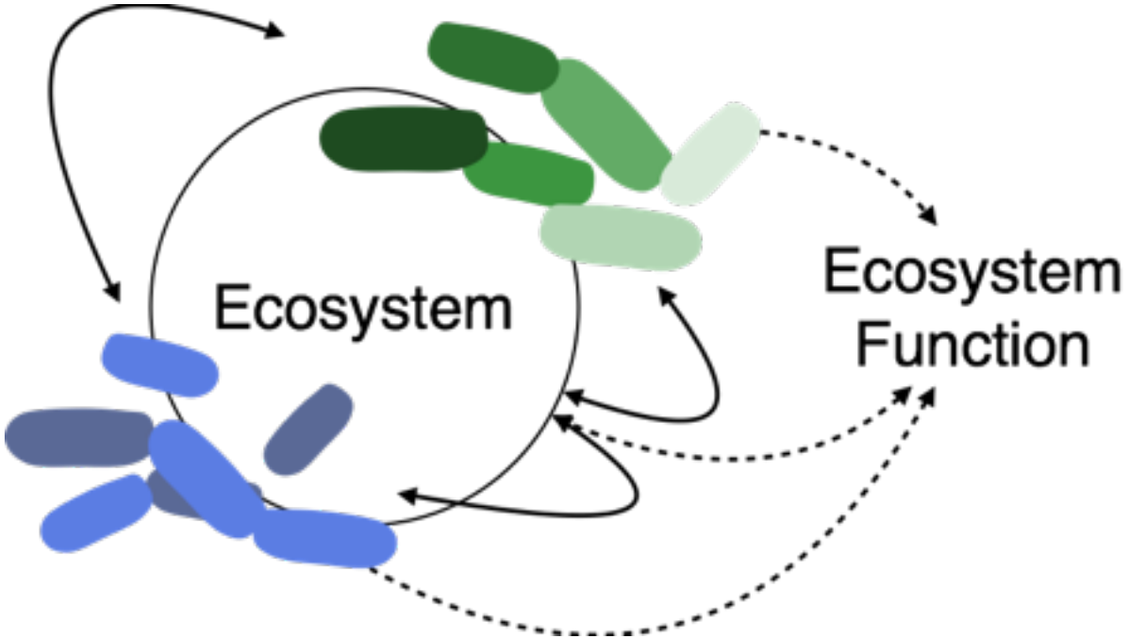
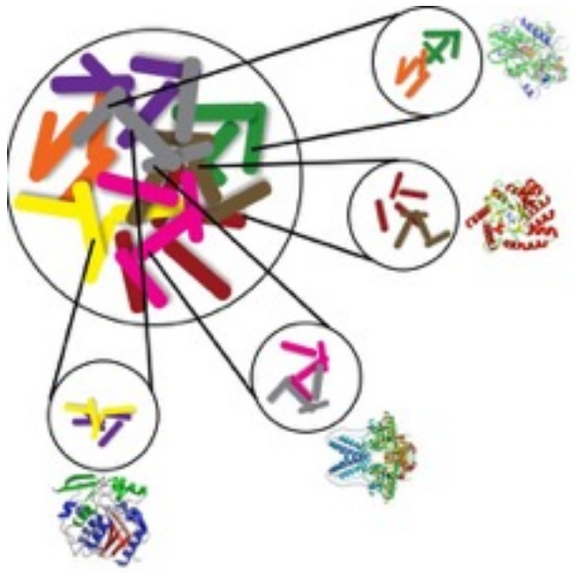
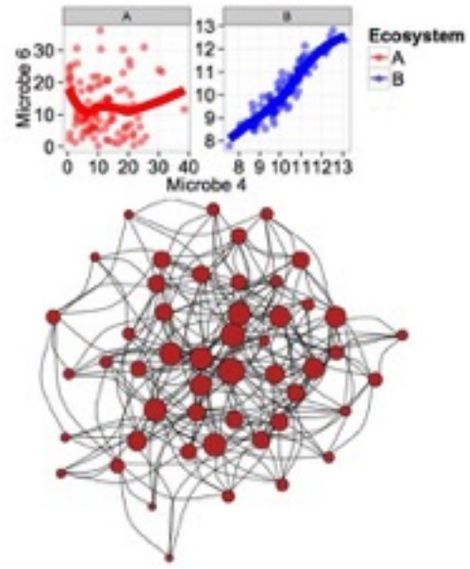






Ryan Williams
Post(Ghost)-doctoral Researcher
Agricultural and Biosystems Engineering
Iowa State University

Modeling and Simulating Data for a Better Understanding of Microbial Ecology

Ryan J. Williams

What is a model?

- Conceptualization of how we understand phenomena in nature.

- All models are wrong but are based on the best information we have.

- Can be a cartoon

- Can be mathematical

- Can be statistical

What is a model?

- Conceptualization of how we understand phenomena in nature.

- All models are wrong but are based on the best information we have.

- Can be a cartoon

- Can be mathematical

- Can be statistical

What is a model?

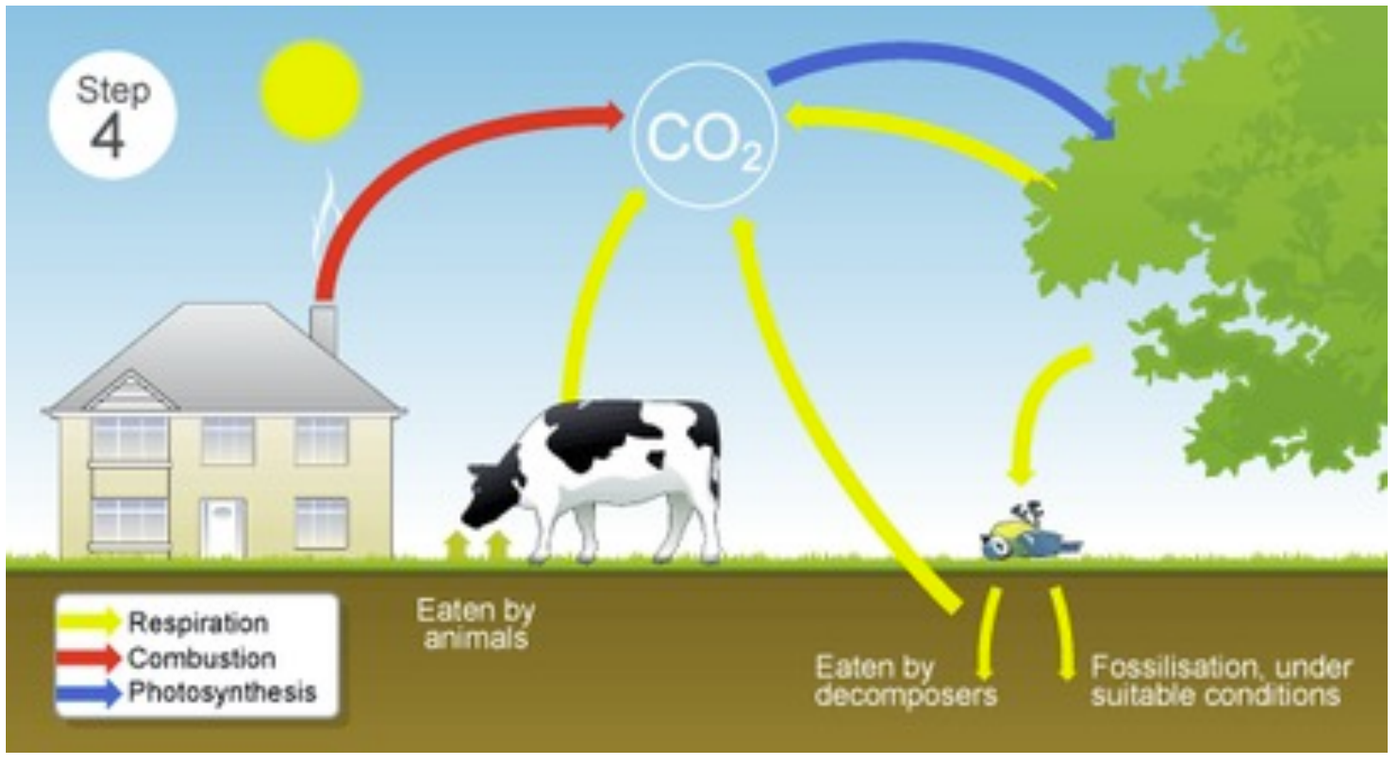
- Conceptualization of how we understand phenomena in nature.

- All models are wrong but are based on the best information we have.

- Can be a cartoon

- Can be mathematical

- Can be statistical



What is a model?

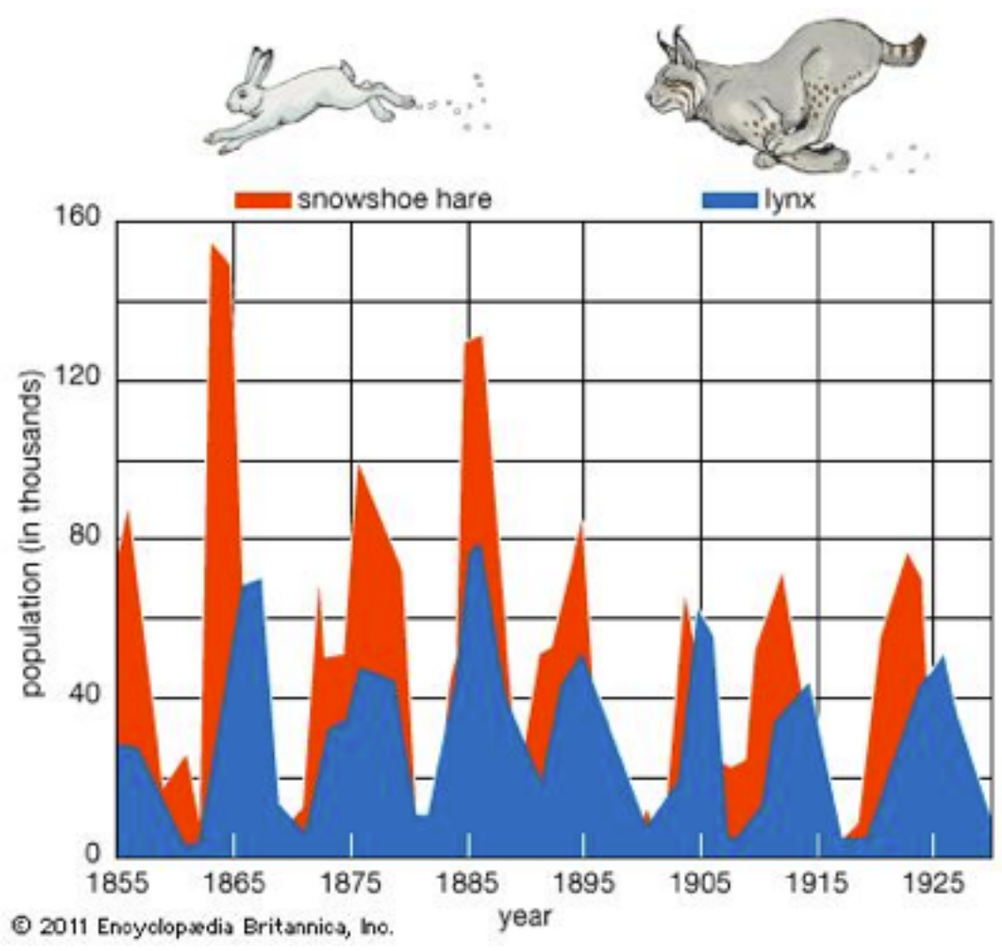
- Conceptualization of how we understand phenomena in nature.

- All models are wrong but are based on the best information we have.

- Can be a cartoon

- Can be mathematical

- Can be statistical



What is a model?

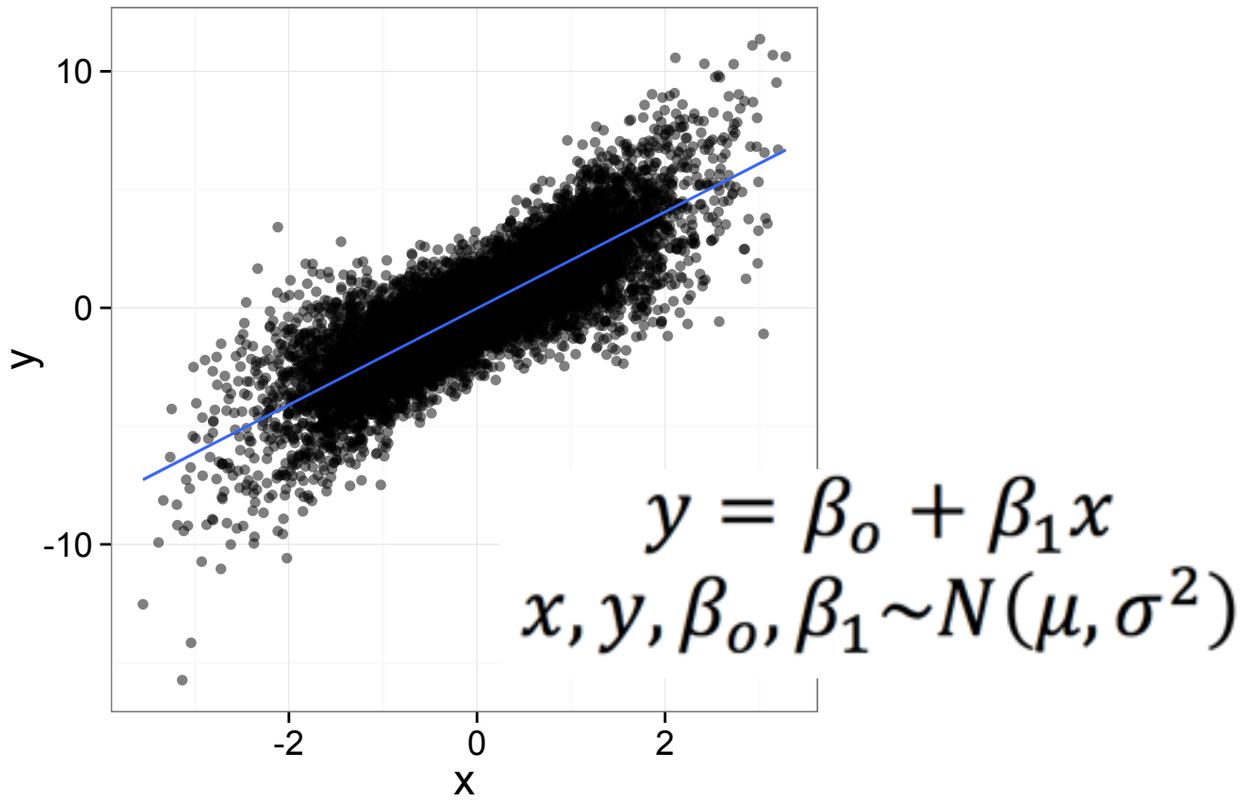
- Conceptualization of how we understand phenomena in nature.

- All models are wrong but are based on the best information we have.

- Can be a cartoon

- Can be mathematical

- Can be statistical



What is simulation?

- Producing information to mimic something we're interested in
  - Simulating sequences
  - Simulating numerical data

What is simulation?

- Producing information to mimic something we're interested in
    - Simulating sequences
    - Simulating numerical data

What is simulation?

- Producing information to mimic something we're interested in
  - Simulating sequences

MetaSim

Daniel H. Huson and Felix Ott

with contributions from:
R. Schmid, A.F. Auch and D.C. Richter

www-ab.informatik.uni-tuebingen.de/software/metasim

OPEN ACCESS Freely available online

PLOS | ONE

NeSSM: A Next-Generation Sequencing Simulator for Metagenomics

Ben Jia[1,3], Liming Xuan[3,4,5], Kaiye Cai[4,5], Zhiqiang Hu[2,4], Liangxiao Ma[4],

1 School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China, 2 Department of Bioinformat Biotechnology, Shanghai Jiao Tong University, Shanghai, China, 3 School of Bioengineering, East China University of Sci Center for Bioinformation Technology, Shanghai, China

Johnson et al. BMC Bioinformatics 2014, 15(Suppl 9):S14
http://www.biomedcentral.com/1471-2105/15/S9/S14

BMC Bioinformatics

PROCEEDINGS Open Access

Grinder: a versatile amplicon and shotg sequence simulator

Florent E. Angly[1,*], Dana Willner[1,2], Forest Rohwer[3], Philip Hugenh Gene W. Tyson[1,5]

A better sequence-read simulator program for metagenomics

Stephen Johnson[1*], Brett Trost[1], Jeffrey R Long[1], Vanessa Pittet[2], Anthony Kusalik[1]

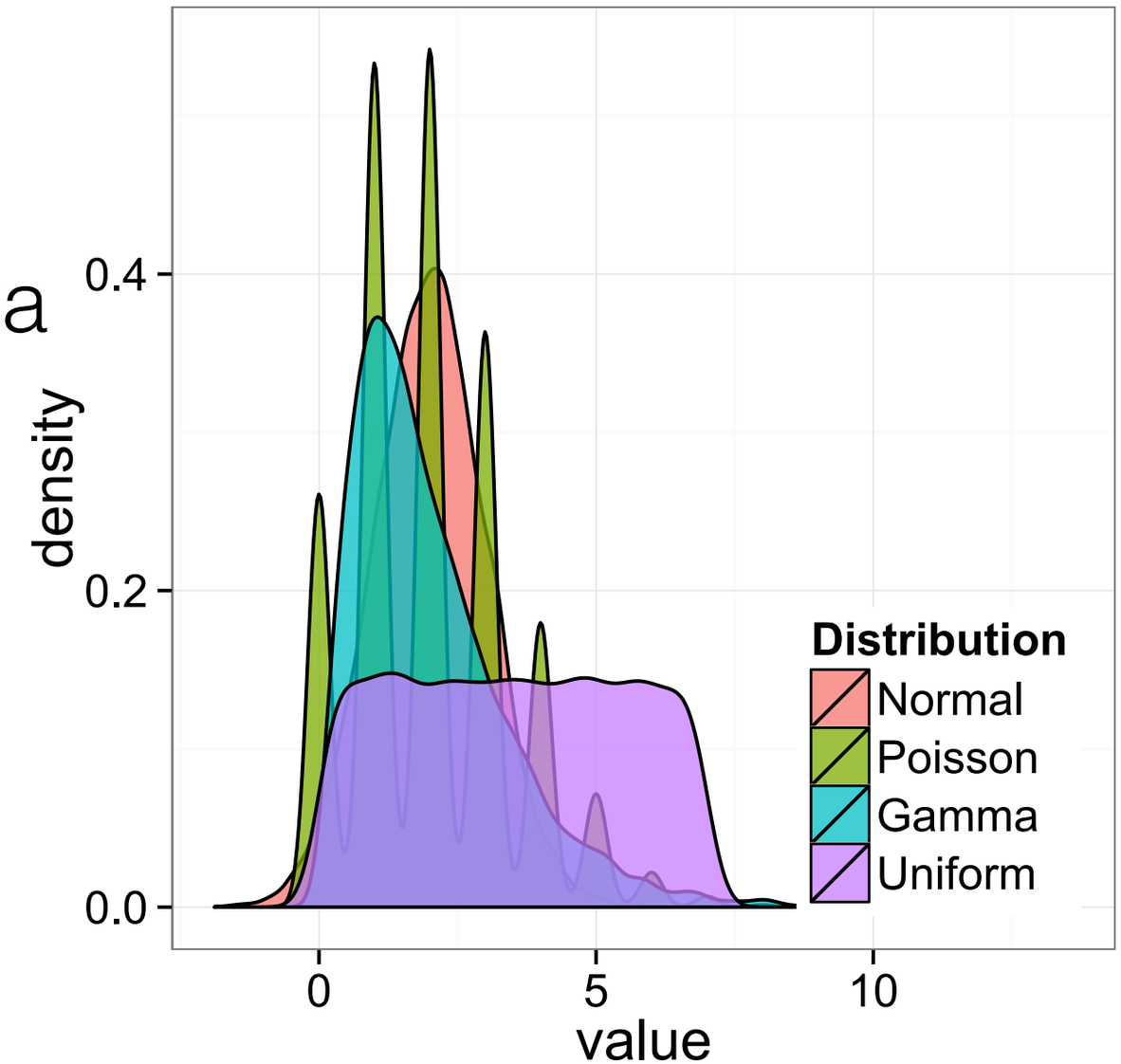
From RECOMB-Seq: Fourth Annual RECOMB Satellite Workshop on Massively Parallel Sequencing Pittsburgh, PA, USA. 31 March - 05 April 2014

What is simulation?

- Producing information to mimic something we're interested in
  - Simulating sequences
  - Simulating numerical data

What is simulation?

- Producing information to mimic something we're interested in
  - Simulating sequences
  - Simulating numerical data



What's the use?

- Microbes are hard to observe in nature, and data describing them can be expensive

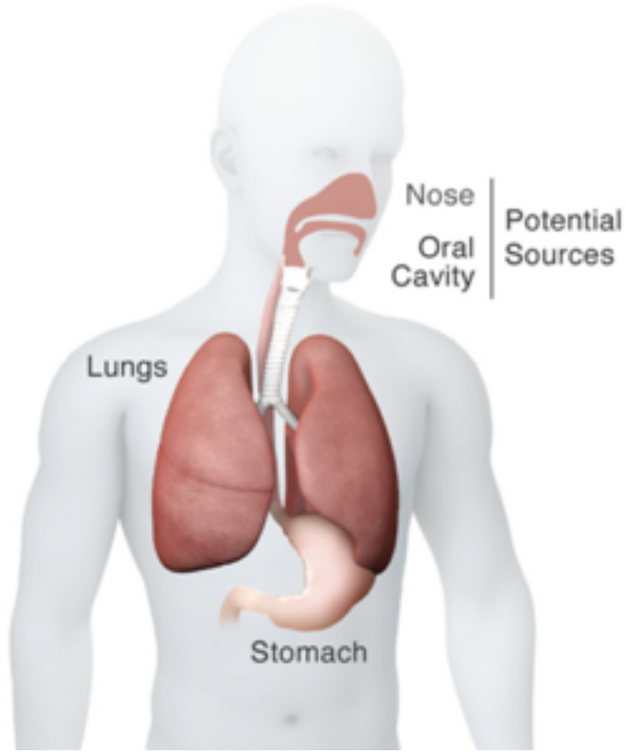
- Microbial data is highly multivariate

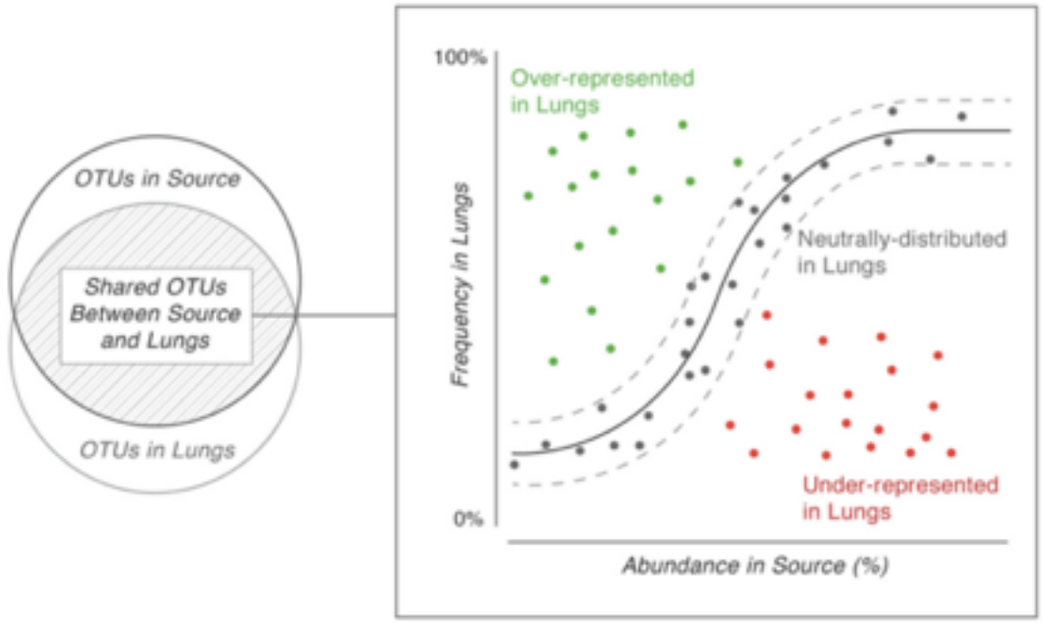
- Microbial ecologists love to use a variety of methods and pre-packaged tools that are not necessarily easy to understand

- Numerical simulation is universal; simulating sequences is applied (both can be very slick but simple tools!)

- Simulation modeling can be used to explain complex natural phenomena

Application of a neutral community model to assess structuring of the human lung microbiome

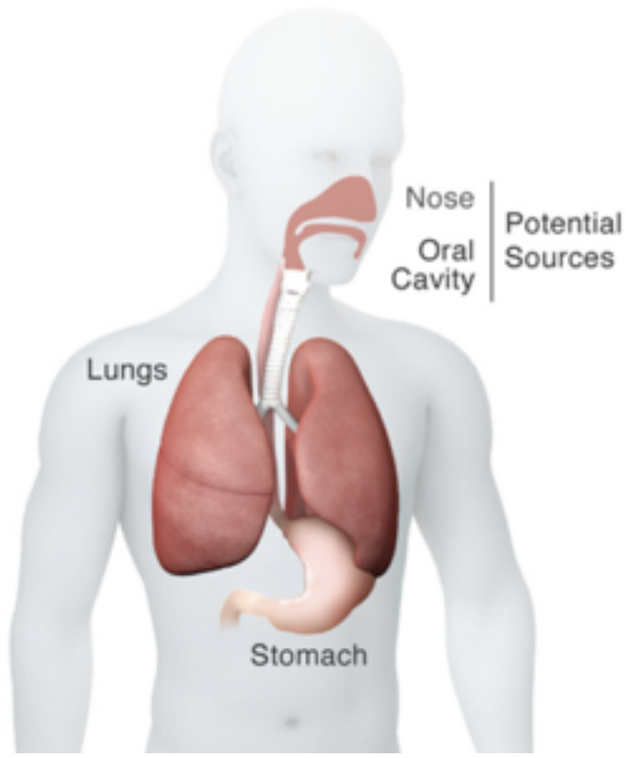


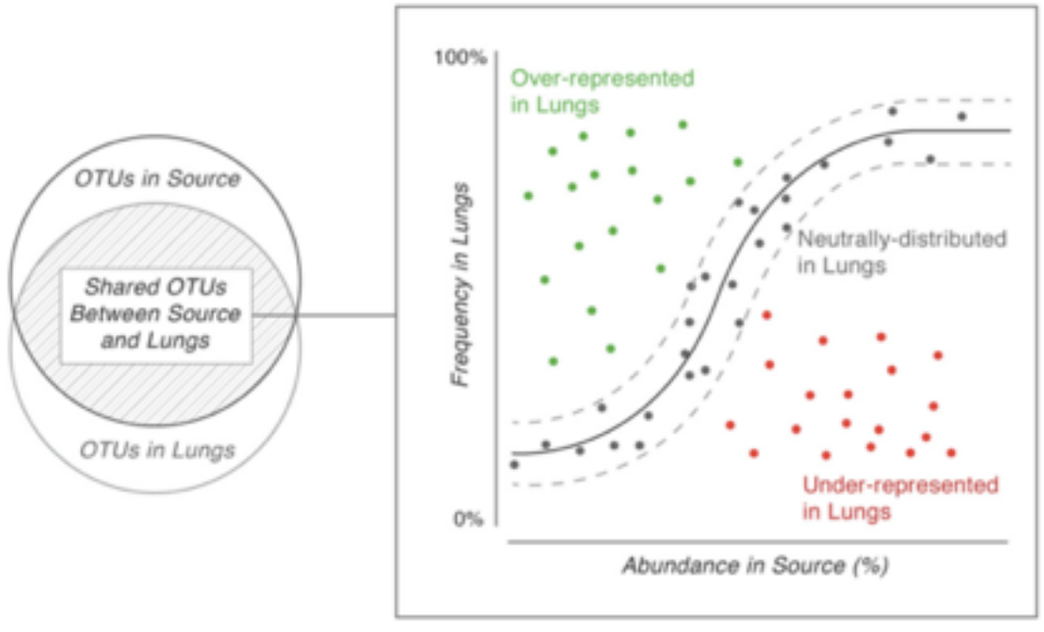


Venkataraman, A., Bassis, C.M., Beck, J.M., Young, V.B., Curtis, J.L., Huffnagle, G.B., and Schmidt, T.M. (2015). MBio 6: e02284–14.

Application of a neutral community model to assess structuring of the human lung microbiome

- Modeled OTU abundances as probabilities of detection in either one environment or another. Similar probabilities across environments suggests neutral community assembly



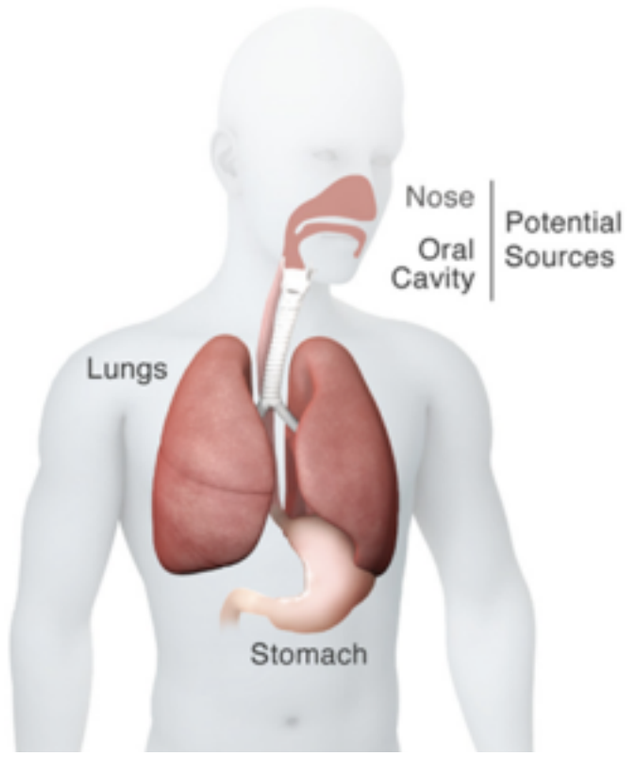


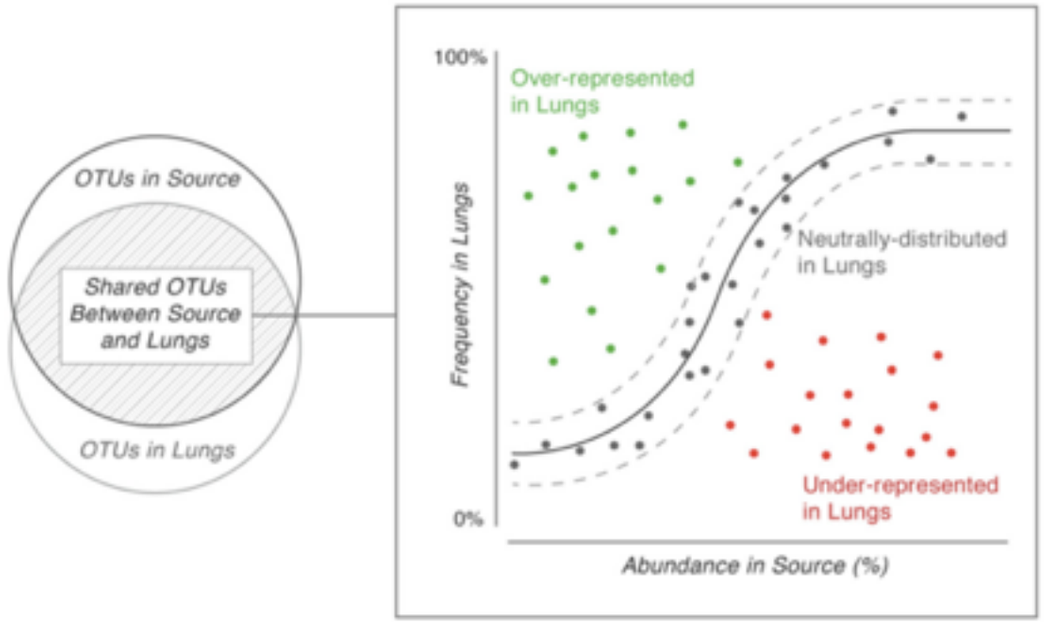
Venkataraman, A., Bassis, C.M., Beck, J.M., Young, V.B., Curtis, J.L., Huffnagle, G.B., and Schmidt, T.M. (2015). MBio 6: e02284–14.

Application of a neutral community model to assess structuring of the human lung microbiome

- Though the model is simple, it represents a statistical expectation for OTU frequencies based on theory that can be simulated from various distributions

- What are pros and cons of this approach? Did you find this approach useful?





Venkataraman, A., Bassis, C.M., Beck, J.M., Young, V.B., Curtis, J.L., Huffnagle, G.B., and Schmidt, T.M. (2015). MBio 6: e02284–14.

Important things to consider when modeling/simulating…

- Processes generate patterns and not the other way around
- How do we determine randomness driven by nature vs. natural variation in a distribution? Am I observing complex natural phenomenon that create a random pattern (neutral theory) or am I observing natural variation? (Google Brian McGill, U of Maine if interested)

Say you are reviewing literature, and you stumble across an interesting analysis…

Say you are reviewing literature, and you stumble across an interesting analysis…

Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in Microbial Diversity

Ashley Shade,[a] Stuart E. Jones,[b] J. Gregory Caporaso,[c,d] Jo Handelsman,[e] Rob Knight,[f,g] Noah Fierer,[h,i] Jack A. Gilbert[c,j]

Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, USA[a]; Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana, USA[b]; Institute for Genomic and Systems Biology, Argonne National Laboratory, Argonne, Illinois, USA[c]; Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona, USA[d]; Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut, USA[e]; Howard Hughes Medical Institute, Boulder, Colorado, USA[f]; Department of Chemistry and Biochemistry and BioFrontiers Institute, University of Colorado, Boulder, Colorado, USA[g]; Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado, USA[h]; Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA[i]; Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, USA[j]

Say you are reviewing literature, and you stumble across an interesting analysis…

Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in Microbial Diversity

Ashley Shade,[a] Stuart E. Jones,[b] J. Gregory Caporaso,[c,d] Jo Handelsman,[e] Rob Knight,[f,g] Noah Fierer,[h,i] Jack A. Gilbert[c,j]

Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, USA[a]; Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana, USA[b]; Institute for Genomic and Systems Biology, Argonne National Laboratory, Argonne, Illinois, USA[c]; Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona, USA[d]; Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut, USA[e]; Howard Hughes Medical Institute, Boulder, Colorado, USA[f]; Department of Chemistry and Biochemistry and BioFrontiers Institute, University of Colorado, Boulder, Colorado, USA[g]; Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado, USA[h]; Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA[i]; Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, USA[j]

- Now I can start identifying cool OTUs that are changing over time.

- What are some potential pitfalls to this analysis? Can I use simulation/modeling to address these concerns?

Walk through R script