



Early Detection and Prediction

Actionable Data/Information



Helen Cui, MD, Ph.D
Los Alamos National Laboratory

University of Missouri
April 13, 2011

An Unprecedented Biological Threat to Society

■ Food Security

- Predicted population growth from 6B to 9B by 2050
- Food scarcity for a growing population driving alternative sources of meat
- Impacts of malevolent and natural infectious disease outbreaks

■ Energy Security

- Peak oil prices and a need for energy independence
- Environmental impacts of biofuels industry

■ Environmental Security

- Effects of climate change
- Effects of global warming on disease transmission
- Change in vector host range

■ Health Security

- Emerging and re-emerging infectious disease
- Decrease in development of new antibiotics

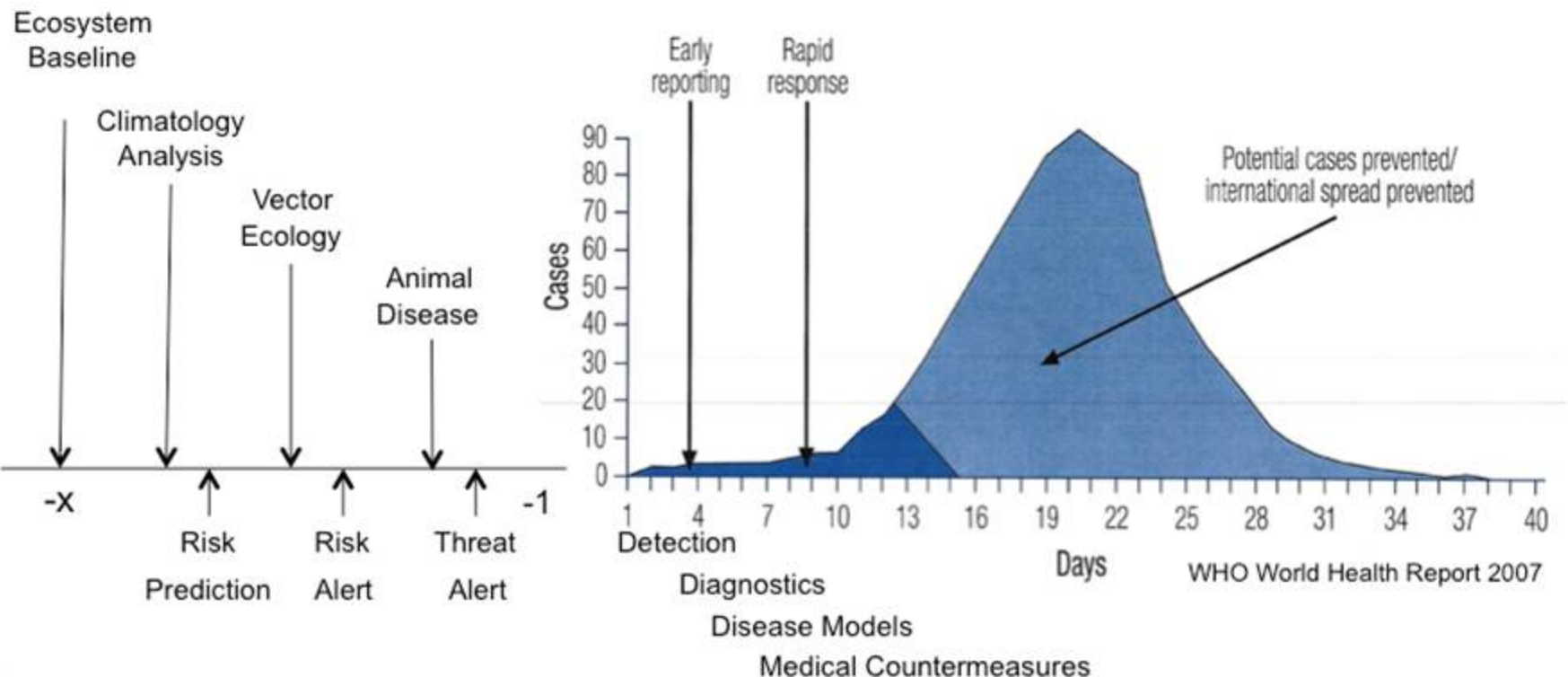
■ Biosecurity:

- Genetically-engineered biothreat
- Inherent threat uncertainty

How Early Can Early Prediction Occur – Courtesy of Bill Huff

HOW MUCH EARLIER CAN WE FORECAST, RECOGNIZE AND CHARACTERIZE AN EVENT?

HOW DO INVESTMENTS TRANSLATE INTO DECREASED MORBIDITY AND MORTALITY?



Enabling S&T Advancements

■ Scientific Foundations

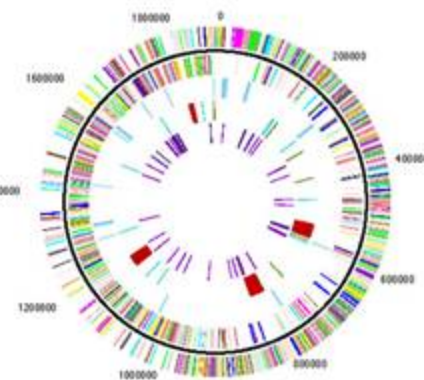
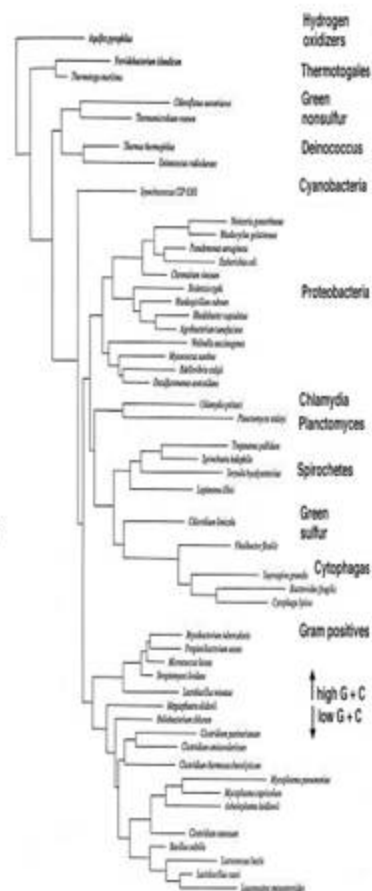
- Genomic science
- Signature science
- Systems biology in post-genomic era

■ Emerging Technology

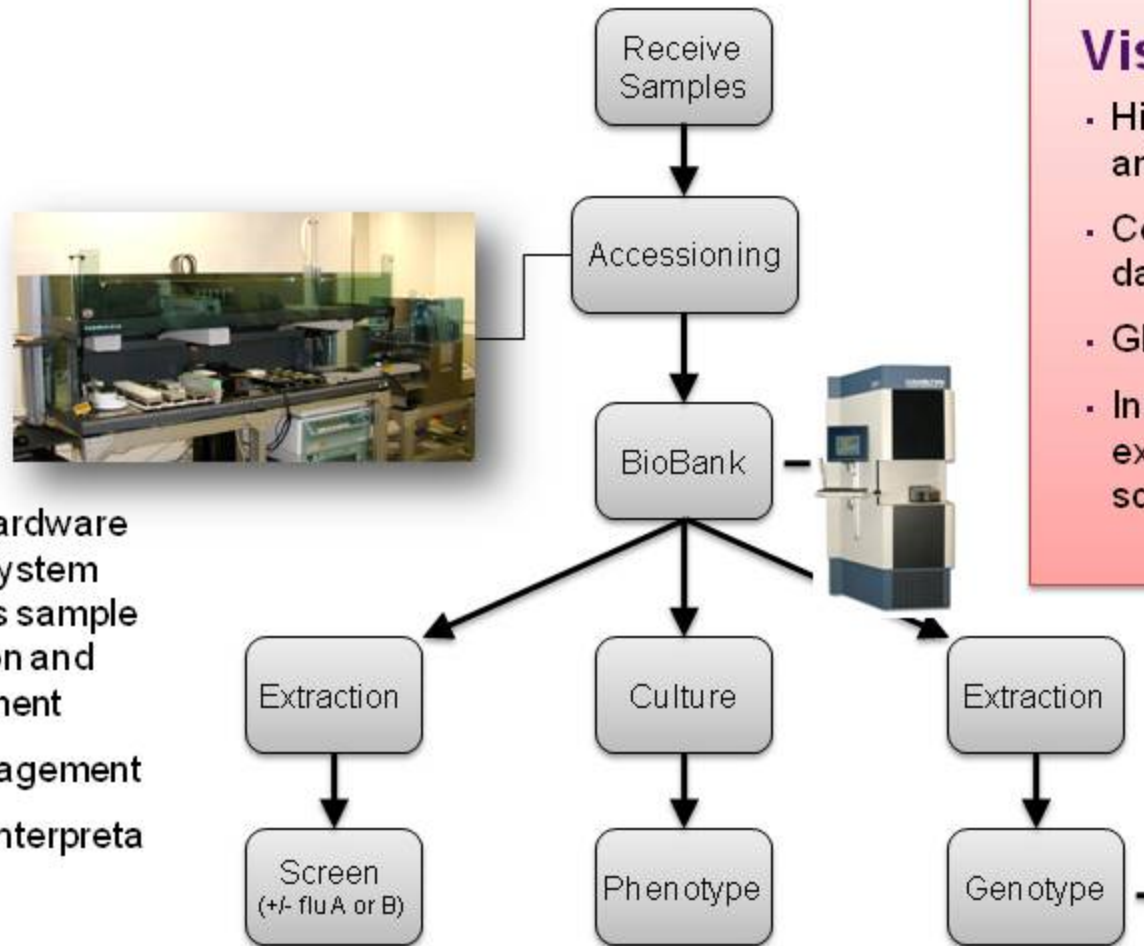
- Multiplexed sensor platforms
- High throughput sample analysis
- Handheld for field application
- Recognition reagents

■ Information Science and Technology

- Data/information management
- Modeling and simulation
- Knowledge generation
- Communication



High Throughput Surveillance Sample Analysis



Vision

- High throughput genotypic and phenotypic analyses
- Common quality control and data management protocols
- Global network
- Initial focus on influenza with expansion to pan-pathogen scope

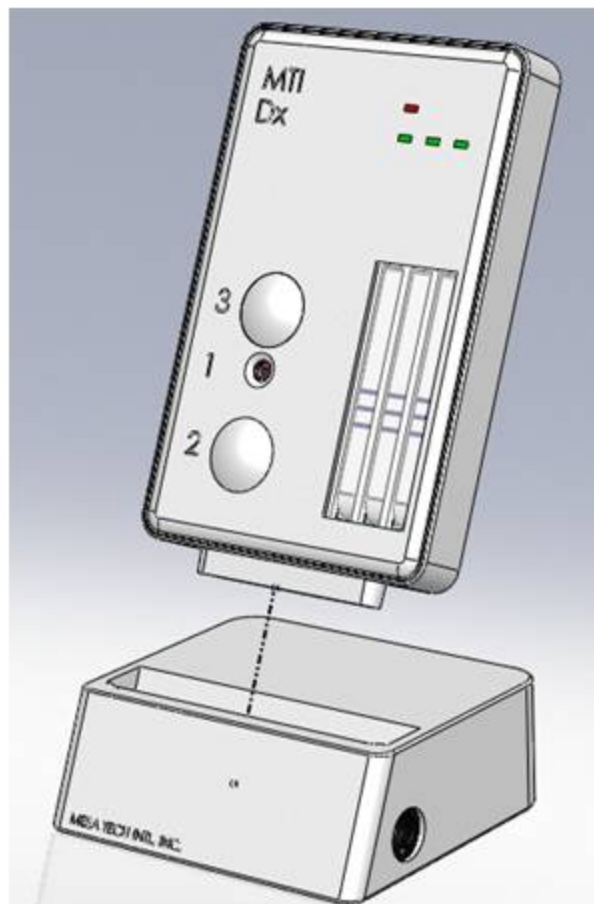
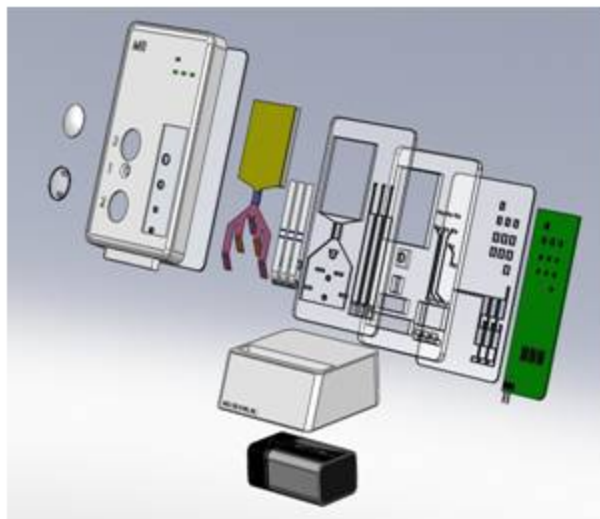
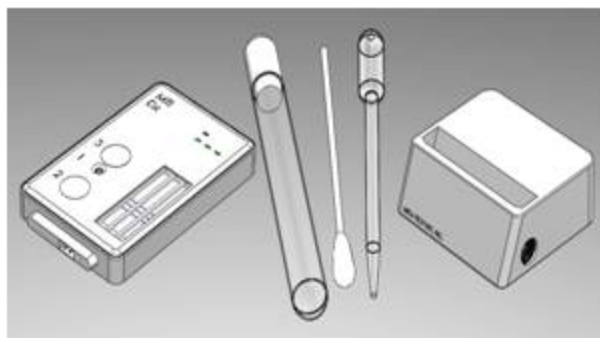
Multiple hardware modular system automates sample preparation and measurement

Data management

Analysis/Interpretation



Handheld for Field Application



Pathogen Feature Identification

If a single gene can indicate a pathogen, it would be an ideal pathogen fingerprint. However, a single “evil” gene doesn’t exist.

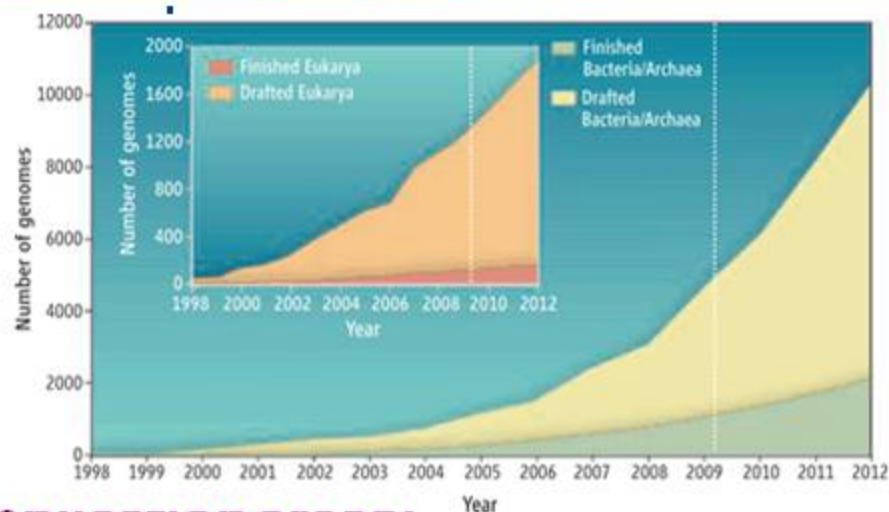
	Protein 1	Protein 2	Protein 3	Protein 4	Protein 5	Protein 6	Protein 7	
Human pathogen	Genome 1	0.94	0.211352	0.879604	0.183404	0.906899	0.125253	0.702583
	Genome 2	0.96	0.255024	0.952227	0.489773	0.361883	0.230016	0.470668
	Genome 3	0.95	0.935249	0.322319	0.711127	0.634149	0.830463	0.219537
Not a human pathogen	Genome 4	0.1	0.913209	0.697955	0.80761	0.92896	0.020984	0.16831
	Genome 5	0.2	0.677969	0.502185	0.930101	0.20112	0.922996	0.155657
	Genome 6	0.01	0.277974	0.607067	0.749366	0.71734	0.414356	0.422842

	Protein 1	Protein 2	Protein 3	Protein 4	Protein 5	Protein 6	Protein 7	
Human pathogen	Genome 1	0.293633	0.373652	0.443052	0.888729	0.308374	0.607774	0.061463
	Genome 2	0.983265	0.111149	0.61294	0.98855	0.691371	0.963415	0.883228
	Genome 3	0.09293	0.673135	0.722787	0.776042	0.117387	0.376356	0.955763
Not a human pathogen	Genome 4	0.885974	0.775081	0.866228	0.923904	0.568658	0.651736	0.000689
	Genome 5	0.48601	0.23065	0.966986	0.84043	0.909809	0.499799	0.051431
	Genome 6	0.114579	0.055455	0.85139	0.503521	0.276715	0.873955	0.299116

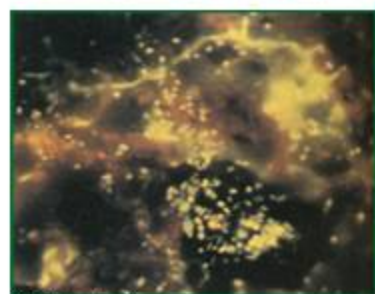
A set of genes may form a “feature vector” that is associated with pathogenicity.

Genomic Science

- Number of high-quality genomic sequences are rapidly growing
- Informatics tools are rapidly advancing to analyze data

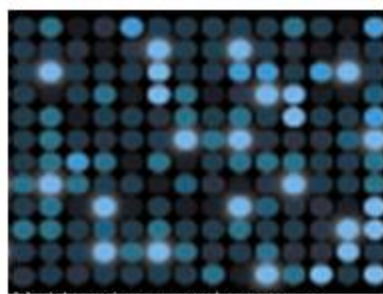


Genomic science impacts many application areas:



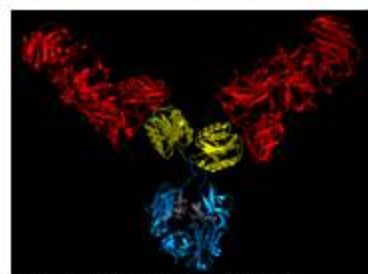
Soil microbes

Microbial community
characterization



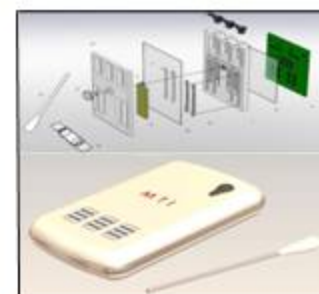
Multi-pathogen micro-array

Detection and forensics
signature design



Anthrax inhibitor protein

Medical
countermeasure
development



Detection 'Dipstick' from MESA Inc.

Development of
field-deployable
detectors

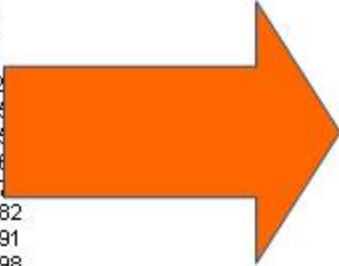
The Potential Feature Space Has 2.2 Million Genes!?

The Potential Feature Space

-15799784
15799892,15801020,15799784
15799993,15801120,15799892,15801020
15800098,15801223,15799993,15801120
15800204,15801326,15800098,15801223,15802248,15800305,15801429,15800204,15801326,15802355
15800305,15801429,15800407,15801531,15802459,15800407,15801634,15800305,15801429,15802568
157999784,15801531,15800699,15801429,15802459,15800407,15801634,15800699,15801531,15802673
157999642,15801634,15800704,15801736,15802782,15800704,15801836,15800704,15801736,15802891
157999642,15801736,15800816,15801836,15802998,15800816,15801936,15802998,15800816,15802000,15803108
15800407,15802000,15803108,15800510,15801634,15802248,15800510,15801736,15802355,15800612,15801836,15802459,15800612,15801936,15802568,15800714,15802041,15801531,15802673,15800714,15802147,15801634,15802782,15800816,15802248,15801736,15802891,15800920,15801836,15802998,15800920,15802041,15801936,15803108,15801020,15802248,15801120,15802355,15801223,15802459,15801326,15802568,15801429,15802673,15801531,15802782,15801634,15802891,15801736,15802998,15801836,15803108,15801936,15803212,15802041,15802147,15802248,15802355,15802459,15802568,15802673,15802782,15802891,15802998,15803108,15803212,15803314,15803419,15803527

Identify A Feature Vector

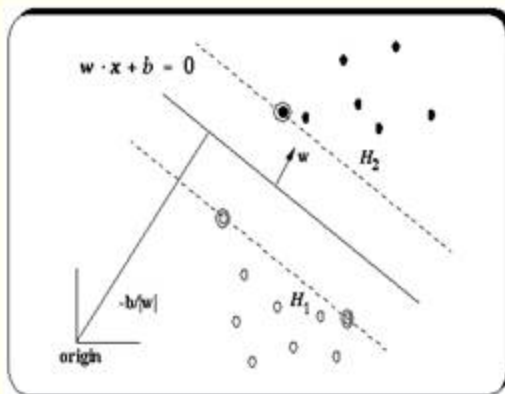
-15602605
15799782
15829356
21233407
26246031
104779614
104780100
39996560
110640311
39997323
110833460
42525857
113869438
50119196
113869451
51892984
56418554
116749888
70732089
120598308
71736092
146283146
74316443
146293614
77166455
146306290
77460629
147676401
78043976
148264838
78221829
148380774
78485931
153930834
82701584
153939013
85858137
157147571
91793895
158424158
92113669
167629470



92117118
94676720
94986798
104779614
104780100
110640311
110833460
113869438
113869451
113969687
116749888
120598308
146283146
146293614
146306290
147676401
148264838
148380774
153930834
153939013
157147571
158424158
167629470

Support Vector Machine (SVM) as the Classifier

Support Vector Machines as Classifiers

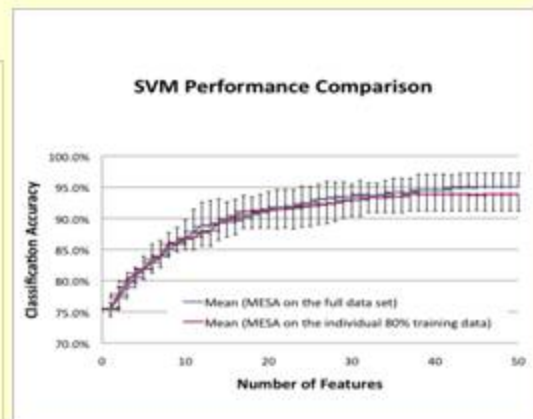


- Maximizes the margin between classes to satisfy structural risk minimization.
- Balances empirical risk and generalization.
- Kernels provide the means for nonlinear classification.

Nonlinear classifier in the original space

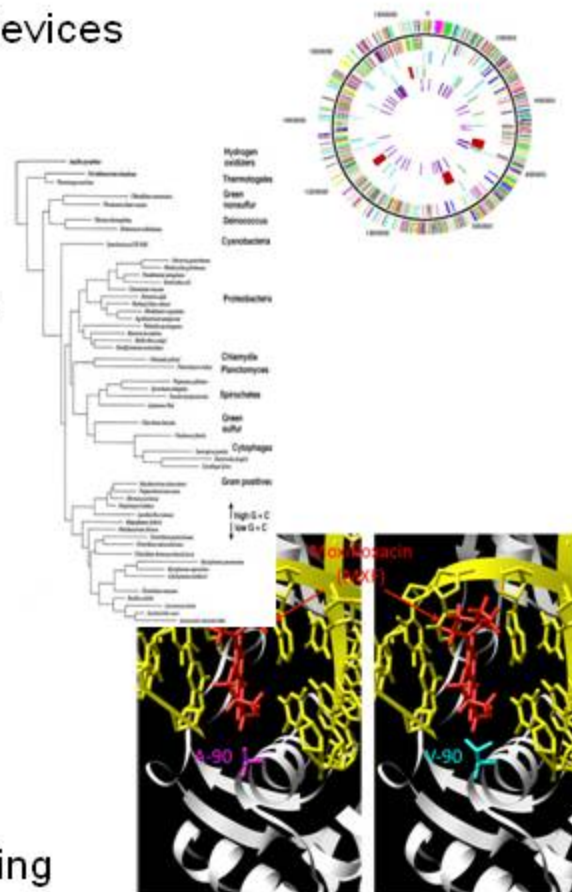
Nonlinear Transformation

Linear classifier in the feature space



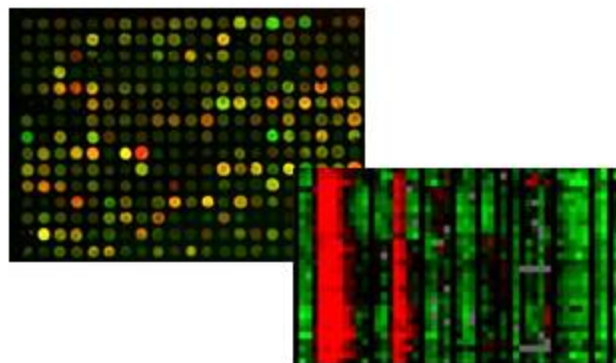
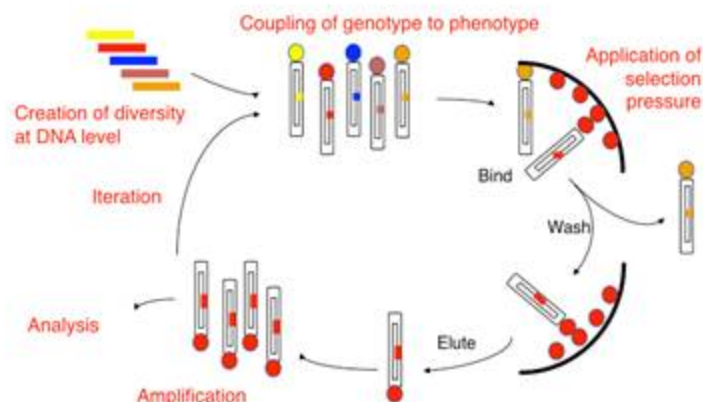
Signature Science

- **Contextual basic knowledge**
 - Fundamental for deriving hypotheses and developing devices
- **Pathogen genetic determinants**
 - Selection, variation, Phylogenetic relationships
 - Need to address pathogen signature erosion
 - From genotypic features to phenotype characterization
- **Genetic manipulation markers**
- **Host-determined pathogenic features**
 - Pathogen-specific Immune response
 - Molecular biomarkers
- **Structural modeling-based validation**
 - TB drug resistance SNP signature as example
 - SNP site identified for TB multi-drug resistance
 - SNP-based gene identification, protein structure modeling
 - Determination of drug binding



Advanced Measurement S&T

- **Tailored ligands on demand--*In vitro* antibody Display**
 - Recognition beyond expressed protein
 - Greater degree of control of specificity and sensitivity
 - High affinity, increased stability, rapid production
- **Multi-analytes:**
 - nucleic acid, protein
- **Highly Multiplexed Arrays**
- **Micro-fabrication**



Strongly Market Driven

Pathogen Identification & Characterization

■ Basic microbiology techniques

- Collection, isolation, culture
- Microscopic examination
- Colonial morphology for presumptive identification
- Biochemical identification

■ Sequence-based techniques

- DNA sequencing, PCR, Gene analysis
- Comparative genomic analysis
- Bioinformatics and functional genomics

■ Host-pathogen interaction studies

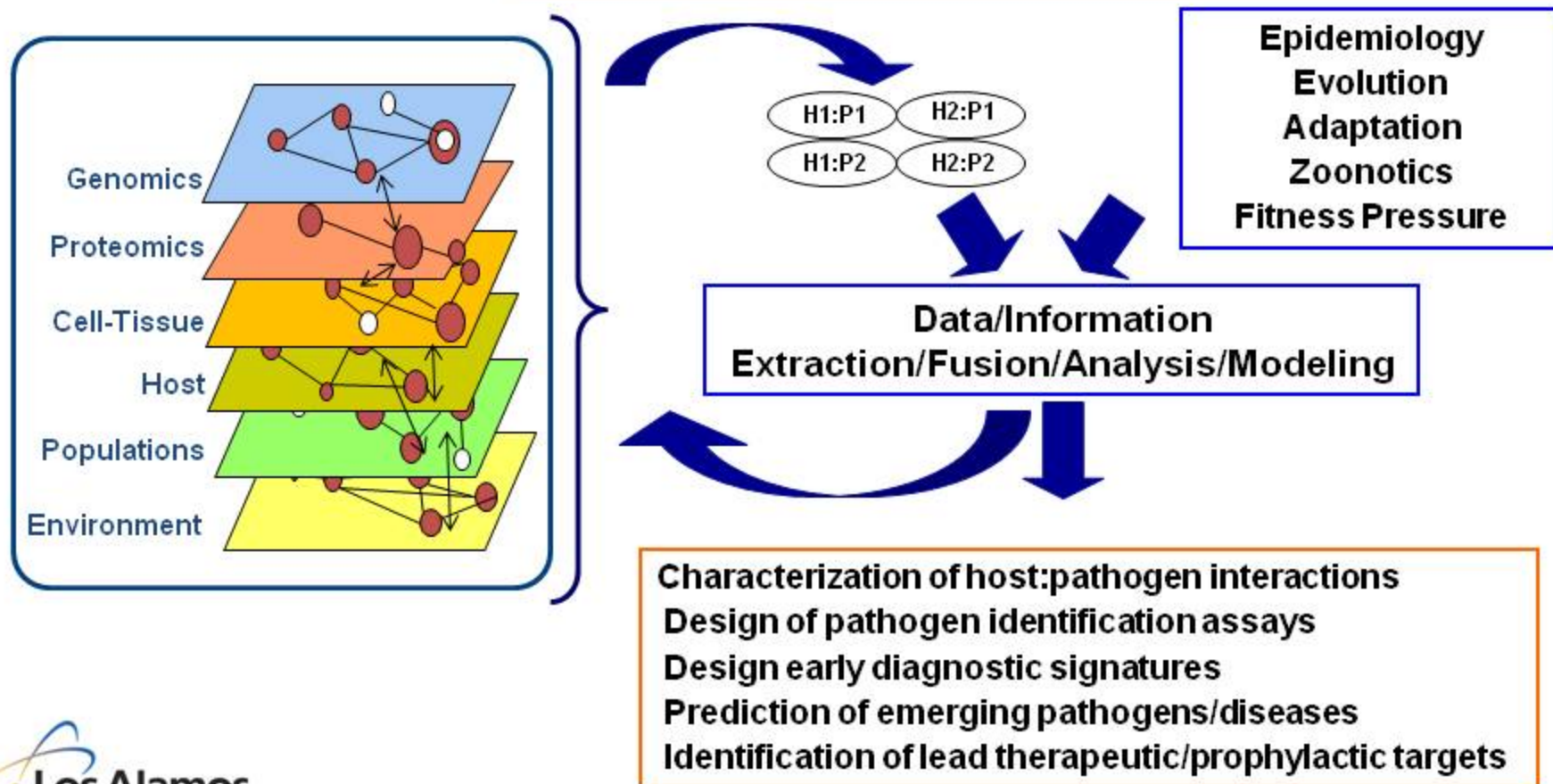
- Host susceptibility

■ Epidemiology studies, simulation

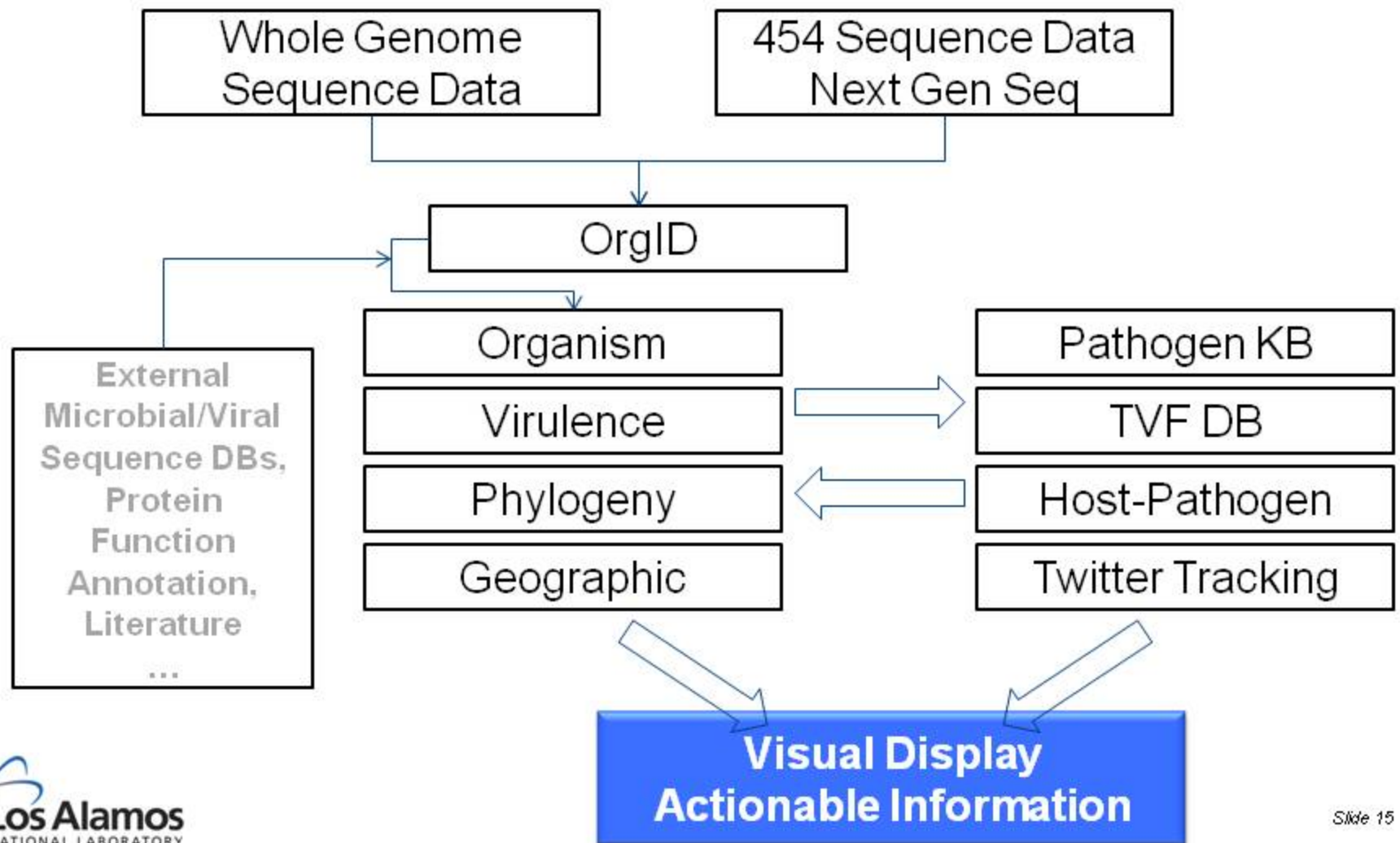


Systems Biology Approaches

Experimental Host:Pathogen:Environment (Hn:Pn:En) Model



System Concept:



Enhanced Global Biosurveillance

Broadening the Scope of Biothreat Detection, by Monitoring for:

- Emerging and re-emerging biothreat agents
- Accidental release
- Engineered biothreat agents employing highly multiplexed sensors
- Pan-pathogens based on fundamental understanding of pathogenicity
- Multi-modes of exposure (air, food, water, carriers)

Moving “Upstream” for Early Warning and Prediction, by:

- Expanding GBSV “upstream” both spatially and temporally
- Monitoring ecological reservoirs (human, agricultural, wildlife, environmental) based on disease ecology
- Monitoring zoonotic interface and high human traffic nodes
- Predicting human virulence/contagion based evolutionary fitness



For GBSV Responsive to:

A fast moving world
Biothreat surprise
Available technology
The National Biosecurity Strategy

Surveillance Data Analysis and Interpretation

- **Data management:**
 - Standardizing data architecture, management
 - Stream line of tools, and automation
- **Omics data integration and multi-scale analysis**
- **Data interpretation**
 - Computer-aided sequence analysis leading to the identification of genetic determinants (e.g. machine learning)
 - Expert guided analysis with automated search
- **Modeling for prediction and simulation**
 - Host-pathogen models
 - Epidemiological models
 - Plume dispersion pattern, path, deposition
 - Emerging threats from globalization and climate changes
- **Hardware**
 - High performance computing
 - Field deployable analysis

Single portal upload and information access



Teams:

■ BioPASS:

- Helen Cui, PI
- Craig Blackhart, systems
- Bob Funkhouser, programming
- Jennifer Harris, infectious diseases
- Chris Stubben, pathogen virulence
- Chen He, twitter map
- Ben MacMahon, phylogeny
- Carla Kuiken, sequence database
- Jian Song, virulence factors
- Patrick Chain: metagenomics
- Amanda Minnich, pathogen knowledge
- Nick Hangartner, proposal development
- Julianna Fessenden, program development
- Gary Resnick, biodefense, strategy

■ Pathogen Feature:

- Jason Gans, sequence analysis
- Jennifer Harris, infectious disease
- Michael Cai, machine learning
- Ruy Ribeiro, host-pathogen

■ EDP Detection and Diagnosis:

- Norman Doggett, signature and assay
- Karen Hill, toxin
- Jennifer Harris, viral pathogens
- Amanda Minnich, bacterial pathogens
- Jason Gans, *in silico* validation

■ Sponsors:

- Department of State
- Department of Defense
- LANL Laboratory Directed Research and Development