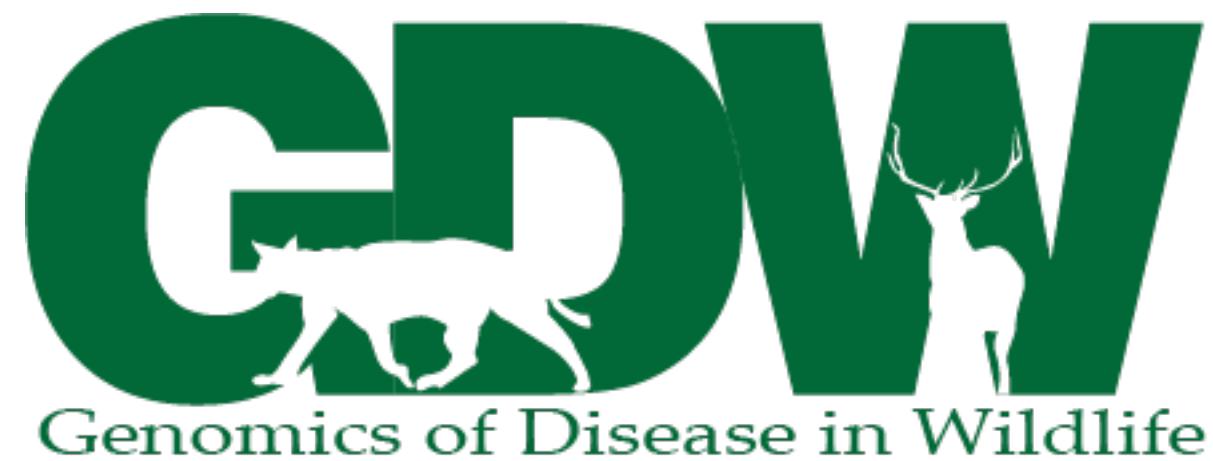


# Metagenomics and disease

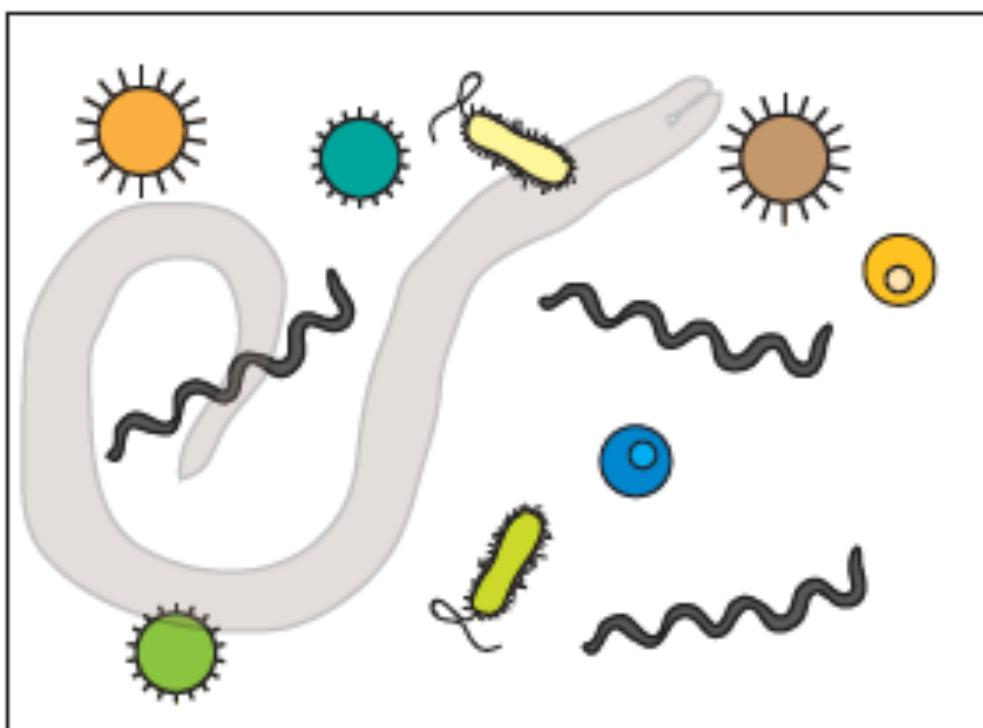
Mark Stenglein, GDW



Metagenomics is the study of >1 genome

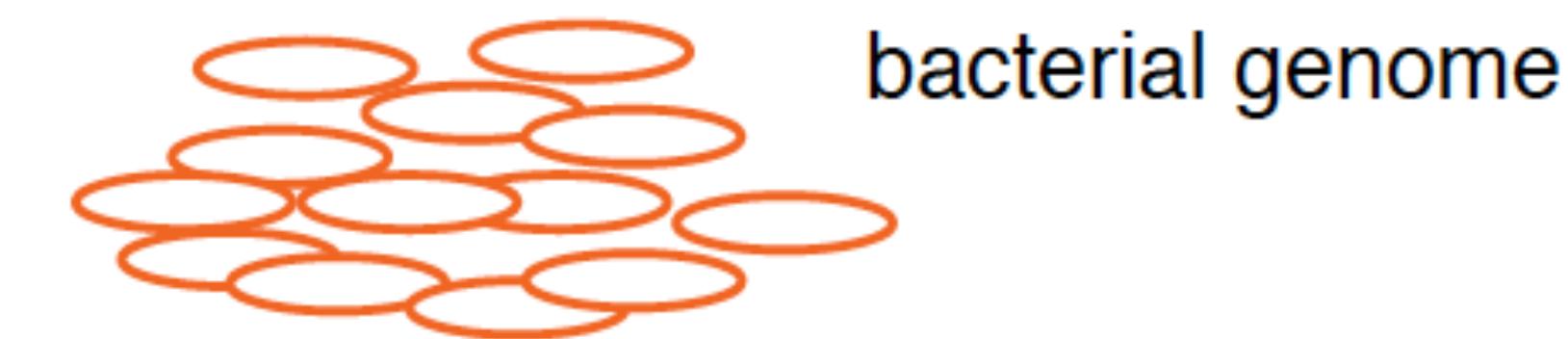
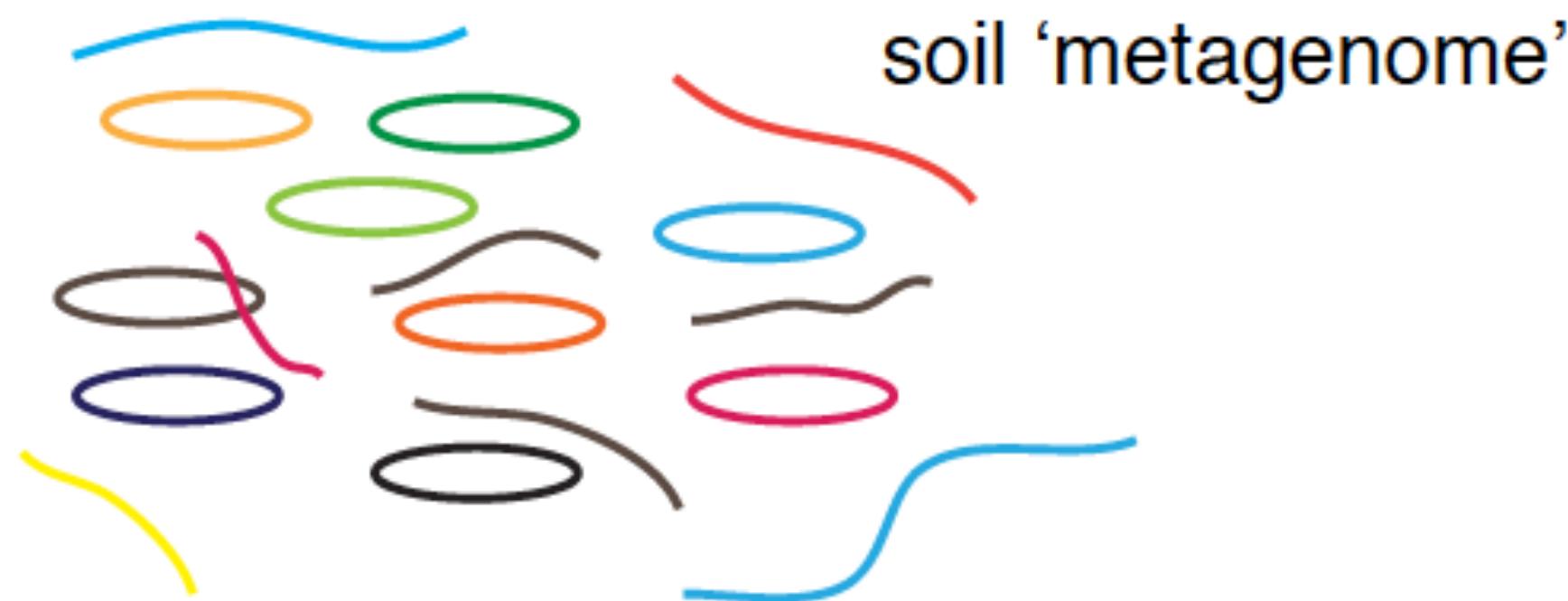
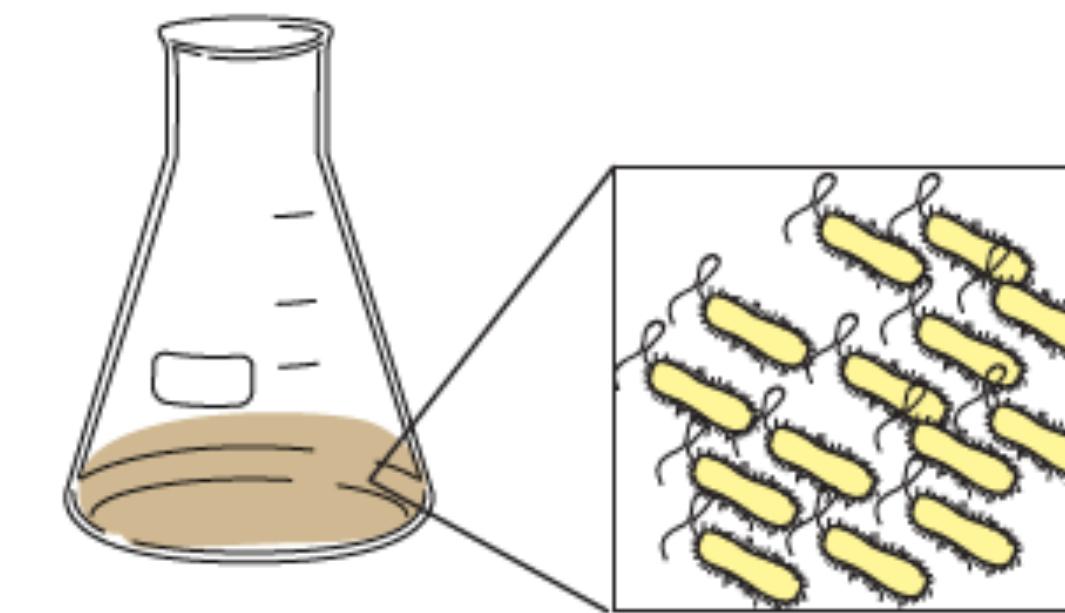
## Many genomes

soil community

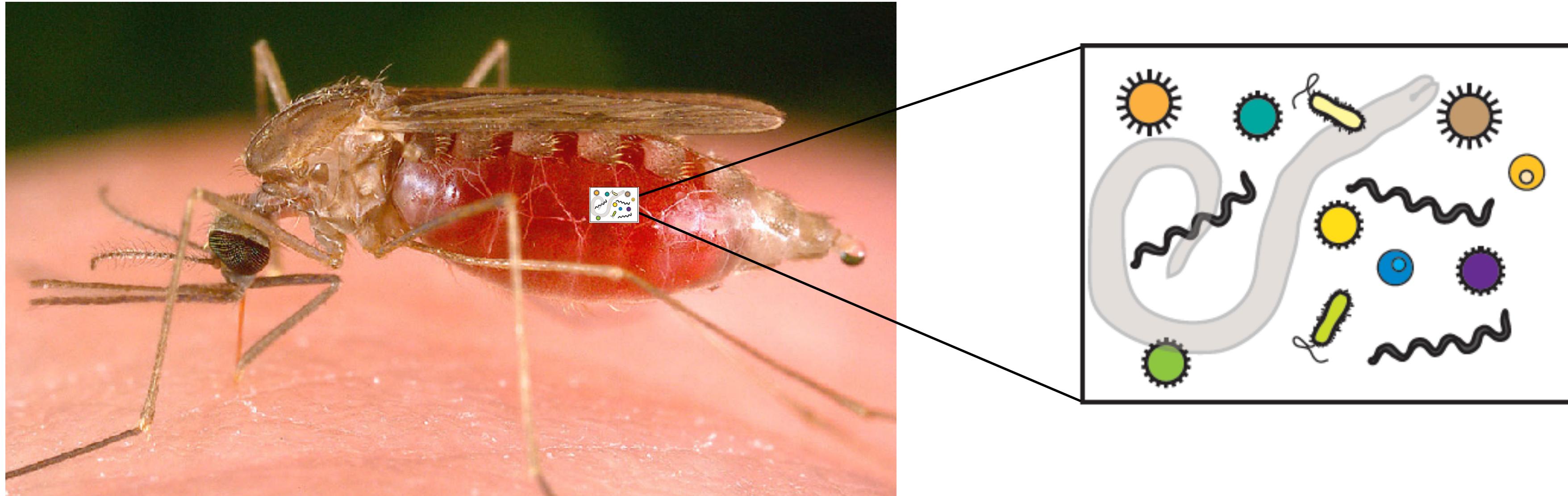


## 1 genome

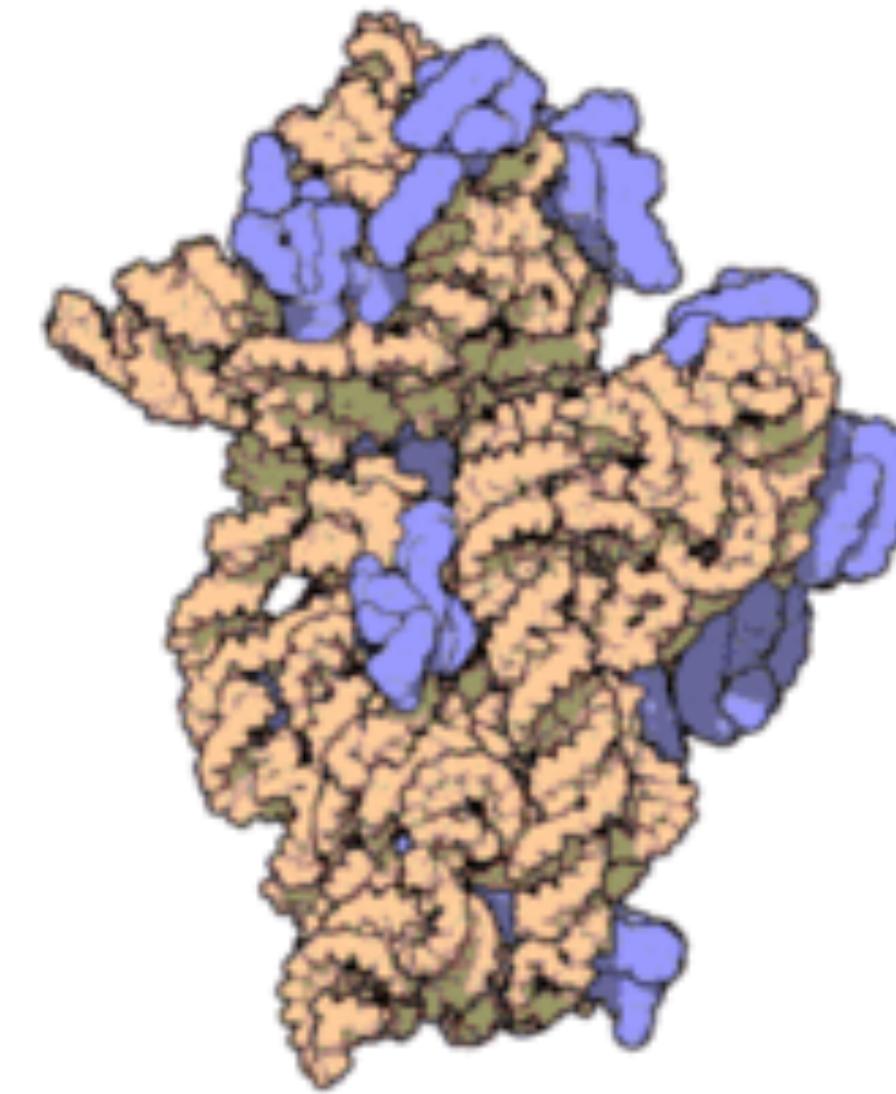
bacterial isolate



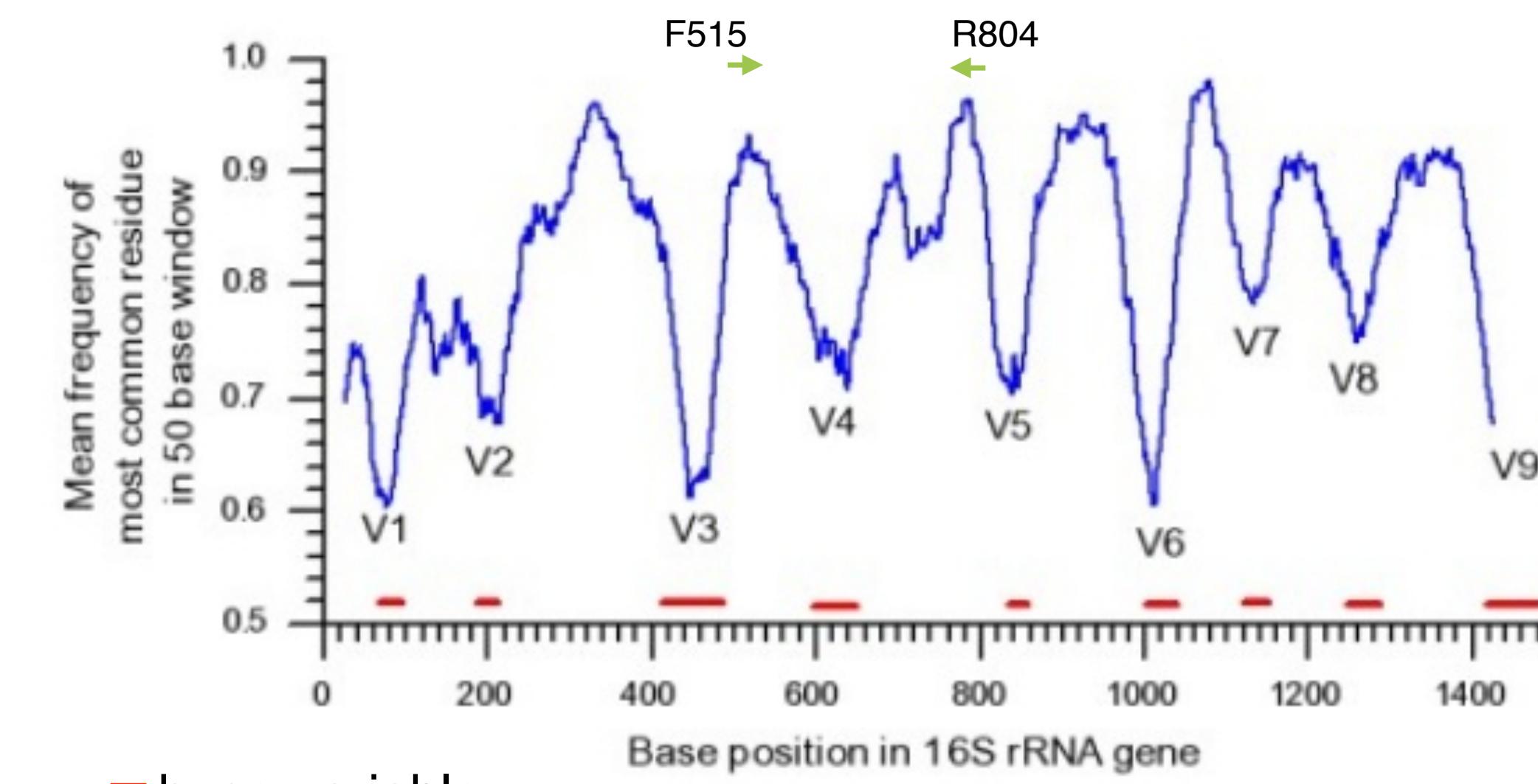
If you sequence total nucleic acid from an intact multicellular organism,  
you are doing metagenomics



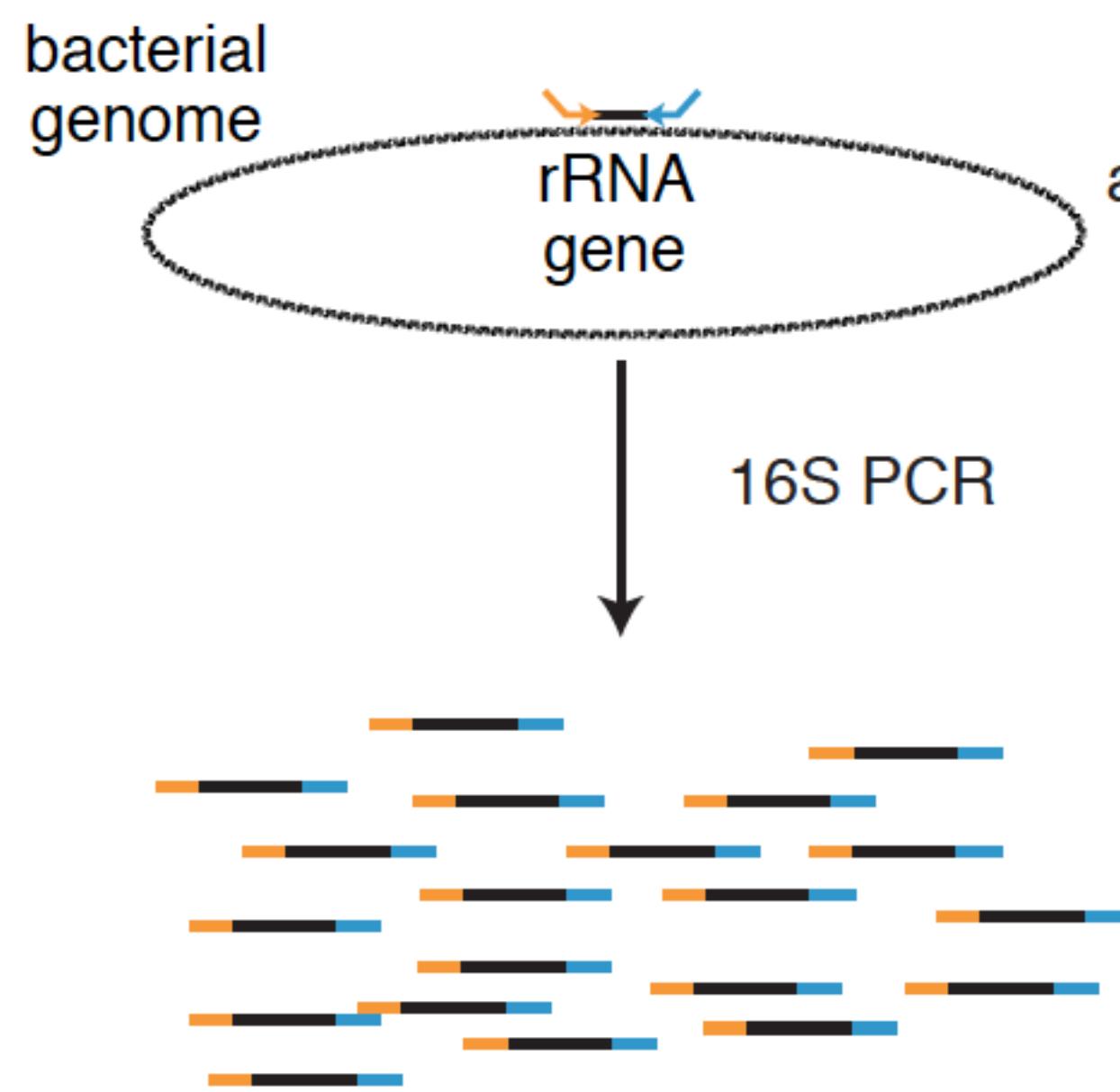
16S sequencing (microbiome sequencing) could be considered a form of metagenomics



bacterial 30S ribosomal subunit  
16S rRNA is in orange  
(purple: ribosomal proteins)  
*image: wikipedia*

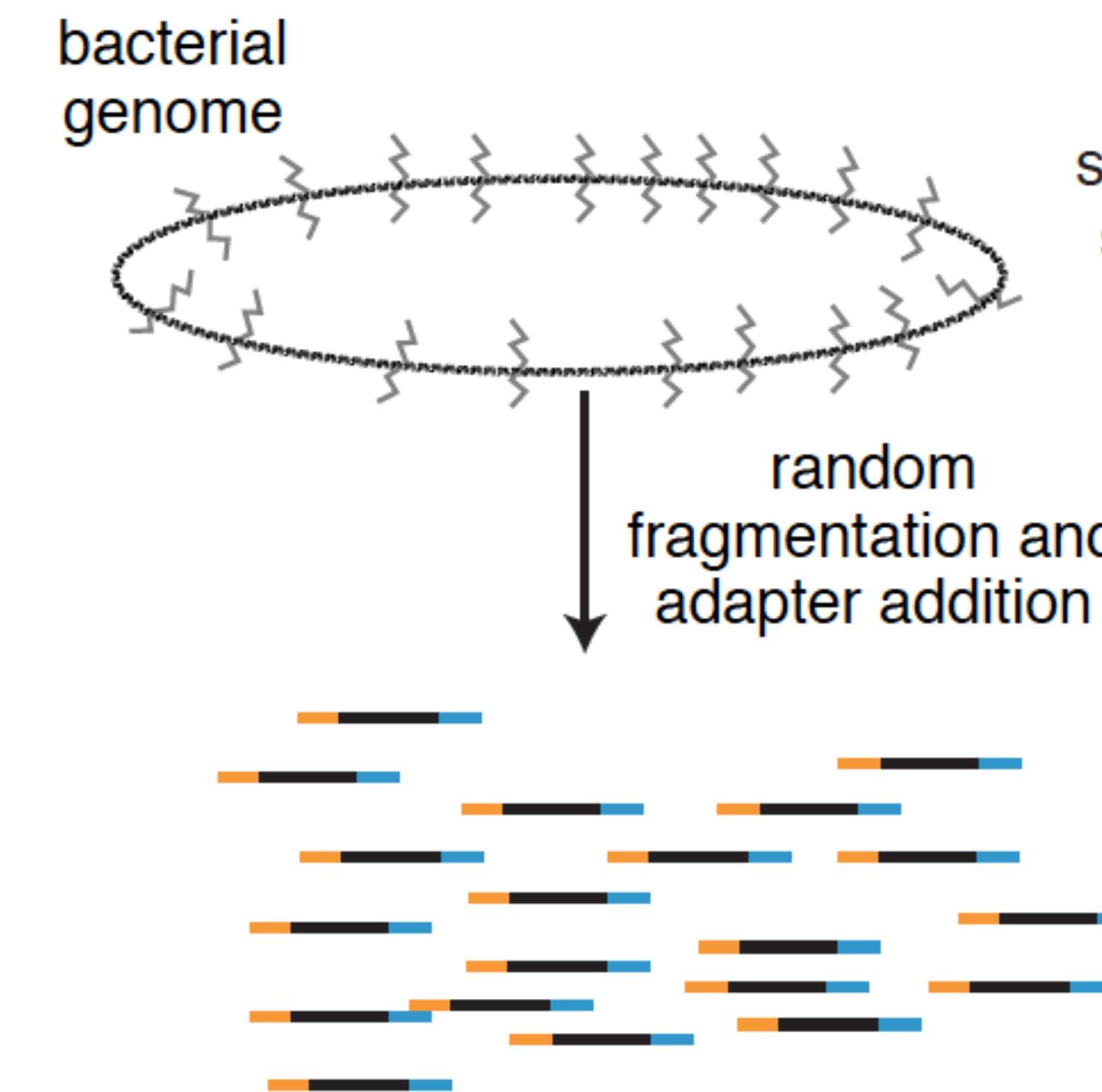


# 16S sequencing vs. shotgun metagenomics



library molecules contain 16S sequences from one or more genomes

16S PCR amplifies ~0.01% of a bacterial genome



shotgun library molecules contain random bits of the genome

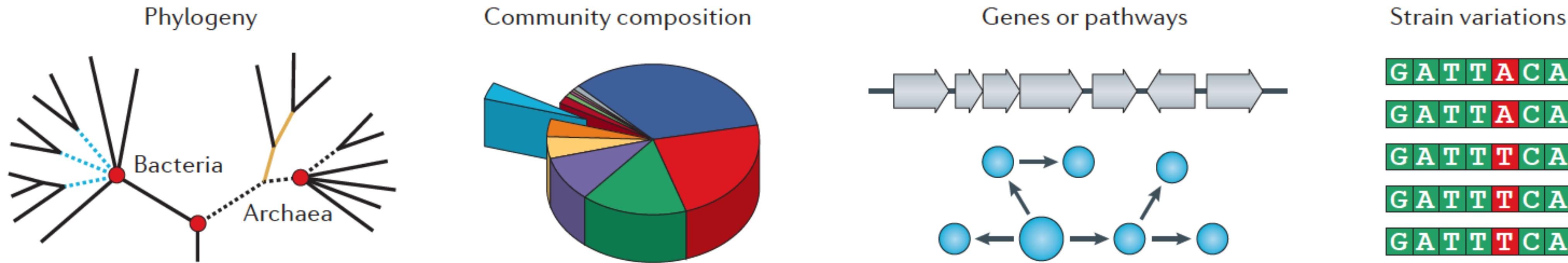
shotgun sequencing samples the entire genome

- Only bacteria and archaea surveyed
- Deeper sampling of bacterial diversity per \$
- Relatively easy to make libraries and interpret results
- Appropriate if all you care about is microbial diversity / ecology

- All organisms studied\*
- Decreased sampling depth per \$
- Enables analysis of other genomic features of organisms, e.g. antimicrobial resistant genes
- Analysis is significantly more difficult

# Applications of metagenomics

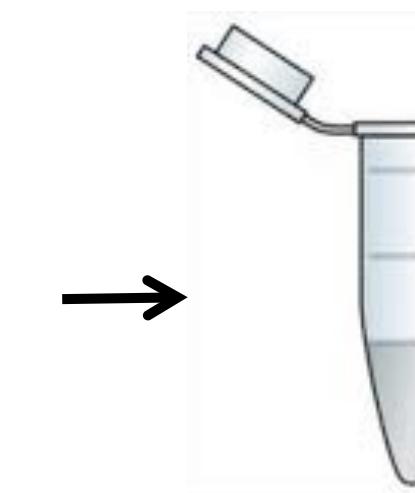
- Community composition analysis (environmental samples, microbiome, ...)
- Characterization of diet, bloodmeal composition
- Identification of genes of interest: AMR genes, industrially useful enzymes
- Pathogen detection and discovery



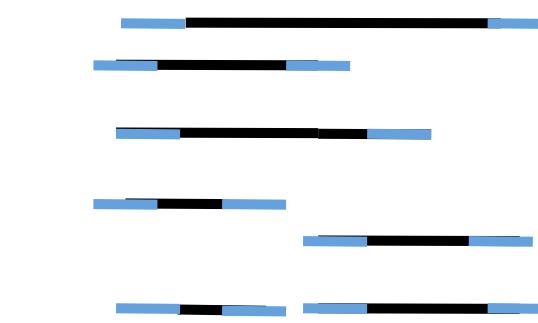
# Pathogen detection and discovery using metagenomic sequencing



case and control  
tissues



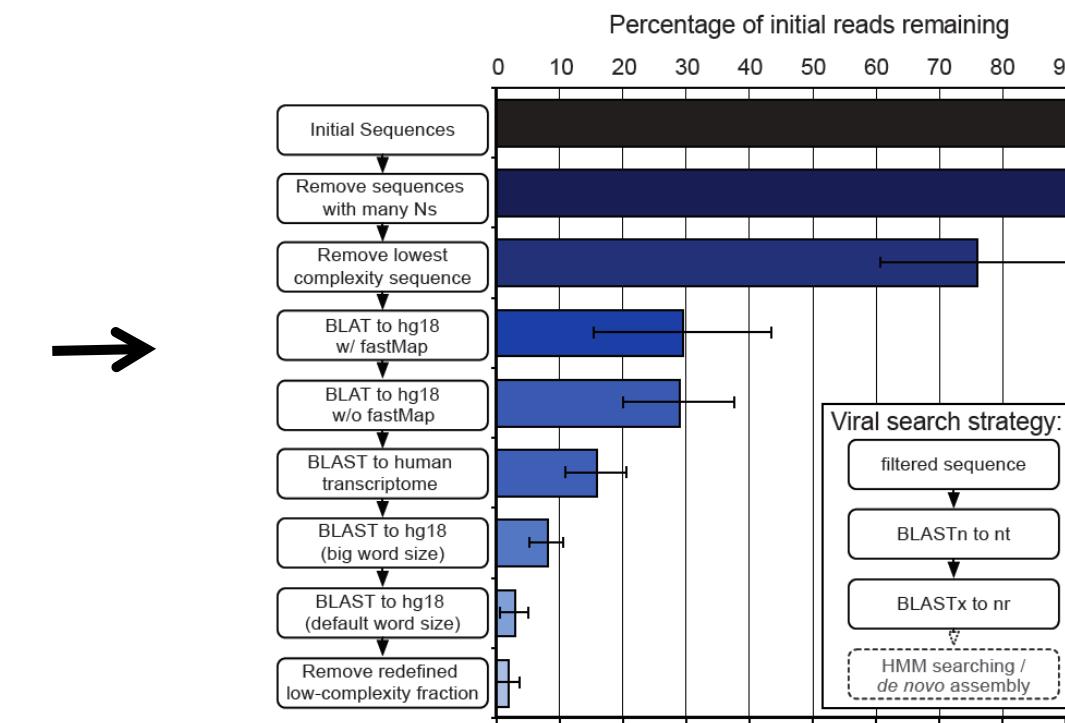
Nucleic acid



Library prep  
/ barcode



Illumina  
sequencing



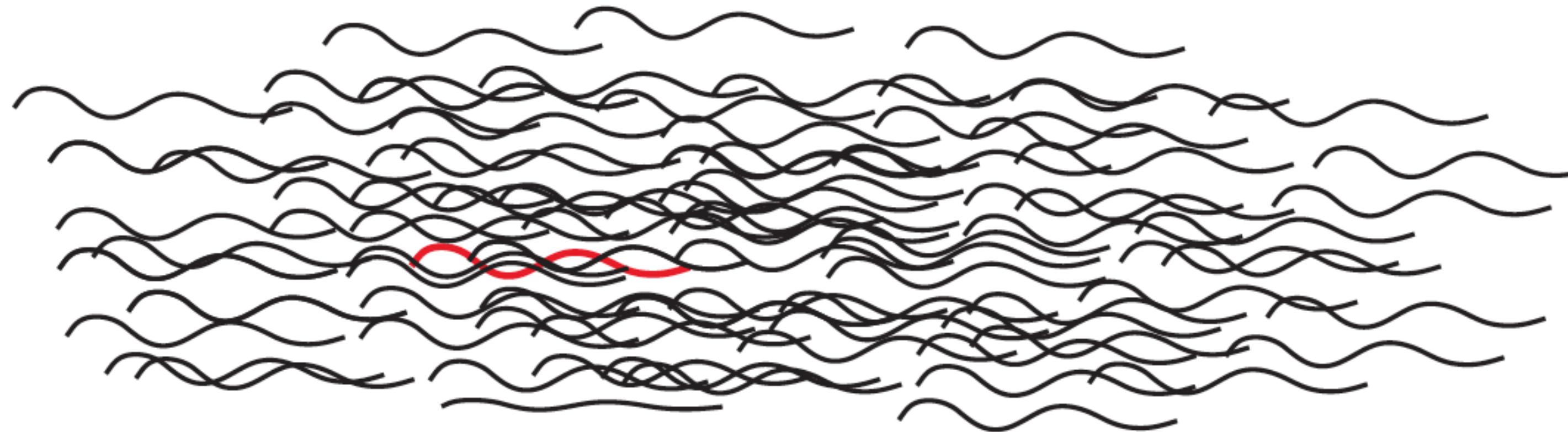
Computational  
Analysis



Follow-up

# Some challenges for metagenomics pathogen discovery

- 1) Pathogen nucleic acid is typically present in a sea of host nucleic acid



~1 viral nucleic acid per  $10^4$ - $10^7$  host nucleic acids

- 2) New pathogens have unknown sequences

**TTTCAG?TTT?ACC????TG??AAA?ACATCC??TATACT??T?**

- 3) Misannotated sequences in databases confound results

How sensitive is NGS for pathogen detection?  
In theory, a single read is sufficient to identify a pathogen  
(but that's cutting it a little close)



Identification of this pathogen completely consistent with histopathology

case had been tested for *ovine herpesvirus 2*



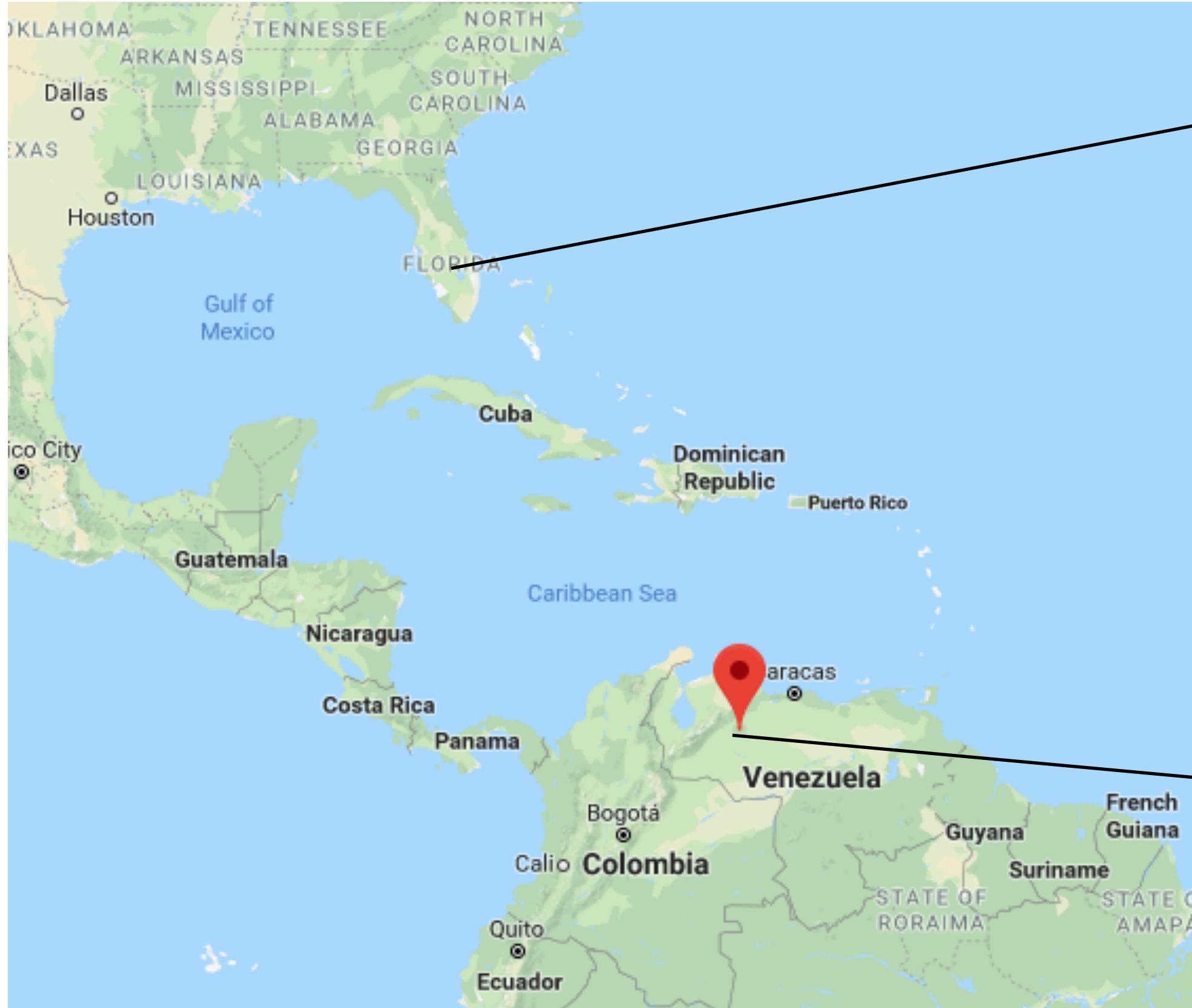
A single read pair aligning to **caprine herpesvirus-2** amongst ~0.5M mule deer reads

PCR is generally more sensitive than NGS for targeted pathogen detection

Laura Hoon-Hanks, DVM

Samples from: Karen Fox DVM, CO Parks & Wildlife

# Example of metagenomic pitfall: Guanarito virus sequence supposedly in a pool of *Culex cedecei* mosquitoes collected in the Florida Everglades



Arenavirus  
Cause of Venezuelan hemorrhagic fever  
Not known to be arthropod borne



*image: American Society of Mammalogists*

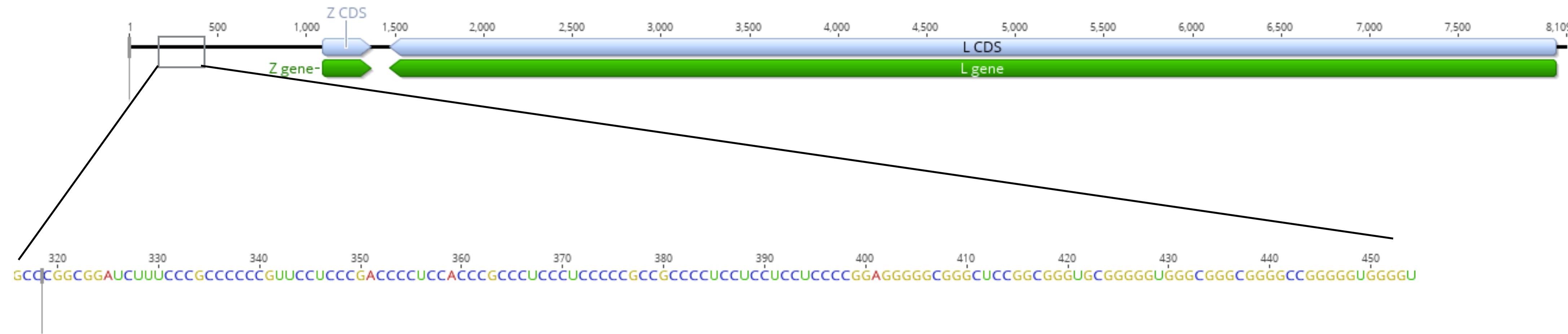
# One of the putative Guanarito virus sequences

```
>NODE_274_length_640_cov_11681_ID_547
GCGGGGGTGGCGGGCGGGCCGGGGTGGGGTCGGCGGGGACCGTCCCCGACCGGCGACCGGCCGCGCCGGC
GCATTTCCACCGCGGCGGTGCGCCGCGACCGGCTCCGGACGGCTGGAAAGGCCGGCGGGAAAGGTGGCTCGGGG
GCCCGTCCCGCCCGTCTTCCCCCGCCCGTCCTCCCCCGGGAGGGCGCGGGTCGGGCGGCGGCGGTGGC
GGCGGGACCACCCCCCGAGTGTTACAGCCCCCGGCAGCAGCACTGCCGAATCCGGGCCGAGGGAGCGAGACCC
GTCGCCGCGCTCTCCCCCTCCCGGCCACCCCCCGCGGGGCCCGGGGTCCCCCGCGCGGGGTCCCCCGCGGGCGCG
CCGGCGGTCTCGTGGGGGCCGGCACCCCTCCCACGGCGCGACCGCTCTCCCACCCCCCTCCCCGCACCCCCGGC
GACGGGGCCCGCGCGGGTGGGGCGGGCGGACTGTCCCCAGTGCGCCCCGGCGGTGCGCCGTCGGGCCGGGG
GGGTTCTCTCGGGGCCACGCGCGTCCCTCGAAGAGGGGACGGCGGAGCGAGCGCACGGGTGGCGCGATGT
CGGCTACCCACCCGACCGTCTTG
```

# BLAST the sequence vs. the NCBI nucleotide database

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Guanarito mammarenavirus isolate CVH-960201 segment L, complete sequence</a>	1081	1081	94%	0.0	99%	KU746283.1
<input type="checkbox"/>	<a href="#">Guanarito mammarenavirus isolate CVH-950801 segment S, complete sequence</a>	1064	1064	90%	0.0	99%	KU746280.1
<input type="checkbox"/>	<a href="#">Chimpanzee 28S ribosomal RNA gene fragment</a>	826	826	100%	0.0	89%	M30950.1
<input type="checkbox"/>	<a href="#">Gorilla 28S ribosomal RNA gene fragment</a>	817	817	99%	0.0	89%	M30951.1
<input type="checkbox"/>	<a href="#">Homo sapiens external transcribed spacer 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribosomal RNA gene, internal transcribed spacer 2, 28S ribosomal RNA gene, and e</a>	808	808	100%	0.0	89%	KY962518.1
<input type="checkbox"/>	<a href="#">Homo sapiens clone BAC JH1 genomic sequence</a>	808	1612	100%	0.0	89%	MF164269.1
<input type="checkbox"/>	<a href="#">Homo sapiens RNA, 45S pre-ribosomal N2 (RNA45SN2), ribosomal RNA</a>	804	804	100%	0.0	89%	NR_146144.1
<input type="checkbox"/>	<a href="#">Homo sapiens RNA, 28S ribosomal N2 (RNA28SN2), ribosomal RNA</a>	804	804	100%	0.0	89%	NR_146148.1
<input type="checkbox"/>	<a href="#">Human DNA sequence from clone CH507-146P16 on chromosome 21, complete sequence</a>	804	804	100%	0.0	89%	CT476837.18
<input type="checkbox"/>	<a href="#">Human ribosomal DNA complete repeating unit</a>	798	798	100%	0.0	89%	U13369.1
<input type="checkbox"/>	<a href="#">Homo sapiens clone BAC JH5 genomic sequence</a>	787	787	100%	0.0	88%	MF164266.1
<input type="checkbox"/>	<a href="#">Homo sapiens RNA, 28S ribosomal N3 (RNA28SN3), ribosomal RNA</a>	784	784	100%	0.0	88%	NR_146154.1
<input type="checkbox"/>	<a href="#">Homo sapiens RNA, 45S pre-ribosomal N3 (RNA45SN3), ribosomal RNA</a>	784	784	100%	0.0	88%	NR_146151.1
<input type="checkbox"/>	<a href="#">Human DNA sequence from clone CH507-528H12 on chromosome 21, complete sequence</a>	784	1707	100%	0.0	88%	FP236383.15

# These Guanarito virus sequences are mis-assembled



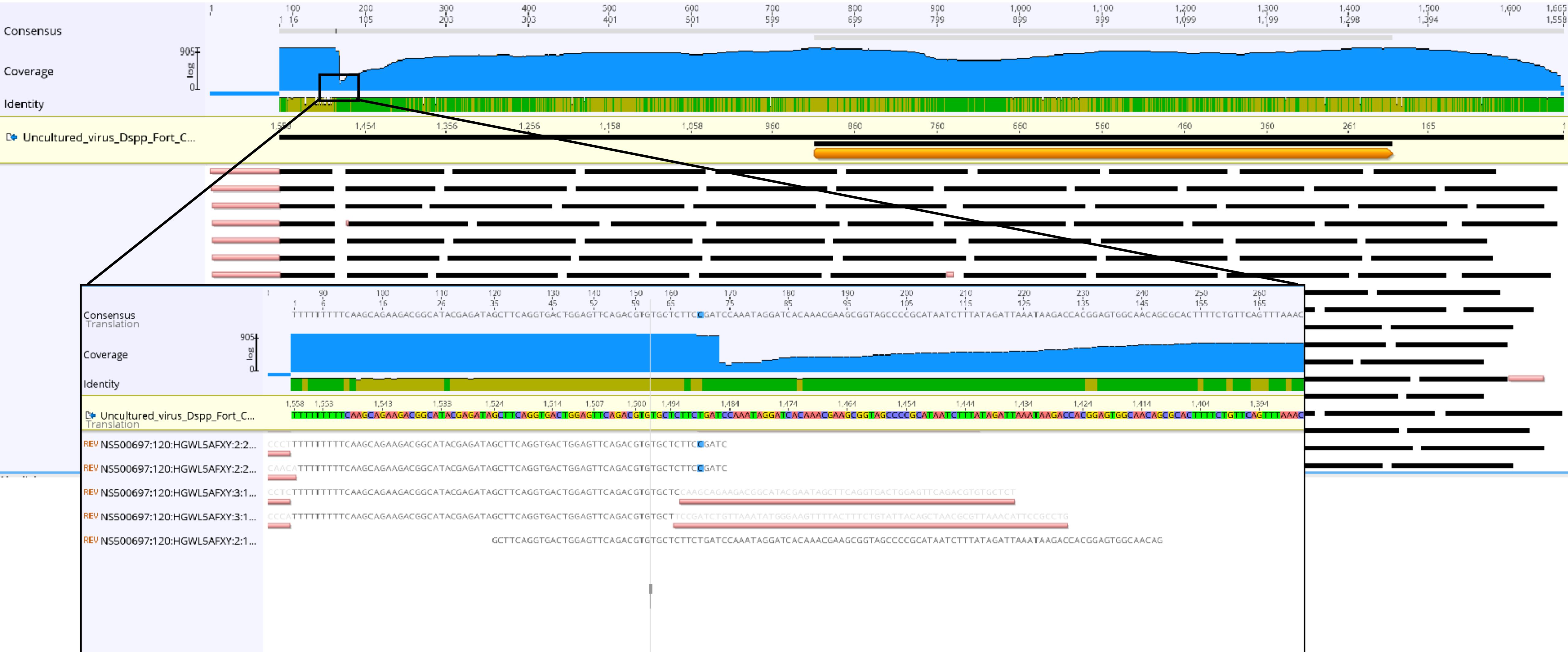
Conclusion: don't blindly trust database annotation nor the output of analysis software

COMMENT  
GenBank Accession Numbers KU746283, KU746284 represent sequences from the 2 segments of Guanarito mammarenavirus CVH-960201.

##Genome-Assembly-Data-START##  
Assembly Method :: Trimmomatic v. 0.32  
SGA v. 0.10.13  
iMetAMUS v. 1.5  
samtools v1.1  
FastQC v. 0.10.0  
Spades v. 3.1.1  
idba v1.1.1  
Pilon v. 1.8  
Quast v. 2.2  
Prokka v. 1.7  
Assembly Name :: GTDV014-SEQ-1-ASM-1  
Genome Coverage :: 6779.96x  
Sequencing Technology :: Illumina HiSeq1500  
##Genome-Assembly-Data-END##.

FEATURES  
source  
Location/Qualifiers  
1..8109  
/organism="Guanarito mammarenavirus"  
/mol\_type="genomic RNA"

# A single (PCR chimera?) read triggered a similar missassembly



Another caveat: using smaller databases (e.g. all viral genomes in RefSeq) is faster but it can produce misleading results

Here: a read was BLASTed against all of the virus nucleotide sequences in Genbank

>a\_sequence  
ATGCAGATCTTCGTGAAGACTCTGACTGGTAAGACCATCACCCCTCGAGGTTGAGCC...

The screenshot shows a BLAST search results window titled "Descriptions". The search query sequence is a partial viral genome. The results table lists 14 significant alignments with bovine viral diarrhea virus genes. The columns in the table are: Description, Max score, Total score, Query cover, E value, Ident, and Accession. All alignments have a Max score of 325, a Total score of 383, a Query cover of 100%, an E value of 5e-87, an Ident of 92%, and an Accession number. The descriptions list various subgenomes and genes of the bovine viral diarrhea virus.

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus T-20 gene for poryprotein, partai cds, strain: T-20</a>	325	383	100%	5e-87	92%	<a href="#">AB111967.1</a>
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus 190cp gene for poryprotein, partai cds, strain: 190cp</a>	325	379	100%	5e-87	92%	<a href="#">AB111966.1</a>
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus-2 subgenome D4 polyprotein mRNA, partial cds</a>	325	536	100%	5e-87	92%	<a href="#">AF104029.1</a>
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus-2 subgenome D1 polyprotein mRNA, partial cds</a>	325	404	100%	5e-87	92%	<a href="#">AF104026.1</a>
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus-2 subgenome C5 polyprotein mRNA, partial cds</a>	325	651	100%	5e-87	92%	<a href="#">AF104025.1</a>
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus-2 subgenome C4 polyprotein mRNA, partial cds</a>	325	518	100%	5e-87	92%	<a href="#">AF104024.1</a>
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus-2 subgenome C3 polyprotein mRNA, partial cds</a>	325	325	100%	5e-87	92%	<a href="#">AF104023.1</a>
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus-2 subgenome C2 polyprotein mRNA, partial cds</a>	325	503	100%	5e-87	92%	<a href="#">AF104022.1</a>
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus-2 subgenome C1 polyprotein mRNA, partial cds</a>	325	408	100%	5e-87	92%	<a href="#">AF104021.1</a>
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus-2 subgenome B polyprotein mRNA, partial cds</a>	325	699	100%	5e-87	92%	<a href="#">AF104020.1</a>
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus-2 subgenome A polyprotein mRNA, partial cds</a>	325	408	100%	5e-87	92%	<a href="#">AF104019.1</a>
<input type="checkbox"/>	<a href="#">Bovine viral diarrhea virus p125 protein gene, partial cds</a>	325	710	100%	5e-87	92%	<a href="#">L13783.1</a>

Cool, looks like a flavivirus! Right?

# Keep analyses as unbiased as possible

The same read was BLASTed against all the nucleotide sequences in Genbank (the 'nt' database):

```
>a_sequence
ATGCAGATCTCGTGAAGACTCTGACTGGTAAGACCATCACCCCTCGAGGTTGAGCC...
```

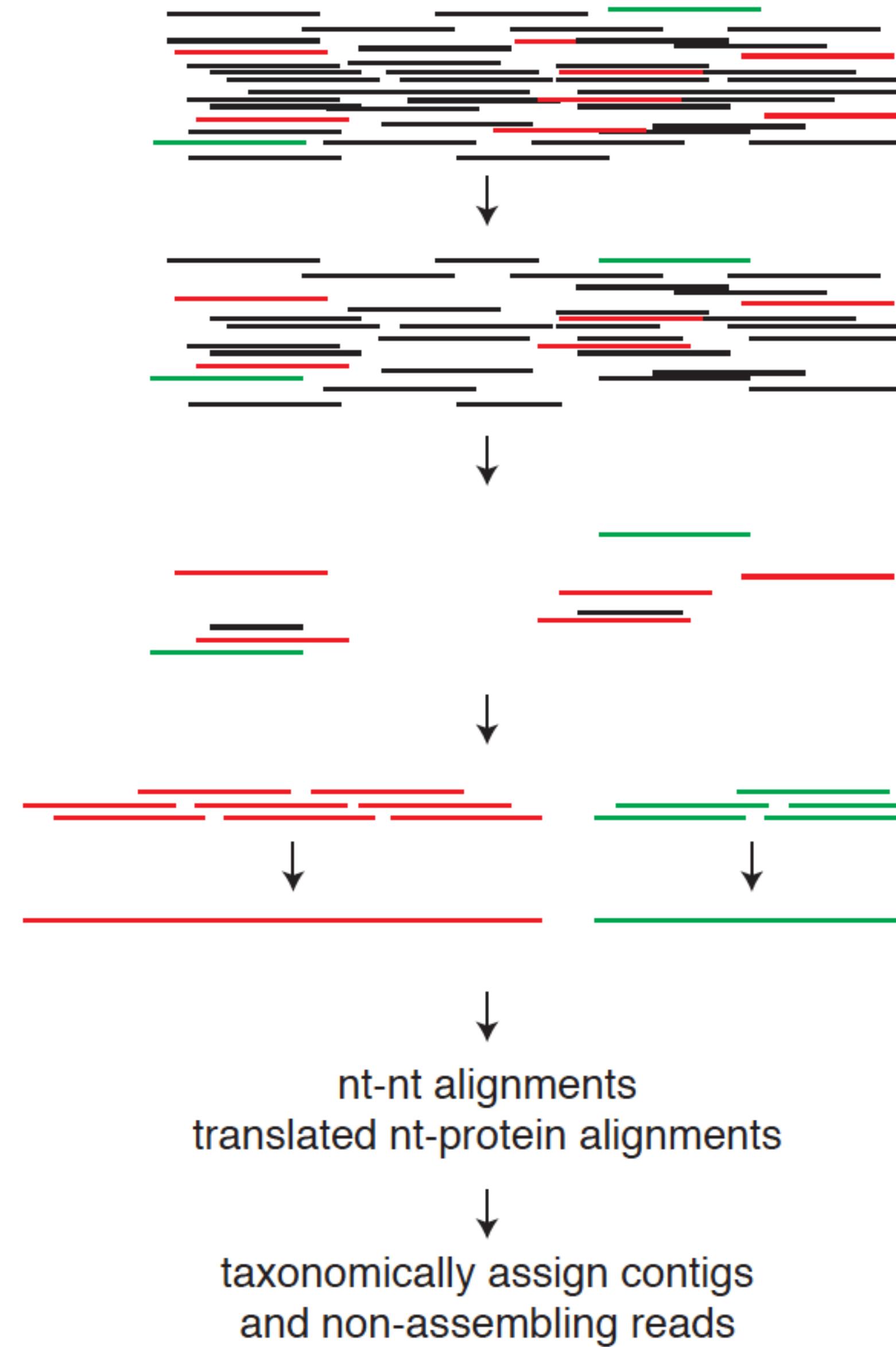
Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Homo sapiens ubiquitin C (UBC), RefSeqGene on chromosome 12</a>	412	3261	100%	2e-111	100%	<a href="#">NG_027722.2</a>
<input type="checkbox"/>	<a href="#">Homo sapiens ubiquitin C (UBC), mRNA</a>	412	3261	100%	2e-111	100%	<a href="#">NM_021009.6</a>
<input type="checkbox"/>	<a href="#">Homo sapiens UbC gene for ubiquitin C, complete cds, clone: MKJ9</a>	412	3241	100%	2e-111	100%	<a href="#">AB643790.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens UbC gene for ubiquitin C, complete cds, clone: DJL8</a>	412	2881	100%	2e-111	100%	<a href="#">AB643789.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens UbC gene for ubiquitin C, complete cds, clone: DJL9</a>	412	3266	100%	2e-111	100%	<a href="#">AB643788.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens UbC gene for ubiquitin C, complete cds, clone: MKS7</a>	412	2558	100%	2e-111	100%	<a href="#">AB643787.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens UbC gene for ubiquitin C, complete cds, clone: MKS9</a>	412	3257	100%	2e-111	100%	<a href="#">AB643786.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens UbC gene for ubiquitin C, complete cds, clone: BHP7</a>	412	2549	100%	2e-111	100%	<a href="#">AB643785.1</a>
<input type="checkbox"/>	<a href="#">Pan troglodytes mRNA for ubiquitin, complete cds, clone: PtsC-51-5_D12</a>	412	1833	100%	2e-111	100%	<a href="#">AK306071.1</a>

Some BVDV genomes contain ubiquitin homologs

# A typical pathogen discovery analysis workflow



Quality filter /  
remove PCR duplicates  
cutadapt  
cd-hit-est

Remove host sequences  
bowtie2

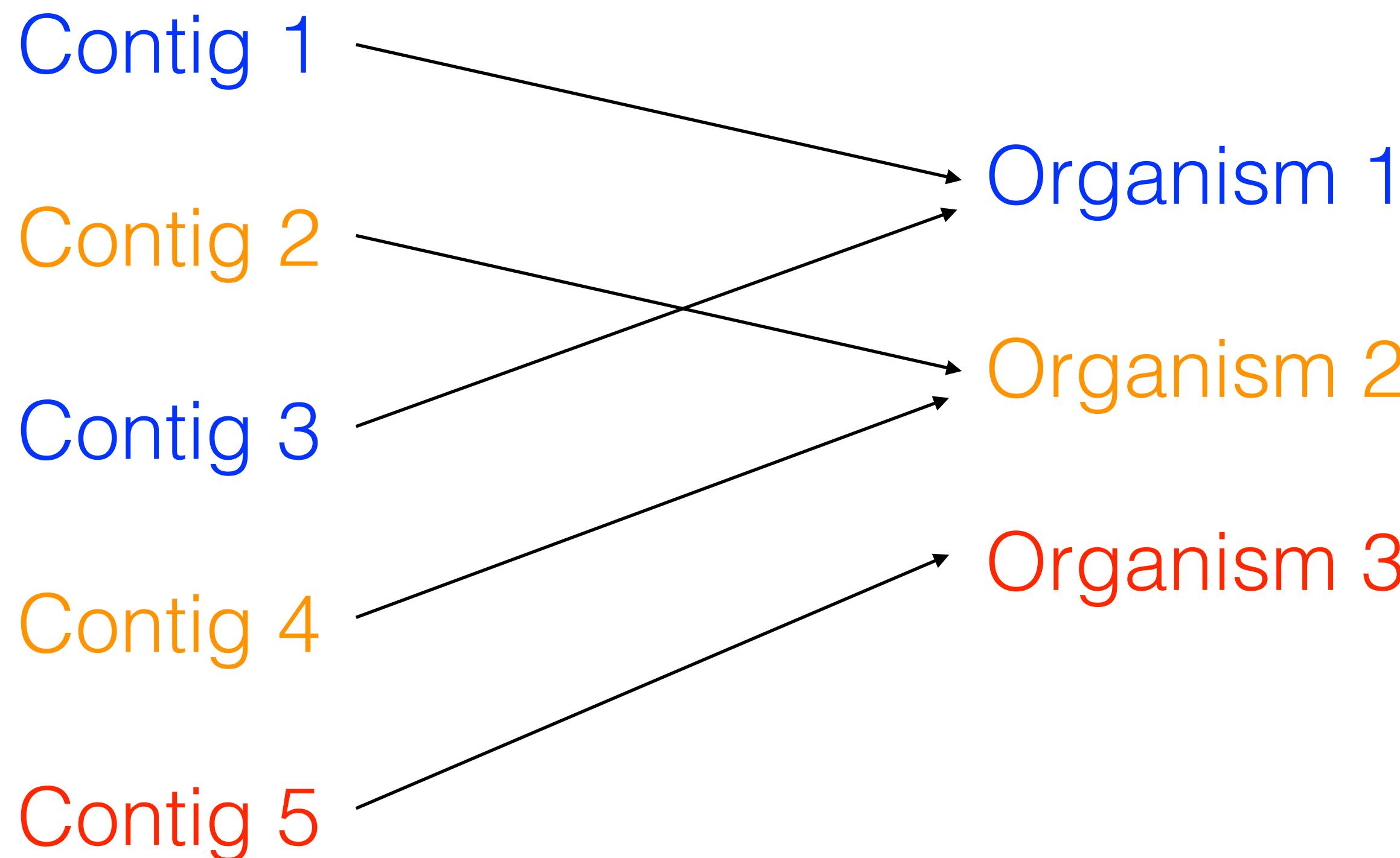
Assemble remaining reads into  
contigs  
SPAdes

Identify most similar sequences  
in NCBI databases  
BLASTN (gsnap),  
BLASTX (diamond)

~40 min for a dataset w/ 6M  
reads

caveat: for a dataset where  
host filtering removed almost  
all of the reads

The goal of metagenomic classification software is to map sequence information to taxonomic information.



Earlier, we did this by BLASTing several contigs on the NCBI website. This is not a practical approach for many contigs.

Nucleotide-level similarity identifies closely related organisms (blastn-like (blastn))  
Protein-level similarity discovers ‘new’ organisms (blastx-like (diamond))

# A nice review of metagenomic classifiers

OXFORD

Briefings in Bioinformatics, 2017, 1–15

doi: 10.1093/bib/bbx120  
Paper

## A review of methods and databases for metagenomic classification and assembly

Florian P. Breitwieser, Jennifer Lu and Steven L. Salzberg

Corresponding author: Steven L. Salzberg, Center for Computational Biology, Johns Hopkins University, 1900 E. Monument St., Baltimore, MD, 21205, USA.  
E-mail: salzberg@jhu.edu

Name	References	URL
CaPSID	Borzenin et al., 2012	<a href="https://github.com/capsid/capsid">https://github.com/capsid/capsid</a>
ClueyHu	Van der Auweret et al., 2014	<a href="http://clueyhu.m-greifswald.de/ClueyHu/query/init">http://clueyhu.m-greifswald.de/ClueyHu/query/init</a>
Clinical PathoScope	Byrd et al., 2014	<a href="https://sourceforge.net/p/pathoscope/wiki/Clinical_PathoScope/">https://sourceforge.net/p/pathoscope/wiki/Clinical_PathoScope/</a>
DUDes	Piro et al., 2016	<a href="http://sf.net/p/dudes">http://sf.net/p/dudes</a>
EnsembleAssembler	Deng et al., 2015	<a href="https://github.com/xutaodang/EnsembleAssembler">https://github.com/xutaodang/EnsembleAssembler</a>
Exhaustive Iterative Assembly (Virus Discovery Pipeline)	Schürch et al., 2014	–
FACSS	Strunzheim et al., 2010	<a href="https://github.com/SciLifeLab/facs">https://github.com/SciLifeLab/facs</a>
GenSeed-HMM	Atwee et al., 2018	<a href="https://sourceforge.net/projects/genseedhmm/">https://sourceforge.net/projects/genseedhmm/</a>
Giant Virus Finder	Kerepesi and Grolmuz, 2016	<a href="http://cgitgroup.org/giant-virus-finder">http://cgitgroup.org/giant-virus-finder</a>
GOTCHA	Freitas et al., 2015	<a href="https://github.com/LANL-Bioinformatics/GOTCHA">https://github.com/LANL-Bioinformatics/GOTCHA</a>
IMSA	Dimon et al., 2013	<a href="https://sourceforge.net/projects/aron-imsa/?source=directory">https://sourceforge.net/projects/aron-imsa/?source=directory</a>
IMSA-A	Cox et al., 2017	<a href="https://github.com/JeremyCoxBML/IMSA-A">https://github.com/JeremyCoxBML/IMSA-A</a>
Kraken	Wood and Salzberg, 2014	<a href="https://github.com/DerrickWood/kraken">https://github.com/DerrickWood/kraken</a>
LMAT	Annes et al., 2013	<a href="https://sourceforge.net/projects/lmat/">https://sourceforge.net/projects/lmat/</a>
MEGAN 4	Huson et al., 2011	<a href="http://ab.inf.uni-tuebingen.de/software/megan4/">http://ab.inf.uni-tuebingen.de/software/megan4/</a>
MEGAN Community Edition	Huson et al., 2016	<a href="http://ab.inf.uni-tuebingen.de/data/software/megan6/download/welcomen.html">http://ab.inf.uni-tuebingen.de/data/software/megan6/download/welcomen.html</a>
MepIC	Takayuki et al., 2014	<a href="https://mepic.nih.go.jp/">https://mepic.nih.go.jp/</a>
MetaShot	Fosso et al., 2017	<a href="https://github.com/bfossou/MetaShot">https://github.com/bfossou/MetaShot</a>
meteMIC	Modha, 2016	<a href="https://github.com/sejmodha/meteMIC">https://github.com/sejmodha/meteMIC</a>
Metavir	Roux et al., 2011	<a href="http://metavir-meb.univ-bpclermont.fr/">http://metavir-meb.univ-bpclermont.fr/</a>
Metavir 2	Roux et al., 2014	<a href="http://metavir-meb.univ-bpclermont.fr/">http://metavir-meb.univ-bpclermont.fr/</a>
MettLab	Nording et al., 2016	<a href="https://github.com/noring/metlab">https://github.com/noring/metlab</a>
NBC	Roux et al., 2011	<a href="https://nbc.ee.ucl.ac.be/">https://nbc.ee.ucl.ac.be/</a>
PathSeq	Kostic et al., 2011	<a href="https://www.broadinstitute.org/software/pathseq/">https://www.broadinstitute.org/software/pathseq/</a>
ProVIDE	Ghosh et al., 2011	<a href="http://metagenomics.cs.uct.ac.za/clinical/Provide/">http://metagenomics.cs.uct.ac.za/clinical/Provide/</a>
QuasQ	Poh et al., 2013	<a href="http://www.statgenexus.edu.sg/~seim\$software/quasq.html">http://www.statgenexus.edu.sg/~seim\$software/quasq.html</a>
READSCAN	Naeem et al., 2013	<a href="http://cbrc.kaust.edu.sa/readscan/">http://cbrc.kaust.edu.sa/readscan/</a>
Rega Typing Tool	Kroneman et al., 2011; Pineda-Peña et al., 2013	<a href="http://egatools.med.kuleuven.be/typing/v3/hiv/typingtool/">http://egatools.med.kuleuven.be/typing/v3/hiv/typingtool/</a>
REMS	Scheuchl et al., 2015	<a href="https://www.flf.de/fileadmin/FLI/IVD/Microarray-Diagnostics/REMS.tar.gz">https://www.flf.de/fileadmin/FLI/IVD/Microarray-Diagnostics/REMS.tar.gz</a>
RINS	Bhaduri et al., 2012	<a href="http://khaverlab.stanford.edu/tools-1/#tools">http://khaverlab.stanford.edu/tools-1/#tools</a>
SLIM	Cotten et al., 2014	*Available upon request*
SMART	Lee et al., 2016	<a href="https://bitbucket.org/ayl/smart">https://bitbucket.org/ayl/smart</a>
SRAA	Iakov et al., 2011	*Available upon request*
SURPI	Neuenschwander et al., 2014	<a href="https://github.com/chitubio/surpi">https://github.com/chitubio/surpi</a>
Taxonomer	Flygare et al., 2016	<a href="https://www.taxonomer.com/">https://www.taxonomer.com/</a>
Taxy-Pro	Klingenberg et al., 2013	<a href="http://gobics.de/TaxyPro/">http://gobics.de/TaxyPro/</a>
"Unknown pathogens from mixed clinical samples"	Gong et al., 2016	–
vFam	Skrwes-Cox et al., 2014	<a href="https://deriskubuscsf.edu/software/vFam/">https://deriskubuscsf.edu/software/vFam/</a>
VIP	Li et al., 2016	<a href="https://github.com/keylabvdo/VIP">https://github.com/keylabvdo/VIP</a>
ViralFusionSeq	Li et al., 2013	<a href="https://sourceforge.net/projects/viralfusionseq/">https://sourceforge.net/projects/viralfusionseq/</a>
Virana	Schelhorn et al., 2013	<a href="https://github.com/eichehcn/Virana">https://github.com/eichehcn/Virana</a>
ViFind	Ho and Tzandilis, 2014	<a href="https://vifind.org/">https://vifind.org/</a>
VIROME	Wommack et al., 2012	<a href="http://virome.dbi.udel.edu/app/#view=home">http://virome.dbi.udel.edu/app/#view=home</a>
ViromeScan	Rampelli et al., 2016	<a href="https://sourceforge.net/projects/viromescan/">https://sourceforge.net/projects/viromescan/</a>
VirGotor	Roux et al., 2015	<a href="https://github.com/simroux/VirGotor">https://github.com/simroux/VirGotor</a>
VirusFinder	Wang et al., 2013	<a href="http://bioinfo.mc.vanderbilt.edu/VirusFinder/">http://bioinfo.mc.vanderbilt.edu/VirusFinder/</a>
VirusHunter	Zhao et al., 2013	<a href="https://www.ibridgenetwork.org/IVD/profiles/905559575893/innovations/103/">https://www.ibridgenetwork.org/IVD/profiles/905559575893/innovations/103/</a>
VirusSeeker	Zhao et al., 2017	<a href="https://wupell.labs.wustl.edu/VirusSeeker/">https://wupell.labs.wustl.edu/VirusSeeker/</a>
VirusSeq	Chen et al., 2013	<a href="http://odin.mdc-berlin.mpg.de/blast/blast/blast/blast/VirusSeq.html">http://odin.mdc-berlin.mpg.de/blast/blast/blast/VirusSeq.html</a>
VirVerSeq	Verblat et al., 2015	<a href="https://sourceforge.net/projects/virverseq/?source=directory">https://sourceforge.net/projects/virverseq/?source=directory</a>
VMGAP	Lorenzi et al., 2011	–

–, No website could be found, the workflow was unavailable.

# Metagenomic classification can be challenging

## Resource intensive

Large databases

Large assemblies

Memory and storage intensive

### Bioinformatics challenges

User-friendly bioinformatics software for analysis of mNGS data is not currently available. Thus, customized bioinformatics pipelines for analysis of clinical mNGS data<sup>56,109–111</sup> still require highly trained programming staff to develop, validate and maintain the pipeline for clinical use. The laboratory can either host computational servers locally or move the bioinformatics analysis and data storage to cloud platforms. In either case, hardware and software setups can be complex, and adequate measures

## Clinical metagenomics

Charles Y. Chiu<sup>1,2\*</sup> and Steven A. Miller<sup>1</sup>

# idseq.net is a new web-based tool that does metagenomic classification

The screenshot shows the homepage of [idseq.net](https://idseq.net). At the top, there's a navigation bar with links to various bioinformatics tools like GenBank, blastn, blastx, etc. Below the navigation is the IDseq logo and a "Join our team" button.

**IDseq is an unbiased global software platform that helps scientists identify pathogens in metagenomic sequencing data.**

**Discover**  
Identify the pathogen landscape

**Detect**  
Monitor and review potential outbreaks

**Decipher**  
Find potential infecting organisms in large datasets

**Learn more about IDseq**  
Already have an account? [Sign in](#).

First Name  Last Name   
Email   
Affiliated Institution or Company   
How would you use IDseq? Optional

**Submit**



The snake sample we analyzed run through idseq

Metagenomic Pipeline v3.7, NT/NR: 2018-12-01 | processed 2 years ago

Snake\_7 > SRR1984309

[Sample Details](#)

[Metagenomic](#) [Antimicrobial Resistance](#)

Taxon name  [Name Type: Scientific](#) [Background: NID Human CSF v3](#) [Categories](#) [Threshold Filters](#) [Read Specificity: All](#)

6153 total rows. [grid](#) [list](#)

> Taxon	Score	Z Score	rPM	r	contig	contig r	%id	L	E value	NT	NR
> Reptarenavirus (11 viral species)	1,042,963,754	99.0 99.0	57,382.5 50,860.0	16,601 14,714	14 20	12,126 12,215	98.8 98.2	1,195.7 386.3	10 <sup>-233</sup> 10 <sup>-266</sup>		
> Clostridium (35 bacterial species) ●● 2	115,659,009	76.3 99.0	916.0 11,921.7	265 3,449	0 1	0 3,412	99.1 36.1	26.1 58.8	10 <sup>-6</sup> 10 <sup>-4</sup>		
> Pogonomyrmex (2 eukaryotic species)	39,646,877	100.0 0.0	3,971.6 0.0	1,149 0	0 0	0 0	100.0 0.0	22.0 0.0	10 <sup>-3</sup> 0		
> Solanum (7 eukaryotic species)	28,728,424	99.0 -100.0	4,704.4 0.0	1,361 0	0 0	0 0	99.0 0.0	22.9 0.0	10 <sup>-4</sup> 0		
> Caenorhabditis (6 eukaryotic species)	27,474,943	99.0 -100.0	3,563.7 0.0	1,031 0	0 0	0 0	96.9 0.0	24.3 0.0	10 <sup>-4</sup> 0		

# Our lab's pipeline is available on GitHub if you want to see how we do it.

stnglein-lab / taxonomy\_pipeline

Code Issues 0 Pull requests 0 Projects 0 Wiki Security

```
61 # ****
62 # first, create contigs using spades
63 # ****
64
65 echo "run spades for $file_base"
66 date
67
68 echo "spades.py -o ${file_base}.spades --pe1-1 $f1 --pe1-2 $f2 -t 24 -m 150"
69 spades.py -o ${file_base}.spades --pe1-1 $f1 --pe1-2 $f2 -t 24 -m 150
70
71 echo "done running spades for $file_base"
72 date
73
```

1 branch 1 release 6M

Initial commit

README.md readme

contig\_based\_taxonomic\_assessm... diamond

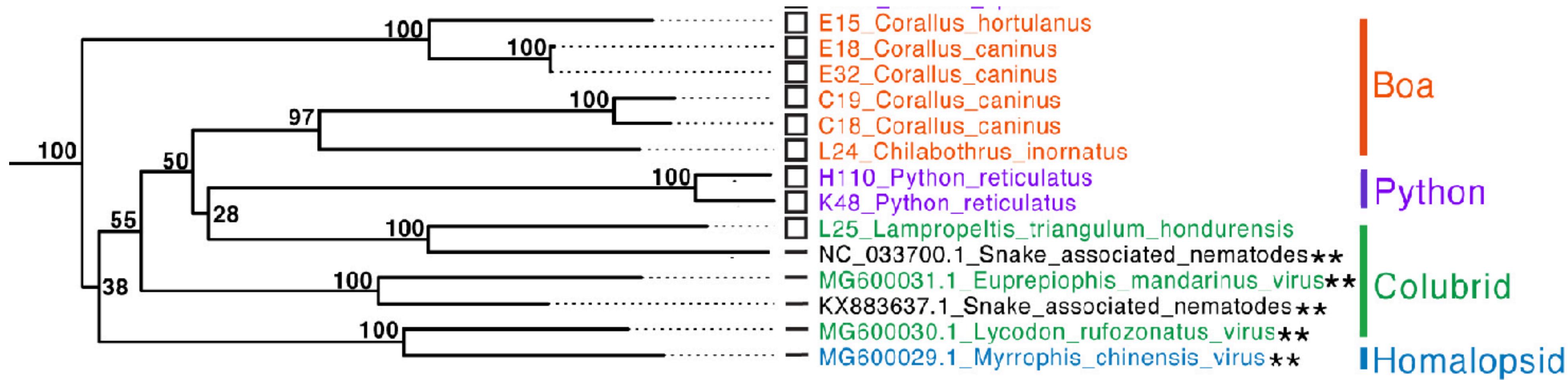
(diamond)

- This pipeline is not easily portable (new nextflow/conda version more portable)
- But at least you can see how we do it, which may be helpful.
- Using 'pipelines' facilitates reproducibility and throughput

[https://github.com/stnglein-lab/taxonomy\\_pipeline](https://github.com/stnglein-lab/taxonomy_pipeline)

# Metagenomic sequencing only gives you sequences

# Serpentoviruses detected in snakes with respiratory disease and also snake-associated nematodes



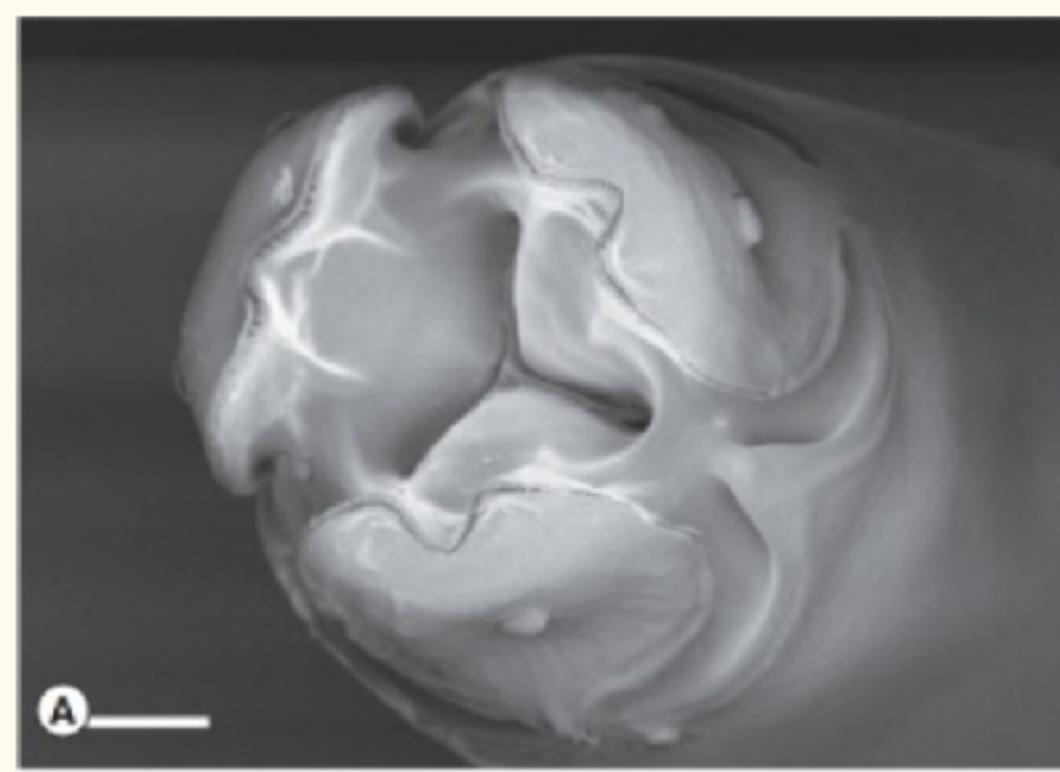
Are these related viruses infecting both nematodes and snakes?

I'd bet that the 'nematode' viruses really infect snakes

Laura Hoon-Hanks

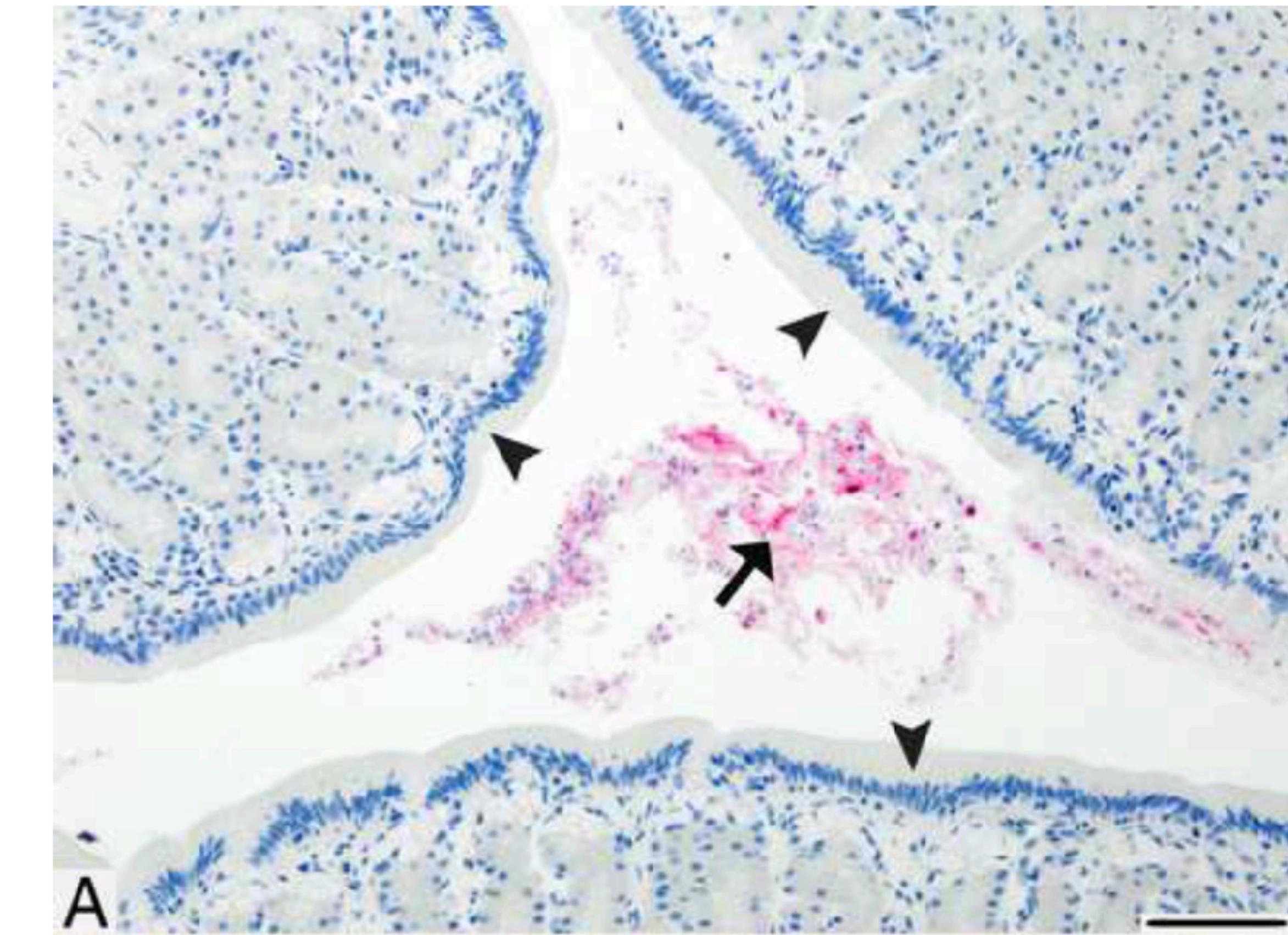


SEM of a snake nematode



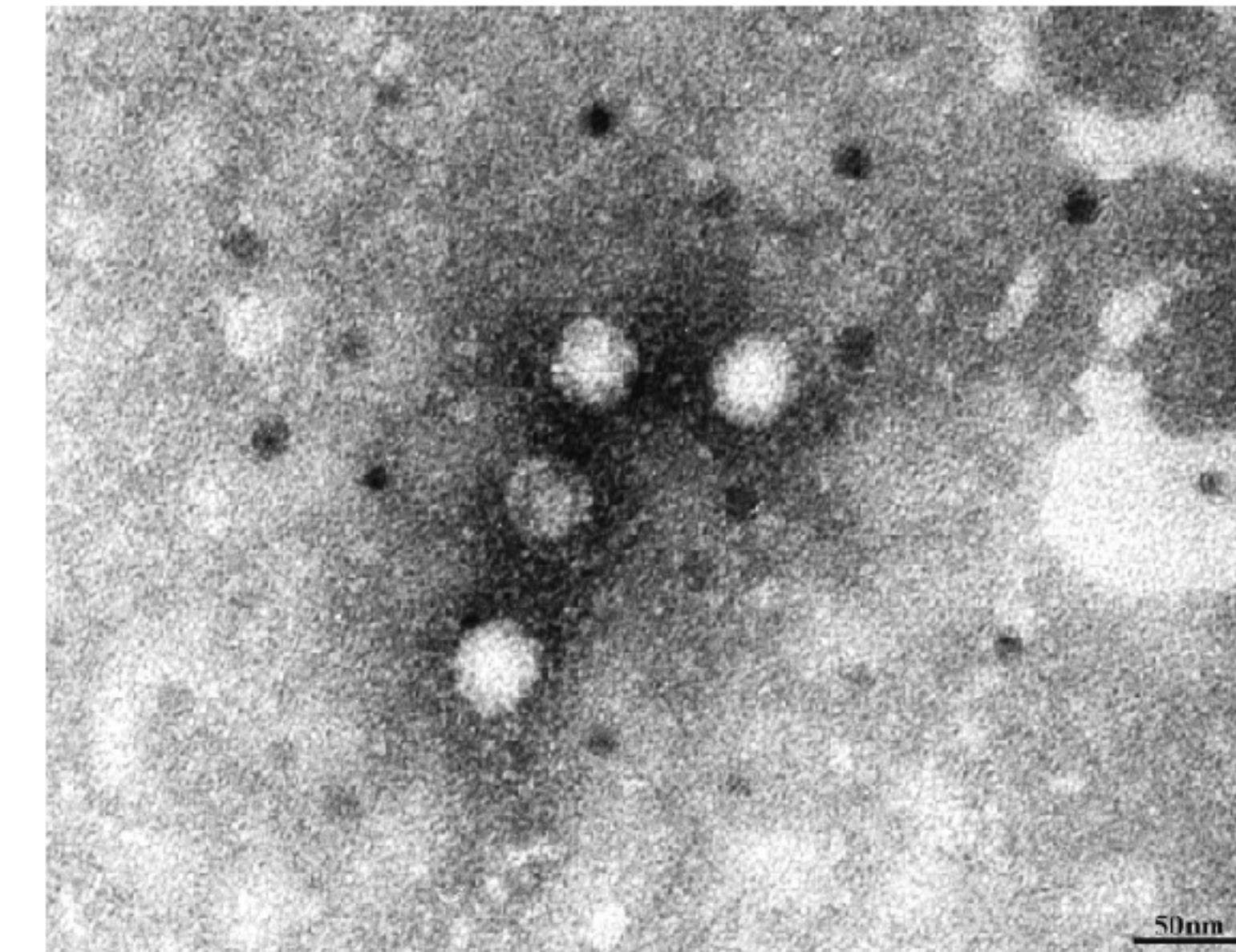
Choe et al (2016)

Serpentovirus antigen detected in python intestinal lumen



# Astroviruses associated with fatal gastroenteritis in rabbits: likely the cause but need additional proof

A rabbit facility in TN  
experienced an outbreak  
of fatal gastroenteritis

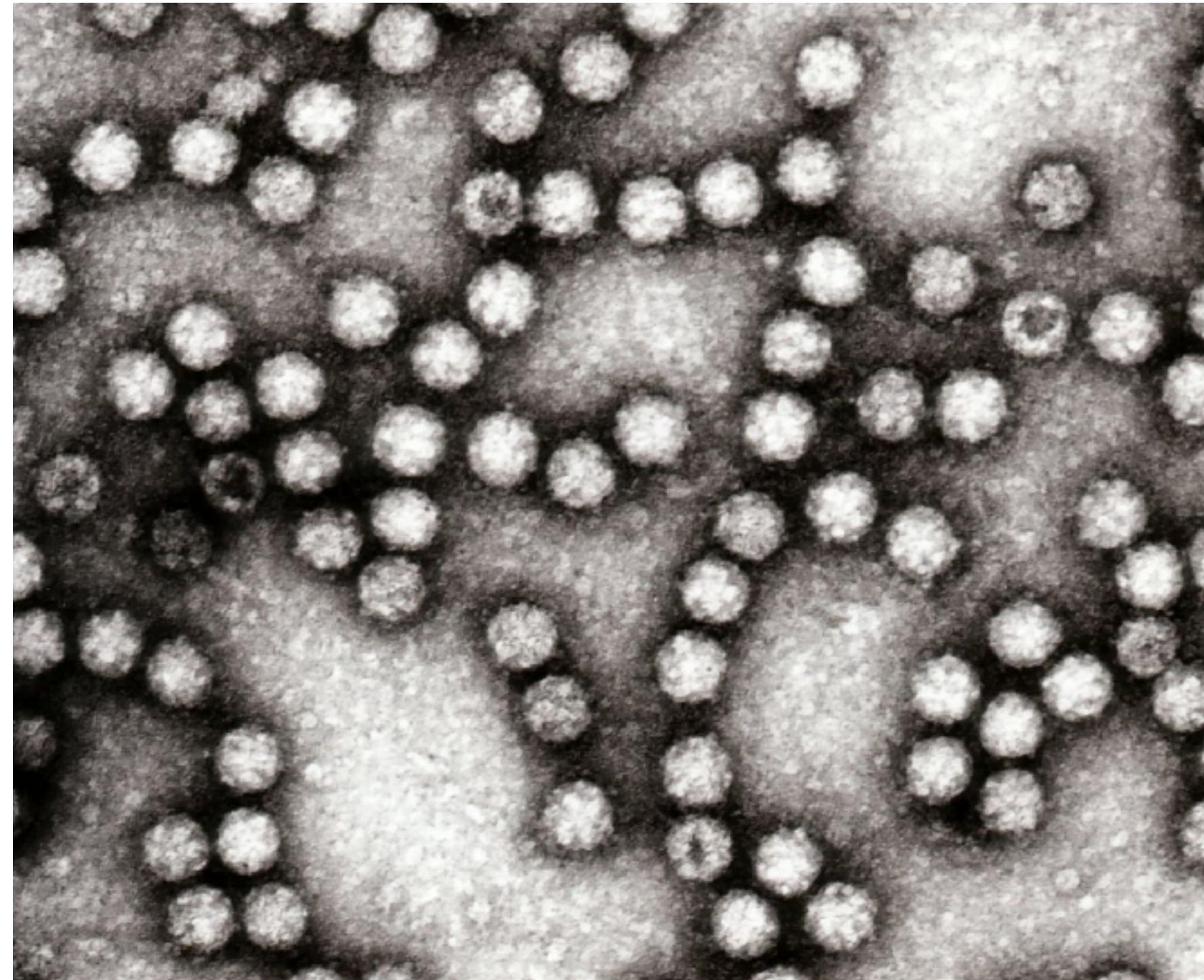


**Figure 1** Electron micrograph of virus like particles in the stool of one animal (Table 1). Scale bar indicates 50 nm.

astrovirus sequences in the stool  
samples from sick rabbits

(Meta)genomics is useful for hypothesis generation but experiments must be done

## Astrovirus particles



JOURNAL OF CLINICAL MICROBIOLOGY, Apr. 1993, p. 955-962  
0095-1137/93/040955-08\$02.00/0  
Copyright © 1993, American Society for Microbiology

Vol. 31, No. 4

### Characterization and Seroepidemiology of a Type 5 Astrovirus Associated with an Outbreak of Gastroenteritis in Marin County, California

KAREN MIDTHUN,<sup>1†\*</sup> HARRY B. GREENBERG,<sup>1‡</sup> JOHN B. KURTZ,<sup>2</sup> G. WILLIAM GARY,<sup>3</sup>  
FENG-YING C. LIN,<sup>4</sup> AND ALBERT Z. KAPIKIAN<sup>1</sup>

### RESULTS

**Volunteer study.** Nineteen adult volunteers were orally administered a filtrate prepared from a 0.1% suspension of stool from one of the ill individuals in the original Marin County outbreak. None of 17 volunteers who received a 1-ml inoculum became ill. Because of this, the amount of inoculum was increased to 20 ml. Of two volunteers who received the larger inoculum, one developed a gastrointestinal illness characterized by nausea, vomiting, diarrhea, and malaise.