Journal of
Emerging Investigators

Using DNA Barcodes to Evaluate Ecosystem Health in the SWRCMS Reserve

Krystal Horton[1], Kim Kemmis[1], Michael Horton[1], Ryan Perry[2], Nick Reeves[3]

[1]Western Center Academy, Hemet, CA 92543

[2]University of California Riverside, Riverside, CA 92521

[3]Mt. San Jacinto Community College, San Jacinto, CA 92583

Summary

The Western Center Academy (WCA) Charter School and Mt. San Jacinto College (MSJC) are located on the Southwestern Riverside County Multi-Species Reserve (SWRCMSR). Developed in 1992, the reserve is home to at least 8 sensitive bird, animal, and plant species. Although the reserve is protected against direct human influence, within a 100-mile circle of SWRCMSR live more than 21 million people with their indirect impact on the environment. Our goal was to determine if the reserve is protecting the species within by analyzing the genetics of those species to determine if there is a healthy amount of genetic diversity in the populations within. We collected insects at MSJC Menifee campus and in the SWRCMSR to evaluate how human impact affects the genetic diversity of insects. MSJC served as the site that was impacted by human activities and SWRCMSR served as the site that was not impacted by human activities. The School Malaise Trap Project (SMTP) selected 160 of our specimens, sequenced their genetic barcodes, and entered the data into the Barcode of Life Database (BOLD) for analysis. Using the BOLD website and a Clustal Omega tool, we compared the genetic sequences in the insects we collected to measure genetic differences in the two populations of insects. The results from the two collections seem to show that the nature reserve insect population is a younger population with genetic changes possibly acquired by gene flow from immigration. The insect population at MSJC appears to be well developed, but isolated genetically.

Received: November 1, 2017; **Accepted:** July 10, 2018; **Published:** September 27, 2018

Introduction

There are 1.2 million square kilometers of protected land in the United States (1). Little research has been done on how well nature reserves work at preserving the health of species within their boundaries beyond the diversity of specimens (2-3) or recovery of endangered species (4-5). One way to measure the health of a population is by quantifying the genetic variation within the population. In order for a population to be healthy, it must be able to evolve in response to changes in the environment. For evolution to occur, there must be variations in the genome for natural selection to act upon. There are four known mechanisms for causing changing allele frequencies, which we have referred to as genetic variations: natural selection, mutation, genetic drift, and gene flow (6).

Gene flow, or gene migration, occurs when gene variations transfer from one population to another. This sharing of genes leads to healthier populations since all of the members of the population will not have identical alleles of certain genes that would make them susceptible to the same diseases and environmental changes. Genetic drift occurs in small populations when random changes (fire, new construction, landslide) modify the allele frequency of one or more genes because a particular allele was randomly and disproportionately affected by the change. This differs from natural selection because it is a random occurrence unrelated to fitness and does not necessarily produce beneficial adaptations. Genetic drift could make the genome of a population more varied or, at other times, less varied (6).

The type of genetic barcoding used in this study is unable to differentiate between genetic drift and mutation. Additionally, the populations studied are too young to have been influenced by natural selection yet. Therefore, our study focuses on mutations and gene flow.

DNA barcoding was developed by Hebert and Stoeckle and uses a segment of approximately 600 base pairs that is highly variable between different species, but consistent within a species (7). If the segment assessed were of low variability, then all species would have the same DNA barcode. If it were of too high variability, then even the same species would look very different. A region called mitochondrial cytochrome c oxidase subunit 1 (COI) was selected that fit all of these requirements. The researchers found that the approximately 600 base pairs in their barcodes were able to correctly recognize species with 98 percent accuracy when validated with

other methods (7).

These barcodes can then be uploaded to the Barcode of Life Database (BOLD) managed by the University of Guelph in Canada. There are millions of sequences representing hundreds of thousands of species in the collection. Once a scientist has properly identified a species in the database, then anyone in the future who uploads a genetically similar sequence will be informed of the probable species identification. Many insects with nearly identical barcodes are noted to be the same species as other insects in the database without knowing the specific species name because no formal identification has been done yet (8). For our research, the actual identity of the species was unimportant, only the knowledge that the insects belonged to the same species.

Through the BOLD website, members of the same BIN (nearly identical DNA, highly likely to be the same species) that were collected by others in surrounding areas could be identified. These could be used to identify if there was any genetic migration (gene flow). If a mutation in one population shows up in another, the probability that there were identical random mutations is much lower than the probability that the mutation migrated from one population to the other.

Through genomic barcoding, the aim of this study is to compare the genetic health of two populations of insects, one in a nature reserve with very limited human influence and the other at a site that was just like the nature reserve before being turned into farms and is now a high-traffic, isolated community college campus. By comparing a virtually no-human-impact group to a high-human-impact group, two things can be determined: how well the nature reserve is preserving the genetic diversity of its occupants and how human impact affects the genetics of specimens in high-traffic areas. As discussed previously, evidence of both gene flow and genetic drift are signs of a healthy population. These genotypic differences give rise to phenotypic differences on which natural selection can act if the environment changes, which may contribute to the long-term survival of the population.

Here, we hypothesized that if, in fact, a nature reserve preserves the genetic health of a population, the gene sequences and genetic relationships reflected by phylogenetic clustering should show more evidence of gene flow and genetic drift in the nature reserve insects than in the community college insects. This would demonstrate more variety in the genome for natural selection to act upon in order to strengthen the genetic diversity of the species.

Results

We compared the genetic diversity of insects at Mt. San Jacinto College (MSJC) and Southwestern Riverside County Multi-Species Reserve (SWRCMSR) using the DNA barcode sequences from the School Malaise Trap Project (SMTP) that we participated in. We chose a suitable insect from the SMTP to analyze. The insect had to be present in high enough numbers at both the MSJC and SWRCMS sites to be able to compare data, and needed to have good DNA barcoding results with a history of clean sequence traces. A large variety of insects were collected and this midge represents only a small portion of the total (**Figure 1**). Any conclusions drawn by this study may only be assumed to be true for

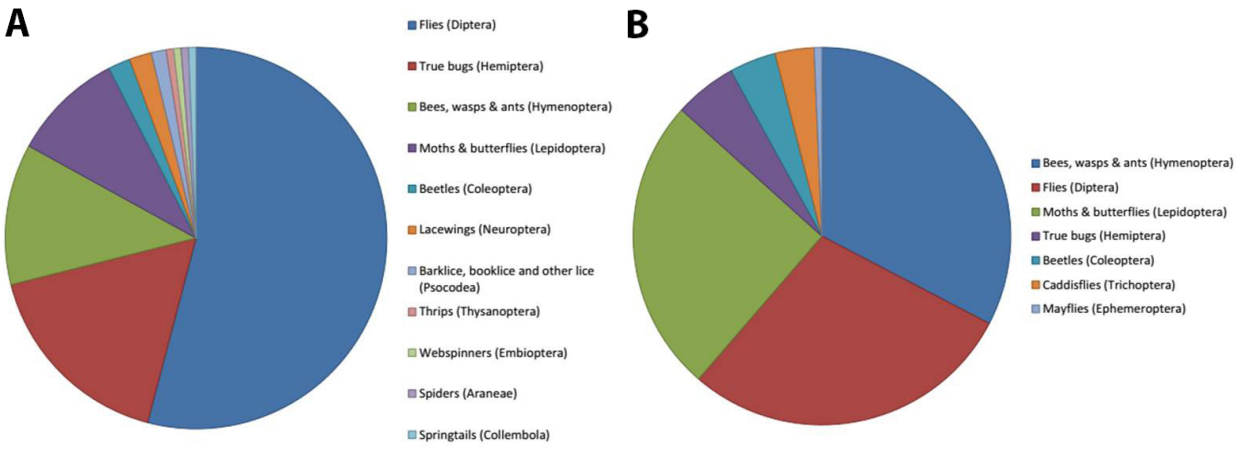

Figure 1. Diversity of Insect Orders at Sample Collection Locations. The variety of insect orders at MSJC, the site that was impacted by human activities (A), is compared to the variety at SWRCMSR, the site not impacted by human activity (B). *Diptera* dominated the variety of orders of insects collected at MSJC, while some crawling insects (spiders) also found their way into the flying insect trap. Caddisflies and mayflies, not present in the dry MSJC environment, reflect the diversity of SWRCMSR and its proximity to water. The larger than expected *Hymenoptera* population the SWRCMSR resulted from a coincidental emergence of winged ants.

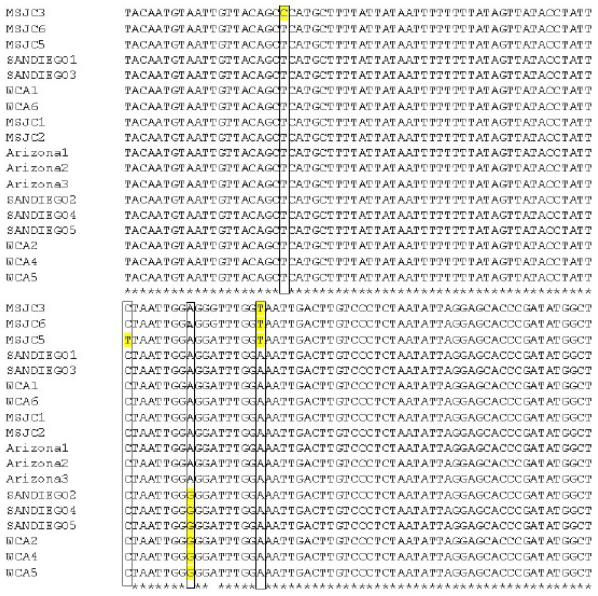

Table 1. An example of the first 60 base pairs of the sequences trimmed, aligned, and compared using Clustal Omega. MSJC specimens were collected at the college, WCA specimens were collected in the nature reserve, and Arizona and San Diego specimens were from the BOLD database. The Clustal Omega tool determined the order of the sequences in the list. Highlights mark mutations. Asterisks at the bottom show no mutations at that base pair. An "n" signals that the quality of the identification of that base pair was low.

this particular midge, not the entire ecosystem. Flies made up a large percentage of the specimens at both sites. An emergence of winged ants in SWRCMS during one of the collections increased the percentage of hymenoptera at that site; otherwise the two populations would have matched even more consistently. Six non-biting midges from both MSJC and SWRCMSR were barcoded and five gave clean sequences. The sequences were entered into the BOLD database for analysis and comparison. The BOLD database was able to identify numerous identical species of midges from surrounding areas. Five specimens were randomly selected from the BOLD database from both San Diego and Arizona to compare to our SWRCMSR and MSJC insects to attempt to evaluate gene flow.

Table 1 shows a small sample of what the DNA barcode sequences for the midge specimens look like when trimmed, aligned, and compared using a Clustal Omega tool. Each sequence represents more than 200 base pairs from an individual midge, but only the first 120 base pairs of one comparison group are shown here. The barcodes were trimmed to remove low-quality data and different start and end points in the sequence, resulting in the approximately 200 perfectly aligned base pairs. The lab used standard barcoding procedures to recover the sequence of the cytochrome c oxidase subunit I (COI) gene.

Comparison	# of differences	# of base pairs	% difference
WCA vs. WCA	4	455	0.88%
MSJC vs. MSJC	13	501	2.59%
MSJC vs. Arizona	11	501	2.20%
WCA vs. Arizona	6	630	0.95%
WCA vs. MSJC	13	455	2.86%

Table 2. The number and percentage of mutations within a group and between different groups. There are relatively few mutations within the WCA population (0.88%) which is very similar to the Arizona population (0.95% differences). MSJC had more mutations within the population (2.59%) and relatively more differences with Arizona and WCA populations (2.20% and 2.86% respectively).

Table 3. Barcodes showing an example mutations. A mutation shared amongst a large number of individual specimens ("SHARED"), suggests a possible gene flow mechanism, and a mutation only found in one specimen ("INDIVIDUAL") suggests a possible genetic drift mechanism.

The vast majority of base pairs (> 97%) were identical for the two samples and the specimens from surrounding areas (**Table 2**). The differences in Table 1 are a mixture of single-nucleotide mutations. We found three single-nucleotide mutations (T to C; C to T, and A to T) that are unique to MSJC's midge specimens (**Table 1**). These changes do not show up in the DNA barcode sequences of the other midge specimens. We also identified single-nucleotide mutations that are seen in the DNA barcode sequences for the SWRCMSR midges (WCA 2, 4, and 5) and in DNA barcodes found in the BOLD database from San Diego (SANDIEGO 2, 4, and 5) midges (**Table 1**, highlighted values). Additional single-nucleotide mutations outside the portion of the DNA barcode sequences shown in **Table 1** were generally shared amongst all of the specimens, but not shared with the MSJC specimens. We also compared sequences between specimens from different populations (WCA specimens from SWRCMS, MSJC specimens, and specimens from Arizona) (**Table 2**). The number of sequence differences is always greater in the MSJC population.

We identified specific DNA barcode sequence changes that occurred in several specimens from several populations (**Table 3**). We observed these

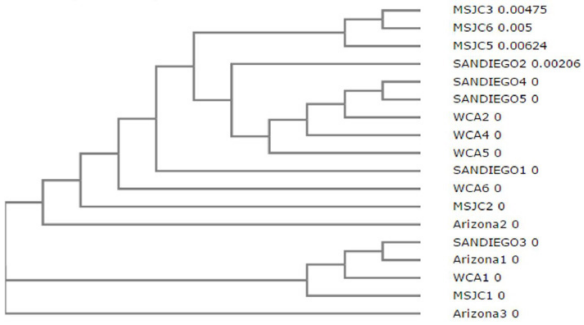

Figure 2. Phylogenetic tree created by the Clustal Omega tool demonstrating the relationship between different specimens. Looking back two lineages on the tree, there are no common ancestors between MSJC specimens and any others. Within two lineages, WCA has common ancestors with Arizona and Arizona has common ancestors with San Diego.

patterns of sequence change—changes seen in several populations and those seen in only one specimen—several times in the midge DNA barcode sequences that we analyzed.

We also created a phylogenetic tree using the Clustal Omega tool (**Figure 2**) based on the sequences we entered into the algorithm. Three of the MSJC specimens (MSJC 3, 5, and 6) are grouped together. The data show that the genetic differences between these specimens is low and the differences comparing these specimens to the others is relatively high. This indicates that the DNA barcode sequences for these MSJC specimens are different from the rest of the specimens.

Discussion

By analyzing gene sequences of midges in Southern California, we found evidence that the SWRCMSR might have a younger, more isolated population of midges than MSJC. Our results indicate that the single-nucleotide changes seen in the SWRCMSR midge population (shown as "WCA" in the tables) are also seen in other midge populations found in the BOLD database, and there are very few DNA barcode changes that are unique to the SWRCMSR midge population. The SWRCMSR is a recently created reserve around the man-made Diamond Valley Lake. This lake and the associated reserve is a new habitat for the midge. Our DNA barcode data analysis demonstrates that the midge population at SWRCMSR was founded by midges that migrated to Diamond Valley Lake either by natural dispersion or dispersion facilitated by humans relatively recently. The founding SWRCMSR population may have recently inherited the DNA barcode changes we observed from other populations like the San Diego and Arizona populations (**Table 1**).

This pattern of DNA barcode sequence similarities suggests a gene flow mechanism, because it appears that these sequence similarities were brought to the SWRCMSR population by migration of midges from other populations. Verifying this hypothesis will require the analysis of additional specimens to determine whether gene flow is the true mechanism of evolution causing the pattern of DNA barcode sequence similarities we see. In our specimens, we see that almost all of the DNA barcode changes in the SWRCMSR midge population are shared changes also seen in one or more of the other analyzed midge populations. We believe that this finding demonstrates that the SWRCMSR midge population is young and was founded by migration recently, because we would expect additional unique DNA barcode sequence changes in this population if it were older and more isolated. We did, indeed, see such unique DNA barcode sequence changes in the MSJC midge specimens that we analyzed. In **Table 1** and **Table 3**, we have highlighted DNA barcode sequence changes that are only seen in the MSJC midge specimens. These unique DNA barcode sequence changes cause three of the MSJC specimens to group separately from the rest of the specimens in our phylogenetic tree (**Table 4**).

This pattern of DNA barcode change seen in the MSJC specimens may suggest a genetic drift mechanism, because it appears that these sequence changes arose randomly in the MSJC midge population and not in the other populations. We think that the MSJC midge population is older and more isolated because we see several of these unique "genetic drift" changes in the MSJC specimens. However, analysis of additional specimens is required to determine whether genetic drift is the true mechanism of evolution causing the pattern of DNA barcode sequence changes we see. Taken together, our results indicate that we have potentially discovered an approach to DNA barcode analysis that may make it possible to distinguish newly founded insect populations from older, isolated insect populations.

By comparing the DNA barcode sequences of several specimens from a newly analyzed population to DNA barcodes from the BOLD database for several other populations using multiple sequence alignment, it is possible to identify shared "gene flow" changes and unique "genetic drift" changes in the new specimens. If there are only "gene flow" changes then the new population may be recently founded by migrants from other populations whereas if there are many "genetic drift" changes then the population may have be older and more isolated from other populations.

The nature reserve surrounds the man-made reservoir Diamond Valley Lake, which was filled in 2003. Because the lake has been in existence for only 14 years, we suspect that further analysis will show that the population will display a good balance of gene flow and

genetic drift given more time. Eventually, the population in SWRCMS may become both well established and healthy in its genetic diversity. At this point in time and with only one set of data, SWRCMSR was advised that the nature reserve seems to be working in protecting the genetic diversity of this insect population. Future follow-up testing will be needed to verify this conclusion.

With the flow of genes from outside the nature reserve boundaries comes the possibility of invasive species and disease. Mixing with other native species is good for a population's genetics, but it also shows that there are paths for detrimental vectors to get in and out of the nature reserve, as well. Naturalists will need to monitor for the presence of invasive pests and disease carefully.

The MSJC Menifee campus has been in existence for 37 years, and the unique genetic mutations in the midge population show that it is well established. Since the area around MSJC is being developed with homes and businesses extensively, it is likely that this population will remain isolated. The small golf course ponds in this area have been in existence since at least 1989. These insects have had time to mutate, but do not appear to share those mutations with surrounding populations.

We studied a small number of only one species of insect. It is possible that a larger number of specimens of this species or a different species could give different results. Because a water-dwelling specimen was selected and the lake is relatively young, a terrestrial specimen may not have shown the lack of mutations we saw in the nature reserve. As such, our conclusions only apply to this one species in this particular location. More work is needed before applying these results to other specimens and other locations. Now that a suitable species has been evaluated for small-scale testing, a larger sample size of midges would lend more credence to the conclusions. Additionally, it is possible that these conclusions only apply to midges. Testing on a variety of insect species would determine if the conclusions are universal.

Our results showed that the midges in SWRCMSR are sharing genes with surrounding populations. We explained to the SWRCMSR board of directors that this is a double-edged sword. Although genetic diversity is healthy, this flux of insects that lead to this diversity also opens the door for disease vectors as well as invasive species infestation in the reserve. The SWRCMSR board is in discussion how to move forward with this information and we will continue our analyses to assist them in their work.

Methods

For this study, two separate collections were conducted. One was part of an international project called the School Malaise Trap Project through the Centre for Biodiversity Genomics and the University of Guelph. For this part of the project, the insects were collected according to the SMTP protocol (copies of which may be requested from bioschoolmalaise@gmail.com) sent to all schools to ensure consistent collection techniques worldwide. Specimens were sent to University of Guelph for analysis. Additionally, we did a second collection of our own and individually sorted the insects and extracted and amplified the DNA. We then sent the DNA to Eton Bioscience Inc. in San Diego for DNA barcoding.

To collect the insects, identical Malaise Traps were set up in the SWRCMSR and MSJC sites at the same time and for the same length of time. A malaise trap is a tent-like structure with an exit to an anesthetic used for trapping flying insects. The same anesthetic was used in both traps (denatured ethanol). One collection lasted a week and the second collection lasted two weeks. The samples from the first collection were hand-identified, sorted, and preserved.

The DNA was extracted from selected specimens, amplified, purified, verified, and sent to Eton for sequencing. The specimens from the second collection were sent directly to Guelph University and 160 specimens of a variety of species were selected and sequenced in their labs. We selected one of these species for our research.

The data from both projects was entered into the BOLD database. The database was used to find identical species in surrounding areas (San Diego and Arizona). The DNA barcode of each specimen was extracted from the database and a Clustal Omega tool was used to trim the sequences to be the same length, and begin and end at the same base pair. The tool was used to find mutations, calculate mutation rates, and create phylogenetic trees. To use the Clustal Omega tool, the sequences were put into FASTA format and pasted into the Multiple Sequence Alignment tool at www.ebi.ac.uk. The site aligns the sequences and marks mutations. In order to get a correct count of the percentage of mutations, we trimmed the sequences to begin and end at the same base pair while maximizing the number of base pairs compared. Depending on which specimens were being compared, therefore, the total number of base pairs could vary. After aligning, we used the site's suite of tools for calculating statistics as well as creating phylogenetic trees.

For the second collection, we extracted genomic DNA from our insects and used PCR to amplify a 650 base pair segment of the mitochondrial cytochrome c oxidase subunit 1 (COI) gene. This segment of the COI

gene has been chosen as the accepted barcode region for insects and other animals. The COI PCR amplicons were purified and verified with gel electrophoresis and a NanoDrop spectrophotometer. The specimens were sent to Eton Bioscience for the barcode sequencing. The specimen details and barcode sequence data was entered into the Barcode of Life Database – Student Data Portal (BOLD-SDP). Within the BOLD-SDP website, the DNA barcode sequences we generated were compared to the Barcode of Life Database. The sequences were then passed through the Clustal Omega tool.

MSJC has perfected this extraction technique in the Honors Biology Laboratory over several years, therefore we used their technique exactly to ensure success. First, we labelled several tubes with the insect ID and put an insect part (leg or wing) in each tube. We added 180 µL of ATL Lysis Extraction Buffer as well as 20 µL of proteinase K enzyme. We ground up the insect part in this solution and mixed well by flicking. We incubated at 56°C in a shaking incubator at 500 rpm. 500 µL of 95% ethyl alcohol was added and the solution was centrifuged at 12,000 rpm for 3 minutes. We then transferred 190 µL of the supernatant to another microcentrifuge tube. Next, 200 µL of AL binding buffer were added and vortexed for 15 seconds after which 200 µL of 100% ethanol were added to the tube and vortexed again. This liquid was then transferred to a DNA extraction column and centrifuged at 8,000 rpm for one minute. We discarded the flow-through, added 500 µL of AW1 wash buffer 1 to the column, centrifuged at 8,000 rpm for 1 minute, and discarded the flow-through again. We repeated this process with AW2 wash buffer and centrifuged at 14,000 rpm for 3 minutes. We stored the samples in the refrigerator until our next session.

We added 50 µL of AE elution buffer to the spin column and let it stand for 1 minute. We then centrifuged at 12,000 rpm for two minutes. We removed the collection tube and discarded the rest of the spin column. The samples were then frozen until they were shipped to the Eton lab.

References

1. *World Database on Protected Areas*. Protected Planet, Jan. 2017, www.protectedplanet.net/c/world-database-on-protected-areas.
2. Zielinska, Janina. “Lichens of the Starożyn Nature Reserve.” *Acta Mycologica* (English translation), vol 5, no. 1, 1969, doi: 10.5586/am.1969.012.
3. Creevy, A., Fisher, J. Puppe, D., Wilkinson, D. “Protist Diversity on a Nature Reserve in NW England.” *Pedobiologia*, vol 59, no 1-2, doi: 10.1016/j.pedobi.2016.02.001.
4. Guan, Tian-Pei, *et al*. “Role of New Nature Reserve in Assisting Endangered Species Conservation - Case Study of Giant Pandas in the Northern Qionglai Mountains, China.” *Plos One*, vol. 11, no. 8, 2016, doi:10.1371/journal.pone.0159738.
5. Zong, Min. “Genetic Diversity and Geographic Differentiation in the Threatened Species Dysosma pleiantha in China as Revealed by ISSR Analysis.” *Biochemical Genetics*; Apr 2008, 180-96, doi: 10.1007/s10528-007-9141-7.
6. Urry, L., *et al*. “*Campbell Biology in Focus*.” Pearson, 2013.
7. Stoeckle, M. and Hebert, P. “Barcode of Life.” *Scientific American* 299, 82-88, doi:10.1038/scientificamerican1008-82.
8. Ratnasingham, S. and Hebert, P. “A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System.” *PLoS ONE* 8(8), doi:10.1371/journal.pone.0066213.