2017

EXAMINING THE ROLE OF HOST USE ON DIVERGENCE IN THE REDHEADED PINE SAWFLY, *NEODIPRION LECONTEI*, ACROSS MULTIPLE SPATIAL SCALES

Robin Kimberly Bagley
University of Kentucky, robinkbagley@uky.edu
Author ORCID Identifier:
https://orcid.org/0000-0003-2209-0521
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.285

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Robin Kimberly Bagley, Student

Dr. Catherine R. Linnen, Major Professor

Dr. David F. Westneat, Director of Graduate Studies

EXAMINING THE ROLE OF HOST USE ON DIVERGENCE IN THE REDHEADED PINE SAWFLY, *NEODIPRION LECONTEI,* ACROSS MULTIPLE SPATIAL SCALES

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Arts and Sciences
at the University of Kentucky

By
Robin Kimberly Bagley

Lexington, Kentucky

Director: Dr. Catherine R. Linnen, Assistant Professor of Biology

Lexington, Kentucky

2017

ABSTRACT OF DISSERTATION

EXAMINING THE ROLE OF HOST USE ON DIVERGENCE IN THE REDHEADED PINE SAWFLY, *NEODIPRION LECONTEI*, ACROSS MULTIPLE SPATIAL SCALES

Phytophagous insects make up over one quarter of described species on Earth, and this incredible diversity seems directly linked to feeding on plants. Comparative studies of sister groups have shown shifts to herbivory are consistently associated with increased species diversity in insects, but the reasons for this diversification remain unclear. While other explanations, such as decreased extinction rates or influences on population structure, exist, one prominent hypothesis suggests shifts and subsequent adaptation to novel host plants can lead to the evolution of reproductive barriers.

Given their extreme specialization on host plants in the genus *Pinus* and intimate, life-long association with their host plants, divergent host use has been suspected to drive speciation in the conifer sawfly genus *Neodiprion*. Previous work showed host shifts coincide with speciation events in the genus; but could not determine if these host shifts initiated speciation or if they occurred after other reproductive barriers arose. Determining the contribution and timing of host shifts relative to speciation will require examination of populations at the earliest stages of divergence, before post-speciation changes amass. If host shifts frequently drive speciation in the genus, there will likely be evidence of host-driven divergence within species occurring on a wide range of host plants.

The goal of this dissertation is to examine populations of the red-headed pine sawfly, *Neodiprion lecontei*, an abundant, well-studied pest species that occurs on multiple hosts throughout its range, for evidence of host-driven divergence. Using a combination of reduced representation genomic sequencing, population genomics, and ecological assays, I specifically look for evidence of 1) genetic differentiation between populations utilizing different host plants, 2) ecological divergence in female oviposition preference, larval performance, and ovipositor morphology between populations on different hosts, and 3) ecologically-driven reproductive isolation between genetically and ecologically divergent populations.

Each chapter of this dissertation examines the role of host use in driving ecological, genetic, and/or reproductive divergence within *N. lecontei* at a different spatial scale. First, I surveyed range-wide patterns of diversity. I identified three genetic clusters, dated the divergence of these clusters to the late Pleistocene, and found evidence

that both dispersal limitation (geography) and host use contribute to genetic differentiation within *N. lecontei*. Next, I looked within one of these genetic clusters for additional evidence of the role of host in driving divergence. Sawflies in this cluster primarily utilize two hosts which differ significantly in needle architecture. Although I found no evidence of neutral genetic differentiation between hosts exists, I did detect spatial and temporal differences in host use, and host-specific differences in ovipositor morphology, a performance-related trait. Finally, I examine a single site where *N. lecontei* utilizes three structurally divergent species of pine. Although there was little genetic structure, no sexual isolation, and no distinct host preferences, the host types were partially temporally isolated and varied in ovipositor morphology and larval performance across on the three hosts. Overall, although divergent host use consistently resulted in divergent ovipositor morphology, a reduction in gene flow via temporal or geographic isolation may be required before additional forms of ecological and genetic differentiation can develop. Together these results suggest host shifts alone may not be enough to drive population divergence and speciation in *Neodiprion*.

KEYWORDS: Host-associated divergence, *Neodiprion* sawflies, isolation-by-distance, isolation-by-environment, ecological speciation

Robin K. Bagley
Student signature

July 24, 2017
Date

EXAMINING THE ROLE OF HOST USE ON DIVERGENCE IN THE REDHEADED PINE SAWFLY, *NEODIPRION LECONTEI*, ACROSS MULTIPLE SPATIAL SCALES

By

Robin Kimberly Bagley

Catherine R. Linnen
Director of Dissertation

David F. Westneat
Director of Graduate Studies

July 24, 2017
Date

For my Poppy.

Master crabber, vacation planner, weather watcher, and the best grandfather ever.

Thank you for always watching over me.

“Around here, however, we don’t look backwards for very long.
We keep moving forward, opening up new doors and doing new things,
because we’re curious… and curiosity keeps leading us down new paths.”

-Walt Disney

Acknowledgements

During my time at UK, I have been lucky to have the opportunity to interact with many people who are not only talented scientists, but also wonderful and inspiring people. First, I would like to thank my advisor, Catherine Linnen. It will always be a bit of a mystery to me, but I will always be thankful that you took a chance on a random, sawfly-loving weirdo from a tiny school in New Jersey. From the very first emails we exchanged in 2009 (!), I have always been struck by your kindness, refreshingly down-to-earth nature, and willingness to share your ideas and knowledge. No matter what else was going on, your door was always open to discuss experimental design, the latest results, manuscript revisions, grant proposals, talks, posters, critter care… anything and everything I needed advice on. Thank you for not only celebrating all the good times, but also for commiserating with me in the not so good times. Working with you has made me a better scientist in every way I can think of; and I look forward to collaborating with you in the future as we expand the sawfly empire.

To my committee – thank you for providing your wisdom, practical advice, and encouragement at each step along this journey. Your feedback and support have both been influential not only on my dissertation, but also on my development as a scientist.

There are many steps in my journey that I would have missed were it not for the support of the incredibly kind and patient staff of the Biology Department. I would especially like to thank Bev Taulbee and Jacquie Burke, both of whom have not only gone out of their way to make sure I file all my paperwork and such on time, but also always made time for a friendly chat.

To me, the best part of a lab is the people in it, a notion exemplified by the lab mates, post-docs and lab managers I have interacted with over the last 6 years, including: Kim Vertacnik, Emily Bendall, John Terbot, Claire O'Quin, Danielle Herrig, Matthew Niemiller, Adam Leonberger, Katie Harper, and Joanna Larsson. In addition to being both excellent and hardworking scientists, they are all also amazing people. I will never forget the times I have spent with you, be they the lowest of lows counting cocoons over tequila shots and scrubbing down virus-ridden growth rooms with bleach and quaternary solution; the long days of driving, radio flipping, DEET spraying, and field work during sawfly hunts; the fleeting moments of feeling normal during the Linnaean games; offkey renditions of "Happy Birthday" (a.k.a. "birthday…birthday…birthday…!") over cake/pie and ice cream; or the wonderful moments celebrating milestones achieved, papers published and awards or grants received over champagne. To all of you: Although my sanity these days is questionable at best, my ability to hang on to any of it at all is largely due to your support (as well as the steady supply of delicious baked goods provided by Kim and Danielle).

I also had the opportunity to interact with many undergraduates and was lucky enough to be the primary mentor for several talented students, including: Taylor Shackleford, Katie Deveau, Rachael Ziembroski, Meghan Pasquale, Mary Collins, Jordan Wolfe, Jeremy Frederick, Melanie Hurst, and Carrie Anderson. Their contributions to projects big and small were invaluable, and often made the difference between a project being manageable and a project being impossible. I would especially like to recognize Jordan and Jeremy, for their invaluable help in nearly all critter-based analyses for the Arboretum project; Carrie, who lead the ovipositor analyses for North; and Melanie, who

piloted ovipositor analyses in the lab, braved both Ontario and the Natchez Trace, participated in a number of insane field surveys and test assays, and performed the ovipositor analyses for the Arboretum. To all of you: Thank you all for putting up with my madness! It has been such a joy to see you achieve great things in the lab, and to share in your successes as you have moved on to chase your dreams. You all have big and bright futures ahead of you!

With the creation of the SuperLab in 2016, my "lab family" also grew to include the members of the Weisrock lab, including: Dave Weisrock, Mary Foley, Nikki Lawrence, Justin Kratovil, Paul Hime, Scott Hotaling, Schyler Nunziata, and Kara Jones. It has been wonderful to have the chance to interact with you all on a more regular basis (and also to have more excuses for birthday cake/pie)!

My research would not have been possible if not for many helpful collaborators I have had the opportunity to work with over the past few years. Vitor Sousa was instrumental in figuring out how to manage rangewide demographic analyses within *N. lecontei*. Our work at the Arboretum was made possible by Todd Rounsaville, who kept us in the loop on sawfly activity and provided invaluable historical information about our field site. Collection information from the Canadian Forest Service's Federal Insect Disease Survey (FIDS) database was provided by S. Barudzija. Over the years, sawflies (and/or leads to potential collecting sites) were generously provided to me by many people, including (but probably not limited to): Michael Hillstrom, Jeff Kastle, Scott Lint, Jen Weimer, Kyle Lombard, A. Zeppa, T. Frontz, Laurie Reid, Kaye Murray, Doug Audley, Rob McFerrin, Bev Taulbee, Jill Bagley, Keith Cals, Duane (Spike) Bagley, Maryanna Bagley, Jackson Mosely, Rebecca Lidster, "Cherry Pie Larry", M. Francis,

Lena van Seggelen, and Eddie Brown. Collecting permits were provided by the Florida Park Service, and the Daniel Boone National Forest Cumberland Ranger District.

I would also like to acknowledge my funding sources: a University of Kentucky Multi-Year Fellowship, a Daniel R. Reedy Quality Achievement Fellowship, a G. Flora Ribble Resarch Fellowship and mini-grant, two Graduate School Incentive Program awards, and a USDA-NIFA Predoctoral Fellowship (2015-67011-22803). Additional funding for my research was provided by the National Science Foundation (DEB-1257739; to Catherine Linnen).

I would also like to extend an extra special thanks to my undergraduate research mentor, Sylvio “Chip” Codella. I have many fond memories of sitting in Chip’s office, the scent of espresso filling the air, quick questions having turned into meandering but enlightening 2 hour discussions. It’s also probably important to mention that Chip was the one who introduced me to sawflies, and I honestly can’t imagine what my life would be like without them. To Chip: Thank you, for everything.

Finally, I would like to thank my family. Although fairly small, I am at times overwhelmed by how much love and support my family has given me through my entire life. To my Aunt “Pimmy”, Uncle Richie, and cousin Allie – although I only get to see you a few times each year, your love and unwavering support are always with me. To my Nanny, thank you for being my nightly phone buddy and cheerleader, for listening to me babble about all things sawflies and science, and for always providing me with octopus clips that are actually big enough to hold up all my hair. To my father, Keith, thank you for not only believing in me, but also for looking the other way when I stored sawflies in

our refrigerator at home and for always hanging up my Christmas lights every Thanksgiving.

Although I know many people make the claim, I can say with 100% conviction that my mother, Jill Bagley, actually IS the best mom in the world. For as long as I can remember, she has done absolutely everything in her power to enable me to succeed. To others, I often describe my mom as being “like me, but cooler,” but really, I just hope one day I could be even a tenth as wonderful a person as my mom is. My mom is my biggest fan, my inspiration, and my best friend. Mommy – ever since I was a little girl, you always told me that, someday, I would go to college. Well it turns out you were right, and then some! I could never have made it this far without you. I hope I have, and will continue to make you proud. Thank you for being my mom, and my best friend.

Finally, I would like to thank my beloved Poppy, Duane Bagley, who passed away over the holidays in 2015. After I found out I was going to start my PhD at UK, Poppy was the second person I told, right after my mom. Almost immediately after this, one of his favorite nicknames for me became “Dr. Robbie,” because he had absolute faith that “my buggies” and I could do anything we set our minds to. You were right, Poppy – we made it.

Table of Contents

List of Tables

List of Figures

Chapter 1 : Introduction

1.1 – PHYTOPHAGOUS INSECTS: WHY ARE THERE SO MANY SPECIALISTS?

Plant-feeding insects are an incredibly diverse group comprising over one quarter of the described species on earth (Strong *et al.* 1984). They are also widespread, with species feeding on every extant vascular land plant (Bernays & Chapman 1994; Schoonhoven *et al.* 2005). This startling variety has attracted considerable scientific attention, but there are still many unanswered questions about this diverse group of organisms.

Most phytophagous insects utilize a limited number of hosts. Traditionally, insects are classed into three groups based on the number and taxonomic breadth of their host plants. Monophagous insects are the most restricted, and feed on a single or handful of hosts within a single genus. Oligophagous insects have a broader diet, feeding on a modest number of host plants from two or more genera, but within the same family. Finally, polyphagous insects feed on many plant species from two or more families (Bernays & Chapman 1994). These terms are somewhat problematic, however, as there is actually continuous variation between insects utilizing a single plant, and those feeding on a large number of hosts. In light of this continuum, it is more useful to distinguish the relative level of specialization, with "specialists" utilizing a relatively limited range of hosts, or broadly feeding "generalists" (Schoonhoven *et al.* 2005; Forister *et al.* 2012).

Definitions aside, why are the majority of insect species specialists? Several hypotheses have been introduced and debated in the literature. One of the oldest and most prominent proposes that insects evolved host specificity as a physiological consequence of adaptation to plant secondary chemicals (Levins & MacArthur 1969). While all plants

are relatively equal in terms of nutritional value (Fraenkel 1953), they also possess a vast array of so called secondary chemicals, which play no known role in primary metabolism. The importance of these chemicals in host choice had been recognized since Verschaffelt (1911) demonstrated that *Pieris* caterpillars were stimulated to feed by the presence of mustard oils characteristic of their cruciferous host plants. Fraenkel (1959) proposed that these "odd" chemicals first evolved as a defense against insect herbivores, but have since been disarmed, and in some cases, exploited by adapted insects as host finding cues. In other words, there is a fitness trade-off; as an insect adapts to the defenses and other chemical properties of a particular host, it becomes less adapted to alternative hosts. Specialists then, are expected to be more efficient at using a given host than generalists (Dethier 1954; Fox & Morrow 1981; Futuyma 1983; Cornell & Hawkins 2003).

Several examples of physiological trade-offs due to host adaptation exist (Rausher 1984; Karban 1989; Mackenzie 1996; Cornell & Hawkins 2003). For example, the specialist swallowtail butterfly *Papilio troilus* was found to be 2-3 times more efficient at utilizing its adapted hosts than its generalist relative *P. glaucus* (Scriber 1979). Looking within species, a spider mite population reared on (and presumably adapted to) a polyculture of bean and mite-resistant cucumber host achieved higher fitness on cucumber, but lower fitness on bean as compared to a sister population reared on a bean monoculture (Gould 1979; also see Fry 1990; Agrawal 2000). Reciprocally transplanted pea aphid clones also obtained higher fitness when reared on the host they were collected from than an alternative host (Via 1991).

Many more studies exist, however, which fail to demonstrate such trade-offs, and cast doubt upon their importance in shaping host range (e.g., Via 1984a; b; Futuyma & Philippi 1987; James *et al.* 1988; Fry 1996; Thompson 1996). Futuyma and Wasserman (1981) found no evidence of greater physiological efficiency in utilization of a common host plant by specialized the tent caterpillar *Malacosoma americanum* versus its generalist relative *M. disstria*. While there were differences in larval survival and growth, no trade-offs in overall fitness were detected between groups of sulfur butterfly reared on their primary host *Medicago sativa* than on a sympatric non-host, *Coronilla varia* (Karowe 1990). In addition, a maple-associated parthenogenic genotype of the moth *Alsophila pometaria* was found to have greater efficiency in utilization of oak foliage than an oak-associated parthenogenic genotype (Futuyma *et al.* 1984).

An additional challenge to the physiological trade-offs hypothesis came with the discovery that many secondary chemicals are not truly toxic, but instead serve as "harmless deterrents." While secondary chemicals certainly deter most insects from feeding on a given plant (Bernays & Chapman 1977), it had long been assumed that this avoidance was an adaptive response to toxicity (Berenbaum 1986). In some cases, however, ingestion of deterrent chemicals has been shown to have no measurable impact on fitness (Boyes 1981; Bernays *et al.* 1981; Usher & Feeny 1983; Cottee 1984; Szentesi & Bernays 1984). Many insects will habituate to hosts containing deterrent chemicals, and will feed on them readily after repeated exposure (Szentesi & Bernays 1984; Jermy 1987). On a whole, these studies suggest physiological adaptation to host chemistry may be a consequence, rather than a driver, of specialization (Jermy 1984; Futuyma &

Philippi 1987; Bernays & Graham 1988; but see Joshi & Thompson 1995; Gompert & Messina 2016).

An alternative to the physiological trade-offs hypothesis proposes pressure from natural enemies as a driver of specialization. The potential role of predators in shaping herbivorous insect host breadth was first offered by Brower (1958), who proposed specialization of cryptic larvae to the matching host would be favored by bird predation. The impact of predators and parasitoids, and escaping them in "enemy free space" in shaping herbivore specialization was also highlighted in early discussions of multi-trophic level interactions (Gilbert & Singer 1975; Price *et al.* 1980; Jeffries & Lawton 1984; Janzen 1985).

Much research exists on the potential for parasites to shape host use (Price *et al.* 1980, 1986; Freelander 1983; Barbosa & Saunders 1985). Avoidance of parasitism by nematodes is thought to have influenced the development of amanitin tolerance, and restriction to amanitin-rich mushrooms, in some mycophagous species of *Drosophila* (Jaenike 1985). Gypsy moth eggs laid on pitch pine as a secondary host suffer significantly less parasitism (and viral infection) than larvae using the primary host. Although the early instar larvae cannot feed on pitch pine, mobility between hosts allows host switching between life stages, and may represent the beginnings of a host shift (Rossiter 1987). After a host shift, mortality due to parasitism was significantly higher in populations of the goldenrod ball gallmaker using the ancestral host. Populations on the novel host avoid parasitism by a major parasitoid whose search pattern focuses on the ancestral host (Brown *et al.* 1995).

Examples of the importance of generalist predator avoidance can also be found in the literature. In choice tests, pyralid moth caterpillars consistently prefer older leaves, which are well suited for leaf rolling, over young leaves, which provide better nutrition, but are inadequate for shelter construction (Damman 1987). Several studies comparing vulnerability of generalist and specialist herbivores to predation by generalist predators demonstrate generalists are significantly more susceptible to generalist predators (Eastop 1973; Heads & Lawton 1984; Bernays 1988, 1989; Bernays & Cornelius 1989).

Bernays and Graham (1988) formally introduced a hypothesis of the role of natural enemies in driving host specialization. They argue that the impacts of predation and parasitism on herbivorous insect populations are underappreciated, and so any change in host use that reduces predation and parasitism would be swiftly favored by natural selection. Once the host shift has occurred, the species may specialize as it acquires additional adaptations to avoid attack by generalist natural enemies. For example, a species may develop cryptic coloration or morphology to match the new host, or evolve a mechanism for defensive sequestration of plant secondary chemicals.

A similar, but less prominent theory for the evolution of narrow host ranges is based on avoidance of interspecific competition. In short, all available niches are subdivided between insect species. Each species then adapts to utilize the host efficiently enough to outcompete and exclude all other plant-feeding species, thereby maximizing their own fitness (Bernays & Chapman 1994). It is generally believed that population densities of phytophagous insects are too low to inspire much interspecific competition (Rathcke 1976; Lawton & Strong. 1981; Strong *et al.* 1984), and most studies have found little evidence for a strong role of interspecific competition in host range evolution

(Davidson & Andrewartha 1948; Connell 1983; Schoener 1983; Jermy 1985; Kaplan & Denno 2007). However, there are a handful of cases that demonstrate a role for interspecific competition in host specialization (e.g., McClure & Price 1975; Stiling 1980; Siemens *et al.* 1991). For example, McClure (1980) demonstrated strong competition, including heterospecific exclusion, between two introduced scale insects sharing a niche on eastern hemlock in Canada. Another study found competition between spittlebugs and plume moth larvae on a shared host. Spittlebugs had significantly lower survival on trees occupied by plume moth larvae, presumably due to destruction of their preferred microhabitat by caterpillar feeding activity (Karban 1986). These finding suggest a need for further exploration of the role of competition in host specialization (Denno *et al.* 1995).

The neural limitations hypothesis offers a unique perspective on the prevalence of specialist herbivores. Insect nervous systems are inherently limited in how much information can be processed at any given time (Levins & MacArthur 1969). Yet when choosing a host, insects are bombarded by a variety of largely irrelevant visual, odorant, gustatory and tactile signals. Generalist insects may be unable to adequately process this complex input to make the best choice between several potential hosts (Fox & Lalonde 1993). One strategy for dealing with this enormous amount of information is specialization: focusing and making decisions based on a reduced set of relevant, reliable, and high contrast stimuli. Specialist insects most often respond to a particular class (e.g., Bartlet *et al.* 1993) or particular blend (e.g., Visser 1986) of secondary chemicals. They also tend to be more strongly deterred by non-host odors than generalists (Bernays & Chapman 1987). This increased focus on high contrast signals offers specialists a variety

of benefits, including rapid and accurate decision-making, increased efficiency of tasks, and greater vigilance against enemies (Bernays & Wcislo 1994; Bernays 1998, 2001).

There is support in the literature for neural limitations in generalists, as well as complementary neural efficiency in specialists. On a whole, generalists spend more time making choices, and these choices are often suboptimal. Ovipositing females of the pipevine swallowtail *Battus philenor* specialize on a single host for a given time, and find host plants at a lower rate if they do not (Papaj 1990). Generalists also seem less equipped to distinguish between suitability of individual plants as hosts. A group of butterflies of varying specialization utilizing the nettle *Urtica dioica* were allowed to choose between healthy and senescing nettle plants. Although the healthy plants were the best choice for all, only the butterflies with narrow host ranges accurately selected healthy plants for oviposition (Janz & Nylin 1997). Specialist populations of the aphid *Uroleucon ambrosiae* were also shown to be better foragers, orienting to suitable hosts more quickly and spending less time on unsuitable hosts than generalist populations (Bernays & Funk 1999). Grasshoppers reared as generalists on mixed flavor diets took three times longer to make host choices than specialists raised on single flavor diets. In addition, the generalist-reared grasshoppers seemed easily distracted by additional host information, and would leave a chosen host mid-meal to inspect additional food options (Bernays 1988). Such distraction extends feeding time and may make generalists more vulnerable to predation, as predation risk is significantly higher during feeding activity than resting (Bernays 1997; Dukas 1998).

Many other hypotheses for the prevalence of narrow host range exist, including but not limited to the sexual rendezvous hypothesis, which posits that narrow host range

is favored due to easier mate finding; plant apparency, where reliable, abundant hosts present greater opportunities for specialization; insect size, which proposes host range may be limited by body size; and phenological timing, which suggests insects specialize based on the phenology of their host plants (Jaenike 1990; Bernays & Chapman 1994). The forces driving specialization likely differ from group to group, and much research is still needed on all fronts, so there is little value in arguing which of these hypotheses are "correct" (Futuyma & Moreno 1988; Jaenike 1990; Bernays & Chapman 1994; Mayhew 1997; Janz 2002).

In addition to the general prevalence of specialization discussed above, there is a strong tendency for related species to utilize related host plants (Ehrlich & Raven 1964; Mitter & Farrell 1991; Bernays & Chapman 1994; Winkler & Mitter 2008). Phytophagous insects are confined to nine insect orders (Strong *et al.* 1984), and instances of specialization and generalization occur in each, but are not evenly distributed. The Orthoptera (grasshoppers) are the least host specific group, with 60-85% considered generalist feeders. The remaining insect orders are primarily composed of specialists. Within each order, there are many examples of tribes and genera associated with taxonomically conserved groups of plants. For example, all sawflies genus *Xyela* feed on developing pollen of trees in the genus *Pinus* (Smith 1993). There is also evidence of phylogenetic conservatism in higher taxa, as older families of insects tend to be associated with older groups of plants, and modern insect families are typically restricted to modern plant groups (Bernays & Chapman 1994). These longstanding observations are supported experimentally by reconstruction of host-use on parsimonious insect phylogenies constructed from other characters. An early study of this kind found,

on average, less than 20% of speciation events involved shifts to a different plant family (Mitter & Farrell 1991). More recent studies (e.g., Winkler & Mitter 2008) have detected higher levels of host shifts (48%), but still support considerable taxonomic host conservation.

These conserved associations are thought to reflect long-term, reciprocal coevolution between insects and their host plants. Coevolution of insects and plants was first proposed by Ehrlich and Raven in their classic 1964 treatise, "Butterflies on plants: A study in coevolution." They performed a detailed survey of the existing literature on the feeding habits of all major groups of butterflies, and searched for patterns in these feeding habits. While there were certainly examples of stochasticity in feeding habits, several large groups of butterflies were almost exclusively associated with a group of chemically (although not always phylogenetically) related plants.

Ehrlich and Raven proposed this pattern was the result of an adaptive radiation following the gradual adaptation and specialization of a founder insect species to a new host plant group. If a plant were to obtain a novel chemical defense that allowed it to escape from herbivore pressure, it could enter a new adaptive zone and multiply into many species. Eventually, an insect species may adapt to one of these novel hosts, presumably making some metabolic changes to overcome the defensive secondary chemicals. After successful adaptation to the novel defenses, the founder insect may shift onto one of the similarly defended hosts produced in the adaptive radiation of the plant. Repetitive shifts onto related hosts may then lead to an insect adaptive radiation (Ehrlich & Raven 1964).

Although several studies have found support for such direct coevolution (e.g., Futuyma & McCafferty 1990; Farrell & Mitter 1994), a flaw of Ehrlich and Raven's theory is the assumption that current interactions imply stepwise evolution is somewhat presumptive. As they readily admit, we are only able to view and study current interactions between herbivorous insects and their host plants. We are largely unaware of what interactions took place in the past. The interactions taking place during the adaptive radiation events would presumably be most important in determining plant secondary chemical composition. And while a plant may have a defense that affords it protection against modern plant feeding insects, such protection does not mean that the defense evolved directly against current insects. In that way, although continued herbivory by insects may *maintain* production of a given secondary chemical within a plant, the herbivore activity itself did not drive the fixation of secondary chemical production, and thus the evolution is not truly stepwise, nor reciprocal, and thus is not truly coevolution.

Janzen (1980) offered a more flexible take on coevolution called "diffuse coevolution." To meet the conditions of Ehrlich and Raven's strict coevolution, a change in one population must drive an evolutionary response in a second population, which in turn must drive an evolutionary response in the first population. Diffuse coevolution is similar, but occurs whenever one or both of the populations is composed of a group of organisms exerting a collective selection pressure on other population. For example, although many herbivorous insects are clearly adapted and specialized to a plant's secondary chemicals, those defenses did not necessarily evolve as a direct defense against that insect. The secondary chemical likely evolved as a defense against another herbivore, and just happened to be effective against non-adapted herbivores on a whole. In this case,

herbivores (the currently-adapted insects and any other previous feeders) are exerting selection pressure on a plant to evolve and maintain secondary chemical defenses (see Berenbaum 1983; Mitter *et al.* 1991).

In addition to phylogenetic host conservation, these studies found some groups of insects track structurally or chemically similar, but taxonomically unrelated groups of plants (Mitter & Farrell 1991; Janz & Nylin 1998; Nosil 2002; Winkler & Mitter 2008). Dethier (1941) was among the first to note this pattern, and described a scenario where *Papilio* switched from feeding on plants in the family Rutaceae to those in Umbelliferae via a series of switches between chemically intermediate host genera. Berenbaum (1983) described similar chemically-mediated host shifts in at least three insect orders (Diptera, Coleoptera and Lepidoptera) between unrelated plants containing various classes of coumarins. In *Timema* walking sticks, host shifts seem to be driven more by suitability for cypsis than chemical similarity, with species shifting onto hosts suiting their preadapted color pattern (Crespi *et al.* 2000).

1.2 – Host use, specialization and speciation: What is the role of host shifts?

While much research has focused on explaining the preponderance of specialized insects, considerable work has also gone into exploring why there are so many plant-feeding insects in general. Despite their numerical dominance, and the relative abundance of terrestrial plants as a potential food source, phytophagous insects are confined to only nine insect orders. This inconsistency led Southwood (1973) to suggest that plants present considerable "hurdles to herbivory," which most insect orders have failed to overcome. Once these barriers are overcome, however, opportunities for radiation are

present due to the sheer number of unexploited niches available for colonization (Strong *et al.* 1984).

As detailed above, plant-feeding insects are usually highly specialized, and related insects generally exploit related host plants. This widespread specialization is commonly cited as a source of diversity in herbivorous insects (e.g., Jaenike 1990). Specialization, however, is largely a pruning process, as it actively narrows the number of host plants available for use, and may reduce genetic variation (Kelley *et al.* 2000; Janz & Nylin 2008). Despite this, major host shifts clearly occur with some frequency, as insects utilize every land plant (Bernays & Chapman 1994; Schoonhoven *et al.* 2005). This suggests that host use is dynamic over evolutionary time. So, despite the general trend for insects to specialize over time (Thompson 1994), this specialization is not always a dead end (Kelley & Farrell 1998; Kelley *et al.* 2000), and insects can expand and contract their host ranges over evolutionary time (Janz & Nylin 1998, 2008; Janz 2002; Nosil 2002).

Comparative analysis of the phylogenies of herbivores with their non-herbivorous sister taxa have consistently linked shifts to phytophagy, but not other forms of parasitism (Wiegmann *et al.* 1993), to increased diversification rates (Mitter *et al.* 1988; Farrell 1998). Sister taxa are, by definition, the same age (Hennig 1966), so the differences in diversity detected between herbivorous and non-herbivorous lineages must be due to differential rates of speciation, extinction, or both (Stanley 1979). The exact mechanism behind this diversification is currently debated (Mitter *et al.* 1988; Nosil *et al.* 2002; Janz *et al.* 2006), but a prominent hypothesis suggests shifts and subsequent specialization to novel host plants may directly promote speciation in herbivorous insects (Bush 1969a;

Funk 1998; Drès & Mallet 2002a; Berlocher & Feder 2002; Nosil *et al.* 2002; Matsubayashi *et al.* 2010).

The potential role of host shifts in herbivorous insect speciation was first proposed by Benjamin Walsh. He determined that a "novel" maggot pest of apple was actually the fruit fly *Rhagoletis pomonella* (Walsh 1864). *R. pomonella* normally infests the small, red fruits of wild hawthorn trees. Later, Walsh (1867) proposed that, upon the introduction of apple trees, some flies had shifted host, colonized apple, and perhaps even diverged as a new species from the ancestral hawthorn population. While his ideas were based solely on conjecture, Walsh had introduced a case for sympatric speciation, or speciation in the absence of geological barriers, nearly a century before Guy Bush would bring the theory into controversial prominence.

1.3 – SYMPATRIC SPECIATION: HOST SHIFTS AS A NON-GEOGRAPHIC ISOLATING MECHANISM

Although he briefly considered a role for geographic isolation in species formation, Darwin's believed species could form from natural selection (1859). In fact, when German naturalist Moritz Wagner formally described a concept of speciation based on geographic isolation, Darwin reportedly labeled the paper as "Most Wretched Rubbish" (Schilthuizen 2001). Wagner's work (1868, 1889) would later be recognized by Ernst Mayr, and incorporated into his theory of allopatric speciation. In short, allopatric speciation occurs when a physical barrier separates two populations, which then, through any number of mechanisms, gradually become genetically distinct and, importantly, reproductively isolated, while geographically separate (Mayr 1942, 1963; Coyne & Orr 2004).

After Mayr introduced the theory of allopatric speciation, and vigorously rejected most non-geographic alternatives, it rapidly became accepted as nearly the only mechanism by which speciation could occur. The implausibility of speciation in the absence of geographic isolation was further heightened when John Maynard Smith published a theoretical model of the genetics of sympatric speciation. For his model, Maynard Smith imagined an environment where two independent niches (niche 1 and niche 2) are available. The first step in sympatric speciation is the generation of a stable polymorphism governing habitat preference, such that individuals possessing genotype *AA* are fitter in and prefer niche 1, and individuals possessing genotype *aa* are fitter in and prefer niche 2. If the two populations were to mate randomly no significant divergence would occur, as genes would pass freely between the two populations. If habitat preference were linked to mate choice, however, gene flow between the two populations would be reduced, and divergence could proceed. Thus the second requirement for sympatric speciation to occur is the development of reproductive isolation. This could occur in two ways. In the first case, there is a second gene that governs mate choice, such that individuals of genotype *BB* preferentially mate with BB, and those of genotype *bb* preferentially mate with *bb*. If habitat preference were associated with mate choice, the reproductively isolated populations *AA BB* and *aa bb* can evolve. This dual polymorphism would be difficult to maintain unless the genes were physically linked, as recombination would otherwise break up this association. In the second, “very unlikely” (pp 643) case, the gene controlling habitat preference could be pleiotropic, and confer both preference for a given niche and preference for mates within that niche (Smith 1966).

However unlikely Mayr and his colleagues viewed sympatric speciation, Guy Bush proposed exactly such a case within Walsh's host-shifted apple maggot, *Rhagoletis pomonella*. As discussed above, an apple-infesting "host race" (Bush 1969a; Drès & Mallet 2002a) of *R. pomella* arose from an ancestral hawthorn-infesting population sometime before 1866. There are slight but consistent morphological differences, including body size, number of bristles, and ovipositor length, between populations infesting apple and hawthorn. The races are also allochronically isolated, as each race emerges when its host fruit is sufficiently developed for oviposition (Pickett & Neary 1940; Bush 1966, 1969a).

After observing their behavior in the field, Bush found that *Rhagoletis* flies tended to mate and lay eggs on their preferred host plant (Bush 1966, 1969a; b). This host fidelity (Feder 1998) translated to assortative mating – i.e. *R. pomonella* flies preferring apple tended to mate with other flies preferring apple. In this way, the gene(s) controls host preference in *Rhagoletis* have the exact sort of pleiotropic effect on mate choice so implausible to Maynard Smith; even though apple and hawthorn plants often occur side-by-side, this assortative mating could reproductively isolated the two populations of *R. pomonella*.

Armed with this evidence from *Rhagoletis* and a number of other systems, Bush proposed a general model by which sympatric speciation via host shift could occur. First, the host shift is initiated when mutations impacting host preference (A → a) and performance (B → b) arise in the ancestral host population (AA BB). At first, these alleles are maintained at low levels in the ancestral population, with only a few homozygous individuals (aa bb) shifting to the new host each generation. Since the

individuals on the ancestral host (AA) tend to mate and lay eggs on the ancestral host, and the ones on the novel host (aa) tend to mate and lay eggs on the novel host, gene flow between these populations would be limited to that brought by newly produced (aa) individuals migrating from the ancestral host. However, as the (AA BB) and (aa bb) population continue to adapt to their respective hosts, they may acquire new mutations that increase their fitness on their hosts. If this were so, any (aa) migrants from the ancestral population would be significantly less fit on the novel host. Divergent selection would act against these hybrids, increasing isolation between the two populations. Eventually, this would lead to complete reproductive isolation, i.e. speciation, between the ancestral and novel host races. And best of all, the apple and hawthorn host races *R. pomonella* seemed to fit these criteria perfectly (Bush 1969b, 1975a; b).

Despite his impressive body of evidence, the scientific community was not yet fully convinced of the reality of sympatric speciation (but see White 1978 for contemporary support). Futuyma and Mayer (1980) criticized non-allopatric speciation in general, but particularly questioned sympatric speciation in *Rhagoletis.* Their primary concern was a lack of evidence for genetic differentiation between *R. pomonella* races. Their primary concern was the lack of evidence of genetic differentiation between the apple and hawthorn races. For example, if oviposition choice were not under genetic control, females from apple might return to hawthorn for oviposition and vice versa, permitting gene flow between the populations and ultimately making reproductive isolation nigh impossible. This was especially concerning since an earlier study suggested females from the apple race seemed to make mistakes in host choice fairly often (Reissig & Smith 1978). This scathing commentary was followed by Jaenike (1981), who

questioned if any differences actually existed between the host races. Without additional proof, the possibility that *R. pomonella* was actually a single entity with early-emerging flies ovipositing on apple, and late-emerging flies utilizing hawthorn could not be ruled out.

Undaunted by these harsh criticisms, Bush and colleagues continued to work on *Rhagoletis*, as well as a number of other systems (e.g., coddling moth – Phillips & Barnes 1975; sockeye salmon – Taylor *et al.* 1996, lake cichlids – Barluenga *et al.* 2006) and found a number of encouraging results. Further behavioral analyses revealed that, while both apple and hawthorn flies preferred hawthorn, apple flies were significantly more likely to accept apple than hawthorn flies (Prokopy *et al.* 1988). Later work found that fruit odor is important for host finding in these flies, and that apple and hawthorn flies preferentially orient to the odor from their respective host plant, providing a basis for assortative mating based on host preference (Linn *et al.* 2003). A follow-up study also revealed that F1 hybrids do not respond to fruit odors of apple or hawthorn, which may reduce their ability to find hosts in the field. So despite relatively high levels of gene flow between the races (~6%, Feder *et al.* 1994), divergent natural selection may act against unfit hybrids, keeping the races distinct (Filchak *et al.* 2000; Linn *et al.* 2004).

A genetic basis for differences in emergence time was confirmed, as apple flies emerge earlier than hawthorn flies, even when reared under identical conditions and a common, artificial diet (Smith 1988). Two independent studies found significant differentiation of allele frequencies at six alloenzyme loci across three chromosomes between the apple and hawthorn races of *R. pomonella* (Feder *et al.* 1988; McPheron *et al.* 1988). Further investigation revealed these loci were linked to differences in timing of

adult emergence, and that they resided on chromosomal inversions (Feder *et al.* 2003b), a feature which supports strong linkage between genes by suppressing recombination (Kirkpatrick 2006). These chromosomal inversions can also be found in the ancestral hawthorn race, and artificial selection experiments can produce a similar advance in adult emergence time, suggesting the initial host shift was at least mediated by existing genetic variation in *R. pomonella* (Feder *et al.* 1997a; b; Feder & Filchak 1999).

While the majority of this process almost certainly occurred in sympatry, an allopatric contribution is revealed in the origins of the chromosomal inversion. A genetic analysis of *R. pomonella* populations on hawthorn across reveals the inversion originated in a geographically isolated and genetically distinct population of hawthorn flies in Mexico. This inversion then entered the ancestral hawthorn population in the United States via introgression, providing the genetic variation in diapause time that facilitated the initial host shift to apple (Feder *et al.* 2003a). Similar allopatric contributions have been discovered, or cannot be ruled out, in several other classic cases of sympatric speciation (Coyne & Orr 2004). Today, sympatric speciation is generally agreed to occur, but the frequency with which it occurs is unknown (Via 2001; Berlocher & Feder 2002; Bolnick & Fitzpatrick 2007). The importance of geography in speciation has also been questioned (Butlin *et al.* 2008; Fitzpatrick *et al.* 2008), particularly since the rise of ecological speciation.

1.4 – Ecological speciation: Speciation driven by natural selection

After the theory of allopatric speciation rose to prominence, the potential for natural selection to drive speciation was largely ignored (but see Muller 1942; Dobzhansky 1951; Mayr 1963). The role of selection as a driver of variation was revived

during the heyday of sympatric speciation, but did not come into full prominence until the 1990s, and since then has continued to rise in popularity (see Nosil 2012 for a detailed treatment). A growing body of evidence, supported heavily by evidence from freshwater fish, plants, and herbivorous insects, now favors the role of natural selection as a general driver of speciation (Rice & Hostert 1993; Coyne & Orr 2004; Via 2009).

There are two main ways natural selection can drive speciation. The first is mutation-order speciation, where two populations experience similar selective pressures, but become reproductively isolated after fixing alternative and incompatible alleles during the process of adaptation (Mani & Clarke 1990; Schluter 2009). Evidence for this kind of speciation is limited, but possible examples include male cytoplasmic sterility in *Mimulus* hybrids (Fishman & Willis 2006; Case & Willis 2008) and sexual conflict in *Drosophila* (Rice *et al.* 2005). The second mechanism is ecological speciation, where reproductive isolation evolves as a byproduct of adaptation to different ecological conditions (Schluter 1998, 2001, 2009; Rundle & Nosil 2005; Nosil 2012).

In order for ecological speciation to occur there must be some form of divergent natural selection (Rundle & Nosil 2005; Nosil 2012). This selection typically stems from one of three sources. The first, and perhaps most intuitive source is from differences between environments. For example, if a population of insects were spread across two host plants characterized by different host chemicals, selection may favor the traits affording maximal physiological efficiency on each host (e.g., Via 1984a; b, 1991). In *Timema* walking sticks, divergent selection is driven by visual predation, and acts on host appropriate cryptic morphology (Sandoval 1994; Sandoval & Nosil 2005; Nosil & Crespi 2006b).

Divergent natural selection may also arise from interactions between populations, such as competition between closely related species. Selection against heterospecific (between population) mating may occur if hybridization reduces the fitness of the parents or offspring. Over time, selection would favor individuals that mated within their own population, and strengthen pre-mating isolation between the two populations (i.e., reinforcement; Dobzhansky 1951; Servedio & Noor 2003). A potential example of divergent natural selection based on host-parasite interactions is the parasitoid *Diachasma alloeum*, which split into two partially reproductively isolated forms that specialize on the apple and hawthorn races of *Rhagoletis pomonella* respectively (Forbes *et al.* 2009).

The final source of divergent selection is environmentally dependent sexual selection. This divergent sexual selection can be due to habitat specific selection on secondary sexual characteristics (e.g., Lande 1982) or on mating/communication systems (Ryan & Rand 1993; Boughman 2002). For example, populations of *Anolis cristatellus* lizards are found in mesic and xeric environments that differ in their light intensity and spectral quality. The design of the dewlap, a fan-like structure used for social and sexual display, has diverged between these populations in a way that increases signal detectability in each habitat. In other words, mesic dewlaps are most noticeable in mesic habitats, but are less noticeable in zeric habitat. The reverse is also true (Leal & Fleishman 2004).

Ecological speciation also requires the presence of some form of reproductive isolation. All three sources of divergent selection necessarily lead two forms of reproductive isolation: immigrant inviability and ecologically-based selection against

hybrids (Nosil 2012). Immigrant inviability predicts that individuals migrating from one environment to the other are likely to be maladapted to, and consequently suffer higher mortality in the non-native environment than adapted individuals. It is thought to reduce gene flow between divergent populations by reducing the number of heterospecific encounters (Funk 1998; Nosil 2004; Nosil *et al.* 2005). Hybrid offspring are also predicted to suffer reduced fitness, in this case to both parental environments, as they will display an intermediate phenotype suited for a usually nonexistent intermediate environment (Rice & Hostert 1993; Wang *et al.* 1997; Rundle & Whitlock 2001; Rundle & Nosil 2005).

Since it promotes adaptations increasing the fit of an organism to its environment, divergent selection also routinely leads to habitat and temporal isolation of diverging populations (Nosil 2012). For example, since many herbivorous insects mate on their preferred host plant, the evolution of strong genetic host preference also reduces gene flow by limiting encounters between individuals on different hosts (Rice & Salt 1990). Temporal isolation also reduces gene flow by limiting the overlap of heterospecific interaction between diverging populations. An example of a system where both habitat and temporal isolation have arisen due to divergent selection is *Rhagoletis pomonella*, where apple and hawthorn populations exhibit considerable host fidelity (Feder *et al.* 1994), and also emerge several weeks apart due to differences in host phenology (Filchak *et al.* 2000).

Sexual isolation, where individuals from different populations are less attracted to, or do not recognize members of other populations as mates, commonly occurs in populations under divergent selection, but the ecological basis for its development is

elusive. One possibility is that the trait under selection has pleiotropic effects on mate choice (Nagel & Schluter 1998; Jiggins *et al.* 2001, 2008; Maan & Seehausen 2011; Servedio *et al.* 2011) As discussed previously, sensory drive, or optimization of social and local signals to the local environment (Boughman 2002), and reinforcement (Dobzhansky 1951; Servedio & Noor 2003) may also promote sexual isolation.

A number of post-mating, prezygotic incompatibilities may occur between populations. For example, although a heterospecific pair may mate, gene flow may be limited by suboptimal transfer of sperm (Price *et al.* 2001), failure of fertilization (Vacquier *et al.* 1997), or within-population sperm or pollen preference (Rieseberg *et al.* 1995; Howard *et al.* 1998). Evidence of ecologically driven incompatibilities of this type is lacking, but one convincing case is known from *Timema cristinae* walking sticks. Although the mechanism is currently unknown, female fecundity is greatly reduced by after mating with different-population individuals utilizing a different host plant, but not those utilizing the same host plant (Nosil & Crespi 2006a).

Finally, reproductive isolation may arise through intrinsic hybrid incompatibilities. Although traditionally not associated with divergent selection, recent work suggests intrinsic incompatibility can actually evolve via divergent selection rapidly, even in the presence of significant gene flow (Gavrilets 2004; Agrawal *et al.* 2011). A possible example of intrinsic incompatibility due to divergent selection due to habitat differences is found in the hybridizing sunflower species, *Helianthus annuus* and *H. petiolaris*. Each parental species' cytoplasm was strongly adapted to their respective environments. In hybrids, however, mismatched cytoplasmic and nuclear genomes sometimes led to reduced hybrid fitness (Sambatti *et al.* 2008).

The final component of ecological speciation is a genetic mechanism linking divergent selection and reproductive isolation. There are two ways selection on ecologically-based traits can be linked to reproductive isolation (Rundle & Nosil 2005; Nosil 2012). The first is through pleiotropy, where reproductive isolation is caused by same gene controlling the ecological trait under selection (Rice & Hostert 1993; Kirkpatrick & Barton 1997). Several examples of pleiotropic reproductive isolation exist, including habitat isolation of insects mating on host plants via divergent host preferences (e.g., Rice & Salt 1990), sensory drive (e.g., Boughman 2002), and immigrant and hybrid inviability (e.g., Via & Hawthorne 2002).

Selection and reproductive isolation can also be linked through non-random association, or linkage disequilibrium. In this case, the gene controlling the ecological trait under selection is separate, but physically linked to, the gene causing reproductive isolation (Rice & Hostert 1993; Kirkpatrick & Barton 1997). Linkage equilibrium is generally difficult to maintain (Felsenstein 1981; Servedio 2009; Feder & Nosil 2010), but there are some factors that may promote linkage disequilibrium during ecological speciation. First, in all cases, strong selection is expected to maintain the association between the selected genes and those causing reproductive isolation (Charlesworth *et al.* 1997; Via & West 2008). This will especially be so if the two genes are under very tight physical linkage (Hawthorne & Via 2001), although it may be difficult to distinguish such linkage from pleiotropic effects (but see Wright *et al.* 2013).

Linkage disequilibrium can also be favored if the genes involved in divergent selection and reproductive isolation are located in a structural feature, such as a chromosomal inversion, that reduces recombination rate (Rieseberg 2001; Noor *et al.*

2001; Ortíz-Barrientos *et al.* 2002; Servedio 2009). If the genes controlling selection and reproductive isolation were located within a chromosomal inversion, they would be protected from being broken up by recombination, even if some hybridization were to occur (Butlin 2005; Hoffmann & Rieseberg 2008).

The maintenance of linkage disequilibrium also depends on the genetic basis of the reproductive isolation. If different mate preference alleles fix in the diverging populations (e.g., preference allele for blue mates in one population, and a different allele for red mates in the other) recombination will tend to break up any linkage disequilibrium linking the genes for divergent selection and reproductive isolation within a given population. If mate preference were controlled by a single allele (e.g., a preference allele that makes individuals prefer mates phenotypically similar to themselves), however, no such problem exists (Felsenstein 1981).

1.5 – METHODS FOR DETECTING ECOLOGICAL SPECIATION VIA HOST SHIFTS IN HERBIVOROUS INSECTS

While there is now evidence for ecological speciation across many groups of organisms, including fish (Schluter 1996; Lu & Bernatchez 1999), lizards (Ogden & Thorpe 2002), and plants (Ramsey *et al.* 2003), plant-feeding insects are arguably the best studied and most promising system for the role of natural selection in speciation (Funk *et al.* 2002; Matsubayashi *et al.* 2010). This is particularly true given the long standing (Walsh 1864, 1867) and now well supported role of host shifts in insect speciation. Many questions, however, remain unanswered, particularly regarding the frequency of, and genetic mechanisms behind, host-shift driven speciation. Answering

these questions will require in-depth investigations of additional groups of plant-feeding insects.

There are several approaches that can be used to determine if host-shift speciation has previously, or is now occurring in a system. One popular method is the comparative approach, which derives from Coyne and Orr's (1989, 1997) innovative framework investigating the relationship between strength of reproductive isolation (RI) and genetic distance (GD) as a proxy for time. Funk and colleagues extended this approach by adding a measure of ecological divergence (ED; Funk & Funk 1998; Funk *et al.* 2002, 2006). Qualitative studies comparing the strength of RI between pairs of allopatric populations on same- versus different-hosts can be used as a test of ecological speciation. If adaptation to the host plant were driving divergence in the group, same-host pairs (experiencing similar selection pressures) would be predicted to have lower levels of RI than different-host pairs (experiencing divergent selection pressures). Tests of this method on *Neochlamisus bebbianae* leaf beetles and *Timema* walking-sticks both supported this prediction, with populations on different hosts consistently displaying more RI than same-host pairs (Funk 1998; Nosil *et al.* 2002).

To test the relative contribution of ED to RI, Funk *et al.* (2002, 2006) plotted time-corrected RI as a function of ED for species pairs from eight diverse taxa and found a consistently positive association between the two. This widespread association suggests that ecology is a general driver of speciation (Funk *et al.* 2006). Funk and Nosil (2008) then applied this approach to four plant-feeding insect taxa and found a positive association between divergent host use and RI in all cases. While the number of taxa examined in these analyses is too small to allow broad generalizations, the results are

consistent with the hypothesis that host shifts regularly contribute to reproductive isolation.

Another approach is to combine phylogenies with host-use data to estimate the proportion of speciation events that were accompanied by host shifts. As described previously, Winkler & Mitter (2008) found that host shifts accompanied about half of the speciation events between 145 phytophagous insect sister-species pairs. However, there is no way of knowing if these host shifts occurred before, during, or after speciation (Futuyma & Mayer 1980; Schluter 2000). In a study of ecological speciation in nematine sawflies, Nyman *et al.* (2010) found that 54% of speciation events were associated with host shifts, but after correcting for post-speciation change, the percentage of sister species pairs with non-overlapping host ranges dropped to 22%. Both of these studies only considered species pairs as having arisen via ecological divergence if they did not share any host taxa and/or feeding habits. This approach is flawed, and may underestimate divergence since it does not consider geographical variation in host use, differences in host preference hierarchies, or within-host niche partitioning. Thus, while these "phylogenetic shift" studies (Nosil 2012) provide some insight into the proportion of insect speciation events that coincide with host shifts, there are important limitations to this approach, including: (1) estimates of ecological overlap are necessarily rough due to a lack of detailed ecological information for many species, (2) post-speciation changes in host use obscure the amount of ecological overlap that was present at the time of speciation, and (3) even when post-speciation changes are accounted for, divergent host use between sister species need not imply that host shifts drove speciation. Despite the

uncertainty within these estimates, however, both of these studies support the idea that host shifts play an important role in herbivorous insect speciation.

Population genomic approaches can offer a complementary glimpse into the genetic variation driving divergent selection and ecological speciation. For a given set of genome-wide markers, there will be a mixture of neutral loci and loci undergoing selection. The expected value for measures of divergence, such as F_{ST}, is the same for all neutral loci, and is determined by the interaction of mutation rate, genetic drift, and gene flow. Stochasticity due to drift at these neutral loci, and of the sampling required to estimating F_{ST} produce a consistent distribution of values around that expectation. Loci under divergent selection will have F_{ST} values higher than the neutral expectation, the exact value determined by the strength of selection. Neutral loci linked to loci under selection are also expected to show increased levels of divergence. Therefore, if sufficient markers are available, it will be possible to detect genomic regions of increased genetic divergence because loci in these regions will appear as "outliers" when compared to the neutral expectation (Lewontin & Krakauer 1973; Luikart *et al.* 2003; Butlin 2010).

Once located, F_{ST} outliers can be used in several ways. Rogers and Bernatchez (2005, 2007) compared the overlap of F_{ST} outliers and previously identified quantitative trait loci (QTL), and found that the outliers were associated with QTL "more often than expected by chance alone." The distribution of outliers can also be used to infer the level of genome wide divergence (e.g., Wood *et al.* 2008; Via & West 2008). Pairwise comparisons of F_{ST} outliers may also reveal candidate regions responsible for adaptation (e.g., Nosil *et al.* 2008; Egan *et al.* 2008).

Patterns of divergence at neutral loci can also be informative. During ecological speciation, divergent selection is expected to reduce gene flow between populations (Piálek & Barton 1997; Gavrilets 2004). The reduction in gene flow between adaptively diverged populations can then permit genome-wide divergence via genetic drift. This ecologically-based, general reduction in gene flow can produce a pattern of isolation-by-environment, where more adaptively diverged populations should exhibit higher levels of genetic differentiation than less adaptively diverged populations (Thibert-Plante & Hendry 2010; Nosil 2012; Shafer & Wolf 2013; Bolnick & Otto 2013; Wang & Bradburd 2014). This pattern is analogous to the isolation-by-distance pattern, where limited dispersal reduces gene flow and increases genetic differentiation between populations as the geographical distance separating the populations increases (Wright 1943; Slatkin 1993). As both geographic and "ecological" distance can influence patterns of neutral divergence, it is important to control for the geographic distance separating population pairs when testing for patterns of isolation-by-environment (Flaxman *et al.* 2012; Shafer & Wolf 2013; Wang & Bradburd 2014).

Population genomics can also be used to assess population structure, and the factors influencing that structure. For example, programs like STRUCTURE and ADMIXTURE use Bayesian and likelihood-based clustering algorithms to determine how many distinct populations are represented within a given multi-locus dataset, and will probabilistically assign individuals to those populations (Pritchard *et al.* 2000; Alexander *et al.* 2009). Genetic variation inferred from marker sets can also be used to perform Principal Component Analyses, which determine how many axes of variation exist within

a given dataset. These axes can then be compared to suspected sources of variation (Patterson *et al.* 2006).

Finally, analyses of molecular variation can be used to assess variation and structuring within subgroups pre-sorted by relevant ecological characteristics (Excoffier *et al.* 1992).

1.6 – *NEODIPRION* SAWFLIES AS A MODEL FOR TESTING THE ROLE OF HOST USE IN SPECIATION

Neodiprion is a Holarctic genus of pine-feeding sawflies (Order: Hymenoptera; Family: Diprionidae). Like many plant-feeding insects, *Neodiprion* sawflies have an intimate and life-long association with their host: adults mate on the host plant, eggs are laid within the host tissue, larvae spend their entire feeding period on the natal host and spin cocoons on the host or directly beneath it (Coppel & Benjamin 1965; Knerer & Atwood 1973; Knerer 1993). This tight association is accompanied by extremely specialized feeding habits, with most species a single or small handful of host-plant species in the genus *Pinus*. While host specialization is a feature common to many insect groups, the extensive life history information available for a large fraction of *Neodiprion* species (many of which are forestry pests; Arnett 1993) is truly unique. In addition to being unusually well studied, *Neodiprion* are abundant in nature, can be reared and crossed under laboratory conditions (personal observation), and vary in many ecologically important traits (e.g., host use, larval color, behavior, overwintering mode). Together, these features make *Neodiprion* an excellent model system in which to uncover the mechanisms driving herbivore adaptation and speciation.

Given their extreme specialization on host plants in the genus *Pinus* and intimate, life-long association with their host plants it has long been hypothesized that divergence in host use is a frequent driver of speciation in *Neodiprion* sawflies (Ghent & Wallace 1958a; Knerer & Atwood 1972, 1973, Bush 1975a; b; Strong *et al.* 1984). In support of this idea, previous work has shown host shifts coincide with speciation events in the genus (Linnen & Farrell 2010). These between-species comparisons of *Neodiprion*, however, were insufficient to establish whether host shifts *drove* speciation (host shift speciation), or developed *after* reproductive barriers evolved. Determining the relationship between host shifts and speciation will require examination of populations at the earliest stages of divergence, before post-speciational changes amass (Coyne & Orr 2004). If host shifts frequently drive speciation is prevalent in the genus, there will likely be evidence of host-driven divergence within species occurring on a wide range of host plants. To that end, in this thesis, I examine populations of *Neodiprion lecontei*, an abundant, well-studied pest species that occurs on multiple hosts throughout its range (Middleton 1921; Benjamin 1955), for evidence of ecological divergence, reproductive isolation, and/or genetic differentiation generated by divergent host use.

1.7 – Thesis Objectives and Overview

In this dissertation, I test the hypothesis that host shifts are a general driver of herbivorous insect speciation using *Neodiprion lecontei* as a model. Specifically, I investigate the role of host use in generating genetic and ecological divergence between populations, as well as assess the contributions of divergent host use to reproductive isolation between populations utilizing different hosts. To do this, I examine the impact of host use in generating divergence at multiple spatial scales.

In Chapter 2, I assess the contributions of historical isolation, geography and host use on range-wide patterns of divergence. I use model and non-model based population structuring methods to identify geographically distinct genetic clusters within *N. lecontei*, and infer the topology and date divergence of these clusters to the late Pleistocene via a composite-likelihood approach based on the site-frequency-spectrum. I then propose potential locations for, and expansion paths from, *N. lecontei* Pleistocene refugia based on knowledge of *Pinus* host refugia, demographic parameter estimates, and patterns of diversity. Finally, using Mantel and partial Mantel tests, I assess the relationship between genetic distance and geography and genetic distance and host use range-wide and within each of the clusters.

In Chapter 3, I focus on a genetic cluster where no relationship between host use and genetic distance was found, and evaluate if any evidence of genotypic or phenotypic host-associated divergence exists within the region. First, using an expanded dataset, I use a model-based population structuring method to identify three genetic clusters, each with additional hierarchical structure, within the region. I then re-evaluate the relationship between genetic distance and geography, and genetic distance and ecology in the entire region, and within each of the three major genetic clusters using Mantel and partial Mantel tests and a locus-by-locus Analysis of Molecular Variance. Next, I describe spatial and temporal patterns of host utilization preference across the region using historical collection data, laboratory choice assays, and eclosion data. Finally, after quantifying differences in host needle architecture, I look for evidence of variation in ovipositor morphology, a performance-related morphological trait.

In Chapter 4, I examine a recently colonized field site that harbors *N. lecontei* on three hosts for evidence of host-associated divergence. First, using both non-model and model based methods, I tested for population structure and genome-wide neutral divergence. Next, based on eclosion records and laboratory assays, I assessed evidence of several reproductive barriers between the host types, including temporal isolation, sexual isolation, and habitat isolation. Finally, I tested for evidence differences in performance-related traits between host types, including ovipositor morphology, survival, and female cocoon weight.

Finally, in Chapter 5, I summarize the results of the previous chapters, and discuss the implications of our findings on the general role of host use in driving population divergence and speciation within *Neodiprion*. Based on lessons learned during my dissertation, I also make several recommendations for future studies of host-associated divergence, and discuss the importance of considering the contributions of both ecology and geography.

Chapter 2 : History, geography, and host use shape genome-wide patterns of genetic variation in the redheaded pine sawfly (*Neodiprion lecontei*)

Previously published as: R.K. Bagley, V.C. Sousa, M.L. Niemiller, and C.R. Linnen. 2017. History, geography and host use shape genomewide patterns of genetic differentiation in the redheaded pine sawfly (*Neodiprion lecontei*). *Molecular Ecology*, 26: 1022-1044.

2.1 – INTRODUCTION

Since the first application of protein electrophoresis to the study of variation in *Drosophila pseudoobscura* (Hubby & Lewontin 1966; Lewontin & Hubby 1966), genetic markers have been a standard tool for making inferences about the processes shaping variation within and between populations. Although the types of polymorphisms employed in population genetic studies have changed over the years (Avise 2004; Hartl & Clark 2007; Gnirke *et al.* 2009; Davey *et al.* 2011), a long-standing assumption has been that the patterns they reveal are largely the product of neutral evolutionary processes (but see McVicker *et al.* 2009; Lohmueller *et al.* 2011; Charlesworth 2012; Phung *et al.* 2016). In isolated populations, for example, genetic drift will give rise to genome-wide genetic divergence. Even in the absence of complete isolation, a decline in dispersal rates, and therefore gene flow, at increasing geographical distances can permit divergence via drift. This process, which has been dubbed “isolation-by-dispersal-limitation” (IBDL; Orsini *et al.* 2013), is expected to produce a pattern of isolation-by-distance (IBD), in

which individuals or populations exhibit increasing genetic divergence as the geographic distance between them increases (Wright 1943).

To date, IBD has been documented in a wide range of taxa, leaving little doubt that dispersal limitation is an important driver of population divergence in nature (Slatkin 1993; Peterson & Denno 1998; Jenkins *et al.* 2010; Meirmans 2012). Over the last decade, however, there has been a growing appreciation that ecology can also play a prominent role in shaping genome-wide patterns of genetic differentiation (Nosil *et al.* 2009a; Wang & Summers 2010; Orsini *et al.* 2013; Wang & Bradburd 2014). For example, if divergent or disruptive natural selection favors different traits in different environments and populations harbor genetic variation for these traits, local adaptation is expected to occur (Williams 1966; Servedio 2004; Kawecki & Ebert 2004; Nosil *et al.* 2005; Räsänen & Hendry 2008; Blanquart *et al.* 2013). Local adaptation may in turn reduce gene flow between populations via multiple mechanisms, including selection against immigrants, habitat-based assortative mating, and reduced hybrid fitness (Nosil 2012; Shafer & Wolf 2013; Wang & Bradburd 2014). This environmentally based reduction in gene flow could then promote drift at neutral loci in a manner analogous to geographical isolation. This process has been dubbed isolation-by-adaptation (IBA; Nosil *et al.* 2008; Funk *et al.* 2011).

IBA is expected to give rise to a pattern of isolation-by-environment (IBE), in which populations or individuals from different environments exhibit greater neutral genetic differentiation than those from the similar environments, independent of geographical distance (Wang & Summers 2010; Bradburd *et al.* 2013; Sexton *et al.* 2014; Wang & Bradburd 2014). Under the right circumstances, IBA may ultimately lead to the

formation of new species (Nosil 2012). As such, the IBE pattern is consistent with (but not unique to, see Wang & Bradburd 2014) incipient ecological speciation, in which reproductive isolation evolves as a byproduct of divergent natural selection (Schluter 2009; Thibert-Plante & Hendry 2010; Nosil 2012). Two recent reviews suggest that IBA is pervasive in nature (Shafer & Wolf 2013; Sexton *et al.* 2014), and comparative work suggests that ecological divergence is a common driver of speciation (Funk *et al.* 2006). It remains to be seen, however, if particular selection pressures (e.g., differences in temperature, humidity, photoperiod, resource availability, predator regimes, etc.) predictably generate patterns of IBE, and what conditions are required for IBA to proceed to full reproductive isolation (Nosil *et al.* 2009b; Nosil 2012; Shafer & Wolf 2013).

One scenario under which IBA may be expected to occur is between populations of herbivorous insects utilizing different host plants. Most plant-feeding insects (~90%) are habitat specialists inextricably linked to their hosts throughout their life (Bernays & Chapman 1994). This parasitic lifestyle provides multiple potential sources of IBA. For example, as many insects mate exclusively on their host, changes in host preference will result in assortative mating (Bush 1975a; Prokopy *et al.* 1988; Drès & Mallet 2002a; Berlocher & Feder 2002; Matsubayashi *et al.* 2010). Additionally, the intimate interaction between an insect and its host plant is expected to generate strong selection for host-associated traits, sometimes at the expense of fitness on other hosts (Via 1991; Feder & Filchak 1999; Cornell & Hawkins 2003; Nosil & Crespi 2006b; Singer 2008; but see Jaenike 1990). When these fitness trade-offs exist, immigrant inviability and poor hybrid performance will also reduce gene flow (Via *et al.* 2000; Rundle & Whitlock 2001; Linn *et al.* 2004; Nosil *et al.* 2005; Matsubayashi *et al.* 2011). Given these numerous

mechanisms for reducing gene flow, specialization on different host plants may frequently produce IBE patterns. However, while intraspecific host specialization is extensively documented in insects (Feder *et al.* 1988; Via 1999; Drès & Mallet 2002a; Nosil *et al.* 2002), only a handful of studies have examined host-related IBE or IBA (Nosil *et al.* 2008; Razmjou *et al.* 2010; Funk *et al.* 2011; Roesch Goodman *et al.* 2012). At present, these studies are still too few to draw general conclusions regarding the relationship between divergent host use and neutral genetic divergence.

In addition to IBDL and IBA, genetic variation within species is also shaped by historical events such as isolation (via vicariance or dispersal) and changes in population size (bottlenecks and range expansions). For example, isolation in different refugia during the last glacial maximum (~20,000–18,000 years ago) and post-glacial range expansions have impacted patterns of genetic variation in many organisms (Hewitt 1996, 1999, 2000; Soltis *et al.* 2006; Jaramillo-Correa *et al.* 2009). Except under specific colonization scenarios (e.g., sequential colonization, Orsini *et al.* 2013), these historical events will give rise to patterns distinct from IBD. For example, when formerly isolated populations come into contact, pairwise genetic divergence can be more strongly influenced by historical isolation than by current geographical distance. Given sufficient time and gene flow, this historical signal will erode. However, if the formerly isolated populations are locally adapted, gene flow could remain low enough that the historical signal becomes permanent (De Meester *et al.* 2002; Orsini *et al.* 2013). Thus, to fully understand extant patterns of differentiation, we must simultaneously consider history, geography, and ecology.

In this study, we utilize double-digest restriction-association DNA sequencing (ddRADseq; Peterson *et al.* 2012) to test the hypothesis that divergent host use contributes to genetic differentiation among populations of the redheaded pine sawfly (*Neodiprion lecontei)*, a widespread pest species that utilizes multiple pine (*Pinus*) species throughout its range in eastern North America (Benjamin 1955; Wilson *et al.* 1992). This hypothesis predicts that population pairs from different hosts will be more genetically differentiated than pairs from the same host, after controlling for the impact of geography. Given that this species has a large geographical range and specializes on host plants that experienced isolation and range changes during the Pleistocene (Webb 1988; Wells *et al.* 1991; MacDonald *et al.* 2000; Walter & Epperson 2001; Godbout *et al.* 2005; Schmidtling 2007; Eckert *et al.* 2010), we first assess overall genetic structure using both model-based (Alexander *et al.* 2009) and model-free (Jombart *et al.* 2010) clustering methods. Next, to gain insight into the historical events that gave rise to observed genetic clusters, we use a composite-likelihood method based on the site frequency spectrum to test alternative divergence scenarios and to estimate demographic parameters (Excoffier *et al.* 2013). Finally, having identified distinct genetic clusters, we evaluate the relationship between genetic divergence, geography, and host use using Mantel and partial Mantel tests (Mantel 1967; Sokal 1979; Smouse *et al.* 1986). Together, our results indicate that historical isolation, dispersal limitation, and ecological divergence contribute to genetic differentiation in this species and support the hypothesis that host use is a common driver of population divergence in host-specialized insects.

2.2 – Materials and Methods

2.2.1 – Study system

Neodiprion is a Holarctic genus of conifer-feeding sawflies (Order: Hymenoptera; Family: Diprionidae). Like many plant-feeding insects, *Neodiprion* sawflies have an intimate and life-long association with their host: adults mate on the host plant, eggs are laid within the host tissue, larvae spend their entire feeding period on the natal host and spin cocoons on or directly beneath the host (Benjamin 1955; Coppel & Benjamin 1965; Wilson *et al.* 1992; Knerer 1993). This tight association is accompanied by extremely specialized feeding habits, with most species utilizing a single or small handful of host-plant species in the genus *Pinus*. Given these features, it has long been hypothesized that host shifting frequently drives population divergence and speciation in this genus (Ghent & Wallace 1958b; Alexander & Bigelow 1960; Knerer & Atwood 1972, 1973, Bush 1975a; b). Consistent with this hypothesis, recent comparative work demonstrates that host shifts are associated with speciation events (Linnen & Farrell 2010). However, these interspecific data cannot distinguish between a scenario in which changes in host use drove speciation and a scenario in which host shifts occurred immediately after speciation was completed via some other mechanism. Thus, *Neodiprion* provides an excellent system for testing the hypothesis that host-related selection is a general driver of differentiation *within* species of host-specialized insects and for connecting IBA within species to the origin of reproductive barriers between species.

2.2.2 – Sample collection and DNA extraction

We sampled *N. lecontei* throughout its range in eastern North America. In total, we sampled 88 individuals from 77 localities and 13 different host-plant species. To

explore broad-scale demographic patterns within this species, we chose a sampling scheme that maximized the number of localities included and sequenced only a single individual per locality/host combination (Table A1.1). We adopted this sampling strategy in part because simulations suggest that prioritizing demes over individuals within demes can produce more accurate demographic parameter estimates, avoiding the potential confounding effects of population sub-structure (Städler *et al.* 2009; Chikhi *et al.* 2010; Sousa *et al.* 2014). In addition, our sampling scheme maximized the range of geographical and ecological distances sampled for this species and, by including individuals sampled from the same site but on different hosts whenever possible, minimized eco-spatial autocorrelation. Together, these features can improve our ability to disentangle the contributions of geography and ecology to genetic differentiation (Shafer & Wolf 2013; Wang & Bradburd 2014). Finally, although our limited sampling of individuals within populations precluded us from estimating population differentiation at individual loci (due to high variance in allele frequencies at each locus), by sampling many independent loci across the genome, we could nevertheless obtain good genome-wide estimates of population differentiation (Patterson *et al.* 2006; Willing *et al.* 2012; Wang & Bradburd 2014).

Individuals were collected as mid- to late-instar feeding larvae and either frozen at -80ºC or placed in 100% ethanol and stored at -20ºC until use. In addition, nine individuals included in this study were reared to adulthood in the lab and preserved at -80ºC upon emergence (Table A1.1). DNA was extracted from preserved larvae and adults using either a Qiagen DNeasy Tissue Kit (Qiagen Inc., Valencia, CA) or a CTAB/Phenol-Chloroform-Isoamyl alcohol method based on Chen *et al.* (2010). A Quant-iT High-

Sensitivity DNA Assay Kit (Invitrogen – Molecular Probes, Eugene, OR, USA) was used to estimate DNA concentrations. DNA quality was assessed by examining A260/280 ratios using a Take3 Micro-Volume Plate (BioTek Instruments, Inc., Winooski, VT, USA). We also visualized each DNA extraction on a 0.8% agarose gel to ensure no samples were degraded.

2.2.3 – ddRAD library preparation and sequencing

To generate a multi-locus dataset for characterizing genetic variation in *N. lecontei*, we employed ddRAD sequencing (Peterson *et al.* 2012). Based on an estimated genome size of 350 Mb for *N. lecontei* (determined using flow cytometry) and fragment recovery via examination of test restriction digests (performed as described in Peterson *et al.* 2012), we selected the enzyme pair *NlaIII* and *EcoRI* (NEB, Ipswich, MA). Libraries were prepared in sets of up to 48 individuals, grouped by DNA yield, and randomized with respect to sampling location, with each sample assigned one of 48 unique 5-basepair (bp) in-line barcode sequences during adapter ligation (Table A1.1, S2). Each set of 48 samples was then pooled for automated selection of a 376-bp fragment (+/- 38 bp) on a PippinPrep (Sage Science, Beverly, MA), followed by 12 rounds of high-fidelity PCR amplification (Phusion High-Fidelity DNA Polymerase, NEB, Ipswich, MA) using PCR primers that included one of 12 unique Illumina multiplex read indices (Table A1.3). After verifying library quality on a Bioanalyzer 2100 (Agilent, Santa Clara, CA), libraries were sent to Beckman Coulter Genomics (Danvers, MA, USA), where they were sequenced using 100bp paired-end reads on an Illumina HiSeq 2000 and an Illumina HiSeq RapidRun 2500.

2.2.4 – Data processing and SNP genotyping

Raw sequence reads were first demultiplexed using the pipeline described in Peterson *et al.* (2012). We then used TRIMMOMATIC (V0.30; Bolger *et al.* 2014) to remove restriction enzyme recognition sites and to trim the forward reads to a minimum 4-base sliding window quality score of 20. After quality filtering, we aligned our forward reads to a high-coverage genome assembly for *N. lecontei* (Vertacnik *et al.* 2016; coverage: 112x; scaffold N50: 244kb; GenBank assembly accession: GCA_001263575.1) using the very sensitive local alignment mode in BOWTIE2 (v2.2.3; Langmead & Salzberg 2012). We then used SAMTOOLS (v0.1.19; Li *et al.* 2009) to exclude reads that mapped to more than one location in our *N. lecontei* reference genome. Next, we used STACKS (v1.37; Catchen *et al.* 2013) to extract loci from the reference alignments, retaining only those loci with at least 10x depth of coverage per individual (*-m* 10). We chose 10x to enable high-confidence genotype calls (Kenny *et al.* 2011; Peterson *et al.* 2012) and to minimize the inclusion of loci impacted by allele dropout (caused by polymorphism within the restriction site), which are expected to have lower coverage on average than loci that lack null alleles (Arnold *et al.* 2013; Gautier *et al.* 2013; but see Schweyen *et al.* 2014).

After an initial round of SNP calling, we excluded two individuals that showed missing data at >70% of SNP loci (Table A1.4). In addition, because *Neodiprion* sawflies employ a haplodiploid sex determination system and as several of the analysis methods we used assume diploid data, we excluded putative haploid males. To infer ploidy, we relied on heterozygosity estimates because (1) the majority of our samples were preserved larvae, which could not be sexed using morphology, and (2) adult sex is not always a reliable indicator of ploidy due to occasional diploid male production (Smith &

Wallace 1971; Harper *et al.* 2016). We estimated the percentage of heterozygous sites for each individual using the *--het* option in VCFTOOLS (v0.1.12b; Danecek *et al.* 2011) and excluded six putatively haploid individuals that had markedly low heterozygosity compared to other individuals sampled within the same geographical region (Table A1.4). The paucity of haploid males in our sample is not surprising given our tendency to use the largest available individuals for extraction (which tend to be diploids) and the fact that *N. lecontei* colonies tend to have heavily female-biased sex ratios (Wilson *et al.* 1992; Craig & Mopper 1993; Harper *et al.* 2016; personal observation).

In total, we excluded eight individuals due to missing data or suspected haploidy, producing a final dataset of 80 *N. lecontei* individuals. To the remaining individuals, we applied additional filters to further reduce the impact of allele dropout, which has the potential to bias population genetic parameters (Arnold *et al.* 2013; Gautier *et al.* 2013). First, we excluded all sites with more than 10% missing data for population structure analyses, or more than 50% missing data for demographic analyses. Second, because sites violating Hardy-Weinberg equilibrium often represent genotyping error (e.g., via undetected paralogs or alignment errors which can lead to an excess of heterozygotes; Hosking *et al.* 2004), we performed exact tests of Hardy-Weinberg equilibrium (Wigginton *et al.* 2005) and excluded sites displaying heterozygote excess with p-values significant at the 0.01 level. For the population structure dataset, we also included only one SNP per RAD locus to minimize linkage disequilibrium (LD) between SNPs. Finally, we note that our analyses assume that RAD markers evolve neutrally and reflect genome-wide patterns, making them useful for inferring demographic history and detecting

IBD/IBE (Sousa & Hey 2013). In the discussion, we consider the impact of violating the assumption of neutrality on our conclusions.

Data processing and all other analyses were performed on either the University of Kentucky's Lipscomb High Performance Computing Cluster or through the University of Texas at Austin's Texas Advanced Computing Center Stampede system, accessed through the NSF XSEDE user portal (Towns *et al.* 2014).

2.2.5 – Detection of population structure

To investigate population structure in *N. lecontei*, we used two individual-based approaches. First, we used a maximum-likelihood-based clustering algorithm, implemented in the program ADMIXTURE (v1.23; Alexander *et al.* 2009), to determine the proportion of genetic ancestry of each individual from a specified number of ancestral populations (*K*) without *a priori* population designation. We tested a range of values for *K* from 1 to 10, and performed 100 independent runs for each value of *K*. The most suitable *K* was determined by comparing 5-fold cross-validation (CV) error values across different values of *K* as described in the ADMIXTURE manual. To assess assignment similarity across 100 replicates for the optimal *K*, we used the Greedy algorithm implemented in CLUMPP (v1.1.2; Jakobsson & Rosenberg 2007) to calculate pairwise matrix similarity statistics (G′). We considered pairs of runs with G′>0.90 to have converged to the same solution, and then averaged ancestry proportions across all runs with the same solution. Following this initial set of cluster analyses, we evaluated evidence for hierarchical structure by performing additional ADMIXTURE analyses within each identified cluster, again testing *K* = 1-10, but with 10 independent runs per *K* (Evanno *et al.* 2005).

Second, we explored population structure using discriminant analysis of principal components (DAPC), which is a multivariate approach that transforms individuals' genotypes using principal components analysis (PCA) prior to a discriminant analysis to maximize differentiation between groups while minimizing variation within groups (Jombart *et al.* 2010). DAPC was implemented using the *dapc* function in the ADEGENET package (v1.3-9.2; Jombart 2008) of the R statistical framework (v.3.0.2; R Core Team 2013). As DAPC requires group assignment *a priori*, we employed a *K*-means clustering algorithm implemented in ADEGENET to identify the optimal number of clusters from *K*=1 to *K*=10. Different clustering solutions were then compared using Bayesian Information Criterion (BIC), following Jombart *et al.* 2010. To avoid over-fitting of discriminant functions, we used α-score optimization to evaluate the optimal number of principle components (PCs) to retain in the analysis. To assess similarity of assignment solutions between different numbers of retained PCs and between DAPC and ADMIXTURE, we used CLUMPP to calculate pairwise similarity statistics.

Final population assignments for use in downstream analyses (demographic modeling and IBD/IBE) were determined considering ADMIXTURE results, DAPC results, and sampling location. When there was disagreement between ADMIXTURE and DAPC assignments, individuals were assigned based on their sampling location. After assigning individuals to clusters, we used ARLEQUIN (v3.5.2; Excoffier & Lischer 2010) to summarize genetic diversity for (1) all individuals assigning to each cluster and (2) only those individuals with >90% ADMIXTURE ancestry for a given cluster.

2.2.6 – Demographic modeling

To gain insight into the historical processes that generated the observed population structure, we compared alternative demographic models and inferred demographic parameters from the site frequency spectrum (SFS), using the composite-likelihood method implemented in FASTSIMCOAL2 (v2.5.2.21; Excoffier *et al.* 2013). Although this method discards LD information, which can be especially useful to disentangle different modes of gene flow [e.g., single pulses of admixture from continuous migration (Harris & Nielsen 2013 and references therein)], SFS-based methods are nevertheless useful for inferring divergence times, population tree topologies, and historic migration rates (e.g., Gutenkunst *et al.* 2009; Lukic & Hey 2012; Sousa & Hey 2013; Excoffier *et al.* 2013; Laurent *et al.* 2015).

2.2.6.1 - Model choice

On the basis of our population structure analyses (see Results), we compared sixteen distinct demographic models with three populations, corresponding to samples from North, Central, and South (Figure A1.1). These included four divergence scenarios: (1–3) all possible bifurcating topologies for the three populations and (4) simultaneous divergence of all three populations (Trifurcation). We modeled each scenario under four conditions: without post-divergence gene flow; with post-divergence gene flow; with post-divergence gene flow and allowing for exponential growth in all populations; and with post-divergence gene flow, allowing for exponential growth in Central and South, and considering a bottleneck in North. All input files and scripts utilized in demographic analyses, including template and parameter estimation files for all models, are available

on DRYAD. A summary of all defined parameters and their search ranges are given in Table A1.5.

For the demographic analyses, we used the filtered SNP data (10x coverage, <50% missing data, removal of loci violating HWE for heterozygous excess) to generate the joint population SFS. We considered the minor allele frequency spectrum (folded SFS) because we lack a good outgroup to determine the ancestral state of each allele, and used a mutation rate of 3.5×10^{-9}, based on an estimate from *Drosophila melanogaster* (Keightley *et al.* 2009). Given that FASTSIMCOAL2 assumes all sites are independent, we investigated the LD patterns within and among the RAD loci. Although we do not have access to a linkage map to orient our scaffolds in relation to each other, we used VCFTOOLS' *--geno-r2* option to determine the correlation among the genotypes (r^2) between RAD loci located on the same scaffold (Danecek *et al.* 2011). Because these analyses suggested RAD loci are largely independent, and that no large linked blocks exist in our data (see Results), we considered sites within RAD loci to be linked, and sites between RAD loci unlinked.

The likelihoods obtained with FASTSIMCOAL2 are an approximation, and become close to the exact value if computed from the joint population SFS and a set of independent (unlinked) SNPs (Excoffier *et al.* 2013). Use of linked markers, however, should not bias parameter estimation, as composite likelihoods converge to the correct parameters that maximize the likelihood (Stephens 2007; Excoffier *et al.* 2013). Therefore, with our patterns of linkage in mind, we took a two-step approach for model choice. First, we estimated the parameters that maximized the likelihood for each model based on the three population SFS (3D-SFS) including linked sites ("all-SNPs" dataset).

To account for local LD patterns, we partitioned the scaffolds into "blocks"; with each block corresponding to a RAD locus. The 3D-SFS was then generated by resampling three individuals from each population per block, keeping only SNPs without missing data across all sampled individuals. The SFS were generated by down-sampling SNPs to ensure no missing data, and hence equal sample sizes across all SNPs, as is usually done to maximize the number of sites in the observed SFS (e.g., Marth *et al.* 2004; Gazave *et al.* 2014; Coffman *et al.* 2015). Second, we used the inferred parameter estimates to re-compute the likelihoods of each model based on a 3D-SFS containing only a single SNP per RAD locus ("single-SNP"), such that we approximate likelihoods with a set of independent (unlinked) SNPs. Because the single-SNP 3D-SFS comprised a set of potentially independent SNPs, the recalculated likelihoods should closely approximate the true likelihood, allowing application of the Akaike information criterion (AIC) for model choice. The single-SNP 3D-SFS was built as the all-SNPs dataset, but by sampling only one SNP at random per RAD locus. These two steps were required due to the limited number of SNPs (<10,000) in the single-SNP dataset (Table A1.6), which would result in a limited power to infer demographic parameters (Excoffier *et al.* 2013).

We further explored support for the most likely population topology in two ways. First, we examined whether differences among models were due to the FASTSIMCOAL2 coalescent approximation by comparing the likelihood distribution for the four asymmetrical migration models, as these were favored over more complex models (see Results). These distributions were obtained by recomputing the likelihood based on 100 expected SFS approximated using 200,000 coalescent simulations under the parameters that maximize the likelihood for each model. These distributions inform us about the

variance due to the FASTSIMCOAL2 approximation, and a considerable overlap among models would indicate no real differences. Second, we considered models with two populations, and estimated the corresponding divergence times and migration rates for each pair of populations (North/Central, North/South, and South/Central). Rather than comparing the likelihoods of the alternative models and the corresponding AIC values, we investigated whether the estimated times of divergence between populations were consistent with the most likely topology according to the AIC values. For these analyses, we used the all-SNPs dataset to generate two-population pairwise SFS (2D-SFS) for each pair of populations. This was done by following the procedure described above for the 3D-SFS, but sampling independently the set of blocks (and SNPs) for each pair of populations, keeping five individuals per population.

2.2.6.2 – Parameter estimation

Once the best model was determined, we performed a final parameter estimation step based on marginal pairwise 2D-SFS. The advantage of using multiple 2D-SFS over a single 3D-SFS for parameter estimation is that the size of the SFS and the number of zero entries are reduced, making it easier to fit the observed SFS (Excoffier *et al.* 2013). To generate the 2D-SFS, we followed the procedure we used to generate the 3D-SFS, using the all-SNPs blocks but down-sampling to seven individuals per population.

While our linked all-SNP dataset should not bias parameter estimates (Stephens 2007; Gutenkunst *et al.* 2009; Excoffier *et al.* 2013), confidence intervals can be too narrow due to pseudo-replication stemming from the use of linked sites (Tang *et al.* 2005; Stephens 2007). We therefore used a block-bootstrap approach to obtain confidence intervals (CI), which accounts for the LD dependency structure in our data (Keinan *et al.*

2007; Bickel *et al.* 2010), and is commonly used to obtain CI in population genomics analyses with linked SNPs (e.g., Keinan *et al.* 2007; Alexander *et al.* 2009). We thus re-sampled with replacement the 1,507 scaffolds of the original dataset (*i.e.*, assuming that each scaffold was a block), such that the total number of resampled sites (including the monomorphic sites) of each bootstrap replicate was approximately the same (± 1%) as in the original dataset.

The likelihood values reflect the fit of our models to the joint-population SFS, but to assess whether the best selected model could reproduce the observed SFS we visually examined the fit of the expected SFSs (mean of 100 SFSs approximated as above) to the observed marginal one dimensional SFS (1D-SFS), marginal pairwise 2D-SFS, and joint 3D-SFS. Additionally, since several migration rate estimates were low (see Results), we performed additional simulations to determine which rates were different from zero. To do so, we compared the likelihood distribution (based on 100 expected SFSs approximated with 200,000 coalescent trees) under the "full" model, where all migration rates could be larger than zero, with the corresponding distribution for eight "nested" models, setting each of the eight migration rates to zero. The rationale of this comparison is similar to the one underlying likelihood ratio tests; i.e., if setting a given migration rate to zero leads to a marked decrease in the likelihood compared to the "full" model, it suggests that migration rate is important, as setting it to zero decreases the fit.

To examine the impact of our filtering and assignment decisions, we repeated the model choice and parameter estimation as described above using two alternative datasets, including (1) only SNPs with less than 10% missing data, and (2) individuals that assign

with 90% or greater ADMIXTURE assignment to each cluster. Due to computational limitations, we did not perform block-bootstrapping for these alternative datasets.

All of the SFS-based analyses were done by pooling SFS entries with less than 10 SNPs (*-C10* option), running 100 independent FASTSIMCOAL2 runs (selecting the parameters of the run attaining the maximum likelihood), each consisting of 40 ECM optimization cycles and using 200,000 coalescent simulations. All joint SFS were generated using a combination of custom scripts (available on DRYAD) and ARLEQUIN (v3.5.2; Excoffier & Lischer 2010).

2.2.7 – Inference of glacial refugia

Because demographic analyses supported a Pleistocene divergence scenario (see Results), we examined relationships between heterozygosity, latitude, and longitude within each cluster to infer possible locations of glacial refugia and patterns of postglacial range expansions. In particular, populations closest to glacial refugia are expected to have the highest genetic diversity (measured in each individual as the percentage of sites that are heterozygous) and diversity is expected to decline in the direction of postglacial range expansion (Hewitt 1996, 1999). The strength and significance of the relationships between diversity and latitude/longitude in different regions were evaluated using Spearman's rank correlation tests implemented in STATA (v13.1; StataCorp LP, College Station, TX).

2.2.8 – Isolation-by-distance and isolation-by-environment

To evaluate the relationship between geography, ecology (host use), and genetic divergence, we performed a series of Mantel tests and partial Mantel tests (Mantel 1967; Sokal 1979; Smouse *et al.* 1986). To account for historical population structure, we

performed separate tests for each of the genetic clusters identified by the population structuring analyses (Kuchta & Tan 2005; Meirmans 2012). These tests required genetic, geographic, and ecological distance matrices.

To describe genetic differentiation between pairs of individuals, we used Rousset's $\hat{a}$, which is analogous to the $F_{ST}/(1-F_{ST})$ ratio (Rousset 2000). Briefly, for a pair of individuals *i* and *j*, Rousset's distance $\hat{a}$ is given by $a_{ij} = (Q_w\text{-}Q_{ij})/(1 - Q_w)$, where Q_{ij} is the probability of identity by state of gene copies between individuals and Q_w is the probability of identity within individuals (estimated from all pairs of individuals in the sample). We calculated pairwise Rousset's $\hat{a}$ (Rousset 2000) range-wide and for each cluster using SPAGEDI (v1.4b; Hardy & Vekemans 2002). We calculated the genetic distance matrix for each cluster considering only those individuals within the cluster. For the geographic matrix, we estimated the linear geographic distance separating each pair of individuals using the Geographic Distance Matrix Calculator (available at http://biodiversityinformatics.amnh.org/open_source/gdmg). To create the ecological distance matrix, we coded host use as a discrete trait: pairs collected on the same host species were assigned a distance of 0, and those collected on different hosts were assigned a distance of 1.

To test for IBD, we performed partial Mantel tests (Smouse *et al.* 1986) in which we examined the correlation between genetic and geographical distance matrices, while controlling for ecological distance. To test for IBE, we performed partial Mantel tests (Smouse *et al.* 1986) examining the correlation between genetic and ecological distance, while controlling for geographical distance. Finally, to investigate the extent of eco-spatial autocorrelation in our data, which can have a strong impact on the performance of

IBE tests, we performed Mantel tests (Mantel 1967) that evaluated the relationship between the ecological and geographical distance matrices (Shafer & Wolf 2013).

To determine the extent to which uncertainty in population assignment impacted our results, we also repeated region-specific IBD/IBE analyses after dropping from each cluster: (1) all individuals with ADMIXTURE ancestry estimates <70% for a given cluster and (2) all individuals with ADMIXTURE ancestry estimates <90% for a given cluster.

All Mantel and partial Mantel tests were performed in PASSAGE2 (v2.0.11.6, Rosenberg & Anderson 2011), and significance was determined via 10,000 permutations.

2.3 – RESULTS

2.3.1 – Sequencing, RAD clustering, and SNP discovery

In total, we obtained an average of 2.2 ± 1.6 (SD; standard deviation) million reads per individual, 2.0 ± 1.4 million of which were retained after quality filtering. After mapping and removing putative paralogs, we retained an average of 1.2 ± 0.8 million reads per individual for SNP calling. STACKS recovered an average of 12,734 ± 4,738 loci per individual, which contained a total of 44,832 SNPs. After dropping two individuals with extensive missing data and six putative haploids (Table A1.4), and applying the <10% missing data filter, the number of SNPs was reduced to 13,990. After a subsequent Hardy-Weinberg filter, we retained a total of 13,946 putatively neutral SNPs. The numbers of retained SNPs under additional filters are given in Tables S6-S7.

2.3.2 – Population genetic structure

Using ADMIXTURE's cross-validation procedure, we found that K=3 was the optimal number of genetic clusters in 100% of the runs (Figure A1.2). Unlike ADMIXTURE, the DAPC method chose K=4 as the optimal number of genetic clusters to

describe our data, although the BIC scores for *K*=3 and *K*=4 were nearly equal (Figure A1.3). The α-score optimization procedure suggested that assignment requires only a small number of retained PCs (Figure A1.4). In particular, the maximum α-score was obtained for one PC, which describes ~5% of the total variance, and α-scores remain high until ~5–10PCs (25–50% variance explained), after which they drop off (Figure A1.5).

Given the apparent disagreement between ADMIXTURE and DAPC analyses, we explored clustering solutions under both *K*=4 and *K*=3. Under *K*=4, both ADMIXTURE and DAPC produce multiple conflicting assignment solutions [mean G′ across 100 ADMIXTURE runs = 0.77 (range: 0.48-1.00); mean G′ across DAPC runs incorporating 1-10 PCs = 0.88 (range: 0.62-1.00)], many of which lacked a clear biological interpretation (e.g., Figure A1.6A, C, D). By contrast, individual admixture proportions for *K*=3 were stable across all 100 ADMIXTURE runs (mean G′ = 0.99; range 0.97-1.00) and across DAPC analyses with different numbers of PCs (mean G′ = 0.96; range 0.90-1.00). Assignment results were also similar, but not identical, across DAPC analyses and ADMIXTURE analyses [mean G′ = 0.94 (range 0.85-1.00); Figure 2.1B; Table A1.8]. Given the unstable and dissimilar assignment solutions under *K*=4 within and between assignment methods, the greater stability and biological interpretability of *K*=3 assignment solutions, and the near identical BIC scores achieved by *K*=3 and *K*=4 in the DAPC analyses; we considered *K*=3 for all subsequent analyses.

Individual population assignments indicate that the three clusters correspond to geographic regions, which we will refer to as “North,” “Central,” and “South” (Figure 2.1A, B; Table A1.9). Although most individuals assign primarily to a single cluster, admixture is evident in some Central individuals, particularly those in close geographic

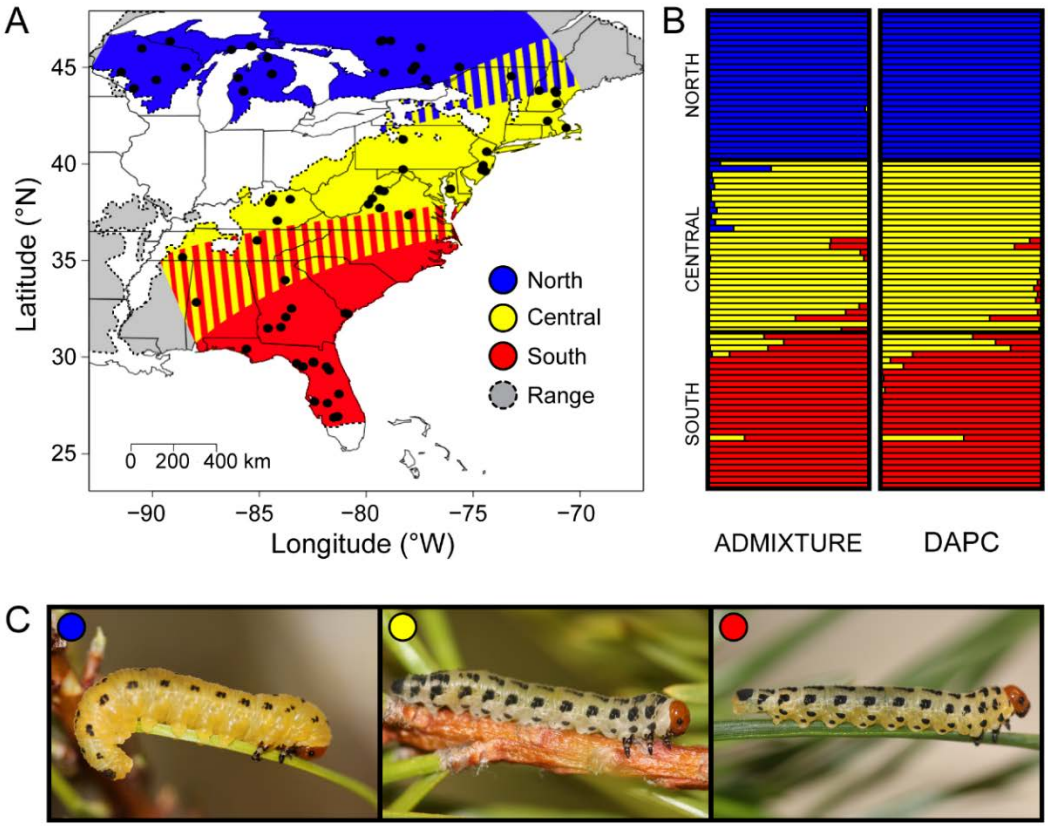


Figure 2.1 – *Neodiprion lecontei* sampling, population structure and morphology. (A) Sampling locations within the estimated range of *N. lecontei* (from Linnen & Farrell 2010). (B) Individual ancestry proportions (ADMIXTURE) and assignment probabilities (DAPC) for $K = 3$, grouped by geographical region. Approximate ranges for each cluster (North, Central and South) and areas of admixture, inferred on the basis of individual assignments in (B), are indicated by shading in (A). Grey shading in (A) indicates uncertainty due to a lack of samples in some portions of *N. lecontei*'s range. (C) Each of the three geographical clusters harbors a unique larval phenotype. Individuals from North tend to have a bright yellow body with reduced spotting. Central harbors a white-bodied, heavily spotted morph. In South, larvae are yellow, heavily spotted, and have a black head capsule, rather than the typical red, early in development (Figure A1.8).

proximity to North and South. Performing additional clustering analyses on individuals assigning to each of the three clusters failed to detect obvious hierarchical structure (i.e., within clusters, *K*=1 always had the lowest CV score; Figure A1.7). Nevertheless, we note that Central is subdivided in some *K*=4 assignment solutions (Figure A1.6). This, combined with the elevated inbreeding coefficient in this cluster (e.g., due to Wahlund effect, Table 2.1), may be indicative of additional population structure in this region. Similarly, although it was never subdivided in our clustering solutions, North also has an elevated inbreeding coefficient that may be indicative of substructure (Table 2.1).

Using the three clusters as populations, F_{ST} estimates from ARLEQUIN suggest moderate to strong differentiation among the three regions, with the highest levels of differentiation observed between North and the other two regions (North/South F_{ST} = 0.47; North/Central F_{ST} = 0.34; South/Central F_{ST} =0.13). North also had reduced genetic diversity compared to the other two regions (Table 2.1). Although Central and South had comparable levels of heterozygosity, Central had considerably more private alleles, suggesting that reduced F_{ST} between South and Central may be due to recent admixture rather than shared ancestry. Genetic summary statistics are similar when considering only those individuals with >90% admixture assignment probability (Table A1.10). These among-region differences in genetic diversity are accompanied by consistent differences in larval morphology (Figures 1C, S8; personal observation).

2.3.3 – Demographic modeling

The analysis of patterns of linkage between SNPs in RAD loci located in the same scaffold showed that LD decays rapidly to values close to zero (r^2<0.05), suggesting that different RAD loci can be considered statistically independent (Figure A1.9).

Table 2.1 – Genetic diversity summary statistics for each population. Population assignments were as in Table A1.9 and ddRAD data were filtered as described in the text [≥10x coverage, ≤10% missing data, and a Hardy Weinberg equilibrium filter excluding sites with heterozygote excess (p-value ≤ 0.01); total number of markers = 5474]. Genome-wide averages of observed heterozygosities (H_o) and inbreeding coefficients (F_{IS}) were calculated using polymorphic loci only. Expected heterozygosities (H_e) for each population were calculated using loci that were polymorphic in any of the populations ("all"); and that were polymorphic within regions ("region")

Population	Polymorphic sites	Private alleles	H_e (all)	H_e (region)	H_o (region)	F_{IS}
North	1112	292	0.049	0.232	0.158	0.237
Central	3994	1865	0.133	0.184	0.134	0.205
South	3102	1174	0.109	0.191	0.162	0.105

Across all levels of complexity, models in which North and Central are sister taxa achieved the greatest likelihood scores. The model allowing asymmetrical migration between groups was favored over all other models, with a relative likelihood of 0.801 (Table 2.2). This topology was also supported by considering models with two populations, as the divergence time of North from Central is the most recent and similar times were estimated for North/South and Central/South divergence (Table A1.11). We also note that there is no overlap in likelihood distributions among models, indicating that likelihood differences across models cannot be explained by the variance in the FASTSIMCOAL2 approximation, further suggesting that the (North, Central), South topology is supported by our data (Figure A1.10).

The maximum-likelihood parameter estimates for our chosen model and the 95% confidence intervals generated by non-parametric block-bootstrapping are given in Table 2.3. ML estimates for population size suggest a large N_e for Central and South and a much smaller N_e in the North (Table 2.3), consistent with the genetic diversity estimates (Table 2.1). All migration rates appeared to be non-zero, excepting those between South and the ancestor of North and Central (Figure A1.11). Estimates suggest moderate gene flow ($2Nm>1.0$) from North into Central and between Central and South, which is consistent with our ADMIXTURE results (Figure 2.1). Assuming a range-wide average of three generations per year for *N. lecontei* , we dated the two divergence times to ~25,000 and ~45,000 years before present (YBP), coinciding with the late Pleistocene and last glacial maximum (Table 2.3, Figure 2.2A). Overall, we found that this model provided a reasonably good fit for the marginal 1D-SFS of all populations (Figure A1.12), and the

Table 2.2 – Summary of the likelihoods for the sixteen demographic models tested. Lhood(ALL SNPs) and Lhood(1 SNP) correspond to the mean likelihood computed with the datasets containing "all SNPs" (including monomorphic sites) and a "single SNP" (without monomorphic sites) per RAD locus, respectively. Mean likelihoods were computed based on 100 expected site frequency spectra simulated according to the parameters that maximized the likelihood of each model. Topology names for each model are as indicated in Figure S1. AIC scores and relative likelihoods (Akaike's weight of evidence) were calculated based on the "single SNP" dataset following Excoffier *et al* 2013.

Topology	Migration allowed?	Exponential growth?	North bottleneck?	$\log_{10}$(Lhood) ALL SNPs	$\log_{10}$(Lhood) 1 SNP	# Parameters	AIC	ΔAIC	Relative likelihood
North-South	No	No	No	-46502.02	-7381.4	7	34006.70	75.69	0.000
North-Central	No	No	No	-46475.82	-7369.0	7	33949.44	18.43	0.000
South-Central	No	No	No	-46502.18	-7381.6	7	34007.60	76.59	0.000
Trifurcation	No	No	No	-46501.54	-7380.4	5	33998.07	67.06	0.000
North-South	Yes	No	No	-46470.49	-7365.0	15	33947.25	16.24	~0.000
North-Central	Yes	No	No	-46462.24	-7361.5	15	33931.01	0.00	0.851
South-Central	Yes	No	No	-46467.69	-7363.8	15	33941.57	10.56	0.004
Trifurcation	Yes	No	No	-46470.28	-7364.7	11	33937.93	6.91	0.027
North-South	Yes	Yes	No	-46469.48	-7362.8	18	33942.91	11.90	0.002
North-Central	Yes	Yes	No	-46461.17	-7361.7	18	33937.82	6.80	0.028
South-Central	Yes	Yes	No	-46463.73	-7363.9	18	33948.15	17.13	~0.000
Trifurcation	Yes	Yes	No	-46467.72	-7363.3	14	33937.39	6.37	0.035
North-South	Yes	Yes	Yes	-46467.45	-7361.5	20	33940.86	9.85	0.006
North-Central	Yes	Yes	Yes	-46461.25	-7362.1	20	33943.82	12.81	0.001
South-Central	Yes	Yes	Yes	-46463.58	-7364.1	20	33953.08	22.07	0.000
Trifurcation	Yes	Yes	Yes	-46466.06	-7362.4	16	33936.93	5.92	0.044

Table 2.3 – Demographic parameters inferred under the asymmetrical migration, North-Central bifurcation model. Maximum-likelihood point estimates for parameters are taken from the run reaching the highest composite likelihood. Migration rates scaled according to population effective sizes (*2Nm*) are given forward in time. The 95% confidence intervals were generated from 100 block-bootstrapped datasets. Estimates of divergence and bottleneck times are given in years, assuming three generations per year (Benjamin 1955; Wilson *et al.* 1992). Estimates of the effective sizes (N_e) are given in number of haploids.

Parameter	ML estimate	95% CI	
		Lower bound	Upper bound
North N_e	35323	30982	43175
Central N_e	315250	279211	362863
South N_e	201161	179538	233121
Ancestral N_e	417251	401164	451283
North+Central Ancestor N_e	1932776	449915	1987435
North/Central divergence time (years)	25675	19987	31668
South/North+Central div. time (years)	45016	37917	62030
2Nm (Central to North)	0.12	0.09	0.16
2Nm (North to Central)	1.10	0.75	1.14
2Nm (South to North)	0.02	0.00	0.05
2Nm (North to South)	0.11	0.04	0.20
2Nm (South to Central)	1.37	1.16	1.85
2Nm (Central to South)	1.01	0.74	1.23
2Nm (Ancestor$_{\text{North+Central}}$ to South)	0.25	0.00	4.69
2Nm (South to Ancestor$_{\text{North+Central}}$)	0.02	0.00	0.12

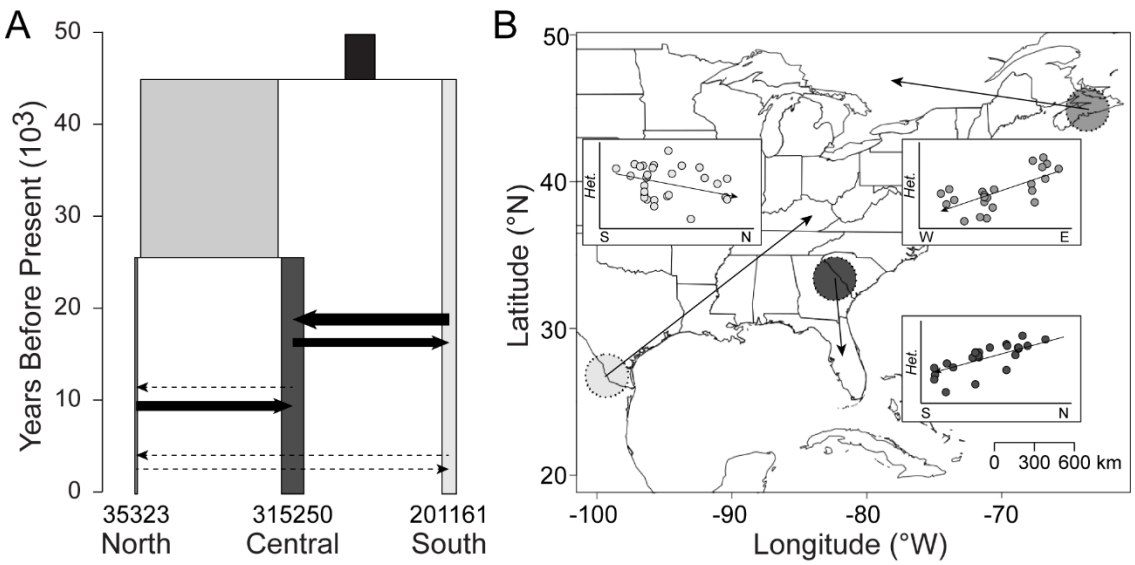


Figure 2.2 – Demographic history for *Neodiprion lecontei*. (A) Favored topology for *N. lecontei*, with divergence times (calculated assuming three generations per year), effective haploid population sizes (indicated by widths of ancestral and extant populations) and average number of immigrants per generation (*2Nm*; indicated by widths of arrows). Dashed lines indicate scaled migration rates <1.0 immigrant per generation (*2Nm* < 1.0). Note that migration rates between the ancestor of North and Central and South were effectively zero (see Table 2.3, Figure A1.11). (B) Proposed glacial refugia (shaded circles) and post-glacial dispersal routes (arrows) for *N. lecontei*, based on patterns of heterozygosity (inset) and known *Pinus* refugia (see text). Shading of proposed refugia locations and heterozygosity plots in (B) match shading of populations in (A).

joint 3D-SFS (Figure A1.13) used for model choice. A similarly good fit was obtained for the pairwise 2D-SFS (Figure A1.14) used for the parameter estimation.

Model choice was largely robust to filtering and assignment decisions, with the (North, Central), South topology always obtaining the highest likelihood score (Tables S12-S13). Additionally, both alternative datasets (>90% assignment probability; <10% missing data) yielded similar parameter estimates to our primary dataset (Table 2.2, S14, S15). We note, however, that the number of SNPs available for the dataset with <10% missing data was almost half that of the other datasets (Table A1.7), and hence we expect a higher degree of uncertainty for those analyses.

2.3.4 – Inference of glacial refugia

Regression analyses revealed a significant reduction in genetic diversity with decreasing longitude (East-to-West) in North ($\rho = 0.48$; $p = 0.016$) and with decreasing latitude (North-to-South) in South ($\rho = 0.75$; $p < 0.0001$). A marginally non-significant reduction in diversity with increasing latitude (South-to-North) was also detected in Central ($\rho = -0.34$; $p = 0.067$; Figures 2B, S15). Based on these patterns, refugia locations and post-glacial colonization routes are proposed in Figure 2.2B (see Discussion).

2.3.5 – Isolation-by-distance and isolation-by-environment

Although we find significant IBD and IBE in the "Whole Range" analysis of *N. lecontei* (Table 2.4), some of this signal is likely an artifact of differentiation among clusters. The impact of cluster membership on genetic differentiation is evident in Figure 2.3: for a given geographical distance, the magnitude of genetic divergence is strongly dependent on where individual pairs were sampled. Nevertheless, Mantel tests also

Table 2.4 – Mantel and partial Mantel test results by geographical region. Pearson's *r*, *P*-value, and Fisher-transformed effect size (Z_r) (Fisher 1921; calculated following Shafer and Wolf [2013]) are given for simple Mantel tests (Matrix 1, Matrix 2) and partial Mantel tests (Matrix 1, Matrix 2 | list of matrices held constant). Population assignments include all individuals assigned to each cluster, as described in Table A1.9.

Comparison	*r*	*P*-value	Z_r
Whole Range			
Geographic, Genetic	0.66	<0.0001	0.80
Geographic, Genetic | Host **(IBD)**	0.61	<0.0001	0.70
Host, Genetic	0.45	<0.0001	0.49
Host, Genetic | Geographic **(IBE)**	0.32	<0.0001	0.34
Geographic, Host **(eco-spatial auto.)**	0.34	<0.0001	0.36
North			
Geographic, Genetic	0.59	<0.0001	0.68
Geographic, Genetic | Host **(IBD)**	0.57	<0.0001	0.65
Host, Genetic	0.20	0.0090	0.20
Host, Genetic | Geographic **(IBE)**	0.10	0.1337	0.10
Geographic, Host **(eco-spatial auto.)**	0.21	0.0122	0.22
Central			
Geographic, Genetic	0.45	<0.0001	0.49
Geographic, Genetic | Host **(IBD)**	0.45	<0.0001	0.49
Host, Genetic	0.16	0.0247	0.16
Host, Genetic | Geographic **(IBE)**	0.14	0.0431	0.14
Geographic, Host **(eco-spatial auto.)**	0.07	0.1339	0.07
South			
Geographic, Genetic	0.79	<0.0001	1.07
Geographic, Genetic | Host **(IBD)**	0.78	<0.0001	1.04
Host, Genetic	0.26	0.0002	0.27
Host, Genetic | Geographic **(IBE)**	0.18	0.0063	0.18
Geographic, Host **(eco-spatial auto.)**	0.19	0.0033	0.20

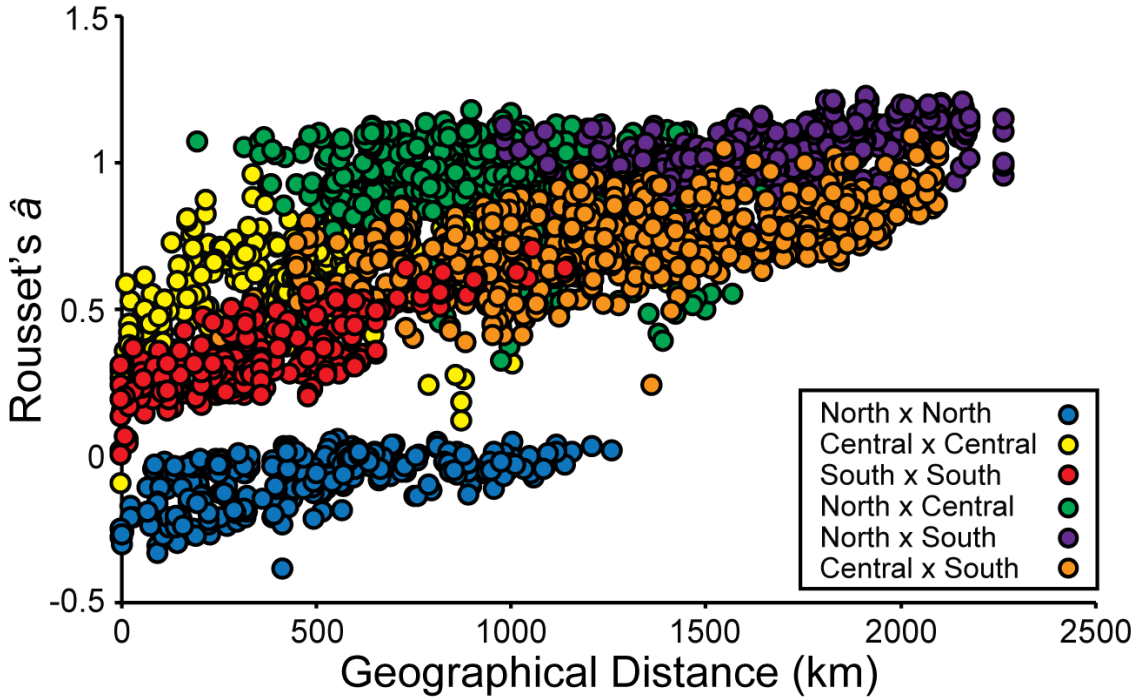


Figure 2.3 – Isolation-by-distance across the range of *Neodiprion lecontei*. Pairwise geographical and genetic distances (measured using Rousset's $\hat{a}$) for all pairwise combinations of 80 individuals. Pairs are colored according to region of origin as indicated in the figure.

revealed significant IBD within each of the three clusters (Table 2.4; Figure 2.4), regardless of the ADMIXTURE cut-off explored (Table A1.16). Of the three clusters, South had the strongest relationship between geographical distance and genetic differentiation, even after accounting for its smaller range overall (Table A1.17). The shape of the IBD relationship also varied among regions (Figure 2.4). In particular, while the Mantel correlation *r* remains high at all examined intervals in South, *r* drops after 450 km in the North and Central clusters (Table A1.17).

Partial Mantel tests indicate significant IBE in the South and Central clusters, but not the North (Table 2.4). This effect is most pronounced in the South, where inspection of the IBD/IBE plot clearly indicates that, for a given geographical distance, genetic divergence is lower for pairs collected on the same host than for those collected on different hosts (Figure 2.4). Considering more stringent assignment cut-offs, significant IBE is still recovered in both Central and South at the 70% ADMIXTURE assignment cut-off; and in South only at the 90% cut-off (Table A1.16).

Examination of the relationship between geography and host use reveals that, within regions, the magnitude of eco-spatial autocorrelation is low enough ($r = 0.07–0.21$; Table 2.4) to permit accurate estimation of IBE correlations (based on simulations described in Shafer & Wolf 2013).

2.4 - DISCUSSION

Like many species, genetic variation in *Neodiprion lecontei* appears to have been shaped by evolutionary processes acting at multiple spatial and temporal scales. Population structure analyses support the existence of three genetic clusters within *N. lecontei*. These clusters correspond to different geographic regions (North, Central, and

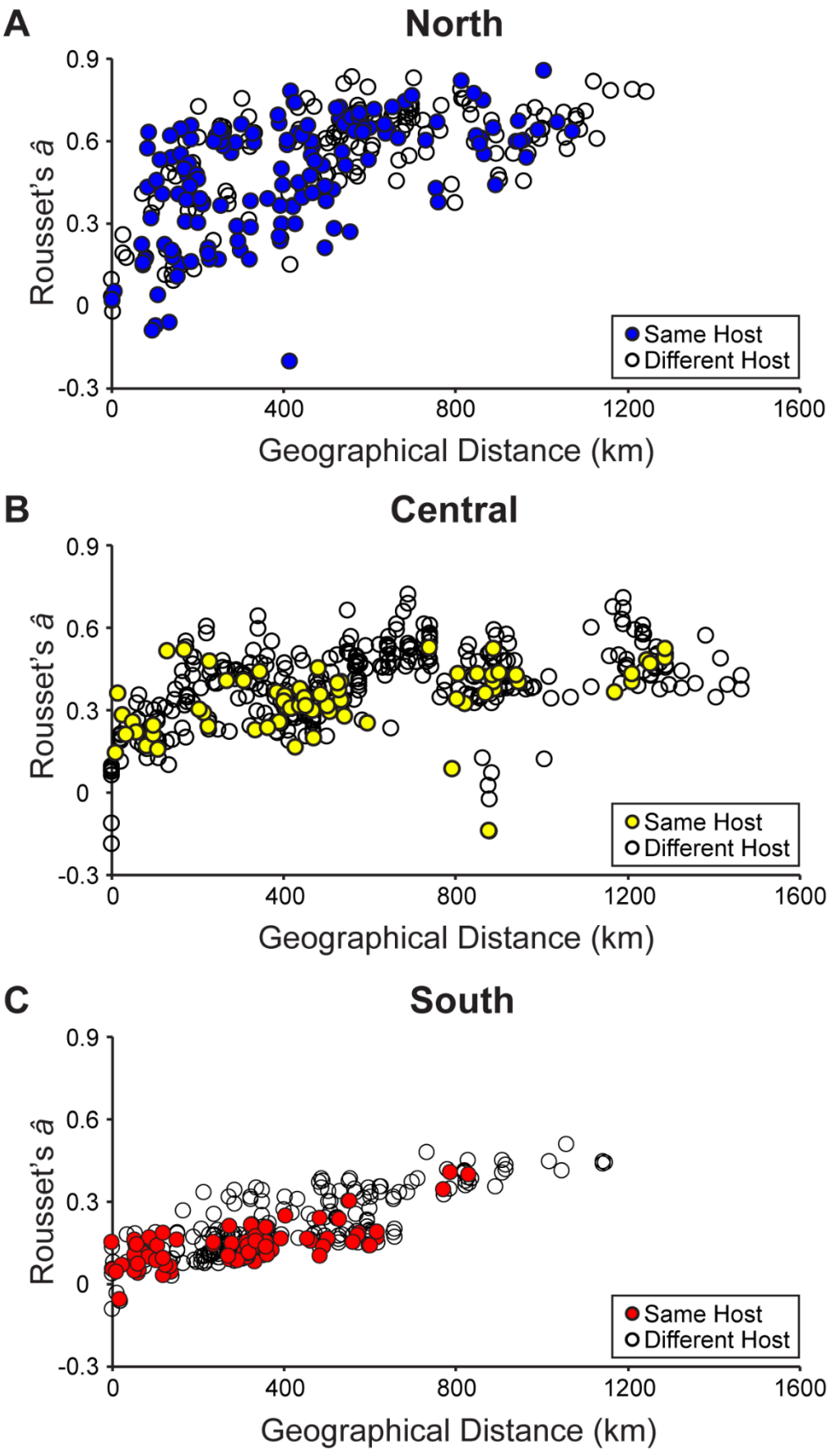

A
North
Rousset's â
0.9
0.6
0.3
0
-0.3
0
400
800
1200
1600
Geographical Distance (km)
Same Host
Different Host
B
Central
Rousset's â
0.9
0.6
0.3
0
-0.3
0
400
800
1200
1600
Geographical Distance (km)
Same Host
Different Host
C
South
Rousset's â
0.9
0.6
0.3
0
-0.3
0
400
800
1200
1600
Geographical Distance (km)
Same Host
Different Host

Figure 2.4 – Isolation-by-distance and isolation-by-environment, by region. Pairwise geographical and genetic distances (measured using Rousset's $\hat{a}$) for individuals sampled from the North (A), Central (B) and South (C) regions (see Figure 2.1 and Table A1.7). Filled circles indicate that a pair of individuals was collected on the same host plant species. To facilitate comparisons among regions, all x-axes and y-axes are displayed on the same scale.

South), each of which harbors a distinct assemblage of host plants. Demographic analyses support Pleistocene divergence followed by post-divergence gene flow. Within each cluster, we detect significant isolation-by-distance, and we also detect significant isolation-by-environment due to host plant within two of the three clusters. Here, we interpret these results in light of *Pinus* phylogeography and *Neodiprion* biology and discuss their implications for population divergence and speciation in plant-feeding insects.

2.4.1 – Influence of historical processes on genetic differentiation in N. lecontei

Demographic analyses suggest that isolation and colonization associated with Pleistocene events have had a lasting impact on genome-wide variation in *N. lecontei*. Specifically, our data suggest that a distinct Southern lineage diverged from the ancestor of North and Central approximately 135,000 generations ago, followed by a split between North and Central 77,000 generations ago. Assuming an average of three generations per year for *N. lecontei* and a mutation rate of 3.5×10^{-9} (Keightley *et al.* 2009), this dates the divergence events to ~45,000 years before present for the split of the South lineage and ~25,000 years before present for the North/Central split. This timing is consistent with glaciation during the middle Wisconsin and the onset of the last glacial maximum, respectively (Richmond & Fullerton 1986). Prior to the onset of the last glacial maximum, the glacial margin was located near the Great Lakes region, and pines likely existed south of their current locations (Delcourt & Delcourt 1981). At peak glaciation, at least four different *Pinus* refugia are thought to have existed. Fossil pollen records indicate that there was a major *Pinus* refugium centered on present day North and South Carolina (Webb 1988; MacDonald *et al.* 2000). In addition, three refugia have been

proposed on the basis of population genetic data from different *Pinus* species: (1) *P. taeda* and *P. palustris* on the western gulf coast, in current day Texas and Mexico (Wells *et al.* 1991; Schmidtling 2007; Eckert *et al.* 2010); (2) *P. banksiana* and/or *P. resinosa* refuge on the exposed northeastern Atlantic coast (Walter & Epperson 2001; Godbout *et al.* 2005); and (3) *P. taeda*, *P. elliotti,* and *P. caribaea* in southern Florida and the Caribbean islands (Wells *et al.* 1991; Schmidtling 2007; Eckert *et al.* 2010). As glaciers retreated, pines expanded from these refugia to their current distributions, with multiple species-specific expansion routes proposed (see references above; Jaramillo-Correa *et al.* 2009).

Based on current knowledge of *Pinus* phylogeography and our demographic modeling results, we propose the following Pleistocene divergence scenario for *N. lecontei* (Figure 2.2; Table 2.3). Prior to the last glacial maximum, a large ancestral population of *N. lecontei* likely existed in a continuous distribution across some portion of the eastern United States. Like other animal and plant taxa (Soltis *et al.* 2006; Jaramillo-Correa *et al.* 2009), the population then split into largely isolated groups during the middle Wisconsin glaciation, with the ancestor of North and Central to the west and the ancestor of South to the east of the Appalachian Mountains. Then, as glaciers advanced across eastern North America during the late Wisconsin, a second split occurred when a portion of the western lineage became isolated in the small Atlantic coast refugium (North) and the remainder (Central) was pushed to the Texas/Mexico refugium. Throughout this time, the eastern lineage (South) likely occupied the much larger *Pinus* refugium near present day North and South Carolina.

As glaciers retreated and pines expanded their range, so too did *N. lecontei*. Examination of the spatial distribution of genetic diversity within each region suggests that North expanded westward towards the Great Lakes region, Central expanded northward across the middle of the United States, and South expanded towards southern Georgia and Florida (Figures 2B, S15). Following range expansion, the three clusters came into contact and began exchanging genes. South and Central are currently experiencing moderate and symmetrical gene flow, and ancestry proportions are indicative of a wide area of admixture (Table 2.3; Figure 2.1). Eventually, the signature of historical isolation between South and Central may be lost altogether.

In contrast, North appears to be resistant to gene flow from the other clusters. Although there is considerable gene flow from North into Central, there is little gene flow into North from either Central or South (Table 2.3). These results are consistent with monopolization, in which colonization of a new area by a small founding population followed by rapid local adaptation yields a resident population that is highly resistant to the successful establishment of migrants (De Meester *et al.* 2002; Orsini *et al.* 2013). This process is expected to produce a persistent pattern of isolation-by-colonization (IBC), in which genetic differentiation reflects historical colonization and founder effects rather than current geographical or ecological distance (Orsini *et al.* 2013; Spurgin *et al.* 2014). Indeed, monopolization may explain why North has maintained markedly lower genetic diversity than the other regions, despite extensive contact with Central (as evidenced by the asymmetric gene flow estimates between North and Central; Table 2.3).

Although we have proposed what we believe to be a realistic historical scenario based on our existing data, we stress that this is a provisional hypothesis that should be

revisited as additional data become available. In particular, it is possible that our results have been impacted by sampling gaps, biases associated with clustering algorithms, and biases associated with ddRAD markers. First, our population sampling has several gaps, most notably at the western and northern extremes of the range (Figure 2.1A). Based on our hypothesized scenario, we predict that samples from these regions, which are closer to our proposed refugia, will harbor greater genetic diversity compared to existing samples. Increased sampling within each of the three regions could also shed light on some unexplained patterns in the data. For example, although genetic diversity is significantly correlated with longitude in North, there are two outliers with markedly higher diversity (Figure A1.15). One possible explanation for these outliers, testable with additional sampling, is that there was a second North refugium and/or colonization route (e.g., as has been proposed for the northern pines *P. banksiana* and *P. resinosa*; Walter & Epperson 2001; Godbout *et al.* 2005). Additional sampling would also give us more power to detect hidden substructure within regions. For example, although our current data indicate that K=1 best explains variation within Central, elevated F_{IS} (Table 2.1) and clustering solutions found under K=4 (Figure A1.6) suggest that there may be undetected structure in this region.

A second important consideration we must take into account is the fact that model-based clustering methods have a tendency to overestimate population structure in the presence of IBD (Frantz *et al.* 2009; Schwartz & McKelvey 2009). When this occurs, sampling gaps can strongly influence cluster assignment (Rosenberg *et al.* 2002; Serre & Pääbo 2004; Schwartz & McKelvey 2009; but see Rosenberg *et al.* 2005). Although there are sampling gaps and significant range-wide IBD in our data (Figure 2.3; Table 2.4),

several lines of evidence support our interpretation that there are three distinct genetic clusters, including: detection of these clusters by both model-based and model-free clustering algorithms, pronounced differences in patterns of genetic variation and morphology among the three clusters, and demographic modeling results that support population divergence during the last glacial maximum and earlier.

In addition, we also have to consider the impact of marker choice. While ddRAD has emerged as a flexible and cost-effective method for generating markers, this method also has several limitations (Davey *et al.* 2013; Puritz *et al.* 2014). Foremost among them is the problem of allele dropout (ADO), which occurs when polymorphism within the restriction site or unequal PCR success masks the presence of a SNP allele (Casbon *et al.* 2011; Arnold *et al.* 2013; Gautier *et al.* 2013; Schweyen *et al.* 2014). Because ADO inflates homozygosity, it can bias estimates of genetic diversity and differentiation (McCormack *et al.* 2013; Arnold *et al.* 2013; Gautier *et al.* 2013). Although we have attempted to minimize the impact of ADO by minimizing PCR cycles and requiring stringent SNP coverage and completeness filters, it is likely that our parameter estimates are impacted to some degree by ADO. Encouragingly, changing the stringency of our completeness filters to allow more ADO (<50% *vs.* <10% missing data) had little impact on our overall conclusions (Table A1.13, S15). Finally, we note that demographic parameters were inferred assuming our SNP markers are neutral, and hence our estimates could be biased if a large proportion of RAD loci were affected by linked selection. Although many details remain to be fleshed out, our existing data strongly suggest that historical events contribute to patterns of genetic differentiation in *N. lecontei*. Less clear are the relative contributions of natural selection and drift to these patterns. On one hand,

the persistent small effective population size in North would have promoted genetic drift. On the other hand, the three geographical regions differ in their assemblages of host plants and in additional factors (e.g., winter duration and intensity) that may generate divergent selection among regions. However, strong correlations between cluster range, geography, and ecology make it almost impossible to disentangle the contribution of these processes to genetic differentiation. Fortunately, we can gain additional insight into the importance of drift and selection by examining how geography and ecology shape genetic differentiation within regions.

Beyond investigating the demographic history of this particular species, we have also introduced a novel approach for demographic analysis with RADseq data. In particular, the generation of RADseq data for non-model organisms is becoming increasingly common (Andrews *et al.* 2016), but the vast majority of researchers either (1) use all SNPs and ignore physical linkage, or (2) use a single SNP per RAD locus. Both approaches can be problematic because (1) with linked SNPs, we cannot appropriately calculate likelihoods or perform non-parametric bootstrap to infer confidence intervals, and (2) sampling a single site per locus considerably reduces the number of SNPs available for demographic inference. Here we suggest a general strategy for dealing with RADseq data, which consists of using all SNPs for parameter estimation, followed by adjustment of likelihood/AIC scores via recomputing likelihoods with a single SNP per RAD locus, and a non-parametric block-bootstrap approach to compute parameter CIs.

2.4.2 – Impact of geography within clusters: IBD

A pattern of isolation-by-distance reflects a balance between divergence due to drift and homogenization via gene flow. When gene flow is too high or low relative to the strength of genetic drift (which is inversely proportional to effective population size), IBD patterns will be absent (Peterson & Denno 1998; Hutchinson & Templeton 1999). These scenarios are unlikely for *N. lecontei* because significant IBD was detected in all three geographical regions. However, the strength and shape of the IBD relationship differed among the three regions (Table 2.4; Figure 2.4). There are several possible explanations for these differences, including differences in the number of generations that have elapsed since an area was colonized, population size, and the presence of barriers to dispersal within clusters (Slatkin 1993; Crispo & Hendry 2005). Although differences in dispersal capabilities and maximal distance sampled are additional explanations that are commonly invoked to explain differences in IBD patterns (Peterson & Denno 1998; Crispo & Hendry 2005; Moyle 2006), we do not consider these further because (1) adult dispersal behavior is similar across all regions (Benjamin 1955; Wilson *et al.* 1992), and (2) the observed differences in IBD persist even after controlling for sampling scale (Table A1.17).Under a scenario in which an ancestral population invades a new area and gives rise to all descendant populations, Slatkin (1993) demonstrated that the IBD pattern arises first between neighboring populations and then expands outwards over time. Thus, the strength of IBD in a non-equilibrium population is expected to correlate positively with the number of generations that have elapsed since colonization (Slatkin 1993; Hutchinson & Templeton 1999). Of the three regions, the South is the only one for which the entire region was unglaciated during the last glacial maximum (~20,000 to 18,000

years ago; Hewitt 1996). The average number of generations per year, which is strongly predicted by the length of the growing season (number of frost-free days), also differs among the regions: 4–5 per year in South, 2–3 per year in Central, and 1–2 per year in North (Benjamin 1955; Wilson *et al.* 1992). Thus, compared to the other regions, the South cluster has most likely been continuously present in the region for many more generations, thus providing more time for IBD to emerge at different spatial scales. Similar differences have been reported in other taxa inhabiting glaciated and unglaciated areas (Hutchinson & Templeton 1999; Rafiński & Babik 2000; but see Crispo & Hendry 2005). Curiously, despite having the northernmost extent and fewest generations per year (and thus likely to have had the least time for IBD to emerge), the strength of the IBD relationship is stronger in North than in Central. One possible explanation for this difference is that undetected population structure (see above) may have impacted the IBD signal in the Central region.

Two additional patterns in the IBD plots require some explanation. First, pairwise Rousset's $\hat{a}$ estimates for North appear much higher in Figure 4 than in Figure 3 (note, however, that the shapes of the IBD relationship are the same). This difference stems from the fact that Rousset's $\hat{a}$ estimates are calculated using (and are positively correlated with) the average level of homozygosity across all sampled individuals. Because average homozygosity is higher in North than in the range-wide data, Rousset's $\hat{a}$ is higher when considering only North individuals (Figure 2.4A). Second, in both the North and Central plots, there are clear outliers in which markedly low genetic divergence is observed between individuals at moderate to large geographical distances. The simplest explanation for these outliers is that they are the result of recent long-range dispersal

facilitated by human planting activity. In particular, just as widespread shipment of seedlings from nurseries for reforestation efforts and ornamental planting has promoted long-range gene flow among *Pinus* populations (Schmidtling 2001), these activities likely promoted *N. lecontei* dispersal as well. In support of this hypothesis, we have first-hand experience with this dispersal mechanism, having received pine seedlings from another state that (unintentionally) bore viable *N. lecontei* eggs (personal observation).

Although finer-scale demographic analyses are needed to investigate structure within regions and to make inferences regarding the prevalence of long-range dispersal, our results demonstrate that dispersal limitation is an important contributor to genetic differentiation in *N. lecontei*. Additionally, our results indicate that populations inhabiting recently glaciated areas may not yet be in regional migration-drift equilibrium (Slatkin 1993; Hutchinson & Templeton 1999; Crispo & Hendry 2005). As such, quantitative estimates of the strength of IBD and IBE in *N. lecontei* from range-wide data will be shaped by a complicated and spatially heterogeneous mix of current and historical processes and should be interpreted with caution (Marko & Hart 2011; Wang & Bradburd 2014). Nevertheless, our data suggest that for distances up to ~450 km, the relationship between geography and genetic distance is monotonic and remarkably similar across the three regions (r = 0.45–0.48; Table A1.17).

2.4.3 – Impact of ecology within clusters: IBE

While IBD has been investigated in a wide range of phytophagous insects (e.g., Peterson & Denno 1998), IBE studies in these organisms are still rare. For example, across a large number of IBE studies compiled in two recent meta-analyses (Shafer & Wolf 2013; Sexton *et al.* 2014), only four studies examined host-associated IBE in

insects. Three of these four studies reported statistically significant IBE (or, if phenotypic divergence was measured, IBA) after controlling for geography (Nosil *et al.* 2008; Razmjou *et al.* 2010; Funk *et al.* 2011), whereas a fourth study did not find evidence of IBE (Roesch Goodman *et al.* 2012). Here, we find similarly mixed results *within* a single species. As was the case of IBD, patterns of IBE are strongest in the South (Table 2.4). By contrast, there is no discernible relationship between host use and divergence in the North. This result does not necessarily indicate, however, that there is a lack of divergent selection on host use in the North (Thibert-Plante & Hendry 2010; Nosil 2012). In fact, reciprocal transplant experiments and host preference assays indicate that there is at least some ecological divergence between North populations associated with *P. banksiana* and those associated with *P. resinosa* (personal observation). Whether or not divergent selection creates a generalized barrier to gene flow that is detectable at neutral loci (thus producing a pattern of IBE) depends on multiple factors, including recombination, migration rate, effective population size, time since colonization, and the strength of divergent selection (Thibert-Plante & Hendry 2010).

Although our results suggest that divergent host use contributes to population differentiation in *N. lecontei*, the methodology we have employed has several limitations. First, Mantel and partial Mantel tests have been criticized due to a lack of power and/or high Type I error rate (Raufaste & Rousset 2001; Harmon & Glor 2010; Legendre & Fortin 2010; Guillot & Rousset 2013; but see Castellano & Balletto 2002; Cushman & Landguth 2010; Diniz-Filho *et al.* 2013). In the context of IBE studies, these tests have demonstrably high false positive rates when there is spatial autocorrelation in the data (i.e., under high levels of IBD and eco-spatial autocorrelation; Guillot & Rousset 2013).

This effect is evident in our own range-wide data: even after controlling for IBD, the range-wide IBE estimate is markedly higher than any of the within-cluster IBE estimates (Table 2.4). Here, we have tried to reduce false positives by employing a sampling design that minimizes eco-spatial autocorrelation within each region. Encouragingly, simulations under comparable levels of IBD and similarly low correlations between environmental and geographic distance (ranging from 0.07–0.21; Table 2.4) yielded reasonable approximations of IBE effect size (Shafer & Wolf 2013). Nevertheless, given the presence of any autocorrelation in the data, results from partial Mantel tests should be interpreted with caution. Fortunately, several recently developed statistical methods offer powerful alternatives to partial Mantel tests for studying IBE (Freedman *et al.* 2010; Bradburd *et al.* 2013; Wang *et al.* 2013; Wang 2013). Although our sampling design and limited ecological data preclude us from using these methods at this time, our Mantel-based results remain useful in that they: (1) provide an initial test of the *a priori* hypothesis that divergent host use contributes to genetic differentiation among populations of *N. lecontei*, (2) provide quantitative estimates of the strength of IBE using a widely used metric (partial Mantel correlation coefficients) that will facilitate comparisons with other taxa (e.g., as in Shafer & Wolf 2013; Sexton *et al.* 2014), (3) include estimates of eco-spatial autocorrelation that will enable better interpretation of IBE results (Shafer & Wolf 2013), and (4) will inform sampling design in future studies seeking to more accurately quantify the relative contributions of IBE and IBD to neutral genetic divergence.

A second limitation in our investigation of IBE is that we have distilled "environment" into a simple dichotomous variable: same or different host plant species.

Generally speaking, continuous scoring is preferable to discrete, and, under some circumstances, discrete scoring appears to inflate effect sizes (Gelman & Hill 2006; Shafer & Wolf 2013). The individuals included in this study were collected on 13 different pine species. Given limited sampling of individuals on each host species, we categorized pairs as having the same or different host as a first step to determining the extent to which host use shapes genetic variation in *N. lecontei*. However, this approach ignores a great deal of variation across *Pinus* species that may impact the strength of divergent selection and, therefore, the signal of IBE. For example, because female sawflies use their saw-like ovipositors to embed their eggs directly into living host plant tissue, one potential source of divergent host-based selection is needle width (Kapler & Benjamin 1960; Knerer & Atwood 1973; Codella & Raffa 2002; Bendall *et al.* 2017). Beyond host use, additional selection pressures, such as the intensity and duration of the winter, could drive neutral genetic divergence among populations. For example, variation in diapause response has been documented among *N. lecontei* populations sampled at different latitudes (Knerer 1983). Thus, accurate quantification of the impact of "ecology" on neutral genetic divergence will require that we quantify host dissimilarity (preferably along as many morphological and chemical axes as possible) and take additional environmental variables into consideration.

Finally, if a large number of our RAD markers are impacted by divergent selection (either directly or via linkage), the pattern of IBE we have detected may be attributable to selection rather than to IBA (which impacts neutral variation). Thus, additional work is needed to assess genome-wide patterns of selection within and between *N. lecontei* populations. If our RAD markers do evolve neutrally as we have

assumed, there are multiple non-mutually exclusive processes that could explain the observed pattern of IBE, including: natural selection against immigrants, selection against hybrids, sexual selection against hybrids, and biased dispersal (Nosil 2012; Wang & Bradburd 2014). Importantly, while some of these processes involve divergent selection, and thus are consistent with local adaptation and incipient ecological speciation, others do not (e.g., divergent sexual selection that is unrelated to the environment or biased dispersal in the absence of divergence in performance). Ultimately, identifying causal mechanisms underlying IBE in *N. lecontei* will require lab- and field-based experiments (Via *et al.* 2000; Nosil & Crespi 2006b; Egan & Funk 2009).

2.4.4 – Implications for speciation

Taken together, our data indicate that geography and history explain the majority of observed genome-wide differentiation within *N. lecontei*. For example, whereas within-region Fisher's transformed effect sizes of IBE range from 0.10-0.18 after controlling for IBD; IBD effect sizes range from 0.49-1.04 after controlling for IBE (Table 2.4). Although these Mantel-based estimates should be interpreted with caution (see above), they nevertheless suggest that IBDL contributes more than IBA to genetic differentiation. This need not imply, however, that isolation and drift are also the main drivers of reproductive isolation. For example, empirical data from other insect systems suggest that considerable reproductive isolation can exist with little to moderate genome-wide divergence (Nosil *et al.* 2009b; Michel *et al.* 2010; Hahn *et al.* 2012; Nadeau *et al.* 2012; Smadja *et al.* 2012; Via *et al.* 2012). Moreover, theoretical models of "genome-wide congealing" suggest that once a population has accumulated sufficient divergently selected variation, populations undifferentiated at neutral loci can rapidly split into

reproductively isolated lineages (Flaxman *et al.* 2013, 2014; Feder *et al.* 2014). To disentangle the contributions of selection and drift to speciation, we must quantify the relationship between reproductive isolation and geography and ecology (Funk *et al.* 2002, 2006). Given the multiple geographical and historical contexts in which host use divergence has occurred, *N. lecontei* provides a particularly powerful system for investigating the contribution of these processes to neutral divergence and reproductive isolation.

Analysis of interspecific data—the products of speciation—provides yet another way to make inferences regarding the contribution of different processes to species formation. Whereas comparative analyses indicate that Neodiprion speciation occurred in multiple geographical contexts, divergence in host use is consistently associated with speciation and with reduced interspecific gene flow (Linnen & Farrell 2010). Although we now know that host use divergence contributes to genetic divergence both within (this study) and between (Linnen & Farrell 2010) species, a causal link between host use divergence and speciation has not yet been established. Thus, a major goal of future work on this system is to determine the sources of divergent selection and the genetic mechanisms linking ecological divergence to reproductive isolation.

2.4.5 – Conclusions

Together, our results support the hypothesis that divergent host use is a general driver of neutral genetic divergence in plant-feeding insects. Coupled with previous comparative work, these data also suggest that, in at least some cases, host divergence leads to the formation of new species. Our results also demonstrate the importance of taking historical processes into account when investigating IBD and IBE. These results

also set the stage for future work on this system that will: (1) more precisely quantify the contributions of IBDL and IBA to neutral differentiation, (2) evaluate the contribution of history, ecology, and geography to the strength of reproductive isolation, and (3) explore causal mechanisms linking divergent host use to population differentiation and speciation. Finally, we note that there are at least five eastern North American *Neodiprion* species (including *N. lecontei*) that have independently evolved similar geographical and host ranges (Linnen & Farrell 2008, 2010). This replication represents a unique opportunity to investigate the repeatability of the historical, geographical, and ecological patterns we have identified in *N. lecontei*, thus providing insight into the predictability of evolution.

2.5 – Data Accessibility

- Short read DNA sequences are available on NCBI SRA (Bioproject PRJNA280451, Biosample numbers SAMN05991526-SAMN05991613)
- The following datasets are available on DRYAD (doi:10.5061/dryad.vh75r):
    - VCF files for all raw SNP datasets
    - Input files for ADMIXTURE (.ped/.map format) and DAPC (.raw format)
    - Input files (2D- and 3D-SFS, TPL, and EST files) for all FASTSIMCOAL2 analyses
    - Distance Matrices for Mantel and partial Mantel tests

2.6 – Author Contributions

RKB and CRL conceived, designed, and sampled specimens for this project. RKB generated the data. RKB, VS, MLN, and CRL analyzed and interpreted the data and wrote the paper.

2.7 – ACKNOWLEDGEMENTS

We thank Brant Peterson, Emily Kay, and Jesse Weber for advice on ddRAD library preparation and demultiplexing; Gideon Bradburd for helpful discussions on IBD and IBE; Laurent Excoffier for his advice regarding demographic analyses; and five anonymous reviewers whose comments helped improve this manuscript. We thank the past and present members of the Linnen lab, especially Joanna Larson, Kim Vertacnik, Emily Bendall, John Terbot II, Danielle Herrig, Rachael Ziembroski, and Mary Collins, for comments on the manuscript, assistance in the field, and maintenance of sawfly colonies. Michael Hillstrom, Jeff Kastle, Scott Lint, Jen Weimer, Kyle Lombard, A. Zeppa, T. Frontz, Laurie Reid, Kaye Murray, Doug Audley, Rob McFerrin, Bev Taulbee, Jill Bagley, Keith Cals, Duane (Spike) Bagley, Maryanna Bagley, and Jackson Mosely provided *N. lecontei* samples. Collecting permits were provided by the Florida Park Service. Funding for this research was provided by the National Science Foundation (DEB-1257739; CRL); a University of Kentucky Multi-Year Fellowship, Daniel R Reedy award, and Ribble mini-grant (RKB); a USDA-NIFA Predoctoral Fellowship (2015-67011-22803; RKB); Swiss NSF grants (31003A-143393 and 310030B-166605; VCS); and national funds through Fundação para a Ciência e a Tecnologia (UID/BIA/00329/2013; VCS). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575.

2.8 – SUPPORTING INFORMATION

The following supporting information can be found in Appendix 1.

- Table A1.1. Sampling locations for all individuals included in this study.

- Table A1.2. 5-bp barcodes with associated P1 adapter sequences.
- Table A1.3. PCR first read indexes.
- Table A1.4. Percent missing data and proportion of heterozygous sites per individual.
- Figure A1.1. Four full-migration divergence scenarios tested in FASTSIMCOAL2.
- Table A1.5. List of demographic parameters and search ranges.
- Table A1.6. Summary of SNP recovery under different data filters and cluster assignment cutoffs (ADMIXTURE and *adegenet*).
- Table A1.7. Summary of SNP recovery under different data filters and cluster assignment cutoffs (FASTSIMCOAL2).
- Figure A1.2. Average CV error scores for each *K* across 100 independent ADMIXTURE runs.
- Figure A1.3. BIC plot for DAPC.
- Figure A1.4. Plots of α-score over 20 PCs with spline interpolation.
- Figure A1.5. Cumulative percent variance explained as a function of the number of retained principal components.
- Figure A1.6. Ancestry proportions (ADMIXTURE) and assignment probabilities (DAPC) for $K = 4$.
- Table A1.8. Pairwise matrix similarity statistics (G′) for K=3.
- Table A1.9. Sampling location, cluster assignment, ancestry proportions, and assignment probabilities for 80 *N. lecontei* individuals.
- Figure A1.7. Average CV error scores for each *K* across 10 independent ADMIXTURE runs for North (A), South (B), and Central (C) clusters.

- Table A1.10. Genetic diversity summary statistics for each population (maximal 0.90 ADMIXTURE ancestry cut-off).
- Figure A1.8. Comparison of mid-instar head capsule coloration.
- Figure A1.9. Correlation among genotypes (r^2) between (A) and within (B) RAD loci on the same scaffold.
- Table A1.11. Demographic parameters inferred under 2-population bifurcation models.
- Figure A1.10. Comparison of Log-likelihood for asymmetric migration demographic models.
- Figure A1.11. Effect of migration rates on the likelihood of chosen model.
- Figure A1.12. Fit of the expected marginal 1D-SFS.
- Figure A1.13. Fit of the expected joint 3D-SFS.
- Figure A1.14. Fit of the expected pairwise 2D-SFS.
- Table A1.12. Summary of likelihood scores for asymmetric migration demographic models (maximal 0.90 ADMIXTURE ancestry cut-off).
- Table A1.13. Summary of likelihood scores for asymmetric migration demographic models (≤10% missing data).
- Table A1.14. Demographic parameters inferred under the full migration, North-Central bifurcation model (maximal 0.90 ADMIXTURE ancestry cut-off).
- Table A1.15. Demographic parameters inferred under the full migration, North-Central bifurcation model ((≤10% missing data).
- Figure A1.15. Relationship between genetic diversity (proportion of heterozygous sites) and geography, by region.

- Table A1.16. Mantel and partial Mantel test results for alternative assignment cutoffs, by geographical region.
- Table A1.17. Mantel correlation coefficients describing the relationship between geographical and genetic distances across different geographical distance intervals within each region.

Chapter 3 : Evidence of host-associated phenotypic divergence, but not host-associated genetic divergence, between populations of the redheaded pine sawfly, *Neodiprion lecontei*, on two northern hosts

3.1 – INTRODUCTION

The field of landscape genetics was formally defined by Manel *et al.* (2003), in an effort to combine population genetics with landscape ecology. Since then, the field has rapidly expanded, with hundreds of studies examining how gene flow varies across genomes, how genetic differentiation is distributed across space, and how gene flow is shaped by geographic and ecological features of landscapes (see Holderegger & Wagner 2006; Storfer *et al.* 2007, 2010; Manel & Holderegger 2013 for reviews of the field). However, many landscape genetic studies are largely exploratory, and do not propose hypotheses based on the ecological and evolutionary history of the organisms under question, making interpretation of spatial patterns in genetic differentiation difficult (Richardson *et al.* 2016). Although landscape genomics on its own will allow description of the underlying structure and patterns of divergence between and within structural units, it is difficult to assess if these patterns are biologically relevant, let alone what mechanisms have generated the observed patterns of genetic variation, without also considering how phenotypes are distributed across the space.

Descriptions variation in both genetic variation and phenotypes across space are central to understanding whether patterns of host use affect genetic differentiation between populations of phytophagous insects. Due to the intimate and often highly specialized relationships insects have with their host plants, host shifts have long been implicated as a potential driver of population differentiation and speciation within insect

lineages (Bush 1969a; Funk 1998; Drès & Mallet 2002a; Berlocher & Feder 2002; Nosil *et al.* 2002; Matsubayashi *et al.* 2010). One of many groups where host use has long been suspected to play a role in population divergence and speciation are *Neodiprion* sawflies (Ghent & Wallace 1958b; Alexander & Bigelow 1960; Knerer & Atwood 1972, 1973, Bush 1975a; b). Comparative analyses by Linnen and Farrell (2010) showed host shifts correlated with speciation events within the genus, but could not determine if the shifts drove speciation events, or occurred after speciation via other mechanisms. Determining if host shifts can initiate speciation events in *Neodiprion*, it is necessary to look for evidence of host-associated divergence *within* species, before post-speciational changes can accumulate (Coyne & Orr 2004).

Recently, Bagley *et al.* (2017) took a landscape genetics approach to evaluate the role of host use in driving neutral genetic divergence within *Neodipron lecontei*, a widely distributed pest species that uses multiple hosts throughout its range (Benjamin 1955; Wilson *et al.* 1992). To do this, they used correlation tests to look for evidence of isolation-by-environment (IBE), where pairs occupying similar environments (in this case, using the same host plant) are less genetically differentiated than pairs in different environments (using different host plants), independent of the geographic distance considered (Wang & Summers 2010; Bradburd *et al.* 2013; Sexton *et al.* 2014; Wang & Bradburd 2014). Analogous to the well-known isolation-by-distance (IBD) pattern (Wright 1943; Slatkin 1993), in both cases opportunities for gene flow are reduced between “distant” populations, allowing genome-wide neutral divergence to accumulate via drift.

Although IBE was identified in two of the three genetic clusters found within *N. lecontei*, this pattern was absent in the third, North cluster (Bagley *et al.* 2017). There are a number of circumstances, however, under which IBE may not be detected, even if there is divergent selection between environments (Thibert-Plante & Hendry 2010; Nosil 2012). Furthermore, in the region where IBE was not detected, there are a number of historical records indicating *N. lecontei* displays strong host preferences in the region (Atwood & Peck 1943; Benjamin 1955). Therefore, in the absence of further ecological and phenotypic data, it is difficult to determine if host use does contribute to divergence in North, or why the pattern of IBE is absent.

The goal of this chapter is to determine what role, if any, host use has on driving patterns of neutral and phenotypic divergence within the North cluster. In North, which is roughly defined by the Canadian province of Ontario and the U.S. states of Michigan and Wisconsin (Bagley *et al.* 2017), *N. lecontei* is found primary on only two host plants – *Pinus banksiana* and *P. resinosa* (Benjamin 1955, personal observation). Using individuals from populations sampled across the cluster and on both primary hosts, we examine how genotypic and phenotypic variation is spatially distributed within the region, and determine if there is evidence of host-associated divergence (HAD) in neutral markers or host-related phenotypic variation. To do so, we assess cluster-wide patterns of neutral genomic divergence, as well as the geographical distribution of adult oviposition preference and a performance-related morphological trait. If HAD is occurring, we predict 1) populations utilizing different hosts will have greater levels of neutral genetic divergence than those on the same host; 2) populations will display distinct preferences for their respective hosts, and/or 3) populations will vary in their performance-related

morphological trait. Alternatively, if host does not contribute to divergence, we predict there will be no evidence of genetic divergence, no variation in host preference, and no morphological variation.

3.2 – Materials and Methods

3.2.1 – Geographic Patterns of Neutral Divergence

3.2.1.1 – Sampling and ddRAD Library Preparation

We sampled populations of *N. lecontei* from 28 locations throughout Ontario, Michigan and Wisconsin. In total, *N. lecontei* utilized only *P. resinosa* at 14 locations, only *P. banksiana* at 7 locations, and both hosts at 7 locations. For most locations, we sampled between 2 and 5 individuals per host per site; but for two of the sympatric sites we increased to 8-15 individuals per host per site. Our final dataset consisted of 185 individuals (Table A2.1).

We generated a genome-wide dataset of putatively neutral markers via ddRAD sequencing (Peterson *et al.* 2012). Following the general protocol of our previous ddRAD protocol (Bagley *et al.* 2017), genomic DNA was extracted from ethanol-preserved larvae and adult females (Table A2.1) using a CTAB/Phenol-Chloroform-Isoamyl alcohol method based on Chen *et al.* (2010). We checked for degradation by visualizing each extraction on a 0.8% agarose gel. The concentration of each sample was then estimated using a Quant-iT High Sensitivity DNA Assay kit (Invitrogen – Molecular Probes, Eugene, OR, USA). The DNA was then fragmented using NlaIII and EcoRI restriction enzymes (NEB, Ipswich, MA).

Individuals were assigned to one of 8 groups of up to 48 individuals. Group assignments were made based on DNA yield and randomized with respect to location.

Each sample was assigned one of 48 unique in-line barcodes during adapter ligation (Table A2.2). We modified the original 48 P1 adapters introduced in Peterson *et al.* 2012 to contain the 5-10 bp variable-length barcodes utilized by Burford Reiskind *et al.* (2016; Table A2.3). Use of variable length barcodes increases sequence diversity over the recognition site of the P1 enzyme, improving sequencing quality and cluster density, while simultaneously allowing a reduction in the amount of PhiX spike-in used (A. Hernandez, personal communication).

Each group of samples was then pooled for automatic size selection of a 379-bp fragment (+/- 76bp) on a PippinPrep (Sage Science, Beverly, MA), followed by 12 rounds of high-fidelity PCR amplification (Phusion High-Fidelity DNA Polymerase, NEB, Ipswich, MA) using PCR primers that included one of 12 unique Illumina multiplex read indices (Table A2.2, A2.4). We included a string of 4 degenerate bases next to the Illumina read index to allow for the detection of PCR duplicates (Schweyen *et al.* 2014).

After verifying library quality on a Bioanalyzer 2100 (Agilent, Santa Clara, CA), libraries were sent to the High-Throughput Sequencing and Genotyping Unit at the University of Illinois, where they were sequenced using 150bp single-end reads on two lanes of an Illumina HiSeq 4000. The sequence of the four degenerate bases in the Illumina index was provided as an additional fastq file.

3.2.1.2 – Data Processing and SNP Genotyping

Raw sequence reads were demultiplexed, quality filtered and trimmed using the *process_radtags* module in STACKS (v1.46; Catchen *et al.* 2013), yielding an average of 1.81 ± 1.90 million high quality single-end reads per individual. The quality-filtered

reads were then aligned to a high-coverage genome assembly for *N. lecontei* (Vertacnik *et al.* 2016; coverage: 112x; scaffold N50: 244kb; GenBank assembly accession: GCA_001263575.1) using the "very sensitive" end-to-end alignment mode in BOWTIE2 (v2.3.1; Langmead & Salzberg 2012). Reads with MAPQ scores <30 or mapping 0 or multiple times to our genome were discarded using SAMTOOLS (v1.3; Li *et al.* 2009). A custom python script was used to identify and remove putative PCR duplicates based on the sequence of the degenerate bases in the index sequence, resulting in 0.95 ± 0.70 million alignments per individual. Finally, we used STACKS' *ref_map.pl* pipeline (v1.46; Catchen *et al.* 2013) to construct RAD loci from the alignments. To enable high-confidence genotype calls (Kenny *et al.* 2011; Peterson *et al.* 2012), we required all loci to have at least 10x depth of coverage per individual (-*m* 10), producing 17,681 ± 7,271 RAD loci with an average coverage of 45.67 ± 25.15x.

After initial assessment of the data, we excluded 5 individuals that received less than 50,000 raw reads each (range: 16,000-46,000), and could not be processed in the STACKS pipeline. Using our initial SNP dataset of 25,999 SNPs, we inferred ploidy of remaining individuals using heterozygosity estimates from VCFTOOLS' --*het* option (v0.1.14b; Danecek *et al.* 2011) as in Bagley *et al.* (2017); and excluded 25 putative haploids based on their strikingly low proportion of heterozygous sites (Table 3.2).

Our final dataset consisted of 11,539 SNPs from 155 putatively diploid individuals, to which we applied several additional quality filters. First, we included only sites where at least 70% of individuals had data. Second, we removed sites displaying deviation from Hardy-Weinberg equilibrium for heterozygote excess significant at the

0.01 level. Finally, we included only one SNP per RAD locus to minimize linkage disequilibrium between SNPs, leaving us with 6,824 total SNPs.

Demultiplexing, alignments and SNP calls were performed on the University of Kentucky's Lipscomb High Performance Computing Cluster.

3.2.1.3 – Population Structure

Population structure within the region was investigated using the maximum-likelihood clustering algorithm implemented in ADMIXTURE (v1.3.0; Alexander *et al.* 2009). This method determines the ancestry of each individual to *K* ancestral populations without *a priori* designation. We performed 100 independent runs for values of *K* from 1 through 15. The fit of each value of *K* to our data was determined by comparing the average 5-fold cross-validation (CV) error of different values of *K.* We then used the main pipeline of CLUMPAK (v1.1; Kopelman *et al.* 2015) to determine assignment stability, and summarize primary and alternate solutions across the 100 replicates of each *K*. These solutions were then visualized using a custom R script.

Based on our region-wide analyses, we also looked for hierarchical structure by performing ADMIXTURE analyses within each of the identified clusters (see Results and Discussion). For each cluster, we again performed 100 independent runs for *K* 1 through 15; and summarized the results using CLUMPAK.

3.2.1.4 – Analysis of Molecular Variance (AMOVA) and Population Differentiation

We used the locus-by-locus AMOVA implemented in ARLEQUIN (v3.5.2.2; Excoffier & Lischer 2010) to assess if host use contributes to differentiation region-wide and within each of the identified clusters. Individuals in each analysis were grouped by

their host plant (*P. banksiana* or *P. resinosa*). Statistical significance was assessed using 10,000 permutations.

We also calculated pairwise F_{ST} between the identified clusters using both the *populations* module of STACKS (v1.46; Catchen *et al.* 2013); and in ARLEQUIN.

3.2.1.5 – Isolation-by-Distance and Isolation-by-Environment

We reevaluated the relationship between neutral divergence, geography and host use throughout the region using Mantel and partial Mantel tests (Mantel 1967; Sokal 1979; Smouse *et al.* 1986). As hierarchical structure can bias Mantel tests, we repeated the tests for each of the clusters we identified in ADMIXTURE (Kuchta & Tan 2005; Meirmans 2012).

For these tests, we produced genetic, geographic and ecological distance matrixes. For the genetic distance matrix, we estimated differentiation between sampling locations using the $F_{ST}/(1-F_{ST})$ ratio. We used SPAGEDI (v1.4b; Hardy & Vekemans 2002) to generate $F_{ST}/(1-F_{ST})$ and linear geographic distance matrixes for the full region and for each of the genetic clusters. Ecological distance matrixes were created by coding host use as a binary trait, with population pairs utilizing the same host assigned a distance of 0, and pairs utilizing different hosts a distance of 1.

We tested for isolation-by-distance (IBD) and isolation-by-environment (IBE) by performing partial Mantel tests (Smouse *et al.* 1986), which examine the correlation between two matrices while controlling for the influence of a third. For IBD tests, we compared genetic and geographical distance, and controlled for ecological distance. For IBE tests, we compared genetic and ecological distance, and controlled for geographical distance. We also performed Mantel tests to examine the relationship between ecological

and geographical distance, as high levels of eco-spatial autocorrelation can impact performance of IBE tests (Shafer & Wolf 2013).

We performed Mantel and partial Mantel tests in PASSAGE2 (v2.0.11.6; Rosenberg & Anderson 2011), and determined significance using 10,000 permutations.

3.2.2 – Geography of Host Preference

3.2.2.1 – Spatial Patterns of Host Affiliation

To determine if host affiliation varies across space, we assessed the relative abundance of *N. lecontei* on *P. banksiana* vs *P. resinosa* across the region by examining collection records. We compiled a list of locations associated with *N. lecontei* collected from *P. banksiana* or *P. resinosa* from multiple sources, including: the Canadian Forest Service Federal Insect Disease Survey (FIDS) database; several museum collections including the National Museum of Natural History (Washington, D.C.), the University of Wisconsin (Madison, WI) and Agriculture Canada (Ottawa, Ontario); and our collection records. Collection records that did not include specific latitude and longitude coordinates were assigned coordinates using the U.S. Geological Survey/U.S. Board on Geographic Names' Geographic Names Information System (GNIS) database (U.S. locations) or the National Geospatial-Intelligence Agency/U.S. Board on Geographic Names' GEOnet Names Server (non-U.S. locations). We then performed two-way ANOVAs to determine if collection host correlated latitude and/or longitude.

3.2.2.2 – Spatial Patterns of Host Preference

To assess if regional differences in host affiliation are influenced by varying host preferences, we assessed female oviposition preference in 6 populations collected across the region and from both hosts (Table 3.1). All assays used females reared from wild-

Table 3.1 – Sampling locations included in female oviposition preference assays. The genetic cluster and collection host are noted for each location.

Location	Cluster	Host Plant	Latitude	Longitude
Laurentian Valley, ON	ONT	*P. resinosa*	45.83	-77.24
Grayling, MI	LP MI	*P. banksiana*	44.66	-84.69
Bitely, MI	LP MI	*P. banksiana*	43.76	-85.74
Necedah, WI	UP MI + WI	*P. banksiana*	44.21	-90.14
Suring, WI	UP MI + WI	*P. resinosa*	44.98	-88.45
Thompson Township, MI	UP MI + WI	*P. resinosa*	45.93	-86.29

caught larvae in the laboratory. Briefly, larvae were returned to the lab and placed in plastic boxes (32.4 cm × 17.8 cm × 15.2 cm) with mesh lids, and provided clippings of their natal host species *ab libitum* until cocoon formation. Cocoons were collected and stored in individual gelatin capsules until adult eclosion. Larvae and cocoons were reared in walk-in environmental chambers maintained at 22°C under an 18:6 light-dark cycle. Live adults were stored at 4°C until use.

Oviposition preference was assessed using choice assays. For each assay, a single virgin female was placed in a mesh cage (35.6cm x 35.6cm x 61cm) and offered two *P. banksiana* seedlings and two *P. resinosa* seedlings. As females are short lived and typically oviposit on a single host, each cage was checked daily until eggs were laid or the female died. If eggs were laid, we recorded which host was selected. Between 29 and 38 females were assayed per population. We excluded females that failed to make a choice (n = 13-23). No females laid eggs on both hosts. To determine if females exhibited preference for *P. banksiana* or *P. resinosa*, we performed exact binomial tests.

3.2.2.3 – Temporal Patterns of Host Preference

As the region covers a relatively large and climatologically variable geographic area, it is difficult to assess temporal variation in host use region-wide. However, it is possible to compare eclosion patterns at sites where *N. lecontei* is found on both host species. We compiled adult eclosion data for wild-caught colonies reared in the lab as previously described across three such sites: Grayling, MI; Frederic, MI; and Mosinee, WI. Total adult emergence was pooled for all colonies collected on each host species per site. We then calculated pairwise estimates of isolation (I) between populations on the two hosts following (Feder *et al.* 1993):

$$1 - \left(\frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \cdot \sum y_i^2}} \right) \cdot 100$$

where x_i and y_i represent the proportion of the total number of live adults from host *x* or *y* on day *i*. Based on survival estimates from Benjamin 1955, we assumed an average female lifespan of 5 days, and an average male lifespan of 4 days.

Significance of differences in adult eclosion at each site were assessed in two ways. First, we calculated mean ordinal date of eclosion and variance for each host. Using this summary data, we performed a one-way summary ANOVA, followed by Tukey's Honest Significant Difference tests. Second, we assessed if the shape of cumulative eclosion curves differed using bootstrapped Kolmogorov-Smirnov (KS) tests with the *ks.boot* function from the *R* module MATCHING (v4.9-2; Sekhon 2008).

3.2.3 – Geography of a Performance-Related Trait

3.2.3.1 – Host Architecture

As *Neodiprion* females deposit eggs within the needle tissue of their host plants, differences in needle architecture between hosts may provide a source of divergent selection. We assessed needle architecture at 7 locations in Michigan and Wisconsin where we had collected *N. lecontei* on *P. banksiana* and/or *P. resinosa*. For each location, we used digital calipers (Mitutoyo CD-6"PMX) to measure the length and width of 10 needles on each of 10 trees of the host species utilized by *N. lecontei* at that site. We then calculated the average needle length and width for each of the ten trees of each host at each site. To analyze differences in needle length and width between the hosts, we

performed two-way ANOVAs with tree as a fixed effect nested within site, followed by a Tukey's Honest Significant Difference test.

3.2.3.2 – Ovipositor Morphology

Variation in ovipositor morphology was assessed across 11 sites distributed through the region. We obtained adults collected from *P. banksiana* at 5 locations, from *P. resinosa* at 3 locations, and from both hosts at 3 locations (Table 3.2). We assessed differences in morphology at three spatial levels: 1) all sites region-wide, 2) only sites within each genetic cluster, and 3) within the three sympatric sites. In general, our analyses considered 2-5 ovipositors per host, per site. However, for one sympatric site (Frederic), we considered a total of 20 ovipositors per host in the local analysis.

Ovipositors were dissected, mounted, and imaged as described in Bendall *et al.* (2017). Briefly, ovipositors were dissected from females obtained from wild-caught colonies reared under the previously described conditions, and preserved at -80°C upon eclosion. One lancet from each female was mounted on a glass slide with an 80:20 permount:tolune solution. After drying, each ovipositor was imaged at 5x magnification and the length and width were measured using the ZEN lite 2012 software package (Carl Zeiss Microscopy, LLC; Thornwood, NY).

We assessed differences in ovipositor shape, length, and width. First, we used a geometric morphometric approach to compare overall ovipositor shape differences while controlling for size. We used IMAGEJ (v1.51; Schneider *et al.* 2012) to define the overall shape of each ovipositor with 30 sliding landmarks (Figure A2.1), and transformed the position of each landmark into Cartesian coordinates. Landmarks of each ovipositor were aligned using a general Procrustes alignment in GEOMORPH (v2.1.4; Adams & Otárola-

Table 3.2 - Cluster identify, host and number of ovipositors per sampling location.

Ovipositors were examined in three groups: all locations, within each cluster (LP MI, ONT, UP MI + WI), and at sympatric sites (†). For Frederic (*), 15 additional ovipositors per host were included in the sympatric analysis.

Location	Cluster	Host	# Ovipositors
Bitely, MI	LP MI	*P. banksiana*	5
Frederic, MI†	LP MI	*P. resinosa*	5*
Frederic, MI†	LP MI	*P. banksiana*	5*
Grayling, MI	LP MI	*P. banksiana*	5
Chelmsford, ON	ONT	*P. banksiana*	2
Harcourt, ON	ONT	*P. resinosa*	5
Laurentian Valley, ON	ONT	*P. resinosa*	5
Mosinee, WI†	UP MI + WI	*P. resinosa*	5
Mosinee, WI†	UP MI + WI	*P. banksiana*	4
Necedah, WI	UP MI + WI	*P. banksiana*	5
Rothschild, WI	UP MI + WI	*P. banksiana*	5
Suring, WI	UP MI + WI	*P. resinosa*	5
Thompson, MI†	UP MI + WI	*P. banksiana*	4
Thompson, MI†	UP MI + WI	*P. resinosa*	5

Castillo 2013). We visualized the shape differences using a principal components analysis; and assessed significance of differences using a Procrustes ANOVA with forewing length, host, and sampling location as fixed factors. Length and width differences were assessed using two-way ANOVAs, again including forewing length, host and sampling as fixed factors. We also assessed the impact of overall size on ovipositor length and width by repeating the two-way ANOVAs without including forewing length.

3.3 – Results and Discussion

3.3.1 – Geographic Patterns of Neutral Divergence

Our population structuring analysis revealed a considerable amount of structure within the region. Similar to previously described range-wide patterns of divergence in *N. lecontei* (Bagley *et al.* 2017), this structure corresponded with geographic regions. Region-wide, CV error scores improved past $K = 3$ and remained stable through $K = 12$ (Figure 3.1). As it was difficult to determine the optimal K, we assessed the stability and biological interpretability across all values of K. At $K = 2$, individuals from the lower peninsula of Michigan (LP MI) split from the rest of the region. At $K = 3$, the group from $K = 2$ was maintained, individuals from Ontario formed a second distinct group (ONT), and the remaining individuals from the upper peninsula of Michigan and Wisconsin formed a third (UP MI + WI). Above $K = 3$, assignment stability declined, with $K \geq 4$ presenting at least two distinct solutions. Despite this assignment stability, however, the three groups produced under $K = 3$ were still easily identified across solutions (Figure 3.2). Given the instability of assignments of $K \geq 4$, and the maintenance of three general

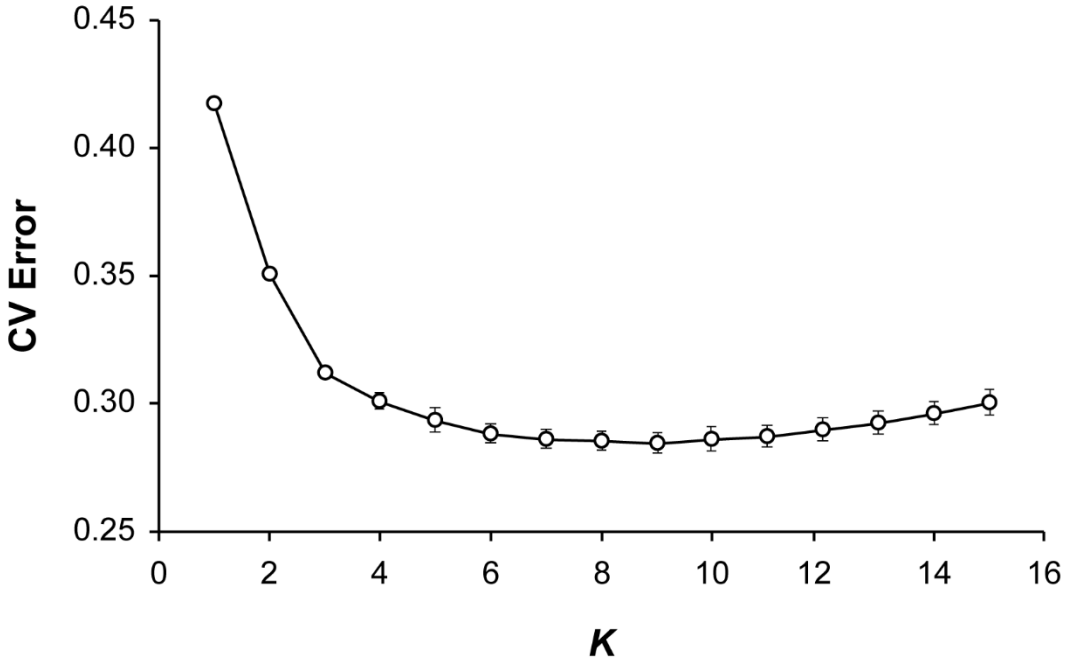


Figure 3.1 – CV error score across 100 independent ADMIXTURE runs. CV scores improve past $K = 3$, and remain stable through $K = 12$.

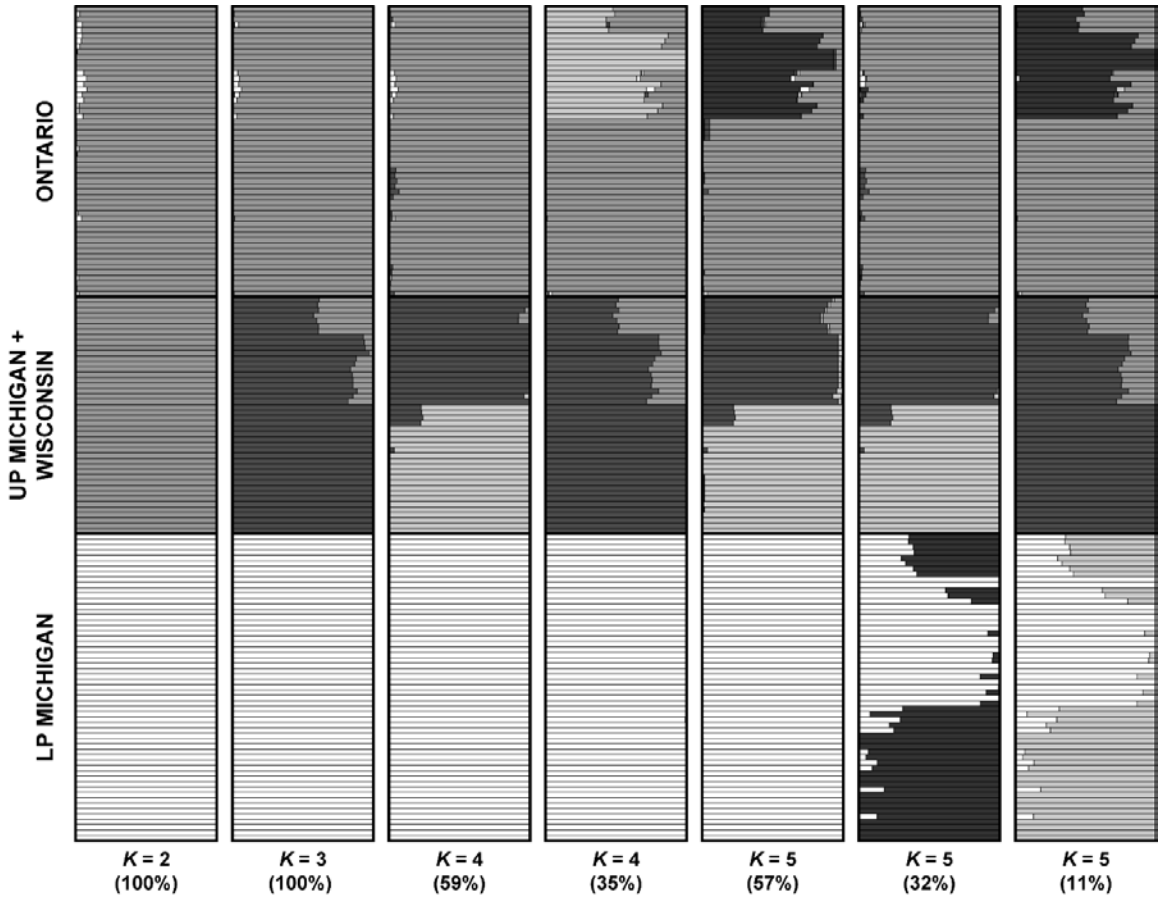


Figure 3.2 – Region-wide assignment of individuals under *K* = 2 through *K* = 5. To facilitate comparison across different solutions and values of *K*, individuals are displayed in the same order across plots. The proportion of runs in which a given assignment solution was identified is listed in parentheses. In each plot, the thick black boxes indicate the three clusters identified under $K = 3$.

groups across several values of *K*, we considered *K* = 3 for all subsequent analyses (Table 3.3).

These three clusters exhibit moderate levels of differentiation, with the highest levels of differentiation existing between LP MI and UP MI + WI (Table 3.4). Within each of these clusters, we found evidence of additional hierarchical structure, often corresponding to areas of uncertainty in the region-wide solutions. To investigate, we visualized all assignment solutions up to, and around the optimal *K* for each cluster. In LP MI (Figure 3.3), *K* = 2 was optimal, with one sampling location forming a distinct group. At *K* = 3, two sampling sites formed distinct groups, and the remaining individuals formed the third. In ONT (Figure 3.4), *K* = 3 was favored. Here, populations in western Ontario separated from those in eastern Ontario, with a geographically intermediate population forming a third distinct cluster. In UP MI + WI (Figure 3.5), *K* = 4 was best. At *K* = 2, populations in Wisconsin separated from those in the upper peninsula of Michigan. At higher values of *K*, individual collecting locations, or groups of geographically close locations, formed distinct populations.

After identifying the underlying structure in the region, we looked for evidence of neutral divergence between populations of *N. lecontei* associated with *P. banksiana* and *P. resinosa* throughout the region, and within each of the three clusters. First, we looked for evidence of IBD and IBE region-wide, and within each of the three clusters. Significant IBD was detected region-wide, in LP MI, and in UP MI + WI; but was not detected in ONT (Table 3.5). Although the shape of the IBD relationship varies depending on the spatial scale investigated, in all cases, there is a high degree of scatter

Table 3.3 – Cluster assignments per individual. Based on our admixture results (Figure 3.2), we assigned all individuals from each sampling into one of three geographically based clusters: lower peninsula Michigan (LP MI), Ontario (ONT), and upper peninsula Michigan/Wisconsin (UP MI + WI).

ID	Collection location	Cluster
CAN061a	Bitely, MI	LP MI
CAN099	Bitely, MI	LP MI
RB080	Bitely, MI	LP MI
RB136.01	Bitely, MI	LP MI
RB136_02	Bitely, MI	LP MI
RB236	Bitely, MI	LP MI
RB237	Bitely, MI	LP MI
RB385	Bitely, MI	LP MI
RB444	Frederic, MI	LP MI
RB445_b	Frederic, MI	LP MI
RB447	Frederic, MI	LP MI
RB448_02	Frederic, MI	LP MI
RB450	Frederic, MI	LP MI
RB451	Frederic, MI	LP MI
RB452	Frederic, MI	LP MI
RB454	Frederic, MI	LP MI
RB456	Frederic, MI	LP MI
RB457	Frederic, MI	LP MI
RB458	Frederic, MI	LP MI
RB460	Frederic, MI	LP MI
RB461	Frederic, MI	LP MI
RB462_02	Frederic, MI	LP MI
RB463	Frederic, MI	LP MI
RB465	Frederic, MI	LP MI
RB466	Frederic, MI	LP MI
RB468	Frederic, MI	LP MI
RB469_02	Frederic, MI	LP MI
RB470	Frederic, MI	LP MI
RB473	Frederic, MI	LP MI
RB476	Frederic, MI	LP MI
RB477	Frederic, MI	LP MI
RB478	Frederic, MI	LP MI

RB479_01	Frederic, MI	LP MI
RB484	Frederic, MI	LP MI
RB485	Frederic, MI	LP MI
CAN043	Grayling, MI A	LP MI
CAN045	Grayling, MI A	LP MI
CAN047_02	Grayling, MI A	LP MI
RB245_02	Grayling, MI A	LP MI
RB247.02	Grayling, MI A	LP MI
CAN048	Grayling, MI B	LP MI
CAN049	Grayling, MI B	LP MI
CAN050	Grayling, MI B	LP MI
CAN054_new	Grayling, MI B	LP MI
CAN055_old	Grayling, MI B	LP MI
CAN056_old	Grayling, MI B	LP MI
RB091_01 ♀	Grayling, MI B	LP MI
RB092_01	Grayling, MI B	LP MI
RB093_01	Grayling, MI B	LP MI
RB249	Grayling, MI B	LP MI
RB250_old	Grayling, MI B	LP MI
RB251_02	Grayling, MI B	LP MI
RB252_old	Grayling, MI B	LP MI
RB253_a	Grayling, MI B	LP MI
RB254_old	Grayling, MI B	LP MI
RB256_old	Grayling, MI B	LP MI
RB257	Grayling, MI B	LP MI
RB258_a	Grayling, MI B	LP MI
RB259b	Grayling, MI B	LP MI
RB260_a	Grayling, MI B	LP MI
RB261	Grayling, MI B	LP MI
RB386	Grayling, MI B	LP MI
RB387_02	Grayling, MI B	LP MI
RB388	Grayling, MI B	LP MI
CAN092	Algoma, ON	ONT
CAN093	Algoma, ON	ONT
CAN094	Algoma, ON	ONT
CAN042	Baldwin, ON	ONT
CAN042_02	Baldwin, ON	ONT
CAN042_03	Baldwin, ON	ONT
CAN042_04	Baldwin, ON	ONT
CAN042_06	Baldwin, ON	ONT
CAN021	Barry's Bay, ON A	ONT

CAN022	Barry's Bay, ON A	ONT
CAN023	Barry's Bay, ON A	ONT
CAN025_02	Barry's Bay, ON A	ONT
CAN026	Barry's Bay, ON B	ONT
CAN030	Barry's Bay, ON B	ONT
CAN031	Barry's Bay, ON B	ONT
CAN032	Barry's Bay, ON B	ONT
CAN034	Barry's Bay, ON B	ONT
CAN063	Blind River, ON A	ONT
CAN069	Blind River, ON A	ONT
CAN070	Blind River, ON A	ONT
CAN071	Blind River, ON A	ONT
CAN072	Blind River, ON A	ONT
CAN064	Blind River, ON B	ONT
CAN065	Blind River, ON B	ONT
CAN066	Blind River, ON B	ONT
CAN067	Blind River, ON B	ONT
CAN068	Blind River, ON B	ONT
CAN080	Chelmsford, ON	ONT
CAN080.02	Chelmsford, ON	ONT
CAN080_03	Chelmsford, ON	ONT
CAN080_04 ♀	Chelmsford, ON	ONT
CAN080_05	Chelmsford, ON	ONT
CAN013	Combermere, ON	ONT
CAN015	Combermere, ON	ONT
CAN016	Combermere, ON	ONT
CAN017	Combermere, ON	ONT
CAN019	Combermere, ON	ONT
RB373_02	Dunlop Lake, ON	ONT
RB374	Dunlop Lake, ON	ONT
RB376	Dunlop Lake, ON	ONT
RB377_02	Dunlop Lake, ON	ONT
RB378	Dunlop Lake, ON	ONT
CAN007	Harcourt, ON	ONT
CAN007_02	Harcourt, ON	ONT
CAN007_03	Harcourt, ON	ONT
CAN007_04	Harcourt, ON	ONT
CAN007_05	Harcourt, ON	ONT
CAN037a	Laurentian Valley, ON	ONT
CAN037a.02	Laurentian Valley, ON	ONT
CAN038	Laurentian Valley, ON	ONT

CAN039_02	Laurentian Valley, ON	ONT
CAN040_01	Papineau-Cameron, ON	ONT
CAN040_02	Papineau-Cameron, ON	ONT
CAN040_03	Papineau-Cameron, ON	ONT
CAN040_04	Papineau-Cameron, ON	ONT
CAN040_06	Papineau-Cameron, ON	ONT
171_02_N	Sebrite, ON	ONT
173-02	Sebrite, ON	ONT
176_02	Sebrite, ON	ONT
CAN002.02	Sebrite, ON	ONT
CAN002_01	Sebrite, ON	ONT
CAN005.01	Sebrite, ON	ONT
CAN005_02	Sebrite, ON	ONT
CAN075	Spanish, ON	ONT
CAN075_02	Spanish, ON	ONT
CAN075_04	Spanish, ON	ONT
CAN076	Spanish, ON	ONT
RB437	Friendship, WI	UP MI + WI
RB438	Friendship, WI	UP MI + WI
RB439_02	Friendship, WI	UP MI + WI
RB440	Friendship, WI	UP MI + WI
RB099_01	Manistique, MI A	UP MI + WI
RB099_03	Manistique, MI A	UP MI + WI
RB099_04	Manistique, MI A	UP MI + WI
RB099_05	Manistique, MI A	UP MI + WI
RB405	Manistique, MI B	UP MI + WI
RB406	Manistique, MI B	UP MI + WI
RB431	Mosinee, WI	UP MI + WI
RB431_01	Mosinee, WI	UP MI + WI
RB431_02	Mosinee, WI	UP MI + WI
RB432	Mosinee, WI	UP MI + WI
RB433	Mosinee, WI	UP MI + WI
RB434	Mosinee, WI	UP MI + WI
RB435_02	Mosinee, WI	UP MI + WI
RB095_02	Naubinway, MI A	UP MI + WI
RB095_03	Naubinway, MI A	UP MI + WI
RB095_04	Naubinway, MI A	UP MI + WI
RB096B	Naubinway, MI B	UP MI + WI
RB098_01	Naubinway, MI B	UP MI + WI
RB480_02	Naubinway, MI B	UP MI + WI
RB480_1	Naubinway, MI B	UP MI + WI

RB481	Naubinway, MI B	UP MI + WI
RB482	Naubinway, MI B	UP MI + WI
RB483	Naubinway, MI B	UP MI + WI
CAN111	Necedah, WI	UP MI + WI
RB284	Necedah, WI	UP MI + WI
RB285	Necedah, WI	UP MI + WI
RB399	Necedah, WI	UP MI + WI
RB402	Necedah, WI	UP MI + WI
RB425	Rothschild, WI	UP MI + WI
RB426	Rothschild, WI	UP MI + WI
RB427	Rothschild, WI	UP MI + WI
RB428	Rothschild, WI	UP MI + WI
RB429	Rothschild, WI	UP MI + WI
RB101	Suring, WI	UP MI + WI
RB104_01	Suring, WI	UP MI + WI
RB390	Suring, WI	UP MI + WI
RB395	Suring, WI	UP MI + WI
RB421	Suring, WI	UP MI + WI
RB407	Thompson Township, MI	UP MI + WI
RB408	Thompson Township, MI	UP MI + WI
RB409	Thompson Township, MI	UP MI + WI
RB410	Thompson Township, MI	UP MI + WI
RB411	Thompson Township, MI	UP MI + WI
RB412	Thompson Township, MI	UP MI + WI
RB413	Thompson Township, MI	UP MI + WI

Table 3.4 – F_{ST} between North clusters. Estimates considering monomorphic and polymorphic sites are given on the upper diagonal, and considering only polymorphic sites on the lower diagonal.

	ONT	**UP MI + WI**	**LP MI**
ONT	-	0.063	0.086
UP MI + WI	0.203	-	0.107
LP MI	0.264	0.316	-

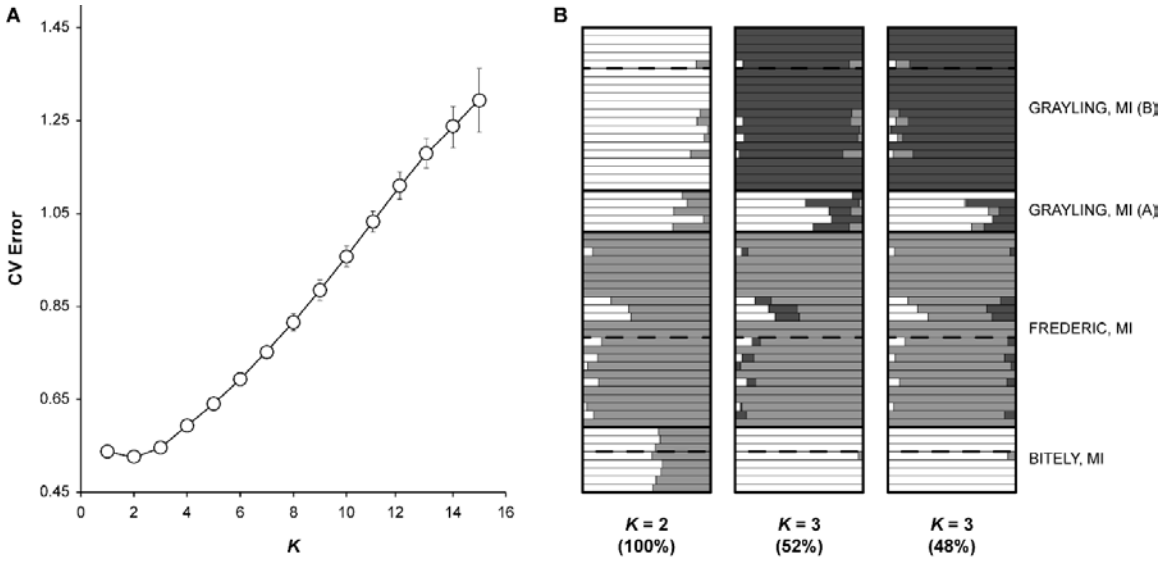


Figure 3.3 – Hierarchical structure within the LP MI cluster. A. CV error scores suggest $K = 2$ is optimal, with scores rapidly worsening past $K = 3$. B. Assignment solutions for $K = 2$ and $K = 3$. Thick black lines are drawn around sampling locations within the cluster. If *N. lecontei* was collected on both hosts at the site, dotted lines separate individuals collected on *P. resinosa* from those collected on *P. banksiana*. Substructure in this cluster is largely within sampling locations.

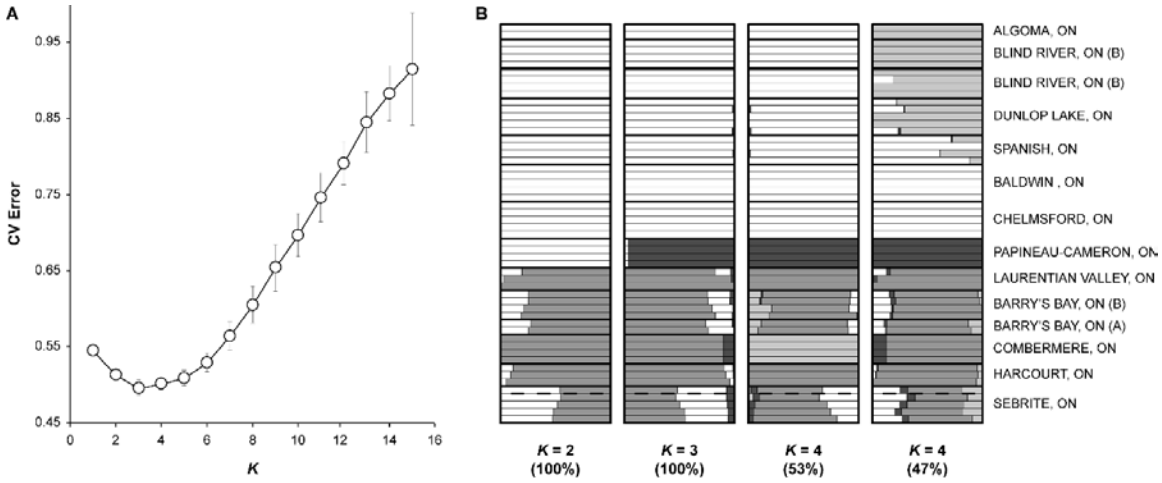


Figure 3.4 – Hierarchical structure within the ONT cluster. A. CV error scores support *K* = 3. B. Assignment solutions for *K* = 2 through *K* = 4. Thick black lines are drawn around sampling locations within the cluster, which are listed in rough East-to-West order. If *N. lecontei* was collected on both hosts at the site, dotted lines separate individuals collected on *P. resinosa* from those collected on *P. banksiana*. Populations in eastern Ontario and western Ontario form distinct groups beginning at *K* = 2.

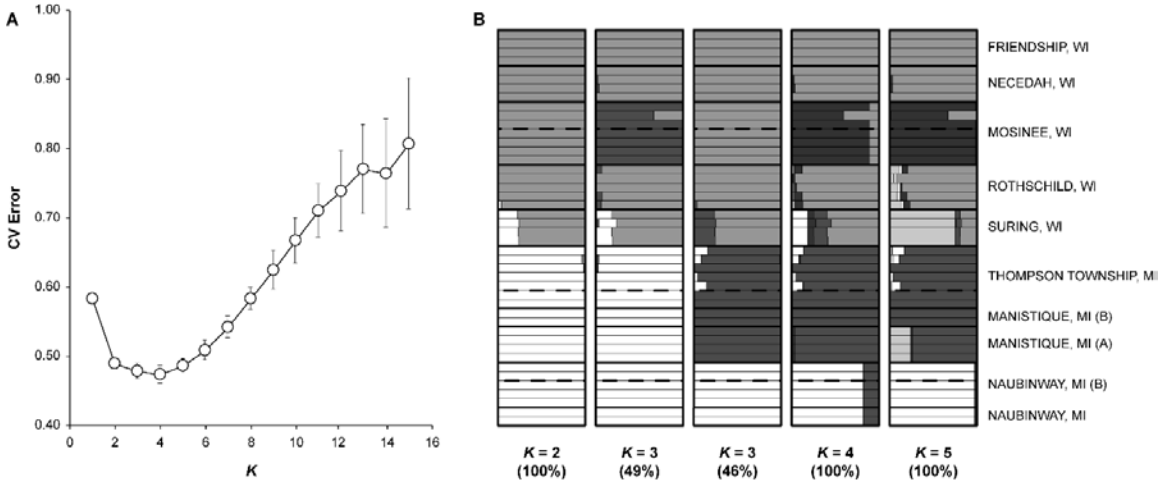


Figure 3.5 – Hierarchical structure within the UP MI + WI cluster. A. The optimal *K* is 4, but CV scores are similar between *K* = 2 and *K* = 5. B. Assignment solutions for *K* = 2 through *K* = 5.

Table 3.5 – Mantel and partial Mantel test results. Pearson's *r* and *P*-value for each Mantel (Matrix 1, Matrix 2) and partial Mantel test (Matrix 1, Matrix 2 | list of matrices held constant) are given for region-wide, and cluster-specific analyses. The individuals and sampling locations assigned to each cluster are indicated in Table 3.3.

Comparison	***r***	***P*-value**
Region-wide		
Geographic, Genetic	0.43	<0.0001
Geographic, Genetic | Host (IBD)	0.43	<0.0001
Host, Genetic	0.05	0.3077
Host, Genetic | Geographic (IBE)	0.00	0.9423
Geographic, Host (eco-spatial auto.)	0.10	0.0348
LP MI		
Geographic, Genetic	0.56	<0.0001
Geographic, Genetic | Host (IBD)	0.55	<0.0001
Host, Genetic	-0.26	0.3211
Host, Genetic | Geographic (IBE)	-0.21	0.2624
Geographic, Host (eco-spatial auto.)	-0.15	0.1364
ONT		
Geographic, Genetic	0.11	0.0892
Geographic, Genetic | Host (IBD)	0.10	0.1122
Host, Genetic	-0.08	0.7399
Host, Genetic | Geographic (IBE)	-0.07	0.7530
Geographic, Host (eco-spatial auto.)	-0.09	0.1571
UP MI + WI		
Geographic, Genetic	0.57	0.0002
Geographic, Genetic | Host (IBD)	0.57	0.0002
Host, Genetic	-0.04	0.7100
Host, Genetic | Geographic (IBE)	0.03	0.8073
Geographic, Host (eco-spatial auto.)	-0.10	0.1909

between points at all geographical distances, even when significant IBD is detected (Figure 3.6).

The amount of scatter between pairwise points reflects the relative influence of gene flow and drift. As gene flow generally reduces differentiation between populations, geographically close populations should exhibit small genetic distances, and have less variability in pairwise genetic distance than geographically distant populations. At migration-drift equilibrium, then, scatter should be narrow at close geographic distances, and widen gradually as geographic distance increases, reflecting a relatively greater role of drift on differentiation as limited dispersal weakens the homogenizing effect of gene flow. When populations are not at equilibrium, however, the relationship between degree of scatter and geographic distance can be used to infer the relative strength of drift and gene flow. When drift has a greater influence than gene flow, there should be greater variability in genetic differentiation at a given geographic distance, producing a wider scatter of points. If gene flow is more influential, however, variation in differentiation will be reduced, producing a narrow scatter (Hutchinson & Templeton 1999).

The wide scatter of points across the region, as well as the lack of significant IBD within ONT, strongly suggest this region has not yet reached equilibrium. This is also consistent with our structure results – populations in this region are likely small and fairly isolated from each other, allowing drift to dominate patterns of genetic differentiation. The shape of the IBD relationship region-wide, and within the two clusters where significant IBD is detected, however, suggests that gene flow is beginning to homogenize differentiation, as there is a positive and somewhat monotonic relationship between genetic and geographic distance (Figure 3.6A, B, D). Given time, the scatter of points in

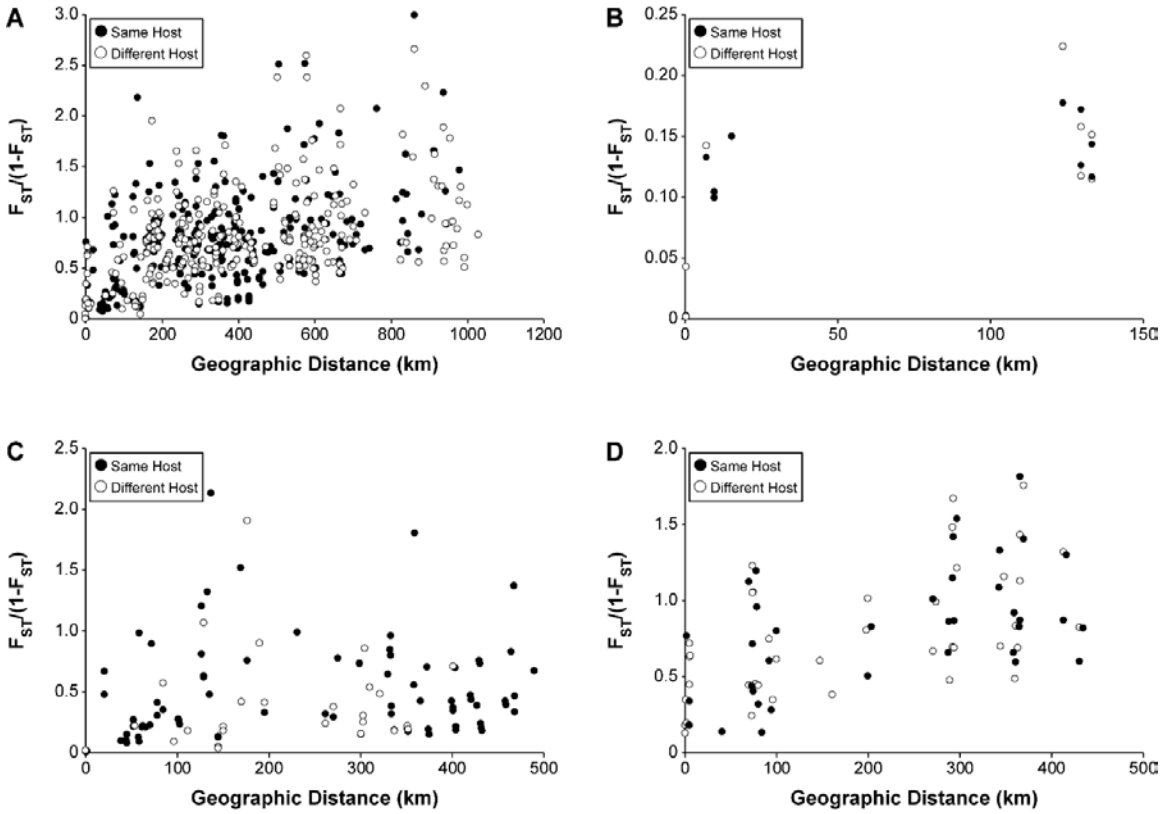


Figure 3.6 - Isolation-by-distance and isolation-by-environment, by region and cluster. Pairwise geographic and genetic distances [measured by the $F_{ST}/(1-F_{ST})$ ratio] for the North region (A), and within the LP MI (B), ONT (C), and UP MI + WI (D) clusters. Colored circles indicate the pair were collected on the same host plant, while open circles represent pairs collected on different host plants.

these regions may narrow, as the strength of IBD in non-equilibrium populations correlates with the number of generations that have passed since colonization (Slatkin 1993; Hutchinson & Templeton 1999).

This may also explain why IBD is not seen in the ONT cluster. The North region spans a considerable geographic and climactic range, so the number of generations per year varies between the identified clusters. While there are typically 2 generations per year in LP MI and UP MI + WI, *N. lecontei* is generally univoltine in ONT (Benjamin 1955; personal observation). Although not significant across the entire cluster, the correlation between genetic and geographic distance is strong and monotonic between pairs separated by less than 250km ($r = 0.44$), but levels off in pairs separated by greater distances ($r = 0.11$). This suggests IBD has begun to emerge in neighboring populations, but has not had enough time to extend to more distant pairs (Slatkin 1993).

The lack of regional mutation-drift equilibrium, and relative isolation of individual populations may contribute to the lack of IBE seen within the region. Consistent with our previous results (Bagley *et al.* 2017), we did not detect significant IBE region wide, or within any of the three clusters (Table 3.5; Figure 3.6). IBE is most consistently detected when migration is intermediate (Thibert-Plante & Hendry 2010). If gene flow is weak, a similar amount of drift could occur between all population pairs, regardless of their ecological distance, obscuring any potential IBE signal.

Further support for a general lack of host-related neutral divergence was provided by the locus-by-locus AMOVAs (Table 3.6). The impact of host on genetic differentiation was significant in ONT and UP MI + WI, but explained less than 6% of variation. In contrast, at least 67% of variation came from within individuals. It should

Table 3.6 – Locus-by-locus AMOVA tables, by cluster. Significance was determined via 10,000 permutations.

Source of Variation	Sum of Squares	Variance Components	Percent Variation	*P*-value
LP MI				
Among hosts	646.98	2.74	0.58	1.0000
Among individuals within hosts	25495.48	43.48	9.22	<0.0001
Within individuals	22054.00	425.29	90.20	<0.0001
ONT				
Among hosts	1293.08	32.54	5.42	<0.0001
Among individuals within hosts	32987.52	138.41	23.06	<0.0001
Within individuals	20922.50	429.30	71.52	<0.0001
UP MI + WI				
Among hosts	1262.91	14.41	2.66	<0.0001
Among individuals within hosts	27272.73	163.89	30.27	<0.0001
Within individuals	15100.00	363.06	67.06	<0.0001

be considered, however, that we grouped together all individuals collected on each host, and did not account for additional structure contributed by sampling location. Future AMOVA tests should be performed to determine how much variation is impacted by sampling location.

It should be noted, however, that a lack of host-related neutral genomic divergence does not mean divergence is not occurring between *N. lecontei* on *P. banksiana* and *P. resinosa*. There are several situations under which patterns of IBE may not be detected, some of which may apply to *N. lecontei* in this region. First, inadequate sampling across geographic and environmental gradients may complicate or distort detection of IBE (Wang & Bradburd 2014). As our increased sampling strategy includes multiple individuals from several populations spread across a large geographic range and from both of the focal host plants (including multiple sympatric locations), however, this is unlikely to have impacted our IBE results.

Second, as with patterns of IBD, not enough time may have passed to allow neutral divergence to accrue between hosts. During ecological speciation, divergence is at first limited only to those loci under direct selection. As the number of loci under selection increases, however, there can be a genome-wide reduction in gene flow, allowing differentiation at neutral loci to accumulate via drift. It may take time for this genome-wide reduction in gene flow to develop, as well as for neutral diverge to accumulate even after the reduction has occurred (Feder *et al.* 2012b). If selection is weak relative to gene flow, however, divergence may remain limited to directly selected loci, and no neutral divergence may accumulate (Saint-Laurent *et al.* 2003; Crispo *et al.* 2006; Yatabe *et al.* 2007; Thibert-Plante & Hendry 2009; Nosil 2012).

Third, IBE may not be detected if the ecological distance considered does not accurately represent the source of selection between populations (Nosil 2012). Our ecological matrix distills differences between host plants to simple 0/1 distances. Although we are generally interested in the overall impact of host use on divergence, and only two host plants are considered here, our simplification may impact detection of IBE. Future work in this and other regions incorporating more specific differences between host species (variation in needle architecture, resin content, volatile profiles, etc.) could be of use in determining the specific source(s) of host-driven divergent selection.

Finally, if geography and host use are too strongly confounded, it may be difficult or impossible to disentangle their relative impact on neutral divergence. This limitation may particularly apply in our case, as we assessed IBE using partial Mantel tests, which have notoriously high Type I error rates when geography and ecology are autocorrelated (Raufaste & Rousset 2001; Harmon & Glor 2010; Legendre & Fortin 2010; Guillot & Rousset 2013; but see Castellano & Balletto 2002; Cushman & Landguth 2010; Diniz-Filho *et al.* 2013). Although we did not detect significant IBE in this case, several more statistically robust methods for disentangling the effects of geography and ecology are available (Freedman *et al.* 2010; Wang *et al.* 2013; Wang 2013) One such example is the program BEDASSLE, which uses a Bayesian framework to quantify the relative contributions of geography and ecology to genetic differentiation (Bradburd *et al.* 2013). As parameters of this model are estimated using Markov chain Monte Carlo algorithm, sampling enough generations for chain convergence can be computationally demanding and time consuming. We are currently in progress of quantifying the contributions of geography and host use (again coded as 0/1 distance) in our dataset using BEDASSLE,

which, when complete, may yield additional insight into geographic and ecological contributions to patterns of neutral divergence. We also plan to repeat BEDASSLE analyses using more precise host differences, including site-specific host needle widths.

3.3.2 – Spatial and Temporal Patterns of Preference

Our analysis of collection data demonstrates a significant geographical bias in host affiliation (Figure 3.7), which is significantly impacted by both latitude ($F_{1,4735}$ = 114.05, $p < 0.0001$) and longitude ($F_{1, 4735}$ = 368.83, $p < 0.0001$). Collection records indicate that *N. lecontei* is more frequently affiliated with *P. resinosa* at more northerly latitudes and more easterly longitudes, and with *P. banksiana* at more southerly latitudes and westerly longitudes. This tendency roughly corresponds with a bias of populations on *P. resinosa* in Ontario (our ONT cluster), and on *P. banksiana* in the U.S. Lake States (Michigan and Wisconsin; our LP MI and UP MI + WI clusters).

Our preference assays suggest the geographic patterns of host affiliation are likely caused by regional host preferences. All populations collected from LP MI and UP MI + WI demonstrated significant preference for *P. banksiana*, regardless of their original collection host; while the population we examined from ONT showed a significant preference for *P. resinosa* (Figure 3.8). Although we only examined a single population from ONT, our observations are consistent with historically noted regional host preferences, well as our own observations during more recent field work (circa 2002-2015). Populations of *N. lecontei* in Ontario are noted to "undoubtedly" prefer *P. resinosa* (Atwood & Peck 1943), while those in Michigan and Wisconsin prefer *P. banksiana* (Benjamin 1955). The preference for *P. banksiana* in the Lake States is so strong that all available trees of this species are sometimes defoliated completely before

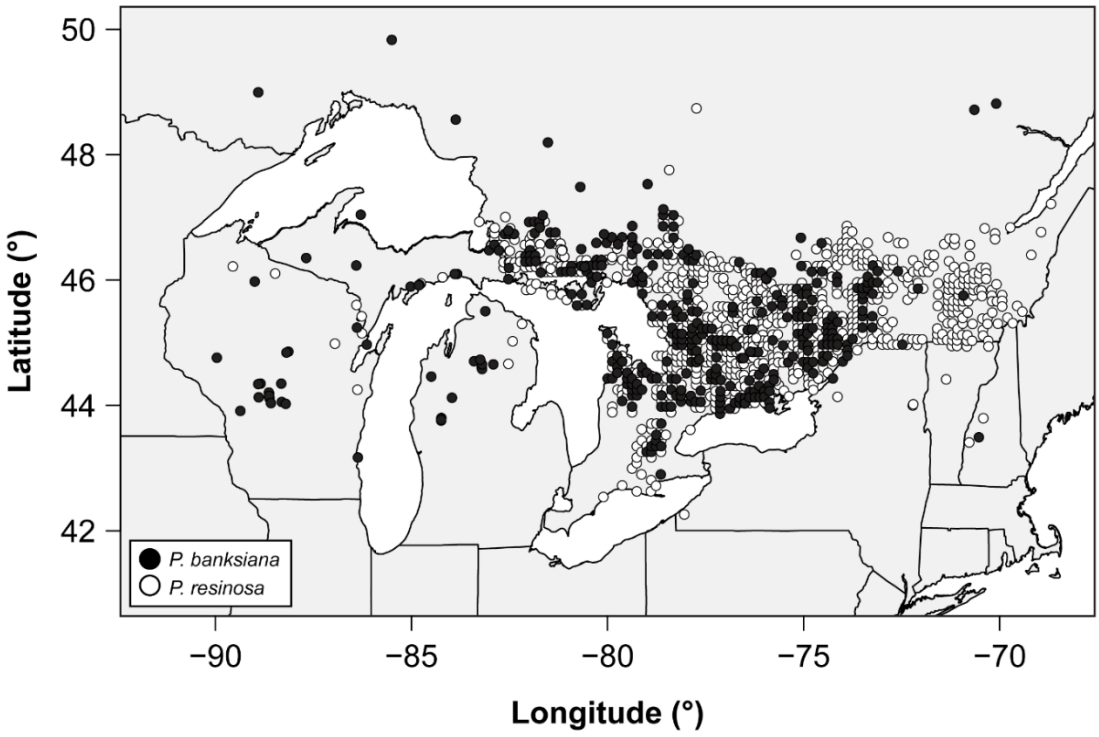


Figure 3.7 – Host affiliation across the region. Each point represents one collection record, with black circles representing records on *P. banksiana* and white circles representing records on *P. resinosa*. For ease of visualization, *P. banksiana* points are displayed on top of *P. resinosa* points.

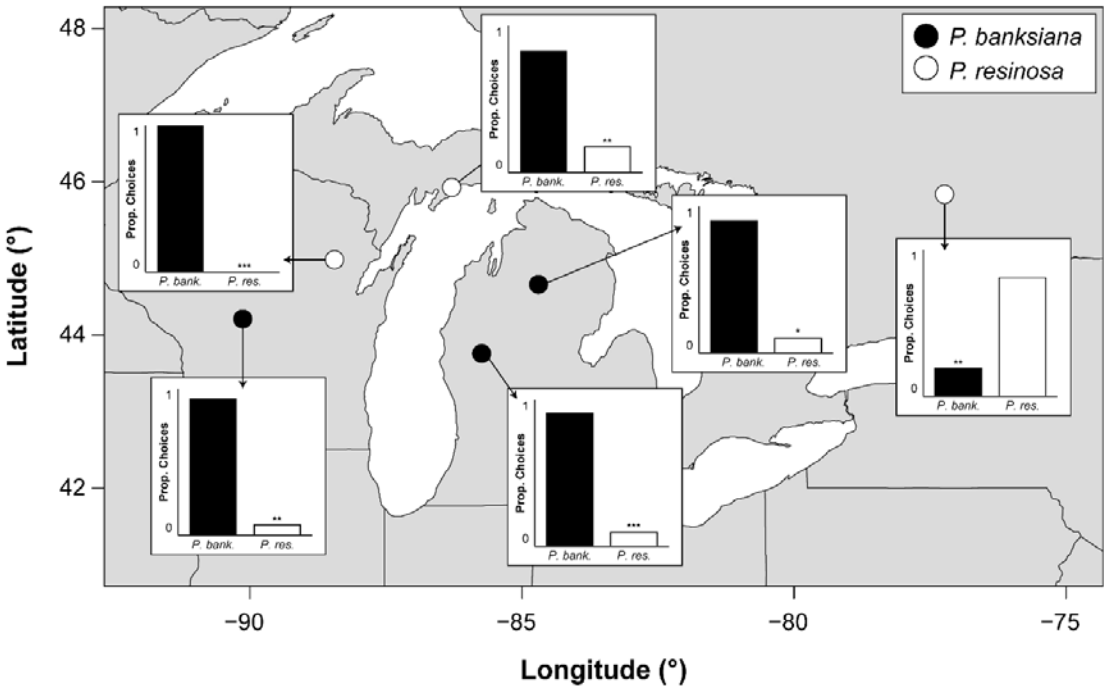


Figure 3.8 – Female oviposition preference across the region. The inset plots display the proportion of choices made on each host, in each population. To facilitate comparison across graphs, all y-axes are presented at the same scale. Differences significant at the 0.05 (*), 0.01 (**), and 0.001 (***) levels are indicated in each plot. Circles are colored to indicate the original host of the population tested, and bars are colored to represent the host selected in choice assays. Black in all cases represents *P. banksiana* and white represents *P. resinosa*.

P. resinosa is utilized, even if only a few trees of *P. banksiana* are available (Benjamin 1955; personal observation).

Although these results confirm general differences in host preference and utilization between regions, they do not explain why this pattern exists. Although our previous work suggested *N. lecontei* colonized this region from a single glacial *Pinus* refugium on the coast of Nova Scotia, the sample size considered was small, and contained outliers that could represent lineages from other refugium (Bagley *et al.* 2017). Both *P. banksiana* (Rudolph & Yeatman 1982; Godbout *et al.* 2005) and *P. resinosa* (Walter & Epperson 2001, 2005) are thought to have occupied several refugium during the last glacial maximum. Refugial populations of *N. lecontei* may have become adapted to their respective hosts in these different refugia, and then accompanied their hosts during post-glacial expansion to their current range. Alternatively, *N. lecontei* in the region may have originated from a single refugium, and the divergent host preferences may have evolved *in situ*, as selection may favor use of different hosts in the different regions. If this were the case, additional research would be needed to determine the underlying reasons for differential host suitability. Future work using demographic modeling will help us to distinguish between these hypotheses.

Regardless of the reasons for general regional preferences, *N. lecontei* is at least occasionally collected on both hosts throughout the region, and in some cases, in sympatry at a single site. At these sympatric locations, we noted that colonies utilizing *P. banksiana* were often at considerably later developmental stages than those utilizing *P. resinosa*. Indeed, our assessments found significant differences in mean eclosion date and shape of eclosion curves at all three locations we examined (Table 3.7, Figure 3.9). The

Table 3.7 – Pairwise temporal isolation at sympatric sites. For each site, we list the pairwise isolation index I (Feder *et al.* 1993), post-hoc Tukey's HSD *P*-values, and bootstrapped Kolmogorov-Smirnov (KS) test *P*-values between populations on *P. banksiana* and *P. resinosa*.

Comparison	**I**	**Tukey's HSD *P*-value**	**Bootstrapped KS test *P*-value**
Thompson Township, MI	0.48	0.008	0.0147
Frederic, MI	0.56	<0.0001	0.0399
Mosinee, WI	0.93	<0.0001	0.0024

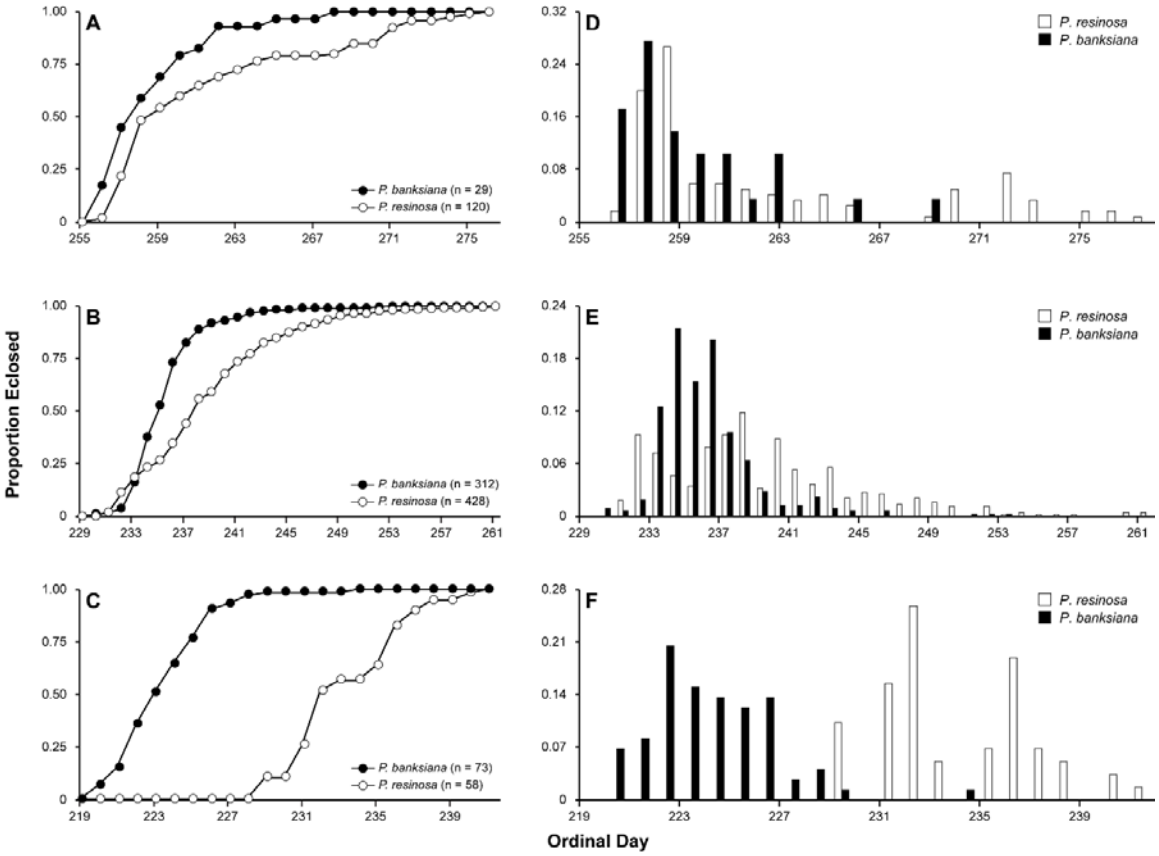


Figure 3.9 – Temporal isolation between *P. banksiana* and *P. resinosa* at sympatric sites. Cumulative eclosion curves for all adults on each host is shown for Thompson Township MI (A), Frederic, MI (B), and Mosinee, WI (C). Daily eclosion at each site is shown in the adjacent panel (D-E). The sample size for each host, per site, is inset in the legends of panels A-C.

degree of isolation between hosts was only moderate at Thompson Township, MI (I = 0.48; $F_{1,148} = 7.22$, $p = 0.008$) and Frederic, MI ($F_{1,739} = 85.74$, $p < 0.0001$), but it was near-complete at Mosinee, WI (I = 0.93; $F_{1,130} = 414.4595$, $p < 0.0001$).

Although our preference assays did not detect significant preference for *P. resinosa* in any U.S. populations, it is possible that the conditions in our lab do not adequately recreate temporal variation in host suitability. Pines are known to vary in moisture content (Van Wagner 1967), volatile emissions (Geron & Arnts 2010), and chemical composition (Nerg *et al.* 1994) throughout the year. As specialist insects may be more capable of discriminating between high and low quality hosts (Janz & Nylin 1997), it is possible that *P. resinosa* is a higher quality host later in the season, leading late emerging adults to utilize *P. resinosa* over *P. banksiana.*

3.3.3 – Spatial patterns in a performance-related trait

Given the strikingly different overall morphologies of *P. banksiana* and *P. resinosa*, it is unsurprising that both needle width ($F_{1,29} = 348.67$, $p < 0.0001$) and length ($F_{1,29} = 645.94$, $p < 0.0001$) vary significantly between hosts (Figure 3.10). As *Neodiprion* sawflies embed their eggs within the needle tissue of their hosts, this variation in needle architecture may serve as a source of divergent selection between hosts. Indeed, oviposition traits were shown to contribute to the evolution of extrinsic post-zygotic reproductive isolation between *N. lecontei* and its sister species *N. pinetum* (Bendall *et al.* 2017).

Region-wide, host did not significantly impact ovipositor shape ($F_{1,37} = 5.08$, $p = 0.102$) or width ($F_{1,52} = 1.34$, $p = 0.252$), but it did impact ovipositor length ($F_{1,52} = 11.73$, $p = 0.001$), with ovipositors from *P. resinosa* females being shorter than those from *P.*

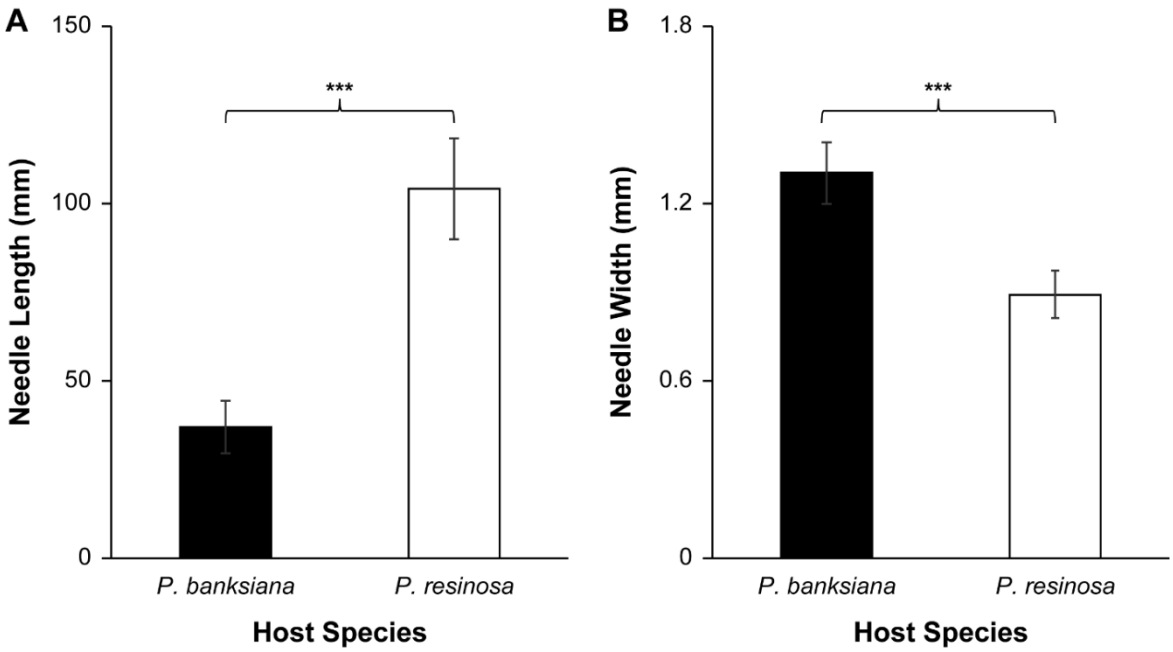


Figure 3.10 – Needle architecture of *P. banksiana* and *P. resinosa*. Both needle length and width vary significantly between hosts.

banksiana females (Figure 3.11). Perhaps unsurprisingly, given the existence of multiple genetic clusters in the region, location also had a strong impact on ovipositor morphology [shape ($F_{10,37} = 2.19$, $p = 0.0003$); length ($F_{10,52} = 16.86$, $p < 0.0001$); width ($F_{10,52} = 27.17$, $p < 0.0001$)]. To address this variation, and assess how it may relate to host plant, we visualized variation in ovipositor length (the primary trait that differed between hosts region-wide) across space, including at sympatric sites. We also directly assessed variation in ovipositor shape, width, and length within each of the genetic clusters, and at three sympatric sites.

Region-wide, there seems to be a geographic pattern in ovipositor length. Although we did not test assess significance of overall length differences between clusters, ovipositors from sites in Ontario and the upper peninsula of Michigan tend to be longer than the region-wide average; while those in Wisconsin and the lower peninsula of Michigan are close to, or shorter than the region-wide averages (Figure 3.12). Considering ovipositors within clusters, host has a significant impact on shape ($F_{1,19} = 2.62$, $p = 0.044$) and length ($F_{1,26} = 24.72$, $p < 0.0001$), but not width ($F_{1,26} = 0.69$, $p = 0.415$) in UP MI + WI (Figure 3.13). Conversely, no differences in ovipositor morphology were observed between hosts in ONT [Figure 3.14; shape ($F_{1,6} = 2.22$, $p = 0.060$); length ($F_{1,8} = 0.83$, $p = 0.389$); width ($F_{1,8} = 0.29$, $p = 0.604$)] or LP MI [Figure 3.15; shape ($F_{1,12} = 0.91$, $p = 0.521$); length ($F_{1,15} = 0.05$, $p = 0.832$); width ($F_{1,15} = 0.73$, $p = 0.407$)]. However, reliability of these results may be impacted by the paucity of samples on the non-preferred host, which are generally from only a single site, in each of these clusters (2/12 from 1 site in ONT; 5/20 from 1 site in LP MI).

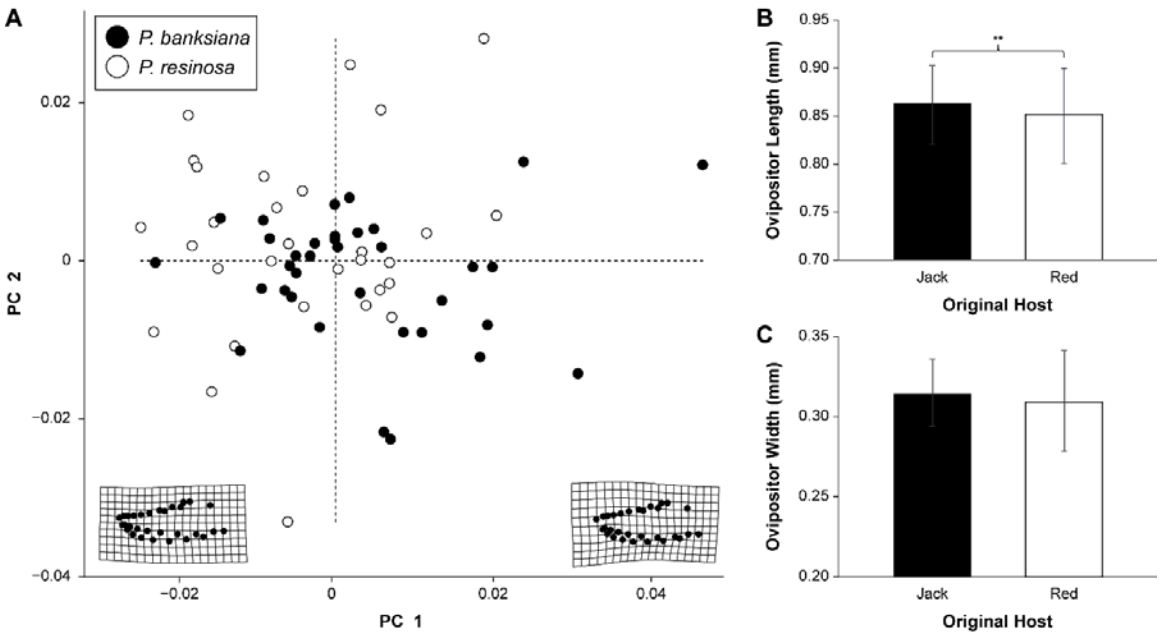


Figure 3.11 – Region-wide variation in ovipositor morphology. A. Principal components analysis of ovipositor shape, with individuals from *P. banksiana* (*n* = 35) shown in black, and those from *P. resinosa* (*n* = 30) shown in white. The inset warp grids show variation in shape along PC1. Ovipositor length (B) differs significantly between hosts, but shape (A) and width (C) do not.

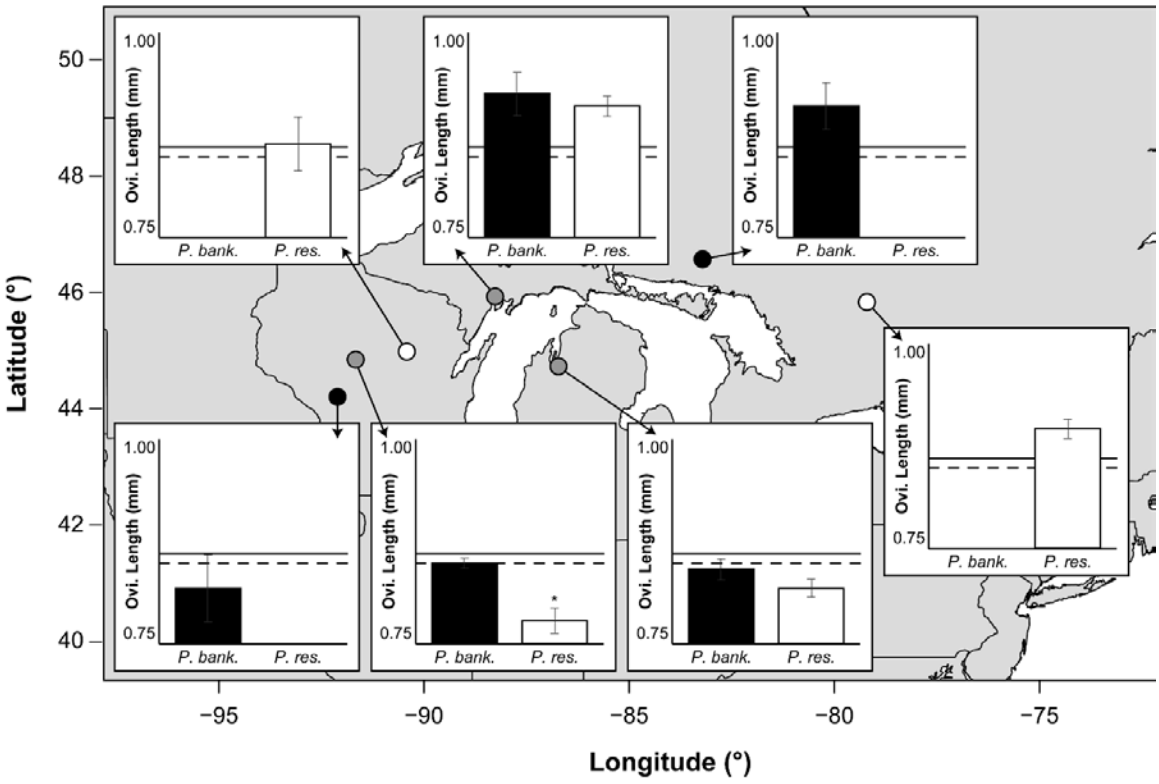


Figure 3.12 – Variation in ovipositor length across the region. Ovipositor length on *P. banksiana* and/or *P. resinosa* is shown for 7 locations throughout the region. To facilitate comparisons, each y-axis is shown on same scale, and the average length of ovipositors on *P. banksiana* (solid line) and *P. resinosa* (dashed line) are shown.

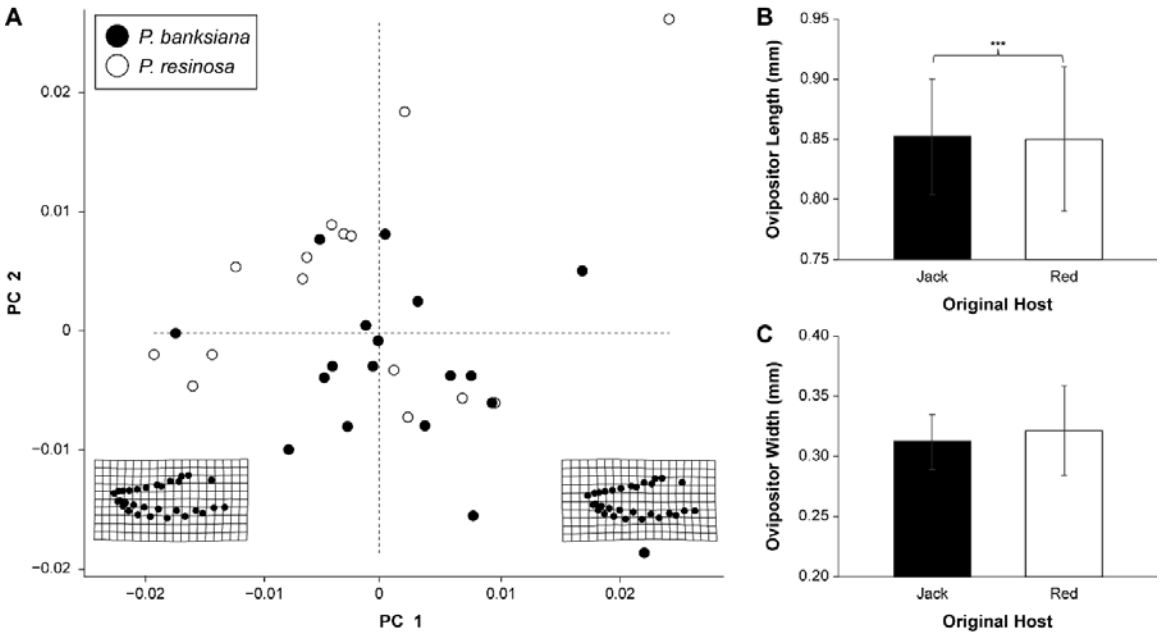


Figure 3.13 – Variation in ovipositor morphology in UP MI + WI. A. Principal components analysis of ovipositor shape, with individuals from *P. banksiana* (*n* = 18) shown in black, and those from *P. resinosa* (*n* = 15) shown in white. The inset warp grids show variation in shape along PC1. Both ovipositor shape (A) and length (B) differ significantly between hosts, but width (C) does not.

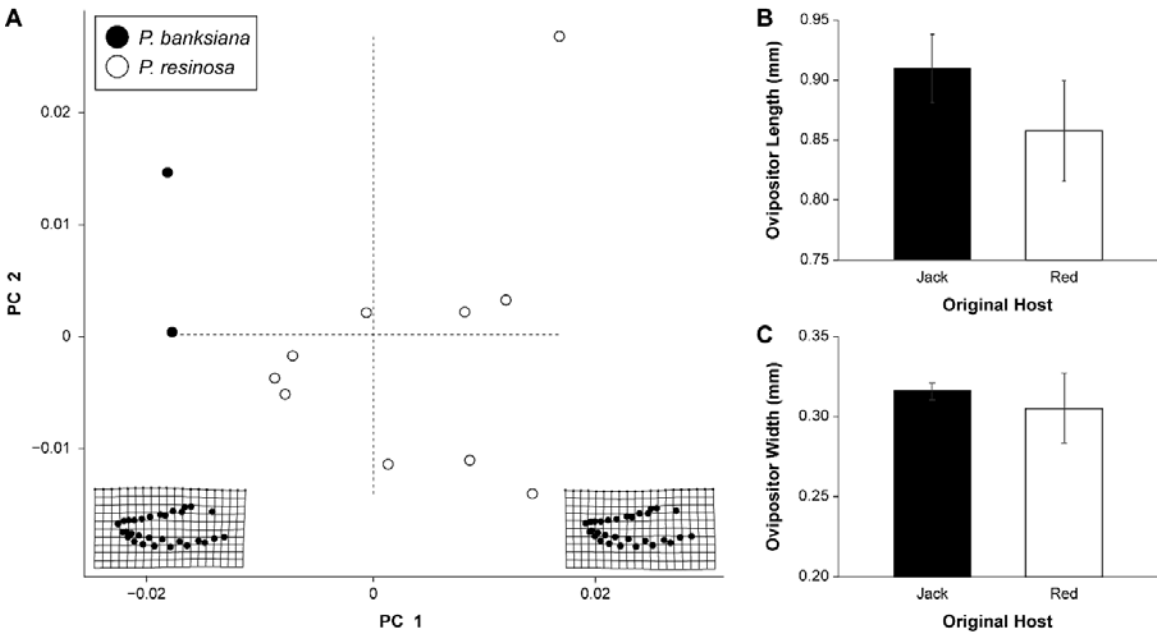


Figure 3.14 – Variation in ovipositor morphology in ONT. A. Principal components analysis of ovipositor shape, with individuals from *P. banksiana* (*n* = 2) shown in black, and those from *P. resinosa* (*n* = 10) shown in white. The inset warp grids show variation in shape along PC1. Ovipositors do not differ between hosts in shape (A), length (B), or width (C).

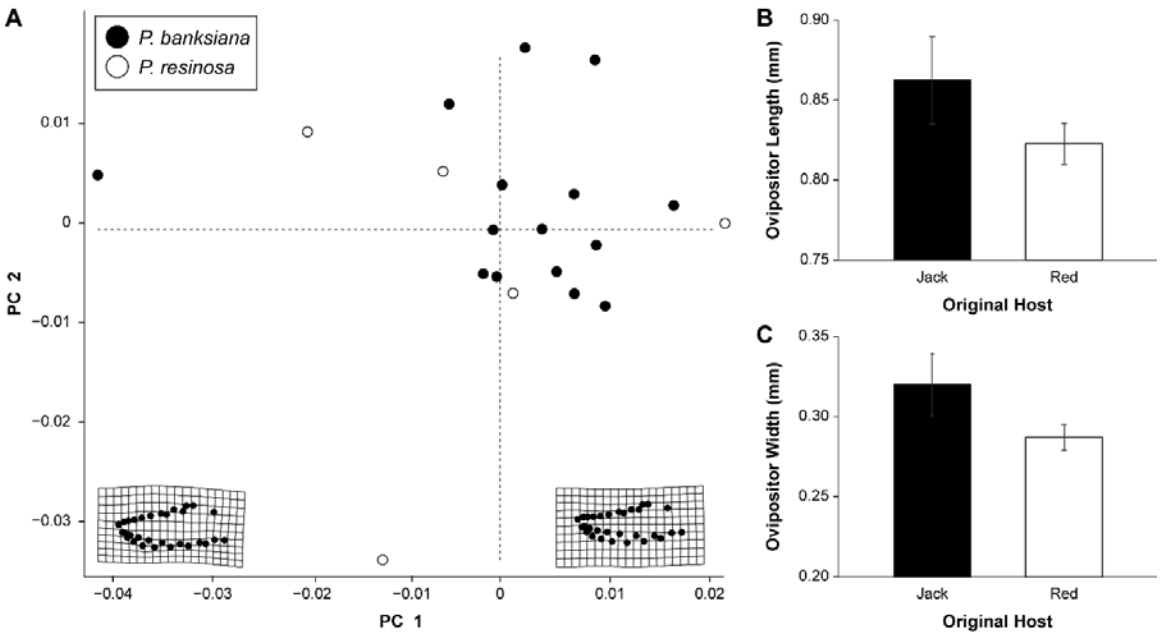


Figure 3.15 – Variation in ovipositor morphology in LP MI. A. Principal components analysis of ovipositor shape, with individuals from *P. banksiana* (n = *5*) shown in black, and those from *P. resinosa* (*n* = 15) shown in white. The inset warp grids show variation in shape along PC1. Ovipositors do not differ between hosts in shape (A), length (B), or width (C) between hosts.

When we examine local scales with our sympatric sites, host does not impact ovipositor morphology at Thompson Township, MI [Figure 3.12, 3.16A, 3.17A; shape ($F_{1,6}$ = 1.47, p = 0.264); length ($F_{1,6}$ = 2.70, p = 0.151); width ($F_{1,6}$ = 0.21, p = 0.663)] or Frederic, MI [Figure 3.12, 3.16B, 3.17B; shape ($F_{1,36}$ = 1.52, p = 0.141); length ($F_{1,37}$ = 2.31, p = 0.137); width ($F_{1,37}$ = 1.52, p = 0.225)]. Host does not impact ovipositor shape (Figure 3.16C; $F_{1,6}$ = 1.36, p = 0.373) or width (Figure 3.17C; $F_{1,6}$ = 0.15, p = 0.711) in Mosinee, WI; but it does impact ovipositor length (Figure 3.12; $F_{1,6}$ = 10.25, p = 0.019). Interestingly, Mosinee, WI is also the site with the strongest temporal isolation between hosts (Table 3.7; Figure 3.9C, F). As gene flow can constrain adaptive divergence, it is possible that the temporal isolation between populations on *P. banksiana* and *P. resinosa* at this site may have facilitated divergence in ovipositor morphology at this site. Alternatively, the increased divergence in ovipositor morphology may have contributed to the development of stronger temporal isolation between the populations (Räsänen & Hendry 2008; Nosil 2012). Determining the exact relationship between ovipositor divergence and temporal isolation will require detailed analysis of additional sympatric locations.

Overall, there seems to be a consistent pattern across the region where *P. resinosa* females have significantly shorter ovipositors than *P. banksiana* females. This is consistent with previous work in *Neodiprion*, as ovipositor length was closely linked to hatching success (and so, with fitness) in a previous study, and shorter ovipositors were favored on a thinly-needled hosts (Bendall *et al.* 2017). Part of the length difference between *P. banksiana* and *P. resinosa* ovipositors may come from a difference in annulus number. Although *N. lecontei* ovipositors typically have 9 annuli, we have observed a

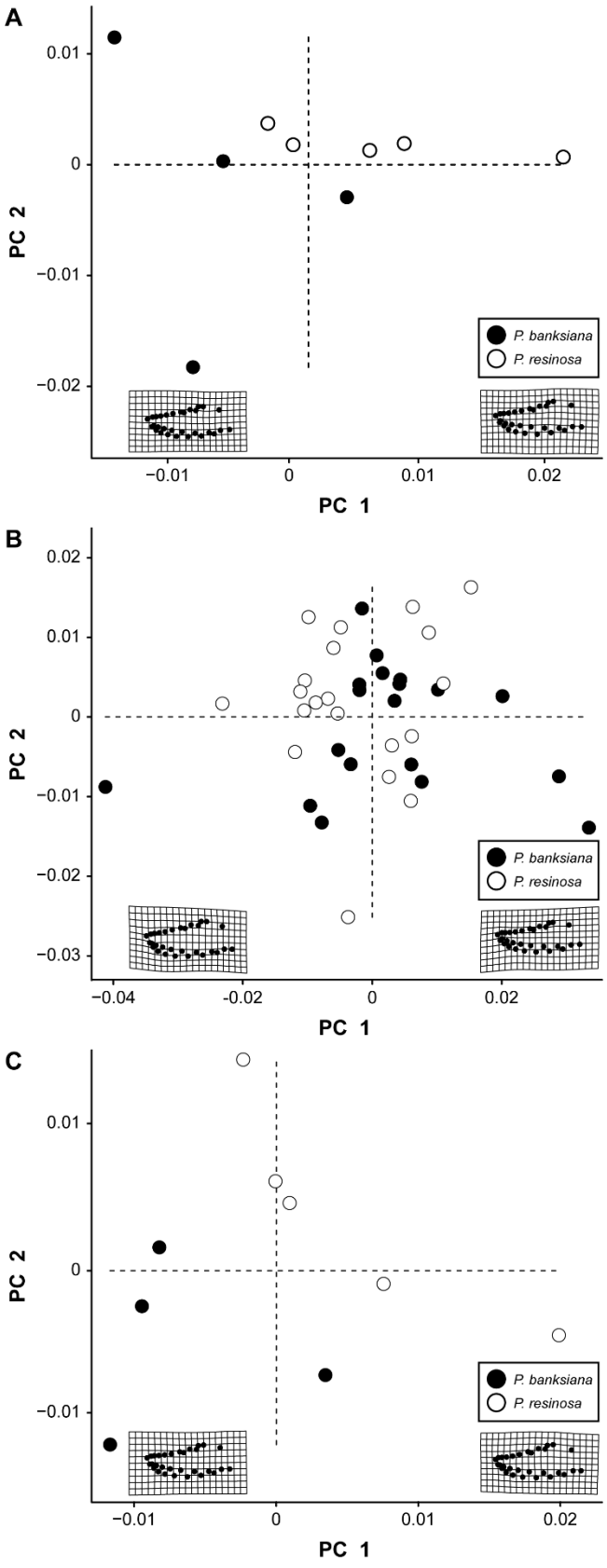

A
0.01
0
−0.01
−0.02
PC 2
P. banksiana
P. resinosa
−0.01
0
0.01
0.02
PC 1
B
0.02
0.01
0
−0.01
−0.02
−0.03
PC 2
P. banksiana
P. resinosa
−0.04
-0.02
0
0.02
PC 1
C
0.01
0
−0.01
PC 2
P. banksiana
P. resinosa
−0.01
0
0.01
0.02
PC 1

Figure 3.16 – Variation in ovipositor shape at sympatric sites. Principal components analysis of ovipositor shape at Thompson Township, MI (A), Frederic, MI (B), and Mosinee, WI (C). In each plot, individuals from *P. banksiana* (*n* = 4, 20, 4) are shown in black, and those from *P. resinosa* (*n* = 5, 15, 5) are shown in white. The inset warp grids show variation in shape along PC1. There are no significant differences in shape at any sites.

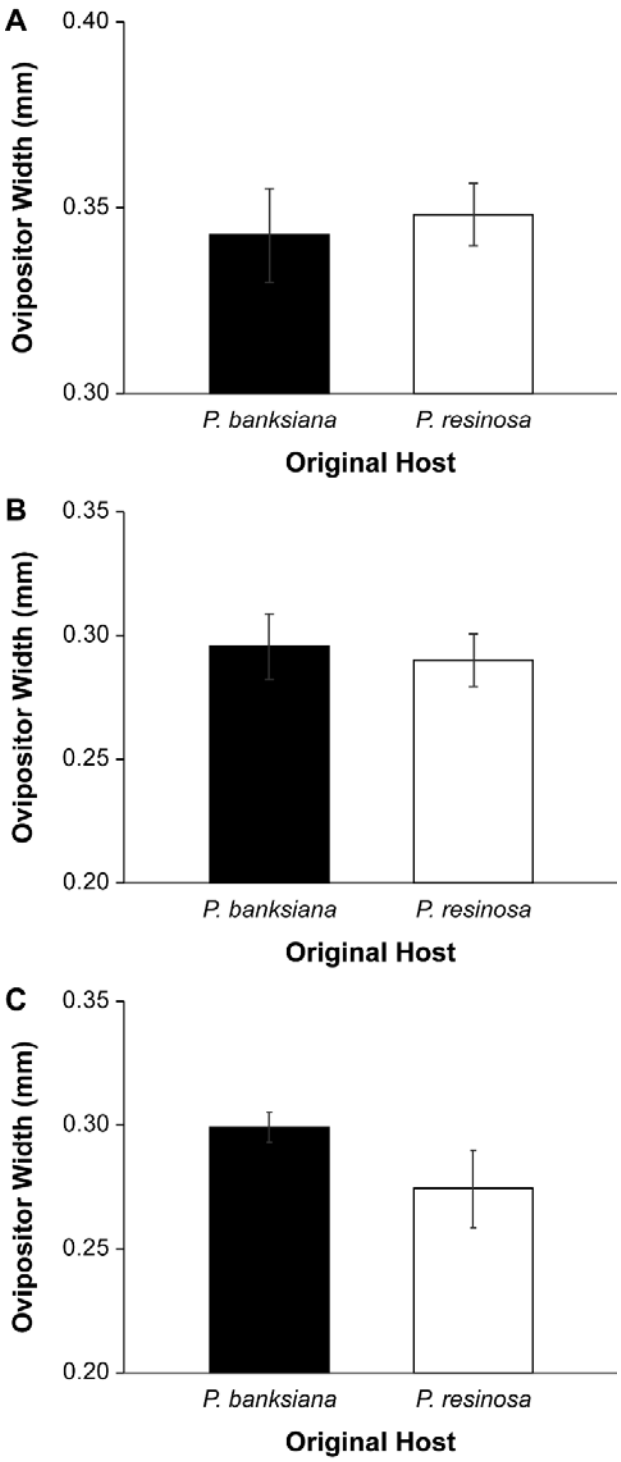


Figure 3.17 – Variation in ovipositor width at sympatric sites. Bar graphs of average ovipositor width at Thompson Township, MI (A), Frederic, MI (B), and Mosinee, WI (C). There are no significant differences in width.

tenth in a small proportion of our specimens, primarily in ovipositors from *P. banksiana* females (Atwood & Peck 1943; unpublished data; Figure A2.3). It is possible, then, that just as a short ovipositor is favored on thinly-needled hosts, that lengthening the ovipositor via acquisition of an additional annulus may be beneficial on thicker-needled hosts.

3.4 – Summary and Conclusions

Overall our data suggest that, although it has not yet left a signature on the neutral genome, host-associated divergence is occurring between populations of *N. lecontei* on *P. banksiana* and *P. resinosa* in North. Although we confirm our previous finding of a lack of IBE in the region, through our increased sampling, we find evidence of considerable substructure, consisting of three genetic clusters that have not yet reached migration-drift equilibrium. Progress towards equilibrium varies between these clusters, potentially due to variation in voltinism across the region. The shape of IBD relationships suggests drift is more influential on patterns of divergence in the region than gene flow at this point, suggesting not enough time has passed for a pattern of IBE to develop.

Despite the lack of neutral divergence between hosts, however, we see strong regional host preferences, with populations in Ontario preferring *P. resinosa* and those in the U.S. preferring *P. banksiana*, as well as host-associated differences in a performance-related trait. Although these differences are likely adaptive, our use of wild-caught individuals in these assays means we cannot rule out the possibility they are plastic responses to host plant (Pfennig *et al.* 2010). We do note, however, that regional host preferences are maintained even after colonies have been reared for multiple generations in the laboratory, suggesting preference is heritable (Figure A2.3). In addition, the

observed association between the degree of temporal isolation and the magnitude of ovipositor differences at local scales would be unusual if the differences were purely plastic. Future studies to confirm the adaptive nature of these differences via reciprocal transplants or multi-generational selection experiments would be beneficial, as would identification of the genetic basis for these and any other adaptive changes.

Along with a handful of other studies (e.g., Nosil 2009; Wang & Summers 2010; Barley *et al.* 2015), our data highlight the importance of considering distribution of both genetic and phenotypic variation when evaluating patterns of divergence. In the absence of phenotypic data, it can be difficult to draw accurate conclusions on biological relevance of genetic structure (Richardson *et al.* 2016), let alone the mechanisms generating the observed patterns of genetic differentiation (Nosil 2012; Shafer & Wolf 2013). Future studies interpreting genotypic variation in the light of relevant phenotypic variation at multiple spatial and temporal scales in this and other species will provide helpful insights into the conditions under which neutral genetic divergence evolves following host shifts, and will further clarify their role in driving population divergence and speciation in *Neodiprion*, and in phytophagous insects as a whole.

3.5 – SUPPORTING INFORMATION

The following supporting information can be found in Appendix 2.

- Table A2.1 – Collection information for individuals used in population genetic analyses.
- Table A2.2 – Individual barcode and Illumina index assignments, and proportion of heterozygous sites.

- Table A2.3 – Sequences for adapters containing variable-length barcodes from Burford Reiskind *et al.* (2016).
- Table A2.4 – Sequences for PCR primers, including degenerate bases.
- Figure A2.1 – Landmark positions for geometric morphometric analysis.
- Figure A2.2 – Variation in annulus number in North *N. lecontei* females.
- Figure A2.3. Cocoon weights from reciprocal transplant analyses.

Chapter 4 : Host-associated divergence in a recently established sympatric population of the red-headed pine sawfly, *Neodiprion lecontei*, on three pine hosts

4.1 – INTRODUCTION

Host shifts, and subsequent adaptation to these novel hosts are a common explanation for divergence and speciation in plant-feeding insects (Matsubayashi *et al.* 2010; Forbes *et al.* 2017). Host shifts are a specific case of ecological speciation, where divergent natural selection between populations leads to the development of reproductive isolation (Schluter 1998, 2001, 2009; Rundle & Nosil 2005; Nosil 2012). In the case of a host shift, the characteristics of the original and novel hosts serve as the source of divergent selection. An initial population of colonists may perform poorly on the new host; initial fitness benefits may be gained through enemy escape or a reduction in interspecific competition (Bernays & Chapman 1994). As long as the colonists survive on the new host, the two populations may phenotypically diverge as selection favors traits that improve fitness on their respective hosts. For example, insects may develop physiological adaptations to improve processing or detoxification of host material (e.g., Rausher 1984; Via 1991; Mackenzie 1996; Cornell & Hawkins 2003); alternative coloring to improve camouflage on the host (e.g., *Timema* walking sticks; Sandoval 1994; Nosil *et al.* 2002; Sandoval & Nosil 2005); or morphology to improve host handling (Moran 1986; Bernays 1991; Soto *et al.* 2008).

Over time, reproductive isolation between populations can evolve as a by-product of divergent selection between the original and novel host plant. For example, when performance differences exist between populations, maladapted immigrants may experience reduced survival on alternative host plants (Funk 1998; Via *et al.* 2000; Nosil

et al. 2005). Similarly, hybrids may exhibit an intermediate phenotype, and thus have poor performance in both parental habitats (Rundle & Whitlock 2001; Rundle & Nosil 2005; Egan & Funk 2009; Kuwajima *et al.* 2010). In both cases, the reduced survival and/or fitness reduces opportunities for gene flow between parental populations. Divergent habitat preferences can result in premating isolation via habitat isolation if adults mate on the host plant (Feder *et al.* 1994; Via 1999; Via *et al.* 2000; Linn *et al.* 2004). Sexual isolation can arise if individuals exhibit preference for same-host *vs.* different host mates regardless of mating environment (Craig *et al.* 1993, 1997; Funk 1998; Nosil *et al.* 2002). Temporal isolation can develop if host phenology differs and insects evolve to match the timing of their host (Feder *et al.* 1993; Groman & Pellmyr 2000; Filchak *et al.* 2000).

As in other cases of ecological speciation, there must be a genetic link between the selection and resulting reproductive isolation (Rundle & Nosil 2005; Nosil 2012). This link is automatic if the trait under selection also pleiotropically impacts reproduction. For example, the evolution of habitat preferences can also confer reproductive isolation when individuals mate in their preferred habitat (Bush 1969a; Drès & Mallet 2002b; Matsubayashi *et al.* 2010; Nosil 2012). Selection on fitness in native environment can also reduce fitness of immigrants and hybrids in non-native environments if genetic trade-offs in performance exist, reducing gene flow between populations (Via 1991; Via & Hawthorne 2002; Nosil 2004). Alternatively, genes conferring reproductive isolation may be in linkage disequilibrium with those under selection. Such associations can be protected from recombination via tight physical linkage with directly selected loci (e.g., Hawthorne & Via 2001), localization with

chromosomal inversions (Noor *et al.* 2001; Feder *et al.* 2003b), or through a collective reduction in genome-wide levels of gene flow via strong selection at multiple loci (Feder *et al.* 2012b; Via 2012; Via *et al.* 2012).

Although initial divergence may be limited to a handful of directly selected loci, host shifts can also impact the level of neutral divergence between populations. Divergent selection is expected to reduce overall levels of gene flow between the diverging populations (Bush 1969b; Drès & Mallet 2002b; Matsubayashi *et al.* 2010; Nosil 2012). The reduction in gene flow allows genome-wide neutral divergence between ecologically-divergent populations (in this case, populations on different hosts) to accumulate via drift. This process, known as isolation-by-adaptation (Nosil *et al.* 2008; Funk *et al.* 2011), produces a pattern analogous to isolation-by-distance (Wright 1943), with “ecologically distant” population pairs having greater levels of neutral divergence than ecologically similar populations, regardless of the geographic scale investigated ("isolation-by-environment"; Wang & Summers 2010; Bradburd *et al.* 2013; Sexton *et al.* 2014; Wang & Bradburd 2014).

Empirical evidence of the phenotypic, reproductive and genetic divergence produced by host shifts have been found in a number of systems [e.g., *Timema* walking-sticks (Nosil *et al.* 2003; Nosil 2007), pea aphid (Via 1999; Hawthorne & Via 2001), *Rhagoletis* (Feder *et al.* 1994, 2003b; Linn *et al.* 2003; Feder & Forbes 2007), *Neochlamisus bebbianae* leaf beetles (Funk 1998; Egan & Funk 2009), etc.]. However, most of these studies examine populations at a single timepoint along the speciation continuum, and after multiple barriers to gene flow have accrued. This makes it difficult

to determine which, if any, traits are the first to diverge and which barriers are first to arise after a host shift.

One system in which the role of host shifts has been investigated at multiple levels of divergence is *Neodiprion* sawflies. Linnen and Farrell (2010) demonstrated that host shifts are correlated with speciation events in the genus, but could not determine if the shifts occurred before or after speciation had initiated. Bendall *et al.* (2017) showed host-related differences in oviposition traits contribute to extrinsic post-zygotic reproductive isolation between the sister species *N. lecontei* and *N. pinetum*. Furthermore, there is evidence of host-associated divergence at neutral loci (Bagley *et al.* 2017) and in preference and performance-related traits (Chapter 3 of this dissertation) within *N. lecontei*. While all of these studies suggest host use contributes to population divergence and speciation within *Neodiprion*, and identify traits that contributing divergence; they examine taxa that are relatively deeply diverged, preventing us from determining the earliest changes following the host shift.

If host shifts frequently drive speciation in *Neodiprion*, we should see evidence of host-driven divergence within species found on a wide range of hosts and, specifically, between populations utilizing different hosts. We have identified a field site at the University of Kentucky's Arboretum and State Botanical Garden which harbors colonies of *N. lecontei* on three host plants native to the broader region: *Pinus echinata* (shortleaf pine), *P. virginiana* (Virginia pine), and *P. rigida* (pitch pine). Spanning an area of ~130m, "The Trail of Pines" has multiple trees of each species planted in close proximity to each other, with the branches of some hosts overlapping. Although the exact date of sawfly colonization at the site is unknown, the majority of sawfly activity observed

occurred on the site's mature trees, which were planted in the mid-to-late 1990s (T. Rounsaville, personal communication).

In this chapter, we leverage this recently colonized field site to gather evidence of host-associated divergence in ecological, reproductive, and genetic traits between sympatric populations and identify which of these traits first arise following a host shift. First, we assess genetic structure at the site using population clustering methods (Alexander *et al.* 2009; Jombart *et al.* 2010) and analyses of molecular variance (AMOVA; Excoffier & Lischer 2010). Next, we look for evidence of ecologically-based temporal, sexual, and habitat isolation between hosts. Finally, we quantify host differences at the site, and assess if sawflies display morphological or physiological adaptations to their hosts. Together, our results suggest sawflies at this site are at the earliest stages of divergence, and that, although divergent host use can generate some physiological and morphological differences between populations, additional isolating mechanisms may be required for speciation to progress.

4.2 – Materials and Methods

4.2.1 – Sample Collection

Sawfly colonies were collected from *P. echinata*, *P. rigida*, and *P. virginiana* at the field site between 2012 and 2015 as early-to-late instar feeding larvae (Table A3.1). A subset of larvae from some colonies were preserved in 100% ethanol for use in genetic assays. The remaining larvae were returned to the lab and reared in plastic boxes (32.4 cm × 17.8 cm × 15.2 cm) with mesh lids, and provided clippings of their natal host species *ab libitum*. Cocoons were collected three times weekly, and stored in individual gelatin capsules until emergence. Larvae and cocoons were kept in walk-in

environmental chambers maintained at 22°C, and an 18:6 light-dark cycle. Cocoons were checked daily for emergence, and live adults were stored at 4°C until use.

For sexual isolation, habitat isolation, and larval performance assays, we propagated multiple families collected from each host species (hereafter, host types "Shortleaf", "Pitch", and "Virginia") for an additional one to two generations in the lab. Briefly, each family was produced by releasing male and female adults from several colonies collected on one of the three host plants into mesh cages containing multiple seedlings of *P. banksiana*. The adults were allowed to mate and oviposit freely. Upon hatching, larvae from these cages were transferred into plastic boxes and reared as described above on clippings of field-collected *P. banksiana*. We chose *P. banksiana* as the oviposition and rearing substrate for propagation as it is an adequate host for most *Neodiprion* (Knerer 1984), to control for the potential impact of maternal effects (Mousseau & Dingle 1991; Mousseau & Fox 1998), and because seedlings of this host are available year-round for purchase.

4.2.2 – DNA Extraction and Library Preparation

DNA was extracted from preserved individuals using a CTAB/Phenol-Chloroform-Isoamyl alcohol method based on Chen *et al.* (2010). Each extraction was visualized on a 0.8% agarose gel to ensure no samples were degraded. The concentration of each intact sample was estimated using a Quant-iT High-Sensitivity DNA Assay Kit (Invitrogen – Molecular Probes, Eugene, OR, USA).

We used a ddRAD sequencing strategy to generate a large dataset of putatively neutral SNP markers (Peterson *et al.* 2012). Following Bagley *et al.* (2017), DNA was fragmented using the enzyme pair NlaIII and EcoRI (NEB, Ipswich, MA). A total of 58

individuals were assigned based on DNA yield into one of 8 groups of up to 48 individuals, and randomized with respect to location. Each sample was assigned one of 48 unique variable-length in-line barcodes during adapter ligation (Table A3.2; A2.3; Burford Reiskind *et al.* 2016).

Each set of samples was then pooled for automatic size selection of a 379-bp fragment (+/- 76bp) on a PippinPrep (Sage Science, Beverly, MA), and amplified over 12 rounds of high-fidelity PCR amplification (Phusion High-Fidelity DNA Polymerase, NEB, Ipswich, MA) using PCR primers containing unique Illumina multiplex read indices and a string of degenerate bases for PCR duplicate detection (Table A3.2; A2.4).

Successful library creation was verified using a Bioanalyzer 2100 (Agilent, Santa Clara, CA), and libraries were sent to the High-Throughput Sequencing and Genotyping Unit at the University of Illinois. Two lanes of 150bp single-end reads from an Illumina HiSeq 4000 were obtained for the libraries.

4.2.3 – Data Processing and SNP Genotyping

Raw sequence reads were quality filtered and trimmed using the *process_radtags* module in STACKS (v1.46; Catchen *et al.* 2013). Surviving reads were then aligned to a high-coverage genome assembly for *N. lecontei* (Vertacnik *et al.* 2016; coverage: 112x; scaffold N50: 244kb; GenBank assembly accession: GCA_001263575.1) using the "very sensitive" end-to-end alignment mode in BOWTIE2 (v2.3.1; Langmead & Salzberg 2012). We then used SAMTOOLS (v1.3; Li *et al.* 2009) to retain only uniquely-mapping reads with MAPQ scores ≥30. Putative PCR duplicates were identified via the sequence of the 4 degenerate bases in the index read (provided as a second fastq file) and removed using a custom python script. We then constructed RAD loci from the filtered alignments in

STACKS' *ref_map.pl* pipeline (v1.46; Catchen *et al.* 2013) To ensure high-confidence genotype calls (Kenny *et al.* 2011; Peterson *et al.* 2012), we kept only those loci with at least 10x depth of coverage per individual.

After an initial round of SNP calling, we excluded seven individuals missing data at >90% of SNP loci (Table A3.3). As in Bagley *et al.* (2017), we inferred the ploidy of the remaining individuals using heterozygosity estimates from vcftools' *--het* option (v0.1.14b; Danecek *et al.* 2011) and excluded two putatively haploid individuals with considerably lower proportion of heterozygous compared to other individuals (Table A3.3).

Our final dataset consisted of 49 individuals (15 Shortleaf, 13 Pitch, and 21 Virginia). We applied several additional filters to these individuals, excluding all sites missing data in 30% or more of individuals and all sites violating Hardy-Weinberg equilibrium for heterozygote excess significant at the 0.01 level. Finally, to minimize linkage disequilibrium between SNPs, we included only one random SNP per RAD locus.

Data processing and all other analyses were performed on the University of Kentucky's Lipscomb High Performance Computing Cluster.

4.2.4 – Detection of Population Structure

We used two individual-based approaches to investigate population structure between host types at the Arboretum. First, we used the maximum-likelihood-based clustering algorithm implemented in the program ADMIXTURE (v1.3.0; Alexander *et al.* 2009) to determine the proportion of ancestry for each individual from *K* ancestral populations without *a priori* designation. We performed 100 independent runs for values

of *K* from 1 through 10. The optimal *K* was selected as described in the ADMIXTURE manual, by comparing the 5-fold cross-validation (CV) error across different values of *K*. To determine assignment stability, and visualize primary and secondary solutions across the 100 replicates of each *K*, we used the main pipeline of CLUMPAK (v1.1; Kopelman *et al.* 2015).

Second, we used a discriminant analysis of principal components (DAPC), a non-model based approach that transforms genotypes using a principal components analysis to maximize the differences between groups while minimizing differences within groups (Jombart *et al.* 2010). We utilized the *dapc* function in the program ADEGENET (v2.0.4; Jombart 2008) of the R statistical framework (v3.4.0; R Core Team 2013). As DAPC requires *a priori* group assignment, we used the *K*-means clustering algorithm in ADEGENET to find optimal number of clusters for *K* 1 through 10; and compared these clustering solutions using Bayesian Information Criterion (BIC), following (Jombart *et al.* (2010). We then used α-score optimization to evaluate the optimal number of principle components (PCs) to retain in the analysis; and visualized the results of the DAPC using a custom R script.

4.2.5 – Analysis of Molecular Variance (AMOVA)

We looked for host-based differentiation using a locus-by-locus AMOVA implemented in ARLEQUIN (v3.5.2.2; Excoffier & Lischer 2010). We grouped individuals by their natal host plant (*P. echinata*, *P. rigida*, or *P. virginiana*). We also assessed the level of differentiation between host by calculating pairwise F_{ST} estimates. Statistical significance of AMOVA and F_{ST} estimates were assessed using 10,000 permutations each.

4.2.6 – Temporal Isolation

To assess if host types are temporally isolated, we tracked the eclosion dates of all individuals returned to the lab in 2012, 2013, and 2014 (Table A3.1). Although *N. lecontei* typically achieves 2-3 generations per year in Kentucky, sawfly abundance varies greatly across generations. Therefore, we focused our analyses on the generation for which we had sampling data for all three hosts available, and tracked eclosion over a 50-day span. For each year, we pooled total adult emergence for all colonies collected from each host species, and calculated pairwise estimates of isolation (I) between populations following Feder *et al.* 1993:

$$1 - \left(\frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \cdot \sum y_i^2}} \right) \cdot 100$$

where x_i and y_i represent the proportion of the total number of live adults from host *x* or *y* on day *i*. We assumed an average lifespan of 5 days for females and 4 days for males based on field estimates (Benjamin 1955).

To assess the significance of differences in mean eclosion date, we calculated the mean ordinal date of eclosion and variance for each host per year. We then performed a one-way ANOVA from the summarized eclosion data per host each year, followed by Tukey's Honest Significant Difference tests. We also assessed if cumulative eclosion curves differed between host types using bootstrapped Kolmogorov-Smirnov (KS) tests with the *ks.boot* function from the *R* module MATCHING (v4.9-2; Sekhon 2008).

4.2.7 – Sexual Isolation

To assess if any sexual isolation exists in the absence of host plant, we performed no-choice mating assays between all pairwise combination of host types (Shortleaf x

Pitch, Shortleaf x Virginia, and Virginia x Shortleaf). For each assay, a single virgin female was placed in a new, plastic 60mm x 12 mm petri dish, and offered a virgin male from either the same or a different line. As *N. lecontei* is known to exhibit inbreeding avoidance (Harper *et al.* 2016), we minimized the likelihood of matched matings by using adults from different propagation cages for same-line assays. 3 same-line and 3 different-line assays were recorded at any given time, and the position of same-line and different-line pairings were switched between assays to minimize positional biases. A total of 30 assays were performed in all directions for each pairwise cross: A♀ x A♂, A♀ x B♂, B♀ x A♂, and B♀ x B♂. All assays were performed under the rearing conditions described above (Table A3.1).

We recorded each set of assays for 75 minutes with a Logitech or Microsoft web camera connected to a Lenovo Ideapad laptop. The footage was then viewed using VLC player, and we recorded if mating occurred or not. We then calculated the Index of Pair Sexual Isolation (I_{PSI}) and assessed deviation from random mating in JMATING (v1.8.0; Carvajal-Rodriguez & Rolan-Alvarez 2006).

4.2.8 – Habitat Preference

To determine if females display preference for their original host plant over the alternative hosts, we performed a series of choice experiments. We placed each female in a mesh cage (35.6cm x 35.6cm x 61cm) with two seedlings of their original host plant and two seedlings of one of the two alternative host species. A total of 30-37 assays were performed per host combination (Table A3.1). The cages were checked daily until eggs were laid or the female died. For each female, we recorded if eggs were laid, and, if so, which host was selected for oviposition. We excluded females who did not make a choice

(*n* = 5-16) or laid on both hosts (*n* = 0-5). To determine if females exhibited preference for their original or the alternative host, we performed exact binomial tests in R.

4.2.9 – Host Characteristics

To assess differences in needle architecture, we measured the width of 10 needles from each mature pine at the Trail of Pines where sawfly activity was observed during the study period. Needle width was measured using digital calipers (Mitutoyo CD-6”PMX). To analyze differences in needle width between hosts, we performed two-way ANOVAs, followed by a Tukey’s Honest Significant Difference test.

4.2.10 – Ovipositor Morphology

After emergence, a subset of females were frozen at -80°C for use in ovipositor morphology analyses (Table A3.1). A total of 55 ovipositors (*n* = 17-19 females per host; 4-9 families per host; 1-7 females per family) were analyzed. Ovipositors were dissected, mounted, and imaged as described in Bendall *et al.* (2017).

Briefly, a single lancet from each female was mounted on a glass microscope slide in an 80:20 permount:toluene solution. Each slide was imaged at 5x magnification and the ovipositor length and width were measured using the ZEN lite 2012 software package (Carl Zeiss Microscopy, LLC; Thornwood, NY). We compared ovipositors from each host using a geometric morphometric analysis, which computes shape differences while controlling for ovipositor size. We used IMAGEJ (v1.51; Schneider *et al.* 2012) to place a total of 30 landmarks defining the overall shape of each ovipositor (Figure A2.1), and transformed the position of each landmark into Cartesian coordinates. We aligned the landmarks of each ovipositor using a general Procrustes alignment in GEOMORPH (v2.1.4; Adams & Otárola-Castillo 2013). Shape differences were visualized via a principle

components analysis; and assessed for significance using Procrustes ANOVA with forewing length, host, and family as fixed factors. We also assessed differences in ovipositor length and width between hosts using ANOVAs, again including forewing length, host and family as fixed factors.

4.2.11 – Larval Performance

To assess if there are differences in larval survival and performance across hosts, we tracked survival and female cocoon weights for families of each host type reared on their original *versus* alternative hosts. Mated females propagated from each host type were mated to a same-type male and offered a seedling of one of the three host plants for oviposition. The number of eggs laid by each female was recorded. Upon hatching, the larvae were reared as described above on clippings of the host they were laid on. The cocoons were collected as they were spun, counted, and sexed by weight. Because it is difficult to determine if cocoons that fail to produce adults are dead or diapausing, we assessed survival using egg-to-cocoon number. For survival analyses, we excluded families that failed to hatch from analyses. We assessed survival of each host type when reared on each host to cocoon using two-way ANOVAs with original and rearing host as fixed effects, followed by post-hoc Z-tests. Survival proportions were arcsine transformed prior to analyses.

For cocoon weight analyses, we discarded families with cocoon weights that could not be confidently sexed or had more than a 10% discrepancy in the number of weights obtained *vs.* the number of cocoons recorded. Female cocoon weights were compared using two-way ANOVAs with original and rearing host as fixed effect, and

family nested within original host, followed by post-hoc Tukey's Honest Significant Difference tests.

4.3 – RESULTS

4.3.1 – Sequencing and SNP calling

We obtained 1.89 ± 2.35 (SD; standard deviation) million single-ended reads per individual; of which 1.88 ± 2.34 million survived quality filtering. After alignment, paralog filtering, and removal of putative PCR duplicates, an average of 0.95 ± 0.91 million alignments survived, and were formed into an average of 15,789 ± 7,271 RAD loci per individual with an average coverage of 45.67 ± 25.15x. These loci contained a total of 33,674 SNPs. After removing the seven individuals with high levels of missing data, two putatively haploid individuals, and enforcing a <30% missing data filter, the number of SNPs was reduced to 17,165. After applying the Hardy-Weinberg filter and subsampling to a single SNP per locus, our final dataset consisted of 8,787 SNPs.

4.3.2 – Detection of Population Structure

ADMIXTURE's cross-validation procedure selected $K = 1$ as the optimal number of clusters across all 100 independent runs, with CV error steadily increasing with K (Figure 4.1A). Higher values of K were also unstable, frequently offering multiple clustering solutions with similar frequencies. The DAPC method offered similar results, favoring $K = 1$, with BIC scores increasing at larger K (Figure 4.1B). The maximum a-score was obtained using 6 PCs (Figure 4.2), which contribute ~30% of variation (Figure 4.3).

Although $K = 1$ was favored, we also examined the clustering solutions identified under $K = 2$ and $K = 3$ to assess patterns of ancestry within and between hosts. No meaningful structure is seen under $K = 2$. In the two admixture solutions, most

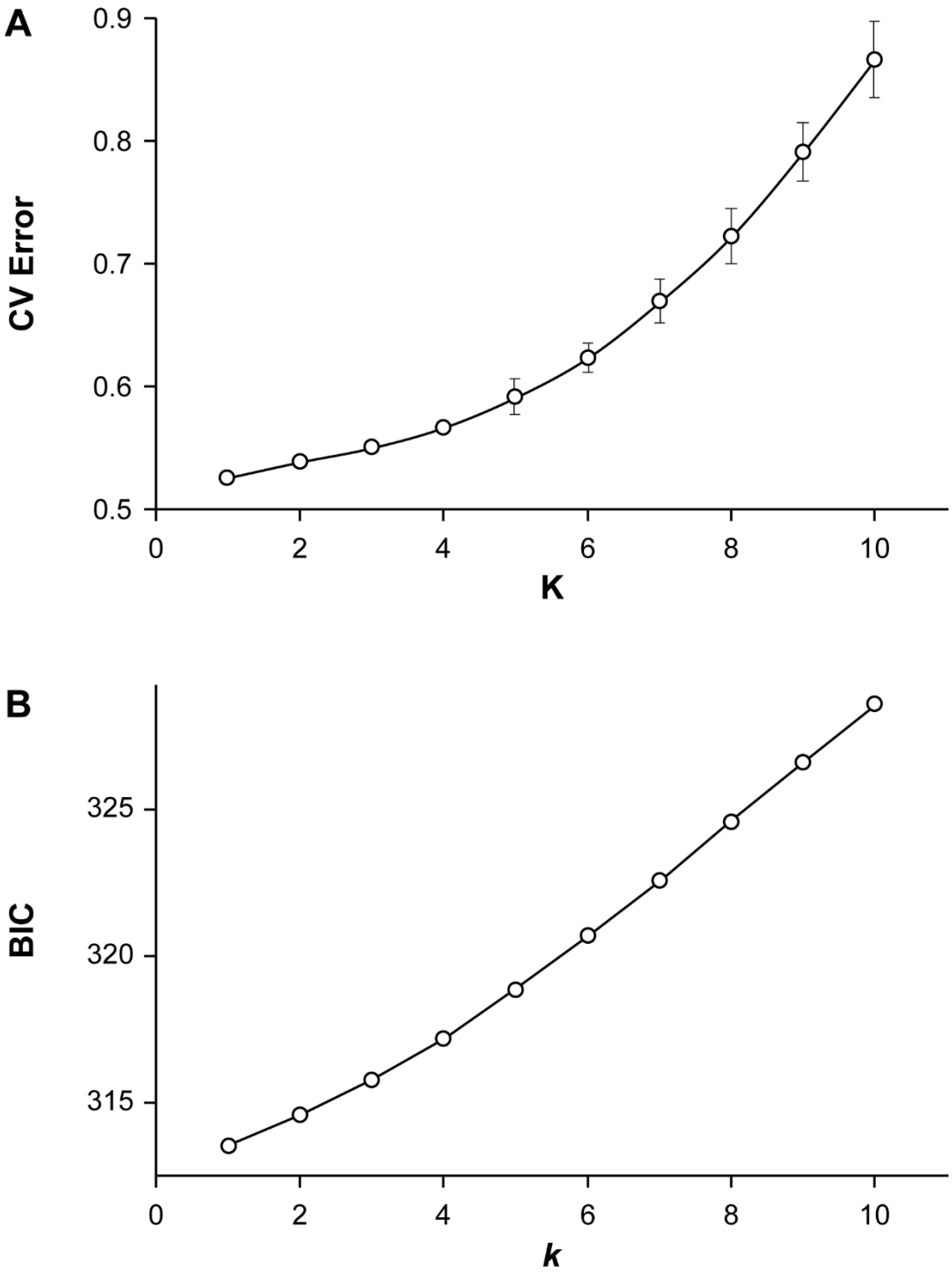


Figure 4.1 – CV error and BIC scores for K 1 through 10.CV error is summarized across 100 independent runs. *K* = 1 is favored by both ADMIXTURE (A) and DAPC (B) clustering methods.

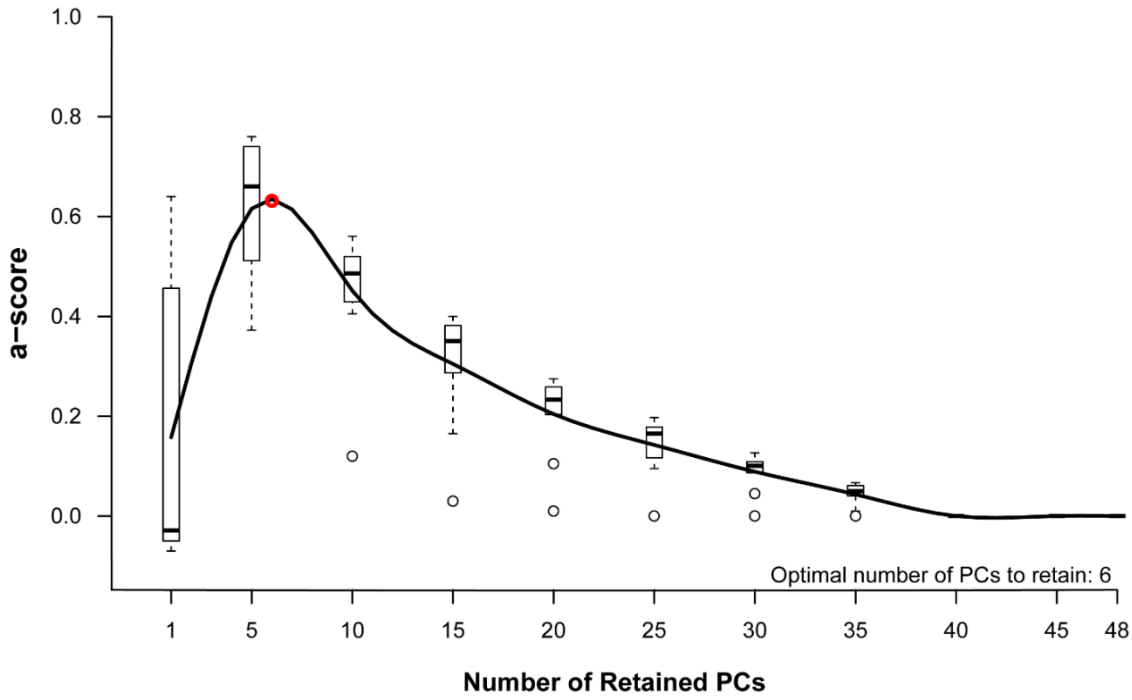


Figure 4.2 – Plot of a-score over 48 PCs, with spline interpolation. The optimal number of PCs to retain is 6.

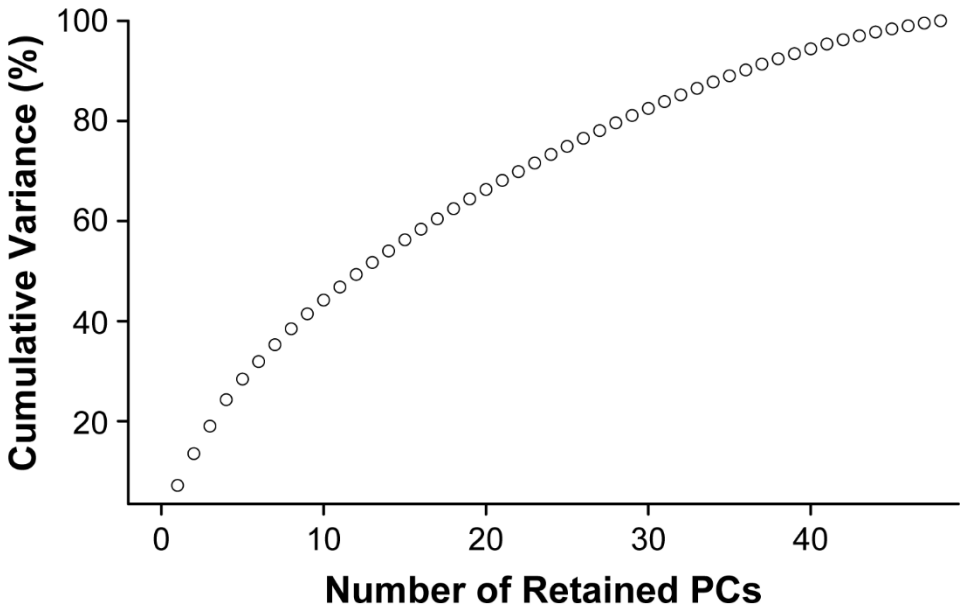


Figure 4.3 – Cumulative percent variance explained as a function of the number of retained principal components. The optimal number of PCs (6; Figure 4.2) correspond to ~30% of total variation.

individuals are admixed to some degree, with no clear differences in the level of admixture between host types (Figure 4.4A, B), while in the DAPC solution, most individuals are assigned to the same cluster (Figure 4.4C). Under $K = 3$, although few individuals assign with 100% confidence to a single cluster, individuals collected on *P. echinata* displays a different pattern of ancestry than *P. rigida* and *P. virginiana* individuals in both the ADMIXTURE and DAPC solutions (Figure 4.5).

4.3.3 – Analysis of Molecular Variance (AMOVA)

The majority of variation comes from within individuals, suggesting little differentiation exists between hosts (Table 4.1). Accordingly, F_{ST} estimates are generally low (*P. echinata vs. P. rigida*: 0.01174, $p = 0.0341$; *P. echinata vs. P. virginiana*: 0.01547, $p = 0.0066$; *P. virginiana vs. P. rigida*: 0.00258, $p = 0.2784$).

4.3.4 – Temporal Isolation

Patterns of eclosion varied between host plants and between years (Figure 4.6; Table 4.2). In 2012, all there was a significant effect of host on mean eclosion date ($F_{2,157} = 67.2959$, $p < 0.0001$), and all host types had significantly different eclosion dates and patterns of eclosion. In 2013, there was a significant effect of host on mean eclosion date ($F_{2,1167} = 237.1679$, $p < 0.0001$), and all host types differed significantly in their eclosion pattern, but Pitch and Shortleaf did not differ in their mean ordinal eclosion date. In 2014, host had a significant effect on eclosion date ($F_{2,958} = 81.0924$; $p < 0.0001$), but Shortleaf and Virginia did not differ in mean eclosion date, and no populations differed in eclosion pattern.

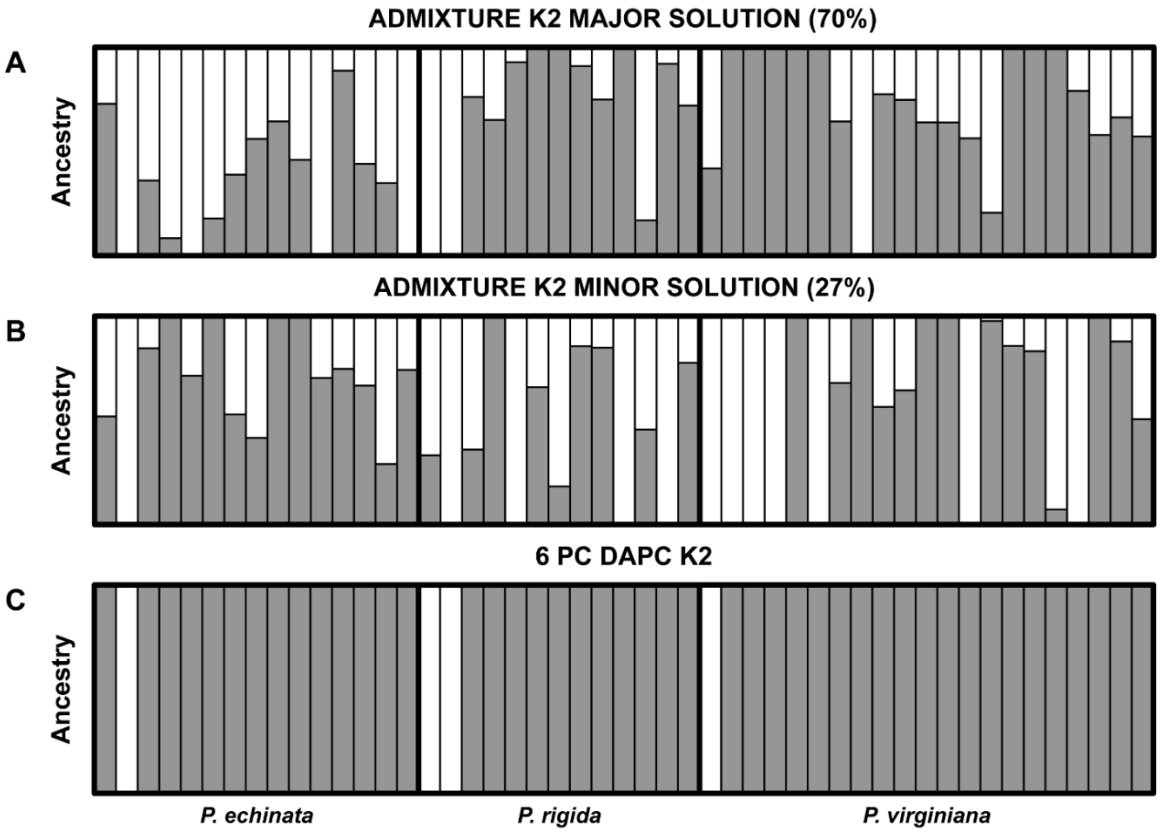


Figure 4.4 – Ancestry solutions for *K* = 2. To allow comparisons across solutions, individuals are displayed in the same order across plots, and grouped by original host. Neither the admixture solutions (70% - A; 27% - B) or the DAPC solution produce particularly meaningful clusters.

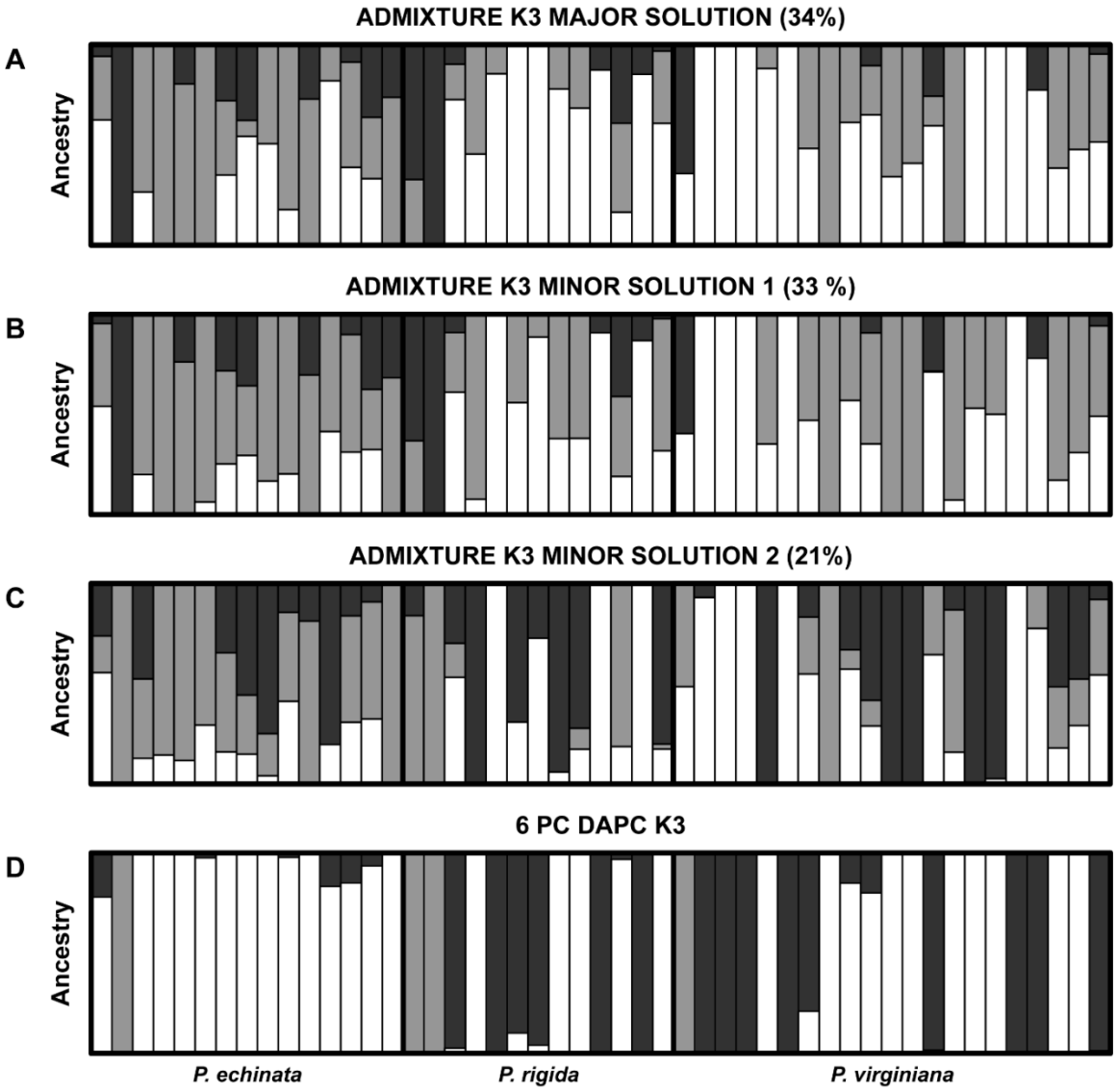


Figure 4.5 – Ancestry solutions for *K* = 3. To allow comparisons across solutions, individuals are displayed in the same order as in Figure 3, and grouped by original host. Although the three ADMIXTURE solutions (A-C) agree that most individuals are considerably admixed, individuals from *P. echinata* display a pattern of ancestry that differs from *P. rigida* and *P. virginiana* individuals. Although the DAPC suggests less individual admixture, *P. echinata* individuals still display a distinct pattern of ancestry.

Table 4.1 – AMOVA table for Arboretum individuals. *P*-values for each source of variation are given in parentheses. The majority of variation is found within individuals, suggesting little population structure between hosts.

Source of variation	Sum of squares	Variance components	Percentage variation
Among hosts (p = 1)	3083.20	10.48	0.89
Among individuals within hosts (p < 0.0001)	51337.56	63.12	5.35
Within individuals (p < 0.001)	49723.50	1105.21	93.76
Total	104144.26	1178.81	

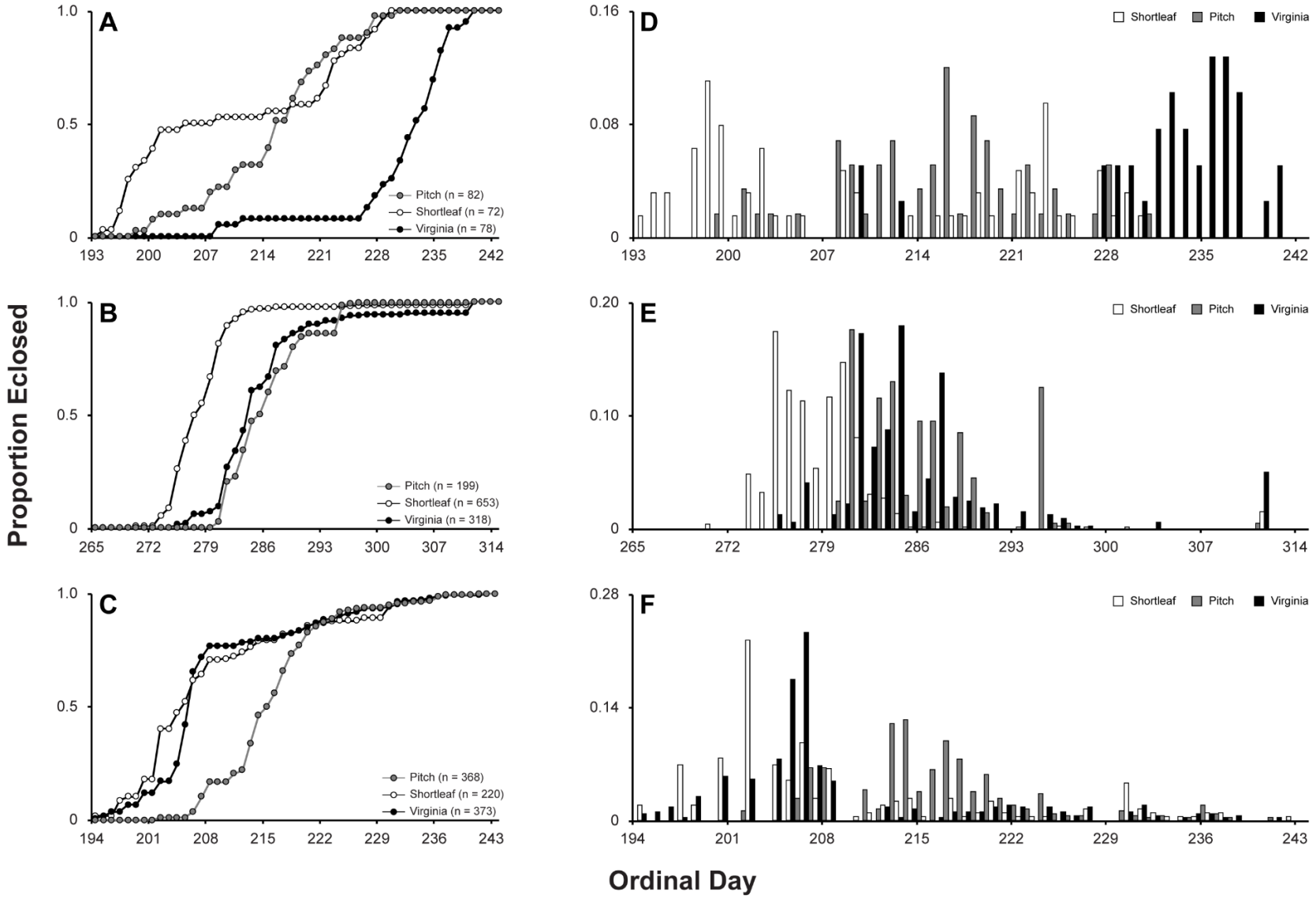

A
B
C
D
E
F
Proportion Eclosed
Ordinal Day
Pitch (n = 82)
Shortleaf (n = 72)
Virginia (n = 78)
Pitch (n = 199)
Shortleaf (n = 653)
Virginia (n = 318)
Pitch (n = 368)
Shortleaf (n = 220)
Virginia (n = 373)
Shortleaf
Pitch
Virginia

Figure 4.6 – Patterns of adult eclosion across host types in 2012, 2013, and 2014. Panels A-C show the cumulative eclosion curves over a 50-day span for each host type in 2012, 2013, and 2014, respectively. A visualization of daily emergence is per host over the same period is shown in panels D-E. The sample size for each host type per year is inset in the legends of panels A-C.

Table 4.2 – Pairwise temporal isolation by year. The pairwise isolation index I (Feder *et al.* 1993), post-hoc Tukey's HSD p-values, and bootstrapped Kolmogorov-Smirnov (KS) test p-values are given for each pairwise combination of host types in 2012, 2013, and 2014.

Comparison	**I**	**Tukey's HSD *P*-value**	**Bootstrapped KS test *P*-value**
2012			
Pitch *vs.* Shortleaf	0.51	0.0008	0.0139
Pitch *vs.* Virginia	0.86	<0.0001	<0.0001
Shortleaf *vs.* Virginia	0.82	<0.0001	<0.0001
2013			
Pitch *vs.* Shortleaf	0.81	<0.0001	0.0084
Pitch *vs.* Virginia	0.51	0.9927	0.0086
Shortleaf *vs.* Virginia	0.70	<0.0001	<0.0001
2014			
Pitch *vs.* Shortleaf	0.63	<0.0001	0.0971
Pitch *vs.* Virginia	0.58	<0.0001	0.1541
Shortleaf *vs.* Virginia	0.17	0.9998	0.8218

4.3.5 – Sexual Isolation

In the absence of host, none of the host types displayed a preference for same- or different-type partners (Figure 4.7, p = 0.43-0.85).

4.3.6 – Habitat Isolation

Overall females showed little evidence of preference for their original *vs.* alternative hosts (Figure 4.8). Neither Shortleaf (p = 1, 1) nor Virginia (p = 0.2101, 0.6291) females displayed a preference for their original host over either alternative host. Pitch females did not discriminate between Pitch and Shortleaf (p = 0.8036), but chose Virginia significantly more often than Pitch (p = 0.0266).

4.3.7 – Host Characteristics

All hosts varied significantly in needle width (Figure 4.9; $p < 0.001$).

4.3.8 – Ovipositor Morphology

Host had a significant impact on overall ovipositor shape (Figure 4.10; $F_{2,32}$ = 1.8382, p = 0.0379); as did family ($F_{19,32}$ = 1.4333, p = 0.0132). Females from Pitch and Virginia have significantly different ovipositor shapes (p = 0.0042); but neither differ significantly from Shortleaf females (p = 0.3108, 0.4482). Family has a significant effect on ovipositor length ($F_{19,32}$= 4.6119, $p < 0.0001$) and width ($F_{19,32}$=5.5986, $p < 0.0001$); but host did not impact either trait (length: $F_{2, 32}$ = 0.7889, p = 0.4630; width: $F_{2,32}$ = 0.0296, p = 0.9708)

4.3.9 – Larval performance

Although survival was not significantly impacted by host types ($F_{2,47}$ = 3.0795, p = 0.05537); it was impacted by rearing host ($F_{2,47}$ = 5.2817, p = 0.0085). Survival differed between sawflies reared on *P. rigida* and *P. echinata* (p = 0.0075). There was no

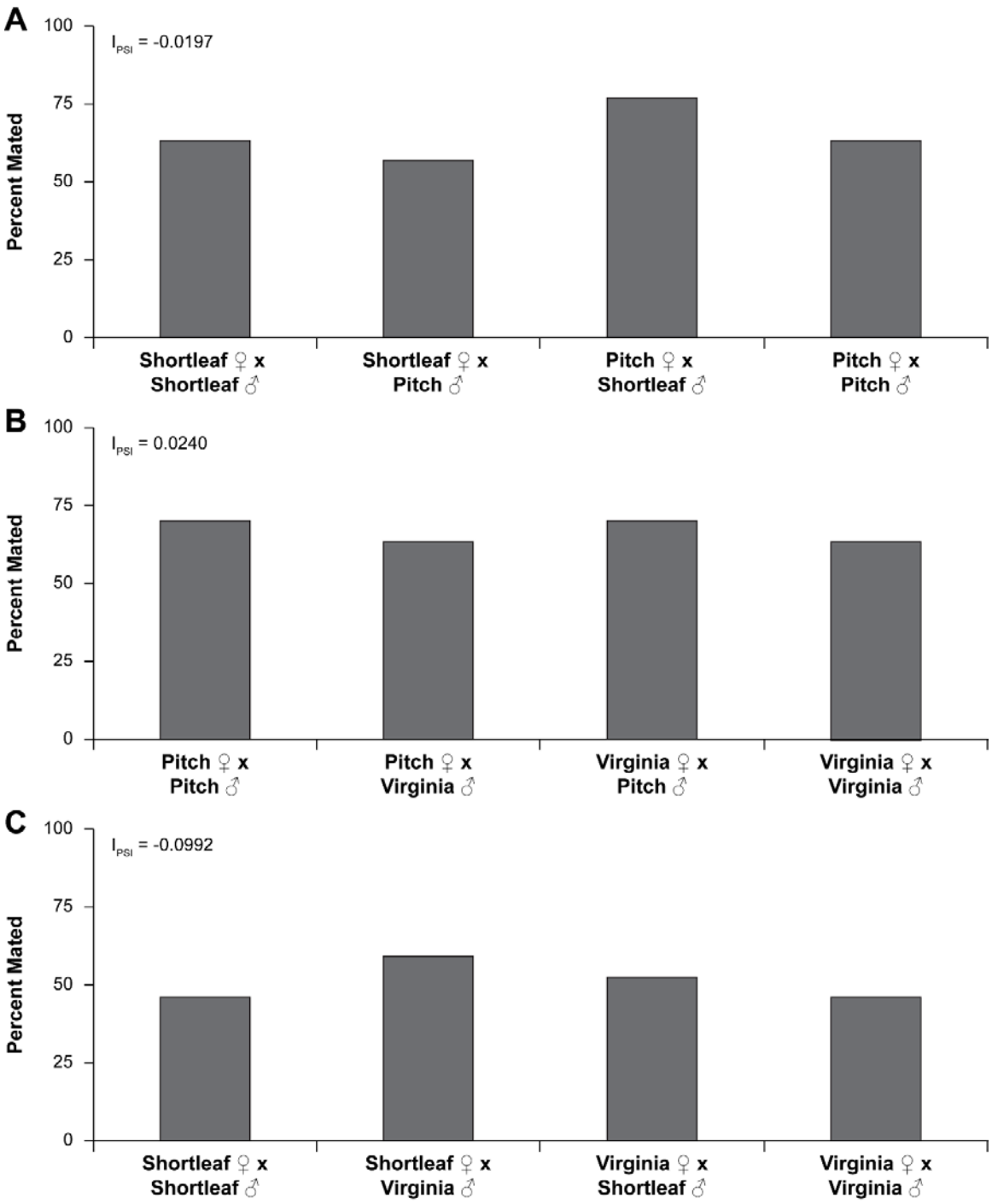


Figure 4.7 – Proportion of same-type *vs.* different-type matings for each pairwise combination of host types. None of the host types showed a preference for same- or different-type mates. I_{PSI} values for each pair of host types is inset.

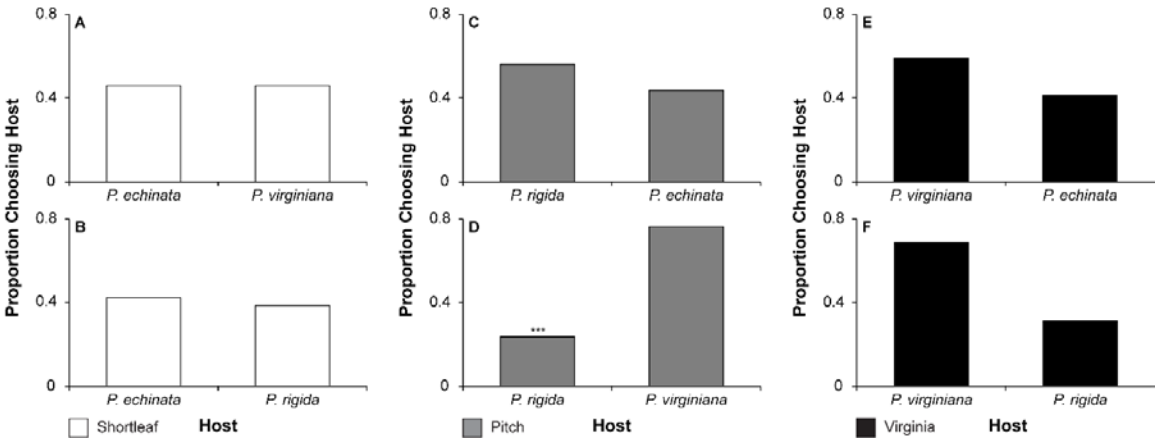


Figure 4.8 – Oviposition preference of each host type in choice cages. Females from each host type were offered a choice between their original host and one of the alternative hosts. Females from the Shortleaf (white bars; A, B) and Virginia (black bars; E, F) host types did not discriminate between their original *vs*. either alternative host plant. Females from Pitch (grey bars) did not discriminate between *P. rigida* and *P. echinata* (C), but preferred to oviposit on *P. virginiana* over their original host *P. rigida* (D).

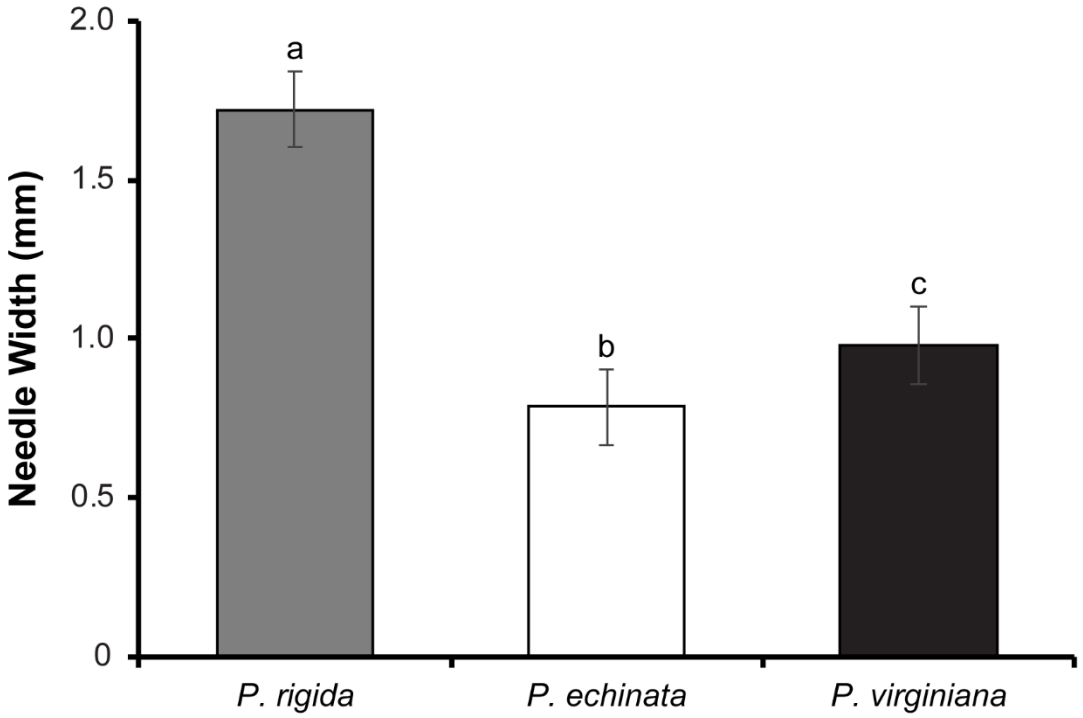


Figure 4.9 – Needle width of the three hosts present on the Trail of Pines. All hosts differ significantly in width, indicated by letters.

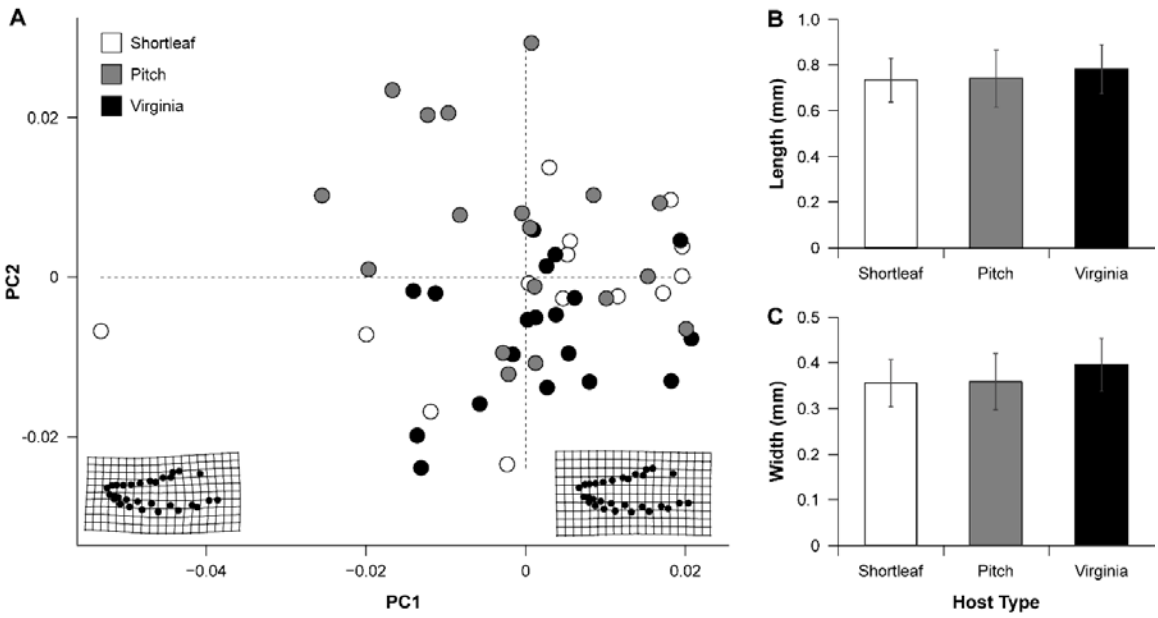


Figure 4.10 – Ovipositor morphology across the three host types. A. Principal Components analysis of overall ovipositor shape of *N. lecontei* females from Shortleaf (white), Pitch (grey) and Virginia (black). Warp grids represent ovipositor shape along PC1. Neither ovipositor length (B) or width (C) differs significantly between host types.

interaction between host type and rearing host ($F_{4,47} = 2.1560$, $p = 0.08862$). Within lines (Figure 4.11), there are no significant differences between Shortleaf ($p = 0.5830$-0.7800) or Pitch ($p = 0.3020$-0.6516) families regardless of rearing host; but Virginia sawflies have significantly reduced survival when reared on *P. echinata* compared to *P. virginiana* ($p = 0.0007$) or *P. rigida* ($p = 0.0003$). Survival of Virginia sawflies did not differ on *P. rigida* compared to *P. virginiana* ($p = 0.9939$).

Host type ($F_{2, 1207} = 781.33$, $p < 0.0001$) and rearing host ($F_{2,1207} = 113.48$, $p < 0.0001$) both had a significant impact on female cocoon weight (Figure 4.12), with all host types and all rearing hosts differing significantly ($p < 0.0001$). There was also a significant interaction between host type and rearing host ($F_{4,1207} = 81.52$, $p < 0.0001$). Shortleaf sawflies achieve significantly lower cocoon weights when reared on *P. echinata* ($p < 0.0001$); but do not differ in weight when reared on *P. virginiana* or *P. rigida* ($p = 0.3264$). Pitch females differ in weight across all rearing hosts ($p < 0.0029$), and obtain the highest cocoon weight on *P. virginiana*. Virginia females also obtain the highest cocoon weight when reared on *P. virginiana* ($p < 0.0001$), but do not differ in weight when reared on *P. echinata* vs *P. rigida* ($p = 0.3730$).

4.4 – DISCUSSION

Overall we see little evidence of genetic divergence between host types of *N. lecontei* at this recently colonized site. Despite the lack of genetic divergence, however, we do see morphological divergence in ovipositor shape, and variation in performance between host types. We see no evidence of sexual or habitat isolation, but the host types are partially temporally isolated. Therefore, although it is possible that there is a single, generalist population of *N. lecontei* at this site; these data suggest the host types may be

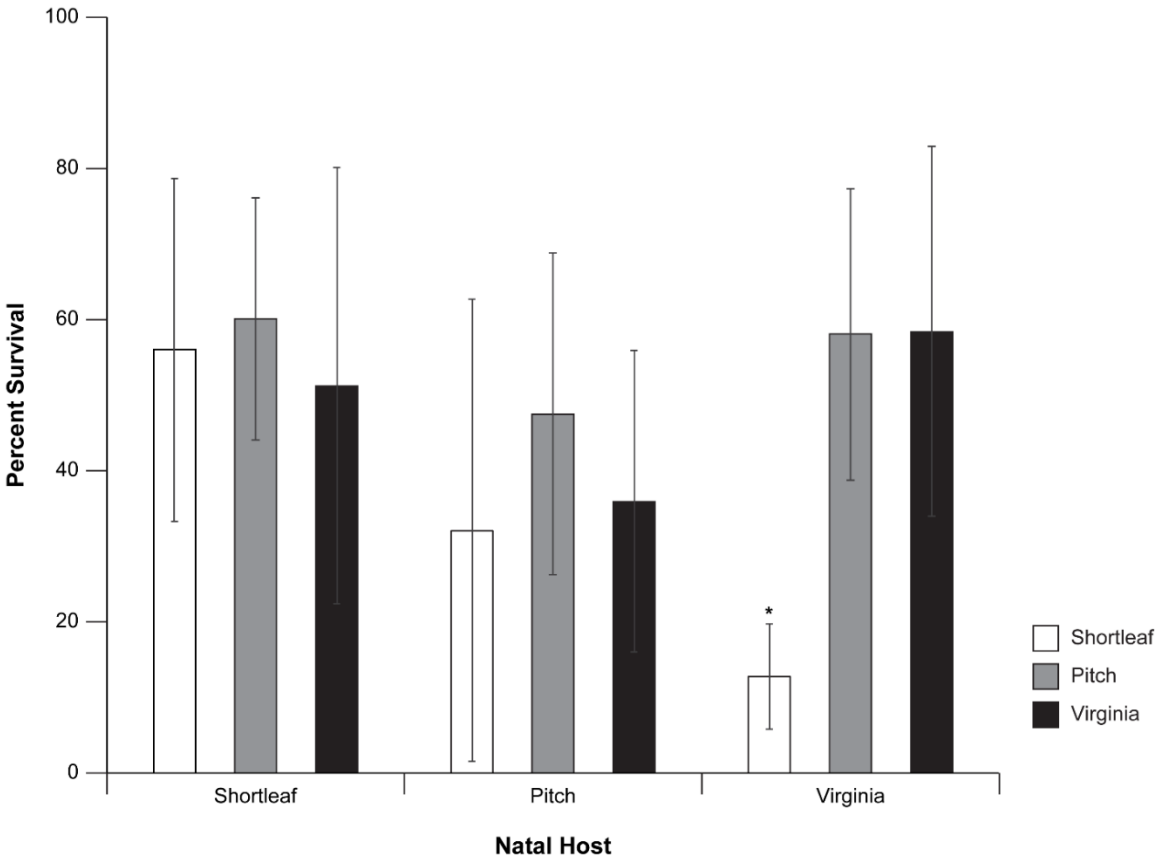


Figure 4.11 - Survival of host types when reared on different hosts. Although rearing host does not impact survival of Shortleaf or Pitch host types; Virginia sawflies survive poorly when reared on Shortleaf pine.

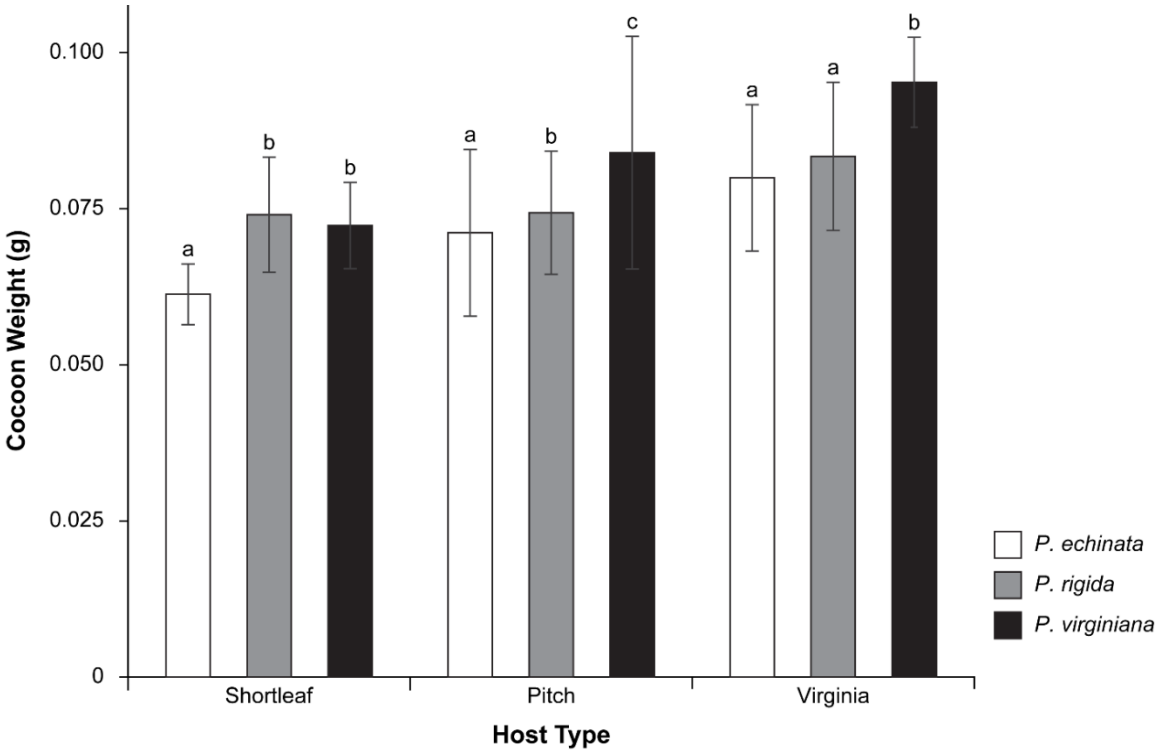


Figure 4.12 – Female cocoon weight of host types reared on different hosts. Host type and rearing effect both impact female cocoon weight. Shortleaf sawflies obtain the lowest cocoon weight when reared on their original host, *P. echinata*. They obtain equal cocoon weights on *P. virginiana* and *P. rigida*. Pitch females obtain significantly different cocoon weights when reared on each host, performing best on *P. virginiana*, and worst on *P. echinata*. Virginia females obtain the highest cocoon weight when reared on their original host, *P. virginiana*, but do not differ in performance on *P. rigida* and *P. echinata.*

at the very earliest stages of speciation. These results offer some interesting insights into the role divergent host use can play in population divergence and speciation in *Neodiprion*, which we discuss below.

4.4.1 – Patterns of Divergence

At the very earliest stages of ecological speciation, when gene flow between populations is high, divergence is likely to be observed solely at directly selected loci. Over time, however, neutral regions closely linked to those under selection may also diverge via divergence hitchhiking, generating elevated F_{ST} peaks around selected loci, and facilitating fixation of new mutations physically linked to selected regions (Via 2009, 2012; Feder & Nosil 2010; Feder *et al.* 2012b). Genome hitchhiking can arise if multiple regions of the genome are under selection, and can reduce gene flow genome-wide. This global reduction in gene flow facilitates fixation of new beneficial mutations genome-wide, and can also allow neutral divergence between populations (Feder *et al.* 2012b; a, Flaxman *et al.* 2012, 2013). As divergently selection regions accumulate across the genome, populations can reach a threshold of divergence where genomic differences "congeal" and the populations rapidly transition into species ("genome-wide congealing"/four stage speciation model; Flaxman *et al.* 2013; Feder *et al.* 2014).

When selection is weak, and migration between populations is high, however, divergent selection may be unable to reduce genome-wide gene flow enough to allow neutral divergence to accumulate (Thibert-Plante & Hendry 2010). Divergence at even a single locus is dependent on establishment and maintenance of polymorphism in the face of gene flow, which may be difficult if selection is weak (Yeaman & Otto 2011). Therefore, it may take a considerable amount of time to accumulate enough variation

across multiple loci involved in adaptation between environments before divergence can proceed (Feder *et al.* 2014). It is perhaps unsurprising then, that we see relatively little evidence of genetic divergence between the host types in our study, as the field site was colonized a maximum of 20 years, or 60 generations ago (Benjamin 1955; Wilson *et al.* 1992). Even if selection on individual loci was strong, there may simply not have been enough time for significant neutral divergence to build up (Funk & Omland 2003; Stireman *et al.* 2005).

Weak selection in the face of gene flow may also explain the performance differences observed between host types. Selection should favor the evolution of traits in each host type that increase performance on their respective host, perhaps at the cost of performance on the alternative hosts (Rausher 1984; Karban 1989; Mackenzie 1996; Cornell & Hawkins 2003). Although there is a considerable amount of variation in performance among the host types when reared on the three potential hosts, the direction of the differences is not always as predicted (Figure 4.12), and frequently does not significantly impact survival (Figure 4.11). In fact, only one host type achieves the highest cocoon weight on its original host (Virginia, on *P. virginiana*). Interestingly, Virginia is also the only host type to display a significant reduction in survival on an alternative host (*P. echinata*).

The relative lack of performance differences between host plants could also contribute to the overall lack of preference differences. Although female preference and larval performance may be decoupled (Friberg *et al.* 2015), the "preference-performance" hypothesis proposes that female preference evolves to match whatever host their larvae perform best on (Jaenike 1978; Thompson 1988; Gripenberg *et al.* 2010). As our host

types have not yet evolved performance differences, there may be little to no selective pressure for specific host preferences to evolve. At this point, even if beneficial combinations of performance and preference alleles were to arise in some individuals, selection may be too weak to prevent these combinations from being broken up by recombination.

It should also be noted, however, that we offered females seedlings of each of the three species. Pines are known to vary in their volatile profile (unpublished data) and needle architecture (Bendall *et al.* 2017) across age groups, so it is possible the host types would demonstrate stronger preferences if offered mature host material. Future studies using clippings of mature hosts, or sleeve cages on planted mature pines would be useful to further assess female preferences.

4.4.2 – Oviposition traits – important to Neodiprion *divergence?*

Despite their strikingly low level of genetic divergence, Pitch and Virginia females have significantly different ovipositor shapes (Figure 4.10). For these morphological differences to emerge and be maintained in the face of presumably high gene flow between these host types, selection on ovipositor morphology is likely strong. This is consistent with previous work in the genus, as mismatches in oviposition-related traits carried a strong fitness penalty, and generate extrinsic postzygotic isolation between *N. lecontei* and its sister species *N. pinetum* (Bendall *et al.* 2017). The fact that differences in ovipositor morphology arise relatively quickly following a host shift, and continue to contribute to reproductive isolation after speciation, suggests oviposition traits may play a generally important role in *Neodiprion* speciation.

Alternatively, morphology may be plastic, and change based on the host an individual is reared upon (Görür 2003; Pfennig *et al.* 2010). Although, the ovipositors included in this study were obtained from wild-caught females reared on their original host, we are currently preparing an additional set of ovipositors from females reared for several generations in the laboratory on *P. banksiana*. If we the differences in ovipositor shape are also found in this additional dataset, it will be compelling evidence that the changes in ovipositor shape are adaptive.

4.4.3 – Contributions of Temporal Isolation

Although $K = 1$ is favored by both clustering methods (Figure 4.1), when considering higher values of K, it appears Shortleaf is beginning to diverge from the other host types (Figure 4.5). Although all pairwise F_{ST} are relatively low, F_{ST} is approximately 5-fold higher between Shortleaf and Virginia or Shortleaf and Pitch than it is between Pitch and Virginia. This result is somewhat surprising as, among the host types, Pitch and Virginia are the most morphologically distinct (Figure 4.10). Shortleaf, however, is arguably the most *temporally* distinct of the host types. Shortleaf adults consistently eclose first (Figure 4.6D-F), and differ in mean eclosion date from at least one of the other host types per year (Table 4.2).

If temporal isolation does contribute to host-associated divergence within *N. lecontei*, it would join a number of other phytophagous insect systems where partial or complete temporal isolation has contributed to divergence via host shifts [e.g., *Terellia fuscicornis* (Sayar *et al.* 2009), *Prodoxus quinquepunctellus* (Groman & Pellmyr 2000), *Eurosta solidaginis* (Craig & Mopper 1993; Craig *et al.* 2001), *Rhagoletis* flies (Feder *et al.* 1993, 1994; Powell *et al.* 2014; Egan *et al.* 2015)]. In these cases, adult eclosion

patterns have shifted to match host phenology, limiting opportunities for between-host matings due to limited adult lifespans. In this way, the temporal isolation acts as a "magic trait" (Gavrilets 2004; Servedio *et al.* 2011) as matching adult eclosion to host phenology simultaneously reduces gene flow between populations. This reduction in gene flow may then facilitate progression along the speciation continuum (Taylor & Friesen 2017).

Although sampling limitations allowed us to consider only 1 of the 2-3 generations typically observed at the Trail of Pines per year in this study, we note that this pattern is upheld regardless of which generations we assessed each year (first generation in 2012 and 2014; second generation in 2013). This suggests differences in eclosion timing may be at least partially heritable. The lack of host preference observed at the site, however, suggests, even if eclosion time is heritable, it is not linked to preference for a given host. An alternate explanation for the consistent pattern of eclosion order could stem from variation in host quality throughout the season. Specialist insects (like *N. lecontei*) are generally thought to be able to distinguish between high- and low-quality hosts (Janz & Nylin 1997). Although we do not know how the individual hosts at the Trail of Pines vary throughout year, nor what host cues *N. lecontei* uses in selecting hosts, seasonal variations in moisture levels (Van Wagner 1967), chemical content (Nerg *et al.* 1994), and volatile profile (Geron & Arnts 2010) have all been noted in pines. If the hosts at the Trail of Pines do vary significantly from each other in chemical or nutritional content such that the "optimal" host varies across the season, the observed patterns may stem from *N. lecontei* females using the best available host at the time of their eclosion.

4.4.4 – Progress towards speciation?

Although our results suggest that the host types at the Trail of Pines are experiencing divergent selection, and perhaps at the very earliest stages of divergence, it is difficult to determine if and how far this divergence will progress. In some situations, populations get "stuck" at intermediate stages of the speciation continuum (e.g., Feder *et al.* 1994; Dopman *et al.* 2005, 2010; Nosil 2007; Kronforst 2008). Alternatively, divergent selection may lead to the maintenance of stable polymorphisms, rather than the evolution of reproductive isolation (Crispo *et al.* 2006; Rueffler *et al.* 2006; Svensson *et al.* 2009). There are several factors that influence how far ecological speciation will progress, such as the geographic context of the populations, the strength of selection, and the underlying genetic architecture of the traits under divergent selection (Rueffler *et al.* 2006; Nosil *et al.* 2009b; Nosil 2012). Determining if, and how often, populations like the Trail of Pines stabilize into reproductively isolated will require long term monitoring, and would benefit from comparisons of other sympatric, parapatric, and allopatric pairs.

4.5 – SUPPORTING INFORMATION

The following supporting information can be found in Appendix 3.

- Table A3.1. Collection and usage information for all samples.
- Table A3.2. List of individuals with their natal host, variable length barcode, and Illumina index.
- Table A3.3 – Missing data and proportion of heterozygous sites per individual.

Chapter 5 : Synthesis

5.1 – Host Shifts: Drivers or Followers of Reproductive Isolation?

The goal of this dissertation was to investigate the role of host shifts in driving phytophagous insect speciation, using *Neodiprion* sawflies as a model. Specifically, I predicted that, if host shifts frequently drive speciation in *Neodiprion*, I would find evidence of host-driven divergence within species utilizing a wide range of hosts. Therefore, I examined populations of the redheaded pine sawfly, *N. lecontei*, a widespread pest species found on multiple hosts across its range (Middleton 1921; Benjamin 1955), across multiple spatial scales, for evidence of incipient divergence.

In Chapter 2, I first looked at range-wide patterns of neutral divergence. I identified thee genetic clusters, each of which corresponded to a distinct geographic area and suite of host plants, and dated the divergence of these clusters to the late Pleistocene. I proposed a Pleistocene divergence scenario for *N. lecontei*, and identified potential refugia for each of the clusters. Finally, using Mantel and partial Mantel tests, I found a significant relationship between genetic differentiation and geographical distance in all three clusters; and a significant relationship between genetic differentiation and ecological distance (host use) in two of the three clusters.

The scale reduced in Chapter 3, where I looked within the single genetic cluster where we did not find a significant relationship between genetic differentiation and host use. Here, despite an expanded dataset, I confirmed the previous finding of little to no relationship between genetic differentiation and host use. However, I did find evidence for strong regional host affiliations, which are driven by spatial variation in host preference. At local scales, there was also temporal variation in host utilization. I also

found evidence of host-associated divergence in ovipositor morphology, a performance-related trait.

Chapter 4 examined the smallest spatial scale, and considered a recently colonized field site harboring *N. lecontei* on three hosts. Here, I again saw little evidence of host-associated divergence in neutral markers. There was also little evidence of sexual isolation, distinct host preference, or of divergence in larval performance across hosts. There was, however, evidence of partial temporal isolation, as well as host-associated differences in ovipositor morphology.

Considering these results together, although there are multiple lines of evidence to suggest local adaptation to hosts is occurring, I find relatively little evidence that this divergent selection has generated reproductive isolation. For example, although divergent host use leads to divergent ovipositor morphology at all spatial scales we examined (region-wide, within regional clusters, and at sympatric sites); I do not consistently see divergent host preferences, larval performance, sexual isolation, or genome-wide neutral divergence. Interestingly, I only observe neutral differentiation and divergent host preferences when the diverging populations are also either geographically and/or temporally isolated. This suggests, at least in the case of *Neodiprion*, although host shifts can generate divergent selection and may contribute to population divergence after additional forms of isolation have arisen, it may be insufficient on its own to *initiate* speciation.

5.2 – BUMPS IN THE ROAD TO SPECIATION

As I found in *N. lecontei*, divergent selection does not always lead to significant population divergence and speciation. Reflecting this, the speciation process is often

described as a continuum, ranging from completely undifferentiated, panmictic populations on one end, to partially isolated intermediates such as ecotypes and host races, to fully isolated species on the other (Darwin 1859; Walsh 1864, 1867; Nosil *et al.* 2009b; Hendry 2009; Gourbière & Mallet 2010; Merrill *et al.* 2011). Populations at all stages of the speciation continuum can be found in nature (Nosil 2012), including a handful of study systems with examples from multiple points of the continuum (Langerhans *et al.* 2007; Seehausen 2008; Nosil & Sandoval 2008; Peccoud *et al.* 2009; Merrill *et al.* 2011). This raises an interesting question: Why does divergent selection lead to complete speciation in some instances, but stall at the earliest stages in others? In other words, what determine how far speciation proceeds?

There are several non-selective factors can facilitate speciation (Nosil *et al.* 2009b; Nosil 2012). For example, both increased time since divergence (Coyne & Orr 1989, 2004) and geographic separation (Coyne & Orr 2004; Gavrilets 2004) are associated with greater levels of reproductive isolation. Speciation may also be promoted if reproductive isolation is pleiotropically influenced (Funk 1998; Bradshaw & Schemske 2003; Coyne & Orr 2004; Gavrilets 2004), or controlled by genes physically linked to directly selected loci (Noor *et al.* 2001; Feder *et al.* 2003b; a; Coyne & Orr 2004; Rundle & Nosil 2005). Finally, speciation may also be promoted if there is a large amount of standing genetic variation for selection to act on (Barrett & Schluter 2008), or via so-called “one-allele assortative mating” mechanisms, where the same assortative mating gene spreads through both diverging populations (Rundle & Nosil 2005; Ortiz-Barrientos & Noor 2005).

Although the number and identity of loci under selection in diverging populations is not yet known in *N. lecontei*, several of these factors likely influence divergence in *N. lecontei*. For example, although the variation in larval performance likely indicates the existence of a good amount of standing genetic variation for host-related traits, divergence of the host types at the Arboretum is likely constrained by their sympatric status, as well as their recent colonization of the site (and therefore recent initiation of divergent selection). Time since divergence may also be a factor in the North cluster, as indicated by their lack of migration-drift equilibrium.

Progression towards speciation can also be influenced by the strength and number of traits under selection (Rueffler *et al.* 2006; Nosil *et al.* 2009b; Nosil 2012). As the strength of selection on a single trait can be limited by the environment (Endler 1986; Kingsolver *et al.* 2001), available genetic variation (Bush 1969a; Futuyma *et al.* 1995; Gavrilets & Vose 2005), and/or functional constraints (Lande 1982; Arnold 1992), multifarious selection acting on many traits may be required to generate enough divergence for speciation to proceed (Nosil *et al.* 2009b; Nosil 2012). Although the strength of selection or contribution to reproductive isolation, on any given trait under multifarious selection may be weak, their combined effects may be strong (Matsubayashi & Katakura 2009). However, multifarious selection may be unable to overcome strong gene flow, so stronger selection on a handful of traits may be more effective at promoting divergence in the face of gene flow(Nosil *et al.* 2009b; Nosil 2012)

Although I did not directly measure the strength or number of traits under selection in *N. lecontei*, it is likely the total selection strength in this system is not strong enough to overcome the level of gene flow between hosts. Selection on individual traits

may be weak as well. For example, although variation in larval performance existed at the Arboretum, it was generally not in the predicted direction and had little impact on survival. One exception to this for *N. lecontei* may be ovipositor morphology, which consistently changed in response to divergent selection, even in the face of likely strong gene flow at local sites.

Another possibility is that divergent selection may lead to evolution of phenotypic plasticity rather than ecological speciation (Rueffler *et al.* 2006; Nosil *et al.* 2008; Pfennig *et al.* 2010; Nosil 2012; Wund 2012). The evolution of plasticity may be favored in unpredictable environments, when costs of plasticity are low, or when levels of gene flow are high (Berrigan & Scheiner 2004; Leimar *et al.* 2006; Svanbäck *et al.* 2009). I have not yet ruled out plasticity in the phenotypic traits I measured in this dissertation, although some evidence, including persistence of host preferences in the lab, suggest at least some differences are genetic. However, it should be noted that plasticity can also promote speciation by facilitating colonization of new environments (Price *et al.* 2003; Pavey *et al.* 2010; Thibert-Plante & Hendry 2011), so the existence of phenotypic plasticity is not necessarily a speciational dead-end.

Ultimately, determining if divergent selection will manifest in reproductive isolation within *N. lecontei,* or if it can initiate speciation events in the genus will require additional research into the contributions of geography, ecology, and additional barriers such as temporal on driving phenotypic and neutral divergence.

5.3 – INTEGRATION IS KEY, OR ADVICE FOR FUTURE RESEARCHERS

While working on this dissertation, I have learned several lessons that would be valuable for future researchers considering similar questions. First, I would recommend

future studies consider multiple spatial scales for their analyses. My initial analyses of range-wide patterns of divergence allowed me to determine overall population structure, and determine the appropriate spatial scale for subsequent questions. In addition, considering local scales in combination with cluster-wide work allowed me to identify which phenotypic traits predictably respond to divergent selection (ovipositor morphology), *vs.* those requiring additional forms of isolation to develop.

My dissertation research has also revealed the importance of integrating phenotypic assays when interpreting population genomic results. It is difficult to assess the biological relevance of patterns of divergence without considering the ecological and evolutionary history of the system (Richardson *et al.* 2016). This is particularly evident in interpreting the "negative" isolation-by-environment result from the North genetic cluster. Incorporating the preference and performance-related phenotypic data allowed us to identify a role of host-associated divergence despite the non-equilibrium state of the cluster.

Finally, my work highlights the need to simultaneously consider multiple sources of reproductive isolation. Historically, speciation research focused almost exclusively on the geographic context of speciation, particularly on arguments for the feasibility of sympatric speciation (Mayr 1942, 1963; Smith 1966; Bush 1969a; Drès & Mallet 2002b; Coyne & Orr 2004). Recently, this focus has almost entirely shifted to consider the role of ecology and selection in driving speciation (Rundle & Nosil 2005; Schluter 2009; Via 2009; Matsubayashi *et al.* 2010; Nosil 2012). However, interpretation of patterns of variation within *N. lecontei* required consideration of not only geography and ecology, but also historical isolation and temporal isolation. Considering the complexity of many

natural systems, future studies should incorporate tests for these additional barriers when possible.

Appendix 1 – History, geography, and host use shape genome-wide patterns of genetic variation in the redheaded pine sawfly (*Neodiprion lecontei*)

Table A1.1 – Sampling locations for all individuals included in this study. Specimens noted with a dagger (†) were adult females, and specimens marked with a double dagger (‡) were adult males. All other specimens were larvae of unknown sex. Full sequences for the adapter-ligated barcodes and Illumina indexes are listed in Tables S2 and S3, respectively.

Specimen ID	Latitude (°N)	Longitude (°W)	Location	Host Plant	Barcode	Illumina Index
003-01	42.229	-71.523	Hopkinton, MA	*P. banksiana*	GCATG	3
017-01_E1	44.544	-73.215	Malletts Bay, VT	*P. resinosa*	TGGAA	2
025-0263_D	45.016	-75.646	Kemptville, ON	*P. resinosa*	AGCTA	2
025-0309_D	44.395	-77.205	Tweed, ON	*P. resinosa*	GTCCG	7
025-0312_D	44.73	-79.169	Orillia, ON	*P. resinosa*	GAGAT	2
025-0355	46.017	-77.45	Chalk River, ON	*P. resinosa*	GAGAT	7
075-04	28.096	-81.275	Canoe Creek, FL	*P. elliottii*	AACCA	1
077-04_D	26.923	-81.336	Palmdale, FL	*P. elliottii*	AAGGA	2
086-04	29.718	-82.457	Gainesville, FL	*P. palustris*	CGTAC	1
087-04_D	29.748	-82.477	Gainesville, FL	*P. taeda*	AACCA	2
088-04_D	29.748	-82.477	Gainesville, FL	*P. palustris*	GGCTC	7
096-04_E1	31.498	-84.593	Morgan, GA	*P. taeda*	TCAGT	2
097-04_E1	31.498	-84.593	Morgan, GA	*P. glabra*	CTGTC	7
102-04_E1	31.555	-83.989	Sylvester, GA	*P. elliottii*	TAGTA	2
106-04_D	32.074	-83.761	Vienna, GA	*P. elliottii*	ACTTC	2
116-04_E1	36.039	-85.109	Crossville, TN	*P. virginiana*	CTTGG	7
125-02_E1	43.115	-71.1	Nottingham, NH	*P. sylvestris*	GTCCG	2
132-04_E1	38.716	-76.064	Trappe, MD	*P. virginiana*	TCACG	2
133-04_E1	38.716	-76.064	Trappe, MD	*P. taeda*	TACGT	2
145-04_D	43.781	-71.17	Ossippee, NH	*P. rigida*	CGTAC	2

Table A1.1 (cont.)

Specimen ID	Latitude (°N)	Longitude (°W)	Location	Host Plant	Barcode	Illumina Index
164-02_D	45.073	-77.71	Bancroft, ON	*P. resinosa*	ATTAC	2
168-02	44.856	-77.859	Apsley, ON	*P. banksiana*	CGTAC	3
168-04	43.685	-71.117	Ossippee, NH	*P. rigida*	GAGTC	1
170-04_D	35.17	-88.592	Selmer, TN	*P. taeda*	TATAC	7
174-03A_D	29.68	-83.257	Dixie Co, FL	*P. taeda*	CGGCT	2
177-02	44.73	-79.169	Sebrite, ON	*P. resinosa*	CGAAT	1
178-03_E1	30.428	-85.603	Crystal Lake, FL	*P. palustris*	TCTGC	2
180-03_D	28.787	-81.982	Lake Co, FL	*P. taeda*	CGATC	2
183-03_D	26.871	-81.521	Glades Co, FL	*P. palustris*	TGCAT	2
185-03_E1	26.871	-81.521	Palmdale, FL	*P. elliottii*	GACAC	7
188-04_D	44.571	-91.635	Eau Claire, WI	*P. banksiana*	ACTGG	2
196-04_E1	43.912	-90.866	Sparta, WI	*P. banksiana*	TCCGG	2
207-04_D	45.975	-90.496	Park Falls, WI	*P. banksiana*	ATACG	2
224-04_D	37.35	-78.016	Amelia, VA	*P. taeda*	AATTA	2
339-02	46.348	-79.334	North Bay, ON	*P. banksiana*	CAACC	1
345-02_E1	46.395	-79.244	North Bay, ON	*P. banksiana*	TACCG	2
349-02	46.378	-78.867	Mattawan, ON	*P. resinosa*	CGTCG	1
372-02_E1	41.874	-70.652	Plymouth, MA	*P. rigida*	GTCGA	2
RB001_D‡	40.68	-74.234	Union, NJ	*P. strobus*	ACGGT	2
RB002_D	38.171	-83.556	Morehead, KY	*P. rigida*	TCGAT	2
RB004_E1	44.35	-89.822	Wisconsin Rapids, WI	*P. banksiana*	TTACC	2
RB008_D‡	32.138	-82.969	Helena, GA	*P. elliottii*	GAGTC	2
RB009_D†	32.523	-83.496	Dry Branch, GA	*P. taeda*	GGATA	2
RB010‡	32.523	-83.496	Dry Branch, GA	*P. echinata*	AAGGA	1

Table A1.1 (cont.)

Specimen ID	Latitude (°N)	Longitude (°W)	Location	Host Plant	Barcode	Illumina Index
RB015_D	44.461	-85.992	Springdale Township, MI	*P. banksiana*	GGTTG	2
RB017_D	37.984	-84.511	Lexington, KY	*P. mugho*	GCTGA	2
RB018	43.797	-71.915	Dorchester, NH	*P. resinosa*	AATTA	1
RB020_D	37.066	-84.159	London, KY	*P. echinata*	CTTGG	2
RB025_D	41.268	-78.28	Jay Township, PA	*P. mugho*	CGTCG	2
RB026_D	37.333	-77.981	Amelia, VA	*P. taeda*	GAGTC	7
RB027_D	33.99	-83.796	Auburn, GA	*P. echinata*	GCCGT	7
RB028‡	39.69	-74.593	Pine Barrens, NJ	*P. rigida*	ATGAG	7
RB042A	38.023	-84.494	Lexington, KY	*P. nigra*	TCGAT	3
RB044_B	27.692	-82.42	Ruskin, FL	*P. elliottii*	GGCCA	7
RB047b‡	27.618	-81.815	Bowling Green, FL	*P. palustris*	ACTGG	7
RB054	29.507	-81.86	Fort McCoy, FL	*P. palustris*	TCCGG	7
RB063B	29.32	-81.727	Fort McCoy, FL	*P. palustris*	CGTAC	7
RB066B†	29.507	-82.96	Chiefland, FL	*P. palustris*	TACGT	1
RB067	29.508	-82.958	Chiefland, FL	*P. palustris*	GTCGA	1
RB071	32.843	-87.952	Eutaw, AL	*P. echinata*	TCCGG	3
RB075_B	32.239	-80.859	Bluffton, SC	*P. palustris*	GTAGT	7
RB077	43.759	-85.741	Bitely, MI	*P. banksiana*	TCAGT	3
RB089_B	43.796	-85.74	Bitely, MI	*P. banksiana*	TCAGT	7
RB090.1	44.657	-84.414	Grayling, MI	*P. sylvestris*	TCGAT	7
RB091	44.657	-84.414	Grayling, MI	*P. banksiana*	CTGAT	3
RB094	45.504	-84.615	Glaque Beach, MI	*P. banksiana*	CGAAT	7
RB095	46.094	-85.339	Naubinway, MI	*P. banksiana*	CTGCG	7
RB096_B	46.096	-85.394	Naubinway, MI	*P. banksiana*	TAGTA	7

Table A1.1 (cont.)

Specimen ID	Latitude (°N)	Longitude (°W)	Location	Host Plant	Barcode	Illumina Index
RB099	45.924	-86.303	Manistique, MI	*P. banksiana*	ATTAC	7
RB100	46.354	-89.179	Watersmeet, WI	*P. banksiana*	ATGAG	3
RB104	44.985	-88.449	Suring, WI	*P. resinosa*	CGTCG	3
RB106	37.913	-79.896	Clifton Forge, VA	*P. virginiana*	AACCA	7
RB107	38.212	-79.719	Mountain Grove, VA	*P. rigida*	GGTTG	7
RB108	38.678	-79.399	Deer Run, WV	*P. rigida*	TGCAT	7
RB110	39.934	-74.533	Brown Mills, NJ	*P. rigida*	TCTGC	3
RB112	39.621	-74.428	Tuckerton, NJ	*P. rigida*	CATAT	7
RB118	38.592	-79.172	Brandywine, WV	*P. virginiana*	CGGTA	7
RB119	37.713	-79.367	Buena Vista, VA	*P. virginiana*	CAACC	7
RB120†	38.209	-84.39	Georgetown, KY	*P. sylvestris*	CGGCT	3
RB124	40.638	-74.368	Scotch Plains, NJ	*P. sylvestris*	GGCTC	1
RB125‡	39.717	-78.28	Sideling Hill, MD	*P. mugho*	CGATC	3
RB129	38.014	-84.504	Lexington, KY	*P. echinata*	GACAC	1
RB130	32.277	-80.983	Bluffton, SC	*P. palustris*	GTAGT	3
RB131	32.277	-80.983	Bluffton, SC	*P. taeda*	CTGTC	3
RB132	38.014	-84.504	Lexington, KY	*P. rigida*	CTGCG	3
RB133	38.014	-84.504	Lexington, KY	*P. virginiana*	GCTGA	1
RB136_B	43.759	-85.741	Bitely, MI	*P. resinosa*	GTCGA	7
RB137	27.618	-81.815	Bowling Green, FL	*P. elliottii*	CTTGG	1

Table A1.2 – 5-bp barcodes with associated P1 adapter sequences. Like many other RAD methods, ddRAD relies upon unique barcode sequences to allow inclusion of multiple individuals per sequencing lane. These barcodes are incorporated into the P1 adapter which is ligated to one of the overhangs left by enzymatic digestion. A P2 adapter is ligated to the other overhang. P1 and P2 adapters are created by annealing a PX.1 adapter with a complementary PX.2 adapter prior to ligation to fragmented DNA. For this study, we utilized 48 of the "flex" P1 adapters described in Peterson *et al* 2012, which were ligated to the ^GTAC overhang left by NlaIII digestion. A single, biotinylated "flex" P2 adapter was ligated to the TTAA^ overhang left by EcoRI digestion (P2.1 sequence: GTGACTGGAGTTCAGAC-GTGTGCTCTTCCGATCT; P2.2 sequence: /5Phos/AATTAGATCGGAAGAGCGAGAACAA/3Bio/).

5-bp barcode	P1.1 adapter sequence	P1.2 adapter sequence
GCATG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCATGCATG	/5Phos/CATGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
AACCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACCACATG	/5Phos/TGGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CGATC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGATCCATG	/5Phos/GATCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TCGAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGATCATG	/5Phos/ATCGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TGCAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCATCATG	/5Phos/ATGCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CAACC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAACCCATG	/5Phos/GGTTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GGTTG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTTGCATG	/5Phos/CAACCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
AAGGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGGACATG	/5Phos/TCCTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
AGCTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCTACATG	/5Phos/TAGCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ACACA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACACACATG	/5Phos/TGTGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
AATTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAATTACATG	/5Phos/TAATTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ACGGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACGGTCATG	/5Phos/ACCGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Table A1.2 (cont.)

5-bp barcode	P1.1 adapter sequence	P1.2 adapter sequence
ACTGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGGCATG	/5Phos/CCAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ACTTC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTTCCATG	/5Phos/GAAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ATACG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATACGCATG	/5Phos/CGTATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ATGAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGAGCATG	/5Phos/CTCATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ATTAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATTACCATG	/5Phos/GTAATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CATAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATATCATG	/5Phos/ATATGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CGAAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGAATCATG	/5Phos/ATTCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CGGCT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGCTCATG	/5Phos/AGCCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CGGTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGTACATG	/5Phos/TACCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CGTAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTACCATG	/5Phos/GTACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CGTCG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTCGCATG	/5Phos/CGACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CTGAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGATCATG	/5Phos/ATCAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CTGCG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGCGCATG	/5Phos/CGCAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CTGTC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGTCCATG	/5Phos/GACAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CTTGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGGCATG	/5Phos/CCAAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GACAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACACCATG	/5Phos/GTGTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GAGAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGATCATG	/5Phos/ATCTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GAGTC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGTCCATG	/5Phos/GACTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GCCGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCGTCATG	/5Phos/ACGGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GCTGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTGACATG	/5Phos/TCAGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GGATA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGATACATG	/5Phos/TATCCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GGCCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCCACATG	/5Phos/TGGCCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GGCTC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCTCCATG	/5Phos/GAGCCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GTAGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAGTCATG	/5Phos/ACTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GTCCG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCCGCATG	/5Phos/CGGACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GTCGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCGACATG	/5Phos/TCGACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Table A1.2 (cont.)

5-bp barcode	P1.1 adapter sequence	P1.2 adapter sequence
TACCG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACCGCATG	/5Phos/CGGTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TACGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACGTCATG	/5Phos/ACGTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TAGTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGTACATG	/5Phos/TACTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TATAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTATACCATG	/5Phos/GTATAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TCACG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCACGCATG	/5Phos/CGTGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TCAGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAGTCATG	/5Phos/ACTGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TCCGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCGGCATG	/5Phos/CCGGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TCTGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCTGCCATG	/5Phos/GCAGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TGGAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGAACATG	/5Phos/TTCCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TTACC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTACCCATG	/5Phos/GGTAAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Table A1.3 – PCR first read indexes. Each set of individually barcoded individuals can be further multiplexed by incorporating different Illumina first read indexes during high-fidelity PCR amplification. Although Peterson *et al* provide 12 different Illumina first read indexes, as incorporated into PCR primer 2, in this study we only utilized four. All libraries used the same PCR primer 1 (AATGATACGGCGACCACCGAGATCTAC-ACTCTTTCCCTACACGACG).

First read index	Sequence
ATCACG (1)	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC
CGATGT (2)	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGC
TTAGGC (3)	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGC
CAGATC (7)	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGC

Table A1.4 – Percent missing data and proportion of heterozygous sites per individual. Individuals are shown in order of increasing heterozygosity within assigned clusters (see Table S9 for assignment details). We determined total percent of SNPs missing before ("Raw") and after ("Filtered") excluding SNPs present in >90% of individuals. Two individuals (†) missing more than 70% of raw SNPs, and more than 25% of filtered SNPs, were dropped from subsequent analyses. After filtering, we determined ploidy by calculating the proportion of heterozygous sites in each individual before and after excluding SNPs that reject Hardy-Weinberg equilibrium for heterozygous excess at the $p < 0.01$ level; as well as after applying all quality filters used in the text ("All filters": heterozygous excess removed, thinned to one SNP per RAD locus). Ploidy and sex are indicated using the following superscripts (following specimen ID): (a) haploid, adult male, (b) diploid, adult female, (c) putative diploid, adult male specimens, and (d) putative haploid, male larva. All other specimens are diploid larvae of unknown sex. Notably, even putative haploid males have non-zero heterozygosity (mean value: 0.001). Possible sources of heterozygous sites in haploids include: somatic mutation, sequencing error (current Illumina error rates are ~0.003-0.008), alignment error, contamination, and barcode/index errors.

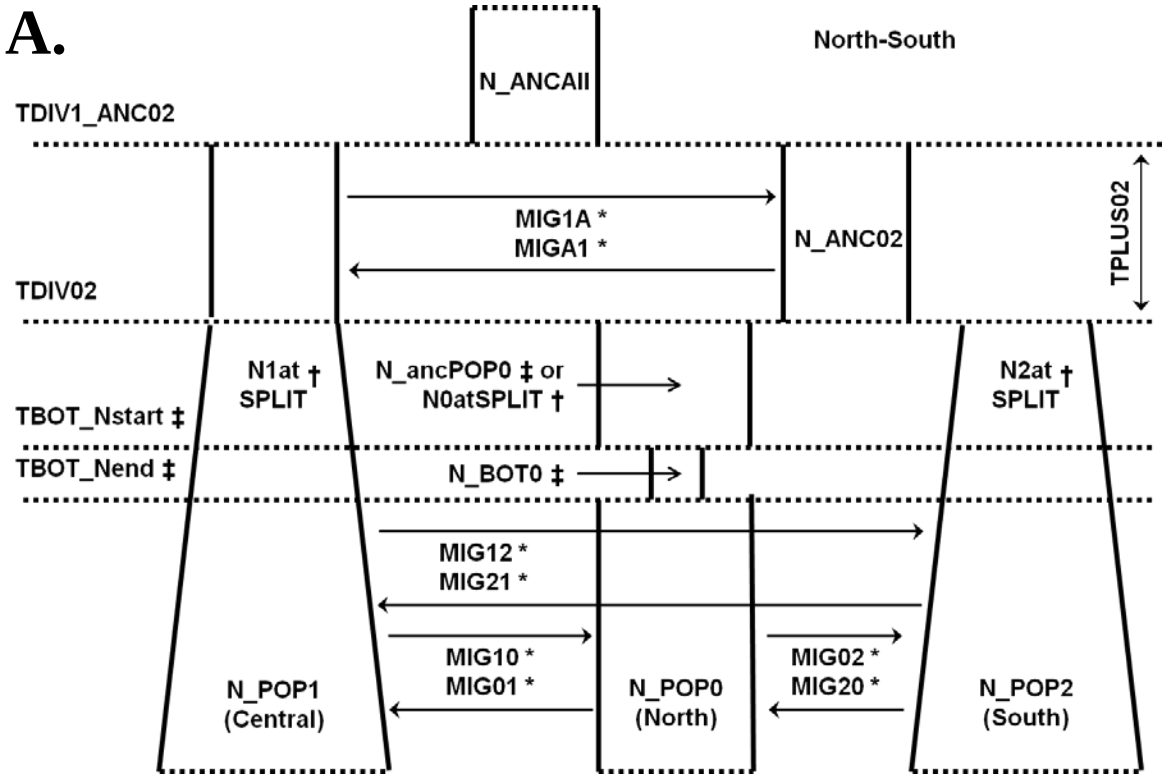
		% Missing SNPs		Proportion heterozygous sites		
ID	**Cluster**	**Raw**	**Filtered**	**Before**	**After**	**All filters**
RB104	North	42.70	0.27	0.019	0.016	0.015
RB096_B	North	39.14	0.37	0.021	0.018	0.016
RB099	North	45.27	0.21	0.022	0.020	0.017
RB094	North	55.24	2.50	0.022	0.020	0.021

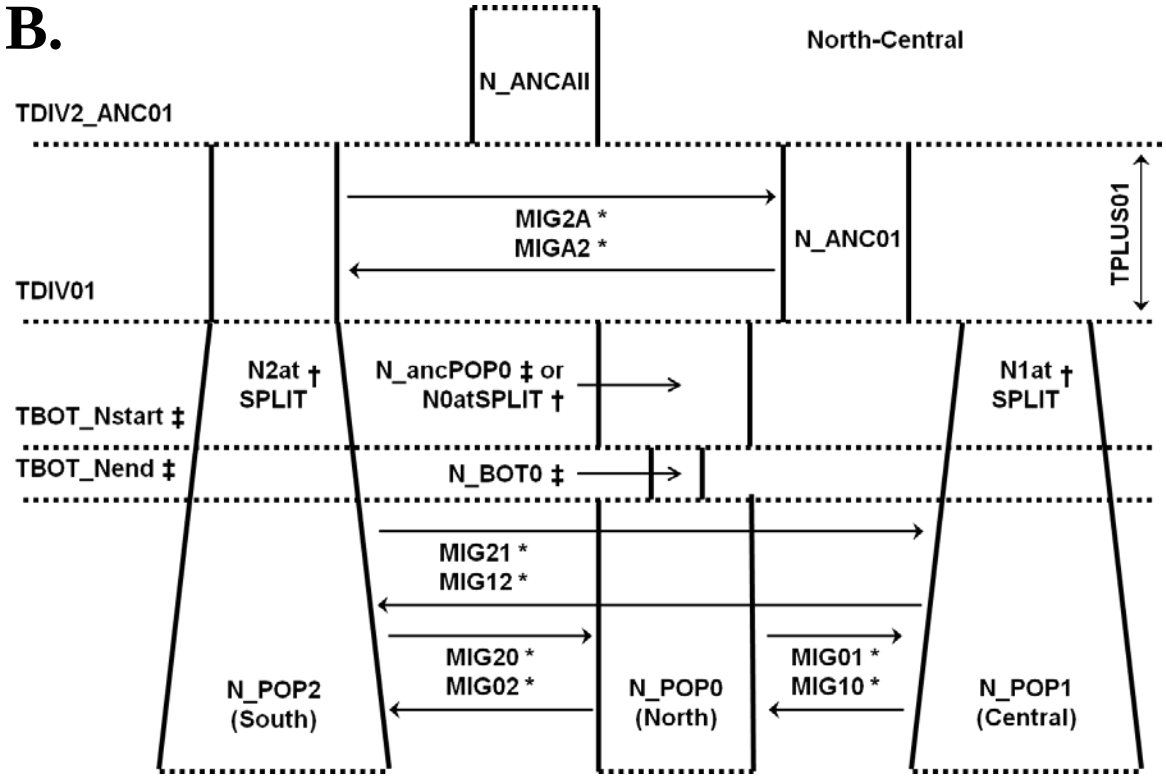
Table A1.4 (cont.)		% Missing SNPs		Proportion heterozygous sites		
ID	**Cluster**	**Raw**	**Filtered**	**Before**	**After**	**All filters**
196-04_E1	North	21.22	0.20	0.026	0.023	0.022
349-02	North	57.83	10.23	0.027	0.025	0.023
RB089_B	North	30.16	0.11	0.030	0.028	0.023
RB004_E1	North	23.93	0.15	0.028	0.025	0.024
RB095	North	45.48	0.11	0.031	0.028	0.025
RB136_B	North	47.41	0.18	0.028	0.026	0.025
RB077	North	58.98	2.33	0.031	0.029	0.026
188-04_D	North	46.42	0.27	0.029	0.027	0.026
RB015_D	North	51.49	0.37	0.030	0.028	0.027
345-02_E1	North	38.93	0.15	0.034	0.031	0.027
RB090.1	North	65.07	13.23	0.034	0.032	0.028
207-04_D	North	49.71	0.12	0.029	0.026	0.028
339-02	North	56.29	6.79	0.035	0.033	0.030
025-0355	North	55.94	0.70	0.037	0.034	0.032
025-0263_D	North	53.44	2.80	0.043	0.040	0.036
168-02	North	49.33	0.30	0.042	0.039	0.037
025-0312_D[†]	North	77.27	36.05	0.043	0.040	0.037
025-0309_D	North	44.28	0.25	0.042	0.040	0.038
177-02	North	42.95	0.15	0.046	0.044	0.040
164-02_D	North	45.74	0.17	0.048	0.046	0.041
RB100	North	44.64	0.23	0.060	0.058	0.055
RB091	North	41.33	0.44	0.078	0.075	0.070
RB026_D[d]	Central	58.23	9.54	0.005	0.002	0.001
RB001_D[a]	Central	64.18	7.42	0.005	0.002	0.002
017-01_E1	Central	44.23	1.21	0.042	0.039	0.036

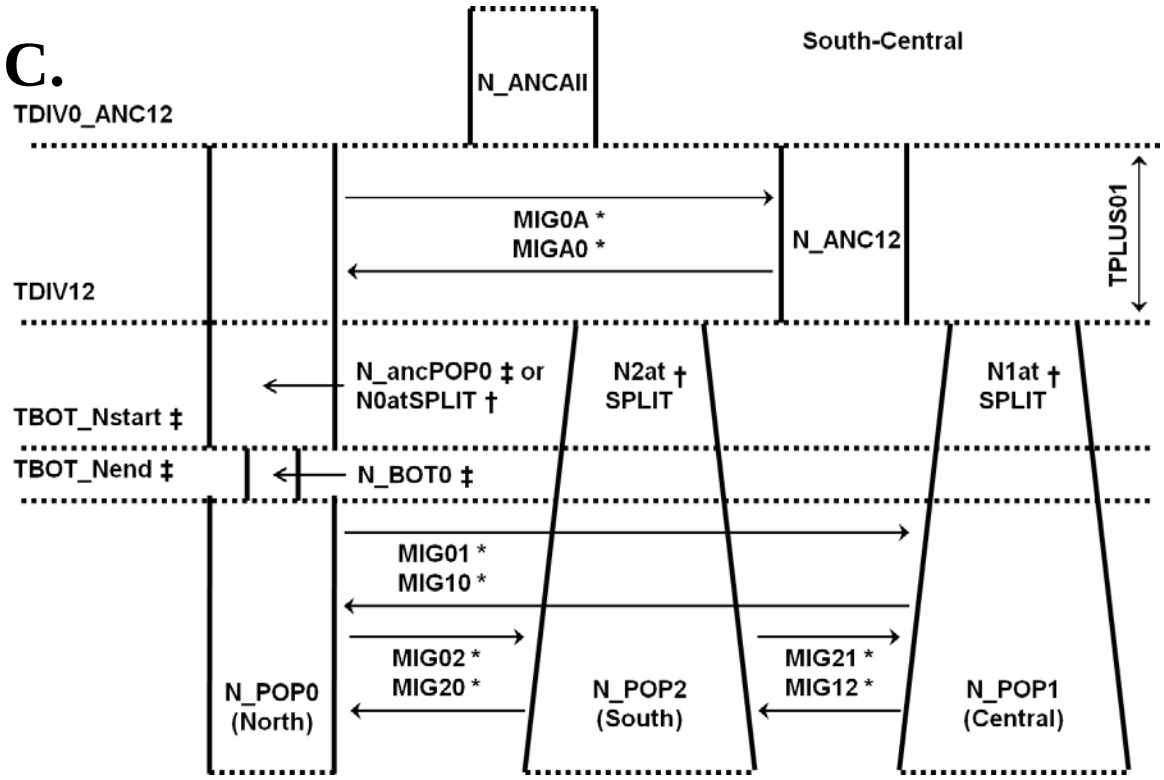
Table A1.4 (cont.)		% Missing SNPs		Proportion heterozygous sites		
ID	**Cluster**	**Raw**	**Filtered**	**Before**	**After**	**All filters**
RB025_D	Central	30.82	1.14	0.052	0.049	0.048
133-04_E1	Central	52.65	1.13	0.071	0.067	0.065
132-04_E1	Central	52.94	1.38	0.081	0.077	0.073
RB120[b]	Central	41.21	1.00	0.085	0.082	0.074
RB018	Central	53.90	5.96	0.083	0.080	0.074
RB132	Central	33.67	0.72	0.087	0.084	0.076
168-04	Central	39.12	0.74	0.090	0.088	0.077
RB112	Central	48.17	0.84	0.088	0.085	0.078
RB125[c]	Central	56.42	1.54	0.085	0.082	0.079
RB129	Central	41.74	0.72	0.090	0.088	0.080
RB133	Central	42.49	0.51	0.097	0.094	0.084
RB042A	Central	40.37	0.46	0.106	0.103	0.092
170-04_D	Central	53.90	2.98	0.105	0.104	0.092
125-02_E1	Central	40.00	0.40	0.107	0.104	0.094
145-04_D	Central	52.76	0.76	0.107	0.105	0.101
003-01	Central	43.47	0.68	0.114	0.112	0.102
RB002_D	Central	43.37	0.34	0.109	0.106	0.102
RB020_D	Central	45.39	0.47	0.116	0.113	0.105
RB107	Central	54.11	0.81	0.118	0.115	0.106
RB110	Central	59.69	3.59	0.119	0.116	0.108
116-04_E1	Central	50.88	0.48	0.129	0.126	0.114
RB118	Central	41.67	0.41	0.131	0.129	0.115
372-02_E1	Central	49.10	0.39	0.123	0.121	0.116
RB119	Central	52.29	0.41	0.131	0.128	0.117
RB124	Central	57.22	8.47	0.130	0.128	0.118

Table A1.4 (cont.)		% Missing SNPs		Proportion heterozygous sites		
ID	**Cluster**	**Raw**	**Filtered**	**Before**	**After**	**All filters**
RB106	Central	54.05	0.77	0.130	0.127	0.118
RB108	Central	46.00	0.24	0.130	0.128	0.118
224-04_D	Central	47.17	0.11	0.132	0.129	0.120
RB028[c]	Central	57.45	3.24	0.151	0.148	0.138
RB017_D[†]	Central	73.55	27.35	0.159	0.156	0.152
RB047b[a]	South	44.25	1.35	0.005	0.001	0.001
180-03_D[d]	South	49.37	1.39	0.005	0.001	0.001
RB008_D[a]	South	39.40	1.95	0.005	0.002	0.002
RB010[a]	South	39.86	0.51	0.007	0.004	0.002
RB137	South	52.93	3.69	0.063	0.061	0.055
RB066B[b]	South	56.43	9.53	0.072	0.069	0.063
185-03_E1	South	47.58	0.46	0.077	0.074	0.068
077-04_D	South	44.99	0.38	0.083	0.080	0.073
097-04_E1	South	54.31	1.10	0.092	0.089	0.078
183-03_D	South	38.35	0.47	0.088	0.085	0.081
075-04	South	43.09	0.52	0.090	0.088	0.081
RB044_B	South	39.36	0.48	0.092	0.089	0.085
RB063B	South	36.86	0.40	0.092	0.089	0.091
088-04_D	South	57.53	4.30	0.104	0.102	0.092
086-04	South	50.52	2.14	0.106	0.104	0.093
174-03A_D	South	38.69	0.40	0.105	0.102	0.094
106-04_D	South	55.63	0.94	0.108	0.106	0.094
RB067	South	54.00	5.72	0.105	0.102	0.097
087-04_D	South	45.39	0.44	0.100	0.097	0.097
RB054	South	33.21	0.30	0.106	0.103	0.097

Table A1.4 (cont.)		% Missing SNPs		Proportion heterozygous sites		
ID	**Cluster**	**Raw**	**Filtered**	**Before**	**After**	**All filters**
RB130	South	59.05	2.50	0.107	0.104	0.101
RB131	South	63.13	7.09	0.113	0.110	0.102
178-03_E1	South	19.37	0.21	0.112	0.109	0.102
RB075_B	South	50.09	0.36	0.116	0.113	0.102
RB071	South	60.30	3.56	0.116	0.113	0.104
102-04_E1	South	41.94	0.51	0.111	0.109	0.104
096-04_E1	South	28.78	0.43	0.118	0.116	0.106
RB027_D	South	45.08	0.30	0.122	0.119	0.111
RB009_D[b]	South	43.32	0.27	0.122	0.119	0.114


A.
North-South
N_ANCAII
TDIV1_ANC02
MIG1A *
MIGA1 *
N_ANC02
TPLUS02
TDIV02
N1at †
SPLIT
N_ancPOP0 ‡ or
N0atSPLIT †
N2at †
SPLIT
TBOT_Nstart ‡
TBOT_Nend ‡
N_BOT0 ‡
MIG12 *
MIG21 *
MIG10 *
MIG01 *
MIG02 *
MIG20 *
N_POP1
(Central)
N_POP0
(North)
N_POP2
(South)


B.
North-Central
N_ANCAII
TDIV2_ANC01
MIG2A *
MIGA2 *
N_ANC01
TPLUS01
TDIV01
N2at †
SPLIT
N_ancPOP0 ‡ or
N0atSPLIT †
N1at †
SPLIT
TBOT_Nstart ‡
TBOT_Nend ‡
N_BOT0 ‡
MIG21 *
MIG12 *
MIG20 *
MIG02 *
MIG01 *
MIG10 *
N_POP2
(South)
N_POP0
(North)
N_POP1
(Central)


C.
South-Central
N_ANCAll
TDIV0_ANC12
MIG0A *
MIGA0 *
N_ANC12
TPLUS01
TDIV12
N_ancPOP0 ‡ or
N0atSPLIT †
N2at †
SPLIT
N1at †
SPLIT
TBOT_Nstart ‡
TBOT_Nend ‡
N_BOT0 ‡
MIG01 *
MIG10 *
MIG02 *
MIG20 *
MIG21 *
MIG12 *
N_POP0
(North)
N_POP2
(South)
N_POP1
(Central)

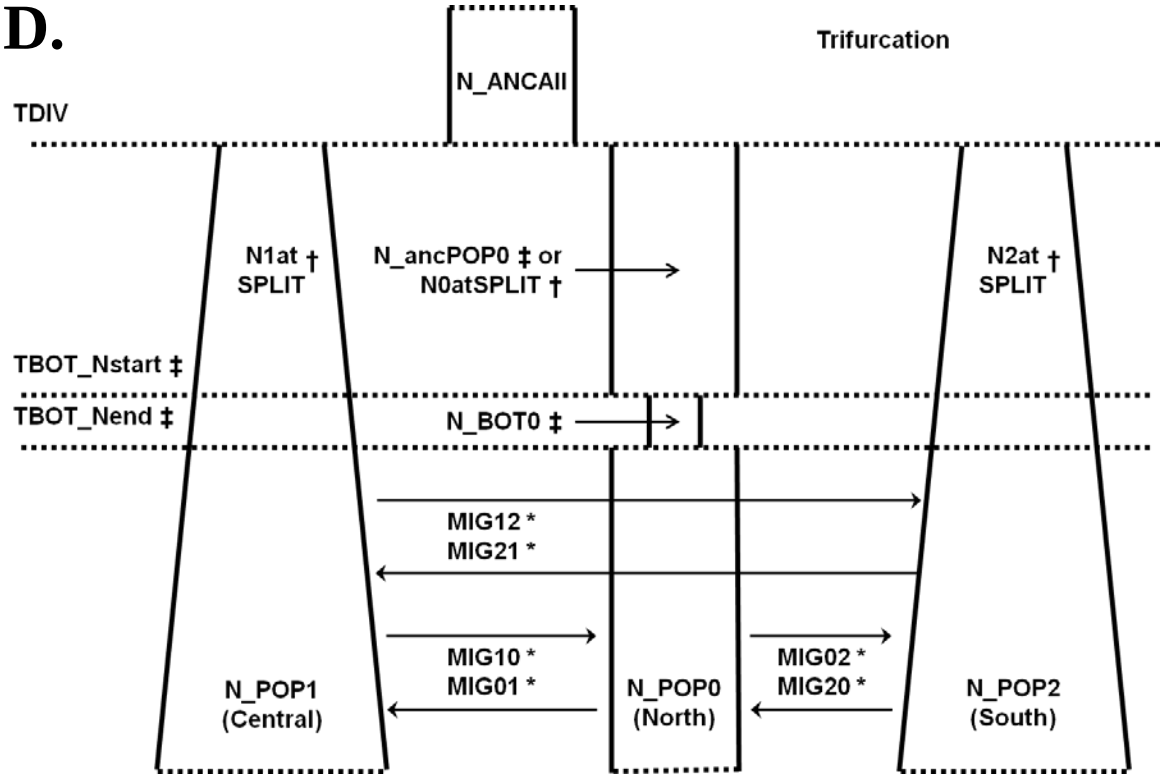

D.
Trifurcation
N_ANCAll
TDIV
N1at †
SPLIT
N_ancPOP0 ‡ or
N0atSPLIT †
N2at †
SPLIT
TBOT_Nstart ‡
TBOT_Nend ‡
N_BOT0 ‡
MIG12 *
MIG21 *
MIG10 *
MIG01 *
MIG02 *
MIG20 *
N_POP1
(Central)
N_POP0
(North)
N_POP2
(South)

Figure A1.1 – Four divergence scenarios tested in FASTSIMCOAL2. The most complex model, which includes a bottleneck in North, is shown here. (A) A bifurcation model in which North and South are sister taxa. (B) A bifurcation model in which North and Central are sister taxa. (C) A bifurcation model in which South and Central are sister. (D) A trifurcation model in which all three populations diverged simultaneously from a common ancestor. Each topology is labeled with the parameters estimated during simulations, using the same names as used in the input files (available on DRYAD) to facilitate comparison. Parameters marked with ‡ are exclusive to models integrating a bottleneck; with † to models integrating exponential population growth; and with * to models allowing migration between populations. Note that in models integrating population growth, but not a bottleneck, the Northern (POP0) lineage would be drawn with an expansion as the others, parameter N_BOT0‡ does not exist, and parameter the N_ancPOP0‡ is replaced with N0atSPLIT†.

Table A1.5 – List of all demographic parameters used in FASTSIMCOAL2 analyses, and their search ranges. Parameter names are as defined in input files available on DRYAD. Each parameter can be stored as either an integer or float, and be pulled from a uniform or log-uniform distribution. Unless noted as bounded, the maximum value of each parameter is soft, and can be exceeded during analyses as needed. Note that only a subset of these parameters was estimated for each model.

Parameter Name	Value Type	Distribution Type	Search Range		Bounded?
			Minimum	Maximum	
N_ANC	Integer	Uniform	10	2.0×10^{6}	No
N_ANCAll	Integer	Uniform	10	2.0×10^{6}	No
N_ANC12	Integer	Uniform	10	2.0×10^{6}	No
N_ANC01	Integer	Uniform	10	2.0×10^{6}	No
N_ANC02	Integer	Uniform	10	2.0×10^{6}	No
N_POP2	Integer	Uniform	10	2.0×10^{6}	No
N_POP1	Integer	Uniform	10	2.0×10^{6}	No
N_ancPOP0	Integer	Uniform	10	2.0×10^{6}	No
N_BOT0	Integer	Uniform	1	1.0×10^{6}	No
N_POP0	Integer	Uniform	10	2.0×10^{6}	No
multTbot	Float	Uniform	1.0×10^{-3}	1	Yes
N0RESIZE	Float	Uniform	1.0×10^{-4}	1	No
N1RESIZE	Float	Uniform	1.0×10^{-4}	1	No
N2RESIZE	Float	Uniform	1.0×10^{-4}	1	No
TDIV	Integer	Uniform	10	2.0×10^{5}	No
TDIV12	Integer	Uniform	10	8.0×10^{4}	No
TDIV02	Integer	Uniform	10	8.0×10^{4}	No
TDIV01	Integer	Uniform	10	8.0×10^{4}	No
TPLUS01	Integer	Uniform	10	8.0×10^{4}	No
TPLUS02	Integer	Uniform	10	8.0×10^{4}	No
NM01	Float	Log-Uniform	1.0×10^{-5}	5	Yes
NM10	Float	Log-Uniform	1.0×10^{-5}	5	Yes
NM02	Float	Log-Uniform	1.0×10^{-5}	5	Yes
NM20	Float	Log-Uniform	1.0×10^{-5}	5	Yes
NM12	Float	Log-Uniform	1.0×10^{-5}	5	Yes
NM21	Float	Log-Uniform	1.0×10^{-5}	5	Yes
NM0A	Float	Log-Uniform	1.0×10^{-5}	5	Yes
NMA0	Float	Log-Uniform	1.0×10^{-5}	5	Yes
NM1A	Float	Log-Uniform	1.0×10^{-5}	5	Yes
NMA1	Float	Log-Uniform	1.0×10^{-5}	5	Yes
NM2A	Float	Log-Uniform	1.0×10^{-5}	5	Yes
NMA2	Float	Log-Uniform	1.0×10^{-5}	5	Yes

Table A1.6 – Summary of SNP recovery under different data filters and cluster assignment cutoffs. Range-wide datasets were used to estimate population structure in ADMIXTURE and *adegenet*. The North, Central, and South datasets were used for estimation of IBD and IBE, and were first filtered to include only individuals assigned to that cluster. All datasets include a filter removing SNPs violating Hardy-Weinberg equilibrium for heterozygous excess ($p < 0.01$).

	Range-wide	**North**	**Central**	**South**
Data filtering				
≤10% missing data; one SNP per RAD locus	5474	2160	4879	3962
ADMIXTURE assignment cutoffs				
Exclude <70% ADMIXTURE assignments	n/a	2160	5009	4052
Exclude <90% ADMIXTURE assignments	n/a	2160	6233	3993

Table A1.7 – Summary of SNP recovery under different data filters and cluster assignment cutoffs used for model choice and parameter estimation in FASTSIMCOAL2. All datasets include a filter removing SNPs violating Hardy-Weinberg equilibrium for heterozygous excess ($p < 0.01$). To build the site frequency spectrum (SFS) we resampled blocks of data without missing data across individuals (each block corresponding to a RAD locus). For the model choice we resampled three individuals per population to generate the three population joint-SFS (3D SFS), and for the parameter estimation we resampled seven individuals per population to generate the three pairwise two population SFS (2D SFS, see Chapter 2 Material and Methods for details).

	Dataset for model choice				Dataset for parameter estimation			
	Range-wide	**North**	**Central**	**South**	**Range-wide**	**North**	**Central**	**South**
Data filtering								
≤50% missing data; all SNPs	11617	3182	7458	6596	15230	3932	10938	8971
≤10% missing data; all SNPs	6668	1897	4266	3826	8924	2331	6443	5328
≤50% missing data; one SNP per RAD locus	4478	1256	2840	2444	n/a	n/a	n/a	n/a
≤10% missing data; one SNP per RAD locus	2393	652	1517	1342	n/a	n/a	n/a	n/a
ADMIXTURE assignment cutoffs								
Exclude <90% ADMIXTURE assignments	11538	3201	7451	6240	15058	3957	10677	8511
Exclude <90% ADMIXTURE assignments; one SNP per RAD locus	4459	1233	2803	2373	n/a	n/a	n/a	n/a

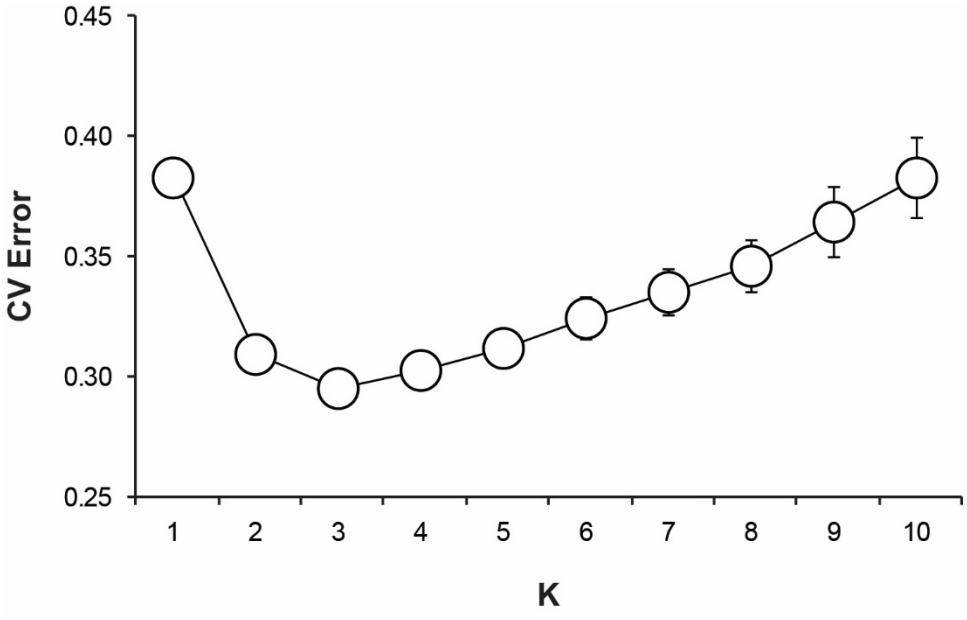


Figure A1.2 – Average CV error scores (+/- standard deviation) for each *K* across 100 independent ADMIXTURE runs. In every run, CV scores indicated that $K = 3$ is the optimal number of clusters.

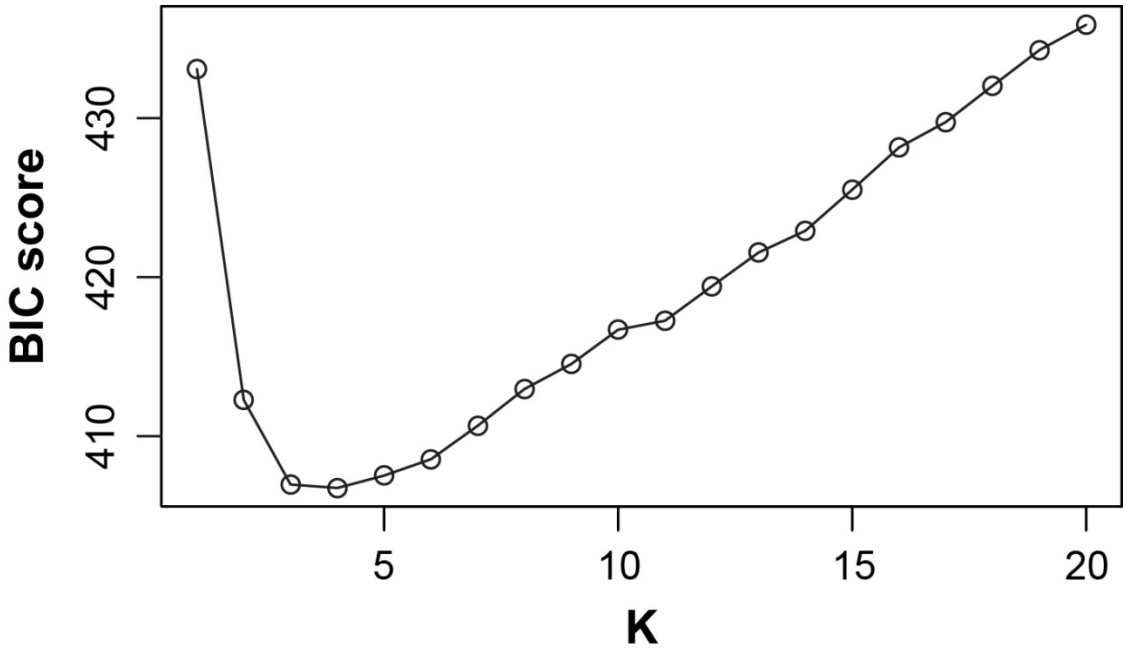


Figure A1.3 – BIC plot for DAPC. *K* = 3 and *K* = 4 achieve nearly the same BIC scores.

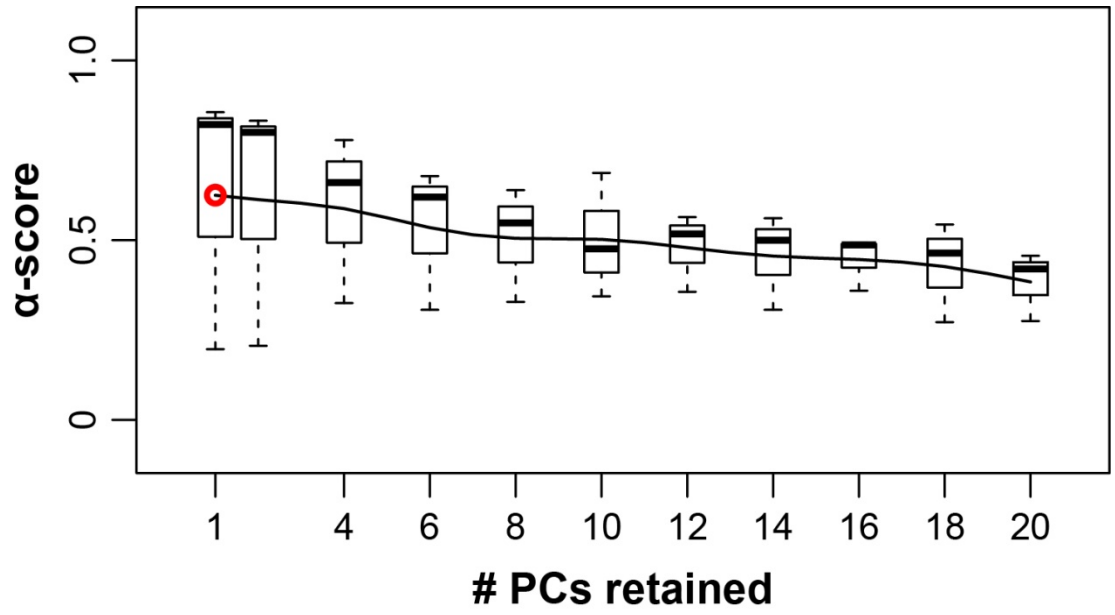


Figure A1.4 – Plots of α-score over 20 PCs, with spline interpolation. The optimal number of PCs to retain is 1.

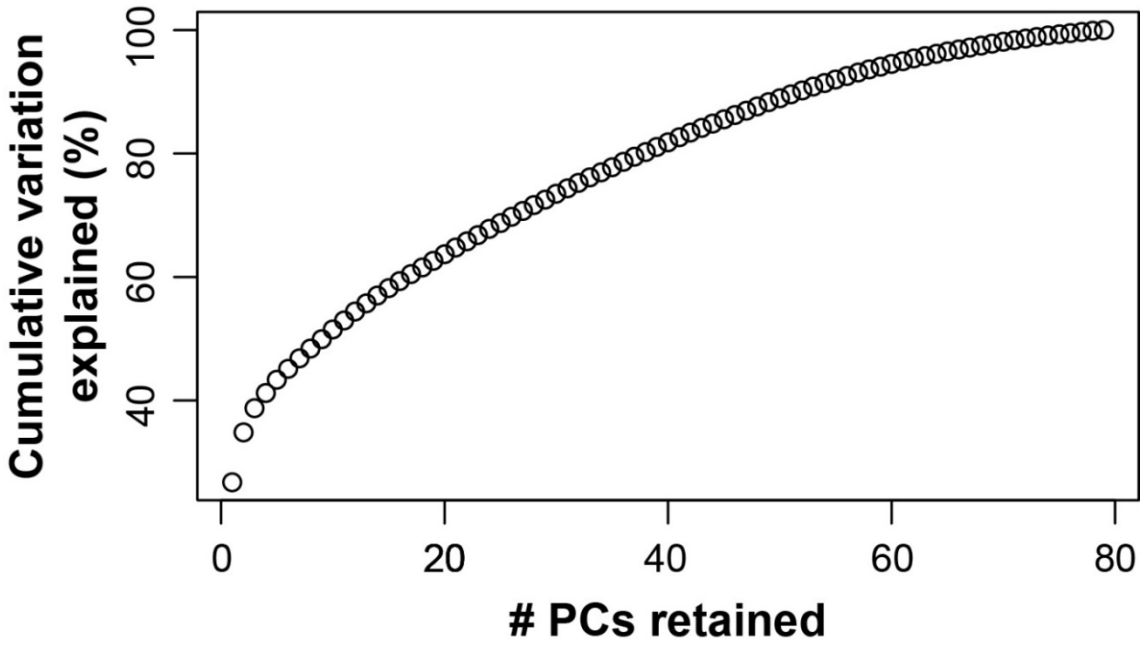


Figure A1.5 – Cumulative percent variance explained as a function of the number of retained principal components.

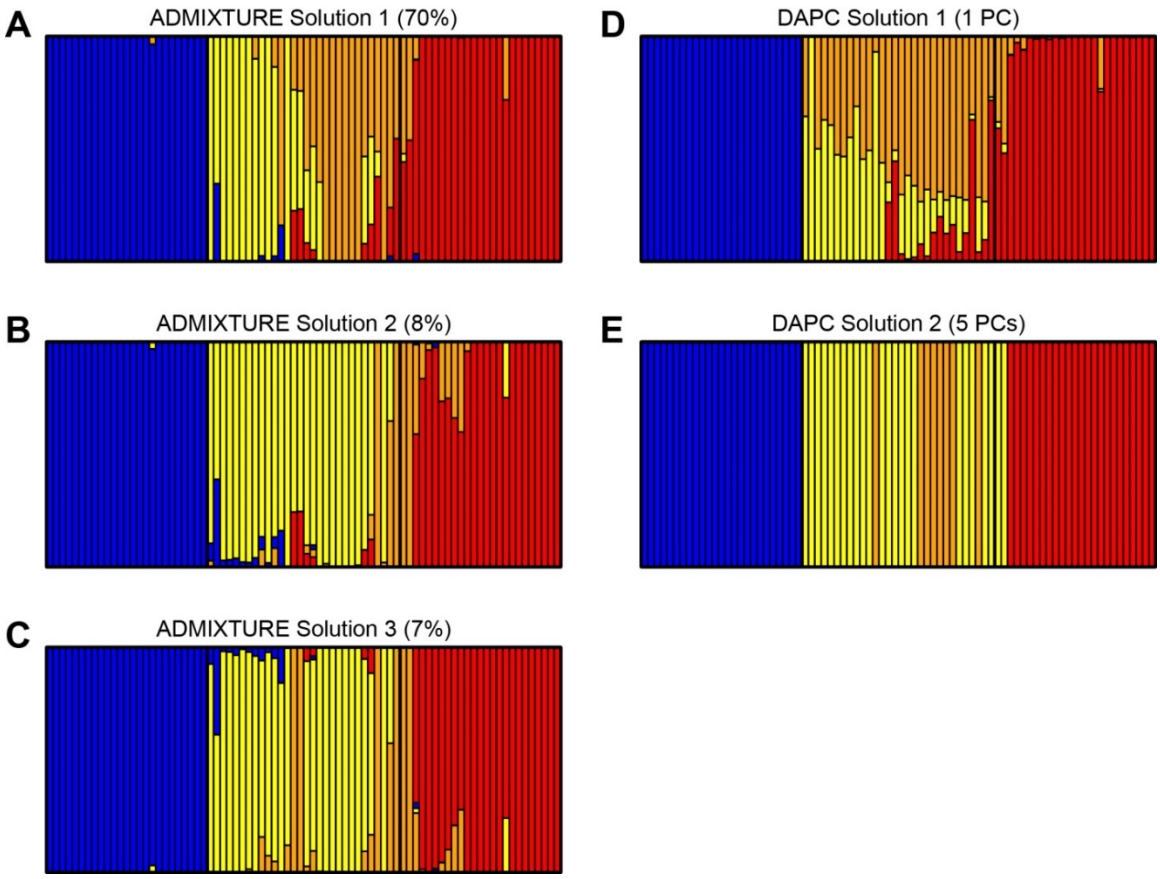


Figure A1.6 – Ancestry proportions (ADMIXTURE) and assignment probabilities (DAPC) for *K* = 4. To facilitate comparisons across solutions, individuals are displayed in the same order, and grouped into three geographical areas, as in Figure 2.1B (northernmost populations are to the left; southernmost to the right). In the majority of runs (70%), ADMIXTURE breaks Central individuals into a northern group and a southern group (A). Two additional ADMIXTURE solutions (8%, 7%) assign a small handful of individuals from Central and South into a fourth cluster (B, C). Using 1 PC, DAPC indicates most Central individuals as being admixed with a fourth population (D). Using 5 PCs, Central is again broken into two groups in a similar pattern to ADMIXTURE solution 1(E).

Table A1.8 – Pairwise matrix similarity statistics (G′) for *K*=3. Values were calculated using assignment matrices summarized across 100 ADMIXTURE runs and for 1-10 PCs included in DAPC.

	ADMIX	DAPC-1	DAPC-2	DAPC-3	DAPC-4	DAPC-5	DAPC-6	DAPC-7	DAPC-8	DAPC-9	DAPC-10
ADMIX	1.00	0.91	0.87	0.88	0.88	0.85	0.85	0.85	0.85	0.85	0.85
DAPC-1	0.91	1.00	0.93	0.92	0.91	0.90	0.90	0.90	0.90	0.90	0.90
DAPC-2	0.87	0.93	1.00	0.97	0.95	0.96	0.96	0.96	0.96	0.96	0.96
DAPC-3	0.88	0.92	0.97	1.00	0.98	0.96	0.96	0.96	0.96	0.96	0.96
DAPC-4	0.88	0.91	0.95	0.98	1.00	0.94	0.94	0.94	0.94	0.94	0.94
DAPC-5	0.85	0.90	0.96	0.96	0.94	1.00	1.00	1.00	1.00	1.00	1.00
DAPC-6	0.85	0.90	0.96	0.96	0.94	1.00	1.00	1.00	1.00	1.00	1.00
DAPC-7	0.85	0.90	0.96	0.96	0.94	1.00	1.00	1.00	1.00	1.00	1.00
DAPC-8	0.85	0.90	0.96	0.96	0.94	1.00	1.00	1.00	1.00	1.00	1.00
DAPC-9	0.85	0.90	0.96	0.96	0.94	1.00	1.00	1.00	1.00	1.00	1.00
DAPC-10	0.85	0.90	0.96	0.96	0.94	1.00	1.00	1.00	1.00	1.00	1.00

Table A1.9 – Sampling location, cluster assignment, ancestry proportions (averaged across 100 ADMIXTURE runs), and assignment probabilities (DAPC, 1 PC) for 80 *N. lecontei* individuals. In all but four cases (*), both methods agreed on cluster assignment ("Cluster Assigned"). When ADMIXTURE and DAPC results were in conflict, assignment was determined on the basis of sampling location. To account for uncertainty in population assignment, we repeated analyses excluding individuals with: (1) maximal ADMIXTURE ancestry proportion <70% (*, †); and (2) maximal ADMIXTURE ancestry proportion <90% (*, †, ‡).

			ADMIXTURE			DAPC		
Specimen	**Location**	**Cluster Assigned**	**North**	**Central**	**South**	**North**	**Central**	**South**
003-01	Hopkinton, MA	Central	0.02	0.98	0	0	1	0
017-01_E1	Malletts Bay, VT	Central	0.08	0.92	0	0	1	0
025-0263_D	Kemptville, ON	North	1	0	0	1	0	0
025-0309_D	Tweed, ON	North	1	0	0	1	0	0
025-0355	Chalk River, ON	North	1	0	0	1	0	0
075-04	Canoe Creek, FL	South	0	0	1	0	0	1
077-04_D	Palmdale, FL	South	0	0	1	0	0	1
086-04	Gainesville, FL	South	0	0	1	0	0	1
087-04_D	Gainesville, FL	South	0	0	1	0	0	1
088-04_D	Gainesville, FL	South	0	0	1	0	0	1
096-04_E1	Morgan, GA	South	0	0	1	0	0.01	0.99
097-04_E1	Morgan, GA	South	0	0	1	0	0.03	0.97
102-04_E1	Sylvester, GA	South	0	0	1	0	0.01	0.99
106-04_D	Vienna, GA	South	0	0	1	0	0.02	0.98

			ADMIXTURE			DAPC		
Specimen	**Location**	**Cluster Assigned**	**North**	**Central**	**South**	**North**	**Central**	**South**
116-04_E1‡	Crossville, TN	Central	0	0.82	0.18	0	0.98	0.02
125-02_E1	Nottingham, NH	Central	0.04	0.96	0	0	1	0
132-04_E1‡	Trappe, MD	Central	0	0.75	0.25	0	0.83	0.17
133-04_E1‡	Trappe, MD	Central	0	0.76	0.24	0	0.92	0.08
145-04_D	Ossippee, NH	Central	0.02	0.98	0	0	1	0
164-02_D	Bancroft, ON	North	1	0	0	1	0	0
168-02	Apsley, ON	North	1	0	0	1	0	0
168-04	Ossippee, NH	Central	0.03	0.97	0	0	1.00	0.00
170-04_D*	Selmer, TN	South	0	0.35	0.65	0	0.57	0.43
174-03A_D	Dixie Co, FL	South	0	0	1	0	0	1
177-02	Sebrite, ON	North	1	0	0	1	0	0
178-03_E1	Crystal Lake, FL	South	0	0	1	0	0.01	0.99
183-03_D	Glades Co, FL	South	0	0	1	0	0	1
185-03_E1	Palmdale, FL	South	0	0	1	0	0	1
188-04_D	Eau Claire, WI	North	1	0	0	1	0	0
196-04_E1	Sparta, WI	North	1	0	0	1	0	0
207-04_D	Park Falls, WI	North	1	0	0	1	0	0
224-04_D†	Amelia, VA	Central	0	0.54	0.46	0	0.67	0.33
339-02	North Bay, ON	North	1	0	0	1	0	0
345-02_E1	North Bay, ON	North	1	0	0	1	0	0
349-02	Mattawan, ON	North	1	0	0	1	0	0
372-02_E1	Plymouth, MA	Central	0.02	0.98	0	0	1	0
RB002_D	Morehead, KY	Central	0	1	0	0	1	0

			ADMIXTURE			DAPC		
Specimen	**Location**	**Cluster Assigned**	**North**	**Central**	**South**	**North**	**Central**	**South**
RB004_E1	Wisconsin Rapids, WI	North	1	0	0	1	0	0
RB009_D‡	Dry Branch, GA	South	0.03	0.11	0.87	0	0.20	0.80
RB015_D	Springdale Township, MI	North	1	0	0	1	0	0
RB018†	Dorchester, NH	Central	0.39	0.61	0	0	1	0
RB020_D	London, KY	Central	0	1	0	0	0.99	0.01
RB025_D	Jay Township, PA	Central	0.04	0.96	0	0	1	0
RB027_D*	Auburn, GA	South	0	0.47	0.53	0	0.71	0.29
RB028‡	Pine Barrens, NJ	Central	0.16	0.84	0	0	1	0
RB042A	Lexington, KY	Central	0	1	0	0	0.97	0.03
RB044_B	Ruskin, FL	South	0	0	1	0	0	1
RB054	Fort McCoy, FL	South	0	0	1	0	0	1
RB063B	Fort McCoy, FL	South	0	0	1	0	0	1
RB066B*	Chiefland, FL	South	0	0.23	0.77	0	0.51	0.49
RB067	Chiefland, FL	South	0	0	1	0	0.01	0.99
RB071*	Eutaw, AL	South	0.01	0.36	0.63	0	0.80	0.20
RB075_B	Bluffton, SC	South	0	0	1	0	0.01	0.99
RB077	Bitely, MI	North	1	0	0	1	0	0
RB089_B	Bitely, MI	North	1	0	0	1	0	0
RB090.1	Grayling, MI	North	1	0	0	1	0	0
RB091	Grayling, MI	North	0.97	0.03	0	1	0	0
RB094	Glaque Beach, MI	North	1	0	0	1	0	0
RB095	Naubinway, MI	North	1	0	0	1	0	0
RB096_B	Naubinway, MI	North	1	0	0	1	0	0
RB099	Manistique, MI	North	1	0	0	1	0	0

			ADMIXTURE			DAPC		
Specimen	**Location**	**Cluster Assigned**	**North**	**Central**	**South**	**North**	**Central**	**South**
RB100	Watersmeet, WI	North	1	0	0	1	0	0
RB104	Suring, WI	North	1	0	0	1	0	0
RB106	Clifton Forge, VA	Central	0	0.93	0.07	0	0.99	0.01
RB107	Mountain Grove, VA	Central	0	1	0	0	1	0
RB108	Deer Run, WV	Central	0	0.94	0.06	0	0.99	0.01
RB110	Brown Mills, NJ	Central	0.02	0.98	0	0	1	0
RB112	Tuckerton, NJ	Central	0	1	0	0	1	0
RB118	Brandywine, WV	Central	0.02	0.95	0.04	0	1	0
RB119‡	Buena Vista, VA	Central	0	0.85	0.15	0	0.97	0.03
RB120	Georgetown, KY	Central	0	1	0	0	0.98	0.02
RB124	Scotch Plains, NJ	Central	0.06	0.94	0	0	1	0
RB125	Sideling Hill, MD	Central	0.05	0.95	0	0	1	0
RB129	Lexington, KY	Central	0	1	0	0	0.95	0.05
RB130	Bluffton, SC	South	0	0	1	0	0.06	0.94
RB131	Bluffton, SC	South	0	0	1	0	0.14	0.86
RB132	Lexington, KY	Central	0	1	0	0	0.97	0.03
RB133	Lexington, KY	Central	0	1	0	0	0.96	0.04
RB136_B	Bitely, MI	North	1	0	0	1	0	0
RB137	Bowling Green, FL	South	0	0	1	0	0.01	0.99

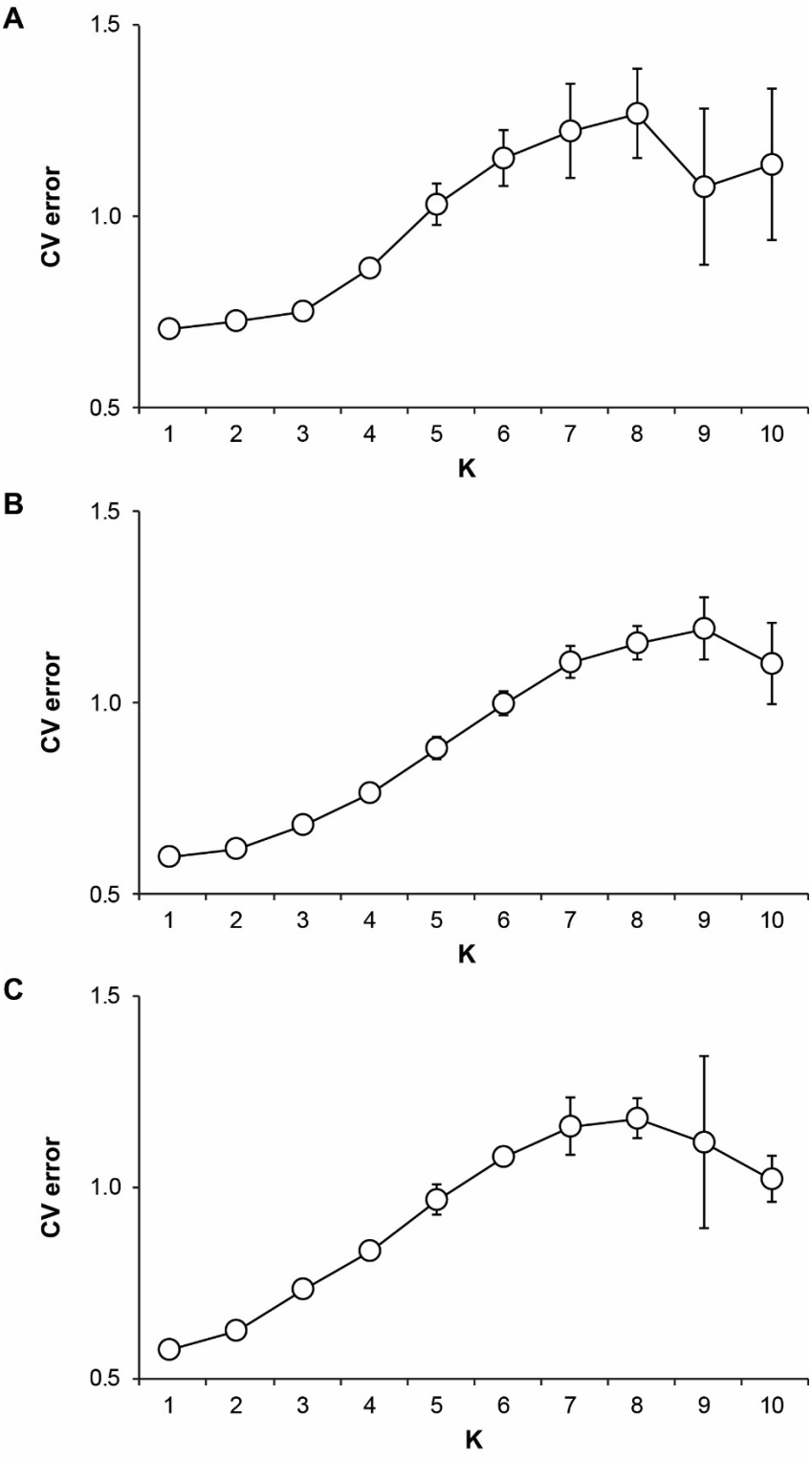

A
B
C
CV error
K
1.5
1.0
0.5
1
2
3
4
5
6
7
8
9
10

Figure A1.7 – Average CV error scores (+/- standard deviation) for each *K* across 10 independent ADMIXTURE runs for North (A), South (B), and Central (C) clusters. In all clusters, all 10 runs selected $K = 1$ as the optimal number of clusters, which is consistent with a lack of hierarchical structure within clusters.

Table A1.10 – Genetic diversity summary statistics for each population, considering only individuals that assign with >90% ADMIXTURE probability to a given cluster (Table A1.7). Data were filtered as described in the text [10x coverage, 90% completeness, and a Hardy-Weinberg equilibrium filter excluding sites with heterozygote excess (p-value ≤ 0.01); total number of markers = 5263]. Genome-wide averages of observed heterozygosities (H_o) and inbreeding coefficients (F_{IS}) were calculated using polymorphic loci only. Expected heterozygosities (H_e) for each population were calculated using loci that were polymorphic in any of the populations ("all"); and that were polymorphic within regions ("region"). Removing admixed individuals has no impact on overall patterns of genetic diversity among the regions.

Population	Polymorphic sites	Private alleles	H_e (all)	H_e (region)	H_o (region)	F_{IS}
North	1179	343	0.055	0.234	0.157	0.248
Central	3746	1842	0.149	0.205	0.147	0.210
South	2830	1153	0.115	0.214	0.188	0.085

Figure A1.8 – Comparison of mid-instar head capsule coloration. (A) Larvae from South retain a dark head capsule until the third or fourth larval instar. The head capsules of larvae from North (B) and Central (C), however, typically become red shortly after hatching.

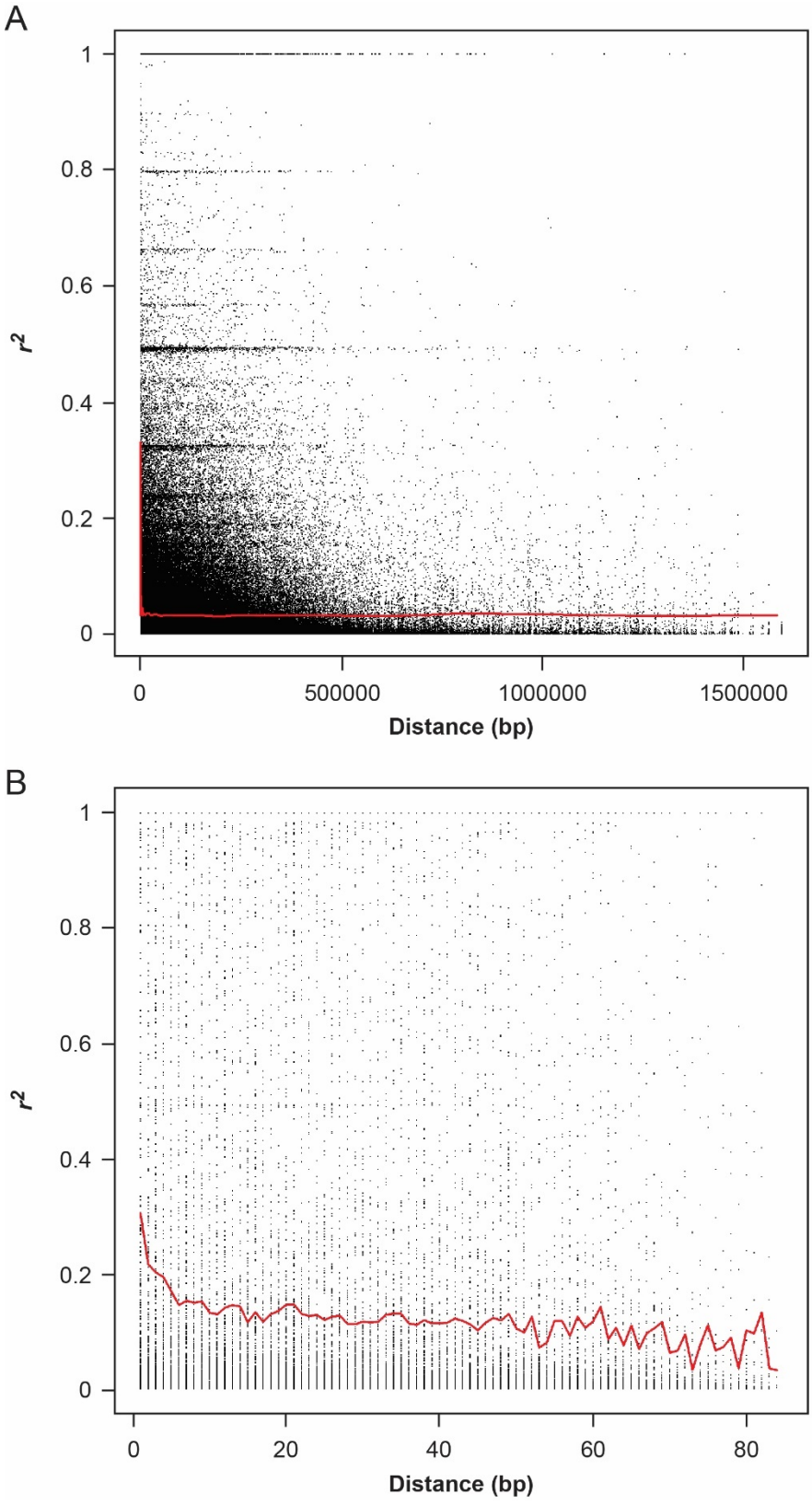

A
1
0.8
0.6
0.4
0.2
0
r^2
0
500000
1000000
1500000
Distance (bp)
B
1
0.8
0.6
0.4
0.2
0
r^2
0
20
40
60
80
Distance (bp)

Figure A1.9 – Correlation among genotypes (r^2) as an approximation to the r^2 based on phased data (as implemented in VCFTOOLS). Prior to constructing SFS, we examined patterns of linkage in our data in order to determine the distance among pairs of SNPs at which the LD is close to zero, and hence can be considered independent. The LD was computed using unphased genotypes and by pooling all individuals from the different clusters together. As such, these estimates will reflect LD caused by both physical linkage and population structure. **A.** LD decays quickly to near-zero when comparing SNPs in different RAD loci within the same scaffold, suggesting that different RAD loci can be considered statistically independent. **B.** Within RAD loci, the LD decays to values lower than 0.2 very quickly, suggesting there is no evidence of large LD blocks. Each point corresponds to a pair of SNPs, and the solid red line corresponds to the mean of pairwise comparisons at a given distance. Based on these results, we assumed that SNPs on different RAD loci could be considered independent

Table A1.11 – Demographic parameters inferred under each 2-population asymmetrical migration, bifurcation model. The maximum-likelihood point estimate for parameters is taken from the run reaching the highest composite likelihood. Migration rates were calculated using current population sizes, and are given forward in time. Estimates of divergence time assume three generations per year (Benjamin 1955; Wilson *et al.* 1992). The divergence date for the North-Central bifurcation model is the most recent (~29,000), and the divergence times of North-South and South-Central are similar (~50,000) supporting a (North, Central), South topology.

NORTH-CENTRAL	
Parameter	ML Estimate
North N_e	39711
Central N_e	463614
Ancestral N_e	456480
North/Central divergence time	29189
2Nm (Central to North)	0.15
2Nm (North to Central)	1.01

NORTH-SOUTH	
Parameter	ML Estimate
North N_e	51891
South N_e	342432
Ancestral N_e	411968
North/South divergence time	49959
2Nm (South to North)	0.14
2Nm (North to South)	0.20

SOUTH-CENTRAL	
Parameter	ML Estimate
Central N_e	410202
South N_e	206874
Ancestral N_e	435560
South/Central divergence time	49796
2Nm (South to Central)	1.65
2Nm (Central to South)	1.00

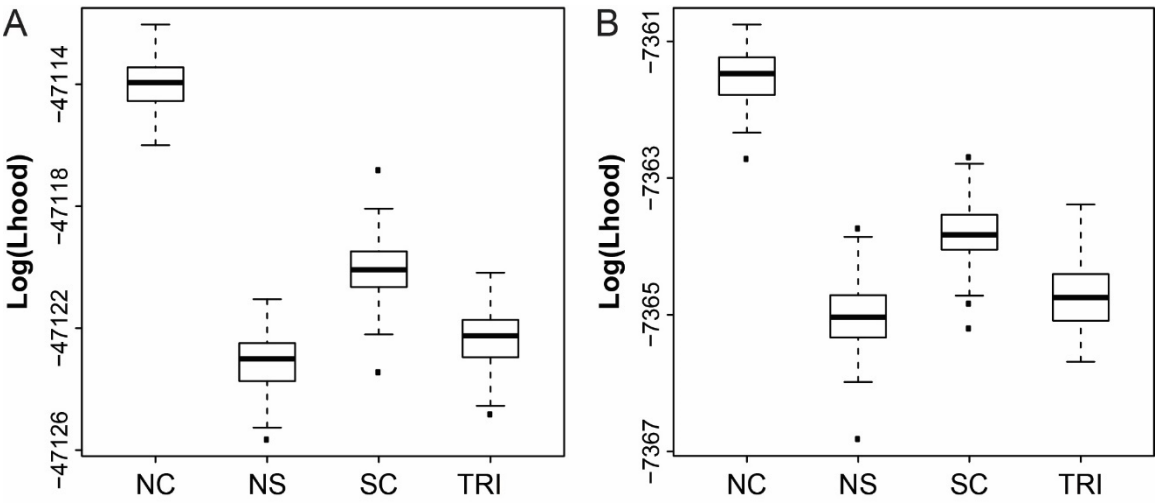


Figure A1.10 – Comparison of the distribution of Log10-likelihood of different topologies for models with asymmetric migration. Plot A shows the Log-likelihood values computed for the 3D-jointSFS, considering all SNPs (11,617 SNPs), and accounting for the number of monomorphic sites across the three populations). Plot B shows the Log-likelihood values considering only 1 SNP per RAD locus (4,478 SNPs). Distributions were obtained from 100 expected SFS approximated with 2×10^5 coalescent simulations with the parameters that maximized the likelihood for each model. NC – (North, Central), South topology; NS – (North, South), Central topology; CS – (Central, South), North topology; TRI – trifurcation topology. Across all analyses, the [(North, Central), South] topology is always favored.

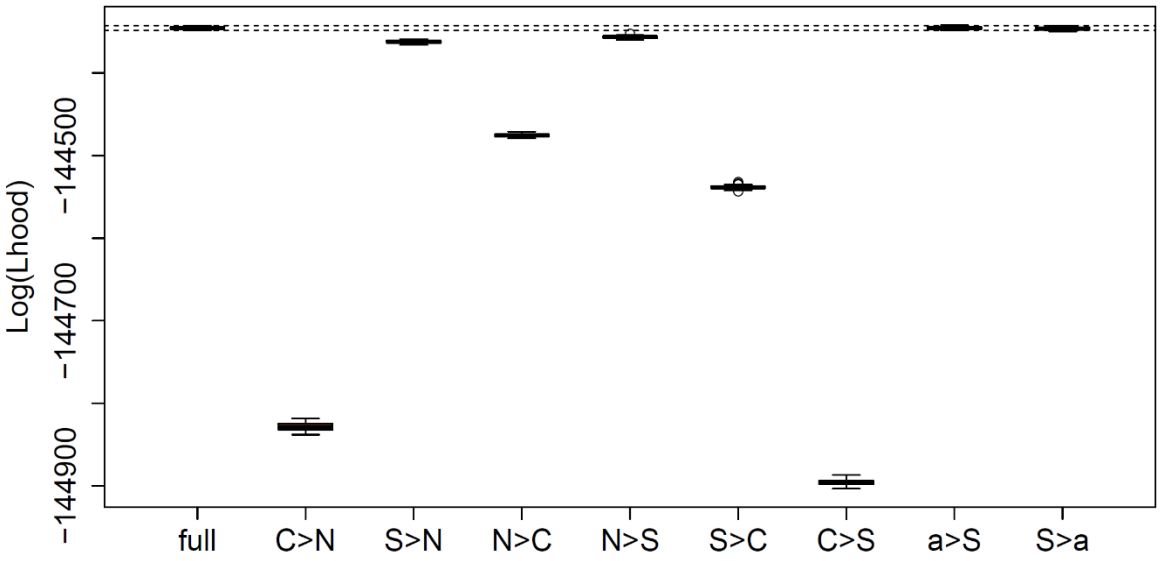


Figure A1.11 – Effect of migration rates on the likelihood. Comparison of the distribution of Log-likelihood for the best model [full migration and the (North, Central), South topology] with corresponding migration rates that maximized the likelihood (full), with the Log-likelihoods obtained by setting one of the migration rates to zero. Migration rates are shown forward in time, as coded as source>sink, where source and sink can correspond to N (North), C (Central), S (South), and a (ancestral population of N and C). Distributions were obtained from 100 expected SFS approximated with 2×105 coalescent simulations. Log-likelihoods were computed accounting for the number of monomorphic sites across the three populations. Dashed lines correspond to the minimum and maximum Log-likelihood values obtained with the full model. Setting most of the migration rates to zero affects the fit of the data, leading to lower likelihoods than the ones obtained with the full model. The only exception are the migration rates involving the ancestral population (S>a and a>S) which, when set to zero, do not lead to lower likelihoods, suggesting that migration between the ancestral populations was limited. Moreover, the migration N>S and S>N do not affect much the likelihood, also suggesting reduced migration between these populations.

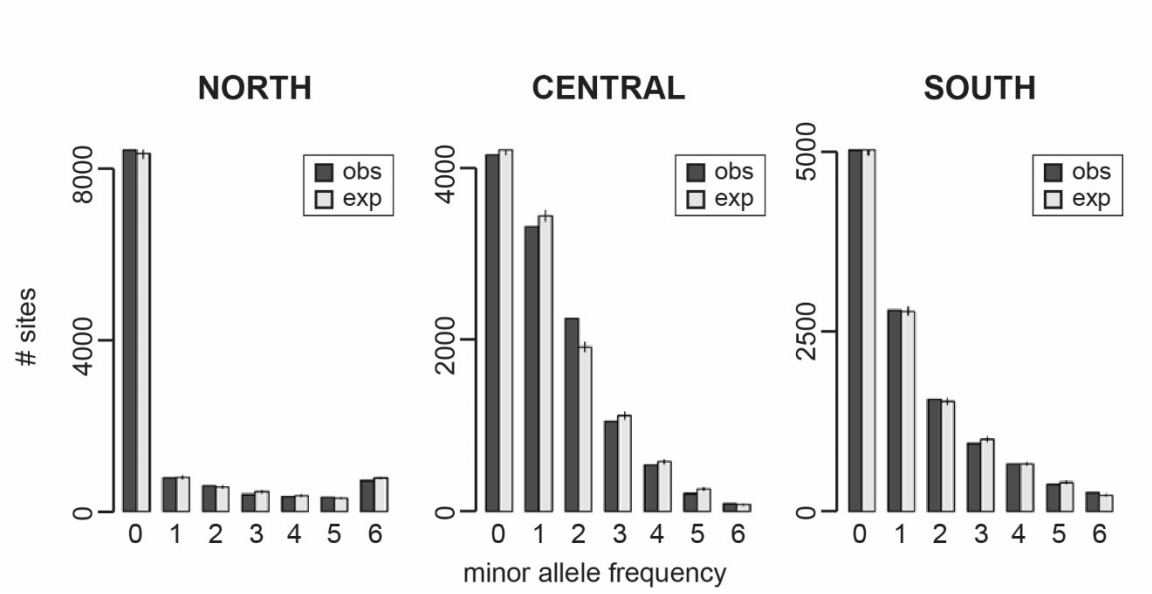


Figure A1.12 – Fit of the marginal 1D expected SFS obtained for the model with the highest relative likelihood [i.e., the model with full migration and the (North, Central), South topology]. (A) North marginal 1D-SFS; (B) Central marginal 1D-SFS; (C) South marginal 1D-SFS. The marginal SFS is obtained by summing all the entries of the joint-SFS with a given frequency in one population and discarding the monomorphic sites across all samples. The expected SFS was obtained as the mean of 100 expected SFS simulated according to the parameters that maximized the likelihood (2×105 coalescent simulations), and then multiplied by the total number of SNPs to be in the same scale (SNP counts) as the observed SFS. Vertical error bars correspond to the range of values obtained across the 100 simulated expected SFS. Overall there is a very good fit for the marginal SFS for the North and South samples. In the Central the fit is slightly worse than for the other populations.

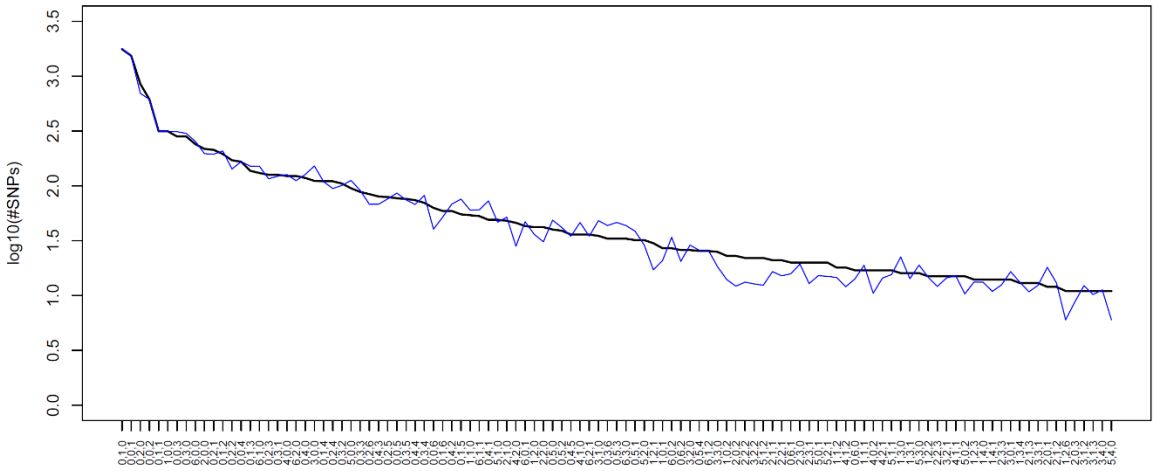


Figure A1.13 – Fit of the expected 3D joint-SFS obtained for the model with the highest relative likelihood [i.e., the model with full migration and the (North, Central), South topology]. Only entries with more than ten SNPs are shown, as all the entries with less than six SNPs were pooled together. Entries are coded in the x-axis as *n*, *c*, *s* where *n* corresponds to the minor allele frequency in the North sample, *c* corresponds to the minor allele frequency in the Central sample, and *s* corresponds to the minor allele frequency in the South sample. The solid black line represents the observed SFS, and the solid blue line represents the expected SFS. The expected SFS was obtained by taking the average of 100 expected SFS simulated according to the parameters that maximized the likelihood (2×10^5 coalescent simulations), and then multiplying by the total number of SNPs to be on the same scale (SNP counts) as the observed SFS. The sites that are monomorphic across all samples were discarded. Overall, there is a good fit for the entries with more SNPs, whereas the fit gets poorer for the entries with only a few SNPs.

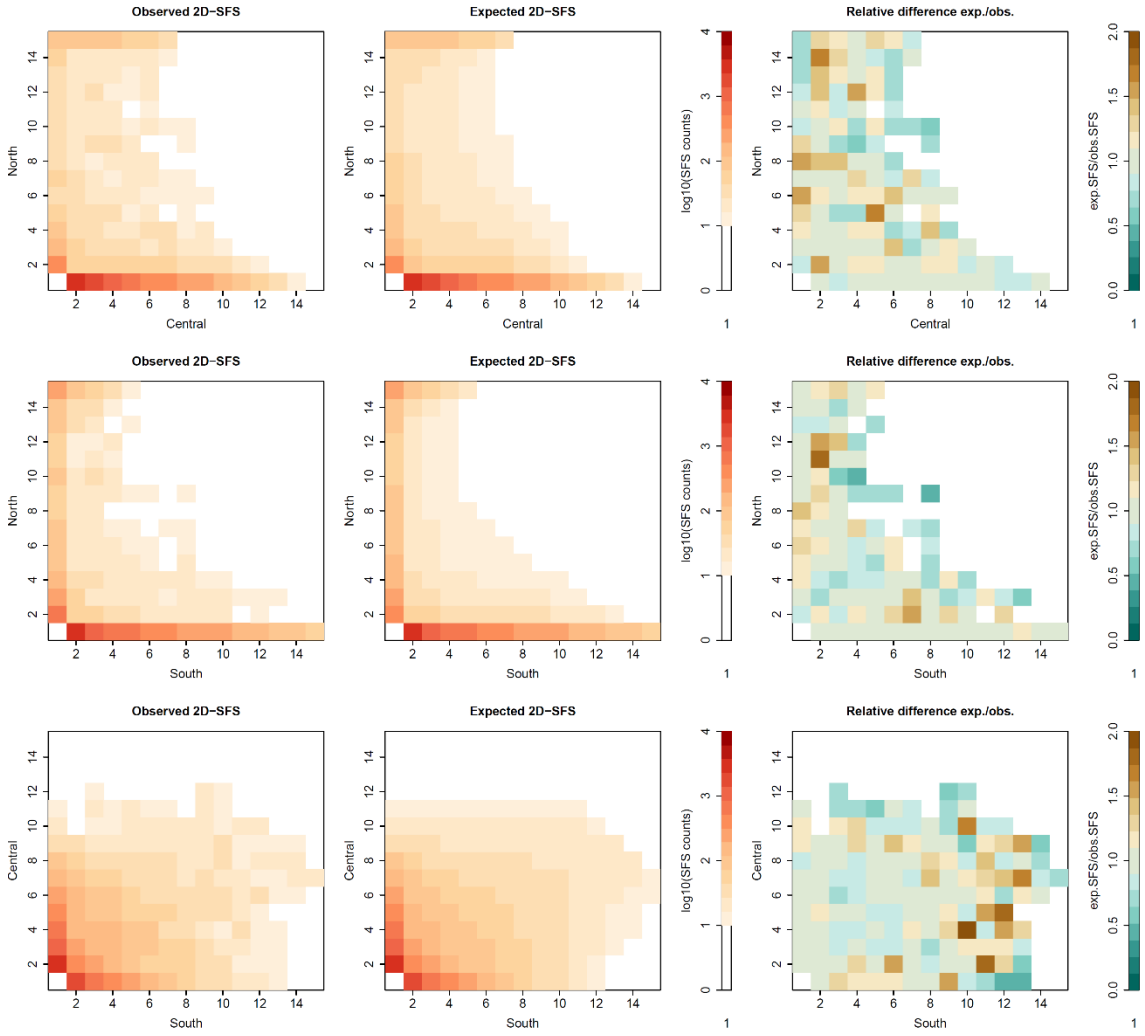

Observed 2D−SFS
Expected 2D−SFS
Relative difference exp./obs.
North
Central
South
log10(SFS counts)
exp.SFS/obs.SFS

Figure A1.14 – Fit of the expected pairwise 2D-SFS obtained for the parameters that maximize the likelihood under the best model [i.e., the model with full migration and the (North, Central), South topology]. (A) Central/North 2D-SFS; (B) South/North 2D-SFS; (C) South/Central 2D-SFS. Each row shows the observed and expected 2D-SFS in log10 scale (left and middle plots), and the relative differences between the observed and expected SFS, defined as the expected/observed SFS in natural scale (right plot). The expected SFS was multiplied by the total number of SNPs to be in the same scale (SNP counts) as the observed 2D-SFS. All the entries with less than ten SNPs were pooled together in the computation of the likelihood, and hence are not shown. The monomorphic sites across all samples were discarded. Overall, there is a good fit for all pairwise comparisons.

Table A1.12 – Summary of likelihood scores for four asymmetric migration demographic models, only individuals with at least 90% ADMIXTURE assignment probability to a cluster. Topology names for each model are as indicated in Figure A1.1. $\log_{10}$(Lhood) (ALL SNPs) and $\log_{10}$(Lhood) (1SNP) correspond to the mean likelihood computed with the datasets containing all SNPs and a single SNP per RAD locus, respectively. Mean likelihoods were computed based on 100 simulated expected site frequency spectra simulated according to the parameters that maximized the likelihood of each model. AIC scores and relative likelihoods (Akaike's weight of evidence) were calculated based on the 1 SNP dataset following Excoffier *et al.* 2013. Removing admixed individuals has no impact on the model chosen.

Topology	$\log_{10}$ (Lhood) (ALL SNPs)	$\log_{10}$ (Lhood) (1 SNP)	# Para-meters	AIC	ΔAIC	Relative likelihood
North-South	-46162.25	-7322.34	15	33750.62	34.51	0.000
North-Central	-46149.78	-7314.85	15	33716.11	0.00	0.996
South-Central	-46161.87	-7319.52	15	33737.64	21.53	~0.000
Trifurcation	-46160.29	-7318.98	11	33727.14	11.03	0.004

Table A1.13 – Summary of likelihood scores for four asymmetric migration demographic models, considering only loci with less than 10% missing data.

Topology names for each model are as indicated in Figure A1.1. $\log_{10}$(Lhood) (ALL SNPs) and log10(Lhood) (1SNP) correspond to the mean likelihood computed with the datasets containing all SNPs and a single SNP per RAD locus, respectively. Mean likelihoods were computed based on 100 simulated expected site frequency spectra simulated according to the parameters that maximized the likelihood of each model. AIC scores and relative likelihoods (Akaike's weight of evidence) were calculated based on the 1 SNP dataset following Excoffier *et al.* 2013. The North-Central model achieves the highest likelihood score. Due to the small number of SNPs available in the 1 SNP dataset, however, the Trifurcation model has the highest relative likelihood as it has a smaller number of parameters.

Topology	**$\log_{10}$ (Lhood) (ALL SNPs)**	**$\log_{10}$ (Lhood) (1 SNP)**	**# Para-meters**	**AIC**	**ΔAIC**	**Relative likelihood**
North-South	-27683.82	-3915.85	15	18063.17	9.30	0.006
North-Central	-27680.58	-3914.02	15	18054.75	0.88	0.380
South-Central	-27683.50	-3915.19	15	18060.13	6.26	0.026
Trifurcation	-27683.61	-3915.57	11	18053.87	0.00	0.589

Table A1.14 –Demographic parameters inferred under the asymmetrical migration, North-Central bifurcation model, considering only individuals with >90% ADMIXTURE assignment probability to a cluster. The maximum-likelihood point estimate for parameters is taken from the run reaching the highest composite likelihood. Migration rates were calculated using current population sizes, and are given forward in time. Estimates of divergence time assume three generations per year (Benjamin 1955; Wilson *et al.* 1992). Estimates of N_e (given in number of haploids) and the date of North/Central divergence are similar to those obtained with the full dataset. Relative migration rates between clusters are also similar, and within the 95% confidence intervals (95% CI). The divergence date of South and the Ancestor of North and Central is deeper than from the full dataset, but close to the upper limit of the 95% CI.

Parameter	ML Estimate
North N_e	37368
Central N_e	343124
South N_e	209649
Ancestral N_e	456209
North+Central Ancestor N_e	377755
North/Central divergence time (years)	25619
South/North+Central div. time (years)	63633
2Nm (Central to North)	0.14
2Nm (North to Central)	0.81
2Nm(South to North)	0.01
2Nm (North to South)	0.10
2Nm (South to Central)	0.90
2Nm (Central to South)	0.67
2Nm (Ancestor$_{\text{North+Central}}$ to South)	4.81
2Nm (South to Ancestor$_{\text{North+Central}}$)	0.02

+

Table A1.15 – Demographic parameters inferred under the asymmetrical migration, North-Central bifurcation model, considering only loci with <10% missing data. The maximum-likelihood point estimate for parameters is taken from the run reaching the highest composite likelihood. Migration rates were calculated using current population sizes, and are given forward in time. Estimates of divergence time assume three generations per year (Benjamin 1955; Wilson *et al.* 1992).. Since we did not have the correct number of monomorphic sites (callable sites with <10% missing data), estimates were re-scaled such that the proportion of polymorphic sites was the same as in the original dataset. Estimates of N_e (given in number of haploids) and divergence times are similar to those obtained with the full dataset, and most are within the 95% confidence interval. Relative migration rates between clusters are also similar.

Parameter	**ML Estimate**
North N_e	34327
Central N_e	370083
South N_e	233137
Ancestral N_e	419668
North+Central Ancestor N_e	417146
North/Central divergence time (years)	22834
South/North+Central div. time (years)	57170
2Nm (Central to North)	0.12
2Nm (North to Central)	1.27
2Nm(South to North)	0.03
2Nm (North to South)	0.10
2Nm (South to Central)	1.93
2Nm (Central to South)	1.07
2Nm ($Ancestor_{North+Central}$ to South)	3.22
2Nm (South to $Ancestor_{North+Central}$)	4.55×10^{-5}

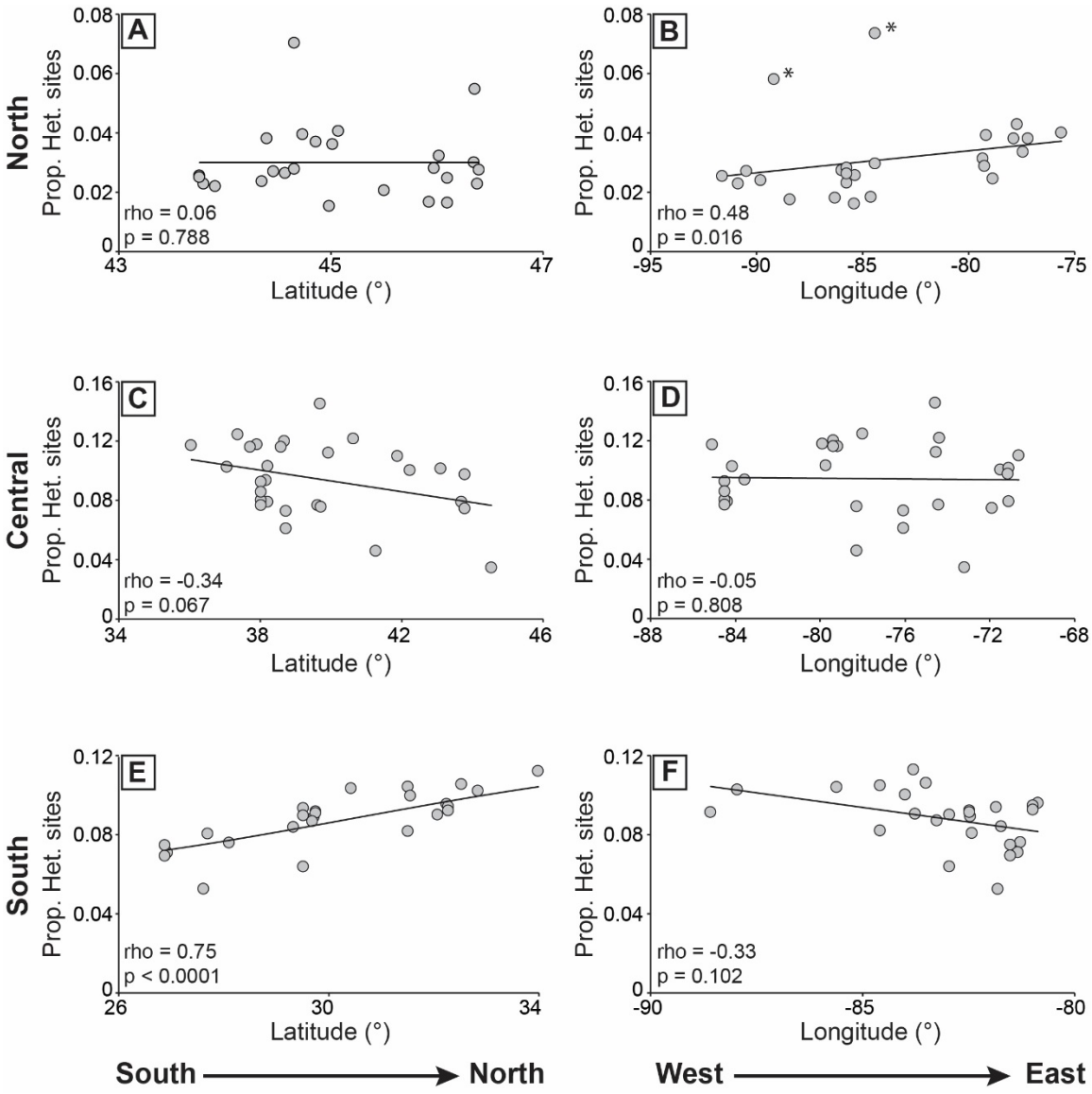


Figure A1.15 – Relationship between genetic diversity (proportion heterozygous sites) and geography, by region. Asterisks in (B) indicate two outlier points that are excluded from Fig. 4 and discussed further in the text (Spearman's rho for North was calculated with these points included). Genetic diversity correlated significantly and positively with latitude in South (E) and with longitude in North (B), and trended negatively with latitude in Central (C).

Table A1.16 – Mantel and partial Mantel test results for alternative ADMIXTURE assignment cutoffs, by geographical region. Pearson's *r* and *P*-value are given for simple Mantel tests (Matrix 1, Matrix 2) and partial Mantel tests (Matrix 1, Matrix 2 | list of matrices held constant). Significant IBD is recovered in all regions under all assignments. Significant IBE is always recovered in South, but is not recovered in Central at the 90% cutoff (*n* = 7 individuals removed).

	70% assignment cutoff		90% assignment cutoff	
Comparison	*r*	*P*-value	*r*	*P*-value
North[a]				
Geographic, Genetic	n/a	n/a	0.59	<0.0001
Geographic, Genetic | Host (IBD)	n/a	n/a	0.57	<0.0001
Host, Genetic	n/a	n/a	0.20	0.0090
Host, Genetic | Geographic (IBE)	n/a	n/a	0.10	0.1337
Geographic, Host (eco-spatial auto.)	n/a	n/a	0.21	0.0122
Central				
Geographic, Genetic	0.43	<0.0001	0.67	<0.0001
Geographic, Genetic | Host (IBD)	0.43	<0.0001	0.67	<0.0001
Host, Genetic	0.16	0.0272	0.10	0.1458
Host, Genetic | Geographic (IBE)	0.15	0.0453	0.13	0.1356
Geographic, Host (eco-spatial auto.)	0.05	0.2007	0.01	0.4261
South				
Geographic, Genetic	0.72	<0.0001	0.68	<0.0001
Geographic, Genetic | Host (IBD)	0.72	<0.0001	0.68	<0.0001
Host, Genetic	0.18	0.0104	0.17	0.0223
Host, Genetic | Geographic (IBE)	0.16	0.0155	0.16	0.0299
Geographic, Host (eco-spatial auto.)	0.10	0.0732	0.08	0.1300

a. All individuals in North have >90% assignment probability.

Table A1.17 – Mantel correlation coefficients describing the relationship between geographical and genetic distances across different geographical distance intervals within each region. To explore the impact of geographical scale on patterns of IBD, we examined patterns of IBD at different spatial intervals, using visual inspection of IBD plots within clusters to guide our choice of intervals (Hutchinson & Templeton 1999). To compare patterns among identified clusters, which differ in their geographical spread, we also restricted these interval analyses to the maximum pairwise distance of the smallest range (1148 km). The chosen intervals include: 0-1148 km, 0-450km, 450-1148km, 450-range maximum. For each interval, Pearson's correlation coefficient, *r*, was calculated using all qualifying pairs in Microsoft Excel. *P*-values were not calculated for these intervals due to the non-independence of data points subsampled from pairwise distance matrices (Hanfling & Weetman 2006). Restricting analyses to the smallest range (1148 km) has little impact on the magnitude of the IBD correlations (*r*) in North and Central. However, *r* drops off substantially above 450 km in North and Central, but not in South. Below 450 km, *r* is very similar across the three regions (range: 0.42-0.49).

Interval	*r*
North	
Full range (0-1244 km)	0.60
0-1148 km	0.59
0-450 km	0.45
450-1148 km	0.21
450-1243 km	0.26
Central	
Full range (0-1465 km)	0.45
0-1148 km	0.39
0-450 km	0.48
450-1148 km	0.07
450-1465 km	0.23
South	
Full range (0-1148 km)	0.79
0-450 km	0.48
450-1148 km	0.71

Appendix 2 – Evidence of host-associated phenotypic divergence, but not host-associated genetic divergence, between populations of the redheaded pine sawfly, *Neodiprion lecontei*, on two northern hosts

Table A2.1 – Collection information for individuals used in population genetic analyses.

ID	Collection Date	Collection location	Latitude	Longitude	Host plant
171_02_N	8/10/2002	Sebrite, ON	44.74	-79.16	*P. resinosa*
173-02	8/10/2002	Sebrite, ON	44.74	-79.16	*P. resinosa*
176_02	8/10/2002	Sebrite, ON	44.74	-79.16	*P. resinosa*
CAN002.02	7/27/2014	Sebrite, ON	44.74	-79.16	*P. banksiana*
CAN002_01	7/27/2014	Sebrite, ON	44.74	-79.16	*P. banksiana*
CAN005.01	7/27/2014	Sebrite, ON	44.74	-79.16	*P. resinosa*
CAN005_02	7/27/2014	Sebrite, ON	44.74	-79.16	*P. resinosa*
CAN007	7/27/2014	Harcourt, ON	44.86	-78.11	*P. resinosa*
CAN007_02	7/27/2014	Harcourt, ON	44.86	-78.11	*P. resinosa*
CAN007_03	7/27/2014	Harcourt, ON	44.86	-78.11	*P. resinosa*
CAN007_04	7/27/2014	Harcourt, ON	44.86	-78.11	*P. resinosa*
CAN007_05	7/27/2014	Harcourt, ON	44.86	-78.11	*P. resinosa*
CAN013	7/28/2014	Combermere, ON	45.32	-77.76	*P. resinosa*
CAN015	7/28/2014	Combermere, ON	45.32	-77.76	*P. resinosa*
CAN016	7/28/2014	Combermere, ON	45.32	-77.76	*P. resinosa*
CAN017	7/28/2014	Combermere, ON	45.32	-77.76	*P. resinosa*
CAN019	7/28/2014	Combermere, ON	45.32	-77.76	*P. resinosa*
CAN021	7/28/2014	Barry's Bay, ON A	45.48	-77.68	*P. resinosa*
CAN022	7/28/2014	Barry's Bay, ON A	45.48	-77.68	*P. resinosa*
CAN023	7/28/2014	Barry's Bay, ON A	45.48	-77.68	*P. resinosa*
CAN025_02	7/28/2014	Barry's Bay, ON A	45.48	-77.68	*P. resinosa*
CAN026	7/28/2014	Barry's Bay, ON B	45.49	-77.67	*P. resinosa*
CAN030	7/28/2014	Barry's Bay, ON B	45.49	-77.67	*P. resinosa*
CAN031	7/28/2014	Barry's Bay, ON B	45.49	-77.67	*P. resinosa*
CAN032	7/28/2014	Barry's Bay, ON B	45.49	-77.67	*P. resinosa*
CAN034	7/28/2014	Barry's Bay, ON B	45.49	-77.67	*P. resinosa*
CAN037a	7/29/2014	Laurentian Valley, ON	45.83	-77.24	*P. resinosa*
CAN037a.02	7/29/2014	Laurentian Valley, ON	45.83	-77.24	*P. resinosa*
CAN038	7/29/2014	Laurentian Valley, ON	45.83	-77.24	*P. resinosa*
CAN039_02	7/29/2014	Laurentian Valley, ON	45.83	-77.24	*P. resinosa*
CAN040_01	7/29/2014	Papineau-Cameron, ON	46.29	-78.81	*P. resinosa*
CAN040_02	7/29/2014	Papineau-Cameron, ON	46.29	-78.81	*P. resinosa*
CAN040_03	7/29/2014	Papineau-Cameron, ON	46.29	-78.81	*P. resinosa*
CAN040_04	7/29/2014	Papineau-Cameron, ON	46.29	-78.81	*P. resinosa*
CAN040_06	7/29/2014	Papineau-Cameron, ON	46.29	-78.81	*P. resinosa*
CAN042	7/30/2014	Baldwin, ON	46.29	-81.79	*P. resinosa*

Table A2.1 (cont.)

ID	Collection Date	Collection location	Latitude	Longitude	Host plant
CAN042_02	7/30/2014	Baldwin, ON	46.29	-81.79	*P. resinosa*
CAN042_03	7/30/2014	Baldwin, ON	46.29	-81.79	*P. resinosa*
CAN042_04	7/30/2014	Baldwin, ON	46.29	-81.79	*P. resinosa*
CAN042_06	7/30/2014	Baldwin, ON	46.29	-81.79	*P. resinosa*
CAN043	7/31/2014	Grayling, MI A	44.60	-84.71	*P. banksiana*
CAN045	7/31/2014	Grayling, MI A	44.60	-84.71	*P. banksiana*
CAN047_02	7/31/2014	Grayling, MI A	44.60	-84.71	*P. banksiana*
CAN048	8/1/2014	Grayling, MI B	44.66	-84.69	*P. resinosa*
CAN049	8/1/2014	Grayling, MI B	44.66	-84.69	*P. resinosa*
CAN050	8/1/2014	Grayling, MI B	44.66	-84.69	*P. resinosa*
CAN054_new	8/1/2014	Grayling, MI B	44.66	-84.69	*P. banksiana*
CAN055_old	8/1/2014	Grayling, MI B	44.66	-84.69	*P. banksiana*
CAN056_old	8/1/2014	Grayling, MI B	44.66	-84.69	*P. banksiana*
CAN061a	8/1/2014	Bitely, MI	43.76	-85.74	*P. banksiana*
CAN063	8/13/2014	Blind River, ON A	46.29	-81.79	*P. resinosa*
CAN064	8/15/2014	Blind River, ON B	46.22	-83.11	*P. resinosa*
CAN065	8/15/2014	Blind River, ON B	46.22	-83.11	*P. resinosa*
CAN066	8/15/2014	Blind River, ON B	46.22	-83.11	*P. resinosa*
CAN067	8/15/2014	Blind River, ON B	46.22	-83.11	*P. resinosa*
CAN068	8/15/2014	Blind River, ON B	46.22	-83.11	*P. resinosa*
CAN069	8/15/2014	Blind River, ON A	46.29	-81.79	*P. resinosa*
CAN070	8/15/2014	Blind River, ON A	46.29	-81.79	*P. resinosa*
CAN071	8/15/2014	Blind River, ON A	46.29	-81.79	*P. resinosa*
CAN072	8/15/2014	Blind River, ON A	46.29	-81.79	*P. resinosa*
CAN075	8/15/2014	Spanish, ON	46.20	-82.36	*P. resinosa*
CAN075_02	8/15/2014	Spanish, ON	46.20	-82.36	*P. resinosa*
CAN075_04	8/15/2014	Spanish, ON	46.20	-82.36	*P. resinosa*
CAN076	8/15/2014	Spanish, ON	46.20	-82.36	*P. resinosa*
CAN080	8/16/2014	Chelmsford, ON	46.57	-81.23	*P. banksiana*
CAN080.02	8/16/2014	Chelmsford, ON	46.57	-81.23	*P. banksiana*
CAN080_03	8/16/2014	Chelmsford, ON	46.57	-81.23	*P. banksiana*
CAN080_04 ♀	8/16/2014	Chelmsford, ON	46.57	-81.23	*P. banksiana*
CAN080_05	8/16/2014	Chelmsford, ON	46.57	-81.23	*P. banksiana*
CAN092	8/17/2014	Algoma, ON	46.72	-83.43	*P. resinosa*
CAN093	8/17/2014	Algoma, ON	46.72	-83.43	*P. resinosa*
CAN094	8/17/2014	Algoma, ON	46.72	-83.43	*P. resinosa*
CAN099	9/11/2014	Bitely, MI	43.76	-85.74	*P. resinosa*
CAN099_02	9/11/2014	Bitely, MI	43.76	-85.74	*P. resinosa*
CAN101	9/12/2014	Manistique, MI A	45.92	-86.30	*P. banksiana*

Table A2.1 (cont.)

ID	Collection Date	Collection location	Latitude	Longitude	Host plant
CAN111	9/13/2014	Necedah, WI	44.21	-90.14	*P. banksiana*
RB080	7/20/2012	Bitely, MI	43.76	-85.74	*P. banksiana*
RB091_01 ♀	7/21/2012	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB092_01	7/21/2012	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB093_01	7/21/2012	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB095_02	7/22/2012	Naubinway, MI A	46.09	-85.34	*P. banksiana*
RB095_03	7/22/2012	Naubinway, MI A	46.09	-85.34	*P. banksiana*
RB095_04	7/22/2012	Naubinway, MI A	46.09	-85.34	*P. banksiana*
RB096B	7/22/2012	Naubinway, MI B	46.10	-85.39	*P. banksiana*
RB098_01	7/22/2012	Naubinway, MI B	46.10	-85.39	*P. banksiana*
RB099_01	7/22/2012	Manistique, MI A	45.92	-86.30	*P. banksiana*
RB099_03	7/22/2012	Manistique, MI A	45.92	-86.30	*P. banksiana*
RB099_04	7/22/2012	Manistique, MI A	45.92	-86.30	*P. banksiana*
RB099_05	7/22/2012	Manistique, MI A	45.92	-86.30	*P. banksiana*
RB101	7/24/2012	Suring, WI	44.98	-88.45	*P. resinosa*
RB104_01	7/24/2012	Suring, WI	44.98	-88.45	*P. resinosa*
RB136.01	10/23/2012	Bitely, MI	43.76	-85.74	*P. resinosa*
RB136_02	10/23/2012	Bitely, MI	43.76	-85.74	*P. resinosa*
RB236	7/16/2013	Bitely, MI	43.76	-85.74	*P. banksiana*
RB237	7/16/2013	Bitely, MI	43.76	-85.74	*P. banksiana*
RB245_02	7/16/2013	Grayling, MI A	44.60	-84.71	*P. banksiana*
RB247.02	7/16/2013	Grayling, MI A	44.60	-84.71	*P. banksiana*
RB249	7/17/2013	Grayling, MI B	44.66	-84.69	*P. resinosa*
RB250_old	7/17/2013	Grayling, MI B	44.66	-84.69	*P. resinosa*
RB251_02	7/17/2013	Grayling, MI B	44.66	-84.69	*P. resinosa*
RB252_old	7/17/2013	Grayling, MI B	44.66	-84.69	*P. resinosa*
RB253_a	7/17/2013	Grayling, MI B	44.66	-84.69	*P. resinosa*
RB254_old	7/17/2013	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB256_old	7/17/2013	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB257	7/17/2013	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB258_a	7/17/2013	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB259b	7/17/2013	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB260_a	7/17/2013	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB261	7/17/2013	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB284	7/20/2013	Necedah, WI	44.21	-90.14	*P. banksiana*
RB285	7/20/2013	Necedah, WI	44.21	-90.14	*P. banksiana*
RB373_02	8/22/2014	Dunlop Lake, ON	46.47	-82.66	*P. resinosa*
RB374	8/22/2014	Dunlop Lake, ON	46.47	-82.66	*P. resinosa*
RB376	8/22/2014	Dunlop Lake, ON	46.47	-82.66	*P. resinosa*

Table A2.1 (cont.)

ID	Collection Date	Collection location	Latitude	Longitude	Host plant
RB377_02	8/22/2014	Dunlop Lake, ON	46.47	-82.66	*P. resinosa*
RB378	8/22/2014	Dunlop Lake, ON	46.47	-82.66	*P. resinosa*
RB385	7/15/2015	Bitely, MI	43.76	-85.74	*P. banksiana*
RB386	7/16/2015	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB387_02	7/16/2015	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB388	7/16/2015	Grayling, MI B	44.66	-84.69	*P. banksiana*
RB390	7/17/2015	Suring, WI	44.98	-88.45	*P. resinosa*
RB395	7/17/2015	Suring, WI	44.98	-88.45	*P. resinosa*
RB399	7/17/2015	Necedah, WI	44.21	-90.14	*P. banksiana*
RB402	7/17/2015	Necedah, WI	44.21	-90.14	*P. banksiana*
RB405	8/16/2015	Manistique, MI B	45.95	-86.26	*P. resinosa*
RB406	8/16/2015	Manistique, MI B	45.95	-86.26	*P. resinosa*
RB407	8/16/2015	Thompson Township, MI	45.93	-86.29	*P. banksiana*
RB408	8/16/2015	Thompson Township, MI	45.93	-86.29	*P. banksiana*
RB409	8/16/2015	Thompson Township, MI	45.93	-86.29	*P. resinosa*
RB410	8/16/2015	Thompson Township, MI	45.93	-86.29	*P. resinosa*
RB411	8/16/2015	Thompson Township, MI	45.93	-86.29	*P. resinosa*
RB412	8/16/2015	Thompson Township, MI	45.93	-86.29	*P. resinosa*
RB413	8/16/2015	Thompson Township, MI	45.93	-86.29	*P. resinosa*
RB421	8/17/2015	Suring, WI	44.98	-88.45	*P. resinosa*
RB425	7/9/2016	Rothschild, WI	44.86	-89.64	*P. banksiana*
RB426	7/9/2016	Rothschild, WI	44.86	-89.64	*P. banksiana*
RB427	7/9/2016	Rothschild, WI	44.86	-89.64	*P. banksiana*
RB428	7/9/2016	Rothschild, WI	44.86	-89.64	*P. banksiana*
RB429	7/9/2016	Rothschild, WI	44.86	-89.64	*P. banksiana*
RB431	7/9/2016	Mosinee, WI	44.84	-89.69	*P. resinosa*
RB431_01	7/9/2016	Mosinee, WI	44.84	-89.69	*P. resinosa*
RB431_02	7/9/2016	Mosinee, WI	44.84	-89.69	*P. resinosa*
RB432	7/9/2016	Mosinee, WI	44.84	-89.69	*P. resinosa*
RB433	7/9/2016	Mosinee, WI	44.84	-89.69	*P. banksiana*
RB434	7/9/2016	Mosinee, WI	44.84	-89.69	*P. banksiana*
RB435_02	7/9/2016	Mosinee, WI	44.84	-89.69	*P. banksiana*
RB437	7/9/2016	Friendship, WI	44.03	-89.71	*P. banksiana*
RB438	7/9/2016	Friendship, WI	44.03	-89.71	*P. banksiana*
RB439_02	7/9/2016	Friendship, WI	44.03	-89.71	*P. banksiana*
RB440	7/9/2016	Friendship, WI	44.03	-89.71	*P. banksiana*
RB444	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB445_b	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB447	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*

Table A2.1 (cont.)

ID	Collection Date	Collection location	Latitude	Longitude	Host plant
RB448_02	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB450	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB451	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB452	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB454	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB456	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB457	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB458	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB460	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB461	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB462_02	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB463	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB464	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB465	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB466	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB468	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB469_02	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB470	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB473	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB474	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB475	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB476	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB477	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB478	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB479_01	7/23/2016	Frederic, MI	44.73	-84.75	*P. banksiana*
RB480_02	7/23/2016	Naubinway, MI B	46.10	-85.39	*P. resinosa*
RB480_1	7/23/2016	Naubinway, MI B	46.10	-85.39	*P. resinosa*
RB481	7/23/2016	Naubinway, MI B	46.10	-85.39	*P. banksiana*
RB482	7/23/2016	Naubinway, MI B	46.10	-85.39	*P. banksiana*
RB483	7/23/2016	Naubinway, MI B	46.10	-85.39	*P. banksiana*
RB484	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*
RB485	7/23/2016	Frederic, MI	44.73	-84.75	*P. resinosa*

Table A2.2 – Individual barcode and Illumina index assignments, and proportion of heterozygous sites. The proportion of heterozygous sites was calculated after excluding sites missing data in >30% of individuals and pruning to 1 SNP per RAD locus. 25 putatively haploid individuals (*) were identified from their strikingly low proportion of heterozygous sites (average = 0.0045) compared to remaining individuals (average = 0.1089), including two confirmed adult female specimens (†).

ID	Illumina index	Barcode	Prop. Het. Sites
171_02_N	CCGTCCCG	TGCTT	0.1183
173-02	CGATGTAT	TCACTG	0.1186
176_02	ATCACGAT	TAGCGGAT	0.1460
CAN002_01*	AGTCAACA	TGGCAACAGA	0.0046
CAN002.02	TTAGGCAT	CGTGGACAGT	0.1534
CAN005.01	AGTCAACA	ATGAGCAA	0.1681
CAN005_02*	TTAGGCAT	ATGAGCAA	0.0053
CAN007	CCGTCCCG	GCGTCCT	0.1213
CAN007_02*	TGACCAAT	CAGATA	0.0053
CAN007_03	GTCCGCAC	CTCGCGG	0.1370
CAN007_04*	TTAGGCAT	TCTTGG	0.0050
CAN007_05	CGATGTAT	TAGCGGAT	0.1147
CAN013	ATCACGAT	TCACTG	0.0872
CAN015	TTAGGCAT	CCTCG	0.0735
CAN016*	AGTCAACA	CCGAACA	0.0044
CAN017	CCGTCCCG	GGATA	0.0961
CAN019	CGATGTAT	TCTTGG	0.0943
CAN021*	CCGTCCCG	TCACTG	0.0039
CAN022*	TGACCAAT	CTCGCGG	0.0044
CAN023	GTGAAACG	CCACTCA	0.1783
CAN025_02	CCGTCCCG	GCCTACCT	0.1594
CAN026	TTAGGCAT	CACCA	0.1074
CAN030	CCGTCCCG	CCTCG	0.1873
CAN031	GTCCGCAC	GAGCGACAT	0.1738
CAN032	AGTCAACA	CCACTCA	0.1651
CAN034*	ATCACGAT	GCCTACCT	0.0047
CAN037a	CGATGTAT	TGCTT	0.1166
CAN037a.02	AGTCAACA	GGATA	0.1318
CAN038	ATCACGAT	ATTAT	0.1262

Table A2.2 (cont.)

ID	Illumina index	Barcode	Prop. Het. Sites
CAN039_02*	TGACCAAT	GCCTACCT	0.0046
CAN040_01	CGATGTAT	GAAGTG	0.0490
CAN040_02	GTCCGCAC	CACCA	0.0743
CAN040_03	ATCACGAT	CTAAGCA	0.0518
CAN040_04*	TGACCAAT	TCACTG	0.0045
CAN040_06	CCGTCCCG	GCAAGCCAT	0.0614
CAN042	AGTCAACA	CTTGA	0.1304
CAN042_02	ATCACGAT	CCGAACA	0.1358
CAN042_03	CCGTCCCG	CTTGA	0.1172
CAN042_04	TTAGGCAT	GGATA	0.0774
CAN042_06	CGATGTAT	ACAACT	0.1319
CAN043	TTAGGCAT	GGTGT	0.1291
CAN045	CCGTCCCG	ATTAT	0.0718
CAN047_02	CGATGTAT	CCTCG	0.0949
CAN048	AGTCAACA	GCGTCCT	0.1171
CAN049*	CCGTCCCG	AAGACGCT	0.0038
CAN050	CGATGTAT	ATGAGCAA	0.1109
CAN054_new	CGATGTAT	GGAACGA	0.0939
CAN055_old	ATCACGAT	GAAGTG	0.1136
CAN056_old	CCGTCCCG	CTCGCGG	0.1095
CAN061a	AGTCAACA	ATAGAT	0.1121
CAN063*	TTAGGCAT	AACTGG	0.0047
CAN064	ATCACGAT	CCACTCA	0.1437
CAN065*	GTGAAACG	ATATCGCCA	0.0037
CAN066	CCGTCCCG	ACAACT	0.1325
CAN067	CGATGTAT	TAGCCAA	0.1337
CAN068	AGTCAACA	ACTGCGAT	0.1421
CAN069	CCGTCCCG	GGTGT	0.1267
CAN070	TGACCAAT	CCACTCA	0.1323
CAN071	CGATGTAT	CTTGA	0.1300
CAN072	ATCACGAT	ATGAGCAA	0.1333
CAN075	AGTCAACA	GAAGTG	0.1434
CAN075_02	ATCACGAT	CACCA	0.1490
CAN075_04	CGATGTAT	CACCA	0.1217
CAN076	CCGTCCCG	ATAGAT	0.1462
CAN080	CCGTCCCG	CTAAGCA	0.1233
CAN080.02	AGTCAACA	CACCA	0.1313
CAN080_03	ATCACGAT	TGCTT	0.1075
CAN080_04†	CGATGTAT	CCACTCA	0.1123

Table A2.2 (cont.)

ID	Illumina index	Barcode	Prop. Het. Sites
CAN080_05	GTCCGCAC	TGACGCCA	0.1269
CAN092*	CCGTCCCG	ACCAGGA	0.0045
CAN093	AGTCAACA	ATTAT	0.0936
CAN094	ATCACGAT	GGTGT	0.0958
CAN099	GTCCGCAC	ATAGAT	0.1140
CAN099_02	CGATGTAT	ACGGTACT	N/A
CAN101	CGATGTAT	GGATA	N/A
CAN111	ATCACGAT	CTTGA	0.0790
RB080	GTCCGCAC	GAAGTG	0.1118
RB091_01†	CGATGTAT	GCGTCCT	0.1010
RB092_01	TTAGGCAT	CTAAGCA	0.1083
RB093_01	ATCACGAT	CCTTGCCATT	0.0921
RB095_02	GTGAAACG	TAGCGGAT	0.0980
RB095_03	CGATGTAT	CAACCACACA	0.1015
RB095_04*	TGACCAAT	TCAGAGAT	0.0046
RB096B	GTGAAACG	TAGCCAA	0.0688
RB098_01	CCGTCCCG	AACGTGCCT	0.1001
RB099_01	ATCACGAT	CAGATA	0.0988
RB099_03	TTAGGCAT	ACAACT	0.0656
RB099_04	CCGTCCCG	TGACGCCA	0.0822
RB099_05	TGACCAAT	AAGACGCT	0.0665
RB101	GTGAAACG	TGGCACAGA	0.0876
RB104_01	GTGAAACG	CAGATA	0.0801
RB136.01	AGTCAACA	ATATCGCCA	0.0955
RB136_02	CCGTCCCG	GGAACGA	0.0943
RB236	ATCACGAT	ACAACT	0.1148
RB237	CGATGTAT	TATGT	0.0980
RB245_02	ATCACGAT	CCTCG	0.1073
RB247.02	AGTCAACA	CCTCG	0.0962
RB249	GTCCGCAC	TATGT	0.1074
RB250_old*	GTGAAACG	TATGT	0.0040
RB251_02	GTCCGCAC	GCGTCCT	0.1089
RB252_old*	TGACCAAT	ACGGTACT	0.0038
RB253_a	TTAGGCAT	CAGATA	0.1062
RB254_old*	TGACCAAT	ACTGCGAT	0.0039
RB256_old	GTGAAACG	GGTGT	0.1212
RB257	GTCCGCAC	ACAACT	0.1174
RB258_a	TTAGGCAT	TATGT	0.1119
RB259b	AGTCAACA	GAGCGACAT	0.1091

Table A2.2 (cont.)

ID	Illumina index	Barcode	Prop. Het. Sites
RB260_a	CCGTCCCG	AACTGG	0.0958
RB261	AGTCAACA	GCCTACCT	0.1091
RB284*	AGTCAACA	GGCTTA	0.0042
RB285	TTAGGCAT	GCGTCCT	0.0971
RB373_02	TGACCAAT	CCGAACA	0.1371
RB374	CGATGTAT	GGCTTA	0.1029
RB376	ATCACGAT	ATAGAT	0.1130
RB377_02	GTGAAACG	ACAACT	0.1408
RB378	AGTCAACA	ACCAGGA	0.1434
RB385	CCGTCCCG	CCACTCA	0.1077
RB386	ATCACGAT	CAACCACACA	0.1119
RB387_02	GTCCGCAC	CTAAGCA	0.1136
RB388	TGACCAAT	GAAGTG	0.1107
RB390	GTCCGCAC	CTCTCGCAT	0.0827
RB395	TTAGGCAT	TATTCGCAT	0.0597
RB399	TGACCAAT	ATGAGCAA	0.0994
RB402	CCGTCCCG	CCGAACA	0.0789
RB405	GTGAAACG	ACTGCGAT	0.1016
RB406	TTAGGCAT	ACAACCAACT	0.1022
RB407	TTAGGCAT	CCACTCA	0.1089
RB408	CGATGTAT	CGTGGACAGT	0.1045
RB409	TTAGGCAT	ACCAGGA	0.1164
RB410	TGACCAAT	GGCTTA	0.0791
RB411	GTCCGCAC	TAGCCAA	0.0940
RB412	AGTCAACA	CGTGGACAGT	0.1022
RB413	CCGTCCCG	GGTGCACATT	0.1085
RB421*	TGACCAAT	ATATCGCCA	0.0050
RB425	GTCCGCAC	ACGGTACT	0.1129
RB426	CGATGTAT	CCTTGCCATT	0.0970
RB427	TTAGGCAT	TCACGGAAG	0.1065
RB428	TGACCAAT	GCGTCCT	0.0997
RB429	CCGTCCCG	CAACCACACA	0.1157
RB431	CCGTCCCG	CGTGGACAGT	0.0791
RB431_01	TGACCAAT	TAGCCAA	0.0799
RB431_02	GTCCGCAC	ATATCGCCA	0.0620
RB432	TTAGGCAT	GAGCGACAT	0.0859
RB433	GTGAAACG	TGCTT	0.0580
RB434	TTAGGCAT	CCTTGCCATT	0.0979
RB435_02	CGATGTAT	ACAACCAACT	0.0583

Table A2.2 (cont.)

ID	Illumina index	Barcode	Prop. Het. Sites
RB437	GTCCGCAC	GGCTTA	0.1036
RB438	CCGTCCCG	TCAGAGAT	0.0980
RB439_02	CGATGTAT	CAGATA	0.0977
RB440	ATCACGAT	TCACGGAAG	0.1000
RB444	ATCACGAT	AACGCACATT	0.1027
RB445_b	TTAGGCAT	AACGTGCCT	0.1213
RB447	TGACCAAT	TGACGCCA	0.1171
RB448_02	ATCACGAT	TGGCAACAGA	0.1193
RB450	CCGTCCCG	GGCTTA	0.1119
RB451*	GTCCGCAC	GCCTACCT	0.0046
RB452	TTAGGCAT	ACTGCGAT	0.1192
RB454	GTCCGCAC	GGATA	0.1031
RB456	CGATGTAT	ACTGCGAT	0.1103
RB457	AGTCAACA	TAGCGGAT	0.1008
RB458	CGATGTAT	CCGAACA	0.1110
RB460*	GTGAAACG	GGATA	0.0045
RB461	TGACCAAT	TAGCGGAT	0.1204
RB462_02*	TTAGGCAT	CAACCACACA	0.0049
RB463	TGACCAAT	CTCTCGCAT	0.1079
RB464	TTAGGCAT	CTCTCGCAT	N/A
RB465	GTGAAACG	ACCAGGA	0.1058
RB466	TGACCAAT	TATTCGCAT	0.1123
RB468	GTGAAACG	CTAAGCA	0.1148
RB469_02	GTCCGCAC	TGCTT	0.1002
RB470	CGATGTAT	AACGCACATT	0.1192
RB473	CCGTCCCG	CCTTGCCATT	0.1190
RB474	CGATGTAT	TCAGAGAT	N/A
RB475	AGTCAACA	ACAACT	N/A
RB476	CCGTCCCG	TGGCACAGA	0.1078
RB477	GTGAAACG	GAAGTG	0.1108
RB478	TTAGGCAT	ATATCGCCA	0.1069
RB479_01	GTCCGCAC	TCTTGG	0.1200
RB480_02	GTCCGCAC	ACAACCAACT	0.0986
RB480_1	TGACCAAT	CGTGGACAGT	0.0964
RB481*	AGTCAACA	CGTCGCCACT	0.0061
RB482*	GTGAAACG	TCTTGG	0.0043
RB483	TGACCAAT	CGTCGCCACT	0.1028
RB484	GTGAAACG	CCTCG	0.1146
RB485	ATCACGAT	CTCTCGCAT	0.1131

Table A2.3 – Sequences for adapters containing variable-length barcodes from Burford Reiskind *et al.* (2016).

Barcode	P1.1 sequence	P1.2 sequence
ATTAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATTATCATG	/5Phos/ATAATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CACCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACCACATG	/5Phos/TGGTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CCTCG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTCGCATG	/5Phos/CGAGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CTTGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGACATG	/5Phos/TCAAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GGATA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGATACATG	/5Phos/TATCCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GGTGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTGTCATG	/5Phos/ACACCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TATGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTATGTCATG	/5Phos/ACATAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TGCTT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCTTCATG	/5Phos/AAGCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
AACTGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACTGGCATG	/5Phos/CCAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ACAACT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACAACTCATG	/5Phos/AGTTGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ATAGAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATAGATCATG	/5Phos/ATCTATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CAGATA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGATACATG	/5Phos/TATCTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GAAGTG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAAGTGCATG	/5Phos/CACTTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GGCTTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCTTACATG	/5Phos/TAAGCCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TCTTGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCTTGGCATG	/5Phos/CCAAGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TCACTG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCACTGCATG	/5Phos/CAGTGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ACCAGGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACCAGGACATG	/5Phos/TCCTGGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CCACTCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCACTCACATG	/5Phos/TGAGTGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CCGAACA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGAACACATG	/5Phos/TGTTCGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CTAAGCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTAAGCACATG	/5Phos/TGCTTAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CTCGCGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCGCGGCATG	/5Phos/CCGCGAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GCGTCCT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCGTCCTCATG	/5Phos/AGGACGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GGAACGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAACGACATG	/5Phos/TCGTTCCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TAGCCAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGCCAACATG	/5Phos/TTGGCTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Table A2.3 (cont.)

Barcode	P1.1 sequence	P1.2 sequence
ACTGCGAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGCGATCATG	/5Phos/ATCGCAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ATGAGCAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGAGCAACATG	/5Phos/TTGCTCATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GCCTACCT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCTACCTCATG	/5Phos/AGGTAGGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TAGCGGAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGCGGATCATG	/5Phos/ATCCGCTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TGACGCCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGACGCCACATG	/5Phos/TGGCGTCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ACGGTACT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACGGTACTCATG	/5Phos/AGTACCGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
AAGACGCT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGACGCTCATG	/5Phos/AGCGTCTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TCAGAGAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAGAGATCATG	/5Phos/ATCTCTGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ATATCGCCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATATCGCCACATG	/5Phos/TGGCGATATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GAGCGACAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGCGACATCATG	/5Phos/ATGTCGCTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GCAAGCCAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCAAGCCATCATG	/5Phos/ATGGCTTGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
AACGTGCCT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACGTGCCTCATG	/5Phos/AGGCACGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TATTCGCAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTATTCGCATCATG	/5Phos/ATGCGAATAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TCACGGAAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCACGGAAGCATG	/5Phos/CTTCCGTGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TGGCACAGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGCACAGACATG	/5Phos/TCTGTGCCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CTCTCGCAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCTCGCATCATG	/5Phos/ATGCGAGAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
AACGCACATT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACGCACATTCATG	/5Phos/AATGTGCGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CCTTGCCATT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTTGCCATTCATG	/5Phos/AATGGCAAGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CGTCGCCACT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTCGCCACTCATG	/5Phos/AGTGGCGACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CGTGGACAGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGGACAGTCATG	/5Phos/ACTGTCCACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GGTGCACATT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTGCACATTCATG	/5Phos/AATGTGCACCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TGGCAACAGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGCAACAGACATG	/5Phos/TCTGTTGCCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
ACAACCAACT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACAACCAACTCATG	/5Phos/AGTTGGTTGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CAACCACACA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAACCACACACATG	/5Phos/TGTGTGGTTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Table A2.4 – Sequences for PCR primers, including degenerate bases. PCR1 is the universal primer used for all reactions, and the PCR2 primers contain their respective Illumina index sequences in their "Name". When ordering, degenerate bases should be hand-mixed to ensure equal base composition.

Name	Sequence
PCR1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG
PCR2_ATCACGAT	CAAGCAGAAGACGGCATACGAGATNNNNATCGTGATGTGACTGGAGTTCAGACGTGTGC
PCR2_CGATGTAT	CAAGCAGAAGACGGCATACGAGATNNNNATACATCGGTGACTGGAGTTCAGACGTGTGC
PCR2_TTAGGCAT	CAAGCAGAAGACGGCATACGAGATNNNNATGCCTAAGTGACTGGAGTTCAGACGTGTGC
PCR2_TGACCAAT	CAAGCAGAAGACGGCATACGAGATNNNNATTGGTCAGTGACTGGAGTTCAGACGTGTGC
PCR2_ACAGTGAT	CAAGCAGAAGACGGCATACGAGATNNNNATCACTGTGTGACTGGAGTTCAGACGTGTGC
PCR2_GGCTACAT	CAAGCAGAAGACGGCATACGAGATNNNNATGTAGCCGTGACTGGAGTTCAGACGTGTGC
PCR2_AGTCAACA	CAAGCAGAAGACGGCATACGAGATNNNNTGTTGACTGTGACTGGAGTTCAGACGTGTGC
PCR2_CCGTCCCG	CAAGCAGAAGACGGCATACGAGATNNNNCGGGACGGGTGACTGGAGTTCAGACGTGTGC
PCR2_GTCCGCAC	CAAGCAGAAGACGGCATACGAGATNNNNGTGCGGACGTGACTGGAGTTCAGACGTGTGC
PCR2_GTGAAACG	CAAGCAGAAGACGGCATACGAGATNNNNCGTTTCACGTGACTGGAGTTCAGACGTGTGC
PCR2_GTGGCCTT	CAAGCAGAAGACGGCATACGAGATNNNNAAGGCCACGTGACTGGAGTTCAGACGTGTGC
PCR2_GTTTCGGA	CAAGCAGAAGACGGCATACGAGATNNNNTCCGAAACGTGACTGGAGTTCAGACGTGTGC

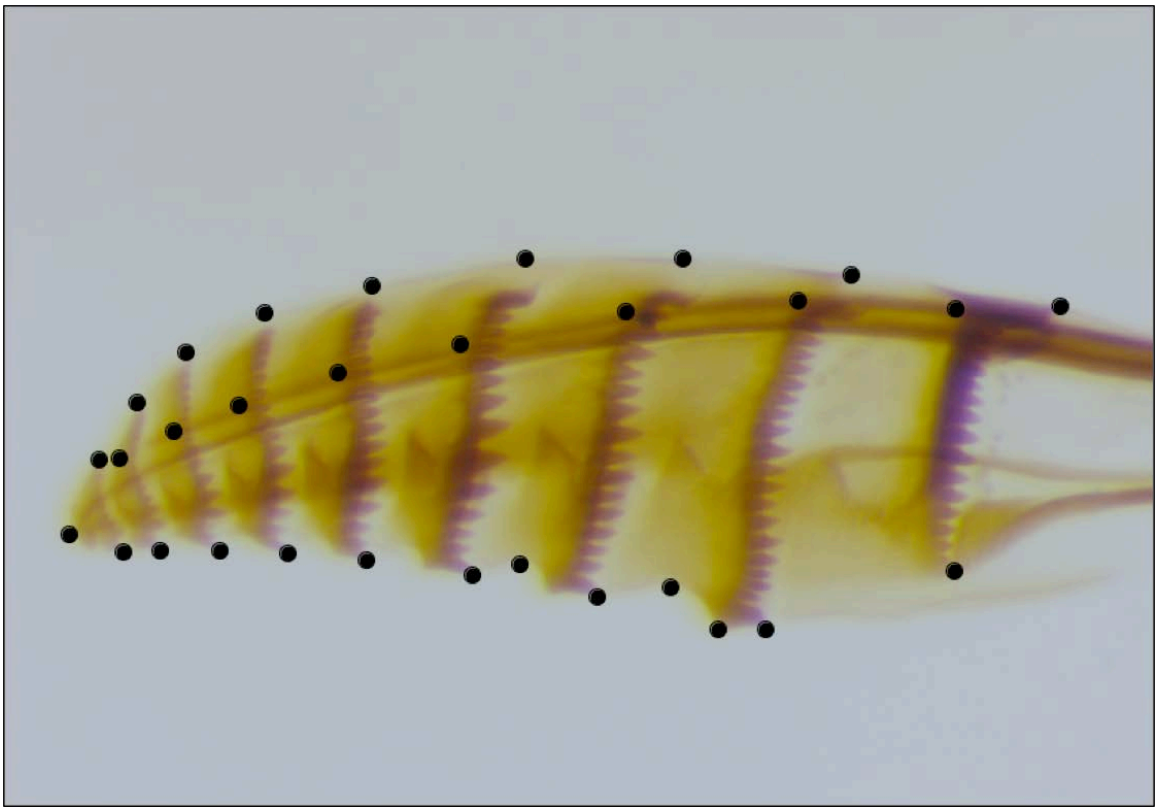


Figure A2.1 – Landmark positions for geometric morphometric analysis. A total of 30 landmarks are laid per ovipositor image.

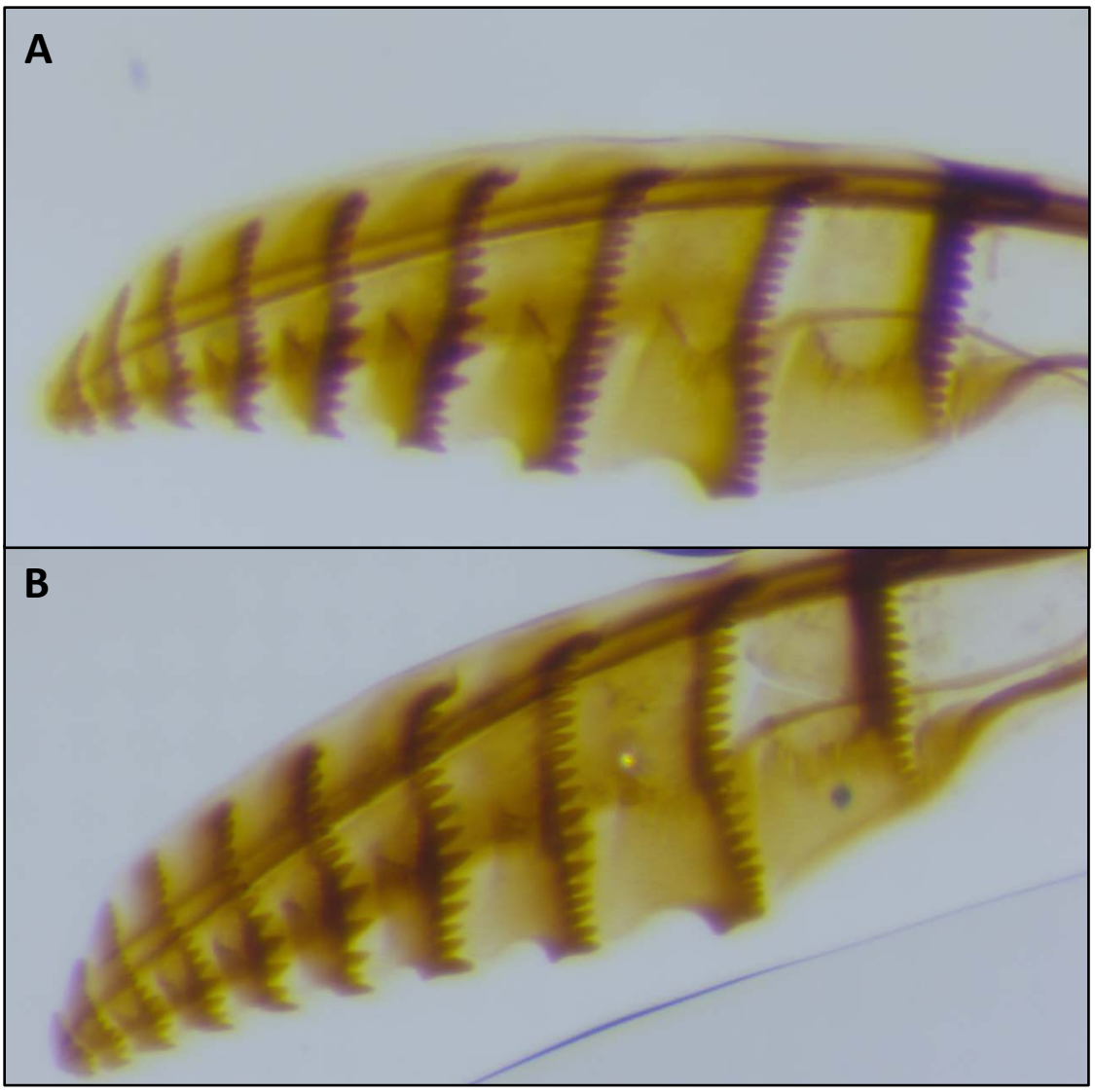


Figure A2.2 – Variation in annulus number in North *N. lecontei* females. Although ovipositors most frequently have 9 annuli (A), a small proportion of ovipositors, mostly from *P. banksiana* females, have 10 annuli (B).

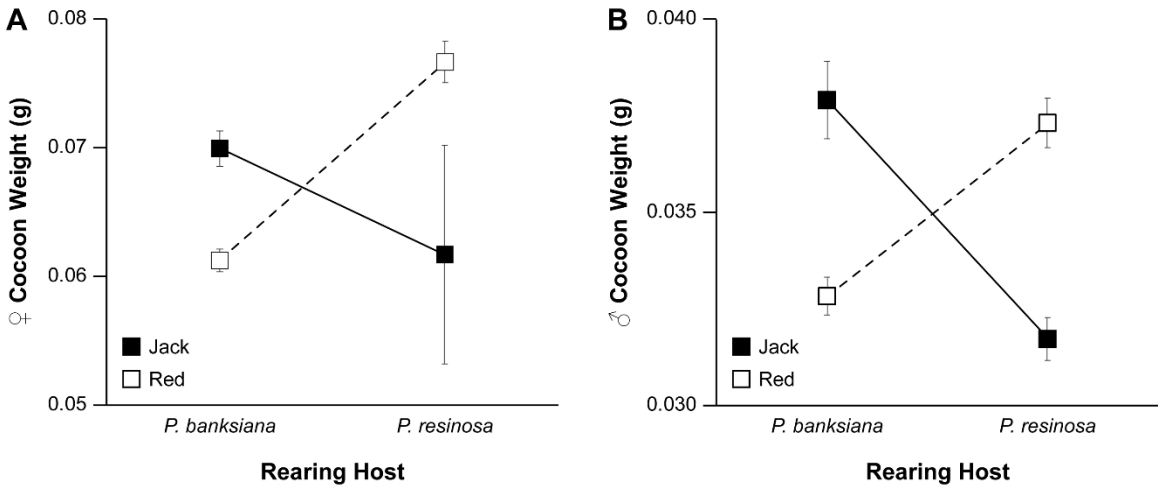


Figure A2.3. Cocoon weights from reciprocal transplant analyses. Both females (A) and males (B) achieve significantly higher cocoon weights when reared on their original host *vs.* the alternative host.

Appendix 3 – Host-associated divergence in a recently established sympatric population of the red-headed pine sawfly, *Neodiprion lecontei*, on three pine hosts

Table A3.1. Collection and usage information for all samples. Specimens were used in one or more of the following assays: population structure and differentiation ("Gen. Assays"), analysis of ovipositor morphology ("Ovipos. Morph."), laboratory tests for sexual isolation, habitat isolation, and larval performance ("Eco. Assays"), or differences in patterns of eclosion and mean eclosion dates ("Temp. Iso."). For each sample, the collection date, host plant, and type of assay(s) specimens were used are indicated.

ID	Collection Date	Host Plant	Gen. Assays	Ovipos. Morph.	Eco. Assays	Temp. Iso.
LL002	6/13/2012	*P. echinata*	X			X
LL003	6/13/2012	*P. echinata*	X			X
LL004	6/13/2012	*P. echinata*	X			X
LL005	6/15/2012	*P. rigida*	X			X
LL006	6/15/2012	*P. rigida*	X			X
LL027	7/25/2013	*P. echinata*	X			
LL028	8/2/2013	*P. echinata*	X			X
LL047	6/9/2014	*P. echinata*		X		X
LL049	6/9/2014	*P. echinata*		X		X
LL050	6/9/2014	*P. echinata*				X
LL051	6/9/2014	*P. echinata*		X		X
LL052	6/9/2014	*P. virginiana*				X
LL053	6/9/2014	*P. virginiana*	X			X
LL054	6/9/2014	*P. virginiana*		X		X
LL055	6/9/2014	*P. virginiana*				X
LL056	6/9/2014	*P. rigida*				X
LL057	6/9/2014	*P. rigida*				X
LL058	6/19/2014	*P. virginiana*	X	X		X
LL059	6/19/2014	*P. virginiana*		X		X
LL060	6/19/2014	*P. virginiana*				X
LL061	6/19/2014	*P. virginiana*	X	X		X
LL062	6/19/2014	*P. echinata*	X	X		X
LL063	6/19/2014	*P. echinata*				X
LL064	6/19/2014	*P. rigida*	X			X
LL066	6/19/2014	*P. rigida*		X		X
LL067	6/19/2014	*P. rigida*		X		X
LL069	6/27/2014	*P. virginiana*	X			

Table A3.1 (cont.)

ID	Collection Date	Host Plant	Gen. Assays	Ovipos. Morph.	Eco. Assays	Temp. Iso.
LL070	6/27/2014	*P. rigida*	X	X		X
LL071	6/27/2014	*P. echinata*	X			X
LL072	6/27/2014	*P. virginiana*	X			X
LL073	6/27/2014	*P. echinata*	X			X
LL074	6/30/2014	*P. echinata*	X	X		X
LL075	6/30/2014	*P. rigida*		X		X
LL076	6/30/2014	*P. virginiana*		X		X
LL077	6/30/2014	*P. virginiana*				X
LL078	6/30/2014	*P. rigida*	X			
LL081	7/24/2014	*P. rigida*				X
LL092	8/28/2014	*P. virginiana*	X			
LL102	9/4/2014	*P. rigida*	X			
LL116	6/2/2015	*P. rigida*		X		
LL121	6/17/2015	*P. rigida*		X		
LL122	6/17/2015	*P. rigida*		X		
LL136	6/19/2015	*P. virginiana*	X			
LL137	6/19/2015	*P. virginiana*	X			
LL139	6/19/2015	*P. echinata*	X			
LL140	6/19/2015	*P. virginiana*	X			
LL179	7/3/2015	*P. echinata*	X			
LL183	7/8/2015	*P. echinata*		X		
LL190	7/14/2015	*P. echinata*		X		
LL191	7/20/2015	*P. echinata*		X		
LL216	8/28/2015	*P. echinata*		X		
RB073	6/21/2012	*P. rigida*				X
RB076	7/5/2012	*P. virginiana*	X			X
RB122	8/20/2012	*P. echinata*	X			X
RB126	8/31/2012	*P. virginiana*	X			X
RB127	8/31/2012	*P. virginiana*	X			X
RB128	8/31/2012	*P. virginiana*	X			
RB129	8/31/2012	*P. echinata*	X			
RB132	9/14/2012	*P. rigida*	X			
RB135	9/14/2012	*P. virginiana*	X			
RB141	6/26/2013	*P. virginiana*				X
RB335	8/22/2013	*P. echinata*			X	X
RB336	8/22/2013	*P. echinata*			X	X

Table A3.1 (cont.)

ID	Collection Date	Host Plant	Gen. Assays	Ovipos. Morph.	Eco. Assays	Temp. Iso.
RB337	8/22/2013	*P. virginiana*	X		X	X
RB338	8/22/2013	*P. echinata*			X	X
RB339	8/22/2013	*P. echinata*	X		X	X
RB341	8/22/2013	*P. virginiana*			X	X
RB342	9/9/2013	*P. echinata*			X	X
RB343	9/9/2013	*P. rigida*	X		X	X
RB344	9/9/2013	*P. rigida*	X		X	X
RB345	9/9/2013	*P. rigida*			X	X
RB346	9/9/2013	*P. virginiana*			X	X
RB347	9/9/2013	*P. virginiana*	X		X	X
RB348	9/9/2013	*P. virginiana*			X	X
RB349	9/9/2013	*P. echinata*	X			
RB353	9/9/2013	*P. echinata*	X			
RB354	9/9/2013	*P. virginiana*			X	X
RB355	9/9/2013	*P. virginiana*	X			
RB357	9/9/2013	*P. virginiana*	X			
RB358	9/9/2013	*P. virginiana*	X			

Table A3.2. List of individuals with their natal host, variable length barcode, and Illumina index. The sequence of the barcoded adapters and primers containing the Illumina indexes are listed in Supplemental Tables A2.3 and A2.4.

ID	Natal Host	Illumina Index	Barcode
LL002_01	*P. echinata*	GTGAAACG	GAGCGACAT
LL003_01	*P. echinata*	GTGAAACG	CCTTGCCATT
LL004_01	*P. rigida*	ATCACGAT	TGGCACAGA
LL005_01	*P. rigida*	TGACCAAT	TGGCAACAGA
LL006_02b	*P. rigida*	GTGAAACG	TCAGAGAT
LL027	*P. echinata*	CGATGTAT	TGGCAACAGA
LL028	*P. echinata*	TTAGGCAT	CTCTCGCAT
LL053	*P. virginiana*	TGACCAAT	TCTTGG
LL058_1	*P. virginiana*	TTAGGCAT	GAAGTG
LL058.2	*P. virginiana*	AGTCAACA	ACGGTACT
LL058_3R	*P. virginiana*	ATCACGAT	GCAAGCCAT
LL061	*P. virginiana*	TGACCAAT	ACCAGGA
LL062_1	*P. echinata*	CGATGTAT	TGACGCCA
LL062_2	*P. echinata*	TGACCAAT	GAGCGACAT
LL062_3R	*P. echinata*	GTCCGCAC	AACGTGCCT
LL064	*P. rigida*	TTAGGCAT	ACGGTACT
LL069_1R	*P. virginiana*	TTAGGCAT	TGCTT
LL069_2R	*P. virginiana*	ATCACGAT	TGACGCCA
LL069_3	*P. virginiana*	GTCCGCAC	GGAACGA
LL070_1	*P. rigida*	CGATGTAT	ATATCGCCA
LL070_2	*P. rigida*	TTAGGCAT	AACGCACATT
LL070_3R	*P. rigida*	CGATGTAT	AACGTGCCT
LL071_1	*P. echinata*	TGACCAAT	GGAACGA
LL071_2R	*P. echinata*	ATCACGAT	TATTCGCAT
LL072	*P. virginiana*	TTAGGCAT	TATGT
LL073_1	*P. echinata*	GTCCGCAC	GCAAGCCAT
LL073_2R	*P. echinata*	CGATGTAT	AAGACGCT
LL074.1R	*P. echinata*	AGTCAACA	TCAGAGAT
LL074_2	*P. echinata*	GTGAAACG	CTTGA
LL074_5	*P. echinata*	ATCACGAT	ACAACCAACT
LL078_1_Redo	*P. rigida*	TTAGGCAT	TCACGGAAG
LL078_2R	*P. rigida*	TTAGGCAT	TGGCAACAGA
LL078_3	*P. rigida*	TTAGGCAT	GGTGCACATT
LL092R_02	*P. virginiana*	GTGAAACG	AACGCACATT

Table A3.2 (cont.)

ID	Natal Host	Illumina Index	Barcode
LL102	*P. rigida*	TTAGGCAT	AACGCACATT
LL136	*P. virginiana*	TGACCAAT	TCACGGAAG
LL137	*P. virginiana*	CGATGTAT	CTCGCGG
LL139	*P. echinata*	CGATGTAT	GAGCGACAT
LL140	*P. virginiana*	GTGAAACG	TCACGGAAG
LL179	*P. echinata*	TTAGGCAT	GGAACGA
RB076_01	*P. virginiana*	GTGAAACG	TGACGCCA
RB122_01	*P. echinata*	TTAGGCAT	TAGCGGAT
RB126_01	*P. virginiana*	GTGAAACG	AACGTGCCT
RB127_01	*P. virginiana*	GTCCGCAC	CAGATA
RB128_01	*P. virginiana*	TTAGGCAT	GGCTTA
RB129_01	*P. echinata*	TTAGGCAT	TCAGAGAT
RB132C	*P. rigida*	ATCACGAT	CGTCGCCACT
RB135b	*P. virginiana*	CGATGTAT	GGTGCACATT
RB337	*P. virginiana*	TTAGGCAT	TCTTGG
RB339	*P. echinata*	GTCCGCAC	CCTTGCCATT
RB343_01	*P. rigida*	ATCACGAT	GGAACGA
RB344	*P. rigida*	AGTCAACA	CCTTGCCATT
RB347	*P. virginiana*	CGATGTAT	ATAGAT
RB349	*P. echinata*	GTGAAACG	ATTAT
RB353	*P. echinata*	AGTCAACA	TATGT
RB355	*P. virginiana*	AGTCAACA	GGTGT
RB357	*P. virginiana*	GTCCGCAC	CCACTCA
RB358	*P. virginiana*	TTAGGCAT	CTTGA

Table A3.3 – Missing data and proportion of heterozygous sites per individual. We determined the total percent of SNPs missing and the proportion of heterozygous sites after excluding SNPs present in >70% of individuals and pruning to 1 SNP per RAD locus. Seven individuals (‡) missing more than 90% of SNPs were dropped from subsequent analysis. We determined ploidy of each individual using the proportion of heterozygous sites, and excluded and two individuals with markedly low proportions of heterozygous sites (mean value: 0.003).

ID	% Missing Data	Proportion Heterozygous Sites
LL002_01	0.42	0.2887
LL003_01*	5.13	0.0033
LL004_01†	1.11	0.1657
LL005_01†	0.64	0.2839
LL006_02b	1.67	0.2826
LL027‡	100	N/A
LL028	0.74	0.1797
LL053	0.92	0.2835
LL058_1	42.57	0.2565
LL058.2	3.00	0.2971
LL058_3R	0.20	0.2625
LL061	0.56	0.2736
LL062_1	29.21	0.2269
LL062_2	0.32	0.2646
LL062_3R	0.45	0.2758
LL064	0.42	0.2544
LL069_1R	4.85	0.2517
LL069_2R	0.29	0.2706
LL069_3	0.69	0.2229
LL070_1	33.08	0.2594
LL070_2	2.53	0.2459
LL070_3R	3.36	0.2515
LL071_1	0.42	0.2879
LL071_2R	0.51	0.2286
LL072	5.89	0.2036
LL073_1	5.17	0.2582
LL073_2R	23.52	0.2379

Table A3.3 (cont.)

ID	% Missing Data	Proportion Heterozygous Sites
LL074.1R	0.36	0.2705
LL074_2	0.31	0.2708
LL074_5	0.40	0.2587
LL078_1_Redo	35.88	0.2643
LL078_2R	58.65	0.2657
LL078_3	0.65	0.2693
LL092R_02	1.47	0.2410
LL102	0.74	0.2749
LL136‡	95.61	0.2784
LL137	4.17	0.2500
LL139‡	97.45	0.2000
LL140	0.30	0.2693
LL179‡	93.37	0.2481
RB076_01	0.81	0.2897
RB122_01*	2.69	0.0035
RB126_01	0.55	0.2101
RB127_01	0.52	0.2427
RB128_01	8.76	0.2250
RB129_01	0.29	0.2433
RB132C	0.59	0.2524
RB135b‡	100	N/A
RB337	7.89	0.1930
RB339	0.81	0.2756
RB343_01	0.37	0.2686
RB344	0.56	0.2208
RB347	27.87	0.1869
RB349‡	96.66	0.2425
RB353‡	100	N/A
RB355	0.36	0.2516
RB357	0.90	0.2570
RB358	3.21	0.2575

References

Adams DC, Otárola-Castillo E (2013) Geomorph: An R package for the collection and analysis of geometric morphometric shape data (E Paradis, Ed,). *Methods in Ecology and Evolution*, **4**, 393–399.

Agrawal AA (2000) Host-range evolution: Adaptation and trade-offs in fitness of mites on alternative hosts. *Ecology*, **81**, 500–508.

Agrawal AF, Feder JL, Nosil P (2011) Ecological divergence and the origins of intrinsic postmating isolation with gene flow. *International Journal of Ecology*, **2011**, 1–15.

Alexander RD, Bigelow RS (1960) Allochronic speciation in field crickets, and a new species, Acheta veletis. *Evolution*, **14**, 334–346.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81–92.

Arnett RH (1993) *American Insects: A Handbook of the Insects of America North of Mexico*. Sandhill Crane Press, Gainesville, FL.

Arnold SJ (1992) Constraints on Phenotypic Evolution. *The American Naturalist*, **140**, S85–S107.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular ecology*, **22**, 3179–3190.

Atwood CE, Peck O (1943) Some native sawflies of the genus *Neodiprion* attacking pines in eastern Canada. *Canadian Journal of Research*, **21d**, 109–144.

Avise JC (2004) *Molecular Markers, Natural History, and Evolution*. Sinauer, Sunderland, MA.

Bagley RK, Sousa VC, Niemiller ML, Linnen CR (2017) History, geography and host use shape genomewide patterns of genetic variation in the redheaded pine sawfly (*Neodiprion lecontei*). *Molecular Ecology*, **26**, 1022–1044.

Barbosa P, Saunders JA (1985) Plant Allelochemicals: Linkages between Herbivores and Their Natural Enemies. In: *Chemically Mediated Interactions between Plants and Other Organisms*, pp. 107–137. Springer US, Boston, MA.

Barley AJ, Monnahan PJ, Thomson RC, Grismer LL, Brown RM (2015) Sun skink landscape genomics: assessing the roles of micro-evolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. *Molecular Ecology*, **24**, 1696–1712.

Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, **439**, 719–723.

Barrett R, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.

Bartlet E, Blight MM, Hick AJ, Williams IH (1993) The responses of the cabbage seed weevil (*Ceutorhynchus assimilis*) to the odour of oilseed rape (*Brassica napus*) and to some volatile isothiocyanates. *Entomologia Experimentalis et Applicata*, **68**, 295–302.

Bendall EE, Vertacnik KL, Linnen CR (2017) Oviposition traits generate extrinsic

postzygotic isolation between two pine sawfly species. *BMC Evolutionary Biology*, **17**, 26.
Benjamin DM (1955) The Biology and Ecology of the Red-Headed Pine Sawfly (F Service, Ed,). , 57.
Berenbaum M (1983) Coumarins and caterpillars: A case for coevolution. *Evolution*, **37**, 163.
Berenbaum M (1986) Postingestive Effects of Phytochemicals on Insects: On *Paracelsus* and Plant Products. In: *Insect-Plant Interactions*, pp. 121–153. Springer New York, New York.
Berlocher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annual Review of Entomology*, **47**, 773–815.
Bernays EA (1988) Host specificity in phytophagous insects: selection pressure from generalist predators. *Entomologia Experimentalis et Applicata*, **49**, 131–140.
Bernays EA (1989) Host range in phytophagous insects: The potential role of generalist predators. *Evolutionary Ecology*, **3**, 299–311.
Bernays EA (1991) Evolution of insect morphology in relation to plants. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **333**.
Bernays EA (1997) Feeding by lepidopteran larvae is dangerous. *Ecological Entomology*, **22**, 121–123.
Bernays E ?A. (1998) The value of being a resource specialist: Behavioral support for a neural hypothesis. *The American Naturalist*, **151**, 451–464.
Bernays EA (2001) Neural limitations in phytophagous insects: Implications for diet breadth and evolution of host affiliation. *Annual Review of Entomology*, **46**, 703–727.
Bernays EA, Chamberlain DJ, Leather EM (1981) Tolerance of acridids to ingested condensed tannin. *Journal of Chemical Ecology*, **7**, 247–256.
Bernays EA, Chapman RF (1977) Deterrent chemicals as a basis of oligophagy in *Locusta migratoria* (L.). *Ecological Entomology*, **2**, 1–18.
Bernays E, Chapman R (1987) The Evolution of Deterrent Responses in Plant-Feeding Insects. In: *Perspectives in Chemoreception and Behavior SE - 10*, pp. 159–173. Springer, New York, NY.
Bernays EA, Chapman RFE (1994) *Host-Plant Selection by Phytophagous Insects*. Chapman & Hall, New York.
Bernays EA, Cornelius ML (1989) Generalist caterpillar prey are more palatable than specialists for the generalist predator *Iridomyrmex humilis*. *Oecologia*, **79**, 427–430.
Bernays EA, Funk DJ (1999) Specialists make faster decisions than generalists: Experiments with aphids. *Proceedings of the Royal Society B: Biological Sciences*, **266**, 151–156.
Bernays E, Graham M (1988) On the evolution of host specificity in phytophagous arthropods. *Ecology*, **69**, 886–892.
Bernays EA, Wcislo WT (1994) Sensory capabilities, information processing, and resource specialization. *The Quarterly Review of Biology*, **69**, 187–204.
Berrigan D, Scheiner S (2004) Modeling the Evolution of Phenotypic Plasticity. In: *Phenotypic Plasticity* (eds DeWitt T, Scheiner S), pp. 82–97. Oxford University Press, Oxford.
Bickel PJ, Boley N, Brown JB, Huang H, Zhang NR (2010) Subsampling methods for

genomic inference. *The Annals of Applied Statistics*, **4**, 1660–1697.
Blanquart F, Kaltz O, Nuismer SL, Gandon S (2013) A practical guide to measuring local adaptation (D Ebert, Ed,). *Ecology Letters*, **16**, 1195–1205.
Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: Models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 459–487.
Bolnick DI, Otto SP (2013) The magnitude of local adaptation under genotype-dependent dispersal. *Ecology and Evolution*, **3**, 4722–4735.
Boughman JW (2002) How sensory drive can promote speciation. *Trends in Ecology & Evolution*, **17**, 571–577.
Boyes HA (1981) Food selection by some graminivorous Acrididae. University of Oxford.
Bradburd GS, Ralph PL, Coop GM (2013) Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, **67**, 3258–3273.
Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, **426**, 176–178.
Brower LP (1958) Bird predation and foodplant specificity in closely related procryptic insects. *The American Naturalist*, **92**, 183–187.
Brown JM, Abrahamson WG, Packer RA, Way PA (1995) The role of natural-enemy escape in a gallmaker host-plant shift. *Oecologia*, **104**, 52–60.
Burford Reiskind MO, Coyle K, Daniels H V. *et al.* (2016) Development of a universal double-digest RAD sequencing approach for a group of nonmodel, ecologically and economically important insect and fish taxa. *Molecular Ecology Resources*, **16**, 1303–1314.
Bush GL (1966) The taxonomy, cytology, and evolution of the genus *Rhagoletis* in North America (Diptera, Tephritidae). *Bulletin of the Museum of Comparitive Zoology*, **134**, 431–562.
Bush GL (1969a) Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera, Tephritidae). *Evolution*, **23**, 237.
Bush GL (1969b) Mating behavior, host specificity, and the ecological significance of sibling species in frugivorous flies of the genus *Rhagoletis* (Diptera-Tephritidae). *The American Naturalist*, **103**, 669–672.
Bush GL (1975a) Sympatric speciation in phytophagous parasitic insects. In: *Evolutionary Strategies of Parasitic Insects and Mites* (ed Price PW), pp. 187–206. Plenum Press, New York.
Bush GL (1975b) Modes of animal speciation. *Annual Review of Ecology and Systematics*, **6**, 339–364.
Butlin RK (2005) Recombination and speciation. *Molecular Ecology*, **14**, 2621–2635.
Butlin RK (2010) Population genomics and speciation. *Genetica*, **138**, 409–418.
Butlin RK, Galindo J, Grahame JW (2008) Sympatric, parapatric or allopatric: The most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 2997–3007.
Carvajal-Rodriguez A, Rolan-Alvarez E (2006) JMATING: A software for the analysis of sexual selection and sexual isolation effects from mating frequency data. *BMC evolutionary biology*, **6**, 40.

Casbon JA, Osborne RJ, Brenner S, Lichtenstein CP (2011) A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic Acids Research*, **39**, e81.
Case AL, Willis JH (2008) Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. *Evolution*, **62**, 1026–1039.
Castellano S, Balletto E (2002) Is the partial Mantel test inadequate? *Evolution*, **56**, 1871–1873.
Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular ecology*, **22**, 3124–40.
Charlesworth B (2012) The role of background selection in shaping patterns of molecular evolution and variation: Evidence from variability on the *Drosophila* X chromosome. *Genetics*, **191**, 233–46.
Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155–74.
Chen H, Rangasamy M, Tan SY, Wang H, Siegfried BD (2010) Evaluation of five methods for total DNA extraction from western corn rootworm beetles. *PloS one*, **5**, e11963.
Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA (2010) The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics*, **186**, 983–995.
Codella SG, Raffa KF (2002) Desiccation of Pinus foliage induced by conifer sawfly oviposition: Effect on egg viability. *Ecological Entomology*, **27**, 618–621.
Coffman AJ, Hsieh PH, Gravel S, Gutenkunst RN (2015) Computationally Efficient Composite Likelihood Statistics for Demographic Inference. *Molecular biology and evolution*, msv255-.
Connell J (1983) On the prevalence and relative importance of interspecific competition: Evidence from field experiments. *American Naturalist*, **122**, 661.
Coppel HC, Benjamin DM (1965) Bionomics of the Nearctic pine-feeding Diprionids. *Annual Review of Entomology*, **10**, 69–96.
Cornell H, Hawkins B (2003) Herbivore responses to plant secondary compounds: A test of phytochemical coevolution theory. *The American Naturalist*, **161**, 507–522.
Cottee PK (1984) A physiological investigation into the role of secondary plant compounds as feeding deterrents to *Locusta migratoria* and *Schistocerca gregaria*. University of Aberdeen.
Coyne JA, Orr HA (1989) Patterns of speciation in *Drosophila*. *Evolution*, **43**, 362–381.
Coyne JA, Orr HA (1997) "Patterns of speciation in *Drosophila*" revisited. *Evolution*, **51**, 295–303.
Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, MA.
Craig TP, Horner JD, Itami JK (1997) Hybridization studies on the host races of *Eurosta solidaginis*: Implications for sympatric speciation. *Evolution*, **51**, 1552–1560.
Craig TP, Horner JD, Itami JK (2001) Genetics, experience, and host-plant preference in Eurosta solidaginis: implications for host shifts and speciation. *Evolution; international journal of organic evolution*, **55**, 773–82.
Craig TP, Itami JK, Abrahamson WG, Horner JD (1993) Behavioral evidence for host-

race formation in *Eurosta solidaginis*. *Evolution*, **47**, 1696–1710.
Craig TP, Mopper S (1993) Sex Ratio Variation in Sawflies. In: *Sawfly Life History Adaptations to Woody Plants*, pp. 61–92. Academic Press, Inc., San Diego.
Crespi, Sandoval, Crespi BJ, Sandoval CP (2000) Phylogenetic evidence for the evolution of ecological specialization in *Timema* walking-sticks. *Journal of Evolutionary Biology*, **13**, 249–262.
Crispo E, Bentzen P, Reznick DN, Kinnison MT, Hendry AP (2006) The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology*, **15**, 49–62.
Crispo E, Hendry AP (2005) Does time since colonization influence isolation by distance? A meta-analysis. *Conservation Genetics*, **6**, 665–682.
Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. *Molecular ecology*, **19**, 3592–3602.
Damman H (1987) Leaf quality and enemy avoidance by the larvae of a Pyralid moth. *Ecology*, **68**, 88–97.
Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
Darwin C (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life*. John Murray, London.
Davey JW, Cezard T, Fuentes-Utrilla P *et al.* (2013) Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology*, **22**, 3151–3164.
Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.
Davidson J, Andrewartha HG (1948) The influence of rainfall, evaporation and atmospheric temperature on fluctuations in the size of a natural population of *Thrips imaginis* (Thysanoptera). *The Journal of Animal Ecology*, **17**, 200.
Delcourt PA, Delcourt HR (1981) Vegetation Maps for Eastern North America: 40,000 Years B.P. To the Present. In: *Geobotany II* (ed Romans RC), pp. 123–165. Plenum, New York.
Denno RF, McClure MS, Ott JR (1995) Interspecific interactions in phytophagous insects: Competition reexamined and resurrected. *Annual Review of Entomology*, **40**, 297–331.
Dethier VG (1941) Chemical factors determining the choice of food plants by *Papilio* larvae. *The American Naturalist*, **75**, 61–73.
Dethier VG (1954) Evolution of feeding preferences in phytophagous insects. *Evolution*, **8**, 33–54.
Diniz-Filho JAF, Soares TN, Lima JS *et al.* (2013) Mantel test in population genetics. *Genetics and molecular biology*, **36**, 475–85.
Dobzhansky T (1951) *Genetics and the Origin of Species*. Columbia University Press, New York.
Dopman EB, Pérez L, Bogdanowicz SM, Harrison RG (2005) Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 14706–11.
Dopman EB, Robbins PS, Seaman A (2010) Components of reproductive isolation

between North American pheromone strains of the European corn borer. *Evolution; international journal of organic evolution*, **64**, 881–902.
Drès M, Mallet J (2002a) Host races in plant–feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **357**.
Drès M, Mallet J (2002b) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **357**, 471–492.
Dukas R (1998) Constraints on Information Processing and their Effects on Behavior. In: *Cognitive Ecology* (ed Dukas R), pp. 89–127. University of Chicago Press, Chicago.
Eastop VP (1973) Deductions from the Present Day Host Plants of Aphids and Related Insects. In: *Insect/Plant Relationships* (ed Vanemden HF), pp. 157–178. Blackwell Scientific Publications, Oxford.
Eckert AJ, van Heerwaarden J, Wegrzyn JL *et al.* (2010) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (Pinus taeda L., Pinaceae). *Genetics*, **185**, 969–82.
Egan SP, Funk DJ (2009) Ecologically dependent postmating isolation between sympatric host forms of Neochlamisus bebbianae leaf beetles. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19426–31.
Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of Neochlamisus bebbianae leaf beetles. *Evolution*, **62**, 1162–1181.
Egan SP, Ragland GJ, Assour L *et al.* (2015) Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow (C Webb, Ed,). *Ecology Letters*, **18**, 817–825.
Ehrlich PR, Raven PH (1964) Butterflies and plants: A study in coevolution. *Evolution*, **18**, 586–608.
Endler JA (1986) *Natural selection in the wild*. Princeton University Press.
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, **14**, 2611–2620.
Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics*, **9**, e1003905.
Excoffier L, Lischer HEL (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.
Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–91.
Farrell BD (1998) “Inordinate Fondness” explained: Why are there So many beetles? *Science*, **281**, 555–559.
Farrell BD, Mitter C (1994) Adaptive radiation in insects and plants: Time and opportunity. *American Zoologist*, **34**, 57–69.
Feder JL (1998) The Apple Maggot Fly, *Rhagoletis pomonella*: Flies in the Face of Conventional Wisdom about Speciation? In: *Endless Forms: Species and Speciation*

(eds Howard DJ, Berlocher SH), pp. 130–144. Oxford University Press, New York.
Feder JL, Berlocher SH, Roethele JB *et al.* (2003a) Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 10314–9.
Feder JL, Chilcote CA, Bush GL (1988) Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature*, **336**, 61–64.
Feder JL, Egan SP, Nosil P (2012a) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.
Feder JL, Filchak KE (1999) It's about time: The evidence for host plant-mediated selection in the apple maggot fly, *Rhagoletis pomonella*, and its implications for fitness trade-offs in phytophagous insects. *Entomologia Experimentalis et Applicata*, **91**, 211–225.
Feder JL, Forbes A a. (2007) Habitat avoidance and speciation for phytophagous insect specialists. *Functional Ecology*, **21**, 585–597.
Feder JL, Gejji R, Yeaman S, Nosil P (2012b) Establishment of new mutations under divergence and genome hitchhiking. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 461–74.
Feder JL, Hunt TA, Bush L (1993) The effects of climate, host plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella*. *Entomologia Experimentalis et Applicata*, **69**, 117–135.
Feder JL, Nosil P (2010) The efficacy of divergence hitchiking in generating genomic islands during ecological speciation. *Evolution*, **64**, 1729–1747.
Feder JL, Nosil P, Wacholder AC *et al.* (2014) Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *The Journal of Heredity*, **105**, 810–820.
Feder JL, Opp SB, Wlazlo B *et al.* (1994) Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 7990–4.
Feder JL, Roethele JB, Filchak K, Niedbalski J, Romero-Severson J (2003b) Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, Rhagoletis pomonella. *Genetics*, **163**, 939–53.
Feder JL, Roethele JB, Wlazlo B, Berlocher SH (1997a) Selective maintenance of allozyme differences among sympatric host races of the apple maggot fly. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 11417–21.
Feder JL, Stolz U, Lewis KM *et al.* (1997b) The effects of winter length on the genetics of apple and hawthorn races of *Rhagoletis pomonella* (Diptera: Tephritidae). *Evolution*, **51**, 1862.
Felsenstein J (1981) Skepticism towards Santa-Rosalia or why are there so few kinds of animals. *Evolution*, **35**, 124–138.
Filchak KE, Roethele JB, Feder JL (2000) Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature*, **407**, 739–742.
Fisher RA (1921) On the "Probable Error" of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3–32.
Fishman L, Willis JH (2006) A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. *Evolution*, **60**, 1372–81.

Fitzpatrick BM, Fordyce JA, Gavrilets S (2008) What, if anything, is sympatric speciation? *Journal of Evolutionary Biology*, **21**, 1452–1459.
Flaxman SM, Feder JL, Nosil P (2012) Spatially explicit models of divergence and genome hitchhiking. *Journal of evolutionary biology*, **25**, 2633–2650.
Flaxman SM, Feder JL, Nosil P (2013) Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution*, **67**, 2577–2591.
Flaxman SM, Wacholder AC, Feder JL, Nosil P (2014) Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, **23**, 4074–4088.
Forbes AA, Devine SN, Hippee AC *et al.* (2017) Revisiting the particular role of host shifts in initiating insect speciation. *Evolution*, **71**, 1126–1137.
Forbes AA, Powell THQ, Stelinski LL, Smith JJ, Feder JL (2009) Sequential sympatric speciation across trophic levels. *Science*, **323**.
Forister ML, Dyer LA, Singer MS, Stireman III JO, Lill JT (2012) Revisiting the evolution of ecological specialization, with emphasis on insect–plant interactions. *Ecology*, **93**, 981–991.
Fox CW, Lalonde RG (1993) Host confusion and the evolution of insect diet breadths. *Oikos*, **67**, 577.
Fox LR, Morrow PA (1981) Specialization: Species property or local phenomenon? *Science*, **211**, 887–893.
Fraenkel G (1953) The nutritional value of green plants for insects. *Transactions of the IX International Congress of Entomology*, **2**, 90–100.
Fraenkel G (1959) Raison d'etre of secondary plant substances. *Science*, **129**, 1466–1470.
Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493–505.
Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular Ecology*, **19**, 3773–3788.
Freelander W (1983) Parasites and the coexistance of animal host species. *The American Naturalist*, **121**, 223–236.
Friberg M, Posledovich D, Wiklund C (2015) Decoupling of female host plant preference and offspring performance in relative specialist and generalist butterflies. *Oecologia*, **178**, 1181–1192.
Fry JD (1990) Trade-offs in fitness on different hosts: Evidence from a selection experiment with a phytophagous mite. *The American Naturalist*, **136**, 569–580.
Fry JD (1996) The evolution of host specialization: Are trade-offs overrated? *The American Naturalist*, **148**, S84–S107.
Funk DJ (1998) Isolating a role for natural selection in speciation: Host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. *Evolution*, **52**, 1744–1759.
Funk DJ, Egan SP, Nosil P (2011) Isolation by adaptation in Neochlamisus leaf beetles: host-related selection promotes neutral genomic divergence. *Molecular Ecology*, **20**, 4671–4682.
Funk DJ, Filchak KE, Feder JL (2002) Herbivorous insects: model systems for the comparitive study of speciation ecology. *Genetica*, **116**, 251–267.
Funk DJ, Nosil P (2008) Comparative analyses of ecological speciation. In:

Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects (ed Tilmon K), pp. 117–135.
Funk DJ, Nosil P, Etges WJ (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 3209–3213.
Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 397–423.
Futuyma DJDJ (1983) Selective factors in the evolution of host choice by phytophagous insects. In: *Herbivorous Insects: Host-Seeking Behavior and Mechanisms* (ed Ahmad S), pp. 227–244. Academic Press, Inc., New York.
Futuyma DJ, Cort RP, van Noordwijk I (1984) Adaptation to host plants in the fall cankerworm (*Alsophila pometaria*) and its bearing on the evolution of host affiliation in phytophagous insects. *The American Naturalist*, **123**, 287–296.
Futuyma DJ, Keese MC, Funk DJ (1995) Genetic Constraints on Macroevolution: The Evolution of Host Affiliation in the Leaf Beetle Genus Ophraella. *Evolution*, **49**, 797.
Futuyma DJ, Mayer GC (1980) Non-allopatric speciation in animals. *Systematic Biology*, **29**, 254–271.
Futuyma DJ, McCafferty SS (1990) Phylogeny and the evolution of host plant associations in the leaf beetle genus *Ophraella* (Coleoptera, Chrysomelidae). *Evolution*, **44**, 1885–1913.
Futuyma DJ, Moreno G (1988) Evolution of ecological specialization. *Annual Review of Ecology and Systematics*, **19**, 207–233.
Futuyma DJ, Philippi TE (1987) Genetic variation and covariation in responses to host plants by *Alsophila pometaria* (Lepidoptera: Geometridae). *Evolution*, **41**, 269–279.
Futuyma DJ, Wasserman SS (1981) Food plant specialization and feeding efficiency in the tent caterpillars *Malacosoma disstria* and *M. americanum*. *Entomologia Experimentalis et Applicata*, **30**, 106–110.
Gautier M, Gharbi K, Cezard T *et al.* (2013) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, **22**, 3165–3178.
Gavrilets S (2004) *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton, NJ.
Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences*, **102**, 18040–18045.
Gazave E, Ma L, Chang D *et al.* (2014) Neutral genomic regions refine models of recent rapid human population growth. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 757–62.
Gelman A, Hill J (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.
Geron CD, Arnts RR (2010) Seasonal monoterpene and sesquiterpene emissions from *Pinus taeda* and *Pinus virginiana*. *Atmospheric Environment*, **44**, 4240–4251.
Ghent AW, Wallace DR (1958a) Oviposition behavior of the Swaine jack-pine sawfly. *Forest Science*, **4**, 264–272.
Ghent AW, Wallace DR (1958b) Oviposition behavior of the swain Jack-pine sawfly.

Forest Science, **4**, 264–272.
Gilbert LE, Singer MC (1975) Butterfly ecology. *Annual Review of Ecology and Systematics*, **6**, 365–395.
Gnirke A, Melnikov A, Maguire J *et al.* (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, **27**, 182–189.
Godbout J, Jaramillo-Correa JP, Beaulieu J, Bousquet J (2005) A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (Pinus banksiana), a broad-range North American conifer. *Molecular Ecology*, **14**, 3497–3512.
Gompert Z, Messina FJ (2016) Genomic evidence that resource-based trade-offs limit host-range expansion in a seed beetle. *Evolution*, **70**, 1249–1264.
Görür G (2003) Phenotypic plasticity of morphological characters in cabbage aphid reared on both radish and cabbage. *Italian Journal of Zoology*, **70**, 301–303.
Gould F (1979) Rapid host range evolution in a population of the phytophagous mite *Tetranychus urticae* Koch. *Evolution*, **33**, 791.
Gourbière S, Mallet J (2010) Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball." *Evolution*, **64**, 1–24.
Gripenberg S, Mayhew PJ, Parnell M, Roslin T (2010) A meta-analysis of preference-performance relationships in phytophagous insects. *Ecology letters*, **13**, 383–93.
Groman JD, Pellmyr O (2000) Rapid evolution and specialization following host colonization in a yucca moth. *Journal of Evolutionary Biology*, **13**, 223–236.
Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods in Ecology and Evolution*, **4**, 336–344.
Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS genetics*, **5**, e1000695.
Hahn MW, White BJ, Muir CD, Besansky NJ (2012) No evidence for biased co-transmission of speciation islands in Anopheles gambiae. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 374–84.
Hanfling B, Weetman D (2006) Concordant genetic estimators of migration reveal anthropogenically enhanced source-sink population structure in the river sculpin, *Cottus gobio*. *Genetics*, **173**, 1487–1501.
Hardy OJ, Vekemans X (2002) SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.
Harmon LJ, Glor RE (2010) Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution*, **64**, 2173–2178.
Harper KE, Bagley RK, Thompson KL, Linnen CR (2016) Complementary sex determination, inbreeding depression and inbreeding avoidance in a gregarious sawfly. *Heredity*, **117**, 326–335.
Harris K, Nielsen R (2013) Inferring demographic history from a spectrum of shared haplotype lengths (JD Jensen, Ed,). *PLoS Genetics*, **9**, e1003521.
Hartl DL, Clark AG (2007) *Principles of Population Genetics*. Sinauer Associates, Sunderland, MA.
Hawthorne DJ, Via S (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature*, **412**, 904–7.

Heads PA, Lawton JH (1984) Bracken, ants and extrafloral nectaries. II. The effect of ants on the insect herbivores of bracken. *The Journal of Animal Ecology*, **53**, 1015.
Hendry AP (2009) Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 1383–1398.
Hennig W (1966) *Phylogenetic Systematics [Translated]* (D Davis, R Zangerl, Eds,). University of Illinois Press, Champaign, IL.
Hewitt G (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.
Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.
Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
Hoffmann AA, Rieseberg LH (2008) Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annual review of ecology, evolution, and systematics*, **39**, 21–42.
Holderegger R, Wagner HH (2006) A brief guide to Landscape Genetics. *Landscape Ecology*, **21**, 793–796.
Hosking L, Lumsden S, Lewis K *et al.* (2004) Detection of genotyping errors by Hardy–Weinberg equilibrium testing. *European Journal of Human Genetics*, **12**, 395–399.
Howard DJ, Gregory PG, Chu J, Cain ML, Apr N (1998) Conspecific sperm precedence is an effective barrier to hybridization between closely related species. *Society*, **52**, 511–516.
Hubby JL, Lewontin RC (1966) A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in Drosophila pseudoobscura. *Genetics*, **54**, 577–594.
Hutchinson DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.
Jaenike J (1978) On optimal oviposition behavior in phytophagous insects. *Theoretical Population Biology*, **14**, 350–356.
Jaenike J (1981) Criteria for ascertaining the existence of host races. *The American Naturalist*, **117**, 830–834.
Jaenike J (1985) Parasite pressure and the evolution of amanitin tolerance in *Drosophila*. *Evolution*, **39**, 1295–1301.
Jaenike J (1990) Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics*, **21**, 243–273.
Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.
James AC, Jakubczak J, Riley MP, Jaenike J (1988) On the causes of monophagy in *Drosophila quinaria*. *Evolution*, **42**, 626–630.
Janz N (2002) Evolutionary Ecology of Oviposition Strategies. In: *Chemoecology of Insect Eggs and Egg Deposition*, pp. 349–376. Blackwell Publishing Ltd, Oxford, UK.
Janz N, Nylin S (1997) The role of female search behaviour in determining host plant range in plant feeding insects: A test of the information processing hypothesis. *Proceedings of the Royal Society B: Biological Sciences*, **264**, 701–707.

Janz N, Nylin S (1998) Butterflies and plants: A phylogenetic study. *Evolution*, **52**, 486.
Janz N, Nylin S (2008) The Oscillation Hypothesis of Host-Plant Range and Speciation. In: *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects*, pp. 203–215. University of California Press, Berkeley, CA.
Janz N, Nylin SS, Wahlberg N *et al.* (2006) Diversity begets diversity: Host expansions and the diversification of plant-feeding insects. *BMC Evolutionary Biology*, **6**, 4.
Janzen DH (1980) When is it coevolution? *Evolution*, **34**, 611–612.
Janzen DH (1985) On ecological fitting. *Oikos*, **45**, 308–310.
Jaramillo-Correa JP, Beaulieu J, Khasa DP, Bousquet J (2009) Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes. *Canadian Journal of Forest Research*, **39**, 286–307.
Jeffries MJ, Lawton JH (1984) Enemy free space and the structure of ecological communities. *Biological Journal of the Linnean Society*, **23**, 269–286.
Jenkins DG, Carey M, Czerniewska J *et al.* (2010) A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography*, **33**, 315–320.
Jermy T (1984) Evolution of insect/host plant relationships. *The American Naturalist*, **124**, 609–630.
Jermy BT (1985) Is there competition between phytophagous insects? *Journal of Zoological Systematics and Evolutionary Research*, **23**, 275–285.
Jermy T (1987) The role of experience in the host selection of phytophagous insects. In: *Perspectives in Chemoreception and Behavior* (eds Chapman RF, Bernays EA, Stoffolano JG), pp. 143–157. Springer, New York.
Jiggins CD, Naisbit RE, Coe RL, Mallet J (2001) Reproductive isolation caused by colour pattern mimicry. *Nature*, **411**, 302–305.
Jiggins CD, OR T, B L *et al.* (2008) Ecological speciation in mimetic butterflies. *BioScience*, **58**, 541.
Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, **11**, 94.
Joshi A, Thompson JN (1995) Trade-offs and the evolution of host specialization. *Evolutionary Ecology*, **9**, 82–92.
Kaplan I, Denno RF (2007) Interspecific interactions in phytophagous insects revisited: A quantitative assessment of competition theory. *Ecology letters*, **10**, 977–94.
Kapler JE, Benjamin DM (1960) The biology and ecology of the red-pine sawfly in Wisconsin. *Forest Science*, **6**, 253–268 .
Karban R (1986) Interspecific competition between folivorous insects on <i>Erigeron glaucus<i>. *Ecology*, **67**, 1063–1072.
Karban R (1989) Fine-scale adaptation of herbivorous thrips to individual host plants. *Nature*, **340**, 60–61.
Karowe DN (1990) Predicting host range evolution: Colonization of *Coronilla varia* by *Colias philodice* (Lepidoptera: Pieridae). *Evolution*, **44**, 1637–1647.
Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Keightley PD, Trivedi U, Thomson M *et al.* (2009) Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Research*, **19**, 1195–1201.
Keinan A, Mullikin JC, Patterson N, Reich D (2007) Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genetics*, **39**, 1251–1255.
Kelley ST, Farrell BD (1998) Is specialization a dead end? The phylogeny of host use in *Dendroctonus* bark beetles (Scolytidae). *Evolution*, **52**, 1731–1743.
Kelley ST, Farrell BD, Mitton JB (2000) Effects of specialization on genetic differentiation in sister species of bark beetles. *Heredity*, **84 (Pt 2)**, 218–27.
Kenny EM, Cormican P, Gilks WP *et al.* (2011) Multiplex target enrichment using DNA indexing for ultra-high throughput SNP detection. *DNA Research*, **18**, 31–38.
Kingsolver JG, Hoekstra HE, Hoekstra JM *et al.* (2001) The Strength of Phenotypic Selection in Natural Populations. *The American Naturalist*, **157**, 245–261.
Kirkpatrick M (2006) Chromosome Inversions, Local Adaptation and Speciation. *Genetics*, **173**, 419–434.
Kirkpatrick M, Barton NH (1997) The strength of indirect selection on female mating preferences. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 1282–6.
Knerer G (1983) Diapause strategies in Diprionid sawflies. *Naturwissenschaften*, **70**, 203–205.
Knerer G (1984) Diprionid Sawflies: Biological Topics and Rearing Techniques (Hymenoptera: Symphyta). *Bulletin of the Entomological Society of America*.
Knerer G (1993) Life history diversity in sawflies. In: *Sawfly Life History Adaptations to Woody Plants* (eds Wagner MR, Raffa KF), pp. 33–60. Academic Press, San Diego, CA.
Knerer G, Atwood CE (1972) Evolutionary trends in the subsocial sawflies belonging to the *Neodiprion abietis* complex (Hymenoptera: Tenthredinoidea). *American Zoologist*, **12**, 407–418.
Knerer G, Atwood CE (1973) Diprionid sawflies: Polymorphism and speciation. *Science*, **179**, 1090–1099.
Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: A program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, **15**, 1179–1191.
Kronforst MR (2008) Gene flow persists millions of years after speciation in *Heliconius* butterflies. *BMC Evolutionary Biology*, **8**, 98.
Kuchta S, Tan A (2005) Isolation by distance and post-glacial range expansion in the rough-skinned newt, Taricha granulosa. *Molecular Ecology*, **14**, 255–244.
Kuwajima M, Kobayashi N, Katoh T, Katakura H (2010) Detection of ecological hybrid inviability in a pair of sympatric phytophagous ladybird beetles (*Henosepilachna* spp.). *Entomologia Experimentalis et Applicata*, **134**, 280–286.
Lande R (1982) Rapid origin of sexual isolation and character divergence in a cline. *Evolution*, **36**, 213.
Langerhans RB, Gifford ME, Joseph EO (2007) Ecological speciation in *Gambusia* fishes. *Evolution*, **61**, 2056–2074.
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature*

Methods, **9**, 357–359.
Laurent S, Pfeifer SP, Settles ML *et al.* (2015) The population genomics of rapid adaptation: disentangling signatures of selection and demography in white sands lizards. *Molecular ecology*.
Lawton JH, Strong. DR (1981) Community patterns and competition in folivorous insects. *The American Naturalist*, **118**, 317.
Leal M, Fleishman L (2004) Differences in visual signal design and detectability between allopatric populations of Anolis lizards. *The American Naturalist*, **163**, 26–39.
Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.
Leimar O, Hammerstein P, Van Dooren TJM (2006) A new perspective on developmental plasticity and the principles of adaptive morph determination. *The American naturalist*, **167**, 367–76.
Levins R, MacArthur R (1969) An hypothesis to explain the incidence of monophagy. *Ecology*, **50**, 910–911.
Lewontin RC, Hubby JL (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. *Genetics*, **54**, 595–609.
Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–95.
Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
Linn CE, Dambroski HR, Feder JL *et al.* (2004) Postzygotic isolating factor in sympatric speciation in *Rhagoletis* flies: Reduced response of hybrids to parental host-fruit odors. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 17753–17758.
Linn C, Feder JL, Nojima S *et al.* (2003) Fruit odor discrimination and sympatric host race formation in *Rhagoletis*. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 11490–3.
Linnen CR, Farrell BD (2008) Comparison of Methods for Species-Tree Inference in the Sawfly Genus *Neodiprion* (Hymenoptera: Diprionidae). *Systematic Biology*, **57**, 876–890.
Linnen CR, Farrell BD (2010) A test of the sympatric host race formation hypothesis in Neodiprion (Hymenoptera: Diprionidae). *Proceedings of the Royal Society B*, **277**, 3131–3138.
Lohmueller KE, Albrechtsen A, Li Y *et al.* (2011) Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome (JM Akey, Ed,). *PLoS Genetics*, **7**, e1002326.
Lu G, Bernatchez L (1999) Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): Support for the ecological speciation hypothesis. *Evolution*, **53**, 1491.
Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Lukic S, Hey J (2012) Demographic inference using spectral methods on SNP data, with an analysis of the human out-of-Africa expansion. *Genetics*, **192**, 619–39.
Maan ME, Seehausen O (2011) Ecology, sexual selection and speciation. *Ecology Letters*, **14**, 591–602.
MacDonald GM, Cwynar LC, Whitlock C (2000) The late Quaternary dynamics of pines in northern North America. In: *Ecology and Biogeography of Pinus* (ed Richardson DM), pp. 122–136. Cambridge University Press, Cambridge.
Mackenzie A (1996) A trade-off for host plant utilization in the black bean aphid, *Aphis fabae*. *Evolution*, **50**, 155–162.
Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution*, **28**, 614–621.
Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.
Mani GS, Clarke BC (1990) Mutational order: A major stochastic process in evolution. *Proceedings of the Royal Society of London B: Biological Sciences*, **240**.
Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.
Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, **26**, 448–456.
Marth GT, Czabarka E, Murvai J, Sherry ST (2004) The Allele Frequency Spectrum in Genome-Wide Human Variation Data Reveals Signals of Differential Demographic History in Three Large World Populations. *Genetics*, **166**, 351–372.
Matsubayashi KW, Kahono S, Katakura H (2011) Divergent host plant specialization as the critical driving force in speciation between populations of a phytophagous ladybird beetle. *Journal of Evolutionary Biology*, **24**, 1421–1432.
Matsubayashi KW, Katakura H (2009) CONTRIBUTION OF MULTIPLE ISOLATING BARRIERS TO REPRODUCTIVE ISOLATION BETWEEN A PAIR OF PHYTOPHAGOUS LADYBIRD BEETLES. *Evolution*, **63**, 2563–2580.
Matsubayashi KW, Ohshima I, Nosil P (2010) Ecological speciation in phytophagous insects. *Entomologia Experimentalis et Applicata*, **134**, 1–27.
Mayhew PJ (1997) Adaptive patterns of host-plant selection by phytophagous insects. *Oikos*, **79**, 417.
Mayr E (1942) *Systematics and the Origin of Species: From the Viewpoint of a Zoologist*. Harvard University Press, Cambridge, MA.
Mayr E (1963) *Animal Species and Evolution*. Belknap Press, Cambridge, MA.
McClure MS (1980) Competition between exotic species: Scale insects on hemlock. *Ecology*, **61**, 1391–1401.
McClure MS, Price PW (1975) Competition among sympatric *Erythroneura* leafhoppers (Homoptera: Cicadellidae) on American sycamore. *Ecology*, **56**, 1388–1397.
McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.
McPheron BA, Smith DC, Berlocher SH (1988) Genetic differences between host races of *Rhagoletis pomonella*. *Nature*, **336**, 64–66.
McVicker G, Gordon D, Davis C, Green P, Yang S (2009) Widespread genomic

signatures of natural selection in hominid evolution (MW Nachman, Ed,). *PLoS Genetics*, **5**, e1000471.
De Meester L, Gómez A, Okamura B, Schwenk K (2002) The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta Oecologica*, **23**, 121–135.
Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology*, **21**, 2839–2846.
Merrill RM, Gompert Z, Dembeck LM *et al.* (2011) Mate preference across the speciation continuum in a clade of mimetic butterflies. *Evolution*, **65**, 1489–1500.
Michel AP, Sim S, Powell THQ *et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9724–9729.
Middleton W (1921) Leconte's sawfly, an enemy of young pines. *Journal of Agricultural Research*, **20**, 741-760–93.
Mitter C, Farrell BD (1991) Macroevolutionary aspects of insect-plant relationships. In: *Insect-Plant Interactions, Volume III* (ed Bernays EA), pp. 35–78. CRC Press, Boca Raton.
Mitter C, Farrell B, Futuyma DJ (1991) Phylogenetic studies of insect-plant interactions: Insights into the genesis of diversity. *Trends in Ecology & Evolution*, **6**, 290–293.
Mitter C, Farrell BD, Wiegmann B (1988) The phylogenetic study of adaptive zones: Has phytophagy promoted insect diversification? *The American Naturalist*, **132**, 107–128.
Moran NA (1986) Morphological adaptation to host plants in*Uroleucon* (Homoptera: Aphididae). *Evolution*, **40**, 1044–1050.
Mousseau TA, Dingle H (1991) Maternal effects in insect life histories. *Annual Review of Entomology*, **36**, 511–534.
Mousseau TA, Fox CW (1998) The adaptive significance of maternal effects. *Trends in Ecology & Evolution*, **13**, 403–407.
Moyle LC (2006) Correlates of genetic differentiation and isolation by distance in 17 congeneric Silene species. *Molecular Ecology*, **15**, 1067–1081.
Muller HJ (1942) Isolating mechanisms, evolution and temperature. *Biological Symposia*, **6**, 71–125.
Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 343–53.
Nagel L, Schluter D (1998) Body size, natural selection, and speciation in sticklebacks. *Evolution*, **52**, 209.
Nerg A, Kainulainen P, Vuorinen M *et al.* (1994) Seasonal and geographical variation of terpenes, resin acids and total phenolics in nursery grown seedlings of Scots pine (Pinus sylvestris L.). *New Phytologist*, **128**, 703–713.
Noor MA, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 12084–8.
Nosil P (2002) Transitions between specialization and generalization in phytophagous insects. *Evolution*, **56**, 1701–1706.

Nosil P (2004) Reproductive isolation caused by visual predation on migrants between divergent environments. *Proceedings. Biological sciences*, **271**, 1521–8.
Nosil P (2007) Divergent host plant adaptation and reproductive isolation between ecotypes of Timema cristinae walking sticks. *The American naturalist*, **169**, 151–62.
Nosil P (2009) Adaptive population divergence in cryptic color-pattern following a reduction in gene flow. *Evolution*, **63**, 1902–1912.
Nosil P (2012) *Ecological Speciation*. Oxford University Press, Oxford.
Nosil P, Crespi BJ (2006a) Ecological divergence promotes the evolution of cryptic reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 991–997.
Nosil P, Crespi BJ (2006b) Experimental evidence that predation promotes divergence in adaptive radiation. *Proceedings of the National Academy of Sciences*, **103**, 9090–9095.
Nosil P, Crespi BJ, Sandoval CP (2002) Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*, **417**, 440–443.
Nosil P, Crespi BJ, Sandoval CP (2003) Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. *Proceedings. Biological sciences / The Royal Society*, **270**, 1911–8.
Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution*, **62**, 316–36.
Nosil P, Funk DJ, Ortiz-Barrientos D (2009a) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.
Nosil P, Harmon LJ, Seehausen O (2009b) Ecological explanations for (incomplete) speciation. *Trends in ecology & evolution*, **24**, 145–56.
Nosil P, Sandoval CP (2008) Ecological niche dimensionality and the evolutionary diversification of stick insects. *PloS one*, **3**, e1907.
Nosil P, Vines TH, Funk DJ (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, **59**, 705–719.
Nyman T, Vikberg V, Smith DR, Boevé J-L (2010) How common is ecological speciation in plant-feeding insects? A "Higher" Nematinae perspective. *BMC evolutionary biology*, **10**.
Ogden R, Thorpe RS (2002) Molecular evidence for ecological speciation in tropical habitats. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 13612–5.
Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, **22**, 5983–5999.
Ortiz-Barrientos D, Noor MAF (2005) Evidence for a one-allele assortative mating locus. *Science*, **310**, 1467–1467.
Ortíz-Barrientos D, Reiland J, Hey J, Noor MAF (2002) Recombination and the divergence of hybridizing species. *Genetica*, **116**, 167–78.
Papaj D (1990) Interference with learning in pipevine swallowtail butterflies: Behavioral constraint or possible adaptation. *Symposia Biologica Hungarica*, **39**, 89–101.
Patterson N, Price AL, Reich D, Reich D, Daly M (2006) Population structure and

eigenanalysis. *PLoS genetics*, **2**, e190.
Pavey SA, Collin H, Nosil P, Rogers SM (2010) The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences*, **1206**, 110–29.
Peccoud J, Ollivier A, Plantegenest M, Simon J-C (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 7495–500.
Peterson MA, Denno RF (1998) The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *The American Naturalist*, **152**, 428–46.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species (L Orlando, Ed,). *PLoS ONE*, **7**, e37135.
Pfennig DW, Wund MA, Snell-Rood EC *et al.* (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution*, **25**, 459–467.
Phillips PA, Barnes MM (1975) Host race formation among sympatric apple, walnut, and plum populations of the codling moth, *Laspeyresia pomonella*. *Annals of the Entomological Society of America*, **68**, 1053–1060.
Phung TN, Huber CD, Lohmueller KE *et al.* (2016) Determining the effect of natural selection on linked neutral divergence across species (JM Akey, Ed,). *PLOS Genetics*, **12**, e1006199.
Piálek J, Barton NH (1997) The spread of an advantageous allele across a barrier: the effects of random drift and selection against heterozygotes. *Genetics*, **145**, 493–504.
Pickett AD, Neary ME (1940) Further studies on *Rhagoletis pomonella* (Walsh). *Scientific Agriculture*, **20**, 551–556.
Powell THQ, Forbes AA, Hood GR, Feder JL (2014) Ecological adaptation and reproductive isolation in sympatry: Genetic and phenotypic evidence for native host races of *Rhagoletis pomonella*. *Molecular Ecology*, **23**, 688–704.
Price PW, Bouton CE, Gross P *et al.* (1980) Interactions among three trophic levels: Influence of plants on interactions between insect herbivores and natural enemies. *Annual Review of Ecology and Systematics*, **11**, 41–65.
Price CS, Kim CH, Gronlund CJ, Coyne JA (2001) Cryptic reproductive isolation in the *Drosophila simulans* species complex. *Evolution; international journal of organic evolution*, **55**, 81–92.
Price TD, Qvarnström A, Irwin DE (2003) The role of phenotypic plasticity in driving genetic evolution. *Proceedings. Biological sciences*, **270**, 1433–40.
Price PW, Westoby M, Rice B *et al.* (1986) Parasite mediation in ecological interactions. *Annual Review of Ecology and Systematics*, **17**, 487–505.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–59.
Prokopy RJ, Diehl SR, Cooley SS (1988) Behavioral evidence for host races in *Rhagoletis pomonella* flies. *Oecologia*, **76**, 138–147.
Puritz JB, Matz M V, Toonen RJ *et al.* (2014) Demystifying the RAD fad. *Molecular Ecology*, **23**, 5937–42.
R Core Team (2013) R: A language and environment for statistical computing.
Rafiński J, Babik W (2000) Genetic differentiation among northern and southern

populations of the moor frog Rana arvalis Nilsson in central Europe. *Heredity*, **84**, 610–618.
Ramsey J, Bradshaw HD, Schemske DW (2003) Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalsis* (Phyrmaceae). *Evolution*, **57**, 1520–1534.
Räsänen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, **11**, 624–636.
Rathcke BJ (1976) Competition and coexistence with a guild of herbivorous insects. *Ecology*, **57**, 76–87.
Raufaste N, Rousset F (2001) Are partial Mantel tests adequate? *Evolution*, **55**, 1703–1705.
Rausher MD (1984) Tradeoffs in performance on different hosts: Evidence from within- and between-site variation in the beetle *Deloyula guttata*. *Evolution*, **38**, 582–595.
Razmjou J, Vorburger C, Moharramipour S, Mirhoseini SZ, Fathipour Y (2010) Host-associated differentiation and evidence for sexual reproduction in Iranian populations of the cotton aphid, Aphis gossypii. *Entomologia Experimentalis et Applicata*, **134**, 191–199.
Reissig WH, Smith DC (1978) Bionomics of *Rhagoletis pomonella* in *Crataegus*. *Annals of the Entomological Society of America*, **71**, 155–159.
Rice WR, Hostert EE (1993) Laboratory experiments on speciation: What have we learned in 40 years? *Evolution*, **47**, 1637.
Rice WR, Linder JE, Friberg U *et al.* (2005) Inter-locus antagonistic coevolution as an engine of speciation: Assessment with hemiclonal analysis. *Proceedings of the National Academy of Sciences of the United States of America*, **102 Suppl**, 6527–34.
Rice WR, Salt GW (1990) The evolution of reproductive isolation as a correlated character under sympatric conditions: Experimental evidence. *Evolution*, **44**, 1140.
Richardson JL, Brady SP, Wang IJ, Spear SF (2016) Navigating the pitfalls and promise of landscape genetics. *Molecular Ecology*, **25**, 849–863.
Richmond GM, Fullerton DS (1986) Summation of quaternary glaciations in the United States of America. *Quaternary Science Reviews*, **5**, 183–196.
Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in ecology & evolution*, **16**, 351–358.
Rieseberg LH, Desrochers AM, Youn SJ (1995) Interspecific pollen competition as a reproductive barrier between sympatric species of *Helianthus* (Asteraceae). *American Journal of Botany*, **82**, 515.
Roesch Goodman K, Welter SC, Roderick GK *et al.* (2012) Genetic divergence is decoupled from ecological diversification in the Hawaiian *Nesosydne* planthoppers. *Evolution*, **66**, 2798–2814.
Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, **14**, 351–361.
Rogers S, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus sp.* Salmonidae) species pairs. *Molecular Biology and Evolution*, **24**, 1423–1438.
Rosenberg MS, Anderson CD (2011) PASSaGE: Pattern analysis, spatial statistics and geographic exegesis. Version 2. *Methods in Ecology and Evolution*, **2**, 229–232.

Rosenberg NA, Mahajan S, Ramachandran S *et al.* (2005) Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genetics*, **1**, e70.
Rosenberg NA, Pritchard JK, Weber JL *et al.* (2002) Genetic structure of human populations. *Science*, **298**.
Rossiter M (1987) Use of a secondary host by non-outbreak populations of the gypsy moth. *Ecology*, **68**, 857–868.
Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.
Rudolph TD, Yeatman CW (1982) Genetics of Jack Pine.
Rueffler C, Van Dooren TJM, Leimar O, Abrams PA (2006) Disruptive selection and then what? *Trends in Ecology & Evolution*, **21**, 238–245.
Rundle HDHD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.
Rundle HD, Whitlock MC (2001) A genetic interpretation of ecologically dependent isolation. *Evolution*, **55**, 198–201.
Ryan MJ, Rand AS (1993) Sexual selection and signal evolution: The ghost of biases past. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **340**.
Saint-Laurent R, Legault M, Bernatchez L (2003) Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchill). *Molecular Ecology*, **12**, 315–330.
Sambatti JBM, Ortiz-Barrientos D, Baack EJ, Rieseberg LH (2008) Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecology letters*, **11**, 1082–91.
Sandoval CP (1994) The effects of the relative geographic scales of gene flow and selection on morph frequencies in the walking stick *Timema christinae*. *Evolution*, **48**, 1866–1879.
Sandoval CP, Nosil P (2005) Counteracting selective regimes and host preference evolution in ecotypes of two species of walking-sticks. *Evolution; international journal of organic evolution*, **59**, 2405–13.
Sayar NP, Smith CA, White IM, Knio KM (2009) *Terellia fuscicornis* (Diptera: Tephritidae): Biological and morphological adaptation on artichoke and milk thistle. *Journal of Natural History*, **43**, 1159–1181.
Schilthuizen M (2001) *Frogs, Flies and Dandelions: The Making of a Species*. Oxford University Press, New York.
Schluter D (1996) Ecological speciation in postglacial fishes [and discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **351**.
Schluter D (1998) Ecological Causes of Speciation. In: *Endless Forms: Species and Speciation*, pp. 114–129. Oxford University Press, New York.
Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, New York.
Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, **16**, 372–380.
Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737–41.

Schmidtling RC (2001) Southern Pine Seed Sources. , 25.
Schmidtling R (2007) Genetic Variation in the southern pines: Evolution, migration and adaptation following the Pleistocene. , 28–32.
Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, **9**, 671–5.
Schoener TW (1983) Field experiments on interspecific competition. *The American Naturalist*, **122**, 240–285.
Schoonhoven LM, van Loon JJA, Dicke M (2005) *Insect-Plant Biology*. Oxford University Press.
Schwartz MK, McKelvey KS (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**, 441–452.
Schweyen H, Rozenberg A, Leese F (2014) Detection and removal of PCR duplicates in population genomic ddRAD studies by addition of a degenerate base region (DBR) in sequencing adapters. *Biological Bulletin*, **227**, 146–160.
Scriber JM (1979) The effects of sequentially switching foodplants upon biomass and nitrogen utilization by polyphagous and stenophagous *Papilio* larvae. *Entomologia Experimentalis et Applicata*, **25**, 203–215.
Seehausen O (2008) Progressive Levels of Trait Divergence along a "Speciation Transect" in the Lake Victoria Cichlid Fish *Pundamilia*. In: *Speciation and Patterns of Diversity* (eds Butlin RK, Bridle J, Schluter D), pp. 155–176. Cambridge University Press, Cambridge.
Sekhon J (2008) Multivariate and propensity score matching software with automated balance optimization: The Matching package for R. *Journal of Statistical Software*, **Forthcomin**.
Serre D, Pääbo S (2004) Evidence for gradients of human genetic diversity within and among continents. *Genome Research*, **14**, 1679–1685.
Servedio MR (2004) The evolution of premating isolation: Local adaptation and sexual selection against hybrids. *Evolution*, **58**, 913–924.
Servedio MR (2009) The role of linkage disequilibrium in the evolution of premating isolation. *Heredity*, **102**, 51–56.
Servedio MR, Doorn GS Van, Kopp M, Frame AM, Nosil P (2011) Magic traits in speciation: "Magic" but not rare? *Trends in Ecology & Evolution*, **26**, 389–397.
Servedio MR, Noor MAFMAF (2003) The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 339–364.
Sexton JP, Hangartner SB, Hoffmann A a (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, **68**, 1–15.
Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, **16**, 940–950.
Siemens DH, Johnson CD, Woodman RL (1991) Determinants of Host Range In Bruchid Beetles. *Ecology*, **72**, 1560–1566.
Singer MS (2008) Evolutionary Ecology of Polyphagy. In: *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects* (ed Tilmon KJ), pp. 29–42. University of California Press, Berkeley, CA.
Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
Smadja CM, Canbäck B, Vitalis R *et al.* (2012) Large-scale candidate gene scan reveals

the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution*, **66**, no-no.
Smith JM (1966) Sympatric Speciation. *The American Naturalist*, **100**, 637–650.
Smith DC (1988) Heritable divergence of *Rhagoletis pomonella* host races by seasonal asynchrony. *Nature*, **336**, 66–67.
Smith DR (1993) Systematics, Life History, and Distribution of Sawflies. In: *Sawfly Life History Adaptations to Woody Plants*, pp. 3–32. Academic Press, Inc., San Diego.
Smith SG, Wallace DR (1971) Allelic sex determination in a lower hymenopteran, Neodiprion nigroscutum Midd. *Canadian Journal of Genetics and Cytology*, **13**, 617–621.
Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.
Sokal RR (1979) Testing statistical significance of geographic variation patterns. *Systematic Zoology*, **28**, 227–232.
Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.
Soto IM, Carreira VP, Soto EM, Hasson E (2008) Wing morphology and fluctuating asymmetry depend on the host plant in cactophilic Drosophila. *Journal of Evolutionary Biology*, **21**, 598–609.
Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data: Modelling gene flow. *Nature Reviews Genetics*, **14**, 404–414.
Sousa V, Peischl S, Excoffier L (2014) Impact of range expansions on current human genomic diversity. *Current opinion in genetics & development*, **29**, 22–30.
Southwood TRE (1973) The insect/plant relationship – An evolutionary perspective. *Symposium of the Royal Entomological Society London*, **6**, 3–30.
Spurgin LG, Illera JC, Jorgensen TH, Dawson DA, Richardson DS (2014) Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? *Molecular Ecology*, **23**, 1028–39.
Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P (2009) The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics*, **182**, 205–216.
Stanley SM (1979) *Macroevolution: Pattern and Process*. WH Freeman and Company, San Francisco.
Stephens M (2007) Inference Under the Coalescent. In: *Handbook of Statistical Genetics* (eds Balding DJ, Bishop M, Cannings C), pp. 878–908. John Wiley and Sons, West Sussex, UK.
Stiling PD (1980) Competition and coexistence among *Eupteryx* leafhoppers (Hemiptera: Cicadellidae) occurring on stinging nettles (*Urtica dioica*). *The Journal of Animal Ecology*, **49**, 793.
Stireman JO, Nason JD, Heard SB (2005) Host-associated genetic differentiation in phytophagous insects: General phenomenon or isolated exceptions? Evidence from a goldenrod-insect community. *Evolution*, **59**, 2573.
Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the “landscape” in landscape genetics. *Heredity*, **98**, 128–142.
Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics:

Where are we now? *Molecular Ecology*, **19**, 3496–3514.
Strong DR, Lawton JH, Southwood TRE (1984) *Insects on Plants: Community Patterns and Mechanisms*. Harvard University Press, Cambridge, MA.
Svanbäck R, Pineda-Krch M, Doebeli M (2009) Fluctuating Population Dynamics Promotes the Evolution of Phenotypic Plasticity. *The American Naturalist*, **174**, 176–189.
Svensson EI, Abbott JK, Gosden TP, Coreau A (2009) Female polymorphisms, sexual conflict and limits to speciation processes in animals. *Evolutionary Ecology*, **23**, 93–108.
Szentesi A, Bernays EA (1984) A study of behavioural habituation to a feeding deterrent in nymphs of *Schistocerca gregaria*. *Physiological Entomology*, **9**, 329–340.
Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology*, **28**, 289–301.
Taylor EB, Foote CJ, Wood CC (1996) Molecular genetic evidence for parallel life-history evolution within a pacific salmon (sockeye salmon and kokanee, *Oncorhynchus nerka*). *Evolution*, **50**, 401.
Taylor RS, Friesen VL (2017) The role of allochrony in speciation. *Molecular Ecology*, **26**, 3330–3342.
Thibert-Plante X, Hendry AP (2009) Five questions on ecological speciation addressed with individual-based simulations. *Journal of Evolutionary Biology*, **22**, 109–123.
Thibert-Plante X, Hendry AP (2010) When can ecological speciation be detected with neutral loci? *Molecular Ecology*, **19**, 2301–2314.
Thibert-Plante X, Hendry AP (2011) The consequences of phenotypic plasticity for ecological speciation. *Journal of Evolutionary Biology*, **24**, 326–342.
Thompson JN (1988) Evolutionary ecology of the relationship between oviposition preference and performance of offspring in phytophagous insects. *Entomologia Experimentalis et Applicata*, **47**, 3–14.
Thompson JN (1994) *The Coevolutionary Process*. The University of Chicago Press, Chicago.
Thompson JN (1996) Trade-offs in larval performance on normal and novel hosts. *Entomologia Experimentalis et Applicata*, **80**, 133–139.
Towns J, Cockerill T, Dahan M *et al.* (2014) XSEDE: Accelerating scientific discovery. *Computing in Science & Engineering*, **16**, 62–74.
Usher BF, Feeny P (1983) Atypical secondary compounds in the family Cruciferae: Tests for toxicity to *Piers rapae*, an adapted crucifer-feeding insect. *Entomologia Experimentalis et Applicata*, **34**, 257–262.
Vacquier VD, Swanson WJ, Lee Y-H (1997) Positive Darwinian selection on two homologous fertilization proteins: What is the selective pressure driving their divergence? *Journal of Molecular Evolution*, **44**, S15–S22.
Verschaffelt E (1911) The cause determining the selection of food in some herbivorous insects. *Proc Acad. Sci. Amsterdam*, **13**, 536–542.
Vertacnik KL, Geib SM, Linnen CR (2016) *Neodiprion lecontei* genome assembly Nlec1.0. (dataset).
Via S (1984a) The quantitative genetics of an insect herbivore. II. Genetic correlations in larval performance within and among host plants. *Evolution*, **38**, 896–905.
Via S (1984b) The quantitative genetics of polyphagy in an insect herbivore. I. Genotype-

environment interaction in larval performance on different host plant species. *Evolution*, **38**, 881–895.
Via S (1991) The genetic structure of host plant adaptation in a spatial patchwork: Demographic variability among reciprocally transplanted pea aphid clones. *Evolution*, **45**, 827–852.
Via S (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution*, **53**, 1446–1457.
Via S (2001) Sympatric speciation in animals: The ugly duckling grows up. *Trends in ecology & evolution*, **16**, 381–390.
Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, **106**, 9939–9946.
Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 451–60.
Via S, Bouck AC, Skillman S (2000) Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution*, **54**, 1626–1637.
Via S, Conte G, Mason-Foley C, Mills K (2012) Localizing F(ST) outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow. *Molecular ecology*, **21**, 5546–60.
Via S, Hawthorne DJ (2002) The genetic architecture of ecological specialization: Correlated gene effects on host use and habitat choice in pea aphids. *The American Naturalist*, **159**, S76–S88.
Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.
Visser JH (1986) Host odor perception in phytophagous Insects. *Annual Review of Entomology*, **31**, 121–144.
Wagner M (1868) *Darwin'sche Theorie und das Migrationsgesetz der Organismen*. Duncker & Humblot, Liepzig.
Wagner M (1889) *Die Entstehung der Arten durch räumliche Sonderung. Gesammelte Aufsätze*. Benno Schwalbe, Basel.
Van Wagner CE (1967) Seasonal variation in moisture content of eastern Canadian tree foliage and the possible effect on crown fires. , **1204**.
Walsh BJ (1864) On phytophagous varieties and phytophagous species. *Proceedings of the Entomological Society of Philadelphia*, **3**, 404–430.
Walsh BJ (1867) The apple-worm and the apple maggot. *Journal of Horticulture*, **2**, 338–343.
Walter R, Epperson B (2001) Geographic pattern of genetic variation in Pinus resinosa: area of greatest diversity is not the origin of postglacial populations. *Molecular Ecology*, **10**, 103–111.
Walter R, Epperson BK (2005) Geographic pattern of genetic diversity in *Pinus resinosa*: Contact zone between descendants of glacial refugia. *American Journal of Botany*, **92**, 92–100.
Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, **67**, 3403–3411.

Wang IJ, Bradburd GS (2014) Isolation by Environment. *Molecular Ecology*, **23**, 5649–5662.
Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. (F Adler, Ed,). *Ecology Letters*, **16**, 175–182.
Wang H, Mcarthur ED, Sanderson SC *et al.* (1997) Narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentata*: Asteraceae). IV. Reciprocal transplant experiments. *Source: Evolution Evolution*, **51**, 95–102.
Wang IJ, Summers K (2010) Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology*, **19**, 447–458.
Webb T (1988) Eastern North America. In: *Vegetation History* (eds Huntley BJ, Webb T), pp. 385–414. Kluwer Academic Publishers, Dordrecht.
Wells O, Switzer G, Schmidtling R (1991) Geographic variation in Mississippi loblolly pine and sweetgum. *Silvae Genetica*, **40**, 105–119.
White MJD (1978) *Modes of Speciation*. WH Freeman and Company, San Francisco.
Wiegmann BM, Mitter C, Farrell BD (1993) Diversification of Carnivorous Parasitic Insects: Extraordinary Radiation or Specialized Dead End? *The American naturalist*, **142**, 737–754.
Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, **76**, 887–893.
Williams GC (1966) *Adaptation and Natural Selection*. Princeton University Press, Princeton, NJ.
Willing E-M, Dreyer C, van Oosterhout C, Poland J, Kawamoto K (2012) Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers (J Chave, Ed,). *PLoS ONE*, **7**, e42649.
Wilson LF, Wilkinson RC, Averill RC (1992) Redheaded Pine Sawfly - Its Ecology and Management. , **694**.
Winkler IS, Mitter C (2008) The Phylogenetic Dimension of Insect/Plant Interactions: A Review of Recent Evidence. In: *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insectsl*, pp. 240–263. University of California Press, Berkeley, CA.
Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123–3135.
Wright S (1943) Isolation by Distance. *Genetics*, **28**, 114–138.
Wright KM, Lloyd D, Lowry DB, Macnair MR, Willis JH (2013) Indirect evolution of hybrid lethality due to linkage with selected locus in <i>Mimulus guttatus<i> (NH Barton, Ed,). *PLoS Biology*, **11**, e1001497.
Wund M a. (2012) Assessing the Impacts of Phenotypic Plasticity on Evolution. *Integrative and Comparative Biology*, 1–11.
Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*, **175**, 1883–93.
Yeaman S, Otto SP (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution*, **65**, 2123–2129.

Vita

Birthplace

Somerville, NJ

Education

University of Kentucky
- Ph.D. Candidate, Department of Biology (Anticipated August 2017)
  - Advisor: Catherine R. Linnen
  - Dissertation title: *Examining the role of host use on divergence in the redheaded pine sawfly, Neodiprion lecontei, across multiple spatial scales*

Kean University
- M.A. Instruction and Curriculum (May 2011)
- B.S. Science and Technology (Biology Option) (May 2010)

Professional Experience

University of Kentucky
- Research Assistant, Linnen lab
  - July 2011-Present

Kean University
- Teaching Assistant, Zoology: Form and Function
  - January 2010-December 2010; May 2011-July 2011
- Teaching Assistant, Animal Behavior
  - January 2010-May 2010
- Research Assistant, Codella lab
  - May 2009-July 2011
- Research Assistant, Fasick lab
  - January 2009-January 2010

Professional Honors and Awards

- 2015-16 Graduate School Incentive Program Award ($1,000)
- 2014-15 Graduate School Incentive Program Award ($1,000)
- 2014 USDA NIFA-AFRI Pre-Doctoral Fellowship ($73,805)
- 2012 G. Flora Ribble Mini-Grant ($500)
- 2011-14 University of Kentucky Multi-Year Fellowship ($18,000/yr)
- 2011-14 Daniel R. Reedy Quality Achievement Fellowship ($18,000/yr)
- 2011 G. Flora Ribble Research Fellowship ($2,000)
- 2011 NSF Graduate Research Fellowship (Honorable Mention)
- 2010 NSF Graduate Research Fellowship (Honorable Mention)

Publications

R.K. Bagley, V.C. Sousa, M.L. Niemiller, and C.R. Linnen. 2017. History, geography and host use shape genomewide patterns of genetic differentiation in the redheaded pine sawfly (*Neodiprion lecontei*). *Molecular Ecology*, 26: 1022-1044. [Cover]

K.E. Harper*, **R.K. Bagley**, K.L. Thompson, and C.R. Linnen. 2016. Complementary sex determination, inbreeding depression and inbreeding avoidance in a gregarious sawfly. *Heredity*, 117: 326-335.

*Denotes undergraduate co-author

Oral Presentations

R.K. Bagley and C.R. Linnen. Host-associated divergence in a sympatric population of the red-headed pine sawfly (*Neodiprion lecontei*). Evolution 2016. Austin, TX. June 17-21, 2016.

R.K. Bagley, V.C. Sousa, M.L. Neimiller, and C.R. Linnen. History, geography, and host use shape genome-wide patterns of genetic variation in the redheaded pine sawfly (*Neodiprion lecontei*). Entomological Society of America Meeting, Minneapolis, MN. 15-18 November 2015.

R.K. Bagley, M.L. Neimiller, and C.R. Linnen. Population structure and demographic history of the redheaded pine sawfly, *Neodiprion lecontei*. Department of Biology 1st Annual Doctoral Candidate Seminar. EREC, University of Kentucky, Lexington, KY. 11 April 2015.

R.K. Bagley, M.L. Neimiller, and C.R. Linnen. Population structure and demographic history of the redheaded pine sawfly, *Neodiprion lecontei*. Department of Biology MCDB Graduate Student Talk Series. University of Kentucky, Lexington, KY. 11 November 2014.

R.K. Bagley, M.L. Neimiller, and C.R. Linnen. Contributions of ecology and geography to genetic differentiation among populations of the red-headed pine sawfly, *Neodiprion lecontei*. Midwest Ecology and Evolution Conference. University of Dayton, Dayton, OH. March 1-2, 2014.

R.K. Bagley and C. Linnen. Host shift speciation in *Neodiprion* sawflies: Preliminary evidence based on host-associated ecological and genetic divergence. Evolution 2012. Ottawa, Ontario, Canada. July 6-11, 2012.

R.K. Bagley and S.G. Codella. Correlated evolution of defensive characters in *Neodiprion* sawflies. 55th Annual New Jersey Academy of Science Meeting (Senior Academy, Undergraduate Session). Kean University, Union, NJ. April 24, 2010.

Poster Presentations

C.N. Anderson*, **R.K. Bagley**, and C.R. Linnen. Morphometric analysis of ovipositor structure between sawfly populations utilizing two northern pines. Showcase of Undergraduate Scholars. University of Kentucky, Lexington, KY. April 26, 2017.

R.K. Bagley and C.R. Linnen. Host-associated divergence in a sympatric population of the red-headed pine sawfly (*Neodiprion lecontei*). USDA NIFA AFRI Project Director's Meeting. Washington, D.C. August 30-31, 2016.

M.N. Hurst*, E.E. Bendall, **R.K. Bagley**, and C.R. Linnen. Morphometric analysis of ovipositor structure in three sympatric populations of the red-headed pine sawfly, *Neodiprion lecontei*. Evolution 2016. Austin, TX. June 17-21, 2016.

M.N. Hurst*, E.E. Bendall, **R.K. Bagley**, and C.R. Linnen. Morphometric analysis of ovipositor structure in three sympatric populations of the red-headed pine sawfly, *Neodiprion lecontei*. Showcase of Undergraduate Scholars. University of Kentucky, Lexington, KY. April 27, 2016.

J. Frederick†*, J. Wolfe†*, **R.K. Bagley**, and C.R. Linnen. Measuring host-shift speciation in *Neodiprion* sawflies: Can ecological divergence drive insect speciation? Showcase of Undergraduate Scholars. University of Kentucky, Lexington, KY. April 29, 2015. (†authors contributed equally)

K.E. Harper*, **R.K. Bagley**, and C.R. Linnen. A test of inbreeding avoidance in multiple populations of the redheaded pine sawfly, *Neodiprion lecontei*. Evolution 2014. Raleigh, NC. June 20-24, 2014.

R.K. Bagley and C.R. Linnen. Host shift speciation in *Neodiprion* sawflies: Preliminary evidence based on host-associated ecological and genetic divergence. Evolution 2013. Snowbird, UT. July 22-25, 2013.

K. Duong, A. Leonberger, **R.K. Bagley**, C.R. Linnen. Chemosensory genes in the draft genome of a generalized specialist, the red-headed pine sawfly (*Neodiprion lecontei*). Evolution 2013. Snowbird, UT. July 22-25, 2013.

M. Collins*, **R.K. Bagley**, C.R. Linnen. The genetic and environmental contributions to oviposition choice in the redheaded pine sawfly, *Neodiprion lecontei*. Evolution 2013. Snowbird, UT. July 22-25, 2013.

R.K. Bagley and S.G. Codella. Phylogenetic analysis of defense characters in conifer sawflies (Hymenoptera: Diprionidae). 47th Annual Meeting of the Animal Behavior Society. University of William and Mary, Williamsburg, VA. July 25-29, 2010.

R.K. Bagley and S.G. Codella. Correlated evolution of defensive characters in *Neodiprion* sawflies. 6th Annual NJCSTM Undergraduate Student Research Symposium. Kean University, Union, NJ. April 29, 2010.

R.K. Bagley and S.G. Codella. Correlated evolution of defensive characters in *Neodiprion* sawflies. Kean Research Day. Kean University, Union, NJ. April 8, 2010.

R.K. Bagley, W. Dong, M. Mongelli, and J.I. Fasick. Bimetallic, Ruthenium-based compounds as chemotherapeutic agents for pancreatic cancer. 12th Annual Undergraduate Research Symposium in the Chemical and Biological Sciences. University of Maryland, Baltimore County, Baltimore, MD. October 10, 2009.

A. Kimsey, **R.K. Bagley**, A. Delao, M. Mongelli, and J.I. Fasick. Asymmetric bimetallic ruthenium complex interactions with plasmid DNA and pancreatic and colon tumor cell lines. Undergraduate Poster Session, 238th American Chemical Society National Meeting, Washington, D.C. August 16-August 20, 2009.

R.K. Bagley, A. Delao, A. Kimsey, M. Mongelli, and J.I. Fasick. Bimetallic Ruthenium complexes inhibit growth and proliferation of CL-188 colon cancer and CRL-1682 pancreatic cancer cells. Kean Research Day. Kean University, Union, NJ. March 25, 2009.

*Denotes undergraduate co-author