

SANDIA REPORT

SAND2014-17260

Unlimited Release

Printed April 2014

Enhanced vector borne disease surveillance of California Culex mosquito populations reveals spatial and species-specific barriers of infection

Stanley Langevin¹, Owen Solberg¹, Sarah Wheeler², Victoria Vandernoot³, Deanna Curtis¹, Chung-Yan Koh³, Ying Fang², Ben Brodsky⁴, Todd Lane¹, and William Reisen²

¹Systems Biology, Sandia National Laboratories, Livermore, California, United States of America. ² Center for Vector-borne Diseases, School of Veterinary Medicine, University of California, Davis, California, United States of America. ³ Biotechnology and Bioengineering, Sandia National Laboratories, Livermore, California, United States of America. ⁴ International Biological Threat Reduction, Sandia National Laboratories, New Mexico, United States of America

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.



Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: <http://www.osti.gov/bridge>

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>



Enhanced vector borne disease surveillance of California *Culex* mosquito populations reveals spatial and species-specific barriers of infection

Stanley Langevin¹, Owen Solberg¹, Sarah Wheeler², Victoria Vandernoot³, Deanna Curtis¹,
Chung-Yan Koh³, Ying Fang², Ben Brodsky⁴, Todd Lane¹, and William Reisen²

¹Systems Biology

Sandia National Laboratories
PO Box 969, Livermore, CA 94551-0969

²Center for Vector-borne Diseases

School of Veterinary Medicine, University of California
Davis, CA

³Biotechnology and Bioengineering

Sandia National Laboratories
PO Box 969, Livermore, CA 94551-0969

⁴International Biological Threat Reduction

Sandia National Laboratories
PO Box 1500, Albuquerque, NM 87185

Abstract

Monitoring infections in vectors such as mosquitoes, sand flies, tsetse flies, and ticks to identify human pathogens may serve as an early warning detection system to direct local government disease preventive measures. One major hurdle in detection is the ability to screen large numbers of vectors for human pathogens without the use of genotype-specific molecular techniques. Next generation sequencing (NGS) provides an unbiased platform capable of identifying known and unknown pathogens circulating within a vector population, but utilizing this technology is time-consuming and costly for vector-borne disease surveillance programs. To address this we developed cost-effective Illumina® RNA-Seq library preparation methodologies in conjunction with an automated computational analysis pipeline to characterize the microbial populations circulating in *Culex* mosquitoes (*Culex quinquefasciatus*, *Culex quinquefasciatus/pipiens* complex hybrids, and *Culex tarsalis*) throughout California. We assembled 20 novel and well-documented arboviruses representing members of Bunyaviridae, Flaviviridae, Ifaviridae, Mesoniviridae, Nidoviridae, Orthomyxoviridae, Parvoviridae,

Reoviridae, Rhabdoviridae, Tymoviridae, as well as several unassigned viruses. In addition, we mapped mRNA species to divergent species of trypanosoma and plasmodium eukaryotic parasites and characterized the prokaryotic microbial composition to identify bacterial transcripts derived from wolbachia, clostridium, mycoplasma, fusobacterium and campylobacter bacterial species. We utilized these microbial transcriptomes present in geographically defined *Culex* populations to define spatial and mosquito species-specific barriers of infection. The virome and microbiome composition identified in each mosquito pool provided sufficient resolution to determine both the mosquito species and the geographic region in California where the mosquito pool originated. This data provides insight into the complexity of microbial species circulating in medically important *Culex* mosquitoes and their potential impact on the transmission of vector-borne human/veterinary pathogens in California.

ACKNOWLEDGMENTS

We would like to thank Minyong Chung at the UC Berkeley Vincent J. Coates DNA sequencing facility for his technical support on the Illumina HiSeq sequencing and Casava software for demultiplexing the RNA-Seq libraries. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. S.A.L., S.S.W., W.K.R, T.W.L. conceived the research. S.A.L., S.S.W., V.V., O.D.S., designed experiments. S.A.L., D.J.C., S.S.W., and Y.F. performed experiments. O.D.S., and S.A.L. analyzed sequencing data. O.D.S. and S.A.L. created figures and tables for the manuscript. S.A.L and O.D.S. wrote the manuscript. All authors participated in the editing of this manuscript

CONTENTS

1.	INTRODUCTION.....	9
1.1.	Materials and Methods	10
1.2.	Mosquito RNA Isolation and RNA-Seq Library Preparation.....	10
1.3.	Rare RNA Transcript Sequence Enrichment	10
1.4.	Deep Sequencing and Raw Sequence QC	11
1.5.	Metagenomic Computational Pipeline	11
1.6.	Quantitative PCR Conditions and Virus Primer Sequences	11
1.7.	Cell Culture and Viral Growth Kinetics	12
1.8.	Viral Genome Analysis and Coverage Maps.....	12
1.9.	Spatial and Phylogenetic Analyses	12
2.	RESULTS	13
2.1.	Arboviral Sequence Enrichment and Genome Coverage	13
2.2.	Culex Microbial Metagenomic Overview	13
2.3.	Viral Contig Coverage Maps and Genome Analysis.....	13
2.4.	Virus Classification and Genome Annotation of Divergent Arboviruses	14
2.5.	Spatial Analysis of California Culex Virome	14
2.6.	Culex Bunyavirus Sequence Distribution and Geophylogeny	14
3.	DISCUSSION	17
4.	REFERENCES.....	19
	Appendix A: FIGURES.....	21
	Distribution.....	29

FIGURES

Figure 1.	Figure Caption.....	23
Figure 2.	Caption	24
Figure 3.	Caption	25
Figure 4.	Caption	26
Figure 5.	Caption	27
Figure 6.	Caption	28
Figure 7.	Caption	29
Figure 8.	Caption	30

NOMENCLATURE

dB	decibel
DOE	Department of Energy
SNL	Sandia National Laboratories

1. INTRODUCTION

The vast majority of emerging/re-emerging pathogens causing epidemics/epizootics worldwide are maintained in zoonotic or vector-borne transmission cycles (refs). Globalization of commerce, altering human migration patterns, and climate change have contributed to the increased outbreak frequency and expanded geographic distribution of vector-borne pathogens worldwide. Mosquito-borne disease surveillance is an important strategy to identify and monitor important human and veterinary vector-borne pathogens circulating in nature prior to reported medical cases (refs). This early warning system can dampen pathogen transmissibility through the rapid implementation of adequate vector-control measures and public awareness strategies prior to reaching epidemic/epizootic levels. Labor-intensive workflows and limited resources for pathogen detection assays that rely on a priori knowledge of the pathogen sequence reduce the overall effectiveness and scope of mosquito surveillance programs.

The California vector control districts in collaboration with the Center for Vector-borne Diseases (CVEC) at UC Davis provides logistical disease surveillance and mosquito population control measures to the state of California. There are vector control districts in California and the majority of them screen culex mosquito pools (n=50 mosquitoes/pool) for pathogenic arboviruses such as West Nile virus, St. Louis encephalitis virus, and Western equine encephalitis virus using quantitative PCR. During the summer of 2012 California experienced a resurgence of West Nile virus activity leading to one of the largest epidemics on record since its introduction in 2004. The mosquito districts reported 2814 WNV positive mosquito pools and 380 confirmed human cases of WNV.

Deep sequencing technologies have been utilized to identify well characterized as well as highly divergent arthropod-borne viruses circulating in mosquito populations [1-3]. Even though RNA viruses have been shown to replicate to high titers (10^6 to 10^{11} viral copies) within their adapted arthropod host, current viral discovery sequencing workflows enrich for intact virions prior to extracting the sample RNA/DNA nucleic acids and generating sequencing libraries [4]. These methods rely on viral propagation in cell culture, ultracentrifugation, sample filtration to remove host cells, and/or nuclease treatments to select for encapsidated viral genomic DNA/RNA [5,6]. While these techniques are effective for enriching for virions they lack the ability to detect non-encapsidated viral nucleic acids and non-viral pathogens (i.e. bacteria, fungi, and parasitic eukaryotic organisms) circulating in targeted mosquito populations.

We applied cost-effective RNA-Seq sample preparation protocols to rapidly generate Illumina® compatible cDNA libraries that are depleted of the abundant host ribosomal RNA background to enrich for rare microbial sequences circulating in vector populations. We selected 40 West Nile virus positive mosquito pools representing 10 mosquito pools from 4 geographically distinct locations (Coachella, Greater Los Angeles, Kern, and Sacramento counties). In addition, an automated computational pipeline was constructed to remove Culex-derived mosquito sequences, de novo assemble microbial contigs, identify the contigs using blastn/blastx, and visualize the microbial metagenomic composition using the Krona hierarchical data browser.

1.1. Materials and Methods

1.2. Mosquito RNA Isolation and RNA-Seq Library Preparation

Pools of 50 *Culex* mosquitoes (n=10) were collected at 4 distinct regions and RNA-Seq libraries for each location were processed on separate days. To extract total RNA 200uL of mosquito supernatant was added to 1mL of RNAzol (MRC), each sample was pulse vortexed for 10 seconds, the samples were allowed to incubate a RT for 10mins, then centrifuged at 16,000 x g at 4C for 15mins, and finally the aqueous layer was removed for subsequent total RNA clean-up using Directzol kit following the manufacturer's instructions (ZymoResearch). Total RNA integrity was assessed using the RNA pico chip on a bioanalyzer 2100 (Agilent technologies). All RNA was quantified by RNA Qubit (Invitrogen) and 50ng of total RNA was used as template for 1st strand cDNA synthesis as previously described [7]. Tagged cDNA was diluted 1:10 in NFH₂O and quantified using soofast qPCR buffer (Bio-Rad) with PCR primers (PP_PCRI-5'CAGGACGCTGTTCCGTTCTATGGG3', PP_PCRII-5'CAGACGTGTGCTCTTCCGATCT3') on a CFX-96 (Biorad). Second-strand synthesis was carried out using Failsafe buffer E and limited PCR cycling conditions were determined by the qPCR ct threshold value for each cDNA reaction (~12 cycles/reaction). All double-stranded cDNA mosquito libraries were eluted in 15uL of NFH₂O and quantified by the high sensitivity DNA Qubit kit (Invitrogen).

1.3. Rare RNA Transcript Sequence Enrichment

All *Culex* mosquito double-stranded cDNA libraries were subjected to Hydroxyapatite chromatography normalization to enrich for low abundant microbial sequences [8]. Briefly, cDNA libraries (100ng) were added to 1x hybridization buffer (insert formula) supplemented with 20% formamide and incubated at 98C for 3 minutes to denature double-stranded cDNA (ds-cDNA) followed by 5 hours at 68C to re-anneal high abundance DNA sequences. All samples were immediately added to a microspin column (Pierce Inc) containing 100uL of rehydrated/defined DNA grade hydroxyapatite HTP Bio-Gel (Bio-Rad), incubated for 3 minutes at 65C on column, washed 2x with 10mM NaPB to remove any unbound cDNA fragments, and the single-stranded cDNA (ss-cDNA) fraction was eluted off the column with 30uL100mM NaPB. All steps involved in the HAC chromatography protocol were performed in an incubator at 65C. The ss-cDNA fraction was cleaned up using 48uL Ampure XP beads (BD), eluted in 25uL NFH₂O, and quantified by qPCR to determine optimal PCR cycling (between 16-20 cycles). Each library was assigned a unique barcode for multiplexing and custom Illumina® compatible sequencing adaptors with unique custom 9mer barcodes were incorporated by limited PCR based on qPCR ct values. 50uL PCR reactions containing 10uM of each adaptor primer (Adaptor I=5'AATGATACGGCGACCACCGAGATCTACACTTCGCTACAGGACGCTGTTCC GTTCTATGGG3' and Adaptor II=5'CAAGCAGAAGACGGCATAACGAGATXXXXXXXXXXGTGGTCAGACGTGTG CTCTTCCGATCT3') and Failsafe buffer E (Epicentre) were used to synthesize final libraries. The 9mer unique barcode sequences implemented in this study for multiplexing are available

All final cDNA libraries were size-selected using 0.7X/0.25X Ampure XP beads (Beckon Dickenson), eluted in 25uL NFH₂O, and quantified by the high sensitivity dsDNA Qubit kit (Invitrogen) according to manufacturer's instructions. 20ng of each mosquito library was combined (n=20) into two multiplexed cDNA libraries, concentrated on a IC spin column (Zymo), and the cDNA size distributions were visualized on the high sensitivity dsDNA chip using the bioanalyzer 2100 (Agilent technologies). Multiplexed cDNA libraries were quantified by Kappa qPCR (Kappa Biosystems) following manufacturer's instructions. All primers used to generate RNA-Seq libraries were HPLC purified (IDT).

1.4. Deep Sequencing and Raw Sequence QC

The 2 multiplexed libraries (n=20 mos pools/library) were loaded at 9PM with custom read 1 (5'ACACTTCGCTACAGGACGCTGTTCCGTTCTATGGG3'), custom read 2 (5'-GTGGTCAGACGTGTGCTCTTCCGATCT-3'), and custom index read (5'-AGATCGGAAGAGCACACGTCTGACCAC-3') primers spiked into the appropriate MiSeq/HiSeq primer wells and sequenced on an Illumina MiSeq (in-house) and HiSeq 2000 (Vincent Coates Genomic Library, UC Berkeley). 150 basepair paired-end (PE) libraries were generated for each MiSeq run and 100 basepair PE sequences were produced for the HiSeq runs. All libraries were inspected to ensure good quality, and custom Illumina compatible adaptors were removed.

1.5. Metagenomic Computational Pipeline

Following quality filter to remove low complexity sequences, all sequences were aligned to all *Culex* mosquito sequences via Bowtie-2 local aligner to remove high abundant eukaryotic and prokaryotic ribosomal and CDS RNA sequences, mosquito derived host sequences. All remaining unidentified sequences were subjected to *de novo* assembly using the Trinity assembler software program, all contigs larger than 500bps were compared to similar sequences found in the blastn/blastx databases. All identifiable contig information and the corresponding sequence contigs analyzed were visualized using Krona metagenomic software[9].

1.6. Quantitative PCR Conditions and Virus Primer Sequences

Total RNA extracted from each *Culex* mosquito pool (n=40) was utilized for all qPCR assays. Briefly, 50ng of total RNA was synthesized into cDNA and amplified using the I-Script 1-step RT-PCR kit with SYBR green under manufacturer's instructions (Bio-Rad). All viral consensus sequences generated by *de novo* assembly were used as templates for the qPCR primers and all virus primers were designed via PrimerQuest software (IDT). RT-qPCR primers were generated to confirm the presence or absence of targeted arboviral RNA in each mosquito pool or C6-36 cell culture supernatant. A select number of arboviral qPCR primer sets were used to screen additional mosquito pool RNA (n=50/site) for the presence of target sequence. All mosquito RNA was tested in duplicate with 50ng total RNA as input per reaction

using 10uL reactions and quantified on a CFX 96-well qPCR machine (BioRad). All samples that generated C(t) values greater than 35 were considered negative or below the detection limit.

1.7. Cell Culture and Viral Growth Kinetics

Select Culex mosquito pool supernatants were inoculated onto an Aedes albopictus mosquito cell line (C6/36 cells) to determine ability of divergent viruses to propagate in cell culture. Since all 39 Culex mosquito pools sequenced were WNV positive according to the qPCR biosurveillance screening assay, we used anti-WNV polyclonal antibodies (1:2 mosquito supernatant:anti-WNV sera), incubated at 4C overnight, and infected C6/36 monolayers in T-25 tissue culture flasks. Viruses were allowed to adsorb for 1.5 hours at 37C with 5% CO₂, the C6/36 cells were washed with PBS, and DMEM supplemented with 5% FBS was added to each flask. Viral supernatant was collect at timepoint 0 (fresh media on cells), 72 hours, and 168 hours frozen down at -80C prior to RNA extraction. All tissue culture experiments were performed in a BSL-3 certified laboratory at UC Davis. Cell culture RNA was extracted using RNAzol as described previously, and 50ng of total RNA per reaction (2X) was used for all viral qPCR assays that targeted RNA viruses that were identified in the original mosquito pool by deep sequencing.

1.8. Viral Genome Analysis and Coverage Maps

Naming of divergent virus genomes (n=13) was based on the various geographic locations where virus positive mosquito pools were collected with the exception of 4 viruses that were named after the mosquito species where the virus contig originated.

1.9. Spatial and Phylogenetic Analyses

All virus genome reference sequences used in the phylogenetic analyses were obtained from the NCBI GenBank sequence database. All amino acid and nucleic acid virus sequences were aligned by MUSCLE using the Geneious R6 software [10,11]. The best-fit DNA or protein model for each alignment was selected using MEGA5 analysis tools [12]. Alignments were visually inspected and misaligned regions were removed from downstream processing. Maximum Likelihood (ML) trees were generated using 1000 bootstrap replicates via the PhyML plugin program in the Geneious R6 software package [13].

2. RESULTS

2.1. Arboviral Sequence Enrichment and Genome Coverage

We compared culex mosquito RNA-Seq libraries with and without HAC normalization to determine the viral sequence enrichment when utilizing this methodology (table ??). Six culex mosquito pools (K3-quinq, K9-tar, S4-pip, S9- tar, G3-quinq, C7-tar) representing three culex species from 4 different locations were used in this analysis. Total sequence reads from each culex RNA-Seq library (-HACN, +HACN) were aligned to viral genomes identified in each pool from virominer deep sequencing results. Mean fold enrichment of virus sequences was 36-fold (range: 6-fold to 85-fold) and the magnitude of enrichment was dependent on the HACN effectiveness for each mosquito RNA-Seq library evaluated. We also tested the overall mean % coverage of each viral genome identified in each mosquito RNA-Seq library (figure ??). In all viral genomes analyzed (n=33), HACN treatment of the mosquito samples increased the raw number of reads that mapped to viruses and the overall mean coverage across each virus genome when compared to untreated mosquito RNA-Seq libraries generated with 3-5 times more sequence reads.

2.2. Culex Microbial Metagenomic Overview

We evaluated the overall microbial composition of culex mosquito species collected from four distinct geographic locations (Coachella, Kern, Greater Los Angeles, and Sacramento) in California. The microbial populations present in these culex mosquitoes were grouped by species and location to obtain a high level assessment of the various bacteria, parasites, and viruses circulating in each region. In addition to West Nile virus we detected arboviruses in all 39 mosquito pools we sequenced on the HiSeq and MiSeq systems. The greater depth obtained from the order of magnitude more sequences obtained on the HiSeq run.

2.3. Viral Contig Coverage Maps and Genome Analysis

We determined sequence homology by blast, genome structure, and genome size for all the viral contigs generated using virominer. Known endemic California arboviruses such as West Nile virus, Culex flavivirus, Umatilla virus, Negev virus, Hart park virus, Culex pipiens pallens densovirus, were found circulating in California culex mosquito populations. In addition, Nam Dinh virus, a member of the genus alphamesonivirus, previously isolated from Vietnam and China, was identified in California culex mosquitoes. By comparing viral contig sizes and blastx results, we identified 13 divergent viral genomes that contained low sequence homology (22% to 56% amino acid identity) to known viruses in GenBank. These viruses spanned 7 different virus families (Bunyaviridae, Iflaviridae, Flaviviridae, Orthomyxoviridae, Reoviridae, Rhabdoviridae, and Tymoviridae) as well as unassigned virus groups that include Negevirus, Sobemovirus, and Nudivirus. Viruses that encompass these families are known to infect invertebrates, vertebrates, and plants. Furthermore, they represent all classes of ssRNA(+ and – sense), dsRNA, ssDNA, dsDNA viruses that were sequenced using a RNA-Seq library preparation and HAC sequence enrichment protocols. All near full-length virus contigs

generated by Virominer that had sequence homology to known virus proteins in GenBank and similar gene structure/size were included in further virome analyses. Sequence coverage maps for each near full length viral contig were constructed by mapping the raw sequences from mosquito pools where consensus viral contigs were assembled by Trinity. No evidence of virus recombination events due to the de novo assembly were observed, the mapped sequences were evenly distributed across the viral contigs, and each virus genome had at least 10X sequence coverage (10X to 1000X range) for any given nucleotide position.

2.4. Virus Classification and Genome Annotation of Divergent Arboviruses

The near-full length virus genomes were translated and compared to all protein sequences available in GenBank using blastx. Based on amino acid sequence similarities, virus contigs were assigned to closest virus family, contig size ranges were determined for a match, and in some cases phylogenetic analyses were performed. The divergent virus genomes (<75% amino acid identity) were annotated in an attempt to further characterize each virus genome structure and to validate de novo assembled sequence contigs.

2.5. Spatial Analysis of California Culex Virome

All viral contigs identified by Blastx that represented near and/or full length genomes in the Cx. mosquito sample set were compiled and raw reads from each sample were mapped to each viral genome. By grouping mosquito pools based on their virome composition, the mosquito pools formed two distinct clades based on mosquito species; the culex tarsalis and the culex quinquefasciatus /pipiens complex groups. Arthropod-borne viruses such as, Salton Sea virus, Mecca virus, Umatilla virus, Coachella virus, Elk Grove virus and Culex tarsalis Rhabdovirus, were only found in Cx. tarsalis mosquitoes while Wilton virus, Culex pipiens densovirus, Echo Park virus and Negev virus were only in Cx. quinquefasciatus/pipiens complex mosquitoes. In contrast, viral sequences that mapped to West Nile virus, Culex bunyavirus, Nam Dinh virus, Box canyon virus, Wilton virus, and Cx. Tymovirus were found in both Cx. quinquefasciatus and Cx. tarsalis mosquito pools, but only from pools collected in Southern California, GRLA and COAV respectively. Culex flavivirus sequences were confined to Cx. mosquito populations in Kern and Sacramento while Culex bunyavirus sequences, Wilton virus and West Nile virus were identified in all samples sequenced.

2.6. Culex Bunyavirus Sequence Distribution and Geophylogeny

In order to determine the prevalence of culex bunyavirus in the California culex mosquito populations we designed Sybr green qPCR primer sets against the S, M, and L segments and tested 50 additional culex mosquito pools from each location. In concordance with the deep sequencing data, all the additional culex mosquito pools analyzed were positive for the culex bunyavirus L segment except two mosquito pools. The S segment was detected only in culex

mosquito pools that originated in Kern (X/50) and SAYO (X/50) while all mosquito pools from COAV (0/50) and GRLA (0/50) were negative. The culex mosquito pools that were positive for the S segment in Kern (X/50) and SAYO (X/50) were also positive for the M segment and all mosquito pools from COAV (0/50) and GRLA (0/50) were negative. A full-length genome alignment of culex bunyavirus L segments (n=36) was generated to determine if these viruses have acquired adaptive mutations that are restricted by species and/or geographic location. Geophylogenetic analysis of all full-length novel bunyavirus L segments generated by the virominer pipeline was performed (Figure ??) . When comparing the full-length L segments isolated in different mosquito species, *Cx. tarsalis*, *Cx. pipiens*, and *Cx. quinquefasciatus*, they formed two distinct overall clades. Viral sequences obtained from *Cx. tarsalis* mosquito pools formed one major clade and all *Cx. quinquefasciatus* /*Cx. pipiens* pools grouped into another clade regardless of location. Although the *Cx. pipiens* mosquito pools that contained full-length viral L segments were limited (n=2), they still formed a subclade that clustered with the Kern *Cx. quinquefasciatus* phlebovirus L segment sequences. Bunyavirus L segments isolated from *Cx. tarsalis* mosquito pools formed distinct subclades within the major *tarsalis* clade indicating adaptive mutations were present. GRLA *Cx. quinquefasciatus* L segments clustered into a distinct subclade as well and the bunyaviruses from Kern *Cx. quinquefasciatus*.

3. DISCUSSION

Mosquito surveillance to monitor vector-borne pathogens circulating in geographically defined regions allows for rapid and targeted vector control measures to lower mosquito populations therefore minimizing the organism's epidemic/epizootic potential.

4. REFERENCES

1. Ng TF, Willner DL, Lim YW, Schmieder R, Chau B, et al. (2011) Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. *PLoS One* 6: e20579.
2. Bishop-Lilly KA, Turell MJ, Willner KM, Butani A, Nolan NM, et al. (2010) Arbovirus detection in insect vectors by rapid, high-throughput pyrosequencing. *PLoS Negl Trop Dis* 4: e878.
3. Hall-Mendelin S, Allcock R, Kresoje N, van den Hurk AF, Warrilow D (2013) Detection of arboviruses and other micro-organisms in experimentally infected mosquitoes using massively parallel sequencing. *PLoS One* 8: e58026.
4. Runckel C, Flenniken ML, Engel JC, Ruby JG, Ganem D, et al. (2011) Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, *Nosema*, and *Crithidia*. *PLoS One* 6: e20656.
5. Hall RJ, Wang J, Todd AK, Bissielo AB, Yen S, et al. (2013) Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *J Virol Methods*.
6. Marklewitz M, Zirkel F, Rwego IB, Heidemann H, Trippner P, et al. (2013) Discovery of a unique novel clade of mosquito-associated bunyaviruses. *J Virol* 87: 12850-12865.
7. Langevin SA, Bent ZW, Solberg OD, Curtis DJ, Lane PD, et al. (2013) Peregrine: A rapid and unbiased method to produce strand-specific RNA-Seq libraries from small quantities of starting material. *RNA Biol* 10: 502-515.
8. Vandernoot VA, Langevin SA, Solberg OD, Lane PD, Curtis DJ, et al. (2012) cDNA normalization by hydroxyapatite chromatography to enrich transcriptome diversity in RNA-seq applications. *Biotechniques* 53: 373-380.
9. Ondov BD, Bergman NH, Phillippy AM (2011) Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12: 385.
10. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647-1649.
11. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792-1797.
12. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.
13. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307-321.

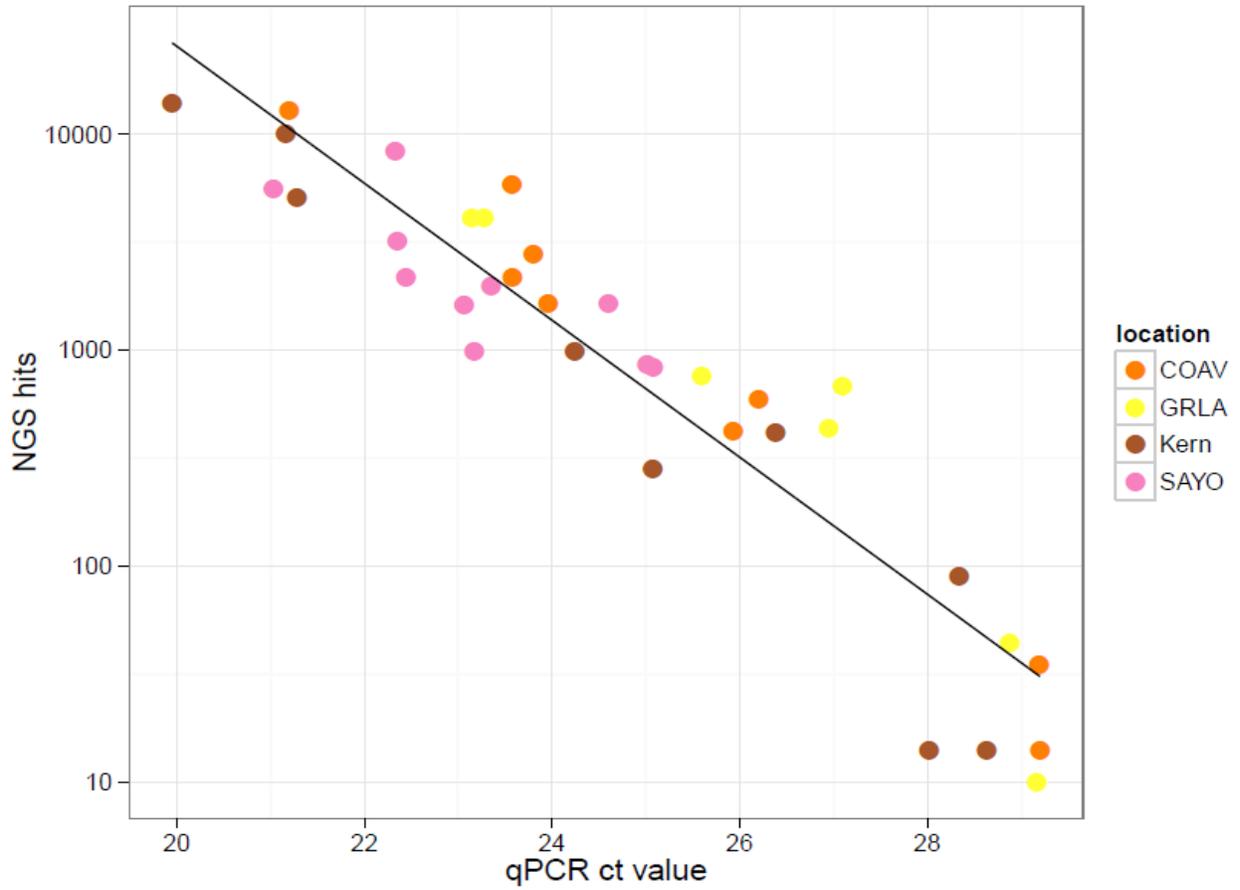


Figure 2. Caption

Sampl	Total reads (million) UT/H	Virus	% reads mapped	% reads mapped (HA)	genome coverage	genome coverage (H)	Fold enrichme
K3-qinq	2.64/0.759	Nam Dinh virus	0.23	3.2	99	100	13.9
K3-qinq		Negev virus	0.02	0.2	96	100	10
K3-qinq		CA Phlebovirus	0.006	0.16	72	100	26.6
K3-qinq		CA Negevirus	0.0023	0.52	91	100	22.6
K3-qinq		CA Ifavirus	0.00026	0.012	6	46	46
K3-qinq		Culex flavivirus	0.00011	0.0032	3	16	29.1
K9-tar	1.87/0.790	CA Sobemovirus	0.003	0.164	65	100	54.6
K9-tar		Flanders virus	0.0069	0.505	52	100	73.2
K9-tar		CA Phlebovirus	0.0028	0.206	48	99	73.6
K9-tar		CA Ifavirus II	0.0051	0.418	44	100	81.9
K9-tar		West Nile virus	0.0015	0.0347	24	78	23.1
S4-pip	2.49/0.849	CA Negevirus	0.0305	0.349	99	100	11.4
S4-pip		Negev virus	0.0778	0.449	99	100	5.8
S4-pip		Nam Dinh virus	0.0548	0.648	95	100	11.8
S4-pip		Culex flavivirus	0.00587	0.0592	61	98	10.1
S4-pip		CA Phlebovirus	0.0042	0.059	50	98	14
S4-pip		West Nile virus	0.00422	0.0429	49	96	10.1
S9-tar	2.98/0.928	CA Phlebovirus	0.0068	0.064	74	100	9.4
S9-tar		Nam Dinh virus	0.0084	0.17	34	98	20
S9-tar		West Nile virus	0.0013	0.014	20	69	10.8
S9-tar		Culex flavivirus	0.0013	0.013	10	50	10
G3-qinq	1.69/0.432	Nam Dinh virus	0.033	2.22	81	100	67.3
G3-qinq		CA Phlebovirus	0.0054	0.3	72	100	55
G3-qinq		CA Chrysovirus	0.0047	0.26	65	100	55.3
G3-qinq		Culex densovirus	0.00053	0.015	19	81	28.3
G3-qinq		CA Negevirus	0.0012	0.068	15	87	56.6
G3-qinq		West Nile virus	0.001	0.02	13	49	20
G3-qinq		CA Ifavirus II	0.00035	0.015	4	34	42.8
C7-tar	1.78/0.485	CA Ifavirus II	0.019	1.34	71	100	70.5
C7-tar		CA Sobemovirus	0.0055	0.382	61	100	69.4
C7-tar		CA Phlebovirus	0.0047	0.402	59	100	85.5
C7-tar		CA Rhabdovirus	0.094	4.94	45	100	52.5
C7-tar		CA Tobamovirus	0.00041	0.016	3	55	39

Figure 3. Caption

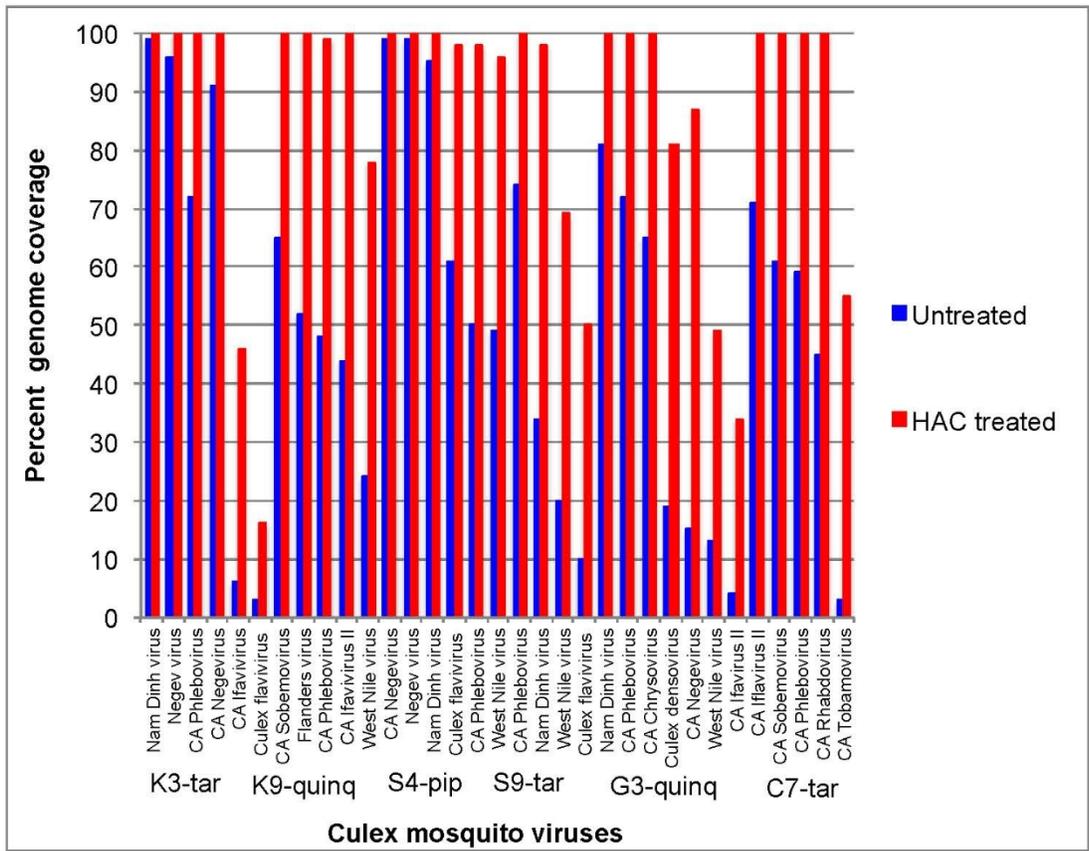


Figure 4. Caption

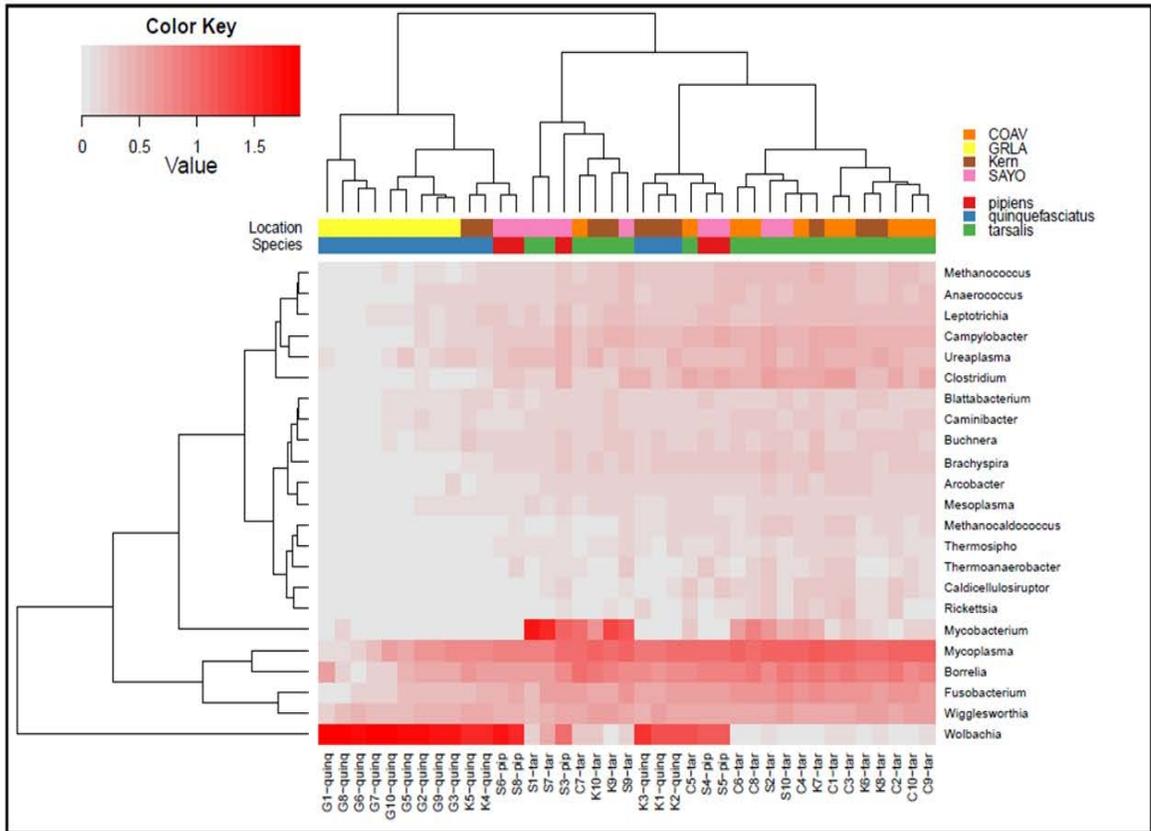


Figure 5. Caption

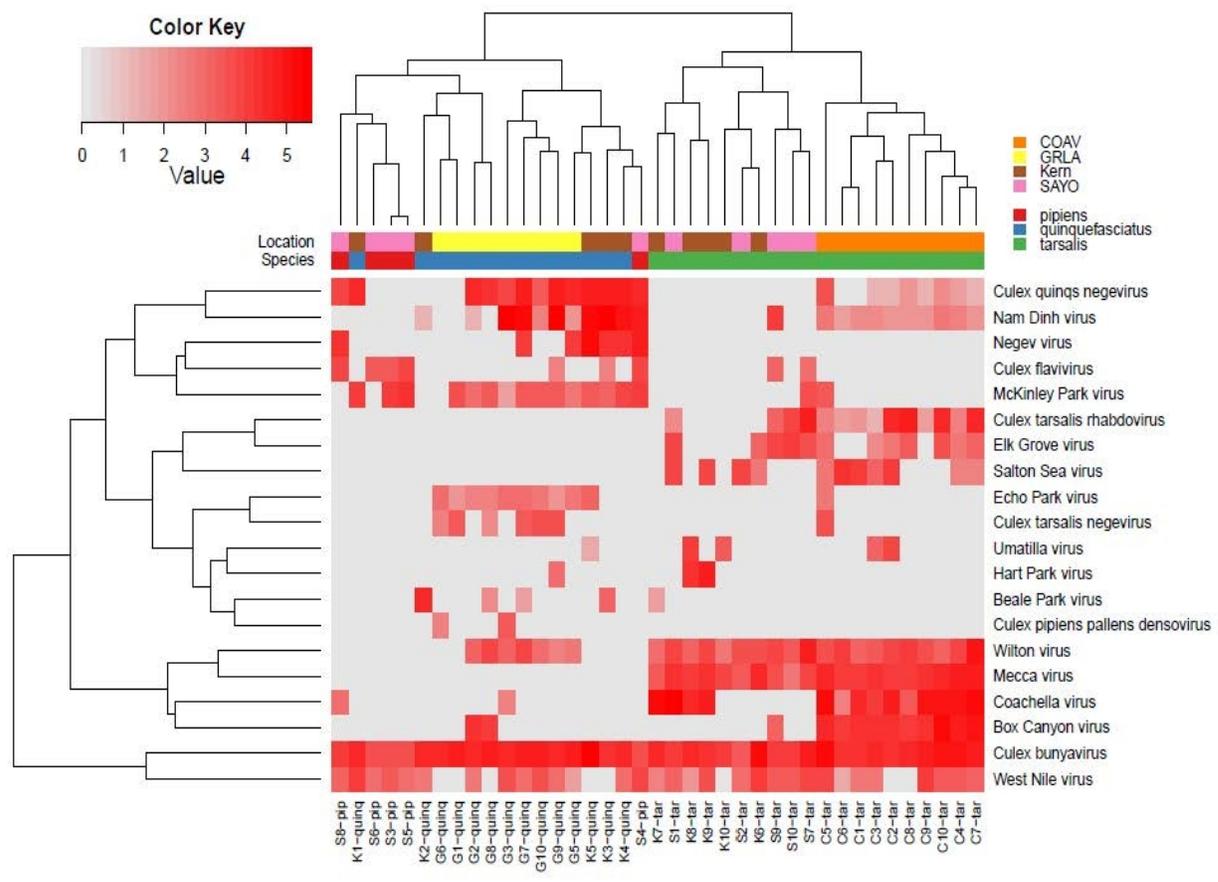


Figure 6. Caption

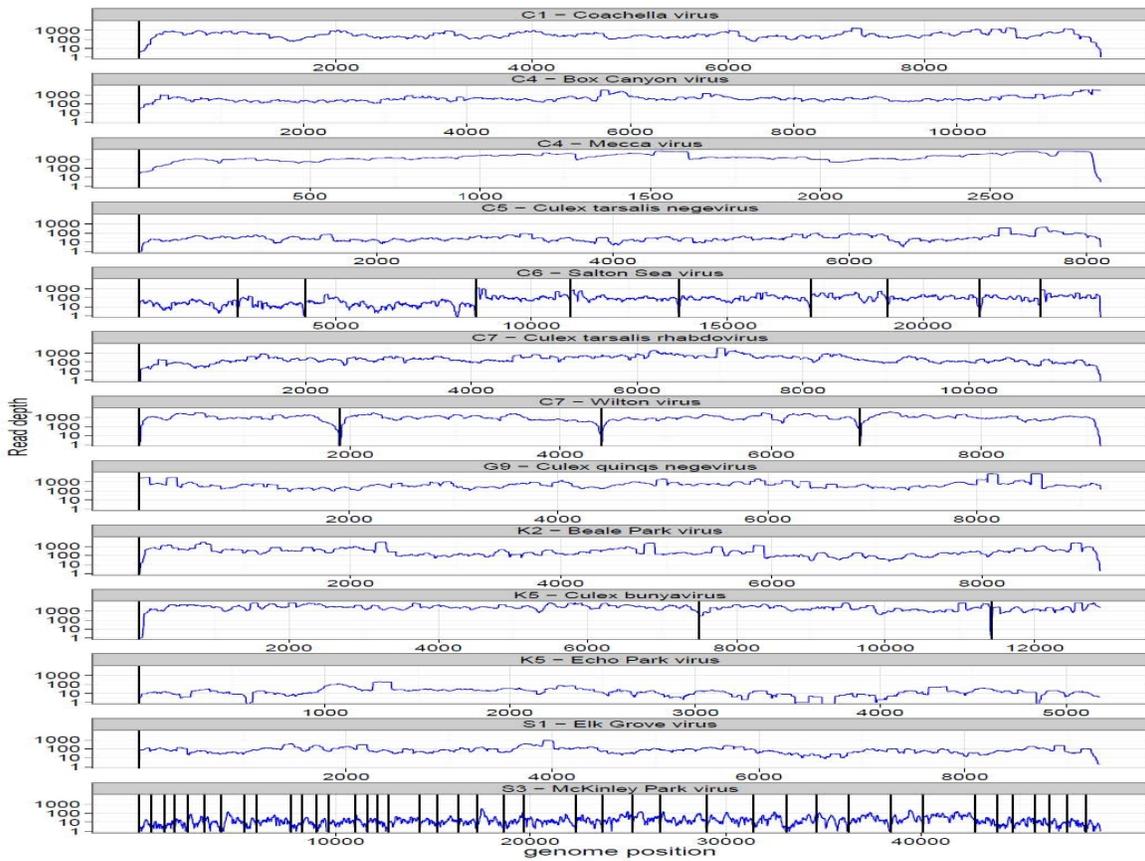


Figure 7. Caption

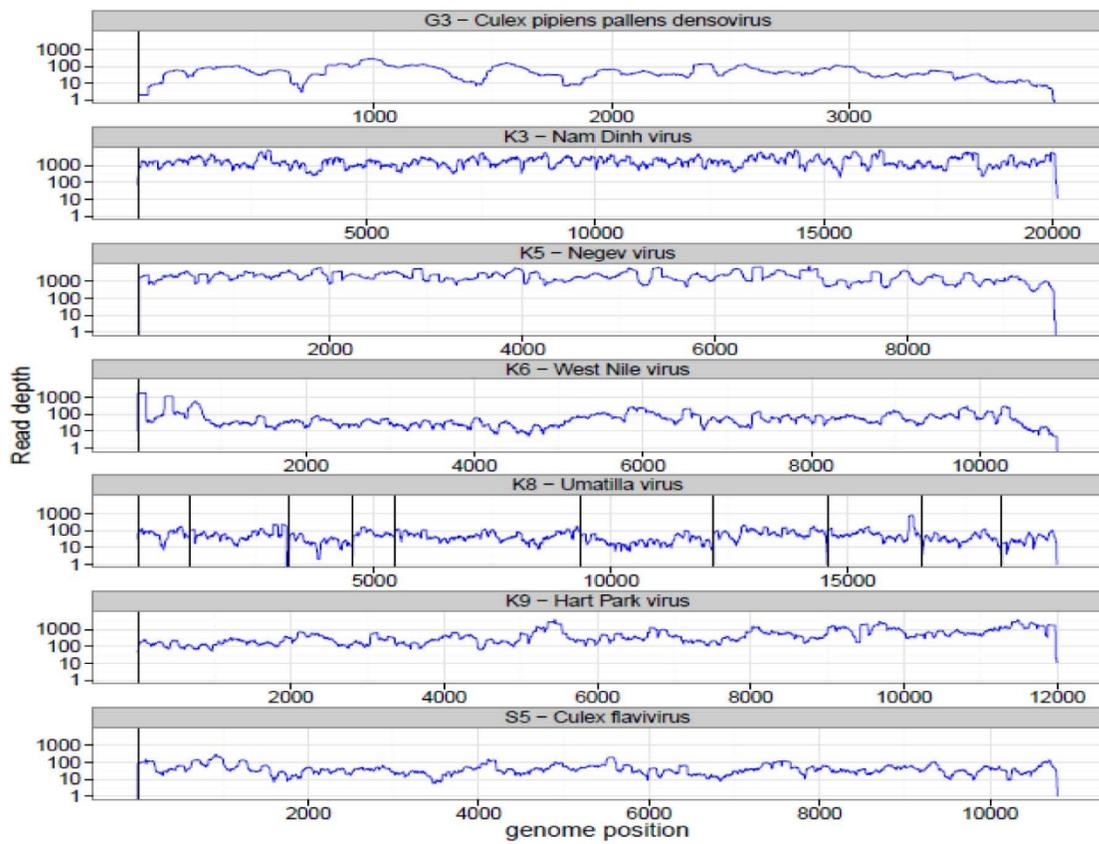


Figure 8. Caption

DISTRIBUTION

1	MS1363	B. Brodsky	6825
1	MS9291	C. Branda	8623
1	MS9291	C. Koh	8621
1	MS9292	T. Lane	8623
1	MS9292	V. VanderNoot	8621
1	MS9671	D. Curtis	8623
1	MS0899	Technical Library	9536 (electronic copy)
For LDRD reports, add:			
1	MS0359	D. Chavez, LDRD Office	1911
For CRADA reports add:			
1	MS0115	OFA/NFE Agreements	10012
For Patent Caution reports, add:			
1	MS0161	Legal Technology Transfer Center	11500



Sandia National Laboratories