**P-2 Modeling and Analytics (M&A) goals for P-2.**

**P2-wide M&A Team:**

* Peter Daszak (EHA) Lead
* Kevin Olival (EHA) M&A Coordinator
* Christine Kreuder-Johnson (CKJ) UC Davis PoC
* Damien Joly – Metabiota PoC

**EHA M&A Team:**

* Peter Daszak (EHA) Lead
* Kevin Olival (EHA) M&A Coordinator
* Carlos Zambrana-Torellio
* Toph Allen
* Andrew Huff
* Lizzie Loh (Behav. risk analysis)
* Parviez Hosseini
* Emily Hagan
* TBD (Kris Murray’s replacement)

**General Management M&A P-2**

* Weekly EHA M&A meetings; regularly liaise with EHA Behavioral Risk Group
* Monthly P2 M&A meetings
* Kevin to sit in on Surveillance calls to represent M&A team.
* Aim to deliver high impact 2-pagers (short, important issues) to USAID every 3-4 weeks. Especially important on topics that Dennis, Andrew, Alisa are thinking through for P2 directions.

**Primary Questions:**

1. What are the BIG goals/projects we want to accomplish by the end of P-2 (like Deep Forest and Hotspots II for P-1)?
2. What data and types of analyses do we need to reach those goals?
3. How should these data be organized to be most useful for analyses (linking with IM team and EIDITH)?
4. How will we best integrate analyses of human behavioral data with wildlife and livestock data?
5. Best ways for M&A to inform surveillance? E.g. What data to collect, number of sites and samples necessary, and selection of pathways in order to test global hypothesis on zoonotic disease emergence.

**Overview: Big, new ideas for P-2**

* **Future projections of hotspot maps** – using projected changes in demography, climate, livestock, and trade.
* **Livestock and Antibiotic-resistant EID hotspots**
  + Do antimicrobial resistant (AMR) EIDs correlate with antibiotic use?
  + Is this driven by veterinary or human use of antibiotics? Look at spatial patterns and evolutionary drivers with veterinary and human use and livestock movement, using global data on antibiotics that we now have.
  + African livestock futures – analyze zoonotic EID risk; create similar analysis for other regions e.g. Asian livestock futures
* **Sicki GRID** – better response variable for hotspot models
  + all outbreaks of an emerging disease if they’re separate events (not just first emergence event, and not just spreading events)
  + Linking in with DoD Biosurveillance Ecosystem (BSVE) and other global datasets and data mining/data scraping activities
* **Modeling to inform Surveillance**
  + Use P-1 data and targeted P-2 hypotheses to help inform site selection and pathway selection
* **DEEP FOREST** – continued in P-2;
  + analysis of big datasets on host and viral communities across the landuse gradients
  + Kris Murray will stay involved for DF
* **What-if scenarios to model disease interventions and human behaviors**
  + Scenario models with regional data to test disease intervention strategies
  + Replacement scenarios; e.g. changing of protein sources and patterns of livestock consumption
  + Predicting likelihood of spread of new high priority PREDICT viruses
  + Spatial models of human behavior, can we map globally using proxy datasets?
* **Evolutionary models** -- starting with existing P-1 sequence data
  + Defining ‘epizones’
  + Recombination hotspots
  + Phylogeography of disease and hosts
  + Modeling of viral evolution with longitudinal datasets
  + Testing nearest neighbor and other phylogenetic hypotheses on determinants of spillover, including molecular mechanisms (e.g. receptors/binding domains) for certain pathogen groups – leverage EHA’s NIH CoV study in China.

**More Details – specific ideas from last EHA brainstorming**

**Future demographic, livestock, and climate changes and disease**

Projecting Hotspots forward 5, 10, and 20 years into the future using modeled changes in pop growth, climate, livestock production, and travel and trade expansion.

Peter on advisory board for US Global Change Research Program – opportunity to collaborate with others doing other climate change.

Carlos – group in Italy is projecting 5000+ mammals into future range. Potential collaboration with Carlo Rondenini, affiliated with FAO.

Need to combine climate data with livestock futures. Scenario based projections – because models aren’t necessarily giving us the picture 5, 10, 20 years into the future. Can predict climate, but harder to predict the human response and behaviors with climate change – that’s where scenario models come in.

Can we do Asian livestock futures? Split into SE Asia, S Asia, Australasia… etc.

What-if modeling of replace protein source, e.g. if you stop eating bats and replace with a livestock protein source, what does this do to risk? Andrew H. has experience with this.

Travel and Trade – What-if: e.g. Future Africa and China trade based on different projections

UN Comtrade data on Livestock trade. Need to be careful because of repackaging and ‘value’ added relabeling of country of origin.

Hotspots with livestock production

Trade movement. Poultry movement between China and Africa over time. Relate this with biosecurity practices by country for risk assessment.

Publicly available troop movement data by country as proxy for military movement. Possible to FOIA additional data.

Look at plant/grain travel also, e.g. did PEDV come in from China in contaminated grain shipments?

**Antibiotic resistance**

Recent paper with data on global antibiotic use.

Can use as proxy, country-level predictor variable – what are the laws on antibiotic use in different countries for humans and veterinary.

Global travel and trade and spread of antibiotic resistance, need to include livestock trade. UN COMTRADE data.

**Sicki GRID** – **better response variable for hotspot models**

Best response will be all outbreaks (not just first emergence event as in Jones et al. 2008), maybe do for subset of top (top 100?) pathogens, or those like Chikungunya, Ebola, Nipah, etc. where outbreaks can be tracked decently.

Leverage EHA’s DTRA work and BSVE to help populate response dataset.

Discuss future API for EIDITH to better link in analyses.

**Modeling to inform Surveillance**

Need to be more clear on what are our units of response, number of replicates, etc. for each gradient.

Should only have one pathway per country. Or one pathway at a given site. Maybe find a few double pathway sites in later years to tease out, focus on single pathways initially. Will allow us to tease out processes of disease emergence better.

What we ideally need is an experiment – e.g. one of Malaysia DF sites was chopped down.

RISK assessment

Does exposure to wildlife increase **risk**? Use human serology to test as a response variable (e.g. if spillover has happened). Predictors in addition to land use change and other variables = how frequent and how diverse the viruses are.

**What-if scenarios** for risk mitigation testing. Maureen role in testing risk mitigation at local scale. What are longterm impacts of different mitigations? E.g. draining swamps in south to reduce malaria risk – has negative ecological impacts.

What’s the consequence we’re trying to calculate risk for. Spillover only, DALYs, pandemic size, economic impact? For reemerging viruses may be easier to calculate; For emerging viruses can do different scenario testing – e.g. for a new Ebola what is the likelihood it will get into an urban area like W. Africa outbreak.

Andrew H and PRH, to collaborate on modeling interventions and policies we can include in what-if scenerios. What effect will different ‘policy cocktails’ have.

**Phylogenetic analysis in P-2**

Predicting viral pathogenicity is exceptionally hard and there are many preconceived notions about it

Test molecular evolution using time-sampled data (longitiudinal data) where available. Can automate and mine GenBank to further populate dataset and add to P-1 molecular data. Can get cell receptor data from Genbank etc.

Test nearest neighbor hypothesis

Can we look at seq we have now (P-1) and where they were collected in pathways – are ones from the same pathway more similar than those from different areas. Does diversity of virus genetics change across the landscape. Target surveillance where viral diversity is predicted to be the highest. From wild to markets, e.g. does diversity decrease across pathway bottleneck – and how does this related to host genetic diversity.

Predict recombination potential across the landscape. P-1 data by site and see how much viral genetic diversity there is.

Phylogenetic relatedness and receptor/host relatedness, 3 hypotheses to test: H0 – do they match; H1 – general decline with phylogenetic distance; H2 – islands of susceptibility scattered across phylogeny.

Test pathogeneicity, is it better than random between viruses of a given clade. E.g. CoVs and pathogenicity – can we predict this? Test a mixed model

**Defining “Epizones”**

Tentative Definition: “A region where a viral group that has or is likely to emerge (including an understanding of viral strain risk) is present in its reservoir, ***and*** where the human risk behaviors and drivers exist for the viral group to spillover and emerge.”

* Operational definition of ‘epizone’ from M&A
  + Specific geographic area where X number of people and X number of animals overlap –
  + e.g SARS – market from where first cases, plus pool where bats came from and wildlife range of those. Need surveillance data on how wide these ranges are – from surveillance team.
  + Are epizones across diseases similar? Do they correlate, or is each disease so unique that there is no strong overlap of epizones. Start with core e.g. areas where Nipah is emerging, and move out from there to see what extent of epizone.
  + Hazard vs Risk. May have wildlife reservoir but no human behavioral risk = hazard, w no risk.
  + Viral evolution plays in to identify likelihood of emergence – some clades may be more likely to emerge/jump than others. E.g. SARS-like viruses are found across Europe. Need to integrate with viral phylogenetic analyses.