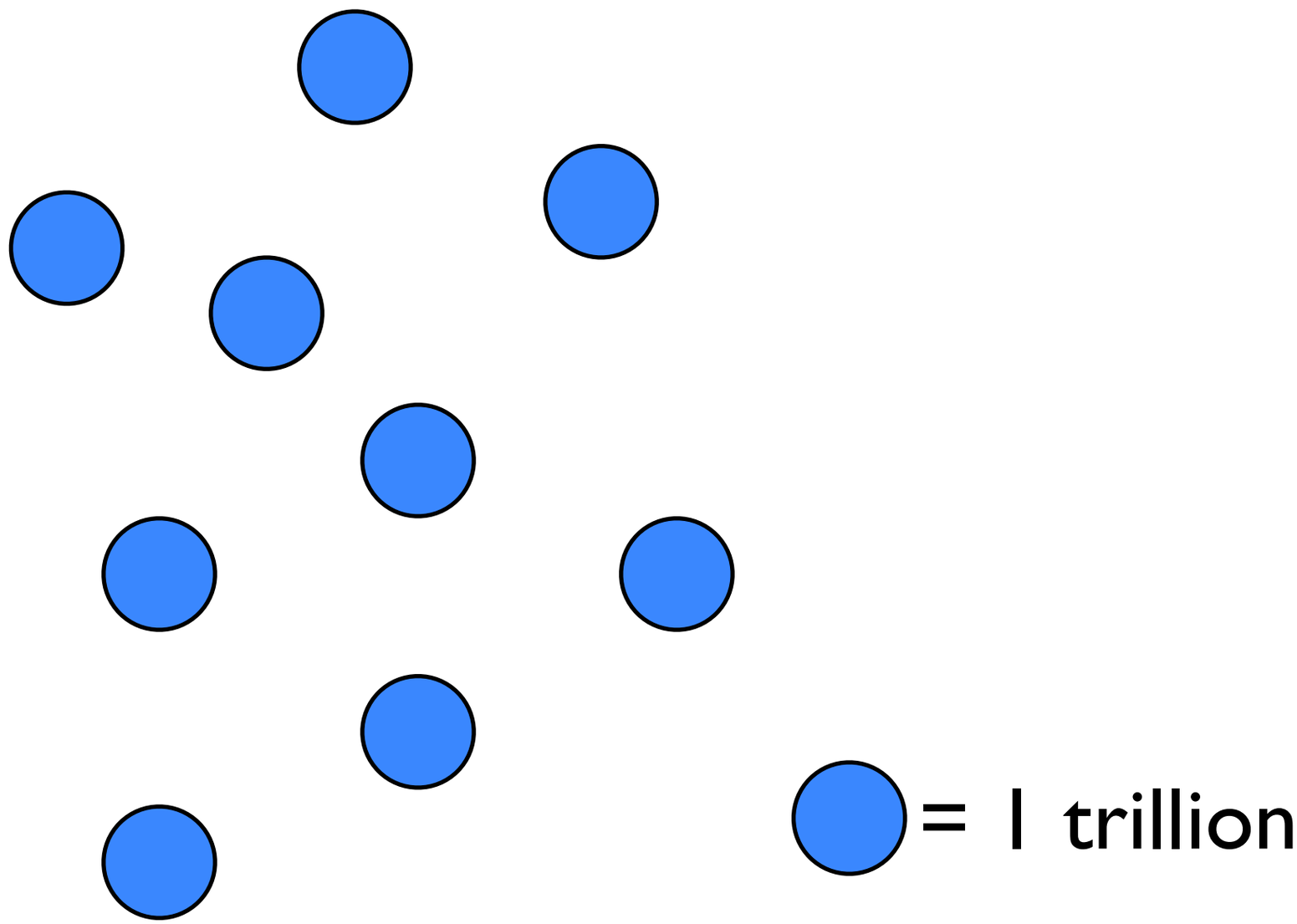
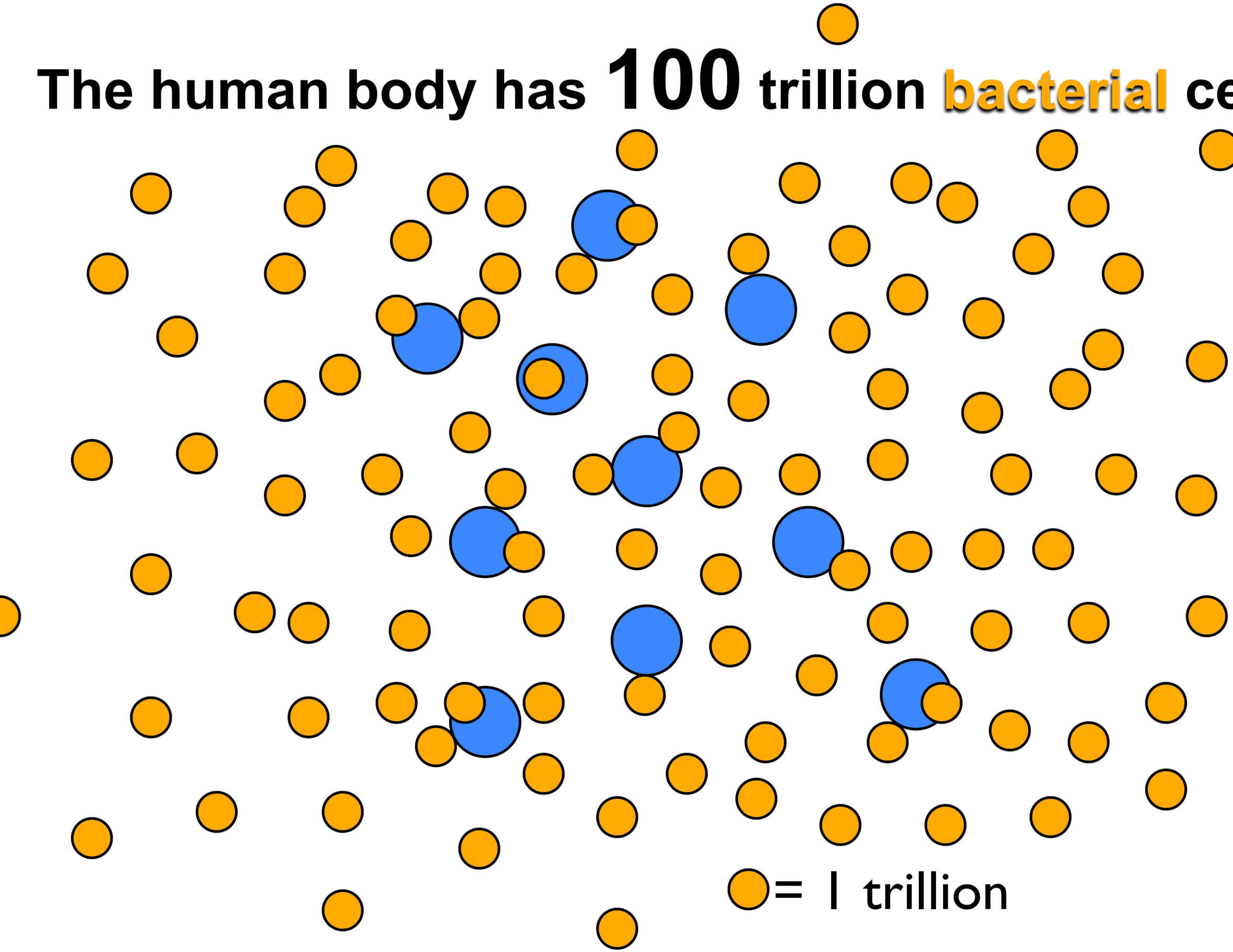
predicting metabolic networks in the human microbiome

- Christopher Smillie

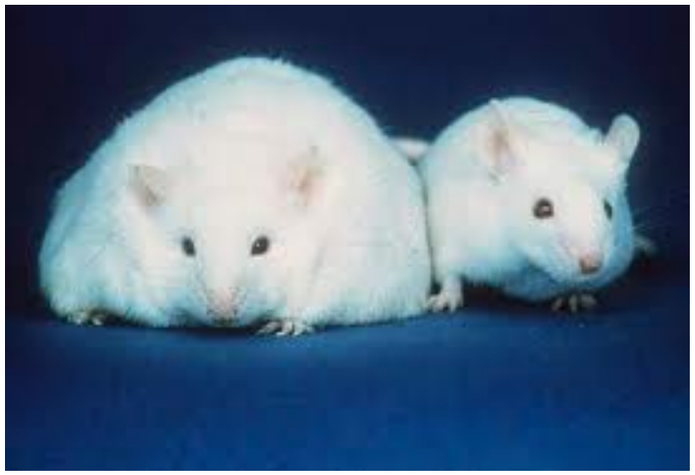
The human body has 10 trillion human cells



The human body has 100 trillion bacterial cells



... with many important functions



- Metabolism
- Drug degradation
- Health (protect against pathogens)
- Disease (obesity, diabetes, cancer)

Can study these functions with DNA sequencing

Human Microbiome Project

- 2.3 Terabytes (compressed) DNA sequences

Amazonas (VE)

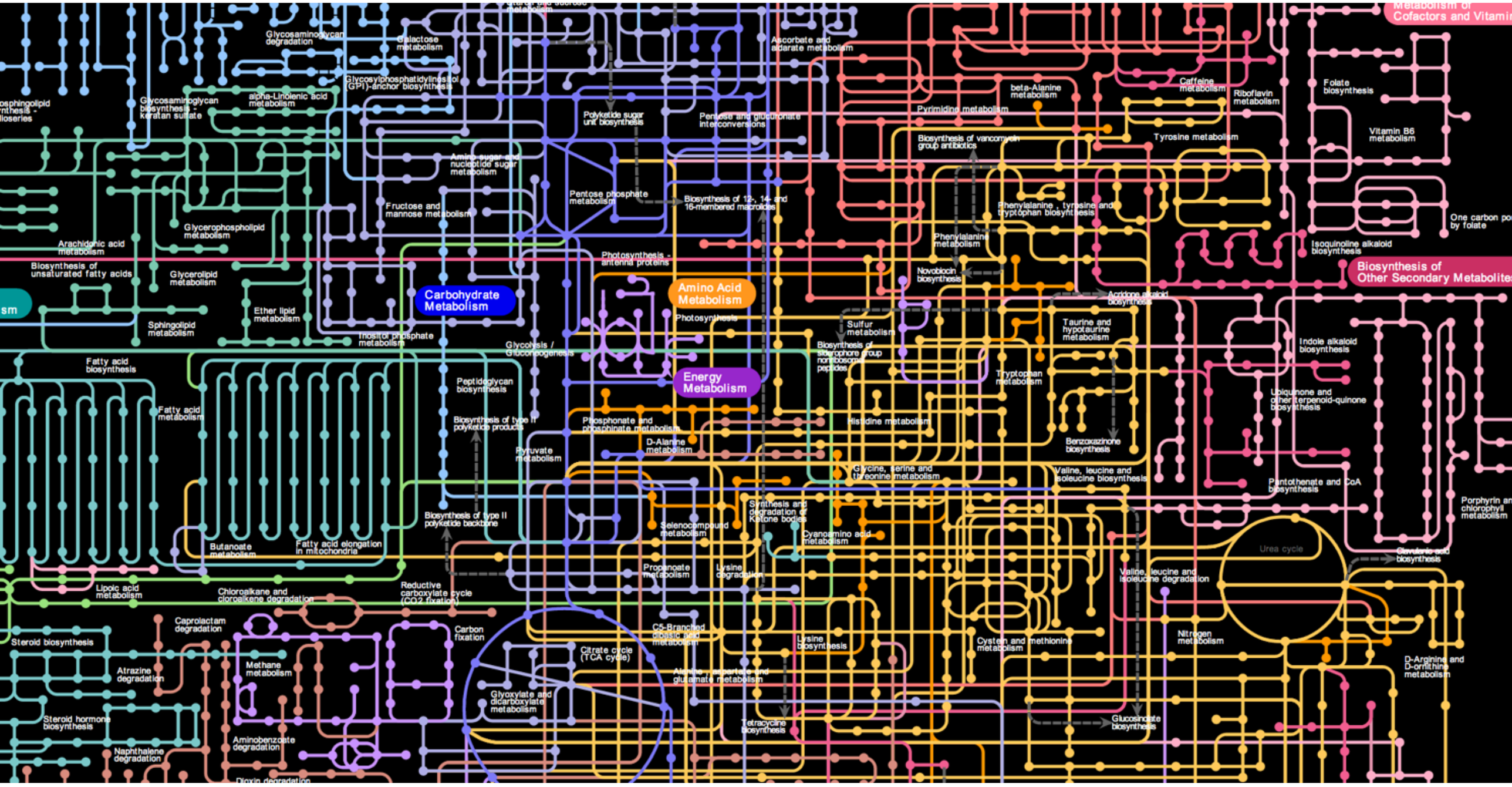
Rural Malawi

USA/Europe/Japan

18.337 Project goals

1. Map DNA to functions (MapReduce)
2. Analyze functions (Julia...)

Step 1. Metabolic functions



(Each dot is an enzyme. Lines connect metabolic pathways)

Step 1. Amazon Elastic MapReduce

1. Upload data to Amazon
2. Search for proteins in DNA
3. Count enzymes per sample
4. Merge enzymes into pathways

(5. Get minimal metabolism)

Step 2. Analysis

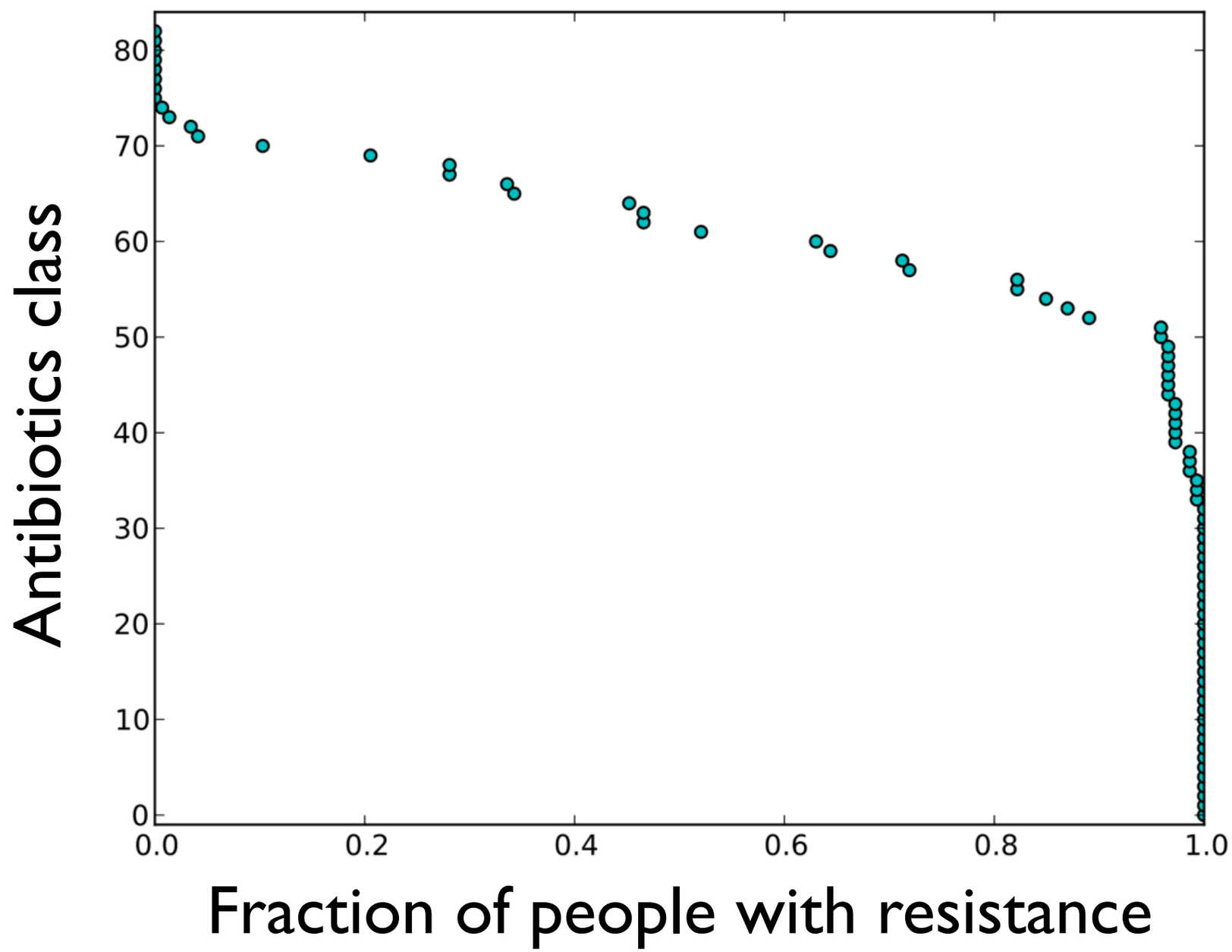
1. Antibiotics resistance

(2. Functional profiles - SVD)

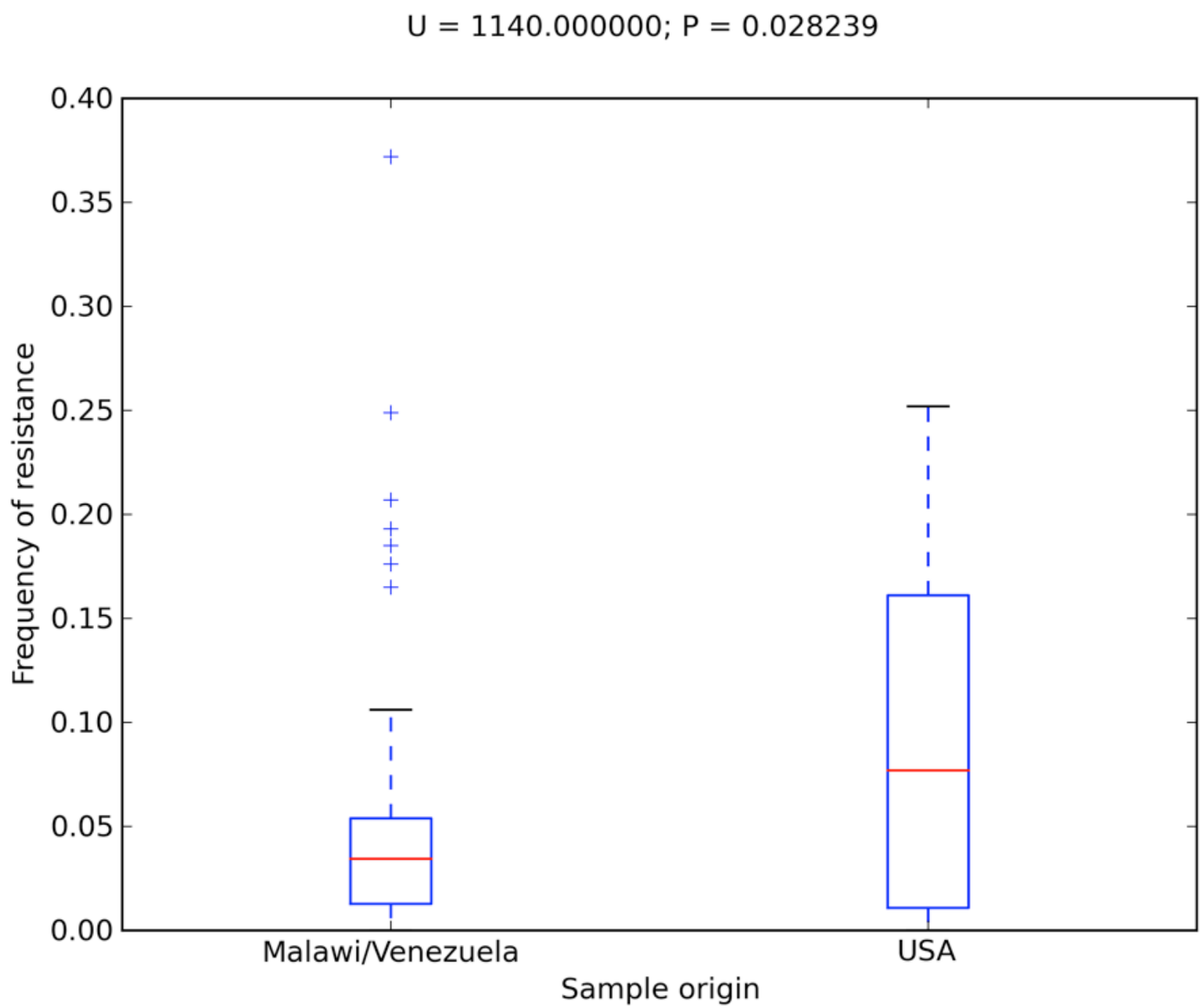
Antibiotics resistance: USA

Antibiotics classes w/ 100% resistant:

- bacitracin
- chloramphenicol
- streptogramin_b
- penicillin
- beta_lactam
- erythromycin
- tetracycline
- fluoroquinolone
- cephalosproin
- streptomycin
- methicillin
- vancomycin

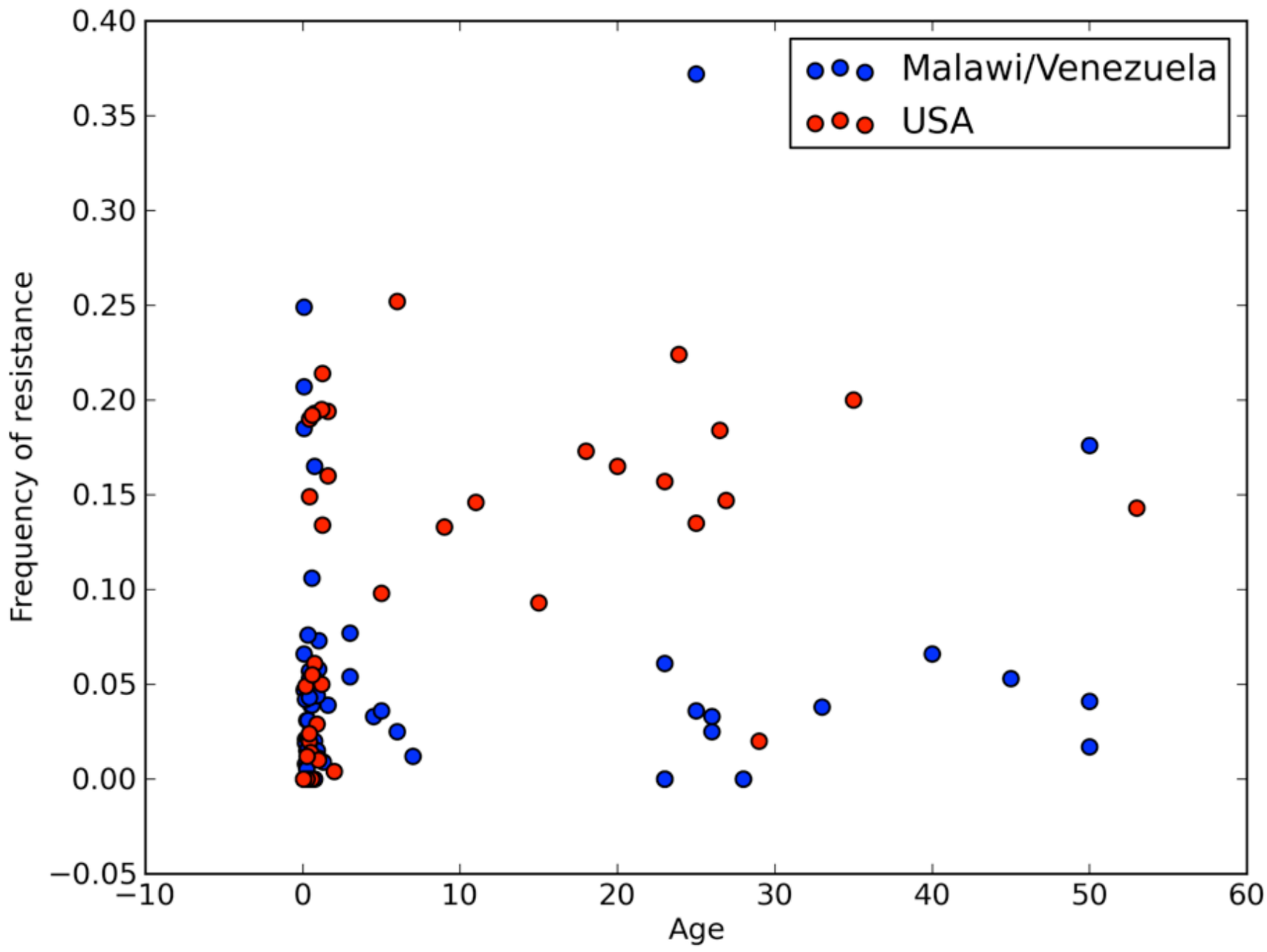


Antibiotics resistance: Malawi/Amazon/USA



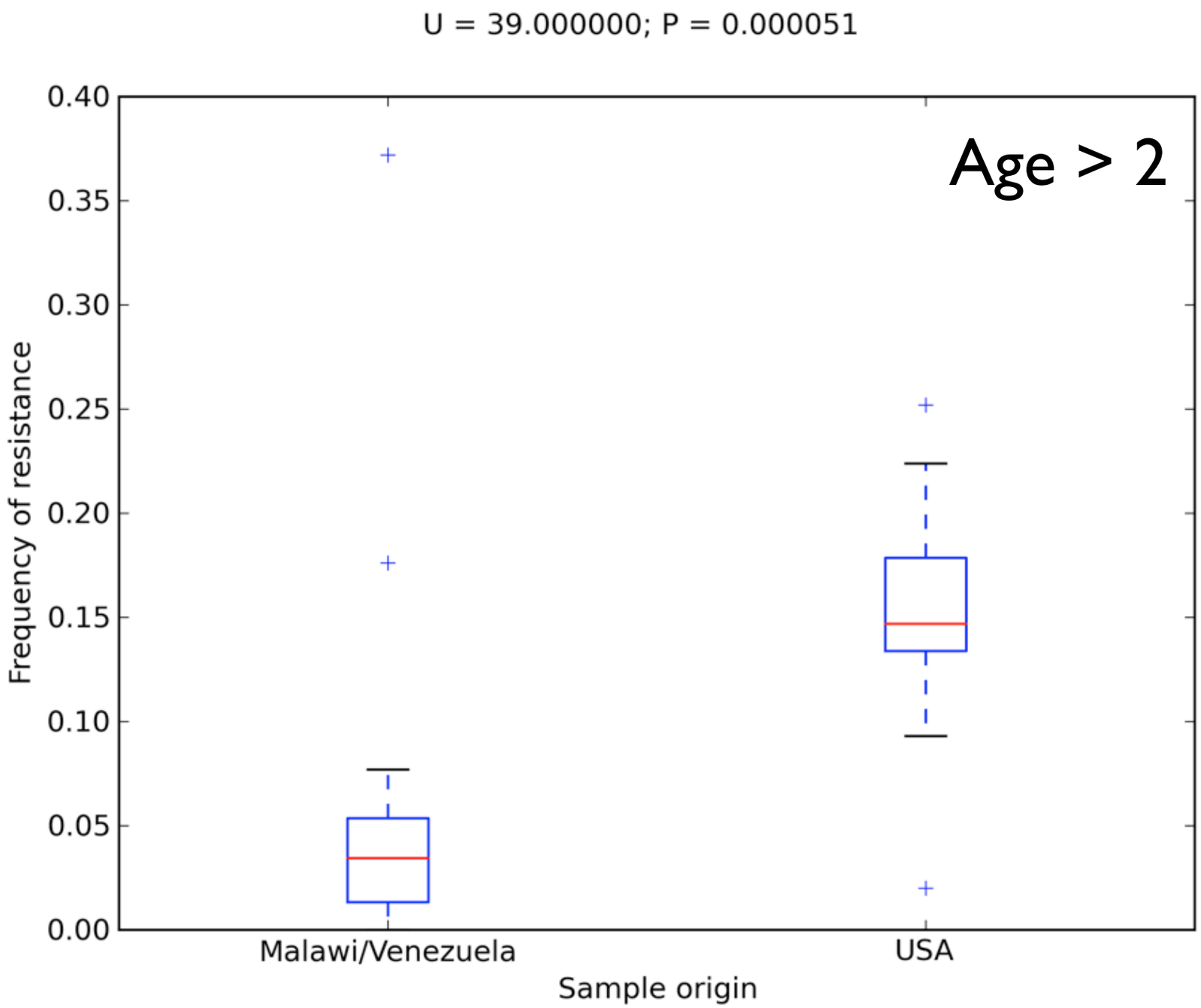
Sample origin has slight effect on AR frequency

Antibiotics resistance: Malawi/Amazon/USA



Early in life, f(AR) is variable. Later in life, it stabilizes?

Antibiotics resistance: Malawi/Amazon/USA



After controlling for age, sample origin is significant (P = 5e-5)