Can we predict the presence of pathogenic bacteria in soils?

Kathryn Atherton[1,2,3], Colin Averill[3,4], Kathryn Wheeler[4], John Foster[4], Michael Dietze[4], Jennifer Bhatnagar[3]

[1]Agricultural and Biological Engineering, Purdue University, [2]Boston University Bioinformatics BRITE REU Program, Summer 2018, [3]Department of Biology, Boston University, [4]Department of Earth and Environment, Boston University

Abstract

Soils are reservoirs for pathogenic bacteria such as *Klebsiella*, *Mycobacterium*, and *Staphylococcus*. Different taxa of bacteria thrive under different environmental conditions such as nutrient availability, temperature, and moisture. It is important to be able to predict the presence of these bacteria in the environment to better prevent infection. We tested the hypothesis that we can use environmental data to predict the abundance of pathogenic bacteria in soils. To do this, we used previously published data on the relative abundance of bacterial taxa from a global soil microbiome study and fit regression models to the data with various combinations of soil properties and climate properties. We found that climate properties are the better predictors for the presence of pathogenic bacteria. In the future, this data will assist us in forecasting the changes in these opportunistic taxa across space and time with environmental change.

Research Questions

1. Are soil properties a better predictor of overall pathogen abundance than climate properties?
2. Can we predict the relative abundance of individual genera that we know track certain environmental properties?

Methods

- Previously published soil sample data[1]: 16S rDNA sequence data (Illumina) and environmental data; 4 sites (Figure 1)
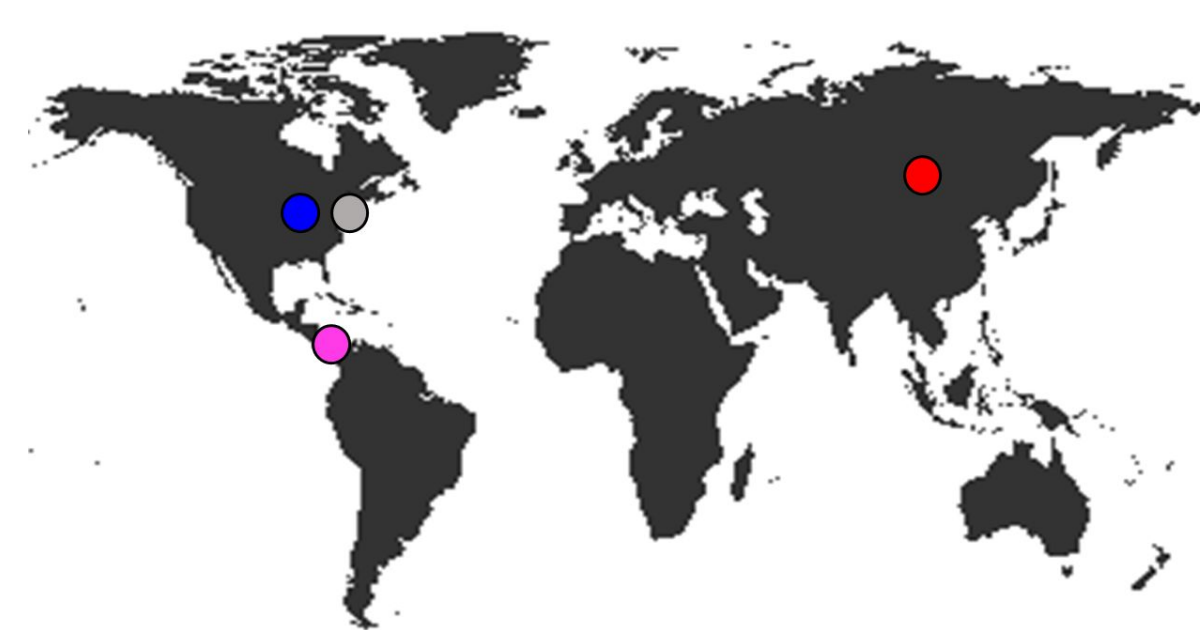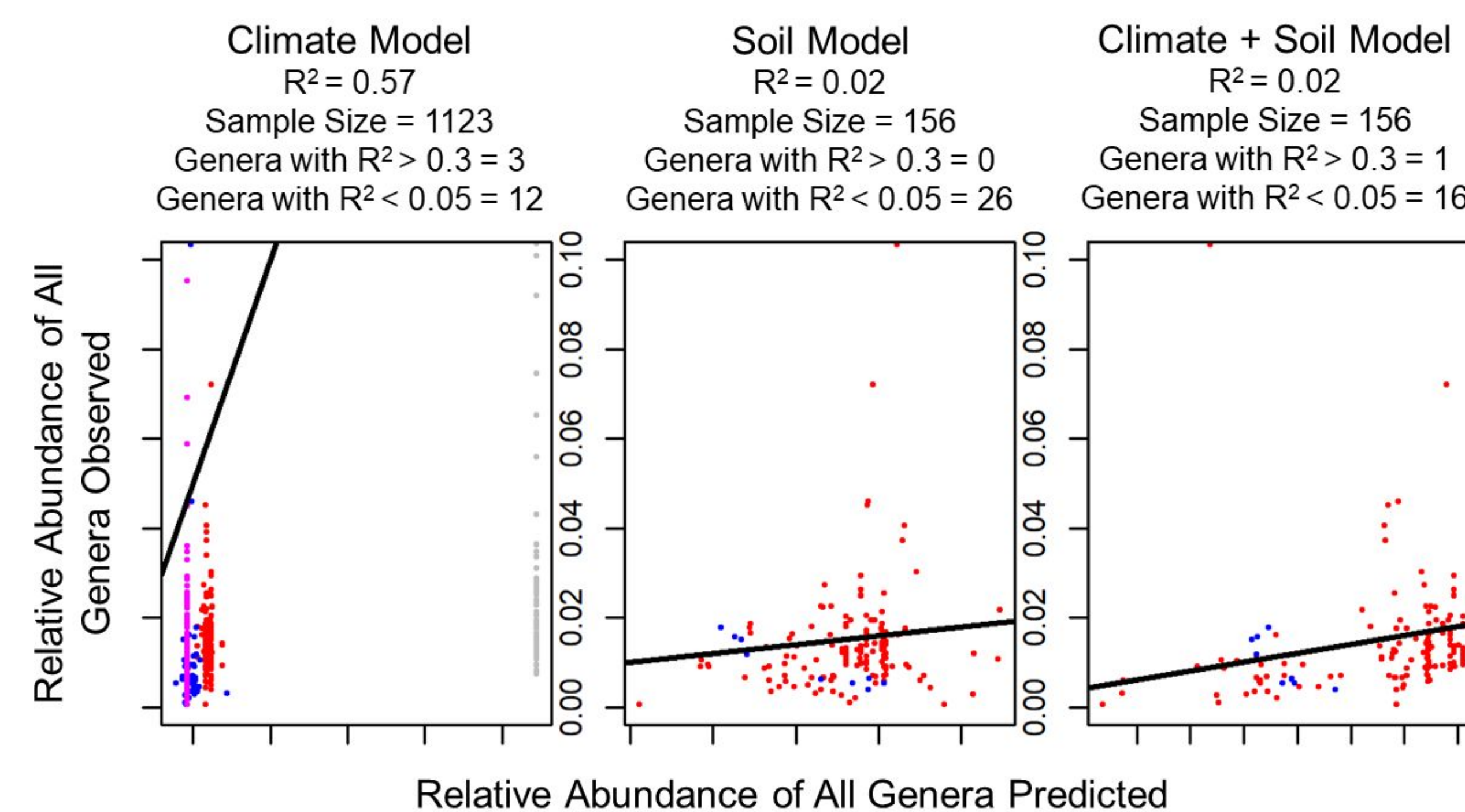

Figure 1: Locations of sample sites analyzed. The colored dots in this figure correspond to the locations of each sample site in Figures 2 and 4.

- Made exact sequence variants tables (count sequences that are exactly the same in each sample)
- Assigned taxonomy with GreenGenes database, aggregated taxonomy counts by genus, found each genus' relative abundance in each sample
- Subset for medically-relevant genera (n = 28)
- Created models with beta regression incorporating climate variables (taken from the WorldClim database), soil variables (from the studies), and both climate and soil variables
  - climate: mean annual temperature/precipitation[2], variance of mean annual temperature/ precipitation, mean diurnal range (mean of monthly (max temp - min temp))
  - soil: pH[2], percent carbon[2], percent nitrogen[3]
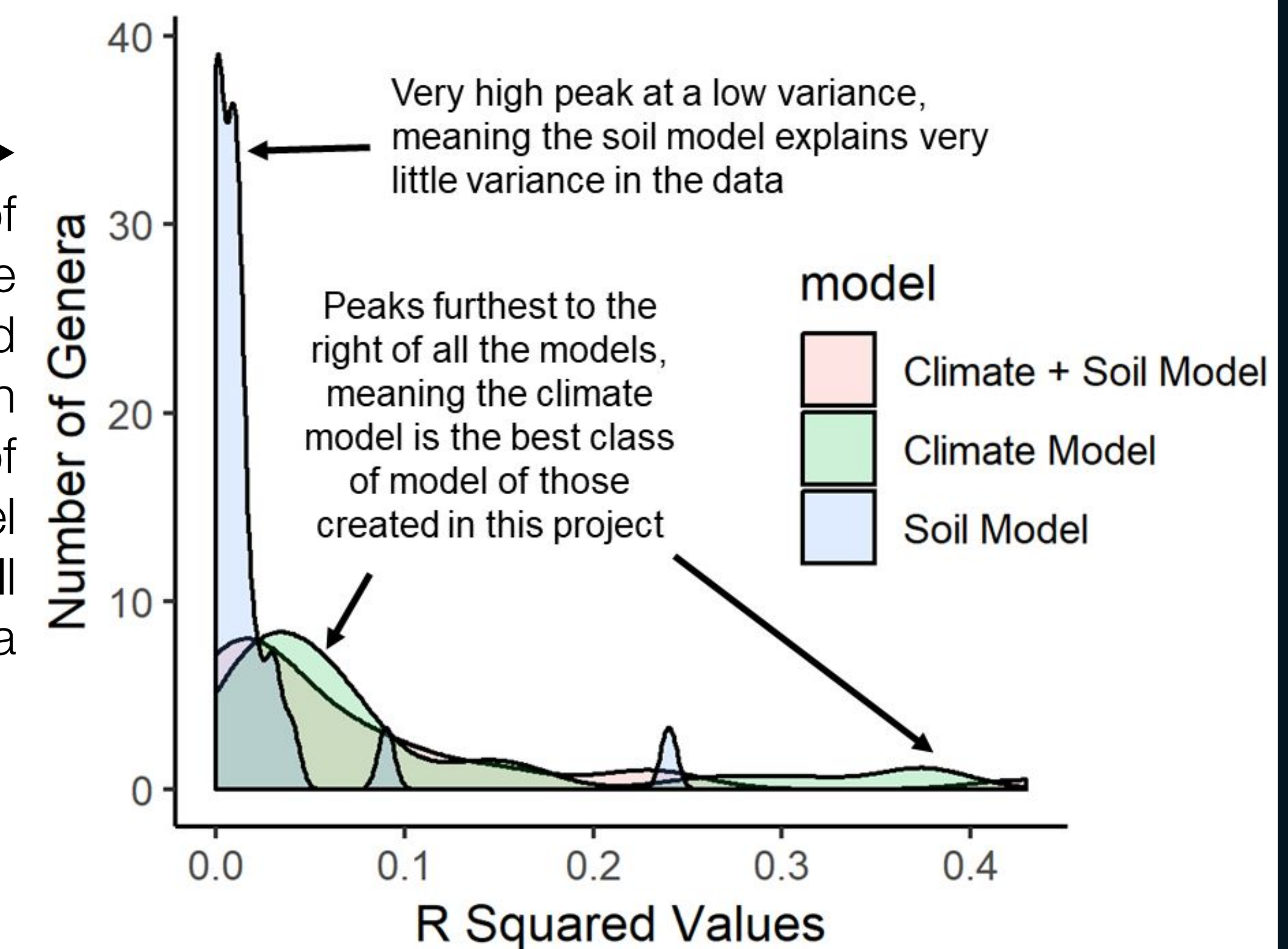- Used R^2 values to determine fit of models; scores above 0.3 are acceptable

Results

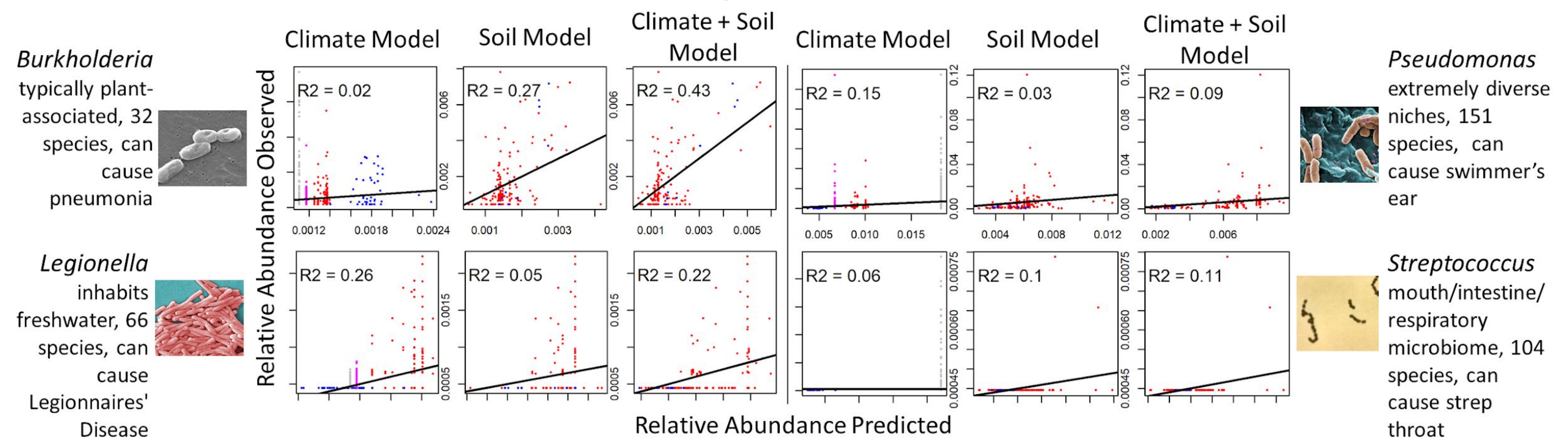**Research Question 1: Climate properties are better predictors of pathogen abundance than soil properties.**

Climate Model
R^2 = 0.57
Sample Size = 1123
Genera with R^2 > 0.3 = 3
Genera with R^2 < 0.05 = 12

Soil Model
R^2 = 0.02
Sample Size = 156
Genera with R^2 > 0.3 = 0
Genera with R^2 < 0.05 = 26

Climate + Soil Model
R^2 = 0.02
Sample Size = 156
Genera with R^2 > 0.3 = 1
Genera with R^2 < 0.05 = 16

Figure 3 ▶ Summary of variance explained by each class of model across all genera

Very high peak at a low variance, meaning the soil model explains very little variance in the data

Peaks furthest to the right of all the models, meaning the climate model is the best class of model of those created in this project

model
- Climate + Soil Model
- Climate Model
- Soil Model

▲ **Figure 2**: Relative abundance across all samples for all aggregated genera (n = 28) observed vs. predicted for each class of model

Research Question 2: Predictive ability for genera depends on breadth of environmental niche.

Burkholderia typically plant-associated, 32 species, can cause pneumonia

Legionella inhabits freshwater, 66 species, can cause Legionnaires' Disease

Climate Model $R2 = 0.02$
Soil Model $R2 = 0.27$
Climate + Soil Model $R2 = 0.43$

Climate Model $R2 = 0.26$
Soil Model $R2 = 0.05$
Climate + Soil Model $R2 = 0.22$

Climate Model $R2 = 0.15$
Soil Model $R2 = 0.03$
Climate + Soil Model $R2 = 0.09$

Climate Model $R2 = 0.06$
Soil Model $R2 = 0.1$
Climate + Soil Model $R2 = 0.11$

Pseudomonas extremely diverse niches, 151 species, can cause swimmer's ear

Streptococcus mouth/intestine/ respiratory microbiome, 104 species, can cause strep throat

▲ **Figure 4**: Examples of individual pathogens that the models predicted well (left) and poorly (right) and information about the pathogens

Discussion and Conclusions

- For all pathogens, the model using only climate variables most accurately predicts pathogen abundance in the environment, meaning we will be able to predict changes in pathogen abundance due to climate change.
- Pathogens that have specific niches in their environments were generally more accurately predicted than more diverse genera. To improve the models, we need more information about the environmental properties that specific pathogens within diverse genera are likely to track.
- We will continue to develop these models to forecast changes in pathogenic bacteria across space and time with environmental changes.

Acknowledgements

[1]Thompson, L.R. et. al. (November 2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*. 551 (23 November 2017), 457-463.
[2]Lauber, C.L. et. al. (August 2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology*.75 (15), 5111-5120.
[3]Ward, B.B. et. al. (1995). Functional and Taxonomic probes for bacteria in the nitrogen cycle. Ed. I. Joint, NATO workshop on *Molecular Ecology of Aquatic Microbes*, pp. 73-86.