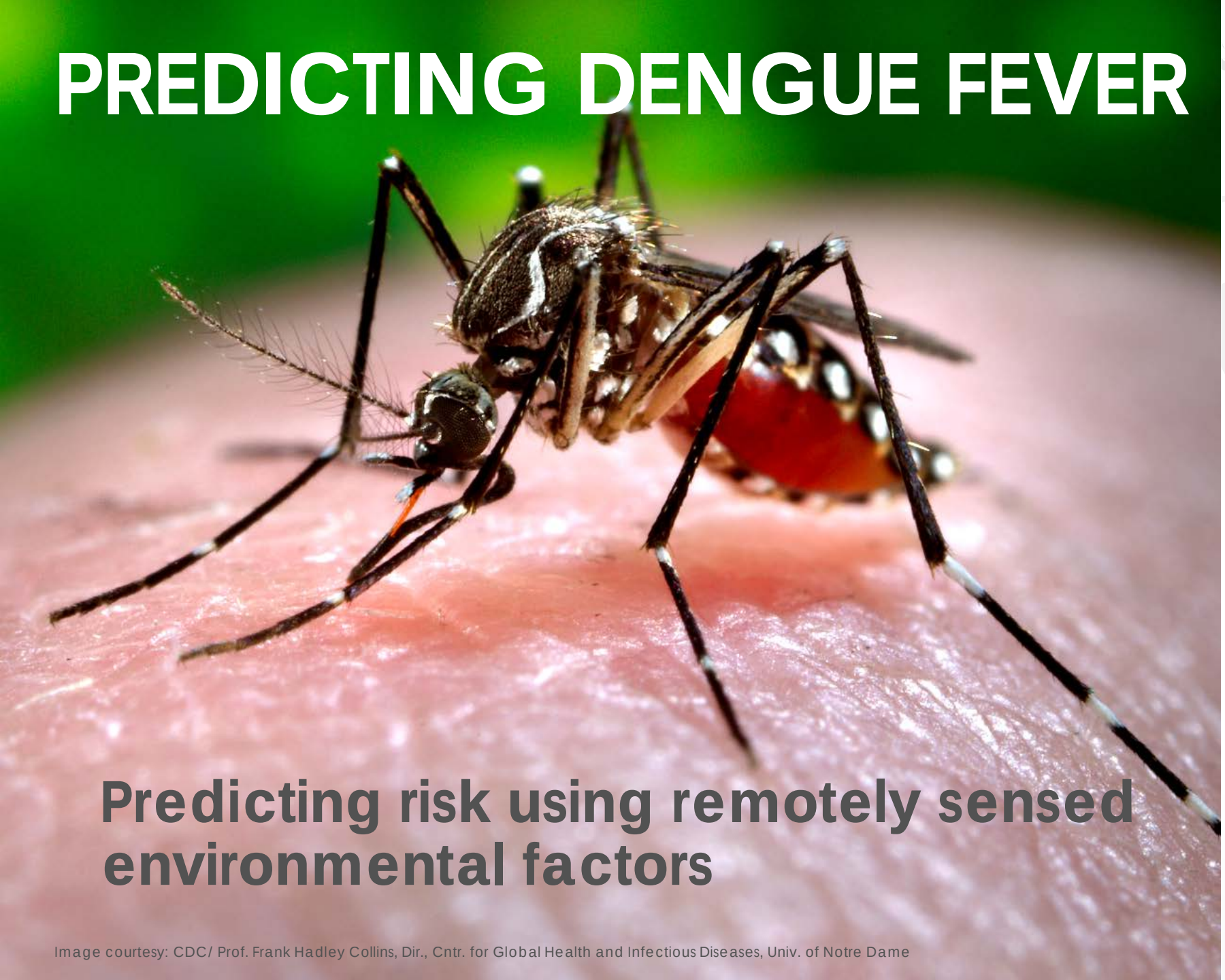
PREDICTING DENGUE FEVER



Predicting risk using remotely sensed environmental factors

Image courtesy: CDC/ Prof. Frank Hadley Collins, Dir., Cntr. for Global Health and Infectious Diseases, Univ. of Notre Dame

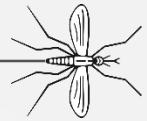
Jeffrey Ashby
Department of Environmental Health Science
Fairbanks School of Public Health

Max J. Moreno-Madriñán
Department of Environmental Health Science
Fairbanks School of Public Health

Austin Stanforth
Department of Geography

IUPUI

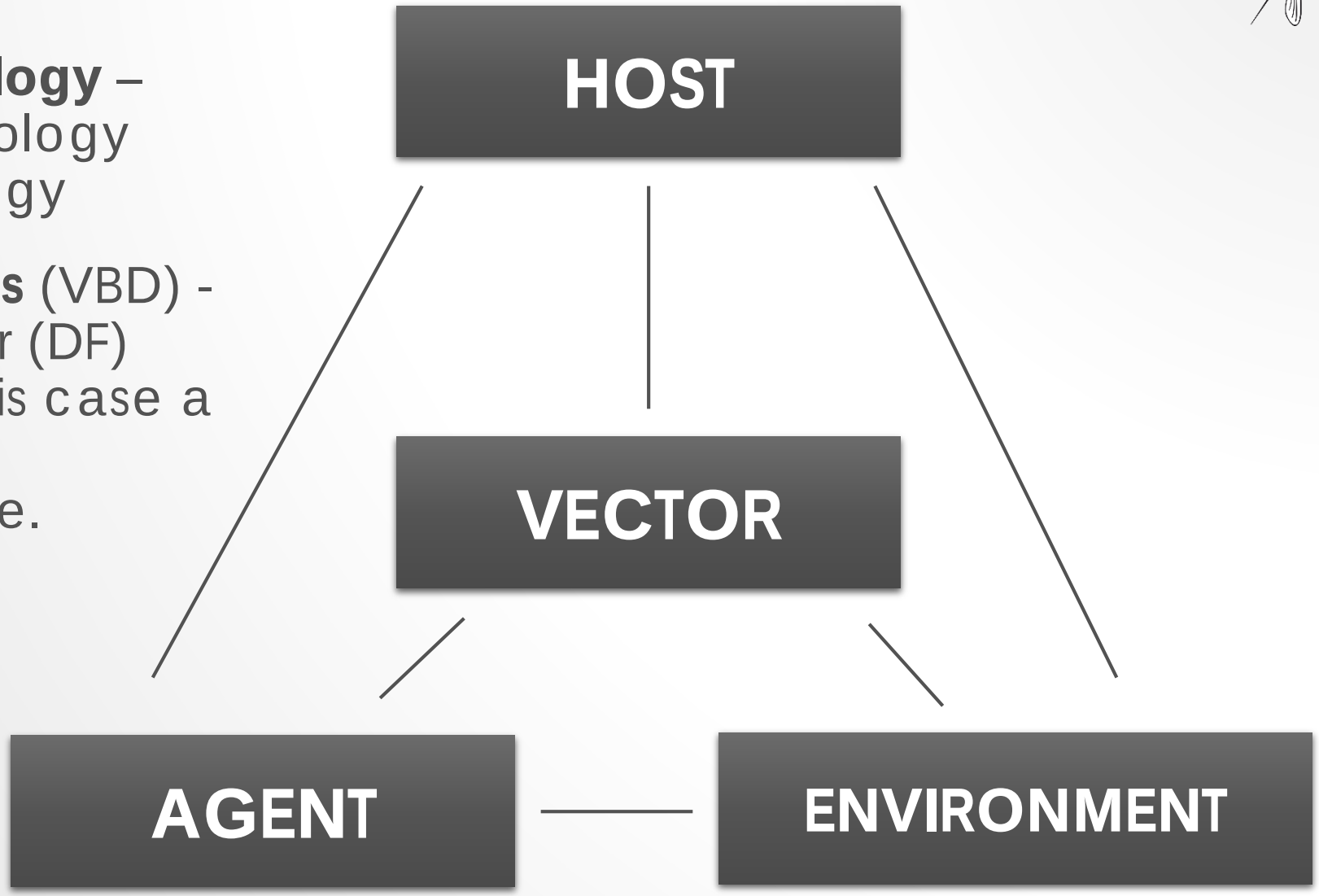
AGENDA



- **Opening Scene** - Dengue Fever and vector-borne diseases
- **The Villain** - the *Aedes aegypti* mosquito
- **The Setting** – Colombia's Magdalena Watershed
- **Supporting Cast** – The Satellites
- **The Hero** – Boosted Regression Trees
- **Roll the Credits** – Results and acknowledgments
- **Audience Commentary** - Questions

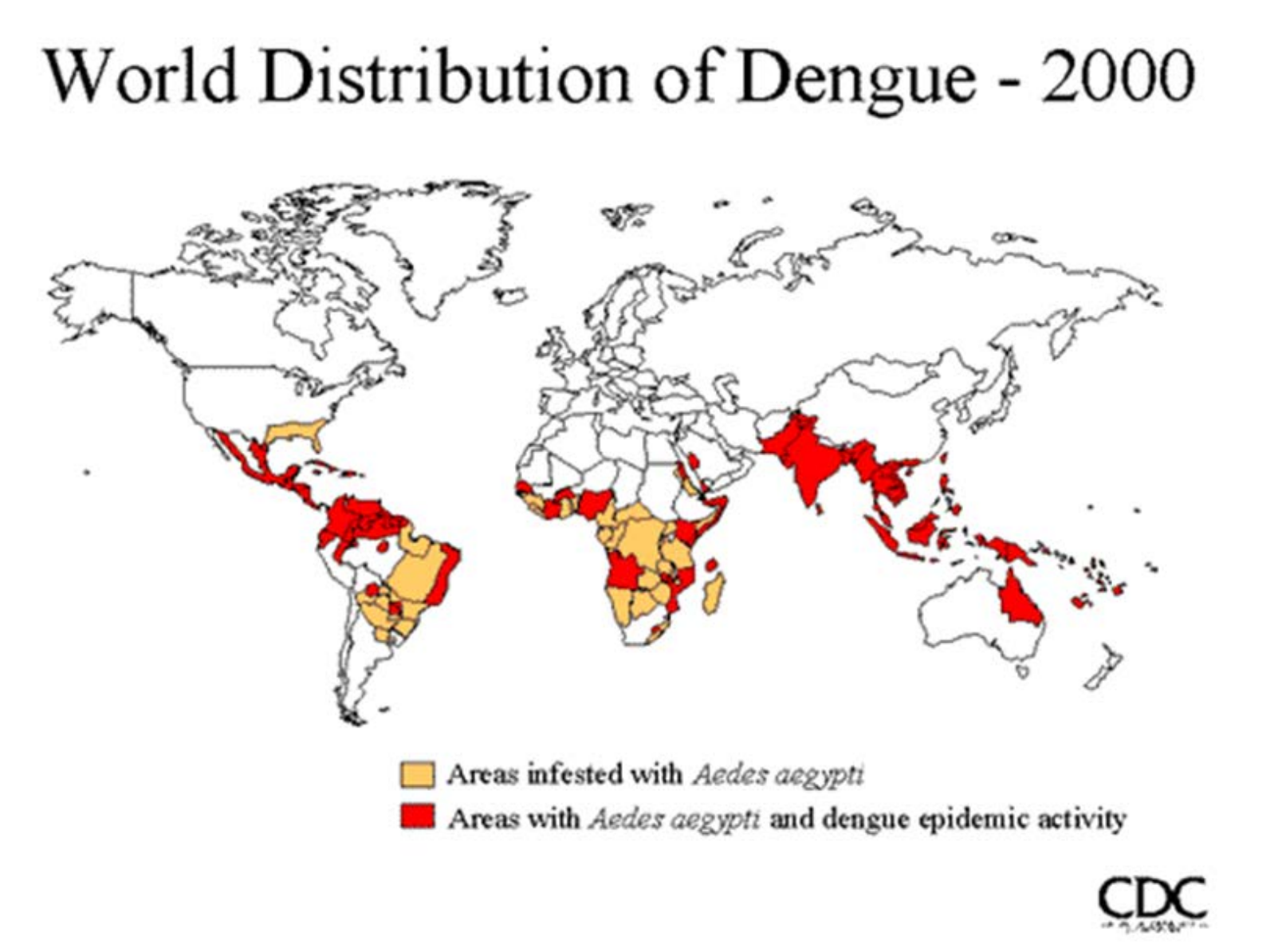
OPENING CREDITS

- **Landscape epidemiology** – combines disease ecology and landscape ecology
- **Vector-borne diseases** (VBD) - such as Dengue Fever (DF) have something, in this case a mosquito, that acts in distributing the disease.



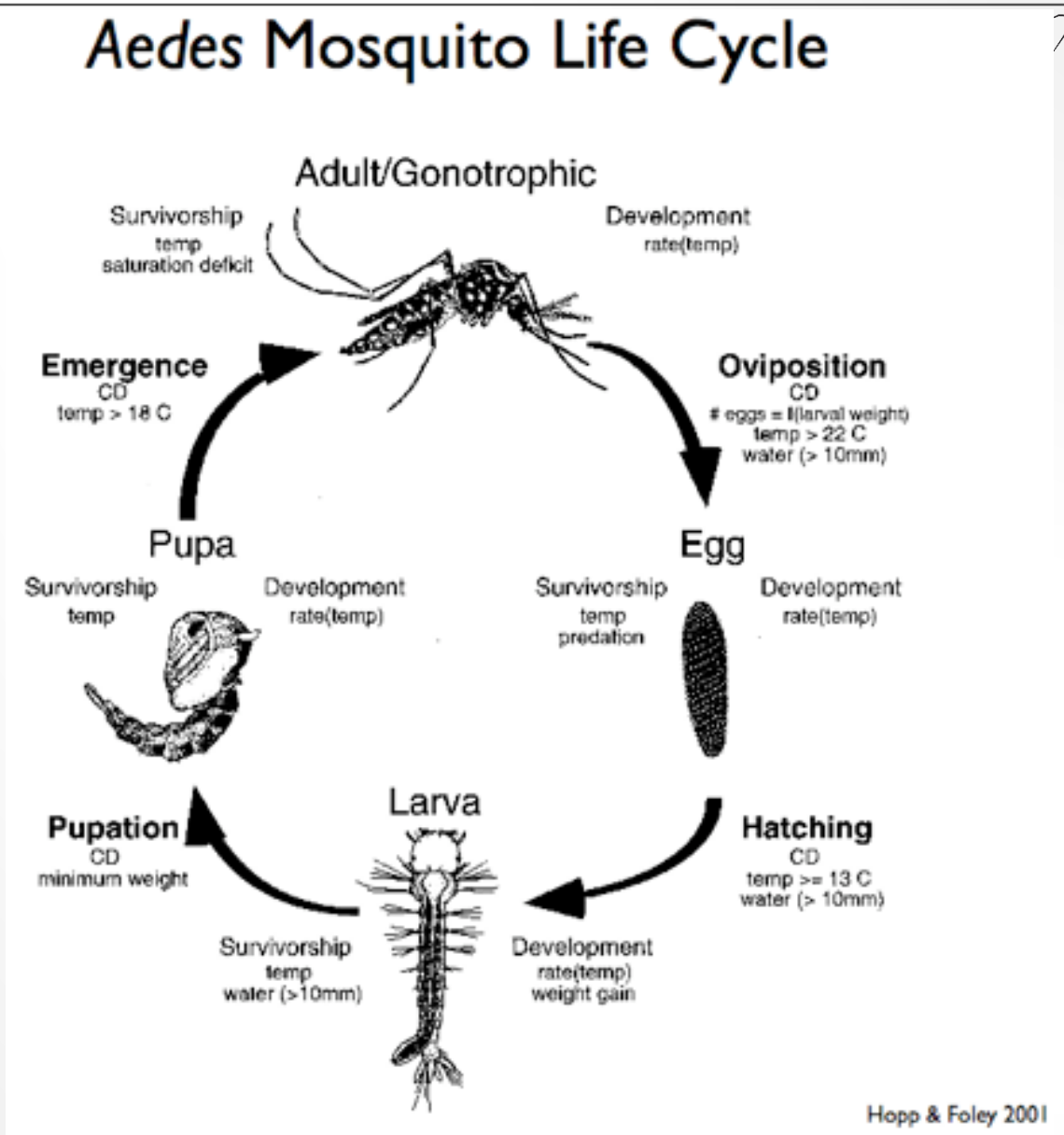
OPENING CREDITS

- **Dengue Fever** – is one of the leading causes of illness and death in tropical regions of the world.
- **One-third of the world's population** – live in areas of risk, with nearly 400 million becoming infected each year
- **Category A** - the U.S. Department of Health and Human Services highest risk category, and the category for Dengue Fever



THE VILLAIN

- The female *Ae. aegypti* mosquito
- Feeds actively during the day
- Closely related to the *Ae. albopictus* mosquito (found in Indiana)
- Highly domesticated
- Strongly anthropophillic

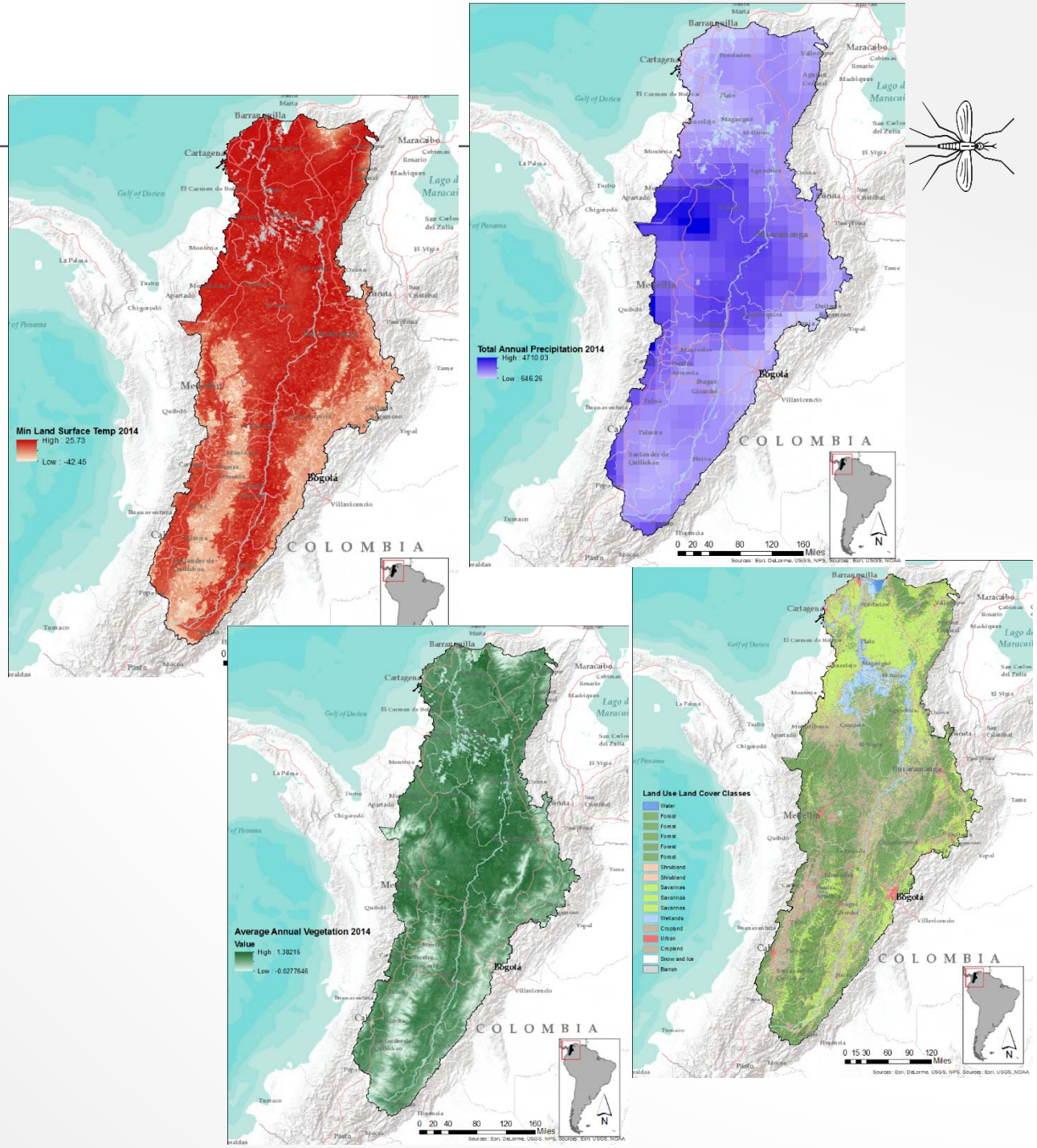


THE SETTING

- **24,949 cases** - of DF in 2014 in the Magdalena River watershed [Instituto Nacional de Salud, INS (National Institute of Health)]
- **Case rate** - 6.9 cases per 10,000 population
- **Natural separation** - from other geographical regions in the country
- **Climatic conditions** - vary widely from seashore to mountains
- **Main urban centers** - 80% of the country's population lives within the watershed

SUPPORTING CAST

- Data acquired from the NASA and USGS data servers directly
- Moderate Resolution Imaging Spectroradiometer (**MODIS**) satellite products
  - MYD11A1 product at 1km resolution – land surface temperature
  - MYD09GQ product at 250m resolution – vegetation index
  - MCD12Q1 product at 500m resolution – land use land cover
- Tropical Rainfall Measuring Mission (**TRMM**) – precipitation



SUPPORTING CAST

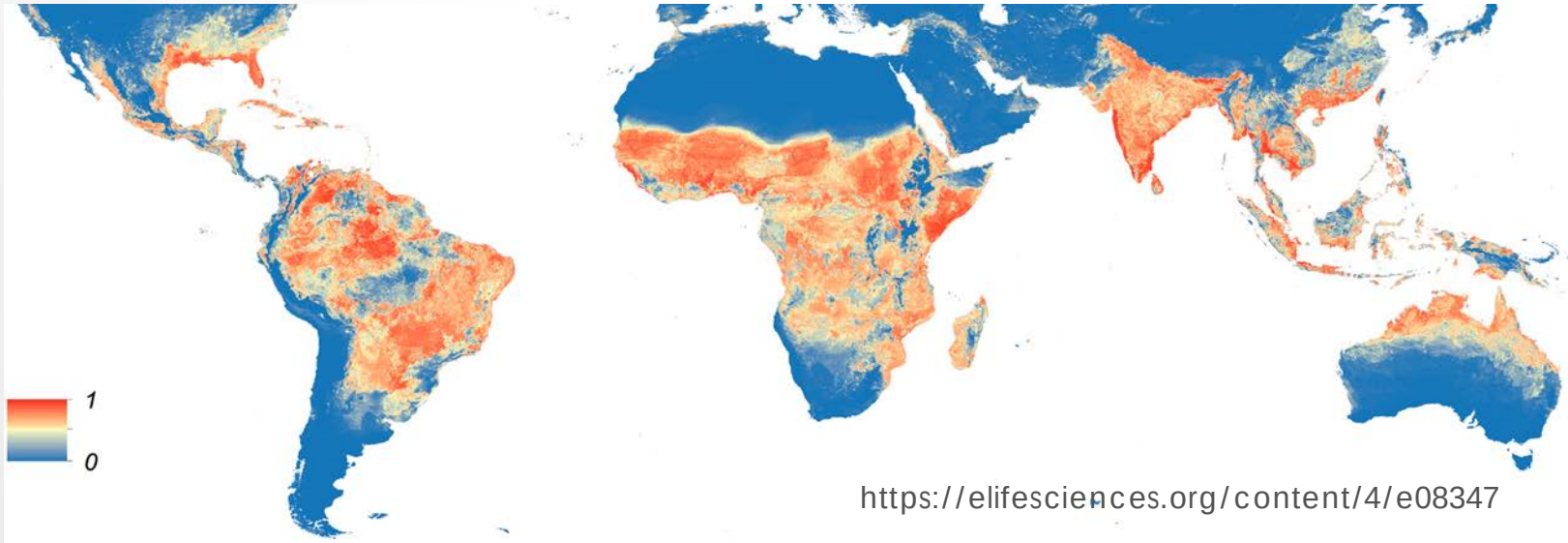
- Shuttle Radar Topography Mission (SRTM) – elevation at 30m resolution
- Dengue Fever cases were collected, processed and confirmed by the Instituto Nacional de Salud of Colombia according to its protocols and downloaded from the INS site
- Environmental variables were all composited weekly using the ISO Week Date format to reduce effect of clouds
- Above composites were aggregated using the ArcGIS 10.3 Spatial Analyst tool - Cell Statistics
  - Created annual minimum, maximum, and mean for each raster cell
- ArcGIS 10.3 Spatial Analyst tool - Zonal Statistics as Table
  - Created annual minimum, maximum, and mean values for each municipality

THE HERO

- Boosted Regression Tree (BRT) models utilize an initial regression tree and then improve upon it repetitively to minimize unexplained variation at each stage

Used in many other areas, including;

- predicting forest productivity
- predicting the properties of wood composites
- analysis of coral reefs
- remotely sensed imagery classification
- predicting crop disease outbreaks
- previous studies with Dengue Fever and other vector-borne diseases
- the habitat of the villain herself

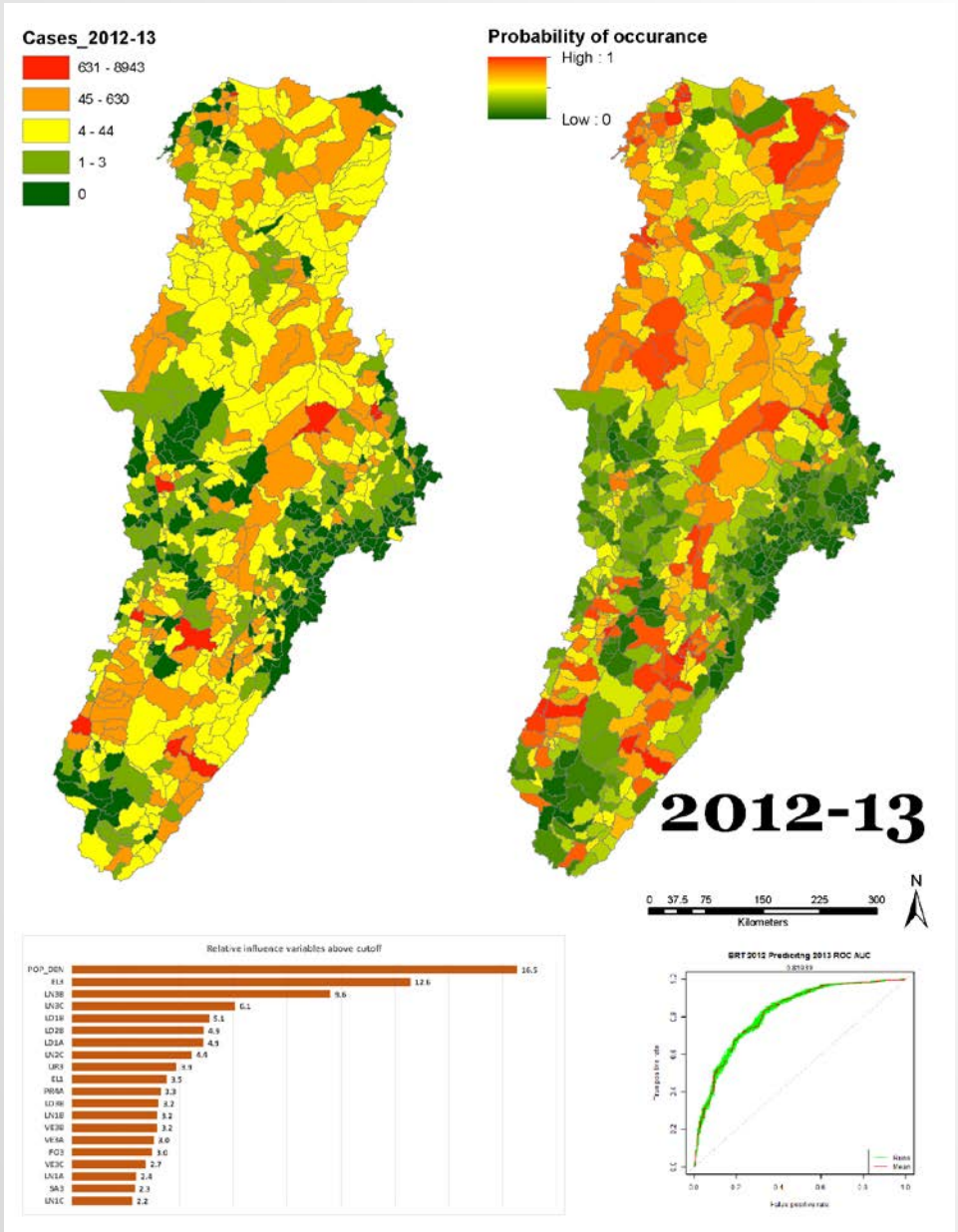


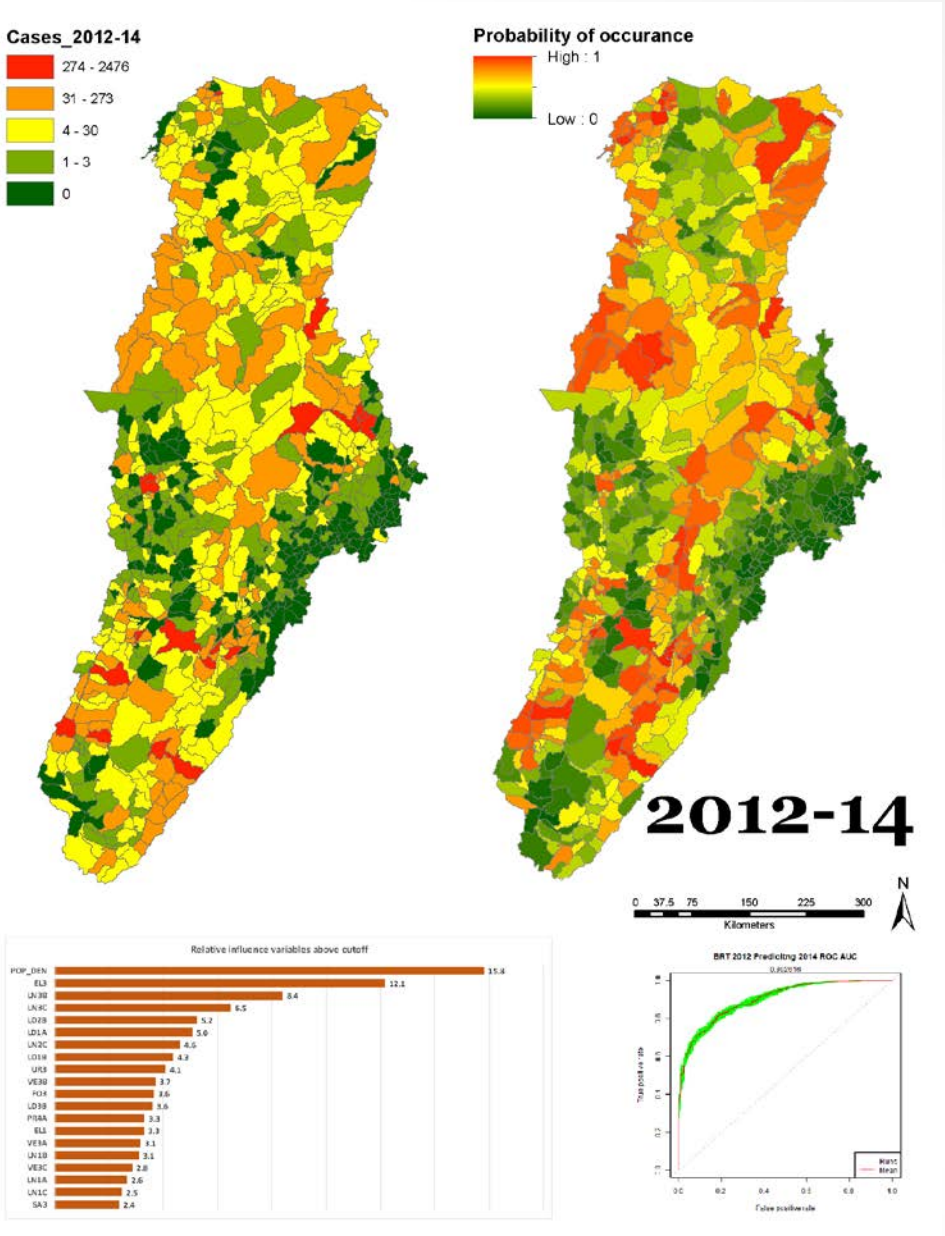
https://elifesciences.org/content/4/e08347

THE HERO

- Any municipality reporting one or more cases of Dengue Fever in a year was considered “presence”, while all others were considered absence of Dengue Fever
- Statistical program R with the gbm package was used to run the model
- Model was run 100 times and the mean for all 100 runs was calculated
- Map created showing a probability of occurrence using a graduated index between 0 and 1
- Model performance was calculated using the Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) between the mean prediction value and the presence/absence

ROLL THE CREDITS





ROLL THE CREDITS

- This study was supported by the IUPUI Chancellor for Research – Research Support Funds Grant.
- The study was also possible because of the accessibility to public health data made available by the National Institute of Health (Instituto Nacional de Salud, INS) of Colombia.
- The authors would like to thank Mary Lefevers for her assistance in the data preprocessing and Victor E. Casallas-Bedoya from INS for his assistance with data information.
- Thank you to the State of Indiana and all of the staff and volunteers that made this event possible

QUESTIONS

Contact information:
Jeffrey Ashby
jlashby@iupui.edu