

Genomics for Bioforensics

Marc Colosimo

781-271-7339 •
mcolosimo@mitre.org



MITRE
Technology
Program

Problem

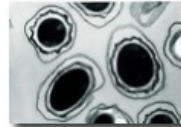
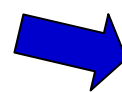
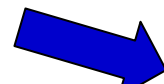
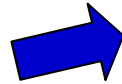
- **How can genomics be used for attribution in a bioterrorism event or biocriminal case**
 - **To help identify the source of a microbial pathogen or toxin**
- **How can genomics be used to distinguish naturally emerging infectious diseases**
 - **From purposefully released pathogens?**
 - **From engineered pathogens?**

Background

“Classical” Forensics Data

- Fingerprints
- Human DNA fingerprinting
- Trace elements

Sample
Collection



Interdiction
Attribution

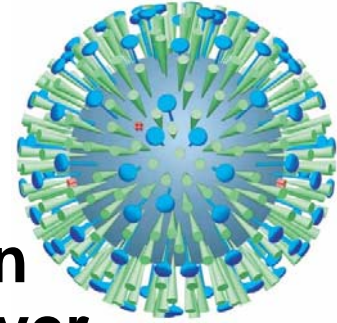
“Bio” Forensics Data

- Primary identification (pathogen or toxin)
- **Strain identification using genomics**

Objective

- **Determine the state-of-the-art in genomic bioforensics (microbial forensics)**
 - **What are the current tools and techniques**
 - **What are the limitations of these tools and techniques**
 - **What type of data resources are necessary**
- **Develop and validate methodology and sample match procedures for attribution**
 - **Initially for one virus (influenza) - year 1**
 - **Extending to other naturally occurring and engineered viruses - years 2 and 3**

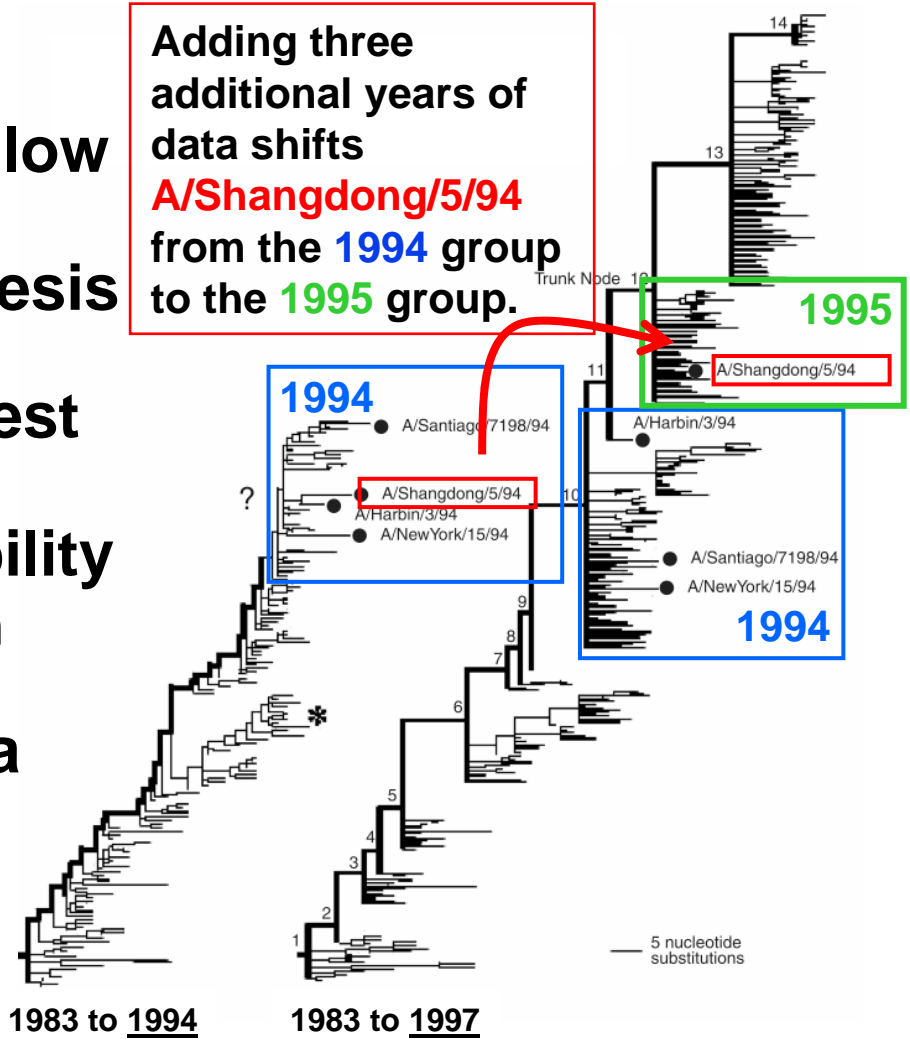
Activities



- We are using genomic data from TIGR on human H3N2 influenza gathered from New York State over 4 influenza seasons (300+ complete sequences)
- We are examining several popular phylogenetic methods and comparing their results (neighbor joining, parsimony, maximum likelihood)
- We are researching new methods that could be used to assign probabilities to assessment of “relatedness”
 - Shape space and multi-dimensional scaling
- Collaborations:
 - We have begun a collaboration with Elodie Ghedin (TIGR and U. Pittsburgh School of Public Health)
 - We have gotten data from Robin Bush (UC Irvine), now funded under the NIGMS MIDAS program

Highlight

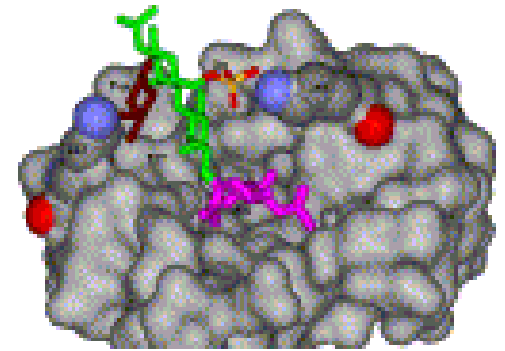
- Current methods do not allow us to assign statistical significance to the hypothesis
 - There are methods to test for the robustness of a tree, but not for probability of evolutionary relation
- Placement of a sample in a group/clade is dependent on the data
 - Geo-spatial/**temporal** sampling is very important



Bush *et al.* 1999

Highlight II

- **Collecting and culturing a viral strain causes the virus to evolve (e.g., to adapt to a new host such as chicken egg), which may leave a genomic fingerprint useful for forensics**
 - **The passage conditions of reference specimens need to be known for proper forensic analysis**
- **There are 22 known codons that undergo egg-adapted changes in the haemagglutinin molecule (HA)**
 - **We have applied machine learning tools to classify egg versus cell passaged viruses**
 - **Preliminary results show 87% accuracy classifying egg vs cell passages strains**



HA receptor binding to sugar molecule from host cell

Impacts

- **Microbial forensics is key to homeland security (e.g., DHS NBACC: National Biodefense Analysis and Countermeasures Center)**
- **Understanding the relationship between a pathogen (genomic variability, temporal-spatial distribution) and structure of the reference sequence library is critical to provide attribution with associated probabilities**
- **Science underlying bioforensics supports related work on the modeling, prediction and management of disease outbreaks, e.g.,**
 - **Epidemiology (e.g., NIGMS MIDAS program)**
 - **Public health surveillance**
 - **Development of new vaccines and drugs**



Future Plans

