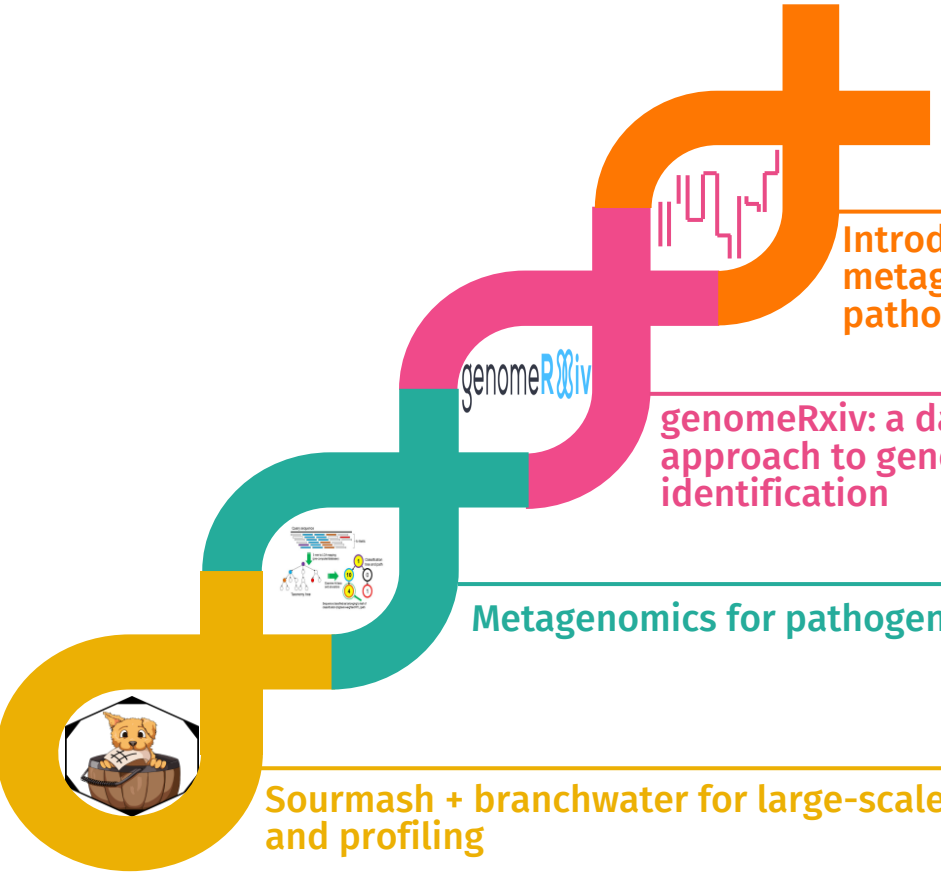
Introduction to metagenome-based pathogen identification

Sahar Abdelrazek, M.Sc., Ph.D.
Department of Biomedical Sciences and Pathobiology
VA-MD College of Veterinary Medicine
Virginia Tech

Genome/metagenome-based pathogen identification



1 Introduction to metagenome-based pathogen identification

Sahar Abdelrazek, Ph.D.
VA-MD College of Veterinary Medicine, Virginia Tech

2 genomeRxiv: a data driven approach to genome taxonomic identification

Mazloom, Reza
Department of Computer Science, Virginia Tech, VA

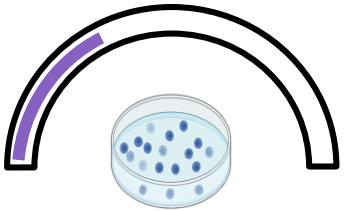
3 Metagenomics for pathogen detection

Parul Sharma, Ph.D.
School of Medicine, Division of infectious diseases
Emory University

4 Sourmash + branchwater for large-scale sequence search and profiling

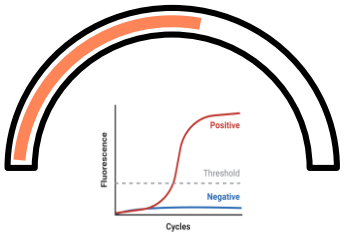
Tessa Ward, Ph.D.
Department of Population Health and Reproduction
University of California, Davis

Methods for Pathogen Identification



Culture-based Methods

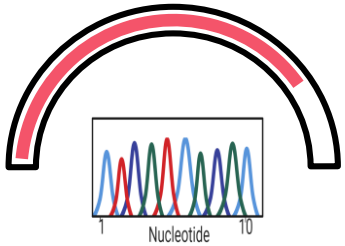
Time-consuming
some bacteria are difficult to culture



qpcr

Molecular Methods

Only pathogens with known sequence data can be detected

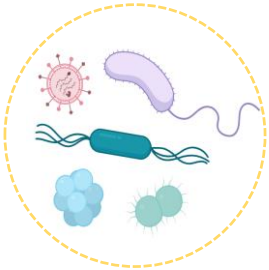


Sequencing Methods

Can detect newly emerging pathogens as well as non-culturable pathogens

Metagenome-based pathogen identification

Metagenomics is the sequencing of a whole community of microorganism DNA as opposed to sequencing just the individual microbes.



Reasons for choosing metagenomics

- No need of culturing the pathogen
- Higher resolution
- Can detect pathogens from newly emerging diseases

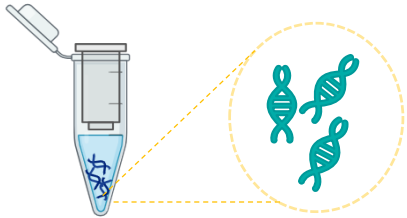
Metagenome-based pathogen identification pipeline

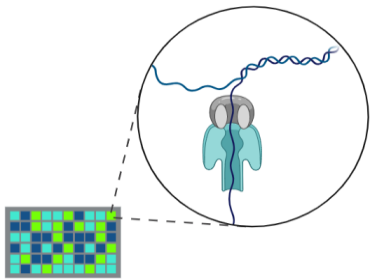
1 Sample Collection

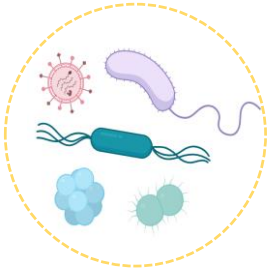
2 DNA extraction

3 Sequencing & Data analysis

4 Pathogen identification







1 Sample Collection

Planning & Preparation

Define Objectives: Study purpose (e.g., disease diagnosis, pathogen diversity, epidemiology).

Select Plant and Tissue Type: Choose specific tissues (e.g., leaves, roots, stems) based on where the pathogen is likely to be present.

Sample Collection

Timing: Collect samples at the right stage of disease development.

Tools and Materials: Sterile gloves, scissors or scalpels and sterile bags or tubes.

Storage &Transport

Transport: Ensure samples remain cool during transport to the lab (ice packs or a portable refrigerator if possible).

Immediate Storage: (-80)

2 DNA extraction

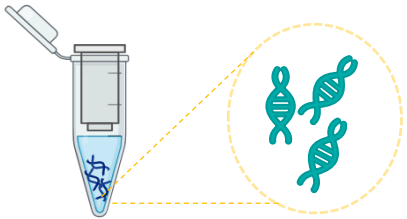
community.nanoporetech.com/docs/prepare/extraction_protocols

- 25 sample types
- 72 protocols
- DNA and RNA
- Sample handling & storage advice

Sample Preparation

Extraction Kit

Quality & Quantity Assessment



Role # 1: You can't sequence what you don't have?

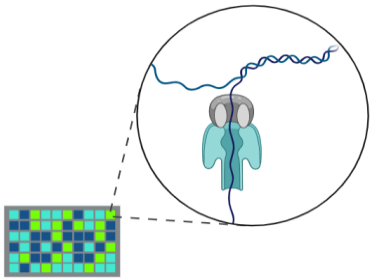
Sample Preparation: In the lab, further clean the samples if needed. Use sterile conditions to prevent contamination.

Extraction Kit: Use a DNA extraction kit suitable for your sample. (modifications to handle each type of sample).

Purity Check: Use Nanodrop to check the DNA purity (A260/A280 ratio should be ~1.8).

Quantification: Use Qubit for accurate DNA quantification.

3 Library preparation



Long Reads

Portability

Real-Time Data

Versatility

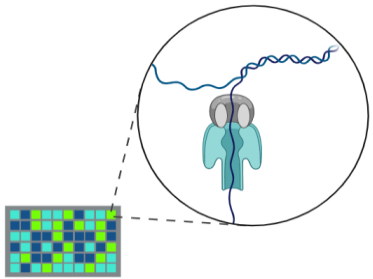
https://nanoporetech.com/products/prepare/dna-library-preparation

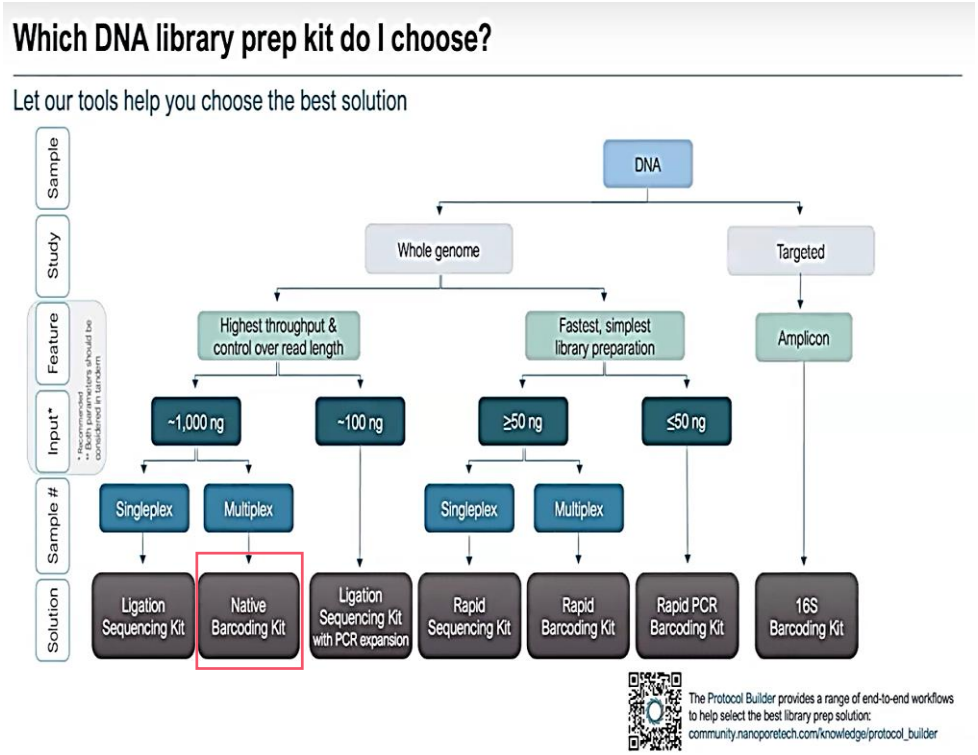
Role # 2: Trade-off between sequencing output and read length

	Any fragment length libraries 20 bases to 4 Mb+	Rapid prep in just 10 minutes	PCR-free protocols to elimate bias
	Output optimised	**Speed optimised**	**Ultra-long reads optimised**
	Ligation Sequencing Kit	Rapid Sequencing Kit	Ultra-long DNA Sequencing Kit
Preparation time	60 minutes	10 minutes	200* mins + 1x O/N incubation
Input recommendation	~1000 ng gDNA or 100-200 fmol for amplicons	~100 ng gDNA	6M cells
Fragmentation	Optional	Transposase-based	Transposase-based
Amplification	No	No	No
Barcode options	Native Barcoding Kit 24 Native Barcoding Kit 96	Rapid Barcoding Kit 24 Rapid Barcoding Kit 96	-
Typical output	●●●	●●○	●●○
Adaptive sampling	Yes	Yes	Yes

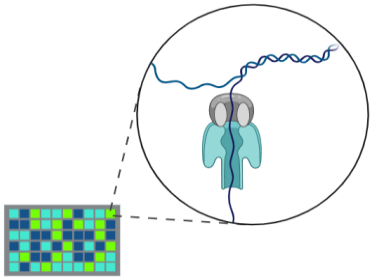
Protocol builder

Provides a range of end to end workflows to help select the best library prep solution





3 Library preparation workflow



Native Barcoding Kit 24 V14

1 High molecular weight gDNA

Optional fragmentation

2 Combined FFPE repair and end-prep

3 3’ Adenine

140 min

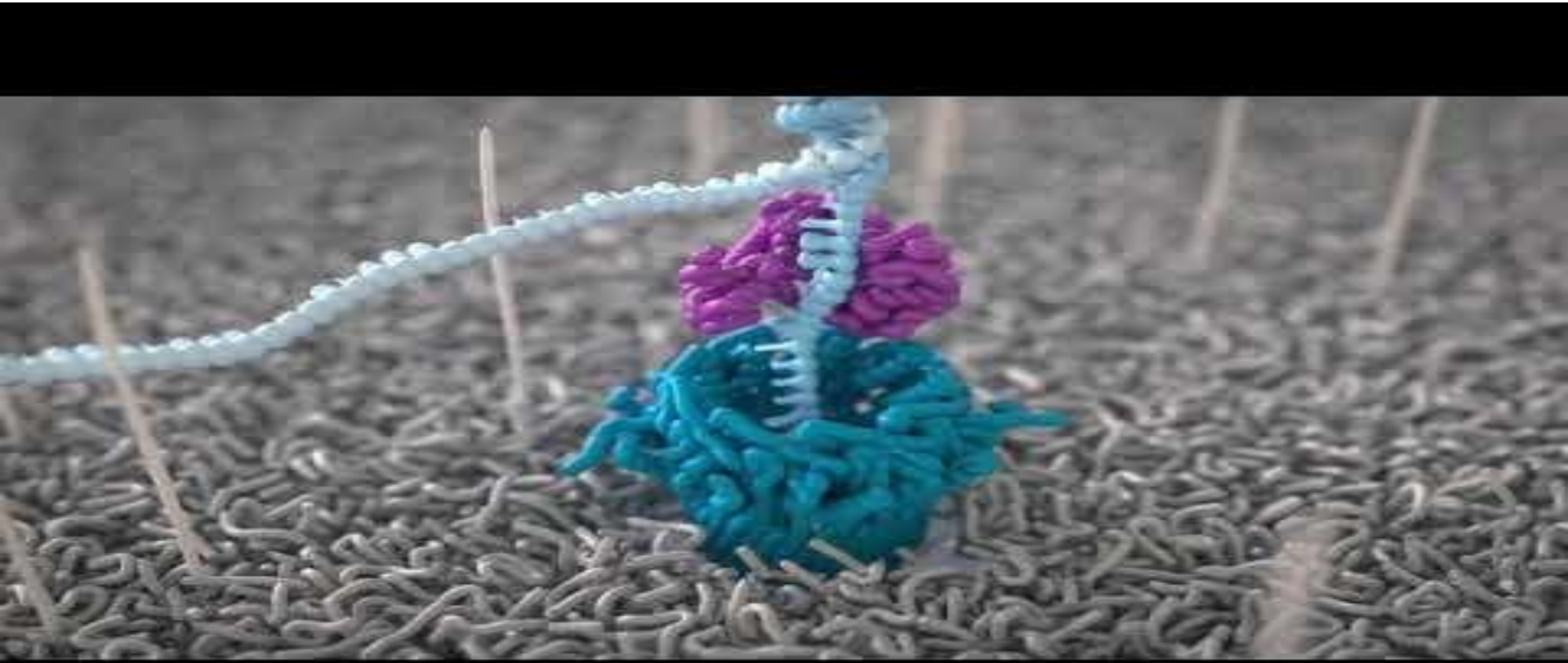
4 Ligation of barcodes — Thymine

5 Ligation of sequencing adapters

Barcoded samples are then pooled together

Priming the flow cell and loading

3 Sequencing



3 Real-time data analysis

EPI2ME

EPI2ME WIMP rapid species identification and quantification from metagenomic samples

3 Example for real-time data analysis using WIMP

Tomato sample with wilt symptoms

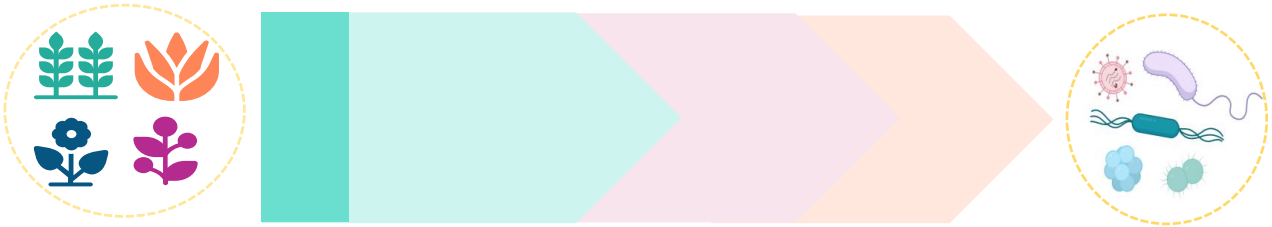
Sample ID	22-263
Collect Date	October-12-2022
Virginia District	Southwestern
County	Buckingham, Virginia
Host Common Name	Tomato
Variety	Red Morning
Host Genus	Solanum
Host Species	lycopersicum
Oxford Nanopore Technologies	PromethION
Samples run per flow cell	20

1 Sample Collection

2 DNA extraction

3 Sequencing & Data analysis

4 Pathogen identification



https://canadagrowsupplies.com/blogs/main/bacterial-wilt-in-tomatoes

3 WIMP Results

4 Pathogen identification

SHOWING TAXA WITH READS

NCBI Taxonomy Tree

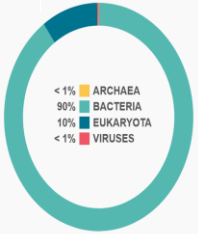
HIDE FILTERS

MINIMUM ABUNDANCE CUTOFF

3% | 1% | 0.5% | 0.1% | 0%

SHOW TOP N TAXA

10 | 20 | 30 | 100 | 200



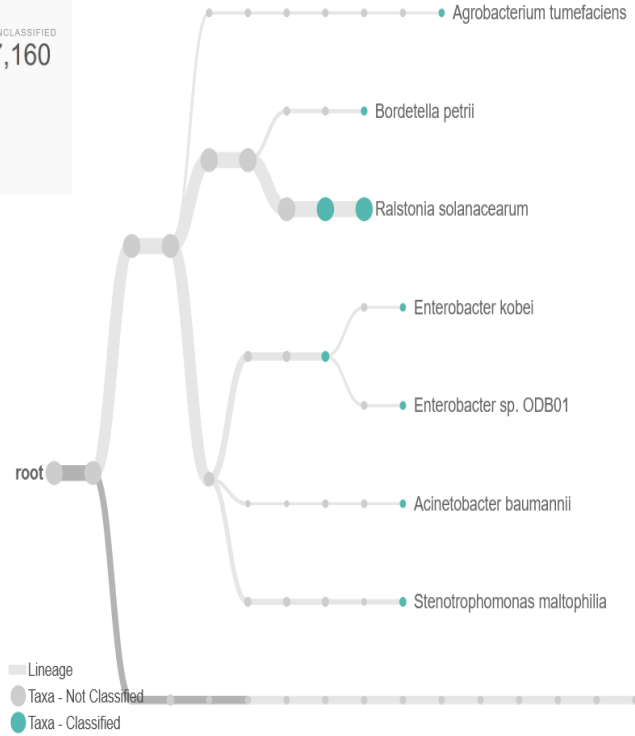
READS ANALYSED 1,088,000

READS CLASSIFIED 520,840

READS UNCLASSIFIED 567,160

EPI2ME

Tomato sample with wilt symptoms



Taxon	Cumulative Reads
Ralstonia solanacearum	288,732
Homo sapiens	48,735
Stenotrophomonas maltophilia	30,867
Bordetella petrii	20,110
Enterobacter sp. ODB01	14,977
Acinetobacter baumannii	11,183
Agrobacterium tumefaciens	9,931
Enterobacter kobei	8,891
Pseudomonas aeruginosa	4,436
Clostridium tetani	3,227
Enterobacter cloacae	3,089

3 | Genome based pathogen identification

Run Kraken2 using PlusPF Kraken2 database

Classified species within our sample were then sorted by z-score, with the species having the highest z-score considered as the pathogen

Extract Ralstonia reads using Kraken

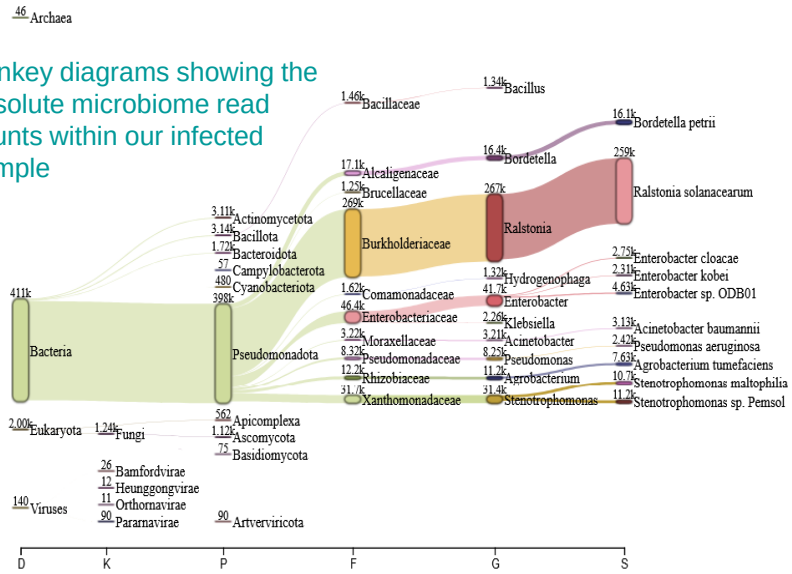
Assemble pathogen genome using Flye

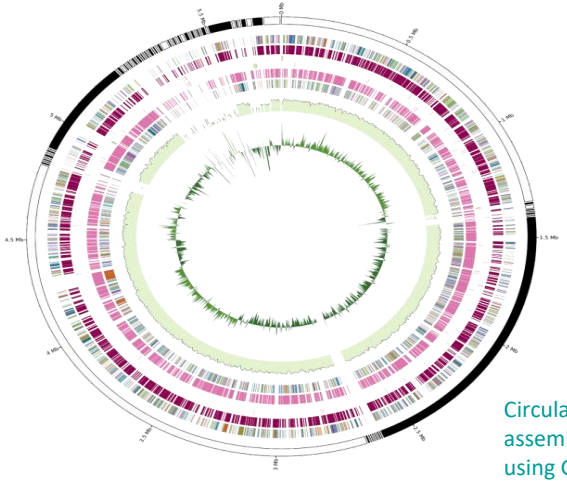
genomeRxiv

Read-based identification still

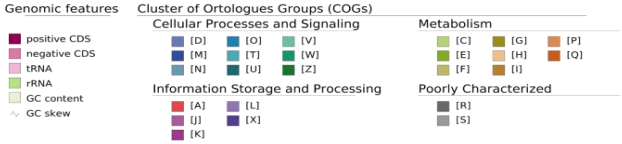
Genome-based identification

Sankey diagrams showing the absolute microbiome read counts within our infected sample





Circular map of *R. solanacearum* assembled draft genomes visualized using GenoVi.



Genome/metagenome-based pathogen identification

GenomeRxiv for deeper genome taxonomic identification

Branchwater for large-scale sequence search and profiling

Core genome tree Phylogenetic analysis

AMRFinderPlus for Virulence and AMR prediction

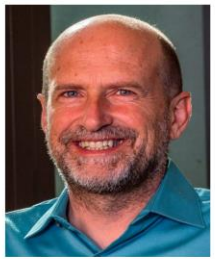
LINtax for read-based strain-level pathogen detection

(Meta)Genomics-Based Pathogen Identification Center for Animal and Plant Disease Diagnostics, Biosecurity, and Pandemic Prevention

Our Team

Kevin Lahmers, DVM, PhD, DACVP–Anatomic Pathology

Clinical Professor of Anatomic Pathology
205 Duck Pond Drive
Blacksburg, VA 24061
klahmers@vt.edu



Boris Vinatzer, PhD

Professor of Plant Pathology
551 Latham Hall
Blacksburg, VA 24061
vinatzer@vt.edu

Lina Rodriguez Salamanca, PhD

Instructor and Plant Disease Clinic Manager and Diagnostician
114 Price Hall
170 Drillfield Drive
Blacksburg, VA 24061
linar@vt.edu

Sahar Abdelrazek, PhD

Research Assistant Professor and
Project Manager
205 Duck Pond Drive
Blacksburg, VA 24061
abdelrazek@vt.edu

Available Services

Comprehensive and Rapid Pathogen Screening

Custom Metagenomic Analysis

Research & Collaboration

Next: genomeRxiv

Questions