Absence of Host-Associated Differentiation in the *Rhopalomyia solidaginis* and *Solidago* Relationship in Western Pennsylvania

Written By
Matthew Douglas Deicke
Spring, 2014



Allegheny College
Department of Biology

Absence of Host-Associated Differentiation in the *Rhopalomyia solidaginis* and *Solidago* Relationship in Western Pennsylvania

Written By
Matthew Douglas Deicke
Spring, 2014

A Senior Comprehensive Project in Partial Fulfillment of the Requirements for a Bachelor of Science Degree from Allegheny College

I hereby recognize and pledge to fulfill my responsibilities as defined in the Honor Code and to maintain the integrity of both myself and the College community as a whole



Pledge

Approval by the Senior Thesis Committee:

Dr. Kristen M. Webb

Prof. Christopher H. Lundberg

Table of Contents

Abstract.....**3**

Introduction.....**4–10**

Materials and Methods.....**10–16**

Sample Collection.....10

DNA Extraction.....11

Amplification of the Cytochrome Oxidase 1 (*COXI*) Gene 12

PCR Product Clean-Up and DNA Sequencing.....13

DNA Sequence Editing.....13

Constructing a Multiple Alignment.....14

Maximum Likelihood Tree Reconstruction and Haplotype Analysis.....14

Analysis of Molecular Variance.....16

Results.....**16–24**

Sample Collection.....16

DNA Extraction and Amplification of the Cytochrome Oxidase 1 (*COXI*) Gene.....17

DNA Sequencing.....18

BLAST and Multiple Alignment Analysis.....18

Haplotype and Variable Site Analysis 19

Maximum Likelihood Tree Reconstruction.....19

Analysis of Molecular Variance.....20

Discussion.....**25–31**

Acknowledgements.....**32**

Literature Cited.....**33–37**

Appendix.....**38–46**

Abstract

A sub-genre of cospecies evolution, known as host-associated differentiation (HAD), involves speciation attributed to a loss of gene flow between subpopulations of parasites due to varying preference for host organisms. Subpopulations of the galling midge, *Rhopalomyia solidaginis*, primarily parasitize both *Solidago gigantea* and *Solidago altissima* goldenrod; however, patterns of HAD have been observed in the Midwest United States by assessing the genetic diversity between host preference subgroups. Studying a previously unsurveyed geographic area helps to generalize the implications that host preference plays in the speciation of this galling midge. *R. solidaginis* samples were taken from old-fields in Pittsburgh, PA and Meadville, PA. Cytochrome oxidase 1 (*COXI*) sequences were then used in bioinformatic comparisons. Statistical tests, including Maximum Likelihood (ML) tree reconstruction and Analysis of Molecular Variance (AMOVA), did not lend support to the presence of HAD and population substructure in Western Pennsylvania. Analyses of variable site degeneracy lacked support for the presence of positive selective pressures imposed on the gall midge to shift hosts, regardless of the existence of *R. solidaginis's* parasitoid wasp, *Platygaster variabilis*, which has been attributed to the host shift observed in the Midwestern United States. This result thus modifies the understanding of HAD as being a universal speciation mechanism, and supports the fact that it may instead occur in more geographically isolated instances.

Introduction

In nature, organisms continuously encounter and interact with other species. While the most common interaction is that of predator and prey, other less aggressive relations also exist (Berryman 1992). Relationships can be classified into a number of different groups, such as parasitic or mutualistic, which are two of the most common types. In fact, it is thought that half of all known species exhibit some form of parasitic tendencies (Stireman et al. 2005). Mutualistic relationships, however, are somewhat less common due to the lack of stability within the shared benefit paradigm (Nunn et al. 2011). Depending on which type of co-species interaction is at play, the organisms involved will adapt and evolve accordingly. The genomic evolution and mutation rates in mutualistic relationships tend to be slower when compared to those in parasitisms due to the absence of selective pressures (Herre et al. 1999; Ewald 1987). Studying the progression of evolution amongst these different types of interactions can broaden the current understanding of how interacting species evolve together, thus forming a more encompassing definition of co-species evolution.

Co-species evolution, as the name suggests, involves the simultaneous evolution of two interacting species. These species can interact in a number of ways, and depending on the type of interaction, evolution can progress at different rates. The rate of evolution depends on the interacting species' exposure to positive selective pressures. If present, mutations acquired at random, which impart an advantage, will be selected for and will accumulate rapidly within the population, thus quickening the process of evolution relative to a system absent of such pressures (Spottiswoode and Stevens 2012). Without selective pressures placed on interacting species, evolution will

occur more slowly due to the sole influence of genetic drift. Positive selective pressures, often seen in parasitic relationships, allow interacting organisms to evolve more rapidly alongside their counterpart in a phenomenon known as the coevolutionary arms race (Spottiswoode and Stevens 2012). Relationships that do not exhibit host preference among interacting species therefore lack a more intimate genic connection, and are thus less likely to display characteristics of co-species evolution (Nunn et al. 2011).

Resistant evolution can occur when a host is repeatedly infected with a parasite. Repeated infection imparts positive selective pressures on the host to increase its resistance to parasitization. Such resistance may take the form of an increased immune response to a given pathogen, or the loss of receptors for that pathogen. If a species is continually infected with a parasite and resistance evolves, then the host's overall fitness increases (Boots and Haraguchi, 1999). An example of this scenario involves the nematode *Caenorhabditis elegans* and the parasitic bacteria *Bacillus thuringiensis*. In a study conducted by Schulte et al. (2010), 48 host generations of *C. elegans* were infected with *B. thuringiensis*. The parasite was seen to become more virulent and aggressive, while the host nematode became more resistant after subsequent infections. Since the acquired evolutionary advantages of both involved species was countered, results were based on mortality rates of control and experimental nematode groups (Schulte et al. 2010). When both the parasite and host evolve to increase their own fitness, it is known as the reciprocal effect. This phenomenon is often seen in relationships involving evolution of resistance (Schulte et al. 2010). Both the nematode and bacterial genomes mutated and evolved in order to, in the nematode's case, be able to survive infection, and in the bacteria's case, become more aggressive. This reciprocal relationship

between the two species led to accelerated evolutionary divergence since both organisms were experiencing positive selective pressures to increase their own fitness. When a parasitic relationship is observed, there is a likelihood that both species involved will be under positive selection, which results in accelerated evolution. If rapid evolution occurs in the presence of species sub-populations, these isolated subgroups may evolve into two genetically differentiated species.

When such evolutionary processes are present alongside barriers to gene flow, a species may experience divergence patterns. Prior to the turn of the twenty-first century, allopatric speciation was generally the accepted model for how one species can diverge into two new distinct and sexually isolated species (Mayr 1963; McCoy 2003). Allopatric speciation involves separating a given species by some type of physical barrier. Examples of barriers include a fallen tree that isolates a population of insects, or a newly formed river that isolates a population of field mice. While these scenarios are quite simplistic, they both involve the same basic principles. One population is isolated into separate groups which experience restrictions to interbreeding. This obstruction to gene flow allows for the potential accumulation of sub-population specific mutations during genome replication due to selective pressures and genetic drift. Such mutations within subgroups can accumulate, thus leading to divergence and eventual speciation (Pennisi 2006).

In the past few decades, evolutionary research has shown that the allopatric paradigm is not a requirement for the process of speciation, especially when explaining parasitic systems due to the fact that a parasite often utilizes multiple hosts in a given location (Dieckmann and Doebeli 1999; McCoy 2003). A phenomenon called sympatric

speciation can also lead to evolutionary divergence, but via a slightly different mechanism. In this scenario, species subgroups are not geographically isolated from one another and have the opportunity to interbreed. This interbreeding does not occur, however, because instead a geographical barrier being present, an ecological barrier to gene flow exists. This barrier may take the form of temporal or behavioral characteristics of an organism, which results in assortative mating (McCoy 2003). Due to the fact that some degree of gene flow has the potential to occur between sub-populations, sympatric speciation is typically a slower process when compared to classical allopatric speciation (Pennisi 2006). One of the first studies to shed light on the concept of sympatric speciation involved a population of cichlid fish living in a Cameroon lake in Western Africa (Schliewen et al. 1994). Varying food sources in the lake allowed different subgroups of cichlid to specialize when feeding. A cichlid was therefore exposed to other fish that mutually preferred a given food source and bred within that group, thus resulting in assortative mating. This behavior of feeding acted to isolate the fish and represents an ecological barrier to gene flow. Currently eleven species of cichlid fish live in Lake Ejagham in Cameroon, which, supported by molecular phylogenetic tree reconstruction, appear to have risen from one common ancestor via the mechanism of sympatric speciation (Schliewen et al. 1994, Barluenga et al. 2006).

Host-associated differentiation (HAD) is a type of speciation that most often occurs during co-species evolution in parasitic relationships. Advantageous and neutral mutations that accumulate during the host's genome replication will not be selected against, and will be passed on to the host species' progeny. If the host in a parasitic relationship mutates is such a way as to alter its relationship with the parasite,

mutations that allow the parasite to follow suit will be favorable in order to maintain the intimate genic connection between host and parasite (Cornell et al. 1998). One example involves *Eurosta solidaginis*, a fly that parasitizes *Solidago altissima*, a common species of goldenrod, resulting in goldenrod gall formation. In a 2011 study, sub-populations of goldenrod and gall flies were observed in their natural habitats of prairies and forests (Craig and Itami, 2011). Flies were then taken from their natural environment and exposed to foreign goldenrod (prairie flies were exposed to forest goldenrod and forest flies were exposed to prairie goldenrod). Oviposition rates of *E. solidaginis* on native goldenrod and foreign goldenrod populations were compared and showed that *E. solidaginis* preferred *S. altissima* from its native environment. When a fly did oviposit on foreign goldenrod, there was a lower chance of larval survival. Gall morphology was also affected, since gall size and shape are results of biochemical interactions between the fly's larva and the host plant's tissue (Gagné 1989; Dreger-Jauffret and Shorthouse 1992; Cornell et al. 1998). This study provides an example of host preference specificity exhibited by a parasite species and shows the positive selective pressures that may be placed on a parasite to maintain compatibility with its host organism (Crutsinger et al. 2008; Craig and Itami, 2011).

When multiple host species are present in a given environment, parasite subgroups may form which allows each group to maintain compatibility with its host, signifying the possibility for HAD to occur in sympatric environments (McCoy 2003). Additional studies involving *E. solidaginis* investigated the implications of host preference in sympatric population of goldenrod consisting of multiple *Solidago* host species (Waring et al. 1990; Craig et al. 1993; Abrahamson et al. 2001). *E. solidaginis*

parasitizes both *Solidago altissima* and *Solidago gigantea* goldenrod. By comparing the *COXI* gene sequences from *E. solidaginis* samples based on the host in which they were collected, significant levels of genetic diversity were observed between subgroups. This result shows the beginnings of sympatric speciation and is congruent with the HAD model. The rationale behind this differentiation was attributed to the presence of a parasitoid wasp, *Eurytoma gigantea*, which infects and kills the *E. solidaginis* larva while it undergoes development inside its gall (Brown et al. 1996; Craig and Itami 2011). The wasp's negative impact on *E. solidaginis's* fitness caused the gall midge to shift from the *Solidago altissima* host to the *Solidago gigantea* host due to *E. gigantea's* lack of preference for the latter (Brown et al. 1996; Craig et al. 2001; Craig and Itami 2011). This scenario, involving a host shift imposed by a parasitoid species, can be applied to other systems with similar components in further investigations of HAD.

Rhopalomyia solidaginis, another gall midge which parasitizes *S. altissima* and *S. gigantea* goldenrod, has displayed similar patterns of HAD in Midwestern regions of the United States (Stireman et al. 2006). Different sub-populations of *R. solidaginis* show signs of sympatric differentiation for the host species of goldenrod on which they choose to oviposit. *R. solidaginis's* host preference is attributed to a host shift imposed by *Platygaster variabilis*, a parasitoid wasp, via a mechanism much like the one previously used to explain the host shift of *E. solidaginis* (Stireman et al. 2006). This has allowed the two sub-populations to diverge independently from one another due to the behavioral barrier obstructing gene flow. If continued, this scenario has the potential to create two genetically distinct and derived species of *R. solidaginis*, thus resulting in complete reproductive isolation between *R. solidaginis* subgroups. While evidence of

HAD in *R. solidaginis* has been documented in the Midwestern United States, additional research is needed in other regions in order to generalize the divergence patterns occurring in this system (Stireman et al. 2005).

This study investigated a previously unsurveyed region of the United States where the interactions between *R. solidaginis* and its host plant had not yet been assessed, i.e. Western Pennsylvania. By evaluating the genetic diversity between *R. solidaginis COXI* sequences collected from varying hosts, the presence of HAD in this region was determined to be absent and therefore contradicted with results found by Stireman et al. (2005). Instead of supporting HAD as a universal phenomenon, newly collected results suggest more isolated occurrences. Information gained may now be applied to other parasitic relationships that show host preference as a means of more fully understanding the limitations of HAD when predicting evolution and speciation.

Materials and Methods

Sample Collection

Rhopalomyia solidaginis samples were collected from one old-field in Meadville, Pennsylvania (41.690476,-80.098318) and one old-field near Pittsburgh (Wexford), Pennsylvania (40.624148,-80.03458) that showed sympatry of *Solidago altissima* and *Solidago gigantea* goldenrod host species. Insects were collected when they were in larval form and still resided in their rosette gall, approximately correlating to the middle of August through the end of September (Newell 1994). Thirty galls from each *S. altissima* and *S. gigantea* host were harvested from each field site in an effort to yield approximately 120 galls total. Harvests therefore oversampled by collecting twice as

many galls as the number of desired *R. solidaginis* larva (60) to help account for any vacant galls. In the lab, galls were dissected to remove larva which were then stored at −80°C to await DNA extraction. Harvests were considered successful if the collected larva had a mass ≥ 4 mg. Samples were named using a two letter, one number system. The first letter represented the species of goldenrod host from which the sample was collected, the second letter represented the location of collection, and the number assigned showed the order of decreasing mass from samples sharing a host species and collection location (Example: GM-1 was the sample with the largest mass that was collected from a *S. **g**igantea* host in **M**eadville, PA).

DNA Extraction

The DNeasy Blood & Tissue Kit Insect Protocol was used for all larval samples following the disposable microtube pestle procedure (Qiagen Germantown, Maryland). The following customizations and modifications were included: step 4 involved the addition of 200 µl 96% ethanol following the addition of AL lysis buffer; step 8/9 consisted of two separate elutions, using 100 µl AE elution buffer for each, in order to maximize the concentration of DNA extracts. Following the extraction, potential presence of DNA degradation was assessed via 1% agarose gel electrophoresis stained with SYBR® Safe DNA Gel Stain (Life Technologies Carlsbad, California), which was run at 95 V for 30 min. Gel results showing bright concentrated bands of non-degraded DNA was a general requirement for each sample prior to using it in a polymerase chain reaction (PCR). Nanodrop spectrophotometry was also used to measure purity, via the 260 nm/280 nm absorbance ratio, and concentration for all extracts. DNA extracts with

an absorbance ratio in the range of 1.6–2.2 and concentration ≥8.0 ng/µl were deemed successful and were then used in PCR.

Amplification of the Cytochrome Oxidase 1 (*COXI*) Gene

The ~600 bp mitochondrial cytochrome oxidase 1 (*COXI*) gene was amplified via PCR using previously published primers (Table 1). The PCR consisted of 100 ng of extracted DNA, 0.16 µM of each primer, and 12.5 µl Taq 2X Master Mix which included 5 mM Tris-HCl, 25 mM KCl, 0.75 mM $MgCl_2$, 0.1 mM deoxyribonucleotides (dNTPs), and 25 µg/ml Hot Start Taq DNA polymerase to each PCR (New Englands Biolabs, Ipswich MA). Lastly, a volume of water was added to give the total reaction mixture a volume of 25 µl to ensure all components were at their optimal concentration. PCRs were subjected to thermal cycling conditions adapted from Stireman et al. (2005) and were then stored at 4°C (Table 2). The sole alteration from the previously published thermal cycling conditions included a reduction of the cycle count from 35 cycles to 30 cycles in an effort to reduce amplification of unwanted products. PCR products were analyzed via 1% agarose gel electrophoresis, stained with SYBR® Safe DNA Gel Stain (Life Technologies, Grand Island, NY), which was run at 70 V for 45 min. Electrophoresis included a low mass DNA ladder (Life Technologies, Grand Island, NY) to determine the success and relative sizes and concentrations of the *COXI* gene amplification products.

Table 1. Primers used in polymerase chain reactions

Primer Name	Primer Sequence (5'→3')
LCO1490 (F)	GGTCAACAAATCATAAAGATATTGG
HCO2198 (R)	TAAACTTCAGGGTGACCAAAAAATCA

Primers used in PCR, which were specific to the amplification of the *COXI* gene in *Rhopalomyia solidaginis* (Folmer et al. 1994).

Table 2. Polymerase chain reaction thermal cycling conditions

Step:	Initial Denature	Repeated for 30 Cycles			Final Extension
		Denature	Annealing	Extension	
Temperature:	94°C	94°C	50°C	72°C	72°C
Time:	2 min	40 s	50 s	1 min	4 min

Thermal cycling conditions used in PCR adapted from those used by Stireman et al. (2005). The sole alteration made from the previously published conditions involved changing the cycle count from 35 to 30 in an effort to reduce amplification of unwanted products.

PCR Product Clean-Up and DNA Sequencing

Excess dNTPs and primers were degraded from all successful PCR products by following the ExoSAP-IT for PCR Product Cleanup protocol (Affymetrix Santa Clara, California). The proper volume of PCR product, based on the size, concentration, and the requirements of Eurofins MWG Operon (http://www.operon.com/services/dna-sequencing/sample-prep.aspx), was included in the purification process. Purified samples were then sent overnight to Eurofins MWG Operon (Louisville, KY) for sequencing. Each PCR product was sequenced unidirectionally using the LCO1490 (F) primer (Table 1).

DNA Sequence Editing

The returned sequences were viewed in FinchTV (http://www.geospiza.com/Products/finchtv.shtml). A sequencing reaction was deemed successful if the generated chromatogram required only minimal trimming at the ends and minor base editing within. During editing, any miscalled bases were corrected with the incorporation of the proper nucleotide or IUB code prior to exportation in FASTA format. Before using sequences in further bioinformatic analyses, each sample's sequence was aligned using NCBI's Basic Local Alignment Search Tool (BLAST) to ensure strong alignment to

previously published *R. solidaginis COXI* gene sequences (Altschul et al. 1990). Those sequences which did not align to previously published *R. solidaginis COXI* sequences were excluded from further investigations of HAD. However, samples which aligned to *Platygaster variabilis*, the parasitoid wasp of *R. solidaginis*, were studied further to establish confidence in the wasp's presence in Western Pennsylvania.

Constructing a Multiple Alignment

Newly collected *R. solidaginis COXI* sequences were aligned using ClustalW (Larkin et al. 2007) along with previously published *R. solidaginis COXI* sequences (Genbank Accession Numbers - DQ241958–DQ241989, DQ267669–DQ267670) from Stireman et al. (2005). Those sequences smaller than 300 bp in length were removed from the alignment before exportation in FASTA format, and were not considered in further bioinformatic techniques. Removing sequences less than 300 bp in length was done to avoid the deletion of informative variable sites in longer *COXI* gene sequences when trimming missing data from the alignment. The alignment was then exported from ClustalW in MEGA6 and FASTA format.

Maximum Likelihood Tree Reconstruction and Haplotype Analysis

An appropriate evolutionary prediction model for the data set was found via the Find Best-Fit Substitution Model (ML) statistical tool in MEGA6 (Nei and Kumar 2000; Tamura et al. 2013). The Tamura 3-parameter model was selected as being the best for the dataset based upon it having the lowest Bayesian Information Criterion (BIC) score, which signifies that the substitution patterns used by this model are most likely to match the actual evolutionary progression of the sample set (Tamura 1992, Tamura et

al. 2013). Overall Mean Distance within host clades was assessed in order to help verify the ancestral host species using the determined model of best-fit (Tamura et al. 2013). Maximum Likelihood (ML) tree reconstruction was performed with 500 bootstrap replicates using previously published *Rhopalomyia pomum COXI* sequences (NC 013063.1, GQ387649.1) as an outgroup to assess where newly collected *R. solidaginis* samples grouped in previously defined host clades (Stireman et al. 2005).

To gain an understanding of the nature of the polymorphisms which contributed to the phylogenetic tree reconstruction, variable sites were identified and sequences present within the FASTA formatted alignment were collapsed into haplotypes via FaBox (Villesen 2007). Variable sites were also analyzed in MEGA6 to determine if they represented synonymous or non-synonymous mutations (Tamura et al. 2013). These variable sites were then used to generate a dn/ds ratio, where dn represents the number of non-synonymous mutations per non-synonymous site and ds represents the number of synonymous mutations per synonymous site. To support the presence of positive selective pressures, a dn/ds value (ω) must be greater than 1, signifying that dn is greater than ds (Yang and Bielawski, 2000).

For those sequences which aligned to the parasitoid wasp, *Platygaster variabilis*, in NCBI's BLAST, a Maximum Likelihood (ML) tree was reconstructed with 500 bootstrap replicates and incorporated previously published *P. variabilis COXI* sequences DQ267640.1–DQ267668.1 (Stireman et al. 2006). *Platygaster lobata* samples (DQ267637.1–DQ267639.1) were used as an outgroup to root the ML tree. This step was completed to determine if the newly collected samples aligning to *P. variabilis* were

in fact genetically and evolutionary similar to the species of parasitoid wasp that has been labeled a potential cause of HAD in the Midwestern United States.

Analysis of Molecular Variance

Analysis of Molecular Variance (AMOVA) was used to assess genetic diversity between *R. solidaginis* samples when grouped corresponding to preferred goldenrod host using Arlequin version 3.5 (Excoffier et al. 2005). AMOVA was run using standard computations for haplotypic format. Data sets assessed by AMOVA included newly collected *R. solidaginis* samples as well as a set containing newly collected sequences alongside previously published *R. solidaginis* sequences (Stireman et al. 2005).

Results

Sample Collection

Gall harvests in both Pittsburgh and Meadville were successful and produced an adequate number of rosette galls. Gall counts following collections included 30 galls from each host species from each location for a total of 120 galls. Dissection of the galls in the laboratory did not produce the desired numbers of larval samples having an adequate mass (≥ 4.0 mg) to be used in subsequent experimental procedures. Harvests counts following dissection included 10 samples from *S. gigantea* and 11 samples from *S. altissima* taken from Pittsburgh, PA and 2 samples from *S. gigantea* and 1 sample from *S. altissima* taken from Meadville, PA exhibiting sufficient masses. While sample counts did not meet the goal of 15 collected specimens from each host species, samples which had masses near the level of adequacy, exhibiting values as low as 1.0 mg, were

used in DNA extraction in an effort to increase the size of the data set and to give a total working sample count of 55.

DNA Extraction and Amplification of the Cytochrome Oxidase 1 (*COXI*) Gene

DNA extractions were successful for 33 of the 55 attempted samples, exhibiting concentrations ≥8.0 ng/µl and 260/280 absorbance ratios in the range of 1.6-2.2 (Appendix A). Some DNA extractions produced results with borderline concentration and purity values; a total of 41 samples were deemed appropriate for PCR having minimum concentrations of 7.0 ng/µl and absorbance ratios in the range of 1.32-3.10.

PCR was successful for 30 of the 41 attempted samples which showed concentrated bands of DNA with an approximate size of ~650 bp when run on a 1% agarose gel (Table 3). PCR product concentrations varied by sample, falling anywhere within the 5 ng/µl to 30 ng/µl range.

Table 3. Success of polymerase chain reactions

Successful PCR Amplification	Unsuccessful PCR Amplification
S. altissima Pittsburgh Samples AP-1, AP-2, AP-3, AP-4, AP-5, AP-6, AP-7, AP-8, AP-9, AP-10	*S. altissima* Pittsburgh Samples AP-11, AP-12
S. gigantea Pittsburgh Samples GP-1, GP-2, GP-3, GP-4, GP-5, GP-6, GP-7, GP-8, GP-9, GP-10, GP-11, GP-12	*S. gigantea* Pittsburgh Samples –
S. altissima Meadville Samples AM-4, AM-5, AM-13	*S. altissima* Meadville Samples AM-1, AM-3, AM-6, AM-11, AM-12, AM-15
S. gigantea Meadville Samples GM-1, GM-4, GM-8, GM-9, GM-12	*S. gigantea* Meadville Samples GM-6, GM-11, GM-13

Samples are grouped as successful or unsuccessful, which was dependent on the presence or absence of an amplified DNA band as seen on an agarose gel. Band intensity was not considered when determining the success of amplification.

DNA Sequencing

Of the 30 PCR products produced, 27 were sequenced successfully and returned sequence data that required only minor end trimming. Edited sequences can be viewed in Appendix B. Samples AM-5, AM-13, and GM-1 all produced uninterpretable chromatograms, and were therefore excluded from further analysis.

BLAST and Multiple Alignment Analysis

Nineteen of the 27 FASTA formatted sequences were identified as *R. solidaginis COXI* based on a strong alignment to previously published *R. solidaginis COXI* sequences deposited in NCBI's Nucleotide Database (www.ncbi.nlm.nih.gov). Samples AM-4, AP-2, GM-8, GM-9, GM-12, AP-8, GP-10, and GP-12 did not align to *R. solidaginis COXI* sequences, and were therefore excluded from those computational assessments of the *R. solidaginis* population. Results of each sequence's BLAST alignment can be viewed in Table 4. Sequences which were identified as *R. solidaginis COXI* were successfully aligned using ClustalW (Larkin et al. 2007) alongside previously published *R. solidaginis COXI* sequences (Stireman et al. 2005). Of the aligned sequences, the majority exhibited sequence lengths greater than 200 bp. Sequence AP-3, with a length of 167 bp, was much shorter than other aligned sequences and was therefore removed from further analyses. This was done in an effort to reduce the amount of trimming needed within the alignment. Trimming is performed to remove missing data which is present when one sequence does not span the length of a longer sequence in an alignment. The need for excessive trimming associated with the incorporation of short sequences has the potential to delete informative variable sites within the longer sequences, thus the

removal of the AP-3 sequence was imperative. Following the removal of short sequences, the alignment had a total length of 583 bp.

Haplotype and Variable Site Analysis

Exported FASTA sequences collapsed into five haplotypes (Table 5). Haplotype 3 had been previously defined by Stireman et al. (2005), while haplotypes 1, 2, 4, and 5 were novel (Table 5). A total of four variable sites were found between the five haplotypes (Table 6). These variable sites existed at positions 214, 256, 297, and 540 of the *COXI* gene. Sites 214 and 256 exhibited synonymous mutations and therefore did not affect the resultant protein's primary structure. Variable sites 297 and 540, however, did exhibit non-synonymous mutations and therefore affected protein structure according to the invertebrate mtDNA genetic code (https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c#SG5). The total number of non-synonymous 0-fold degenerative sites was 364 while the total number of synonymous 2-fold and 4-fold degenerative sites was 196 (Tamura et al. 2013). By using the dn/ds ratio equation, an ω value of 0.538 was calculated. This ω value, being less than 1, indicates that dn is less than ds and does not support the presence of positive selective pressures in the studied system (Yang and Bielawski, 2000).

Maximum Likelihood Tree Reconstruction

Maximum Likelihood (ML) tree reconstruction, which included *R. solidaginis* *COXI* sequences previously published by Stireman et al. (2005), produced a cladogram with two main clades diverging from a common ancestor (Tamura 1992, Tamura et al. 2013) (Figure 1). These clades consisted of the *Solidago altissima* host clade and the

Solidago gigantea host clade. Newly collected *R. solidaginis COXI* sequences all grouped within the *S. altissima* host clade, which contains members of the ancestral clade as shown by further statistical analyses explained below (Figure 1).

Genetic diversity within each clade of the ML tree was calculated in order to infer the ancestral host species by using MEGA6 (Nei and Kumar 2000, Tamura 1992, Tamura et al. 2013). The *S. altissima* host clade, including all newly collected sequences, exhibited a diversity value of 0.0125, which is approximately four fold greater than the diversity observed in the *S. gigantea* host clade at a value of 0.0032. This increased diversity measured in the *S. altissima* host clade, as well as the branching patterns observed in the ML tree, lend support for it being the ancestral host species (Figure 1).

Analysis of Molecular Variance

The fixation index (Φ_{ST}) value calculated for newly collected *R. solidaginis COXI* sequences was 0.07507 ($p = 0.14956 \pm 0.00899$), and therefore showed no significant variation between *S. altissima* and *S. gigantea* host populations. The Φ_{ST} produced when analyzing newly collected *R. solidaginis COXI* sequences alongside previously published sequences by Stireman et al. (2005) was a value of 0.42796 ($p = 0.00000 \pm 0.00000$), which does indicate significant variation between host-associated populations. This significant variation, however, is attributed solely to the variation existing between *R. solidaginis* host populations of previously published data from the Midwestern United States, as the non-significant Φ_{ST} value for newly collected sequences does not support the presence of host-associated variation between Western Pennsylvanian *R. solidaginis* subgroups.

Table 4. BLAST alignment results for newly collected *R. solidaginis* samples

Sequence	Best Alignment	Accession Number	e-value
AM4	*Rhipicephalus pumilio COXI*	HM193878.1	2.0E-04
AP1	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
AP2	*Hymenoptera spp. COXI*	JX833433.1	4.0E-46
AP3	*Rhopalomyia solidaginis COXI*	DQ267670.1	4.0E-77
AP4	*Rhopalomyia solidaginis COXI*	DQ241966.1	0.0
AP5	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
AP6	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
AP7	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
AP8	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
AP9	*Rhopalomyia solidaginis COXI*	DQ267670.2	0.0
AP10	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
GM4	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
GM8	*Baryscapus coerulescens COXI*	HQ106702.1	0.0
GM9	*Baryscapus spp. COXI*	HQ106774.1	0.0
GM12	*Platygaster variabilis COXI*	DQ267646.1	0.0
GP1	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
GP2	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
GP3	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
GP4	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
GP5	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
GP6	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
GP7	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
GP8	*Platygaster lobata*	DQ267637.1	2.0E-145
GP9	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
GP10	*Platygaster lobata*	DQ267637.1	4.0E-142
GP11	*Rhopalomyia solidaginis COXI*	DQ267670.1	0.0
GP12	*Platygaster variabilis COXI*	DQ267643.1	0.0

NCBI Basic Local Alignment Search Tool (BLAST) results showing best alignment for each of the newly collected sequences (Altschul et al. 1990). All sequences aligned to a *COXI* gene, however not all alignments were to the *COXI* gene of *R. solidaginis*. Each sample's strongest alignment from the GenBank database is shown along with the associated accession number. The included e-values are statistical measures generated by NCBI which signify the probability that the alignment was due to chance. E-values have an inverse relationship to the likelihood of an alignment occurring by chance. Those sequences which aligned to the *R. solidaginis* parasitoid wasp (*Platygaster variabilis*) were further studied using cladogram analysis.

Table 5. *COXI* haplotype distribution of newly collected *R. Solidaginis* samples

	Haplotypes				
	1	**2**	**3***	**4**	**5**
Samples	AP-1	AP-4	AP-5	AP-8	GP-1
	AP-6		GM-4	AP-9	
	AP-7		GP-7		
	AP-10				
	GP-2				
	GP-3				
	GP-4				
	GP-5				
	GP-6				
	GP-9				
	GP-11				

Newly collected *R. solidaginis COXI* sequences were collapsed into five haplotypes using the FaBox Online Toolbox (Villesen 2007). Haplotype 3 (*) has been previously defined by Stireman et al. (2005), while haplotypes 1, 2, 4, and 5 are novel.

Table 6. Variable sites existing within newly collected *R. solidaginis COXI* sequences

		COXI Loci			
		214	256	297	540
Haplotypes	1	G*	T*	C	T
	2	.	.	.	A
	3	A*	.	.	.
	4	A*	C*	.	.
	5	.	.	A	.

Variable sites within newly collected *R. solidaginis COXI* samples can be observed at positions 214, 256, 297, and 540 of the COXI gene. The presence of an asterisks (*) adjacent to a variable site nucleotide denotes a synonymous mutation while the lack of an asterisk signifies a non-synonymous mutation. These variable sites were contained inside the 583 bp alignment of newly collected *R. solidaginis COXI* samples, and were analyzed using FaBox Online Toolbox (Villesen 2007) and MEGA6 (Tamura et al. 2013). Numbered haplotypes coordinate with those presented in Table 5.

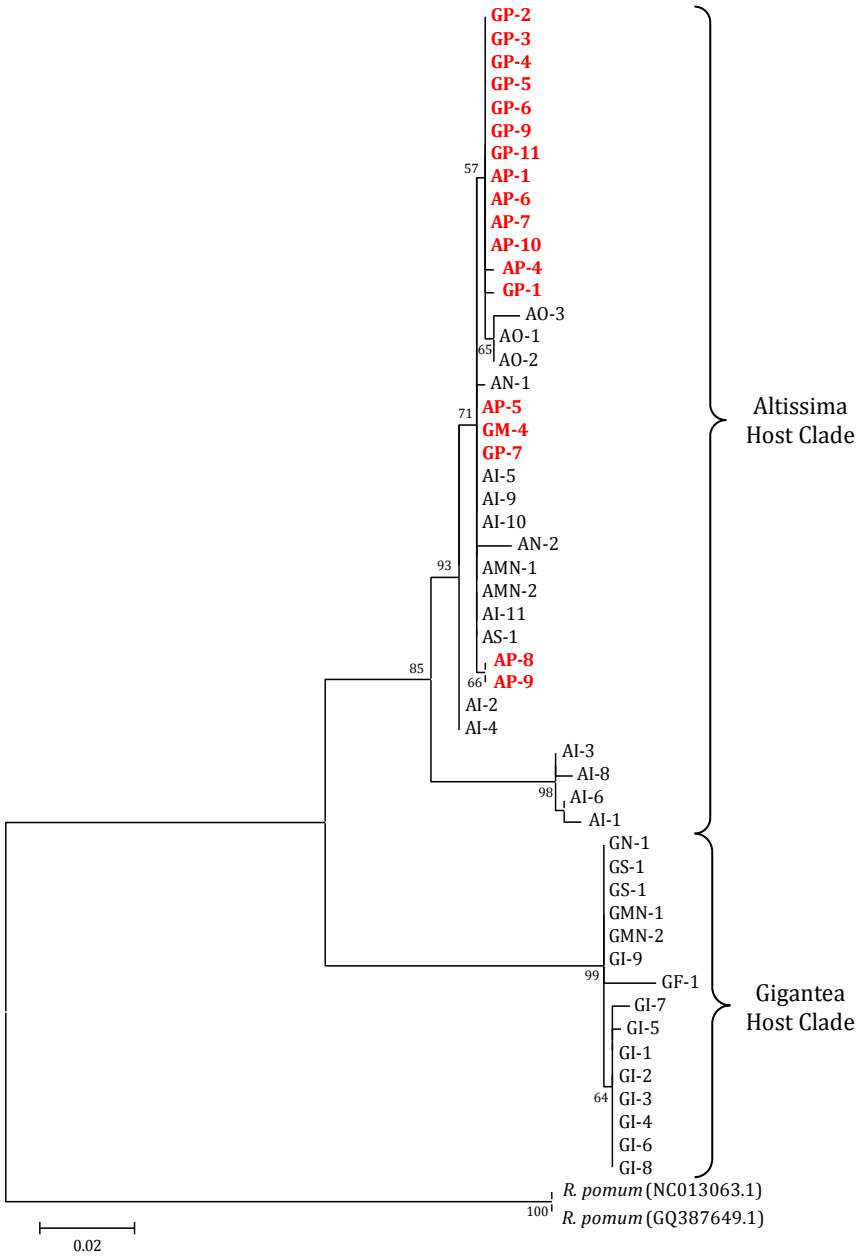


Figure 1. *Rhopalomyia solidaginis* Maximum Likelihood (ML) tree with the highest log likelihood score (-507.6197) was reconstructed using the Tamura 3-Parameter (g = +2.00) model (Tamura 1992, Tamura et al. 2013). This tree shows two main clades which were previously defined by Stireman et al. (2005), and include the *S. altissima* host clade (top) and the *S. gigantea* host clade (bottom). Newly collected sequences all fell within the *S. altissima* host clade and can be seen in bold red. Branch lengths are drawn to scale with the number of substitutions per site represented by the incorporated scale bar. The ML tree was rooted with *Rhopalomyia pomum COXI* sequences (NC013063.1, GQ387649.1) as an outgroup.

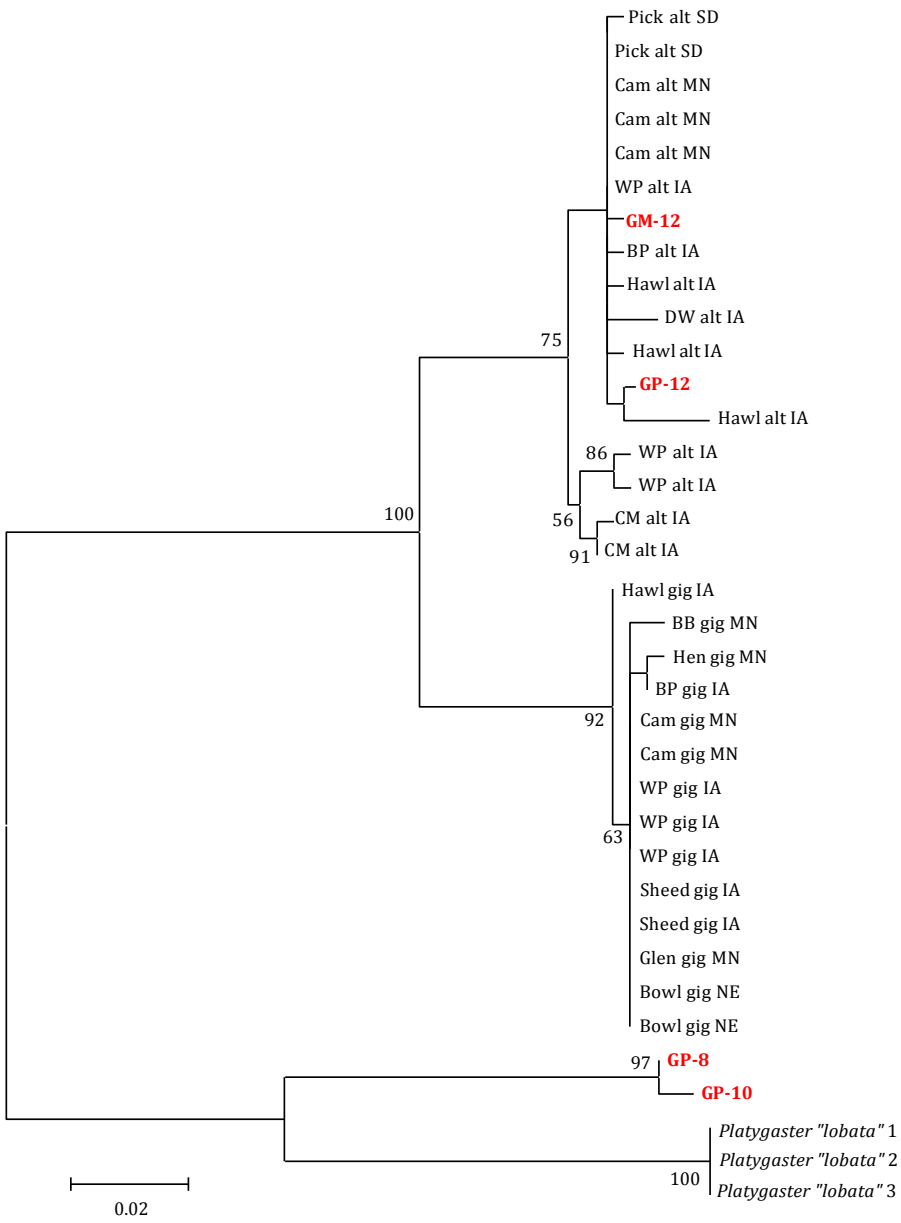


Figure 2. *Platygaster variabilis* Maximum Likelihood (ML) tree with the highest log likelihood score (-1090.2523) was reconstructed using the Tamura 3-Parameter (g = +0.32) model (Tamura 1992, Tamura et al. 2013). This tree shows two main clades which were previously defined by Stireman et al. (2006), and coordinate with the observed patterns of HAD in *P. variabilis*. Newly collected samples can be seen in bold red. Two of the newly collected *P. variabilis* samples, GM-12 and GP-12, fell within the *S. altissima* host clade. Samples GP-8 and GP10 do not group with either host clade, but show a closer evolutionary relationship to the *Platygaster lobata* outgroup. Branch lengths are drawn to scale with the number of substitutions per site represented by the incorporated scale bar. The ML tree was rooted with *Platygaster lobata COXI* sequences (DQ267637.1-DQ267639.1) acting as an outgroup.

Discussion

Assessing the presence of host-associated genetic differentiation (HAD) in the relationship between the parasitic fly *Rhopalomyia solidaginis* and its *Solidago spp.* host requires the use of cladogram analysis along with molecular variance statistics. The Maximum Likelihood (ML) tree reconstruction showed that all newly collected *R. solidaginis* samples grouped with the *Solidago altissima* host domain previously defined by Stireman et al. (2005) independent of which host the gall midge larva was collected from. This result does not support the presence of HAD in western Pennsylvania, as separation of flies based on their parasitized host would be necessary. Analysis of molecular variance (AMOVA) was also not able to support the presence of HAD in this newly studied region. The percentage of molecular variation between host groups of newly collected samples was insignificant; therefore, population substructure is not present in the observed locations. Lastly, haplotype analysis does not support HAD, as *R. solidaginis* samples from both *S. gigantea* and *S. altissima* hosts were found to contain identical *COXI* sequences. These findings are somewhat unexpected, as previous research has definitively observed HAD in the *R. solidaginis* and *Solidago spp.* system in Midwestern United States (Stireman et al. 2005). The reasoning behind this opposing conclusion could stem from a number of different rationales. At this time, each explanation is purely academically based speculation, and would require supplemental investigations in order to be supported.

One potential explanation to why HAD was not observed in Western Pennsylvania would be if host goldenrod plants were misidentified during gall harvests. Throughout preliminary field work, *Solidago altissima* was distinguished from *Solidago*

gigantea based primarily on the presence of pubescence. *S. altissima* characteristically has a scabrous and fuzzy texture covering both the undersides of mature leaves as well as the main stalk; *S. gigantea* has a smooth glabrous coating on the stalk and leaves (Abrahamson and Heinrich 2013). By observing the presence of pubescence, gall harvest adopted a systematic approach. Nevertheless, the amount of "fuzz" present on stem and leaf structures fit more of a gradient, thus adding additional complication to host identification practices (M. Deicke, pers. obs.). However, to avoid collecting *R. solidaginis* samples from an incorrect host, goldenrods exhibiting an intermediate degree of pubescence were disregarded. This selectivity during harvest lends confidence that host goldenrods were properly identified, yet there is always the possibility for human error. If in fact host-associated divergence was present and plants were misidentified during harvests, numerous variations of expected results would be produced for each potential scenario.

In a scenario where all samples were collected from the *Solidago altissima* host species but varied in assigned host identity, expected results would show all *R. solidaginis* samples grouping with the previously defined *S. altissima* clade (Stireman et al. 2005). This expected result does match actual findings; therefore, this type of host misidentification cannot truly be dismissed. If a similar scenario ensued where all samples were instead collected from the *S. gigantea* host species, all *R. solidaginis* samples would then be expected to group with the previously defined *S. gigantea* host clade. In collections where *R. solidaginis* samples were harvested from both *Solidago* hosts but were haphazardly misidentified, samples would be expected to group with each previously defined host clade in an unpredictable pattern. In any case, to

definitively exclude the potential of host misidentification, further research would need to utilize additional strategies during field work. Such strategies could include rhizome root observation techniques as well as genetic analysis of the host organism for truly unquestioned host plant identification (Abrahamson and Heinrich 2013).

If the assumption is made that *R. solidaginis* samples were collected from properly identified goldenrod host plants, the relative abundances of available host species in sympatric environments may play a role in why HAD was not observed in Western Pennsylvania. The old-fields which were visited in Pittsburgh, PA and Meadville, PA both housed greater numbers of *S. altissima* plants than *S. gigantea* plants (C. Lundberg and M. Deicke, pers. obs.). This fact alone would imply that *R. solidaginis* flies aimlessly in search of a host plant for purposes of oviposition have a greater probability of finding and landing on *S. altissima*. It is also true that gallmakers show preference for the host in which they oviposit due to the biochemically specific interactions that occur between the larvae and plant tissue (Gagné 1989; Dreger-Jauffret and Shorthouse 1992; Cornell et al. 1998). The maintenance of this intimate relationship is especially important when the parasite organism feeds on the plant due to the availability of nutrients and compounds present within the host, which pertains to the *R. solidaginis* and *Solidago spp.* relationship (Medina et al. 2013). Combining these two points suggests that any *R. solidaginis* fly which showed preference for the *S. gigantea* host species would be at an inherent disadvantage due to the sheer lack of available host plants in the old-fields of Western Pennsylvania. Such a disadvantage would favor *R. solidaginis* flies to remain as generalists in this region, allowing them the ability to parasitize both *S. gigantea* and *S. altissima* hosts. If this generalistic approach

was in fact utilized by *R. solidaginis* to avoid disadvantageous circumstances, ecological barriers to gene flow would not be established and host-associated differentiation within a *R. solidaginis* population would not occur. Further studies involving observations of Western Pennsylvanian *R. solidaginis* host preference and relative host abundances may lend support for this potential explanation. One such investigation could be completed by comparing the ratio of parasitized *S. gigantea* to total *S. gigantea* versus the ratio of parasitized *S. altissima* to total *S. altissima* in a defined area. By comparing these relative parasitized host frequencies, a greater understanding of host preference can be established for the *R. solidaginis* and *Solidago* system in Western Pennsylvania.

Diversity statistics used to assess *R. solidaginis* samples based on their host plant showed increased levels of diversity for flies which parasitized *S. altissima* relative to flies that parasitized *S. gigantea*. This result, along with previously published support, suggests that *S. altissima* is the ancestral host species while *S. gigantea* is derived (Brown et al. 1995; Stireman et al. 2005). The basic rational behind this assumption of ancestry is that a subpopulation of *R. solidaginis* which parasitized an ancestral *Solidago spp.* host would have a longer amount of time to accumulate genetic variations due to influences such as genetic drift and any present selective pressures. The variation that existed in a derived parasite subgroup would therefore be less when compared to that of the ancestral group. This is due to the fact that the derived subgroup arose from a subset of the ancestral group, and consequently has less genetic diversity due to the decreased duration of genetic drift and selective pressure exposure. Divergence events can take place when a parasite shifts to a novel host (Medina et al. 2013; Bush 1969;

Winkler and Mitter 2008). Reasons behind this host shift can vary, but ecological selective forces, as well as predation by enemy organisms, may play a role in forcing an organism to find a new host species in order to maintain its fitness (Craig et al. 1997; Itami et al. 1998). This paradigm is thought to explain why *Eurosta solidaginis*, another gall producing fly, began parasitizing the *S. gigantea* host plant (Brown et al. 1995; Craig et al. 2001). *E. solidaginis* plays victim to the attack of *Eurytoma gigantea*, a parasitoid wasp which infects the galls created by *E. solidaginis*, causing larval death (Abrahamson et al. 2001). The presence of this predator is thought to be the cause of *E. solidaginis's* host shift from *S. altissima* to *S. gigantea* due to the wasp's lack of preference for the derived goldenrod host. This host shift was primarily responsible in forming host-associated subgroups of *E. solidaginis*, thus establishing barriers to gene flow between groups and leading to the occurrence of HAD (Brown et al. 1995; Craig et al. 2001).

Like *E. solidaginis, R. solidaginis* also has a predator wasp, known as *Platygaster variablis*, which attacks the rosette galls made by the midge (Stireman et al. 2006). The presence of *P. variablis* has been speculated as a potential cause for a host preference shift in *R. solidaginis* seen in the Midwestern United States (Stireman et al 2005). In a possible scenario where there was a regional absence of this predator wasp, it is possible that there would be no selective pressures placed on *R. solidaginis* to utilize the *S. gigantea* host over the *S. altissima* host, thus HAD would not be observed. The existence of this specific species parasitoid wasp in Western Pennsylvania was supported by BLAST alignment data and ML tree reconstruction, as one sample from each collection location showed significant genetic similarity to previously published *P.*

variabilis samples. While this data does support the wasp's presence, the data set was too small to assess genetic variation among samples, and therefore could not comment on the duration of the wasp's presence in this region. If this parasitoid wasp has only recently colonized this area, selective pressures influencing *R. solidaginis's* host preference may have not been in place long enough to cause detectable variation between subgroups of the gall midge. This rational is supported by the data set, as degeneracy analysis showed no imposed positive selective pressures placed on *R. solidaginis* to evade a parasitoid species. Still, confounding variables are most likely present which may be masking a longstanding and influential existence of *P. variabilis* in Western Pennsylvania.

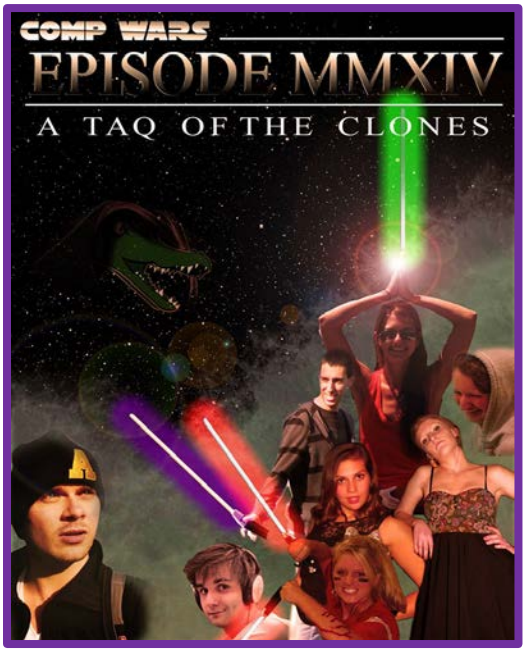
To investigate the potential influence *P. variablis* has had on Western Pennsylvania *R. solidaginis* flies, analyses of genetic diversity between newly collected *P. variabilis* samples can be compared to those studied in the Midwestern United States by Stireman et al. (2006). Higher levels of genetic diversity among a regional population of *P. variabilis* would suggest a longer existence in that given location. This would then suggest that the parasitoid wasp has inhabited that location for a longer period of time, and therefore has had the ability to impose selective pressures on *R. solidaginis* for the duration. A second approach, which assumes the brief presence of selective pressures not detectable by *COXI* analysis, could investigate different genetic sequences in *R. solidaginis* that have higher rates of mutation in order to identify early signs of differentiation. One potential region that might offer additional insight would be the mitochondrial control region (D-loop) due to its mutation prone nature and abundance of hypervariable sites (Stoneking et al. 2000). Outside of mitochondrial DNA,

microsatellite length analysis can also be used to compare *R. solidaginis* specimens, as samples with similar lengths of microsatellite tandem repeats are typically more closely related (Bowcock et al. 1994). Utilizing these techniques during investigations of HAD has the potential to provide refinement in showing the very beginnings of a speciation event which was unable to be shown by analyses solely involving *COXI*.

In conclusion, HAD in the *R. solidaginis* and *Solidago spp.* relationship was not supported in Western Pennsylvania based on the assessment of genetic variability in the *COXI* gene. The reasoning behind why HAD appears to be present in Midwestern regions of the United States but not in Western Pennsylvania is currently unclear, however potential explanations involving the abundance of the preferred host species and the lack of selective pressures necessary to cause a host shift may point to likely scenarios. In any case, further research is required to investigate this unexpected outcome. This study does support the fact that HAD should be treated like a localized phenomenon, as an isolated occurrence seen in a given location does not generate an all-encompassing conclusion for the organism of interest. As research continues to test the waters of evolutionary processes, a greater understanding of speciation can be formed. By opposing the universal understanding of HAD in the *R. solidaginis* and *Solidago spp.* system set forth by previously published research, this study comments on HAD's localized presence and illuminates its inherent limitations as a mechanism for worldwide speciation.

Acknowledgements

I would first like to thank Allegheny College, Department of Biology, for generously funding my research endeavor. Secondly, I would like to thank my comprehensive project advisor, Dr. Kristen Webb, who was truly an invaluable resource and mentor throughout this entire experience. Next, I would like to thank my second reader, Prof. Chris Lundberg, as well as Dr. Ron Mumme for the conversations and guidance they provided me in regard to the *Rhopalomyia* and *Solidago* system. I also thank my senior seminar "comp group" who has critiqued my work to help ensure the production of a quality research piece. Lastly, I would like to thank my family and friends for their support throughout the duration of this project; specifically I must acknowledge Annie Altman, Sean Loose, and Karyl Deicke for their assistance in the goldenrod fields during gall collections. All of these mentioned individuals positively impacted my project in some shape or form, and for that I am truly grateful.



The Beloved "Comp Group"

Literature Cited

Abrahamson, W., & Heinrich, P. The Solidago Eurosta Gall homepage. Retrieved July, 2013, from http://www.facstaff.bucknell.edu/abrahmsn/solidago/main.html

Abrahamson, W. G., Eubanks, M. D., Blair, C. P., & Whipple, A. V. (2001). Gall flies, inquilines, and goldenrods: A model for host-race formation and sympatric speciation. *American Zoologist, 41*(4), 928-938.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "*f*." J. Mol. Biol. 215:403-410. PubMed

Barluenga, M., Stölting, K., Salzburger, W., Muschick, M., & Meyer, A. (2006). Sympatric speciation in nicaraguan crater lake cichlid fish. *Nature, 439*, 719. doi:doi:10.1038/nature04325

Berryman, A. A. (1992). The orgins and evolution of predator-prey theory. *Ecology*, 73(5), 1530-1535. doi:10.2307/1940005

Boots, M., & Haraguchi, Y. (1999). The evolution of costly resistance in Host-Parasite systems. *The American Naturalist, 153*(4), 359-370.

Bowcock, A., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J., & Cavalli-Sforza, L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature, 368*, 455-457.

Brown, J. M., Abrahamson, W. G., & Way, P. A. (1996). Mitochondrial DNA phylogeography of host races of the goldenrod ball gallmaker, eurosta solidaginis (diptera: Tephritidae). *Evolution, 50*(2), 777-786.

Bush, G. (1969). Sympatric host race formation and speciation in frugivorous flies of genus Rhagoletis (Diptera: Tephritidae). *Evolution* 23: 237–251.

Cornell, H. V., B. A. Hawkins, and M. E. Hochberg. (1998). Towards an empirically-based theory of herbivore demography. Ecol. *Entomol* 23:340–349.

Craig, T. P., Itami, J. K., Abrahamson, W. G., & Horner, J. D. (1993). Behavioral evidence for host-race formation in eurosta solidaginis. *Evolution, 47*(6), 1696-1710.

Craig, T. P., J. D. Horner, and J. K. Itami. (1997). Hybridization studies on the host races of Eurosta solidaginis: implications for sympatric speciation. *Evolution* 51:1552–1560.

Craig, T. P., Horner, J. D., & Itami, J. K. (2001). Genetics, experience, and host-plant preference in eurosta solidaginis: Implications for host shifts and speciation.*Evolution, 55*(4), 773-782. doi:10.1111/j.0014-3820.2001.tb00813.x

Craig, T. P., & Itami, J. K. (2011). Divergence of eurosta solidaginis in response to host plant variation and natural enemies. *Evolution, 65*(3) doi:10.1111/j.1558-5646.2010.01167.x

Crutsinger, G. M., Habenicht, M. N., Classen, A. T., Schweitzer, J. A., & Sanders, N. J. (2008). Galling by rhopalomyia solidaginis alters solidago altissima architecture and litter nutrient dynamics in an old-field ecosystem. *Plant and Soil, 303*(1-2), 95-103. doi:http://dx.doi.org/10.1007/s11104-007-9490-3

Dieckmann, U., & Doebeli, M. (1999). On the origin of species by sympatric speciation. *Nature, 400*(6742), 354-357.

Dreger-Jauffret, F. and J. D. Shorthouse. (1992). Diversity of gall-inducing insects and their galls. Pp. 8–33 in J. D. Shorthouse and O. Rohfritsch, eds. *Biology of insect-induced galls*. Oxford Univ. Press, Oxford, U.K.

Ewald, P. W. (1987). Transmission modes and evolution of the parasitism-mutualism continuuma. *Annals of the New York Academy of Sciences, 503*(1), 295-306. doi:10.1111/j.1749-6632.1987.tb40616.x

Excoffier, L. G. Laval, and S. Schneider (2005) *Arlequin ver. 3.0: An integrated software package for population genetics data analysis*. Evolutionary Bioinformatics Online 1:47-50.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek R. (1994). DNA primers for amplification of mitochondrial cytochrome coxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology, 3*(5), 294-299.

Gagné, R. D. (1989). The plant-feeding gall midges of North America. Cornell Univ. Press, Ithaca, NY.

Herre, E. A., Knowlton, N., Mueller, U. G., & Rehner, S. A. (1999). The evolution of mutualisms: Exploring the paths between conflict and cooperation. *Trends in Ecology & Evolution, 14*(2), 49-53. doi:http://dx.doi.org/10.1016/S0169-5347(98)01529-8

Itami, J. K., T. P. Craig, and J. D. Horner. (1998). Factors affecting gene flow between the host races of Eurosta solidaginis. Pp. 375–407 in S. Mopper and S. Strauss, eds., Genetic variation and local adaptation in natural insect populations: effects of ecology, life history and behavior. Chapman and Hall, New York.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007) *ClustalW and ClustalX version 2*. Bioinformatics 23(21): 2947-2948.

Mayr, E. (1963). *Animal species and evolution*. Cambridge: Belknap Press of Harvard University Press.

McCoy, K. D. (2003). Sympatric speciation in parasites – what is sympatry? *Trends in Parasitology, 19*(9), 400-404. doi:http://dx.doi.org/10.1016/S1471-4922(03)00194-6

Medina, R. F., Szendrei, Z., Harrison, K., Isaacs, R., Averill, A., Malo, E. A., & Rodriguez-Saona, C. (2014). Exploring host-associated differentiation in the north american native cranberry fruitworm, acrobasis vaccinii, from blueberries and cranberries. *Entomologia Experimentalis Et Applicata, 150*(2), 136-148.

Nei M. and Kumar S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Newell, S. J. (1994). Occurrence of goldenrod galls: Study of insect ovipositing behavior. *The American Biology Teacher, 56*(1), 51-54.

Nunn, C. L., Ezenwa, V. O., Arnold, C., & Koenig, W. D. (2011). Mutualism or parasitism? using a phylogenetic approach to characterize the oxpecker-ungulate relationship. *Evolution, 65*(5) doi:10.1111/j.1558-5646.2010.01212.x

Pennisi, E. (2006). speciation standing in place. *Science, 311*, 1372. doi:10.1126

Schliewen, U. K., Tautz, D., & Paeaebo, S. (1994). Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature, 368*(6472), 629-632.

Schulte, R. D., Makus, C., Hasert, B., Michiels, N. K., & Schulenburg, H. (2010). Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. *Proceedings of the National Academy of Sciences of the United States of America, 107*(16) doi:10.1073/pnas.1003113107

Spottiswoode, C. N., & Stevens, M. (2012). Host-parasite arms races and rapid changes in bird egg appearance. *American Naturalist, 179*(5) doi:10.1086/665031

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

StiremanIII, J. O., Nason, J. D., & Heard, S. B. (2005). Host-associated genetic differentiation in phytophagous insects: General phenomenon or isolated exceptions? evidence from a goldenrod-insect community. *Evolution, 59*(12), 2573-2587. doi:10.1111/j.0014-3820.2005.tb00970.x

Stoneking, M. (2000). Hypervariable sites in the mtDNA control region are mutational hotspots. *The American Journal of Human Genetics, 67*(4), 1029-1032. doi:http://dx.doi.org/10.1086/303092

Tamura K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Molecular Biology and Evolution* 9:678-687.

Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30: 2725-2729.

Villesen, P (2007), FaBox: an online toolbox for fasta sequences, *Molecular Ecology Notes* 7 (6), 965–968. doi:10.1111/j.1471-8286.2007.01821.x

Waring, G. L., Abrahamson, W. G., & Howard, D. J. (1990). Genetic differentiation among host-associated populations of the gallmaker eurosta solidaginis (diptera: Tephritidae). *Evolution, 44*(6), 1648-1655.

Winkler I., Mitter C. (2008). The phylogenetic dimension of insect-plant interactions: a review of recent evidence. Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects (ed. by KJ Tilmon), pp. 240–263. University of California Press, Berkeley, CA, USA.

Yang, Z., & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution, 15*(12), 496-503. doi:http://dx.doi.org/10.1016/S0169-5347(00)01994-7

Appendix

Appendix A. Success of DNA extractions

Sample	Concentration (ng/µl)	260/280 Ratio	Band Presence
AP-1	**173.2**	**2.15**	**Yes**
AP-7	**156.2**	**2.07**	**Yes**
AP-5	**125.9**	**2.08**	**Yes**
AP-8	**112.6**	**2.13**	**Yes**
AP-6	**110.8**	**1.92**	**Yes**
AP-3	**108.8**	**2.04**	**Yes**
AP-2	**106.5**	**1.99**	**Yes**
AP-4	**101.6**	**1.98**	**Yes**
GP-12	**101.5**	**2.05**	**No**
GP-10	**96.3**	**2.08**	**No**
GM-1	95.4	1.32	No
GP-11	**95.3**	**2.09**	**Yes**
AP-9	**94.3**	**2.04**	**Yes**
GP-5	**83.6**	**2.01**	**Yes**
GP-6	**81.6**	**1.96**	**Yes**
GP-9	**77.3**	**2.12**	**Yes**
GP-4	**76.3**	**1.84**	**Yes**
AP-10	**49.9**	**2.08**	**No**
GP-2	**47.8**	**2.07**	**Yes**
GP-3	**47.0**	**2.02**	**Yes**
GP-1	**43.2**	**1.99**	**Yes**
AM-5	**41.0**	**1.86**	**Yes**
GP-8	**38.6**	**2.08**	**No**
GM-4	**31.7**	**2.19**	**Yes**
GP-7	**23.1**	**2.06**	**Yes**
AM-13	**18.3**	**1.63**	**No**
GM-9	17.6	1.46	No
GM-8	**17.3**	**1.92**	**No**
GM-13	**13.4**	**1.89**	**No**
AP-11	**12.0**	**1.79**	**No**
AP-12	**11.8**	**1.66**	**No**
AM-4	10.8	2.26	No
AM-12	**10.3**	**1.80**	**Yes**
GM-11	**9.6**	**1.71**	**No**
AM-15	**8.9**	**1.67**	**No**
GM-12	**8.7**	**1.88**	**No**

Sample	Concentration (ng/µl)	260/280 Ratio	Band Presence
GM-6	8.2	5.54	No
AM-1	7.8	2.29	No
AM-11	7.7	2.01	Yes
AM-6	7.6	2.41	No
AM-3	7.1	3.10	Yes
GM-14	6.4	1.50	No
GM-10	5.7	1.29	No
AM-8	5.2	1.39	No
AM-16	4.3	1.75	No
AM-7	3.8	1.21	No
GM-7	3.8	4.27	No
AM-14	3.7	3.74	No
AM-17	3.4	2.06	No
AM-2	3.3	6.10	No
AM-9	3.2	1.93	No
AM-10	2.8	1.41	No
GM-3	-3.8	0.76	Yes
GM-5	-4.0	0.83	Yes
GM-2	-7.4	1.40	No

Samples are ordered by decreasing concentration value as determined by Nanodrop spectrophotometry. Band presence was determined if a concentrated band of DNA was observed when the agarose gel was exposed to UV light. The double barred line indicates the concentration threshold for being deemed successful (≥8.0 ng/µl). Samples that had both sufficient concentrations and absorbance ratios (1.6-2.2) are bolded. Unsuccessful extractions can be seen shaded in blue.

Appendix B. Newly collected *COXI* Sequences in FASTA format (edited)

>AM-4
TTTATACAAGATGTAAAACTAAGAACCTTACARATAGTTAGTGATTCGCTATGAACCATT
GTAATTCATCATTCGCATACCATTAGTTCTGTATGTAGTGTGATTTTTTGGTCACCCTGA
AGTTTAAAA

>AP-1
TTGGATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAG
GAGTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCAT
TTATTATAATTTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAA
TTCCTATCATATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTTT
GATTATTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAG
GTACAGGATGAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAG
TTGATTTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTA

ATTTTATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCAT
TATTTATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAG
CGGGAGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTA
TAGGAGGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGG

>AP-2
AAATGATCAAATTTATAATTCAATTGTKACTACTCATGCATTTACGATAATTTTTTTTTT
TGTTATACCTGTAATAATAGGKGGATTKGGAAATTATTTAATTCCATTATTTTTAGGAGT
CCCAGATATAGCATTTCCACGGATA

>AP-3
GGATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGA
GTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTT
ATTATAATTTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGG

>AP-4
ATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGAGT
AAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTTAT
TATAATTTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAATTCC
TATCATATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTTTGATT
ATTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGGTAC
AGGATGAACTGTTTATCCCCCCCTATCGWCTAGAATTGYCCATTCAGGAACATCAGTTGA
TTTTTCAATTTTTTCATTACATATAGCANGAATTTCATCTATTTTAGGAGCAATTAATTT
TATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATTATT
TATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGAATTAGCGGG
AGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATAGG
AATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATAGGAGGGGGAGA
TCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACC

>AP-5
GATCTGAAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGA
GTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTT
ATTATAATTTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAATT
CCTATCATATTAGGAGCACCAGATATAGCATTTCCTCGAATAAATAATTTAAGATTTTGA
TTATTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGGT
ACAGGATGAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAGTT
GATTTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAAT
TTTATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATTA
TTTATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAGCG
GGAGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATA
GGAGGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAAGTT
TTAA

>AP-6
GAATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGAGTAAAAAATT
TAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTTATTATAATTT
TTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAATTCCTATCATAT

TAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTTTGATTATTACCTC
CTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGGTACAGGATGAA
CTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAGTTGATTTTTCAA
TTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAATTTTATCTCAA
CTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATTATTTATTTGAT
CAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAGCGGGAGCTATTA
CTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATAGGAGGGGGAG
ATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAAGTTTAA

>AP-7
GAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGAGTAAAAAA
TTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTTATTATAAT
TTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAATTCCTATCAT
ATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTTTGATTATTACC
TCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGGTACAGGATG
AACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAGTTGATTTTTC
AATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAATTTTATCTC
AACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATTATTTATTTG
ATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAGCGGGAGCTAT
TACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATAGGAGGGGG
AGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAAAGTTTAA

>AP-8
AGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGAGTAAAAA
ATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTTATTATAA
TTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAATTCCTATCA
TATTAGGAGCACCAGATATAGCATTTCCTCGAATAAATAATTTAAGATTTTGATTATTAC
CCCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGGTACAGGAT
GAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAGTTGATTTTT
CAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAATTTTATCT
CAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATTATTTATTT
GATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAGCGGGAGCTA
TTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATAGGAGGGG
GAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAAGTTTAA

>AP-9
ATCTGAGAGGAATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGAG
TAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTTA
TTATAATTTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAATTC
CTATCATATTAGGAGCACCAGATATAGCATTTCCTCGAATAAATAATTTAAGATTTTGAT
TATTACCCCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGGTA
CAGGATGAACTGTTTATCCCCCCCTATCGTCTAGAANTGCCCATTCAGGAACATCAGTTG
ATTTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAATT
TTATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATTAT
TTATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAGCGG
GAGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATAG

GAGGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAAAGTT
AAA

>AP-10
AGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGAGTAAAAA
ATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTTATTATAA
TTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAATTCCTATCA
TATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTTTGATTATTAC
CTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGGTACAGGAT
GAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAGTTGATTTTT
CAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAATTTTATCT
CAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATTATTTATTT
GATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAGCGGGAGCTA
TTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATAGGAGGGG
GAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAAAGTTTAA

>GM-4
GGATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGA
GTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTT
ATTATAATTTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAATT
CCTATCATATTAGGAGCACC
AGATATAGCATTTCCTCGAATAAATAATTTAAGATTTTGATTATTACCTCCTTCATTAAT
ATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGGTACAGGATGAACTGTTTATCC
CCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAGTTGATTTTTCAATTTTTTCATT
ACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAATTTTATCTCAACTATATTAAA
TATAAAAATTAAATTTATTAAATTTGATCAAATTTCATTATTTATTTGATCAATTTTAAT
TACAGCTATTTTATTATTATTATCATTACCAGTATTAGCGGGAGCTATTACTATATTATT
AACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATAGGAGGGGGAGATCCAATTTT
ATATCAACATTTATTTTGATTTTTTGGTCACCCTGAAGTTTAA

>GM-8
GATATGATCCGGATTTTAGGAATATCAATAAGAATAATTATTCGAATAGAATTAGGTAAC
CCAGGTTCAATAATTGGAAATGATCAAATTTATAATGCTATTGTTACAACACATGCATTT
ACTATAATTTTTTTTTTTTGTTATACCTGTTATAATAGGAGGATTTGGTAATTATTTAATT
CCTATAATAATAGGTGTTCCTGATATAGCTTTCCCTCGAATAAATAATATAARATTTTGA
CTATTACCTCCTARATTAATATTATTATTATCAAGAATATTTATTGGTAATGGGACTGGT
ACTGGGTGAACTGTATATCCCCCTTTATCAGGGAATTTATCACATARAGGACCTTCAGTT
GATTTATCAATTTTTTCATTACATATTGCAGGAGTTTCTTCAATTATAGCTTCAATTAAT
TTTATTTCAACCATTTTAAATATAAAAATTTATAAATTAGAAATAATTTCATTATTTTCA
TGATCGATAATATTGACAGCAATTTTATTATTATTATCATTACCAGTATTGGCTGGAGCT
ATTACTATATTATTATTTGATCGGAATTTAAATACATCATTTTTTGATCCATCTGGGGGC
GGARACCCAATTTTATATCAGCATTTATTTTGATTTTTTGGTCACCCTG

>GM-9
GATATGATCCGGATTTTAGGAATATCAATAAGAATAATTATTCGAATAGAATTAGGTAAC
CCAGGTTCAATAATTGGAAATGATCAAATTTATAATGCTATTGTTACAACACATGCATTT
ACTATAATTTTTTTTTTTTGTTATACCTGTTATAATAGGAGGATTTGGTAATTATTTAATT

CCTATAATAATAGGTGTTCCTGATATAGCTTTCCCTCGAATAAATAATATAARATTTTGA
TTATTACCTCCTARATTAATATTATTATTATCAAGAATATTTATTGGTAATGGAACTGGT
ACTGGGTGAACTGTATATCCCCCTTTATCAGGGAATTTATCACATARAGGACCTTCAGTT
GATTTATCAATTTTTTCATTACATATTGCAGGAATTTCTTCAATTATAGCTTCAATTAAT
TTTATTTCAACCATTTTAAATATAAAAATTTATAAATTARAAATAATTTCATTATTTTCA
TGATCGATAATATTAACAGCAATTTTATTATTATTATCATTACCAGTATTGGCTGGAGCT
ATTACTATATTATTATTTGATCGGAATTTAAATACATCATTTTTTGATCCATCTGGGGGC
GGAGACCCAATTTTATATCAGCATTTATTTTGATTTTTTGGTCACCC

>GM-12
TTAAGAATACTTATTCGTCTAGAACTAGGAACACCCTCATCTCTTATCGGAAACGATCAA
ATTTATAACTCTATTGTTACAGCTCACGCCCTAGTTATAATTTTYTTCATAGTAATACCA
ATTATACTGGGAGGATTCGGAAATTGATTAATTCCTCTTATACTATCAGCACCTGATATA
GCATTCCCACGAATAAATAATATAAGATTCTGACTTTTGCCGCCATCAATAATACTACTC
TTATACAGAAGRATTTTTAGTCTAGGAACAGGAACCGGATGAACACTATACCCTCCTTTA
TCCCTATCTCTCAATTCCTCAATTGATCTAAGAATTTTTTCTTTACACCTTGCAGGAATC
TCTTCAATCATAGGATCAATTAATTTTATTTGCACRATTTTAAACATAACTCCAATAGGA
TTTAAAATAGAAAAAAACTCATTATTTACTTGATCAATTTTCATTACTACAATTCTTCTA
CTATTGTCCCTTCCTGTATTAGCAGGAGCAATCACTATACTACTAACTGATCGAAATTTA
AATACATCATTCTTTGACCCAGCTGGAGGGGGAGACCCAATCCTATACCAACATTTATTT
TGATTTTTTGGTCACCCTGGAAGTTTAA

>GP-1
TTTGGATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGA
GGAGTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCA
TTTATTATAATTTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTA
ATTCCTATCATATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTT
TGATTATTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAAATGGAACA
GGTACAGGATGAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCA
GTTGATTTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATT
AATTTTATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCA
TTATTTATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTA
GCGGGAGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCT
ATAGGAGGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAA
GTTTAA

>GP-2
GGATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGA
GTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTT
ATTATAATTTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAATT
CCTATCATATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTTTGA
TTATTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGGT
ACAGGATGAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAGTT
GATTTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAAT
TTTATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATTA
TTTATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAGCG

GGAGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATA
GGAGGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACC

>GP-3
TTTGGATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGA
GGAGTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCA
TTTATTATAATTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTA
ATTCCTATCATATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTT
TGATTATTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACA
GGTACAGGATGAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCA
GTTGATTTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATT
AATTTTATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCA
TTATTTATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTA
GCGGGAGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCT
ATAGGAGGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACC

>GP-4
TTTGGATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGA
GGAGTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCA
TTTATTATAATTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTA
ATTCCTATCATATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTT
TGATTATTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACA
GGTACAGGATGAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCA
GTTGATTTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATT
AATTTTATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCA
TTATTTATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTA
GCGGGAGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCT
ATAGGAGGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAA
GTTTAA

>GP-5
TTTGGATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGA
GGAGTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCA
TTTATTATAATTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTA
ATTCCTATCATATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTT
TGATTATTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACA
GGTACAGGATGAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCA
GTTGATTTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATT
AATTTTATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCA
TTATTTATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTA
GCGGGAGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCT
ATAGGAGGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAA
GTTTAA

>GP-6
GGATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGA
GTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTT

ATTATAATTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAATT
CCTATCATATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTTTGA
TTATTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGGT
ACAGGATGAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAGTT
GATTTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAAT
TTTATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATTA
TTTATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAGCG
GGAGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATA
GGAGGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAAGTT
TAA

>GP-7
TCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGGAGTA
AAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATTTATT
ATAATTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAATTCCT
ATCATATTAGGAGCACCAGATATAGCATTTCCTCGAATAAATAATTTAAGATTTTGATTA
TTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGGTACA
GGATGAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAGTTGAT
TTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAATTTT
ATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATTATTT
ATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAGCGGGA
GCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTATAGGA
GGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAAGTTTAA

>GP-8
GATTTGATCAGGATGGTTGGTAGAAGATTAAGAATACTTATTCGTCTTGAACTGGGAACT
CCGTCATCTCTTATCGGAAATGATCAAATCTATAACTCTATTGTTACAGCTCATGCATTG
GTCATAATTTTTTTTATAGTTATACCTATTATACTAGGGGGATTTGGAAATTGACTAATC
CCTCTAATACTATCAGCCCCTGATATAGCTTTCCCACGAATAAATAACATAAGATTTTGA
CTTTTACCCCCCTCTATAATACTCCTTTTATATAGTAGAATTTTCAGACTAGGTACTGGA
ACAGGATGAACATTATACCCACCCCTATCTTTATCTTTAAATTCTTCAATTGATTTAAGA
ATTTTCTCATTACATCTAGCAGGTATTTCTTCAATCATAGGGTCAATTAATTTTATTTGC
ACAATTCTAAATATAACACCTATAAGATTTAAAATAGAAAAAAACTCTCTTTTTACTTGA
TCAATTTTCATTACTACAATTTTATTATTACTTTCTCTACCTGTATTAGCAGGAGCTATT
ACTATATTATTAACTGACCGTAACTTAAACACATCATTCTTTGATCCAGCAGGTGGGGGA
GACCCAATCCTATACCAACACTTATTTTGATTTTTTGGTCACCCTGAAGTTTAA

>GP-9
TGGATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGG
AGTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATT
TATTATAATTTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAAT
TCCTATCATATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTTTG
ATTATTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGG
TACAGGATGAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAGT
TGATTTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAA
TTTTATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATT
ATTTATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAGC

GGGAGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTAT
AGGAGGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAAGT
TTAA

>GP-10
GATGGTTGGTAGAAGATTAAGAATACTTATTCGTCTTGAACTGGGAACTCCGTCATCTCT
TATCGGAAATGATCAAATCTATAACTCTATTGTTACAGCTCATGCATTGGTCATAATTTT
TTTTATAGTTATACCTATTATACTAGGGGGATTTGGAAATTGACTAATCCCTCTAATACT
ATCAGCCCCTGATATAGCTTTCCCACGAATAAATAACATAAGATTTTGACTCTTACCCCC
CTCTATAATACTCCTTTTATATAGTAGAATTTTCAGACTAGGTACTGGAACAGGATGAAC
ATTATACCCACCCCTATCTTTATCTTTAAATTCTTCAATTGATTTAAGAATTTTCTCATT
ACATCTAGCAGGTATTTCTTCAATCATAGGGTCAATTAATTTTATTTGCACAATTCTAAA
TATAACACCTATAAGATTTAAAATAGAAAAAAACTCTCTTTTTACTTGATCAATTTTCAT
TACTACAATTTTATTATTACTTTCCCTACCTGTATTAGCAGGAGCTATTACTATATTATT
AACTGACCGTAACTTAAACACATCATTCTTTGATCCAGCAGGTGGGGGAGACCCAATCCT
ATACCAACACTTATTTTGATTTTTTGGTCACCCTGAAGTTTAA

>GP-11
TGGATCTGAGAGGATAATTGGAACATCTTTAAGAATTTTGATTCGAATAGAATTAAGAGG
AGTAAAAAATTTAATTGGAAGAGATCAAATTTATAATGTAATTGTTACTGCTCATGCATT
TATTATAATTTTTTTTTATAGTTATACCTATTATAATTGGGGGATTTGGAAATTGATTAAT
TCCTATCATATTAGGAGCACCAGATATGGCATTTCCTCGAATAAATAATTTAAGATTTTG
ATTATTACCTCCTTCATTAATATTACTATTAATTAGAAGAATAGTAGAAACTGGAACAGG
TACAGGATGAACTGTTTATCCCCCCCTATCGTCTAGAATTGCCCATTCAGGAACATCAGT
TGATTTTTCAATTTTTTCATTACATATAGCAGGAATTTCATCTATTTTAGGAGCAATTAA
TTTTATCTCAACTATATTAAATATAAAAATTAAATTTATTAAATTTGATCAAATTTCATT
ATTTATTTGATCAATTTTAATTACAGCTATTTTATTATTATTATCATTACCAGTATTAGC
GGGAGCTATTACTATATTATTAACTGATCGAAATTTAAATACATCTTTTTTTGATCCTAT
AGGAGGGGGAGATCCAATTTTATATCAACATTTATTTTGATTTTTTGGTCACCCTGAAGT
TTAA

>GP-12
GATAGTAGGAAGAAGATTAAGAATACTTATTCGTCTAGAACTAGGAACACCCTCATCTCT
TATCGGGAACGATCAAATTTATAACTCTATTGTTACAGCTCACGCCCTAGTTATAATTTT
TTTCATAGTAATACCAATTATACTGGGAGGATTCGGAAATTGATTAATTCCTCTTATACT
ATCAGCACCTGATATAGCATTCCCACGAATAAATAATATAAGATTCTGACTTTTGCCGCC
ATCAATAATACTACTCCTATACAGAAGGATTTTTAGTCTAGGAACAGGAACCGGATGAAC
ACTATACCCTCCTTTATCCCTATCTCTCAATTCCTCAATTGATCTAAGAATTTTTTCTTT
ACACCTTGCAGGAATCTCTTCAATCATAGGATCAATTAATTTTATTTGCACAATTTTAAA
CATAACTCCAATAGGATTTAAAATAGAAAAAAACTCATTATTTACTTGATCAATTTTCAT
TACTACAATTCTTCTACTATTGTCCCTTCCTGTATTAGCAGGAGCTATCACTATACTACT
AACTGATCGAAATTTAAATACATCATTCTTTGACCCAGCTGGAGGGGGAGACCCAATCCT
ATACCAACATTTATTTTGATTTTTTGGTCACCCTGAAGTTTAA