

Influenza Infectious Disease Ontology (Influenza-IDO)

IDO Conference - December 9, 2010

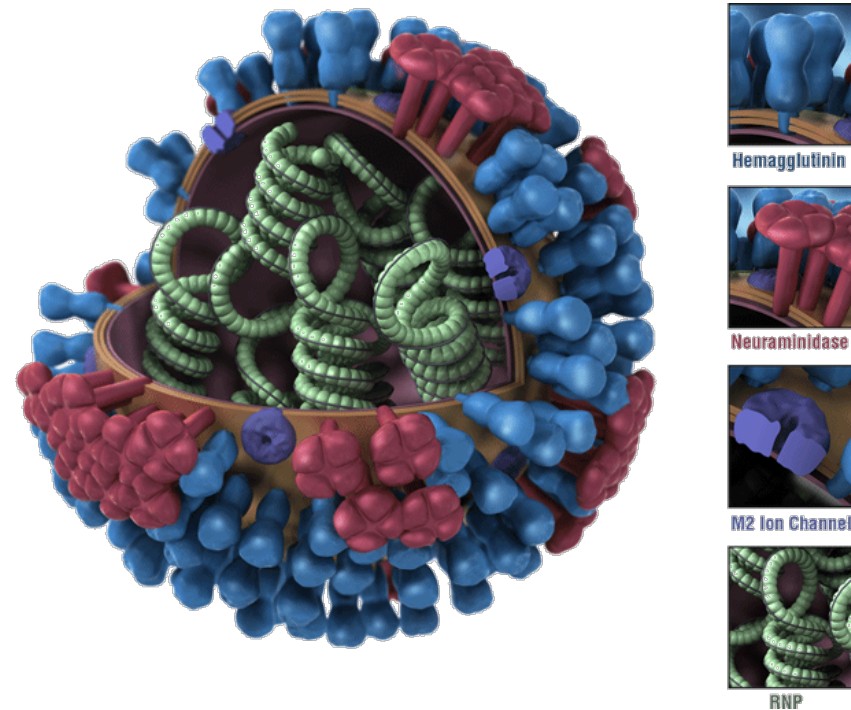
R. Burke Squires
Influenza Research Database
U.T. Southwestern Medical Center

Influenza-IDO

- Influenza-IDO
- Present Needs / Potential Benefits

Influenza Virus

- Single stranded, negative sense RNA virus
- 8 RNA Segments
- 11 Proteins
- Epidemics
- Pandemics



Influenza-IDO Authors

- Ryan Brinkman, BCCA
- **Melanie Courtot, BCCA**
- Meredith Keybl, MITRE
- Joanne Luciano, RPI
- Richard Scheuermann, Influenza Research Database (IRD) / UT Southwestern
- Lynn Schriml, Univ. Maryland
- R. Burke Squires, IRD / UT Southwestern

Influenza-IDO

Domain Requirements

- Influenza ontology is an extension of IDO
- Host and Pathogen:
 - Anatomy (Host and pathogen)
 - Taxonomy (Host and pathogen)
- Biological Material
- Infectious Disease / Clinical
- Laboratory Experiments
- Surveillance (outside of IDO)
 - Disease transmission and specimen collection

Influenza-IDO V0.9

- Includes (Thanks to OntoBee!):
 - No. of terms (including imported) – 1160
 - No. of terms (excluding imported) – 315
 - Terms shared by BO & FLU: 587
- PURL:
 - <http://purl.obolibrary.org/obo/flu/dev/flu.owl>
 - <http://purl.obolibrary.org/obo/flu.owl>

Example Influenza-IDO Terms

- Surveillance (Avian)
 - Wild / Domestic
 - Migratory / non-migratory
- Virus Experimental Infection Event
 - Route of delivery
 - Amount delivered (Volume)
 - Amount delivered (Titer)
 - Passage time period
- Host Anatomy / Age
 - Cloaca
 - Oral/Pharyngeal
 - Hatch year
 - Adult
- Sequence Changes (Analysis)
 - Virus sequence residue type
 - Virus sequence residue number
 - Virus specimen source residue
 - Post passage residue

**PRESENT NEEDS /
POTENTIAL BENEFITS**

CEIRS & Minimum Information Data Standards

- Center for Excellence in Influenza Research and Surveillance (CEIRS)
- Minimum Information Data Standards
 - Facilitate submission of data to IRD
 - Facilitate query, visualization and analysis
 - Separate document for each type of data
 - Avian surveillance (finalized)
 - Non-human mammalian surveillance (finalized)
 - Virus isolate data (finalized)
 - Phenotype
 - Human surveillance
 - Passage experiment

Animal Surveillance Search

CEIRS MDS

Minimum Data Fields for Wild and Domestic Bird Surveillance
Finalized October 3, 2008 by DMID/NAID
Updated April 30, 2010

- Location (REQUIRED)**
Definition: Descriptor of local sampling environment, such as name of wild life refuge, name of zoo, city name, name of retail market. Examples: Bass Lake Refuge; Memphis
Format: Text
Curation: None
- Country (REQUIRED)**
Definition: Country in which original sample was collected
Format: Value must be found on the approved list used by NCBI and found at the NCBI web site: <http://www.ncbi.nlm.nih.gov/projects/collab/country.html>
Curation: Must appear in NCBI country codes as above
- State/Province (REQUIRED)**
Definition: Descriptor of sampling location that is one gradation finer than country, for example Minnesota, Alberta, Guangxi. Must spell out the name of the state/province.
Format: Text
Curation: Must appear in list of approved names; combined list from all CEIRS maintained and updated by IRD
- Latitude (REQUIRED)**
Definition: Latitude by GPS of site at which original sample was collected. If not available, a course estimate is acceptable.
Format: WGS84, decimal degrees, signed, five digit accuracy when available
Curation: Checked for within-range
- Longitude (REQUIRED)**
Definition: Longitude by GPS of site at which original sample was collected. If not available, a course estimate is acceptable.
Format: WGS84, decimal degrees, signed, five digit accuracy when available
Curation: Checked for within-range
- Host Species (REQUIRED)**
Definition: An accurate name of host genus and species, without abbreviations or unknown. If genus is known but species is unknown, then use the genus followed by "sp" (for example domestic duck would be reported as "Anas sp"). If both the genus and species are unknown, then use the scientific family name, without abbreviation.
Format:

IRD Implementation

IRD - Influenza Research Database

SEARCH DATA ANALYZE & VISUALIZE ACCESS WORKBENCH SUBMIT DATA HOME

Animal Surveillance Search

Search influenza surveillance data deposited in IRD. Worldwide surveillance data from both avian and non-human mammalian studies has been collected. Both flu positive and flu negative samples have been found. Search results can be shown on a world map.

A keyword search is also available.

SURVEILLANCE DATA TYPE

Avian
 Non-human mammalian (Beta)

SAMPLE SELECTION

Only tested samples
 Only flu-positive samples

Subtype of flu positive samples
 Use comma to separate multiple entries

SPECIFY TAXONOMY INFORMATION

Order

All
 Anseriformes
 Apodiformes
 Caprimulgiformes
 Charadriiformes
 Ciconiiformes
 Columbiformes

DISPLAY TYPE

List of host animals surveyed
 List of surveillance samples
 Customizable summary table

ADVANCED OPTIONS

Select Advanced Option

Latitude/longitude range

LATITUDE/LONGITUDE RANGE

Latitude Min. Max.
 Longitude Min. Max.

SAMPLING LOCATION

All
 Bangladesh
 Benin
 Brazil
 Cambodia
 Canada
 China

HOST SELECTION

All host species
 Select order from list
 Select family from list
 Select species from list
 Specify species scientific or common name
 Select host from avian taxonomy

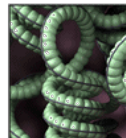
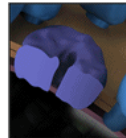
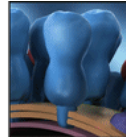
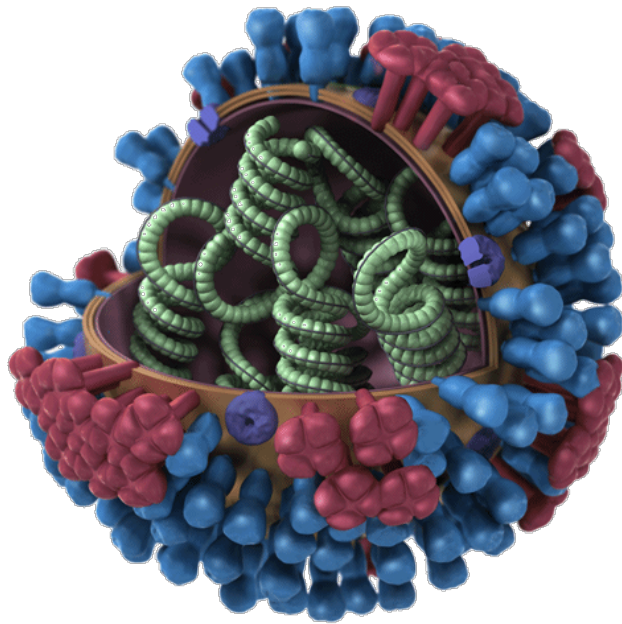
Tip: To select multiple items, CTRL-click on Windows or CMD-click on Mac OS

Clear Search

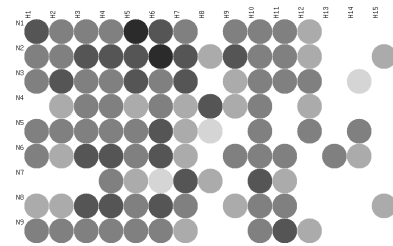
Site IRD Tutorials Report a Bug Request Web Training Contact Us
 This project is funded by the National Institutes of Health and the National Science Foundation (NIH) under Contract No. H49N016006400410 and is a collaboration between MCHRC, Guannan Health IT, University of Texas Southwestern Medical Center, Virona Technologies, RAGE Analytics and Los Alamos National Laboratory

Release Date: Jul 23, 2010

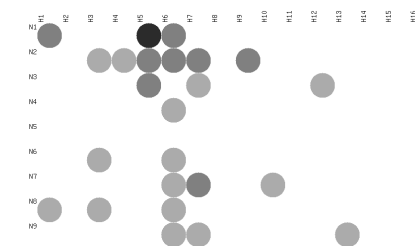
Subtype (H1N1)



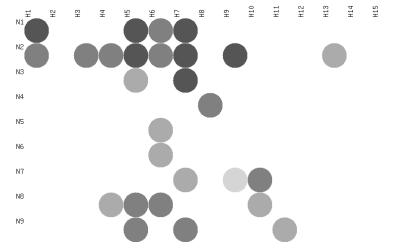
Duck



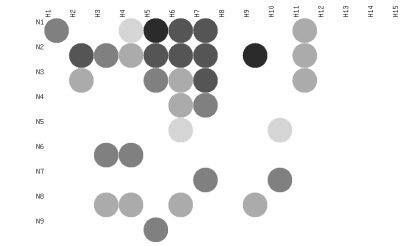
Goose



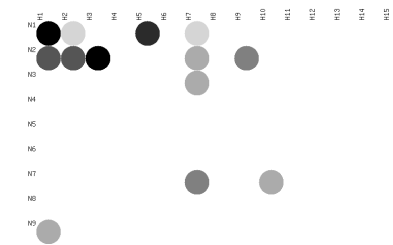
Turkey



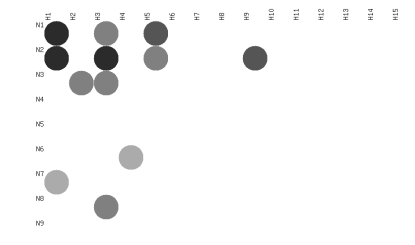
Chicken



Human



Swine

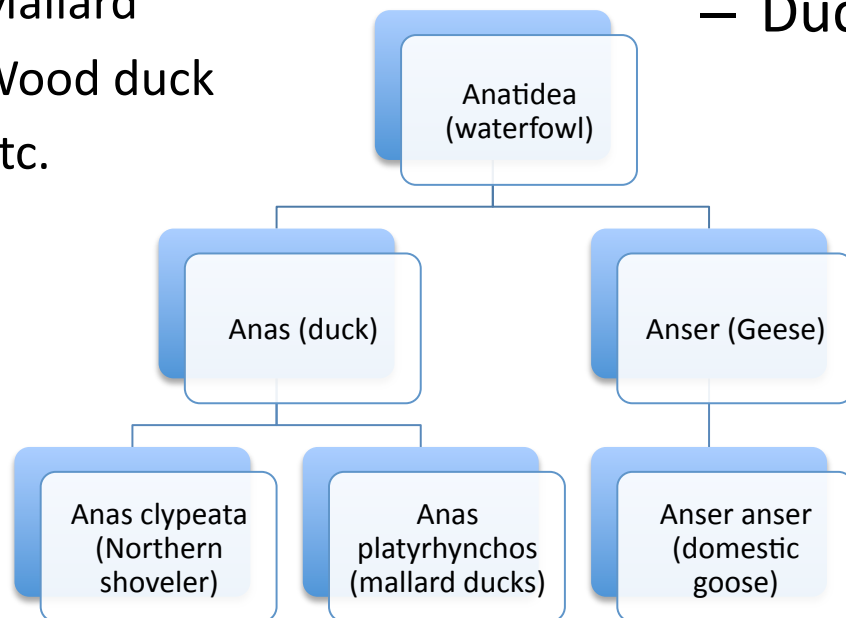


Ontology Enabled Research: Precise Search & Analysis

- Without ontologies:
 - Search for “duck”: 6681
 - Host searches:
 - Include non-relevant data
 - Exclude relevant data
- False positives:
 - Ruddy shelduck
 - Australian shelduck
- False negatives:
 - Mallard: 5236
 - Dk: 95
- With ontologies:
 - Search for “duck” will:
 - include all duck species
 - NOT include non-duck species
 - Cleaner data means cleaner statistics

Ontology Enabled Research: Quick Searches Across Species

- Without ontologies:
 - Search for each species independently:
 - Shovler
 - Mallard
 - Wood duck
 - Etc.
- With ontologies:
 - Search by order, family, genus and get all species below.
 - Duck, NOT goose



Location

- Current location
 - Local location
 - Country
 - State/Province
 - Latitude
 - Longitude
- Use of Gaz
- Challenges
 - US by state
 - US by city
 - Infectious disease to not respect geopolitical boundaries
 - Regions / Continents

Q & A