Spatial and genetic structure of host-associated differentiation in the parasitoid *Copidosoma gelechiae*

C. R. KOLACZAN,* S. B. HEARD,* K. A. SEGRAVES,† D. M. ALTHOFF† & J. D. NASON‡

*Department of Biology, University of New Brunswick, Fredericton, NB, Canada
†Department of Biology, Syracuse University, Syracuse, NY, USA
‡Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA

**Introduction**

Phytophagous insects and their associated parasitoids constitute a large fraction of the Earth’s total biodiversity, and explaining the astounding diversity of phytophagous and parasitic clades is an important challenge for evolutionary ecologists. A large and rapidly growing body of literature has focused on the evolution of new insect diversity via host shifting and subsequent host-associated differentiation (HAD) by both phytophagous insects (e.g. Stireman *et al.*, 2005; Feder & Forbes, 2008; Funk & Nosil, 2008) and their natural enemies (e.g. Jager & Menken, 1994; Kankare *et al.*, 2005; Smith *et al.*, 2006; Stireman *et al.*, 2006; Marussich & Machado, 2007; Abrahamson & Blair, 2008; Forbes *et al.*, 2009). Mounting evidence suggests that HAD can proceed rapidly, perhaps even in sympatry, and has occurred or is occurring in numerous and ecologically diverse insect lineages (Berlocher & Feder, 2002; Dres & Mallet, 2002). However, it remains unclear what ecological factors determine whether an insect lineage associated with multiple hosts undergoes HAD rather than remaining a host generalist, and in particular, how often the occurrence of HAD in a phytophagous insect species results in ‘cascading’ HAD among parasitoids of that species (Stireman *et al.*, 2006).

A powerful model system for the study of HAD among phytophagous insects and parasitoids is provided by the insect community associated with the goldenrods *Solidago altissima* and *Solidago gigantea*. Among the 100+ insect herbivores feeding on one or both goldenrods (Root & Cappuccino, 1992; Fontes *et al.*, 1994) there are at least four cases of HAD leading to the occurrence of HAD in a phytophagous insect species results in ‘cascading’ HAD among parasitoids of that species (Stireman *et al.*, 2006).

**Abstract**

Host-associated differentiation (HAD) appears to be an important driver of diversification in the hyperdiverse phytophagous and parasitoid insects. The gallmaking moth *Gnorimoschema gallaesolidaginis* has undergone HAD on two sympatric goldenrods (*Solidago*), and HAD has also been documented in its parasitoid *Copidosoma gelechiae*, with the intriguing suggestion that differentiation has proceeded independently in multiple populations. We tested this suggestion with analysis of Amplified Fragment Length Polymorphism (AFLP) markers for *C. gelechiae* collections from the midwestern and northeastern United States and eastern Canada. AFLP data were consistent with the existence of HAD, with between-host *F*<sub>ST</sub> significant before Bonferroni correction in two of seven sympatric populations. *Amova* analysis strongly rejected a model of HAD with a single historical origin, and thus supported the repeated-HAD hypothesis. *Copidosoma gelechiae* shows significant host-associated divergence at a number of allozyme loci (Stireman *et al.*, 2006), but only weak evidence via AFLPs for genome-wide differentiation, suggesting that this species is at a very early stage of HAD.

**Keywords:**
AFLPs; allozymes; *Copidosoma*; host-race formation; host specialization; *Solidago*. 

**Correspondence:** Stephen B. Heard, Department of Biology, University of New Brunswick, Fredericton, NB, Canada, E3B 1R3. Tel.: 506 452 6047; fax: 506 453 3583; e-mail sheard@unb.ca
is the most common parasitoid of the gallmaking moth Gnorimoschema gallasolidaginis. Gnorimoschema gallasolidaginis has strongly differentiated host races, or possibly cryptic species, on its two hosts (Nason et al., 2002; Stireman et al., 2005). Stireman et al. (2006) found that collections of C. gelechiae from G. gallasolidaginis larvae on the two goldenrod hosts showed subtle but significant allozyme differentiation. Furthermore, Stireman et al. (2006) suggested that C. gelechiae had undergone HAD repeatedly, with independent differentiation at three widely separated sites (in Minnesota, Ontario and New Brunswick). This suggests that HAD might occur remarkably easily, and so has major implications for our understanding of insect diversification.

Stireman et al.’s (2006) conclusions were limited to three C. gelechiae populations and nine allozyme markers, and therefore our picture of host specialization and HAD in C. gelechiae is incomplete. Here, we greatly expanded sampling both of the species’ geographic range and of its genome, collecting material from 21 sites across a rough east–west transect over 2,000 km long. We generated 940 AFLP (amplified fragment length polymorphism; Vos et al., 1995) markers for our collections. Together, the geographic and genetic scale of our assays allowed us to test for host-associated genetic differentiation across much of the C. gelechiae’s range and across broad regions of its genome. We asked two general questions about geographic and host-associated genetic structure in C. gelechiae. First, we asked whether AFLP markers indicated genome-wide differentiation between collections from S. altissima and S. gigantea, either at locations where Stireman et al. (2006) documented allozyme differentiation or at newly sampled sites. Second, we asked whether any genetic structure present was more consistent with a single historical origin of HAD (origin of a novel host form, followed by its dispersal across the sampled range) or with repeated local origins of HAD. To aid comparison of the results of our AFLP analyses to those obtained by Stireman et al. (2006) using allozymes, we also subjected their allozyme data to several additional analyses. Based on the joint AFLP and allozyme results we suggest that C. gelechiae is in the earliest stages of HAD, and offers an exciting opportunity to study the genetic and geographic structure of nascent biodiversity in an insect parasitoid.

Materials and methods

Study organisms

Copidosoma gelechiae Howard (Hymenoptera: Encyrtidae) is a primary parasitoid of Gnorimoschema gallasolidaginis Riley (Lepidoptera: Gelechiidae), which in turn is a gallmaking herbivore of the two goldenrod species Solidago altissima L. (late goldenrod) and S. gigantea Ait. (tall goldenrod). The goldenrod hosts are closely related members of the S. canadensis complex (Asteraceae: Astereae). They are widespread perennials and are common, and often sympatric, across much of North America (Semple & Cook, 2006). These plants are often abundant in prairies, at forest edges, along roadsides, in abandoned fields and in empty commercial lots.

Gnorimoschema gallasolidaginis is a common univoltine gallmaker on S. altissima and S. gigantea through most of their range. Larvae bore into host-plant stems and induce spindle-shaped galls; pupation begins in mid-summer with emergence occurring around the beginning of September. Gnorimoschema gallasolidaginis populations on S. altissima and S. gigantea are genetically distinct and represent either well-established host races or young cryptic species (Nason et al., 2002; Stireman et al., 2005). Moths on S. gigantea are sometimes treated as G. jecelynæ Miller (Miller, 2000).

Copidosoma gelechiae is the most common primary parasitoid of G. gallasolidaginis (on both goldenrod hosts) across most of its range (S.B. Heard, unpublished). The wasp oviposits in G. solidaginis eggs, but grows most actively late in the larval development of its host. Copidosoma gelechiae is polyembryonic, with up to about 200 larvae arising from a single egg (Patterson, 1915). Although larvae from each C. gelechiae egg are clones, occasionally superparasitism results in a mixed brood (all male, all female, or mixed sex, depending on whether one or both eggs were fertilized). Parasitized G. gallasolidaginis larvae do not pupate, instead continuing to feed and growing significantly larger than unparasitized larvae. Copidosoma gelechiae larvae puate in late summer and adult wasps, which are 0.5–2 mm long, emerge in the last week of August.

Copidosoma spp. can also be found in Gnorimoschema hosts on other host plants, including Gnorimoschema salinaris on Solidago sempervirens. The taxonomic identity of Copidosoma on G. salinaris has not yet been documented. Other Copidosoma also parasitize nongallmaking Lepidoptera; among these is the well-studied Copidosoma floridanum, which parasitizes Trichoplusia ni. The complete host ranges of C. gelechiae and of the genus Copidosoma are unknown.

Collections

In August of 2005, we collected C. gelechiae from sites in the mid-western and northeastern US and two Canadian provinces (Fig. 1), roughly along an east–west transect from Minnesota to New Brunswick. This does not span the entire range of C. gelechiae; although that range is poorly known, we suspect it covers most or all of the range of the host insect G. gallasolidaginis (Miller, 2000). Exact locations for all collection sites are reported in Appendix I. Sites from Stireman et al. (2005) were re-sampled. We preferred sites where we could collect C. gelechiae from Gnorimoschema galls on both goldenrod hosts (henceforth referred to as ‘sympatric sites’), but in order to achieve appropriate geographical sampling, we
also included sites with only one host plant present (or with both plants present, but *Gnorimoschema* galls on only one). We refer to sites of the latter kind as ‘allopatric’ sites, but we acknowledge that there may be individuals of the alternate host within the dispersal range of emerging adults.

At each collection site, we collected and opened *G. gallaesolidaginis* galls, removing the gallmaker larvae and inspecting them for the presence of parasitoid (*C. gelechiae*) pupae. Parasitized larvae were immediately preserved in 95% ethanol. Some gallmaker larvae showed signs of infestation with *C. gelechiae* (much larger body size, reduced mobility and flexibility), but did not yet show visual evidence of *C. gelechiae* pupae. These larvae were held alive until *C. gelechiae* pupae developed and were then preserved in 95% ethanol. Preserved samples were stored at ambient temperatures (20–30°C) until DNA extraction. Representatives of two additional *Copidosoma* species were also obtained to serve as outgroups in phylogeographic analyses. We collected *Copidosoma* sp. from *Gnorimoschema salinaris* larvae galling *Solidago sempervirens* (from Ninagret Conservation Area, RI, USA). Preserved *C. floridanum* were obtained from Dr. Michael R. Strand (University of Georgia).

**DNA extraction, AFLP reactions, and electrophoresis**

We extracted DNA using Qiagen® (Valencia, CA, USA) DNeasy blood and tissue DNA extraction kits. A single *C. gelechiae* pupa was dissected from each host larva, in order to eliminate the possibility of combining genetic material from nonclones in the case of a mixed brood. Samples were air dried for 15 min and then ground in liquid nitrogen. DNA extraction followed the kit protocol, except that samples were incubated at 55°C overnight, eluted just once with 25 μL of elution buffer AE, and the elution buffer was allowed to sit on the membrane for 5 min before the final centrifugation. An aliquot from each sample was run on a 1.5% agarose electrophoresis gel and stained with ethidium bromide to determine DNA quality. Samples showing fragmented or degraded DNA were re-extracted.

Procedures for AFLP analysis were based on those of Vos et al. (1995) following the modifications of Segraves & Pellmyr (2004). Genomic DNA was cut with the *Eco*RI and *Mse*I restriction endonucleases and then adapter sequences were ligated onto the fragments with T4 DNA ligase (Promega, Madison, WI, USA). Preselective amplification was conducted using the primers *Mse*+C and *Eco*A, and a second selective amplification step was performed using six primer combinations in two multiplex reactions (*Mse*+CAC or *Mse*+CTT with each of *Eco*AAG, *Eco*+AAC and *Eco*+ACA; see Segraves & Pellmyr, 2004). *Eco*+3 primers were labelled with one of three ‘lightsaber’ dyes (black/blue/green; Synthegen, Houston, TX, USA). Fragment lengths for the resulting labelled PCR products were determined via electrophoresis on a Beckman Coulter (Fullerton, CA, USA) CEQ6000 sequencer.

**AFLP data analysis**

Raw fragment length data were filtered with software developed by Abdo et al. (2006). This software identifies fluorescence peaks associated with AFLP fragments automatically, enabling quick and consistent scoring of
large data sets. The filtering algorithm first standardizes
the dataset by calculating relative peak areas. Next, the
standard deviation of peak areas is calculated. Any peaks
more than three standard deviations above the mean
fluorescence are scored as representing true AFLP frag-
ments and are excluded from further calculations. This
process is iterated until no more peaks can be scored, and
any remaining fluorescence peaks are considered as
background noise (Abdo et al., 2006).

Amplified Fragment Length Polymorphism fragment
data were used in three complementary analyses. First, we
conducted analyses of molecular variance (amovas)
using Arlequin® V 2.0 (Schneider et al. 2000). Amova
analyses apportion genetic variation to levels of hierar-
chical organization – in this case, host affiliation and site.
We compared amova analyses specifying sites nested
within host affiliations and specifying host affiliations
nested within sites. The former hierarchical structure is
appropriate if a single HAD event was followed by
dispersal of the novel host race, whereas the latter
structure is appropriate either if HAD proceeded inde-
pendently in each population or if there is no host-
associated differentiation, but some geographical differ-
entiation. We compared the explanatory power of the
alternative models using the corrected Akaike Informa-
tion Criterion, following calculations outlined in Halver-
son et al. (2008). The Amova analysis used data for seven
sympatric sites. Collections from Ft. Wayne, IN, included
only two broods from S. altissima and were excluded from
the Amova analyses.

Second, we calculated $F_{ST}$ values between C. gelechiae
collections from S. altissima and S. gigantea at the same
seven sympatric sites used in the Amova, using AFLP-
Surv (Vekemans et al., 2002). Estimates of allelic frequen-
cies were computed using the ‘Bayesian method with
nonuniform prior distribution of allelic frequencies’
(Zhivotovsky, 1999), and Hardy–Weinberg genotypic
proportions were assumed. $F_{ST}$ values significantly differ-
ent from zero would identify populations with barriers to
gene flow between host-associated types. In order to
reduce inflation of type I error when performing these
tests on multiple locations, we checked $P$-values for signifi-
cance following sequential Bonferroni correction. This
procedure has been criticized as being too conservative (Moran,
2003; Nakagawa, 2004) so cautious interpretation of these
results would be appropriate. As an alternative way to test
for significant local differentiation, we also used Amova to
calculate the between-hosts genetic variance component
$V_o$ using permutation (1000 replicates) to test for $V_o$ values
significantly greater than zero. The results of this proce-
dure differed only slightly from the tests using $F_{ST}$, and so
are not reported.

Third, as a way to visualize phylogeographic structure
in our data we calculated Nei-Li genetic distances among
pairs of populations using AFLP-Surv (Vekemans et al.,
2002). We then used the NEIGHBOR program of Phylip
3.67 (Felsenstein, 2005) to create a neighbour-joining
tree based on the Nei-Li distances. This analysis included
data from all our collections, both sympatric and allo-
patric, although one using only sympatric collections
(not shown) leads to similar interpretation. Analyses
using the two alternative outgroups generally resulted in
identical ingroup topologies, so we report only results
using Copidosoma from S. salinaris. Under the single-HAD-
event hypothesis, we would expect populations from one
of the two hosts to cluster together (e.g. Stireman et al.,
2005, tree for G. gallaesolidaginis); in contrast, under the
multiple-HAD-events hypothesis, we would expect popu-
lations from the two hosts at each site to cluster. We
calculated bootstrap support for each node as a measure of
confidence in tree structure.

Allozyme data analysis

Sampling and allozyme genotyping of C. gelechiae popula-
tions are described in Stireman et al. (2006) for sym-
patric collections from S. altissima and S. gigantea at Fredericton
(nine polymorphic loci), Milaca (eight polymorp-
ghic loci) and Toronto (seven polymorphic loci). For
each sympatric site, Stireman et al. (2006) tested for
significant differentiation at individual loci using exact
genic tests (Raymond and Rousset 1995) and over all loci
using an exact genotypic test (Goudet et al., 1996),
finding significant genic tests for a subset of loci (different
subsets at each site) and a significant genotypic test at
each site. For more direct comparison to our AFLP
analyses, we used Stireman et al.’s (2006) allozyme data
to estimate $F_{ST}$ (and 95% bootstrap confidence limits)
between sympatric populations using Weir & Cocker-
ham’s (1984) $\theta$ as calculated by the program GDA (Lewis
& Zaykin, 1999). Stireman et al. (2006) reported site-
specific estimates of $F_{ST}$ but not their statistical signifi-
cance.

Stireman et al.’s (2006) finding that genic tests were
significant for different subsets of loci at different sites
indicates HAD at multiple genomic regions and suggests
that overall significance was not the result of selection
afflicting genetic variation at a few specific loci or closely
linked genomic regions. To more directly evaluate the
potential for selection to explain differentiation at indi-
vidual allozyme loci, we subjected $F_{ST}$ estimates for each
locus and at each site to the Lewontin & Krakauer (1973)
test of neutrality, as advocated in several recent papers
(Vitalis et al., 2003; Whitlock, 2008). Because the L–K
test uses the expected neutral distribution of single-locus
$F_{ST}$ values in the range 0–1, we constrained any negative
variance components obtained from GDA to 0 to obtain,
for each site, single- and multi-locus estimates of $F_{ST}$ in
the range 0–1. Using the constrained multi-locus $F_{ST}$ as
the mean $F_{ST}$ value in the L–K test (with 1 degree of
freedom), for each site we rejected neutrality for an
individual locus if its estimated $F_{ST}$ was greater than or
equal to the upper 95% critical value of the expected
neutral distribution. Because we employed the L–K test
AIC analysis indicating essentially no support for model (a) among sites and host plants. Model (b) is strongly favoured, with \( \Delta AIC_c = 9.5 \). This analysis suggests that while there is relatively little structure in the genetic data, the structure present is best accommodated by a model grouping collections from the two host plants together within sites.

### Local \( F_{ST} \)s

Amplified Fragment Length Polymorphism genetic differentiation between sympatric \( C. \) gelechiae collections from \( S. \) altissima and \( S. \) gigantea was generally subtle (Table 2), with an average \( F_{ST} \) of 0.012. Before Bonferroni correction, \( F_{ST} \) values were significantly greater than zero at two sites (Missisquoi, \( F_{ST} = 0.043 \); and Pellston, \( F_{ST} = 0.028 \)). Following Bonferroni correction, none of the seven \( F_{ST} \) values was significant, although the highly conservative nature of the correction makes interpretation of this result difficult. For the sites re-sampled from Stireman et al. (2006); Fredericton, Milaca, and Toronto), our nonsignificant AFLP estimates of \( F_{ST} \) do not confirm the significant differentiation observed for allozyme markers using genic and genotypic exact tests. However, allozyme estimates of \( F_{ST} \) for these sites were also less conclusive than the genic and genotypic tests, with \( F_{ST} \) significantly greater than zero at only one site (Toronto, \( F_{ST} = 0.0510 \) with 95% limits 0.0060–0.0677).

### Phylogeography

The neighbour-joining tree of all \( C. \) gelechiae populations (Fig. 2) shows no strong grouping based on either sites or host affiliations. We were not able to infer with confidence the relationships between populations (bootstrap support for all nodes < 70%; data not shown). This lack of recognizable genetic structure was surprising and led to some concern that the AFLP data were too noisy or provided a negative variance component for host plants (Table 1A), and the alternative model nesting host plants within sites was strongly favoured (\( \Delta AIC_c = 9.5 \)).
otherwise unreliable. However, when analyzed as individuals (UPGMA tree not shown) the outgroup individuals clustered together, whereas individuals from other populations were intermixed. This suggests that our markers can reveal population structure, but that AFLP differentiation between sympatric collections from *S. altissima* and *S. gigantea* is absent or extremely subtle at this (genetic) scale.

### L–K tests of allozyme neutrality

Lewontin–Krakauer tests of single-locus $F_{ST}$ estimates (Table 3) were significant for one of nine polymorphic alloyme loci at Fredericton (PGI $F_{ST} = 0.0570$, $P = 0.016$) and one of eight polymorphic loci at Milaca ($ACOH F_{ST} = 0.0727$, $P = 0.040$). No tests were significant after sequential Bonferroni correction for multiple tests within sympatric sites.

#### Table 3

<table>
<thead>
<tr>
<th>Locus</th>
<th>Fredericton</th>
<th>Milaca</th>
<th>Toronto</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_{ST}$</td>
<td>$P$</td>
<td>$F_{ST}$</td>
</tr>
<tr>
<td>ACOH</td>
<td>0.0237</td>
<td>0.120</td>
<td>0.0727</td>
</tr>
<tr>
<td>G3PDH</td>
<td>0.0000</td>
<td>1.000</td>
<td>0.0000</td>
</tr>
<tr>
<td>HADH3</td>
<td>0.0000</td>
<td>1.000</td>
<td>0.0254</td>
</tr>
<tr>
<td>HK</td>
<td>0.0082</td>
<td>0.362</td>
<td>0.0085</td>
</tr>
<tr>
<td>IDH</td>
<td>0.0157</td>
<td>0.207</td>
<td>0.0588</td>
</tr>
<tr>
<td>LDH</td>
<td>0.0000</td>
<td>1.000</td>
<td>0.0000</td>
</tr>
<tr>
<td>MDH</td>
<td>0.0000</td>
<td>1.000</td>
<td>0.0343</td>
</tr>
<tr>
<td>PGI</td>
<td>0.0570</td>
<td>0.016</td>
<td>0.0000</td>
</tr>
<tr>
<td>PGM</td>
<td>0.0161</td>
<td>0.200</td>
<td>0.0054</td>
</tr>
</tbody>
</table>

### Discussion

Our AFLP results provide at most weak evidence of host-associated divergence (HAD) at the whole-genome scale in *C. gelechiae*, but in combination with evidence of HAD from allozymes (Stireman et al., 2006) they do shed useful light on the geographic structure of differentiation. Our AMOVA analysis strongly favours a model of genetic structure with host affiliation nested within sites (Table 1; as did Stireman et al.’s (2006) similar analysis). This is clearly inconsistent with a model of HAD involving a single host shift followed by dispersal of a novel host form – that is, the model that was supported for *C. gelechiae*’s host, *G. gallaesolidaginis* (Nason et al., 2002; Stireman et al., 2005). Instead, it is consistent either with the absence of HAD but the presence of some geographic differentiation, or with the occurrence of HAD independently at multiple sites (as argued by Stireman et al., 2006). These two possibilities are best separated by site-by-site analysis of reproductive isolation (Table 2). The results of such analyses, calculating Wright’s $F_{ST}$ from our AFLP data, are suggestive of HAD at Missiquoi and Pellston ($F_{ST}$ significant before, but not after, Bonferroni correction) but provide no evidence for genome-wide differentiation at the other sites.

![Fig. 2](image-url)

Estimated phylogeographic relationships among sampled *Copidosoma gelechiae* populations. Tree was constructed via neighbour-joining based on Nei-Li distances (indicated as scale bar). Bootstrap support for all nodes < 70%. The outgroup (at which the root is forced) is *Copidosoma* sp. collected from *Solidago salinaris*. Filled circles indicate collections from *S. altissima* and open circles indicate collections from *S. gigantea*.
These AFLP results contrast with Stireman et al.’s (2006) analysis of differentiation at three sympatric sites based on allozyme markers (7–9 polymorphic loci/site). Stireman et al. (2006) found strong evidence for HAD in C. gelechiae at all sites using both genic and genotypic exact tests. This, along with AMOVA analyses favouring grouping of populations by site and then by host affiliation, suggested that C. gelechiae had undergone HAD, and that differentiation had occurred independently at least three times. We expected that our $F_{ST}$-based analyses of AFLP and allozyme markers would confirm the occurrence of local differentiation, but they do so only weakly: when assessed using $F_{ST}$, differentiation based on AFLP markers was nonsignificant at all three of Stireman et al.’s (2006) sites whereas allozyme differentiation was significant only at one of them (Table 2). The apparent conflict between the strong allozyme results reported by Stireman et al. (2006) and the weak AFLP and allozyme results of this study has at least three potential explanations: the differences might be attributable to (i) the nature of expected differentiation during early stages of HAD, (ii) issues of statistical power arising from the inheritance and sampling of the two classes of genetic markers, or (iii) issues of power arising from the different analyses of differentiation used.

First, discrepancies between allozyme and AFLP-based measures of differentiation might not be unexpected at early stages of HAD, when divergence may often be driven by strong selective pressures and consequent changes at a relatively small number of loci (Coyne & Orr, 2004; Gavrilets, 2006) – loci that encode phenotypically important information, or are closely linked to other loci that do. These few divergent regions of the genome will be surrounded by large areas of neutral or nearly neutral DNA (coding and noncoding) that have not undergone sufficient divergence to be easily distinguished. Over time, and once a sufficient level of reproductive isolation has evolved, divergence is expected to spread across the rest of the genome (Coyne & Orr, 2004), including loci under weak or no selection and including noncoding regions (which are sampled by AFLP but not by allozyme markers). Because allozymes reflect the genetic variation only of coding regions, it is possible that the relatively strong differentiation revealed by the exact genic and genotypic tests of Stireman et al. (2006) reflect greater average effects of selection on allozyme vs. AFLP loci. However, for two reasons this seems unlikely to be the case for C. gelechiae. First, Stireman et al.’s genic tests were significant for different subsets of loci at different sites, suggesting that if selection is operating on these loci it is doing so inconsistently across geographic locations. Second, Lewontin and Krakauer tests conducted here on Stireman et al.’s (2006) allozyme data revealed only two potential cases of selection (Table 3), neither of which was significant after correction for multiple tests.

Second, the difference between studies may be attributable to differences in the markers employed and the sampling of genetic information. Simulations by Gouyet et al. (1996) indicate that when genetic differentiation between populations is low, as is the case for host-associated populations of C. gelechiae, the power to detect differentiation using $F_{ST}$ benefits strongly from sampling many individuals per populations. They suggest that getting better estimates of local allele frequencies increases the power of the test to detect small differences in allele frequencies. While we assayed a large number of AFLP loci, the number of individuals sampled per host-associated population was modest (4–22; in accord with many AFLP studies). In contrast, Stireman et al. (2006) assayed far fewer allozyme loci but sampled 44–71 individuals per population. Furthermore, estimates of local allele frequencies can be affected by the mode of inheritance of the marker system. While allele frequencies at codominant loci (such as allozymes) can be estimated without bias, allele frequency estimates at dominant loci (such as AFLPs), can be biased in small populations (Lynch & Milligan, 1994). Still, these observations are unlikely to entirely explain the observed differences between AFLP and allozyme loci for HAD in C. gelechiae. One reason is that AFLP $F_{ST}$ estimates were highest and significant (before sequential Bonferroni correction) not at the sites with the largest sample sizes (Capac and Fredericon) but at two sites (Missisquoi and Pellston) with barely half as many sample individuals (Table 2). Another reason is that AFLP and allozyme estimates of $F_{ST}$ were in fact quite similar (Table 2). Therefore, inherent differences between AFLP and allozyme markers are unlikely to explain the apparent conflict between our results and those of Stireman et al. (2006).

The third, and most likely, explanation for the contrast between our results and those of Stireman et al. (2006) is a difference in analyses of differentiation. In this study, we focussed on $F_{ST}$ analyses amenable to both AFLPs and allozymes, but Stireman et al. (2006) used exact genic and genotypic tests of differentiation using allozymes (such tests cannot be applied to dominant markers such as AFLPs). Gouyet et al. (1996) tested the power of exact allelic and genotypic tests and $F_{ST}$-estimator tests of population differentiation, and found them to have similar power for balanced data sets. Therefore, we expected to be able to expand on the well-supported allozyme evidence of geographically independent HAD reported by Stireman et al. (2006) with our $F_{ST}$-estimator tests of differentiation using AFLPs and allozymes. Instead, whether employing AFLP or allozyme data, our $F_{ST}$-based tests appear to have less power to detect differentiation than Stireman et al.’s (2006) exact allelic and genotypic tests. The reason for this is unclear but may be related to $F_{ST}$-estimator tests being less sensitive than the exact tests to variation in the frequency of low frequency alleles.
The wide geographic scope of our study (Fig. 1) was designed to test Stireman et al.’s (2006) hypothesis that HAD in C. gelechiae had proceeded independently at the three sites they sampled. Although AFLP markers provided only weak evidence for HAD itself, our AMOVA analysis strongly rejects a hypothesis of HAD with a single historical origin. As the existence of HAD is strongly supported by Stireman et al.’s (2006) allozyme data, we interpret our AFLP results as supporting the independent-origins hypothesis. Intriguingly, our results even suggest that HAD may have begun independently at more sites than the three studied by Stireman et al. (2006). The two sympatric sites at which we found weak evidence of HAD lie midway between pairs of sites from Stireman et al., 2006 (Pelliston, MI between Milaca and Toronto, and Missisquoi, VT between Toronto and Fredericton; Fig. 1), and in neither case do we see any evidence that host-associated collections at our new sites group with collections from the same hosts at the sites of Stireman et al. (2006) (Fig. 2). A high priority for future research will be to re-sample the Pelliston and Mississquoi sites with much larger sample sizes, in order to examine host-associated divergence at the allozyme loci that were informative at other sites for Stireman et al. (2006).

Our emerging picture of host evolution in C. gelechiae is one of subtle differentiation in a species at the very earliest stages of HAD. This offers an exciting opportunity, because the same pair of host plants supports both herbivores (Stireman et al., 2005) and parasitoids (Stireman et al., 2006) at stages of differentiation ranging from host races to cryptic host-specialist species. Differentiation in C. gelechiae thus offers a view of the earliest stages of a process that can be studied in its later stages via other insect species in the same ecological setting. Taken together, studies to date of the goldenrod-insect system (e.g. Abrahamson & Weis, 1997, Stireman et al., 2005, 2006; Abrahamson & Blair, 2008) paint a picture of repeated origin of new insect diversity through HAD via processes that are complex and include interesting variation in extent, timing and mechanism.

Acknowledgments

We thank Linley Jesson, Judy Loo, Dan Quiring, Gary Saunders and two anonymous reviewers for comments on the manuscript. John Semple provided advice on Solidago systematics and the outline map used for Fig. 1. The City of Fredericton, Tommy Thompson Park, Rhode Island Department of Environmental Management and several private landowners graciously permitted us to collect galls on their properties. Ramesh Raina provided assistance with and access to the Beckman sequencing facility. This work was supported by grants to SBH from the NSF (USA; DEB-0107752) and NSERC (Discovery Grant), to JDN from the NSF (DEB-0107938) and to DMA from the NSF (DEB-032129).

References


Schneider, S., Roessli, D. & Excoffier, L. 2000. *Arlequin*: a software for population genetics data analysis. Ver. 2.000. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.


**Appendix I. Collection locations**

<table>
<thead>
<tr>
<th>Site name</th>
<th>State/ Province</th>
<th>No. broods from <em>Solidago altissima</em></th>
<th>No. broods from <em>Solidago gigantea</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Akron</td>
<td>OH</td>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>Brimfield</td>
<td>OH</td>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>Brownville</td>
<td>NY</td>
<td>6</td>
<td>–</td>
</tr>
<tr>
<td>Calcium</td>
<td>NY</td>
<td>9</td>
<td>–</td>
</tr>
<tr>
<td>Capac</td>
<td>MI</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>Casnovia</td>
<td>MI</td>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>Deerfield</td>
<td>WI</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Diamond</td>
<td>OH</td>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>Fort Wayne</td>
<td>IN</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Fredericton</td>
<td>NB</td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td>Grant</td>
<td>MI</td>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>Hiawatha</td>
<td>MI</td>
<td>9</td>
<td>–</td>
</tr>
<tr>
<td>Iowa City</td>
<td>IA</td>
<td>7</td>
<td>–</td>
</tr>
<tr>
<td>Menominee</td>
<td>MI</td>
<td>9</td>
<td>–</td>
</tr>
<tr>
<td>Milaca</td>
<td>MN</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Mississipi</td>
<td>VT</td>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>Ninigrat</td>
<td>RI</td>
<td>–</td>
<td>*</td>
</tr>
<tr>
<td>Normal</td>
<td>IL</td>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>Peltson</td>
<td>MI</td>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>Pulaski</td>
<td>NY</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Toronto</td>
<td>ON</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Valparaiso</td>
<td>IN</td>
<td>10</td>
<td>–</td>
</tr>
</tbody>
</table>

*Collections (10 broods) from Ninigret were of *Cepidosaema* sp. from *Gnorimoschema salinaris* galling *Solidago sempervirens*.*

Received 16 July 2008; revised 11 March 2009; accepted 16 March 2009