Characterization of a Novel North American Bumble Bee Virus

Brianna Flynn

May 19, 2019

Supervising Professor: _____

Biology Honors Advisor: _____

Abstract:

Bee species around the world are experiencing increasing mortality rates, causing concern over the ecological impact their decline will have on the environment. While the primary focus of bee viruses research has been on North American honey bee (*Apis mellifera*) viruses, native bee pathogens are also responsible for increased bee mortality. Our research focused on characterizing the viruses that infect these understudied native populations. In our study, we characterized the diversity of viruses infecting bees of two distinct North American regions, Colorado and New Jersey. From our results, we observed 25 viruses that infect 12 bee species. We were also able to identify and characterize one novel virus that infects five bee species, including honey bees (*Apis mellifera*). We found that this novel virus is most likely a member of the newly described Negevirus family, and is closely related to another recently reported bee virus infecting Negevirus found in Belgium. We found that the novel virus is widely distributed, infecting both Colorado and New Jersey individuals. The results of our study show that the novel virus can infect multiple host species, and confirm that native bee pathogens should be taken into account to fully understand the total amount of viruses capable of infecting Apis and non Apis species.

Introduction:

Native bee populations around the globe are experiencing significant population declines. There are a variety of factors contributing to increased bee mortality. While agricultural pesticides and herbicides (neonicotinoids, glyphosate, BT transgene pollen, etc.) have been investigated as potential causes of colony collapse, natural bee pathogens may also explain increased population instability and decline (Galbraith et. al 2018). Because of their essential role in maintaining a stable agricultural economy, honey bees and their pathogens are well researched. In contrast, the prevalence and virulence of native-bee species pathogens is comparatively understudied.

It has been shown that honey bees and native bees can be infected with the same viral strains (Schoonvaere et. al 2018). However, many potential pathogens are specific to just wild pollinator species and may still be undiscovered (Galbraith et. al, 2018). Viruses found to be pathogenic in *A. mellifera* (such as Deformed Wing Virus (DWV), Israeli Acute Paralysis Virus, Acute Bee Paralysis Virus, and Kashmir Bee Virus) have been shown to also infect wild bee species (Galbraith et al 2018). Which bee species serve as hosts to a particular virus are still mostly unknown. Characterizing non-honey bee viruses is not only important for understanding the mechanisms of wild pollinator population declines, but may also lead to the detection of viruses that may also be pathogenic to honey bees. The data from this and other analyses can be used to characterize the health and pathogen load of wild pollinator populations outside of the scope of known honey bee viruses.

The samples used for this project were taken from two regions of the United States. Characterizing current native American bee pathogens is not only important for establishing what viruses currently plague these bees, but for developing a baseline for detecting pathogens

that invade North America due to increased globalization (Galbraith et. al 2018). Exotic bee species possibly carry pathogens that are transmittable to North American species. Having a more thorough understanding of the current pathogenic diversity of native and non-native bees will help researchers now and in the future assess the effect invasive species have on native pollinator pathogen load.

With high throughput sequencing and metagenomic analysis, it is possible to detect entire viral communities efficiently and relatively inexpensively, without targeting a specific virus of interest (Grozinger et. al 2015). This meta-transcriptome analysis pipeline can be used to identify novel viral species and strains (Galbraith et. al 2018). Most bee viruses are single stranded RNA viruses. Thus, isolating double stranded RNA can be used to identify the presence of actively replicating viruses, since non-replicating viruses present in the food or as contaminants will be excluded (Roossinck et al 2010). Thus, dsRNA is formed in the host from the replication of single stranded RNA viruses. Several recent studies have used this approach to identify novel viruses in bees collected from sites in Northern and Western Europe, Africa, and the Middle East (Remnant et. al 2017, Schoonvaere et. al 2016, Levin et. al 2016, and Schoonvaere et. al 2018).

To identify the distribution of viruses in North American bee species, we sequenced, assembled and analyzed the single- and double-stranded RNA found within wild bumble bees collected from several sites in New Jersey and Colorado. We used this approach to characterize the viral communities of 12 bee species collected from two states. With this approach, we were able to search for novel viruses and determine the distribution of viral families in wild bees across the United States. As a result, we were able to identify and characterize a novel virus that can infect many species of North American bumble bees and also honey bees.

Methods:

Samples were collected by Dan Cariveau, Jessica Tung, and Nancy Moran from locations in New Jersey and Colorado. The bee species, latitude and longitude position were recorded for each specimen (Table 1). Guts were extracted and stored in Thermo Fisher RNAlater. Double-stranded RNA was extracted from guts using a phenol:chloroform extraction protocol from Roossinck et al 2010. Isolated dsRNA was amplified by cloning using methods from Stenger et al 2009. Total RNA was extracted using the Epicentre Masterpure kit as described in Powell et al 2016. cDNA was produced pooled dsRNA and Total RNA samples using a qSCRIPT kit. Library preparation and sequencing of cDNAs made from pooled total RNA and pooled dsRNA samples was performed at the University of Texas Genome Sequencing and Analysis Facility with paired end 2x150 on HiSeq4000. Reads were assembled in CLC Workbench with default parameters. After assembly, the contigs derived from the sequencing of total RNA and of double-stranded RNA were BLASTed against the full viral non-redundant protein sequence NCBI databases. The query contigs that obtained viral BLAST hits were compiled into a list of all viral species present. For each virus contained in the unique list, its genome was retrieved from the GenBank database. Contigs were aligned with their full genome sequence using Geneious. This was used to determine whether the contigs we found represented complete or partial genomes.

Phylogenetic analysis was used to determine the relationship of this strain to other known viruses. The conserved protein database on NCBI was used to find the novel virus's RNA Dependent RNA Polymerase (RDRP2) gene. An alignment was made from each virus's RDRP2 sequence using MAFFT (Nakamura et. al 2018). The novel virus's RDRP2 region as the query in

a BLASTx search, and the top BLASTx results were included in our phylogenetic analysis. This protein alignment was used as input for IQtree to make a maximum likelihood phylogeny with bootstraps (1000 samples) (Nguyen et. al 2015, Hoang et. al 2018).

NCBI's ORF finder was used to identify ORFs in the new virus, and the Pfam database was used to identify potential homologous protein families in known viruses (Schoonvaere et. al 2018). From these, two major conserved protein families were found in the new virus (RDRP2 and SP24). RDRP2 is involved in viral replication and is used often for reconstructing phylogenetic relationships among viruses (Galbraith et. al 2018). SP24 is a structural integral membrane protein also found in chronic bee paralysis virus (Solovyev et. al 2017).

To identify individual samples containing the new virus or close relatives of the new virus, a PCR screen was performed. cDNA from the 43 dsRNA samples was produced using a qSCRIPT kit. This cDNA was then used as the template in our PCR screen. Primers were designed based on conserved regions in the alignment for the new virus and one closely related virus (*A. haemorrhoa* nege-like virus, Schoonvaere et. al 2018). These primer sequences were AATAATTTTGCCGACCATGC forward and GTCTGCCCCGTTTAAAATCA reverse, with an expected amplicon size of 201 base pairs

CATGTCATATGCCTTCACATCCGGAGTTTCATAGTAATTTATAAAAGTACGATATGG
TCGAATAGAATATTGCACTTGTTTTTGATATTCTTCACGAAAATTACTTAAAAAATAT
CCGTTAGCAACAGCAACAAGAAAAAAGAGATTAATTAGCGACATTTTGATTTTAAA
CGGGGCAGACTTTATTTAATATAATATGAT.

Table 1. Library, Label, Location, Species, and Date information for each individual bee collected and used in our analysis.

dsRNA	Label on Tube	State	Species	Location	Date	Longitude	Latitude
	RV 43	CO	Bombus huntii	Cimarron Colorado	Aug-15	38.44225	-107.5567
	BB BF 3 8/20	NJ	Bombus bimaculatus	Beuchagh Fork	Jun-13	40.509024	-74.574077
	RV 32 11/27	CO	Apis Mellifera	Cimarron Colorado	Aug-15	38.44225	-107.5567
	RV 39 10/2/18	CO	Bombus huntii	Cimarron Colorado	Aug-15	38.44225	-107.5567
	BI-RG 2 8/19	NJ	Bombus impatiens	Rutgers Garden	Aug-13	40.509024	-74.574077
	RV-33	CO	Apis Mellifera	Cimarron Colorado	Aug-15	38.44225	-107.5567
	RV-38	CO	Apis Mellifera	Cimarron Colorado	Aug-15	38.44225	-107.5567
	RV 35 11/27	CO	Bombus huntii	Cimarron Colorado	Aug-15	38.44225	-107.5567
	BI-RG 2 8/20	NJ	Bombus impatiens	Rutgers Garden	Aug-13	40.509024	-74.574077
	RV 41 12.18	CO	Apis Mellifera	Cimarron Colorado	Aug-15	38.44225	-107.5567
	RV 40 12.18	CO	Apis Mellifera	Cimarron Colorado	Aug-15	38.44225	-107.5567
	RV 44 12.18	CO	Bombus huntii	Cimarron Colorado	Aug-15	38.44225	-107.5567
	T.8.2.13 dsRNA 8.26	CO	Bombus huntii	Cimarron Colorado	Aug-15	38.44225	-107.5567
	8.8.2.2 dsRNA 8.26	CO	Apis Mellifera	Grand Junction Colorado	Aug-15	39.095557	-108.550214
	RV 37 11/27	CO	Bombus bifarius	Cimarron Colorado	Aug-15	38.44225	-107.5567
	8.8.2.11 dsRNA 8.26	CO	Bombus appositus	Grand Junction Colorado	Aug-15	39.095557	-108.550214
	RV 34 11.27	CO	Bombus huntii	Cimarron Colorado	Aug-15	38.44225	-107.5567
	RV 36 11.27	CO	Bombus huntii	Cimarron Colorado	Aug-15	38.44225	-107.5567

	RV 46 12.8	CO	Bombus bifarius	Cimarron Colorado	Aug-15	38.44225	-107.5567
	RV 47 12.8	CO	Bombus huntii	Cimarron Colorado	Aug-15	38.44225	-107.5567
	BB CP 1 8.19	NJ	Bombus bimaculatus	Colonial Park	May-13	40.509024	-74.574077
	8.8.2.3 dsRNA 8/26	CO	Apis mellifera	Grand Junction Colorado	May-13	39.095557	-108.550214
	BC1.22 8/20	CO	Bombus appositus	Brush Creek	Jul-14	38.861033	-106.92284
	CC-3 8/20	CO	Bombus ccidentalis	Copper Creek	Jul-14	39.0188798	-106.8092052
	BC1-21 8.20	CO	Bombus californicus	Brush Creek	Jul-14	38.861033	-106.92284
	RV 42 12/8	CO	Bombus bifarius	Cimarron Colorado	Aug-15	38.44225	-107.5567
	RV 45 12/8	CO	Bombus bifarius	Cimarron Colorado	Aug-15	38.44225	-107.5567
	BB CP 2 8.19	NJ	Bombus bimaculatus	Colonial Park	May-13	40.509024	-74.574077
	8.8.2.1 dsRNA 8.26	CO	Apis mellifera	Grand Junction Colorado	Aug-15	39.095557	-108.550214
	BB BF-1 8.18	NJ	Bombus bimaculatus	Beuchagh Fork	Jun-13	40.509024	-74.574077
	BCI 23 8.20	CO	Bombus appositus	Brush Creek	Jul-14	38.861033	-106.92284
	BB BF 1 8.20	NJ	Bombus bimaculatus	Beuchagh Fork	Jun-13	40.509024	-74.574077
	BB BF 2 8.18	NJ	Bombus bimaculatus	Beuchagh Fork	Jun-13	40.509024	-74.574077
	BI-RG 1 8.20	NJ	Bombus impatiens	Rutgers Garden	Aug-13	40.509024	-74.574077
	RG BB CO B dsRNA	NJ	Bombus bimaculatus	Rutgers Garden	May-13	40.509024	-74.574077
	RG Bimac dsRNA	NJ	Bombus bimaculatus	Rutgers Garden	May-13	40.509024	-74.574077
	BI RG 1 818	NJ	Bombus impatiens	Rutgers Garden	Aug-13	40.509024	-74.574077
	BI-RG 3 8.19	NJ	Bombus impatiens	Rutgers Garden	Aug-13	40.509024	-74.574077

	8.8.2.12 dsRNA 8.26	CO	Bombus sp.	Grand Junction Colorado	Aug-15	39.095557	-108.550214
	BI RG 1 8.19	NJ	Bombus impatiens	Rutgers Garden	Aug-13	40.509024	-74.574077
	BI RG 3 8.18	NJ	Bombus impatiens	Rutgers Garden	Aug-13	40.509024	-74.574077
	BI RG 2 8.18	NJ	Bombus impatiens	Rutgers Garden	Aug-13	40.509024	-74.574077
	B3 BF 2 8.20	NJ	Bombus bimaculatus	Beuchagh Fork	Jun-13	40.509024	-74.574077
Total RNA							
	APP.1	CO	Bombus appositus	N/A	7/22/14	38.9818008	-107.0185265
	APP.3	CO	Bombus appositus	N/A	7/25/14	38.9866424	-107.0115847
	BC1.1	CO	Bombus appositus	N/A	7/26/14	38.864147	-106.913549
	BC1.14	CO	Bombus appositus	N/A	7/26/14	38.864147	-106.913549
	BC1.20	CO	Bombus appositus	N/A	7/26/14	38.864147	-106.913549
	BC1.9	CO	Bombus appositus	N/A	7/26/14	38.864147	-106.913549
	BC2.5	CO	Bombus appositus	N/A	7/26/14	38.859786	-106.921874
	BC2.7	CO	Bombus appositus	N/A	7/26/14	38.859786	-106.921874
	EC.3	CO	Bombus appositus	N/A	7/27/14	38.860373	-107.067812
	EC.4	CO	Bombus appositus	N/A	7/27/14	38.860373	-107.067812
	KP.16	CO	Bombus appositus	N/A	7/26/14	38.95598	-106.975978
	KP.20	CO	Bombus appositus	N/A	7/26/14	38.95598	-106.975978
	KP.24	CO	Bombus appositus	N/A	7/26/14	38.95598	-106.975978
	KP.7	CO	Bombus appositus	N/A	7/26/14	38.95598	-106.975978
	BC1.10	CO	Bombus bifarius	N/A	7/26/14	38.864147	-106.913549
	BC1.15	CO	Bombus bifarius	N/A	7/26/14	38.864147	-106.913549

	BC1.16	CO	Bombus bifarius	N/A	7/26/14	38.864147	-106.913549
	BC1.19	CO	Bombus bifarius	N/A	7/26/14	38.864147	-106.913549
	BC1.3	CO	Bombus bifarius	N/A	7/26/14	38.864147	-106.913549
	BC1.7	CO	Bombus bifarius	N/A	7/26/14	38.864147	-106.913549
	BC2.2	CO	Bombus bifarius	N/A	7/26/14	38.859786	-106.921874
	BIF.2	CO	Bombus bifarius	N/A	7/25/14	38.9866383	-107.0115847
	BIF.3	CO	Bombus bifarius	N/A	7/25/14	38.9970783	-107.0200787
	EC.5	CO	Bombus bifarius	N/A	7/27/14	38.860373	-107.067812
	KP.10	CO	Bombus bifarius	N/A	7/26/14	38.95598	-106.975978
	KP.22	CO	Bombus bifarius	N/A	7/26/14	38.95598	-106.975978
	KP.5	CO	Bombus bifarius	N/A	7/26/14	38.95598	-106.975978
	KP.8	CO	Bombus bifarius	N/A	7/26/14	38.95598	-106.975978
	NJ.11	NJ	Bombus bimaculatus	N/A	5/16/13	40.509024	-74.574077
	NJ.18	NJ	Bombus bimaculatus	N/A	7/22/13	40.471154	-74.522877
	NJ.16	NJ	Bombus bimaculatus	N/A	5/16/13	40.509024	-74.574077
	NJ.19	NJ	Bombus bimaculatus	N/A	7/22/13	40.471154	-74.522877
	NJ.2	NJ	Bombus bimaculatus	N/A	5/15/13	40.473177	-74.422896
	NJ.23	NJ	Bombus bimaculatus	N/A	6/11/13	40.503477	-74.46007
	NJ.30	NJ	Bombus bimaculatus	N/A	6/11/13	40.503477	-74.46007
	NJ.6	NJ	Bombus bimaculatus	N/A	5/15/13	40.473177	-74.422896
	NJ.8	NJ	Bombus bimaculatus	N/A	5/15/13	40.473177	-74.422896
	BC1.6	CO	Bombus fervidus	N/A	7/26/14	38.864147	-106.913549
	BC2.4	CO	Bombus fervidus	N/A	7/26/14	38.859786	-106.921874

	CAL.1	CO	Bombus fervidus	N/A	7/27/14	38.969921	-106.9976957
	CC.2	CO	Bombus fervidus	N/A	7/28/14	38.968228	-106.966637
	KP.19	CO	Bombus fervidus	N/A	7/26/14	38.95598	-106.975978
	KP.9	CO	Bombus fervidus	N/A	7/26/14	38.95598	-106.975978
	NJ.32	NJ	Bombus impatiens	N/A	7/22/13	40.509024	-74.574077
	NJ.34	NJ	Bombus impatiens	N/A	7/22/13	40.509024	-74.574077
	NJ.36	NJ	Bombus impatiens	N/A	7/22/13	40.509024	-74.574077
	NJ.37	NJ	Bombus impatiens	N/A	7/22/13	40.509024	-74.574077
	NJ.41	NJ	Bombus impatiens	N/A	7/22/13	40.509024	-74.574077
	NJ.42	NJ	Bombus impatiens	N/A	7/22/13	40.471154	-74.522877
	NJ.43	NJ	Bombus impatiens	N/A	7/22/13	40.471154	-74.522877
	NJ.44	NJ	Bombus impatiens	N/A	7/22/13	40.471154	-74.522877
	NJ.45	NJ	Bombus impatiens	N/A	7/22/13	40.471154	-74.522877
	NJ.51	NJ	Bombus impatiens	N/A	7/22/13	40.471154	-74.522877
	NJ.59	NJ	Bombus impatiens	N/A	8/2/13	40.473177	-74.422896
	NJ.60	NJ	Bombus impatiens	N/A	8/2/13	40.473177	-74.422896
	NJ.63	NJ	Bombus impatiens	N/A	8/2/13	40.473177	-74.422896
	NJ.64	NJ	Bombus impatiens	N/A	8/2/13	40.473177	-74.422896
	NJ.65	NJ	Bombus impatiens	N/A	8/3/13	40.473177	-74.422896
	NJ.69	NJ	Bombus impatiens	N/A	8/2/13	40.473177	-74.422896
	NJ.74	NJ	Bombus impatiens	N/A	8/2/13	40.473177	-74.422896
	NJ.75	NJ	Bombus impatiens	N/A	8/26/13	40.473177	-74.422896
	NJ.76	NJ	Bombus impatiens	N/A	8/26/13	40.473177	-74.422896

	NJ.77	NJ	Bombus impatiens	N/A	8/26/13	40.473177	-74.422896
	NJ.78	NJ	Bombus impatiens	N/A	8/26/13	40.473177	-74.422896
	NJ.81	NJ	Bombus impatiens	N/A	8/26/13	40.473177	-74.422896
	BC1.13	CO	Bombus nevadensis	N/A	7/26/14	38.864147	-106.913549
	BC2.6	CO	Bombus nevadensis	N/A	7/26/14	38.859786	-106.921874
	KP.1	CO	Bombus nevadensis	N/A	7/26/14	38.95598	-106.975978
	KP.2	CO	Bombus nevadensis	N/A	7/26/14	38.95598	-106.975978
	KP.13	CO	Bombus nevadensis	N/A	7/26/14	38.95598	-106.975978
	KP.14	CO	Bombus nevadensis	N/A	7/26/14	38.95598	-106.975978

Results:

Identification of viral sequences

We analyzed the sequences from the dsRNA and ssRNA samples, each representing RNA pooled from multiple individual bees and bee species. On average, 2.9 % of the assembled contigs in each sample showed evidence of being viral in origin based on BLASTx amino acid homology. We identified 25 unique contigs among the 338 contigs that shared sequence homology with previously identified insect-specific viruses (Table 2). We found 17 unique contigs that showed homology with plant pathogens. Of the five contigs that corresponded to insect-specific viruses, four contigs showed homology with viruses characterized in bees and one with a virus previously characterized only in mosquitos.

Within our dataset, we identified four well studied bee viruses: Slow Bee Paralysis Virus, Lake Sinai Virus, Black Queen Cell Virus, and Honeybee Slow Paralysis Virus. Viruses related to Black Queen Cell Virus were the most abundant virus group as represented by these contigs, comprising 59.2% of all viral BLAST hits. Viruses related to Tobacco Ringspot were the second most prevalent group, and the most prevalent plant virus group, comprising 9.76% percent of our contigs. All identified plant and insect infecting viral sequences are positive single strand RNA (+ssRNA) viruses.

Table 2. List of all viral species found in both the total and ds RNA datasets, with 25 unique species in total.

Unique Virus	Host	Genome (complete or partial)	DS or SS library
Dezidougou virus	Insect	complete	DS
Slow bee paralysis	Insect	complete	DS
Cycas necrotic stunt virus	Plant	partial	SS
Lake Sinai virus 2	Insect	complete	DS
Xenopus tropicalis human immunodeficiency virus type I	Frog	partial	SS
Equine arteritis virus	Horse	complete	SS
Black queen cell virus strain	Insect	complete	DS
Picoides pubescens	Animal	partial	SS
Tobacco ringspot virus	Plant	complete	SS
Potato black ringspot virus	Plant	complete	SS
White clover cryptic virus 2	Plant	complete	SS
Red clover cryptic virus 2	Plant	complete	SS
Cannabis cryptic virus isolate	Plant	partial	SS
Hop trefoil cryptic virus 2 isolate	Plant	partial	SS
Crimson clover cryptic virus 2	Plant	complete	SS
Chiltepin yellow mosaic virus	Plant	complete	SS
Tomato blistering mosaic virus	Plant	complete	SS
Watercress white vein virus	Plant	complete	SS
Tobacco streak virus	Plant	complete	SS
Parietaria mottle virus	Plant	complete	SS
Ageratum latent virus isolate	Plant	partial	SS
Blackberry chlorotic ringspot virus isolate	Plant	partial	SS
Potexvirus white clover mosaic virus	Plant	complete	SS
Carnation ringspot virus	Plant	partial	SS
Sweet clover necrotic mosaic virus	Plant	partial	SS

Honey bee slow paralysis virus isolate R4LY 47 RDRP gene, partial cds	Insect	partial	DS

Identification of novel virus

Dezidougou-like virus. We identified one Dezidougou-like contig that shared sequence homology with known viruses from the recently described *Negevirus* family, a viral family that includes mostly known mosquito-infecting viruses such as Biggievirus, Tanay Virus, and Culex Virus.

The contig length of this newly described virus was 3,014 nucleotides, which included a putative RNA dependent RNA polymerase region. A maximum likelihood phylogenetic analysis of this region revealed that this particular contig was most closely related to a newly reported bee virus of the *Sandwavirus* sub-group of Negeviruses reported in *Andrena haemorrhoa* with a percent identity of 72.9% (Schoonvaere et. al 2018). All other viruses in our phylogenetic analysis were mosquito-infecting and had percent identities of 60.8% or fewer for amino acids. Because *Bombus* and *Andrena* are distantly related species, the Negeviruses can possibly infect a wide array of hosts (Hedtke et. al 2013).

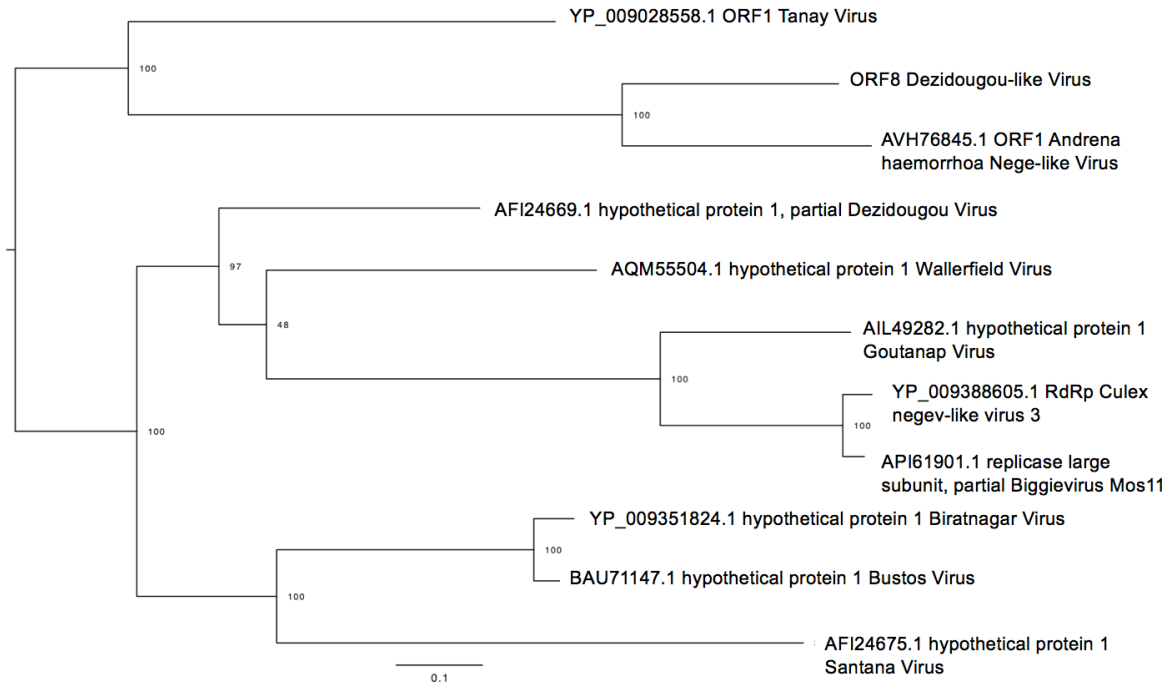
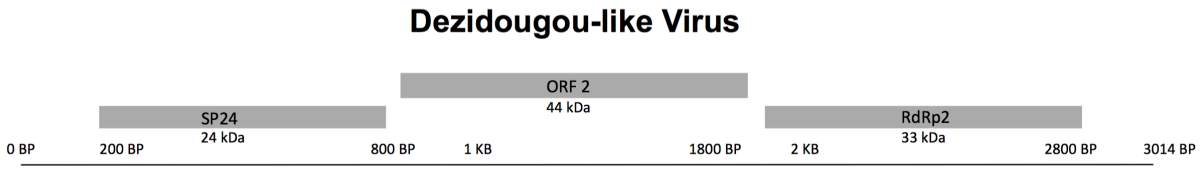


Figure 1. Maximum likelihood phylogenetic tree showing the evolutionary relationships between the novel virus (ORF8 Dezidougou-like Virus). and other members of the Negevirus family. Phylogenetic analysis based on amino acid sequence of the RdRp or replicase protein.

Genome of the Newly Discovered Dezidougou-Like Virus

The genome of the Dezidougou-like contig contains three open reading frames containing a protein that showed greatest sequence homology to ORF 1 of the recently described *Andrena haemorrhoa* virus. The polyprotein also encodes a structural protein, SP24, a recently described member of a superfamily of plant and insect viruses (Pfam 16504). This protein is most likely an integral membrane protein and is commonly found in the *Negevirus* family, first characterized in Chronic Bee Paralysis Virus. Lastly, the contig sequence contains an RNA dependent RNA polymerase 2 gene (RDRP2), a highly conserved protein superfamily necessary for (+)ssRNA virus replication.



SP24 Superfamily Pfam 16504	606 NT | 201 aa	Homologous structural protein found in Negeviruses and first characterized in Chronic Bee Paralysis Virus. Most likely an integral membrane protein.
ORF 2	1098 NT | 365 aa	Showed greatest sequence homology to ORF 2 of Andrena haemorrhoa nege-like virus. ORF2 in Negeviruses may be homologous proteins.
RdRp2 Pfam 00978	837 NT | 278 aa	RNA Dependent RNA polymerase gene, highly conserved protein family in viruses necessary for viral replication

Figure 2. Genome of the newly discovered Dezidougou-like virus.

Confirmation of Novel Virus Presence using PCR screen

The novel virus was found in multiple different *Bombus* species and *Apis mellifera* in both the Colorado and New Jersey samples (Table 3). We used PCR to confirm the presence of our novel virus or closely related viruses among our individual samples. For our screen, we used cDNA produced from 43 dsRNA samples collected from Colorado and New Jersey. From this PCR screen, we detected the presence of the novel virus or closely related viruses in eight individuals, across five different bee species, in both Colorado and New Jersey specimen.

Table 3. Individuals infected with the novel Dezidougou-like or closely related virus, identified by a PCR screen.

Label on Tube	State	Species	Location	Date	Longitude	Latitude
RV 39 10/2/18	CO	Bombus huntii	Cimarron Colorado	Aug-15	38.44225	-107.5567
BI-RG 2 8/20	NJ	Bombus impatiens	Rutgers Garden	Aug-13	40.509024	-74.574077
BC1-21 8.20	CO	Bombus californicus	Brush Creek	Jul-14	38.861033	-106.92284

RV 42 12/8	CO	Bombus bifarius	Cimarron Colorado	Aug-15	38.44225	-107.5567
RV 45 12/8	CO	Bombus bifarius	Cimarron Colorado	Aug-15	38.44225	-107.5567
8.8.2.1 dsRNA 8.26	CO	Apis mellifera	Grand Junction Colorado	Aug-15	39.095557	-108.550214
BI RG 1 8.19	NJ	Bombus impatiens	Rutgers Garden	Aug-13	40.509024	-74.574077
BI RG 2 8.18	NJ	Bombus impatiens	Rutgers Garden	Aug-13	40.509024	-74.574077

Discussion

The objectives of this study were to characterize the viral diversity of North American bumble bee species, to find potential novel viruses, and characterize their genomes and distribution. We found a total of five insect viruses, seven plant pathogens, four bee viruses and one previously undiscovered bee virus, referred to as Dezidougou-like virus. Total RNA from 116 samples and double stranded RNA from 43 samples were pooled and sequenced via RNA seq. From these two pooled libraries, we assembled reads into viral contigs and used these as queries for BLAST. Our queries were run against the non-redundant protein database to identify viruses based on their amino acid sequence homology.

Samples were collected from various locations in Colorado and New Jersey. One limitation in our search for any potential new viruses was the limited amount of geographical regions sampled from for our study. Though these two regions are distant, the small sampling size and limited sampling locations are not sufficient to get a complete picture of North American bee viral diversity. Future work should strive to incorporate a greater range of sampling sites and to collect more individuals. Further, the elevational range between both locations was vastly different. Elevation plays a large role in what bee species populate a certain location, and what pathogens they are exposed to (Widhiono et. al 2017). While the mean

elevation of Colorado is 6,800 feet above sea level with the lowest point being 3,315 feet above sea level, the mean elevation of New Jersey is 1,803 feet. It is unclear whether the diversity of pathogens found in either Colorado or New Jersey were representative of ecological differences between the two regions, elevational differences, or both.

We found that a significant percentage of our contig queries BLASTed to plant pathogens. While its assumed that some of these plant viruses are contamination from pollen alone, it has been recently shown that plant viruses such as Tobacco Ringspot virus are able to infect and replicate within honey bee hosts (Li et. al 2014). To better understand the wild pollinator's role as a carrier, and possibly host for these viruses, further work should be done to screen for plant viruses that are actively replicating in the bee host using the dsRNA extraction and screening method we used to identify actively replicating viruses.

Despite previously having found the Dezidougou-like Virus in a New Jersey specimen, we found the novel virus or closely related viruses in not only the New Jersey individuals but in the Colorado specimen as well. This suggests that the novel virus could be more widespread than what we previously predicted. This opens up some interesting future questions to address because of the distinct differences between the two distinct geographical regions we sampled from. Follow up research should be done to screen individuals for what viruses each carries. This would help determine whether the widespread distribution of the novel virus is consistent with the distribution of other previously identified bee viruses.

The majority of bees used for our viral diversity RNAseq analysis were found foraging when collected. Because we primarily studied bees that were foraging, we may have only captured a sub-set of the actual viral diversity of the species in that region. Diseased bees have been shown to forage significantly less (Koch et. al 2017). The samples we collected therefore

may carry a lower pathogen load or be un-representative of the viral community present and actively infecting these species.

A study by Shooenvaere et al. (2018) characterized the first known bee virus from the Negevirus family in *A. haemorrhoa* bees. Our phylogenetic analysis suggests that the virus we observed is distinct from the previously described bee virus in this family with a relatively low percent amino acid sequence identity (72.9%). Its other closest phylogenetic relatives of the virus are Negevirus members found in mosquito hosts. A nucleotide BLAST showed closest homology with the Dezidougou virus, another mosquito virus in the same family. Because the BLASTx results showed that the virus was most closely related with another bee virus of the same family gives us some evidence that the novel virus, though closely related to mosquito viruses, does infect bee hosts. Our results show that these two bee viruses are very closely related despite their significant distance apart (with the *A. haemorrhoa*-infecting virus reported in Belgium and the Dezidougou-like virus observed in the Northeastern and Western United States). Because members of this viral family infect multiple species *(Bombus impatiens, Bombus huntii, Bombus californicus, Bombus bifarius, Apis mellifera*) and have been found two distant geographic locations (Belgium and United States), the Negevirus family is widespread and even globally distributed.

Despite the popular focus on North American *A. mellifera* viruses, our study shows that bee viruses can infect multiple host species, and thus wild pollinator populations should be taken into account when quantifying the total amount of viruses that infect Apis and non Apis species. Additional work needs to be done to observe what effect the novel virus has on the bee species it infects.

References

Brown, M.J.F.F. and Paxton, R.J. (2009) The conservation of bees: a global perspective. apidologie, 40, 410–416. https://doi.org/10.1051/apido/2009019.

D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh (2018) UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol., 35, 518–522. https://doi.org/10.1093/molbev/msx281

Galbraith, D. A., Fuller, Z., Brockmann, A., Frazier, M., Gikungu, M. W., Kapheim, K. M., . . . Grozinger, C. M. (2018). Investigating the viral ecology of global bee communities with high-throughput metagenomics. Scientific Reports, 8, 8879. doi:10.1101/243139

Grozinger, C.M. and Robinson, G.E. (2015) The power and promise of applying genomics to honey bee health. Current Opinion in Insect Science, 10, 124–132. https://doi.org/10.1038/nm.2399.The.

Hedtke, Shannon M.; Patiny, Sébastien; Danforth, Bryan M. (2013). The bee tree of life: a supermatrix approach to apoid phylogeny and biogeography. BMC Evolutionary Biology, 13, 138. doi:10.1186/1471-2148-13-138.

Koch, H., Brown, M. J. & Stevenson, P. C. (2017). The role of disease in bee foraging ecology. Curr. Opin. Insect Sci. 21, 60–67

Lam-Tung Nguyen, Heiko A. Schmidt, Arndt von Haeseler, Bui Quang Minh. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution, 32, 268–274, https://doi.org/10.1093/molbev/msu300

Levin, S., Sela, N. and Chejanovsky, N. (2016) Two novel viruses associated with the Apis mellifera pathogenic mite varroa destructor. Scientific Reports, 6, 37710. https://doi.org/10.1038/srep37710.

Li, J. L. et al. (2014) Systemic spread and propagation of a plant-pathogenic virus in European honeybees, Apis mellifera. MBio 5, e00898–13.

Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", Nucleic Acids Res. 45, 200-203

Nakamura, Yamada, Tomii, Katoh (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics, 34, 2490–2492

Ncbi.nlm.nih.gov. (2015). ORF Finder. http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi

Remnant, E.J., Shi, M., Buchmann, G., Blacquie`re, T., Holmes, E.C., Beekman, M. and Ashe, A. (2017) A diverse range of novel RNA viruses in geographically distinct honey bee populations. Journal of Virology, 91, e00158-17, https://doi.org/10.1128/JVI.00158-17.

Roossinck, M. J., Saha, P., Wiley, G. B., Quan, J., White, J. D., Lai, H., . . . Roe, B. A. (2010). Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. Molecular Ecology, *19*, 81-88. doi:10.1111/j.1365-294x.2009.04470.x

Schoonvaere, K., Smagghe, G., Francis, F., & Graaf, D. C. (2018). Study of the metatranscriptome of eight social and solitary wild bee species reveals novel viruses and bee parasites. Frontiers in Microbiology, 9, 117, doi:10.3389/fmicb.2018.00177

Schoonvaere, K., De Smet, L., Smagghe, G., Vierstraete, A., Braeckman, B.P. and De Graaf, D.C. (2016) Unbiased RNA shotgun metagenomics in social and solitary wild bees detects associations with eukaryote parasites and new viruses. PLoS ONE, 11, 1–18. https://doi.org/10.1371/journal.pone.0168456.

Solovyev Andrey G., Morozov Sergey Y. (2017). Non-replicative integral membrane proteins encoded by plant alpha-like viruses: emergence of diverse orphan ORFs and movement protein genes. Frontiers in Plant Science, 8, 1820.

Vasilakis, N., Forrester, N. L., Palacios, G., Nasar, F., Savji, N., Rossi, S. L., . . . Tesh, R. B. (2012). Negevirus: a proposed new taxon of insect-specific viruses with wide geographic distribution. Journal of Virology, 87, 2475-2488. doi:10.1128/jvi.00776-12

Widhiono, I., Sudiana, E., & Darsono, D. (2017). Diversity of wild bees along elevational gradient in an agricultural area in central Java, Indonesia. Psyche: A Journal of Entomology, 2017, 1-5. doi:10.1155/2017/2968414