



The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title: Application of the Human Virome to Touched Objects and Hair Shafts

Author(s): Michael Adamowicz, Ph.D.

Document Number: 306957

Date Received: June 2023

Award Number: 2019-75-CX-0017

This resource has not been published by the U.S. Department of Justice. This resource is being made publicly available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Final Report

DOJ Grant 2019-75-CX-0017

Project Titled; Application of the Human Virome to Touched Objects and Hair Shafts

Michael Adamowicz, Ph.D.

madamowicz2@unl.edu

103 Agriculture Hall

Lincoln, NE 68583

402-472-7932

Board of Regents, University of Nebraska-Lincoln

151 Whittier Research Center

2200 Vine Street

Lincoln, NE 68583

Grant Period 01/01/2020 – 12/31/2022

Award \$443,931

Summary of project

This project was designed to address the ongoing need to create forensically relevant linkages between persons, places, and objects by continuing to develop the heretofore untapped potential of the human viral microbiome (virome). The human virome is a source of rich genetic diversity that needs to be examined to determine if it is stable, transferable, and provides a sufficient power of discrimination to be used as an alternative to traditional human forensic deoxyribonucleic acid (DNA) tests when such tests are infeasible. The human bacterial microbiome is already being examined as an alternative method for human identification in forensically relevant cases. The human virome offers some advantages as the viral genomes are even smaller than those of bacteria, and thus are potentially more physically stable, have a variety of morphologies (double- and single-stranded) increasing the possible number of discriminating markers, and is present throughout the human body, including the skin and body fluids, making it transferable. Also, the copy number of viral genomes, compared to the copy number of human genomes, in a given volume is substantially higher, increasing the likelihood of isolating a sufficient quantity for successful testing. Prior to this project, there was a limited amount of empirical data on the human virome regarding its suitability to forensic applications and none regarding the potential to transfer viral material from the skin of an individual to an object. The work described here has begun to address this gap in our knowledge by generating data in the previously mentioned areas of stability, transfer, and discrimination as applied to various types of evidentiary material. The project hypothesized that the genetic diversity contained in each human being's particular virome, as characterized by our previous work, can be transferred to physical objects and detected in situations where standard human DNA samples will not yield a usable profile. The potential for virome profiles developed from cut hairs was also examined in detail. The work was completed with protocols, instrumentation, and analysis methods that are either already in forensic DNA laboratories or can be readily assimilated as described in our previous work. The results of the research have also leveraged

the publicly accessible and expandable database on the human virome that we have developed to demonstrate the practical use of the information. We hope this project can lead to an additional DNA comparison method for cases in which human DNA is too scant or degraded to derive a statistically useful profile from, thus adding individualizing information where there would have been none.

The successful objectives of the project have been to develop new technology in the following areas: 1) examine the potential for virome samples to transfer from humans to touched objects, 2) assess the forensic significance of the profiles in terms of their efficacy in separating individual humans from those objects, 3) characterize the role that background viral particles play in deriving human virome profiles from touched objects, and 4) examine shed and cut hairs value as a source of human virome data. Addressing these aims has furthered our understanding of the human virome, examined the critical barrier of individual identification when conditions for human DNA testing are poor, and developed new technology for the use of the human virome as a tool for forensic applications.

The project was designed to determine the individualization, stability, and transfer of the human skin viral microbiome (virome) in contexts relevant to forensic investigation. We recruited ten adult participants (≥ 19 years of age), five male and five female. Direct samples of the hair virome (ten replicate hair samples) and transfer samples (five total objects, each touched for 60 seconds at two separate time points) were collected from each subject. This is summarized in the table below.

Sample Set	Target # of Samples	Source	Subsample Set	Service
Hair from Head	100	Subject Collection	5 Female, 5 Male, 10 hairs each	Viral DNA Sequencing
Touch Objects	160	Subject Collection	5 Female, 5 Male, 5 objects + controls (all in at least 2 replicates)	Viral DNA Sequencing

Hairs were collected by cutting several centimeters of each hair shaft. Hair roots were avoided to minimize the collection of unwanted human DNA. The objects selected for touch sampling were a ceramic mug, a plastic cup, a glass cup, a small kitchen knife, and a cell phone. All are objects commonly

found and handled in a wide variety of contexts and were considered to be appropriate facsimiles of potential crime scene evidence. All could also be easily cleaned using bleach solutions and ultraviolet (UV) light. Each object was handled by each participant for 60 seconds on each of two separate collection sessions. The collection sessions were separated by seven to ten days to allow their skin viromes to normalize and to collect data regarding any short-term variations in virome content. Each participant had their hands swabbed for viral DNA collection prior to each collection session, allowing us to develop a consensus skin virome profile for each person which was shown to be valuable in our previous research. At each sampling the participants completed a questionnaire to gather information on travel, grooming, lifestyle, and other information that could help identify features affecting the microbiome. During each sampling, negative control swabs were collected to evaluate any contamination that may result from environmental factors including the phosphate buffered saline (PBS) used. The swabs collected were stored at -20°C until used for viral enrichment and sequencing.

Viral DNA was extracted, purified, and isolated from each sample prior to massive parallel sequencing (MPS) using the Illumina Novaseq platform. At the time of writing this report, sequence data has just been completed and received. The resultant data is in the process of being analyzed for identification and quantification of viral species. Once all the sequence data is available, these viral profiles will be analyzed for their ability to discriminate between individuals and the level of transference to each of the objects.

Collection swabs carrying viral particles were saturated with 200µl of 0.02µm filtered sterile 1x PBS and were placed in a 2ml tube containing a spin basket. Swabs were centrifuged at 16000 x g for 10 minutes to elute viral particles from the swab into the PBS solution. The filtrate containing the viral particles was further filtered using a 0.22µm filter to remove non-viral cellular contaminants. An aliquot of each sample from each individual and day of sampling was pooled together, concentrated by ultracentrifugation and treated with DNase I to remove “naked” contaminant DNA from lysed cells. The

pooled samples and the swab samples were then used for viral DNA extraction using the QIAamp Ultra-Sensitive Virus Kit according to the manufacturer's protocol. Viral DNA from hairs was extracted directly from the hair shaft, using the same DNA extraction kit described above. As previous work showed, adding a filtering step to this kind of sample retained most of the usable DNA in the hair shaft. The resulting viral DNA was subjected to whole genome amplification (WGA) based on multiple displacement amplification (MDA) using the TruePrime WGA Kit (4BaseBio) and associated protocol. Following WGA, the samples were quantified using the DeNovix dsDNA High Sensitivity Kit.

One hundred nanograms of the amplified viral DNA were used for each library preparation. The DNA was sheared using sonication to 600bp in length. The resulting sheared DNA was used for library preparation using the NEBNext Ultra II Library preparation kit (New England Biolab) with a modification in the manufacturer's protocol to use custom adaptors. Size selection of 500-700bp fragments was performed with magnetic bead-based purification. Final library preps were evaluated using an Agilent Bioanalyzer with high sensitivity chips to identify sample base pair distribution and sample concentration. Additionally, libraries were quantified using the DeNovix dsDNA High Sensitivity Kit. Resulting libraries were sequenced using the 150 bp paired-end sequencing strategy on the Illumina Novaseq platform.

The subsequent sequence data will be trimmed and filtered using the trimming tool Trimmomatic/0.39 to remove low quality reads using a quality filter threshold of Q30 and a length threshold of 75 bp. Reads resulting from the positive control Phi X DNA will be removed from trimmed reads using the BBDuk. Following quality filtering, bacterial contamination will be assessed by mapping trimmed reads to the Silva 16S ribosomal database v.138.1 with BBDuk using parameters described for high precision mapping of contamination detection. Additionally, all sample reads will be mapped to the human genome (GRCh38.p14) using BBDuk as per BBDuk Guide, for high precision mapping with low sensitivity to lower the risk of false positive mapping. All mapped reads will be removed to ensure the

viromes are devoid of contamination from bacterial sources or the source human host. Metagenome assemblies will be performed using Megahit v.1.2.8. Assemblies will be performed using two approaches, 1) assembly within each sample and 2) a master meta-assembly using all reads. Assembly quality will be assessed using Quast v.5.0.2. In addition, a meta-assembly using all negative control reads will be performed to identify potential contaminants in the dataset that may arise from reagents. The virome assemblies generated will be mapped to the negative control contigs greater than 1000bp using BWA-mem to remove any reads that may result from contamination. Subsequent contigs greater than 1000bp will then be utilized for downstream analysis for viral identification, diversity analysis, and assessment of stability of the virome.

Putative viral contigs containing viral genes will be identified using the tool CheckV v.0.8.0. Contigs that are determined to have viral genes, as per CheckV results, will be identified as viral sequences and classified using various viral annotation and viral classification tools as described below. Viral contigs will be classified using both nucleotide-based classification tools Kraken2 v.2.0.8-beta, Demovir, Blastn (with a >10% query coverage cut-off) and using a protein-based classification tool Kaiju v.1.8. After assessing the results from all the classification tools, the resulting classification having the lowest e-value or highest percent confidence will be used. The least common ancestor will be used for classification results having similar e-values or percent confidences but different results to reduce misclassification of viral contigs.

Raw sample reads will be mapped to the meta-assembly consisting of contigs from the sample-by-sample based assembly. Read mapping will be performed using Bowtie 2 v.2.4 and subsequent Samtools v.1.15 manipulation to identify the abundance of each viral gene/contig within each sample. The resulting viral gene abundances will be further analyzed using "R" v.4.2. Unique contigs will be used to identify viral diversity and changes over-time. A phyloseq object will be created using viral/gene abundance data and then used for diversity analysis. Annotation of viral contigs will be performed using

the classification tools described above. The count table and mapping file information will be used as input for phyloseq object generation, allowing for both denovo and reference-based analysis of the hair shafts and “touch objects” virome data.

Viral gene contig stability will be assessed at family, genus, and species level. Additionally, an assembly independent method will be employed using the identified stable viral families for further refinement of viral taxonomic identification at the species and genus level and investigation into putative human identification makers. The bioinformatic pipeline described above was worked out and refined in our previous project. While there are quite a few steps involved, the analysis proceeds fairly quickly, as opposed to the completed wet-lab benchwork, as it is computationally based and aided by our access to the high throughput processors at the Holland Computing Center on the University of Nebraska-Lincoln campus.

Gathering data on these objectives has furthered our understanding of the human viral microbiome, addressing the critical barrier of individual identification when conditions for human DNA testing are poor. The project has contributed to developing new technology for the use of the human virome as a tool for forensic applications. The advantages of opening another area of forensic biological testing are significant to the criminal justice system. Having the potential to generate a type of DNA profile when the normal methods for human identification are not working could be the only individualizing evidence in a case. There are some types of cases that do not lend themselves to the current human DNA testing methods, as the nature of the biological material transferred is problematic, such as low-template DNA cases. In these situations, there is simply not enough human DNA of sufficient quality to yield a usable short tandem repeat (STR) profile. There are also many cases that yield only partial human DNA profiles, stochastic mixed profiles, or a combination of the two. These can be very difficult to interpret, often requiring advanced probabilistic genotyping software to gain any interpretive value. The addition of human virome markers could complement the existing data and

provide added value, particularly in making a clear exclusion in an otherwise inconclusive interpretation. The results of this project have begun to demonstrate that at least some additional information for human identification can be obtained by recovering and processing the skin virome markers. While our results indicate that the recovered information is limited, it is still useful and more importantly, we have shown that human skin viral markers can be isolated from touched objects and sometimes from hair shafts.

Participants

Name: Jennifer Clarke

Project Role: Co-PI

Name: Samodha Fernando

Project Role: Co-PI

Name: Joshua Herr

Project Role: Co-PI

Name: Carlos Riera Ruiz

Project Role: Graduate Student (Ph.D. candidate)

Changes in approach from original design

This project's progress was severely hampered by the COVID-19 pandemic. The University of Nebraska-Lincoln decided to send all students, faculty, and staff to work remotely in March 2020 and all laboratory research was suspended until late summer, when faculty and staff were allowed to return to campus. This set the project back many months. However, we were able to complete all the preliminary experiments and move to complete full participant sample collections by the end of 2022 (including a one year no-cost extension). There were no substantive changes in the project from the original design. Small alterations to the extraction protocol were made as we performed preliminary experiments to optimize the recovery of viral DNA from the hairs and touched objects. This was

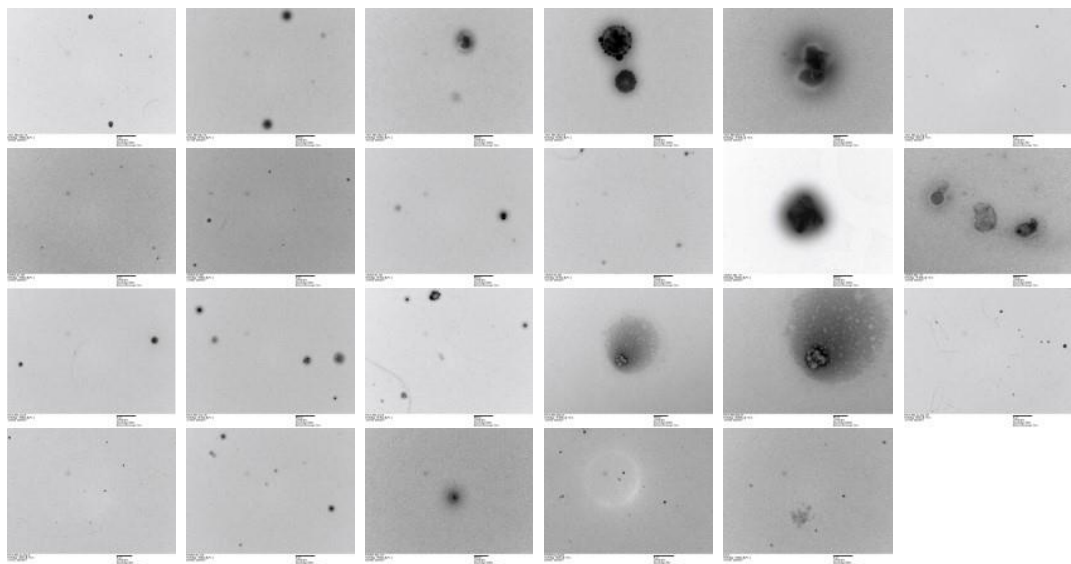
expected, as our previous work was only with collections taken directly from people's skin, which were rich sources of viral sample. While we have developed a working extraction protocol (detailed above), we expect that it will further evolve as more work is done in future projects. Analysis of the data collected will go on for some time, as previously mentioned, as the dataset is large, and we are still discovering new questions to ask regarding its possibilities.

Outcomes

Skin virome samples were collected from the hands of ten individuals three separate times to establish each person's baseline virome profile. Ten separate head hairs were collected by cutting the hair shafts, but not including the root material. Each participant then handled the five common objects for 60 seconds each on two separate occasions. The head hair and objects were chosen due to their potential to be of forensic relevance as they are similar to evidence that may be found at a crime scene. The hands were swabbed at the time of each collection to act as a replicate and to evaluate differences in virome composition and diversity across similar skin types but having differing environmental contact due to things such as the activities that each participant had previously been involved in during the ordinary course of their day. Prior to the handing collections, each of the objects was cleaned in a process that used a strong bleach solution (~20%) and exposure to ultraviolet light for 900 seconds. In a series of preliminary experiments, these conditions were shown to virtually eliminate background viral particles. We also performed a series of preliminary experiments to assess the level of viral contamination on uncleaned objects and objects that were cleaned but then left out in the open in a room over time. We found that that the amount of viral DNA detectable on an object that was not cleaned was low, but still potentially detectable and therefore a cleaning regimen should be employed prior to use in the controlled collection experiments. Cleaned items that were left uncovered on a laboratory bench under an air vent in a room with regular human traffic for up to week displayed very little viral DNA when swabbed and the resultant extracts were quantified. While the first result

(uncleaned object) was not surprising, the second (cleaned and exposed) was. In the interest of minimizing potential contamination, we only used freshly cleaned objects in our collection study, however further work needs to be done to examine if purposely deposited viral particles will remain detectable on an object if left exposed to the environment.

To date, the experimental results with hair shafts have been mixed, although there is still a lot of sequence data to examine. During preliminary experiments designed to optimize the collection and extraction protocol for hair shafts, very low viral DNA yields were encountered. We decided to use a transmission electron microscope to test whether we could actually visualize any viral particles collected from hair shafts by incubating samples overnight in 10% PEG or saline magnesium buffer. The figure below shows some of those TEM micrographs.



These micrographs indicated that there very few intact viral particles on the surface of a cut hair shaft and led us to develop a new collection protocol of using total DNA extraction from the hair without a swab collection step. Data from the preliminary experiments to establish the most effective protocol to collect viral DNA from hair shafts is shown in the table below. The number of detected viral DNA

contigs, listed in the rightmost column is the significant information, with more viral contigs indicating more viral DNA present. The “6_P19_SC_3” sample shows data gathered using the swab collection

A e b l y	N . f n t	N . P v I n t	N . V I n t
LIB_1_1-	19451	4	19447
LIB_2_5-	181889	2	181887
LIB_3_w b	17417	6	17411
LIB_4_n t l	4435	0	4453
LIB_5_w b2	5691	1	5690
LIB_6_P19_SC_3	13065	1	13064

directly from a person’s scalp from our previous study and is included as a comparison value to assess the yield from samples collected from hairs. The “1_1-hair” sample consisted of a single cut hair strand processed by adding the hair directly to extraction buffer. The “2_5-hair” sample was processed in the same manner, but five cut hairs from a single person were extracted together to examine if a greater yield could be obtained. The “3_swabs” sample included a step in which a collection swab wetted with 1X PBS was rubbed along the length of several hair shafts and then added to cell lysis buffer, followed by sample harvesting using centrifuge concentration. The “5_swab2” sample also used a collection swab step, however the viral particles were harvested by centrifuge concentration. The flow though was then passed through 0.2 µm filter and the resulting flow though was used for DNA extraction with the kit previously described. These experiments indicated that either direct exposure of the hair to extraction

or of the collection swab to extraction yielded substantially more viral contigs than just the PBS collected from sample swab alone. As the direct exposure method was simpler, faster, and required less manipulation of the hair it was chosen as the final protocol for the 100 hairs used in the project. Those samples were collected, extracted, and whole genome amplification performed on the viral DNA. Sequencing has been completed and data analysis begun.

Preliminary experiments were also conducted on the touched objects and data collected to ascertain the most effective protocol for viral sample collection and extraction. The table below shows the number of detected viral contigs for two of the experimental objects (knife and plastic cup) that had been handled for one minute, plus the associated negative control. The number of viral contigs was

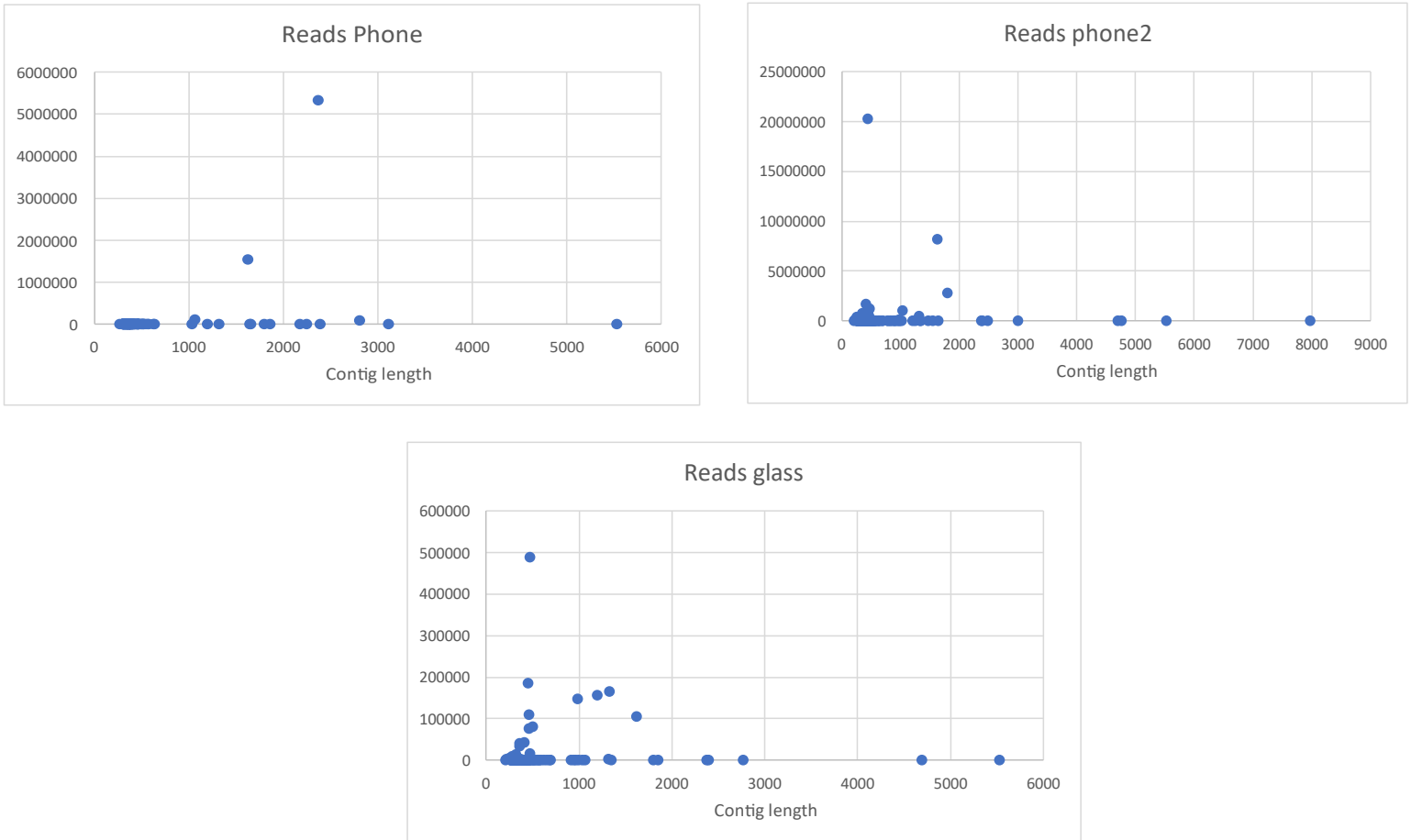
Contigs with viral signal according to CheckV

Assembly	No. of contigs	Proviral contigs	Viral contigs (>10% confidence)	Viral contigs (>50% confidence)
LIB_CR_22_Knife_220408_touch	80	0	1	0
LIB_CR_23_Plastic_220408_touch	85	0	3	2
LIB_CR_25_Neg_C_4_220408	-	-	-	-

very low, however this was expected. A very large amount of published work has shown that human “touch DNA” samples are low template (≤ 100 pg) and touch viral samples were also expected to be low yield. With this in mind, additional preliminary samples were collected to refine the

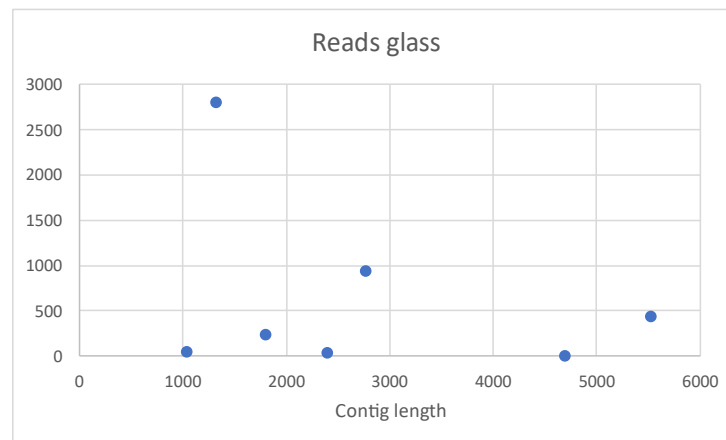
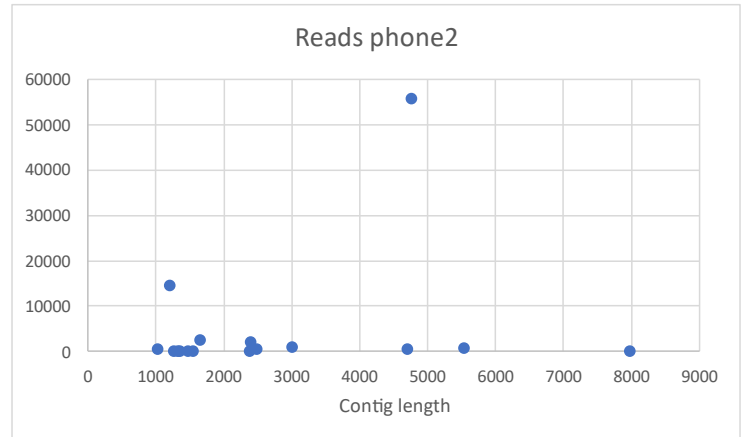
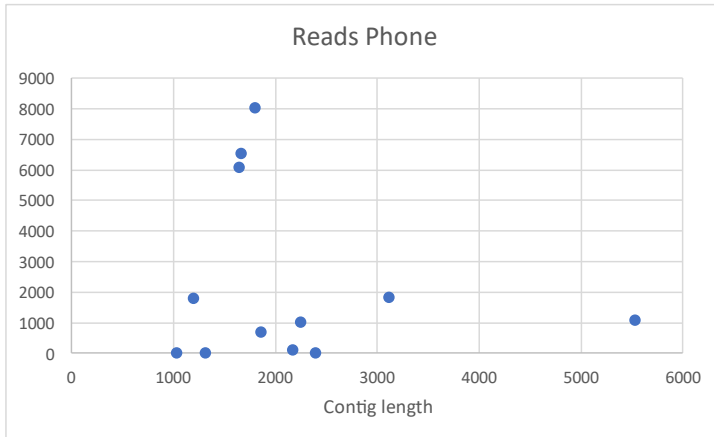
collection/extraction process and examine the length of the contigs putatively identified as viral. Some of this data is shown in the figure below.

Used Object



The contig data for two sample collections from the cell phone (Reads phone and Reads phone 2) from two different people, and the glass cup were plotted with the contig length (bp) on the x-axis and number of sequence reads on the y-axis. The longer contigs have a greater probability of mapping to an identified viral taxon and are more desirable. To better examine the data shown above, any reads of <1000 bp were filtered out and the data was replotted as shown below.

Used Object (>1000 b)



After filtering for contig reads >1000 bp, we observed approximately 27,000 reads in 9 contigs for cell phone collection 1, approximately 78,000 reads in 8 contigs for cell phone collection 2, and approximately 4400 reads in the glass cup collection. Not all contigs can be linked to specific viral taxa, partly because some may be from uncharacterized viruses and some may be from DNA sequences that share similarities to viral genomes but are not actually viral DNA. Some contigs did from the data shown above did align to such viruses as Enterobacteria phage DE3 and a mutant Salmonella phage seszw, indicating that identifying viral markers is possible using the protocol and data analysis pipeline employed.

With this information, we proceeded to collect experimental data from the ten participants as described earlier. These collected samples were extracted, whole genome amplified, and sequenced. The resulting data are now being bioinformatically processed to remove low quality reads and contamination. Bacterial contamination is assessed by mapping trimmed reads to 16S reference reads obtained from the Silva database v.1.3.7. Percentages of trimmed reads per sample mapped to 16S reads ranged from 0% to 0.15%, showing lowered levels of bacterial contamination in samples and thus showing sufficient virome sequencing preparation sample processing. Trimmed and processed reads will be assembled for further processing by removing negative control mapped contigs from the overall assembly. All contigs 1000bp in length and larger will be retained. Contigs will then be run through CheckV and any identified as having a known viral gene and considered to be viral in origin. These will then be compared to our current library of 59 previously identified viral contig “markers” to assess their potential power of identification.

The primary limitation of this project, as with any low-template transfer DNA sample, has been the small amount of testable genetic material deposited by relatively brief periods physical contact. This was expected from the beginning and much of the effort in this project was to 1) develop an efficient viral DNA collection/extraction protocol for touched objects and 2) test our previous bioinformatic pipeline to determine if it was as viable for these types of evidentiary items as it was for samples collected directly from peoples’ skin. Our experimental results indicate that both goals have been successful to a limited extent. The bioinformatic pipeline performed well with the collected viral amplification products, identifying viral DNA sequences. The collection and extraction protocols yielded DNA that led to substantial numbers of sequencing reads, but relatively few useful contigs, as compared to samples collected directly from human skin. This result is still very significant however, as it clearly demonstrated for the first time that forensically useful information can be gathered from viral material transferred from a person’s skin to the surface of an object.

Artifacts

Conference Papers:

E. Graham, M. Adamowicz, J. Clarke, S. Fernando, J. Herr, 2022. The Identification and Characterization of Viral Biomarkers. Presented at the AAFS 74th Annual Scientific Meeting, Seattle, WA.

C. Riera-Ruiz, M.S. Adamowicz, J. Clarke, S. Fernando, and J. Herr, 2022. Application of the Human Virome to Touched Objects and Hair Shafts. Presented at Northeastern Association of Forensic Scientists Annual Meeting, 10/18/22.

Publications:

E.H. Graham, M.S. Adamowicz, P.C. Angeletti, J.L. Clarke, S.C. Fernando, J.R. Herr. 2022. Genome Sequence of Feline Papillomavirus Strain P20 Assembled from Metagenomic Data from the Skin of a House Cat Owner. MRA, 11(7), e01070-21.

E.H. Graham, J.L. Clarke, S.C. Fernando, J.R. Herr, M.S. Adamowicz. 2022. The Application of the Skin Virome for Human Identification. Forensic Science International: Genetics 57:102662.