POPULATION LEVEL GENETIC DIFFERENTIATION AMONG TEMPERATE *NEOCONOCEPHALUS* KATYDIDS

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

GIDEON NEY

Dr. Johannes Schul, Dissertation Supervisor

July 2017

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled

POPULATION LEVEL GENETIC DIFFERENTIATION AMONG TEMPERATE *NEOCONOCEPHALUS* KATYDIDS

presented by Gideon Ney,

a candidate for the degree of Doctor of Philosophy

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Johannes Schul

Professor Sarah Bush

Professor Chris Pires

Professor Gavin Conant

ACKNOWLEDGEMENTS

There are numerous people without whom this work would not have been possible. These people have provided both intellectual, as well as moral encouragement over the last six years and I will be forever grateful for their support. First I need to thank the members of the Schul Lab (Katy Frederick, Rhett Hartman, Megan Murphy, Nathan Harness, Luesoni Johnson, Anna Scharnagl, Kai Murphie, Nathan Thompson, and Montrai Spikes) who have been involved in nearly every part of this work. I sincerely appreciate the time we spent together in the field, in the lab, and out socializing.

I would like to thank the faculty and staff at the University of Missouri. Without who's logistical, intellectual, and moral support this work would not be possible. In particular I owe a debt of gratitude to the members of my dissertation committee (Johannes Schul, Sarah Bush, Chris Pires, and Gavin Conant), who have all taken the time to advise and encourage me both academically and personally. The staff of the Division of Biological Sciences has provided crucial support that allowed me to complete this dissertation.

Finally, my family and friends have supported me personally and contributed to this work in various ways. I cannot thank my family (Dad, Mom, Robin, Cara, Connor, Helen, and Mekonen) enough for the emotional and financial support through my many years in higher education. Immense gratitude to my parents whose example and encouragement is what has made me the person I am today. Thank you to my friends, both in the department and at large, who have kept me sane throughout this process. Lastly, I owe a huge debt of gratitude to my soon to be wife, Heather, for her endless support. Finding and marrying you has been my life's greatest accomplishment.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF FIGURES v

LIST OF TABLES vii

ABSTRACT viii

Chapter

1. POPULATION STRUCTURE WITHIN THE ONE-DIMENSIONAL RANGE OF A COASTAL PLAIN KATYDID 1

   Abstract 1

   Introduction 2

   Methods 4

   Results 10

   Discussion 15

   References 22

2. LOW GENETIC DIFFERENTIATION BETWEEN POPULATIONS OF AN ENDEMIC PRAIRIE KATYDID DESPITE HABITAT LOSS AND FRAGMENTATION 35

   Abstract 35

   Introduction 36

   Methods 38

   Results 45

   Discussion 48

   References 55

3. EPIGENETIC AND GENETIC VARIATION BETWEEN TWO BEHAVIORALLY ISOLATED SPECIES OF *NEOCONOCEPHALUS* KATYDID 74

Abstract 74

Introduction 75

Methods 78

Results 85

Discussion 88

References 93

4. A POST-PLEISTOCENE CALIBRATED MUTATION RATE FROM INSECT MUSEUM SPECIMENS 107

Abstract 107

Introduction 108

Methods 111

Results 114

Discussion 115

References 120

VITA 134

LIST OF FIGURES

Chapter

1. POPULATION STRUCTURE WITHIN THE ONE-DIMENSIONAL RANGE OF A COASTAL PLAIN KATYDID

Figure 1 26

Figure 2 28

Figure 3 29

Figure 4 30

Figure 5 31

Figure 6 33

2. LOW GENETIC DIFFERENTIATION BETWEEN POPULATIONS OF AN ENDEMIC PRAIRIE KATYDID DESPITE HABITAT LOSS AND FRAGMENTATION

Figure 1 62

Figure 2 63

Figure 3 64

Figure 4 65

Figure 5 68

Supplemental Figure 1 70

Supplemental Figure 2 71

3. EPIGENETIC AND GENETIC VARIATION BETWEEN TWO BEHAVIORALLY ISOLATED SPECIES OF *NEOCONOCEPHALUS* KATYDID

Figure 1 98

Figure 2 100

Figure 3 103

Figure 4 104

Figure 5 105

4. A POST-PLEISTOCENE CALIBRATED MUTATION RATE FROM INSECT MUSEUM SPECIMENS

Figure 1 127

Figure 2 129

Figure 3 130

Figure 4 131

Supplemental Figure 1 133

LIST OF TABLES

Chapter

1. POPULATION STRUCTURE WITHIN THE ONE-DIMENSIONAL RANGE OF A COASTAL PLAIN KATYDID

Table 1 27

Table 2 32

Supplemental Table 1 34

2. LOW GENETIC DIFFERENTIATION BETWEEN POPULATIONS OF AN ENDEMIC PRAIRIE KATYDID DESPITE HABITAT LOSS AND FRAGMENTATION

Table 1 66

Supplemental Table 1 69

Supplemental Table 2 73

3. EPIGENETIC AND GENETIC VARIATION BETWEEN TWO BEHAVIORALLY ISOLATED SPECIES OF *NEOCONOCEPHALUS* KATYDID

Table 1 99

Table 2 101

Table 3 102

Supplemental Table 1 106

4. A POST-PLEISTOCENE CALIBRATED MUTATION RATE FROM INSECT MUSEUM SPECIMENS

Table 1 128

Supplemental Table 1 132

POPULATION LEVEL GENETIC DIFFERENTIATION
AMONG TEMPERATE *NEOCONOCEPHALUS* KATYDIDS

Gideon Ney

Dr. Johannes Schul, Dissertation Advisor

ABSTRACT

Recent work indicates that radiation events may play a significant role in shaping species diversity across entire continents. Here we quantify population level genetic differentiation in several species of *Neoconocephalus* katydids in order to determine the mechanisms of genetic isolation across a continental scale. Patterns of genetic isolation can be generalized into four types: isolation by resistance (IBR), isolation by barrier (IBB), isolation by distance (IBD), and epigenetic incompatibility.

N. melanorhinus is a salt marsh specialist restricted to a narrow corridor along the Atlantic and Gulf coasts. IBD was the predominant pattern of variation across their range. In addition, we saw evidence of two possible biogeographic barriers to gene flow (IBB), one at the Atlantic-Gulf divide and the other along the Gulf coast.

We investigated the impact of IBR by comparing genetic differentiation between a habitat specialist, *N. bivocatus*, and a habitat generalist, *N. robustus*. Similar levels of genetic diversity and genetic differentiation were present within populations of both species. Genetic variation and epigenetic changes can diverge between populations in isolation. We found significant variation in total methylation levels between *N. bivocatus* and *N. robustus*. Genetic differentiation did a better job of explaining species-specific phenotypes than epigenetic differentiation. Epigenetic differentiation, although present between species, is likely the result of an interaction between genetic and epigenetic loci.

Chapter 1

POPULATION STRUCTURE WITHIN THE ONE-DIMENSIONAL RANGE OF A COASTAL PLAIN KATYDID

Published in PLOS One with Johannes Schul as coauthor

Abstract

Biogeography plays a significant role in species' dispersal, and in turn population structure, across the landscape. The North American katydid *Neoconocephalus melanorhinus* belongs to a genus with high mobility. Unlike other members of the genus, *N. melanorhinus* is a salt marsh specialist restricted to a narrow corridor along the Atlantic and Gulf coasts. In addition, their range crosses at least one known biogeographic barrier and possesses biogeographic characteristics of the stepping-stone as well as the hierarchical island model of dispersal. Using AFLP markers we searched for areas that conform to the predictions of isolation by distance and for areas of non-uniform increases in genetic variance, indicative of isolation by barrier. We found significant genetic differentiation between all twelve sampled sites. Isolation by distance was the predominant pattern of variation across their range. In addition, we saw possible evidence of two biogeographic barriers to gene flow, one at the Atlantic-Gulf divide and the other along the Gulf coast. We also observed a change in body size across the range. Body size, as measured by male hind femur length, correlated closely with latitude, a possible indication of differential selection across the species range.

Introduction

The phylogeography of organisms, i.e. the distribution of genetic lineages of a species across landscapes (Avise et al. 1987), is strongly affected by their dispersal ability (Peterson and Denno 1998). Isolation by distance (IBD) and isolation by barrier (IBB) are two common patterns found among numerous species and taxonomic groups (Avise 2000). Most species show dispersal limitations across some spatial scale (Greenwood 1980), i.e. individuals cannot disperse across their entire species' range during their lifetime. Such reduced dispersal can result in a pattern of uniformly increasing genetic variation with geographic distance (= IBD; Wright 1943). Alternatively, IBB occurs if biogeographic barriers to dispersal reduce gene flow. This often leads to a pattern of genetic discontinuity, i.e. a non-uniform increase in genetic differentiation across the landscape (Slatkin and Voelm 1991). While the difference between IBD and IBB is likely an artificial dichotomy, their study allows us to contemplate the relative importance that various factors play in shaping population structure.

A number of models explain the genetic structure resulting from different dispersal limitations (Wright 1943; Slatkin and Voelm 1991) such as distance or barriers. Within the stepping-stone model dispersal is limited by geographic distance and individuals disperse only between neighboring sites. In its most restricted form, sites fall along a line. Within this one-dimensional stepping-stone model, dispersal is limited to, at most, two adjacent sites. This produces a strong signature of IBD as gene flow decreases with distance along a narrow corridor (Wright 1943; Kimura and Weiss 1964).

In an alternative model, the hierarchical island model, barriers between certain neighboring sites reduce the rate of dispersal and gene flow. This produces a non-uniform increase in genetic differentiation between these sites, the signature of IBB (Slatkin and Voelm 1991). This divides the population structure into distinct subpopulations. Within subpopulations there is an assumption of little or no genetic differentiation (panmixia). This is the result of significant gene flow within, and reduced gene flow between, subpopulations. Unlike in the stepping-stone model, geographic distance does not play a role in this pattern of genetic differentiation.

Neoconocephalus is a diverse group of New World katydids with variation in habitat preference and species distributions (Greenfield 1990). All species are highly mobile with strong flying capabilities and adult life spans of several months (review in Greenfield 1990). Most species occupy large ranges, some of which encompass much of the Midwestern and Eastern United States. The North American ranges of these species seem to show little population level genetic structure (Frederick 2013). For example, *N. bivocatus* and *N. robustus* show a lack of genetic structure across more than 450 km of grassland habitat (Ney and Schul 2017).

Neoconocephalus melanorhinus is a habitat specialist found only in North American salt marshes along the Atlantic and Gulf coasts (Rhen and Hebard 1915). This preference results in a species distribution, unique among *Neoconocephalus* katydids, that is more than 2,000 km long, but no more than a few kilometers wide in most localities. This one-dimensional arrangement of populations, which is similar to those in a one-dimensional stepping-stone model, may limit dispersal to neighboring sites and reduce gene flow across the range. *Neoconocephalus melanorhinus'* range also falls along

several known biogeographic divides, including one dividing the range on either side of the Florida peninsula (Fig 1). This Atlantic-Gulf divide is a significant barrier to gene flow in many marine taxa found along the Southern United States coast (Avise 2000; Soltis et al. 2006). In addition, their range crosses numerous other geographic features, including estuaries, developed shoreline, and stretches of coastline lacking saltmarsh habitat, that could potentially function as biogeographic barriers. *Neoconocephalus melanorhinus'* narrow range may decrease individuals' ability to avoid barriers by going around them and result in greater genetic differentiation among subpopulations on either side.

Here we describe the population genetic structure of *N. melanorhinus*, a highly mobile species, in a one-dimensional range, which possesses characteristics of both a stepping-stone and a hierarchical island model. We sampled from twelve locales across the range. Using molecular markers and statistical tools, we searched for areas that conform to the predictions of IBD and for areas with non-uniform increases in genetic variance indicative of IBB. We note the confounding effect signals of IBD and IBB may have and utilized tools that control for these effects. In addition, we collected measurements for body size to determine whether this phenotype co-varies with the observed pattern of genetic differentiation across *N. melanorhinus'* range.

Methods

Sampling:

All necessary permissions were obtained prior to the collection of specimens (Cape May NWR; permit 14007). We collected a total of 227 *N. melanorhinus* from 12 salt

marshes sites (Table 1, Fig 1) along the US Atlantic and Gulf coasts. We localized males using their calls and collected them by hand within the 2-3 hours following sunset. Calling activity was greatest at dusk, with individuals ceasing calling activity within an hour of sunset. *Neoconocephalus melanorhinus* were found either on or in close proximity to *Juncus roemerianus* in the high marsh habitats South of the Chesapeake Bay. In the three most Northern sites (BV, CM, and SU) *J. roemerianus* was less common and individuals were predominantly found calling from cord grass. We identified individuals as belonging to the target species through their call and morphological features, including cone pigmentation and body shape (Rhen and Hebard 1915; Walker SINA).

In 2014, we collected 188 *N. melanorhinus* from all the Atlantic coast sites and the Gulf coast sites DH and CK (Table 1, Fig 1). After finding evidence of significant genetic differentiation along the Gulf, two additional sites were sampled (IB and CR) in 2015. One additional sample was added to site DH in 2015 as well. The thirty-nine samples added in 2015 were run with a subset of the 2014 samples (eight DH, ten CK), and of these, eight 2014 samples were run in replicate in 2015 (see below).

We removed one hind femur from each *N. melanorhinus* sampled and placed them in 95% EtOH for DNA preservation. We later extracted DNA from the hind femurs using the DNeasy Blood and Tissue Kit (Qiagen Inc., Valencia, CA, USA). DNA quantification was performed on each sample by spectrophotometry (NanoDrop 1000, Thermo Scientific, Wilmington, DE). We used amplified fragment length polymorphism (AFLP) markers to quantify genetic variation.

Quantification of AFLP loci and alleles:

AFLP bands were generated using the method described in Snyder et al. (2009). Briefly, we performed a double digest on whole genomic material using the enzymes EcoRI and MseI, followed by ligation of adapters onto the fragments' sticky ends. We then performed two rounds of selective amplification. Pre-selective polymerase chain reaction (PCR) was carried out using primers Eco+A and Mse+A (Snyder et al. 2009). We performed selective PCR with one of four MseI primers (Mse+ATA, Mse+AGA, Mse+AAC, Mse+AGC) and one of two fluorescently labeled EcoRI primers (Eco+AAC, 6FAM; Eco+AGC, PET) (as described in (Snyder et al. 2009)). The selective PCR products were genotyped at the University of Missouri's DNA Core Facility on an ABI 3730 genetic analyzer (Applied Biosystems Corporation, Foster City, CA, USA). We called AFLP bands using GeneMarker v1.6 (Softgenetic Corp, State College, PA, USA) using an automated peak-calling scheme (as described in (Holland et al. 2008)). Alleles were called between 50-500 bp with a minimum peak intensity of 50.

AFLP error analysis:

We tested the rate of genotyping errors between collection years to determine the viability of combining AFLP datasets. In order to test repeatability, eight 2014 samples were rerun in replicate with the 2015 samples. Fragment lengths were recalled for the original as well as both sets of rerun samples as described previously and the mismatch error rate was calculated within and between AFLP runs (Pompanon et al. 2005). Mismatch error rates between-runs in excess of within-run error rates would indicate additional between-run biases in fragment amplification or fragment calling, suggesting

the need for an independent treatment of datasets (Bonin et al. 2007).

The 2014 AFLP analysis produced 2170 distinct fragments from 188 individuals. The 2015 AFLP run produced 1175 fragments from fifty-seven individuals. Mismatch error rates averaged 6.1% for within-run peak calls and 14.0% for between-run peak calls. Because of the increase in mismatch error rates between AFLP runs we chose to not combine 2014 and 2015 datasets. By treating datasets independently we avoid added error incorporated by uncontrolled variations in fragment amplification or peak intensities between-runs.

Automated peak calling schemes produce a larger number of total fragments, but tend to produce larger error rates than manual peak calling schemes. Larger AFLP datasets, despite added noise, reveal greater genetic differentiation than smaller, less error prone, datasets (Zhang and Hare 2012). As genetic variation tends to be relatively low within *Neoconocephalus*, the automated peak-calling scheme is preferable due to its sensitivity in detecting lower levels of genetic variation. Significant patterns of genetic structure found in this study would indicate that these levels of within-run noise do not obscure the overall genetic signal.

Phylogenetic analysis:

A fifty percent rule Bayesian tree was constructed in MrBayes v.3.2 from individuals collected in 2014 (Ronquist et al. 2012). The MCMC settings for the MrBayes analysis were: two runs, each with eight chains for 20,000,000 generations, sampled every 10,000 generations, and sampled for a total of 2,000 trees per run. As AFLP fragments are dominant markers scored as present/absent, we utilized the binary

restriction site model and the 'noabsencesites' substitution model. Runs were determined to have reached a stationary distribution by viewing the log-likelihoods in Tracer v.1.5 (Rambaut and Drummond 2010).

Genetic differentiation:

We used four methods to quantify genetic differentiation across *N. melanorhinus'* range. First, genetic differentiation among localities was calculated using a Bayesian non-uniform allele frequency distribution method for each year's dataset independently (Zhivotovsky 1999) as implemented in the program AFLPsurv v1.0 (Vekemans 2002). Expected F_{ST} was estimated from 100,000 permutations (10 bootstrapped runs) of individuals among groups.

Second, a Bayesian clustering analysis of individuals from 2014, as implemented in the program STRUCTURE v.2.3.4 (Pritchard et al. 2000), produced assignment scores for individuals to a hypothesized number of clusters based on AFLP fragment data. In STRUCTURE the admixture model was implemented with a local prior, allele frequencies correlated, with a run length of 100,000 (Burnin=10,000) for 10 replicates each of K=1-12. The best supported K was estimated as described in (Evanno et al. 2005) and implemented in Structure Harvester v.0.6.94 (Earl 2012). The program Clumpp v.1.1.2 (Jakobsson and Rosenberg 2007) was used to align the 10 repetitions of the best supported number of clusters.

Third, multiple AMOVAs were performed using the 2014 and 2015 AFLP datasets. AMOVA is a method for testing hypotheses of population structure against genetic evidence (Cockerham 1969; Cockerham 1973). Sites were split into two groups at

every possible location along the coast resulting in nine AMOVAs among the ten sites sampled in 2014 and three AMOVAs among the four Gulf coast sites from the 2015 dataset. AMOVAs were conducted in Arlequin v.3.5 (Excoffier and Lischer 2010) using the locus-by-locus approach, averaged over all polymorphic loci.

Fourth, standard and partial Mantel tests were conducted for populations sampled in 2014 using the program GenoDive v.2.0 (Meirmans and Tienderen 2004) to examine the relationship between genetic differentiation ($F_{ST}/(1-F_{ST})$), genetic structure, and geographic distance. Geographic distance was measured as the simplified shoreline distance, the distance along a path connecting neighboring sites by straight lines. This path is likely similar to the path *N. melanorhinus* must disperse through, not necessarily following all of the contours of the coast, but also not crossing large stretches of non-marsh habitat. In addition, we conducted a linear model analysis of variance in R v.3.2.0 (R Core 2015) to test for differences in dispersal across predicted barrier sites by testing for significant differences in slopes and y-intercepts.

We produced three matrices representing the most well supported models of population structure formulated from the other analyses. Significance was calculated by 1,000 permutations of samples between groups. To directly take into account population structure, partial Mantel tests were performed measuring the relationship between genetic and geographic distance, while using the hypothesized genetic clustering as a covariate. In addition, partial Mantel tests examining the relationship between genetic differentiation and the three hypothesized population structures were run with geographic (simplified shoreline) distance as a covariate.

Phenotypic differentiation:

We measured body size using the length of the hind femur. After removing a single hind leg from each sampled individual, the length from the proximal end of the leg to the first joint was measured using a ruler to the nearest half-millimeter. As there is sexual dimorphism in *Neoconocephalus*, only the body measures from males were used in this analysis. In addition, as there may be an effect between years, we chose to compare only males collected in 2014. Hind femur length was correlated with latitude and the significance evaluated using Pearson's correlation coefficient. In addition, the difference in average male hind femur length between sites was correlated with genetic differentiation and evaluated using a Mantel test.

Results

Phylogenetic analysis:

A phylogenetic analysis of individuals collected from all sites in 2014 allowed us to determine a hypothesis of the ancestry among all individuals. Sites whose individuals share ancestry with individuals from other sites have likely seen significant gene flow in their past. Sites comprised of monophyletic clusters of individuals have likely remained isolated and have an ancestry less affected by gene flow. Individuals from South Atlantic sites did not form a monophyletic clade, but a large clade made up almost exclusively of North Atlantic (sites SU and CM) individuals was nested within the larger cluster of South Atlantic sites (Fig 2). This suggests that range expansion likely proceeded South to North along the Atlantic coast. The presence of individuals from the two Northern sites (SU and CM) intermixed with individuals from more Southern sites may indicate the

presence of a genetic source population in the North Atlantic, from which individuals may emigrate, but to which Southern individuals do not commonly immigrate.

All but two individuals from the Gulf coast sites (DH, CK) formed a monophyletic clade (Fig 2), nested within the larger Atlantic cluster of animals. This suggests that the Gulf population was likely established following a dispersal event from an ancestral Atlantic population. The Atlantic-Gulf divide likely acts as a barrier, reducing gene flow and maintaining genetic differentiation following dispersal to the Gulf (Soltis et al. 2006). Two CK individuals, however, fell outside of the Gulf clade and may represent recent migrants from the Atlantic. This suggests limited recent dispersal across the Atlantic-Gulf divide. Within the Gulf clade, individuals from the Western Gulf site DH all fell into a monophyletic cluster. The monophyletic clustering of individuals by site within both Gulf sites suggests the presence of reduced gene flow via IBB. A pattern of IBD better explains the intermixing of individuals observed between neighboring sites along the South Atlantic portion of the range.

Bayesian clustering analysis:

Our Bayesian clustering analysis, utilized to independently evaluate the grouping of individuals seen in the phylogeny, revealed a peak ΔK (Evanno et al. 2005) for $K=8$ genetic clusters among the 2014 individuals (Fig 3). We observed one genotype cluster found almost exclusively among individuals from the two Gulf sites (red in Fig 3). The lack of this genetic cluster among Atlantic individuals is an indication of hierarchical genetic differentiation between Gulf and Atlantic sites. Reduced dispersal across the Atlantic-Gulf divide is the most likely cause for this pattern.

We observed a second cluster (dark blue in Fig 3) mainly to individuals from Northern Atlantic sites. Its occurrence decreases to the South. This pattern may indicate the presence of a genetic cline and continued gene flow, rather than a barrier within the Atlantic range. The other six clusters did not assign to individuals from any particular site or region (Fig 3). The results of the STRUCTURE analysis showed support for the significance of the Atlantic-Gulf divide on *N. melanorhinus'* population structure, as well as an additional barrier along the Gulf. Genetic differentiation along the Atlantic appears to be clinal, with IBD alone shaping genetic structure.

AMOVA test of population structure:

To test our hypotheses of population structure against genetic differentiation, we tested for the grouping of sites into subpopulations, as would be expected under a model of IBB. Twelve hypothetical genetic structures were produced by dividing sites into two subpopulations at each break between sampled sites. Each hypothetical population structure, implemented from 2014 and 2015 samples, was tested using a separate AMOVA. For the 2014 dataset, among-group genetic variance was highest when Gulf site DH was assigned to one group and all other locations to another (12.86%, Fig 4). Thus, DH was genetically distinct from all the other sites and dispersal was limited along the Gulf coast. The addition of two Gulf sites in 2015 allowed us to determine that most of the Gulf's genetic variance is the result of reduced dispersal between sites IB and CR (5.88% among group variance, Fig 4). This pointed to the presence of a barrier between these sites, rather than a gradual increase in genetic variance with distance across the length of the Gulf coast, as would be expected under IBD.

In addition, we observed a relatively high level of among-group genetic variance at the site of the Atlantic-Gulf divide. All genetic structures subdividing sites along the Atlantic explained relatively little among-group variance. These patterns of genetic variance indicated that biogeographic barriers occurred between Gulf sites IB and CR, and at the Atlantic-Gulf divide.

To examine the relationship between population genetic distance and geographic distance, we examined the correlation between genetic differentiation ($F_{ST}/(1- F_{ST})$) and geographic distance, often used as an indication of IBD, and found a significant correlation (Mantel $R = 0.795$, $p < 0.001$) among the 2014 sites (Table 2). We subdivided pairwise F_{ST} values (Supplementary Table 1) into three groups based on sampling localities: within-Atlantic pairs (no barriers between, as indicated by previous analyses), Atlantic v. CK pairs (one barrier between), and Atlantic v. DH pairs (two barriers between). Within all three groups, genetic variation increased with geographic distance (Fig 5).

The slopes of the three linear correlations did not differ significantly from one another, based on a linear model analysis of variance. However, based on linear models, the groups' linear trend lines differed significantly in their y-intercepts, with the within-Atlantic line having the smallest and DH v. Atlantic the largest y-intercept. This evidence suggests the existence of barriers along the coast at the sites of both the proposed Atlantic-Gulf and Gulf divides. The similarity in slopes among the three subgroups indicates an impact of IBD across the entire range. We also examined the relationship between genetic differentiation and latitude. While this relationship was significant

(Mantel R = 0.3793; p = 0.010), geographic distance described more of the variation in genetic differentiation between sites.

We utilized Mantel tests to determine the significance of correlations between genetic distance, geographic distance, and several hypothesized population structures. We produced three hypothesized population structures based on the presence of one or both hypothesized barriers (Atlantic-Gulf divide and/or within Gulf barrier). We observed a significant correlation between genetic differentiation and the population structure model built with both barriers (Table 2 row 2). The Atlantic-Gulf divide only hypothesis also showed a significant correlation with genetic differentiation (Table 2 row 4), however the Gulf divide only hypothesis did not meet significance with a $p < 0.05$ criterion (Table 2 row 3).

Because of these significant correlations of genetic differentiation with both distance and hierarchical population structures described above, we performed several partial Mantel tests to control for the confounding effects between them. Partial Mantels allowed for the comparison of two variables, while controlling for the effects of a third. Genetic differentiation and geographic distance maintained a significant relationship when population structure was controlled for (Table 2 rows 5,7,9). However, the correlation between genetic differentiation and the hypothesized population structures lost significance when geographic distance was controlled for (Table 2 rows 6,8,10). The partial Mantel analyses indicated that a prediction of IBD alone was better supported than any of our hierarchical population structures.

Phenotypic differentiation:

We observed a change in body size across *N. melanorhinus'* range. Hind femur length showed a strong linear relationship with latitude (Fig 6, Pearson's p = -0.8233). Gulf and South Atlantic individuals, found at similar latitudes, were of a similar size. On the other hand, phenotypic variation, as measured by the difference in average male hind femur length, was not significantly correlated with genetic differentiation (Mantel R = 0.297, p = 0.074).

Discussion

Neoconocephalus melanorhinus showed significant genetic differentiation among all sampled sites. IBD was the predominant pattern of variation across the range. However, some of our analyses found evidence for at least two biogeographic barriers to gene flow; at the site of the Atlantic-Gulf divide and a second along the Gulf coast, resulting in a hierarchical genetic structure.

Genetic isolation by distance:

We observed a relationship between genetic differentiation ($F_{ST}/(1-F_{ST})$) and geographic distance along both the Atlantic and Gulf coasts (Fig 5). All Mantel tests comparing genetic to geographic distances showed significance, an indication of IBD (Fig 5, Table 2). In most cases neighboring sites were the most closely related, resulting in a genetic cline along the coast. This would indicate that *N. melanorhinus* follows a stepping-stone like model of dispersal (Wright 1943; Kimura and Weiss 1964). *Neoconocephalus melanorhinus'* largely one-dimensional distribution limits movement

to corridors connecting neighboring sites, decreasing dispersal to the least number of paths possible, and reducing gene flow at greater distances.

Ecological differences may also increase with increasing geographic distance, especially with changing latitude along the Atlantic coast. The observed genetic differentiation could be due to differences in selection, resulting in partial reproductive isolation between locally adapted populations. Determining whether reproductive isolation correlates more closely with geographic distance or microhabitat could differentiate these alternative explanations. This, however, could be difficult as distance and ecological differences are themselves correlated (Coyne and Orr 2004).

In the case of *N. melanorhinus*, genetic differentiation increases both along the North/South axis of the range, where variation in ecology is likely to be greatest, and also along their East/West axis (at its southernmost limit), where many ecological factors, especially those correlated with season length or temperature are likely to be more conserved. As such, geographic distance explained genetic differentiation (Fig 5) better than latitude alone. Further analyses investigating specific ecological differences and their correlation with genetic differentiation may be a worthwhile area of future investigation.

Hierarchical genetic structure:

We found evidence for two barriers to gene flow in the phylogenetic pattern (Fig 2), the STRUCTURE analysis (Fig 3), and the Mantel test (Fig 4), while the partial Mantel tests (Table 2) failed to detect these barriers. Below we argue that this failure of

the partial Mantel tests is an artifact of our sampling scheme (see section *Correcting for relative effects of IBD and IBB).*

Increased genetic differentiation in vertebrates, invertebrates, and plant species occurs along the Atlantic and Gulf coasts, specifically at the site of the Atlantic-Gulf divide (reviewed in Avise 2000; Soltis et al. 2006)). The fact that so many diverse taxa (e.g. horseshoe crab (Saunders et al. 1986), dusky seaside sparrow (Avise and Nelson 1989), and diamond-back terrapin (Lamb and Avise 1992)) show similar phylogeographic patterns across this region likely indicates a shared history. While no single geologic or environmental event has been identified, changing environmental conditions and the shifting presence of tropical and subtropical habitats in this region during the Pleistocene likely have played a substantial role in population level diversification in the region (Avise 2000).

The contemporary lack of obligatory salt marsh habitat along most of the Florida peninsula seems a plausible barrier to dispersal for *N. melanorhinus*. Contemporary salt marsh habitats became established in North America in the past 3,000 to 4,000 years (Warren and Niering 1993). Given *N. melanorhinus'* single generation per year, the recent biogeography of the region is the most likely source of their population structure. Thus, we assume that contemporary biogeographic factors were sufficient to produce the patterns of IBB observed in *N. melanorhinus*.

In addition to the Atlantic-Gulf divide, we found evidence of a second barrier to dispersal along the Florida panhandle between localities IB and CR. This stretch of coast possesses few barrier islands and reduced salt marsh habitat. The salt marsh habitat that is present is primarily fragmented patches found upstream in estuaries. It is likely that the

reduced connectivity of salt marsh habitat along the Gulf has resulted in the observed pattern of genetic differentiation.

Correcting for relative effects of IBD and IBB:

The ability to differentiate the affects of IBD and IBB is important for the proper interpretation of population structure. Some of the most commonly used tools (e.g. STRUCTURE) to investigate patterns of hierarchical genetic clustering may falsely detect hierarchical population structure in datasets influenced only by IBD (Cushman and Landguth 2010; Meirmanns et al. 2011; Blair et al. 2012; Meirmanns 2012). Similarly, datasets with only hierarchical genetic structure can show results indicative of IBD (standard Mantel tests, Fig 5) (Meirmanns 2012).

Partial Mantel tests were used to separate the affects of IBD and IBB on *N. melanohrinus'* observed genetic variance. The results of the partial Mantel tests indicated that a pattern of IBD alone explained a significant proportion of the genetic variation. However the partial Mantel tests did not indicate an effect of IBB. This result that contradicted the findings of several of our previous analyses (e.g. phylogeny, AMOVA, structure, and stratified Mantel) that indicated hierarchical genetic structure explained a large proportion of the total genetic variance. We find it likely that the partial Mantel test failed to find the true signal of IBB where the other analyses did, as apposed to a type one error being present in all the other analyses. Multiple facets of this study's sampling scheme may have made the detection of a signal of IBB difficult. Inconsistences between our sampled data and our model of IBB may have led to type two errors in the partial Mantel tests. Within-group variance, like that observed across the Atlantic range,

has a significant impact on *N. melanorhinus'* genetic structure and may have obscured a signal of between-group variance (Cushman and Landguth 2010; Cushman 2006). The partial Mantel test may have therefore failed to find support for IBB because excessive within-group variance broke the assumption of within-subpopulation panmixia, suggestive of a pattern of hierarchical genetic structure (Cushman 2006). In addition, the uneven sampling of sites from the Atlantic and Gulf coasts, likely exaggerated the skewed variance between subpopulations. Because of the likely bias in genetic variance between subpopulations and its affect on the partial Mantel, we conclude that both IBB and IBD still likely play a role in the observed pattern of genetic structure.

Neoconocephalus melanorhinus' strong signal of IBD is likely, at least in part, the result of its strict habitat specificity and the linear arrangement of its coastal habitat. Habitat specialists often show decreased dispersal across complex, heterogeneous landscapes because of the uneven distribution of their resources (Stoner and Joern 2004). Populations surrounded by largely inhospitable environments are likely to incur higher "risk costs" (Bonte et al. 2012), as the likelihood of locating required resources at neighboring sites is low penalizing dispersal. Higher risk costs are likely to increase the selective pressure against dispersal, even in highly mobile species with the physiological capability for long distance dispersal (Papadoulou et al. 2009).

Phenotypic variation:

We observed significant correlation between hind femur length and latitude. This variation in body size could be due to differences in either genetic or environmental effects. Ectotherms, reliant on a limited warm period for development, grow larger at

lower latitudes where the summer seasons are longer ("the converse of Bergmann's rule" (Park 1949; Mousseau 1997)). When raised in common garden conditions, populations of Orthopterans tend to show the same pattern of differential growth as wild caught populations (Masaki 1967; Masaki 1983; Mousseau and Roff 1989; Dingle et al. 1990; Orr 1996). This indicates a genetic mechanism underlying this pattern (Mousseau 1997). However, we observed no significant relationship between body size and overall genetic differentiation between collection sites. The AFLP dataset used in this study represents a genome-wide measure of genetic variation. Specific genes responsible for size determination could show a divergent pattern of variance from the genome as a whole if under selection. A similar common garden experiment with *N. melanorhinus* could clarify the genetic contribution to size determination.

Conclusions:

Neoconocephalus melanorhinus' population structure showed characteristics in common with both the stepping-stone and hierarchical island (barrier) models. While IBD has a larger overall effect, biogeographic barriers have a measureable impact on genetic differentiation in *N. melanorhinus*. As the loss of North American salt marsh habitat continues, anthropogenic barriers to dispersal may become more common and of larger effect. Changes that reduce dispersal among salt marsh sites are likely to decrease genetic diversity within populations and increase the risk of local extinction (51). Habitat specialists, like those that rely on salt marsh habitat, are often the most susceptible to the effects of habitat loss. Species with lower levels of mobility are also at a higher risk than more mobile species like *N. melanorhinus*, which may be able to maintain dispersal

across a more highly fragmented landscape. The significance of biogeographic changes on genetic variation should be of continued concern as our world is further shaped by anthropogenic habitat loss.

References

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst. 1987 Nov;18(1):489-522.

Peterson MA, Denno RF. The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. Am Nat. 1998 Sep;152(3):428-46.

Avise JC. Phylogeography: the history and formation of species. Harvard University Press; 2000.

Greenwood PJ. Mating systems, philopatry and dispersal in birds and mammals. Anim Behav. 1980 Nov 30;28(4):1140-62.

Wright S. Isolation by distance. Genetics. 1943 Mar;28(2):114.

Slatkin M, Voelm L. FST in a hierarchical island model. Genetics. 1991 Mar 1;127(3):627-9.

Kimura M, Weiss GH. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics. 1964 Apr;49(4):561.

Greenfield MD. Evolution of acoustic communication in the genus Neoconocephalus: discontinuous songs, synchrony, and interspecific interactions. The tettigoniidae: Biology, systematics and evolution. 1990:71-97.

Frederick KH. Investigating an adaptive radiation in temperate Neoconocephalus (Orthoptera: Tettigoniidae). Dissertation, University of Missouri; 2013.

Ney, G. and Schul, J. Low genetic differentiation between populations of an endemic prairie katydid despite habitat loss and fragmentation. *Conserv Genetics.* 2017: 10.1007/s10592-017-0987-x.

Rehn JA, Hebard M. Studies in American Tettigoniidae (Orthoptera). Trans. Amer. Entomol Soc. 1915[1914];40:377.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS. Comparative phylogeography of unglaciated eastern North America. Mol Ecol. 2006 Dec 1;15(14):4261-93.

Walker TJ. Singing Insects of North America. Neoconocephalus [Internet]. 2017 Jan. Available from http://entnemdeptufledu/walker/buzz/g185ahtml.

Snyder RL, Frederick-Hudson KH, Schul J. Molecular phylogenetics of the genus Neoconocephalus (Orthoptera, Tettigoniidae) and the evolution of temperate life histories. PloS one. 2009 Sep 25;4(9):e7203.

Holland BR, Clarke AC, Meudt HM. Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. Systematic Biology. 2008 Jun 1;57(3):347-66.

Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors: causes, consequences and solutions. Nat Rev Genet. 2005 Nov 1;6(11):847-6.

Bonin A, Ehrich D, Manel S. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. Mol Ecol. 2007 Sep 1;16(18):3737-58.

Zhang H, Hare MP. Identifying and reducing AFLP genotyping error: an example of tradeoffs when comparing population structure in broadcast spawning versus brooding oysters. Heredity. 2012 Jun 1;108(6):616-25.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology. 2012 May 1;61(3):539-42.

Rambaut A, Drummond AJ. Tracer v 1.5. Program distributed by the authors. 2010.

Zhivotovsky LA. Estimating population structure in diploids with multilocus dominant DNA markers. Mol Ecol. 1999 Jun 1;8(6):907-13.

Vekemans X. Aflp-surv. Software for Estimating Genetic Diversity and Population Genetic Structure from Population Samples (Laboratoire de Gé né tique et Ecologie Végétale, Université Libre de Bruxelles, Belgium), version. 2002;1.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000 Jun 1;155(2):945-59.

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005 Jul 1;14(8):2611-20.

Earl DA. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012 Jun 1;4(2):359-61.

Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007 Jul 15;23(14):1801-6.

Cockerham CC. Variance of gene frequencies. Evolution. 1969 Mar 1:72-84.

Cockerham CC. Analyses of gene frequencies. Genetics. 1973 Aug 1;74(4):679-700.

Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010 May 1;10(3):564-7.

Meirmans PG, Van Tienderen PH. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes. 2004 Dec 1;4(4):792-4.

R Core Team. R: A Language and Environment for Statistical Computing (Version 3.1. 2): R Foundation for Statistical Computing. Vienna, Austria. Available from URL http://www. R-project. org. 2015.

Coyne JA, Orr HA. Speciation. Sunderland, MA; 2004.

Saunders NC, Kessler LG, Avise JC. Genetic variation and geographic differentiation in mitochondrial DNA of the horseshoe crab, *Limulus polyphemus*. Genetics. 1986;112, 613–627.
Avise JC, Nelson WS. Molecular genetic relationships of the extinct dusky seaside sparrow. Science. 1989;243, 646–648.

Lamb T, Avise JC. Molecular and population genetic aspects of mitochondrial DNA variability in the diamondback terrapin, *Malaclemys terrapin*. J. Hered. 1992;83, 262–269.

Warren RS, Niering WA. Vegetation Change on a Northeast Tidal Marsh: Interaction of Sea - Level Rise and Marsh Accretion. Ecology. 1993 Jan 1;74(1):96-103.

Cushman SA, Landguth EL. Spurious correlations and inference in landscape genetics. Mol Ecol. 2010 Sep 1;19(17):3592-602.

Meirmans PG, Goudet J, Gaggiotti OE. Ecology and life history affect different aspects of the population structure of 27 high - alpine plants. Mol Ecol. 2011 Aug 1;20(15):3144-55.

Blair C, Weigel DE, Balazik M, Keeley AT, Walker FM, Landguth E, Cushman SA, Murphy M, Waits L, Balkenhol N. A simulation - based evaluation of methods for inferring linear barriers to gene flow. Mol Ecol Resour. 2012 Sep 1;12(5):822-33.

Meirmans PG. The trouble with isolation by distance. Mol Ecol. 2012 Jun 1;21(12):2839-46.

Cushman SA. Effects of habitat loss and fragmentation on amphibians: a review and prospectus. Biol Conserv. 2006 Mar 31;128(2):231-40.

Stoner KJ, Joern A. Landscape vs local habitat scale influences to insect communities from tallgrass prairie remnants. Ecol Appl. 2004; 14(5):1306-1320

Bonte D, Van Dyck H, Bullock JM, Coulon A, Delgado M, Gibbs M, Lehouck V, Matthysen E, Mustin K, Saastamoinen M, Schtickzelle N. Costs of dispersal. Biol Rev. 2012; 87:290-312.

Papadopoulou A, Anastasiou I, Keskin B, Vogler AP. Comparative phylogeography of tenebrionid beetles in the Aegean archipelago: the effect of dispersal ability and habitat preference. Mol Ecol. 2009; 18:2503-2517.

Ho SY, Phillips MJ, Cooper A, Drummond AJ. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol. 2005 Apr 6;22(7):1561-8.

Park O. Application of the converse Bergmann principle to the carabid beetle, Dicaelus purpuratus. Physiol Zool. 1949 Oct 1;22(4):359-72.

Mousseau TA. Ectotherms follow the converse to Bergmann's rule. Evolution. 1997 Apr 1;51(2):630-2.

Mousseau TA, Roff DA. Adaptation to seasonality in a cricket: patterns of phenotypic and genotypic variation in body size and diapause expression along a cline in season length. Evolution. 1989 Nov 1:1483-96.

Dingle H, Mousseau TA, Scott SM. Altitudinal variation in life cycle syndromes of California populations of the grasshopper, Melanoplus sanguinipes (F.). Oecologia. 1990 Sep 1;84(2):199-206.

Orr MR. Life-history adaptation and reproductive isolation in a grasshopper hybrid zone. Evolution. 1996 Apr 1:704-16.

Masaki S. Geographic variation and climatic adaptation in a field cricket (Orthoptera: Gryllidae). Evolution. 1967 Dec 1:725-41.

Masaki S. Climatic speciation in Japanese ground crickets. GeoJournal. 1983 Nov 1;7(6):483-90.

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I. Inbreeding and extinction in a butterfly metapopulation. Nature. 1998 Apr 2;392(6675):491-4.

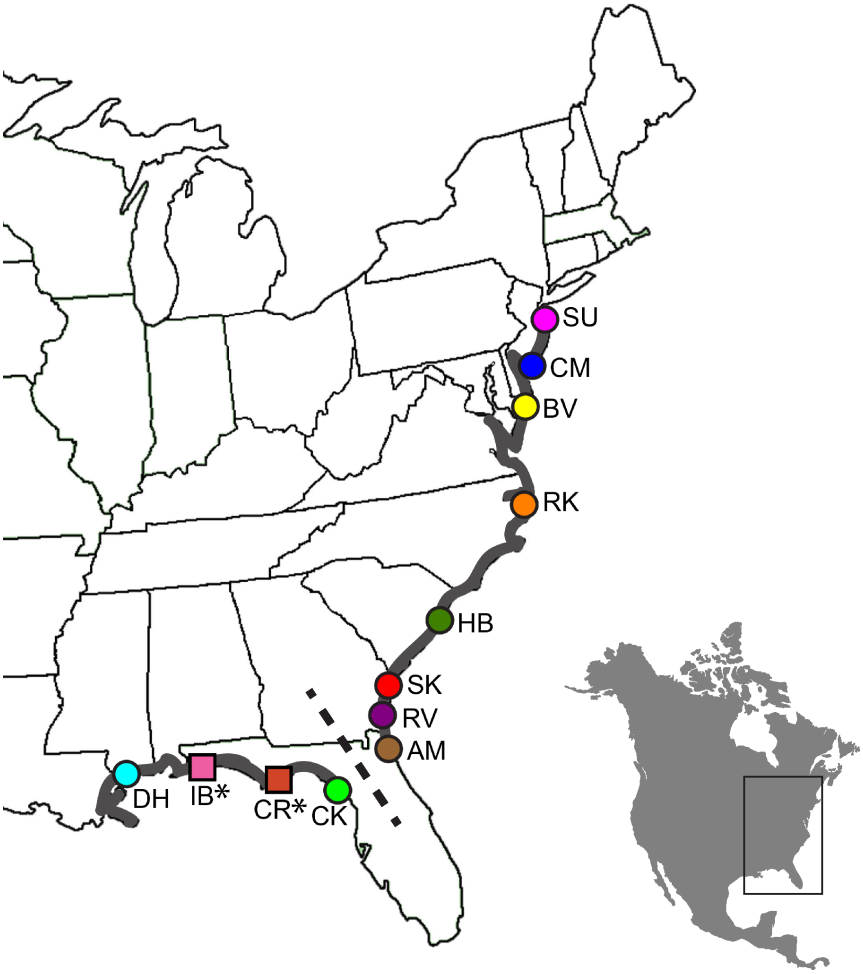


Figure 1: Map of collection localities.

Shaded in area represents the hypothesized range of *N. melanorhinus* (modified from Walker SINA) based on literature and collection records. Squares and * represent sites sampled only in 2015. The dashed line indicates the site of the hypothesized Atlantic-Gulf divide.

Table 1. ***Neoconocephalus melanorhinus*** **sample list.**

Sampling site	Code	Latitude	Longitude	Shoreline	N	PL 5%	Hj
2014							
Diamond Head, MS	DH	30.363	-89.375	0.0	8	1205	0.193
Cedar Key, FL	CK	29.210	-83.064	622.3	18	1119	0.184
Amelia Island, FL	AM	30.630	-81.473	842.3	25	1025	0.172
Ridgeville, GA	RV	31.407	-81.396	929.1	23	1027	0.178
Skidaway Island, GA	SK	31.948	-81.055	997.3	21	1140	0.176
Huntington Beach, SC	HB	33.513	-79.057	1252.6	19	1091	0.170
Roanoke Island, NC	RK	35.838	-75.620	1659.6	27	1023	0.166
Bishopville, MD	BV	38.404	-75.125	1948.2	9	1125	0.174
Cape May Court House, NJ	CM	39.108	-74.891	2029.1	22	1103	0.168
Surf City, NJ	SU	39.685	-74.208	2116.0	16	1061	0.174
2015							
Diamond Head, MS	DH	** Eight samples collected in 2014			9	728	0.218
Indian Bayou, Milton, FL	IB	30.528	-87.101	218.8	13	689	0.237
Carrabelle, FL	CR	29.845	-84.670	456.3	25	705	0.224
Cedar Key, FL	CK	** All samples collected in 2014			10	731	0.203

Table includes samples from 2014 and 2015, sampling site, locality code, latitude, longitude, relative simplified shoreline distance from site DH (km), sample size (N), polymorphic loci at 5% level (PL 5%), and expected heterozygosity under Hardy-Weinberg genotype proportions (Hj). Simplified shoreline distance is measured as the distance along a path connecting neighboring sites by straight lines. (**) Eight DH samples and all ten CK samples from 2014 were added to the 2015 dataset (reflected in sample sizes listed).

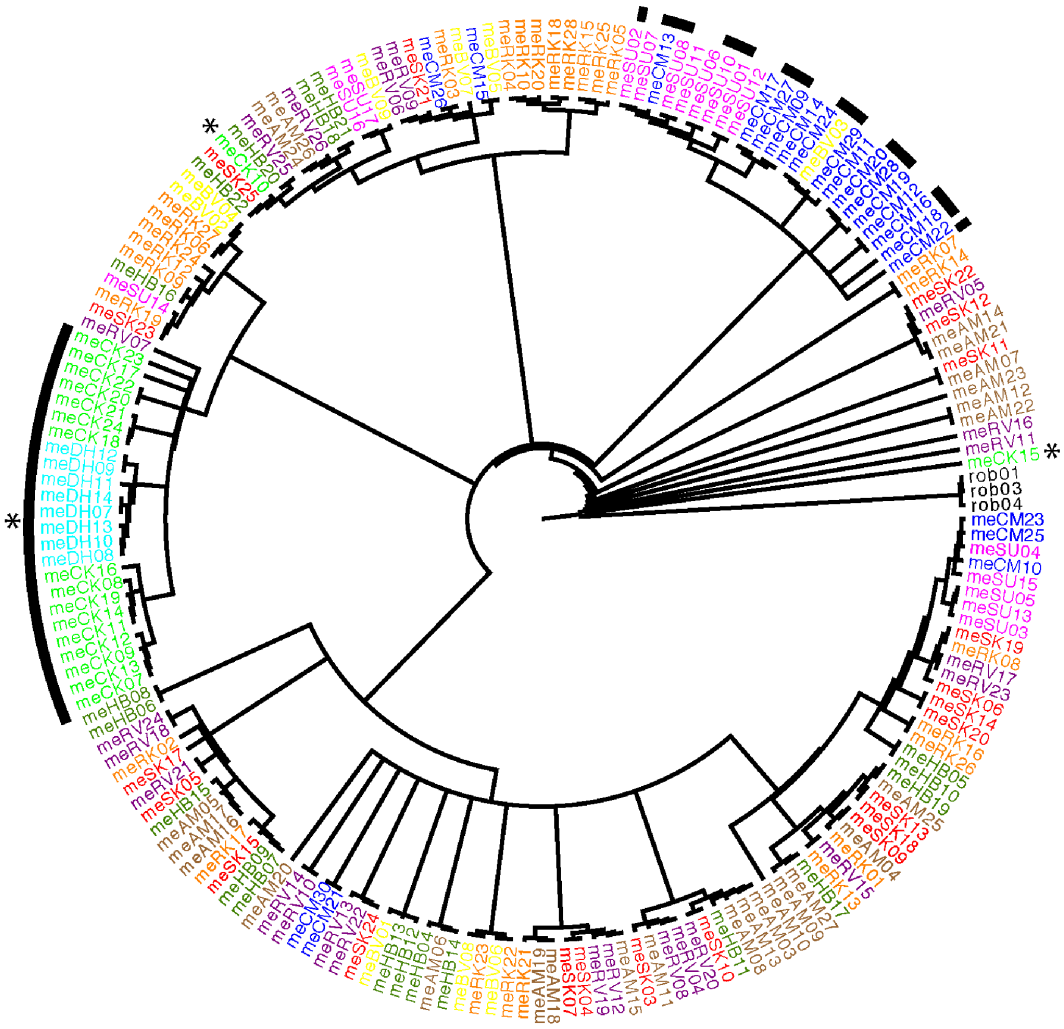


Figure 2: Fifty percent rule Bayesian tree of 2014 individuals.

Tree was constructed using MrBayes v.3.2 (2 runs, 3 chains, and 20 million generations). Three samples of *N. robustus* were used as an out-group. Site localities are indicated by colors corresponding to Fig 1 and site abbreviation in taxa labels. * Indicates Gulf coast individuals; solid line indicates a clade consisting entirely of Gulf coast individuals; dashed line indicates the large North Atlantic clade.

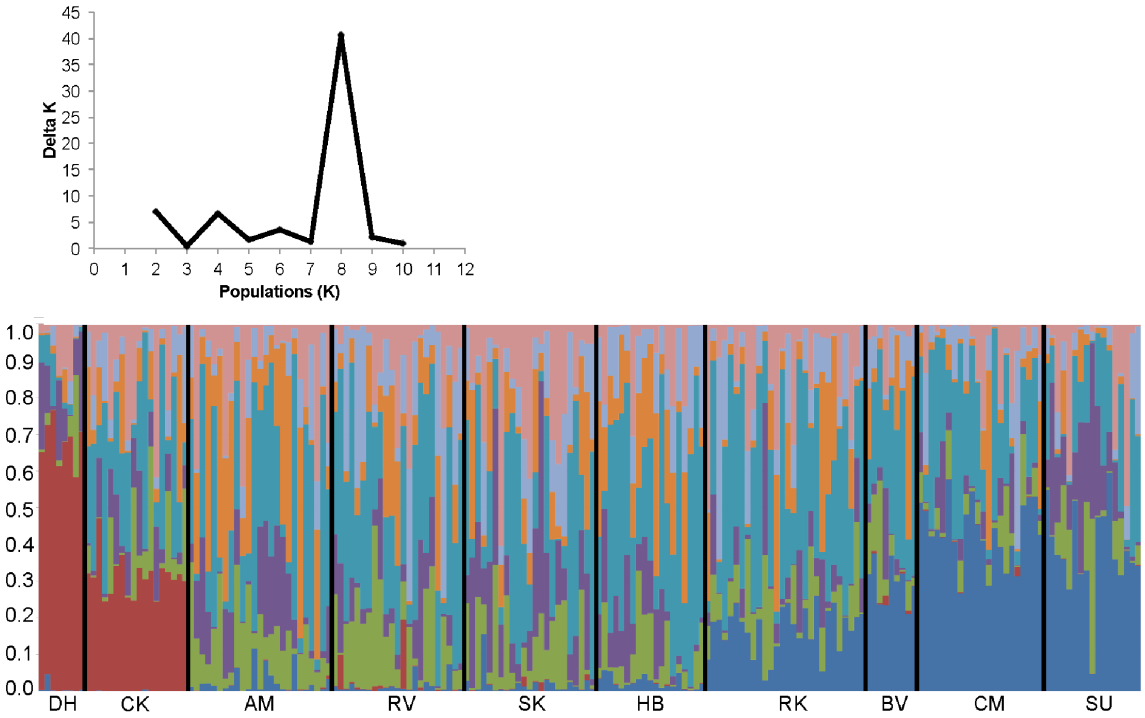


Figure 3: Consensus shared ancestry population structure for the 2014 dataset (K=8).

(Top) Delta K calculated using the Evanno et al. (2005) method. The modal value of this distribution is the accepted K or the uppermost level of structure, here K=8 genetic clusters. (Bottom) Each bar represents an individual grouped by sampling sites. Sites are listed next to their nearest neighbors (West to East; South to North).

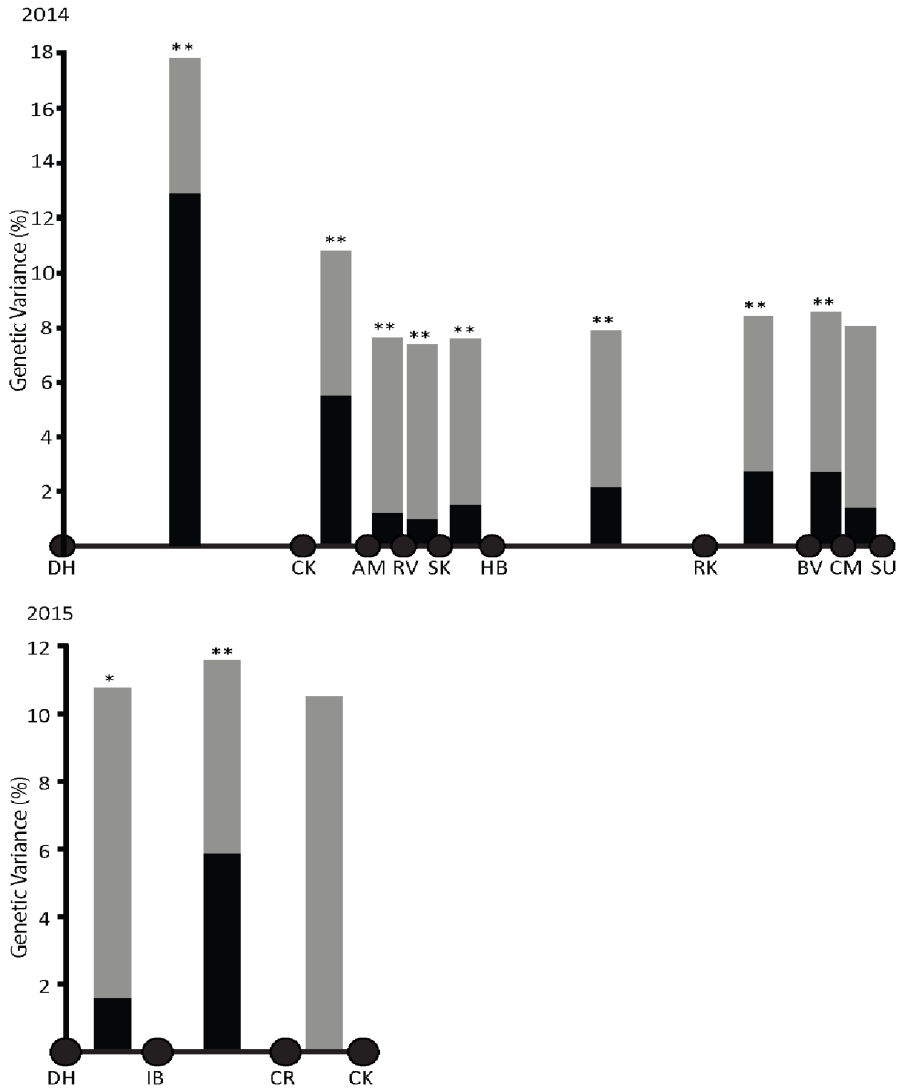


Figure 4: Locus-by-locus analysis of molecular variance (AMOVA) conducted grouping sites into adjacent groups at each site along the coast.

AMOVAs are shown for 2014 and 2015 datasets. Bars indicate the results of an AMOVA conducted by grouping sites on either side of the bar into two separate groups. Black bars represent the genetic variance explained among hypothesized groups split between those two localities and gray bars the genetic variance explained within these groups. Significance of among-group genetic variance sums of squares indicated as * ($p < 0.05$) or ** ($p < 0.00001$). Among-group genetic variance was negative for CR to CK in 2015, shown as zero.

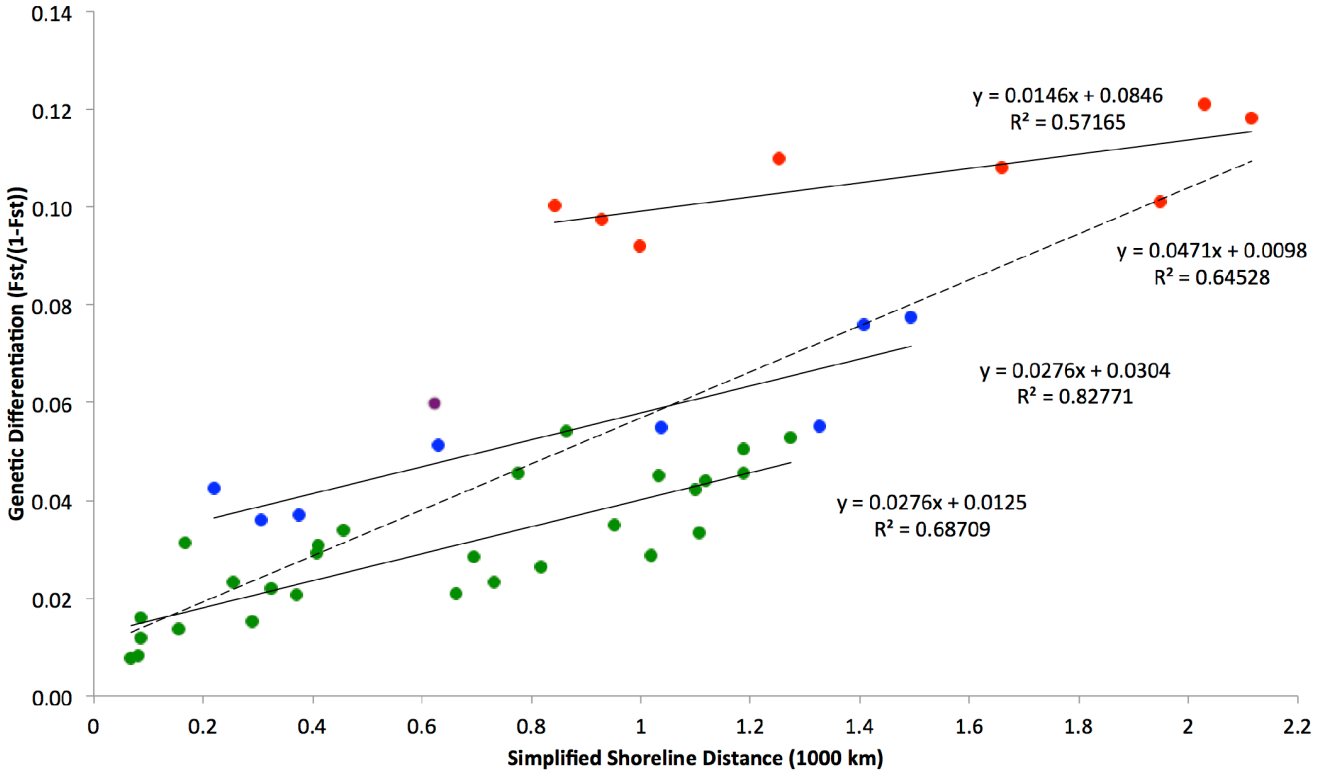


Figure 5: Relationship between genetic differentiation and geographic distance within 2014 dataset.

Between-Atlantic site comparisons (green), between CK and Atlantic sites (blue), and between DH and Atlantic sites (red). Purple indicates the pairwise differences between Gulf sites DH and CK (the westernmost and easternmost Gulf coast sites). The dotted line represents the linear correlation between all points. The equation and R^2 are provided for each line.

Table 2: Results of standard and partial Mantel tests distinguishing the effects of the 2014 dataset's genetic differentiation, geographic distance, and population clustering.

Matrix A	Matrix B	Adjustment	Mantel's R	P value
1) Genetic	Geographic distance	-	0.795	0.001
2) Genetic	Gulf and Atlantic-Gulf divide clusters	-	0.891	0.009
3) Genetic	Gulf divide only clusters	-	0.832	0.074
4) Genetic	Atlantic-Gulf divide only clusters	-	0.750	0.027
5) Genetic	Geographic distance	Partial: Gulf and Atlantic-Gulf divide clusters as covariate	0.812	0.001
6) Genetic	Gulf and Atlantic-Gulf divide clusters	Partial: geographic distance as covariate	0.894	0.748
7) Genetic	Geographic distance	Partial: Gulf divide only clusters as covariate	0.784	0.003
8) Genetic	Gulf divide only clusters	Partial: geographic distance as covariate	0.813	0.751
9) Genetic	Geographic distance	Partial: Atlantic-Gulf divide only clusters as covariate	0.773	0.001
10) Genetic	Atlantic-Gulf divide only clusters	Partial: geographic distance as covariate	0.705	0.753

P values are estimated from 1,000 permutations of samples within matrices.

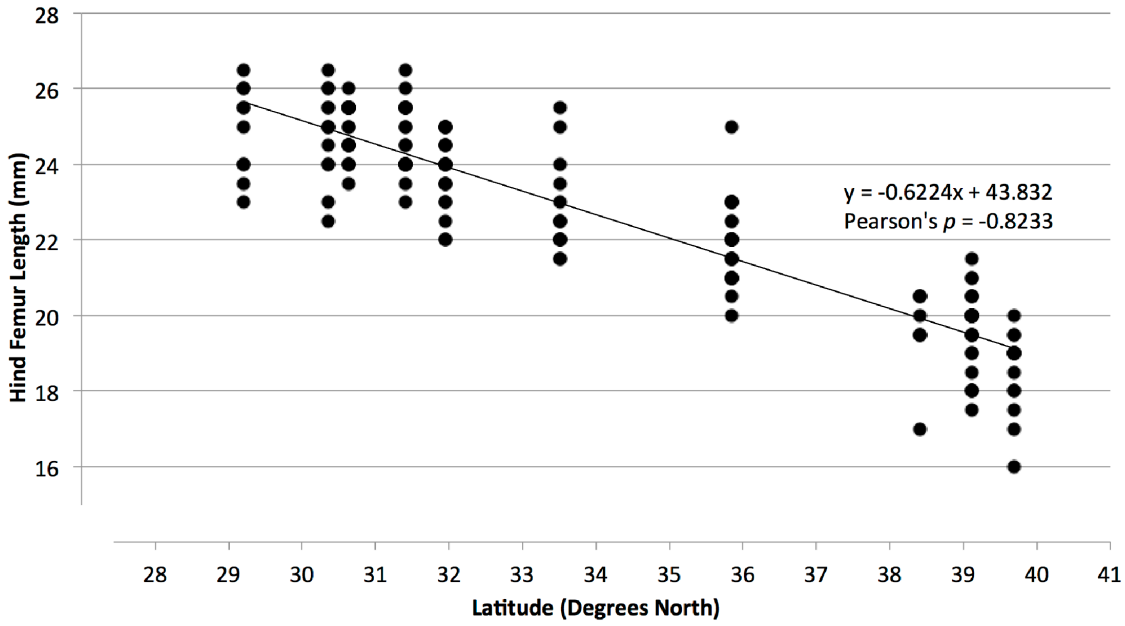


Figure 6: Relationship between hind femur length and latitude among 2014 males.

The equation of the line of best fit and the Pearson's correlation coefficient are provided.

2014	MS	CK	AM	RV	SK	HB	RK	BV	NJ	SC
MS	0.0000	0.0565	0.0912	0.0887	0.0842	0.0991	0.0976	0.0919	0.1080	0.1057
CK	0.0565	0.0000	0.0408	0.0347	0.0356	0.0488	0.0520	0.0523	0.0706	0.0719
AM	0.0912	0.0408	0.0000	0.0117	0.0134	0.0299	0.0257	0.0322	0.0435	0.0502
RV	0.0887	0.0347	0.0117	0.0000	0.0077	0.0214	0.0227	0.0280	0.0406	0.0481
SK	0.0842	0.0356	0.0134	0.0077	0.0000	0.0227	0.0206	0.0337	0.0431	0.0421
HB	0.0991	0.0488	0.0299	0.0214	0.0227	0.0000	0.0283	0.0276	0.0435	0.0513
RK	0.0976	0.0520	0.0257	0.0227	0.0206	0.0283	0.0000	0.0150	0.0202	0.0328
BV	0.0919	0.0523	0.0322	0.0280	0.0337	0.0276	0.0150	0.0000	0.0081	0.0304
NJ	0.1080	0.0706	0.0435	0.0406	0.0431	0.0435	0.0202	0.0081	0.0000	0.0157
SC	0.1057	0.0719	0.0502	0.0481	0.0421	0.0513	0.0328	0.0304	0.0157	0.0000

2015	MS	IB	CR	CK
MS	0.0000	0.0335	0.0547	0.0618
IB	0.0335	0.0000	0.0451	0.0767
CR	0.0547	0.0451	0.0000	0.0150
CK	0.0618	0.0767	0.0150	0.0000

Supplemental Table 1: 2014 and 2015 Pairwise F_{ST} values based on AFLP markers as calculated in AFLPsurv. All pairwise F_{ST} values were significant ($P < 0.05$). Color gradient shows range of pairwise F_{ST} values in dataset, green the least divergent and red the most.

Chapter 2

LOW GENETIC DIFFERENTIATION BETWEEN POPULATIONS OF AN ENDEMIC PRAIRIE KATYDID DESPITE HABITAT LOSS AND FRAGMENTATION

Published in Conservation Genetics with Johannes Schul as coauthor

Abstract:

Tallgrass prairie habitats within North America have suffered severe fragmentation and habitat loss as land has been converted for agricultural purposes. Habitat loss and fragmentation can affect gene flow and the genetic structure of insect populations. *Neoconocephalus bivocatus* is a prairie obligate katydid found only in isolated prairie patches. We compared genetic diversity and population differentiation using AFLP markers in *N. bivocatus* and *N. robustus,* a grassland generalist that is not isolated to prairie fragments and occupies a more contiguous range. Similar levels of genetic diversity were present within populations of both species. While population genetic structure was found in both species, there was no relationship between assigned genotypes and sampling localities. This genetic structure may instead be evidence of a past barrier to gene flow that has since been removed. Genetic differentiation within both species was low, with no evidence of a correlation with geographic distance, indicating neither species is dispersal limited at these distances. We see no significant reduction in genetic diversity or genetic differentiation within *N. bivocatus* when compared to *N. robustus*. We therefore conclude that while *N. bivocatus* utilizes a fragmented landscape, long-distance dispersal likely maintains gene flow between isolated prairie patches.

Introduction

An estimated 87% of North American tallgrass prairies have been lost since 1770, more than any other North American ecosystem (Rickletts et al. 1999; Samson et al. 2004). This habitat loss is largely due to anthropogenic modifications for agricultural purposes (Samson et al. 2004). The remaining tallgrass prairie habitat is distributed patchily across a complex heterogeneous landscape. Restoration efforts have been successful in many areas, but restored and remnant patches are often isolated from one another (Dobson et al. 1997). Species that have evolved in historically contiguous habitats may not be capable of dispersal across a landscape of increasing heterogeneity, raising concerns about their ability to persist within such a fragmented ecosystem (Stoner and Joern 2004; Gauffre et al. 2015).

Reduced dispersal rates increase local extinction risks through a decreased capacity for the colonization of isolated habitat fragments (Brown and Kodric-Brown 1977; Lande 1993; Saccheri et al. 1998). For this reason maintaining an ability to move between patches is important for the persistence of species in fragmented landscapes (Stevens et al. 2006). Measurements of gene flow and genetic structure, using molecular tools, allow for the quantification of species dispersal over time and the resulting effects on genetic diversity (Roach et al. 2001).

A broad range of insect taxa play key roles as consumers and pollinators in prairie ecosystems and for that reason have been commonly used to monitor fragment restoration success (Burger et al. 2003). Studies investigating the effects of fragmentation on prairie insects have largely focused on species-richness or abundance (e.g. Tscharntke 1995;

Tscharntke et al. 2002; Collinge and Palmer 2002). Less understood are the effects of fragmentation on the dispersal of insect populations between isolated prairie patches.

Habitat fragmentation may affect insects differently depending on their feeding guild and their specificity for particular food sources (Golden and Crist 1999; Kruess and Tscharntke 2000). Specialists and generalists may show differences in their dispersal ability across a complex heterogeneous landscape because of the uneven distribution of the specialists' resources across the environment (Stoner and Joern 2004). Because of this, the fragmentation of prairie ecosystems may restrict dispersal between patches more severely in specialist than in generalist species.

Neoconocephalus is a diverse group of North American katydids with variation in habitat preferences, as well as call phenotypes (Greenfield 1990). *Neoconocephalus bivocatus'* habitat preference is restricted to remnant and revitalized prairie patches, preferring dry habitats composed primarily of tall prairie grasses (Walker et al. 1973). *Neoconocephalus robustus* is the sister taxon of *N. bivocatus* (Snyder et al. 2009). The two species are morphologically cryptic, but possess divergent male call traits and corresponding female call preferences (Deily and Schul 2004, 2006). Unlike *N. bivocatus*, *N. robustus* is a grassland generalist, found in habitats with a wide range of flora and environmental conditions including roadside ditches, agricultural, and urbanized areas (Walker et al. 1973; Frederick 2006). The two species maintain divergent preferences in habitat across a broadly overlapping range occupying much of the Eastern United States (Whitesell 1969; Walker 2000).

The goal of this study is to assess how habitat fragmentation affects the genetic diversity and population structure of an insect species specializing in prairie habitats. We

compare population structure and genetic diversity among populations of a prairie specialist (*N. bivocatus*) and a generalist (*N. robustus*). Increased genetic structure among populations of *N. bivocatus,* beyond that observed in *N. robustus,* would indicate reduced dispersal associated with the loss and fragmentation of prairie habitats. Males of both species were collected from prairie sites across Missouri and neighboring states. We used amplified fragment length polymorphism (AFLP) markers to measure genetic diversity, population structure, and gene flow among thirteen prairie sites.

Methods

Specimen Collection:

We collected 377 male *N. bivocatus* and *N. robustus* from thirteen grassland sites within Missouri and several surrounding states during the Summers of 2013 and 2015 (Supplementary Table 1). Distances between sites ranged from 4 to 773 km. Our sample sizes were sufficient for comparisons between sites that were between 4 and 448 km apart; these pairwise distances are given in Table 1. We used male calls to localize individuals in the field and collected them by hand after sunset. We identified males in the field as belonging to members of the target group through their call and morphological features, including cone pigmentation and body size, prior to collection (Walker et al. 1973). As the two species are morphologically cryptic, we identified individuals as either *N. bivocatus* or *N. robustus* based on their divergent call characteristics.

Call recordings:

We recorded male calls in 2013 within three days of collection using an Audiotechnica ATR 55 microphone and a Marantz PMD-671 solid-state recorder (16 bit, 48 kHz sampling rate). Recordings were made outdoors at ambient temperatures (25-28° C) with males placed in individual mesh cages (approximately 10x20x10 cm) spaced at least 3 m apart. In 2015, we recorded male calls in the field immediately preceding collection using a Tascam DR-40 linear PCM recorder (16 bit, 48 kHz sampling rate). Ambient temperatures were in the range of 22-27° C.

Temporal call analysis:

We marked each sound pulse (produced during the closing movement of the wings; Walker 1975) using custom software. In short, the recordings were rectified and the envelope extracted with a temporal resolution of 0.125 ms. Pulses were detected automatically using a threshold based algorithm and manually checked before saving their data in text-files. We analyzed about 2s of each male's call containing 150-250 pulses. Further analyses were conducted on the text files in MS Excel.

Male calls of *N. robustus* have a single pulse rate of about 200/s equivalent to a pulse period of about 5 ms (Walker 1975). Females recognize this pattern by the absence of silent gaps longer than about 2 ms (Deily and Schul 2004). In contrast, male calls of *N. bivocatus* have two alternating pulse periods of about 4.5 ms and 7 ms (Deily and Schul 2004), resulting in a 'galloping rhythm' or 'pulse pairs.' Female *N. bivocatus* recognize this pattern by the rate of the pulse pairs: pulse pair rates around 87/s were attractive and attractiveness decreased towards faster and slower rates. For the pulse pairs to be

detectable, the two alternating pulse periods must differ sufficiently (J. Schul, unpublished).

We quantified the ratio of the means of the alternating pulse periods (longer pp / shorter pp). For the single pulse calls of *N. robustus*, this should result in values close to one, while in *N. bivocatus* significantly larger values result (Bush and Schul 2010). We found among the calls of 354 males analyzed, one group between 1 and 1.17 and a second group with ratios > 1.3. A natural break occurred in the data between 1.17 and 1.3. We classified ratios < 1.17 as '*N. robustus*', and > 1.3 as '*N. bivocatus'* in temporal pattern.

Spectral analysis:

We analyzed the amplitude spectrum of male calls using a fast Fourier transformation (FFT, Hamming window, frame length 256) in Audacity v.1.3 (Audacity 2008). Spectra were averaged over two seconds of the call. The main energy in the spectrum of *Neoconocephalus* calls is concentrated in a narrow band between 7 and 15 kHz, with the peak frequency differing among species (Schul and Patterson 2003). We measured the center frequency of this low frequency band as the geometric mean of the upper and lower cut-off frequencies at -3dB from the peak amplitude.

In *N. bivocatus*, the center frequencies of this band have a mean of about 10 kHz among individuals, ranging from 7 to 15 kHz. Females have little spectral selectivity in this frequency range (Deily and Schul 2006). The center frequencies of calls with a pulse period ratio of >1.3 all fell into this range and were accordingly classified as *N. bivocatus* calls.

In *N. robustus*, the low frequency band is narrower than in *N. bivocatus* and are typically limited to 10 kHz and below (Schul and Patterson 2003). Indeed, frequencies above 10 kHz have an inhibitory effect on female phonotaxis (Deily and Schul 2006). Of the individuals with pulse period ratios < 1.17, center frequencies of all but four individuals clustered between 6 and 9 kHz. These individuals were classified as *N. robustus*. The four remaining individuals had center frequencies well above 10 kHz, which would be significantly less attractive for *N. robustus* females. Since their temporal patter fell into the *N. robustus* range, we classified these calls as 'intermediate phenotypes.'

Molecular analysis:

We removed males' hind femurs and placed them in 95% EtOH for DNA preservation. We later extracted DNA from the hind femurs using the DNeasy Blood + Tissue Kit (Qiagen Inc., Valencia, CA, USA). DNA quantification was performed on each sample by spectrophotometry (NanoDrop 1000, Thermo Scientific, Wilmington, DE). W e used AFLP markers to measure genetic variation among species, populations, and collection sites.

AFLP loci were generated using the method described in Snyder et al. 2009. Pre-selective PCR was carried out using primers Eco+A and Mse+A. We performed selective PCR with six primer pairs. Fluorescently labeled Eco primers (Eco+AAC (6FAM), Eco+AGC (PET)) were used in selective PCR (as described in Snyder et al. 2009) and fragments were separated in an ABI 3730 genetic analyzer at the DNA Core Facility, University of Missouri. We called AFLP bands using GeneMarker v.1.6 (Softgenetic

Corp, State College, PA, USA) using an automated peak-calling scheme (as described in Holland et al. 2008); alleles were called between 50-500 bp with a minimum peak intensity of 50. All non-informative, non-variable, bands were removed prior to further analyses. Automated peak-calling schemes produce larger numbers of total fragments but have a higher mismatch error rates than manual peak calling. Larger AFLP datasets reveal, despite higher error rates, greater population differentiation than smaller, completely error free data sets (Zhang and Hare 2012). The same AFLP amplification and peak-calling scheme has been successfully used in a closely related species to reconstruct significant patterns of interspecific genetic differentiation and isolation by distance (Ney and Schul 2017).

We tested the rate of genotyping errors between AFLP runs to determine the viability of combining AFLP data sets from both years. To test for peak calling repeatability, eight 2013 samples were rerun in duplicate with the 2015 dataset. Fragment lengths were recalled for the original as well as both sets of rerun samples as described previously and the mismatch error rate was calculated within and between AFLP runs (Bonin et al. 2007).

2013 samples produced 1235 AFLP loci and 2015 samples 2461 AFLP loci from six selective primer pairs. When the 2013 and 2015 data sets were analyzed separately, mismatch error rates of peak calling averaged 6.1%. When the data of the two years were combined, mismatch error rate increased to 9.7%. To avoid this increase in error rate, we chose to analyze the data of the two years separately.

We estimated the phylogenetic relationship between individuals using the neighbor-joining algorithm as implemented in SplitsTree v.4.13 (Huson and Bryant 2006)

and calculated posterior probabilities using a bootstrapping method (1000 permutations) for each dataset independently. The population structure analysis was implemented for all individuals using the program Structure v.2.3.3 (Pritchard et al. 2010). The admixture model was used, allele frequencies correlated, with a run length of 100,000 (Burnin=10,000) for 10 replicates each of K=1-15 (clusters). The estimate of the most well supported K was calculated as described in Evanno et al. (2005) and implemented in Structure Harvester v.0.6.94 (Earl and vonHoldt 2012). The program Clumpp v.1.1.2 (Jakobsson and Rosenberg 2007) was used to align the 10 repetitions of the most well supported number of clusters. Separate structure analyses, following the same settings as before, were run for individuals belonging to 2015 *N. bivocatus* and *N. robustus* independently.

To double-check the grouping of individuals in the AFLP neighbor-joining network and population structure analyses, we sequenced a mitochondrial gene (CO1) for the 2013 individuals. Briefly, we amplified 841 bp of the gene COI using the primers Ron and Calvin1 as described in Snyder et al. 2009. We performed sequencing at the DNA Core Facility, University of Missouri, on an ABI 3730 genetic analyzer, then edited and aligned using Geneious v6.0.5 (Kearse et al. 2012). Finally we constructed a median-joining haplotype network (Bandelt et al. 1999) in Popart (http://popart.otago.ac.nz).

Genetic differentiation was calculated using the AFLP loci and a Bayesian non-uniform allele frequency distribution method (Zhivotosky 1999) as implemented in the program AFLPsurv v.1.0 assuming Hardy-Weinberg genotypic proportions (Vekemans 2002). Genetic diversity was estimated following the calculation of allele frequencies as outlined in Lynch and Milligan (1994). Expected F_{ST} was estimated from 100,000

permutations (10 bootstrapped runs) of individuals among assigned groups and compared to observed F_{ST} values in order to determine significance (Lynch and Milliagan 1994). Males were first assigned to groups based on phenotypic species assignments. Individuals possessing intermediate call phenotypes were not included in species-specific analyses. We then in separate analyses assigned individuals of each species to groups based on their collection site. Collection sites possessing less than five males of either species were removed from the analysis. Mantel tests for the correlation between genetic differentiation (pairwise F_{ST} values) and geographic distance were conducted in GenAlEx v.6.5 (Peakall and Smouse 2006).

Within-run error rate controls:

To test for the importance of the within run error rate on the neighbor-joining network and the Structure analyses, we generated two data sets, each consisting of the 2015 data and one of the replicate sets of eight individuals collected in 2013 (see above). We repeated the neighbor-joining analyses with these two data sets and compared the cluster assignment of the eight individuals (Supplementary Fig 1). The cluster assignment of none the eight 2013 individuals nor any of the 2015 individuals differed between the two analyses. Nodal support remained >0.99 between the species designating clusters in both analyses. As support beyond species designating clusters was low, we only use these clustering analyses for species assignment and not as a measure of interspecific variation.

In addition we ran these two data sets independently through Structure, the Bayesian clustering analyses. The replicate datasets in both cases resulted in K=2 being the most well supported number of clusters. The estimated membership coefficient (Q)

differed by an average of only 2.56% (0.00-14.22%) between the eight replicated samples (Supplementary Table 2).

Results

Network Analysis:

We included a total of 96 individuals collected in 2013 in this analysis. Of those, 73 had call recordings of sufficient quality to allow spectral and temporal call analysis. Twenty-six had the *N. robustus* call type, forty-four the *N. bivocatus* call type, and three an intermediate call type (Fig 1). We collected a total of 281 individuals during the summer of 2015. Of these 195 individuals had the *N. robustus* call type, 85 the *N. bivocatus* call type, and one an intermediate call type (Fig 1).

The neighbor-joining network of 2013 individuals was composed of two clusters with high bootstrap support (0.95). One consisted of individuals primarily possessing the *N. robustus* call type, the other included primarily individuals with the *N. bivocatus* call type (Fig 2b). Four *N. bivocatus* appeared in the primarily *N. robustus* cluster and two *N. robustus* appeared in the primarily *N. bivocatus* cluster. In order to check the reliability of these mismatched species assignments we investigated the mitochondrial clustering assignments of these mismatched individuals. We found that they also possessed heterospecific haplotype assignments and that there were no inconsistencies in genetic assignments based on these two genetic markers (Supplementary Fig 2). These genotype/phenotype-mismatched individuals came from several localities, with mismatches of both possible types present at sites T and D.

The 2015 neighbor-joining network grouped, as in 2013, into two clusters with high bootstrap support (0.99; Fig 2c). One cluster was made up exclusively of *N. bivocatus* individuals, while the second cluster was comprised of *N. robustus* individuals and the single intermediate caller (Fig 2c). Nodal support within species designating clusters was low among all branches, < 0.04, indicating low genetic differentiation beyond the major species clusters.

Genetic Structure Analysis:

The population structure analysis (Fig 3) supported the genetic structure observed in the neighbor-joining network. Within 2013 individuals K=2 populations had the highest change in log likelihood (Fig 3a). All individuals, but one possessed a Q (proportion of ancestry from population) for one of the two clusters >0.90 (Fig 3e). Additional analyses of population structure run within each species showed no evidence of additional intraspecific genetic structure (not shown).

Within the 2015 dataset the structure analysis supported the presence of three clusters (K=3, Fig 3b). These species differentiating clusters aligned with the species assignments produced in the neighbor-joining network analysis (see above, Fig 2c). The intermediate phenotype individual again assigned predominantly to one of the two *N. robustus* populations (Q=0.99). Analyzing the genetic structure of the 2015 *N. bivocatus* individuals (Fig 3a) revealed two sub populations (K=2) not evident in the combined 2015 analysis. The analysis of the 2015 *N. robustus* individuals did not reveal any additional population structure (K=2, Fig 2d, 2h).

Population assignment scores within the 2015 *N. bivocatus* and *N. robustus* structure analyses showed major variation between individuals collected at the same site (Fig 3g, Fig 3h). Within *N. bivocatus*, the locality M possessed a higher proportion of membership to population two than any other *N. bivocatus* locality. Within *N. robustus*, localities K, O, and A showed a majority assignment to population two, while all other localities had a majority assignment to cluster one (Fig 4). While there was evidence of genetic differentiation between localities, this variation did not appear to have a strong relation to the geographic arrangement of the localities (Fig 4).

Genetic differentiation:

In 2013, *N. bivocatus* localities showed mean within-population heterozygosity, Hj (equivalent to mean Nei's gene diversity; Nei 1987) of 0.180 (0.167-0.191 between sites), while *N. robustus* localities showed a mean expected heterozygosity of 0.190 (0.150-0.210 between sites). In 2015, *N. bivocatus* localities had a mean within-population heterozygosity estimate of 0.213 (0.210-0.215 between sites), while *N. robustus* localities had a mean value of 0.184 (0.172-0.200 between sites).

The analysis of genetic differentiation between 2013 species resulted in an observed F_{ST} value of 0.0864. The estimation of observed F_{ST} for 2013 *N. bivocatus* and *N. robustus* individuals grouped by collection site, both resulted in slightly negative values (-0.0066 and -0.0041 respectively), which we interpreted as not differing from zero. Few pairwise F_{ST} values between sites within either species differed from zero (Table 1). Those 2013 sites that showed genetic differentiation from one another tended

to be those at the geographic extremes of the sampling range. Localities T in *N. bivocatus* and D in *N. robustus* were responsible for all the non-zero pairwise F_{ST} values.

The analysis of genetic differentiation between 2015 species resulted in an F_{ST} value of 0.1079. In 2015, *N. bivocatus* and *N. robustus* analyses resulted in estimations of observed F_{ST} values of 0.0063 and 0.0027 respectively. A higher proportion of 2015 pairwise F_{ST} values between sites were significantly different than zero ($p < 0.0001$) when compared to 2013 samples, within both *N. bivocatus* and *N. robustus*. Within 2015 *N. bivocatus* all but one locality pair had a paired F_{ST} diverging from zero, while in *N. robustus* only half the locality pairs showed any genetic differentiation. Genetic differentiation between sites within both species, while significant, remained low. In 2015, no relationship between genetic differentiation and geographic distance was evident within either species (Fig 5).

Discussion

Neoconocephalus bivocatus and *N. robustus* cluster strongly based on call characters and show significant levels of species specific differentiation. Beyond the species level, both showed evidence of additional within species hierarchical genetic structure. However, site-specific genetic diversity differed little between species. While slightly higher in *N. bivocatus*, genetic differentiation was low in both species and not correlated with geographic distance.

Independent genetic contrast:

The 2013 dataset showed evidence of potential genetic introgression in the form of several intermediate calling and mismatched genotype/phenotype individuals (Frederick 2013). However, 2013 followed the most severe year of drought in the Midwest since 1895 (Hoerling et al. 2014). Changes in population size and environmental disturbances, such as those conditions following a severe drought, can increase the rates of hybridization between closely related species (Lamont et al. 2003; Seehausen et al. 2008), especially those that have retained divergent ecological niches (Seehausen 2006). Sampling two years later did not show the same evidence of genetic introgression and may point to the possible short-lived effects of such events on long-term genetic structure (Gilman and Behm 2011). While the presence of intermediate phenotype and mismatch genotype/phenotype individuals may indicate past genetic introgression, it should be noted that we did not find evidence of mixed genetic ancestry within these individual's AFLP profiles. AFLP markers, while a useful measure of genome wide diversity, only sample a small fraction of the total genome, meaning that older, more dilute, genetic introgression could be missed. We conclude that *N. bivocatus* and *N. robustus*, despite potential genetic introgression, represent independent genetic groups and that an independent contrast of the two is valid.

Genetic Consequences of habitat fragmentation:

Population genetic theory suggests that fragmentation will lead to a decrease in gene flow between isolated sites, which will result in reduced genetic diversity and increasing genetic differentiation through genetic drift within isolated populations (Young et al. 1996). A number of other studies investigating effects of habitat

fragmentation on insect taxa have found evidence of an effect on population level genetic diversity (Berwaerts et al. 1998; Williams et al. 2003; Dhuyvetter et al. 2005; Keyghobadi et al. 2005). Despite the contemporary habitat loss and fragmentation of prairie habitats, *N. bivocatus* retained approximately equal genetic diversity at all our sampled sites, differing little from the generalist, *N. robustus*. As a whole, genetic diversity for both species was relatively low, but within the range observed in other insect studies utilizing this method (Ming and Wang 2006; Chaput-Brady et al. 2008; Ming and Alymes et al. 2009; Paris et al. 2010; Taylor et al. 2011; Paris and Despres 2012).

Genetic differentiation, like the levels of genetic diversity measured, did not support a substantial effect of fragmentation on dispersal within this system. *Neoconocephalus bivocatus*, the prairie specialist, possessed slightly higher levels of genetic differentiation than the generalist, *N. robustus*. However, both species exhibited such low genetic differentiation as to infer substantial contemporary gene flow. The observed genetic differentiation was significantly lower than that observed in nearly any other grassland insect (Schmitt and Seitz 2002, Williams et al. 2003, Krauss et al. 2004, Stireman et al. 2005, Weisner et al. 2014, Saarinen et al. 2016). The flightless grasshopper *Pezotettix giornaeI,* for example, showed greater genetic differentiation (F_{ST} = 0.045) between grassland sites across a fragmented agricultural landscape at much shorter distances (20 km), likely owing to their decreased dispersal ability (Gauffre et al. 2015). On the other hand, the moth *Heliothis virescens* possessed genetic differentiation on the same level observed in this study (F_{ST} = 0.002) at a similar geographic scale (Korman et al. 1993). The low genetic differentiation found in this species is consistent with its high level of dispersal. Individuals of *H. virescens* have been observed to

disperse on average 10 km, with some individuals dispersing nearly 50 km, in their adult life (Schneider 1999). As *H. virescens* and *Neoconocephalus* possess similar levels of genetic differentiation the two groups may show similar levels of dispersal.

Neoconocephalus' maintenance of genetic homogeneity across a heterogeneous environment was unique among insects utilizing a variety of fragmented habitat types (Monaghan et al. 2001, Garnier et al. 2004, Vandergast et al. 2007, Phillipsen and Lytle 2013, Watts et al. 2016). For example habitat fragmentation had a significantly greater impact on genetic differentiation in the grasshopper *Mioscirtus wagneri,* which is limited to highly fragmented hypersaline habitats in Spain. *M. wagneri* showed genetic differentiation approximately one-hundred times greater (Fst ≈ 0.2) than *N. robus*tus and *N. bivocatus* at similar geographic distances (average 250 km between sites; Ortego et al. 2010). The beetle *Bolitophagus reticulatus* showed genetic differentiation between forest fragments at least ten times higher (F_{ST} = 0.069) than our estimates, across a much smaller geographic scale (maximum distance between fragments = 4.4 km; Knusten et al. 2000). Alternatively, the aquatic insect *Boreonectes aequinoctialis*, known as a strong flyer and found in isolated aquatic habitats, showed levels of genetic differentiation (F_{ST} = 0.006) similar to that observed in this study and lower than that observed in two less mobile aquatic insects from the same region (Phillipsen et al. 2015).

The results of this investigation suggest that both our species possessed sufficient gene flow between grassland sites as to prevent genetic divergence over time. Recent fragmentation might lead to a similar pattern and is thus an alternative explanation for our findings. However it is unlikely that sufficient time has not passed since fragmentation to detect genetic differentiation. Twenty generations were sufficient in a modeling study to

result in genetic differentiation detectable with tools, similar to those used here (Blair et al. 2012). *Neoconocephalus* katydids have one generation/year and habitat loss leading to prairie fragmentation has been occurring for well over 100 years (Samson et al. 2004).

Our study also found no relationship between genetic and geographic distance. Isolation by distance (IBD) is a common pattern observed within taxa possessing limited dispersal ability (Wright 1943). We did not detect a significant pattern of IBD in either species, suggesting that neither shows evidence of distance-limited dispersal. Both *N. bivocatus* and *N. robustus* occupy large ranges, from which this study only sampled a portion. It is possible that an investigation at a larger geographic scale could find a pattern of IBD not present at the distances measured here.

In agreement with the high levels of between population gene flow detected, no meaningful clustering of populations was detected in either species. Genotypes assigning to divergent genetic clusters were shared among all sites and are therefore unlikely to be due to the effects of contemporary gene flow. This cryptic genetic structure alternatively could be due to one or more events of significant genetic isolation in the past, followed by secondary contact (Austin et al. 2002). This would have occurred prior to dispersal into the study area, resulting in the prevalence of both genotypes across all sites (Avise 2000). Significant levels of contemporary gene flow could act to maintain both genotypes among sites. The 2013 dataset may have detected no evidence of intraspecific genetic structure because of reduced sample sizes and power (Evanno et al. 2005). The combination of high genetic diversity, low genetic differentiation, and no evidence of population structure or IBD all point to substantial gene flow and significant contemporary dispersal across the study area.

Our results suggest that *N. bivocatus* dispersal is not hindered by habitat loss or fragmentation at the distances measured in this study (< 450 km). High levels of gene flow between sites would suggest *N. bivocatus* is able to move from prairie fragment to prairie fragment with little difficulty, aided by its high dispersal ability. Little direct evidence is available as to the dispersal abilities of *Neoconocephalus* species, however their large body size, large relative wing size, and large flight muscles would indicate a propensity for long distance flight (Greenfield 1990). The related genus *Ruspolia* possesses a similar wing and body morphology and has been observed flying 1000 km out to sea and as much as 43 km per night (Bailey and McCrae 1978; Greenfield 1990). *Neoconocephalus bivocatus* established residence within a newly restored prairie patch uninhabited for at least five years (Schul personal obs.). *Neoconocephalus robustus* has a disjunct range located in California that likely required extreme long distance dispersal in order to found, if unaided by human introduction (Greenfield 1990).

Conservation message:

In terms of prairie restoration efforts these findings indicate that highly mobile prairie obligate insects, like *N. bivocatus*, have the ability to find and maintain gene flow between isolated prairie patches. Prairie restoration sites do not need to be located within a particular distance of one another for taxa that can retain significant dispersal across a fragmented landscape. Widely dispersing insects, such as the katydids investigated here, will likely pave the way for higher-level prairie consumers and the assemblage of a more complex prairie ecosystem, despite the isolation of prairie sites. The conservation of these prairie katydids and other non-dispersal limited prairie specialists will require less active

management strategies. With this information in mind land managers can make more informed decisions on how to best use resources to better benefit more dispersal limited species.

References

Audacity Team. Audacity (version 1.3. 4-beta)[computer program].

Austin JD, Lougheed SC, Neidrauer L, Chek AA, Boag PT. Cryptic lineages in a small frog: the post-glacial history of the spring peeper, Pseudacris crucifer (Anura: Hylidae). Molecular Phylogenetics and Evolution. 2002 Oct 31;25(2):316-29.

Avise JC. Phylogeography: the history and formation of species. Harvard University Press, Cambridge, MA. 2000.

Aymes ND, Mavoungou JF, De Stordeur E, Duvallet G. Landscape, population structure and genetic diversity of Stomoxys calcitrans. Parasite. 2009 Mar 1;16(1):37-41.

Bailey WJ, McCrae AW. The general biology and phenology of swarming in the East African tettigoniid Ruspolia differens (Serville)(Orthoptera). Journal of natural history. 1978 Jun 1;12(3):259-88.

Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Molecular biology and evolution. 1999 Jan 1;16(1):37-48.

Berwaerts K, Van Dyck H, Van Dongen S, Matthysen E. Morphological and genetic variation in the speckled wood butterfly (Pararge aegeria L.) among differently fragmented landscapes. Netherlands Journal of Zoology. 1997 Jan 1;48(3):241-53.

Blair C, Weigel DE, Balazik M, Keeley AT, Walker FM, Landguth E, Cushman SA, Murphy M, Waits L, Balkenhol N. A simulation‐based evaluation of methods for inferring linear barriers to gene flow. Molecular Ecology Resources. 2012 Sep 1;12(5):822-33.

Bonin A, Ehrich D, Manel S. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. Molecular ecology. 2007 Sep 1;16(18):3737-58.

Brown JH, Kodric-Brown A. Turnover rates in insular biogeography: effect of immigration on extinction. Ecology. 1977 Mar 1;58(2):445-9.

Bush SL, Schul J. Evolution of novel signal traits in the absence of female preferences in Neoconocephalus katydids (Orthoptera, Tettigoniidae). PloS one. 2010 Aug 27;5(8):e12457.

Burger JC, Redak RA, Allen EB, Rotenberry JT, Allen MF. Restoring arthropod communities in coastal sage scrub. Conservation Biology. 2003 Apr 1;17(2):460-7.

Chaput‐Bardy A, Lemaire C, Picard D, Secondi J. In‐stream and overland dispersal across a river network influences gene flow in a freshwater insect, Calopteryx splendens. Molecular Ecology. 2008 Aug 1;17(15):3496-505.

Collinge SK, Palmer TM. The influences of patch shape and boundary contrast on insect response to fragmentation in California grasslands. Landscape Ecology. 2002 Oct 1;17(7):647-56.

Deily JA, Schul J. Recognition of calls with exceptionally fast pulse rates: female phonotaxis in the genus Neoconocephalus (Orthoptera: Tettigoniidae). Journal of Experimental Biology. 2004 Sep 15;207(20):3523-9.

Deily JA, Schul J. Spectral selectivity during phonotaxis: a comparative study in Neoconocephalus (Orthoptera: Tettigoniidae). Journal of Experimental Biology. 2006 May 1;209(9):1757-64.

Dhuyvetter H, Gaublomme E, Verdyck P, Desender K. Genetic differentiation among populations of the salt marsh beetle Pogonus littoralis (Coleoptera: Carabidae): a comparison between Atlantic and Mediterranean populations. Journal of heredity. 2005 Feb 24;96(4):381-7.

Dobson AP, Bradshaw AD, Baker AÁ. Hopes for the future: restoration ecology and conservation biology. Science. 1997 Jul 25;277(5325):515-22.

Earl DA. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources. 2012 Jun 1;4(2):359-61.

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology. 2005 Jul 1;14(8):2611-20.

Frederick KH. Phylogenetic structure of katydid (Orthoptera: Tettigoniidae) communities in northeast Missouri grasslands.

Frederick KH. *Investigating an adaptive radiation in temperate Neoconocephalus* (Doctoral dissertation, University of Missouri--Columbia).

Garnier S, Alibert P, Audiot P, Prieur B, Rasplus JY. Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, Carabus solieri. Molecular Ecology. 2004 Jul 1;13(7):1883-97.

Gauffre B, Mallez S, Chapuis MP, Leblois R, Litrico I, Delaunay S, Badenhausser I. Spatial heterogeneity in landscape structure influences dispersal and genetic structure: empirical evidence from a grasshopper in an agricultural landscape. Molecular ecology. 2015 Apr 1;24(8):1713-28.

Gilman RT, Behm JE. Hybridization, species collapse, and species reemergence after disturbance to premating mechanisms of reproductive isolation. Evolution. 2011 Sep 1;65(9):2592-605.

Golden DM, Crist TO. Experimental effects of habitat fragmentation on old-field canopy insects: community, guild and species responses. Oecologia. 1999 Mar 4;118(3):371-80.

Greenfield MD. Evolution of acoustic communication in the genus Neoconocephalus: discontinuous songs, synchrony, and interspecific interactions. The tettigoniidae: Biology, systematics and evolution. 1990:71-97.

Hoerling M, Eischeid J, Kumar A, Leung R, Mariotti A, Mo K, Schubert S, Seager R. Causes and predictability of the 2012 Great Plains drought. Bulletin of the American Meteorological Society. 2014 Feb;95(2):269-82.

Holland BR, Clarke AC, Meudt HM. Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. Systematic Biology. 2008 Jun 1;57(3):347-66.

Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution. 2005 Oct 12;23(2):254-67.

Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007 May 7;23(14):1801-6.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012 Apr 27;28(12):1647-9.

Keyghobadi N, Roland JE, Strobeck C. Genetic differentiation and gene flow among populations of the alpine butterfly, Parnassius smintheus, vary with landscape connectivity. Molecular Ecology. 2005 Jun 1;14(7):1897-909.

Korman AK, Mallet J, Goodenough JL, Graves JB, Hayes JL, Hendricks DE, Luttrell R, Pair SD, Wall M. Population structure in Heliothis virescens (Lepidoptera: Noctuidae): an estimate of gene flow. Annals of the Entomological Society of America. 1993 Mar 1;86(2):182-8.

Knutsen H, Rukke BA, Jorde PE, Ims RA. Genetic differentiation among populations of the beetle Bolitophagus reticulatus (Coleoptera: Tenebrionidae) in a fragmented and a continuous landscape. Heredity. 2000 Jun 1;84(6):667-76.

Krauss J, Schmitt T, Seitz A, Steffan‐Dewenter I, Tscharntke T. Effects of habitat fragmentation on the genetic structure of the monophagous butterfly Polyommatus coridon along its northern range margin. Molecular Ecology. 2004 Feb 1;13(2):311-20.

Kruess A, Tscharntke T. Species richness and parasitism in a fragmented landscape: experiments and field studies with insects on Vicia sepium. Oecologia. 2000 Jan 4;122(1):129-37.

Lamont BB, He T, Enright NJ, Krauss SL, Miller BP. Anthropogenic disturbance promotes hybridization between Banksia species by altering their biology. Journal of evolutionary biology. 2003 Jul 1;16(4):551-7.

Lande R. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. The American Naturalist. 1993 Dec 1;142(6):911-27.

Lynch M, Milligan BG. Analysis of population genetic structure with RAPD markers. Molecular ecology. 1994 Apr 1;3(2):91-9.

MING QL, WANG CZ. Genetic differentiation of Helicoverpa armigera (Hübner) and H. assulta (Guenée)(Lepidoptera: Noctuidae) based on AFLP markers. Insect Science. 2006 Dec 1;13(6):437-44.

Monaghan MT, Spaak P, Robinson CT, Ward JV. Genetic differentiation of Baetis alpinus Pictet (Ephemeroptera: Baetidae) in fragmented alpine streams. Heredity. 2001 Apr 1;86(4):395-403.

Ney G, Schul J. Population structure within the one-dimensional range of a coastal plain katydid. PloS one. 2017 Jun 9;12(6):e0179361.

Ortego J, Aguirre MP, Cordero PJ. Population genetics of Mioscirtus wagneri, a grasshopper showing a highly fragmented distribution. Molecular Ecology. 2010 Feb 1;19(3):472-83.

Paris M, Boyer S, Bonin A, Collado A, DAVID JP, Despres L. Genome scan in the mosquito Aedes rusticus: population structure and detection of positive selection after insecticide treatment. Molecular Ecology. 2010 Jan 1;19(2):325-37.

Paris M, Despres L. Identifying insecticide resistance genes in mosquito by combining AFLP genome scans and 454 pyrosequencing. Molecular Ecology. 2012 Apr 1;21(7):1672-86.

Peakall RO, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Resources. 2006 Mar 1;6(1):288-95.

Phillipsen IC, Kirk EH, Bogan MT, Mims MC, Olden JD, Lytle DA. Dispersal ability and habitat requirements determine landscape‐level genetic patterns in desert aquatic insects. Molecular ecology. 2015 Jan 1;24(1):54-69.

Phillipsen IC, Lytle DA. Aquatic insects in a sea of desert: population genetic structure is shaped by limited dispersal in a naturally fragmented landscape. Ecography. 2013 Jun 1;36(6):731-43.

Pritchard JK, Wen X, Falush D. Documentation for STRUCTURE software, version 2.3. University of Chicago, Chicago, IL.

Ricketts TH. Terrestrial ecoregions of North America: a conservation assessment. Island Press; 1999.

Roach JL, Stapp P, Van Horne B, Antolin MF. Genetic structure of a metapopulation of black-tailed prairie dogs. Journal of Mammalogy. 2001 Nov;82(4):946-59.

Saarinen EV, Reilly PF, Austin JD. Conservation genetics of an endangered grassland butterfly (Oarisma poweshiek) reveals historically high gene flow despite recent and rapid range loss. Insect Conservation and Diversity. 2016 Nov 1;9(6):517-28.

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I. Inbreeding and extinction in a butterfly metapopulation. Nature. 1998 Apr 2;392(6675):491-4.

Samson FB, Knopf FL, Ostlie WR. Great Plains ecosystems: past, present, and future. Wildlife Society Bulletin. 2004 Mar;32(1):6-15.

Schmitt T, Seitz A. Postglacial distribution area expansion of Polyommatus coridon (Lepidoptera: Lycaenidae) from its Ponto-Mediterranean glacial refugium. Heredity. 2002 Jul 1;89(1):20-6.

Schneider JC. Dispersal of a highly vagile insect in aheterogeneous environment. Ecology. 1999 Dec 1;80(8):2740-9.

Schul J, Patterson AC. What determines the tuning of hearing organs and the frequency of calls? A comparative study in the katydid genus Neoconocephalus (Orthoptera, Tettigoniidae). Journal of Experimental Biology. 2003 Jan 1;206(1):141-52.

Seehausen O. Conservation: losing biodiversity by reverse speciation. Current Biology. 2006 May 9;16(9):R334-7.

Seehausen OL, Takimoto G, Roy D, Jokela J. Speciation reversal and biodiversity dynamics with hybridization in changing environments. Molecular ecology. 2008 Jan 1;17(1):30-44.

Snyder RL, Frederick-Hudson KH, Schul J. Molecular phylogenetics of the genus Neoconocephalus (Orthoptera, Tettigoniidae) and the evolution of temperate life histories. PloS one. 2009 Sep 25;4(9):e7203.

Stevens VM, Verkenne C, Vandewoestijne S, Wesselingh RA, Baguette M. Gene flow and functional connectivity in the natterjack toad. Molecular Ecology. 2006 Aug 1;15(9):2333-44.

Stireman III JO, Nason JD, Heard SB. Host-associated genetic differentiation in phytophagous insects: general phenomenon or isolated exceptions? Evidence from a goldenrod-insect community. Evolution. 2005;59(12):2573-87.

Stoner KJ, Joern A. Landscape vs. local habitat scale influences to insect communities from tallgrass prairie remnants. Ecological Applications. 2004 Oct 1;14(5):1306-20.

Taylor SJ, Downie DA, Paterson ID. Genetic diversity of introduced populations of the water hyacinth biological control agent Eccritotarsus catarinensis (Hemiptera: Miridae). Biological Control. 2011 Sep 30;58(3):330-6.

Tscharntke T, Greiler HJ. Insect communities, grasses, and grasslands. Annual review of entomology. 1995 Jan;40(1):535-58.

Tscharntke T, Steffan-Dewenter I, Kruess A, Thies C. Contribution of small habitat fragments to conservation of insect communities of grassland–cropland landscapes. Ecological Applications. 2002 Apr 1;12(2):354-63.

Vandergast AG, Bohonak AJ, Weissman DB, Fisher RN. Understanding the genetic effects of recent habitat fragmentation in the context of evolutionary history: phylogeography and landscape genetics of a southern California endemic Jerusalem cricket (Orthoptera: Stenopelmatidae: Stenopelmatus). Molecular ecology. 2007 Mar 1;16(5):977-92.

Vekemans X, Beauwens T, Lemaire M, Roldán‐Ruiz I. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Molecular ecology. 2002 Jan 1;11(1):139-51.

Walker TJ. Stridulatory movements in eight species of Neoconocephalus (Tettigoniidae). Journal of Insect Physiology. 1975 Mar 1;21(3):595-603.

Walker TJ. Stridulatory movements in eight species of Neoconocephalus (Tettigoniidae). Journal of Insect Physiology. 1975 Mar 1;21(3):595-603.

Walker TJ, Whitesell JJ, Alexander RD. The robust conehead: two widespread sibling species (Orthoptera: Tettigoniidae: Neoconocephalus" robustus").

Watts K, Vanhala T, Connolly T, Cottrell J. Striking the right balance between site and landscape-scale conservation actions for a woodland insect within a highly fragmented landscape: A landscape genetics perspective. Biological Conservation. 2016 Mar 31;195:146-55.

Whitesell JJ. *Biology of the United States Coneheaded Katydids of the Genus Neoconocephalus (Orthoptera: Tettigoniidae)* (Doctoral dissertation, University of Florida).

Wiesner KR, Habel JC, Gossner MM, Loxdale HD, Köhler G, Schneider AR, Tiedemann R, Weisser WW. Effects of habitat structure and land-use intensity on the genetic structure of the grasshopper species Chorthippus parallelus. Royal Society open science. 2014 Oct 1;1(2):140133.

Williams BL, Brawn JD, Paige KN. Landscape scale genetic effects of habitat fragmentation on a high gene flow species: Speyeria idalia (Nymphalidae). Molecular Ecology. 2003 Jan 1;12(1):11-20.

Wright S. Isolation by distance. Genetics. 1943 Mar;28(2):114.

Young A, Boyle T, Brown T. The population genetic consequences of habitat fragmentation for plants. Trends in Ecology and Evolution. 1996 Oct 1;11(10):413-8.

Zhang H, Hare MP. Identifying and reducing AFLP genotyping error: an example of tradeoffs when comparing population structure in broadcast spawning versus brooding oysters. Heredity. 2012 Jun 1;108(6):616-25.

Zhivotovsky LA. Estimating population structure in diploids with multilocus dominant DNA markers. Molecular Ecology. 1999 Jun 1;8(6):907-13.

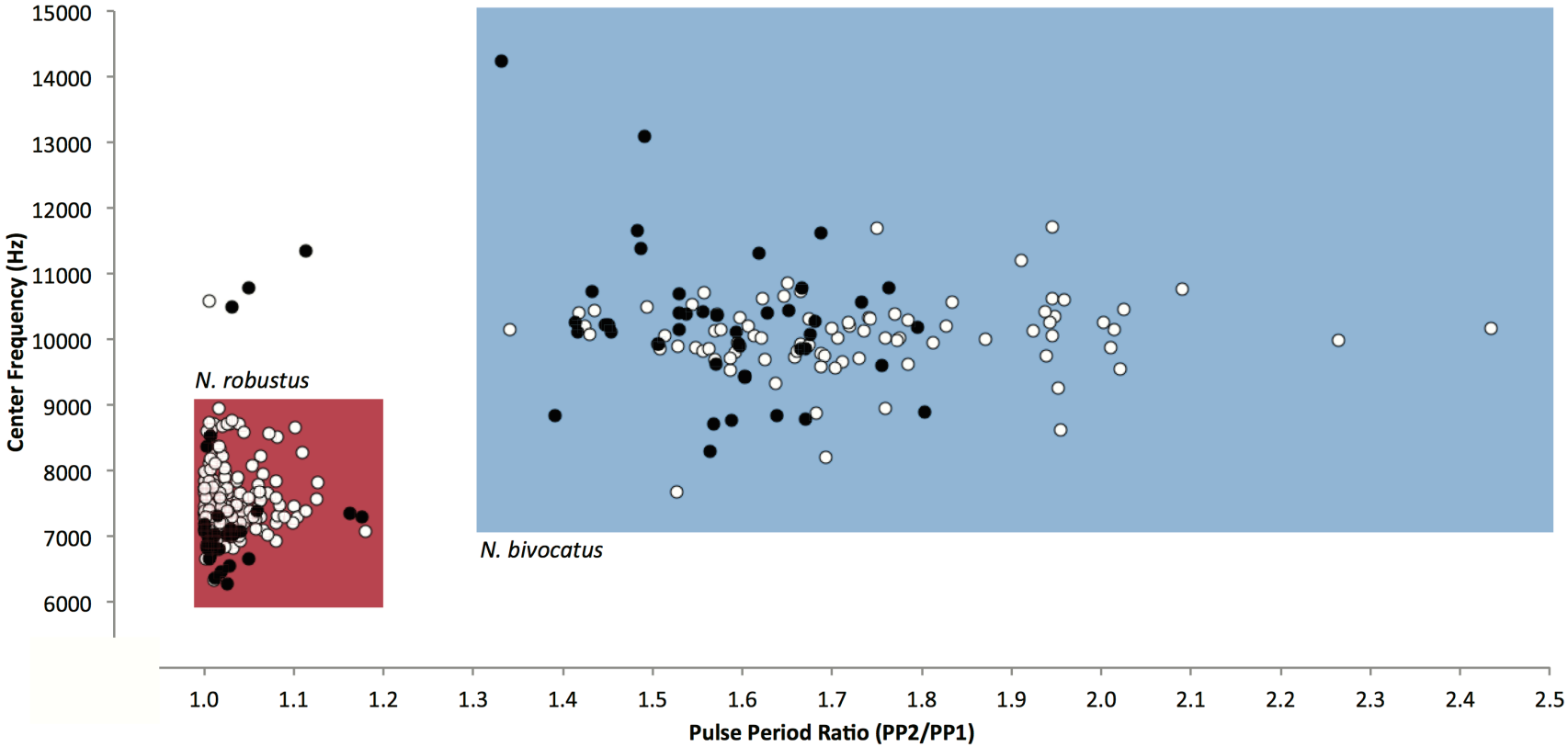


Figure 1: Species assignment based on call pulse period ratio and center frequency. Labeled boxes indicate the calls classified as *N. robustus* and *N. bivocatus*. Individuals sampled in 2013 represented by black circles and 2015 by white circles.

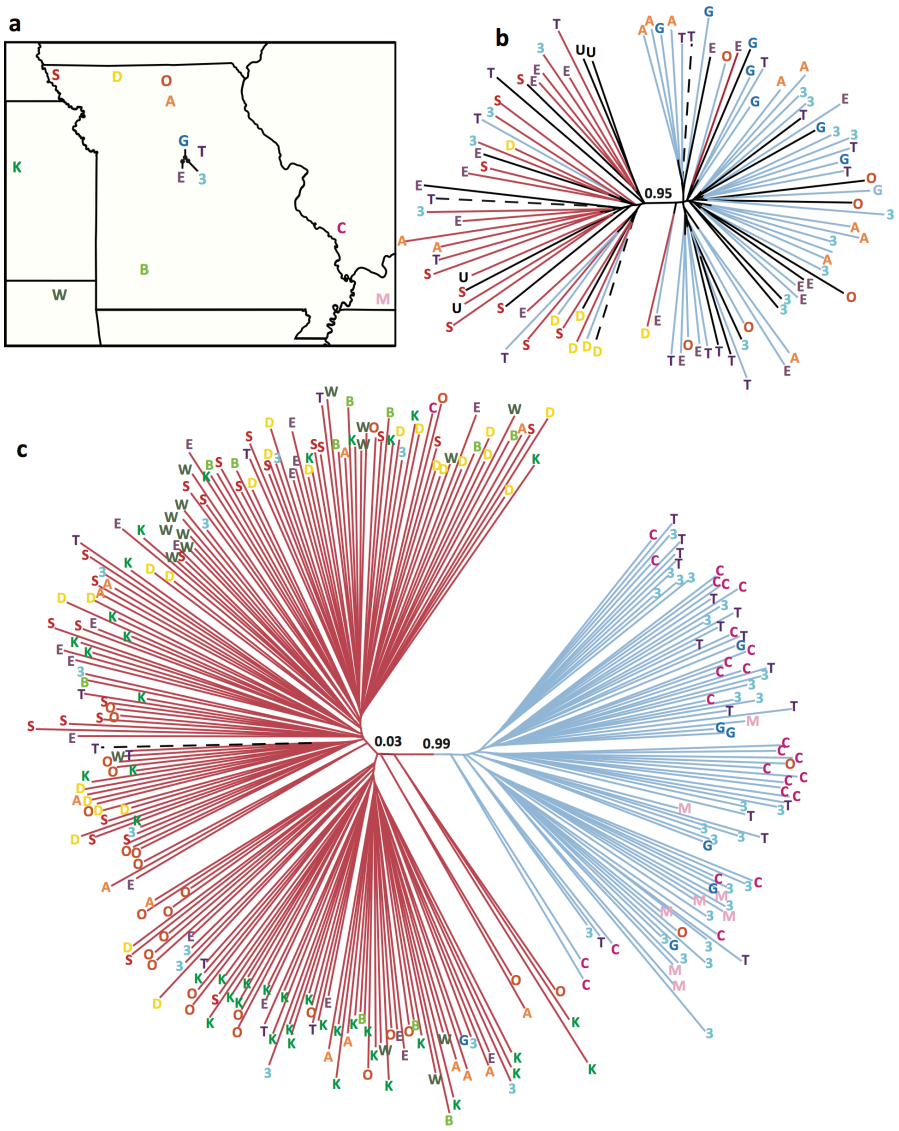


Figure 2: (A) Map of collection localities. Neighbor-joining network analysis implemented in SplitsTree v.4.13 of (B) 2013 individuals and (C) 2015 individuals. Branches leading to individuals are colored according to phenotypic assignment (light *N. robustus* call type, dark *N. bivocatus* call type, dashed-line intermediate call type, black unknown call type) and abbreviations at the end of branch collection sites. Bootstrap support is based on 1,000 replicates of network topology.

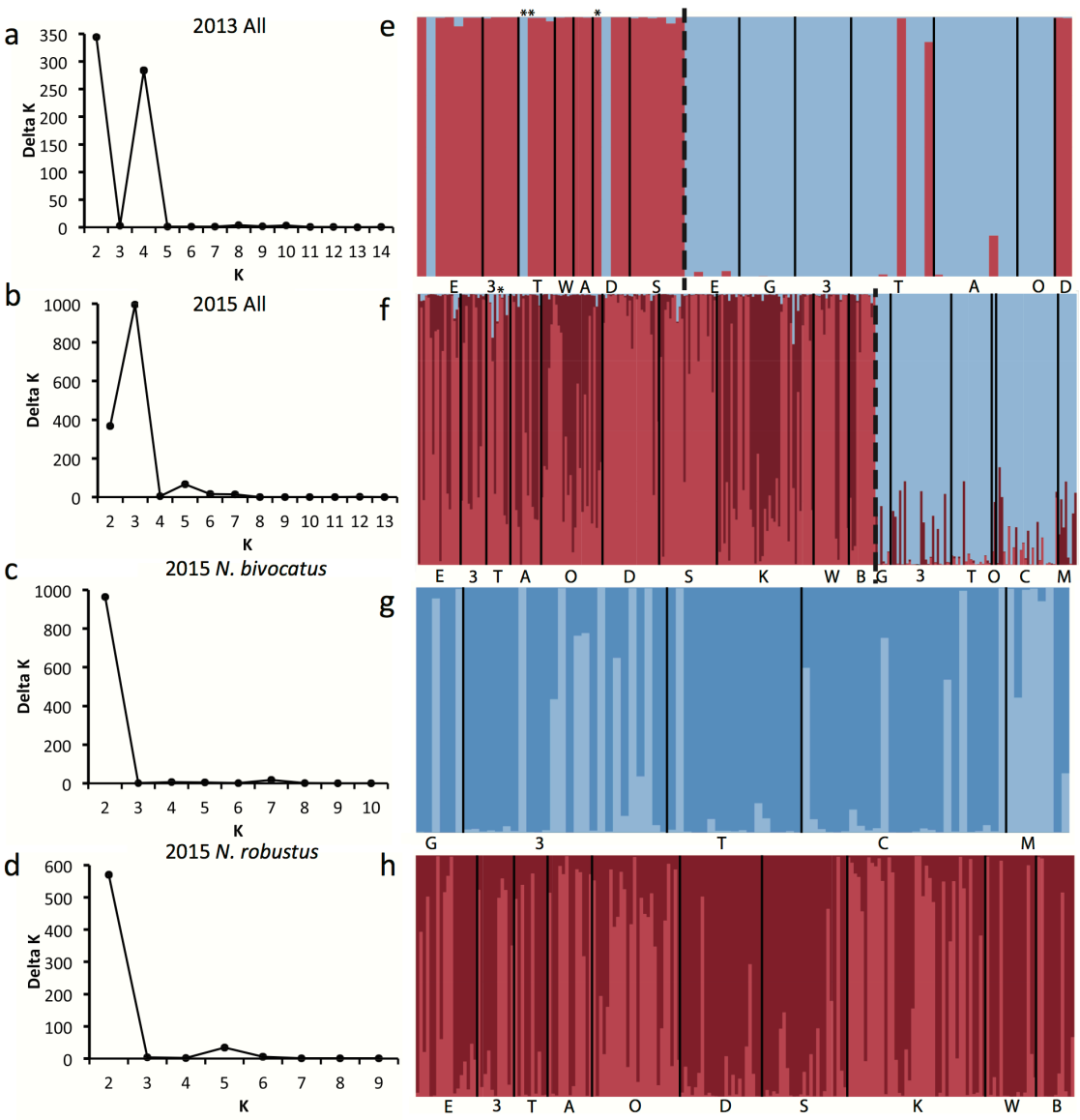


Figure 3: Delta K values for combined Structure runs for (A) all 2013 individuals (B) all 2015 individuals (C) 2015 individuals with a pulse period ratio > 1.3 (*N. bivocatus*) and (D) 2015 individuals with a pulse period ratio < 1.17 (*N. robustus*). Structure analysis bar plots for (E) 2013 individuals K=2, (F) All 2015 individuals K=3, and (G) 2015 individuals with pulse period ratios > 1.3 (*N. bivocatus*) K=2 and (H) 2015 individuals with a pulse period ratio < 1.17 (*N. robustus*). Each bar represents an individual colored by the proportion they assign to each inferred population. Dotted line indicates pulse period ratio cut off of 1.17-1.3. (*) Denotes individuals possessing an intermediate call.

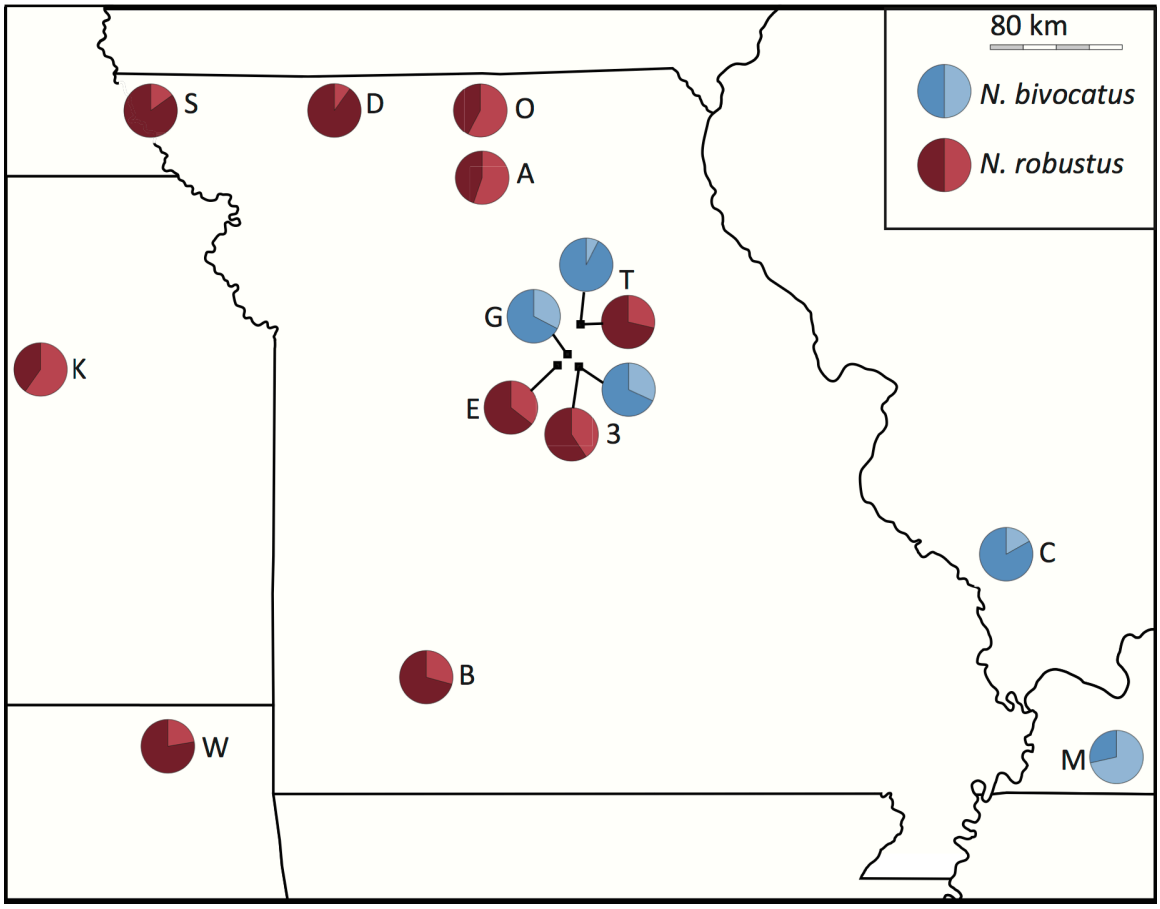


Figure 4: Proportion of population genotypes present at each collection site within 2015 *N. bivocatus* and *N. robustus* and (Fig 3, g and h). Sites possessing five or fewer individuals of the species are not represented.

Table 1: 2013 and 2015 Pairwise F_{ST} values calculated in AFLPsurv v.1.0. Numbers below the diagonal are pairwise F_{ST} values and numbers above the diagonal the geographic distance between sites in kilometers. Bold values represent significant pairwise F_{ST} values. Localities with less than five individuals are not shown.

2013 N. bivocatus	A	E	G	O	3	T
A	-	111.6	114.6	34.4	116.9	113.1
E	0.0000	-	12.3	140.7	16.2	41.0
G	0.0000	0.0000	-	145.3	4.0	29.8
O	0.0000	0.0000	0.0000	-	148.0	146.6
3	0.0000	0.0000	0.0000	0.0000	-	27.0
T	**0.0021**	0.0000	0.0000	0.0000	**0.0075**	**-**

2013 *N. robustus*	D	E	S	3	T
D	**-**	228.5	128.1	241.4	250.0
E	**0.0025**	-	327.9	16.2	41.0
S	0.0000	0.0000	-	343.0	357.8
3	**0.0159**	0.0000	0.0000	-	27.0
T	**0.0028**	0.0000	0.0000	0.0000	-

2015 *N. bivocatus*	C	G	M	3	T
C	**-**	296.3	147.4	292.2	275.2
G	**0.0024**	-	430.1	4.0	29.8
M	**0.0145**	**0.0095**	-	426.1	412.7
3	**0.0029**	**0.0029**	**0.0051**	-	27.0
T	**0.0009**	0.0000	**0.0244**	**0.0011**	**-**

Table 1 continued: 2013 and 2015 Pairwise F_{ST} values calculated in AFLPsurv v.1.0. Numbers below the diagonal are pairwise F_{ST} values and numbers above the diagonal the geographic distance between sites in kilometers. Bold values represent significant pairwise F_{ST} values. Localities with less than five individuals are not shown.

2015 *N. robustus*	A	B	D	E	K	O	S	3	T	W
A	-	302.9	152.5	111.6	361.9	34.4	275.0	116.9	113.1	436.3
B	0.0000	-	361.9	200.8	334.9	323.5	406.9	205.3	228.3	172.7
D	**0.0106**	0.0000	-	228.5	264.4	122.0	128.1	241.4	250.0	443.1
E	0.0000	**0.0010**	**0.0042**	-	357.4	140.7	327.9	16.2	41.0	352.8
K	0.0000	**0.0095**	**0.0205**	**0.0016**	-	348.6	183.5	373.5	397.2	296.5
O	0.0000	**0.0094**	**0.0178**	0.0000	**0.0004**	-	247.6	148.0	146.6	447.5
S	**0.0074**	0.0000	0.0000	**0.0010**	**0.0138**	**0.0129**	-	343.0	357.8	438.1
3	0.0000	0.0000	**0.0073**	0.0000	0.0000	0.0000	**0.0023**	-	27.0	361.6
T	0.0000	**0.0057**	**0.0081**	0.0000	**0.0051**	**0.0062**	**0.0055**	0.0000	-	387.4
W	**0.0033**	0.0000	**0.0045**	**0.0006**	**0.0085**	**0.0120**	**0.0015**	**0.0027**	0.0000	-

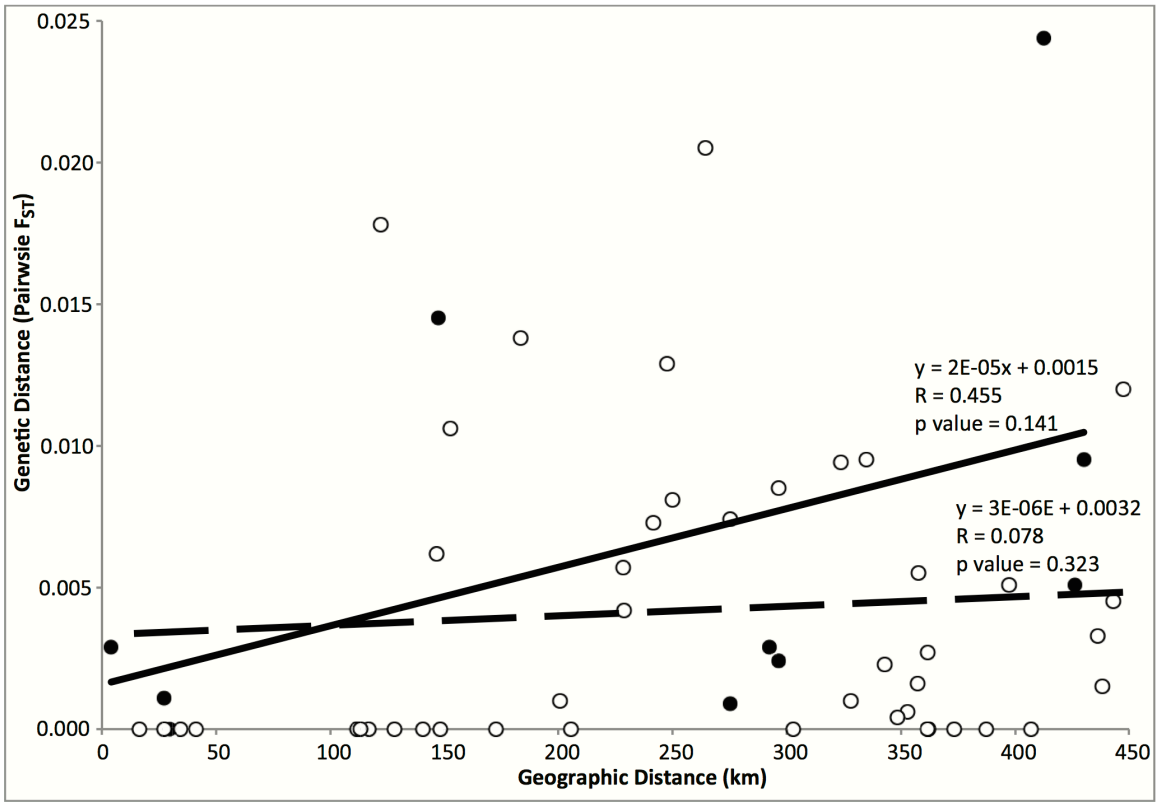
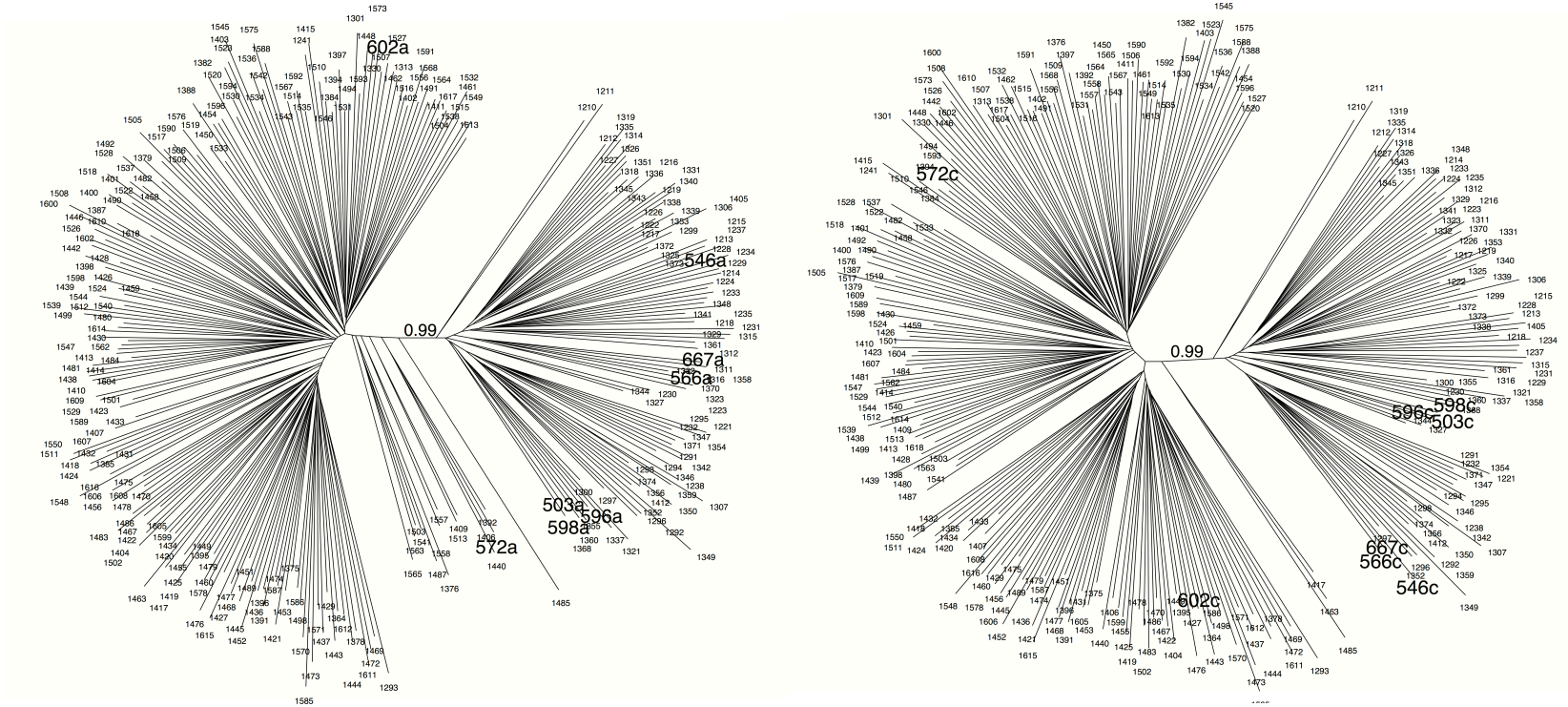


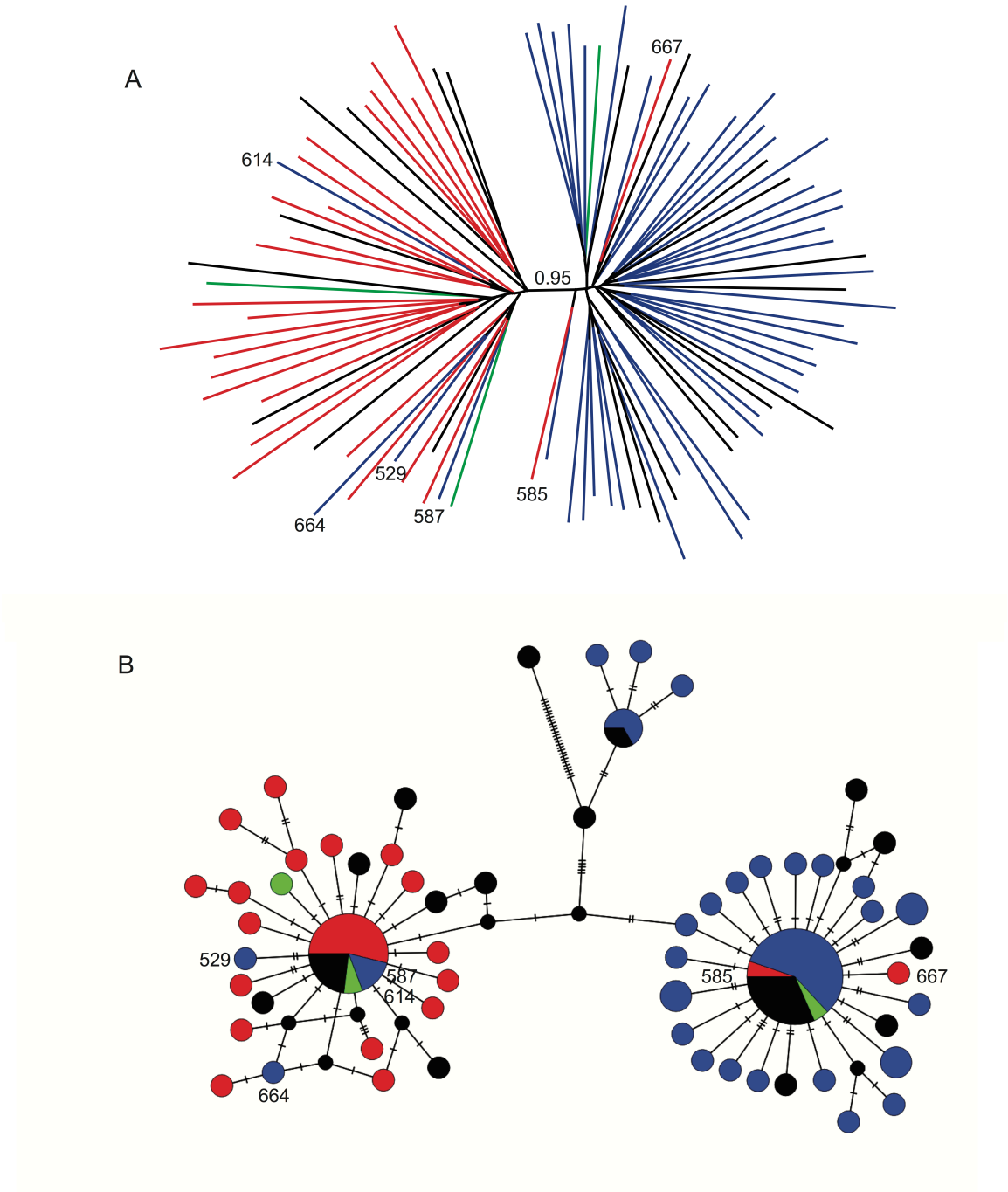
Figure 5: Relationship between pairwise genetic differentiation (pairwise F_{ST}) and geographic distance in 2015 *N. bivocatus* (filled circles, solid line) and *N. robustus* (empty circles, dashed line). Independent Mantel tests were performed for each species; Mantel R and p values for each species are noted.

	Locality	Coordinates	Collection Dates	*N. robustus*	*N. bivocatus*
A	Atlanta State Wildlife Area, MO	39.886, -92.504	7/30/13	2	9
			8/12/15 - 8/13/15	13	0
B	Bois D'Arc Conservation Area, MO	37.279, -93.514	9/10/15	11	0
C	Carbondale, IL	37.717, -89.257	7/24/15	1	26
D	Dunn Ranch Prairie, MO	40.500, -94.110	8/8/13	4	2
			8/31/15	24	0
E	Eagle Bluffs Conservation Area, MO	38.883, -92.461	8/15/13	6	7
			8/10/15	18	0
G	Grassland Trailhead Columbia, MO	38.865, -92.321	8/15/13	0	6
			8/9/15	1	6
K	Konza Prairie Preserve, KS	39.076, -96.589	8/20/15 - 8/21/15	41	0
M	Murray, KY	36.605, -88.351	7/30/15	0	8
O	Oberle Prairie, Liberty Township, MO	40.128, -92.756	8/5/13	0	4
			8/11/15	26	2
S	Star School Hill Prairie Conservation Area, MO	40.550, -95.624	8/9/13	6	0
			8/22/15	25	0
3	Three Creeks Conservation Area, MO	38.849, -92.279	9/4/13	5	6
			8/9/15, 9/17/15	11	26
T	Tucker Prairie Natural Area, MO	38.948, -91.994	8/22/13	4	10
			8/8/15, 9/16/15	11	17
W	White Oak, OK	36.620, -95.274	9/2/15 - 9/3/15	15	0

Supplementary Table 1: *N. robustus* and *N. bivocatus* collection site abbreviations and names, coordinates, dates of collection, and numbers of each species collected.



Supplemental Figure 1: AFLP neighbor-joining network analyses produced from data sets each consisting of all 2015 individuals and one of the replicate sets of eight individuals form 2013. Letters ‘a’ and ‘c’ denote which plate samples were run on. Replicate samples always fell with their matching species cluster. Networks showed no reduction in the between species bootstrap results (>0.99). Bootstrap values within species clusters were low (<0.04), resulting in replicate samples falling in different places within the two analyses.



Supplemental Figure 2: (A) Neighbor-joining AFLP network and (B) median-joining haplotype map of 2013 samples. Colors indicate species-specific call assignments, *N. robustus* (red), *N. bivocatus* (blue), intermediate phenotype (green), and unknown call type (black). Numbered individuals are those that possess a mismatched AFLP cluster

assignment and call phenotype. Within the median-joining CO1 haplotype network three primary haplotype clusters were observed with little genetic differentiation within clusters. Two were made up of primarily *N. bivocatus* individuals and one was made up of primarily *N. robustus* individuals. All individuals found in one of the two CO1 *N. bivocatus* clusters fell within the *N. bivocatus* AFLP cluster and all individuals in the *N. robustus* CO1 cluster were found in the *N. robustus* AFLP cluster. This concordance of nuclear and mitochondrial relationships held true even among individuals possessing the heterospecific call type within genetic clusters.

Supplemental Table 2: Estimated membership coefficient (Q) assignments for population 1 as measured from two independent Structure runs. Data sets consisted of all 2015 individuals and one of the replicate sets of eight individuals form 2013.

Sample	Run 1	Run 2	Delta
503	0.0092	0.0103	0.0011
546	0	0	0
566	0	0.0019	0.0019
572	0.998	0.9994	0.0014
596	0.002	0.001	0.001
598	0.001	0.0379	0.0369
602	0.9788	0.999	0.0202
667	0.1432	0.001	0.1422
		Average =	0.0256

Chapter 3

EPIGENETIC AND GENETIC VARIATION BETWEEN TWO BEHAVIORALLY ISOLATED SPECIES OF *NEOCONOCEPHALUS* KATYDID

Submitted to PLOS One with Johannes Schul as coauthor

Abstract

Epigenetic variation allows for rapid changes in phenotypes without alterations to nucleotide sequences. These epigenetic signatures may diverge over time among isolated populations. Epigenetic incompatibility following secondary contact between these populations could result in the evolution of reproductive isolating mechanisms. If epigenetic incompatibility drove the evolution of species isolating mechanisms, we expect to see significant epigenetic differentiation between these species. Alternatively, epigenetic variation could be the result of predominantly environmental variables and not align along species boundaries. A methylation sensitive amplified fragment length polymorphism analysis was performed on individuals of the closely related katydid species *Neoconocephalus robustus* and *N. bivocatus* collected from eight localities and from three different years. We assessed the diversity and differentiation of epigenetic and genetic markers between species, years, and localities. We observed significant variation in total methylation levels between species. However, genetic differentiation remained larger than epigenetic differentiation between species groups. Methylation levels between individuals collected in 2006 differed significantly from those collected in 2013 and 2014. We measured a significant correlation between the epigenetic and genetic distance between individuals. Epigenetic differentiation is therefore likely the result of an

interaction between genetic and epigenetic loci, and not a mechanism for species differentiation.

Introduction

Epigenetic variation can lead to changes in phenotypic expression without any change to the nucleic acid sequence (Berger et al. 2009). Unlike genetic variation, epigenetic changes can produce reversible, heritable phenotypic changes within a lineage (Jablonka and Lamb 1998). Epigenetic controlled phenotypic plasticity may therefore play an important role in rapid, adaptive changes (Ledón-Rettig et al. 2013; Burggren and Crews 2014). DNA methylation, the most commonly studied form of epigenetic modification, involves the addition of methyl groups, usually to CpG dinucleotides that regulate gene expression (Boyes and Bird 1991). Methylation regulated phenotypic plasticity has been documented across multiple taxonomic groups (Braam and Davis 1990; Rapp and Wendel 2005; Varriale and Bernardi 2006; Elango et al. 2009; Herrera and Bazaga 2010). Until recently epigenetic dependent variation has been largely overlooked as an evolutionary mechanism contributing to speciation (Smith et al. 2016). Facultative phenotypes can be maintained in a population as adaptive alternatives to divergent environmental conditions. These alternative phenotypes may ultimately become fixed in a population by way of genetic assimilation (Pál and Miklós 1999; West-Eberhard 2005).

Phenotypic variation can also evolve in response to epigenetic incompatibility between groups. Following the epigenetic diversification of groups in isolation, intermediate forms may show reduced fitness following secondary contact (Jablonka and

Lamb 1995 1998). Aphids moved to a new host plant and allowed to reproduce asexually quickly developed a preference for the new host and produced morphological characteristics similar to conspecifics utilizing this host. Backcrossing to the parental line results in the production of non-viable offspring (Shaposhinkov 1966). Heritable phenotypic variation in aphids has been correlated with changes in methylation (Field et al. 1989; Walsh et al. 2010). Methylation dependent incompatibility could be reinforced by reproductive isolation (Pál and Miklós 1999).

In many animals, behavioral isolation plays a significant role in maintaining species boundaries. Acoustic communication has been studied as a mechanism for reproductive isolation in both vertebrate and invertebrate groups (Gerhardt and Huber 2002). *Neoconocephalus*, a New World genus of katydids, possesses a diverse range of habitat preferences (Walker et al. 1973; Frederick 2013) and call phenotypes (Schul and Patterson 2003; Beckers and Schul 2008; Deily and Schul 2009). Females utilize these calls in mate recognition and in phonotaxis, the directional movement towards a calling male (Greenfield 1990; Beckers and Schul 2010; Triblehorm and Schul 2009). The acoustic communication system of *Neoconocephalus* allows for reproductive isolation among the multiple species that may be found in sympatry (Schul et al. 2014).

The temperate species *N. robustus* and *N. bivocatus* have divergent habitat preferences and call types, but little genetic differentiation (Snyder et al. 2009). *Neoconocephalus robustus* is a grassland generalist and utilizes grasslands with a wide range of flora and environmental conditions. *Neoconocephalus bivocatus,* on the other hand, is a prairie obligate preferring dry habitats composed primarily of tall prairie grasses (Walker et al. 1973). While morphologically cryptic they maintain divergent call

characters and call preferences (Deily and Schul 2004, 2006). The lack of substantial genetic differentiation between *N. robustus* and *N. bivocatus* could point to an underlying epigenetic incompatibility driving species differentiation.

DNA methylation is common in many insect groups (Field et al. 2004). The presence of significant genome wide cytosine methylation has been measured in multiple Orthopteran species (Robinson et al. 2011; Sarkar et al. 1992). Methylation may play a role in regulating phenotypic variation in some insect groups. For example, DNA methylation may play a role in caste determination within some species of social Hymenoptera (Weiner et al. 2013; Elango et al. 2009).

Few molecular techniques are available for the analysis of epigenetic variation in non-model systems (Ledón-Rettig 2012). Methylation sensitive amplified fragment length polymorphism (MS-AFLP) analysis is one of the few techniques that allows for the quantification of genome wide patterns of cytosine methylation without any previous knowledge of genome sequences (Xiong et al. 1999). The MS-AFLP technique uses two isoschizomer restriction enzymes (MspI and HpaII) that recognize the same restriction site (5'-CCGG-3'), but have different sensitivities to the presence of cytosine methylation. By comparing the presence/absence of fragments produced by both enzymes, the methylation status at each restriction site can be evaluated. In addition, genetic polymorphisms can be evaluated between individuals that lack the restriction site in both digestions as an indication of a change in nucleotide sequence.

If the phenotypic variation between species is the result of epigenetic mediated changes, we predict substantial epigenetic variation among heterospecifics. *Neoconocephalus robustus* and *N. bivocatus* provide the opportunity to investigate the

influence of epigenetic, as well as genetic, variation on divergent behavioral phenotypes. In this study, we address three questions. First we ask to what extent the epigenetic and/or genetic differentiation align with the divergent call types. Second we test whether methylation patterns differed among collection years and/or collection sites. Such differences would suggest environmental influences on gene methylation. Third, we analyzed whether the patterns of genome-wide DNA methylation are correlated with genetic variation. We utilized a MS-AFLP technique to differentiate between methylation sensitive loci (MSL) and genetic loci in *N. robustus* and *N. bivocatus* collected from three different years in grassland sites across Missouri.

Methods

Specimen Collection:

We utilized 94 male *N. bivocatus* and *N. robustus* collected in the Summers of 2006, 2013, and 2014 from eight grassland sites around the state of Missouri (Supplemental Table 1). We used males' calls to localize individuals in the field and collected them by hand after sunset. We identified males in the field as belonging to members of the target group through their call and morphological features, including cone pigmentation and body size, prior to collection (Walker et al. 1973). As the two species are morphologically cryptic, we identified individuals as either *N. bivocatus* or *N. robustus* based on their divergent call characteristics as described below.

Call recordings:

We recorded male calls in 2006 and 2013 within three days of collection using an Audiotechnica ATR 55 microphone and a Marantz PMD-671 solid-state recorder (16 bit, 48 kHz sampling rate). Recordings were made outdoors at ambient temperatures (25-28° C) with males placed in individual mesh cages (approximately 10x20x10 cm) spaced at least 3 m apart. In 2014, we recorded male calls in the field immediately preceding collection using a Tascam DR-40 linear PCM recorder (16 bit, 48 kHz sampling rate). Ambient temperatures ranged from 22-27° C.

Temporal call analysis:

We marked each sound pulse [produced during a single closing movement of the wings (Walker 1975)] using custom software. In short, the recordings were rectified and the envelope extracted with a temporal resolution of 0.125 ms. Pulses were detected automatically using a threshold based algorithm and manually checked before saving the data to text-files. We analyzed about 2 s of each male's call containing 150-250 pulses. Further analyses were conducted on the text files in MS Excel.

Male calls of *N. robustus* have a single pulse rate of about 200/s, equivalent to a pulse period of about 5 ms (Walker 1975). Females recognize this pattern by the absence of silent gaps longer than about 2 ms (Deily and Schul 2004). In contrast, male calls of *N. bivocatus* have two alternating pulse periods of about 4.5 ms and 7 ms (Deily and Schul 2004), resulting in a 'galloping rhythm' or 'pulse pairs'. Female *N. bivocatus* recognize this pattern by the rate of the pulse pairs: pulse pair rates around 87/s were attractive and attractiveness decreased towards faster and slower rates. For the pulse pairs to be

detectable, the two alternating pulse periods must differ sufficiently (Schul, unpublished data).

We quantified the ratio of the means of the alternating pulse periods (longer pp / shorter pp). For the single pulse calls of *N. robustus*, this should result in values close to one, while in *N. bivocatus* significantly larger values result (Bush and Schul 2010). Among the calls of 94 males analyzed, there were two distinct groups of values, one between 1 and 1.18 and a second group with ratios > 1.38. A natural break occurred in the data between 1.18 and 1.38. We classified ratios < 1.18 as '*N. robustus*', and > 1.38 as '*N. bivocatus'* in call type.

Spectral call analysis:

We analyzed the amplitude spectrum of male calls using a fast Fourier transformation (FFT, Hamming window, frame length 256) in Audacity v.1.3 (Audacity 2008). Spectra were averaged over two seconds of the call. The main energy in the spectrum of *Neoconocephalus* calls is concentrated in a narrow band between 7 and 15 kHz, with the peak frequency differing among species (Schul and Patterson 2003). We measured the center frequency of this low frequency band as the geometric mean of the upper and lower cut-off frequencies at -3dB from the peak amplitude.

In *N. bivocatus*, the center frequencies of this band have a mean of about 10 kHz among individuals, ranging from 7 to 15 kHz. Females have little spectral selectivity in this frequency range (Deily and Schul 2006). The center frequencies of calls with a pulse period ratio of >1.38 that also fell into this range of center frequencies were classified as *N. bivocatus*. One individual had a pulse period ratio greater than 1.38 and a center

frequency lower than seven kHz; this male was classified as an intermediate caller, as its call characteristics differed from that of both species.

In *N. robustus*, the low frequency band is narrower than in *N. bivocatus* and are typically limited to 10 kHz and below (Schul and Patterson 2003). Indeed, frequencies well above 10 kHz have an inhibitory effect on female phonotaxis (Deily and Schul 2006). Of the individuals with pulse period ratios < 1.18, center frequencies of all but four individuals clustered between 6 and 10 kHz. The four remaining individuals had center frequencies ranging from 10.4 to 11.1 kHz, which would be less attractive for *N. robustus* females. These individuals were classified as possessing an intermediate call type. All intermediate callers were removed from the analysis of epigenetic and genetic differentiation.

Molecular analysis:

We removed the hind femurs of collected males and placed them in 95% EtOH for DNA preservation. We later extracted DNA from the hind femurs using the DNeasy Blood and Tissue Kit (Qiagen Inc., Valencia, CA, USA). DNA quantification was performed on each sample by spectrophotometry (NanoDrop 1000, Thermo Scientific, Wilmington, DE). Genomic DNA was stored at -80° C prior to molecular analysis.

We used a MS-AFLP assay modified from Xu et al. (2000). DNA (55 ng) from each sample was digested in two separate double digest reactions (EcoRI/HpaII and EcoRI/MspI). EcoRI selectively cuts the sequence 5'-GAATTC-3'. HpaII and MspI are isoschizomers, meaning they selectively cut at the sequence (5'-CCGG-3'), but differ in their sensitivity to cytosine methylation at those sites. HpaII will only cut if the external

cytosine is hemimethylated. MspI cuts when the internal cytosine is either hemi or fully methylated. Both enzymes will cut if the target site is completely unmethylated. Using this method we can evaluate the CpG methylation of restriction sites by comparing the fragments produced by both digests.

Digestion and ligation were carried out together to prevent regeneration of restriction sites. Synthetic double stranded DNA adaptors (Xu et al. 2000) were ligated to the cleaved ends of restriction sites. The EcoRI/HpaII and EcoRI/MspI digestion/ligation reaction (11 ul final volume) is comprised of 1.1 μL 10X CutSmart^{TM} buffer (New England Biolabs, USA), 0.55 μg/μL of Bovine Serum Albumin (BSA) (New England Biolabs), 0.3 μL water, 5 U of EcoRI HF, 1 U of either HpaII or MspI, 1 U T4 ligase (New England Biolabs), 1 μL ATP (10 μM), EcoRI adaptors (5 μM), either HpaII or MspI adaptors (50 μM), and 5.5 μL genomic DNA (10 ng/ μL). The reaction was incubated at 37 °C for 2 hours. Preselective PCR was conducted with 1:10 dilute digestion/ligation products and the Eco+A (5'-GACTGCGTACCAATTCA-3') and the HpaII/MspI+A (5'-GATGAGTCTAGAACGGA-3') primers using thermocycler settings as described in Snyder et al. (2009). We performed selective PCR independently with three primer pairs (Table 1) and 1:100 dilute preselective PCR products from both the HpaII and MspI reactions. Fluorescently labeled Eco primers (Eco+AAC (6FAM), Eco+AGC (PET)) were used in selective PCR (as described in Snyder et al. 2009) and the products multiplexed and diluted to produce a 1:10 final dilution of each product. Fragments were separated in an ABI 3730 genetic analyzer at the DNA Core Facility, University of Missouri. We called MS-AFLP bands using GeneMarker v.1.6 (Softgenetic Corp, State College, PA, USA) using an automated peak-calling scheme (as described in

Holland et al. 2008) and called alleles between 75-500 bp with a minimum peak intensity of 50.

Data analysis:

We obtained presence/absence fragment data for both EcoRI/HpaII and EcoRI/MspI datasets from GeneMarker. Presence of a fragment in both the EcoRI/HpaII and EcoRI/MspI digestions (1/1) was described as unmethylated. The presence of a fragment in one digestion, but not the other (0/1 or 1/0), was defined as methylated (either hemimethylated or internal cytosine methylated respectively). If fragments were absent in both digestions (0/0) the loci was considered uninformative as this state could be due to either full methylation at the target site (methylation of both the inner and outer cytosine) or the absence of the fragment due to variation in the nucleotide sequence between individuals (Schulz et al. 2013; Fulneček and Kovařík 2014).

The methylation sensitivity of each locus was identified using the MSAP Package (Pérez-Figueroa 2013) in R v.3.2.0. Loci were classified as methylation sensitive loci (MSL) if there was evidence of methylation in at least 5% of the sampled individuals at that locus. Genetic markers were extracted from the MS-AFLP loci. Fragments that were present in one or both MspI and HpaII analyses were scored as present. Fragments that were absent in both reactions were scored as absent. This method of using MS-AFLP loci to estimate genetic parameters has been found to produce similar results to standard AFLP markers (Smith et al. 2016). Only polymorphic MSL and genetic, loci that differed among sampled individuals, were used in further analyses. We evaluated variation for each epigenetic (MSL) and genetic locus individually then calculated the mean diversity

of MSL and genetic loci using Shannon's diversity index (*I*). The relative frequency of total methylation (internal cytosine methylated and hemimethylated fragments) and unmethylated MSL was compared between species, years, and collection sites. Significant variation between groups was tested using a Kruskal-Wallis test and Mann-Whitney U tests for post hoc analyses.

We tested for differences in MSL and genetic differentiation using two-way analyses of molecular variance (AMOVA; Excoffier et al. 1992) that grouped individuals by species, collection years, and collection locality. Significance of the test statistic's (Φ_{ST}) deviation from zero was estimated based on 1000 permutations of individuals among groups. Principal coordinate analyses (PCoA) of both epigenetic and genetic loci were performed using the R stats package v.3.2.0 (cmdscale; R Core Team 2015) to visualize the Euclidean distance between species, collection years, and collection localities. A population structure analysis was implemented for all individuals using the program STRUCTURE v.2.3.3 (Pritchard et al. 2010). The admixture model was used, allele frequencies correlated, with a run length of 100,000 (Burnin=10,000) for 10 replicates each of K=1-10 (genetic clusters). The estimate of the most well supported K was calculated as described in Evanno et al. (2005) and implemented in Structure Harvester v.0.6.94 (Earl 2012). The program Clumpp v.1.1.2 (Jakobsson and Rosenberg 2007) was used to align the 10 repetitions of the most well supported number of clusters.

If similar signals exist for the MSL and genetic structure then epigenetic and genetic distance may be significantly correlated. We estimated the Euclidean distance between individuals for both epigenetic and genetic datasets using the R stats package v.3.2.0 (R Core Team 2015) and compared the distance matrices using a Mantel test

(Mantel 1967; ade4 package v.1.7.2; Dray et al. 2007). We utilized 10,000 permutations of the design matrix to determine the significance of the correlation coefficient.

Results

Genome wide variation in methylation:

We included 94 males in our analysis. We determined the males' species assignments based on the temporal and spectral frequency call preferences of *N. robustus* and *N. bivocatus*. The call analysis classified the 94 males as 31 *N. bivocatus*, 57 *N. robustus*, and 6 having an intermediate call type; five of the intermediates had the *N. robustus* temporal pattern and one the *N. bivocatus* pattern, with frequency spectra that fell outside of these respective species' specific pattern (Fig 1). Individuals possessing intermediate call phenotypes were removed from further analyses. The MS-AFLP analysis yielded 227, 277, and 318 fragments in each of three selective PCR reactions (Table 1). Of these fragments, a total of 364 were polymorphic for their methylation status among sampled individuals and were classified as MSL. A loci-calling scheme, utilizing the MS-AFLP dataset, allowed for the inference of the genetic state of fragments among individuals, similar to a traditional AFLP analysis. In total 668 polymorphic genetic loci were produced from 822 MS-AFLP fragments.

We compared species, years, and localities to determine whether these groups differed significantly in their proportion of genome-wide methylated sites. The relative frequency of genome wide methylation (combined hemimethylation and internal cytosine methylation) shows low (Fig 2; *N. robustus* = 0.272; *N. bivocatus* = 0.240) but significant variation between species (Mann-Whitney U, $W = 587$, $p = 0.001243$). The frequency of

unmethylated MSL in 2006 was significantly different than 2013 (Fig 2; Mann-Whitney U, W = 257.5, p = 0.002168) and 2014 (Fig 2; Mann-Whitney U, W = 137.5, p = 0.04183). The variation between collection localities was not significantly different in either overall frequency of methylated or unmethylated sites.

Epigenetic and genetic structure:

We examined epigenetic and genetic diversity as they relate to species assignment. The within species epigenetic Shannon diversity index was 5.0616 ± 0.2054 and 5.0094 ± 0.2087, within *N. robustus* and *N. bivocatus* respectively. *Neoconocephalus robustus* and *N. bivocatus'* genetic Shannon diversity indexes were 5.2802 ± 0.2242 and 5.2179 ± 0.1943 respectively. Genetic diversity was significantly greater than epigenetic diversity in both species (Wilcoxon rank sum test; *N. robustus*, W = 2760, p < 0.0001; *N. bivocatus*, W = 779, p = 0.0001).

Epigenetic and genetic structure was evaluated using a two-level AMOVA. The between species epigenetic divergence was Φ_{ST} = 0.0504 (P < 0.0001). This is slightly less than one-third of the genetic divergence observed between species, Φ_{ST} = 0.1591 (P < 0.0001). Greater genetic variation between species would suggest that genetic mechanisms are underlying species differentiation. Significant genetic and epigenetic variation was observed between collection years in both species (Table 2). In particular, 2006 showed significant differentiation in both MSL and genetic loci from collection years 2013 and 2014 in *N. robustus* (Table 3). In *N. bivocatus* only genetic differentiation was significant between years 2006 and 2013. Neither genetic nor epigenetic markers showed differentiation between 2013 and 2014 (Table 3). As only a single 2014 *N.*

bivocatus was present in the dataset, no comparisons were made with *N. bivocatus* from other collection years.

We observed no significant epigenetic or genetic differentiation between collection sites in either species (data not shown). While samples from most sites were composed of predominantly one species or the other, at two sites we collected enough individuals of each species to compare genetic and epigenetic differentiation between species collected from the same locality. Individuals from Three Creeks Conservation Area (Boone County, Missouri. USA) showed significant genetic differentiation between species ($\Phi_{ST} = 0.1310$; $p < 0.0001$), but no significant epigenetic differentiation ($\Phi_{ST} = 0.0017$; $p = 0.3987$). A second site, Atlanta State Wildlife Area (Macon County, Missouri. USA), showed a similar pattern of genetic, but not epigenetic differentiation (genetic $\Phi_{ST} = 0.506$, $p = 0.0275$; MSL $\Phi_{ST} = -0.022$, $p = 0.3842$); with genetic differentiation approaching significance (Bonferroni-corrected alpha level of $p < 0.025$ for two tests).

The principal coordinate analyses (PCoA), calculated using MSL and genetic profiles, showed that genetic variance was smaller than epigenetic variance within species (Fig 3). The epigenetic PCoA between species elucidated little meaningful differentiation among groups, with ellipses (95% confidence intervals) overlapping almost entirely. The PCoA of genetic data between species explained 17.7% of the genetic variance in the first two coordinates (Fig 3A) and showed moderate variation between individuals based on call phenotype, with species showing divergence in Euclidean space. The between year PCoA of epigenetic loci showed little evidence of divergence in the first two coordinates (Fig 3D). The PCoA of genetic loci (Fig 3C)

showed that individuals grouped predominately along species lines, with some variance in the amount of genetic variation present between years. Three *N. robustus* individuals clearly fall into the *N. bivocatus* cluster within the genetic PCoA (Fig 3A), indicating a possible mismatch between phenotypic and genotypic assignments.

The Bayesian analysis of genetic structure revealed the best-supported number of genetic cluster to be K=2, based on ΔK values (Fig 4B; Evanno et al. 2005). These two genetic clusters aligned closely to species boundaries (Fig 4A). As in the PCoA the same three individuals identified as *N. robustus* assigned primarily to the *N. bivocatus* genetic cluster, indicating that they possess the *N. robustus* call type, but a *N. bivocatus* genotype. As in the PCoA the epigenetic structure analysis did not show support for significant epigenetic (MSL) structure between species groups. ΔK values did not rise above 30 for any permutation in the number of clusters (Fig 4D), an indication of low support for epigenetic structure. In addition, the best supported number of clusters, K=4, showed no grouping of epigenetic clusters by species assignment (Fig 4C).

In every instance, when significant epigenetic variation was observed between species or years, significant genetic variation was also detected. We investigated the correlation between inter-individual genetic and epigenetic Euclidean distance. Between individual epigenetic and genetic distance showed a strong positive correlation (Fig 5 Mantel $R = 0.8448$, $p = 0.0001$).

Discussion

We did not see evidence of epigenetic mediated phenotypic plasticity between *N. robustus* and *N. bivocatus*. While MSL differentiation was significant between species,

genetic differentiation was larger. The AMOVA and PCoA analyses indicate that methylation patterns varied among individuals, but showed less differentiation between species than genetic markers. In all cases where epigenetic differentiation was observed between species or between years, genetic differentiation was as well. A correlation between epigenetic and genetic patterns could explain low levels of epigenetic variation between groups as correlative and having resulted from being pulled along with genetic differentiation.

Genetic differentiation:

Genetic differentiation between species was low, but significant, as would be expected between two closely related taxa (Snyder et al. 2009). Genetic structure aligned with species boundaries, with the exception of the three phenotypically *N. robustus* individuals that showed a majority assignment to the primarily *N. bivocatus* genotypic cluster. Similar occurrences of mismatched genotype/phenotype individuals have been found between *N. robustus* and *N. bivocatus* previously (Frederick 2013; Ney and Schul, In press). These mismatched genotype/phenotype individuals are likely the result of recent hybridization events.

Individuals from 2006 showed genetic differentiation from both 2013 and 2014. We propose this may be the result of significant genetic drift occurring between the years 2006 and 2013. 2013 followed the most severe year of drought in the Midwest since 1895 (Hoerling et al. 2014). A severe reduction in population size, as observed during the drought (Schul, personal observation), could account for the shift in genetic patterns observed between years (Nei et al. 1975). Following a bottleneck event the effects of

demographic changes remain in the population, as observed by the lack of significant genetic differentiation between 2013 and 2014 samples.

Epigenetic variation:

Neoconocephalus robustus showed a significantly higher level of genome wide methylation than *N. bivocatus* (Fig 2). This may be due to variation in the species' habitat preference. Variable environmental conditions can cause shifts in epigenetics, observed between diverging natural habitats (Lira-Medieros 2010), as well as between natural and altered habitat types (Gao et al. 2010). *Neoconocephalus robustus*, unlike *N. bivocatus*, is a grassland generalist and may show epigenetic patterns resulting from exposure to a more variable set of environmental conditions during a lifetime and across generations. Because of this *N. robustus* may possess a larger repertoire of adaptive genes for the varied environmental conditions it may utilize. The higher level of methylation found in *N. robustus* may therefore be the result of higher levels of methylation required to silence the large repertoire of adaptive genes when not in use.

While epigenetic differentiation between species was significant, it remained lower than genetic differentiation. This study did not show support for the epigenetic regulation of species-specific phenotypes; however this does not eliminate a possible epigenetic mechanism underlying phenotypic differentiation. The MS-AFLP technique has many benefits; however there are also some inherent limitations to its application. MS-AFLPs can underestimate genome-wide levels of methylation due to several limitations (Fulneček and Kovařík 2014). As a genome wide scanning method MS-AFLP only detects methylation at 5′-CCGG-3′ sites. In addition it is unable to discriminate

between full methylation states (hypermethylation of both cytosines) and sequence variation at or near the restriction site (Xiong et al. 1999). This technique is most valuable as a first step in investigating genome wide methylation patterns. Phenotypic variation controlled by a smaller number of methylated sites is unlikely to be detected. A similar investigation into genome-wide methylation found no significant differentiation in methylation between the solitarious and gregarious phases of *Locusta migratoria* (Robinson et al. 2011), despite divergent phenotypes and the identification of differentially expressed genes between morphs (Kang et al. 2004).

Correlation between epigenetic and genetic diversity:

MS-AFLP variation often shows correlations with genetic variation (Liu et al. 2012; Lira-Mereiros et al. 2010; Zhang et al. 2010). We observed that individuals showing greater genetic variation on average varied more epigenetically as well (Fig 5). Changes in DNA methylation over time are correlated with genetic relatedness and suggest that DNA methylation maintenance may be under genetic control (Bjornsson et al. 2008). In addition, genetic variation in retrotransposons (i.e. their presence/absence) could affect the methylation state of retroelements, requiring more methylation the more repetitive elements are present (Michaud et al. 1994). Repetitive elements have been estimated to make up thirty percent of the genome of the Orthopteran *L. migratoria* (Wilmore and Brown 1975). In addition, the retrotransposon SINE is differentially methylated between the solitarious and gregarious phases of the species (Guo et al. 2010).

Mechanisms of neutral drift could also account for the correlation between epigenetic and genetic variation. In the event of substantial gene flow between species, strong divergent selection would be needed to maintain divergent epigenetic variation between species. Gene flow between groups would reduce differentiation accumulated via drift. Evidence from this study however suggests that genetic differentiation between species is significant and gene flow, therefore, relatively low (Table 2). Stochastic processes of drift then could allow epigenetic patterns to diverge between species in parallel, resulting in correlations between epigenetic and genetic variation, without a causal link (Richards et al. 2010).

Conclusion:

While our findings did not suggest a mechanism of genome wide methylation in regulating phenotypic variation, we observed clear evidence of genomic methylation in *Neoconocephalus.* Epigenetics may still play a key role in phenotypic differentiation within *Neoconocephalus* katydids through the differential regulation of a relatively small number of genes of large effect; a mechanism not detectable with the methods used here. Further work identifying differentially expressed genes between call types could allow for the targeted analysis of methylation patterns at these sites.

References

Beckers OM, Schul J. Developmental plasticity of mating calls enables acoustic communication in diverse environments. Proc R Soc London B. 2008 Jun 7;275(1640):1243-8.

Beckers OM, Schul J. Female adaptation to developmental plasticity in male calling behavior. Behav Ecol Sociobiol. 2010 Aug 1;64(8):1279-90.

Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. An operational definition of epigenetics. Genes Dev. 2009 Apr 1;23(7):781-3.

Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, et al. Intra-individual change over time in DNA methylation with familial clustering. Jama. 2008 Jun 25;299(24):2877-83.

Boyes J, Bird A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell. 1991 Mar 22;64(6):1123-34.

Braam J, Davis RW. Rain-, wind-, and touch-induced expression of calmodulin and calmodulin-related genes in Arabidopsis. Cell. 1990 Feb 9;60(3):357-64.

Burggren WW, Crews D. Epigenetics in comparative biology: why we should pay attention. ICB. 2014 Apr 9:icu013.

Bush SL, Schul J. Evolution of novel signal traits in the absence of female preferences in Neoconocephalus katydids (Orthoptera, Tettigoniidae). PloS one. 2010 Aug 27;5(8):e12457.

Deily JA, Schul J. Recognition of calls with exceptionally fast pulse rates: female phonotaxis in the genus Neoconocephalus (Orthoptera: Tettigoniidae). J Exp Biol. 2004 Sep 15;207(20):3523-9.

Deily JA, Schul J. Selective phonotaxis in Neoconocephalus nebrascensis (Orthoptera: Tettigoniidae): call recognition at two temporal scales. J Comp Physiol A. 2009 Jan 1;195(1):31-7.

Deily JA, Schul J. Spectral selectivity during phonotaxis: a comparative study in Neoconocephalus (Orthoptera: Tettigoniidae). J Exp Biol. 2006 May 1;209(9):1757-64.

Dray S, Dufour AB, Chessel D. The ade4 package-II: Two-table and K-table methods. R news. 2007 Oct;7(2):47-52.

Earl DA. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012 Jun 1;4(2):359-61.

Elango N, Hunt BG, Goodisman MA, Soojin VY. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, Apis mellifera. Proc Natl Acad Sci. 2009 Jul 7;106(27):11206-11.

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005 Jul 1;14(8):2611-20.

Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992 Jun 1;131(2):479-91.

Field LM, Devonshire AL, Ffrench-Constant RH, Forde BG. Changes in DNA methylation are associated with loss of insecticide resistance in the peach-potato aphid Myzus persicae (Sulz.). FEBS lett. 1989 Jan 30;243(2):323-7.

Field LM, Lyko F, Mandrioli M, Prantera G. DNA methylation in insects. Insect Mol Biol. 2004 Apr 1;13(2):109-15.

Frederick FH. Phylogenetic structure of katydid (Orthoptera: Tettigoniidae) communities in Northeast Missouri grasslands. MS thesis, Truman State University. 2013.

Fulneček J, Kovařík A. How to interpret methylation sensitive amplified polymorphism (MSAP) profiles?. BMC Genet. 2014 Jan 6;15(1):2.

Gao L, Geng Y, Li BO, Chen J, Yang JI. Genome‐wide DNA methylation alterations of Alternanthera philoxeroides in natural and manipulated habitats: implications for epigenetic regulation of rapid responses to environmental fluctuation and phenotypic variation. Plant, Cell Environ. 2010 Nov 1;33(11):1820-7.

Gerhardt HC, Huber F. Acoustic communication in insects and anurans: common problems and diverse solutions. University of Chicago Press.

Greenfield MD. Evolution of acoustic communication in the genus Neoconocephalus: discontinuous songs, synchrony, and interspecific interactions. The tettigoniidae: Biology, systematics and evolution. 1990:71-97.

Guo W, Wang XH, Zhao DJ, Yang PC, Kang L. Molecular cloning and temporal–spatial expression of I element in gregarious and solitary locusts. J Insect Physiol. 2010 Aug 31;56(8):943-8.

Herrera CM, Bazaga P. Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis. New Phytologist. 2010 Aug 1;187(3):867-76.

Hoerling M, Eischeid J, Kumar A, Leung R, Mariotti A, Mo K, et al. Causes and predictability of the 2012 Great Plains drought. Bull Am Meteorol Soc. 2014 Feb;95(2):269-82.

Holland BR, Clarke AC, Meudt HM. Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. Syst Biol. 2008 Jun 1;57(3):347-66.

Hulce D, Li X, Snyder-Leiby T, Johathan Liu CS. GeneMarker® Genotyping Software: Tools to Increase the Statistical Power of DNA Fragment Analysis. *Journal of Biomolecular Techniques : JBT*. 2011;22(Suppl):S35-S36.

Jablonka E, Lamb MJ. Epigenetic inheritance and evolution: the Lamarckian dimension. Oxford University Press on Demand; 1999.

Jablonka E, Lamb MJ. Epigenetic inheritance in evolution. J Evol Biol. 1998 Mar 1;11(2):159-83.

Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007 Jul 15;23(14):1801-6.

Kang L, Chen X, Zhou Y, Liu B, Zheng W, Li R, et al. The analysis of large-scale gene expression correlated to the phase changes of the migratory locust. Proc Natl Acad Sci USA. 2004 Dec 21;101(51):17611-5.

Ledón-Rettig CC, Richards CL, Martin LB. Epigenetics for behavioral ecologists. Behav Ecol. 2012 Sep 25:ars145.

Lira-Medeiros CF, Parisod C, Fernandes RA, Mata CS, Cardoso MA, Ferreira PC. Epigenetic variation in mangrove plants occurring in contrasting natural environment. PLoS One. 2010 Apr 26;5(4):e10326.

Liu S, Sun K, Jiang T, Ho JP, Liu B, Feng J. Natural epigenetic variation in the female great roundleaf bat (Hipposideros armiger) populations. Mol Genet Genomics. 2012 Aug 1;287(8):643-50.

Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Res. 1967 Feb 1;27(2 Part 1):209-20.

Michaud EJ, Van Vugt MJ, Bultman SJ, Sweet HO, Davisson MT, Woychik RP. Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. Genes Dev. 1994 Jun 15;8(12):1463-72.

Nei M, Maruyama T, Chakraborty R. The bottleneck effect and genetic variability in populations. Evol. 1975 Mar 1:1-0.

Ney G, Schul J. Low genetic differentiation between populations of an endemic prairie katydid despite habitat loss and fragmentation. Conserv Genet. 2017. doi: 10.1007/s10592-017-0987-x.

Pál C, Miklós I. Epigenetic inheritance, genetic assimilation and speciation. Journal of Theor Biol. 1999 Sep 7;200(1):19-37.

Pérez‐Figueroa A. msap: a tool for the statistical analysis of methylation‐sensitive amplified polymorphism data. Mol Ecol Resour. 2013 May 1;13(3):522-7.

Pritchard JK, Wen X, Falush D. Documentation for STRUCTURE software, version 2.3. University of Chicago, Chicago, IL.

Rapp RA, Wendel JF. Epigenetics and plant evolution. New Phytol. 2005 Oct 1;168(1):81-91.

Richards CL, Bossdorf O, Verhoeven KJ. Understanding natural epigenetic variation. New Phytol. 2010 Aug 1;187(3):562-4.

Robinson KL, Tohidi-Esfahani D, Lo N, Simpson SJ, Sword GA. Evidence for widespread genomic methylation in the migratory locust, Locusta migratoria (Orthoptera: Acrididae). PLoS One. 2011 Dec 5;6(12):e28167.

Sarkar S, Rao SR, Gupta VS, Hendre RR. 5-Methylcytosine content in Gryllotalpa fossor (Orthoptera). Genome. 1992 Feb 1;35(1):163-6.

Schul J, Bush SL, Frederick KH. Evolution of call patterns and pattern recognition mechanisms in Neoconocephalus katydids. In Insect hearing and acoustic communication. 2014 (pp. 167-183). Springer Berlin Heidelberg.

Schul J, Patterson AC. What determines the tuning of hearing organs and the frequency of calls? A comparative study in the katydid genus Neoconocephalus (Orthoptera, Tettigoniidae). J Exp Biol. 2003 Jan 1;206(1):141-52.

Schulz B, Eckstein RL, Durka W. Scoring and analysis of methylation‐sensitive amplification polymorphisms for epigenetic population studies. Mol Ecol Resour. 2013 Jul 1;13(4):642-53.

Shaposhnikov GK. Origin and breakdown of reproductive isolation and the criterion of the species. Entomol Rev. 1966;45:1-8.

Smith TA, Martin MD, Nguyen M, Mendelson TC. Epigenetic divergence as a potential first step in darter speciation. Mol Ecol. 2016 Feb 1.

Snyder RL, Frederick-Hudson KH, Schul J. Molecular phylogenetics of the genus Neoconocephalus (Orthoptera, Tettigoniidae) and the evolution of temperate life histories. PloS one. 2009 Sep 25;4(9):e7203.

Team A. Audacity (version 1.3)[computer program]. 2008.

Team RC. R: A Language and Environment for Statistical Computing (Version 3.1. 2): R Foundation for Statistical Computing. Vienna, Austria. Available from URL http://www.R-project.org. 2015.

Triblehorn JD, Schul J. Sensory-encoding differences contribute to species-specific call recognition mechanisms. J Neurophysiol. 2009 Sep 1;102(3):1348-57.

Varriale A, Bernardi G. DNA methylation in reptiles. Gene. 2006 Dec 30;385:122-7.

Walker TJ, Whitesell JJ, Alexander RD. The robust conehead: two widespread sibling species (Orthoptera: Tettigoniidae: Neoconocephalus" robustus"). Ohio J Sci. 1973 Nov;73(6):321-30.

Walker TJ. Stridulatory movements in eight species of Neoconocephalus (Tettigoniidae). J Insect Physiol. 1975 Mar 1;21(3):595-603.

Walsh TK, Brisson JA, Robertson HM, Gordon K, Jaubert-Possamai S, Tagu D, et al. A functional DNA methylation system in the pea aphid, Acyrthosiphon pisum. Insect Mol Biol. 2010 Mar 1;19(s2):215-28.

Weiner SA, Galbraith DA, Adams DC, Valenzuela N, Noll FB, Grozinger CM, et al. A survey of DNA methylation across social insect species, life stages, and castes reveals abundant and caste-associated methylation in a primitively social wasp. Naturwissenschaften. 2013 Aug 1;100(8):795-9.

West-Eberhard MJ. Developmental plasticity and the origin of species differences. Proc Natl Acad Sci. 2005 May 3;102(suppl 1):6543-9.

Wilmore PJ, Brown AK. Molecular properties of orthopteran DNA. Chromosoma. 1975 Dec 1;51(4):337-45.

Xiong LZ, Xu CG, Saghai Maroof MA, Zhang Q. Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. MGG. 1999 Apr 4;261(3):439-46.

Xu M, Li X, Korban SS. AFLP-based detection of DNA methylation. Plant Mol Biol Rep. 2000 Dec 1;18(4):361-8.

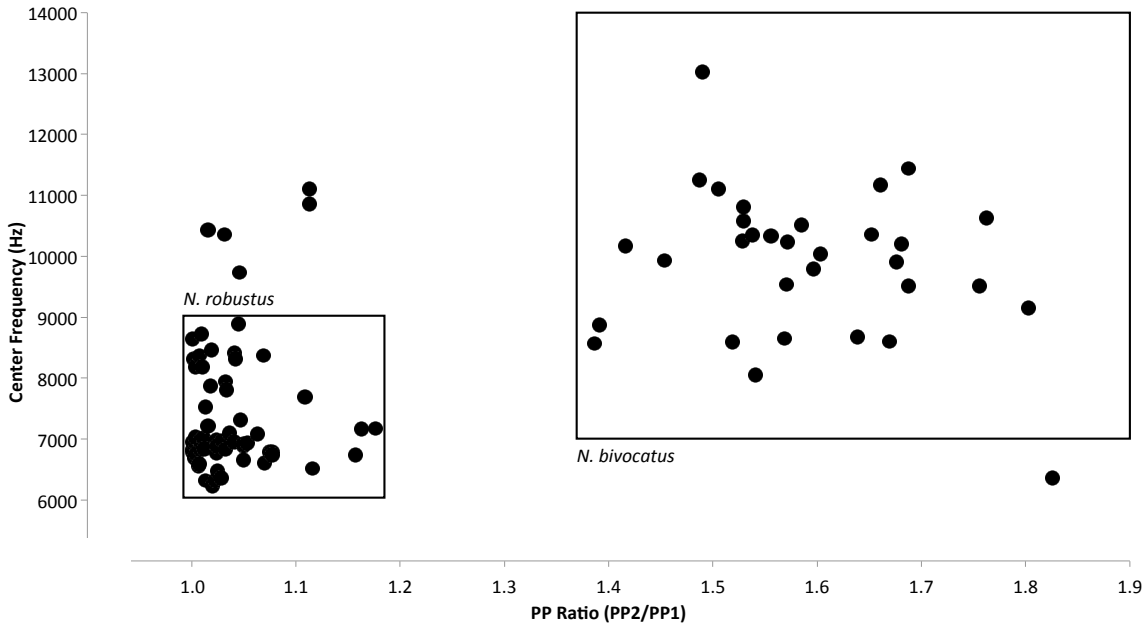


Figure 1: Species assignment based on call pulse period ratio and center frequency. Labeled boxes indicate the calls classified as *N. robustus* and *N. bivocatus*. Individuals that fall outside of species classifications were removed from further epigenetic and genetic analyses.

Table 1: Selective primer combinations used in the MS-AFLP analysis, number of scored bands per primer pair, and the number of those bands classified as polymorphic methylation sensitive loci (MSL) and polymorphic genetic loci.

***Eco*RI**	***Msp*I/*Hpa*II**	**Bands**	**MSL**	**Genetic loci**
-AAC	-ATC	277	183	265
-AGC	-AAT	227	49	102
-AGC	-ATC	318	170	301
	Total	822	402	668

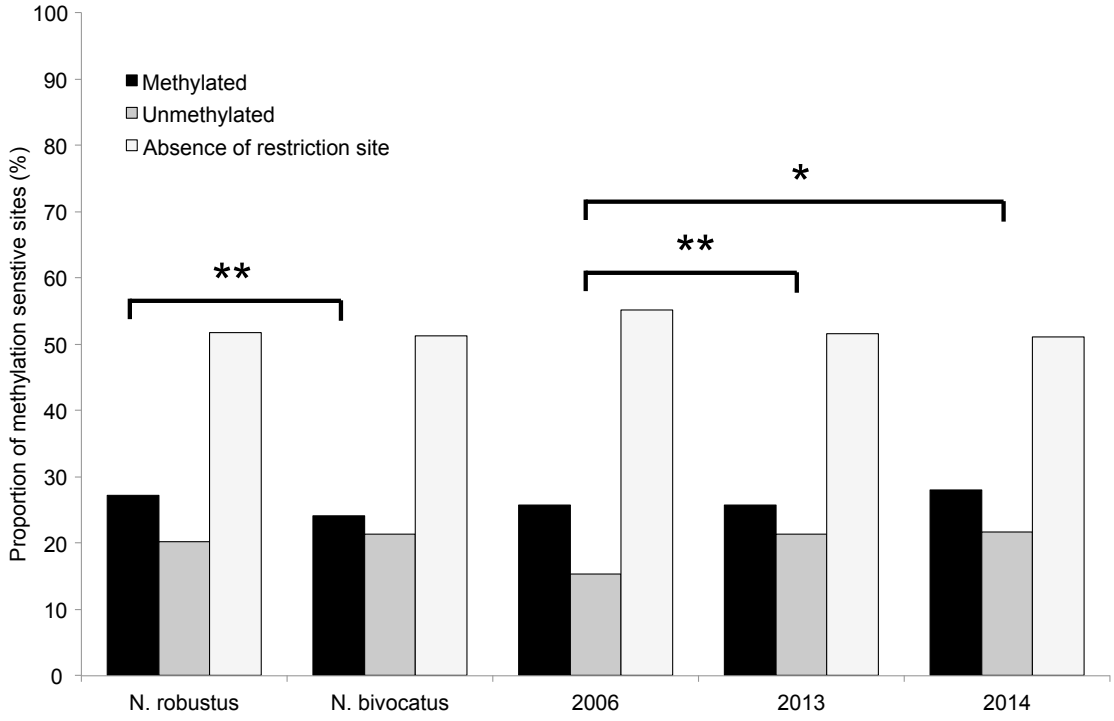


Figure 2: Comparison of genome wide methylation levels between species (left) and collection years (right). Mann-Whitney U test; $p < 0.005$(**) $p < 0.05$ (*). Between-species significant variation is in total methylation (internal cytosine methylated and hemimethylated fragments). Between-year significant variation is in total unmethylated MSL.

Table 2: Two-level AMOVA of methylation sensitive loci (MSL) or genetic loci produced from MS-AFLP markers among populations grouped by species assignment and collection year. Included is genetic and epigenetic variance between groups, Φ_{ST}, and the corresponding p value.

	Between group variance	**Within group variance**	**Φ_{ST}**	**p value**
Between species				
MSL (epigenetic) Genetic loci	3.429 (5.04%) 15.38 (15.91%)	64.54 (94.96%) 81.28 (84.09%)	0.0504 0.1591	< 0.0001 < 0.0001
Between years (*N. robustus*)				
MSL (epigenetic) Genetic loci	2.473 (3.00%) 2.553 (2.99%)	62.82 (97.00%) 82.93 (97.01%)	0.0379 0.0300	0.0004 < 0.0001
Between years (*N. bivocatus*)				
MSL (epigenetic) Genetic loci	1.073 (2.23%) 5.726 (7.27%)	47.05 (97.77%) 73.03 (92.73%)	0.0968 0.0727	0.0019 0.0014

Table 3: Φ_{ST} values computed from AMOVA of MSL (epigenetic) and genetic loci. Significant values ($p < 0.05$) in bold. As only a single 2014 *N. bivocatus* is represented in the dataset, these comparisons have been omitted.

	MSL Φ_{ST}	Genetic Φ_{ST}
N. robustus		
2006 / 2013	**0.068**	**0.055**
2006 / 2014	**0.057**	**0.055**
2013 / 2014	0.011	-0.001
N. bivocatus		
2006 / 2013	0.022	**0.073**

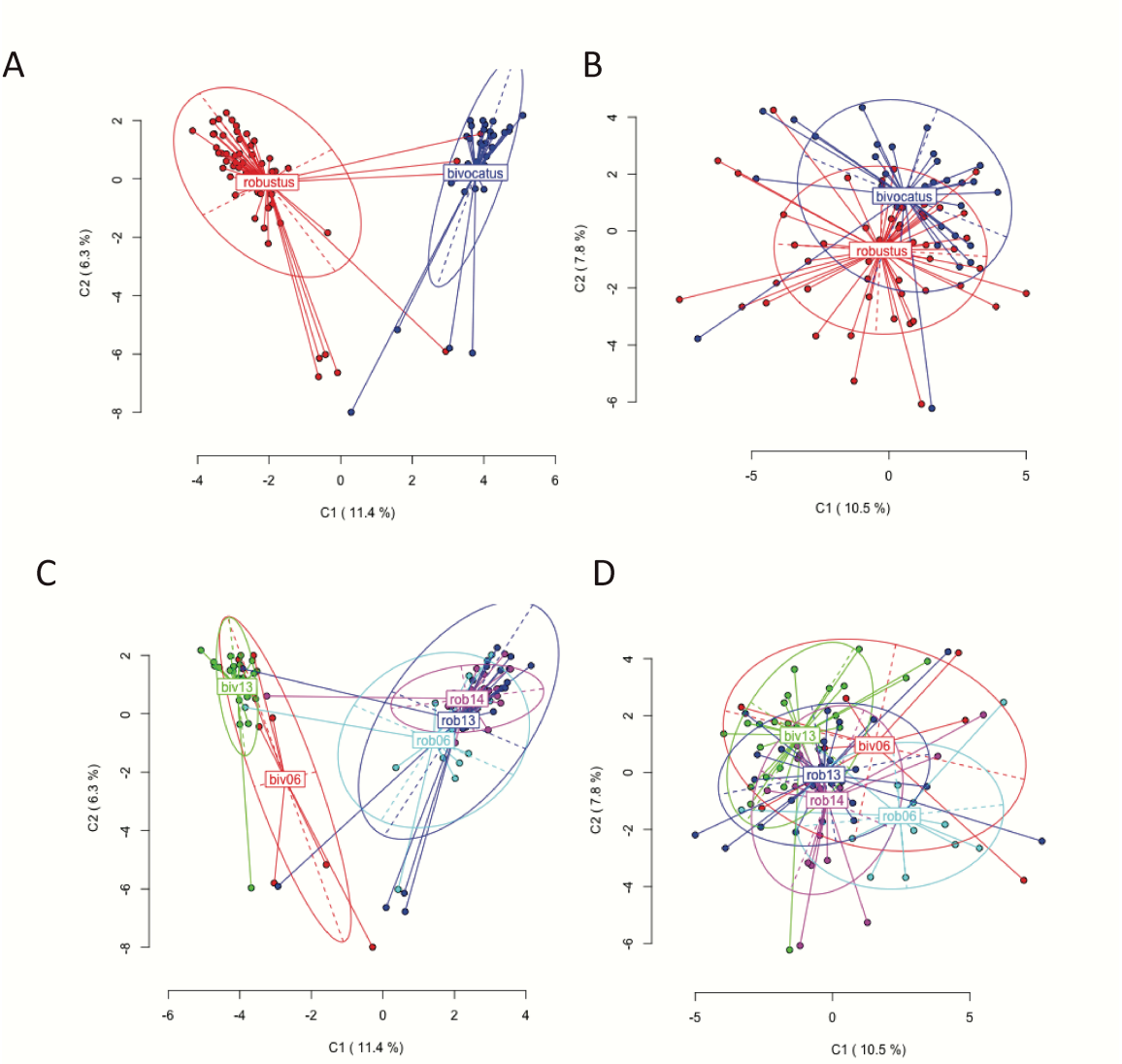


Figure 3: PCoA of *N. robustus* and *N. bivocatus* utilizing genetic and epigenetic data separately. A and C: Genetic Euclidean distance with individuals grouped by species and by year respectively. B and D: Epigenetic Euclidean distance with individuals grouped by species and by year respectively. Group labels show the centroid of the points for each group. The long axis of the ellipse represents the direction of maximum dispersion and the short axis the direction of minimum dispersion.

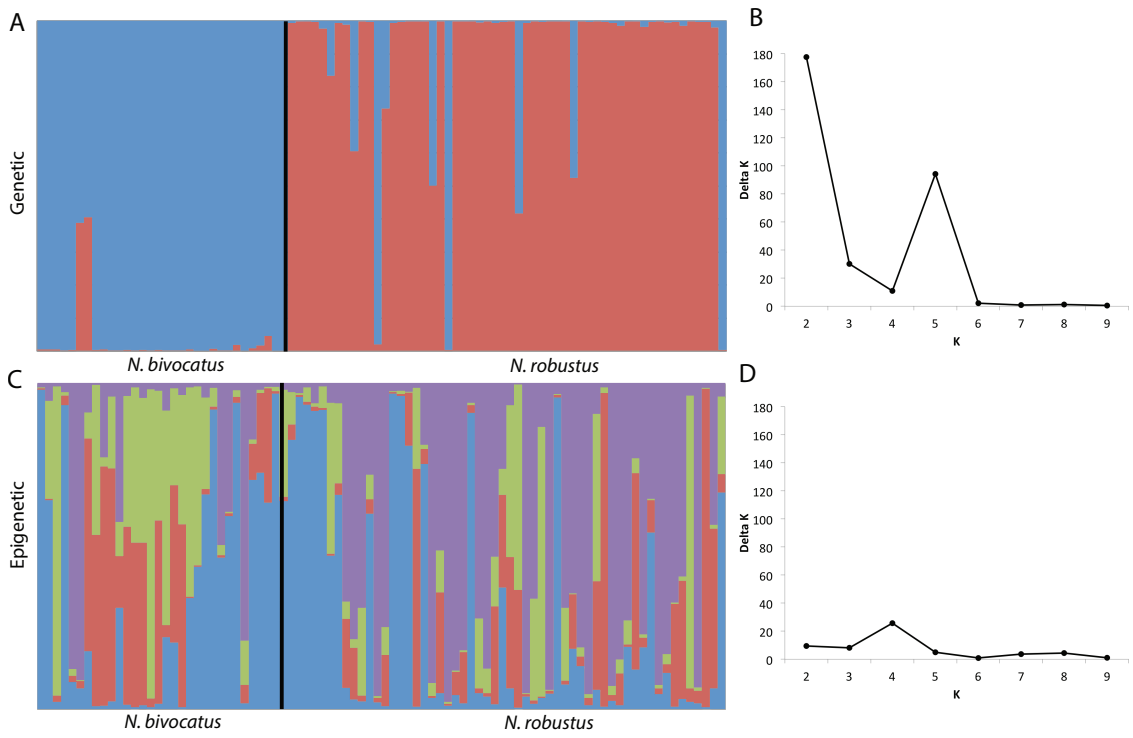


Figure 4: (A and C) Bar plot using MS-AFLP loci to estimate genetic and epigenetic structure among *N. robustus* and *N. bivocatus* using the software package STRUCTURE. (B and D) Delta K graphs for K=1-10 genetic clusters showing moderate support for K=2 genetic clusters and low support for K=4 epigenetic clusters.

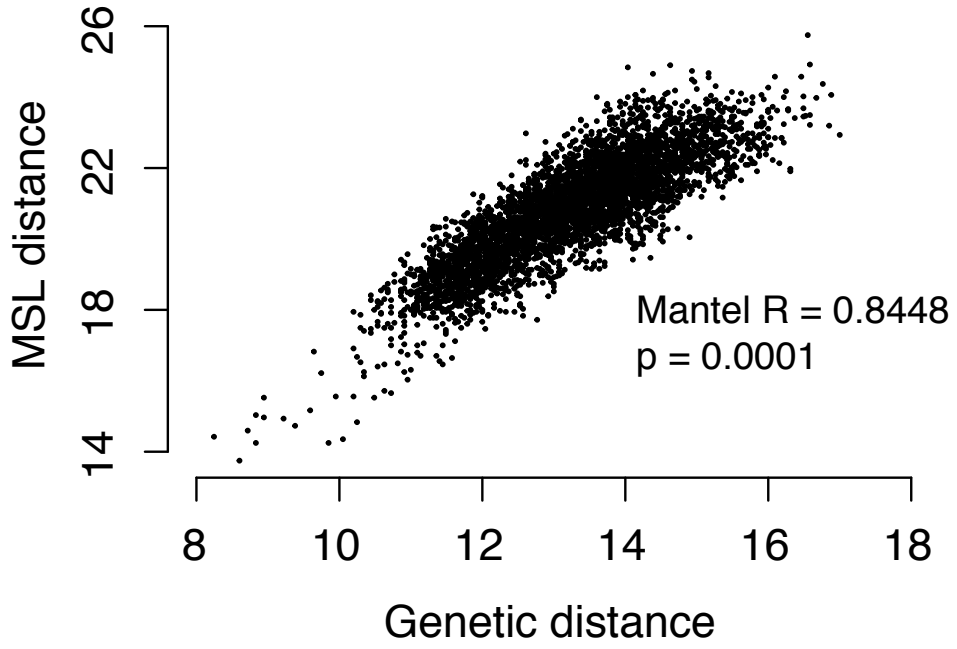


Figure 5: Scatterplot of between individual Euclidean genetic and epigenetic distance showing significant positive correlation between genetic and epigenetic differentiation. The correlation was tested using a Mantel test and 10,000 permutations of the design matrix to determine significance.

Supplemental Table 1: Sample collection localities, locality coordinates, and number of each species sampled in each year (*N. robustus* / *N. bivocatus*). Species assignments were determined based on call phenotypes as show in Figure 1 and described in the methods. The six individuals possessing intermediate phenotypes are not included.

Locality	Coordinates	2006	2013	2014
Atlanta State Wildlife Area	39.886, -92.504	-- / --	2 / 7	-- / --
Dunn Ranch Prairie	40.500, -94.110	-- / --	4 / 1	-- / --
Eagle Bluffs Conservation Area	38.883, -92.461	8 / --	8 / 2	17 / --
Grassland Trailhead	38.865, -92.321	-- / --	-- / 5	-- / --
Oberle Prairie	40.128, -92.756	-- / --	-- / 4	-- / --
Star School Hill Prairie Conservation Area	40.550, -95.624	-- / --	5 / --	-- / --
Three Creeks Conservation Area	38.849, -92.279	2 / 7	5 / 4	3 / --
Tucker Prairie Natural Area	38.948, -91.994	1 / --	2 / --	-- / --

Chapter 4

A POST-PLEISTOCENE CALIBRATED MUTATION RATE FROM INSECT MUSEUM SPECIMENS

To be submitted to PLOS Currents: Tree of Life

with Katy Frederick and Johannes Schul as coauthors

Abstract

Quantifying the age of recent species divergence events can be challenging in the absence of calibration points within many groups. The katydid species *Neoconocephalus lyristes* provides the opportunity to calibrate a post-Pleistocene, taxa specific mutation rate using a known biogeographic event, the Mohawk-Hudson Divide. DNA was extracted from pinned museum specimens of *N. lyristes* from both Midwest and Atlantic populations and the mitochondrial gene COI sequenced using primers designed from extant specimens. Coalescent analyses using both strict and relaxed molecular clock models were performed in BEAST v1.8.2. The assumption of a strict molecular clock could not be rejected in favor of the relaxed clock model as the distribution of the standard deviation of the clock rate strongly abutted zero. The strict molecular clock model resulted in an intraspecific calculated mutation rate of 14.4-17.3 %/myr, a rate substantially higher than the common rates of sequence evolution observed for insect mitochondrial DNA sequences. The rate, however, aligns closely with mutation rates estimated from other taxa with similarly recent lineage divergence times.

Introduction

In recent years, many examples of rapid speciation and diversification occurring during the last glacial cycle (i.e. within 500 kyr BP, Rohling et al. 1998), or even after the last glacial maximum (LGM, 19 kyr BP, Yokoyama et al. 2000) have been described. Arguably, the most impressive examples of rapid diversification are the cichlid radiation events within the African Rift Valley, where a small number of founding species diversified into hundreds of species after the LGM (Meyer et al. 2015, 2006, Nagl et al. 2000). Other examples include the old world pea aphids (Peccoud et al. 2009), North American songbirds (Weir et al. 2008, Milá et al. 2007, Johnson and Cicero 2004), and the threespine sticklebacks of British Columbia (Taylor and McPhail 1999); in some cases, significant diversification arose in as little as 50 years (Lescak et al. 2015).

Estimates of nucleotide evolutionary rates vary greatly dependent with the age of the calibrating point, with younger calibration points resulting in higher rate estimates (Ho et al. 2005, Ho and Lo 2013, Papadopoulou et al. 2010). Fossils and most biogeographic events are ancient (millions of years old) and are appropriate for the dating of similarly ancient events. The few available rate estimates using very young age calibration points (< 200kyrs; Clarke et al. 2001, Gratton et al. 2008, Haag-Liautrard et al. 2008) suggest an exponential increase of estimated rates (Ho et al. 2007, 2008, 2015); additional data is needed to support this pattern. The exponential pattern of estimates is likely an artifact of the estimation methods and does not reflect true differences in rates on nucleotide evolution (Ho et al. 2005). One reason for the small number of estimates for recent lineage divergences is that suitably recent calibration points are scarce (Ho et al. 2008), since these events are too recent to use fossil evidence.

Here we use a postglacial vicariance event to calibrate a lineage specific mutation rate for North American *Neoconocephalus* katydids. At a time following the LGM, water from the North American Great Lakes drained through the Mohawk-Hudson Outlet to the Atlantic coast (Teller 2003). Wetland habitats formed within the Hudson and Mohawk Valleys, which allowed coastal plain species to expand their ranges into the wetlands surrounding the Great Lakes (Reznicek 1994). The opening of the St. Lawrence Seaway (10,750-10,600 ^{14}C yr BP; Rayburn et al. 2005), diverted melt water and led to the drying of the wetlands in the Mohawk-Hudson outlet. This vicariance event left disjunct wetland habitats in the Midwest (mainly bogs and fens) and along the Atlantic Coast (bogs and marsh habitat). Such disjunct ranges matching this pattern are found in plant, reptile, amphibian, and insect species possessing a coastal plain affinity (Reznicek 1994, Thomas 1951, Thomas 1933). *Neoconocephalus lyristes* is an example of such a habitat specialist, limited to bog and fen wetlands. The species' described range follows the pattern of the Mohawk-Hudson Divide, with isolated populations in the Great Lakes area (Thomas 1933) and North Atlantic Coast (Rehn and Hebard 1915; Fig. 1).

The eleven North American *Neoconocephalus* katydid species possess markedly little genetic variation despite their high diversity of species-specific call patterns and may be an example of a recent species radiation (Snyder et al. 2009, Frederick 2013). The accurate timing of this radiation will help identify the evolutionary mechanisms leading to rapid species diversification observed in this group. *Neoconocephalus lyristes* provides a unique opportunity among species of *Neoconocephalus* for the calibration of a post-Pleistocene mutation rate as gene flow between these two disjunct ranges likely ceased with the draining of the Mohawk-Hudson Outlet (10,750-10,600 ^{14}C yr BP, Rayburn et

al. 2005). Here we sequenced mtDNA from museum specimens representing both populations and estimated an intraspecific mutation rate using a coalescent Bayesian method.

Methods

Over three years of searching previous collection sites we found only a single extant population of *N. lyristes*, in Cedar Bog Nature Preserve, Urbane, OH. Due to apparent local extinction of *N. lyristes* from most of its Midwest and its entire Atlantic range, we used museum samples collected in the first half of the 20th century. We selected 18 dried *N. lyristes* specimens, from the Hebard Collection at the Academy of Natural Sciences of Drexel University for DNA extraction and analysis. Specimens represent samples from both Atlantic Coastal and Midwest populations (Fig. 1), with collection dates ranging from 1905-1932. We used a non-destructive method for DNA extraction (modified from Gilbert et al. 2007). A hind leg was removed and placed in a 1.5 ml microcentrifuge tube fully submerged in one ml of digestion buffer: 3 mM CaCl2, 2% sodium dodecyl sulphate (SDS), 40 mM dithiothreitol (DTT), 250 mg/ml proteinase K, 100 mM Tris buffer pH 8 and 100 mM NaCl (quantities represent molarity of final concentrations). Hind legs were incubated overnight (17-19 hrs.) at 55°C. Following digestion we removed the hind legs from buffer and placed them in 100% EtOH for two hours to stop enzymatic activity. Extraction of DNA contained in the buffer was completed using the standard Qiagen DNeasy Blood + Tissue Kit (Qiagen Inc., Valencia, CA, USA) extraction method.

Amplification took place in a laboratory without prior exposure to DNA that could be amplified by primers used in this study. Polymerase chain reaction (PCR) prep was performed in a UV hood. All equipment and surfaces were sanitized with a 10% bleach solution and tools were sanitized in a UV Stratalinker 1800. For this study we designed six overlapping primer pairs (Supplemental Table 1) around non-variable regions of the mitochondrial gene cytochrome oxidase I (COI). These primers were based upon extant *N. lyristes*, *N. robustus*, and *N. bivocatus* COI sequences and designed using the Primer3 (Untergasser et al. 2007) plugin in Geneious v6.0.5 (Kearse et al. 2012). Each primer pair amplified approximately 150 bp; combined, they provide complete coverage of the 743 bp target region.

PCR amplification was performed on an Eppendorf Mastercycler gradient (Eppendorf-Brinkman Instruments Inc., Westbury NY, USA) using Taq DNA polymerase (Platinum Taq, Invitrogen Inc., Carlsbad, CA). All primers were used at a concentration of 10 mM. Thermocycling conditions for all six primer-sets are as follows: Hot start at 94°C 2 min, denaturation at 94°C 30 sec, annealing at 56°C 30 sec, extension 72°C 40 sec, repeated 40x, with a final 72°C extension for 7 min. Amplified PCR products were prepared for sequencing using a ExoI/SAP enzymatic cleanup (2.75 µl 10x SAP buffer, 0.5 µl SAP, 0.25 µl ExoI per 20 µl of PCR product) incubated at 37°C for 30 min, followed by 80°C for 15 min to inactivate enzymes. Sequencing was performed at the DNA Core Facility, University of Missouri, Columbia, MO on an ABI 3730 DNA Analyzer, using standard Big Dye Terminator cycle sequencing chemistry (Applied Biosystems, Foster City, CA, USA). Sequences were edited, aligned and trimmed in Geneious v6.0.5 (Kearse et al. 2012). We used a global alignment with free end gaps and

70% similarity rule. Regions of sequence with high ambiguity were labeled as missing. One individual, with greater than ten percent ambiguity, was removed from the analysis (m017). Individual m007 failed to amplify. We successfully sequenced COI from 16 individuals.

We evaluated substitution models using jModel Test v0.1.1 (Posada 2008) and found GTR+Γ to be a suitable model. Phylogenetic analyses were conducted using a coalescent method as implemented in BEAST v1.8.2 xml (Drummond et al. 2012); input files were formatted using BEAUti v1.7.4 (Drummond et al. 2012). Our analysis assumed a constant population size for the coalescent inferences (Kingman 1982). We performed ten runs with twenty million generations sampled every two thousand trees. This analysis was performed using both a strict, (Drummond et al. 2006) as well as a relaxed molecular clock model (Zuckerkandl and Pauling 1965). The Midwest individuals were run both unconstrained as well as constrained to monophyly. The constrained run assured that the age calibration point was assigned to the correct node in all trees (Papadopoulou et al. 2010). To evaluate the influence of the prior settings on the posterior samples, we repeated the analysis as above but without any sequence data.

Using the radiocarbon date of 10,750±150 ^{14}C yr BP, the end of 150-300 year period of steady melt water flow following the final large flood through the Hudson Valley at 10,900 ^{14}C yr BP (Rayburn et al. 2005), we calibrated the calendar age of the Mohawk-Hudson Divide. We performed the radiocarbon to calendar age conversion using the IntCal13 curve in OxCal v4.2 online (Reimer et al. 2013). The age estimate was fixed to the highest likelihood value within the 95% confidence interval; yielding a calibrated date of 10,739.5 cal BP. Being a known biogeographic barrier we allowed the

node age prior probability of the Midwest clade to vary along a normal distribution, with the calibrated date as the mean age and a standard deviation of one-thousand years. This allows for the possibility of lineage divergence prior to the biogeographic event, as well as the overestimation of the events age (Ho 2007). The Euclidean mean and standard deviation priors were set to exponential with mean values of 10 and 0.3 respectively. Convergence of MCMC runs was visualized using Tracer v1.5 (Rambaut and Drummond 2007). With Tracer v1.5 we ascertained the average mutation rate between populations of *N. lyristes* based on the Mohawk-Hudson calibration. Runs were combined in LogCombiner v1.8.2 (Drummond et al. 2012) and a maximum clade credibility consensus tree was formed in TreeAnnotator v1.7.4 (Rambaut and Drummond 2013).

Results

We successfully sequenced 743 bp of the mitochondrial gene COI from sixteen individuals (5 from Midwest and 11 from Atlantic Coast populations, Table 1). Sequence similarity among the 16 samples ranged from 92.0% to 99.8%. We found the greatest diversity within the Atlantic population. The Midwest clade fell within the larger clade of Atlantic Coast *N. lyristes* (Fig. 2). This observation is congruent with the hypothesized biogeographic history of the species where the Midwest populations diverging from the ancestral Atlantic population.

Using the unconstrained coalescence model, four out of five Midwestern individuals formed a clade within the larger clade of Atlantic Coast *N. lyristes*. One Midwest individual (m010) grouped among Atlantic individuals (Supplementary Fig. 1). In order to prevent the age calibration point from being assigned to the wrong nodes in

some trees we constrained the Midwest clade to monophyly in further analyses (Papadopoulou et al. 2010). The resulting constrained consensus tree (Fig. 2) is congruent with the hypothesized biogeographic history of the species, with the Midwest population diverging from the ancestral Atlantic population.

Using a relaxed clock model, we obtained branch specific mutation rates between 14.4 and 37.5 %/myr from the consensus of the ten runs. The average rate of mutation among branches was 15.8 %/myr, ranging from 15.7-15.9 %/myr between the ten independent runs. The distribution of the standard deviation of the clock rate strongly abutted zero when the relaxed molecular clock was used (Fig. 3). This indicates support for a constant rate of substitution and a strict molecular clock was used (Heath 2015). The strict molecular clock analysis produced a tree (Fig. 2) with a similar, but not identical, topology to the relaxed clock's consensus tree. The relationship between Midwest animals and their relationship to the Atlantic clade remained unchanged, with minor changes in the relationships between Atlantic individuals. The strict consensus tree possessed an average mutation rate of 17.3 %/myr, with mutation rates between the ten runs. Predictably, a slower rate of 14.4 %/myr was obtained when the same analysis was run with individual M010 removed. These two rates, while diverging slightly, both indicate a rate of mutation significantly faster than most reported in the literature [Ho et al. 2005, Ho and Lo 2013].

Discussion

Here, we focused on the calibration of an intraspecific mutation rate at a very recent timescale. Evolutionary rates calibrated across divergent timescales can be

markedly different (Ho et al. 2008), with younger calibration dates (< 1 Mya) showing substantially higher estimates of rates divergence than older lineages (Ho et al. 2005, Ho and Lo 2013). In mammals, for example, the age of the calibration dates shows a negative relationship with estimates for molecular evolutionary rates (Ho et al. 2005, Shapiro et al. 2004). Metastudies utilizing insect mtDNA rates estimated from both inter- and intraspecific calibrations show a similar pattern to that observed in mammals (Papadopoulou et al. 2010, Ho and Lo 2013). Available data suggest an exponential increase of estimated rates (Clarke et al. 2001, Gratton et al. 2008, Haag-Liautard et al. 2008) with decreasing calibration age (Fig. 4). The exponential pattern of estimates is likely an artifact of the estimation methods and does not reflect true differences in rates on nucleotide evolution (Ho et al. 2005).

The sequence variation among populations has two components, fixed substitutions between them that have accumulated since divergence and current within population variation (Ho et al. 2005). The fixed substitutions among lineages represent the actual evolutionary divergence. Most of the within population genetic variation will be removed over time by genetic drift and selection and therefore only a small fraction will ultimately contribute to lineage divergence (Ho et al. 2005, 2007). For young divergence times, the within species variation will contribute a much larger fraction of the total nucleotide differences, as only few fixed substitutions have accumulated. For ancient divergence times, in contrast, the same amount of within population variation would be dwarfed by fixed substitutions accumulated since divergence. Thus, short calibration times should lead to gross overestimations of evolutionary divergence rates,

while ancient calibration times (>1 Mya, Ho et al. 2005) should provide much more realistic estimates.

Insect mtDNA rates of mutation

We estimated a mutation rate for COI at 14.4-17.3 %/myr, using the strict molecular clock model and a very recent calibration time. Our estimate is significantly higher than the commonly assumed mtDNA mutation rate of 1.15 %/myr (Knowles 2000, Brower 1994), which were based on much older divergence times. Estimates of substitution rates calibrated from the age of the Mid-Aegean Trench (9-12 Mya), for example, within an insect model range from 1.0-2.7 %/myr dependent upon application of various substitution and clock models (Papadopoulou et al. 2010). Our estimated rate of 14.4-17.3 %/myr, on the other hand, aligns with estimates found using similarly recent calibration dates. A mutation rate of 19.2 %/myr was estimated for the European butterfly *Parnassius mnemosyne*. As in our analysis, a coalescent model was run in BEAST utilizing a vicariance event calibrated at 10,000 years BP (Gratton et al. 2008). Intermediate calibration dates resulted in an intermediate estimate of evolutionary rates. The mutation rate for the North American ground beetle (*Nebria*) was estimated at 5.7 %/myr, using a vicariance event dated to 150,000 years BP (Clarke et al. 2001). Our estimates for *N. lyristes* fit into the exponential pattern previously described (Fig. 4). Thus, this study agrees with the slower estimates of Orthopteran mtDNA sequence evolution and may serve as an internal calibration point for *Neoconocephalus* diversification.

As more evidence accumulates supporting the occurrence of postglacial species diversification, the greater the need for appropriate tools for timing these events. This will in part include the utilization of young vicariance events for molecular clock calibration. Geologically supported postglacial vicariance events within North America are lacking for many taxa groups (Brooks and McClennan 2012, Avise 2000). The Mohawk-Hudson Divide provides a recent biogeographic vicariance event, with the potential for the calibration of lineage specific mutation rates for a number of plant, amphibian, reptile, and insect groups.

Use of museum samples

The use of ancient DNA (aDNA) samples can be hindered by severe degradation (Gilbert et al. 2007, Andersen and Mills 2012). In this study two of the eighteen samples could not be sequenced successfully. These two samples were not the oldest, nor from the same locality. Severe degradation of DNA, beyond that in the other sixteen samples, or a mismatch in primer binding sites may account for failed amplification (Table 1). In those samples that were sequenced successfully ambiguities were high, while this is likely due to the degraded nature of aDNA, the coamplification of nuclear pseudogenes could also led to such ambiguities. The amplification of relatively short (150 bp) segments increases the likelihood of amplifying pseudogenes, not amplified when targeting longer sequences. Nuclear pseudogenes of COI, while not noted in *Neoconocephalus*, have been found in other Orthopterans (Song et al. 2008). While internal stop codons, common within pseudogenes, were absent from our sequences we cannot rule out the presence of pseudogenes completely.

In this study museum specimens replaced extant samples, necessitated by the rarity, or likely local extinction, of *N. lyristes* from most of its known range. Despite the additional challenges of working with museum specimens, aDNA can replace extant specimens when collection is either not possible because of extinction (Zedane et al. 2016, Rasmussen et al. 2009) or broad resampling is untenable (Dentinger et al. 2016).

With advances in the amplification of ancient DNA (Bakker et a. 2016, Timmermans et al. 2016, Staats et al. 2013), museum collections are also opening up areas of study that are not possible with extant data alone (Heyduk et al. 2016, Zedane 2016, Besnard 2014). Ancient DNA can be utilized in the calibration of molecular clocks through dating tip ages (Debruyne et al. 2008). Samples from multiple time points, can provide additional information about the genetic and demographic changes in groups over time (Fulton et al. 2013). Ancient DNA has been used in the reconstruction and timing of many mammal groups (Shapiro et al. 200 4, Wisely et al. 2004, Fulton et al. 2013, Seguin-Orlando et al. 2013), but remains underutilized in the timing of insect lineages despite the abundance of specimens in museums.

Next generation sequencing is able to vastly improve the sequencing of aDNA. High throughput sequencing has the capability to target short and degraded DNA samples, a major issue in the sequencing of aDNA from museum specimens (Burrell et al. 2015). Using next generation techniques the sequencing of whole genomes from aDNA has been accomplished in plants, animals, and fungi (Dentinger et al. 2016, Der Sarkissian et al. 2015, Staats et al. 2013). Genomes as old as 735 kyr BP have been sequenced from bones of a prehistoric horse species preserved in arctic regions (Seguin-Orlando et al. 2013). Less destructive sampling techniques are also possible because of

the need for smaller amounts of starting material (Wolinsky 2010, Tin et al. 2014, Thomsen et al. 2009). The expanded use of next generation sequencing in recent years, along with the abundance of insect specimens held in museum collections, has and will continue to increase the questions being addressed using aDNA.

References

Andersen JC, Mills NJ. DNA extraction from museum specimens of parasitic Hymenoptera. PLoS One. 2012 Oct 15;7(10):e45549.

Avise, JC. Phylogeography: the history and formation of species. Harvard university press. 2000.

Bakker FT, Lei D, Yu J, Mohammadin S, Wei Z, van de Kerke S, Gravendeel B, Nieuwenhuis M, Staats M, Alquezar-Planas DE, Holmer R. Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. Biological journal of the Linnean Society. 2015 Aug 7;117(1):33-43.

Besnard G, Christin PA, Malé PJ, Lhuillier E, Lauzeral C, Coissac E, Vorontsova MS. From museums to genomics: old herbarium specimens shed light on a C3 to C4 transition. Journal of experimental botany. 2014 Sep 25;65(22):6711-21.

Brooks DR, McLennan DA. The nature of diversity: an evolutionary voyage of discovery. University of Chicago Press; 2012 Apr 26.

Brower AV. Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. Proceedings of the National Academy of Sciences. 1994 Jul 5;91(14):6491-5.

Burrell AS, Disotell TR, Bergey CM. The use of museum specimens with high-throughput DNA sequencers. Journal of human evolution. 2015 Feb 28;79:35-44.

Clarke TE, Levin DB, Kavanaugh DH, Reimchen TE. Rapid evolution in the Nebria gregaria group (Coleoptera: Carabidae) and the paleogeography of the Queen Charlotte Islands. Evolution. 2001;55(7):1408-18.

Debruyne R, Chu G, King CE, Bos K, Kuch M, Schwarz C, Szpak P, Gröcke DR, Matheus P, Zazula G, Guthrie D. Out of America: ancient DNA evidence for a new world origin of late quaternary woolly mammoths. Current Biology. 2008 Sep 9;18(17):1320-6.

Dentinger BT, Gaya E, O'Brien H, Suz LM, Lachlan R, Díaz-Valderrama JR, Koch RA, Aime MC. Tales from the crypt: genome mining from fungarium specimens improves resolution of the mushroom tree of life. Biological Journal of the Linnean Society. 2015 May 20;117(1):11-32.

Dentinger BT, Gaya E, O'Brien H, Suz LM, Lachlan R, Díaz-Valderrama JR, Koch RA, Aime MC. Tales from the crypt: genome mining from fungarium specimens improves resolution of the mushroom tree of life. Biological Journal of the Linnean Society. 2015 May 20;117(1):11-32.

Der Sarkissian C, Allentoft ME, Ávila-Arcos MC, Barnett R, Campos PF, Cappellini E, Ermini L, Fernández R, da Fonseca R, Ginolhac A,Hansen AJ. Ancient genomics. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2015 370:1660:20130387.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS biology. 2006 Mar 14;4(5):e88.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS biology. 2006 Mar 14;4(5):e88.

Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution. 2012 Feb 25;29(8):1969-73.

Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution. 2012 Feb 25;29(8):1969-73.

Frederick KH. *Investigating an adaptive radiation in temperate Neoconocephalus* (Doctoral dissertation, University of Missouri--Columbia).

Fulton TL, Norris RW, Graham RW, Semken HA, Shapiro B. Ancient DNA supports southern survival of Richardson's collared lemming (Dicrostonyx richardsoni) during the last glacial maximum. Molecular ecology. 2013 May 1;22(9):2540-8.

Gaunt MW, Miles MA. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Molecular biology and evolution. 2002 May 1;19(5):748-61.

Gilbert MT, Moore W, Melchior L, Worobey M. DNA extraction from dry museum beetles without conferring external morphological damage. PLoS One. 2007 Mar 7;2(3):e272.

Gratton P, Konopiński MK, Sbordoni V. Pleistocene evolutionary history of the Clouded Apollo (Parnassius mnemosyne): genetic signatures of climate cycles and a 'time-dependent'mitochondrial substitution rate. Molecular Ecology. 2008 Oct 1;17(19):4248-62.

Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD. Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. PLoS biology. 2008 Aug 19;6(8):e204.

Heath TA. Divergence Time Estimation using BEAST v2.

Heyduk K, Trapnell DW, Barrett CF, Leebens-Mack J. Phylogenomic analyses of species relationships in the genus Sabal (Arecaceae) using targeted sequence capture. Biological Journal of the Linnean Society. 2015 May 13;117(1):106-20.

Ho SY, Duchêne S, Molak M, Shapiro B. Time‐dependent estimates of molecular evolutionary rates: evidence and causes. Molecular ecology. 2015 Dec 1;24(24):6007-12.

Ho SY, Larson G. Molecular clocks: when timesare a-changin'. TRENDS in Genetics. 2006 Feb 28;22(2):79-83.

Ho SY, Lo N. The insect molecular clock. Austral Entomology. 2013 May 1;52(2):101-5.

Ho SY, Phillips MJ, Cooper A, Drummond AJ. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Molecular biology and evolution. 2005 Apr 6;22(7):1561-8.

Ho SY, Phillips MJ, Cooper A, Drummond AJ. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Molecular biology and evolution. 2005 Apr 6;22(7):1561-8.

Ho SY, Saarma U, Barnett R, Haile J, Shapiro B. The effect of inappropriate calibration: three case studies in molecular ecology. PLoS One. 2008 Feb 20;3(2):e1615.

Ho SY, Shapiro B, Phillips MJ, Cooper A, Drummond AJ. Evidence for time dependency of molecular rate estimates. Systematic biology. 2007 Jun 1;56(3):515-22.

Ho SY. Calibrating molecular estimates of substitution rates and divergence times in birds. Journal of Avian Biology. 2007 Jul 1;38(4):409-14.

Johnson NK, Cicero C. New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. Evolution. 2004 May;58(5):1122-30.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012 Apr 27;28(12):1647-9.

Kingman JF. On the genealogy of large populations. Journal of Applied Probability. 1982 Jan;19(A):27-43.

Knowles LL. Tests of Pleistocene speciation in montane grasshoppers (genus Melanoplus) from the sky islands of western North America. Evolution. 2000 Aug;54(4):1337-48.

Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics. 2007 Mar 22;23(10):1289-91.

Lescak EA, Bassham SL, Catchen J, Gelmond O, Sherbick ML, von Hippel FA, Cresko WA. Evolution of stickleback in 50 years on earthquake-uplifted islands. Proceedings of the National Academy of Sciences. 2015 Dec 29;112(52):E7204-12.

Mendelson TC, Shaw KL. Sexual behaviour: rapid speciation in an arthropod. Nature. 2005 Jan 27;433(7024):375-6.

Meyer BS, Indermaur A, Ehrensperger X, Egger B, Banyankimbona G, Snoeks J, Salzburger W. Back to Tanganyika: a case of recent trans-species-flock dispersal in East African haplochromine cichlid fishes. Royal Society open science. 2015 Mar 1;2(3):140498.

Milá B, McCormack JE, Castañeda G, Wayne RK, Smith TB. Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus Junco. Proceedings of the Royal Society of London B: Biological Sciences. 2007 Nov 7;274(1626):2653-60.

Müller J, Reisz RR. Four well‐constrained calibration points from the vertebrate fossil record for molecular clock estimates. BioEssays. 2005 Oct 1;27(10):1069-75.

Nagl S, Tichy H, Mayer WE, Takezaki N, Takahata N, Klein J. The origin and age of haplochromine fishes in Lake Victoria, East Africa. Proceedings of the Royal Society of London B: Biological Sciences. 2000 May 22;267(1447):1049-61.

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, Johnson PL. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature. 2013 Jul;499(7456):74-8.

Papadopoulou A, Anastasiou I, Vogler AP. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. Molecular Biology and Evolution. 2010 Feb 18;27(7):1659-72.

Peccoud J, Ollivier A, Plantegenest M, Simon JC. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. Proceedings of the National Academy of Sciences. 2009 May 5;106(18):7495-500.

Penny D. Evolutionary biology: relativity for molecular clocks. Nature. 2005 Jul 14;436(7048):183-4.

Posada D. jModelTest: phylogenetic model averaging. Molecular biology and evolution. 2008 Apr 8;25(7):1253-6.

Rambaut A. Tracer v1. 4. http://beast.bio.ed.ac.uk/Tracer. 2007.

Rambaut, A., and Drummond, A. J. TreeAnnotator v1. 7.0. University of Edinburgh, Institute of Evolutionary Biology. 2013.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature. 2010 Feb 11;463(7282):757-62.

Rayburn JA, Knuepfer PL, Franzi DA. A series of large, Late Wisconsinan meltwater floods through the Champlain and Hudson Valleys, New York State, USA. Quaternary Science Reviews. 2005 Nov 30;24(22):2410-9.

Rehn JA, Hebard M. A synopsis of the species of the genus Neoconocephalus found in North America north of Mexico. Trans. Amer. Ent. Soc. 1915;40(1914):365-413.

Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Ramsey CB, Buck CE, Cheng H, Edwards RL, Friedrich M, Grootes PM. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. Radiocarbon. 2013 Jan;55(4):1869-87.

Reznicek AA. The disjunct coastal plain flora in the Great Lakes region. Biological Conservation. 1994 Jan 1;68(3):203-15.

Rohling, E. J., et al. "Magnitudes of sea-level lowstands of the past 500,000 years." *Nature* 394.6689 (1998): 162-165.

Seehausen O. African cichlid fish: a model system in adaptive radiation research. Proceedings of the Royal Society of London B: Biological Sciences. 2006 Aug 22;273(1597):1987-98.

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MT, Barnes I, Binladen J, Willerslev E. Rise and fall of the Beringian steppe bison. Science. 2004 Nov 26;306(5701):1561-5.

Snyder RL, Frederick-Hudson KH, Schul J. Molecular phylogenetics of the genus Neoconocephalus (Orthoptera, Tettigoniidae) and the evolution of temperate life histories. PloS one. 2009 Sep 25;4(9):e7203.

Song H, Buhay JE, Whiting MF, Crandall KA. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proceedings of the national academy of sciences. 2008 Sep 9;105(36):13486-91.

Staats M, Erkens RH, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PLoS One. 2013 Jul 29;8(7):e69189.

Taylor EB, McPhail JD. Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (Gasterosteus): insights from mitochondrial DNA. Biological Journal of the Linnean Society. 1999 Mar 1;66(3):271-91.

Teller JT. Controls, history, outbursts, and impact of large late-Quaternary proglacial lakes in North America. Developments in Quaternary Sciences. 2003 Dec 31;1:45-61.

Thomas ES. Distribution of Ohio animals. Ohio Journal Science. 1951 51(4):154-167.

Thomas ES. Neoconocephalus Lyristes (Rehn and Hebard) in the Middle West (Tettigoniidae, Orthopt.). Annals of the Entomological Society of America. 1933 Jun 1;26(2):303-8.

Thomsen PF, Elias S, Gilbert MT, Haile J, Munch K, Kuzmina S, Froese DG, Sher A, Holdaway RN, Willerslev E. Non-destructive sampling of ancient insect DNA. PLoS One. 2009 Apr 1;4(4):e5048.

Timmermans MJ, Viberg C, Martin G, Hopkins K, Vogler AP. Rapid assembly of taxonomically validated mitochondrial genomes from historical insect collections. Biological Journal of the Linnean Society. 2015 May 20;117(1):83-95.

Tin MM, Economo EP, Mikheyev AS. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. PloS one. 2014 May 14;9(5):e96793.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3—new capabilities and interfaces. Nucleic acids research. 2012 Jun 21;40(15):e115-.

Walker TJ, Moore TE. Singing insects of north america. University of Florida; 2000.

Weir JT, Schluter D. Calibrating the avian molecular clock. Molecular ecology. 2008 May 1;17(10):2321-8.

Whitesell JJ. *Biology of the United States Coneheaded Katydids of the Genus Neoconocephalus (Orthoptera: Tettigoniidae)* (Doctoral dissertation, University of Florida).

Wisely SM, Maldonado JE, Fleische RC. A technique for sampling ancient DNA that minimizes damage to museum specimens. Conservation Genetics. 2004 Jan 1;5(1):105-7.

Wolinsky H. History in a single hair. EMBO reports. 2010 Jun 1;11(6):427-30.

Yokoyama Y, Lambeck K, De Deckker P, Johnston P. correction: Timing of the Last Glacial Maximum from observed sea-level minima. Nature. 2001 Jul 5;412(6842):99.

Zedane L, Hong-Wa C, Murienne J, Jeziorski C, Baldwin BG, Besnard G. Museomics illuminate the history of an extinct, paleoendemic plant lineage (Hesperelaea, Oleaceae) known from an 1875 collection from Guadalupe Island, Mexico. Biological Journal of the Linnean Society. 2015 Mar 30;117(1):44-57.

Zuckerkandl E, Pauling L. Evolutionary divergence and convergence in proteins. Evolving genes and proteins. 1965;97:97-166.

Zuckerkandl E, Pauling L. Molecular disease, evolution and genetic heterogeneity.

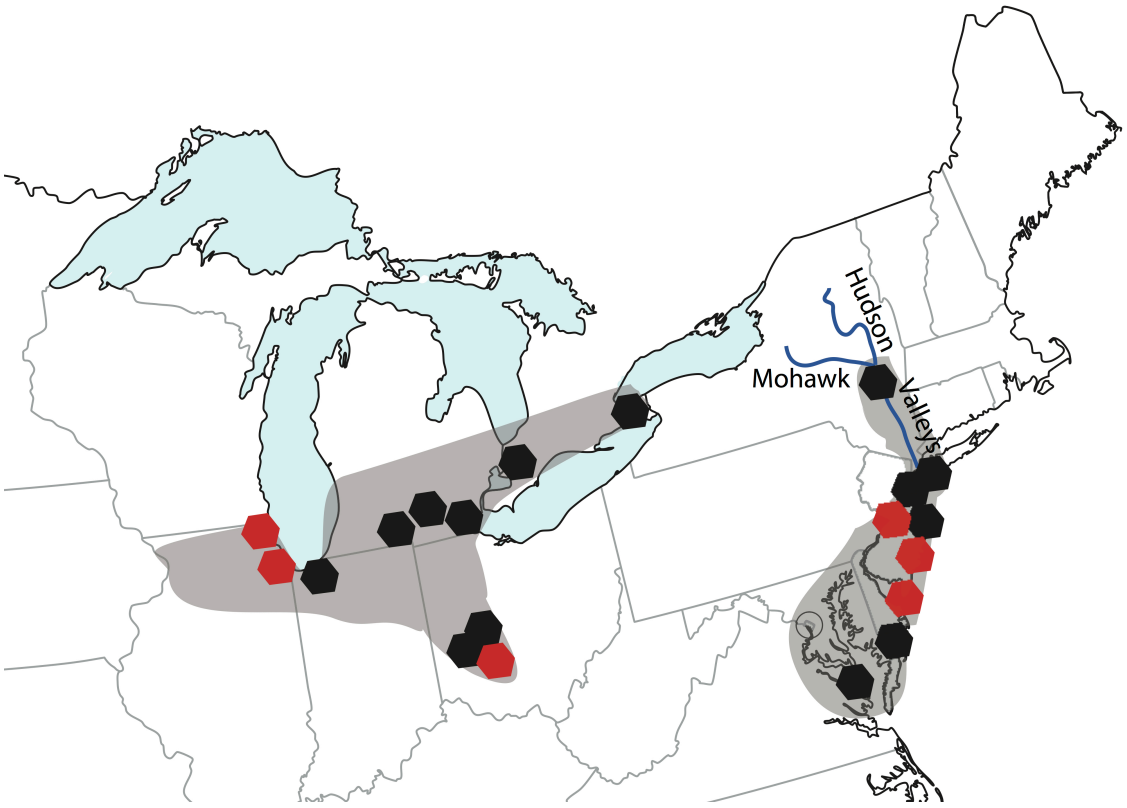


Figure 1: Collection records for *N. lyristes* overlaid with hypothesized range: Hypothesized *N. lyristes* range (modified from Walker 2000, Whitesell 1969) based on literature and collection records. The collection localities of museum samples used in this study are indicated in red.

Table 1: Museum specimen list: *N. lyristes* pinned specimens obtained from the Hebard collection at the Academy of Natural Sciences of Drexel University. Included is all relevant data from specimen label, as well as the ambiguities present in final sequences. (*) denotes samples removed from analysis for failed amplification or excess ambiguity.

Study reference #	Locality	Collection Date	Collected/ID By	Ambiguities (#/743 bp)
m001	Cape May Court House, NJ	1914	Hebard	0
m002	Cape May Court House, NJ	1914	Hebard	1
m003	Cape May Court House, NJ	1914	Hebard	0
m004	Cape May Court House, NJ	1914	Hebard	1
m005	Cape May Court House, NJ	1914	Hebard	0
m006	Cape May Court House, NJ	1914	Hebard	24
m007	Cedar Swamp, OH	1929	Unknown	N/A*
m008	Cedar Swamp, OH	1932	Edward S. Thomas	25
m009	Cedar Swamp, OH	1932	Edward S. Thomas	0
m010	Chicago, IL (Beach IL)	1906	Unknown	2
m011	Chicago, IL (S. of Jackson Park)	1905	Unknown	18
m012	Chicago, IL (S. of Jackson Park)	1905	Unknown	15
m013	Whitesbog, NJ	1923	Det. D.C. Rentz (1974)	0
m014	Whitesbog, NJ	1923	H. Fox	0
m015	Whitesbog, NJ	1923	Unknown	0
m016	Whitesbog, NJ	1923	H. Fox	0
m017	Whitesbog, NJ	1923	Unknown	103*
m018	Whitesbog, NJ	1923	Unknown	1

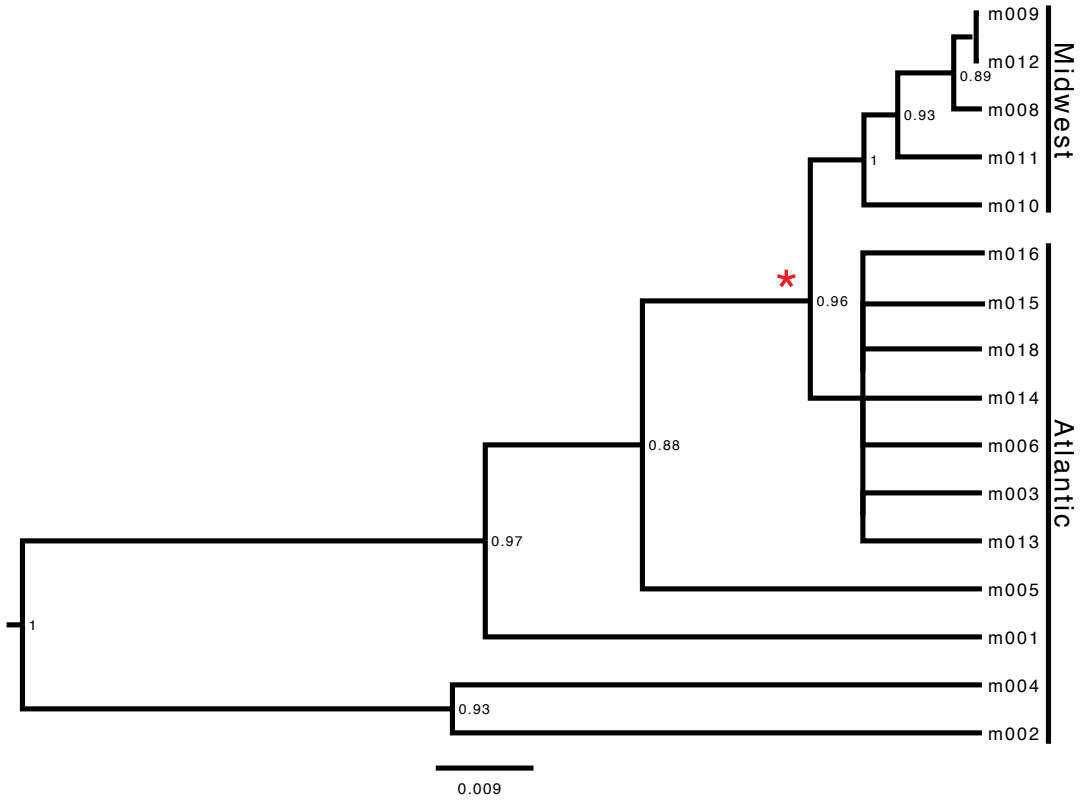


Figure 2: Consensus tree from coalescence analysis using a strict molecular clock model and Midwest clade constrained to monophyly: Nodes possessing < 0.85 posterior probabilities were collapsed. Red star represents the Mohawk-Hudson Divide, with the prior of the node age set to a normal distribution with a mean age of 10,739.5 cal BP. The Midwest specimen m010 fell outside of the Midwest clade prior to constraining the group to monophyly.

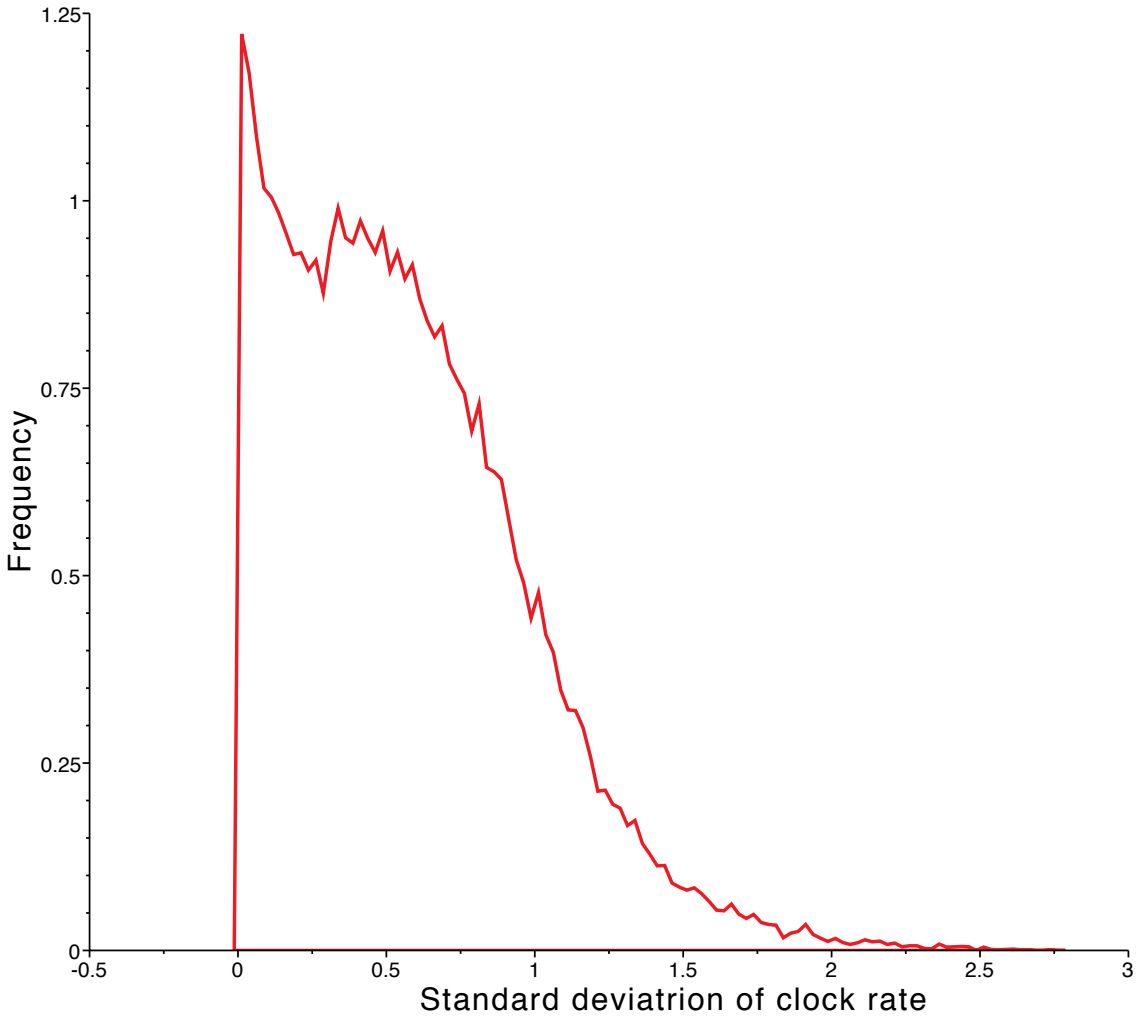


Figure 3: Distribution of the standard deviation rates from relaxed clock analysis: Includes data from ten combined runs (twenty million generations sampled every two thousand trees) using a relaxed molecular clock model. Units for the clock rate are in substitutions per site per million years. The distribution strongly abuts zero, indicating support for a strict molecular clock (Heath 2015).

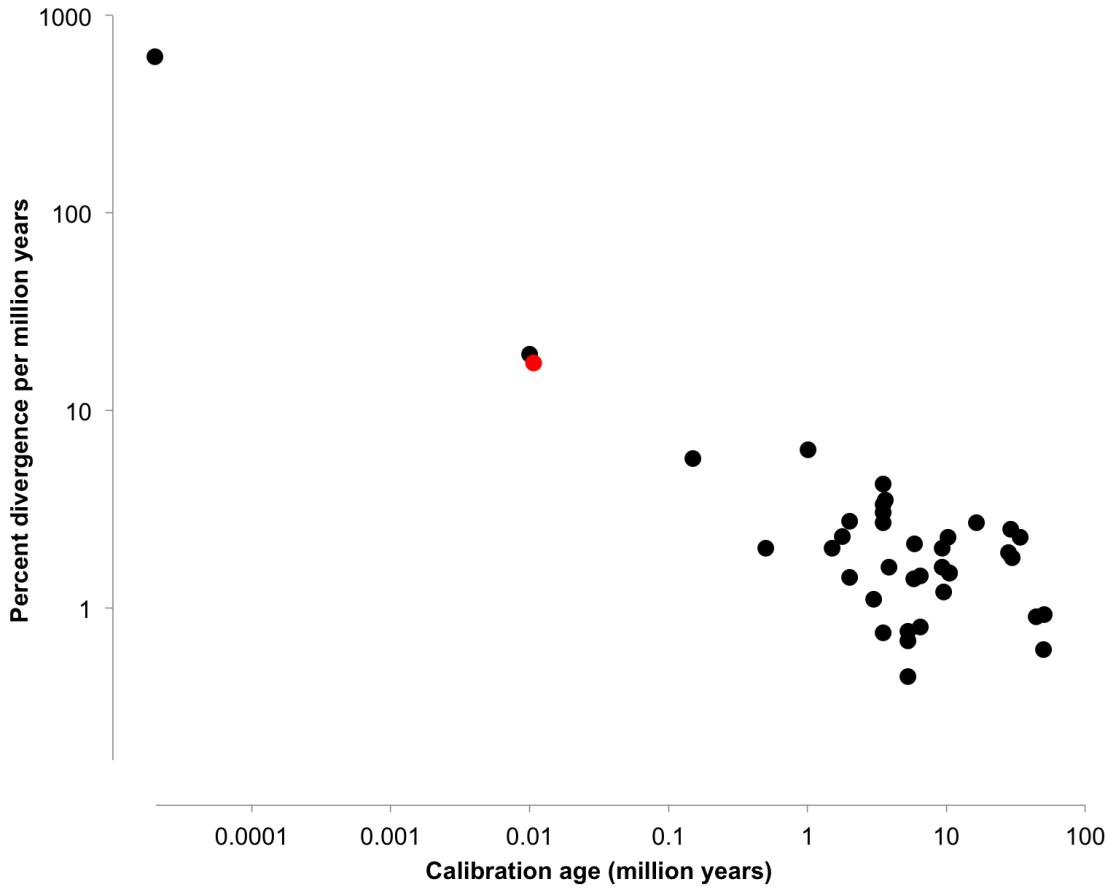
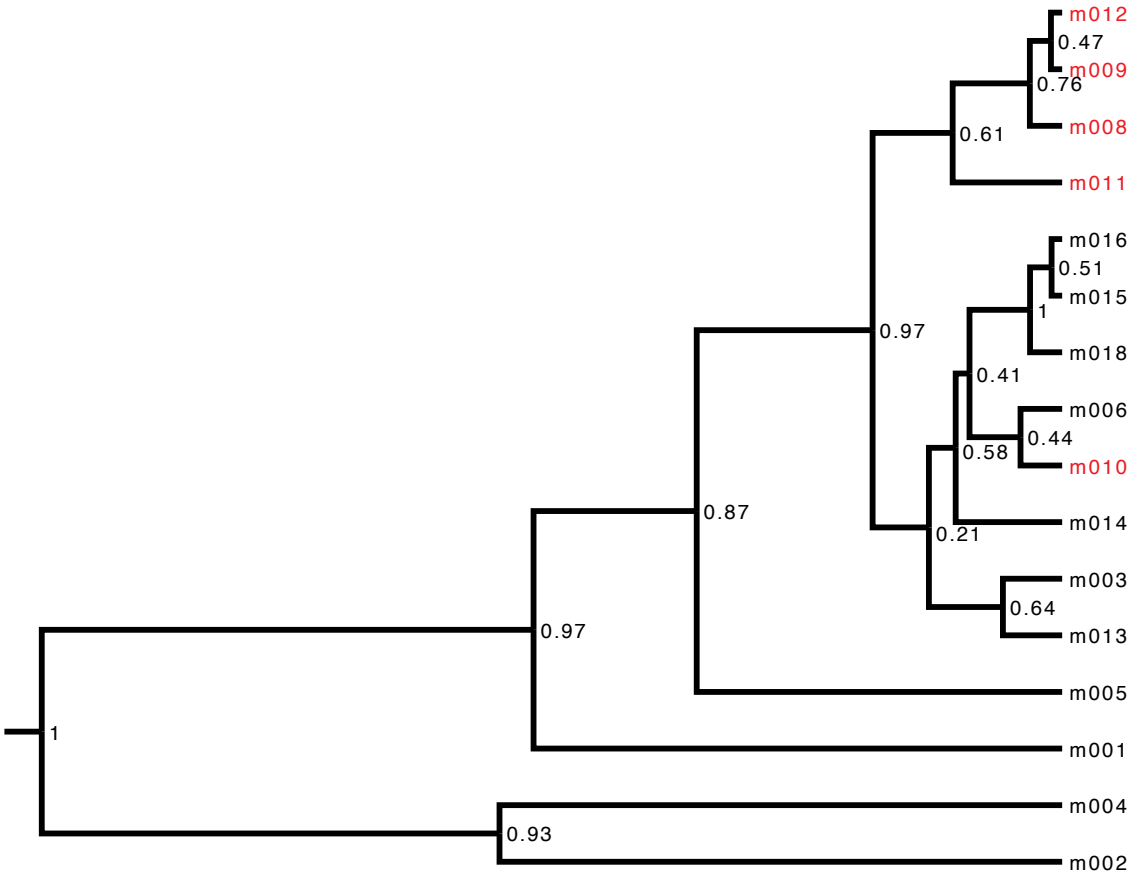


Figure 4: Estimates of evolutionary rates (%/myr) plotted against calibration age (myr): The black data points were obtained directly from Papadopoulou et al (2010) and Ho and Lo (2013). The red point represents the mutation rate estimate from this study. Note both axes are in log scale.

Supplemental Table 1: Table of primers designed for amplification of *N. lyristes* COI sequences: Primers were designed from reference sequences of extant *N. lyristes, N. bivocatus,* and *N. robustus*.

Primer name	Primer sequence
lyF68 (forward)	5'-GGA ATT GCA CAT GCT GGA GC-3'
lyR197 (reverse)	5'-GTG ATA TTC CTG GGG CAC GT-3'
lyF187 (forward)	5'-ACG TGC CCC AGG AAT ATC AC-3'
lyR336 (reverse)	5'-CCG GCA GGA TCA AAG AAT GA-3'
lyF317 (forward)	5'-TCA TTC TTT GAT CCT GCC GGA-3'
lyR466 (reverse)	5'-GGC TTC CTT TTT CCC ACT TTC T-3'
lyF440 (forward)	5'-AGT CAA GAA AGT GGR AAA AAG GA-3'
lyR589 (reverse)	5'-AGC TGA AGT AAA ATA RGC TCG TG-3'
lyF545 (forward)	5'-ACA GTA GGA ATG GAT GTT GAT ACA C-3'
lyR694 (reverse)	5'-GCC TAG AGC TCA TAA AAG GGA AG-3'
lyF666 (forward)	5'-ACA GTC CTT CCC TTT TAT GAG CT-3'
lyR811 (reverse)	5'-AGA TAG AAC ATA ATG GAA ATG GGC T-3'



Supplemental Figure 1: Consensus tree using a strict molecular clock and the Midwest clade unconstrained: Node values represent posterior probabilities calculated from eighteen million total trees. Red taxa represent Midwest samples and black taxa Atlantic samples.

VITA

Gideon Ney was born October 28, 1986 in Wichita, Kansas. Growing up in Wichita, Kansas and Honolulu, Hawaii, he graduated Wichita East High School in May 2005. From there, Gideon attended The University of Missouri in Columbia, Missouri. He graduated in May 2009 with a B.A. in biological sciences and a B.S. in secondary education. Following graduation Gideon accepted a position as a high school science instructor for the Wichita Public Schools, where he taught anatomy and physiology and general biology for two years. In 2011 he returned to The University of Missouri in Columbia, Missouri where he worked under the advisement of Dr. Johannes Schul. He earned his Ph.D. in biological sciences in July 2017. Gideon has accepted a faculty teaching position at Kirkwood Community College in Cedar Rapids, Iowa to start in August 2017.