Final Report

RFP No. 207c for Endangered Species Research Projects for Freshwater Mussels, Region 2, East Texas: Extension for Pigtoe Genomic Research

John S. Placyk, Jr., Ava Laszlo, Ashley Broadbent, Kate L. Hertweck, Lance R. Williams, Marsha B. Williams, and Joshua A. Banta

Department of Biology,
The University of Texas at Tyler

Submitted to Colin McDonald and Melissa Salmon, Texas Comptroller of Public Accounts
Sunday, June 30th, 2019

Introduction

Freshwater mussels (Unionidae) have experienced dramatic declines in both abundance and distribution throughout North America over the last century, due in part to their extreme sensitivity to pollution and other environmental changes. Of the 297 currently recognized species, 12% may now be extinct, 23% are listed as threatened or endangered, and ~43% are in need of conservation status (Williams et al. 1993; Neves et al. 1997; Vaughn 1997). Texas is home to 52 of the 297 species (Howells et al. 1996). Eleven of those 52 are currently under review by the U.S. Fish and Wildlife Service to determine if federal threatened or endangered listings are warranted. About half of those 11 occur only in Texas. The demise of many species has been linked to a lack of critical information on life-histories, habitat use, and taxonomic confusion (Warren Jr. et al. 2000). While much of our previous work (Bertram et al. 2017; Marshall et al. 2018) focused on elucidating the life-histories of Texas freshwater mussels by determining likely fish hosts for glochidia or assessing habitat use via ecological niche modeling (e.g., Walters et al. 2017), one question that remains regards the taxonomic relationship between the Texas Pigtoe (*Fusconaia askewi*) and the Critically Imperiled Triangle Pigtoe (*F. lananensis*) (G1S1), both of which are listed as threatened in Texas. This work resolves the relationship between these two species using a large genomic dataset, which is then used to make ecological niche models based on the new data. This is important because the habitat needs of the aggregated species complex may be different from the habitat needs of the two groups when they are parsed out into different species (Walters et al. 2017).

The Texas Pigtoe is a big river species endemic to east Texas rivers and is abundant in

the Sabine and Neches Rivers. Morphologically, it is very difficult to distinguish from the Triangle Pigtoe, which overlaps its distribution in the southern part of its range. Specifically, the Triangle Pigtoe is endemic to the Neches drainage basin of east Texas and has been recorded in the Angelina River, Attoyac Bayou and southern tributaries of the Neches such as Village Creek (Vidrine 1990; Howells 2006; Karatayev and Burlakova 2007). Given the morphological similarities between these two pigtoes, Burlakova et al. (2012) and Pieri et al. (2018) explored their genetic similarities in an attempt to determine if the they are indeed separate species. They both found low enough interspecific variation that they recommended folding the two into one single species; however, their recommendations is based on only two mitochondrial genes and part of a nuclear gene, which represent a very small fraction of the genomes of these species, and may show divergence patterns incongruent with other markers (e.g., nuclear sequences, AFLPs, SNPs, microsatellites, etc.). Further, even if we were to accept their genetic analyses, there is still much morphological and ecological data that support the current classification of the two as the separate species they are currently considered (Howells et al. 1996; Ford 2013). Therefore, the debate as to their taxonomic status lingers. The goal of this project was to provide a thorough sampling of genomic data to resolve the taxonomic relationship between the Texas and Triangle Pigtoe. We then created an ecological niche model informed by the genomics. To accomplish this task, we used the company SNPSaurus[1] to perform Nextera-tagmented, reductively-amplified DNA (nextRAD) genotyping (Russello et al. 2015), which allowed us to obtain genome-wide single-nucleotide DNA polymorphisms (SNPs) for each of our specimens. Genome-wide SNPs provide more reliable phylogenetic and population genetic information than one or a few gene sequences (Brito and Edwards 2009). This allows large numbers of genetic

[1] https://www.snpsaurus.com/

variations such as single nucleotide polymorphisms (SNPs) to be readily identified (Miller et al. 2007; Baird et al. 2008; Davey et al. 2011). Because nextRAD markers are not based on any single individual genes, but rather are an aggregation of loci genome-wide, they provide an unbiased perspective on the overall patterns of relatedness of individuals to one another (Steane et al. 2011) and can therefore be key to resolving taxonomic questions (Rubin et al. 2012). In fact, very closely related taxa such as the Triangle Pigtoe and Texas Pigtoe are perfect for a nextRAD phylogenetic approach, which has been shown to be reliable in such circumstances (Rubin et al. 2012).

Our project addressed the following three questions: (1) What is the relationship between Texas Pigtoes and Triangle Pigtoes? Are they separate species/subspecies or a single species? If they are separate species/subspecies, are there misidentified specimens revealed by the genome-wide taxonomic analysis that can now be properly classified? (2) What are the habitat requirements for these taxa, in light of the information about their taxonomic statuses, when their habitat preferences/tolerances are modeled in accordance with this new information? (3) Does habitat suitability accurately predict the presence/absence of these mussels?

Our project had three specific objectives, which match the research questions listed above. They are: (1) Resolve the taxonomy of the state threatened Texas Pigtoe (*Fusconaia askewi*) and the state threatened and critically imperiled Triangle Pigtoe (*F. lananensis*); (2) Refine niche models based on phylogenomic data; (3) Verify whether the mussels are found at locations where the habitat is predicted to be the most suitable and not found at less suitable locations.

Materials and Methods

Genomics. Our collections for genomic analysis were from throughout East Texas, from the Sulphur River, Big Cypress Creek, Sabine River, Neches/Angelina Rivers, Trinity River, and San Jacinto River watersheds (Figure 1). The collection consisted of 27 Texas Pigtoe and 13 Triangle Pigtoe specimens, as well as eight inconclusive specimens that were either Texas Pigtoe, Triangle Pigtoe, or Texas/Triangle Pigtoe hybrids (Table 1). In order to understand the Texas-Triangle Pigtoe relationship in a broader context, the collection also included 14 Trinity Pigtoe (*Fusconaia chunii*), seven Wabash Pigtoe (*Fusconaia flava*), and three Louisiana Pigtoe (*Pleurobema riddellii*) specimens.

We identified thousands of single nucleotide polymorphism (SNP) markers genome-wide for each specimen. Genome-wide SNPs provide more reliable phylogenetic and population genetic information than one or a few gene sequences (Brito and Edwards 2009). We extracted DNA from the specimens using illustra™ tissue & cells genomicPrep mini spin kits. Then we used the company SNPSaurus[2] to perform Nextera-tagmented, reductively-amplified DNA (nextRAD) genotyping (Russello et al. 2015), which allowed us to obtain genome-wide single-nucleotide DNA polymorphisms (SNPs) for each of our specimens. The SNPSaurus methods were as follows. Genomic DNA was converted into nextRAD genotyping-by-sequencing libraries as in Russello et al. (2015). Genomic DNA was first fragmented with Nextera reagent (Illumina, Inc), which also ligates short adapter sequences to the ends of the fragments. The Nextera reaction was scaled for fragmenting 15 ng of genomic DNA. Fragmented DNA was then amplified for 26 cycles at 73 degrees, with one of the primers matching the adapter and extending 9 nucleotides into the genomic DNA with the selective sequence GTGTAGAGC.

[2] https://www.snpsaurus.com/

Thus, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer will be efficiently amplified. The nextRAD libraries were sequenced on a HiSeq 4000 with one lane of 150 bp reads (University of Oregon). The genotyping analysis used custom scripts (SNPsaurus, LLC) that trimmed the reads using bbduk (BBMap tools, http://sourceforge.net/projects/bbmap/)[3]. Next, a de novo reference was created by collecting 10 million reads in total, evenly from the samples, and excluding reads that had counts fewer than 8 or more than 1000. The remaining loci were then aligned to each other to identify allelic loci and collapse allelic haplotypes to a single representative. All reads were mapped to the reference with an alignment identity threshold of 90% using bbmap (BBMap tools). Genotype calling was done using Samtools[4] and bcftools[5]. The vcf was filtered to remove alleles with a population frequency of less than 15%. Loci were removed that were heterozygous in all samples or had more than 2 alleles in a sample (suggesting collapsed paralogs). The resulting dataset had 10,500 genome-wide SNPs for each specimen, although we thinned it down one SNP per nextRAD fragment (for a total of 483 SNPs) in order to reduce redundant genomic information (Takeuchi et al. 2005).

We used two approaches for phylogenetic analysis of our specimens. First, we used a straightforward, "total evidence" approach by concatenating loci into a single matrix and inferring a maximum likelihood phylogeny using RAxML (Stamatakis 2014). This approach provided a single tree that averages the phylogenetic signal from across the genome, but may obscure more nuanced variation from individual loci, such as might arise from evolutionary phenomena like incomplete lineage sorting. Second, we used the program Bayesian Evolutionary

[3] The settings were: bbmap/bbduk.sh in=$file out=$outfile ktrim=r k=17 hdist=1 mink=8 ref=bbmap/resources/nextera.fa.gz minlen=100 ow=t qtrim=r trimq=10
[4] The settings were: samtools mpileup -gu -Q 12 -t DP,DPR -f ref.fasta -b samples.txt
[5] The settings were: bcftools call -cv - > genotypes.vcf

Analysis by Sampling Trees (BEAST) (Bryant et al. 2012), with the add-on package SNP and AFLP Package for Phylogenetic analysis (SNAPP) (Bryant et al. 2012) to infer species trees and species demographics from independent (unlinked) biallelic markers such as well spaced SNPs (Bryant et al. 2012). This program implements a full coalescent model, but uses a novel algorithm to integrate over all possible gene trees, rather than sampling them explicitly. Following Yoder et al. (2013), we analyzed our SNP data using a multispecies coalescent approach in SNAPP version 1.3.0 within BEAST2 v2.3.2. The analysis used the GTR model of evolution and proceeded for 10,000,000 generations with 1,000,000 (10%) discarded as burnin. The full SNP data were converted to a 0, 1, 2 format for analysis, with 1 representing a heterozygous genotype. Once the program completed, the results were analyzed in Tracer (Drummond and Rambaut 2007) for performance and accuracy (not shown).

In addition to phylogenetics, we used the program STRUCTURE (Pritchard et al. 2000; Porras-Hurtado et al. 2013) to estimate the admixture (in other words, the hybrid ancestry) of each specimen. Because STRUCTURE does not require a priori placement of taxa into groups, this program is appropriate for uncovering historic patterns of gene flow across the genome. We determined the optimal number of inferred ancestral groups, or clusters, by trying different numbers of clusters, ranging from 1-20. We found that the increase in the likelihood (i.e., the lowest log-likelihood; Sokal and Rohlf 1995) levelled off at four groups (data not shown). STRUCTURE uses the chosen number of clusters, in this case four, to assign "membership coefficients" to each specimen. The coefficients correspond to the proportion of the genome contributed to that specimen by each inferred ancestral group, or cluster, and they sum to one.

Ecological Niche Modeling. Next we created a habitat suitability map (also called ecological niche model) for Texas Pigtoe. We did not create an ecological niche model for

Triangle Pigtoe, because the genomics results told us that it is synonymous with Texas Pigtoe (see results below). Therefore, we modeled them together as one species, which we termed *F. askewi* (Texas Pigtoe). We used the Maxent method for ecological niche modeling, which uses a general-purpose machine learning method that estimates the probability of a species distribution by finding the probability of a distribution that is closest to uniform and then altering one environmental variable at a time repeatedly to maximize the likelihood of the occurrence dataset (Hernandez et al. 2006; Phillips et al. 2006). Maxent produces a heat map that visualizes a fitted cloglog link function relating the environmental data to the habitat suitability of every parcel of the landscape (at the grain size of the environmental data) (Phillips 2017). The habitat suitability scores range on a scale from zero (most unsuitable) to one (most suitable).

This study focused on the associations of environmental factors with mussel distributions at a relatively fine-grained resolution (100 m2 x 100 m2 resolution), where local habitat parameters including water velocity, depth, and substrate type are commonly thought to influence mussel abundance and distribution (Vannote and Minshall, 1982; Holland-Bartels, 1990; Strayer and Ralley, 1993; Strayer et al., 1994). The analysis was restricted primarily to the Trinity, Cypress, Sulphur, Sabine, Neches, and Angelina rivers and their associated watersheds (Figure 2). We used only those Texas Pigtoe specimens that were less than 10% admixed (hybridized) with other species, as inferred from the STRUCTURE analysis. This is so we would model "pure" representatives of the species and not hybrids to any appreciable degree. To minimize autocorrelation at 1km, we used the 'thin' function of the package spThin (Aiello-Lammens et al. 2015) in R version 3.5.2 (R Core Team 2019).

We modeled the suitability of habitat for the Texas Pigtoe as a function of continuous soil variables from the SSURGO database (United States Department of Agriculture et al. 1995),

continuous hydrology variables from the NHDFlowline database (US Environmental Protection Agency (USEPA) and US Geological Survey (USGS) 2005), and continuous climate variables from the WorldClim database (Hijmans et al. 2005). We also included geology characteristics (Horton 2017), landform (Hammond 1964), and vegetation type (U.S. Geological Survey Gap Analysis Program 2016) as categorical variables. Finally, we included the distributions of the two confirmed fish hosts for Texas Pigtoe, the Blacktail Shiner (*Cyprinella venusta*) and the Red Shiner (*Cyprinella lutrensis*) (Bertram et al. 2017) as continuous variables. The full list of environmental variables is in the supplementary data.

We followed the approach outlined in Walters et al. (2017), wherein the layers are clipped to constrain them to a 100-meter distance around rivers and streams, delineated by the NHDFlowline dataset (US Environmental Protection Agency (USEPA) and US Geological Survey (USGS) 2005). The resolution of the models was 100m x 100m. We ended up with nine distinct GPS locations of genomically confirmed specimens of Texas Pigtoe (not hybrids) (Figure 2) (Table 2); two sites had to be thrown out because they lay outside of the lotic buffer for which we had environmental data (not shown). All layers and data points were converted to the Universal Transverse Mercator coordinate system (Buchroithner and Pfahlbusch 2017) using the projection for zone 15 North and the datum North American Datum 1983 (Junkins and Garrard 2006).

In order to include the distributions of the Blacktail Shiner and the Red Shiner in our niche model for the Texas Pigtoe, we first performed ecological niche modeling on the individual fish species and used the resulting habitat suitability maps of each species as layers in the Texas Pigtoe model. Therefore, the ecological niche model for the Texas Pigtoe contains two more environmental layers than the ecological niche models for the Blacktail Shiner and Red

Shiner (namely, the habitat suitability maps from the ecological niche models of the two fish species). Otherwise the modeling procedures for the fish and the Texas Pigtoe were generally the same. The locations used for ecological niche modeling of the fish species came from the Fishes of Texas database (Hendrickson and Cohen 2015) and included only those locations since 1990 that fell within our study area. We thinned the fish locations to be at least 1 km apart from one another as described above. The locations of the fish points used for modeling are included in the supplementary data as shape file packages ("red shiner.zip" and "blacktail shiner.zip").

To avoid model overfitting and other problems, we included only continuous environmental layers that had Pearson correlation coefficients of |0.60| or less (see the supplementary data). We performed the correlation analysis using ENMTools version 1.4.4 (Warren et al. 2010). This left us with 25 continuous variables, and three categorical variables, that we used for modeling both the fish and Texas Pigtoe. We also added the two fish distribution layers derived from this modeling approach to the model for Texas Pigtoe. Thus there were 28 layers used for modeling the fish species, and 30 layers used for modeling Texas Pigtoe (i.e., the Texas Pigtoe model used the same 28 layers that were used for the fish models, plus two fish distribution layers derived from the two fish models) (see the supplementary data).

Models were validated using the test AUC, or the area under the operator receiving curve. AUC measures the probability that a randomly chosen presence site will be ranked above a randomly chosen pseudoabsence site (Pearson et al. 2007). The test AUCs represent the average percentage of the pseudoabsence data with lower habitat suitability scores than "test" presence locations left out of the model building process for each model fold. Importantly, this model validation procedure is based on data points (test data) that were naïve to the model building process for each model fold.

The final models were confirmed with cross-validation, where the data are split into k independent subsets, and for each subset, the model is trained with $k - 1$ subsets and evaluated on the kth subset (Merow et al. 2013). For the fish species, which had very large numbers of occurrences, we used $k = 10$ equal-sized model folds, where all samples are included in the training and test sets at least once, iteratively. Model statistics were then averaged across the k folds. For Texas Pigtoe, the number of genomically-informed presence locations was small (only nine). Therefore, we used a "leave-one-out" or "n-1" cross-validation method, as previously described by Pearson et al. (Pearson et al. 2007), which is appropriate for small sample sizes. Specifically, we set k to equal the number of samples, so that each of the k folds contained n -1 observations, where n is the total sample size. This means that each fold only had a single test data point, and that each observation was the test data point, in turn, for a separate fold. Model statistics were then averaged across the k folds.

To quantify the relative importance of the individual environmental variables to the models, the fit of each full model was compared to reduced univariate models (Phillips and Dudik 2008). If an environmental variable accounted for a substantial portion of the model fit when modeled by itself (as compared to the full model that was based on all the environmental variables), then the environmental variable was considered important in determining the varying habitat suitability of the landscape for that model (Phillips and Dudik 2008).

Model fit was measured with the gain statistic. Gain is a likelihood (deviance) statistic that measures the model performance compared to a model that assigns equal habitat suitabilities to all areas of the landscape. Taking the exponent of the final gain gives the (mean) probability of the presence sample(s) compared to the pseudoabsences. For instance, a gain of 3 means that an average presence location has a habitat suitability of $e^3 = 20.1$ times higher than an average

pseudoabsence site. Test gains should be positive; a negative test gain means that the variable modeled by itself does a worse job at predicting species occurrence points than a uniform distribution (i.e., one that assigns every pixel the same habitat suitability score).

Ground-truthing. We verified that the genomically-informed habitat suitability map for Texas Pigtoe accurately assigns higher habitat suitabilities to areas where Texas Pigtoe is found than to areas where it is not found. Importantly, this effort was based upon independent data that was newly collected and was not used in the ecological niche modeling effort. This means that different data was used to make the habitat suitability map than was used to verify it. The new sites were never visited before as part of any other modeling efforts. We visited a total of 25 sites (Figure 3), which included ones the habitat suitability map deemed highly suitable as well as ones deemed less suitable, such that we had a diversity of sites with different levels of habitat suitability represented in our ground-truthing efforts (Table 3).

We analyzed the ground-truth data using two approaches: *t*-tests comparing the habitat suitability of areas where the Pigtoes were found to the areas where the Pigtoes were not found (Sokal and Rohlf 1995), and logistic regression of the Pigtoes' presence/absence as a function of habitat suitability (Cox 1958). The habitat suitabilities were extracted from the habitat suitability map for Texas Pigtoe. The *t*-test provided a *P*-value for the hypothesis that areas where the Pigtoes are found have higher habitat suitabilities than areas where they are not found. The logistic regression provided a *P*-value for the hypothesis that areas with higher habitat suitabilities are more likely to have Pigtoes than areas with lower habitat suitabilities.

Results

Genomics. The maximum likelihood and Bayesian coalescence-based phylogenies yielded different topologies and had many unresolved nodes (Figures 4 and 5, respectively). This means that the phylogenetic analyses did not yield clear pictures of the relationships among species. This is not surprising, however, given that there is evidence of hybridization among them (see below). Phylogenetic software does not handle hybridization and heterozygosity well (Lischer et al. 2014). But whereas the genomic results were inconclusive, the results based on Bayesian cluster analysis, which are more appropriate in this circumstance, were more definitive.

The Bayesian cluster results (STRUCTURE) show that Texas and Triangle Pigtoe are not differentiated genomically (Figure 6, top). They cluster together as the same species, which is indicated by the color red. The Trinity Pigtoe specimens cluster together as indicated by the color yellow, the Wabash Pigoes cluster together as indicated by the color blue, and the Louisiana Pigtoe specimens cluster together as indicated by the color green. The results suggest that there is hybridization among all of the species with one another, as indicated by the diversity of ancestors (colors) comprising many of the individuals (columns in Figure 6; Table 4). Other noteworthy results are as follows:

- Louisiana Pigtoe: the analysis reveals admixture between Louisiana Pigtoe and the other mussels analyzed. Traces of its ancestry (represented by the green color) can be seen in several specimens collected from Neches and Angelina Rivers (Figure 6 bottom), and conversely Texas and Trinity Pigtoe ancestries can be seen in one of the Louisana Pigtoe specimens (Figure 6; Table 4).
- Trinity Pigtoe: the analysis reveals admixture between Trinity Pigtoe and the other mussels analyzed. High amounts of Wabash Pigtoe and Texas Pigtoe ancestry are found in several Trinity Pigtoe specimens. Specifically, high amounts of Wabash Pigtoe

ancestry are found in Trinity Pigtoe specimens in the Trinity and San Jacinto watersheds; high amounts of Texas Pigtoe ancestry are found in Trinity Pigtoe specimens in the Trinity watershed (Figure 6 bottom; Table 4).

- Specimens from the San Jacinto watershed were hybrids of Trinity Pigtoe and Wabash Pigtoe, rather than just Texas Pigtoe specimens as we had assumed (Figure 6, Table 4). Trinity Pigtoe has not been identified in the San Jacinto watershed before to our knowledge.

Ecological niche modeling. The ecological niche modeling results for the fish species are available in the supplementary data. We will focus here on the ecological niche modeling results for Texas Pigtoe. We found that the test gain for the model with only gypsumcombinedraster had a highly negative test gain (see the supplementary data), indicating that it was dragging down the model fit. Therefore we removed this layer and re-ran the model and that is what we report here.

The average test AUC for the full model was 0.87, indicating that the model is useful (Elith 2002). The model agrees with our previous research (Walters et al. 2017), showing that the most suitable Texas Pigtoe habitat is in the upper Sabine, upper Neches, and Angelina Rivers (Figure 7). The most important variables contributing to the test gain of the full model were: geologic type, soil erodibility, base flow, Blacktail shiner distribution, landform, stream velocity, soil calcium carbonate concentration, and stream volumetric flow rate (Table 5). Response curves are available in the supplementary data. The geologic types associated with Texas Pigtoe habitat suitability were sand and alluvium. Along with this, their habitat suitability increased with increasing erodibility of the surrounding soil. Their habitat suitability increased as the percentage of the stream's total volume comprised of groundwater increased relative to

precipitation runoff. They were associated with locations that were more suitable for the Blacktail shiner, one of their hosts. They were associated with one particular type of landform: one particular type of irregular plains, where 50-75% of gentle slope is in upland. They are associated with higher stream velocities, lower soil calcium carbonate levels, and higher stream volumetric flow rates corresponding to larger rivers.

Ground-truthing. On average, the ground truth locations where Texas Pigtoe were found had higher habitat suitability than the ground truth locations where it was not found (Figure 8). The trend was suggestive but was not statistically significant, either by the *t*-test *P*-value or by the logistic regression *P*-value.

Discussion

Genomics. In summary, the results presented here show that Texas and Triangle Pigtoe are synonymous. This has been suggested by other researchers (Burlakova et al. 2012; Pieri et al. 2018), but the prior research was inferred from the analysis of one or a few genes, which can be inconclusive or misleading when taxa are closely related (Pamilo and Nei 1988; Maddison 1997; Shaw 2002; Brito and Edwards 2009; Wiens et al. 2010; Swenson and El-Mabrouk 2012; Mitchell and Gonder 2013). In other words, while phylogenies based on one or a few genes can tell genera apart, they cannot necessarily tell species apart. Phylogenies based on one or a few genes may inappropriately lump species together in ways that do not reflect their true ancestries. Thus our study with genome-wide data is more definitive, and we believe it is now warranted to collapse the two species into one. The Bayesian cluster analyses in particular demonstrated that the two entities are synonymous. They did not have different ancestral groups that distinguish

them, although the other species did have different ancestral groups that distinguish them from Texas/Triangle Pigtoe. We will refer solely to Texas Pigtoe for the rest of this report, and it will include both Texas Pigtoe and Triangle Pigtoe specimens.

The Bayesian cluster analysis also revealed evidence of hybridization among species. Hybridization is a risk factor for extinction of a rare species, because its distinct gene pool may become lost (Rhymer and Simberloff 1996; Fitzpatrick et al. 2015; Todesco et al. 2016; Galaverni et al. 2017). We found that most of our Texas Pigtoe specimens, which were the focus of this study, showed evidence of hybridization with other species. The unique gene pool of Texas Pigtoe may be at risk of genetic swamping, and this has not been documented before. Uniquely it is our genome-wide approach that allowed us to come to this conclusion. The next steps regarding what to do about this hybridization lie outside the scope of this grant, but we believe we have presented conclusive genomic evidence that it is occurring among all of the East Texas endemic species included in our dataset. We believe plans should be made to mitigate hybridization among species, by, for instance, finding areas of allopatry among species and monitoring the populations over time, both genetically with traditional field observations, to ensure that other species or their genes do not move in to those areas.

While the focus of our efforts was on Texas Pigtoe, we had some noteworthy hybridization results involving Louisiana Pigtoe. Louisana Pigtoe is endemic to East Texas and listed as threatened by the Texas Parks and Wildlife Department,[6] and is also under review by the US Fish & Wildlife Service for listing as threatened or endangered.[7] Our results suggest that hybridization is occurring between the Louisiana Pigtoe and the Texas Pigtoe and between the Louisiana Pigtoe and the Trinity Pigtoe. This is surprising, because the Louisana Pigtoe is in a

[6] https://tpwd.texas.gov/huntwild/wild/wildlife_diversity/nongame/listed-species/invertebrates.phtml
[7] https://ecos.fws.gov/ecp0/profile/speciesProfile?spcode=Q3RF

different genus from the Texas Pigtoe and Trinity Pigtoe, and also because it was not known that the Trinity Pigtoe and the Louisiana Pigtoe come into contact. Genomic testing is needed on a larger number of Louisiana Pigtoe specimens in order to determine how often hybridization is occurring and with which other species the hybridization is most frequently occurring. This research can be complemented by laboratory test crosses of Louisiana Pigtoe with Texas Pigtoe and Triangle Pigtoe to determine how cross-fertile they are. While it is clear that Louisiana Pigtoe is hybridizing, the question remains of how big a threat this is to the species.

We also had some tangential results of interest about the Trinity Pigtoe. Contrary to several sources (Vidrine 1993; Howells et al. 1996; Haag 2010; Burlakova et al. 2012; Williams et al. 2017), our results suggest that the Trinity Pigtoe (*Fusconaia chunii*) is in fact distinct from other species in the *Fusconaia* genus. Previous sources have claimed that the Trinity Pigtoe is synonymous with either the Texas Pigtoe (Burlakova et al. 2012; Williams et al. 2017) or the Wabash Pigtoe (Vidrine 1993; Howells et al. 1996; Haag 2010). But we found it to genomically cluster separately from the other Pigtoes. This agrees with the studies by Inoue et al. (2018) and Pieri et al. (2018), who also suggested that the Trinity Pigtoe is distinct based on its genetic ancestry. But we emphasize that our study is more definitive, because it is based on multiple specimens using genome-wide SNPs, in contrast to the individual DNA sequences used in the previous studies. We believe it is now possible to conclude with confidence that Trinity Pigtoe is a separate species. Yet we know very little about the Trinity Pigtoe. Specifically, we do not know whether the populations of Trinity Pigtoe are stable or are in decline, we do not know its habitat affinities, we do not have a comprehensive knowledge of its biogeography (such as what rivers it is found in), and which species pose threats to its existence through hybridization. More research is warranted on these topics.

That said, the Trinity Pigtoe appears to be hybridizing with other species. We found high amounts of hybridization with Wabash Pigtoe evident in the ancestries of the Trinity Pigtoe specimens in the San Jacinto River (Figure 6, bottom). This matches Pieri et al.'s (2018) finding that the Wabash and Trinity Pigtoe coexist. Our study extends this finding, suggesting that not only do they coexist but they are also hybridizing. Hybridization with the widespread Wabash Pigtoe could be a threat to the Trinity Pigtoe, but more genomic research is needed on the extent of this problem in both the San Jacinto and Trinity Rivers. We also found high amounts of Trinity Pigtoe ancestry in specimens from the Neches/Angelina and Sabine River watersheds (Figure 6, top). We do not have conclusive evidence of the Trinity Pigtoes in those watersheds, but it appears that contact between the Trinity Pigtoe and other species in these watersheds was recent enough to leave a signature in the genomes. Since the Trinity Pigtoe is morphologically similar to the Wabash Pigtoe (Pieri et al. 2018), future sampling in the Neches, Angelina, and Sabine Rivers should include Wabash Pigtoe-looking specimens to see if they may, in fact, be Trinity Pigtoes. Such studies should be performed using multiple specimens of the Texas, Trinity, and Wabash Pigtoes for comparative purposes, and using genome-wide SNPs. We emphasize that the genomic signature of Trinity Pigtoe was revealed to be present in other watersheds only because of our genome-wide SNPs. It had not been observed before.

Finally, we found that two of the specimens from the Trinity River contain high amounts of Texas Pigtoe ancestry (Figure 6, bottom). This suggests that either the Texas Pigtoe occurs in the Trinity River watershed or that the specimens in the Trinity River were hybridizing with the Texas Pigtoe recently enough for it to leave a signature in their genomes. Again, we emphasize that the biogeography of the Trinity Pigtoe has not been studied in depth, so it is unclear how the Texas Pigtoe and the Trinity Pigtoe would have come into contact and when. But more research

is warranted on more specimens to determine if there are, in fact, Texas Pigtoe specimens in the Trinity River. Such studies should be focus on specimens from the Trinity River, but should also include multiple specimens of the Texas, Trinity, and Wabash Pigtoes for comparative purposes, and using genome-wide SNPs. We emphasize that the genomic signature of Texas Pigtoe was revealed to be present in Trinity Pigtoe specimens only because of our genome-wide SNPs. It had not been observed before.

Ecological niche modeling. The niche model for Texas Pigtoe that we created broadly agreed with our previous efforts (Walters et al. 2017), showing highest habitat suitability in the middle Neches and Angelina Rivers, and the Upper Sabine River. But there was also some suitable habitat found in the Trinity River, which could explain why some Trinity Pigtoe specimens have a genomic signature of hybridization with the Texas Pigtoe in the Trinity River watershed. Our ground-truthing results showed that Texas Pigtoes tend to occur in areas of higher habitat suitability than areas of lower suitability on the niche model map; however, there was a lot of variability around this trend, rendering it not statistically significant. This could be due to the fact that "pure" Texas Pigtoe is rarer than we thought it was, giving us relatively data points to work with for the modeling effort, as well as due to inherent "noise" in the ecological dataset arising from the inherent complexity of natural environments (Osenberg et al. 1994; Moller and Jennions 2002).

We believe our map is a useful tool for land managers to point them towards areas where Texas Pigtoe is more likely to occur versus not occur, but field verification of individual sites is still warranted. This was the most intensive niche modeling effort to date on any East Texas freshwater mussel, to our knowledge, incorporating the newest and largest number of environmental layers from diverse sources, covering the land, water, and air (climate) at high

resolution (100 m x 100 m). We did not include sites on the Trinity River in our ground-truthing effort, despite the appearance of some suitable habitat for Texas Pigtoe in that basin and the indication that hybridization with Texas Pigtoe is occurring in that basin, because no morphological specimens of Texas Pigtoe have been reported in the Trinity River to date. Therefore, we surmised that any Texas Pigtoe specimens in that basin would be morphologically cryptic and require genotyping by genomic markers; but the genomics portion of the study occurred before the niche modeling and ground-truthing portions, so no more genomic work was possible. That said, future field and genomic work is warranted in the Trinity River to search for more signatures of Texas Pigtoe occurring in that basin. Texas Pigtoe and Trinity Pigtoe pose a threat to each other, as evidenced by our genomic results, when they come into contact because they hybridize.

We found that suitable habitat for the Texas Pigtoe is associated with the suitability of the same habitat for the Blacktail Shiner (see supplementary data). Interestingly, however, the Red Shiner's habitat affinities did not affecting the Texas Pigtoe's distribution. For some reason, the distribution of the Texas Pigtoe was more reliant on the Blacktail Shiner's distribution than on the Red Shiner's Distribution. This suggests that the Blacktail Shiner is a more important host fish for the Texas Pigtoe in the wild than the Red Shiner. Mechanistic research is needed on the ecological interactions among the Texas Pigtoe, Blacktail Shiner, and Red Shiner to confirm these findings as well as to help understand why the Blacktail Shiner would be a more important to Texas Pigtoe than the Red Shiner. This is the first study documenting that the Blacktail Shiner is more important to the Texas Pigtoe's distribution than its other putative host fish, the Red Shiner.

Also new to the latest modeling effort were the inclusion of a base flow layer, representing the component of streamflow that can be attributed to ground-water discharge into streams. This layer was downloaded from the US Geological Survey (Wolock 2003). We found that Texas Pigtoes are associated with areas where groundwater makes up more of the river's total flow as compared to precipitation. This makes sense in light of the lack of available calcium in East Texas. Available calcium in the water column, associated with a more basic pH, has traditionally been regarded as a crucial determinant of mollusc distributions (Boycott 1936; Clarke and Berg 1959). Yet soils in East Texas are naturally low in calcium and naturally acidic, like much of the south-eastern United States (Beck and Reuter 1974). This has made it unclear where East Texas mussels are acquiring the calcium that they need for shell formation (Walters et al. 2017). Our results suggest that they may be getting this calcium from mineral-rich groundwater upwellings. This has phenomenon has never been documented before in East Texas mussels, to our knowledge. The correlative findings warrant further study on a mechanistic level. For instance, the dissolved ion contents of different rivers could be measured, and Texas Pigtoes could be grown in the laboratory under those different dissolved ion concentrations, with the effect on shell formation documented. This would be a fruitful area of future study.

Acknowledgements

We thank Dr. Charles Randklev at the Texas A&M AgriLife Dallas Center for providing tissue samples. UT-Tyler undergraduate students Samantha Rowe and Emmanuel Guitierrez contributed to the field work, as did Biology faculty member Dr. Neil Ford. We are appreciative of Kim Horndeski, Megan Hope, Colin McDonald, and Melissa Salmon at the Texas

Comptroller of Public Accounts for making this work possible and for their continued guidance, support, meeting, and efforts. We gratefully acknowledge the thoughtful feedback we have received from Clint Robertson at the Texas Parks and Wildlife Department River Studies Program, and Gary Pandolfi, Jacob Lewis, and Erik Orsak at the US Fish & Wildlife Service Arlington Ecological Services Field Office. Finally, we thank the Pigtoe Genomics Technical Advisory Panelists Dr. Amanda Rosenberger at Tennessee Technological University and Dr. Michael Gangloff at Appalachian State University, who helped formulate the research plan for this project and who provided constructive critiques along the way that greatly improved the quality of the work.

Literature Cited

Aiello-Lammens, M. E., R. A. Boria, A. Radosavljevic, B. Vilela, and R. P. Anderson. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography 38:541-545.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376.

Beck, K. C. and J. H. Reuter. 1974. Organic and inorganic geochemistry of some coastal plain rivers of the southeastern United States. Geochemica et Cosmochimica Acta 38:341-364.

Bertram, E. P., J. S. Placyk, Jr., M. Williams, and L. Williams. 2017. Verification of two Cyprinid host fishes for the Texas Pigtoe, *Fusconaia askewi*. Freshwater Mollusk Biology and Conservation 20:65-70.

Boycott, A. E. 1936. The habitats of the fresh-water molluscs in Britain. J. Anim. Ecol. 5:116-186.

Brito, P. H. and S. V. Edwards. 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica 135:439-455.

Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol Biol Evol 29:1917-1932.

Buchroithner, M. F. and R. Pfahlbusch. 2017. Geodetic grids in authoritative maps – new findings about the origin of the UTM Grid. Cartography and Geographic Information Science 44:186-200.

Burlakova, L. E., D. Campbell, A. Y. Karatayev, and D. Barclay. 2012. Distribution, genetic analysis and conservation priorities for rare Texas freshwater molluscs in the genera Fusconaia and Pleurobema (Bivalvia: Unionidae). Aquatic biosystems 8:12.

Clarke, A. H. and C. O. Berg. 1959. The freshwater mussels of central New York. Memoirs of the Cornell University Agricultural Experiment Station 367:1-79.

Cox, D. R. 1958. The regression analysis of binary sequences. Journal of the Royal Statistical Society Series B-Statistical Methodology 20:215-242.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12:499-510.

Drummond, A. J. and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7.

Elith, J. 2002. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. Pp. 39 - 58 *in* S. Ferson, and M. Burgman, eds. Quantitative methods for conservation biology. Springer, New York.

Fitzpatrick, B. M., M. E. Ryan, J. R. Johnson, J. Corush, and E. T. Carter. 2015. Hybridization and the species problem in conservation. Curr. Zool. 61:206-216.

Ford, D. 2013. Ground truthing Maxent in East Texas rivers. M.S. Thesis, University of Texas at Tyler.

Galaverni, M., R. Caniglia, L. Pagani, E. Fabbri, A. Boattini, and E. Randi. 2017. Disentangling timing of admixture, patterns of introgression, and phenotypic indicators in a hybridizing wolf population. Molecular Biology and Evolution 34:2324-2339.

Haag, W. R. 2010. A hierarchical classification of freshwater mussel diversity in North America. J. Biogeogr. 37:12-26.

Hammond, E. 1964. Classes of land-surface form in the United States. U.S. Geological Survey, Reston, VA. https://water.usgs.gov/lookup/getspatial?na70_landfrm . Accessed 29 July 2019.

Hendrickson, D. A. and A. E. Cohen. 2015. Fishes of Texas Project Database (Version 2.0). doi:10.17603/C3WC70. Accessed 29 June 2019.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773-785.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965-1978.

Horton, J. D. 2017. The State Geologic Map Compilation (SGMC) geodatabase of the conterminous United States (ver. 1.1, August 2017): U.S. Geological Survey data release, https://doi.org/10.5066/F7WH2N65 . Accessed 29 July 2019.

Howells, R. G. 2006. Statewide freshwater mussel survey. State Wildlife Grant Final Report T-15-P, Texas Parks and Wildlife Department, Austin.

Howells, R. G., R. W. Neck, and H. D. Murray. 1996. Freshwater mussels of Texas. Texas Parks and Wildlife Department.

Inoue, K., D. M. Hayes, J. L. Harris, N. A. Johnson, C. L. Morrison, M. S. Eackles, T. King, J. W. Jones, E. M. Hallerman, A. D. Christian, and C. R. Randklev. 2018. The Pleurobemini (Bivalvia: Unionida) revisited: Molecular species delineation using a

mitochondrial DNA gene reveals multiple conspecifics and undescribed species. Invertebrate Systematics 32:689-702.

Junkins, D. and G. Garrard. 2006. Demystifying reference systems: a chronicle of spatial reference systems in Canada. Geomatica 60:151-164.

Karatayev, A. Y. and L. E. Burlakova. 2007. East Texas Mussel Survey. Final Report. Texas Parks and Wildlife State Wildlife Grant.

Lischer, H. E. L., L. Excoffier, and G. Heckel. 2014. Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of microtus voles. Molecular Biology and Evolution 31:817-831.

Maddison, W. P. 1997. Gene trees in species trees. Syst. Biol. 46:523-536.

Marshall, N. T., J. A. Banta, L. R. Williams, M. G. Williams, and J. S. Placyk. 2018. DNA Barcoding Permits Identification of Potential Fish Hosts of Unionid Freshwater Mussels. American Malacological Bulletin 36:42-56.

Merow, C., M. J. Smith, and J. A. Silander Jr. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36:1058-1069.

Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17:240-248.

Mitchell, M. W. and M. K. Gonder. 2013. Primate Speciation: a case study of African apes. Nature Education Knowledge 4:1.

Moller, A. P. and M. D. Jennions. 2002. How much variance can be explained by ecologists and evolutionary biologists? Oecologia 132:492-500.

Neves, R. J., A. E. Bogan, J. D. Williams, S. A. Ahlstedt, and P. W. Hartfield. 1997. Status of the aquatic mollusks in the southeastern United States: a downward spiral of diversity *in* G. W. Benz, and D. E. Collins, eds. Aquatic Fauna in Peril: The Southeastern Perspective Special Publication 1. Southeast Aquatic Research Institute.

Osenberg, C. W., R. J. Schmitt, S. J. Holbrook, K. E. Abu-Saba, and A. R. Flegal. 1994. Detection of environmental impacts: natural variability, effect size, and power analysis. Ecological Applications 4:16-30.

Pamilo, P. and M. Nei. 1988. Relationships between gene trees and species trees. Mol Biol Evol 5:568-583.

Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. 34:102-117.

Phillips, S. J. 2017. A brief tutorial on Maxent. Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2017-05-30.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231-259.

Phillips, S. J. and M. Dudik. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161-175.

Pieri, A. M., K. Inoue, N. A. Johnson, C. H. Smith, J. L. Harris, C. Robertson, and C. R. Randklev. 2018. Molecular and morphometric analyses reveal cryptic diversity within freshwater mussels (Bivalvia: Unionidae) of the western Gulf coastal drainages of the USA. Biological Journal of the Linnean Society 124:261-277.

Porras-Hurtado, L., Y. Ruiz, C. Santos, C. Phillips, A. Carracedo, and M. V. Lareu. 2013. An overview of STRUCTURE: applications, parameter settings, and supporting software. Front Genet 4:98.

Pritchard, J. K., M. Stephens, N. A. Rosenberg, and P. Donnelly. 2000. Association mapping in structured populations. American Journal of Human Genetics 67:170-181.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rhymer, J. M. and D. Simberloff. 1996. Extinction by hybridization and introgression. Annual Review of Ecology and Systematics 27:83-109.

Rubin, B. E. R., R. H. Ree, and C. S. Moreau. 2012. Inferring phylogenies from RAD sequence data. PLoS ONE 7:e33394.

Russello, M. A., M. D. Waterhouse, P. D. Etter, and E. A. Johnson. 2015. From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ 3:e1106-e1106.

Shaw, K. L. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proceedings of the National Academy of Sciences 99:16122.

Sokal, R. R. and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research. W.H. Freeman and Company, New York.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

Steane, D. A., D. Nicolle, C. P. Sansaloni, C. D. Petroli, J. Carling, A. Kilian, A. A. Myburg, D. Grattapaglia, and R. E. Vaillancourt. 2011. Population genetic analysis and phylogeny

reconstruction in Eucalyptus (Myrtaceae) using high-throughput, genome-wide genotyping. Molecular Phylogenetics and Evolution 59:206-224.

Swenson, K. M. and N. El-Mabrouk. 2012. Gene trees and species trees: irreconcilable differences. BMC bioinformatics 13 Suppl 19:S15-S15.

Takeuchi, F., K. Yanai, T. Morii, Y. Ishinaga, K. Taniguchi-Yanai, S. Nagano, and N. Kato. 2005. Linkage disequilibrium grouping of single nucleotide polymorphisms (snps) reflecting haplotype phylogeny for efficient selection of tag SNPs.

Todesco, M., M. A. Pascual, G. L. Owens, K. L. Ostevik, B. T. Moyers, S. Hubner, S. M. Heredia, M. A. Hahn, C. Caseys, D. G. Bock, and L. H. Rieseberg. 2016. Hybridization and extinction. Evolutionary Applications 9:892-908.

U.S. Geological Survey Gap Analysis Program. 2016. GAP/LANDFIRE National Terrestrial Ecosystems 2011. U.S. Geological Survey, https://doi.org/10.5066/F7ZS2TM0 . Accessed 29 July 2019.

United States Department of Agriculture, National Resources Conservation Service, and National Soil Survey Center. 1995. Soil Survey Geographic (SSURGO) Data Base: Data use information. Miscellaneous Publication Number 1527. Downloaded from http://www.fsl.orst.edu/pnwerc/wrb/metadata/soils/ssurgo.pdf on 13, September 2018.

US Environmental Protection Agency (USEPA) and US Geological Survey (USGS). 2005. National hydrography dataset plus. Horizon Systems Corporation. www.horizon-systems.com/NHDPlus. (Accessed May 7 2019).

Vaughn, C. C. 1997. Catastrophic decline of the mussel fauna of the Blue River, Oklahoma. The Southwestern Naturalist 42:333-336.

Vidrine, M. E. 1990. Fresh-water mussel-mite and mussel-Ablabesmyia associations in Village Creek, Hardin County, Texas. Proceedings Louisiana Academy of Science 53:1-4.

Vidrine, M. F. 1993. The historical distributions of freshwater mussels in Louisiana. Gail Q. Vidrine Collectibles, Eunice, LA.

Walters, A. D., D. Ford, E. T. Chong, M. Williams, N. B. Ford, L. R. WIlliams, and J. A. Banta. 2017. High-resolution ecological niche modeling of threatened freshwater mussels in east Texas, USA. Aquatic Conservation: Marine and Freshwater Ecosystems, in press.

Warren, D. L., R. E. Glor, and M. Turelli. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. Ecography 33:607-611.

Warren Jr., M. L., B. M. Burr, S. J. Walsh, H. L. Bart Jr., R. C. Cashner, D. A. Etnier, B. J. Freeman, B. R. Kuhajda, R. L. Mayden, H. W. Robison, S. T. Ross, and W. C. Starnes. 2000. Diversity, distribution, and conservation status of the native freshwater fishes of the southern United States. Fisheries 25:7-31.

Wiens, J. J., C. A. Kuczynski, and P. R. Stephens. 2010. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. Biological Journal of the Linnean Society 99:445-461.

Williams, J. D., A. E. Bogan, R. S. Butler, K. S. Cummings, J. T. Garner, J. L. Harris, N. A. Johnson, and G. T. Watters. 2017. A revised list of the freshwater mussels (Mollusca: Bivalvia: Unionida) of the United States and Canada. Freshwater Mollusk Biology and Conservation 20:33-58.

Williams, J. D., M. L. Warren Jr., K. S. Cummings, J. L. Harris, and R. J. Neves. 1993. Conservation status of freshwater mussels of the United States and Canada. Fisheries 18:6-22.

Wolock, D. M. 2003. Base-flow index grid for the conterminous United States. Available at: https://water.usgs.gov/lookup/getspatial?bfi48grd. Accessed May 7, 2019. US Geological Survey, Reston, VA.

Yoder, J. B., R. Briskine, J. Mudge, A. Farmer, T. Paape, K. Steele, G. D. Weiblen, A. K. Bharti, P. Zhou, G. D. May, N. D. Young, and P. Tiffin. 2013. Phylogenetic signal variation in the genomes of *Medicago* (Fabaceae). Syst Biol 62:424-438.