EDNA Methodology: E-probe Diagnostic Nucleic acid Analysis

William Schneider

USDA-ARS Foreign Disease/Weed Science Research Unit

Fort Detrick, Maryland

Plant pathogen diagnostic challenges

- **Multiple types of plant pathogens**
  - **Fungal**
  - **Oomycetes**
  - **Bacterial**
  - **Phytoplasma**
  - **Viral**
- **Even more non-pathogenic microbes in the plant biosphere**
  - **Need to be able to distinguish good guys from the bad guys**

Common plant pathogen detection technologies

- Protein based immunological assays (ELISA, Immunostrips, etc…)
- Nucleic acid based assays
  - Conventional PCR
  - Real-time PCR
  - Multiplex PCR
- Not capable of detecting more than a few pathogens at the same time
- All require previous characterization of the pathogen

The one assay….

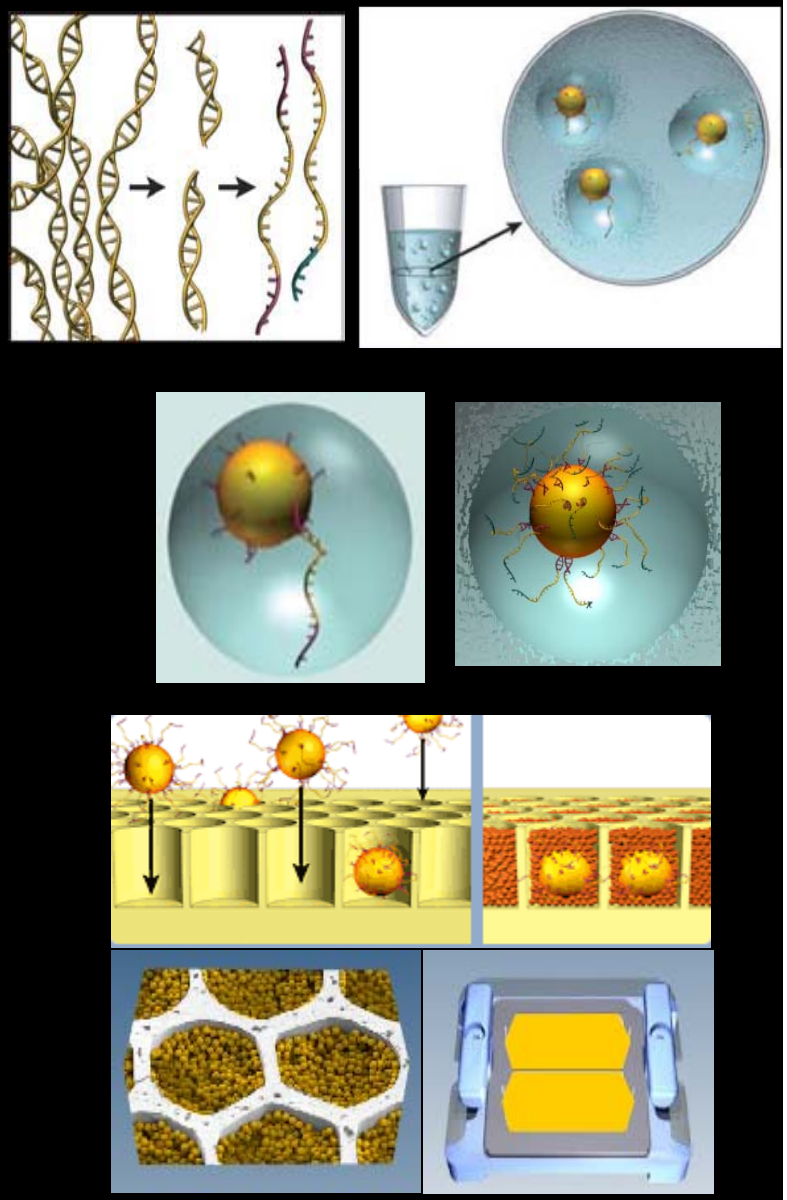
- Can detect any and all pathogens/microbes
  - RNA virus, DNA virus, prokaryotic and eukaryotic
- Can detect from any background
  - Soil, plant, water, insect, etc…
- Can detect both known and unknown organisms
- Easy to use
- Easy to interpret
- Flexible
- Cheap

Nextgen Sequencing

Thousands and thousands of short sequences generated for a given DNA sample (e.g. Roche 454, AB SOLiD, Solexa)

Comprehensive picture of the entire organismal profile for any sample

Metagenomics



NGS diagnostics

The power of NGS is the sheer volume of sequence data generated

The problem with NGS is the sheer volume of sequence data generated

For pathogen and strain identification, full genome sequences are not necessary

The goal is to find a way to ignore all irrelevant sequences and limit the bioinformatic processing of sequences that are of use

Metagenomics based diagnostics

Typical metagenome analysis involves:

- **Sample extraction**
- **NGS**
- **Quality screens of resulting sequence**
- **Assembly**
- **BLAST against Genbank (or some subset)**

The sizes of NGS runs are increasing exponentially

The size of the reference database (Genbank) is increasing exponentially

The reference database is prone to bias

EDNA:
E-probe Diagnostic Nucleic acid Analysis

Bioinformatics tool designed to ignore irrelevant sequences and limit processing

Control the size of the reference database: Dump raw non-assembled sequence data into its own database (create a mini-genbank).

Control the size of the query set: Query the raw sequence data base with a series of signature diagnostic sequences ("e-probes").

Don't ask/don't tell diagnostic tool

100mg

DNA Extraction

Host + Pathogen DNA

WTA/WGA (amplification)

MPS: (454, Illumnia, SOLiD)

SSD

The Sample Sequence Database (SSD) contains a phyto-metagenomic profile

Query with e-probes

Hits

Pathogen is present

Further testing for forensic analysis & looking for genetic modifications

No Hits

STOP

Stobbe et al., Journal of Microbiological Methods
doi: 10.1016/j.mimet.2013.07.002

EDNA diagnostics

First pass query: Identification

Second pass query: Forensic analysis

Third pass query: Evidence of genetic manipulation/engineering

The same "sample" can be assayed bioinformatically in as many ways as the researcher can imagine

The keys to success are proper selection of query sequences and the minimization of post sequencing analysis: Tools for Fingerprint Identification (TOFI)

EDNA Viral detection results

- Detects Plum pox potyvirus and Bean golden mosaic geminivirus from infected plant tissue
- Capable of detecting mixed infections
- Able to strain type Plum pox potyvirus
- By changing the e-probe set to general virus family sequences we were able to identify a novel tombusvirus from switchgrass

Stobbe et al., Phytopathology 104: 1125-1129

EDNA bacterial detection results

- Detects *Psuedomonas syringae*, *Ralstonia solanacearum* and *Serratia marcescens in planta*
- Capable of detecting mixed infections
- Detects GFP modified *S. marcescens*

Daniels et al., BMC Bioinformatics (in prep)

Human pathogens on plants detection results

- Successful detection *E. coli* o157
- Successful detection of *Salmonella* sp.

Blagden et al., J. of Food Science (submitted)

EDNA eukaryotic detection results

- Successful detection of *Phytophthora ramorum* and *Pythium aphinadermatum*
- Successful detection of *Puccinia graminis* and *Phakopsora pachyrhizi*
- Strain typing of *P. aphinadermatum*

Espindola et al., Int. J. of Data Mining and Bioinformatics 12: 115-128

Testing EDNA detection of vectors and pathogens

Proof of concept using single species samples:

Acyrthosiphum pisum* + *Soybean dwarf Luteovirus

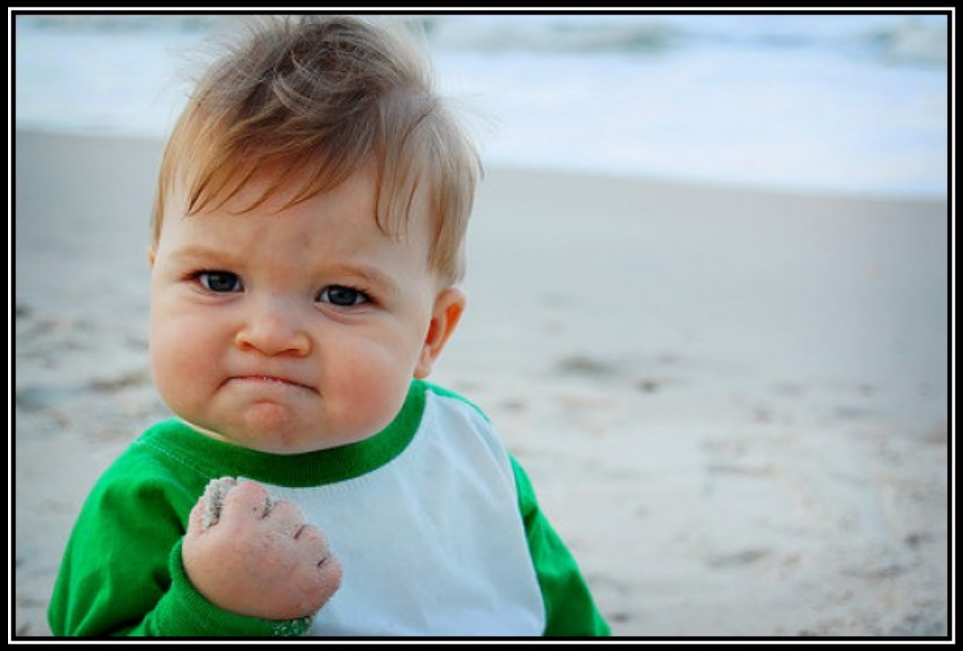
Myzus persicae* + *Plum Pox Potyvirus

***Diaphorina citri* + *Ca.* Liberibacter asiaticus (LAS)**

***Blowflies* + *E. coli* O157**

EDNA results with pathogens

- Detects RNA and DNA viruses
- Detects bacteria
- Detects oomycetes and fungi
- Detects food borne human pathogens
- Useful in pathogen discovery



Comparison

EDNA	Time (HH:MM:SS)	"Traditional"	Time (HH:MM:SS)
Extract fastA	00:00:00	Extract fastQ	00:00:56
EDNA Pipeline	00:00:14	FastQC	00:01:07
		Filter & Trim reads	00:00:58
		BLASTn - GenBank nt	09:18:04
		MEGAN	00:00:00
Total	**00:00:14**		**09:21:05**

- **Over 2400 times faster!**

When does EDNA make sense?

- Situations where diagnostics are needed for a large number of pathogens
- Situations where a wide variety of pathogens are a possibility

- Plant quarantine facilities
- Insect traps/vector surveys
- Introductions of new crops into new ecosystems

EDNA in the Real World*

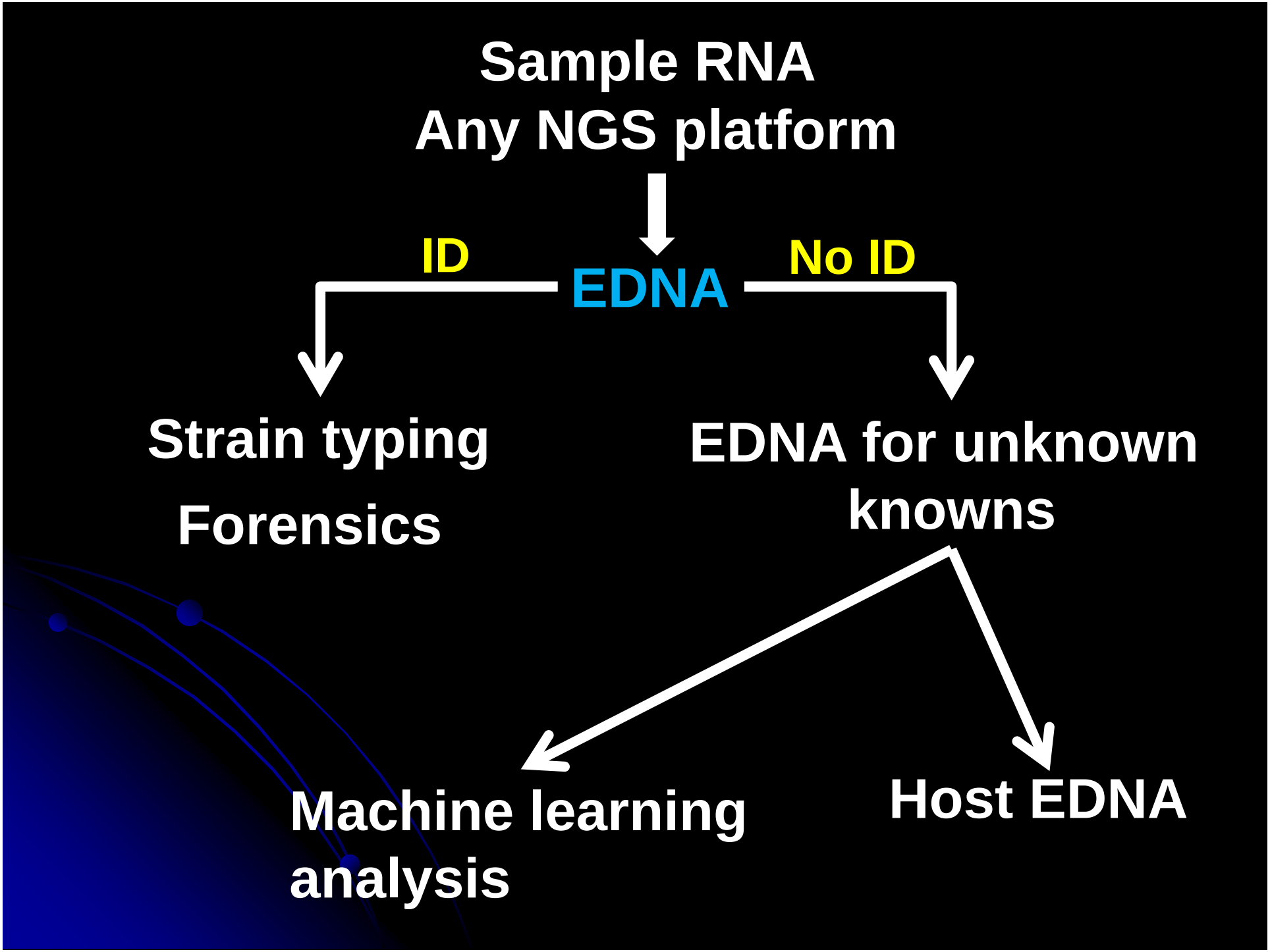
(*Not MTV)

Proof of concept using imported switchgrass accessions:

Discovered a new tombusvirus

Proof of concept using potato samples suspected positive (PCR) for *R. solanacearum* R3BV2:

Determined that samples were positive for a *Ralstonia*, but definitely not R3BV2


Sample RNA
Any NGS platform
ID
EDNA
No ID
Strain typing
Forensics
EDNA for unknown knowns
Machine learning analysis
Host EDNA

Next generation sequencing (NGS) for pathogen detection and discovery

Oklahoma State University
NIMFFAB

Tony Stobbe
Jon Daniels
Andres Espindola
Ruchi Verma
Trenna Blagden
Ulrich Melcher
Francisco Ochoa-Corona
Carla Garzon
Jacque Fletcher

Acknowledgements

Oklahoma State NIMFFAB

Sharon Andreason

Chris Timmons

Maria Ma

Astri Wayadande

USDA-APHIS

Mark Nakhla

John Rascoe

Jorge Abad

USDA Systematic Mycology Lab

Martha Malapi-Wight

Jo Anne Crouch

This project was supported by the Agriculture and Food Research Initiative Competitive Grants Program Grant No. 2010-85605-20542 from the National Institute of Food and Agriculture.

Thanks for listening…