SHARE WITH WILDLIFE

Project Title: Investigating Natural Defenses in New Mexico Bats
Principal Investigator(s): Jennifer Hathaway, M.S. and Diana E. Northup, Ph.D.
Share with Wildlife Award Number 151203 / **UNM Project Number** A16-0333**State Wildlife Grant Number:** T-32-4, 9

Final Report February 15, 2016 through May 15, 2017

Hathaway SWW Final Report May 2017

Executive Summary:

This study investigated the diversity of a gene, Type II Polyketide Synthase (PKS II) from bacterial isolates cultured off of bats captured in New Mexico and Arizona. PKS II is the gene involved in one of the three main pathways for secondary metabolite production in bacteria. Secondary metabolites include antibacterial and antifungal compounds. We used the gene as an indicator of the antimicrobial potential of the external microbiota of bats. We looked at six bat species from New Mexico, from four regions across the state, as a representative sample of the bats in the state. We also looked at isolates that have been shown to inhibit the growth of *Pseudogymnoascus destructans* (*Pd*), the fungus that causes white-nose syndrome (WNS), and is decimating bat populations in eastern and central North America.

We screened 50 isolates from each of the six species to determine how wide spread this gene is in the population. We found differences in the number of isolates that were positive for the gene among the different bat species. We found that *Myotis ciliolabrum* (western small-footed Myotis) had the fewest number of positive isolates while *Eptesicus fuscus* (big brown bat) had the most. However, PKS II can have multiple copies of the same gene in a single bacterium, so this information is not enough to draw conclusions. We sequenced 13 isolates from each species in order to determine the copy number of the gene and the overall gene diversity. We found differences among bat species in terms of both copy number and diversity. This indicates a potential difference among species in the antimicrobial capability of the external microbiota of bats. We found that the diversity of the gene was wide spread in each species. When the region the bat was caught in was considered, we found that the bats from El Malpais National Monument had the most isolates that were potentially producing novel secondary metabolite compounds.

In addition to the 300 isolates from the six study species, 17 additional isolates that were previously identified as inhibiting *Pd* were screened for the presence of the PKS II gene. These isolates were from *Tadarida brasiliensis* (Brazilian free-tailed bat) from New Mexico, and *Antrozous pallidus* (pallid bat) and *Myotis evotis* (long-eared myotis) from Arizona. We wanted to better characterize these inhibitors and determine if any of them warranted further investigation as possible biocontrol agents against WNS. We found several inhibitors that potentially were producing novel secondary metabolites based on the sequence similarity to known PKS II products. We also found a group of inhibitors from *Corynorhinus townsendii* (Townsend's big-eared bat) bats captured in three different caves in El Malpais that were clustered by themselves with no closely related known sequence. This indicates that perhaps there is a shared bacterial type, or least a functional type, among *C. towsendii* at El Malpais that may be capable of inhibiting *Pd*. These isolates should be further investigated.

Management Implications:

This study found that El Malpais had the greatest potential for novel secondary metabolite products and that the caves and bats in the park should be protected as a potential source of novel antimicrobials. Additionally, *M. ciliolabrum* had the fewest number of isolates positive for the gene PKS II, as well as the second fewest number of isolates that were potentially making novel compounds. Moreover, *M. ciliolabrum* had only one isolate that was shown to be an inhibitor of *Pd*, but this isolate was negative for the PSK II gene. These factors combined would suggest that *M. ciliolabrum* is perhaps more vulnerable to WNS than other species sampled in this study and should be monitored for signs of WNS. This is especially critical now that *Pd* has been detected in Texas and WNS confirmed in Oklahoma. With the

National Speleological Society's National Convention being held in Rio Rancho, NM in June 2017, cavers from across the country, including WNS positive areas, will be visiting New Mexico. With this increased visitation, monitoring for *Pd* will be crucial in New Mexico for the next several years.

Background:

The discovery of white-nose syndrome (WNS) in the state of New York in the winter of 2006-2007 made bat conservation a critical area of research in the United States. Before 2006, bats in the United States faced threats such as insecticide use, habitat loss, and climate change. But the emergence of WNS is the most critical threat to bats in the United States today. Bats are vital to ecosystems, playing roles in pollination and pest control (Kunz et al., 2011), and have an estimated economic value of $22.9 billion a year in agriculture alone (Boyles et al., 2011). WNS is caused by a cold-loving fungus, *Pseudogymnoascus destructans* (*Pd*). Bats infected with WNS develop lesions on their wings and a fungal growth in their nose. The fungal infection causes bats to awaken from hibernation early, and thus lose vital fat resources. This ultimately leads to death, with more than 99% mortality rate in some bat species (Frank et al., 2014). More than 6 million bats in eastern and central North America have died as a result of this fungus (USFWS, 2012). The fungus is able to over-summer in caves and therefore can infect the colonies year after year (Lorch et al., 2013). Since its emergence, WNS has been progressing westward and by the end of the 2016-2017 hibernating season was in 31 states and 5 Canadian provinces. The fungus *Pd* (but not WNS) was confirmed in Texas for the first time in 2017 and WNS was confirmed in Oklahoma in 2017, two years after *Pd* was first detected there. With the disease and fungus in two neighboring states, New Mexico is now on the front lines of the advancing disease.

New Mexico is home to a wide diversity of bats, which play many roles ranging from pest control to pollination to tourism. New Mexico is currently *Pd* free (Northup and Buecher, personal communication). However, the climate inside cave hibernacula, cold and moist in winter (Buecher and Northup, 2012), and the species of bats present in the state put the area at high risk of developing WNS. Thus, it is critical that studies of healthy bats be performed now to determine the microbial flora of a healthy bat, which in turn will lead to a better understanding of what natural defenses bats may have. New Mexico is home to species of bat, *Corynorhinus townsendii* (Townsend's big-eared bat), that is closely related to *Corynorhinus townsendii virginianus* (Virginia big-eared bat) that have been shown to be colonized by *Pd*, but do not acquire WNS, such as. It is also home to some species that we suspect will be vulnerable based on their close relationship to especially vulnerable species in the eastern USA, such as *Myotis ciliolabrum* (western small-footed bat), the western analog to the threatened species *Myotis leibii* (eastern small-footed bat) and *Myotis evotis* (long-eared myotis), the western analog to the now newly federally endangered *Myotis septentrionalis* (northern long-eared bat). This makes New Mexico an excellent place to study the microbial flora of bats and to try and understand how the microbes may be aiding bats in combating WNS.

Currently there is no cure or treatment for WNS. Work is being done on finding antimicrobial agents (biocontrols) that could perhaps be used to treat bats before hibernation in order to inhibit the growth of *Pd* during torpor (Hoyt et al, 2015). However, not enough is known about the kinds of antimicrobial agents that may be able to inhibit the fungus, or how such an agent would alter the ecosystem in which the bats live. Moreover, little is known about the role a bat's own natural defenses may play in combating *Pd.*

Starting in 2013, Dr. Northup and her team have been conducting surveys on the microbial flora of bats in New Mexico. Six species of bats, *C. townsendii, Eptesicus fuscus* (big brown bat), *M. ciliolabrum, M. evotis, Myotis velifer* (cave myotis), and *Myotis thysanodes* (fringed myotis), from four locations across the state (El Malpais National Monument (ELMA), Fort Stanton-Snowy River Cave National Conservation Area (FS), Carlsbad Caverns National Park (CC), and Bureau of Land Management High Grassland caves near Roswell, New Mexico (HGL)), were swabbed for both DNA analysis of the microbiota of bats, as well as to culture Actinobacteria living on the bats. Actinobacteria are a group of bacteria known for their ability to produce antimicrobial agents and are often found in high abundance in caves, including those where bats are known to hibernate (Groth and Saiz-Jimenez, 1999; Northup et al., 2011). By culturing this type of bacteria, we are able to determine which, if any, of the bacteria naturally found on bats are able to inhibit the growth of *Pd*. The original cultures were then sub-cultured until a presumptive pure isolate was obtained, uniform in appearance. This culture collection was used to screen isolates for an ability to inhibit *Pd*, as described in Hamm et al. (2017). Any isolate that was able to inhibit the growth of *Pd* will be referred to as an inhibitor.

The next step in determining what natural defenses the bats have to WNS, as well as other bacterial or fungal infections, is to screen these isolates for a group of genes call polyketide synthase type II (PKS II). PKS II is a group of genes that are involved in the production of secondary metabolites such as antimicrobial agents. Using the unique bacterial culture collection from bats of over 3300 isolates from 14 bat species from New Mexico and Arizona, we were able to screen a subset of six bat species from New Mexico. The six New Mexico bat species had a pool of over 900 isolates to choose from. This allowed for the following objectives to be completed: 1) Determine the diversity of PKS II genes living on bats in New Mexico; 2) Determine if there are differences in the diversity these genes among the different species of bats; 3) Determine the diversity of PKS II genes associated with isolates that have been identified as inhibiting *Pd*. Knowing these facts about the ability of the bacteria to make natural defenses for the bats would help conservation efforts on several fronts and could contribute to determining which actinobacterial inhibitor from New Mexico bats would be most effective in combating *Pd*.

Methods, Results, and Conclusions:

Summary of Methods:

Six bat species (*C. townsendii, E. fuscus, M. ciliolabrum, M. evotis, M. velifer,* and *M. thysanodes*) were selected for study (Figure 1). A previous study had swabbed these and other species of bats to generate cultures of Actinobacteria, which are known secondary metabolite producers. Fifty isolates from the culture collection from each bat species were screened using polymerase chain reaction (PCR) to amplify the gene PKS II, thus demonstrating its presence or absence in a given isolate. After the PCR screening was complete, 13 isolates from each of the six species were randomly selected from among the positive isolates and the PSK II gene was sequenced. Sequencing was initially performed at the University of New Mexico's Molecular Biology Facility to determine if the isolates had one or more copies of PKS II. Those isolates that did not return clean sequences in the initial sequencing attempt were considered to have multiple copies and were then cloned, a process allowing for the separation of multiple copies of genes present in a single individual. Sequencing of the clones was performed at Genewiz, Boston, MA. In addition to the 300 isolates from the six study species, 17 additional isolates that were inhibitors were screened for the presence of the PKS II gene. These isolates were from *Tadarida*

brasiliensis (Brazilian free-tailed bat) from New Mexico, and *Antrozous pallidus* (pallid bat), and *M. evotis* from Arizona.

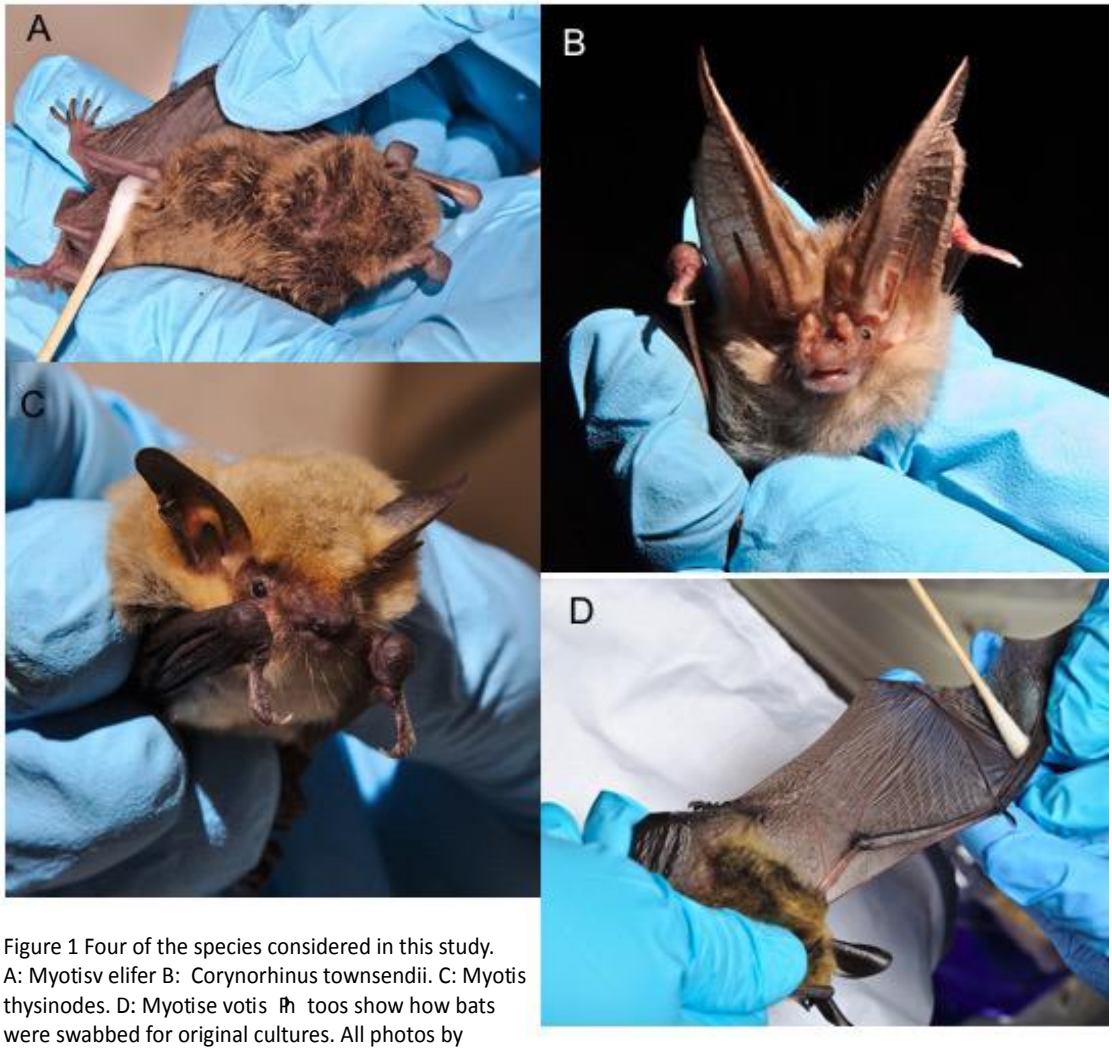

Figure 1 Four of the species considered in this study. A: Myotisv elifer B: Corynorhinus townsendii. C: Myotis thysinodes. D: Myotise votis Ph toos show how bats were swabbed for original cultures. All photos by Kenneth Ingham

Objective 1: Determine the diversity of PKS II genes living on bats in New Mexico:

Fifty isolates were selected for each of six species (*C. townsendii* (COTO), *E. fuscus* (EPFU), *M. ciliolabrum* (MYCI), *M. evotis* (MYEV), *M. velifer* (MYVE), and *M. thysanodes* (MYTH)) from the pool of over 900 isolates. Selection was not entirely random as all isolates that had been shown to be *Pd* inhibitors in a separate study were selected. Additionally, the percentage of bats by geographic area and by sex was determined based on the total culture collection. The selected isolates were checked to make sure they were representative of all areas in the state where the bats were captured. The percentages of the selected isolates for each area and sex were calculated and adjusted as needed to roughly match the percentages of the total culture collection for these factors. Additionally, any isolate that had been shown to inhibit *Pd*

from other species, such as *T. brasiliensis* (TABR), and *A. pallidus* (ANPA), were selected for screening.

PCR with PKS II specific primers were performed on all selected isolates. The results of the PCR screen will be discussed under Objective 2. From the positive isolates, 13 were randomly chosen with the following exception: all inhibitor isolates were selected.
Isolates were first sequenced at the University of New Mexico's Molecular Biology Facility in order to determine the copy number of the PKS II gene. PKS II can have from 1 to 5 copies of the gene, with most having 1 to 3 copies. Any sequence that was noisy, or did not sequence cleanly, was considered to have multiple copies of the gene. Of the 85 isolates sequenced (78 from the six species and seven additional *Pd* inhibitors), 49 had a single copy of the gene and 36 had multiple copies.

The isolates with multiple copies were then cloned with the TOPO TA cloning kit. Molecular cloning is a process that takes individual PCR products and places them in a plasmid (small circle of DNA), which is then placed into an *E. coli* cell. Each *E. coli* only has one plasmid, and each plasmid only has one copy of the PSK II gene. The *E. coli* then replicates through cellular division, allowing for many copies of an individual PSK II gene to be generated, from which sequencing can be performed. From each isolate cloning reaction, 12 clones were sequenced, thus giving a representative sample of the genes in the isolates. Clones were sequenced at Genewiz.

Sequences were edited in Sequencher v 5.1 for quality. Clones were grouped at 97% similarity (any sequence that matched at 97% was put into a group and the consensus sequences were used for further analysis). These grouping are called Operational Taxonomic Units (OTUs). There were 127 OTUs generated from the sequencing efforts.

Nucleotide sequences were then compared using the GENBANK BLAST (https://blast.ncbi.nlm.nih.gov/) tool to determine closest taxonomy. BLAST is a tool to compare a sequences to all the known published sequences based on sequence similarity. The cultured and uncultured BLAST hit to each sequence was recorded for future comparisons, resulting in 197 different BLAST hit comparison sequences. The gene copy numbers, represented by the number of OTUs per isolate, are summarized in Figure 2. As previously indicated, a majority (69%) of the isolates had only one copy of the gene, while 19% had two copies and 12% had three copies. Good's coverage (an indication of what percentage of the total species in a sample has been represented) ranged from 0.75 to 0.92, indicating that we have probably captured the diversity in our samples.

When nucleotide sequences were compared to the 197 BLAST hits, there were 58 unique hits that were the top matches. When the nucleotide sequences were compared to the DoBISCUIT database of secondary metabolites (http://www.bio.nite.go.jp/pks/), 21 different metabolite products were identified.

Nucleotide sequences were then translated into protein sequences using Prodigal (Hyatt et al., 2010) and then hand checked to ensure proper frame translations. A group of three DNA nucleotides, called codons, encode for each of the 26 amino acids. The nucleotide sequences can be read in three different frames, depending on if you start with the first second or third nucleotide, and each frame will result in a different protein translation. In order to ensure we were translating in the proper frame, each nucleotide translation was checked against reference protein sequences to confirm conserved areas were correctly translated.

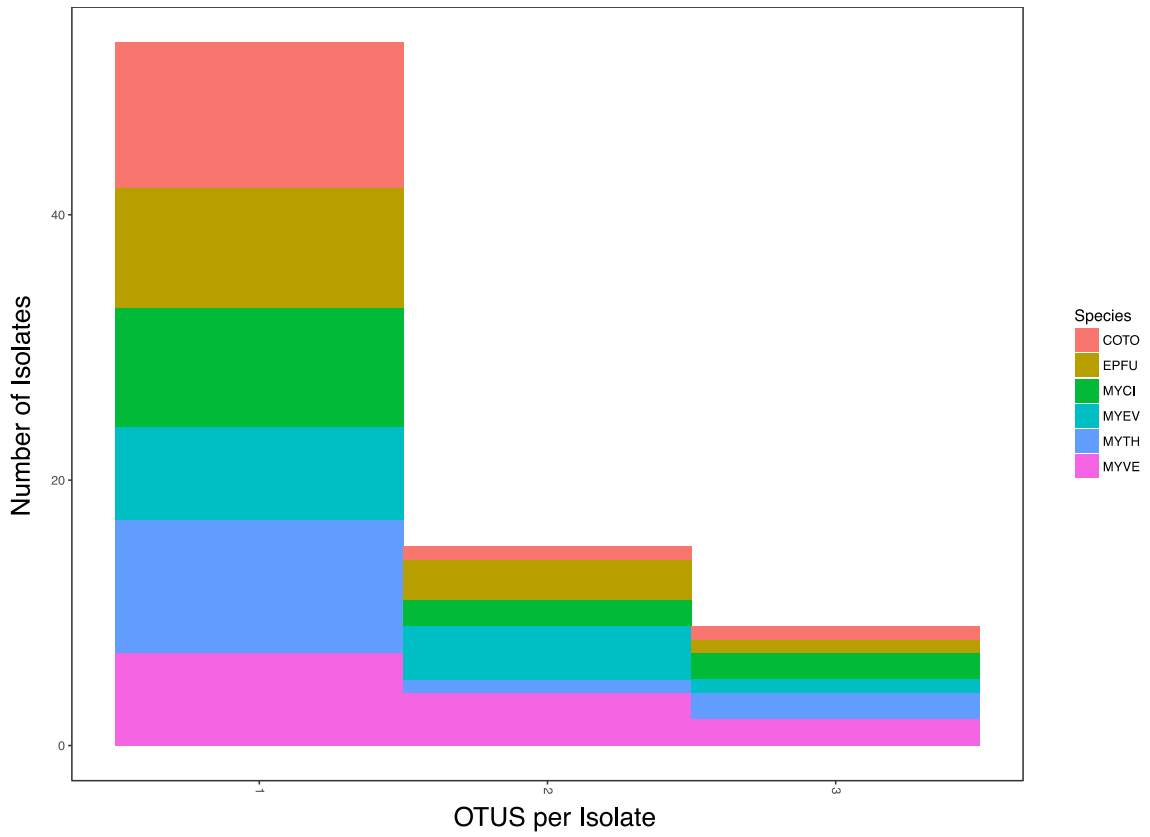

Figure 2: The number of copies of the PKS II gene, represented by Operational Taxonomic Units (OTUs) grouped at 97% sequence similarity, are shown and are color-coded by species.

We used a reference set of 120 protein PKS II sequences, which included 82 representative sequences for known PKS II secondary metabolite products, as well as 38 from environmental sequences downloaded from BLAST hits. Protein sequences were aligned using Clustal Omega (Sievers et al., 2011). Aligned sequences were trimmed using BioEdit.

Identity matrices were made for just experimental sequences (sequences generated in this study) and for experimental sequences compared to reference sequences of known PKS II producing bacteria. Identity matrices use aligned sequences to determine how many positional differences a sequence has from any other sequence in the dataset. This allows for determination of how different or similar experimental data are to other sequences. Histograms of the top percent identification are shown in Figures 3 and 4. When compared to other sequences from this study, we find that a majority of the sequences are producing compounds similar to others found in this dataset. However, there are 14 sequences have less than 85% sequence similarity. These represent the most novel or different compounds within our dataset.

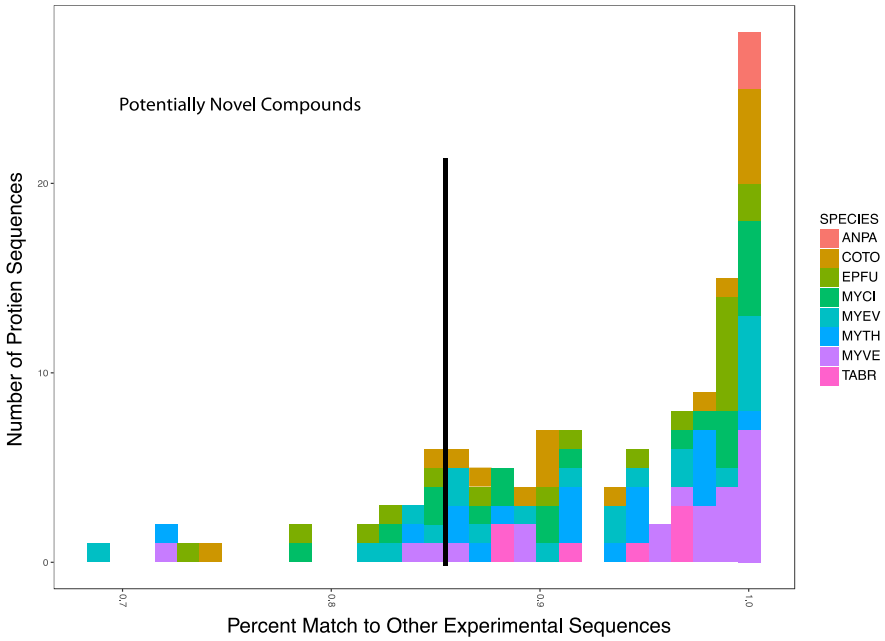

Figure 3. Percent match to other experimental sequences, colored by species. Percent match is based off of an identity matrix made from aligned protein sequences. Black line indicates 85% sequence similarity.

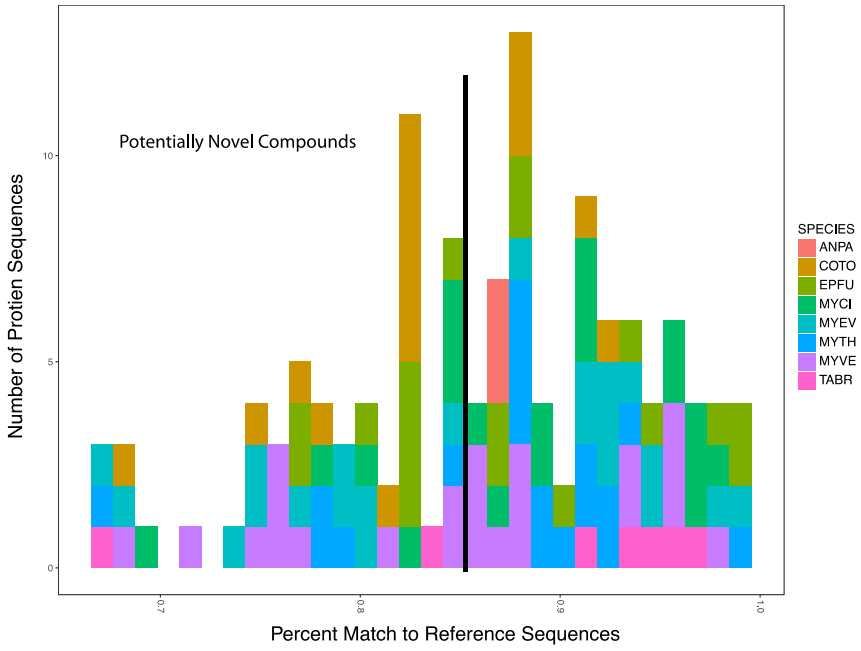

Figure 4. Histogram of top percent match to reference sequences based on aligned protein sequences. Black line indicates 85% sequence similarity.

When analyzing PKS II protein sequences, sequences that have less than 85% sequence identity with known sequences are considered to be novel genes. In other words, any sequences that shares less than 85% similarity with known sequences may represent a new gene/protein. In our dataset, we have 49 of 127 sequences (39%) that share less than 85% identity to the reference sequences (Figure 4). This large group of novel proteins should be investigated further to determine if they are in fact producing novel compounds. Because of the growing number of pathogens that show high levels of antibiotic resistance, new compounds with novel mechanisms of killing pathogens are critically important. Our novel proteins are worth investigating further to determine if they have any novel mechanisms, as well as novel compounds.

To better determine the phylogenetic relationship among the 127 experimental protein sequences and the reference set, a phylogenetic tree was made. A maximum likelihood tree was made using IQ tree with the default setting of auto detection for substitution models. The resulting tree was then annotated using the Interactive Tree of Life. The tree (Figure 5) shows a number of interesting points. First there are several areas of the tree that have no experimental sequences. This indicates that we did not find the potential to make the compounds made by bacteria on these parts of the tree on the bats. This is not unusual as not all PKS II genes will be found in every environment. What is more notable is that, based on the relationship of sequences on the tree, there is the potential to produce 23 known compounds, which is about 1/3 of all the compounds currently known to be produced by PKS II genes. There are several clades where the closest previously identified sequence is also unknown, and therefore the potential secondary metabolite being produced is still unknown.

A second observation is the inhibitors (indicated by red stars or dots on the tree at the base of sequence names, Figure 5) are spread around the tree. However, there are clades that are just inhibitors. One such case of note is the clade at the top of the tree (Figure 5, 6), indicated by the green triangle, where there are four inhibitors from four different isolates from three caves all in a single clade, along with one sequence that is not an inhibitor. Each of these bats is from ELMA, and are all COTO. The isolates in this clade deserve further exploration to see if they are in fact producing Oxytetracycline (the broad spectrum antibiotic produced by their closest cultured relative on the tree), or if they are producing a novel compound. Since these sequences share less than 85% identity with any known sequence (those indicated in light blue on the tree), it is likely that they are producing a novel compound. Moreover, Hamm et al (2017) showed these isolates to be a new species of *Streptomyces*. Characterizing the compound produced by these isolates might reveal a new type of secondary metabolite product.

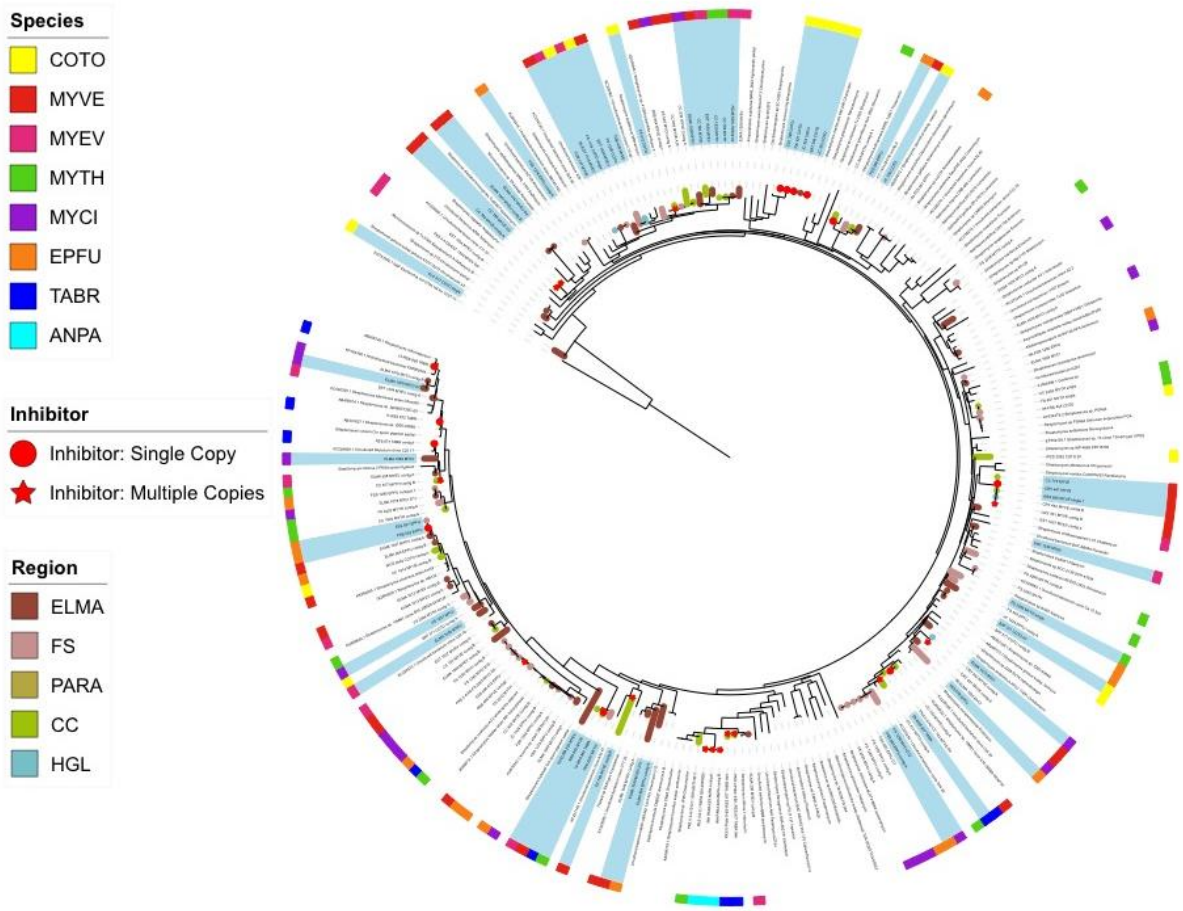

Figure 5. Maximum likelihood tree of PKS II gene protein sequences. Species are indicated by color on the outer ring. Inhibitors are indicated by red circles (single copy of gene) or red stars (multiple copies of gene from one isolate). Geographic regions are colored on the branches of the tree. Those sequences that share less than 85% similarity to the reference sequences (i.e., potentially novel sequences) are shaded in light blue. Green triangle indicates a clade of interest, and is highlighted in Figure 6.

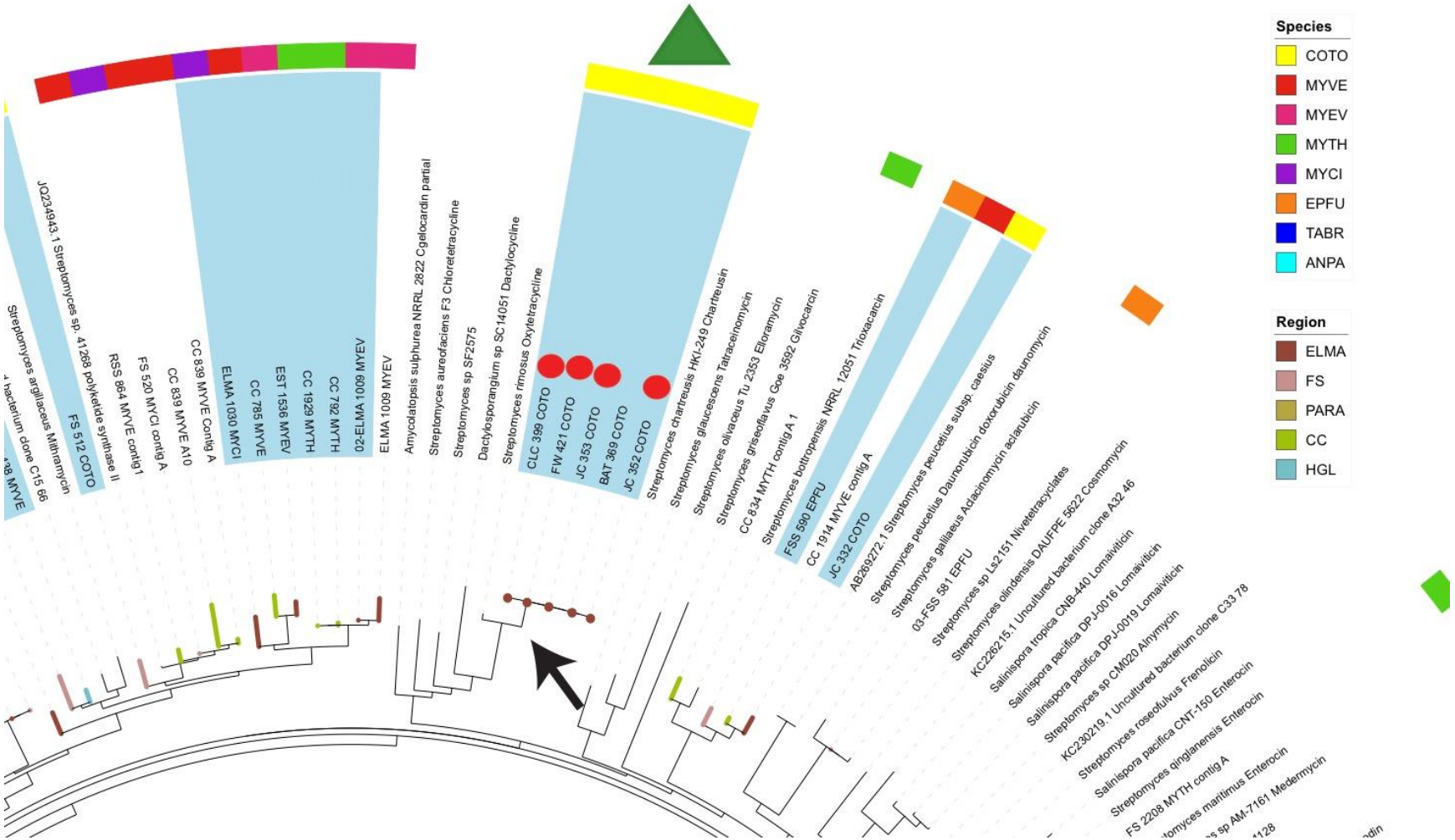

Figure 6: Section of maximum likelihood tree (Figure 5) showing a clade of isolates all from COTO from ELMA from three different caves that share less than 85% sequence similarity with the known *Streptomyces ramous*, which produces oxytetracycline.

It is important to note that the presence of PKS II genes in an isolate does not necessarily mean that these the PSK II genes are making the antifungal that inhibits *Pd*. There are several other gene clusters, such as PKS I and NRPS (nonribosomal peptide synthase), which could be producing the antifungal compound. However, further investigation is warranted to see if any of the inhibitors are producing novel compounds. This will be discussed further under Objective 3.

Third, there is a large section of the tree, to the left hand side, where there are many experimental sequences, and very few reference sequences. Most of the reference sequences in this part of the tree are also from unknown bacteria, or bacteria whose compounds have not been identified. The identified compounds in this part of the tree are spore pigments. Spore pigments have a variety of roles, from coloration, to antibiotics, to spore formation. Without further investigation, it is difficult to assign functions to these sequences.

Fourth, the sequences that are potentially producing novel compounds (i.e., those sharing less than 85% similarity to any known protein sequence) are spread throughout the tree. This indicates that the potential for new secondary metabolite compounds is widespread, and not just concentrated in one type of compound. This is unsurprising as we are finding many new species of *Streptomyces* in the culture collection (Hamm et al. 2017)

The protein sequences were clustered using CD-HIT (Ying et al, 2010) at 85% similarity in order to determine how many unique protein sequences were in the dataset. The program automatically picks a representative sequence and then calculates the percent match to that representative sequence for each cluster. Thirty-five clusters were formed (Supplemental Table 1). There were 14 clusters that only contained one sequence (singleton) and five that contained two sequences (doubletons). Of the singletons, two were from CC, five were from FS and seven were from ELMA. Two were COTO, four were EPFU, two were MYCI, two were MYEV, three were MYTH, and one was MYVE. The 14 singletons matched to 12 different reference sequences with a percent match identity of between 67 and 98%. This indicates that the singletons represent a group of secondary metabolite compounds, with the potential for novel compounds.

Of the doubletons, three were from the same isolate, indicating that the sequences had some divergence, but they still encoded the same protein sequence. The other two both had one sequence from CC and one from ELMA, but differed in the species, one doubleton had a MYTH and a COTO and the other had MYVE and MYEV.

The other 16 clusters contained three or more sequences. The cluster with the most sequences, cluster 13, had 15 sequences. Two of the largest clusters, clusters 3 and 4, both grouped within clades on the tree belonging to spore forming pigments.

Conclusions from Objective1:

The overall diversity of PKS II genes from bats was extensive. There is the potential to produce at least 23 different known compounds, and many additional unknown compounds. The tree gives a good indication of what isolates might be producing the most novel compounds and future studies should look at these isolates for potential novel antimicrobials.

Objective 2: Determine if there are differences in the diversity of these genes among the different species of bats:

To determine if there are any differences among the bat species in relation to the diversity of PKS II genes and gene products, we looked at a number of factors. First, we looked at the number of positive isolates from each species as a result of screening 50 isolates per species (Figure 7). EPFU had the largest number of positive isolates at 30, while MYCI had the smallest at 19. We compared the number of positive isolates to the number of bats the isolates represented to determine if this was an influencing factor. Although there is a slight positive correlation between

the number of bats the isolates represent and the number of positive isolates, it was not statically significant (Adjusted R-squared: 0.08932, p-value: 0.2632) (Figure 8).

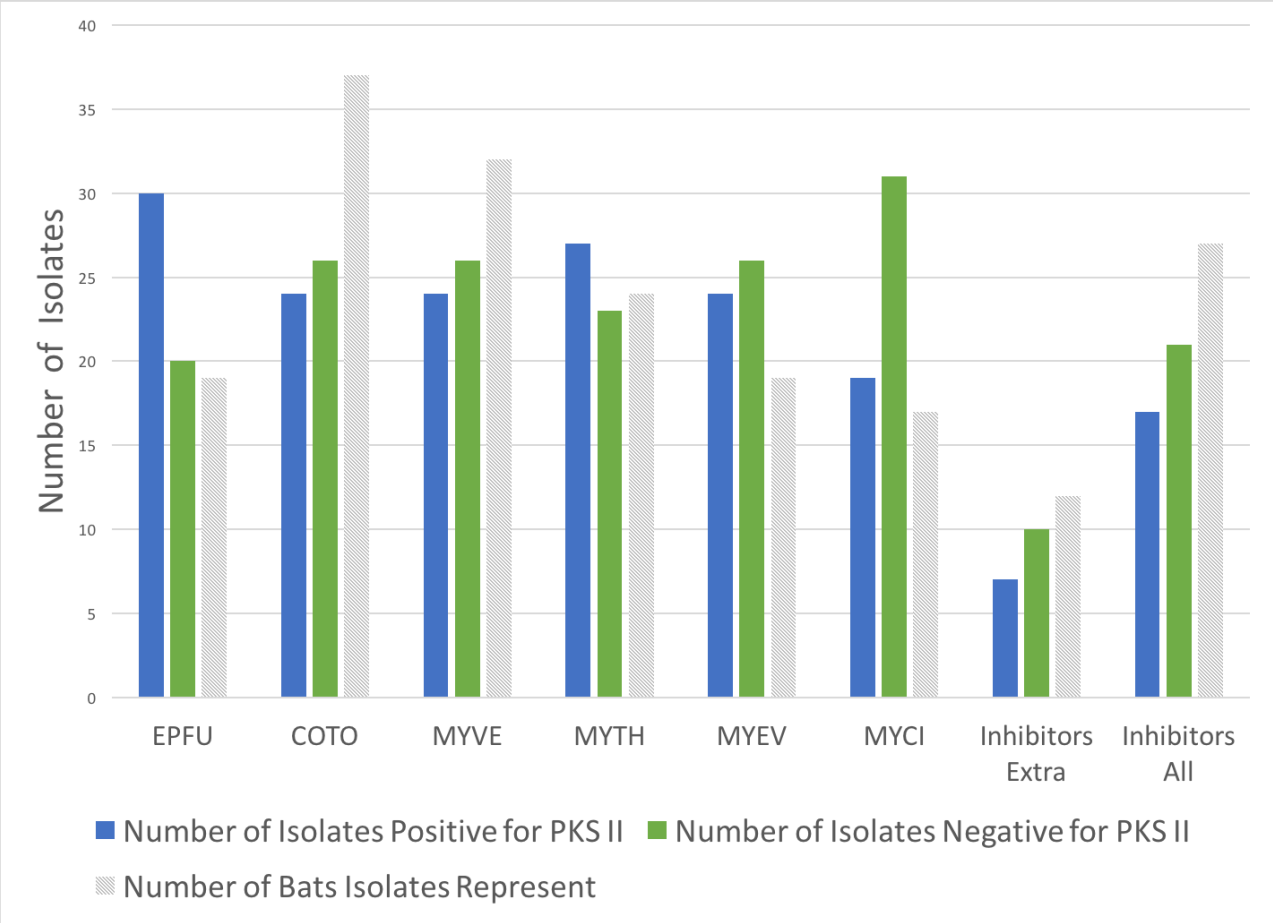

Figure 7: Number of isolates positive and negative for PKS II from PCR screen. Also shown is the number of bats the isolates came from. Additionally, the positive and negative isolates for the inhibitors from species not previously screened for PSK II (Inhibitor Extra) and from all the inhibitors screened regardless of species or location of capture (Inhibitors All) are shown.

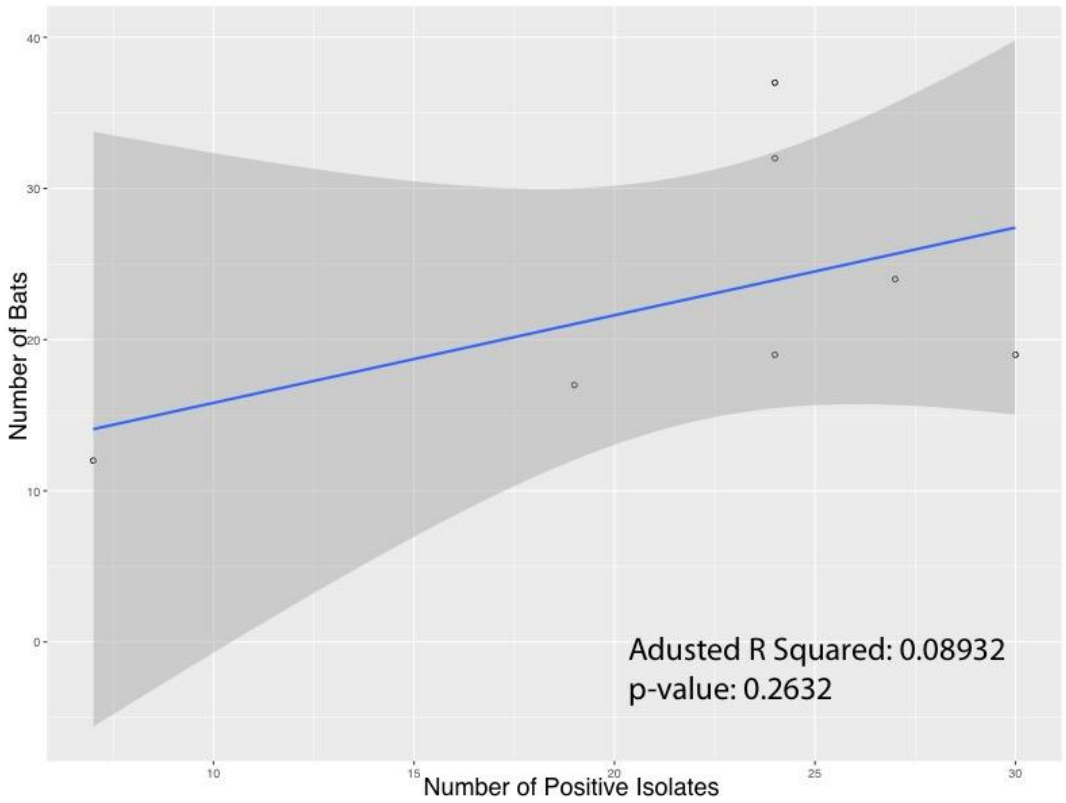

Figure 8. Scatter plot showing the number of positive isolates per species of bat vs. the number of bats those isolates represent.

These results indicate that there are differences among bat species in the number of isolates that have the potential to make secondary metabolites. However, since each isolate can have multiple copies of the gene, further analysis is necessary.

We then looked at the number of copies of the gene per isolate by species for nucleotide sequences (Figure 9). We see that the mean of each species is different, with COTO being the lowest at 1.307 genes per isolate and MYEV being the highest at 1.692 gene per isolate. An ANOVA was performed in which we could not reject the null hypothesis that the averages were equal (F= 0.4759, p-value 0.7931). While this indicates that we did not find statistically significant differences in the numbers of copies of PKSII genes among the species, this result has to be taken with caution, as the data were not normally distributed. A generalize linear model does not require data to be normally distributed. When species was modeled as a predictor of gene copy number, we found species was a weak predictor of gene copy number (Supplemental Table 2). COTO were the most predictive species, with a negative influence on gene copy number. EPFU had the most positive influence on gene copy number. If there are more copies of the gene in a bacterium, they would have a greater potential to produce different types of antimicrobials.

When nucleotide sequences were compared to the database DoBuscuit, we saw that the number of unique, known compounds the sequences matched to ranged from 7 (COTO) to 11 (EPFU). MYTH and MYEV each had 9 unique matches, while MYCI and MYVE had 8.

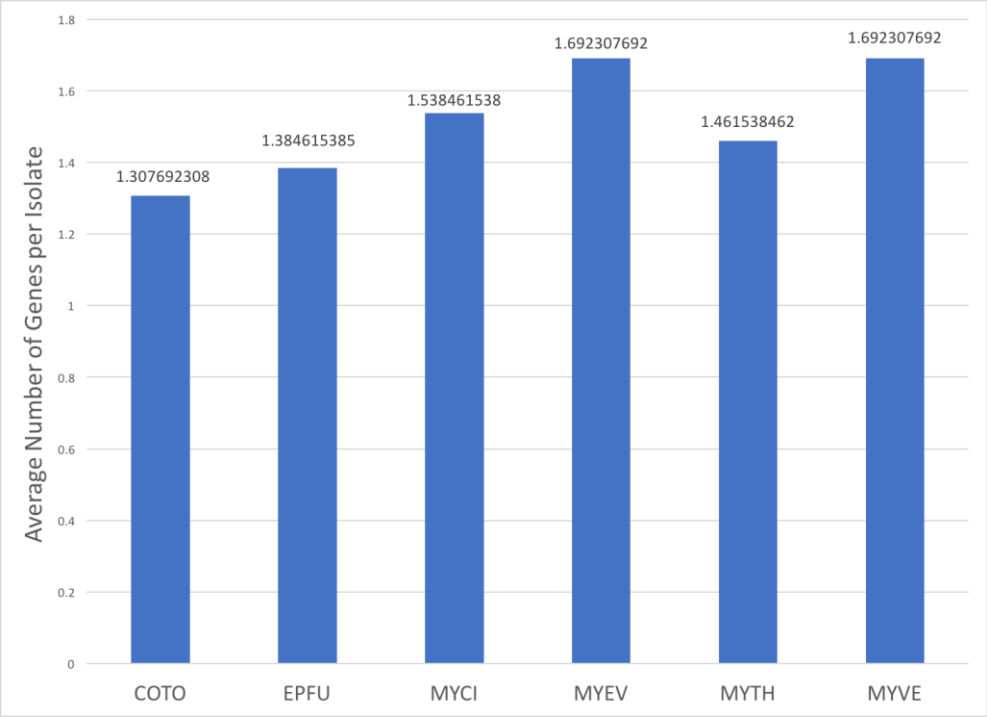

Figure 9: Mean of number of genes per isolate by species for nucleotide sequences with a similarity cut off at 97% similarity.

Using the protein translations, we saw that MYTH had the fewest novel sequences (those with less than 85% similarity to known sequences) at 4, while COTO had the most at 11 (Figure 10). This indicates that MYTH and MYCI may have fewer novel secondary metabolites, while COTO may have the potential for more novel compounds and should be seen as a possible source of potentially novel antimicrobials.

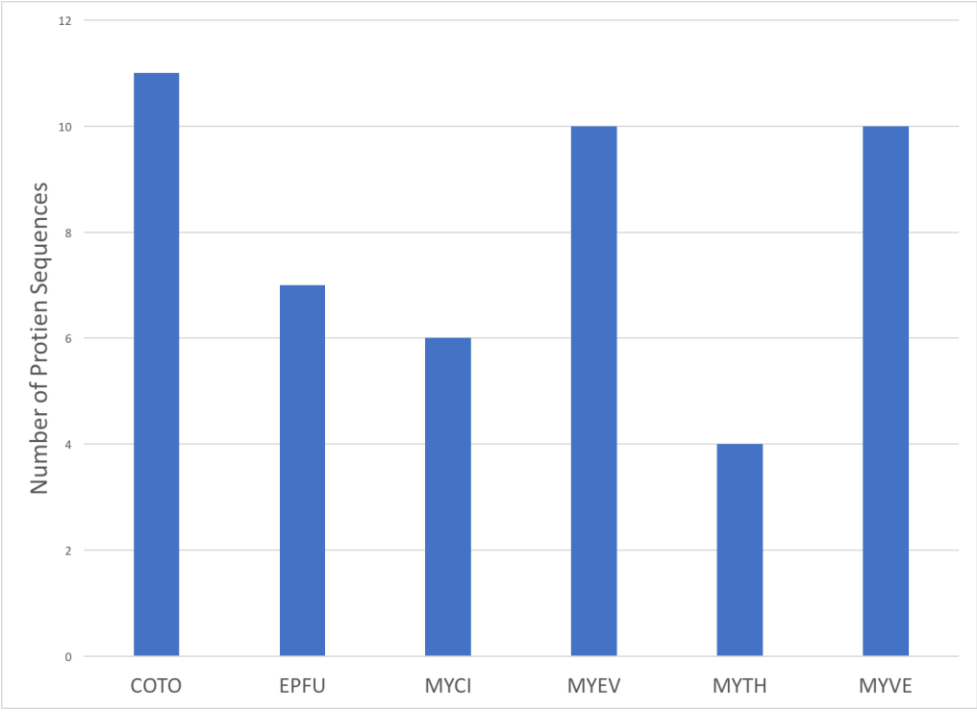

Figure 10. Number of protein sequences per species with less than 85% similarity to reference sequences.

Aligned sequences from the six bat species were compared to a reference set of known PKS II genes, plus unique hits from BLAST based on nucleotide searches. A sequence identity matrix was created in BioEdit, and the top hit and percent match were recorded for each experimental sequence. COTO had the fewest number of matches to unique reference sequences, with 10 of the 16 (62.5%) sequences being unique and the other 6 being duplicate sequences. EPFU had 13 of 18 (72.2%) unique matches, MYCI had 14 of 20 (70%), MYEV had 15 of 22 (68.2%), MYTH had 14 of 17 (72.2%), and MYVE had 14 of 22 (63.6%). The clade on the tree (green triangle, Figure 5), with five highly similar sequences, is all from COTO. These sequences, where four of the five are from isolates that can inhibit *Pd*, make up 31.25% of the COTO sequences in this study. COTO is one of the most common species represented in the culture collection, comprising almost 25% of the total isolates in the culture collection. We were limited in the number of isolates we were able to sequence from each species. It is possible if we sequenced more isolates from COTO, or more isolates from a single COTO bat, that greater diversity would be seen. However, in a parallel study of bats in Arizona, we found that only 17 of 50 COTO bats were positive for PKS II genes. Given the results of that study, we have greater confidence in the conclusions of this study. Overall, we saw that although COTO had mostly single copies of the PKS II gene, and low overall diversity, the genes of COTO were much more novel than those of other species.

When the protein sequences were clustered, the 16 sequences from COTO clustered into 8 separate groups, the 18 sequences from EPFU clustered into 12 groups, the 20 sequences from MYCI into 9 groups, the 22 sequences from MYEV into 13 groups, the 17 sequences from MYTH into 13 groups, and the 22 sequences from MYVE into 14 groups. The small number of clusters for MYCI, despite having a larger number of sequences, indicates that the diversity of genes on MYCI is lower than other species. MYVE had both a large number of sequences and the largest of clusters, perhaps indicating that this bat species has a greater potential to produce secondary metabolites.

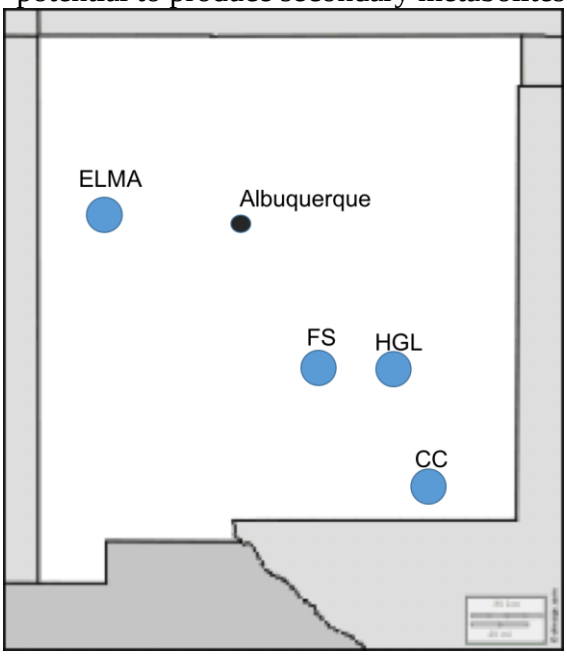

Figure 11. Map of New Mexico showing locations where bats were swabbed (ELMA, FS, HGL, and CC).

We also looked at regional difference. In New Mexico, we had four regions in which the bats were captured: Carlsbad Caverns National Park (CC), El Malpais National Monument (ELMA), Fort Stanton Cave Area (FS) and BLM High Grassland caves (HGL) (Figure 11). We were interested in finding out if any of the regions had lower diversity of PKS II genes. This might indicate a weakness in defenses against WNS when it enters the state. Any regions with fewer PSK II genes could be more vulnerable to the disease.

Most of the regions had between 44-53% of the isolates sequenced being positive for the PKS II gene, with the exception of HGL, which had 70% of isolates positive for the gene. However, this region is a bit of an outlier, as it only had 10 isolates from a single bat species (MYVE) included in the study. All the bats from this region are cave-caught, and these bats having been shown to have a significantly different microbiota associated with them, even when compared to bats of the same species caught in CC (Winter et al., 2017). Perhaps if we had sequenced more isolates from this region, the number would have been closer to 50%.

Looking at regional difference in the diversity of the protein sequences found, ELMA had the most novel sequences with 49% of the sequences sharing less than 85% with the reference set of 120 sequences (representative of the known compounds and environmental samples, Figure 12), while the other three regions ranged from 32 to 38%. We have found the microbiota of bats at ELMA to also be very diverse and unique (Winter et al., 2017; unpublished data).

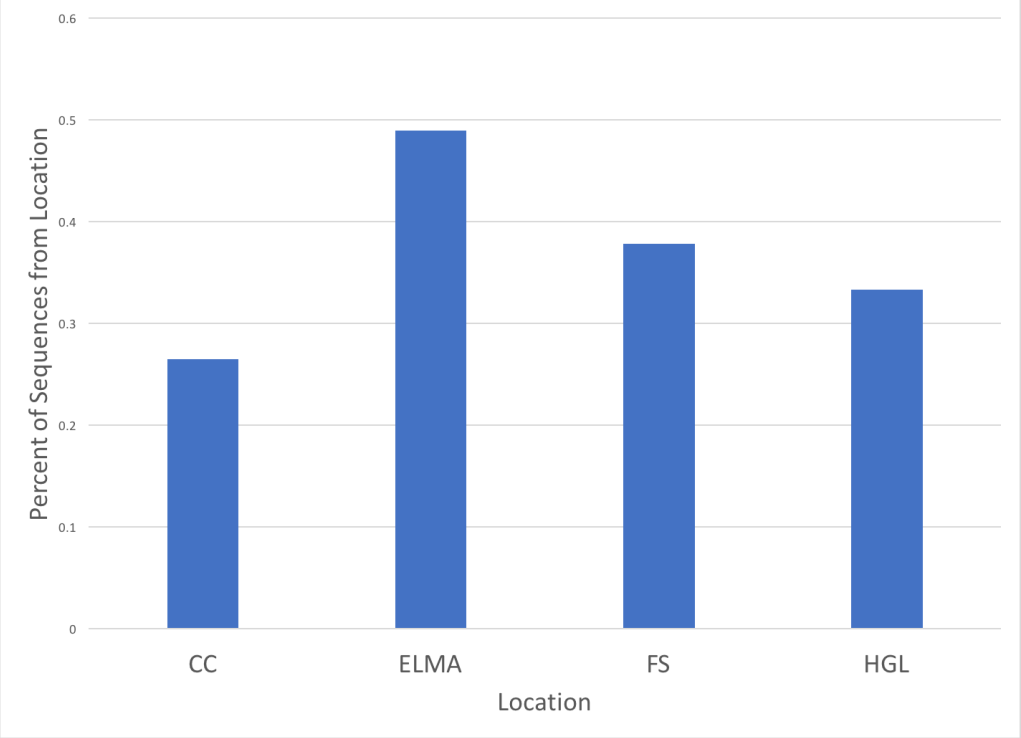

Figure 12: Percent of protein sequences, based on number of sequences per region, that share less than 85% sequence similarity to reference set of 120 known PKS II protein sequences.

Conclusions of Objective 2:

Overall, we saw differences in the diversity of the PKS II genes on bats of different species. COTO bats more often had single copies of the gene, and several of these have the

potential to produce novel compounds. A group of five sequences that all came from COTO bats from ELMA is driving some of results we are seeing. These five sequences all share less than 85% similarity with the reference sequences. MYCI has the fewest number of positive isolates and one of the smaller numbers of protein sequences clusters, indicating that it may have a lower potential to produce a diversity of secondary metabolites. ELMA is the region where there is the most potential to find novel secondary metabolite compounds based on the sequence similarity with known protein sequences.

Objective 3: Determine the diversity of PKS II genes associated with isolates that have been identified as inhibiting *Pd*.

Within our culture collection, we have identified over 45 isolates to date that are able to inhibit the growth of Pd. In order to better understand the potential of these inhibitors, and to help inform decisions on which to further investigate, a total of 39 inhibitors were screened for the PKS II gene. Twenty-two of the inhibitors were found in the six bat species in this study. An additional eight were from TABR from CC, one from ANPA from Arizona, and eight from other bats swabbed in Arizona, mostly MYEV. Of the inhibitors screened, 17 were positive for PSK II, representing 43.5% of the inhibitors being positive for the gene. This was comparable to the percent positive in the overall population of isolates at 49%. There was a group of isolates from bats from Arizona that had PCR products of the wrong size in the PCR screen. When sequenced, the products returned fatty acid synthase genes, a group of related, but functionally divergent, genes that are commonly used as an outgroup when looking at PKS II. These isolates were not investigated further.

There were 26 unique nucleotide sequences from the 17 inhibitor isolates that were PKS II positive at 97% similarity. When looking at the cluster analysis previously described in Objective 1, the inhibitor sequences clustered into 15 clusters using translated proteins at 85% similarity (Supplemental Table 1). Only one cluster was a singleton, and one was a doubleton. Six clusters had only one inhibitor, but other sequences in the cluster were not inhibitors, while three had only inhibitors in the cluster, and six had multiple inhibitors and other sequences. From the tree (Figure 5), we can see that the inhibitors (red dots and stars) are spread out across the different classes of PKS II compounds. This indicates that the PKS II genes in the inhibitors represent a variety of classes of secondary metabolites. The protein sequences matched to the following known compounds: Kinamycin, Mithramycin, Kisinostatin, Pradimicin, X26, Urdamycin, Nogalamycin, Oxytetracycline, Griseorhodon, Trioxcarcin, PGA64, Rubromycin, and three separate spore pigments.

When the sequences were looked at for novelty, 42% of the inhibitors shared less than 85% similarity to known sequences (Figure 13).

Conclusions for Objective 3:

The inhibitors share many of the same trends as the isolates do as a whole. The sequences are spread out across the tree and match to 15 different known PKS II products. Those inhibitors that have less than 85% identity to known sequences, and do not match to spore pigment compounds, should be further characterized. The clade of four inhibitor isolates from COTOs from ELMA should also be investigated further.

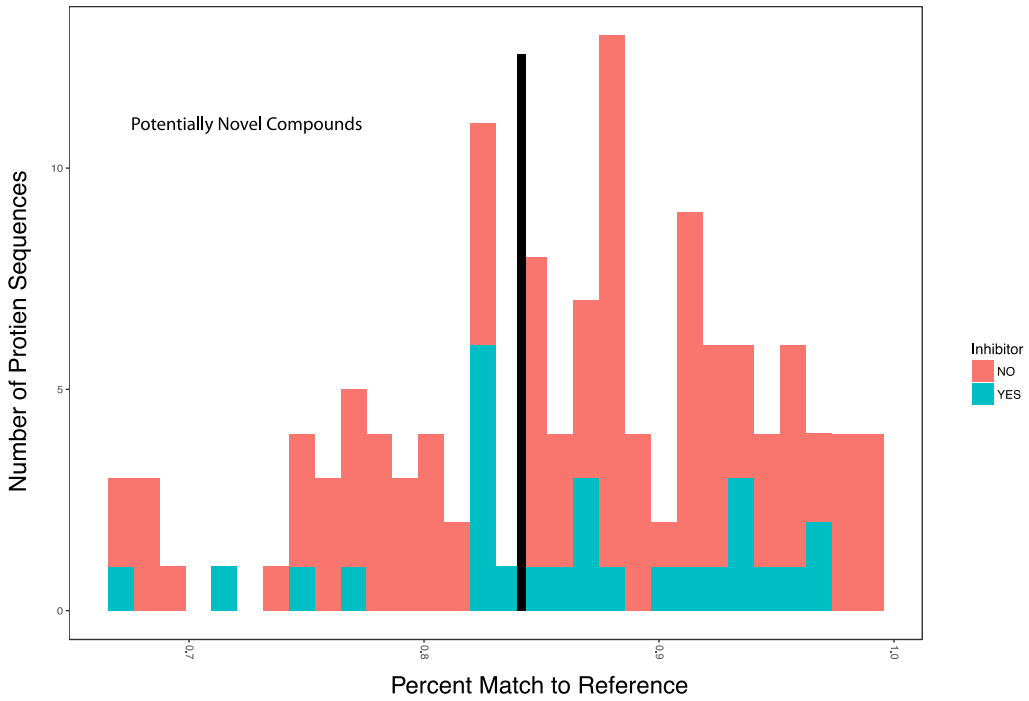

Figure 13: Percent match to Reference sequences based on aligned protein sequences. Inhibitors are indicated in turquoise. Black line indicates 85% sequence similarity.

Implications for Management

This study has several implications for management.

We have shown there are differences in the potential to produce, and the types of, secondary metabolite compounds among the bats species in New Mexico. Specifically, MYCI appears to be vulnerable to disease based on the lower number of positive isolates, and the lower number of isolates with the potential to produce novel compounds. We recommend closer monitoring of this species for WNS in New Mexico. MYTH also had a low potential of producing novel compounds and warrants close monitoring.

Carlsbad Caverns may be a possible route of entry for WNS in New Mexico given that it has a lower number of isolates producing novel compounds. Many of the inhibitors in this study are from Carlsbad, but many of those were isolated from TABR bats, which do not hibernate overwinter. These bats could be bringing different bacteria able to inhibit *Pd* from their winter roosts on their skin microbiota.

We also found that the isolates from bats captured in ELMA were very diverse. ELMA was also the region with the most sequences with the potential to produce novel compounds. It is also where the cluster of novel inhibitors (green triangle Figure 5) were from. ELMA should be considered a location for a source of novel antimicrobial compounds, and the caves and bats protected in such a manner to preserve this resource.

References:

Boyles, JG, Cryan, PM, McCracken, GF, and Kunz, TH. 2011. Economic Importance of Bats in Agriculture. *Science.* 332(6025): 41-42.

Buecher, D.C. and Northup, D.E. 2012. Characterization of roost microclimate in bat hibernacula in the lava tubes of El Malpais National Monument with microbial sampling for the presence of Geomyces destructans (White Nose Syndrome) on bats. Unpublished report to WNPA. 10pp.

Frank CL, Michalski A, McDonough AA, Rahimian M, Rudd RJ, et al. 2014. The Resistance of a North American Bat Species (*Eptesicus fuscus*) to White-Nose Syndrome (WNS). *PLoS ONE.* 9(12): e113958. doi:10.1371/journal.pone.0113958

Hamm, P.S., Caimi, N.A., Northup, D.E., Valdez, E.W., Buecher, D.C., Dunlap, C.A., Labeda, D.P., Lueschow, S. and Porras-Alfaro, A., 2017. Western Bats as a Reservoir of Novel Streptomyces Species with Antifungal Activity. *Applied and Environmental Microbiology*, *83*(5), pp.e03057-16.

Hoyt JR, Cheng TL, Langwig KE, Hee MM, Frick WF, Kilpatrick AM. 2015. Bacteria Isolated from Bats Inhibit the Growth of *Pseudogymnoascus destructans*, the Causative Agent of White-Nose Syndrome. *PLoS ONE*. 10(4): e0121329. doi:10.1371/ journal.pone.0121329

Groth, I., and Saiz-Jimenez, C. (1999) Actinomycetes in hypogean environments. Geomicrobiol J 16: 1–8.

Kunz, TH, Braun de Torrez, E, Bauer, D, Lobova, T, and Fleming TH. 2011. Ecosystem services provided by bats. *Ann. N.Y. Acad. Sci.* 1223 : 1–38.

Lorch JM, Muller LK, Russel RE, O'Connor M, Lindner DL, et al. (2013). Distribution and environmental persistence of the causative agent of white-nose syndrome, Geomyces destructans, in bat hibernacula of the Eastern United States. *Appl. Environ. Microbiol.* 74: 1293–1301.

Northup, DE, Melim, LA, Spilde, MN, Hathaway, JJM, Garcia, MG, Moya, M, Stone, FD, Boston, PJ, Dapkevicius, MENL, Riquelme, C. (2011) Lava cave microbial communities within mats and secondary mineral deposits: Implications for life detection on other planets. *Astrobiology*. 11(7): 601-618.

Ridley, C.P., Lee, H.Y. and Khosla, C., 2008. Evolution of polyketide synthases in bacteria. *Proceedings of the National Academy of Sciences*, *105*(12), pp.4595-4600.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J. and Thompson, J.D., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, *7*(1), p.539.

USFWS, 2012. North American bat death toll exceeds 5.5 million from white-nose syndrome. Press Release. 17 January 2012. Available: http://www.batcon.org/pdfs/USFWS_WNS_Mortality_2012_NR_FINAL.pdf.

Winter, A.S., Kimble, J.C., Young, J.M., Buecher, D.C., Valdez, E.W., Hathaway, J.J., Porras-Alfaro, A., Read, K.J. and Northup, D.E., 2016. External bacterial diversity on bats in the southwestern United States: Changes in bacterial community structure above and below ground. *PeerJ Preprints*, *4*, p.e2493v1.

Ying H, Beifang N, Ying G, Limin F and Weizhong L., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, 26:680

Supplemental Table 1. Clusters of protein sequences based on 85% similarity. Match % is the percent similarity to the representative (REP) sequence in each cluster. Inhibitor column indicates that the sequence is an inhibitor.

Cluster Name	Sequences in Cluster	Match %	Inhibitor
>Cluster 0			
0	193aa, >INH_PARA_629_ANPA_contig_A	REP	YES
1	193aa, >INH_PARA_629_ANPA_contig_B	100%	YES
2	193aa, >PKS_II_8C1PARA_629_ANPAD6	100%	YES
>Cluster 1			
0	193aa, >PKS_II_4A10CC_1920_MYTH_B11	REP	
>Cluster 2			
0	192aa, >ELMA_1049_MYEV_contig_B	REP	
1	192aa, >ELMA_1049_MYEV_D11	86%	
2	192aa, >EST_1537_MYEV_contig_A_	85%	
3	192aa, >CC_763_MYVE_contig_B	85%	
4	192aa, >CC_817_MYTH_contig_B_	85%	YES
>Cluster 3			
0	192aa, >BAT_371_COTO_contig_A_		
1	192aa, >ELMA_1013__MYEV_contig_A_	86%	
2	192aa, >ELMA_1013__MYEV_contig_B_	86%	
3	192aa, >WCS_2062_COTO_contig_A_	85%	
4	192aa, >CC_1914_MYVE_contig_B	85%	
5	192aa, >ELMA_1047_MYEV_contig_A	86%	
6	192aa, >ELMA_964_EPFU_contig_B_	86%	
7	171aa, >FSS_552_EPFU_	85%	YES
8	171aa, >FSS_557_EPFU	86%	
9	170aa, >ELMA_1605_MYEV_	89%	

10	192aa, >FS_2208_MYTH_contig_C_	87%	
11	170aa, >FS_1251_MYCI_	89%	
12	175aa, >ELMA_1082_MYCI_	85%	
13	192aa, >ELMA_208_MYEV_contig_A	87%	
>Cluster 4			
0	192aa, >ELMA_1079_MYCI_contig_B_	REP	
1	151aa, >ELMA_1079_MYCI_A12_	98%	
2	180aa, >11RSS_872_TABR	87%	YES
3	192aa, >EST_1534_MYEV_contig_A	91%	
4	183aa, >13RSS_880_TABR	89%	YES
5	192aa, >ELMA_1080_MYCI_contig1	85%	
6	192aa, >PKS_II_4H6FS_2283_MYCI_D5	85%	
7	192aa, >FS_1249_MYCI_B10	85%	
8	192aa, >FS_1250_MYCI_contig_B	85%	
9	192aa, >FSS_1284_EPFU_contig_A	85%	
10	192aa, >FSS_1278_EPFU_contig_B_	86%	
11	192aa, >CC_839_MYVE_Contig_B_	86%	
>Cluster 5			
0	192aa, >FSS_1285_EPFU_contig_A_1	REP	
1	192aa, >ELMA_1078_MYCI_D11	100%	
2	192aa, >FS_2223_MYTH_contig_A	94%	
3	192aa, >CC_1920_MYTH_contig_A	93%	
>Cluster 6			
0	165aa, >RSS_864_MYVE_contig2	96%	YES
1	170aa, >FSS_600_A12_EPFU	97%	
2	175aa, >FS_2233_MYTH_	97%	
3	192aa, >JC_1424_EPFU_contig_B	REP	
4	192aa, >RSS_877_TABR_contig1	87%	YES
>Cluster 7			
0	192aa, >FS_2208_MYTH_contig_A	REP	
>Cluster 8			
0	192aa, >CC_789_MYVE_contig_B	REP	YES
>Cluster 9			
0	192aa, >ELMA_1049_MYEV_contig_A	REP	
>Cluster 10			
0	192aa, >ELMA_964_EPFU_contig_A	REP	
>Cluster 11			
0	192aa, >CC_834_MYTH_contig_A_1	REP	YES
1	171aa, >FSS_590_EPFU_	85%	

2	192aa, >CC_1914_MYVE_contig_A	85%	
>Cluster 12			
0	170aa, >RSS_856_MYVE_single_1		
1	192aa, >CRY_451_MYVE_contig_B	REP	YES
2	192aa, >CRY_442_MYVE_contig_B	100%	
3	174aa, >CRY_441_MYVE	99%	YES
4	171aa, >CC_773_MYVE_	99%	
>Cluster 13			
0	192aa, >BAT_371_COTO_contig_B	REP	
1	173aa, >BAT_371_COTO_A7	100%	
2	174aa, >FS_2258_MYTH_single_	87%	
3	183aa, >FS_563_EPFU	87%	
4	192aa, >CC_763_MYVE_contig_A	86%	YES
5	192aa, >PKS_II_8H2CC_763_MYVE__D6	86%	
6	192aa, >JC_1424_EPFU_contig_A	85%	
7	175aa, >06FSS_1280_EPFU_	86%	
8	175aa, >ELMA_1069_MYCI	86%	
9	171aa, >04FSS_465_COTO_	85%	
10	174aa, >FS_461_MYTH_single_	85%	
11	179aa, >08ELMA_1026_MYCI_	87%	
12	175aa, >ELMA_1616_MYCI	86%	
13	192aa, >FS_2208_MYTH_contig_B	85%	
14	192aa, >EST_1537_MYEV_contig_b	89%	
15	167aa, >EST_1540_MYEV	86%	
>Cluster 14			
0	192aa, >WCS_2062_COTO_D1	REP	
1	172aa, >CC_2265_MYTH_single	86%	
>Cluster 15			
0	170aa, >09RSS_907_TABR	92%	
1	192aa, >CC_817_MYTH_contig_A	REP	YES
2	192aa, >FSS_600_EPFU_D7_	91%	
3	192aa, >FS_1249_MYCI_contig_A	91%	
4	192aa, >FS_2283_MYCI_contig_A	91%	
5	192aa, >FS_1250_MYCI_contig_A_	91%	
6	192aa, >CRY_442_MYVE_contig_A	86%	
7	192aa, >CRY_451_MYVE_contig_A	86%	YES
>Cluster 16			
0	192aa, >FS_1250_MYCID12	REP	
1	170aa, >FSS_600_EPFU_contig_A	93%	

2	192aa, >CC_839_MYVE_A10	85%	
>Cluster 17			
0	192aa, >PKS_II_Plate_9B3RSS_877_TABR_INH	REP	YES
1	192aa, >ELMA_208_MYEV_contig_B	95%	
2	192aa, >PKS_II_Plate_9B1RSS_877_TABR_INH	96%	YES
>Cluster 18			
0	192aa, >ELMA_1047_MYEV_contig_B	REP	
1	192aa, >ELMA_1047_MYEV_D4	100%	
>Cluster 19			
0	175aa, >ELMA_1030_MYCI	85%	
1	174aa, >FS_512_COTO_	88%	
2	192aa, >RSS_864_MYVE_contig1	REP	YES
3	192aa, >FS_520_MYCI_contig_A	95%	
4	192aa, >CC_839_MYVE_Contig_A	93%	
5	174aa, >EST_1539_MYEV_	87%	
6	171aa, >FS_1200_COTO_	86%	
7	171aa, >TOR_438_MYVE_	90%	
>Cluster 20			
0	192aa, >FSS_1278_EPFU_contig_A	REP	
>Cluster 21			
0	192aa, >ELMA_1078_MYCI_contig_A	REP	
>Cluster 22			
0	192aa, >ELMA_1079_MYCI_contig_A	REP	
>Cluster 23			
0	192aa, >CC_789_MYVE_contig_A	REP	YES
1	192aa, >CC_789_MYVE_G3_	100%	YES
>Cluster 24			
0	192aa, >EST_1534_MYEV_contig_B_	REP	
1	192aa, >PKS_II_4D9EST_1534_MYEV_D9	93%	
>Cluster 25			
0	171aa, >FSS_2267_MYTH	93%	
1	175aa, >RSS_862_MYVE	REP	
2	174aa, >14RSS_884_TABR	97%	YES
>Cluster 26			
0	174aa, >05EST_1716_MYEV_	REP	
1	174aa, >FS_1176_COTO_single_	100%	
2	171aa, >TOR_137_MYVE	100%	
>Cluster 27			
0	151aa, >02ELMA_1009_MYEV_	96%	

1	171aa, >CC_1929_MYTH	99%	
2	174aa, >CC_792_MYTH	REP	
>Cluster 28			
0	172aa, >CLC_399_COTO_	REP	YES
1	171aa, >BAT_369_COTO	100%	
2	171aa, >FW_421_COTO_	100%	YES
3	171aa, >JC_352_COTO_	100%	YES
4	171aa, >JC_353_COTO	100%	YES
>Cluster 29			
0	171aa, >03FSS_581_EPFU	REP	
>Cluster 30			
0	171aa, >JC_332_COTO_	REP	
>Cluster 31			
0	171aa, >FSS_578_EPFU	REP	
>Cluster 32			
0	171aa, >FS_2255_MYTH_	REP	
>Cluster 33			
0	171aa, >CC_785_MYVE_	REP	
1	157aa, >EST_1536_MYEV_	91%	
>Cluster 34			
0	170aa, >CLC_417_COTO_single	REP	
>Cluster 35			
0	157aa, >02ELMA_204_MYEV	REP	

Supplemental Table 2: Results of General Linear Model of Species as a Predictor of Gene Copy Number

	mean	mcse	sd	2.50%	25%	50%	75%	97.50%	n_eff	Rhat
(Intercept)	1.0137	0.0628	7.2401	-13.2268	-3.8583	1.0114	5.8985	15.1129	13291	1.000176888
COTO	-0.3978	0.0628	7.2412	-14.5006	-5.2697	-0.3937	4.4696	13.8140	13299	1.000176541
EPFU	0.2909	0.0628	7.2426	-13.7956	-4.5845	0.3053	5.1674	14.5316	13288	1.000173885
MYTH	-0.0479	0.0628	7.2427	-14.1657	-4.9343	-0.0473	4.8194	14.1652	13308	1.000174692
MYCI	0.1043	0.0628	7.2444	-14.0276	-4.7872	0.1029	4.9719	14.3048	13300	1.000175666
MYEV	0.1859	0.0628	7.2434	-13.9350	-4.7010	0.1819	5.0558	14.4136	13299	1.000173587
MYVE	-0.2631	0.0628	7.2420	-14.3845	-5.1436	-0.2625	4.6154	13.9635	13282	1.000174365
sigma	1.1059	0.0003	0.0641	0.9890	1.0615	1.1031	1.1468	1.2410	35122	1.000177227
mean_PPD	0.9312	0.0005	0.1240	0.6885	0.8481	0.9312	1.0144	1.1743	52634	1.000027756
log-posterior	-253.5687	0.0124	2.0242	-258.3787	-254.7014	-253.2438	-252.0840	-250.6048	26777	1.000086496