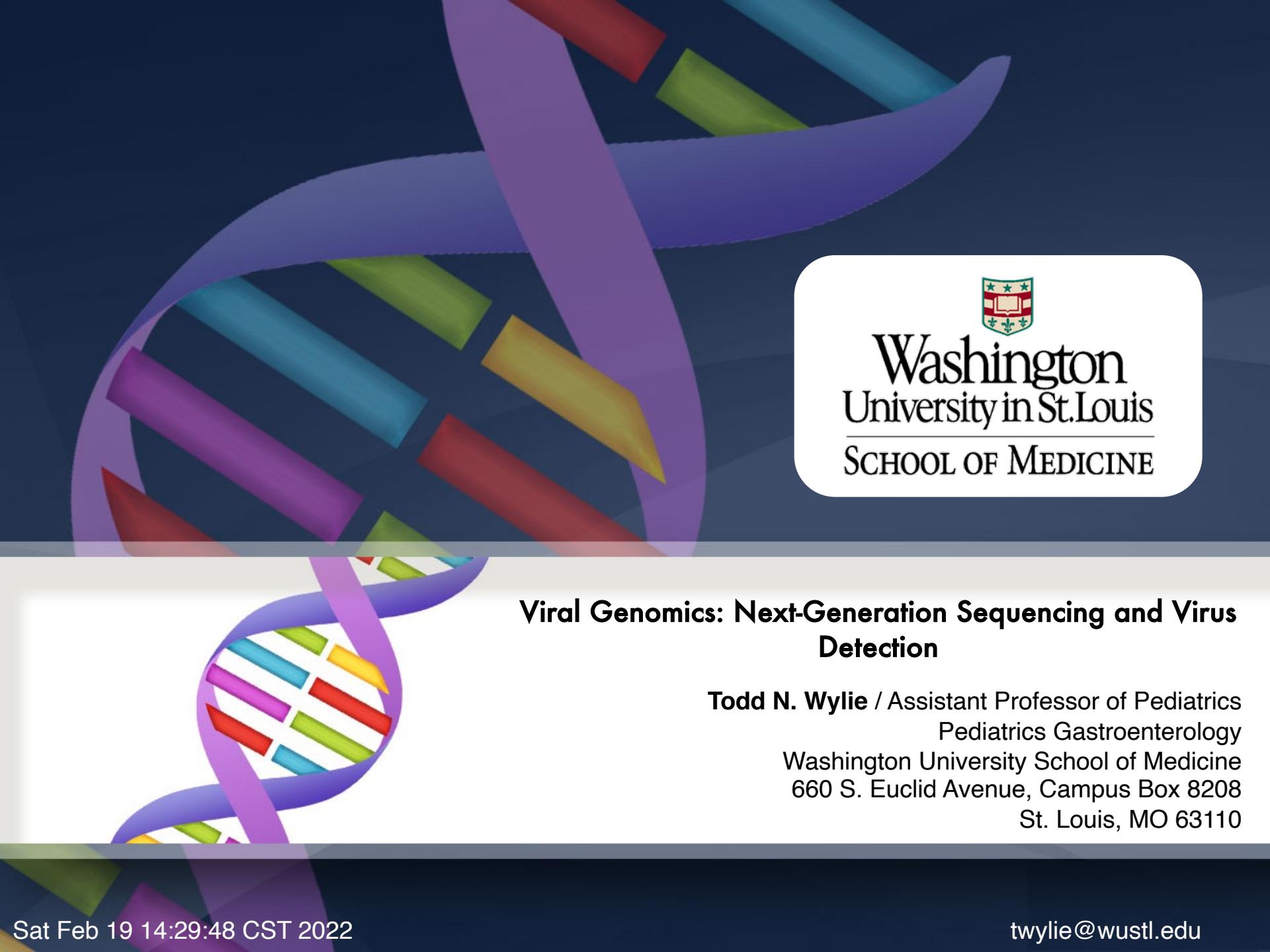




Washington
University in St.Louis
SCHOOL OF MEDICINE

Viral Genomics: Next-Generation Sequencing and Virus Detection

Todd N. Wylie / Assistant Professor of Pediatrics
Pediatrics Gastroenterology
Washington University School of Medicine
660 S. Euclid Avenue, Campus Box 8208
St. Louis, MO 63110



Disclosures

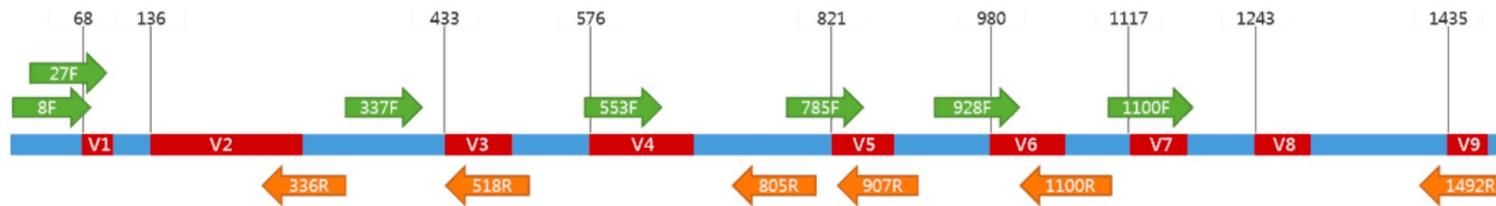
ViroCap: Compositions and methods for detecting viruses in a sample	Patent: US20170218465A1
Enterovirus D68 real-time PCR	Patent: US20160355897A1
VarScan: variant detection in massively parallel sequencing of individual and pooled samples	License: WUSTL Case Number 009883
Hypha Life Sciences and Biogenerator	Collaboration (ViroCap and ViroMatch)

2021: Week #14

Week #14 (2021)	ViroMatch: A Computational Pipeline for Detection of Viral Sequences from Complex Metagenomic Data
Author	Todd N. Wylie
Slides	https://github.com/genome/bfx-workshop/blob/master/archive/lectures/week_14/week_14_viromatch_wylie.pptx
Video	https://wustl.box.com/s/hwerlzg4t3qi3bly1s4bejg74nypume7

16S rRNA Sequencing

Sequence variation among bacterial 16S is known to be not uniformly distributed. Nine hypervariable regions were identified among Bacteria, which are named V1 to V9.



- All members of Bacteria and Archaea are known to have 16S gene
- We may target by amplicon sequencing the 16S gene
- Subsequently, we classify bacteria based on variable sequence regions:
software (UPARSE, Qime, Mothur, DADA2)
- Cheap & targeted; well-established with many, many studies; community comparisons; only targets bacteria; not all bacteria can be identified down to the species/strain level; still useful
- Gut microbiome studies (non-invasive — i.e. stool samples)
- Many commercially available 16S sequencing library prep kits available, as well as service-on-demand companies (e.g. Mr. DNA)
- GTAC/MGI 16S sequencing service — GTAC has their own method for sequencing and analyzing 16S sequencing called MVRSION

Week #20 (2021)	16S analysis: from OTU to ASVs
Author	Brigida Rusconi, PhD
Media	https://github.com/genome/bfx-workshop/tree/master/archive/lectures/week_20

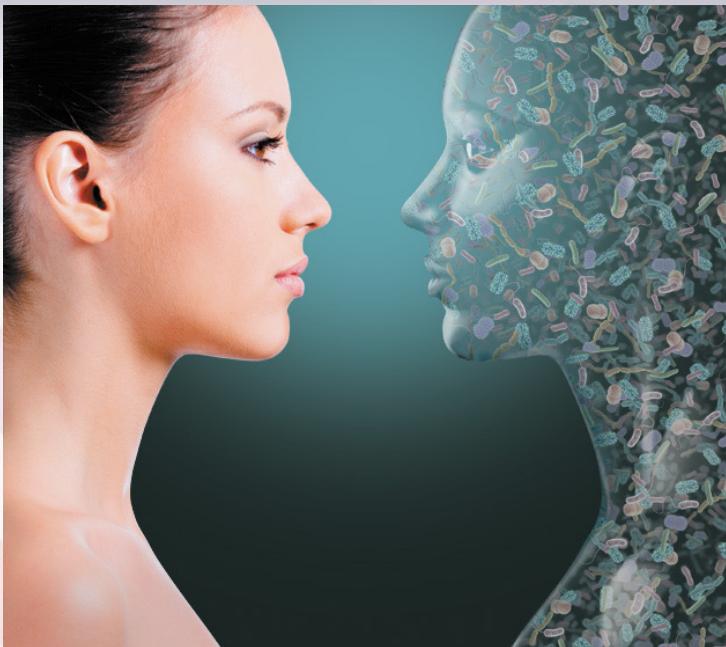
Metagenomics

- However ... 16S sequencing is not metagenomics!
- We want to assess all of the underlying **microbiota** (organisms) within a given sample – not just bacteria.
- Also, we wish to characterize all of the genomes for these species (**microbiome**) – not a single gene, as with 16S sequencing.
- Thus, we use NGS metagenomic shotgun sequencing (**MSS**) for metagenomic studies.
- This has the potential to give us access to the **virome**; however, not without challenges.

The Virome

Virome refers to the assemblage of viruses that is often investigated and described by **metagenomic sequencing** of viral nucleic acids that are found associated with a particular ecosystem, organism or holobiont. The word is frequently used to describe environmental **viral shotgun metagenomes**.

Metagenomics Challenges



R01 Funded Projects

¹Microbial community disruption following topical antimicrobial application in *Staphylococcus aureus*. **PIs:** Fritz, K. Wylie

²The vaginal microbiome, maternal response, and preterm birth. **PIs:** Stout, K. Wylie

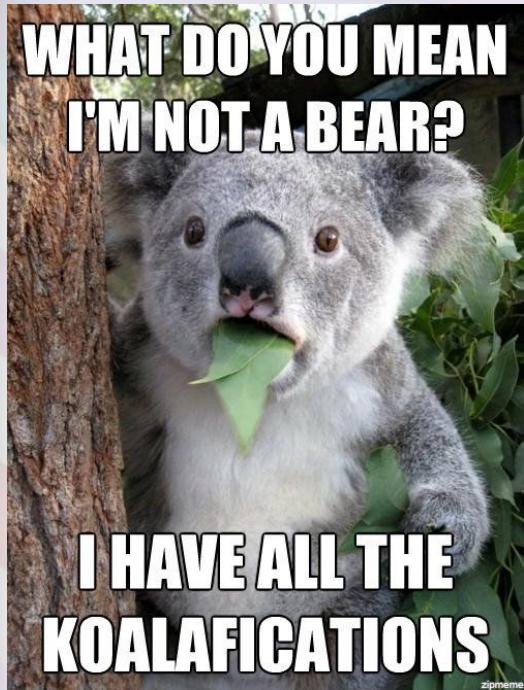
³Metagenomic shotgun microbial sequencing in post-transplant lymphoproliferative disorders (PTLD). **PI:** Dharnidharka

⁴Viral pathogenesis of chronic inflammatory lesions of the placenta. **PIs:** K. Wylie, Ernst

CHALLENGES:

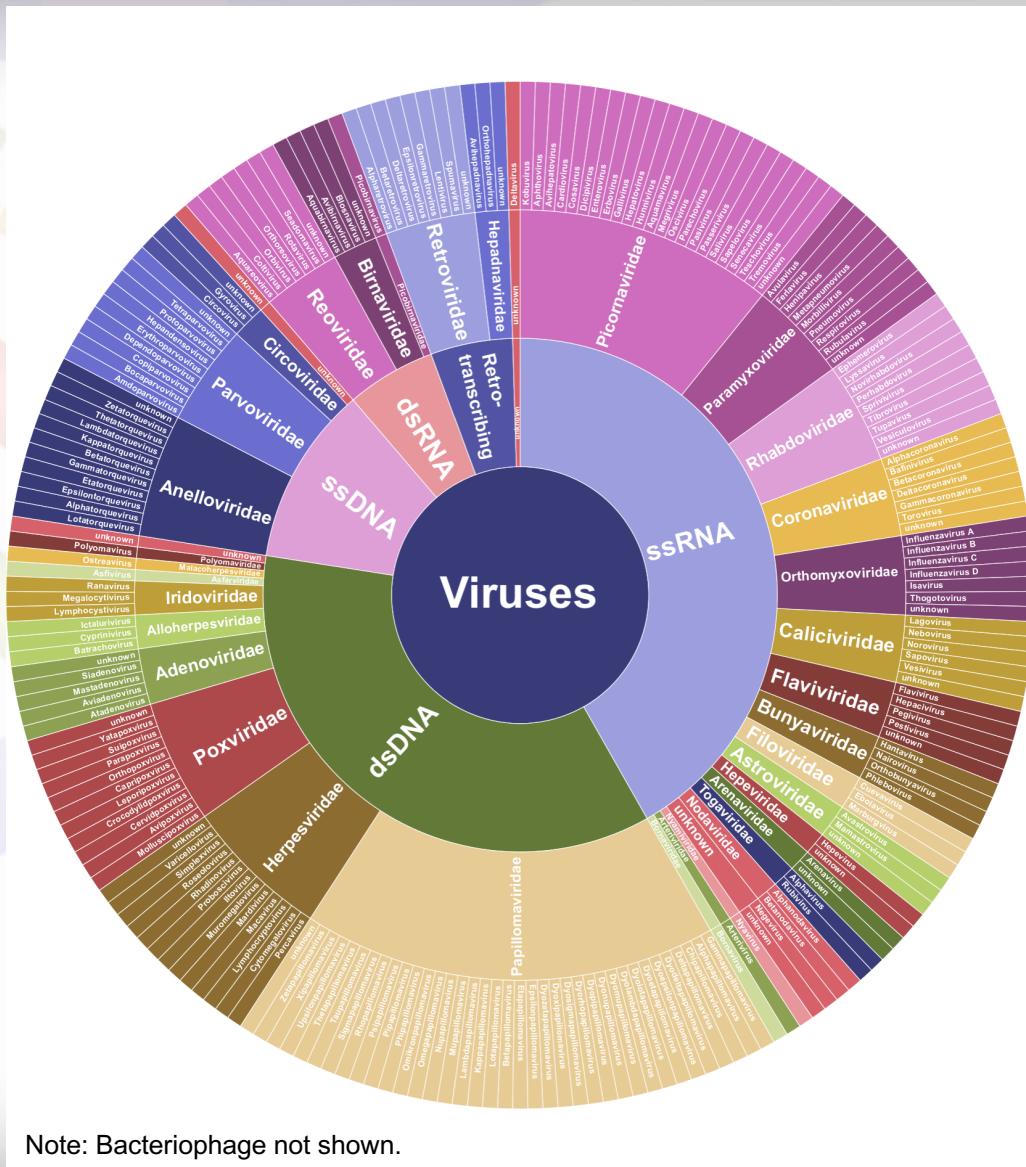
- Not a single reference genome, but thousands of *potential* references
- Expanded sample types — e.g. stool, skin¹, CSF, NP swabs, blood, vaginal swabs², post-transplant tissues³, tumors, bowels, placenta⁴, etc.
- Contamination concerns
- Sample collection and isolation techniques can vary results
- Microbiome is not static
- Expectation vs unknown
- Microbial taxonomy not always consistent; not heterogenous
- Not all microorganisms studied at same level
- Not a “solved problem” in many areas

Virome Analysis Challenges

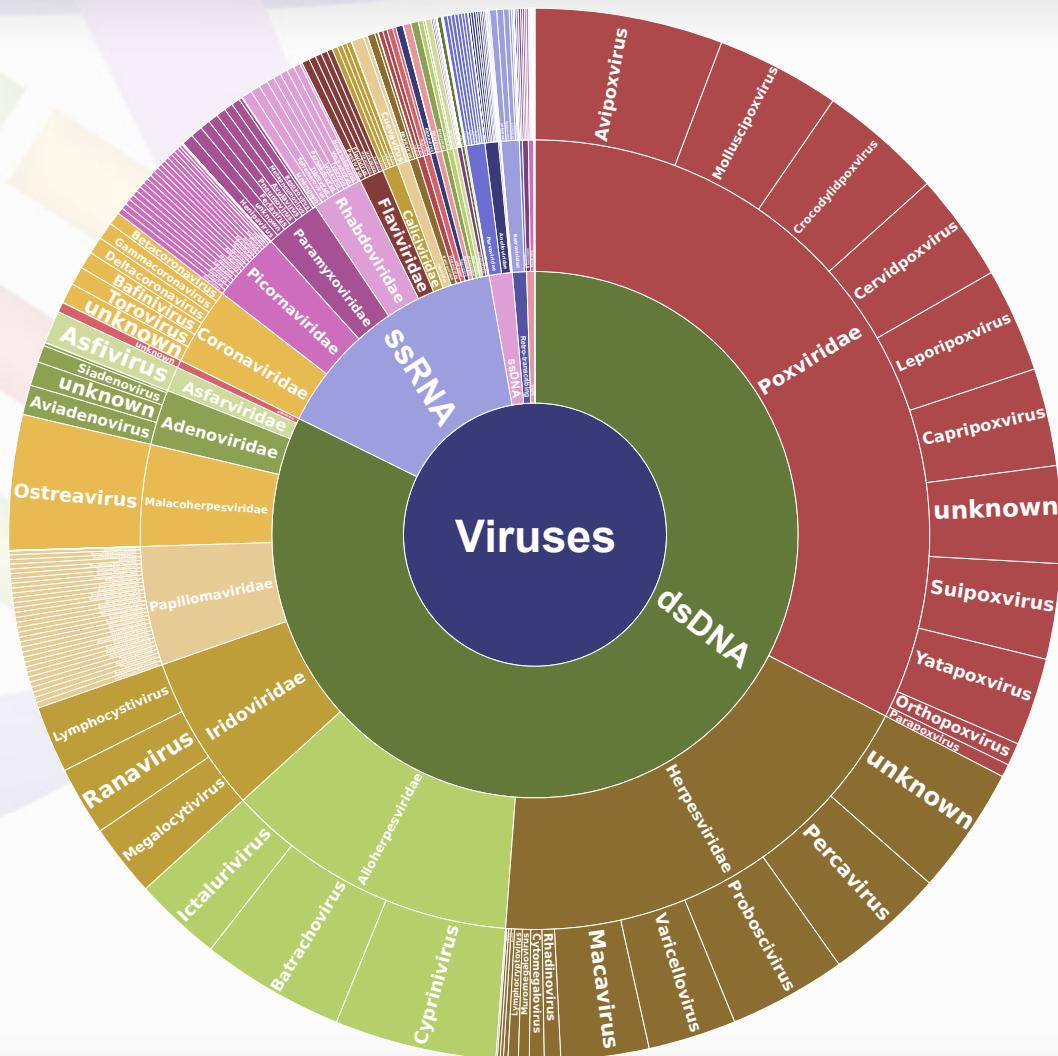


No gene shared among all groups of viruses (ala 16S)	<ul style="list-style-type: none">• Impacts discovery by sequencing
Viral genomes are relatively small (i.e. compared to host genome size)	<ul style="list-style-type: none">• Impacts discovery by sequencing
Viruses have high heterogeneity	<ul style="list-style-type: none">• Impacts classification and taxonomy
Alignment parameters have to serve conflicting goals	<ul style="list-style-type: none">• Identify highly conserved viral genome sequences (e.g. Herpesviruses)• Identify divergent viral genome sequences (e.g. HCV)• Identify viruses without a good reference sequence available (e.g. Enterovirus subtypes)
Virus groups have different criteria for distinction (viral taxonomy is a mess)	<ul style="list-style-type: none">• Some genera contain viruses with high nucleotide sequence similarity• Some genera contain viruses that primarily share gene content, but not nucleotide sequence
Meaningful virus classifications are virus-specific & may require <i>post hoc</i> analyses	<ul style="list-style-type: none">• Rotavirus and flu typed based on specific genes that can re-assort (so you may type based on alignment to multiple reference genomes)• EBV-1 and -2 typed on a very specific genomic region

Viral Taxonomy (variation)



Viral Taxonomy (genome size)



Note: Bacteriophage not shown.

Virus Genome Sizes

Viruses are much smaller than other components of the microbiome!

Genome	Size (bp)	bp to seconds
Human	3,099,706,404	98.2 years
<i>E. coli</i>	5,498,578	63.6 days
SARS-CoV-2	29,903	8.3 hours

Viral Sequencing

Viral detection with next-generation sequencing sounds challenging ...
what to do?

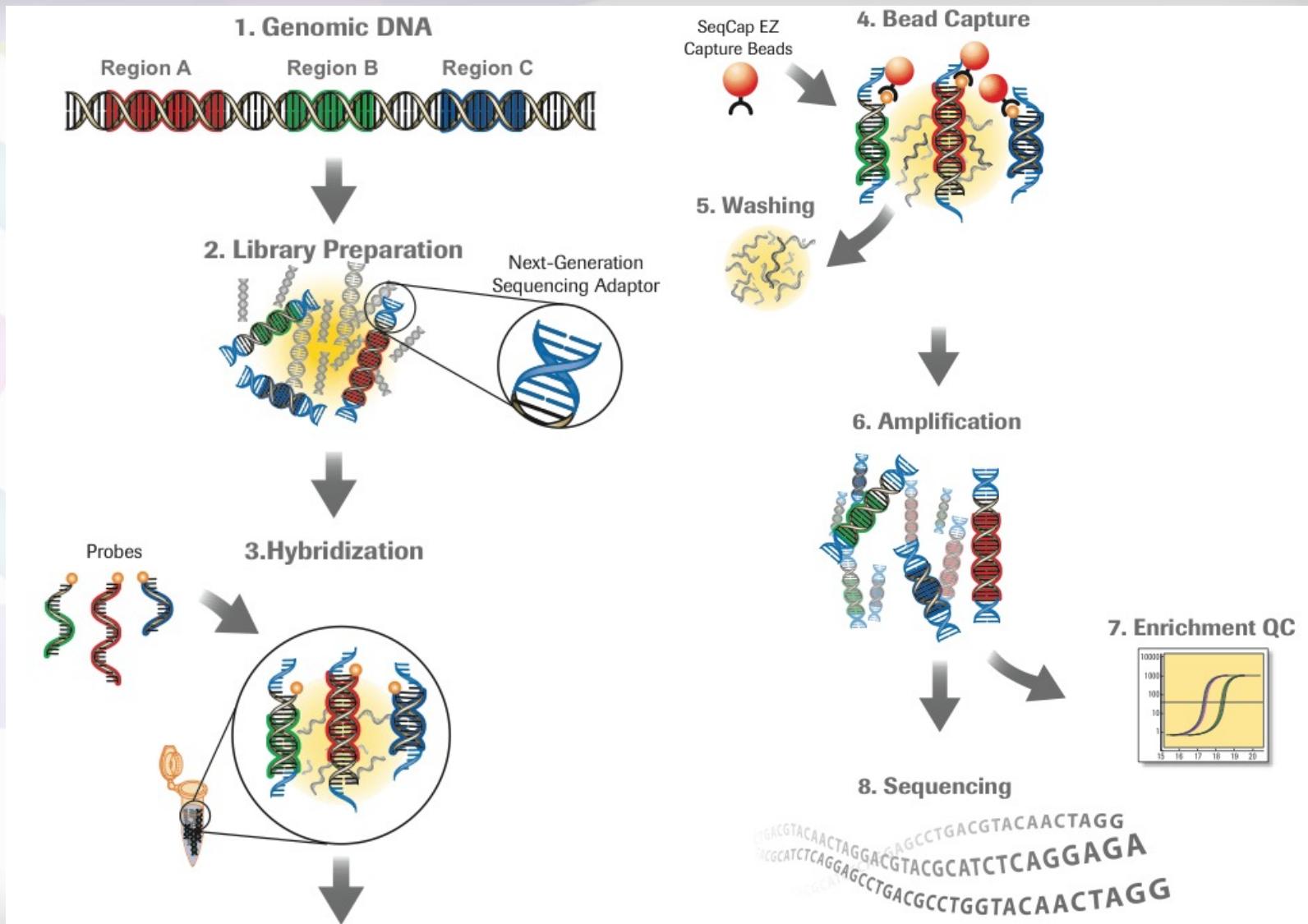
Metagenomic Shotgun Sequencing (MSS)

Analysis of the viral data we've generated has been challenging:

- Small numbers of sequences are harder to interpret/believe
- Harder (or impossible) to subtype based on few sequence
- We want our assays to be highly sensitive so we can accurately detect low level viruses (particularly important for control groups)
- In our experience, we have detected **fewer than 10 viral sequences per 25 million sequence reads** generated for a virus that was detected in a sample by a molecular assay. In other instances we have failed to detect viruses known to be present based on molecular assays (*Wylie, K. et al. 2012*).
- *What to do...?*

Why not use sequence capture?

Roche NimbleGen SeqCap EZ Library



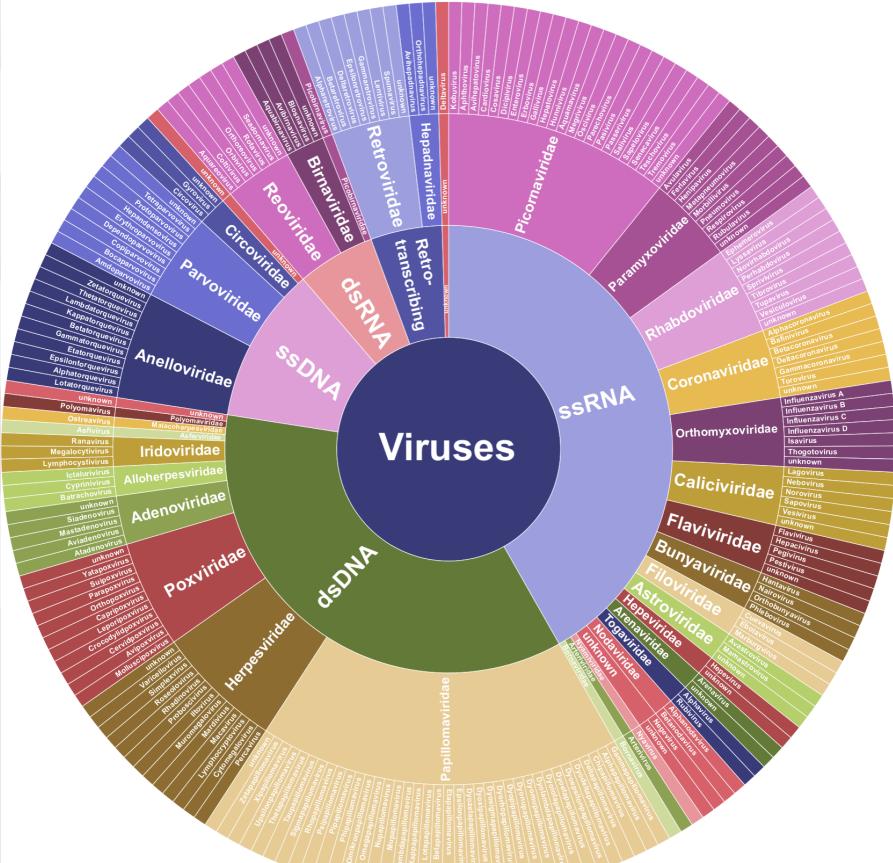
Viral Capture Goals

Our aim was to develop a virus-specific sequence capture reagent that could be used to:

(1) Assess **all known viruses that can infect human cells**, making *a priori* knowledge of the viruses within a sample unnecessary and (2) detect **novel or divergent human viruses**.

To this end, we targeted sequence capture platforms capable of enhancing the detection of a comprehensive set of viruses with vertebrate and invertebrate (insect) hosts.

ViroCap



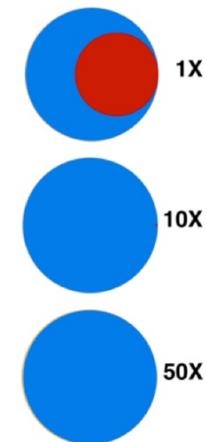
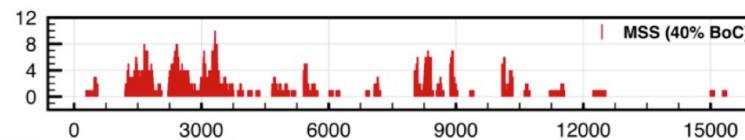
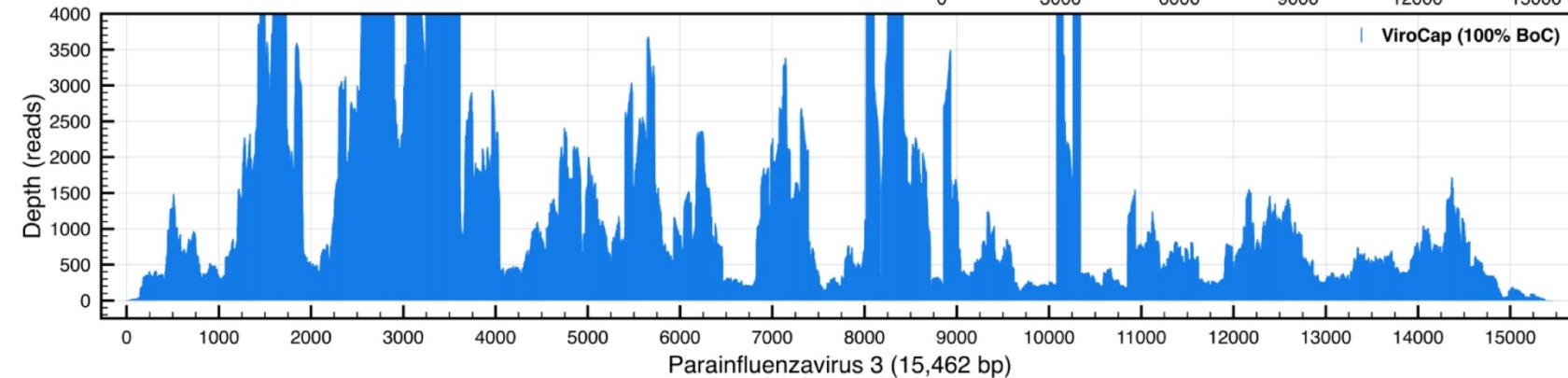
ViroCap (v.1) includes targets from 34 viral families, which consist of 190 genera and 337 species.

The viruses included represent **all viruses** with sequenced genomes from **vertebrate and invertebrate hosts**.

Nearly **1 billion bases** of viral genome sequence was condensed into less than **200 million bases** of target sequence using *k*-mer and clustering analyses to select a unique set of reference sequences.

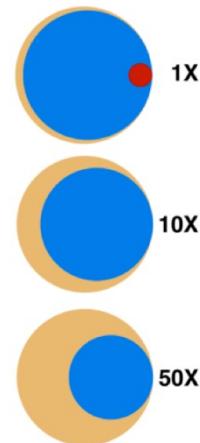
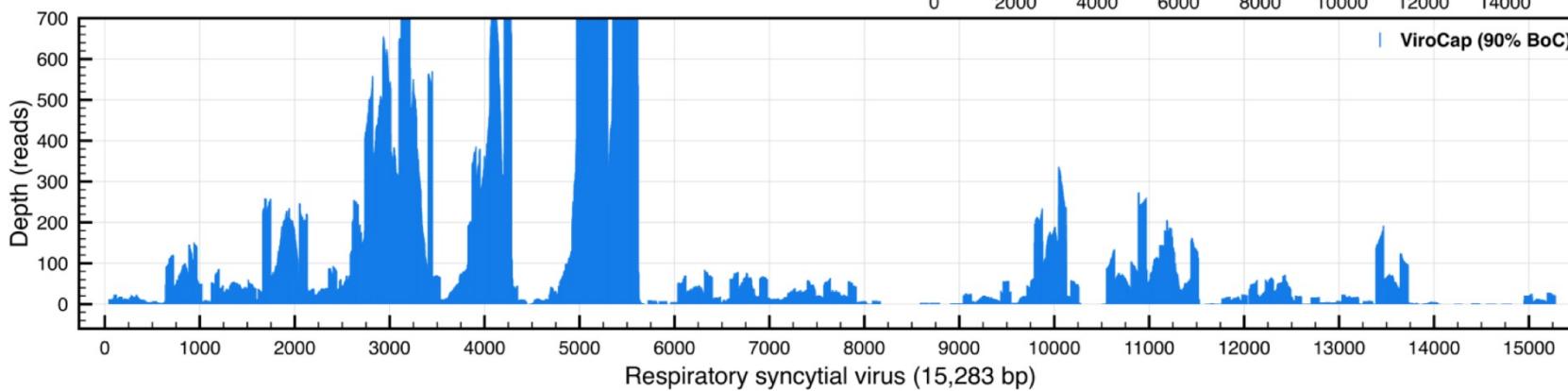
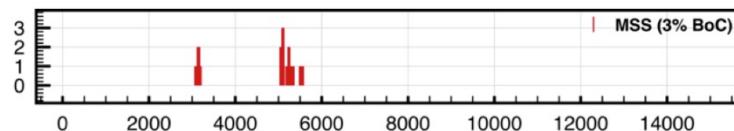
100% Coverage with ViroCap

Parainfluenzavirus 3
60% BoC gain; 1,918 PVR fold increase



Low Abundance Detection

Respiratory syncytial virus (pooled)
87% BoC gain; 6,042 PVR fold increase



Viral Capture Summary

We've designed and tested a large scale viral capture assay, ViroCap.

ViroCap provides viral sequence enrichment hundreds to thousands of fold increase over MSS.

Divergent viruses were captured with as low as 65% overall sequence similarity in our tests.

Bioinformatics approaches to capture sequence selection reduced potential sequence target pools 3–29x.

Limitations include library representation, very low input samples, and cost.

We often routinely use ViroCap enrichment in our metagenomic studies.

Currently using ViroCap v.2 and working on v.3 for 2022 release.

Further Reading...

Wylie KM, Wylie TN, Buller R, Herter B, Cannella MT, Storch GA. *Detection of Viruses in Clinical Samples by Use of Metagenomic Sequencing and Targeted Sequence Capture*. J Clin Microbiol. 2018 Nov 27;56(12):e01123-18. doi: 10.1128/JCM.01123-18. PMID: 30232133; PMCID: PMC6258860.

Wylie TN, Wylie KM, Herter BN, Storch GA. *Enhanced virome sequencing using targeted sequence capture*. Genome Res. 2015 Dec;25(12):1910-20. doi: 10.1101/gr.191049.115. Epub 2015 Sep 22. PMID: 26395152; PMCID: PMC4665012.

Viral Identification

Great, we can sequence viruses now; but, how do we know which ones are present?

ViroMatch

ViroMatch is virus characterization software that evaluates NGS nucleotide reads. It screens/removes host sequences and performs nucleotide and translated nucleotide alignments in an iterative fashion. Candidate virus reads are validated against all current references in NCBI nt/nr.

In our virus-focused metagenomic studies, **ViroCap** followed by **ViroMatch** is our one-two punch!

ViroMatch History



Kristine M. Wylie, Ph.D Assistant Professor of Pediatrics

Early Iterations of the HMP Viral Discovery Pipeline (ca. 2010–2011)

Holtz LR, **Wylie KM**, Sodergren E, Jiang Y, Franz CJ, Weinstock GM, Storch GA, Wang D. *Astrovirus MLB2 viremia in febrile child*. Emerg Infect Dis. 2011 Nov;17(11):2050-2.

Wylie KM, Mihindukulasuriya KA, Sodergren E, Weinstock GM, Storch GA. *Sequence analysis of the human virome in febrile and afebrile children*. PLoS One. 2012;7(6):e27735. doi: 10.1371/journal.pone.0027735. Epub 2012 Jun

McElvania TeKippe E, **Wylie KM**, Deych E, Sodergren E, Weinstock G, Storch GA. *Increased prevalence of anellovirus in pediatric patients with fever*. PLoS One. 2012;7(11):e50937. doi: 10.1371/journal.pone.0050937. Epub 2012 Nov 30. PubMed PMID: 23226428; PubMed Central PMCID: PMC3511395.

Wylie KM, Mihindukulasuriya KA, Zhou Y, Sodergren E, Storch GA, Weinstock GM. *Metagenomic analysis of double-stranded DNA viruses in healthy adults*. BMC Biol. BioMed Central Ltd; 2014;12(1):71.

Wylie TN, **Wylie KM**, Herter BN, Storch GA. *Enhanced virome sequencing using targeted sequence capture*. Genome Res. 2015 Dec;25(12):1910-20. doi: 10.1101/gr.191049.115. Epub 2015 Sep 22.

ViroMatch Development Goals:

- Research not clinical • One-step automation • Replacement of proprietary components
- Efficient usage of resources • Map reduce for database splitting/handling • CLI options (not hardcoded)
 - Taxonomy API • OOP code base for extensibility • Docker deployment (HPC) • Distribution & documentation

Version	Code Base	Automation	Processing
1	Multiple shell scripts	Manually executed (multi-step)	Linear
2	OOP Perl	One-step	Linear
3	OOP Python	One-step	Linear
4	Snakemake (OOP Python back-end)	One-step	Linear or Parallel

<https://hub.docker.com/r/twylie/viromatch>

The screenshot shows the Docker Hub interface for the `twylie/viromatch` repository. At the top, there's a search bar and navigation links for Explore, Pricing, Sign In, and Sign Up. Below the header, the repository path `Explore > twylie/viromatch` is shown. The main card features a blue 3D cube icon, the repository name `twylie/viromatch` with a star icon, and a download count of `10K+`. It also shows the author `twylie` and the last update time. A brief description states: "ViroMatch analyzes millions of short next-generation sequence reads to identify viral sequences." Below the card, there are tabs for Overview and Tags, with "Overview" being active. The "Overview" section contains a large heading "ViroMatch" and a detailed text block explaining the tool's purpose and workflow. To the right, a "Docker Pull Command" box contains the command `docker pull twylie/viromatch`.

twylie/viromatch ☆

By [twylie](#) • Updated 5 months ago

ViroMatch analyzes millions of short next-generation sequence reads to identify viral sequences.

Container

Overview Tags

ViroMatch

We developed ViroMatch to analyze datasets of millions of short next-generation sequence reads to identify viral sequences. The ViroMatch workflow incorporates both nucleotide and translated amino acid sequence alignment against a comprehensive database of viral reference genomes, which allows us the sensitivity to detect highly conserved and divergent viral sequences. Specifically, metagenomic sequences are first screened for putative viral reads by nucleotide alignment with BWA-MEM and translated alignment with Diamond against a database of viral genomes (downloaded from NCBI). This first screen is fast, but the hits include many false positives. Therefore, the putative viral hits are then aligned to the comprehensive NCBI nt nucleotide database using BWA-MEM and the comprehensive NCBI nr protein database using Diamond. Only sequences with an unambiguous alignment to a viral reference are counted as viral hits. Ambiguous hits (those that have alignments with similar scores to viruses and human, bacteria, etc.) are not counted. Ambiguous hits include those that map to repetitive regions that are not suitable for determining virus positivity. This pipeline has been used primarily for analysis of vertebrate viruses, including human viruses found in clinical samples.

Docker Pull Command

```
docker pull twylie/viromatch
```

twylie@wustl.edu

<https://github.com/twylie/viromatch>

Why GitHub? Team Enterprise Explore Marketplace Pricing

Search Sign in Sign up

twylie / viromatch Watch 1 Star 0 Fork 0

Code Issues Pull requests Actions Projects Security Insights

main 1 branch 0 tags Go to file Code

tnwylie Proofread docs for web site. 5630d32 26 days ago 8 commits

bin	First commit for the distribution.	last month
db	First commit for the distribution.	last month
docs	Proofread docs for web site.	26 days ago
hugo	Proofread docs for web site.	26 days ago
viromatch	Added support for 'results dir' symlink in parallel-processing.	29 days ago
.gitignore	First commit for the distribution.	last month
AUTHOR	First commit for the distribution.	last month
CONTACT	First commit for the distribution.	last month
Dockerfile	First commit for the distribution.	last month
LICENSE	First commit for the distribution.	last month
README.md	First commit for the distribution.	last month
VERSION	First commit for the distribution.	last month

About

ViroMatch analyzes millions of short next-generation sequence reads to identify viral sequences.

Readme View license

Releases

No releases published

Packages

No packages published

Languages

Python 94.4% HTML 4.6% Dockerfile 1.0%

README.md

ViroMatch

twylie@wustl.edu

<https://twylie.github.io/viromatch/>

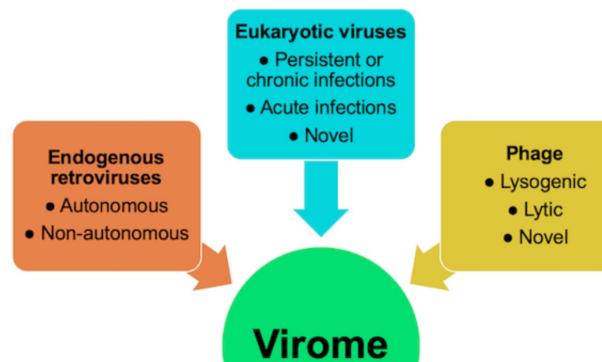
The screenshot shows the ViroMatch website with a red header bar containing a virus icon and the word "VIROMATCH". Below the header is a search bar with placeholder text "Search...". A sidebar on the left lists navigation links: "Home", "1. Quick Start", "2. Download and Installation", "3. Running ViroMatch", "4. Pipeline Overview", and "5. Project Info". A funding section mentions NIH CTSA UL1TR002345 and NIH 1R01HD095986. A disclaimer states the resource is for research purposes only. A Creative Commons BY-NC license logo is present, along with copyright information for T.N. Wylie and K.M. Wylie from 2019-2020.

VIROMATCH

A computational pipeline for detection of viral reads from complex metagenomic data.

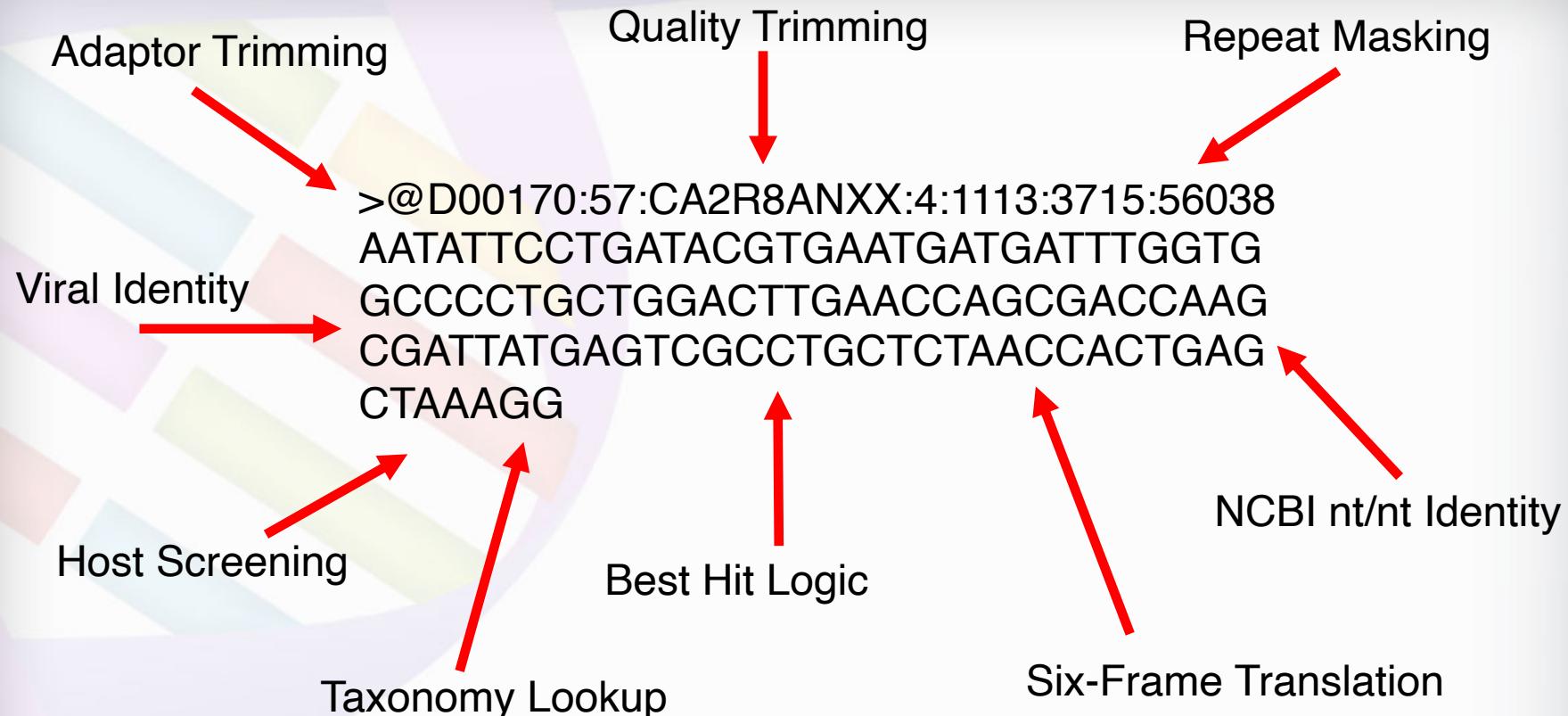
The Virome

The virome is the viral component of the microbiome. Viruses are a diverse group of microbes. They require a host cell for the production of new viral particles, but the hosts range from microbial to eukaryotic cells. There is no single gene that is common among the group. In fact, viral genomes are strikingly varied. They can be composed of DNA or RNA; double stranded or single stranded; positive sense or negative sense; non-segmented or segmented. The diversity in viruses adds complexity to genomic analysis of the virome. Next generation sequencing is well-suited to virome analysis, as it enables culture-independent assessment of any type of viral nucleic acid in samples ranging from sea water to human clinical samples. Sequencing allows the comprehensive characterization of the virome and discovery of novel viruses. We developed **ViroMatch** to analyze datasets of millions of short sequence reads to identify viral sequences.



Wylie TN, Wylie KM. ViroMatch: A Computational Pipeline for the Detection of Viral Sequences from Complex Metagenomic Data. Microbiol Resour Announc. 2021 Mar 4;10(9):e01468-20. doi: 10.1128/MRA.01468-20. PMID: 33664143; PMCID: PMC7936641.

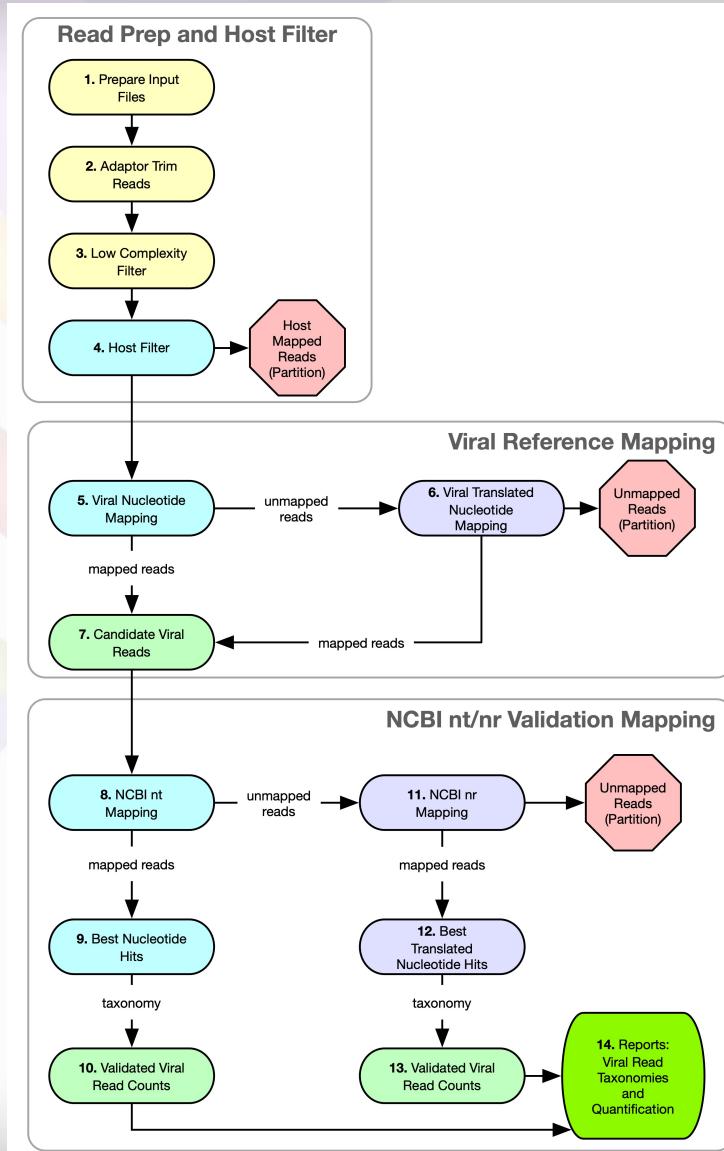
Micro-transactions



0.0005 sec/transaction
9 transactions = 0.0045

50,000,000 transactions = 225,000 seconds
225,000 seconds = **62.5 hours**

ViroMatch Overview



Wylie TN, Wylie KM.

ViroMatch: A Computational Pipeline for the Detection of Viral Sequences from Complex Metagenomic Data.
Microbiol Resour Announc. 2021 Mar 4;10(9):e01468-20.
doi: 10.1128/MRA.01468-20.
PMID: 33664143; PMCID: PMC7936641.

Taxonomy Reports

NUCLEOTIDE ALIGNMENT READ COUNTS

LINEAGE (R1 + R2)

[1] read count
[2] percent
[3] lineage

149	0.634	Viruses --> Caudovirales --> Podoviridae --> unclassified Podoviridae --> crAss-like viruses --> environmental samples --> uncultured crAssphage
36	0.153	Viruses --> Anelloviridae --> Gyrovirus --> Chicken anemia virus
12	0.051	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D
5	0.021	Viruses --> Anelloviridae --> Gyrovirus --> unclassified Gyrovirus --> Avian gyrovirus 2
4	0.017	Viruses --> Anelloviridae --> Gyrovirus --> unclassified Gyrovirus --> Gyrovirus GyV3
4	0.017	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus D9
4	0.017	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus D10
3	0.013	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus 45
3	0.013	Viruses --> Anelloviridae --> Gyrovirus --> unclassified Gyrovirus --> Gyrovirus Tu789
3	0.013	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus 60a
2	0.009	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus 17
2	0.009	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus 25
2	0.009	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus 58
1	0.004	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus 38
1	0.004	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus 26
1	0.004	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus 20
1	0.004	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus 19
1	0.004	Viruses --> Adenoviridae --> Mastadenovirus --> Human mastadenovirus D --> Human adenovirus D13

235 TOTAL

GENUS (R1 + R2)

[1] read count
[2] percent
[3] genus

149	0.634	unknown genus uncultured crAssphage
48	0.204	Gyrovirus
38	0.162	Mastadenovirus

235 TOTAL

SPECIES (R1 + R2)

[1] read count
[2] percent
[3] species

149	0.634	uncultured crAssphage
38	0.162	Human mastadenovirus D
36	0.153	Chicken anemia virus
5	0.021	Avian gyrovirus 2
4	0.017	Gyrovirus GyV3
3	0.013	Gyrovirus Tu789

235 TOTAL

ViroMatch DBs

Database	Description
Adaptor	Sequences used for trimming adaptor.
Host	Human reference genome.
Viral DB (nucleotide)	Database of viral sequences from GenBank & RefSeq. (1 file)
Viral DB (translated nucleotide)	Six-frame translated version of Viral DB. (1 file)
NCBI nt – 75 files, <4 GB each	Split version of NCBI nt.
NCBI nr – 41 files, <4 GB each	Split version of NCBI nr.

ViroMatch Usage

```
LSF_DOCKER_VOLUMES="/storage1/fs1/kwylie/Archive/2020_09_21_AHA_RAW_DATA_ONLY/RAW_DATA:/storage1/fs1/kwylie/Archive/2020_09_21_AH  
A_RAW_DATA_ONLY/RAW_DATA /storage1/fs1/kwylie/Active/viromatchDBs:/storage1/fs1  
/kwylie/Active/viromatchDBs /storage1/fs1/kwylie/Active/2020_09_21_AHA:/storage1/fs1/kwylie/Active/2020_09_21_AHA/  
/scratch1/fs1/kwylie/testPP:/scratch1/fs1/kwylie/testPP" \  
bsub \  
-M 16G -R "select[mem>16G] rusage[mem=16G]" \  
-G compute-kwylie \  
-q general \  
-e $PWD/LSF.err \  
-o $PWD/LSF.out \  
-a 'docker(twylie/viromatch:latest)' \  
viromatch \  
--input /storage1/fs1/kwylie/Archive/2020_09_21_AHA_RAW_DATA_ONLY/RAW_DATA/gerald_HG3LNDXY_4_GGTTGGAC-TACAGGAT.bam \  
--outdir /scratch1/fs1/kwylie/testPP/test \  
--host /storage1/fs1/kwylie/Active/viromatchDBs/host/GRCh38_latest_genomic.fna \  
--adaptor /storage1/fs1/kwylie/Active/viromatchDBs/adaptor/adaptor.fqtrim \  
--nt /storage1/fs1/kwylie/Active/viromatchDBs/ncbi/nt/nt.pp.fofn \  
--nr /storage1/fs1/kwylie/Active/viromatchDBs/ncbi/nr/nr.pp.fofn \  
--sampleid 'GAGATTCC-CCTATCCT_S58_L001' \  
--viralfna /storage1/fs1/kwylie/Active/viromatchDBs/viral-only/nuc/ncbi_viral_refseq_gn_20191120.fna \  
--viralfaa /storage1/fs1/kwylie/Active/viromatchDBs/viral-only/trans_nuc/ncbi_viral_refseq_gn_20191120.dmdn \  
--taxid /storage1/fs1/kwylie/Active/viromatchDBs/taxonomy/taxonomy.tsv \  
--keep \  
--wustlconfig $PWD/run.yaml
```

Snakemake

Snakemake is a workflow management system (WMS)

- Other WGS: Galaxy, CWL (Cromwell), GNU Make, NextFLow, Taverna, Arvados, DNAexus, SevenBridges, etc.

Snakemake was written by an bioinformatician (Johannes Köster) and has been around since 2011—2012; approx. 230,000+ downloads and 5 new citations a week

Cross-platform (*nix, Mac, Windows) and portable

Free and open source (MIT license)

Snakemake uses pure Python syntax. There is no tool specific-language to learn like in GNU Make, NextFlow, WDL, etc.

Anything that you can do in Python, you can do with Snakemake (since you can pretty much execute arbitrary Python code anywhere inline)

Highlights:

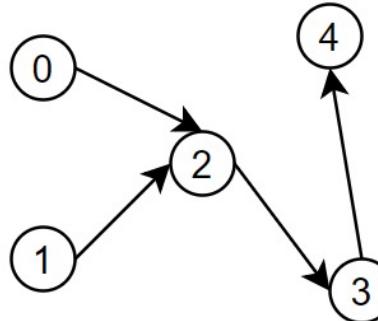
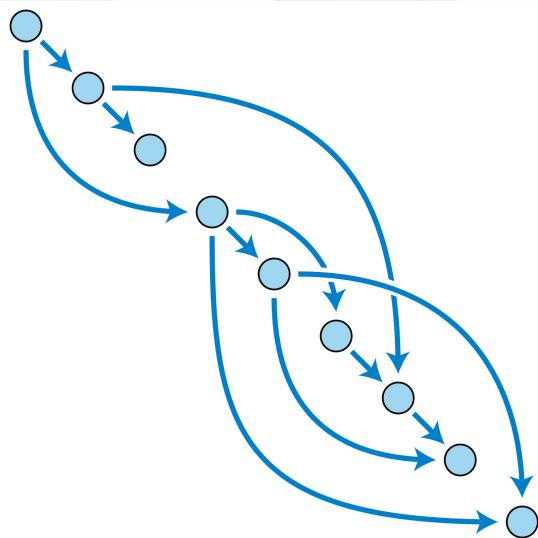
- Reproducibility
- Readability
- Parallel processing and multi-core support
- Scalable to server, cluster, and grid/cloud
- Friendly for Python and R users
- Use of “fuzzy” wildcards



Directed acyclic graph (DAG)

"In mathematics, particularly graph theory, and computer science, a directed acyclic graph (DAG) is a directed graph with no directed cycles. That is, it consists of vertices and edges (also called arcs), with each edge directed from one vertex to another, such that following those directions will never form a closed loop. A directed graph is a DAG if and only if it can be topologically ordered, by arranging the vertices as a linear ordering that is consistent with all edge directions. DAGs have numerous scientific and computational applications, ranging from biology (evolution, family trees, epidemiology) to sociology (citation networks) to computation (scheduling)."

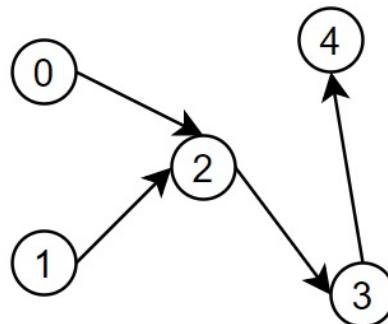
— Wikipedia



Graph 1

Directed acyclic graph (DAG)

```
rule bwa_map:  
    input:  
        "data/genome.fa",  
        "data/samples/A.fastq"  
    output:  
        "mapped_reads/A.bam"  
    shell:  
        "bwa mem {input} | samtools view -Sb - > {output}"
```



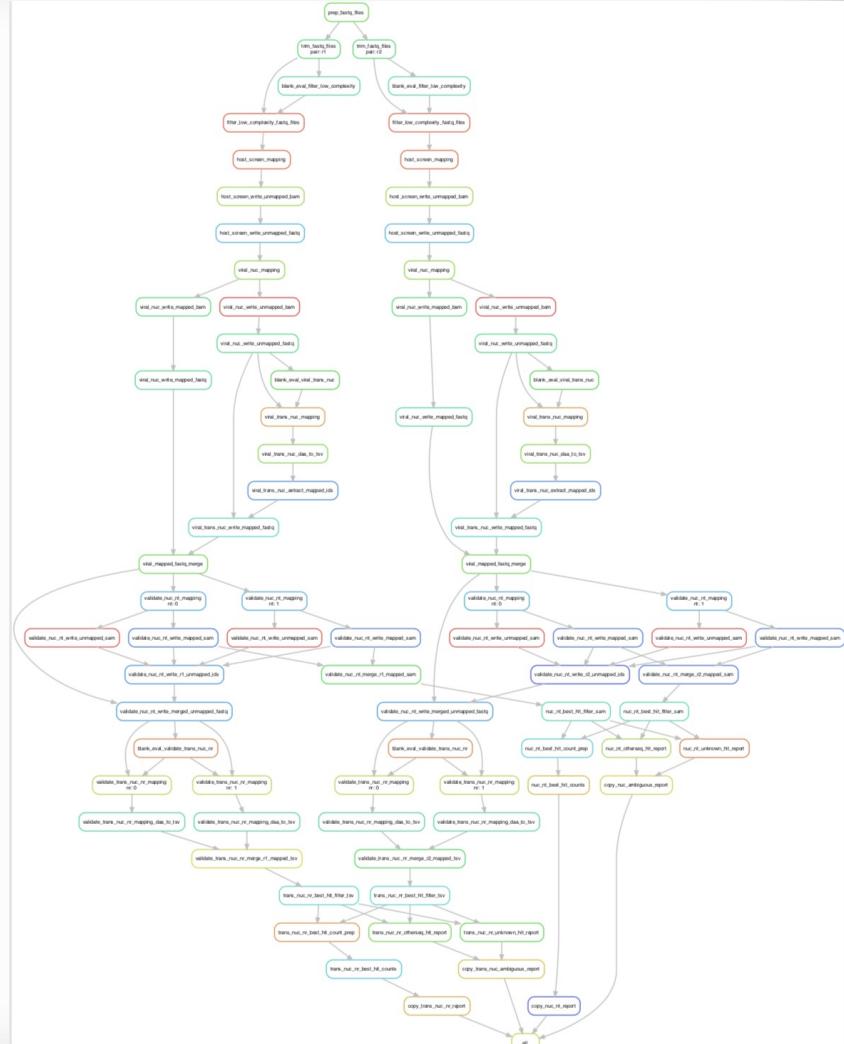
Graph 1

ViroMatch

Snakemake DAG Rulegraph



Snakemake I/O DAG



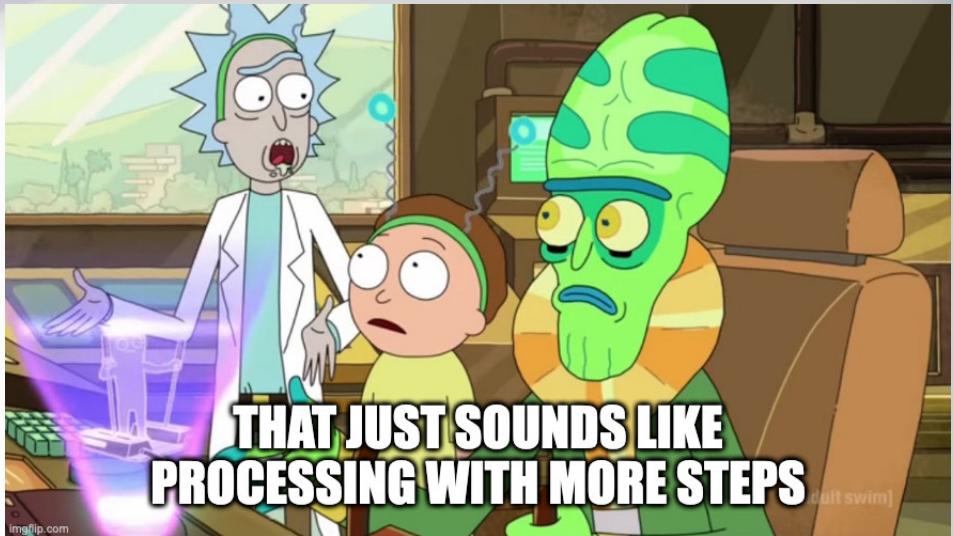
Parallel Processing (LSF at RIS)

Is Snakemake parallel processing possible on the LSF compute server at RIS?

A *qualified* yes

--wustlconfig

```
docker:  
  image: 'twylie/viromatch:latest'  
  volumes:  
    - '/scratch1/fs1/kwylie/testPP:/scratch1/fs1/kwylie/testPP'  
    - '/storage1/fs1/kwylie/Active/viromatchDBs:/storage1/fs1/kwylie/Active/viromatchDBs'  
    - '/storage1/fs1/kwylie/Active/2020_09_21_AHA/redoPP2:/storage1/fs1/kwylie/Active/2020_09_21_AHA/redoPP2'  
    - '/storage1/fs1/kwylie/Archive/2020_09_21_AHA_RAW_DATA_ONLY/RAW_DATA:/storage1/fs1/kwylie/Archive/2020_09_21_AHA_RAW_DATA_ONLY/RAW_DATA'  
  
lsf:  
  memory: '16G'  
  results dir: '/storage1/fs1/kwylie/Active/2020_09_21_AHA/redoPP2/mirror/'  
  cores: '50'  
  local cores: '1'  
  compute group: 'compute-kwylie'  
  queue: 'general'  
  latency wait: '100'  
  restart times: '3'
```



ViroMatch & Snakemake

Live Demo

Acknowledgments

Collaborators	
Vikas Dharnidharka, M.D., M.P.H.	Washington University School of Medicine
Linda M. Ernst, M.D.	Northshore University HealthSystem
Stephanie A. Fritz, M.D., M.S.C.I.	Washington University School of Medicine
Gregory A. Storch, M.D.	Washington University School of Medicine
Molly J. Stout, M.D., M.S.C.I.	University of Michigan Health
Philip I. Tarr, M.D.	Washington University School of Medicine
Krisine M. Wylie, Ph.D	Washington University School of Medicine

ViroMatch Funding

National Institute of Health Clinical and Translational Science

NIH CTSA UL1TR002345

Institute of Clinical and Translational Sciences Award (ICTS)

ICTS GRANT ID#: CTRFP2019-17 (CSTA1215)

Tools for Automated Interpretation of Viruses from Next-Generation Sequencing Data

ICTS/Washington University School of Medicine

PI: Todd N. Wylie

National Institute of Health Research Project (R01)

1R01HD095986

The Vaginal Microbiome, Maternal Response, and Preterm Birth

Washington University School of Medicine

PIs: Kristine M. Wylie; Molly Stout

Kristine M. Wylie Lab — WashU

Kristine M. Wylie

Aaron Aanestad

McKenzie Anderson

Madison Eschbach

Sam Jeyasingh

Lauren King

Hunter Olson

Jane Schrimpf

Galileu Barbosa Costa*

Haley Gula*

Brandi N. Herter*

Stephanie Lushniak*

Renee Rubiano*

Dylan Stoutenburg*

Claire Wild*

*Former members Wylie Lab

Contact



Todd N. Wylie

Assistant Professor of Pediatrics
Department of Pediatrics
Washington University School of Medicine
660 S. Euclid Avenue
St. Louis, MO 63110
Campus Box 8208
Office: (314) 273-9078
E-mail: twylie@wustl.edu
Twitter: @ToddNWylie