Journal of the American Mosquito Control Association, 41(1):11–19, 2025

COMPARISON OF NANOPORE AND CLASSICAL SANGER SEQUENCING TO IDENTIFY MOSQUITO BLOODMEAL HOSTS

LINDA KOTHERA, SPENCER J. HILLS, NICHOLAS A. PANELLA AND AARON C. BRAULT

U.S. Centers for Disease Control and Prevention, Division of Vector-Borne Diseases, 3156 Rampart Road, Fort Collins, CO 80521

ABSTRACT. The tools available to vector control districts (VCDs) to collect mosquito surveillance data are constantly evolving. As more VCDs obtain real-time polymerase chain reaction (PCR) instruments and the costs associated with computing power and next-generation sequencing continue to decrease, the option of generating useful molecular data in-house becomes more viable. Measures such as arbovirus testing and genotyping for insecticide resistance mutations using RT-qPCR, and identifying species used for mosquito bloodmeals with next-generation sequencing or Sanger sequencing are examples. In this study we identify mosquito host bloodmeal species using Nanopore sequencing from Oxford Nanopore Technologies. We used MinION and Flongle flow cells and a Mk1C device to sequence 96 barcoded amplicon samples in a single sequencing run, and share details of data analysis using the free-to-use Galaxy bioinformatics platform. After sequencing the same samples with Sanger sequencing, we conclude that Nanopore sequencing is better at identifying species in mixed bloodmeals. This work demonstrates a potential use of nanopore sequencing by VCDs with basic biology laboratory and computing equipment.

KEY WORDS Barcode gene, bloodmeal ID, COI gene, mixed species bloodmeals, mosquito bloodmeal host species identification, Nanopore sequencing, Sanger sequencing

INTRODUCTION

Diseases caused by arboviruses such as West Nile virus, Saint Louis encephalitis virus, and LaCrosse virus negatively impact human health, and the control of mosquito vectors of these viruses often relies on the use of chemical insecticides. Continued effective use is constrained by factors such as the financial resources needed to purchase and properly apply the product, as well as the need to manage insecticide resistance. Resistance can arise when insecticides are overused or misused, and its presence makes vector control challenging. Effective decisions regarding chemical control are the result of conducting surveillance to procure data such as the infection rate in mosquitoes, the presence and degree of resistance, and whether humans are being used as bloodmeal hosts. When robust surveillance data are in the hands of vector control districts (VCDs), decisions about where and when to spray can be made that slow the development of resistance.

Many VCDs possess the capacity to procure quality surveillance data. As well, partnerships sometimes exist with entities such as state public health laboratories, or local universities who offer assistance with obtaining data on the metrics named above. However, such partnerships are not ubiquitous. As the need for generating good surveillance data has increased, the costs associated with performing molecular work such as qPCR (quantitative polymerase chain reaction) and deoxyribonucleic acid (DNA) sequencing have come down. Thus, a case can be made that conducting this kind of work has become more feasible for VCDs over time.

Here, we focus on assays to conduct bloodmeal host species identification. Mosquito feeding behavior, including host preference, is thought to have a genetic basis (Hamer et al. 2009, Takken and Verhulst 2013, Noreuil and Fritz 2021). However, some species are also opportunistic feeders and will take a bloodmeal from a human when their preferred hosts are absent. Making host bloodmeal identifications can help determine which species are amplifying hosts for emerging or under-studied arboviruses (Hannon et al. 2019) and can also contribute to estimates of risk of arboviral transmission to humans (Riccetti et al. 2022).

Early methods to conduct host bloodmeal species identification used a variety of serological techniques reflecting that DNA sequencing was mostly unavailable (Kent 2009). Modern approaches generally utilize PCR-based assays that detect DNA sequence differences among vertebrate species to make bloodmeal host identifications, although other techniques are also used such as mass spectrometry (Tandina et al. 2020) and enzyme-linked immunosorbent assay (ELISA; Lardeux et al. 2007). Among the targets of PCR-based assays are mitochondrial genes and regions such as the Cytochrome B gene (Kocher et al. 1989), 12S and 16S regions (Cawthorn et al. 2012), the NADH1 gene (Saccone et al. 1999), the d-loop (Nicholls et al. 2014) and the cytochrome oxidase subunit 1 (COI) gene, also referred to as the Barcode Gene (Folmer et al. 1994, Hebert et al. 2003). Advantages of using the COI gene include a phylogenetically informative rate of mutation, a lack of indels that could hamper alignments (Doyle et al. 2000), and an extensive database of curated sequences [the Barcode of Life Database (BOLD); (Ratnasingham and Hebert 2007)] readily available for comparison.

Methods for sequencing the results of host bloodmeal identification assays include Sanger sequencing

and next-generation sequencing (NGS). Sanger sequencing is methodologically straightforward, but requires instrumentation beyond a regular PCR thermal cycler, as well as trained personnel to run and maintain the instrument. Sanger sequencing produces one forward and one reverse read per sample, which are computationally assembled into a consensus sequence. That sequence is Basic Local Alignment Search Tool (BLAST) searched on the NCBI (National Center for Biotechnology Information) website (Altschul et al. 1990) to look for matches to sequences in the GenBank database. Ideally, BLAST searches return a result with a high percent identity, indicating the sequence in the bloodmeal DNA closely matches a species in the database. A drawback of directly Sanger sequencing PCR amplicons is its inability to detect mixed bloodmeals, that is, samples where a mosquito has fed on more than one host. One way to address this issue is to clone amplicons into plasmid vectors and sequence several colonies per sample. However, the process requires additional time, equipment, and reagent resources. Host bloodmeal identification data are relevant to surveillance efforts. Samples with mixed bloodmeals are often uninterpretable or do not sequence properly because there are two organisms' DNA in the sample, and thus are failed samples.

Nanopore sequencing is an emerging NGS platform from Oxford Nanopore Technologies (ONT) that lends itself well to multiplexed amplicon sequencing (Makunin et al. 2022, Whitford et al. 2022). Analysis tools are available on several free or mostly-free to use platforms including Linux command line, the free-to-use bioinformatics platform Galaxy (Galaxy Community 2024) and ONT's own Epi2Me platform available on GitHub (Oxford Nanopore Technologies 2024a). Although several sequencing devices are available from ONT, we focus here on MinION R10.4.1 flow cells and their associated devices in this study. Samples for amplicon sequencing are prepared from an initial PCR reaction, followed by library preparation with proprietary library kits and loaded onto flow cells for sequencing. MinION devices accept either the standard sized MinION flow cells, which can be washed and reused for up to approximately 72 h of run time, or the smaller Flongle R10.4.1 flow cells, which have approximately one tenth of the capacity and cost, and are discarded after use. Requirements to conduct amplicon sequencing on the ONT platform after purchasing a device include a basic molecular biology laboratory equipment, a regular thermal cycler and a computer that meets ONTs IT requirements (Oxford Nanopore Technologies 2024b). The low cost of entry to using Nanopore sequencing, as well as support and resources available from ONT aimed at new users, represents an opportunity for VCDs to conduct a portion of their own molecular surveillance.

We demonstrate an example of such surveillance with host bloodmeal species identification and Nanopore sequencing. We then share a data analysis workflow using Galaxy to create *de novo* assemblies of Nanopore sequencing reads to identify species in vector mosquito bloodmeal homogenates. Our objective was to demonstrate that Nanopore sequencing produced results equivalent or better than Sanger sequencing.

MATERIALS AND METHODS

Sample collection and processing

Deoxyribonucleic acid was extracted from 90 field collected, single specimen bloodmeal homogenates left over from previous surveillance efforts and stored at −80° (six of these were used twice to make 96 total samples that would utilize all available barcodes in the kit). Samples from New Jersey were part of a previously published study (Fagre et al. 2024), whereas samples from Arizona and Texas were from unpublished surveillance work and had no metadata associated with them besides year and state in which they were collected. Samples were prepared from individual blooded abdomens that had been removed from the rest of the body and placed into individual 2ml tubes. To each sample was added one copper BB and 1ml of BA-1 diluent (made in-house at CDC), and samples were homogenized for 4 min at 20 shakes per min on a Qiagen TissueLyser. The DNA was extracted from 300µl of cleared homogenate using a Kingfisher Flex instrument with a Mag Max DNA Ultra 2.0 Multi Sample kit (Thermo, Waltham, MA) and the kit's protocol to perform DNA purification with 96 deep well format plates. Purified DNA was eluted with 80ul of EB buffer.

Bloodmeal host identification

Host bloodmeal identification PCR of COI gene amplicons proceeded with primers described initially by Ivanova et al. (2007; Supplemental File 1) and PCR as in Crabtree et al. (2013) with the following minor changes: we used Amplitaq Gold 360 2x Hot Start PCR master mix (Thermo), and PCR was performed on a Bio-Rad T1000 thermal cycler (Bio-Rad, Hercules, CA) with the protocol adjusted to include a 3 min hot start at 95°C. The PCR products were cleaned using a 0.7x bead ratio of Ampure XP beads (Beckman Coulter, Brea, CA) and the purified PCR product was eluted in a final volume of 23ul EB buffer from the DNA Ultra 2.0 extraction kit. Sample concentration was assayed with a Qubit Flex fluorometer with DNA BR or HS reagents (Thermo) and a portion was diluted to the proper concentration for subsequent Nanopore sequencing, or used undiluted for Sanger sequencing.

Sanger sequencing followed a standard protocol where 3ul bead-cleaned PCR product served as template in each of two 20µl sequencing reactions consisting of 8ul Big Dye master mix, 0.65µl 10uM forward or reverse primer cocktails as per Ivanova et al. (2007) and 8.35µl molecular grade water. Sequencing PCR reactions were conducted on a Bio-Rad T1000 thermal cycler

Table 1. Collection information for bloodmeal homogenates used in this study.

Poulation abbreviation	State	N	Year collected
AZ	Arizona	15	2015
NJ	New Jersey	42[1]	2021
TX	Texas	33	2013

[1] Six NJ specimens were used with two different barcodes to make N = 96 samples.

using the following program: 96° C for 1 min, followed by 25 cycles of [96° C for 10 sec, 50° C for 5 sec, 60° C for 4 min] and a 4° C hold. Sequencing was performed on an ABI 3500XL instrument (Advanced Biosystems Inc., Foster City, CA). The resulting forward and reverse sequencing reads per sample were assembled into a contig using SeqMan Pro (Lasergene, Madison, WI) and exported as a fasta file, which was BLAST searched to make host bloodmeal species identifications. The top hits were examined to eliminate closely related species not expected to be present in the collection location, and the remaining top hit was deemed the result.

We performed Nanopore sequencing using the ONT Native Barcoding 96 kit (SQK-NBD114.96; ONT, London, UK) and a protocol (Supplemental File 2) specifically for sequencing pooled, barcoded amplicons, which required the purchase of reagents from New England Biolabs (NEB; Ipswich, MA). We prepared libraries for 96 samples (Table 1) and pooled them in a single sequencing run. Briefly, 200 fmol of purified PCR product was end-prepped using the NEB kit reagents, followed by barcode ligation, pooling into one sample, and ligation of sequencing adapters. Libraries were diluted to obtain the recommended mass of 5-20 fmols (depending on flow cell size), loaded onto flowcells and sequenced. We used a Mk1C device and a MinION flow cell, and repeated the run with the same library on the same device using a smaller capacity Flongle flow cell. Each sequencing run was performed with the default settings (i.e., the Fast basecalling option). The MinION flow cell was run for 4 h and the Flongle for 15 h. Data were output in the proprietary POD5 file format.

Nanopore sequencing data analysis

Data were rebasecalled with the command line version of the basecaller *Dorado* (Oxford Nanopore Technologies 2024c) to increase accuracy of the results on a system with faster processing speed than was locally available. We used the default HAC (high accuracy basecalling) model with options to retain reads with quality scores ≥ 15. We did not guide basecalling with a reference sequence, as we wanted to be able to detect mixed bloodmeals. The rebasecalled data were output initially as a single BAM file, which was subsequently demultiplexed by barcode with *Dorado* using the *demux* command that included the *–emit-fastq* option to output one fastq file per barcode for downstream analyses. The *demux* command proceeds using a modified Needleman-Wunsch method, which examines each read for each 24 bp barcode sequence, scoring along the way for similarity. The reads having the highest scores for a particular barcode are binned together as being from the same sample. As a further quality metric, we selected the option *–barcode-both-ends*, which carries forward only reads that have the same barcode on either end of the read. The last step of the *demux* command trimmed barcodes and adapters from all reads. Finally, we used the command *summary* to generate a tab-delimited text file that contained run metrics. Details about the code used for analyses is in Supplemental File 3.

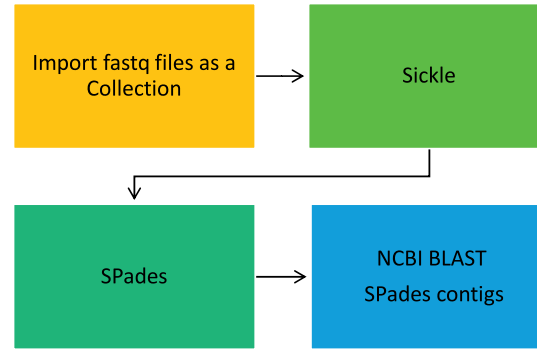

Fig. 1. Representation of the data analysis pipeline developed for this study.

The basecalled fastq files were uploaded to Galaxy, along with the summary file. The tool *pycoQC* (Leger and Leonardi 2019) was used to generate run metrics from the *Dorado* summary file. We made a Galaxy workflow (Fig. 1; link) with the following steps to generate *de novo* assembly contigs. First, the tool *Sickle* (Joshi and Fass 2001) was used to remove short (<100bp) reads. Next, the trimmed reads were submitted to the *de novo* assembler *SPades* (Bankevich et al. 2012) with the following options: careful correction, automatic selection of k-mers and a user-specified coverage cutoff of 20, which produced contigs with at least 20x coverage.

The first 10 contigs for each barcode were submitted to NCBI BLAST, using the megablast option. The results were examined for species identifications with the highest percent identity with alignment lengths close in size to the PCR amplicon (approximately 700 bp). As well, results were examined for evidence of mixed bloodmeals, where more than one species was identified with ≥98% identity across the 10 contigs.

RESULTS

Nanopore *de novo*

The results were similar between the two different sized flow cells (data not shown) in terms of number of reads generated and read quality. Thus, going

forward we present results for the run conducted on a MinION flow cell. Because Flongle flow cells are disposable, we let them run overnight and retrieved the data the next day. Some of the run metrics, specifically the translocation speed and the read quality, stayed constant for the first 6–8 h of the run and then became more variable. Therefore, data were retrieved from the prior time frame only. By comparison, it took approximately 4 h of running the MinION flow cell to obtain sufficient reads to generate unambiguous consensus sequences.

The *pycoQC* tool indicated a total of 883,574 reads were generated for the set of 96 samples with an N50 of 708 bp, a median read length of 704 bp and a median PHRED score of 17. The number of reads per barcode after the initial trim steps during *Dorado* (basecalling that removed reads with a quality score of <15 and a second trim after the *demux* command to remove barcodes and adapters) ranged from 3,373 to 14,569. After trimming with *Sickle*, the number of reads per barcode ranged from 2,871 to 14,078. *SPades'* ability to put distinct species into separate contigs, which facilitated the detection of mixed bloodmeals, appeared to decrease when fewer, larger contigs were made by increasing the depth of coverage parameter past 25, so the original parameters stated above were retained. The BLAST searching more than the first 10 contigs did not identify additional bloodmeal host species in our samples.

Supplemental File 4 presents length and coverage data for the contigs generated by *SPades*, as well as additional details of the BLAST hits including accession numbers, E-values and lengths of alignments. Contig size in samples where a positive species identification was made ranged from 273 bp to 1694 bp, and coverage ranged from 15X to 4989X.

Nanopore versus Sanger sequencing

A direct comparison of bloodmeal host species identifications from Sanger and Nanopore sequencing is shown in Table 2. All identified species were birds. New Jersey and Arizona had approximately the same diversity of species, where 15 species were observed among the 42 New Jersey samples, and five species identified for the 14 Arizona samples. There was slightly less diversity in the Texas samples, with 10 species identified for 34 samples. There were differences between the two sequencing platforms in terms of samples from which no identifications could be made. A total of 15 Sanger sequenced samples failed to produce viable BLAST identifications, whereas fewer (five) samples failed with Nanopore sequencing. All of the samples that failed with Nanopore sequencing also failed with Sanger sequencing, suggesting the presence of several degraded bloodmeals from which viable DNA was not obtainable. Between 5% and −20% of the New Jersey samples, collected in 2021, failed (2 with Nanopore sequencing and 9 with Sanger sequencing), suggesting that the age of the sample was not necessarily a factor in whether a sample would fail. By comparison the Texas samples, collected in 2013 had 4 failures with Sanger sequencing (12%) and three failures with Nanopore sequencing (9%).

For samples where at least one bloodmeal host was identified, the percent identity values obtained with Sanger sequencing ranged from 80.7 to 99.8%. The Nanopore sequencing values were higher, ranging from 93.7 to 100%. Interestingly, the two lowest percent identities were for the same species, *Poecile carolinensis* (Audobon) with the remaining values $\geq$ 98%. A total of 26 Nanopore sequenced contigs returned BLAST hits with 100% identity. The length of the BLAST alignment using Sanger sequencing ranged from 247bp-712bp, and for Nanopore sequencing the length ranged from 145bp-725bp. The E-values for BLAST hits associated with Nanopore sequencing were generally closer to 0.0 than for Sanger sequenced samples. The species identifications were congruent between sequencing platforms where an identification was made.

Nanopore sequencing and subsequent *de novo* assemblies detected mixed bloodmeals in 24/90 unique homogenates (27%; barcodes 91–96 were repeats of samples already in the study). In mixed bloodmeals, a 2nd species identification was well supported with a high percent identity value. For example, across the 1st and 2nd species determinations, only 2/116 Nanopore identifications had a percent identity $<$ 98%. In 1 case (barcode23), 3 species were identified.

DISCUSSION

Host bloodmeal identifications are one of the factors VCDs take into account when making decisions about chemical insecticide application locations and timing. The degree to which humans are at risk for contracting arboviral diseases can be estimated from these and other kinds of surveillance data, which in turn allows a more targeted and responsive approach to using chemical insecticides that minimizes the development of resistance. In this study, we demonstrated a workflow using Nanopore amplicon sequencing coupled with straightforward data analysis to identify the bloodmeal host species present in a set of mosquito homogenates. Nanopore sequencing was superior to Sanger sequencing in terms of its ability to detect mixed bloodmeals, and it produced contigs that returned BLAST hits with higher percent identities.

Although most of the samples returned BLAST hits with high percent identities, some samples failed to yield any viable sequence. This could be because of the length of time between when the mosquito took the bloodmeal, and when it was collected. Since blood is digested to promote egg development, it degrades over time. Another reason a sample may have failed is that the mosquito took a small bloodmeal the volume of which was insufficient to serve as template for our PCR reaction (Kent and Norris 2005).

Table 2. Mosquito host bloodmeal species identifications made with Sanger sequencing (columns 3–5) and Nanopore sequencing (columns 6–15). An expanded version of this table that includes E- values, further contig coverage details, and additional BLAST hits can be found as Supplemental File 4.

		Sanger sequencing			ONT sequencing								
					First contig blast hit[1]				Second contig blast hit				
ONT Barcode	Original Sample ID	Scientific name	% ID	Alignment length	Scientific name	% ID	Alignment length	Length of contig	Coverage (read depth)	Scientific name	% ID	Alignment length	Length of contig	Coverage (read depth)
1	NJ-B001	No Hit			Cardinalis cardinalis	99.1	453	837	15.3					
2	NJ-B002	Cardinalis cardinalis	**97.0**[2]	374	Cardinalis cardinalis	99.3	720	1313	28.4	Hylocichla mustelina	99.5	638	1146	249.5
3	NJ-B003	Setophaga pinus	**93.1**	710	Setophaga pinus	100.0	444	445	1418.3	Cardinalis cardinalis	98.5	720	984	22.2
4	NJ-B004	No Hit			Cardinalis cardinalis	99.7	707	1061	18.5					
5	NJ-B005	Catharus fuscescens	98.8	**323**	Catharus fuscescens	100.0	545	670	498.4	Cardinalis cardinalis	99.6	707	1178	20.1
6	NJ-B006	Cardinalis cardinalis	**95.7**	705	Cardinalis cardinalis	99.6	459	461	2203.4					
7	NJ-B009	No Hit			Poecile carolinensis	**93.7**	681	1064	34.6					
8	NJ-B010	Cardinalis cardinalis	**97.3**	707	Cardinalis cardinalis	99.8	426	433	1095.1					
9	NJ-B011	Zenaida macroura	**97.6**	247	Zenaida macroura	99.3	677	1073	281.0					
10	NJ-B013	Sayornis phoebe	**97.6**	673	Sayornis phoebe	99.8	442	442	3811.9					
11	NJ-B014	Hylocichla mustelina	**95.9**	639	Hylocichla mustelina	99.4	530	572	1570.9	Passer domesticus	99.0	711	956	23.5
12	NJ-B016	Thryothorus ludovicia	**95.3**	711	Thryothorus ludovicia	100.0	453	453	1205.5					
13	NJ-B017	No Hit			Zenaida macroura	98.6	708	1109	20.2					
14	NJ-B019	Catharus fuscescens	98.2	511	Catharus fuscescens	100.0	431	431	1173.5	Cardinalis cardinalis	99.3	720	972	16.9
15	NJ-B020	Empidonax alnorum	**96.9**	682	Empidonax alnorum	98.3	**287**	317	1704.0	Cardinaliscardinalis	98.7	739	838	16.4
16	NJ-B021	Setophaga pinus	**94.4**	695	Setophaga pinus	100.0	400	400	1350.3					
17	NJ-B022	Catharus fuscescens	99.1	436	Catharus fuscescens	99.9	703	831	662.7	Cardinalis cardinalis	99.7	707	1110	17.5
18	NJ-B023	No Hit			Cardinalis cardinalis	98.8	658	837	24.0					
19	NJ-B024	No Hit			No Hit	0.0								
20	NJ-B025	No Hit			No Hit	0.0								
21	NJ-B026	Cardinalis cardinalis	99.3	705	Cardinalis cardinalis	99.4	**166**	354	374.0					
22	NJ-B027	Cardinalis cardinalis	99.4	707	Cardinalis cardinalis	100.0	474	474	1468.9					
23[3]	NJ-B028	Empidonax alnorum	**80.7**	**341**	Empidonax alnorum	98.8	501	501	936.9	Cardinalis cardinalis	99.9	705	1062	18.5
24	NJ-B029	Cardinalis cardinalis	99.1	708	Cardinalis cardinalis	99.9	700	707	2730.7					
25	NJ-B030	Setophaga pinus	**95.6**	**249**	Setophaga pinus	99.6	716	1694	786.4	Cardinalis cardinalis	99.0	606	839	16.8
26	NJ-B031	Catharus fuscescens	**96.4**	449	Catharus fuscescens	100.0	454	454	988.7					
27	NJ-B033	Catharus fuscescens	**96.8**	708	Catharus fuscescens	100.0	553	553	2050.5					
28	NJ-B034	No Hit			Vireo olivaceus	98.9	710	1033	23.6					
29	NJ-B036	Setophaga pinus	**95.4**	439	Setophaga pinus	99.7	665	665	1056.8					
30	NJ-B037	No Hit			Cardinalis cardinalis	99.3	720	1104	21.1					
31	NJ-B038	Cardinalis cardinalis	99.3	705	Cardinalis cardinalis	100.0	436	439	2267.7					
32	NJ-B039	Cardinalis cardinalis	99.0	**286**	Cardinalis cardinalis	99.3	705	820	526.2					
33	NJ-B040	Baeolophus bicolor	**97.7**	681	Baeolophus bicolor	99.7	378	378	425.0					
34	NJ-B041	Cardinalis cardinalis	99.7	663	Cardinalis cardinalis	100.0	547	571	1179.0					
35	NJ-B042	Cardinalis cardinalis	99.1	701	Cardinalis cardinalis	99.8	397	398	1611.4					
36	NJ-B043	Sayornis phoebe	**95.8**	674	Sayornis phoebe	99.9	660	762	1470.3					
37	NJ-B044	Vireo olivaceus	98.6	438	Vireo olivaceus	99.4	679	782	700.1	Cardinalis cardinalis	99.2	724	964	20.5
38	NJ-B045	Vireo olivaceus	**97.2**	542	Vireo olivaceus	100.0	587	587	1562.9					

Table 2. Continued.

		Sanger sequencing			ONT sequencing								
					First contig blast hit[1]				Second contig blast hit				
ONT Barcode	Original Sample ID	Scientific name	% ID	Alignment length	Scientific name	% ID	Alignment length	Length of contig	Coverage (read depth)	Scientific name	% ID	Alignment length	Length of contig	Coverage (read depth)
39	NJ-B046	Cardinalis cardinalis	99.2	706	Cardinalis cardinalis	99.8	504	505	1395.9					
40	NJ-B047	Setophaga ruticilla	**92.1**	**256**	Setophaga ruticilla	99.3	714	1670	282.1					
41	NJ-B048	Catharus ustulatus	**97.7**	700	Catharus ustulatus	99.9	686	709	1633.6					
42	NJ-B049	Spizella passerina	99.2	707	Spizella passerina	100.0	388	388	1994.1	Cardinalis cardinalis	99.4	720	837	16.9
43	TX13-488	No Hit			Cardinalis cardinalis	100.0	412	412	2011.1					
44	TX13-489	Cardinalis cardinalis	98.8	426	Cardinalis cardinalis	100.0	546	659	2200.2					
45	TX13-492	No Hit			Cardinalis cardinalis	98.9	720	861	17.7	Zenaida macroura	100.0	539	539	2053.6
46	TX13-495	No Hit			No Hit	0.0								
47	TX13-497	Passer domesticus	98.7	626	Passer domesticus	99.4	529	636	1512.2	Cardinalis cardinalis	99.3	721	1004	24.7
48	TX13-498	Parus carolinensis	**95.4**	680	Poecile carolinensis	**95.7**	423	444	1863.8					
49	TX13-499	Zenaida macroura	99.0	709	Zenaida macroura	100.0	504	504	2042.4					
50	TX13-500	Zenaida macroura	99.2	707	Zenaida macroura	100.0	362	362	1512.3					
51	TX13-503	Cardinalis cardinalis	99.0		Cardinalis cardinalis	99.5	388	388	1154.1					
52	TX13-505	Zenaida macroura	99.0	708	Zenaida macroura	99.1	463	470	3984.7					
53	TX13-506	Zenaida asiatica	99.8	641	Zenaida asiatica	99.9	647	679	50.3	Cyanocitta cristata	99.6	695	699	63.9
54	TX13-507	Mimus polyglottos	99.0	711	Mimus polyglottos	98.9	**185**	370	22.3					
55	TX13-508	Zenaida macroura	99.2	707	Zenaida macroura	99.7	361	489	3522.4					
56	TX13-509	Mimus polyglottos	99.6	671	Mimus polyglottos	99.6	555	682	921.4					
57	TX13-514	Turdus migratorius	99.2	710	Turdus migratorius	99.7	**335**	369	1170.5	Zenaida macroura	99.2	468	658	20.9
58	TX13-519	Cardinalis cardinalis	98.7	707	Cardinalis cardinalis	99.8	669	589	2598.5					
59	TX13-521	No Hit			No Hit	0.0								
60	TX13-523	Zenaida macroura	99.1	562	Zenaida macroura	99.7	689	690	2130.6					
61	TX13-524	Cardinalis cardinalis	99.1	705	Cardinalis cardinalis	100.0	**304**	664	69.1					
62	TX13-525	Zenaida macroura	**95.9**	712	Zenaida macroura	99.8	539	649	1950.7					
63	TX13-526	Cardinalis cardinalis	99.1	438	Cardinalis cardinalis	99.4	720	862	853.9					
64	TX13-527	No Hit			No Hit	0.0								
65	TX13-528	Cardinalis cardinalis	99.0	707	Cardinalis cardinalis	99.8	455	462	1812.1					
66	TX13-531	Cardinalis cardinalis	99.1	705	Cardinalis cardinalis	99.7	665	665	2236.5					
67	TX13-532	Passer domesticus	99.0	711	Passer domesticus	100.0	407	407	1296.4					
68	TX13-536	Cardinalis cardinalis	99.1	705	Cardinalis cardinalis	99.8	457	457	2689.6					
69	TX13-537	Zenaida macroura	99.3	707	Zenaida macroura	99.7	364	364	2378.7	Cardinalis cardinalis	99.9	708	1001	22.5
70	TX13-538	Cardinalis cardinalis	99.1	705	Cardinalis cardinalis	99.8	502	502	4218.9					
71	TX13-541	Cyanocitta cristata	98.4	615	Cyanocitta cristata	99.6	473	573	892.8	Thryothorus ludovicianus	99.6	715	715	100.2
72	TX13-542	Cardinalis cardinalis	99.4	705	Cardinalis cardinalis	99.8	618	619	2389.2					
73	TX13-544	Mimus polyglottos	99.2	710	Mimus polyglottos	99.8	464	464	498.2					
74	TX13-545	Cardinalis cardinalis	99.5	561	Cardinalis cardinalis	100.0	517	521	2263.8					
75	TX13-547	Cyanocitta cristata	99.0	711	Cyanocitta cristata	99.8	405	404	2196.2	Cardinalis cardinalis	99.4	720	1048	22.2
76	AZ15-376	Passer domesticus	98.3	709	Passer domesticus	99.0	713	1532	318.0	Cardinalis cardinalis	99.0	725	1415	26.0
77	AZ15-377	Passer domesticus	**97.5**	440	Passer domesticus	100.0	435	435	487.8	Cardinalis cardinalis	99.3	720	1052	21.4
78	AZ15-378	Passer domesticus	**97.4**	614	Passer domesticus	99.5	571	895	218.1	Cardinalis cardinalis	99.6	708	1174	17.0

Table 2. Continued.

		Sanger sequencing			ONT sequencing								
					First contig blast hit[1]					Second contig blast hit			
ONT Barcode	Original Sample ID	Scientific name	% ID	Alignment length	Scientific name	% ID	Alignment length	Length of contig	Coverage (read depth)	Scientific name	% ID	Alignment length	Length of contig	Coverage (read depth)
79	AZ15-379	Passer domesticus	98.4	703	Passer domesticus	98.9	**278**	383	70.5					
80	AZ15-383	Passer domesticus	**97.1**	**277**	Passer domesticus	99.2	711	1132	104.6					
81	AZ15-386	Zenaida macroura	99.3	700	Zenaida macroura	99.4	**321**	322	1187.3					
82	AZ15-391	No Hit			Passer domesticus	99.3	710	1133	23.2					
83	AZ15-392	Passer domesticus	98.3	682	Passer domesticus	99.2	500	500	429.8					
84	AZ15-393	Passer domesticus	98.2	711	Passer domesticus	100.0	422	422	1106.6	Cardinalis cardinalis	99.4	720	1024	25.2
85	TX13-512	Setophaga petechia	99.3	703	Setophaga petechia	99.8	582	582	2407.8					
86	AZ15-395	Passer domesticus	98.7	711	Passer domesticus	99.3	**145**	293	433.4	Cardinalis cardinalis	99.4	668	837	22.9
87	AZ15-396	Passer domesticus	98.2	563	Passer domesticus	99.5	584	710	526.3					
88	AZ15-402	Passer domesticus	98.5	697	Passer domesticus	100.0	494	494	1769.1	Cardinalis cardinalis	99.0	720	1186	32.8
89	AZ15-404	Passer domesticus	99.0	711	Passer domesticus	99.2	715	1241	221.5					
90	AZ15-405	Passer domesticus	98.9	657	Passer domesticus	98.0	**149**	273	52.9					
91	same as bc02				Cardinalis cardinalis	99.4	702	837	15.8	Hylocichla mustelina	99.8	520	773	212.0
92	same as bc08				Cardinalis cardinalis	99.9	689	691	1067.3					
93	same as bc21				Cardinalis cardinalis	100.0	400	400	4989.1					
94	same as bc22				Cardinalis cardinalis	100.0	432	432	1773.2					
95	same as bc24				Cardinalis cardinalis	100.0	530	530	2405.2					
96	same as bc31				Cardinalis cardinalis	99.5	370	370	1540.1					

[1] What was called "First" and "Second" ONT Blast hits was arbitrary, and arranged here for ease of comparison to Sanger hits.
[2] Values in bold indicate either alignment lengths that are less than half the length of the 700 bp amplicon, or a percent ID of less than 98%.
[3] The de novo assembly results showed support for a third species, shown in Supplemental File 4.

A somewhat unexpected result was the degree to which mixed bloodmeals were present in our samples. Over 25% of the samples contained COI gene sequences from more than one bird species. In one case, more than two species in different genera were detected. The bloodmeal homogenates from New Jersey were from *Culiseta melanura* (Coquillett), a species documented to take multiple bloodmeals, thus the results are consistent with the biology of the mosquito (Lorenz and Scott 1996, Mahmood and Crans 1997). The mosquito species associated with the other samples are unknown. Our results suggest routine Sanger sequencing of purified PCR products lacks the ability to detect multiples species in a bloodmeal in a way that compares to Nanopore sequencing.

Nanopore sequencing thus is appealing for several reasons. With reasonable molecular biology laboratory and IT requirements, a VCD can conduct mosquito host bloodmeal identifications in-house, which can significantly decrease the time between sample collection and species identification because samples do not have to be transferred to another location for sequencing. As well, ONT provides support for new users through instructional videos and opportunities to communicate with their technical support staff and other users through the Nanopore Community on ONT's website. In addition, although we presented data from a full-sized MinION flow cell run, we found that the less expensive Flongle flow cells perform in an equivalent manner and can accommodate 96 barcodes worth of amplicon sequencing data in one 6–8-h run. Unlike Illumina or Ion Torrent sequencing instruments, Nanopore sequencing devices can be stopped when sufficient data are generated. Because of this, it is possible to go from extracted DNA to basecalled Nanopore data in a day. Finally, Nanopore sequencing is less expensive per sample than Sanger sequencing, partly because only one sequencing reaction is needed per sample. The cost of off-site Sanger sequencing services is approximately $7–$15 for the 2 reads needed per sample (internet search of companies that offer mail-in Sanger sequencing services). In contrast, the cost per sample for the Nanopore sequencing kit, plus the NEB reagents was calculated here to be $5.29.

A small number of studies have also targeted the COI gene for mosquito bloodmeal host identification, demonstrating the wide applicability of Nanopore sequencing to this area of vector biology. Mirza et al. (2024) included host bloodmeal identification in a workflow that also sequenced mosquito species and arboviral sequences in one sample. Their metagenomic approach showed potential for that type of broad molecular surveillance approach, with the drawback of there being low odds of detecting virus in individual mosquitoes. A recent study by Kipp et al. (2023) used a feature of Nanopore sequencing called adaptive sampling in an assay that started with genomic DNA, not PCR product. Reference sequences were uploaded into the basecalling software, and sequences that matched were preferentially sequenced, resulting in the desired COI gene sequences. It is likely that the number of these kinds of studies and assays will grow.

Improvements to assays that inform decisions about vector control are an important component of reducing the spread of arboviral diseases. We demonstrated that Nanopore sequencing is a practicable alternative to Sanger sequencing to identify mosquito bloodmeal host species, and offers several advantages including the ability to multiplex up to 96 samples in one sequencing run, in addition to time and cost savings. It is our hope that VCDs with an interest in generating surveillance data such as bloodmeal species identification take advantage of some of the tools available.

ACKNOWLEDGMENTS

We acknowledge previous Centers for Disease Control and Prevention (CDC) staff who collected the Texas and Arizona samples for this study, and current CDC insectary staff for providing bloodfed colony specimens used in the early stages of the project. This work was partly funded by a CDC Advanced Molecular Detection (AMD) Incubator Grant in 2022 and 2023 (FY22-23 208inc).

REFERENCES CITED

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.[Internet] [accessed May 28, 2024]. Available at https://www.ncbi.nlm.nih.gov/

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. https://doi.org/10.1089/cmb.2012.0021

Cawthorn DM, Steinman HA, Witthuhn RC. 2012. Evaluation of the 16S and 12S rRNA genes as universal markers for the identification of commercial fish species in South Africa. *Gene* 491:40–48.

Crabtree MB, Kading RC, Mutebi J-P, Lutwama JJ, Miller BR. 2013. Identification of host blood from engorged mosquitoes collected in western Uganda using cytochrome oxidase I gene sequences. *J Wildl Dis* 49:611–626.

Doyle JJ, Gaut BS. 2000. Evolution of genes and taxa: A primer. *Plant Mol Biol* 42:1–23.

Fagre AC, Soto RA, Magleby R, Cuadera MKQ, Sun A, Cervantes K, Crans SC, Panella NA, Kenney JL, Angelus A, Burkhalter KL, Woell D, Horiuchi K, Biggerstaff BJ, Staples JE, Connelly R, Martin SW, Komar N. 2024. Enhancing Eastern Equine Encephalitis Virus Surveillance in New Jersey: Optimized Collection of *Culiseta melanura*. *J Am Mosq Control Assoc* 40:92–101. https://doi.org/10.2987/23-7148

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3:294–299.

Galaxy Community. 2024. The. The Galaxy platform for accessible, reproducible, and collaborative data analyses:

2024 update, Nucleic Acids Research 2024. gkae410, https://doi.org/10.1093/nar/gkae410
Hamer GL, Kitron UD, Goldberg TL, Brawn JD, Loss SR, Ruiz MO, Hayes DB, Walker ED. 2009. Host selection by *Culex pipiens* mosquitoes and West Nile virus amplification. *Am J Trop Med Hyg* 80:268–278.
Hannon ER, Jackson KC, Biggerstaff BJ, Raman V, Komar N. 2019., Bloodmeal Host Selection of *Culex quinquefasciatus* (Diptera: Culicidae) in Las Vegas, Nevada, United States. *J Med Entomol* 56:603–608. https://doi.org/10.1093/jme/tjy237
Hebert PD, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proc Biol Sci* 270:313–321.
Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN. 2007. Universal primer cocktails for fish DNA barcoding. *Mol Ecol Notes* 7:544–548.
Joshi NA, Fass JN. 2011. Sickle: A windowed adaptive trimming tool for FASTQ files using quality. https://github.com/najoshi/sickle
Kent RJ, Norris DE. 2005. Identification of mammalian blood meals in mosquitoes by a multiplexed polymerase chain reaction targeting cytochrome B. *Am J Trop Med Hyg* 73:336–342.
Kent RJ. 2009. Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. *Mol Ecol Res* 9:4–18. https://doi.org/10.1111/j.1755-0998.2008.02469.x
Kipp EJ, Lindsey LL, Milstein MS, Blanco CM, Baker JP, Faulk C, Oliver JD, Larsen PA. 2023. Nanopore adaptive sampling for targeted mitochondrial genome sequencing and bloodmeal identification in hematophagous insects. *Parasit Vectors*. 16:68. https://doi.org/10.1186/s13071-023-05679-3
Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc Natl Acad Sci* 86:6196–6200.
Lardeux F, Loayza P, Bouchité B, Chavez T. 2007. Host choice and human blood index of *Anopheles pseudopunctipennis* in a village of the Andean valleys of Bolivia. *Malar J* 22:6–8. https://doi.org/10.1186/1475-2875-6-8
Leger A, Leonardi T. 2019. pycoQC, interactive quality control for Oxford Nanopore Sequencing. *J Open Source Softw* 4:1236. https://doi.org/10.21105/joss.01236
Lorenz LH, Scott TW. 1996. Detection of multiple bloodfeeding in *Culiseta melanura* using a histologic technique. *J Am Mosq Control Assoc* 12:135–136.
Mahmood F, Crans WJ. 1997. Observations on multiple bloodfeeding in field-collected *Culiseta melanura*. *J Am Mosq Control Assoc* 13:156–157.
Makunin A, Korlević P, Park N, Goodwin S, Waterhouse RM, von Wyschetzki K, Jacob CG, Davies R, Kwiatkowski D, St Laurent B, Ayala D, Lawniczak MKN. 2022. A targeted amplicon sequencing panel to simultaneously identify mosquito species and *Plasmodium* presence across the entire *Anopheles* genus. *Mol Ecol Resour* 22:28–44. https://doi.org/10.1111/1755-0998.13436
Mirza JD, de Oliveira Guimarães L, Wilkinson S, Rocha EC, Bertanhe M, Helfstein VC, de-Deus JT, Claro IM, Cumley N, Quick J, Faria NR, Sabino EC, Kirchgatter K, Loman NJ. 2024. Tracking arboviruses, their transmission vectors and potential hosts by nanopore sequencing of mosquitoes. *Microb Genom* 10:001184. https://doi.org/10.1099/mgen.0.001184
Nicholls TJ, Minczuk M. 2014. In D-loop: 40 years of mitochondrial 7S DNA. *Exp Gerontol* 56:175–181.
Noreuil A, Fritz ML. 2021. Differential gene expression in the heads of behaviorally divergent *Culex pipiens* mosquitoes. 2021. *Insects* 12:271. https://doi.org/10.3390/insects12030271
Oxford Nanopore Technologies. 2024a. EPI2ME labs: Bioinformatics resources from Oxford Nanopore Technologies. [Internet] [Accessed July 8, 2024] Available from https://github.com/epi2me-labs
Oxford Nanopore Technologies. 2024b. IT requirements. [Internet] [Accessed May 16, 2024]. https://community.nanoporetech.com/requirements_documents/minion-it-reqs.pdf.
Oxford Nanopore Technologies. 2024c. Dorado open source basecaller for Oxford Nanopore reads. [Internet] [Accessed May 16, 2024] Available from: https://github.com/nanoporetech/dorado
Ratnasingham S, Hebert PD. 2007. BOLD: The Barcode of Life Data System. *Mol Ecol Notes* 7:355–364.
Riccetti N, Fasano A, Ferraccioli F, Gomez-Ramirez J, Stilianakis NI. 2022. Host selection and forage ratio in West Nile virus-transmitting *Culex* mosquitoes: Challenges and knowledge gaps. *PLoS Negl Trop Dis*. 16:e0010819. https://doi.org/10.1371/journal.pntd.0010819
Saccone C, De Giorgi C, Gissi C, Pesole G, Reyes A. 1999. Evolutionary genomics in Metazoa: The mitochondrial DNA as a model system. *Gene* 238:195–209.
Takken W, Verhulst NO. 2013. Host preferences of bloodfeeding mosquitoes. *Annu Rev Entomol* 58:433–453. https://doi.org/10.1146/annurev-ento-120811-153618
Tandina F, Niare S, Almeras L, Davoust B, Doumbo OK, Raoult D, Parola P, Laroche M. 2020. Identification of mixed and successive blood meals of mosquitoes using MALDI-TOF MS protein profiling. *Parasitol* 147:329–339. https://doi.org/10.1017/S003118201900163X
Whitford W, Hawkins V, Moodley KS, Grant MJ, Lehnert K, Snell RG, Jacobsen JC. 2022. Proof of concept for multiplex amplicon sequencing for mutation identification using the MinION Nanopore sequencer. *Sci Rep* 12:8572. https://doi.org/10.1038/s41598-022-12613-7