Spring 2011

Multilocus Genetic Divergence Within and Among Three Species of Timena Walking Sticks

Mathew Sommers
University of Colorado Boulder

Multilocus Genetic Divergence Within and Among Three Species of *Timema* Walking Sticks

Mathew Sommers
University of Colorado at Boulder
Department of Ecology and Evolutionary Biology
Patrik Nosil (Primary Advisor)
Defense Date: April 4, 2011

Committee Members:

Patrik Nosil – *Department of Ecology and Evolutionary Biology*
Barbara Demmig-Adams – *Department of Ecology and Evolutionary Biology*
Andrew Martin – *Department of Ecology and Evolutionary Biology*
Matthew Wilsey-Cleveland – *Program for Writing and Rhetoric*

Abstract: The theory of ecological speciation suggests that adaptation to different habitats promotes the formation of new species (speciation), but debates persist about the extent to which this can occur when populations are in geographic contact and exchange genes (i.e. gene flow). Multilocus genetic analysis (using multiple genes) can inform this debate by comparing levels of genetic divergence between taxa (groups such as species) at different points in the speciation process and with differing geographic arrangements. For example, taxa undergoing less gene flow are predicted to exhibit stronger genetic divergence than taxa undergoing more gene flow. Furthermore, multilocus data can be used to test for genealogical discordance—a scenario where different genes tell different evolutionary histories, due to being differentially affected by evolutionary processes such as selection, mutation, and gene flow. The present study uses multilocus genetic data (four nuclear genes and a mitochondrial gene) to compare the genetic structure of six populations of herbivorous *Timema* walking-sticks with different geographic relationships: two diverging host-associated ecotypes (phenotypically and ecologically different groups) of *T. cristinae* in parapatry (geographic contact) and allopatry (geographic separation), and two closely related species, *T. californicum* and *T. poppensis*, in parapatry and allopatry. As expected, we observe that species are generally more genetically differentiated than ecotypes. Interestingly, however, allopatric populations are not always significantly more differentiated than parapatric populations. Also, genealogical discordance was observed. Mitochondrial differentiation among allopatric *T. cristinae* population was stronger than nuclear differentiation. In contrast, mitochondrial differentiation between the *T. californicum* and *T. poppensis* was weak, whereas nuclear genes were differentiated to the point of reciprocal monophyly, indicative of genetically distinct species. While these data

strongly suggest differing levels of gene flow among genes and taxa, further analyses using coalescent-based models should be done to estimate gene flow specifically.

Glossary

Ecological Speciation – *the process in which reproductive isolation evolves between populations through ecologically based divergent selection*

Ecologically Based Divergent Selection – *selection arising from environmental differences and/or ecological interactions, which act in contrasting directions on two populations*

Reproductive Isolation – *a reduction or lack of genetic exchange (gene flow) between populations*

Gene Flow – *the transfer of genes or alleles from one population to another population*

Ecotype – *phenotypically and ecologically different groups or populations within the same species*

Multilocus Genetic Analysis – *analysis utilizing genetic sequence data at multiple different locations across the genome*

Speciation Genes – *genes whose divergence made a significant contribution the evolution of reproductive isolation between populations or any gene that reduces hybrid fitness*

Neutral Genes – *genes that have no significant role in the speciation process*

Adaptive Genetic Divergence – *the process of genetic differentiation between two populations arising from different sets of environmental conditions to become more suited for*

Assortative Mating – *mating of individuals which share more common traits (e.g. individuals of the same ecotype) than likely from random mating*

Amplified Fragment Length Polymorphism (AFLP) – *a molecular method for identifying DNA polymorphisms, used to detect loci that exceed neutral expectations (speciation genes)*

Reciprocal Monophyly – *mutually exclusive groups of individuals, characteristic of distinct species*

Introduction

There are historical and current debates regarding the underlying mechanisms and geographic modes of speciation (Coyne & Orr, 2004). Some assert that divergence can only proceed if gene flow (intermixing of genes between populations) is absent or very limited, while others claim that divergence can indeed take place in the face of gene flow. Furthermore, it remains unclear what the implications of gene flow are for the rate of speciation or for how far the process of divergence proceeds (Nosil, Harmon, & Seehausen, 2009; Via, 2001). Another more rudimentary debate is simply how to define a species. This issue is known as the "species problem", comprised of numerous working definitions for a species (Hey, 2006a). For the purposes of this project, I will use Ernst Mayr's (1942) biological species concept that a species is an interbreeding group of individuals, reproductively isolated from other groups. Here, speciation is most simply the evolutionary process of forming new "biological" species, trending toward reduced gene flow and increased genetic divergence between taxa as speciation unfolds. In this context, speciation may refer to the evolution from one species into another, or the divergence of one ancestral species into two or more new species (Dobzhansky, 1940). Many have argued that the species problem exists, at least in part, because speciation occurs along a continuum of divergence rather than as a discrete event (Mallet, 2007; Mallet, Beltran, Neukirchen, & Linares, 2007). Over time, differentiation gradually accumulates until two populations become strongly or fully reproductively isolated species. The extended and continuous nature of the speciation process thus makes it difficult to define the exact point at which two diverging populations have become distinct species (Mallet, 2007).

The geographic arrangement of populations is also important when discussing the processes of evolutionary divergence. Speciation was originally thought to require populations

in allopatry (geographic isolation). Mayr, a large contributor to this field, claimed that speciation is unlikely to happen under conditions other than strict allopatry (Futuyma & Mayer, 1980; Mayr, 1942, 1947, 1954). This view implies that geographic isolation is required to act as the barrier to gene flow and allows for the build-up of genetic differences between populations, which incidentally cause reproductive isolation. Under this hypothesis, isolated subpopulations adapt to different environmental conditions and experience random genetic mutation and genetic drift (random change in allele frequencies), resulting in genetic differentiation and thus the eventual formation of new species over time. Intuitively, this makes sense, especially under the assumption that gene flow would break down the process of genetic divergence and speciation (Mayr, 1963; Rice & Hostert, 1993; Wright, 1931). However, with new technology focused on generating large genetic datasets, and new theoretical and conceptual arguments being made, there have been many challenges to the view of strict allopatric speciation (Futuyma & Mayer, 1980; Nosil, 2008; Rice & Hostert, 1993; White, 1978). It is now clear that speciation happens (at least some of the time) in sympatry, where populations occur in the same geographic area, and parapatry, where adjacent populations share a marginal area of overlap (Rice & Hostert, 1993; Rundle & Nosil, 2005). In non-allopatric models, barriers to gene flow no longer need arise from geographic constraints but rather are initiated through diversifying evolutionary processes (Nosil & Rundle, 2009). For example, reproductive isolation might evolve as a by-product of adaptive genetic divergence, stemming from ecological differences between populations and subsequent ecologically based divergent selection (i.e. ecological speciation) (Rundle & Nosil, 2005), or via the process of "reinforcement", where a reduction in hybrid offspring fitness causes selection against hybridization and for assortative mating (Rice & Hostert, 1993; Rundle & Nosil, 2005).

With a growing understanding of the complex nature of speciation, many historical assumptions have come into question; particularly, that gene flow must be absent during the process of speciation. More specifically, speciation scenarios where gene flow *may* be present involve sympatric or parapatric populations that undergo ecological speciation and/or reinforcement. Although the end of speciation is marked by complete reproductive isolation (no genetic exchange), it does not imply that gene flow must not have occurred at some point during divergence. In fact, the possibility of divergence with gene flow has been shown in a number of theoretical scenarios (Bolnick & Fitzpatrick, 2007; Nosil, 2008), but the number of empirical examples of speciation with gene flow remains limited (Emelianov, Marec, & Mallet, 2004; Faure, Jollivet, Tanguy, Bonhomme, & Bierne, 2009; Niemiller, Fitzpatrick, & Miller, 2008). Recent advances in technology have increased the ease of acquisition of genetic sequence data, and have allowed for the use of such data for population level analyses. Because genetic information provides the best insight into the evolutionary histories of extant taxa and because of the statistical power gained from large sample sizes, population level genetic data are necessary to make accurate inferences about evolutionary divergence and gene flow. Despite substantial improvements, evolutionary geneticists are still troubled with reconstructing the genetic histories that involve gene flow during divergence, leaving the role of gene flow in speciation not yet fully understood.

Gene Flow and Population Differentiation:

Gene flow, selection, and mutation are three key processes involved in genetic differentiation and subsequent speciation (Coyne & Orr, 2004). Mutation generates new genetic variants, which can then be acted upon by selection or genetic drift. These latter processes work together to establish genetic differentiation between diverging species, while

gene flow may break down or reinforce divergence (Tajima, 1983). Genetic differentiation itself can be assessed through the use of F_{ST}, the inbreeding coefficient. An F_{ST} value of 0 (or values not significantly different from 0) indicates little to no genetic differentiation and an F_{ST} value of 1 reflects complete genetic differentiation (i.e. fixed differences) (Wright, 1965). For example, low F_{ST} values might reflect populations experiencing strong gene flow, while high F_{ST} values might suggest that little or no gene flow is occurring. However, a major difficulty here is disentangling the effects of gene flow versus divergence time on levels of differentiation. For example, low F_{ST} values might arise from high gene flow, recent divergence (i.e. insufficient time to accumulate differentiation), or a combination of these processes (Hey, 2006b).

Genetic Differentiation and Genealogical Discordance:

In addition to varying among different pairs of populations, levels of genetic differentiation can vary among regions of an organism's genome. By using multilocus genetic data, one can compare the amount of differentiation at each locus. When different loci (or genes) show different amounts of genetic differentiation, and thus illustrate different evolutionary histories, it is known as genealogical discordance (Degnan & Rosenberg, 2009). Variation among F_{ST} estimates between different loci can provide useful information about how selection and gene flow differentially affect gene regions (Beaumont, 2005). For example, genes subject to divergent selection are expected to be more differentiated than other genes, because selection is essentially pulling apart allele frequencies between populations at selected loci and preventing alleles at these loci from moving between populations.

While mutation, selection, and gene flow all influence genetic differentiation, they are also the root causes of genealogical discordance. Mutation, a stochastic process, does *not occur* at equal rates across the genome (Neigel & Avise, 1986). Selection then plays a major role in the differential rates at which mutations (or genetic differentiation) *accumulate* in a population. For example, if a mutation occurs in a gene that is directly involved in establishing reproductive isolation (speciation gene), selection will remove that mutation from the population if it proves to be disadvantageous, while mutations in genes not involved in speciation (neutral genes) may persist. Also, genes do not flow at equal rates between populations (Wang, Wakeley, & Hey, 1997). Similar to the accumulation of mutation, divergent selection plays a critical role in determining which genes will or will not "flow". For speciation genes, some alleles will be advantageous for one ecotype and disadvantageous for another ecotype. These alleles will be selected differentially and will therefore "flow" less readily when introgression (mating between diverging populations) occurs (Wang, et al., 1997). Furthermore, while neutral genes that are not physically linked to genes under divergent selection may flow readily between populations, gene flow at neutral loci that are linked to speciation genes will be resisted (Barton, 1998; Smith & Haigh, 1974). For these reasons, it is common for different genes to reveal remarkably different evolutionary histories, which is shown as highly variable levels of genetic divergence between populations across different loci. One such example of genealogical discordance is the contrast between mitochondrial genes and nuclear genes. Mitochondrial DNA is estimated to evolve 5 to 10 times faster than nuclear DNA (Brown, George, & Wilson, 1979). There are many potential explanations for this pattern including: (1) mitochondrial DNA replicates more frequently than nuclear DNA and therefore has a greater likelihood of mutating, (2) mitochondrial DNA is only inherited maternally,

resulting in a smaller effective population size and (3) mitochondrial DNA does not undergo recombination events, causing mutations to accumulate more quickly in a population (Brown, et al., 1979; Dopman, Perez, Bogdnowicz, & Harrison, 2005).

In sum, although genealogical discordance is expected to occur even in the absence of gene flow, simply due to the stochasticity of genetic drift, discordance is expected to be greater when gene flow occurs, because gene flow is affected at some regions (e.g. speciation genes or divergently selected regions) more than others (e.g. neutral regions) (Hey, 2006b).

Study System: Timema *walking sticks*

Timema walking sticks are wingless, herbivorous insects found primarily in the chaparral of western North America (Vickery, 1993). The genus *Timema* is comprised of approximately 20 different species of walking stick and is estimated to have emerged 20 million years ago in conjunction with the spread of the chaparral biome (Crespi & Sandoval, 2000; Sandoval, Carmean, & Crespi, 1998). *Timema* are highly cryptic insects, resting on leaves and stems during the day and feeding at night (Crespi & Sandoval, 2000). A large body of research supports the conclusion that ecological speciation is occurring in different populations of the species *Timema cristinae* (Nosil, 2007). Notably, *T. cristinae* is in the early stages of ecological speciation and found as two different host-plant associated "ecotypes", or phenotypically and ecologically different groups. These ecotypes are distinguished by numerous morphological differences, with one of the most pronounced being the presence or absence of a dorsal stripe (along the back of the insect). The stripe is common on the host *Adenostoma fasciculatum*, a plant with thin needle-like leaves. In contrast, the unstriped ecotype is common on *Ceanothus spinosus*, a broad-leaved plant (Nosil, 2007). It is clear from past experimental and genotypic (e.g. amplified fragment length polymorphisms [AFLP]) data

that these two ecotypes are undergoing ecological speciation in the presence of gene flow though host-plant adaptation and ecologically based divergent selection (Nosil, 2007; Nosil, Crespi, & Sandoval, 2003; Nosil & Rundle, 2009). However, multilocus DNA sequences data have not yet been analyzed to make, for example, independent estimates of divergence time versus gene flow (Nosil, et al., 2003). Other members of the genus, namely *Timema poppensis* and *Timema californicum*, have unclear phylogenetic histories (Law & Crespi, 2002). Based on sequence divergence in the mitochondrial gene cytochrome oxidase 1 (CO1), a gene commonly used for inferring insect phylogenies, the latter two "species" appear to be closely related or even the *same species*, but are phenotypically highly distinct (Law & Crespi, 2002). Since genealogical discordance can provide information on gene flow and divergence time, further work on these taxa utilizing multiple loci is necessary.

Hypotheses and Questions:

I predict that the species pairs will exhibit greater genetic divergence than the ecotype pairs, given that species are further along the speciation continuum than ecotypes. Also, I predict that estimates of F_{ST} will differ among genes, especially so between mitochondrial genes and nuclear genes. Furthermore, it is likely that mitochondrial DNA will show a different evolutionary history than nuclear DNA, given what is known about genealogical discordance and mitochondrial evolution. Also, I will examine the role of geographic arrangement during speciation and more specifically, see if geographic arrangement is playing a role in the differentiation of *T. cristinae* or if ecologically based divergent selection is driving factor.

Methods

Study Populations and Sampling:

In order to make comparisons of genetic divergence for varying degrees of speciation (how much divergence has occurred), we compared divergence between ecotypes of *Timema cristinae* to divergence between the species pair, *Timema californicum* and *Timema poppensis.* To also allow for the examination of the effects of geographic arrangement, we examined two adjacent parapatric populations ("parapatric pair" hereafter) and two geographically separated populations ("allopatric pair" hereafter) for both ecotype and species level comparisons (*see Table 1*). The recent and incomplete ecological speciation of *T. cristinae* ecotypes could represent a model of recent divergence, while *T. poppensis* and *T. californicum* could fit a model of more ancient divergence. By using both parapatic and allopatric populations for comparisons we are able to assess the level of genetic divergence given these geographic scenarios. Previous experimental and genetic research indicates that *Timema* do not disperse over long distances, resulting in limited interbreeding between geographically separated populations (Crespi & Sandoval, 2000; Nosil, et al., 2003). Thus, these allopatric populations can be considered genetically isolated from their respective parapatric counterparts, making them a sufficient comparison with the parapatric population pairs to discern genetic divergence due to isolation and assess the impacts of gene flow at multiple loci.

Approximately 25 specimens from each of the six populations were collected during their breeding season, between March and June, in 2010. The parapatric ecotypes of *T. cristinae* were collected from the populations HVC and HVA (N34 29.309 W119 47.180 and N34 29.305 W119 47.191 respectively) and the allopatric *Adenostoma* ecotype was collected at LA (N34 30.464 W119 47.694). Individuals from parapatric populations of *T. poppensis*

(PSM1RW) and *T. californicum* (CSMQ) were collected along Summit Road (N37 133.43 W122 05.271) and individuals from the allopatric population of *T. poppensis* were collected at PTBDF (N38 37.100 W123 17.322). Once collected, all samples were stored in 100 percent ethanol, and shipped to the University of Colorado at Boulder to be stored at -40°C.

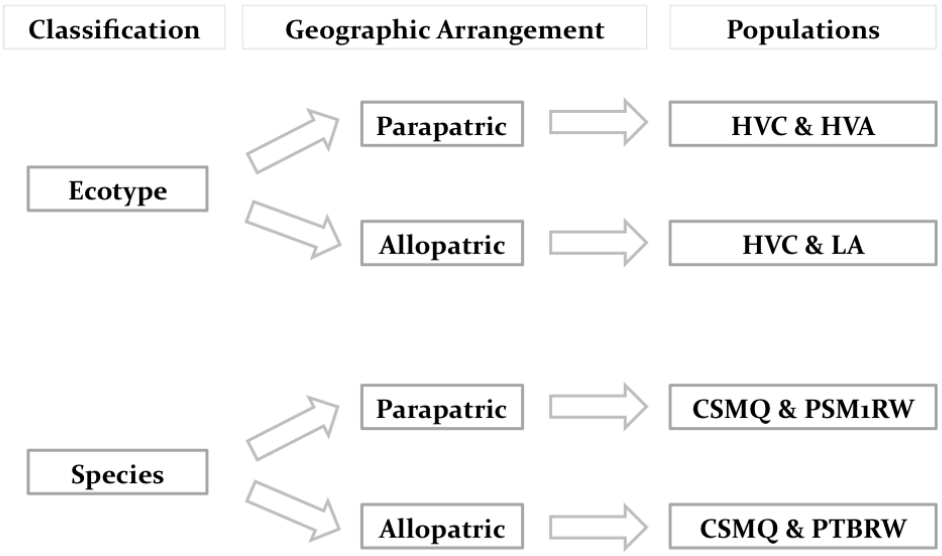


Figure 1. Study Design: Ecotype and species comparisons by populations based on geographic arrangement. The ecotypes of *Timema cristinae* are represented by three populations, two on the plant-host *Adenostoma* (HVA and LA) and one on *Ceanothus* (HVC). *Timema poppensis* is represented by two populations (PSM1RW and PTBRW) and *Timema californicum* is represented by one popuations (CSMQ).

Molecular Methods:

DNA was extracted from *Timema* legs using Quiagen DNEasy Blood and Tissue Kits, following the DNEasy protocol (Quiagen, Valencia, CA). Depending on the size and quantity of the legs available, samples were eluted to a final volume of 100 - 150 µl, so as to obtain optimal DNA concentrations. All extractions were checked using a NanoDrop (micro-sample

quantification instrument for nucleic acid concentration) (Thermo Scientific) to determine DNA concentration. Extraction products were kept frozen at -20°C for later use.

Nuclear and mitochondrial primers (short nucleic acid strands that designate the beginning and end of the amplification region) were developed based on mRNA sequences of *T. cristinae* obtained from 454 pyrosequencing of normalized cDNA libraries (P. Nosil, unpublished). Four nuclear loci (N1, N2, N3 and N4) and one mitochondrial locus, cytochrome oxidase 1 (CO1), were amplified in the six populations of *Timema* using polymerase chain reaction (PCR) techniques with the following reaction conditions: 5µl of 5x buffer, 0.25µl dNTPs (10mM), 2.5µM $MgCl_2$, 0.2µM forward and reverse primers, 0.25µl Taq polymerase; 23µl of this master mix and 2µl template DNA were used per reaction. All nuclear genes were amplified with a PCR cycling profile of 95°C for 3 minutes followed by 35 cycles of 95°C for 1 minute, 50°C for 40 seconds and 72°C for 45 seconds, finished with 72°C for 7 minutes. Mitochondrial CO1 was amplified with a PCR cycling profile of 94°C for 5 minutes followed by 30 cycles of 94°C for 1 minute, 43°C for 1 minute and 72°C for 50 seconds and finished with 72°C for 5 minutes. All PCR amplifications were executed using Mastercycler pro S (Eppendorf). The amplified products were run on 2% agarose gels, in combination with cybergreen and loading dye using standard electrophoresis techniques to check for successful amplification via ultra-violet imaging.

Table 1. Locus details and PCR conditions.
(L) indicates the sequence length

Locus Name	L	Primer Sequences (5' - 3')
Nuclear		
N1	425	F-ATCCTGGAATTCACGCACTTAC R-CTTACCCTTCTCCAAAATGTCG
N2	421	F-AAATTGCATTATCCAGGTCAGC R-GTTTATTCAGAAGCCCGTTGTC
N3	380	F-AGCCTCTATTGATCGTCGAAAC R-TCAATTTCACGATTTTCACTCG
N4	455	F-AAAGCACCCAAGAAGAATGTTG R-CAGCATGGTTGGTTCATAAGTG
Mitochondrial		
CO1	612	F-GGTCAACAAATCATAAAGATATTGG R-TAAACTTCAGGGTGCCAAAAAATCA

DNA sequencing was performed using traditional Sanger sequencing methods at Arizona State University (Tempe, AZ) (Sanger, Nicklen, & Coulson, 1977). Forward and reverse sequences were assembled into contigs, aligned and edited using the program Geneious 5.3.4 (Drummond, et al., 2010). Ambiguous, heterozygous loci in nuclear genes were parsed using Phase 2.1.1, which uses Bayesian methods for haplotype reconstruction from genetic sequence data (Stephens & Donnelly, 2003; Stephens, Smith, & Donnelly, 2001). It is important to note here that genetic sequence data for CO1 in the allopatric population of *T. poppensis* was unobtainable. This is likely due to the primers being developed based on *T. cristinae* mRNA sequences and potential mutations in *T. poppensis* at the priming region causing irregular or inconsistent amplification. This problem was also observed for 13 individuals from the *T. poppensis* population (PSM1RW) resulting in a markedly reduced sample size (n=15) for CO1.

Population Statistics:

Within population summary statistics, including sample size (n) the numbers of haplotypes (h), genetic diversity, percent of polymorphic sites (%S) and Tajima's D, a test of selective neutrality, were all calculated using Arliquin 3.5.1.2 (Excoffier & Lischer, 2010) (*see Table 2*). Tajima's D indicates whether a gene is evolving via random neutral processes (i.e. drift) or via nonrandom mechanisms (i.e. divergent selection). Values greater than two or less than minus two roughly indicate differentiation via nonrandom processes such as divergent selection or other non-random forms of differentiation (Tajima, 1989). Pairwise F_{ST} values were estimated using the same software to provide information about the degree of genetic differentiation between populations (Weir & Cockerham, 1984).

Phylogenetic Analyses:

A priori neighbor-joining trees were constructed for each locus using all populations of *Timema*. For the construction of these unrooted trees, the resampling method of bootstrapping with 1000 replicates was utilized in the program MEGA 5.0 (Tamura, et al., 2011). For all trees (e.g. at every locus) *T. cristinae* was reciprocally monophyletic from *T. poppensis/T. californicum* with 100% confidence, justifying subsequent phylogenetic analyses within in each clade (*T. cristinae* clade and *T. poppensis/T. californicum* clade) separately. In these subsequent analyses, *T. cristinae* was used to root the "*T. poppensis/T.californicum* trees" and *T. poppensis* was used to root "*T. cristinae* trees". The rooted gene trees for ecotype and species comparisons were constructed for all four nuclear loci and mitochondrial CO1 using Bayesian Markov chain Monte Carlo (MCMC) methods. These phylogenetic analyses were conducted to reconstruct phylogenies using MrBayes v3.0b4 (Ronquist & Huelsenbeck, 2003), implementing the GTR + I + Γ model of evolution.

Two simultaneous runs of four MCMC analyses, consisting of one cold and three incrementally heated chains, were initiated with random trees for a total of 5.0×10^5 generations (sampling every 100 generations). The first 2.5×10^5 generations from each run were discarded as burn-in.

Population Structure:

The program Structure 2.3.3 was used to visualize the genetic structure of the six populations. An admixture model of ancestry was used so as to assume each individual consists of DNA from any of the defined populations. Additionally, a model of independent allele frequencies was assumed in order to allow for different populations to have different allelic frequencies (Pritchard, Stephens, & Donnelly, 2000). The parameter (K) indicates the number of predefined genetic clusters. For these analyses K = 3 and K = 6 were used due to the biological justifications: three species and six populations, respectively. Analyses were executed using 100,000 replicates and a 10,000 burn-in.

Results

Within Population Genetic Diversity:

Within population summary statistics reveal significant genetic variability within each population, and also some differences in the level of variability among the different gene loci (*see Table 2).* For example, percent polymorphic sites (number of variable sites over sequence length) were highest in mitochondrial locus CO1 and lowest in nuclear loci N3 and N4. Also, the number of haplotypes for each locus and population were respectively higher in nuclear genes (e.g. N1 and N2) than the mitochondrial gene. Also, Tajima's D values generally fell within the range of expectation for neutrally evolving loci (i.e. between two and minus two, *Table 2*).

Table 2. Within population descriptive statistics including sample size (n), number of haplotypes (h), genetic diversity, percent polymorphic sites (%S) and Tajima's D.

Locus	Population	n	*h*	Genetic Diversity	%S	Tajima's D
Nuclear						
235	HVC	50	26	0.93 ± 0.02	4.71	-1.24
	HVA	52	24	0.94 ± 0.02	4.24	-0.94
	LA	46	17	0.93 ± 0.02	5.41	-1.28
	CSMQ	50	5	0.65 ± 0.04	0.94	1.79
	PSM1RW	54	2	0.11 ± 0.06	0.47	-0.90
	PTBDF	48	1	0.00 ± 0.00	0.00	0.00
9172	HVC	44	8	0.63 ± 0.08	2.61	-1.48
	HVA	50	9	0.67 ± 0.07	2.85	-1.14
	LA	44	8	0.76 ± 0.05	2.38	-1.10
	CSMQ	46	4	0.24 ± 0.08	0.95	-1.45
	PSM1RW	54	2	0.07 ± 0.05	0.24	-0.88
	PTBDF	48	1	0.00 ± 0.00	0.00	0.00
10854	HVC	48	3	0.20 ± 0.07	0.53	-1.00
	HVA	52	3	0.21 ± 0.07	0.53	-0.89
	LA	44	3	0.28 ± 0.08	0.53	-0.66
	CSMQ	50	2	0.18 ± 0.07	0.26	-0.24
	PSM1RW	52	2	0.27 ± 0.07	0.26	0.27
	PTBDF	48	2	0.51 ± 0.02	0.53	2.27
11144	HVC	48	1	0.00 ± 0.00	0.00	0.00
	HVA	52	2	0.04 ± 0.04	0.22	-1.10
	LA	42	2	0.09 ± 0.06	0.22	-0.85
	CSMQ	48	2	0.04 ± 0.04	0.88	-1.87
	PSM1RW	48	3	0.45 ± 0.06	0.44	1.64
	PTBDF	44	1	0.00 ± 0.00	0.00	0.00
Mitochondrial						
CO1	HVC	25	7	0.81 ± 0.05	8.006536	1.44
	HVA	26	7	0.85 ± 0.03	8.006536	2.10
	LA	21	5	0.64 ± 0.10	7.51634	0.51
	CSMQ	25	8	0.71 ± 0.09	1.960784	-0.06
	PSM1RW	15	4	0.72 ± 0.08	1.470588	1.54
	PTBDF	N/A	N/A	N/A	N/A	N/A

Between Population Genetic Differentiation:

Estimates of population differentiation are presented in Table 3. For nuclear genes, F_{ST} values (indicating the level of genetic differentiation) between both the parapatric pair and allopatric pair of *T. cristinae* illustrate no significant differentiation. However, for

mitochondrial CO1, the F_{ST} values indicate significant differentiation between allopatric *T. cristinae* populations (0.34), but no differentiation between parapatric populations (not significantly different from 0). In contrast to what was observed between ecotypes of *T. cristinae*, F_{ST} estimates between the species pair (*T. californicum* and *T. poppensis*) illustrate significant and very strong genetic differentiation at all nuclear loci between both the parapatric and the allopatric pairs (ranging from 0.72 to 0.93). In contrast to nuclear differentiation between species, mitochondrial CO1 was weakly differentiated between *T. californicum* and *T. poppensis (see Table 3).*

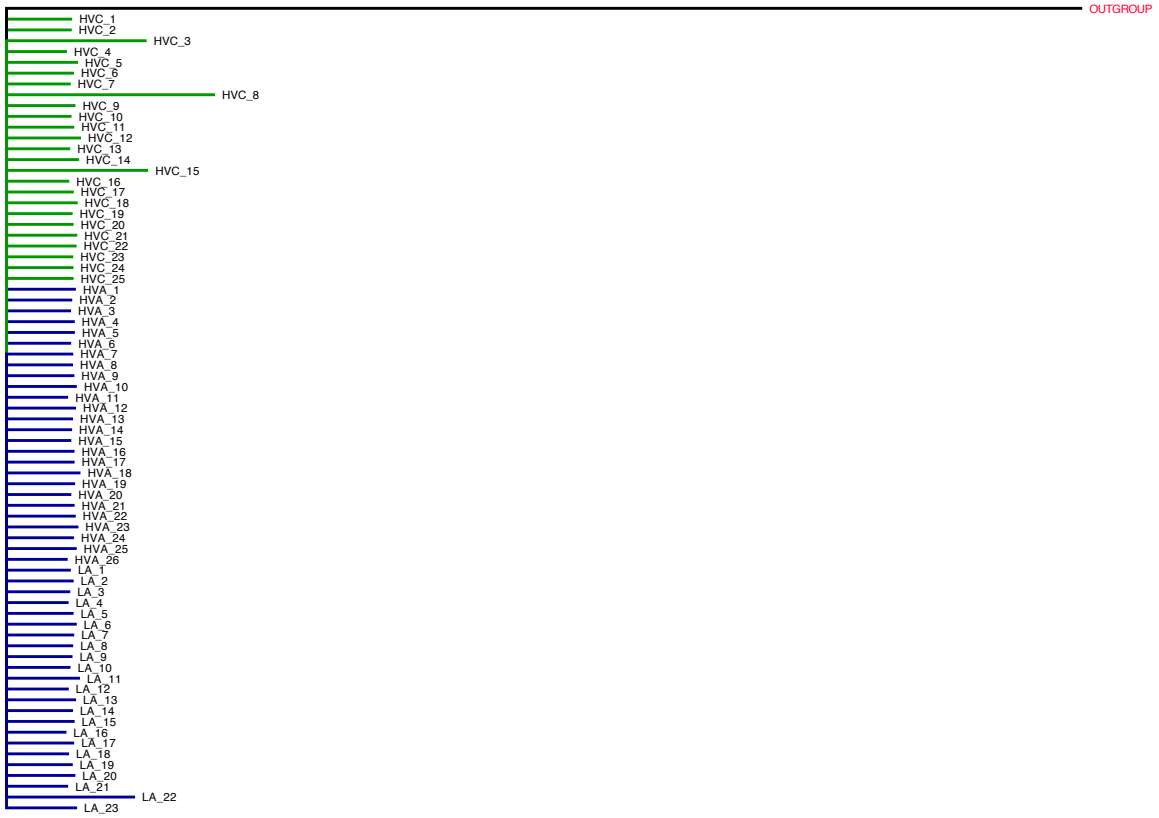
Table 3. F_{ST}: The inbreeding coefficent, a measure of genetic divergence.
* indicates significance (e.g. significantly different from zero) (p-value < 0.05)

Population Pair	N1	N2	N3	N4	CO1
HVC x HVA	0.016	-0.008	-0.004	-0.002	0.03
HVC x LA	-0.013	0.026	0.004	0.029	0.343*
CSMQ x PSM1RW	0.924*	0.904*	0.884*	0.896*	0.331*
CSMQ x PTBDF	0.929*	0.917*	0.717*	0.982*	N/A

Gene Trees:

Consistent with our findings from F_{ST}, nuclear gene trees show no significant differentiation among all *T. cristinae* populations, illustrated by a comb-like tree structure at all nuclear loci (*see Figures 2-5*). In contrast, there was some differentiation among conspecific populations of *T. cristinae* in the mitochondrial gene tree, grouping most of the allopatric population (LA) together, but with both parapatric populations and LA being nonetheless mixed, suggesting the early stages of genetic divergence at CO1 in allopatry.

Nuclear gene trees for the *T. poppensis/T. califronicum* species pair comparison show complete differentiation, or species-level reciprocal monophyly, for all nuclear loci *(see Figures 2-5)*. In contrast, the mitochondrial gene tree for CO1 suggests a mixed relationship between the species, *T. californicum* and *T. poppensis*, consistent with the findings of Law and Crespi, 2002 *(see Figure 6)*.


OUTGROUP
HVC_1
HVC_2
HVC_3
HVC_4
HVC_5
HVC_6
HVC_7
HVC_8
HVC_9
HVC_10
HVC_11
HVC_12
HVC_13
HVC_14
HVC_15
HVC_16
HVC_17
HVC_18
HVC_19
HVC_20
HVC_21
HVC_22
HVC_23
HVC_24
HVC_25
HVA_1
HVA_2
HVA_3
HVA_4
HVA_5
HVA_6
HVA_7
HVA_8
HVA_9
HVA_10
HVA_11
HVA_12
HVA_13
HVA_14
HVA_15
HVA_16
HVA_17
HVA_18
HVA_19
HVA_20
HVA_21
HVA_22
HVA_23
HVA_24
HVA_25
HVA_26
LA_1
LA_2
LA_3
LA_4
LA_5
LA_6
LA_7
LA_8
LA_9
LA_10
LA_11
LA_12
LA_13
LA_14
LA_15
LA_16
LA_17
LA_18
LA_19
LA_20
LA_21
LA_22
LA_23
0.2

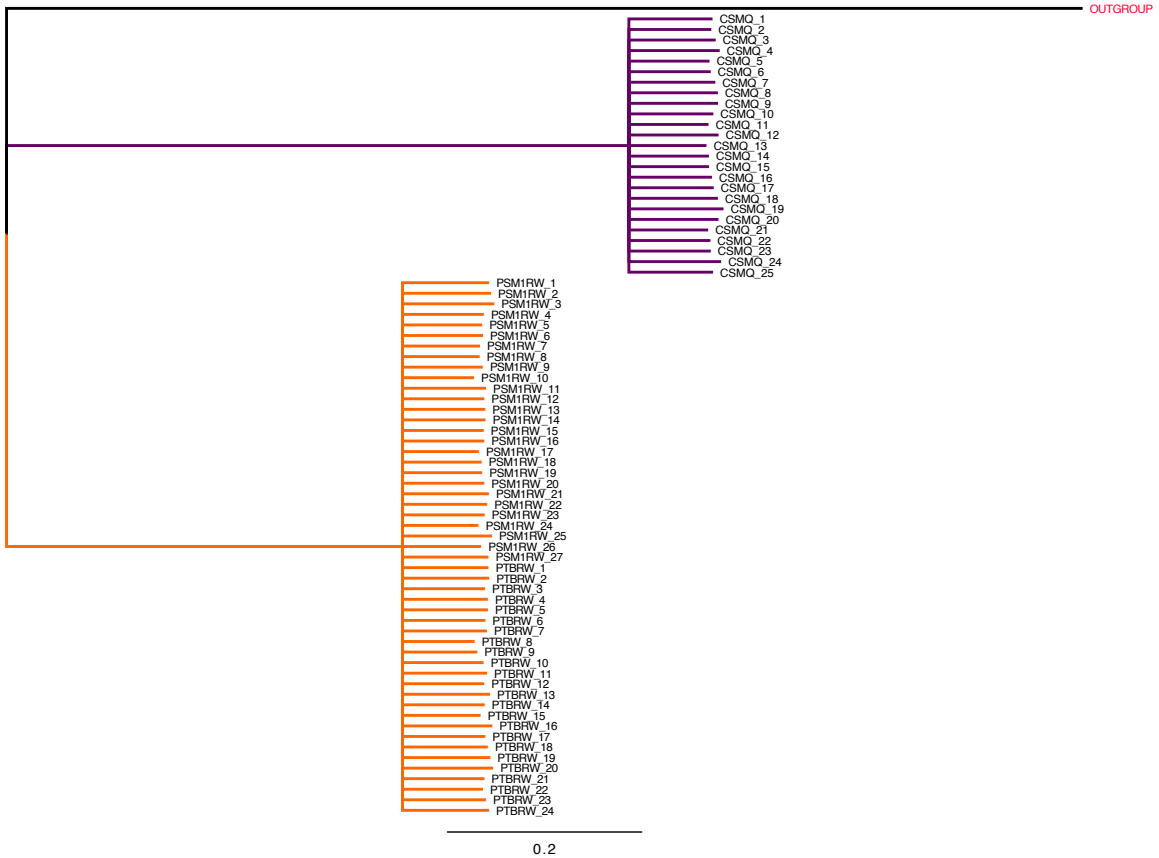

OUTGROUP
CSMQ_1
CSMQ_2
CSMQ_3
CSMQ_4
CSMQ_5
CSMQ_6
CSMQ_7
CSMQ_8
CSMQ_9
CSMQ_10
CSMQ_11
CSMQ_12
CSMQ_13
CSMQ_14
CSMQ_15
CSMQ_16
CSMQ_17
CSMQ_18
CSMQ_19
CSMQ_20
CSMQ_21
CSMQ_22
CSMQ_23
CSMQ_24
CSMQ_25
PSM1RW_1
PSM1RW_2
PSM1RW_3
PSM1RW_4
PSM1RW_5
PSM1RW_6
PSM1RW_7
PSM1RW_8
PSM1RW_9
PSM1RW_10
PSM1RW_11
PSM1RW_12
PSM1RW_13
PSM1RW_14
PSM1RW_15
PSM1RW_16
PSM1RW_17
PSM1RW_18
PSM1RW_19
PSM1RW_20
PSM1RW_21
PSM1RW_22
PSM1RW_23
PSM1RW_24
PSM1RW_25
PSM1RW_26
PSM1RW_27
PTBRW_1
PTBRW_2
PTBRW_3
PTBRW_4
PTBRW_5
PTBRW_6
PTBRW_7
PTBRW_8
PTBRW_9
PTBRW_10
PTBRW_11
PTBRW_12
PTBRW_13
PTBRW_14
PTBRW_15
PTBRW_16
PTBRW_17
PTBRW_18
PTBRW_19
PTBRW_20
PTBRW_21
PTBRW_22
PTBRW_23
PTBRW_24
0.2

Figure 2. Gene trees for the ecotypes of *T. cristinae (top)* and the species pair *T. californicum and T. poppensis (bottom)* constructed for the nuclear locus 1 (N1). Green – *Ceanothus* ecotype; Blue – *Adenostoma* ecotype; Purple – *T. californicum*; Orange – *T. poppensis.*

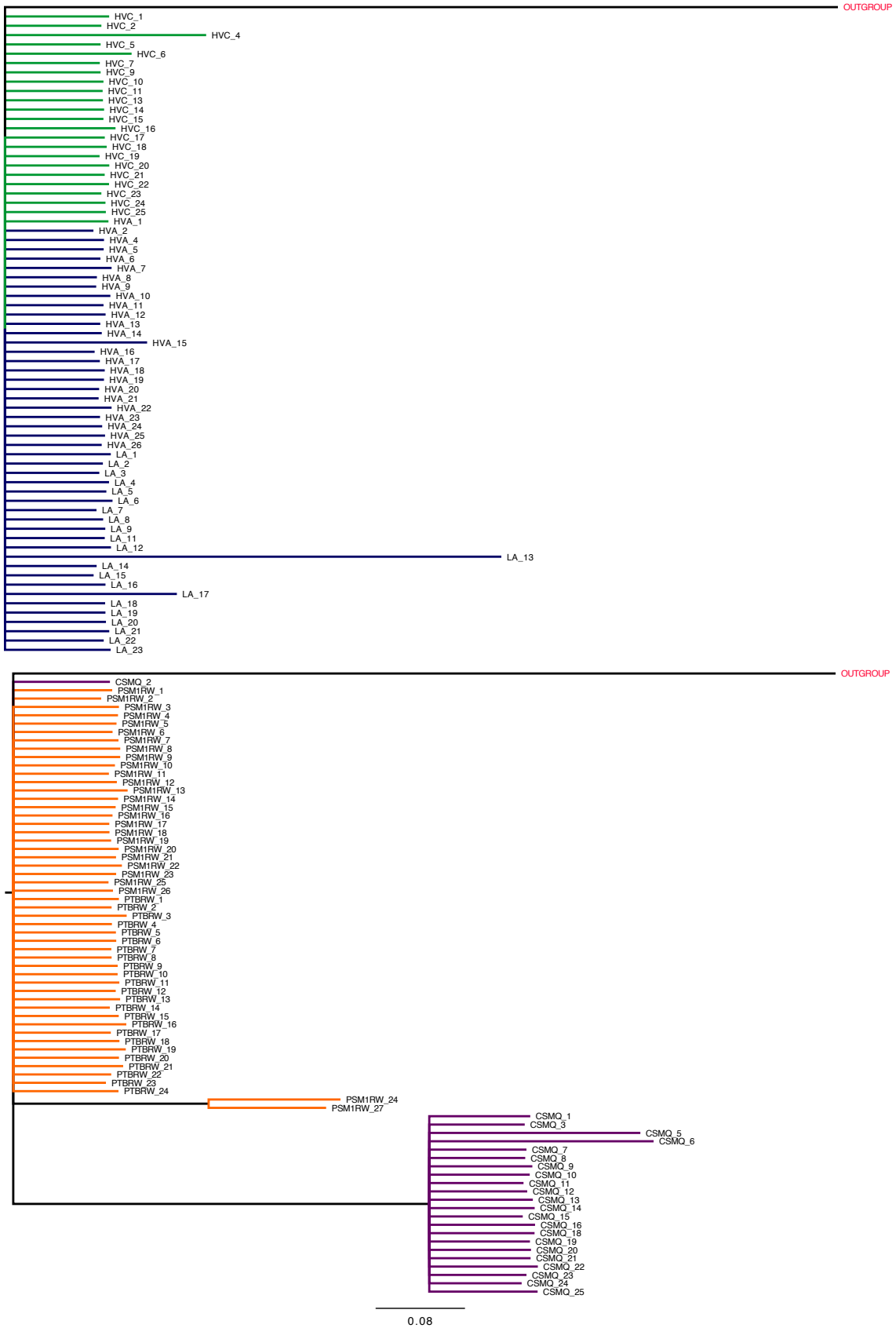
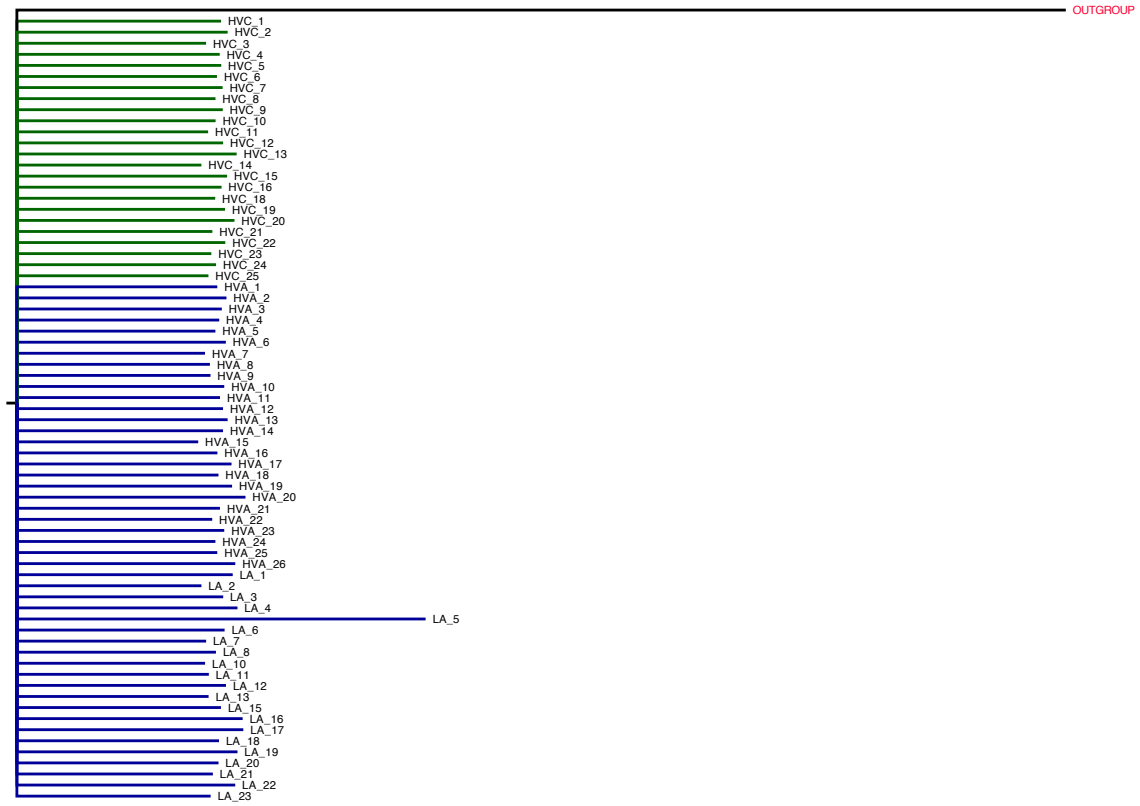

OUTGROUP
HVC_1
HVC_2
HVC_4
HVC_5
HVC_6
HVC_7
HVC_9
HVC_10
HVC_11
HVC_13
HVC_14
HVC_15
HVC_16
HVC_17
HVC_18
HVC_19
HVC_20
HVC_21
HVC_22
HVC_23
HVC_24
HVC_25
HVA_1
HVA_2
HVA_4
HVA_5
HVA_6
HVA_7
HVA_8
HVA_9
HVA_10
HVA_11
HVA_12
HVA_13
HVA_14
HVA_15
HVA_16
HVA_17
HVA_18
HVA_19
HVA_20
HVA_21
HVA_22
HVA_23
HVA_24
HVA_25
HVA_26
LA_1
LA_2
LA_3
LA_4
LA_5
LA_6
LA_7
LA_8
LA_9
LA_11
LA_12
LA_13
LA_14
LA_15
LA_16
LA_17
LA_18
LA_19
LA_20
LA_21
LA_22
LA_23
OUTGROUP
CSMQ_2
PSM1RW_1
PSM1RW_2
PSM1RW_3
PSM1RW_4
PSM1RW_5
PSM1RW_6
PSM1RW_7
PSM1RW_8
PSM1RW_9
PSM1RW_10
PSM1RW_11
PSM1RW_12
PSM1RW_13
PSM1RW_14
PSM1RW_15
PSM1RW_16
PSM1RW_17
PSM1RW_18
PSM1RW_19
PSM1RW_20
PSM1RW_21
PSM1RW_22
PSM1RW_23
PSM1RW_25
PSM1RW_26
PTBRW_1
PTBRW_2
PTBRW_3
PTBRW_4
PTBRW_5
PTBRW_6
PTBRW_7
PTBRW_8
PTBRW_9
PTBRW_10
PTBRW_11
PTBRW_12
PTBRW_13
PTBRW_14
PTBRW_15
PTBRW_16
PTBRW_17
PTBRW_18
PTBRW_19
PTBRW_20
PTBRW_21
PTBRW_22
PTBRW_23
PTBRW_24
PSM1RW_24
PSM1RW_27
CSMQ_1
CSMQ_3
CSMQ_5
CSMQ_6
CSMQ_7
CSMQ_8
CSMQ_9
CSMQ_10
CSMQ_11
CSMQ_12
CSMQ_13
CSMQ_14
CSMQ_15
CSMQ_16
CSMQ_18
CSMQ_19
CSMQ_20
CSMQ_21
CSMQ_22
CSMQ_23
CSMQ_24
CSMQ_25
0.08

Figure 3. Gene trees for the ecotypes of *T. cristinae (top)* and the species pair *T. californicum and T. poppensis (bottom)* constructed for the nuclear locus 2 (N2). Green – *Ceanothus* ecotype; Blue – *Adenostoma* ecotype; Purple – *T. californicum*; Orange – *T. poppensis.*


OUTGROUP
HVC_1
HVC_2
HVC_4
HVC_5
HVC_6
HVC_7
HVC_8
HVC_9
HVC_10
HVC_11
HVC_12
HVC_13
HVC_14
HVC_15
HVC_16
HVC_17
HVC_18
HVC_19
HVC_20
HVC_21
HVC_22
HVC_23
HVC_24
HVC_25
HVA_1
HVA_2
HVA_3
HVA_4
HVA_5
HVA_6
HVA_7
HVA_8
HVA_9
HVA_10
HVA_11
HVA_12
HVA_13
HVA_14
HVA_15
HVA_16
HVA_17
HVA_18
HVA_19
HVA_20
HVA_21
HVA_22
HVA_23
HVA_24
HVA_25
HVA_26
LA_1
LA_2
LA_3
LA_4
LA_5
LA_6
LA_7
LA_8
LA_9
LA_10
LA_11
LA_13
LA_14
LA_15
LA_16
LA_17
LA_18
LA_19
LA_20
LA_21
LA_22
LA_23
0.06
OUTGROUP
PSM1RW_1
PSM1RW_2
PSM1RW_3
PSM1RW_4
PSM1RW_6
PSM1RW_7
PSM1RW_8
PSM1RW_9
PSM1RW_10
PSM1RW_11
PSM1RW_12
PSM1RW_13
PSM1RW_14
PSM1RW_15
PSM1RW_16
PSM1RW_17
PSM1RW_18
PSM1RW_19
PSM1RW_20
PSM1RW_21
PSM1RW_22
PSM1RW_23
PSM1RW_24
PSM1RW_25
PSM1RW_26
PSM1RW_27
PTBRW_1
PTBRW_2
PTBRW_3
PTBRW_4
PTBRW_5
PTBRW_6
PTBRW_7
PTBRW_8
PTBRW_9
PTBRW_10
PTBRW_11
PTBRW_12
PTBRW_13
PTBRW_14
PTBRW_15
PTBRW_16
PTBRW_17
PTBRW_18
PTBRW_19
PTBRW_20
PTBRW_21
PTBRW_22
PTBRW_23
PTBRW_24
CSMQ_1
CSMQ_2
CSMQ_3
CSMQ_4
CSMQ_5
CSMQ_6
CSMQ_7
CSMQ_8
CSMQ_9
CSMQ_10
CSMQ_11
CSMQ_12
CSMQ_13
CSMQ_14
CSMQ_15
CSMQ_16
CSMQ_17
CSMQ_18
CSMQ_19
CSMQ_20
CSMQ_21
CSMQ_22
CSMQ_23
CSMQ_24
CSMQ_25
0.06

Figure 4. Gene trees for the ecotypes of *T. cristinae (top)* and the species pair *T. californicum and T. poppensis (bottom)* constructed for the nuclear locus 3 (N3). Green – *Ceanothus* ecotype; Blue – *Adenostoma* ecotype; Purple – *T. californicum*; Orange – *T. poppensis.*


OUTGROUP
HVC_1
HVC_2
HVC_3
HVC_4
HVC_5
HVC_6
HVC_7
HVC_8
HVC_9
HVC_10
HVC_11
HVC_12
HVC_13
HVC_14
HVC_15
HVC_16
HVC_18
HVC_19
HVC_20
HVC_21
HVC_22
HVC_23
HVC_24
HVC_25
HVA_1
HVA_2
HVA_3
HVA_4
HVA_5
HVA_6
HVA_7
HVA_8
HVA_9
HVA_10
HVA_11
HVA_12
HVA_13
HVA_14
HVA_15
HVA_16
HVA_17
HVA_18
HVA_19
HVA_20
HVA_21
HVA_22
HVA_23
HVA_24
HVA_25
HVA_26
LA_1
LA_2
LA_3
LA_4
LA_5
LA_6
LA_7
LA_8
LA_10
LA_11
LA_12
LA_13
LA_15
LA_16
LA_17
LA_18
LA_19
LA_20
LA_21
LA_22
LA_23
0.05

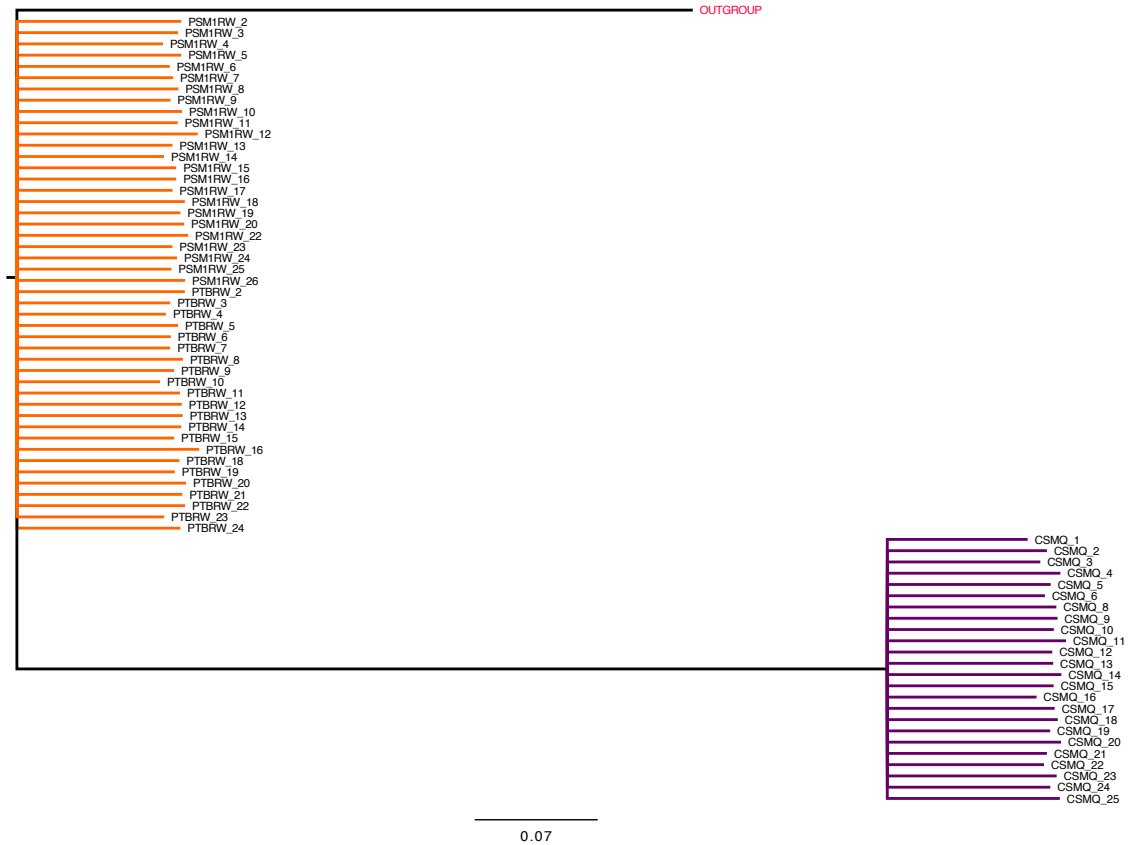
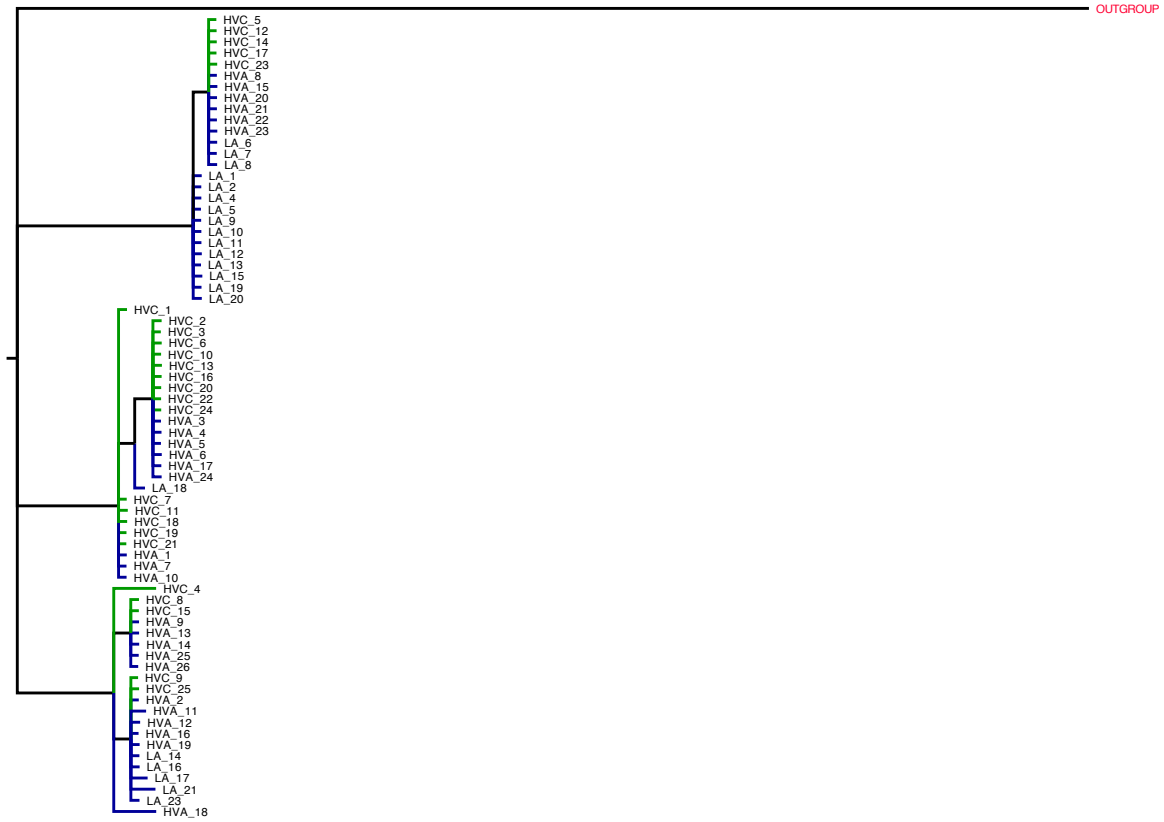

OUTGROUP
PSM1RW_2
PSM1RW_3
PSM1RW_4
PSM1RW_5
PSM1RW_6
PSM1RW_7
PSM1RW_8
PSM1RW_9
PSM1RW_10
PSM1RW_11
PSM1RW_12
PSM1RW_13
PSM1RW_14
PSM1RW_15
PSM1RW_16
PSM1RW_17
PSM1RW_18
PSM1RW_19
PSM1RW_20
PSM1RW_22
PSM1RW_23
PSM1RW_24
PSM1RW_25
PSM1RW_26
PTBRW_2
PTBRW_3
PTBRW_4
PTBRW_5
PTBRW_6
PTBRW_7
PTBRW_8
PTBRW_9
PTBRW_10
PTBRW_11
PTBRW_12
PTBRW_13
PTBRW_14
PTBRW_15
PTBRW_16
PTBRW_18
PTBRW_19
PTBRW_20
PTBRW_21
PTBRW_22
PTBRW_23
PTBRW_24
CSMQ_1
CSMQ_2
CSMQ_3
CSMQ_4
CSMQ_5
CSMQ_6
CSMQ_8
CSMQ_9
CSMQ_10
CSMQ_11
CSMQ_12
CSMQ_13
CSMQ_14
CSMQ_15
CSMQ_16
CSMQ_17
CSMQ_18
CSMQ_19
CSMQ_20
CSMQ_21
CSMQ_22
CSMQ_23
CSMQ_24
CSMQ_25
0.07

Figure 5. Gene trees for the ecotypes of *T. cristinae (top)* and the species pair *T. californicum and T. poppensis (bottom)* constructed for the nuclear locus 4 (N4). Green – *Ceanothus* ecotype; Blue – *Adenostoma* ecotype; Purple – *T. californicum*; Orange – *T. poppensis.*


OUTGROUP
HVC_5
HVC_12
HVC_14
HVC_17
HVC_23
HVA_8
HVA_15
HVA_20
HVA_21
HVA_22
HVA_23
LA_6
LA_7
LA_8
LA_1
LA_2
LA_4
LA_5
LA_9
LA_10
LA_11
LA_12
LA_13
LA_15
LA_19
LA_20
HVC_1
HVC_2
HVC_3
HVC_6
HVC_10
HVC_13
HVC_16
HVC_20
HVC_22
HVC_24
HVA_3
HVA_4
HVA_5
HVA_6
HVA_17
HVA_24
LA_18
HVC_7
HVC_11
HVC_18
HVC_19
HVC_21
HVA_1
HVA_7
HVA_10
HVC_4
HVC_8
HVC_15
HVA_9
HVA_13
HVA_14
HVA_25
HVA_26
HVC_9
HVC_25
HVA_2
HVA_11
HVA_12
HVA_16
HVA_19
LA_14
LA_16
LA_17
LA_21
LA_23
HVA_18
0.6

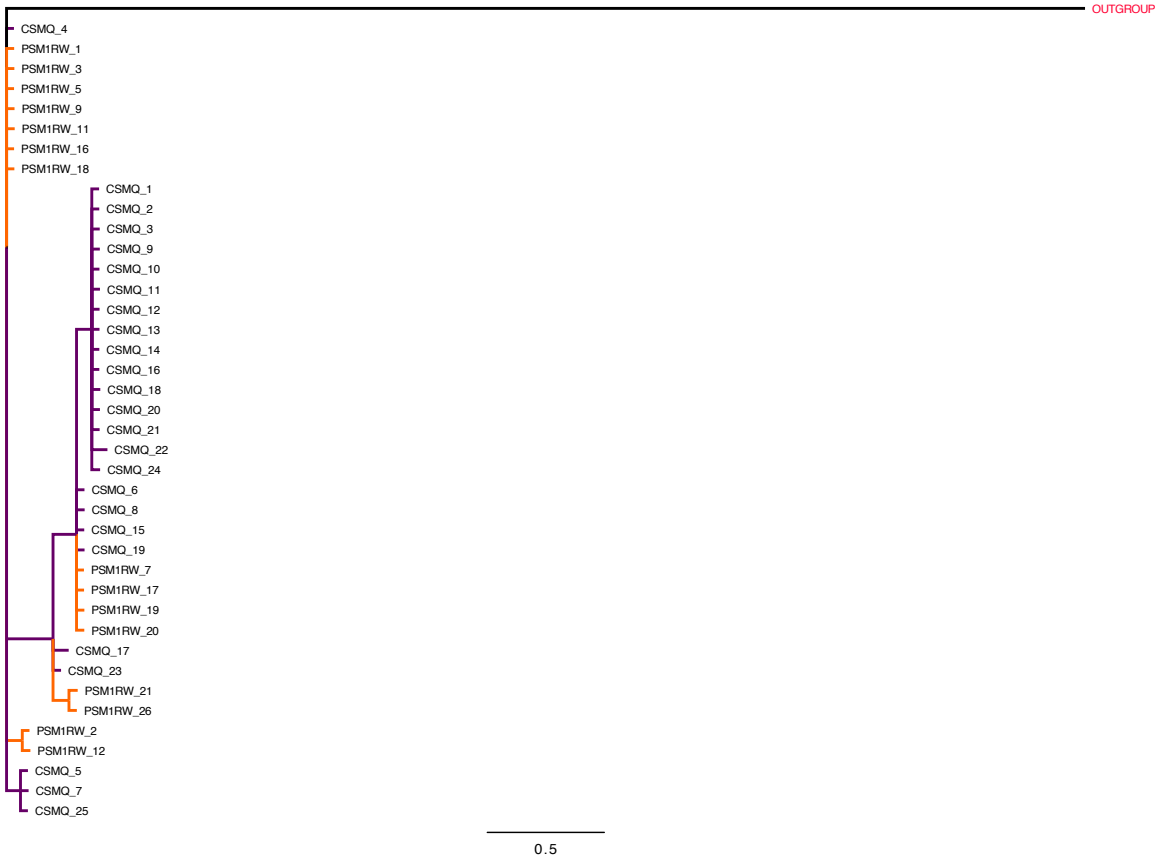
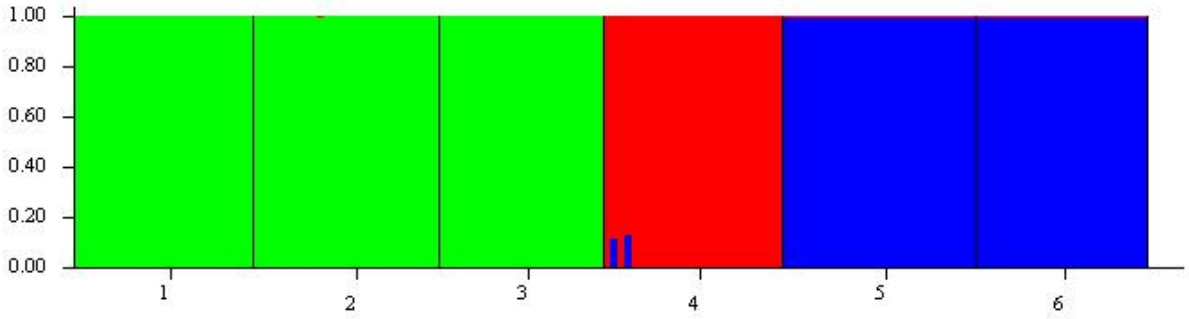

OUTGROUP
CSMQ_4
PSM1RW_1
PSM1RW_3
PSM1RW_5
PSM1RW_9
PSM1RW_11
PSM1RW_16
PSM1RW_18
CSMQ_1
CSMQ_2
CSMQ_3
CSMQ_9
CSMQ_10
CSMQ_11
CSMQ_12
CSMQ_13
CSMQ_14
CSMQ_16
CSMQ_18
CSMQ_20
CSMQ_21
CSMQ_22
CSMQ_24
CSMQ_6
CSMQ_8
CSMQ_15
CSMQ_19
PSM1RW_7
PSM1RW_17
PSM1RW_19
PSM1RW_20
CSMQ_17
CSMQ_23
PSM1RW_21
PSM1RW_26
PSM1RW_2
PSM1RW_12
CSMQ_5
CSMQ_7
CSMQ_25
0.5

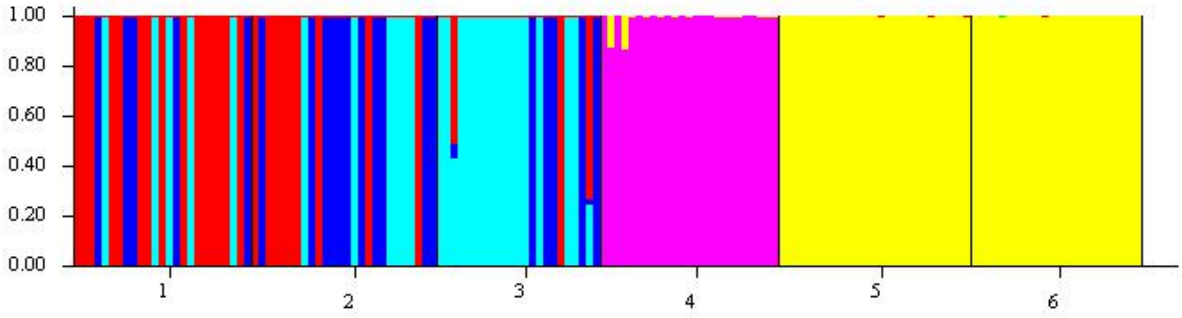
Figure 6. Gene trees for the ecotypes of *T. cristinae (top)* and the species pair *T. californicum and T. poppensis (bottom)* constructed for the mitochondrial locus CO1. Green – *Ceanothus* ecotype; Blue – *Adenostoma* ecotype; Purple – *T. californicum*; Orange – *T. poppensis.*

Population Structure:

Using the prior of K = 3, the program Structure strongly groups all species together. Notably, two individuals of *T. californicum* exhibit some genetic variation that is reflective of variation predominately observed in *T. poppensis.* Using the prior K = 6, the program Structure reflects the same grouping for *T. poppensis and T. californicum* as observed for K = 3, but attempts to break up the individuals of *T. cristinae* into multiple genetic clusters. Interestingly, the allopatric population of *T. cristinae* (LA) somewhat forms its own cluster (light blue). However, the ecotypes of *T. cristinae* nonetheless remain genetically mixed *(see Figure 7)*.



K = 3



K = 6

Figure 7. Structure output for the number of predefined genetic clusters of K = 3 clusters and K = 6 clusters. Each individual is represented by a single bar. Each bar reflects the proportion of that individual's genotype (based on aggregated polymorphic site data), which are derived from any given genetic group. There were 254 total variable loci across all populations and all genes. The population individuals originally derived from are labeled below each plot: (1) HVC, (2) HVA, (3) LA, (4) CSMQ, (5) PSM1RW, (6) PTBRW.

Discussion

Past genetic analyses of speciation in *Timema* focused primarily on a single species, *T. cristinae* (Nosil, Egan, & Funk, 2008), and were based on genotypic or single gene DNA data (Law & Crespi, 2002), rather than multilocus genetic data. I presented here a multilocus genetic analysis of the genetic divergence within and between three species of *Timema* walking sticks. Variation was observed both among genes and among taxa. I first discuss patterns of variation within populations and then turn to the focus of the study: genetic divergence between populations and species. I conclude by highlighting particularly pressing avenues for future research.

Within-population genetic variation

Levels of within-population genetic diversity varied quite substantially both among genes and among the populations of *Timema* walking sticks examined in this study. In relation to the variation among genes, the high levels of genetic diversity in the mitochondrial gene CO1 are consistent with prior empirical findings in other study systems (Brown, et al., 1979; Lin & Danforth, 2004) and with the argument that mitochondrial DNA will tend to be more variable than nuclear DNA because mitochondrial DNA mutates faster and accumulates mutation faster (Brown, et al., 1979; Lin & Danforth, 2004). These data illustrate the root of genealogical discordance through differential variation across multiple loci and especially so between CO1 and the nuclear loci.

Between-population genetic differentiation

The differences between F_{ST} estimates at mitochondrial CO1 and nuclear genes as well as differences based on the geographic arrangement of populations reveal variable patterns of genetic differentiation and potential gene flow. The high levels of differentiation at CO1

between allopatric populations of *T. cristinae* compared with low levels of differentiation at CO1 between parapatric *T. cristinae* are likely due to two reasons: (1) the high rates of mitochondrial differentiation when gene flow is limited or absent, (2) in conjunction with low levels of differentiation between parapatric populations due to homogenizing gene flow (Nosil, et al., 2003). The lack of nuclear differentiation between both allopatric and parapatric pairs indicate that the ecotypes of *T. cristinae* are not highly differentiated, at least for the sample of nuclear genes examined in this study. Thus, genetic differentiation between phenotypically differentiated and partially-reproductively isolated ecotypes is likely restricted to a few genes, especially (or perhaps only) "speciation genes" subject to ecologically based divergent selection (Coyne & Orr, 2004; Nosil & Schluter, 2011; Wu, 2001). Consistent with this argument, previous research on heterogeneous genomic divergence among more than 500 AFLP loci detected outlier loci (using F_{ST}) whose genetic differentiation exceed neutral expectations (Nosil, et al., 2008). These outlier loci are likely affected by divergent selection and are thus driving the ecological speciation of *T. cristinae* and are the cause of initial reproductive isolation and assortative mating.

F_{ST} estimates for *T. californicum* and *T. poppensis* indicate high levels of genetic differentiation between species at all nuclear loci in both parapatry and allopatry. Assuming that mitochondrial DNA diverges more quickly, one would expect equivalent or higher F_{ST} values for CO1 versus nuclear genes. However, we see lower F_{ST} at CO1 for *T. californicum* and *T. poppensis* in parapatry. This finding is highly suggestive of mitochondrial gene flow between these species; gene flow that results in reduced differentiation of the mitochondrial genome. As mentioned above, sequence data for CO1 from the allopatric population of *T. poppensis* and some individuals from the parapatric population of *T. poppensis* were unobtainable. This is

likely due to mutations in the priming region (since primers were developed from *T. cristinae* mRNA sequences) in most individuals from *T. poppensis.* However, since some sequences were obtainable for the parapatric population, and none for the allopatric population, it appears that there may be more differentiation in the allopatric population than the parapatric population, but these results remain inconclusive.

Gene trees and analyses of genetic structure both provide further support for our findings based on F_{ST}. Nuclear gene trees illustrate that species are more diverged than ecotypes. Also, population structure analyses with K=3 illustrate this point very clearly, strongly grouping individuals of the same *species* together. With K=6, we see the initiation of divergence between allopatric populations of *T. cristinae.* While previous research suggests ecologically based divergent selection as the driving force of genetic differentiation, past work shows that this process of ecological speciation can be constrained by gene flow and thus promoted by geographic separation (Nosil, 2007; Nosil & Crespi, 2004; Nosil, et al., 2008). The result of increased clustering of the allopatric population of *T. cristinae* is exaggerated by (1) the fact that mitochondrial DNA contributed the majority of polymorphic sites used to conduct the genetic structure analyses; (2) and the elevated rate of mitochondrial evolution. These two factors are indicated by the high F_{ST} value for the allopatric ecotype comparison *(Table 3)* and greater percent polymorphic site (%S) *(Table 2).* One final notable result is the appearance of two individuals from the population of *T. californicum* with mixed genetic structure *(see Figure 7).* This further suggests a marginal amount of gene flow between parapatric *T. californicum* and *T. poppensis,* also indicated by a low mitochondrial F_{ST} estimate. Additionally, one individual showed genetic characteristics of a hybrid, exhibited by the presence of haplotypes

from both populations. Future analyses using the program "NewHybrids" can test this explicitly (Anderson & Thompson, 2002).

Conclusions, Limitations and Future Directions

As one would expect from the general assumption of increased genetic differentiation as speciation unfolds, we found ecotypes to be significantly less differentiated than species. Interestingly however, taxon pairs subject to geographic isolation (i.e. allopatry) did not predict higher levels of nuclear differentiation than their parapatric counterparts (for both species and ecotypes). This finding supports the hypothesis that ecologically based divergent selection is the key driver of genetic differentiation between host-plant associated ecotypes of *Timema cristinae* and also that selection is only acting on a few speciation genes. Additionally, the mitochondrial gene CO1 appears to be "flowing" readily in parapatry for diverging ecotypes and recently diverged species. This finding is not uncommon. Ferris *et al*. (1983) found mitochondrial DNA flowing between different species of mice and comments that, if this is common, it may require a redefinition of a biological species to specify the use of only nuclear genes. Nearly 30 years have passed since the latter paper was published and despite the fact that multiple other studies have demonstrated the same phenomenon, no redefinition of a biological species has been made, much less the formation of a single cohesive and all inclusive definition of a "species".

Genealogical discordance was observed here, especially between mitochondrial and nuclear gene tress, thus emphasizing the importance of multilocus genetic data for

phylogenetic purposes. While the species level mitochondrial gene tree is consistent with the finding of Law and Crespi (2002), which indicates a mixed phylogenetic relationship between *T. califronicum* and *T. poppensis*, the inclusion of four nuclear genes provides sufficient evidence for the separation of these taxa as distinct species. The results for CO1 are consistent with the general patterns observed in animal taxa. Funk and Omland (2003) reviewed studies of mitochondrial DNA divergence using 2319 assayed species and found that 23 percent of examined *distinct species pairs* were not reciprocally monophyletic based on mitochondrial DNA. Overall, these results indicate that genetic studies relying on a single locus may result in misleading conclusions.

Further analyses with the taxa employed here are necessary for clarifying our understanding of ecological speciation. Multilocus analyses with a larger sample from populations of *T. cristinae* are requisite for more precisely investigating how much genetic differentiation is reflected at neutral loci. Although genetic differentiation may be minimal at these loci, obtaining significant F_{ST} estimates will provide important information about background differentiation along the speciation continuum for genes not directly involved in initial divergence but that may play a role in the eventual complete reproductive isolation of two diverging populations. Additionally, statistically based estimates of gene flow need to be made using "Isolation with Migration" models (Nielsen & Wakeley, 2001). While F_{ST} can provide useful information about genetic divergence, it does not allow us to differentiate between recent divergence, gene flow or some combination of the two. Another important direction for further research is the development of primers for CO1 based on *T. poppensis.* The unsuccessful sequencing of CO1 in most individuals of *T. poppensis* confers the need for new primers, which would allow for complete analyses across all populations and all genes.

Finally, although it is a lofty goal, we need to reach a better solution to the "species problem." The use of a singular working definition of a "species" would immensely benefit the fields of evolutionary biology, ecology, and conservation (Mallet, 2007) and would especially be useful when studying evolutionary divergence and how species are formed. The most common approach to a solution is to aggregate many of the working concepts of a species (Mallet, 1995). For example, one could use information about morphology, ecological role, and amount of genomic differentiation to define a species. However, all of these realms of distinction require a fairly arbitrary line to be drawn. How different do *species* have to look, or behave? How genetically differentiated must *species* be? After conducting this study, it is clear to me why the definition of a species is so unclear, especially on the genetic level. Depending on where in the genome one looks, or what genome one looks at, varying levels of genetic differentiation will likely be observed. However, by conducting studies using multilocus *nuclear* data (on the order of hundreds or thousands of loci) across a wide range of taxa, we may be able to more precisely define where to draw a line of distinction based on genetic differentiation. This new definition should take into consideration key factors such as ecological roles and general morphological difference, but fundamentally, genes are driving the processes of speciation and an extensive survey of genetic differentiation will ultimately contribute the most information to our understanding of evolutionary divergence and thus how to define a "species".

Acknowledgements

I would like to thank my advisor, Patrik Nosil, for his help with conceiving and funding this project, for the use of his lab equipment and for his feedback and support along the way. I am also indebted to Aaron Comeault for his guidance and assistance with molecular techniques and help with the statistical programs used in this project. Additional thanks go to Tim Farkas for his assistance with DNA extractions and to Tom Parchman for his Taq polymerase. Thank you to Robert Jadin for conducting the Bayesian phylogenetic analyses and to Rebecca Safran for her permission to put ethidium bromide in her UV imager. And of course, a huge thank you to my Honors Committee, Patrik Nosil, Barbara Demmig-Adams, Andrew Martin and Matthew Wilsey-Cleveland for their time, help and support along the way.

References:

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics, 160*, 1217-1229.

Barton, N. H. (1998). The effect of hitch-hiking on neutral genealogies. *Genetical Research, 72*, 123-133.

Beaumont, M. A. (2005). Adaptation and speciation: what can Fst tell us? *Trends in Ecology and Evolution, 20*(8), 435-440.

Bolnick, D. I., & Fitzpatrick, B. M. (2007). Sympatric Speciation: Models and empirical evidence. *Annu. Rev. Ecol. Evol. Syst., 38*, 459-487.

Brown, W. M., George, M. J., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America, 76*(4), 1967-1971.

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates, Inc.

Crespi, & Sandoval. (2000). Phylogenetic evidence for the evolution of ecological specialization in Timema walking-sticks. *Journal of Evolutionary Biology, 13*(2), 249-262.

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution, 24*(6), 332-340.

Dobzhansky, T. (1940). Speciation as a stage in evolutionary divergence. *The American Naturalist, 74*(753), 312-321.

Dopman, E. B., Perez, L., Bogdnowicz, S. M., & Harrison, R. G. (2005). Consequences of reproductive barriers for genealogical discordance in European corn borer. *PNAS, 102*(41), 14706-14711.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., et al. (2010). Geneious v5.3.4 Available from http://www.geneious.com.

Emelianov, I., Marec, F., & Mallet, J. (2004). Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 271*(1534), 97-105.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources, 10*, 564-567.

Faure, B., Jollivet, D., Tanguy, A., Bonhomme, F. o., & Bierne, N. (2009). Speciation in the Deep Sea: Multi-Locus Analysis of Divergence and Gene Flow between Two Hybridizing Species of Hydrothermal Vent Mussels. *PLoS ONE, 4*(8), e6485.

Ferris, S. D., Sage, R. D., Huang, C.-M., Nielsen, J. T., Ritte, U., & Wilson, A. C. (1983). Flow of mitochondrial DNA across a species boundary. *Proceedings of the National Academy of Sciences of the United States of America, 80*, 2290-2294.

Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondiral DNA. *Annu. Rev. Ecol. Evol. Syst., 34*, 397-423.

Futuyma, D. J., & Mayer, G. C. (1980). Non-allopatric speciation in animals. *Systematic Zoology, 29*(3), 254-271.

Hey, J. (2006a). On the failure of modern species concepts. *Trends in Ecology and Evolution, 21*, 447-450.

Hey, J. (2006b). Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development, 16*, 592-596.

Law, J. H., & Crespi, B. J. (2002). The evolution of geographic parthenogenesis in *Timema* walking-sticks. *Molecular Ecology, 11*, 1471-1489.

Lin, C.-P., & Danforth, B. N. (2004). How do insect nuclear and mitochondrial gene substitution patters differ? Insights from Bayesian analyses of combined datasets. *Molecular Phylogenetics and Evolution, 30*(3), 686-702.

Mallet, J. (1995). A species definition for the modern synthesis. *Trends in Ecology and Evolution, 10*, 294-299.

Mallet, J. (Ed.) (2007) Encyclopedia of Biodiversity. Elsevier Inc.

Mallet, J., Beltran, M., Neukirchen, W., & Linares, M. (2007). Natural hybridization in *heliconiine* butterflies: the species boundary as a continuum. *BMC Evolutionary Biology, 7*(28).

Mayr, E. (1942). *Systematics and the Origin of Species*. New York: Columbia University Press.

Mayr, E. (1947). Ecological factors in speciation. *Evolution, 1*(263-288).

Mayr, E. (1954). Changes of genetic environment and evolution. In J. Huxley, A. C. Hardy & E. B. Ford (Eds.), *Evolution as a process*: Allen & Unwin.

Mayr, E. (1963). *Animal Species and Evolution*. Cambridge, Massachusetts: Harvard University Press.

Neigel, J. E., & Avise, J. C. (1986). *Phylogenetic relationships of motochondrial DNA under various demographic models of speciation*. New York: Academic.

Nielsen, R., & Wakeley, J. (2001). Distinguishing Migration from Isolation: A Markov Chain Monte Carlo Approach. *Genetics, 158*, 885-896.

Niemiller, M. L., Fitzpatrick, B. M., & Miller, B. T. (2008). Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: *Gyrinophilus)* inferred from gene genealogies. *Molecular Ecology, 17*, 2258-2275.

Nosil, P. (2007). Synthesis: Divergent host plant adaptation and reproductive isolation between ecotypes of *Timema cristinae* walking sticks. *The American Naturalist, 169*(2), 151-162.

Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology, 17*, 2103-2106.

Nosil, P., & Crespi, B. J. (2004). Does gene flow constrain trait divergence or vice-versa? A test using ecomophology and sexual isolation in *Timema cristinae* walking-sticks. *Evolution, 58*, 101-112.

Nosil, P., Crespi, B. J., & Sandoval, C. P. (2003). Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270*, 1911-1918.

Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: 'isolation by adaptation' and multiple roles of divergent selection. *Evolution, 62*, 316-336.

Nosil, P., Harmon, L., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution, 24*, 145-156.

Nosil, P., & Rundle, H. (2009). Ecological Speciation: natural selection and the formation of new species *The Princeton Guide to Ecology*. Princeton, NJ: Princeton University Press.

Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology and Evolution, 26*(4), 160-167.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inferences of population structure using mutilocus genotype data. *Genetics, 155*, 945-959.

Rice, W. R., & Hostert, E. E. (1993). Laboratory experiments on speciation: what have we learned in 40 years? *Evolution, 47*(6), 1637-1653.

Ronquist, F., & Huelsenbeck, J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics, 19*(12), 1572-1574.

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters, 8*(3), 336-352.

Sandoval, C., Carmean, D. A., & Crespi, B. J. (1998). Molecular phylogenetics of sexual and parthenogenic *Timema* walking-sticks. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 265*, 589-595.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America, 74*(12), 5463-5467.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favorable gene. *Genetical Research, 23*, 23-35.

Stephens, M., & Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction. *American Journal of Human Genetics, 73*, 1162-1169.

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics, 68*, 978-989.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics, 105*, 437-460.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*, 585-595.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution, (submitted)*.

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology and Evolution, 16*(7), 381-390.

Vickery, V. R. (1993). Revision of *Timema* Scudder (Phasmatoptera: Timematodea) including three new species. *The Canadian Entomologist, 125*, 657-692.

Wang, R. L., Wakeley, J., & Hey, J. (1997). Gene flow and natrual selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics, 147*, 1091-1106.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution, 38*, 1358-1370.

White, M. J. D. (1978). Chain Processes in Chromosomal Speciation. *Systematic Biology, 27*(3), 285-298.

Wright, S. (1931). Evolution of Mendelian populations. *Genetics, 16*, 97-159.

Wright, S. (1965). The interpretation of population structure by F-Statistics with special regard to system of mating. *Evolution, 19*(3), 395-420.

Wu, C.-I. (2001). The genic view of the process of speciation. *The Journal of Evolutionary Biology, 14*(6), 851-865.