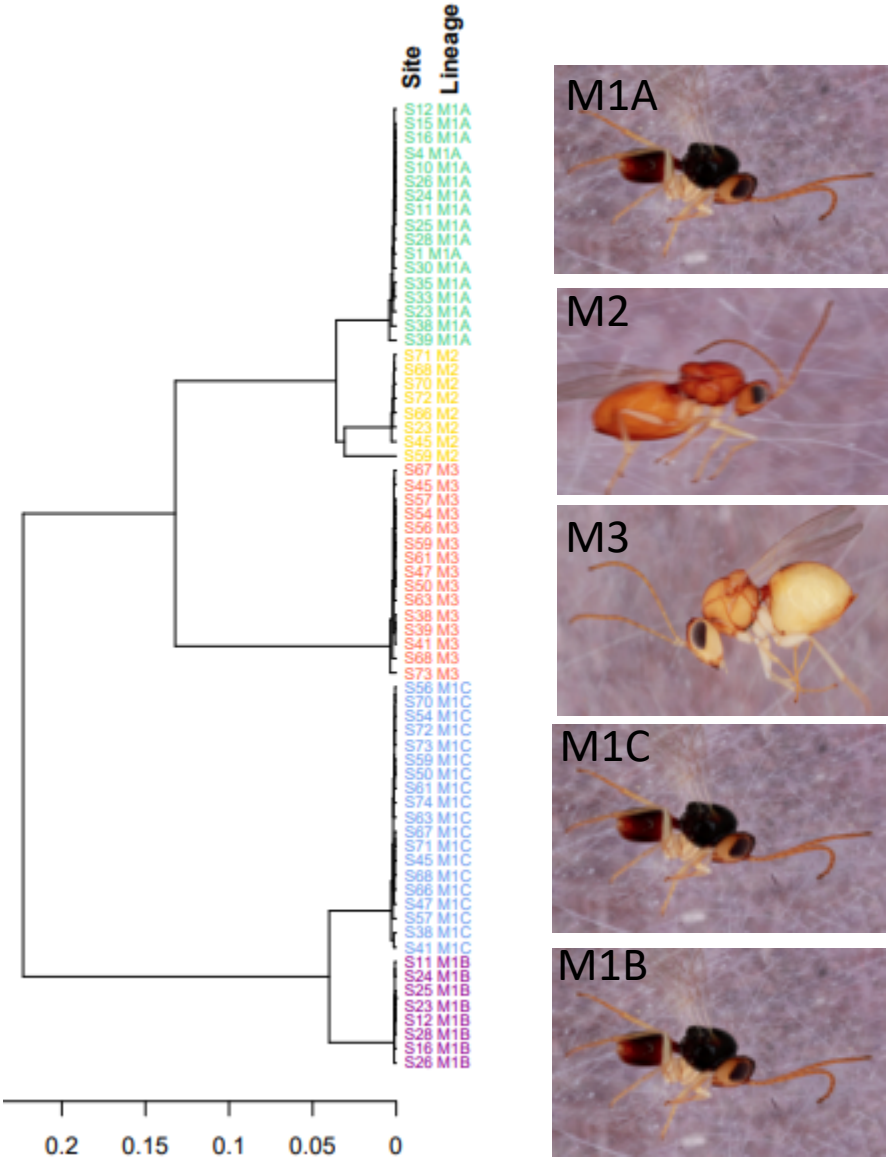

Figure 6. Nei's D_A dendrogram summarizing pairwise genetic distances among *Synergus* sampled from all 35 sites and the five lineages identified by ENTROPY.

Comparing the Nei's D_A of *Synergus* to the Nei's D_A of *B. treatae* presents another way to look for evidence of HAD (Fig. 7). If evidence of HAD is present, morphs within populations that are more genetically similar in *Synergus* would map to populations that are also more genetically similar in *B. treatae*. This pattern does not

occur as M1B and M1A, which are on opposite sides of the dendrogram, and therefore genetically distant, both map to the same lineage in *B. treatae*. The same pattern is present across the dendrogram.

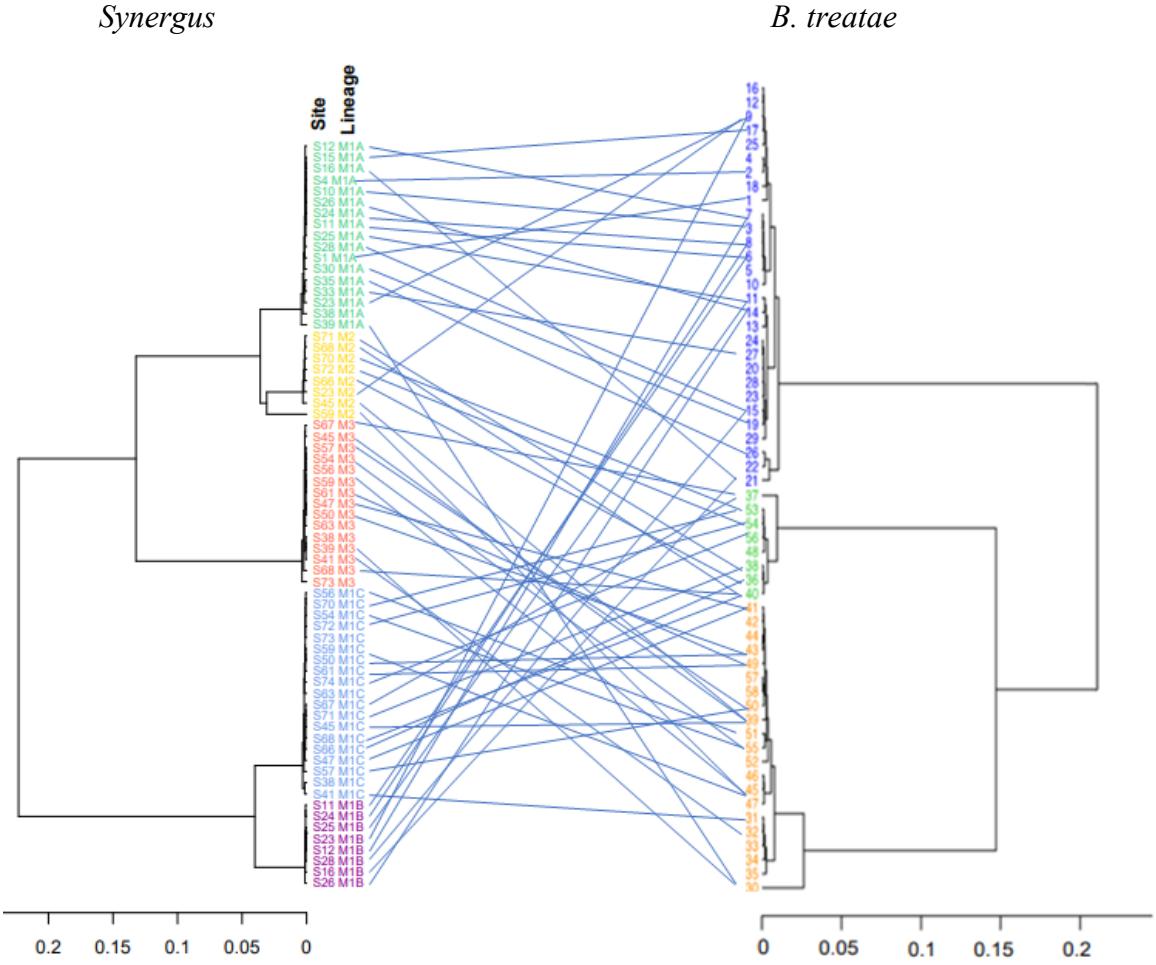

Figure 7. Mapping Nei's D_A dendrogram of *Synergus* to *B. treatae*'s dendrogram. Different site numbers were used for *B. treatae* collection and *Synergus* collection, so each line connects the site that *Synergus* were collected from relative to *B. treatae*.

IV. DISCUSSION

Patterns of Genetic Structure

This study looked for evidence of HAD across trophic levels focusing on the inquiline parasitoid *Synergus sp*. yet found none. This is supported by previous evidence as *Synergus* are a generalist parasitoid, and generalists typically do not demonstrate patterns of HAD (Forbes *et al*., 2017). However, two studies that looked at evidence of HAD in generalist cotton flea hoppers and bird-winged grasshoppers, respectively, found evidence of localized specialization—suggesting a mosaic of specialists that comprise a generalist (Antwi *et al*., 2015; Sword, & Hillis, 2002). We looked for evidence of HAD relative to both the three live oak host plant species that *Synergus* develop on and the influence of genetic substructure with *B. treatae* on the genetic structure of *Synergus*. At neither level was HAD found to be a major driver of genetic variation in *Synergus*. A previous study also had similar findings which looked at *Synergus* in the Western Palearctic (Bihari *et al*., 2011). This is likely because the strength of *Synergus*' association with their hosts is not enough to drive their differentiation.

Figure 4 shows that all lineages identified are present on at least two host plants. M1A and M2 are more genetically similar that are on opposite sides of the central geographic division in Mississippi and M1B and M1C form a second genetically similar pair that are also on opposite sides of the division. This implies two ancestor lineages that were geographically split into the four lineages present here. This longitudinal division was also present in Driscoe *et al*. (2019 in review), which suggests that this division represents a biogeographic barrier that functions to structure genetic variation in this system. Previous biogeographical research on organisms with limited dispersal ability has

also found the same genetic structure corresponding to this break in the Southeastern United States across taxa s(Gonzalez *et al*. 2008, Walker & Avise, 1998).

Based on this longitudinal break, *Synergus* and *B. treatae* share similar evolutionary patterns of genetic differentiation, but this pattern is most parsimoniously attributed to shared biogeographical history rather than parallel HAD in response host plants. This is supported by Figure 7 where the differentiation in *Synergus* doesn't correspond to the differentiation in *B. treatae*. This indicates frequent host shifts in *Synergus* which supports that these organisms use a generalist approach to be viable across a variety of hosts instead of a specific one (Crawley, & Aktheruzzaman, 1988).

These results raise questions about what could lead to the genetic differentiation of *Synergus* beyond the geographical division and the lack of HAD. Further research could investigate how the biological patterns could have driven this genetic structure which would provide insight into other ecological causes of genetic differentiation.

Lineages of *Synergus*

The three morphotypes of *Synergus* originally identified by Busbee (2018) correspond to five genetic lineages. Morph 1 corresponds to three cryptic lineages (genetic clusters) M1A, M1B, and M1C. Of these, M1A and M1B are the most genetically divergent but co-occur in the same on the same host plant in the western range of *Qv* and *Qf* while M1C is present on the opposite side of the geographical divide in the east and is found on *Qv* and *Qg*. M1B and M1C appear to share a common ancestor before geographic division, as do M1A and M2.

Previous research (Forbes *et al*., 2015; Busbee, 2018) found that the distribution of the relative abundances of insect natural enemies of *B. treatae* that emerge from galls induced by *B. treatae* was highly skewed. *Synergus* is by far the most of abundant natural enemy and drives the extreme skewness of the natural enemy community. Our research demonstrates that *Synergus* developing with the galls of *B. treatae* in the United States is, in fact, composed of 5 distinct lineages with multiple lineages present at many sample sites distributed across the three host plant species. Thus, our results will allow the natural enemy community relative abundances estimated by Busbee (2018) to be refined to more accurately reflect the relative composition of the natural enemy community and the geography of variation in the composition of natural enemy communities centered on *B. treatae*.

Most organisms assign neatly to one cluster or another (Fig. 3) indicating the absence of gene exchange among lineages. This raises the question: what leads to the lack of hybrids, especially if there are various lineages that are present on the same plant that look morphologically similar (such as M1A and M1B)? The evidence herein finds that these are independently lineages.

The sub-analyses helped investigate finer scale differentiation within a morphotype which is better suited for this type of analysis, and the results corroborate our previous analyses. Figure S2 depicts the M1C sub-analysis that showed some evidence of genetic similarity between organisms collected from various sites that are geographically closer with decreasing similarity with increased distance, which comports with isolation by distance. At $k = 6$, there is some degree of parallel genetic structure as *Synergus* collected from sites 38 and 41 form a cluster where *B. treatae* also formed a cluster

denoted as the *Qv_ gulf* lineage. However, this is only clear at this higher k's, which suggests that this division is much less significant division than if it appeared at $k = 2$. Based on the predictions of isolation by distance, we would expect to see greater genetic similarity between organisms collected from the same site than from other sites. At a high enough *k*, isolation by distance predicts a cluster for each site if no other processes are present, and this is what the ENTROPY analysis starts to show.

Figure S3 shows a similar pattern of isolation by distance where organisms collected from nearby sites are more similar independently of host plant. This analysis demonstrates parallel structure between *B. treatae* and *Synergus*, but this is most likely due to the geographical division that shaped the trajectory of many organisms in the Southeastern US, which is also evident in the higher degree analysis that we observed. These analyses demonstrate that HAD plays an insignificant role in the structure of *Synergus* even at a more refined level.

This study looked for evidence of host associated differentiation. Instead, we found evidence for a biogeographical driver of differentiation of the structure seen in *Synergus*. The strength of their association with their host was not strong enough to drive the genetic structure of these organisms, which is likely the result of having options of alternative host plants and their generalist strategy.

APPENDIX A

Supplemental Table 1. Sampling localities of *Synergus* in Southeastern US.

Site	*Synergus* site *number*	*B. treatae* site number	Host Plant	Latitude (°N)	Longitude (°W)	# of Individuals Retained N = 758
Quartz Mts, OK	1	1	*Q. fusiformis*	34.89008	-99.3011	22
Irion County, TX	4	2	*Q. fusiformis*	31.21474	-100.842	22
Mason County, TX	10	3	Q. fusiformis	29.87511	-100.109	21
Rocksprings, TX	11	6	*Q. fusiformis*	29.93731	-98.0099	22
Freeman Ranch, TX	12	7	*Q. fusiformis*	27.85438	-97.2105	22
Live Oak Park, TX	15	17	*Q. fusiformis*	27.8543833	-97.210494	22
Encino, TX	16	21	*Q. fusiformis*	26.8941667	-98.135194	22
Altair, TX	23	9	*Q. fusiformis*	29.5625444	-96.504961	25
Luling, TX	24	8	*Q. fusiformis*	29.5625444	-96.504961	22
Rice, TX	25	11	*Q. virginiana*	29.7173889	-95.402278	22
High Island, TX	26	14	*Q. virginiana*	29.5611667	-94.391806	22
Sulpher, LA	28	15	*Q. virginiana*	30.2346111	-93.360639	22
Oak Grove highway, LA	30	19	*Q. virginiana*	29.7668333	-92.975	22
Golden Meadow, LA	33	27	*Q. virginiana*	29.3938889	-90.272861	11
Picayune, MS	35	26	*Q. virginiana*	30.5271444	-89.681253	22
Ocean Springs, MS	38	NA	*Q. virginiana*	30.4100278	-88.755722	12
Gautier, MS	39	30	*Q. virginiana*	30.3802917	-88.610369	9
Dauphin Island, AL	41	32	*Q. virginiana*	30.2504	-88.132525	30
North Highland View, FL	45	39	*Q. virginiana*	29.8382778	-85.316556	24
Perry, FL	47	41	*Q. virginiana*	30.1161	-83.589542	23
High Springs, FL	50	43	*Q. virginiana*	29.8354167	-82.631894	24
Kissimmee River, FL	54	55	*Q. virginiana*	27.3779722	-81.096778	20
Oak Hill, FL	56	51	*Q. virginiana*	28.8957778	-80.854639	36
Palm Coast, FL	57	50	*Q. virginiana*	29.595	-81.195028	15
Jekyll Island, GA	59	45	*Q. virginiana*	31.0174444	-81.429722	22
Charleston, SC	61	49	*Q. virginiana*	32.7687778	-79.973389	22
Topsail, NC	63	NA	*Q. virginiana*	34.4642778	-77.479861	24
Inlet Beach, FL	66	36	*Q. geminata*	30.2743139	-86.003869	21
Parker, FL	67	37	*Q. geminata*	30.1123889	-85.603556	24
Ochlocknee, FL	68	40	*Q. geminata*	29.9600833	-84.385111	23
Lake Lizzie, FL	70	53	*Q. geminata*	28.2276722	-81.179989	23
Oceanside Village, FL	71	38	*Q. geminata*	29.9542222	-85.427722	25
Archbold, FL	72	54	*Q. geminata*	27.1846111	-81.352111	23
Fort Pierce, FL	73	NA	*Q. geminata*	27.4644167	-80.330194	15
Dickinson State Park, FL	74	56	*Q. geminata*	27.0261111	-80.109028	22

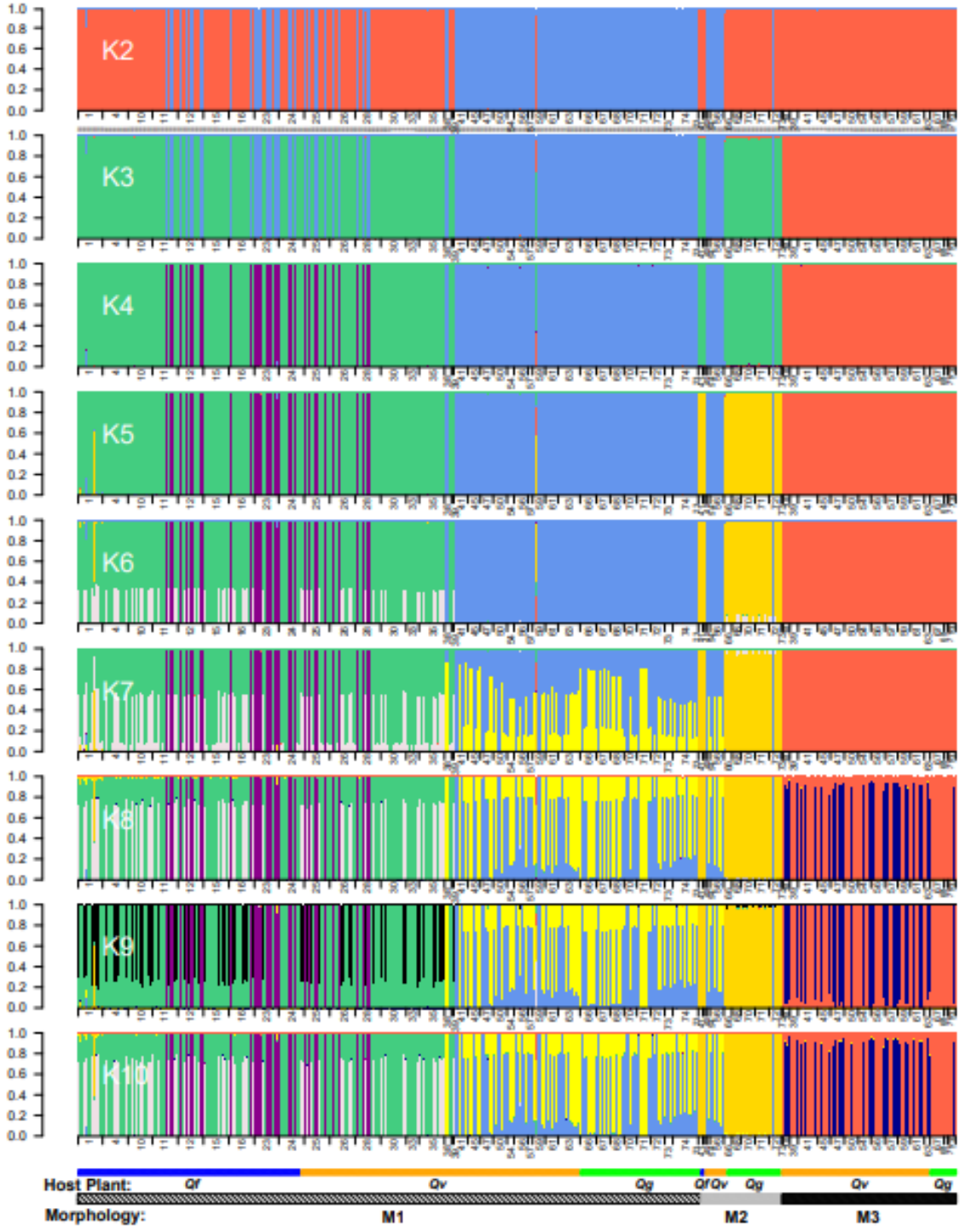

Figure S1. Ancestry proportions (q) based on 758 individuals estimated in ENTROPY (MCMC settings: 75,000 steps, 5,000 burn in and thinning every 10th step; *k* = 2– 10. Each bar corresponds to a sample individual. Colored segments within each bar depict the proportion of an individual's genome inherited from one of *k* inferred source populations (admixture proportion). Individuals are ordered first by a priori

morphology, bottom-lowest bar, then host plant, then sample sites arranged from east to west.

APPENDIX B

Methods for Sub-Analyses

To run the two sub-analyses, individuals were assigned to the clusters based on majority assigned ancestry by ENTROPY at $k = 5$. We then reran the process on each of the two lineages. Individuals from both the M1C and M3 lineages were assembled to their own artificial genomes using the dDocent de novo assembly with minor modifications following the methods of Puritz *et al*. (2014a) and Puritz *et al*. (2014b) using the retained reads from the first analysis of the individuals. For sequences to be considered for inclusion into the reference genome, we required 4 or more reads represented by ≥4 individuals with a minimum of 80% sequence homology. Sequences were assembled to contigs using BWA ver. 0.7.13 (Burrows-Wheeler Aligner, Li, & Durbin, 2009). SAMtools ver. 0.1.19 and BCFtools ver. 0.1.19 (Li *et al*., 2009) were used to index, sort and merge the individual alignments, and then identify bi-allelic SNPs for which at least 50% of all individuals had reads. We used a full prior for variant calling and set the threshold probability for identifying a variant site at P = 0.05. We incorporated genotype uncertainty due to sequencing and alignment errors in downstream analysis by retaining genotype likelihoods (Li, 2011; Skotte *et al*., 2013). SNPs with more than one alternative allele were removed to eliminate potential paralogs. One SNP per fragment was chosen at random and retained for analysis to minimize linkage disequilibrium among SNPs. We then sorted variants by minor allele frequencies (MAF),

retaining SNPs with a MAF > 5%. In total, we retained 232 individuals of the M1C lineage identified 30,427 loci, and 149 individuals from the M3 with 23,136 loci.

We ran ENTROPY models for $k = 2$ through 6. Posterior estimates of genotype probabilities were obtained by running two chains of 15,000 Markov Chain Monte Carlo (MCMC) steps with a 5,000 step-burn in and thinning by retaining every 5th value. MCMC mixing and convergence were checked by estimating effective sample size (ESS) and by examining Gelman-Rubin diagnostics using coda in R. Mean assignment probabilities (q) were averaged between the two chains run for each k model.

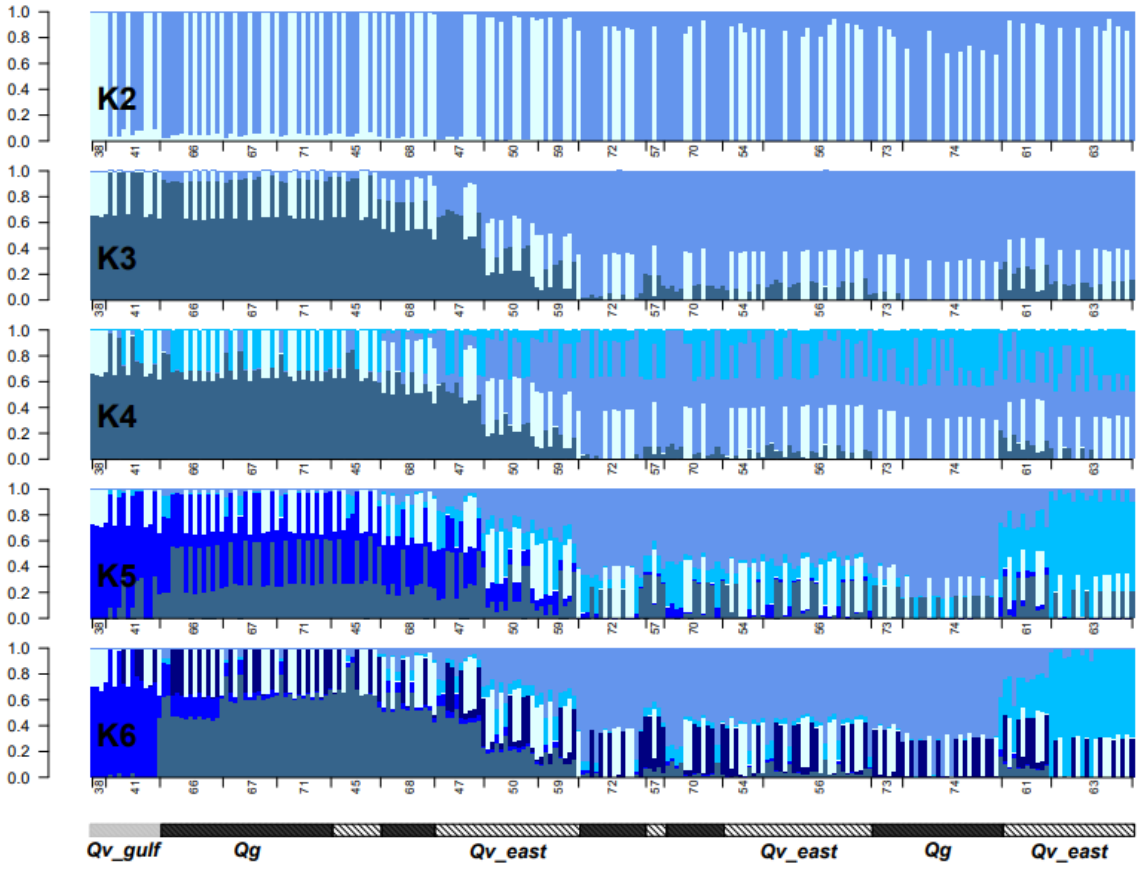

Figure S2. ENTROPY sub-analysis of morphotype 1C. Shown are admixture proportions (q) for k2 – k6 based on analysis of 232 individuals collected and 30,427 loci. Sites are arranged by longitude and labels indicate the corresponding *B. treatae* lineage from the same sample site. Patterns of co-geographic structure appears at $k = 3$.

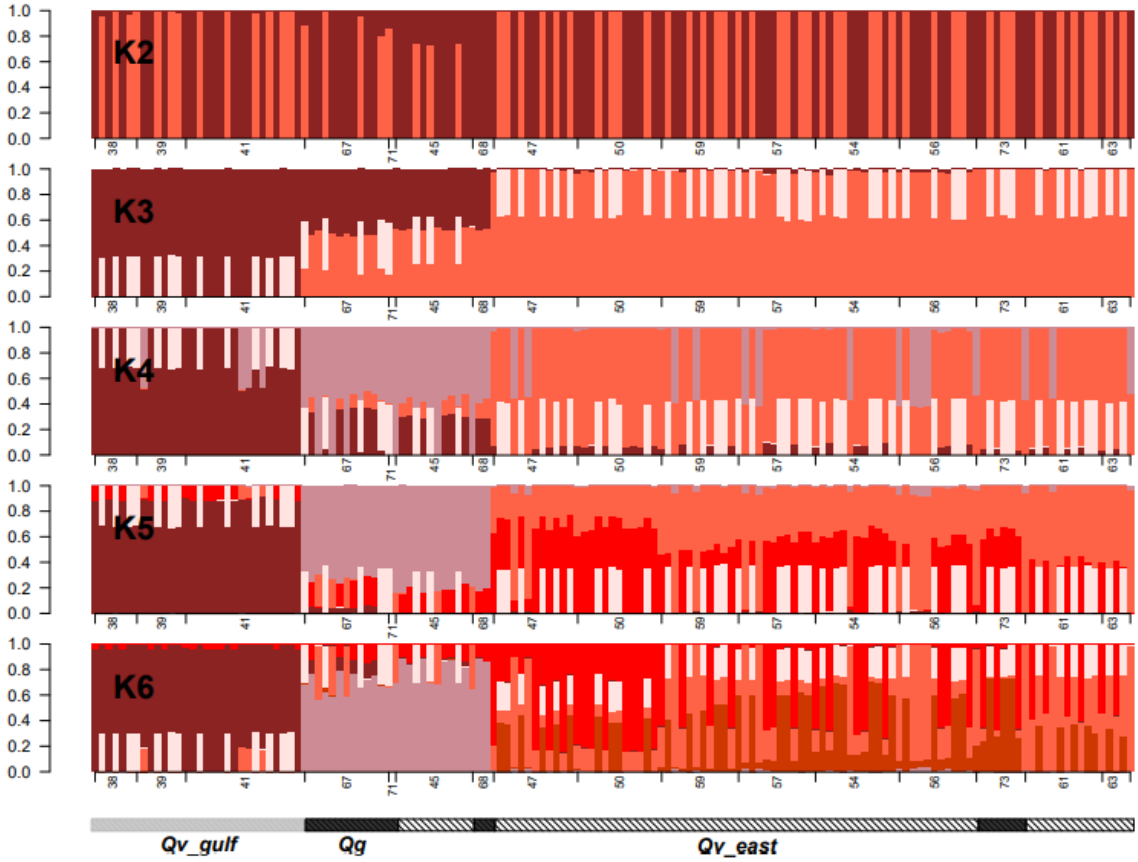

Figure S3. ENTROPY analysis of the Morph 3 subdivision. Shown are admixture proportions (q) for k2 – k4 based on analysis of 149 individuals collected from 23,136 loci. Sites are arranged by longitude and labels indicate the corresponding *B. treatae* lineage from the same sample site. Patterns of co-geographic structure appears at $k = 2$.

REFERENCES

Acs Z, Challis R, Bihari P, Penzes Z, Blaxter M, Hayward A, Melika G, Pujade-Villar J, NIeves-Aldrey JL, Schonrogge K, Stone GN. (2010). Phylogeny and DNA barcoding of inquiline oak gall wasps (Hymenoptera: Cynipidae) of the Western Palaearctic. *Molecular Phylogenetics and Evolution* 55: 210-225.

Althoff, D. M. (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*(17), 3917-3927.

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.

Askew, R. R. (1961). On the biology of the inhabitants of oak galls of Cynipidae (Hymenoptera) in Britain. *Trans Soc Bri Entomol*, *14*, 237-268.

Askew, R. R. (1984). The biology of gall wasps. *Biology of gall insects*, 223-271.

Bailey, R., Schönrogge, K., Cook, J. M., Melika, G., Csóka, G., Thuróczy, C., & Stone, G. N. (2009). Host niches and defensive extended phenotypes structure parasitoid wasp communities. *PLoS biology*, *7*(8).

Barman, A. K., Parajulee, M. N., Sansone, C. G., Suh, C. P. C., & Medina., R. F. (2012). Geographic pattern of host associated differentiation in the cotton fleahopper, *Pseudatomoscelis seriatus. Entomol. Exp. Appl.* 143:31–41

Bernays, E., & Graham, M. (1988). On the evolution of host specificity in phytophagous arthropods. *Ecology*, *69*(4), 886-892.Bush, G. L. (1969). Sympatric host race formation and speciation in frugivorous flies of genus Rhagoletis (Diptera, Tephritidae). *Evolution* 23:237.

Buerkle, C.A, Gompert, Z., (2013). Population genomics based on low coverage sequencing: how low should we go? *Molecular Ecology*, 22:11, 3028–3035.

Busbee, R. W. (2018). *Host plant and spatial influences on the natural enemy community structure of a host specific insect herbivore* (Master's dissertation, Texas State University).Bush, G. L. (1969). Sympatric host race formation and speciation in frugivorous flies of the genus Rhagoletis (Diptera, Tephritidae). *Evolution*, *23*(2), 237-251.

Çabej, N. (2012). *Epigenetic principles of evolution*. Elsevier.

Carton, Y., Poirié, M., & Nappi, A. J. (2008). Insect immune resistance to parasitoids. *Insect Science*, *15*(1), 67-87.

Cook, J. M., & Segar, S. T. (2010). Speciation in fig wasps. *Ecological Entomology*, *35*, 54-66.

Csóka, G., Stone, G. N., & Melika, G. (2005). Biology, ecology, and evolution of gall-inducing Cynipidae. *Biology, ecology, and evolution of gall-inducing arthropods*, *2*, 573-642.

Crawley, M., J. & Aktheruzzaman, M. (1988). Individual variation in the phenology of oak trees and its consequences for herbivorous insects. *Functional Ecology* 2: 409–415.

Dobzhansky T (1951). Genetics and the Origin of Species. Third Edition. New York, NY, USA: Columbia University Press

Dicke, M., & Sabelis, M. W. (1987). How plants obtain predatory mites as bodyguards. *Netherlands journal of zoology*, *38*(2-4), 148-165.

Dopman, E. B., Sword, G. A., & Hillis, D. M. (2002). The importance of the ontogenetic niche in resource-associated divergence: evidence from a generalist grasshopper. *Evolution*, *56*(4), 731-740.

Dres, M., & Mallet, J. (2002). Host races in plant–feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1420), 471-492.

Driscoe, A. L. (2018). Host plant association and spatial autocorrelation as drivers of genetic differentiation among populations of a regionally host-specific insect herbivore (Master's dissertation, Texas State University).

Driscoe A. D., Nice, C. C., Busbee, R. W., Hood, G. R., Egan, S. P, & Ott, J. R. (2019). Host associations and geography interact to shape diversification in a specialist insect herbivore. *Molecular Ecology, in review.*

Evanno G., Regnaut S., & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611–2620.

Forbes, A. A., Powell, T. H. Q., Stelinski, L. L., Smith, J. J., & Feder J. L., (2009). Sequential Sympatric Speciation across Trophic Levels. *Science*, *323*(5915), 776.

Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and plants: a study in coevolution. *Evolution, 18*(4), 586-608

Forbes, A. A., Devine, S. N., Hippee, A. C., Tvedte, E. S., Ward, A. K., Widmayer, H. A., & Wilson, C. J. (2017). Revisiting the particular role of host shifts in initiating insect speciation. *Evolution*, *71*(5), 1126-1137.

Forbes, A. A., Hall, M. C., Lund, J., Hood, G. R., Izen, R., Egan, S. P., & Ott, J. R. (2015). Parasitoids, hyperparasitoids, and inquilines associated with the sexual

and asexual generations of the gall former, Belonocnema treatae (Hymenoptera: Cynipidae). *Annals of the Entomological Society of America*, *109*(1), 49-63.

Forbes, A. A., Powell, T. H., Stelinski, L. L., Smith, J. J., & Feder, J. L. (2009). Sequential sympatric speciation across trophic levels. *Science*, *323*(5915), 776-779.

Fox, L., & Morrow, P. (1981). Specialization: Species Property or Local Phenomenon? Science, 211(4485), 887-893.

Funk, D. J., Filchak, K. E., & Feder, J. L. (2002). Herbivorous insects: model systems for the comparative study of speciation ecology. In *Genetics of Mate Choice: From Sexual Selection to Sexual Isolation* (pp. 251-267). Springer, Dordrecht.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457-472.

Gilbert, K. J., Andrew, R. L., Bock, D. G., Franklin, M. T., Kane, N. C., Moore, J. S., …Vines, T. H. (2012). Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. *Molecular Ecology*, 21, 4925–4930.

Godfray, H.C.J. (1994). Parasitioids: behavioral and evolutionary ecology. *Princeton University Press*

Gonzales, E., Hamrick, J. L., & Chang, S. M. (2008). Identification of glacial refugia in south-eastern North America by phylogeographical analyses of a forest understorey plant, Trillium cuneatum. *Journal of Biogeography*, *35*(5), 844-852.

Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular ecology*, *23*(18), 4555-4573.

Hayward, A., & Stone, G. N. (2005). Oak gall wasp communities: evolution and ecology. *Basic and Applied Ecology*, *6*(5), 435-443.

Hood, G. R., & Ott, J. R. (2010). Developmental plasticity and reduced susceptibility to natural enemies following host plant defoliation in a specialized herbivore. *Oecologia*, *162*(3), 673-683.

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K= 2 conundrum. *Molecular Ecology*, *26*(14), 3594-3602.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987-2993.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, *25*(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Lill, J. T., Marquis, R. J., & Ricklefs, R. E. (2002). Host plants influence parasitism of forest caterpillars. *Nature*, *417*(6885), 170.

Lund, J. N., Ott, J. R., & Lyon, R. J. (1998). Heterogony in Belenocnema treatae Mayr (Hymenoptera: Cynipidae). *Proceedings-Entomological Society of Washington*, *100*, 755-763.

Mandeville, E. G., Parchman, T. L., McDonald, D. B., & Buerkle, C. A. (2015). Highly variable reproductive isolation among pairs of Catostomus species. *Molecular ecology*, *24*(8), 1856-1872.

Mayr E. 1942. Systematics and the origin of species. New York, NY: Columbia University Press.

Mayr, G. L. (1872). *Die Einmiethler der mitteleuropäischen Eichengallen* (No. 16). Selbstverl..Mayr, G. L. (1881). *Die Gerera der gallenbewohnenden Cynipiden...* Alfred Hölder.

Medina, R. F. (2012). Implications of host-associated differentiation in the control of pest species. *Insect outbreaks revisited*, 291-310.

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular ecology*, *24*(13), 3223-3231.

Nosil, P. (2008). Ernst Mayr and the integration of geographic and ecological factors in speciation. *Biological Journal of the Linnean Society*, *95*(1), 26-46.

Nosil, P., Crespi, B. J., & Sandoval, C. P. (2002). Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*, *417*(6887), 440.

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular ecology*, *21*(12), 2991-3005.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news*, *6*(1), 7-11.

Price, P. W. (2002). Resource-driven terrestrial interaction webs. *Ecological Research*, *17*(2), 241-247.

Price, P. W., Bouton, C. E., Gross, P., McPheron, B. A., Thompson, J. N., & Weis, A. E. (1980). Interactions among three trophic levels: influence of plants on interactions

between insect herbivores and natural enemies. *Annual review of Ecology and Systematics*, *11*(1), 41-65.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909

Price, P. W., & Pschorn-Walcher, H. (1988). Are galling insects better protected against parasitoids than exposed feeders?: a test using tenthredinid sawflies. *Ecological Entomology*, *13*(2), 195-205.

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, *2*, e431.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014). Demystifying the RAD fad. *Molecular ecology*, *23*(24), 5937-5942.

Quicke, D. L. (1997). *Parasitic wasps*. Chapman & Hall Ltd.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rosenthal, G. A., & Berenbaum, M. R. (2012). *Herbivores: their interactions with secondary plant metabolites: ecological and evolutionary processes* (Vol. 2). Academic Press.

Rohfritsch, O. (1992). Patterns in gall development, pp. 60-86 *In* J. D. Shorthouse and O. Rohfritsch (eds.), Biology of insect-induced galls. Oxford University Press, New York, NY.

Ronquist, F. (1994). Evolution of Parasitism among Closely Related Species: Phylogenetic Relationships and the Origin of Inquilinism in Gall Wasps (Hymenoptera, Cynipidae). *Evolution,48*(2), 241-266.

Schluter, D. (2000). *The ecology of adaptive radiation*. OUP Oxford.

Schuler, H., Egan, S. P., Hood, G. R., Busbee, R. W., Driscoe, A. L., & Ott, J. R. (2018). Diversity and distribution of Wolbachia in relation to geography, host plant affiliation and life cycle of a heterogonic gall wasp. *BMC evolutionary biology*, *18*(1), 37.

Singer, M. S., & Stireman III, J. O. (2005). The tri-trophic niche concept and adaptive radiation of phytophagous insects. *Ecology Letters*, *8*(12), 1247-1255.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693-702.

Stireman, J.O., Nason, J.D., Heard, S.B., & Seehawer, J.M. (2006) Cascading host-associated genetic differentiation in parasitoids of phytophagous insects. *Proc Biol Sci* 273(1586): 523–530

Stone, G. N., Hernandez-Lopez, A., Nicholls, J. A., Di Pierro, E., Pujade-Villar, J., Melika, G., & Cook, J. M. (2009). Extreme host plant conservatism during at least 20 million years of host plant pursuit by oak gallwasps. *Evolution: International Journal of Organic Evolution*, *63*(4), 854-869.

Stone, G. N., Schonrogge, K., Atkinson, R. J., Bellido, D., & Pujade-Villar, J. (2002). The Population Biology of Oak Gall Wasps (Hymenoptera: Cynipidae). *Annual Review of Entomology*, *47*(1), 633.

Strand, M. R., & Noda, T. (1991). Alterations in the haemocytes of Pseudoplusia includens after parasitism by Microplitis demolitor. *Journal of Insect Physiology*, *37*(11), 839-850.

Strand, M. R., & Pech, L. L. (1995). Immunological basis for compatibility in parasitoid-host relationships. *Annual review of entomology*, *40*(1), 31-56.

Sword, G. A., Joern, A., & Senior, L. B. (2005). Host plant-associated genetic differentiation in the snakeweed grasshopper, Hesperotettix viridis (Orthoptera: Acrididae). *Molecular Ecology*, *14*(7), 2197-2205.

Takezaki, N., & Nei, M. (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, *144*(1), 389–399.

Tuomi, J., Niemelä, P., Chapin, F. S., Bryant, J. P., & Sirén, S. (1988). Defensive responses of trees in relation to their carbon/nutrient balance. In *Mechanisms of woody plant defenses against insects* (pp. 57-72). Springer, New York, NY.

Turelli, M., Barton, N. H., & Coyne, J. A. (2001). Theory and speciation. *Trends in ecology & evolution*, *16*(7), 330-343.

Turlings, T. C., Tumlinson, J. H., & Lewis, W. J. (1990). Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. *Science*, *250*(4985), 1251-1253.

Vamosi, S. M. (2005). On the role of enemies in divergence and diversification of prey: a review and synthesis. *Canadian Journal of Zoology*, *83*(7), 894-910.Via, S. (2002). The Ecological Genetics of Speciation. *The American Naturalist, 159*(S3), S1-S7.

Via, S. (2002). The ecological genetics of speciation. *The American Naturalist*, *159*(S3), S1-S7.

Walker, D., & Avise, J. C. (1998). Principles of phylogeography as illustrated by freshwater and terrestrial turtles in the southeastern United States. *Annual Review of Ecology and Systematics*, *29*(1), 23-58.

Walsh, B. D. (1861). *On phytophagic varieties and phytophagic species*.

Wiebes-Rijks, A. A. (1979). A character analysis of the species *Synergus* Hartig, section II (Mayr, 1872)(Hymenoptera, Cynipidae). *Zoologische Mededelingen*, *53*(28), 297-321.