Predicting primate sources of Zika virus spillover

Barbara A. Han[1], Subho Majumdar[2,3], A. Mojsilović[3], Kush Varshney[3]

[1]Cary Institute of Ecosystem Studies, [2]Univ. Florida Informatics Institute, [3]IBM Watson Research Lab

BACKGROUND

- Non-human primates maintain mosquito-borne flaviviruses in the wild (reservoirs).
- Successful control of Zika virus relies on spillover prevention from wild reservoirs[1].
- Primate diversity & abundance is exceptionally high in the neotropics, especially Brazil.
- Identifying potential primate reservoirs will require highly targeted surveillance.
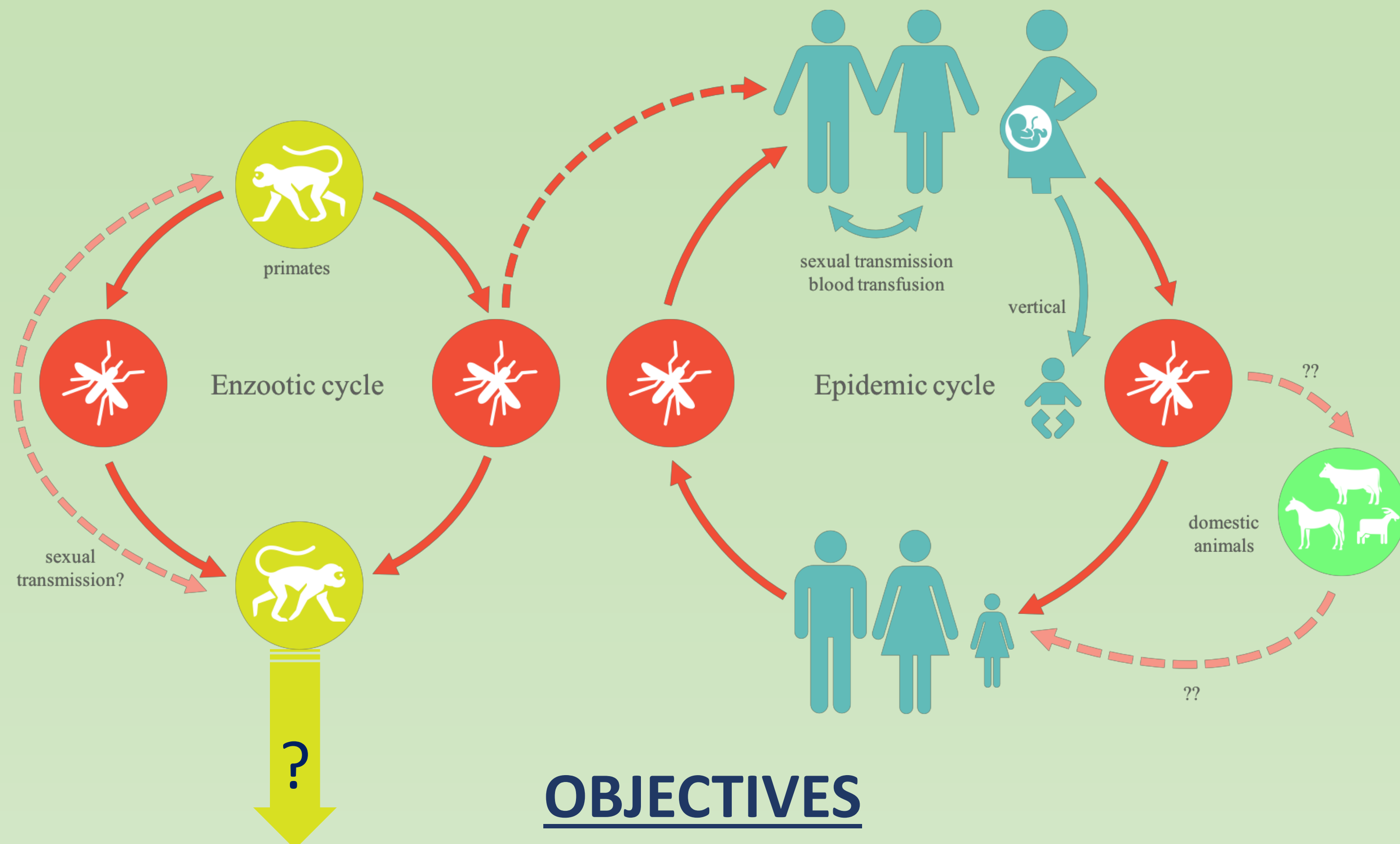
OBJECTIVES

- Which primate species are most likely to carry Zika virus in the Neotropics?
- Where are these species distributed?
- Where should human and non-human primate surveillance be focused?

METHODS

- **Collect primate trait data** to identify primate species with the highest probability of carrying ZIKV
  - **Why trait data?** Organismal trait data are useful because:
    - They are immediately available (unlike ZIKV surveillance data)
    - They reliably distinguish one host species from another
    - They underpin species' intrinsic capacity to harbor zoonoses[2-3]
- **Apply a machine learning method on imputed trait data** to recognize primate species resembling known flavivirus hosts to target ZIKV surveillance.

DATA

RESPONSE VARIABLE: Host status for ZIKV & other mosquito-borne flaviviruses (YFV, JEV, SLEV, DENV, WNV) for all primates (N=376); 20 positive species in 376 x 6 matrix; only 2 ZIKV+ species

PROBLEM: POSITIVE SAMPLES VERY RARE

SOLUTION:

BAYESIAN MULTI-LABEL MACHINE LEARNING[6]

Simultaneously models reservoir status for all 7 mosquito-borne flaviviruses across all primates by considering their species-level features as predictors

PREDICTOR VARIABLES: 50 organismal traits (e.g., body mass, longevity, metabolic rate, geographic range area, population density; compiled from [4-5])
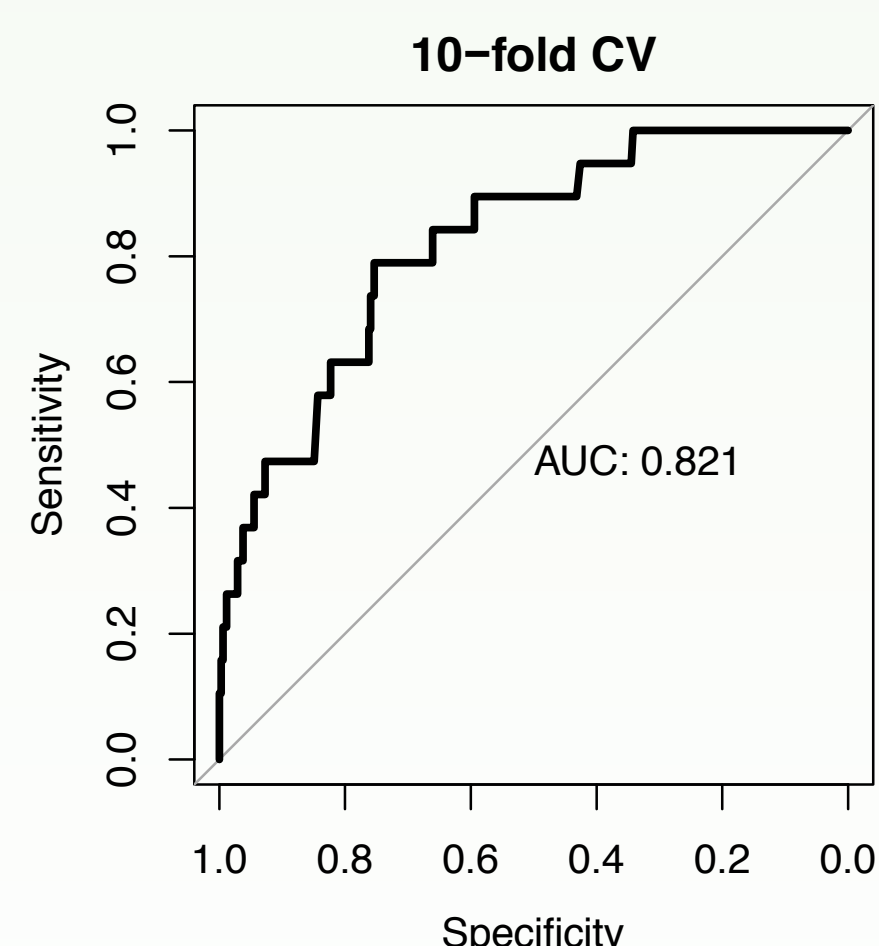
PROBLEM: INCOMPLETE TRAITS FOR MANY SPECIES

SOLUTION:

MULTIPLY CHAINED IMPUTATION[7]

MICE method predicts missing entries by iteratively leveraging information available across the other variables. Imputation process is repeated until the entries across a number of imputed datasets reach a stable distribution

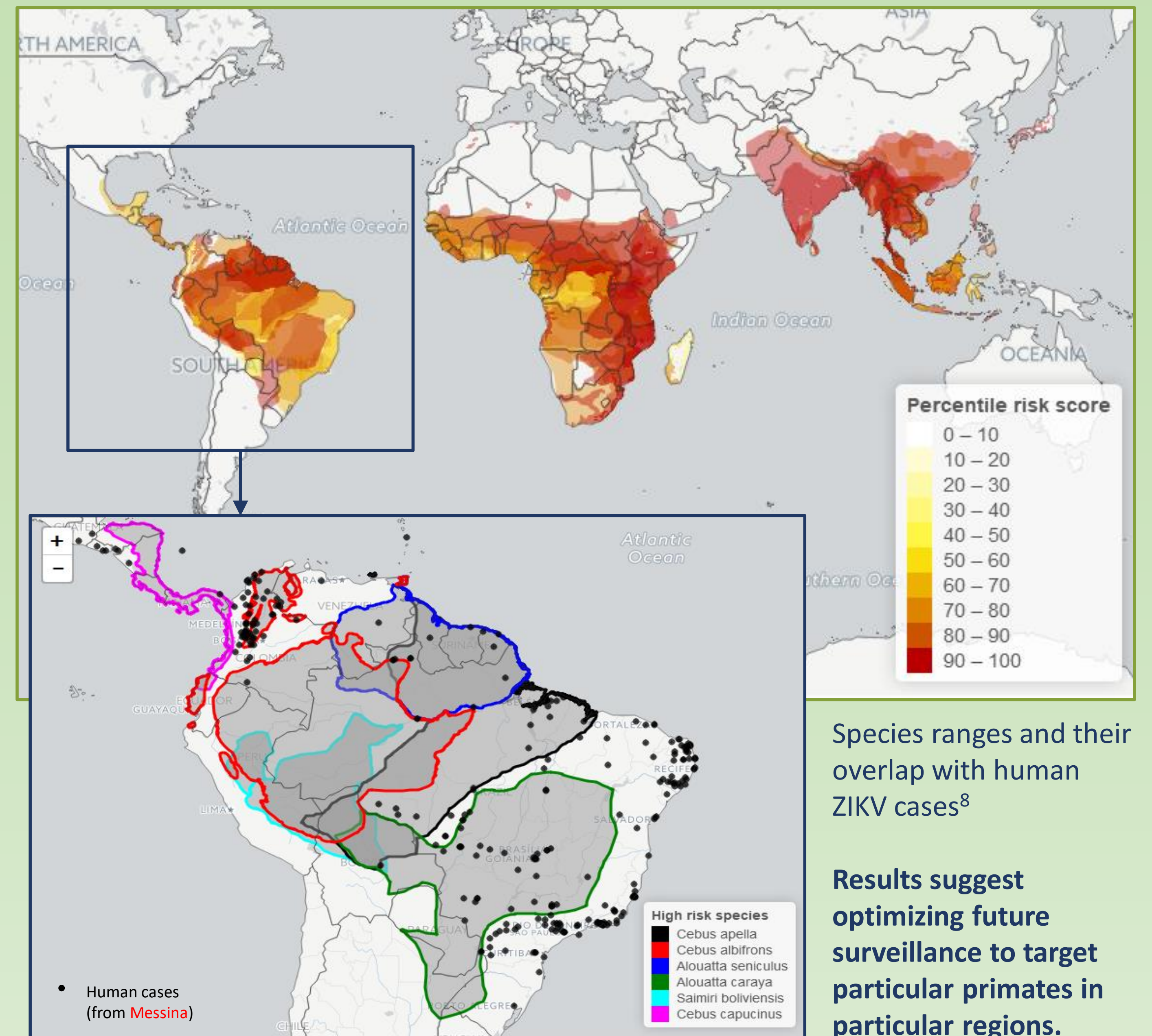## MODEL PERFORMANCE

10-fold CV

AUC: 0.821

- Model classified flavivirus+ primates with 82% accuracy
- Model identified particular primate species as likely ZIKV hosts
- Model identified primates as likely hosts for other flaviviruses

REFERENCES

1. Althouse BM, Vasilakis N, Sall AA, Diallo M, Weaver SC, Hanley KA. Potential for Zika virus to establish a sylvatic transmission cycle in the Americas. PLoS Negl Trop Dis 2016; 10: e0005055.
2. Han BA, Schmidt JP, Bowden SE, Drake JM. Rodent reservoirs of future zoonotic diseases. Proc Natl Acad Sci USA 2015; 112: 7039–44.
3. Han BA, Schmidt JP, Alexander LW, Bowden SE, Hayman DTS, Drake JM. Undiscovered bat hosts of filoviruses. PLoS Negl Trop Dis 2016; 10: e0004815.
4. Jones KE, Bielby J, Cardillo M, et al. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology 09/2009; 90: 2648–2648.
5. Wilman H, Belmaker J, Simpson J, de la Rosa C, Rivadeneira MM, Jetz W. EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. Ecology 2014; 95: 2027–2027.
6. Rai P, Hu C, Henao R, Carin L. Large-scale Bayesian multi-label learning via topic-based label embeddings. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, eds. Advances in Neural Information Processing Systems 28. Curran Associates, Inc., 2015: 3222–30.
7. Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. Surv Methodol 2001; 27: 85–96.
8. Messina JP, Kraemer MU, Brady OJ, et al. Mapping global environmental suitability for Zika virus. Elife 2016; 5: e15272.
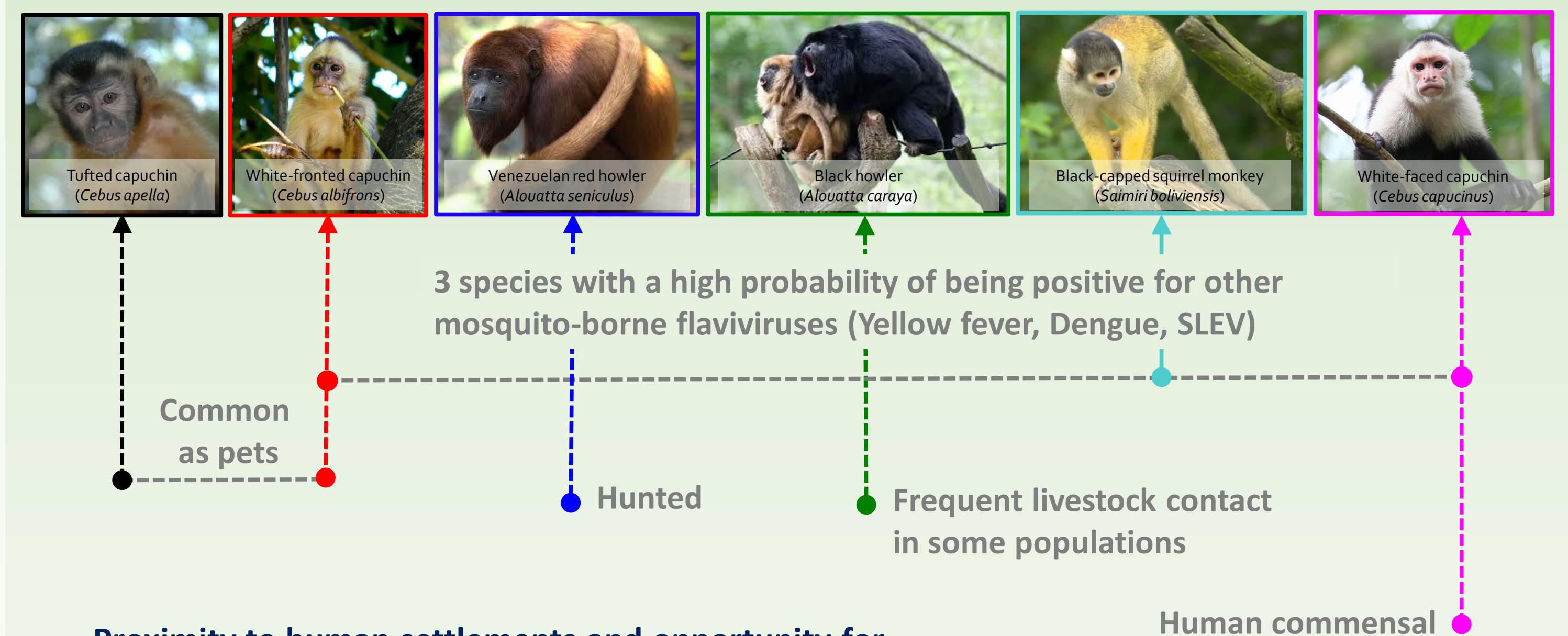
RESULTS & CONCLUSIONS

A map depicting overlapping species ranges of global primates and their probability of being ZIKV+.

Percentile risk score
- 0 – 10
- 10 – 20
- 20 – 30
- 30 – 40
- 40 – 50
- 50 – 60
- 60 – 70
- 70 – 80
- 80 – 90
- 90 – 100

Human cases (from Messina)

High risk species
- Cebus apella
- Cebus albifrons
- Alouatta seniculus
- Alouatta caraya
- Saimiri boliviensis
- Cebus capucinus

Species ranges and their overlap with human ZIKV cases[8]

Results suggest optimizing future surveillance to target particular primates in particular regions.

6 primate species in Central and South America in the 90th percentile probability of being ZIKV+

Tufted capuchin (Cebus apella) · White-fronted capuchin (Cebus albifrons) · Venezuelan red howler (Alouatta seniculus) · Black howler (Alouatta caraya) · Black-capped squirrel monkey (Saimiri boliviensis) · White-faced capuchin (Cebus capucinus)

3 species with a high probability of being positive for other mosquito-borne flaviviruses (Yellow fever, Dengue, SLEV)

Common as pets

Hunted

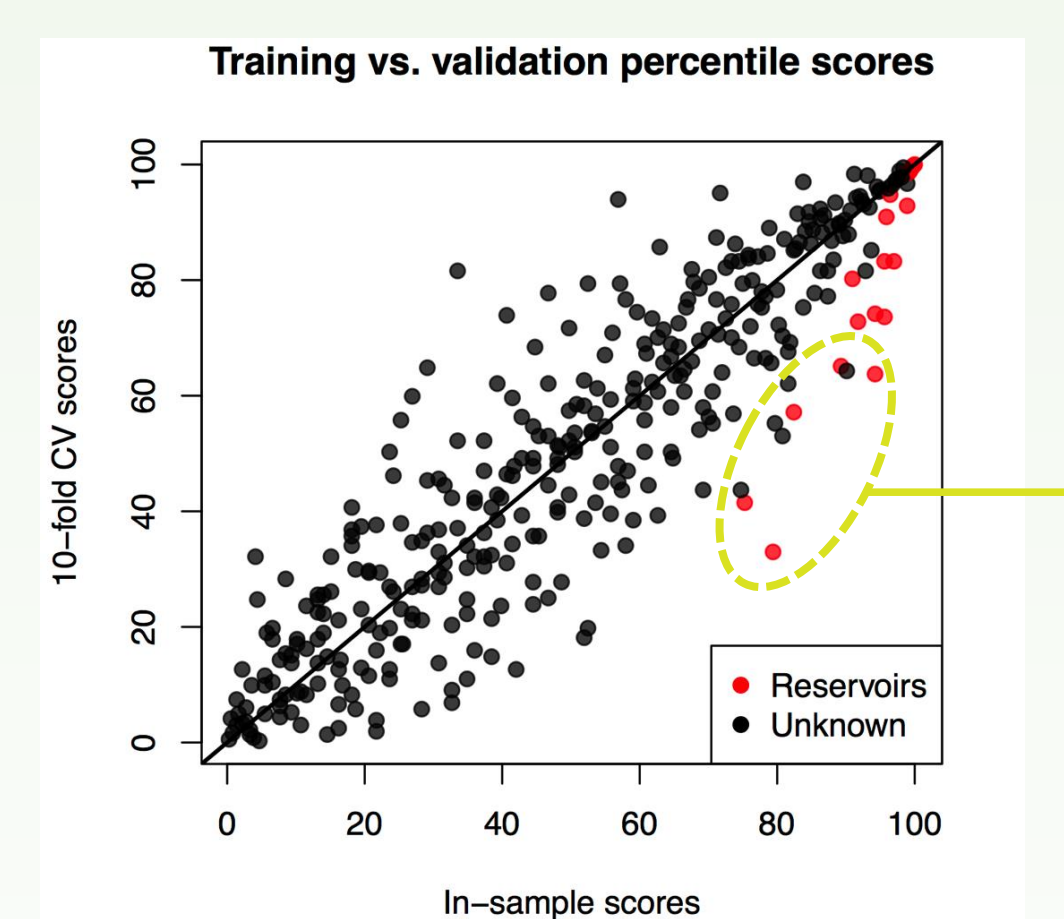Frequent livestock contact in some populations

Human commensal

Proximity to human settlements and opportunity for frequent human contact suggests prioritizing these species for ZIKV surveillance in Central America.

MODEL VALIDATION

Out-of-sample validation: Reassigned virus-positive primates to negative, and re-ran the model to obtain risk scores.

- Model assigned high risk scores to majority of known primate hosts.
- Model assigned low risk of flavivirus positivity to some known primate hosts.
- Data scarcity, even for very common species, may preclude our capacity to predict flavivirus reservoirs.

Training vs. validation percentile scores
- Reservoirs
- Unknown

In-sample scores · 10-fold CV scores

Black howler (Alouatta caraya) · Midas tamarin (Saguinus midas) · Ring-tailed lemur (Lemur catta) · White-faced saki (Pithecia pithecia) · African green (Chlorocebus aethiops)

ACKNOWLEDGEMENTS

IBM Social Good Fellowship