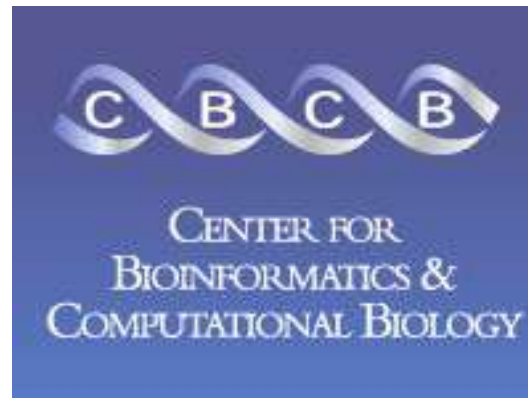
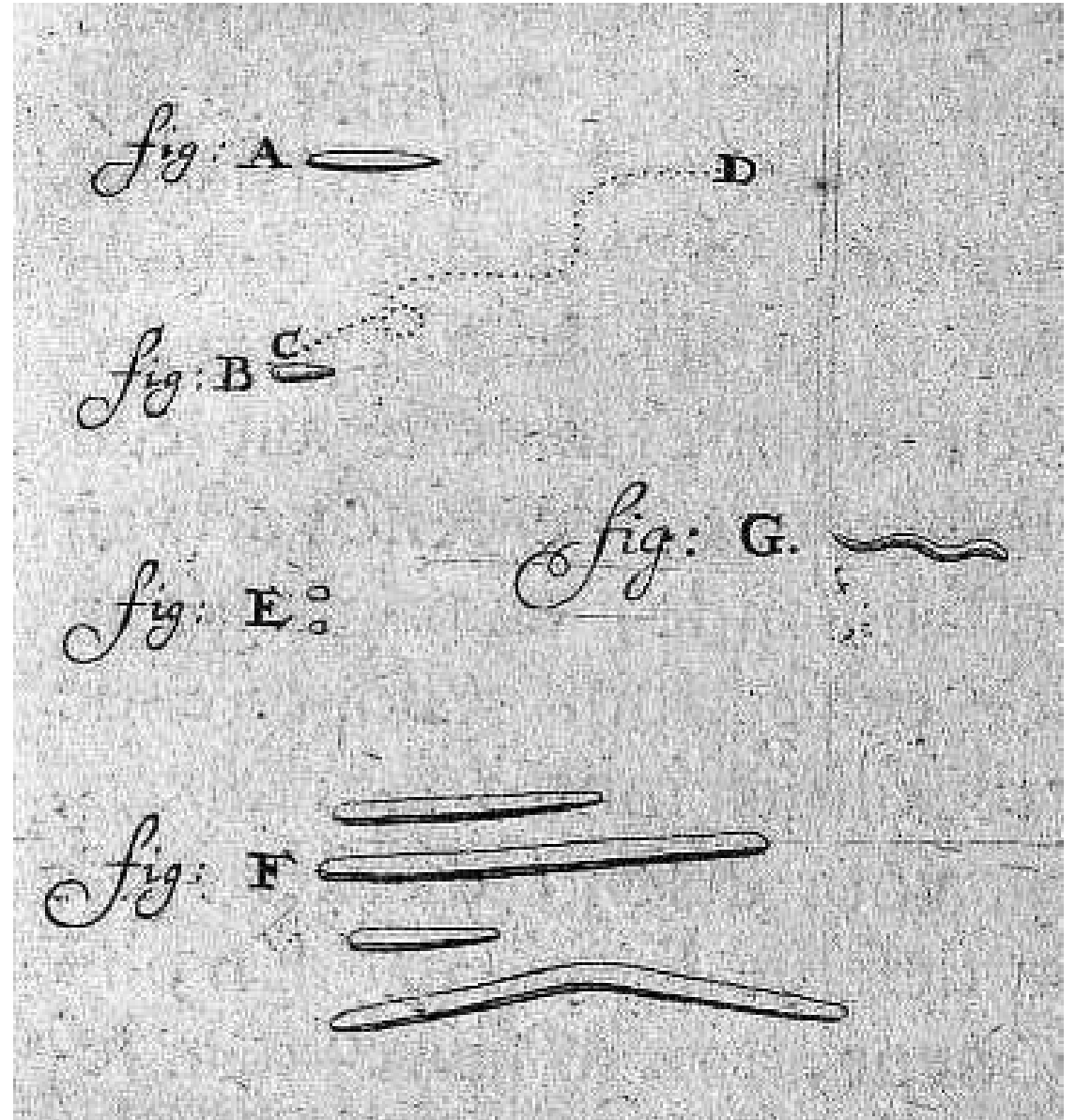


Challenges and opportunities in characterizing microbiomes

Mihai Pop

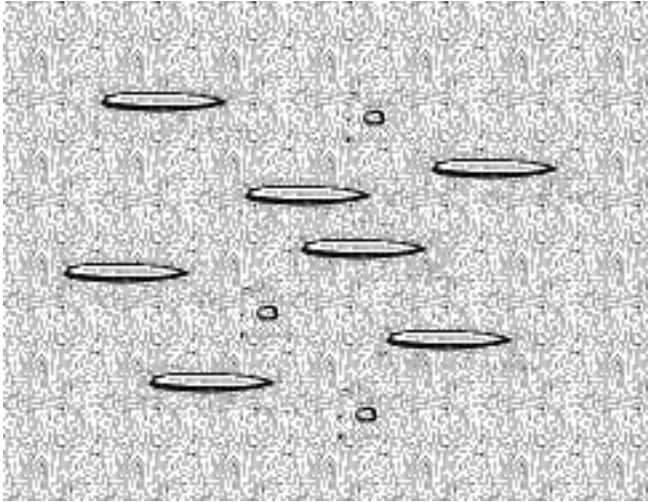


17th century biology

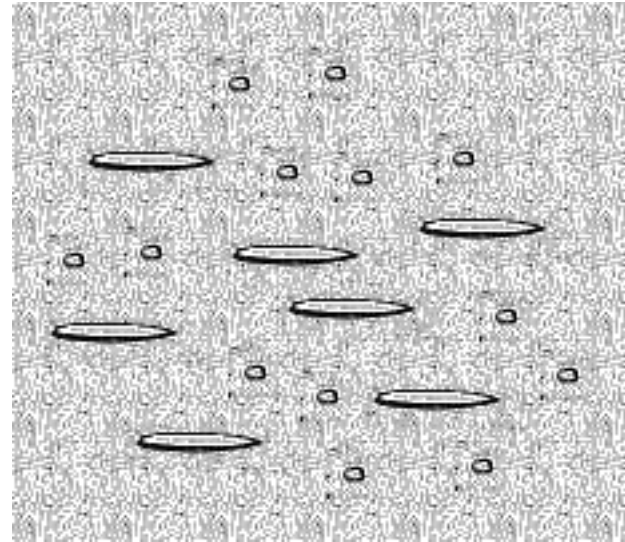


What's the difference?

A



B



21st century biology

```
>F4BT0V001CZSIM rank=0000138 x=1110.0 y=2700.0 lengt
ACTGCTCTCATGCTGCCTCCCGTAGGAGTGCCTCCCTGAGCCAGGATCAAA
>F4BT0V001BBJQS rank=0000155 x=424.0 y=1826.0 lengtl
ACTGACTGCATGCTGCCTCCCGTAGGAGTGCCTCCCTGCGCCATCAA
>F4BT0V001EDG35 rank=0000182 x=1676.0 y=2387.0 leng:
ACTGACTGCATGCTGCCTCCCGTAGGAGTCGCCGTCCTCGACNC
>F4BT0V001D2HQQ rank=0000196 x=1551.0 y=1984.0 leng:
ACTGACTGCATGCTGCCTCCCGTAGGAGTGCCGTCCTCGAC
>F4BT0V001CM392 rank=0000206 x=966.0 y=1240.0 lengtl
AANCAGCTCTCATGCTCGCCCTGACTTGGCATGTGTAAAGCCTGTAGGCTA
>F4BT0V001EIMFX rank=0000250 x=1735.0 y=907.0 lengtl,--v
ACTGACTGCATGCTGCCTCCCGTAGGAGTGTGCGGCCATCAGACTG
>F4BT0V001ENDKR rank=0000262 x=1789.0 y=1513.0 length=56
GACTGTGCATGCTGCCTCCCGTAGGAGTGCCTCCCTGAGCCAGGATCAAAC
CTG
>F4BT0V001D91MI rank=0000288 x=1637.0 y=2088.0 length=56
ACTGCTCTCATGCTGCCTCCCGTAGGAGTGCCTCCCTGAGCCAGGATCAAAC
CTG
>F4BT0V001D0Y5G rank=0000341 x=1534.0 y=866.0 length=75
GTCTGTGACATGCTGCCTCCCGTAGGAGTCTACACAAGTTGTGGCCCAGAACC
ACTGAGCCAGGATCAAAC
CTG
>F4BT0V001EMLE1 rank=0000365 x=1780.0 y=1883.0 length=84
ACTGACTGCATGCTGCCTCCCGTAGGAGTGCCTCCCTGCGCCATCAATGCTGCATGCTGCTCCCTGAGCCAGGATCAAAC
CTG
```



3000 samples
~1000 clinical variables

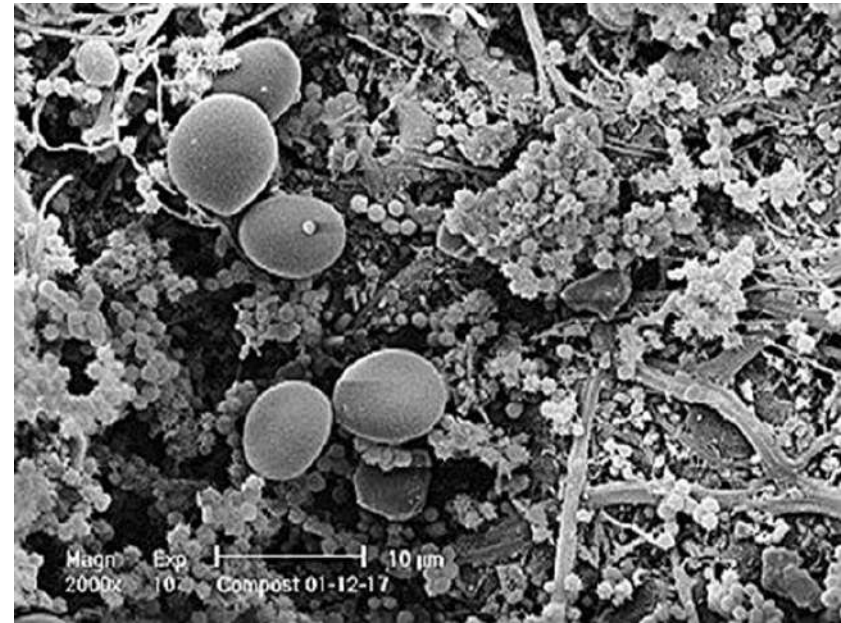
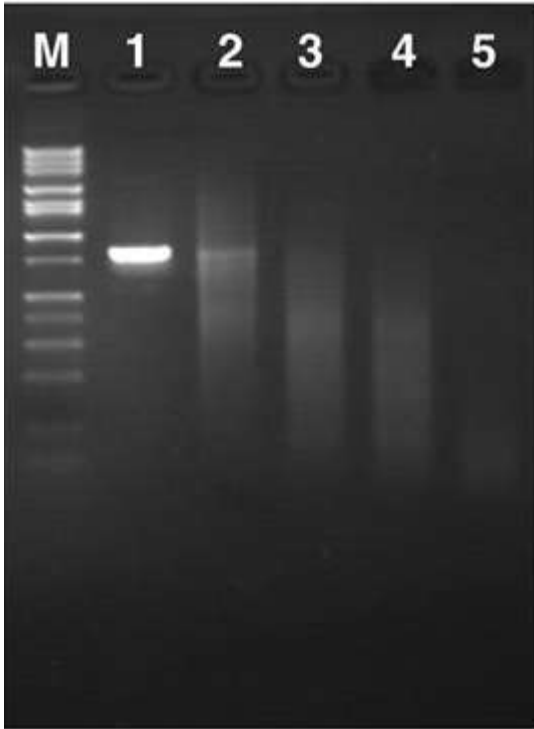
Healthy				Sick					

~60,000 "organisms"



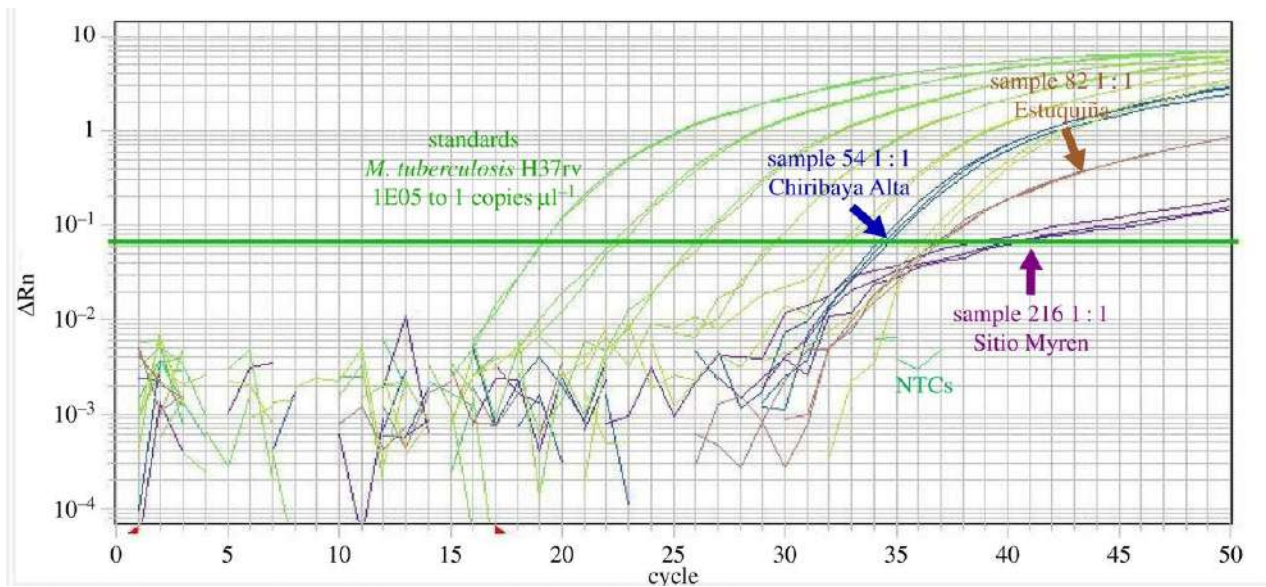
~10,000 sequences/sample

Data from study of diarrhea

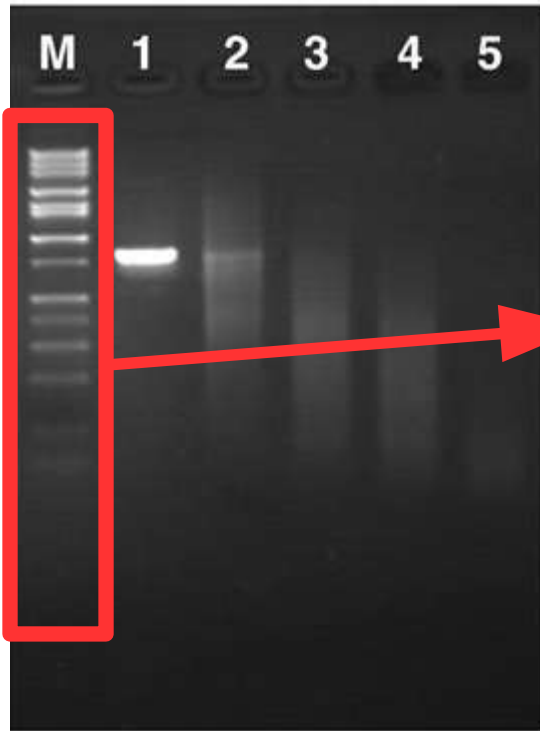


<http://www.forgreenies.com/growing-healthy-soils>

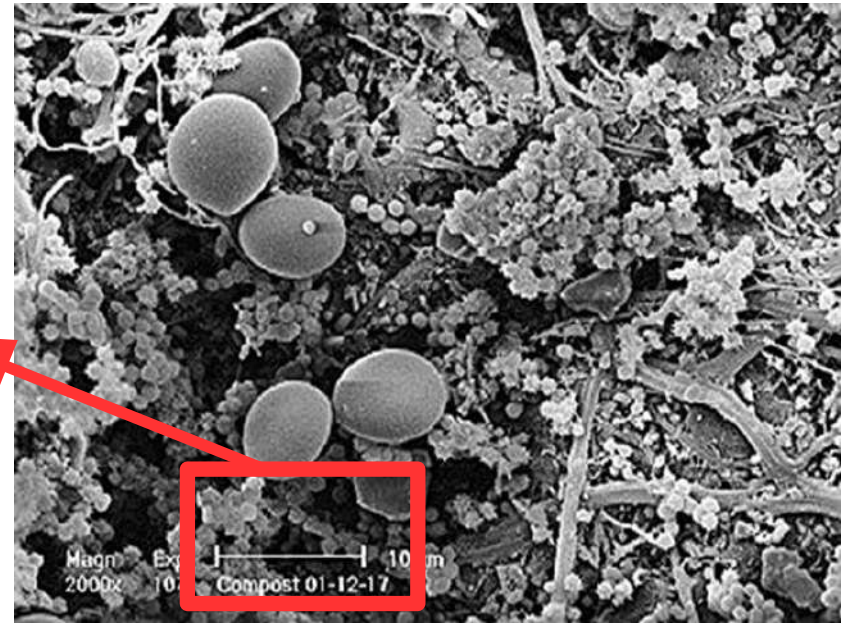
<https://academic.oup.com/nar/article/30/24/e139/1077910/Random-DNA-fragmentation-with-endonuclease-V>



<http://rstb.royalsocietypublishing.org/content/370/1660/20130622>

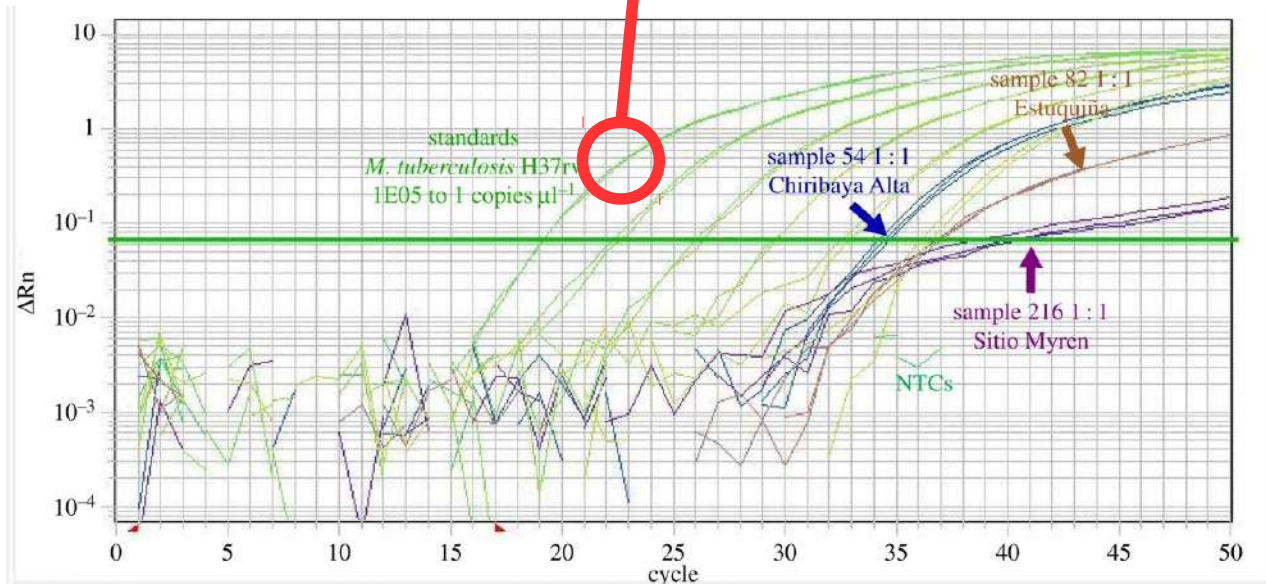


STANDARDS



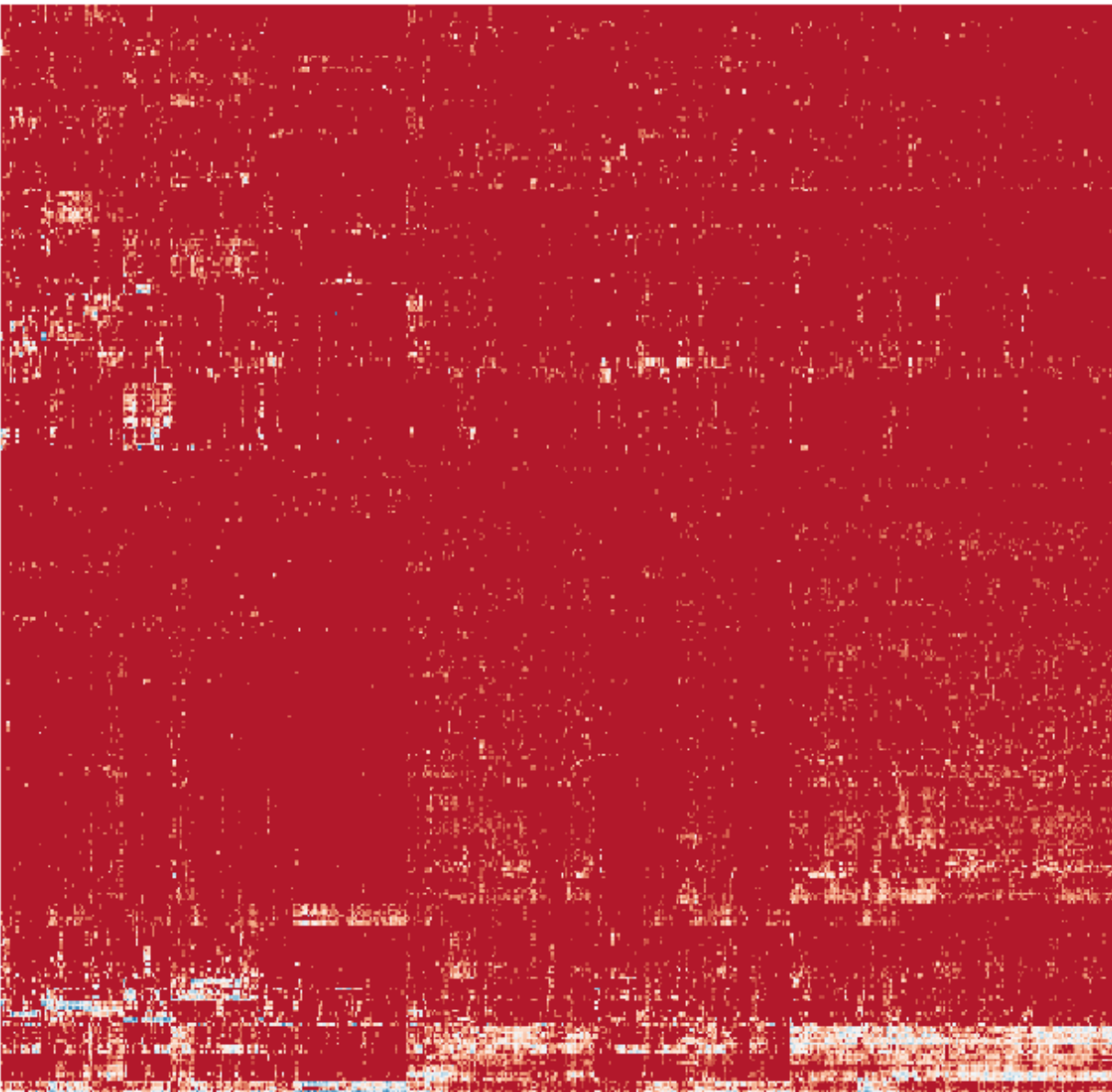
<http://www.forgreenies.com/growing-healthy-soils>

<https://academic.oup.com/nar/article/30/24/e139/1077910/Random-DNA-fragmentation-with-endonuclease-V>



<http://rstb.royalsocietypublishing.org/content/370/1660/20130622>

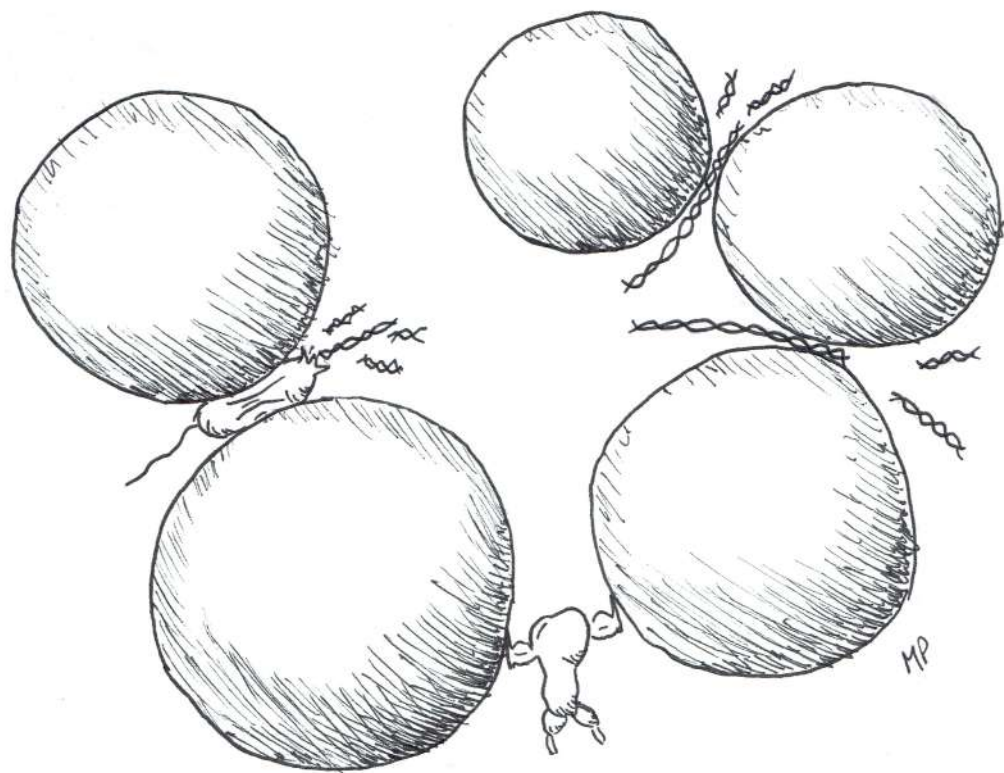
Healthy				Sick					



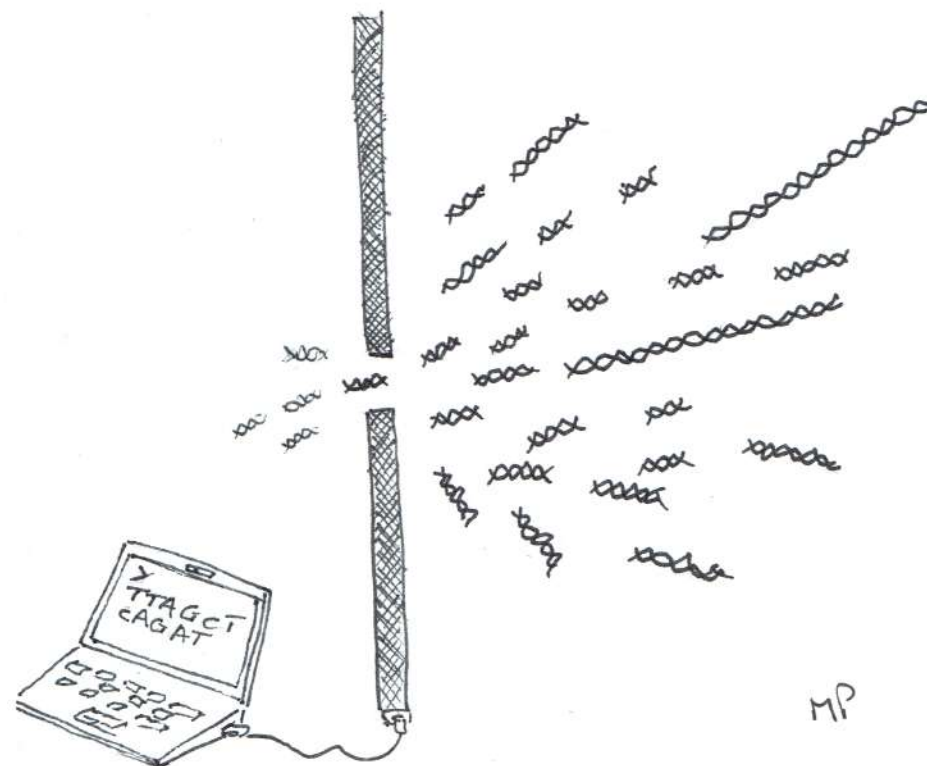
standard microbe?
standard community?

?

Biased cell lysis



Biased sequencing



How about the bioinformatics?

- $1 + 1 + 1 = 3$?

```
double x = 0.3;
```

```
double y = 0.1 + 0.1 + 0.1;
```

```
if (x == y)  
    System.out.println("Hurray");
```

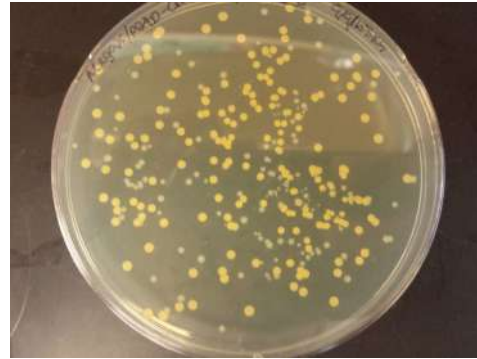
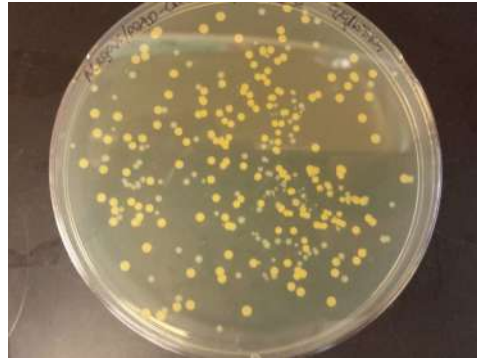
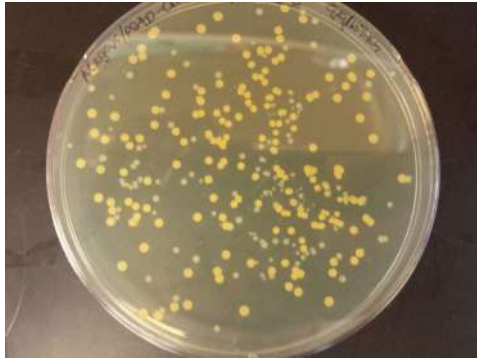


Savage Chickens

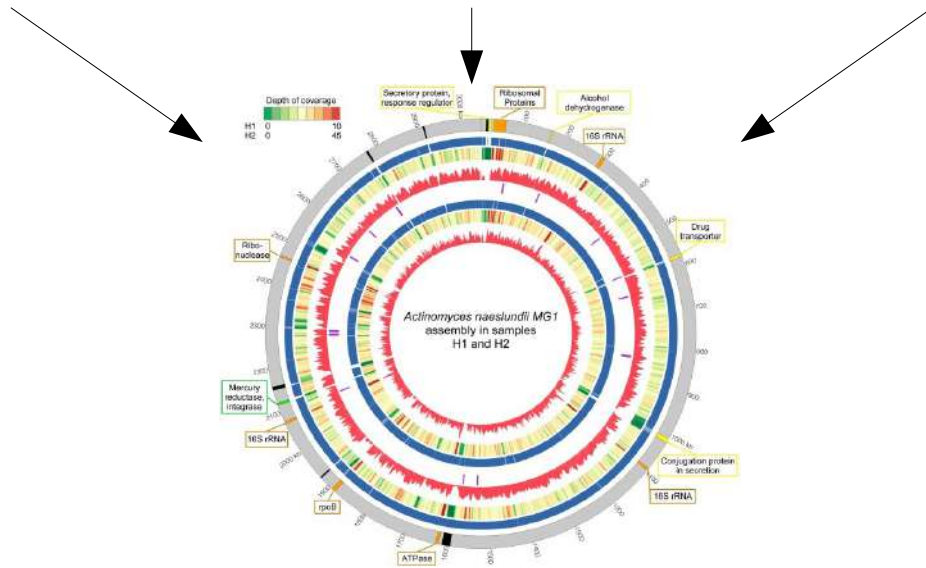
by Doug Savage



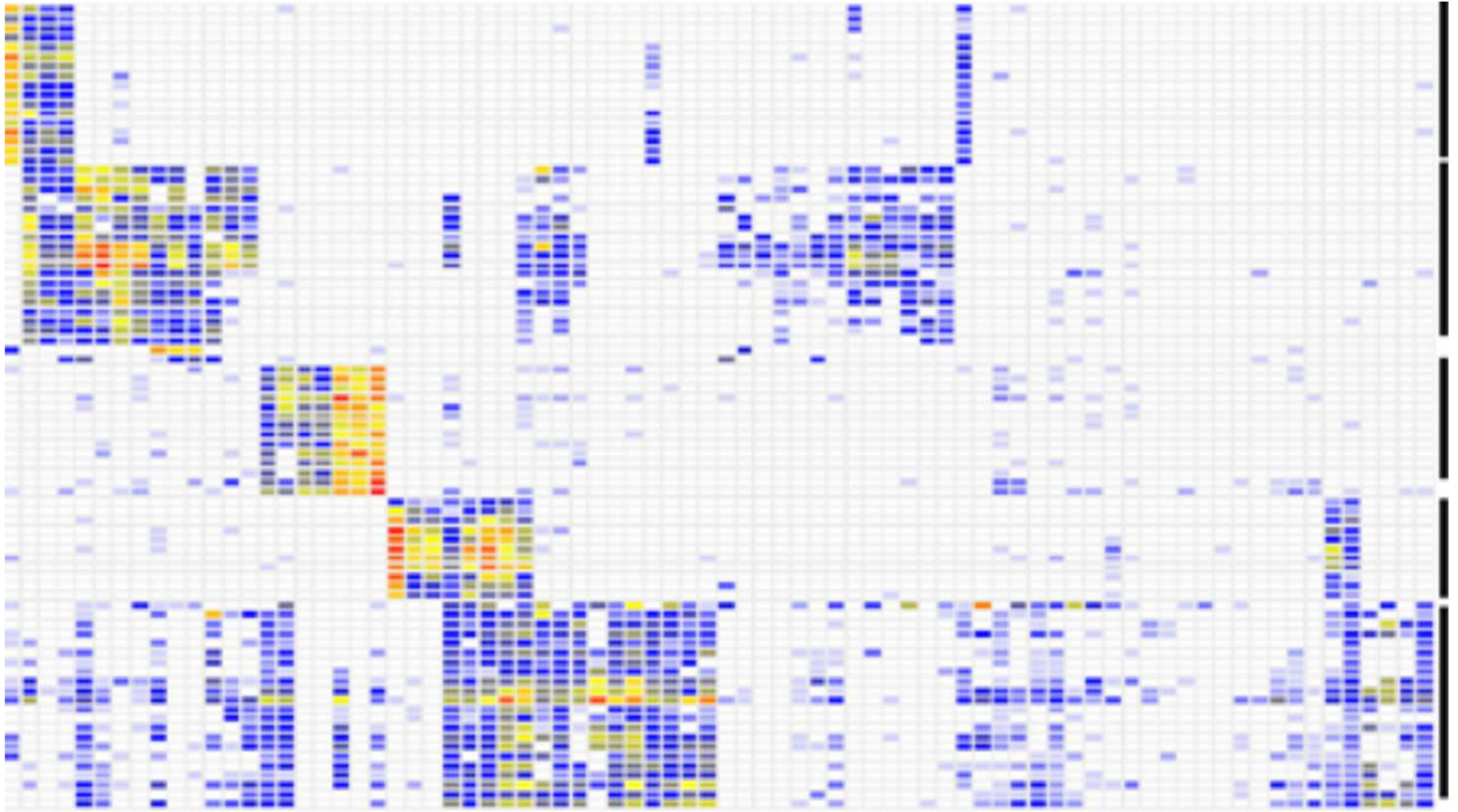
www.savagechickens.com



...



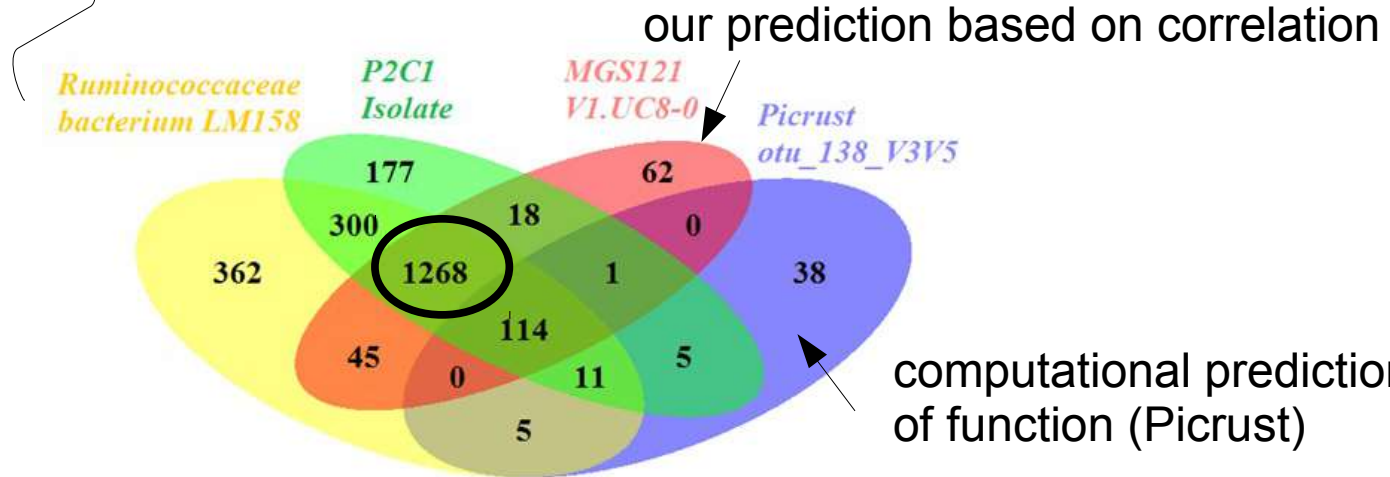
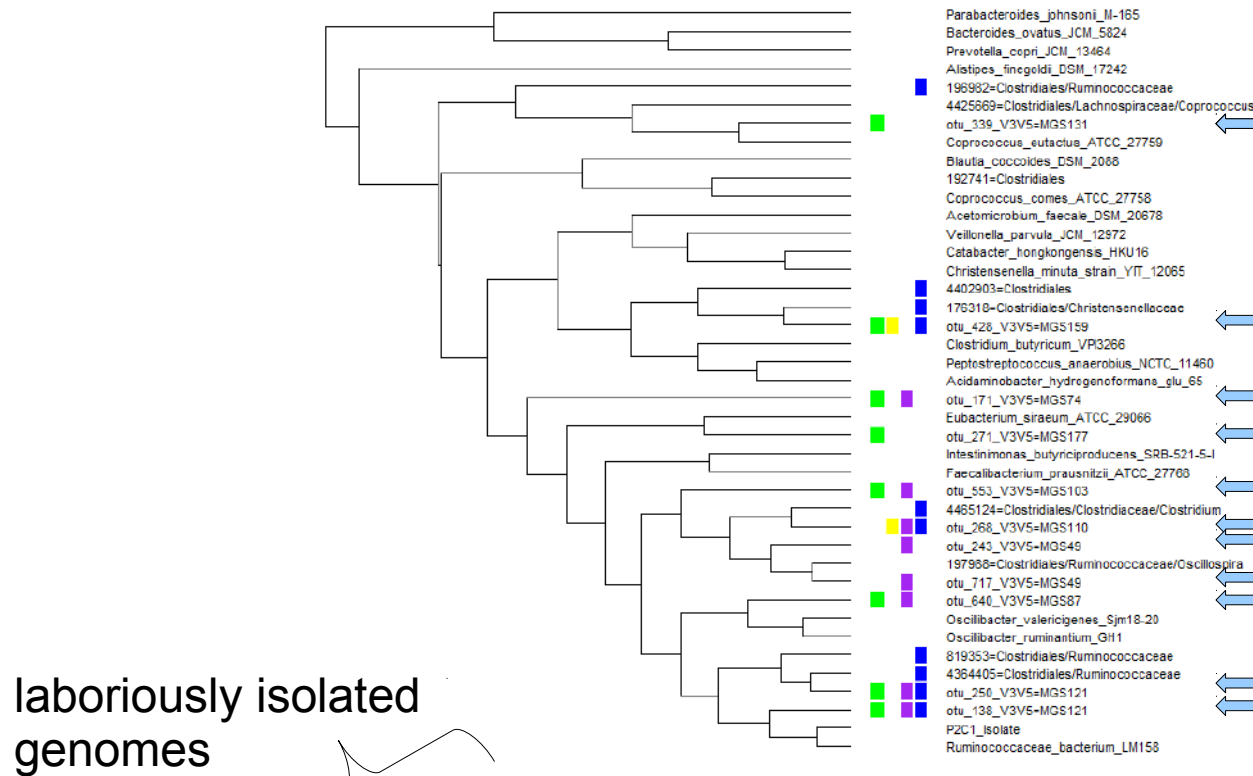
Proxy for function: correlation across samples



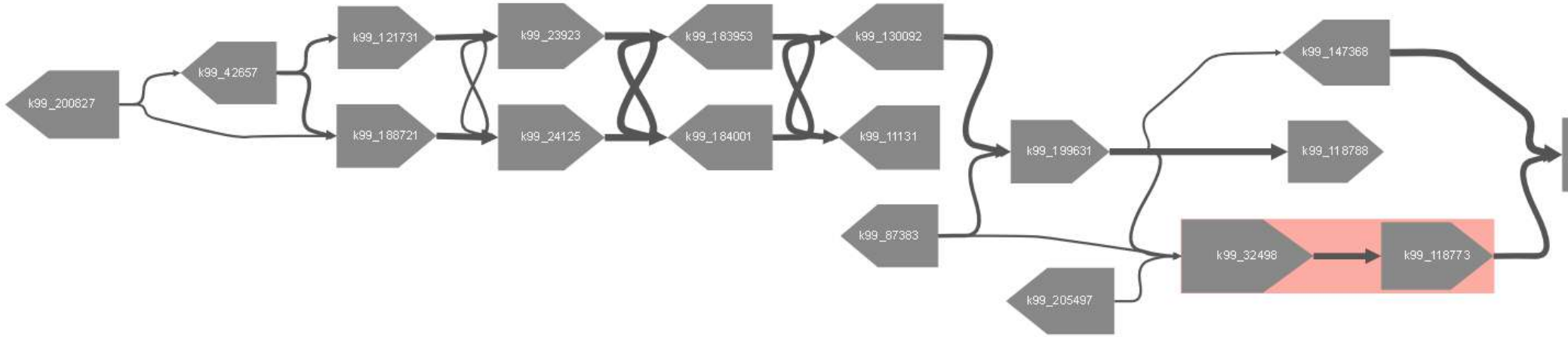
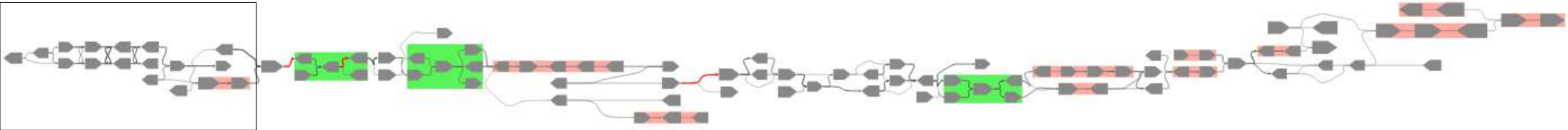
Quince – Concoct

Borenstein – Metagenomic deconvolution

Associating most-wanted OTUs with genes

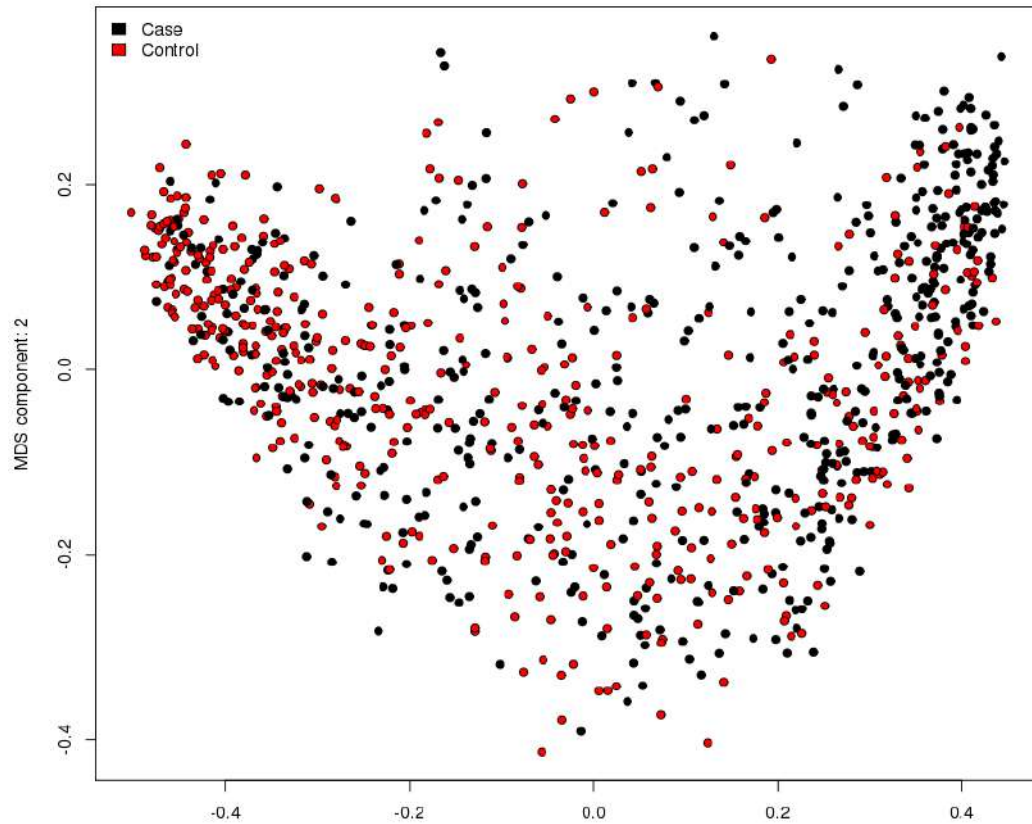


Biology hidden in complexity

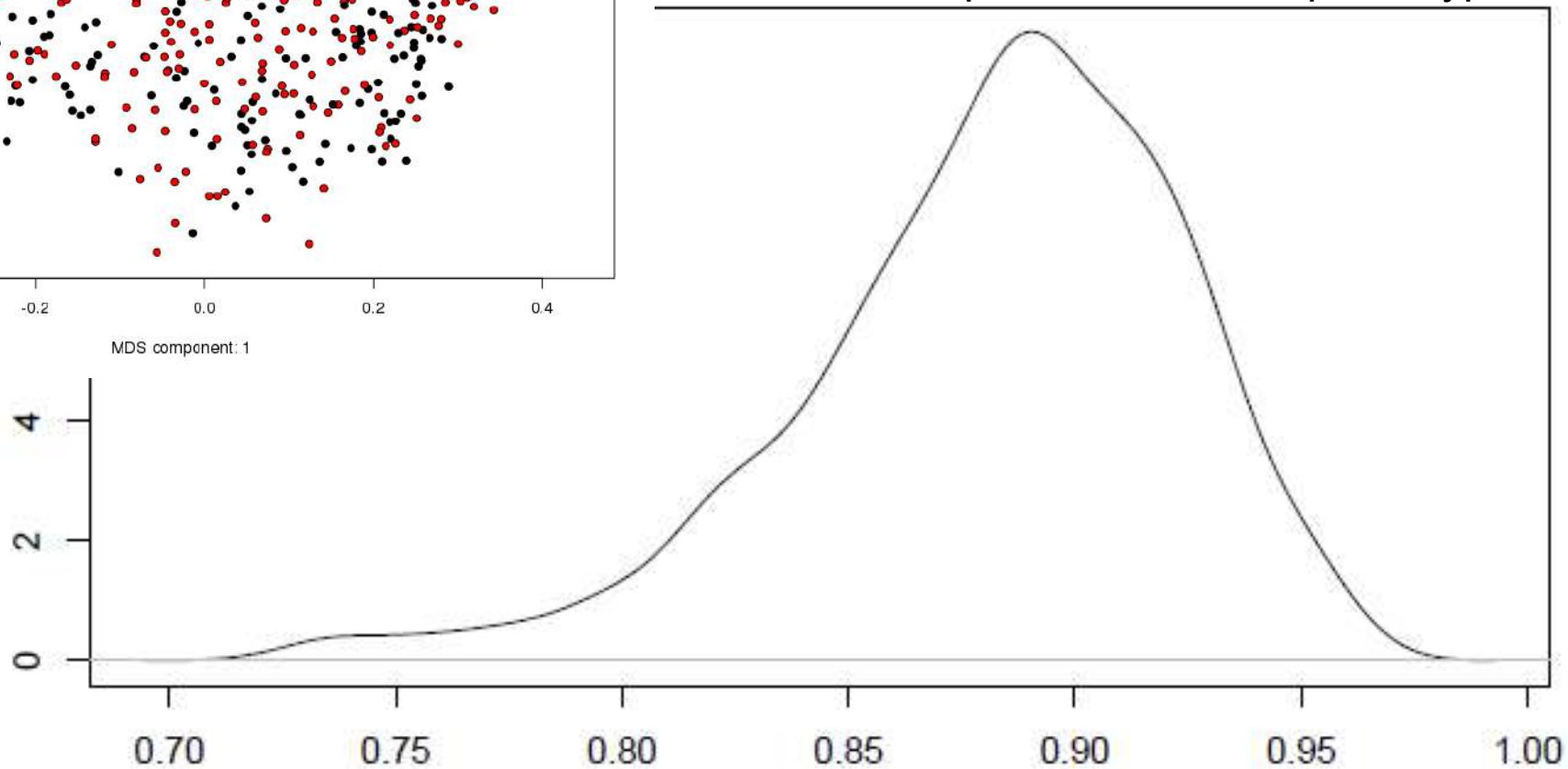


"blind" data science is not the answer

diarrhea or not?



microbiome predicts random "phenotype"



A vision for the future

- Biological "signal" often stronger than experimental biases
- Better experiments are better than more data
- Close integration between measurement devices, bioinformatics, analytics, and biology is critical
- Education needs to be cross-disciplinary