Introduction to sequencing microbiota

Spencer Glantz, Ph.D.

Postdoctoral Associate, Oh Lab

spencer.glantz@jax.org

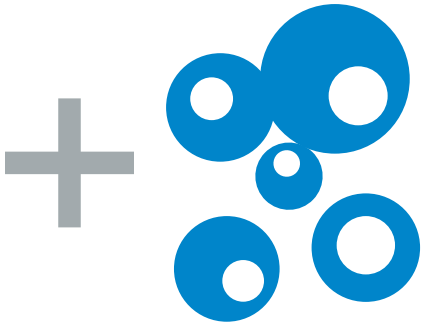
The Jackson Laboratory

Leading the search for tomorrow's cures

Big data to knowledge

Data science *Biological question*

```
101010101000
101010110101
010010001001
001001010101
```



New insights and hypotheses

→ Computational approaches can address biological questions nearly impossible to address in the wet lab alone…but having *all* the skills is challenging!

Pedagogical strategy

Overall goal: Use bioinformatics technologies/techniques to deliver insight into a biological question of interest

→ First we'll divide the overall goal into its component pieces:

1) What is the biology? How can we study it by sequencing?
2) What is a typical experimental & computational workflow?
3) What kind of biological insights can we derive?
4) How do we program at the command line?

→Then we'll try to put *everything* together

5) Synthesis: hands on module

1) What is the biology?
And how can we study it by sequencing?

Microbial communities are all around

human epithelia

soil

ocean

hot springs

Upon a closer look…

Microbiome

- fungi
- viruses
- bacteria
- archaea
- protists
- micro-animals

How do we identify members of microbiomes?

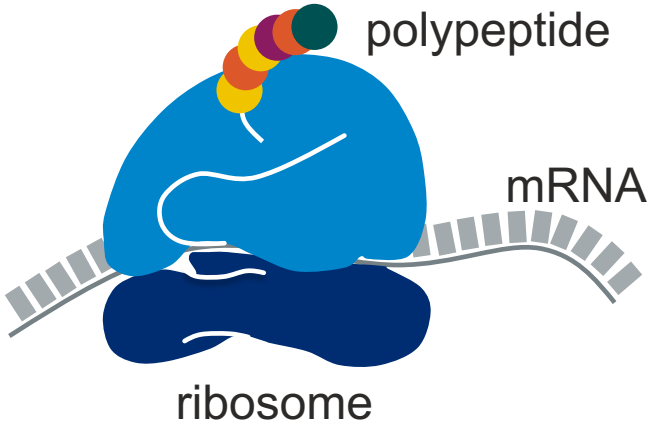
We need a “microbial fingerprint”

- Fingerprints are *both* **universally present** on all people *and* **unique**

The 16S ribosomal RNA as a microbial fingerprint

- Fingerprints are *both* **universally present** on all people *and* **unique**

- The **ribosome** is essential for survival across all kingdoms of life and is thus **highly conserved**



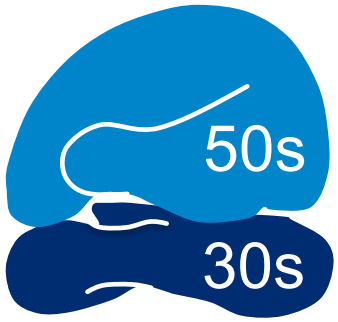
The 16S ribosomal RNA as a microbial fingerprint

- Fingerprints are *both* **universally present** on all people *and* **unique**

- Specifically, the **16S rRNA** component of the ribosome is highly **conserved** among bacteria/archaea, yet contains **hypervariable** regions

23S rRNA
5S rRNA
31 proteins

16S rRNA
21 proteins



ribosome

Hypervariable 16S regions can be used for species identification

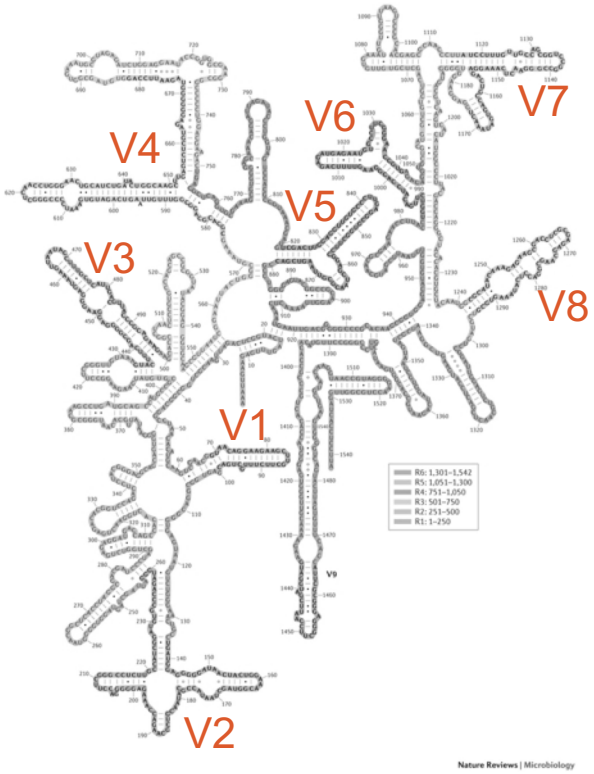
16S rRNA contains 9 variable regions

conserved region (non-specific)

variable region (species-specific)

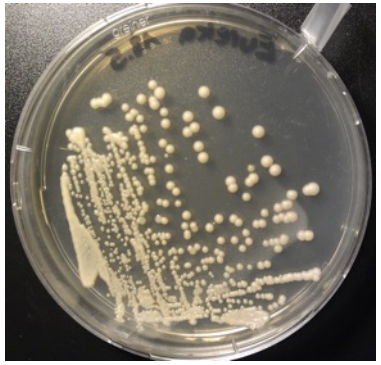
→ More distantly related species exhibit more divergent 16S RNA sequences

16S rRNA secondary structure



Microbial genomics suffers from lack of cultivation approaches

Isolate



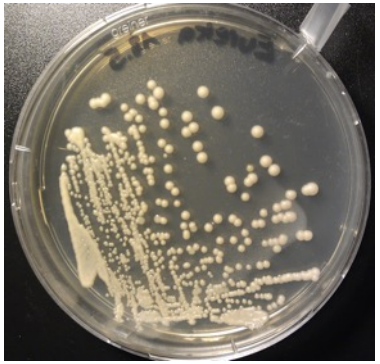
Genomics

"The estimate that fewer than 1% of the prokaryotes in most environments can be cultivated in isolation has produced a quandary: what is the significance of the field of modern microbial genomics if it is limited to culturable organisms?"

Schloss et al, Genome Biology, 2005

Metagenomics has revolutionized microbiome studies

Isolate



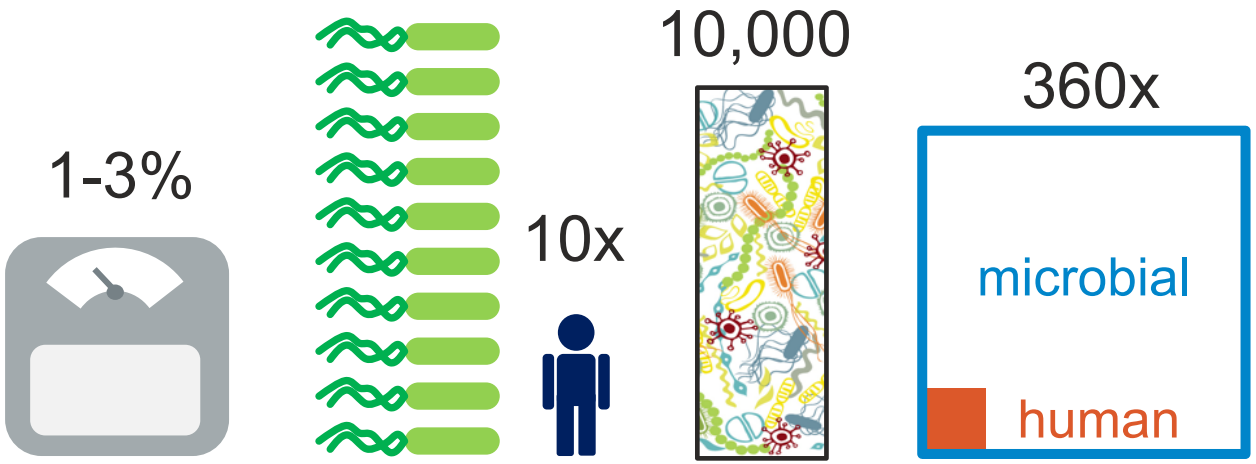
Genomics

Direct sequencing

Metagenomics

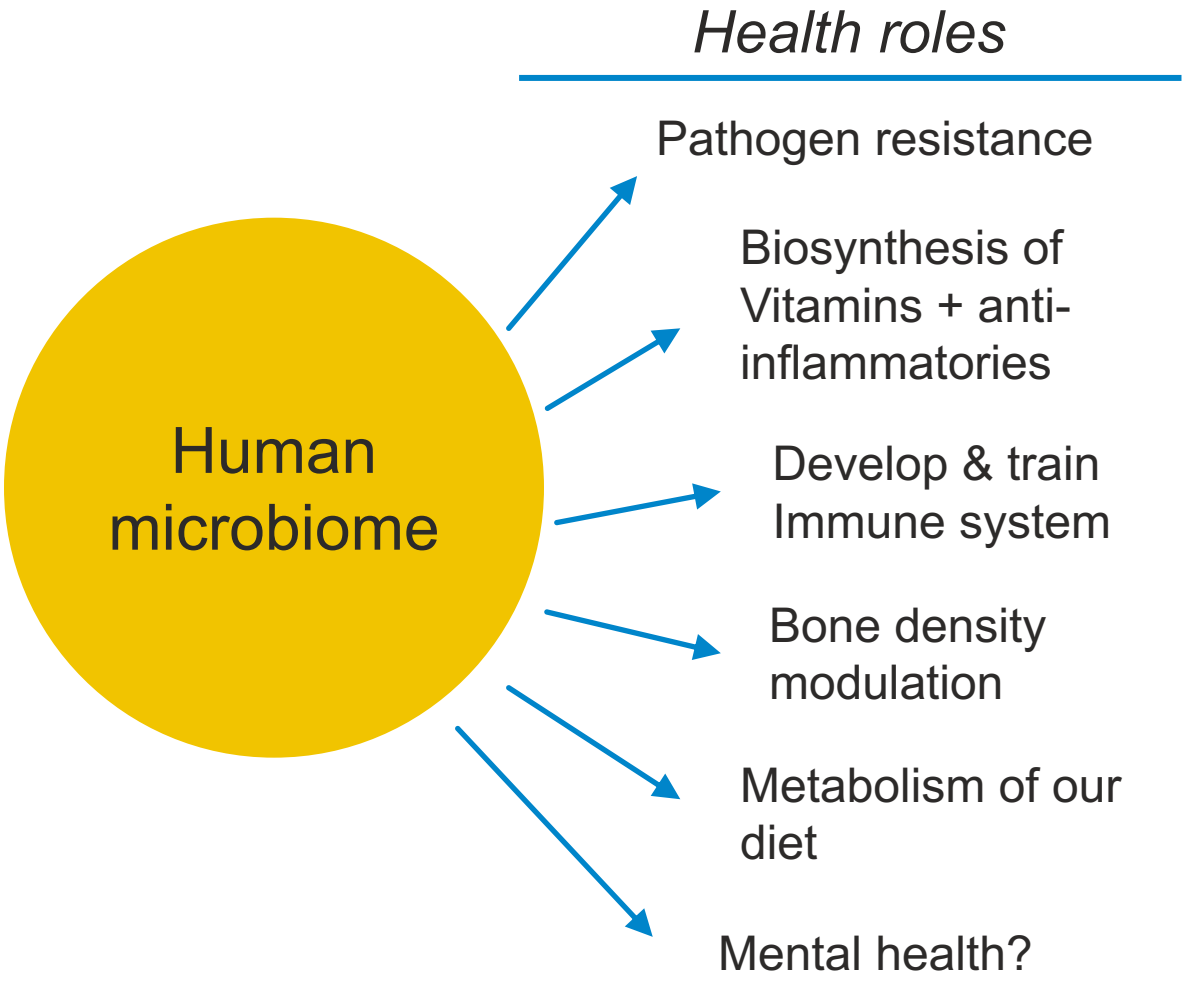
BIG data

The human microbiome by the numbers



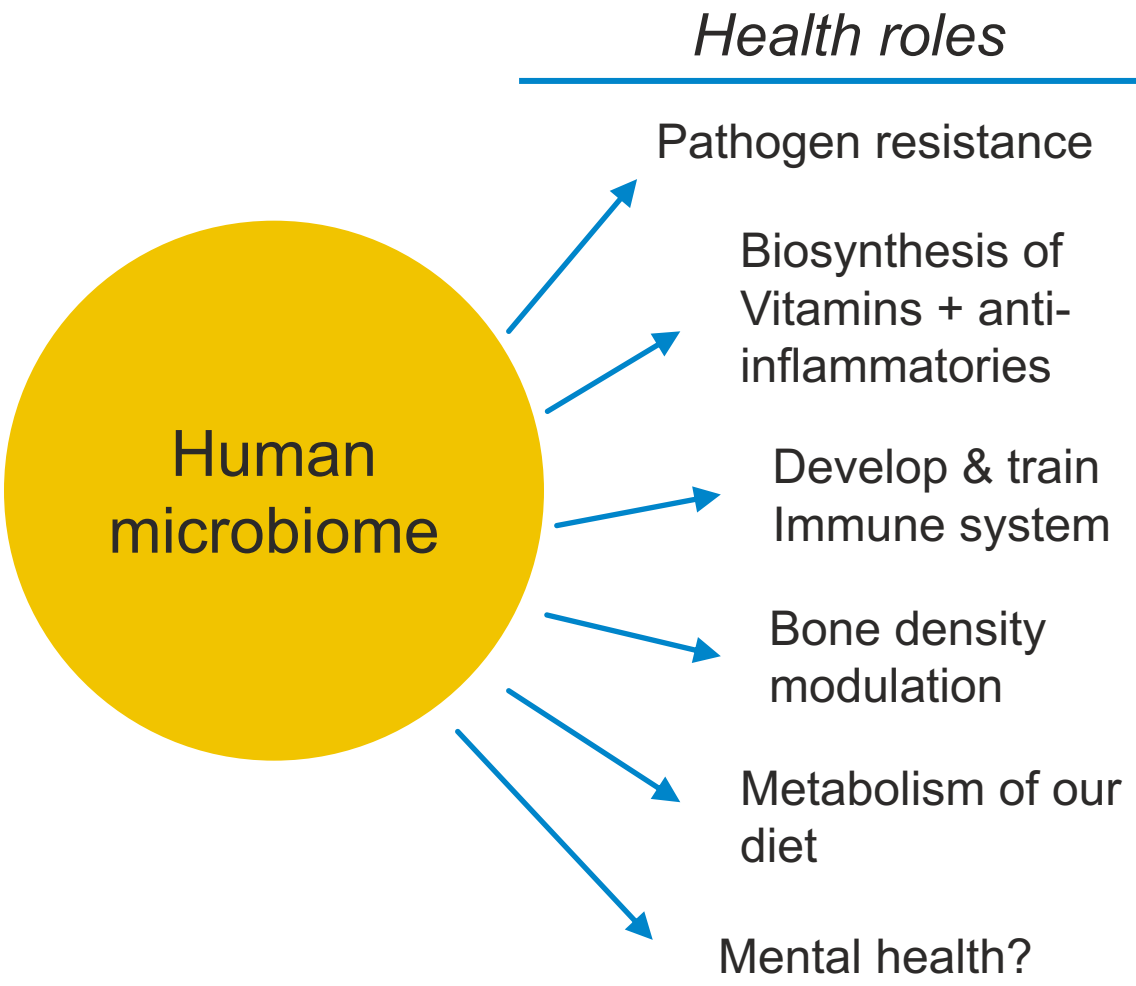
What are all these microbes doing??

As a community, microbes actually play diverse physiological roles

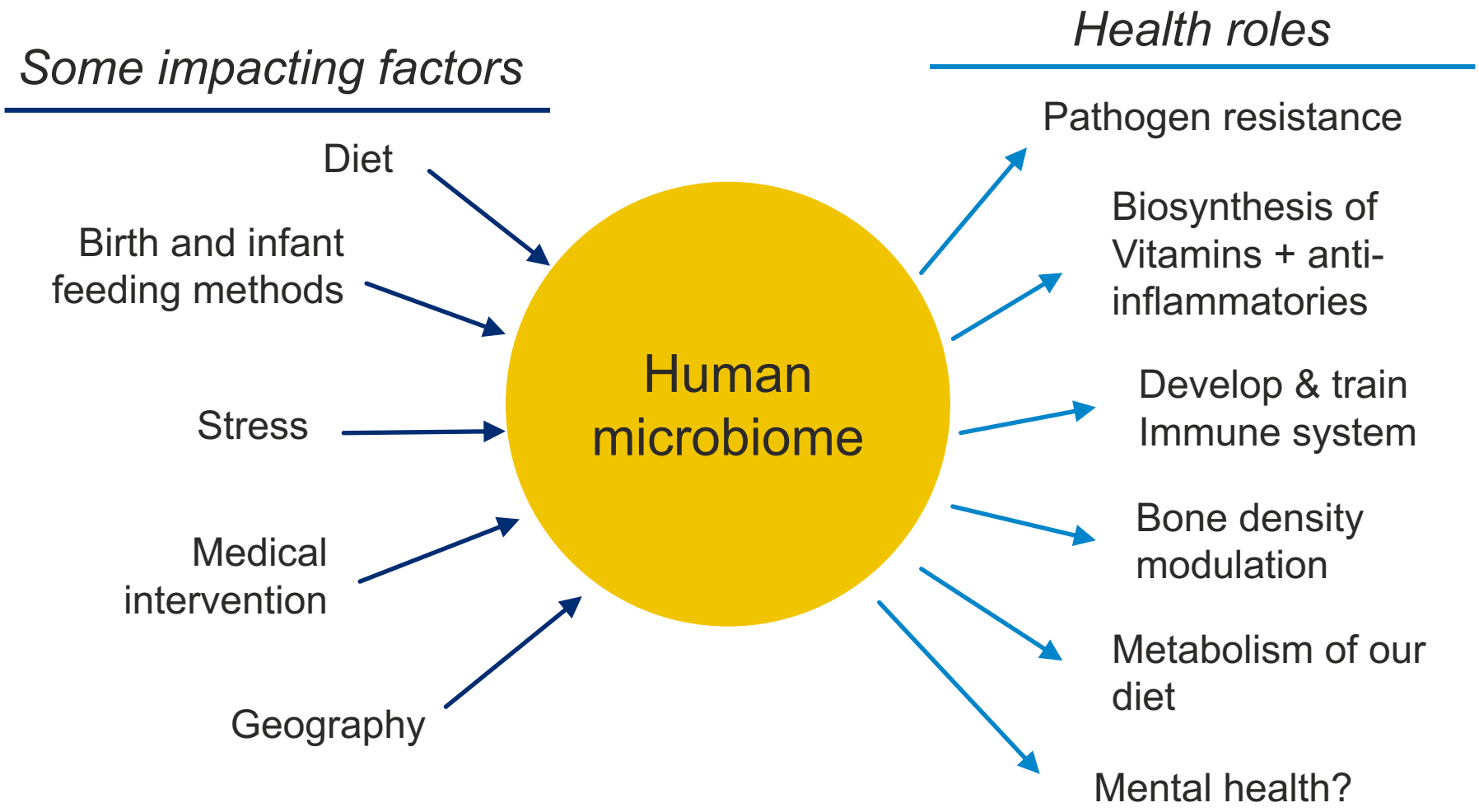


As a community, microbes actually play diverse physiological roles

Do we all have the same microbiome?



The human microbiome is dynamic



How can we better understand our microbiomes?

2) What is a typical experimental & computational workflow?

Key elements of a microbiome study

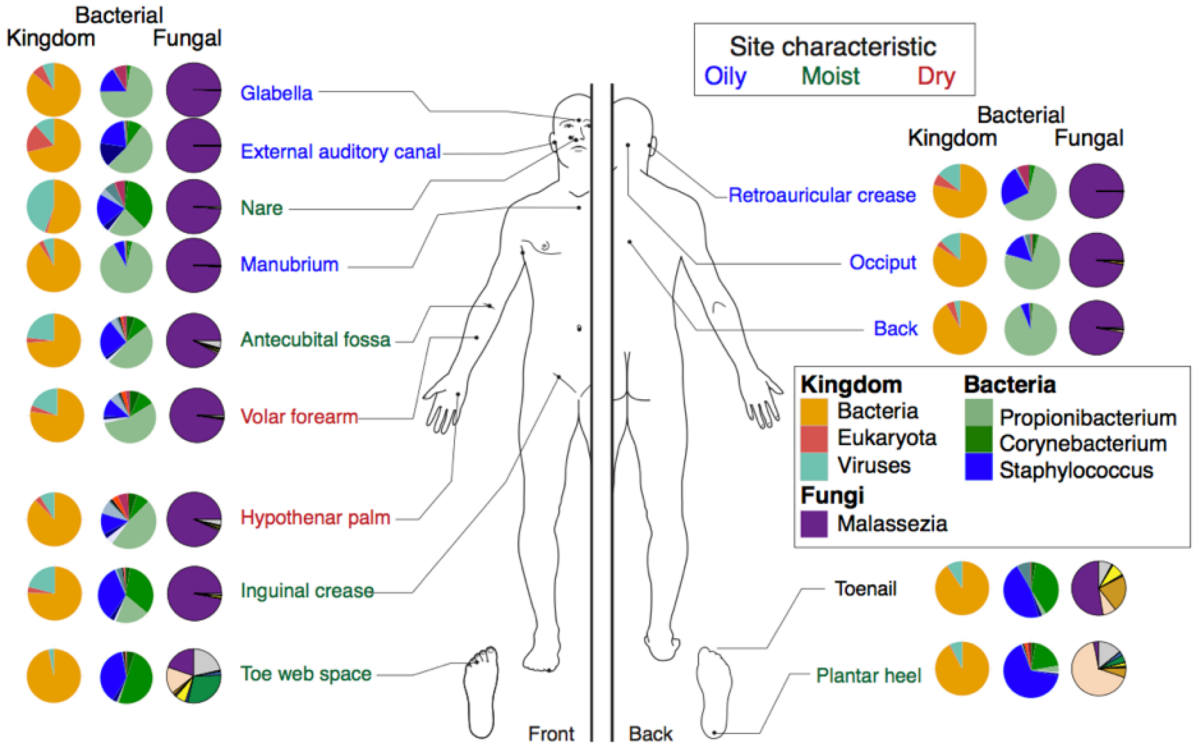
Goodrich et al, Cell (2014)

Key elements of a microbiome study

Goodrich et al, Cell (2014)

Studies can be descriptive

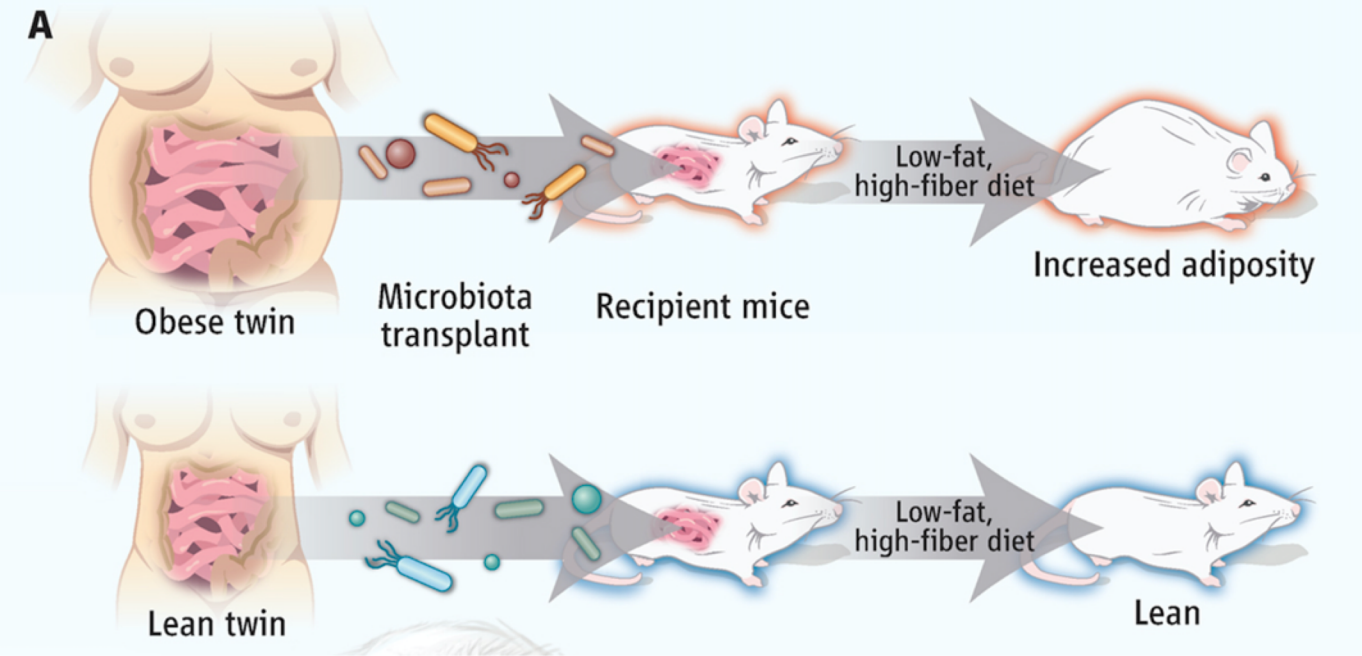
How do bacterial communities differ depending on the physical characteristics of different body sites?



Oh et al., *Nature* 2014; *Cell* 2016

Or studies can be case-control

What happens to germ-free mice fed a consistent diet, but transplanted with microbiota from lean v. obese donors?



Walker and Parkhill, Science, 2013

What biological questions can we address with our big data approach?

A sampling of microbiome questions

What microbes live where?

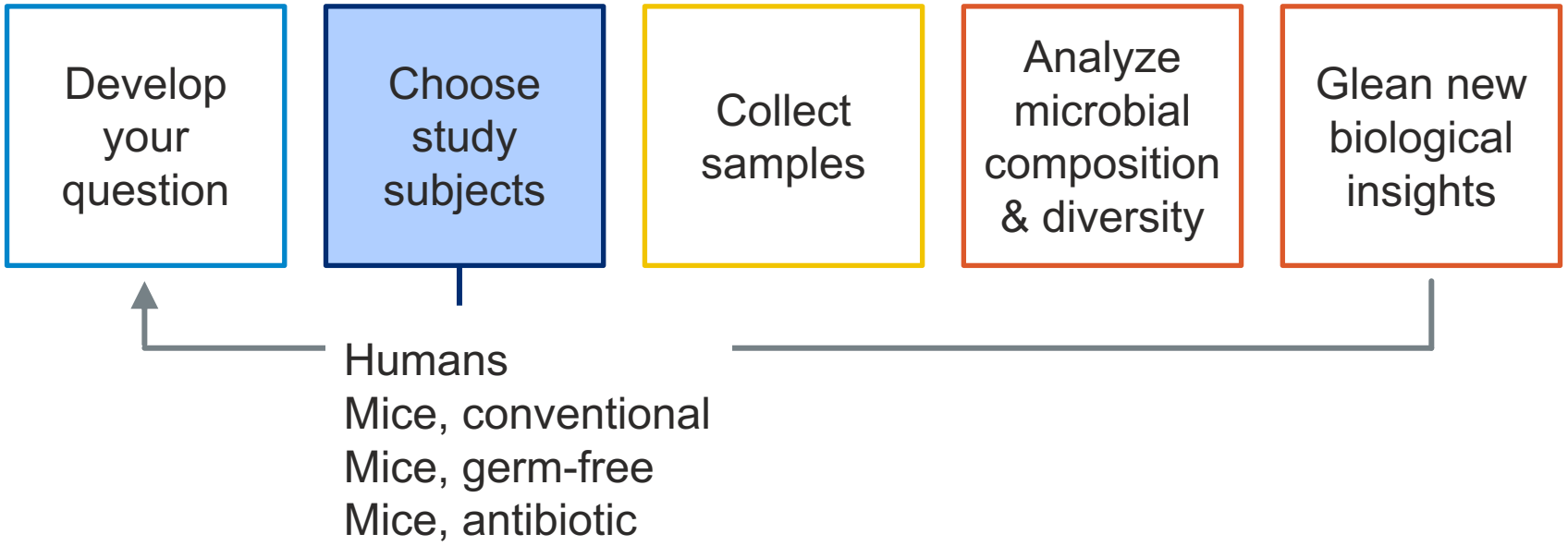
What drives resilience of a microbiome and enables resistance to change?

How do microbes interact with the host (e.g. immune system)?

How does microbiome composition affect microbiome function?

Are changes in the microbiome causative of disease or result from disease or both?

Key elements of a microbiome study



Goodrich et al, Cell (2014)

Key elements of a microbiome study

Humans
Mice, conventional
Mice, germ-free
Mice, antibiotic

How many subjects do we need?

Goodrich et al, Cell (2014)

Sample size is related to our ability to make a claim

→ **Example: Is this coin biased or not?**

4 flips: 2 heads, 2 tails

1000 flips: 333 heads, 667 tails

Sample size is related to our ability to make a claim

→ **Example: Is this coin biased or not?**

4 flips: 2 heads, 2 tails

1000 flips: 333 heads, 667 tails

4 flips: 1 heads, 3 tails

1000 flips: 492 heads, 508 tails

Sample size is related to our ability to make a claim

- Sample size = # of subjects (e.g. number of individual coin flips)

- Null hypothesis = hypothesis that there is no difference from the expected mean (e.g. fair coin)

- Statistical power = ability to detect a true difference from the null hypothesis – depends in part on the sample size (e.g. can we actually discover when a coin is not fair?)

Sample size ↑ Statistical power ↑

Study design

There are 4 parameters that will help us determine our required sample size to ensure that we observe a certain experimental effect with high probability:

Coin flip analogy	Statistical term	Sample size needed
• How weighted is the coin? (55% heads or 90% heads)	• Effect size (δ) ↑	↓
• How far does the result typically deviate from 50% heads?	• Variation of the data (σ) ↑	↑
• What should be our tolerance for saying the coin is biased when it actually isn't? (5%)	• Significance level (α) ↑	↓
• To what degree do we want to be able to detect a biased coin when it is actually biased? (80%)	• Statistical power (β) ↑	↑

Study design

Example:

We want to test the effect of diet on weight gain for two groups of germ-free mice – one of which will be transplanted with a gut microbiome from an obese donor group and another with a gut microbiome from a lean donor group.

*How many mice do we need to study to detect at least a **10% difference** in mean weight between the two groups at a **significance level of 5%** (α) and at a **statistical power (β) of 80%**? From past experimentation you know that typical **mouse weight standard deviation** (σ) **is 10%**.*

Study design

Example:

*How many mice do we need to study to detect at least a **10% difference** in mean weight between the two groups at a **significance level of 5%** (α) and at a **statistical power (β) of 80%**? From past experimentation you know that typical **mouse weight standard deviation** (σ) **is 10%**.*

(1) Compute the effect size:

$$\text{Effect size } (\delta) = \frac{\text{Variation between groups}}{\text{Random variation}} = \frac{\text{mean difference}}{\text{standard deviation}} = \frac{\mu - \mu_0}{\sigma}$$

$$\textbf{Effect size } (\boldsymbol{\delta}) = \frac{0.1}{0.1} = 1$$

Study design

Example:

*How many mice do we need to study to detect at least a **10% difference** in mean weight between the two groups at a **significance level of 5%** (α) and at a **statistical power (β) of 80%**? From past experimentation you know that typical **mouse weight standard deviation** (σ) **is 10%**.*

(1) Compute the effect size:

$$\text{Effect size } (\delta) = \frac{\text{mean difference}}{\text{standard deviation}}$$

$$\textbf{Effect size } (\boldsymbol{\delta}) = \frac{0.1}{0.1} = 1$$

(2) Compute the required sample size*:

$$\text{Sample size (n)} = \frac{2*(Z_{\alpha/2} + Z_{\beta})^2}{\delta} = \frac{15.68}{1}$$

You will need at least 16 mice per group

*assumes normally distributed data

Study design

Example:

*How many mice do we need to study to detect at least a **10% difference** in mean weight between the two groups at a **significance level of 5%** (α) and at a **statistical power (β) of 80%**? From past experimentation you know that typical **mouse weight standard deviation** (σ) **is 10%**.*

(1) Compute the effect size:

$$\text{Effect size } (\delta) = \frac{\text{mean difference}}{\text{standard deviation}}$$

$$\textbf{Effect size } (\boldsymbol{\delta}) = \frac{0.1}{0.1} = 1$$

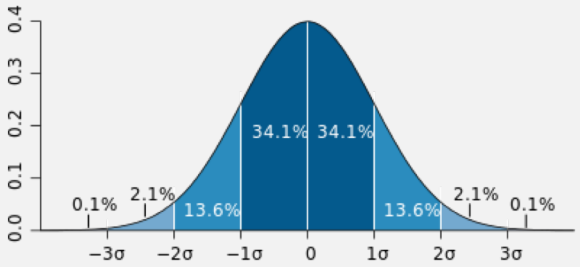
(2) Compute the required sample size*:

$$\text{Sample size (n)} = \frac{2*(Z_{\alpha/2} + Z_{\beta})^2}{\delta} = \frac{15.68}{1}$$

You will need at least 16 mice per group

*assumes normally distributed data

A Z test gives critical values for each significance level assuming normally distributed data around a mean of 0

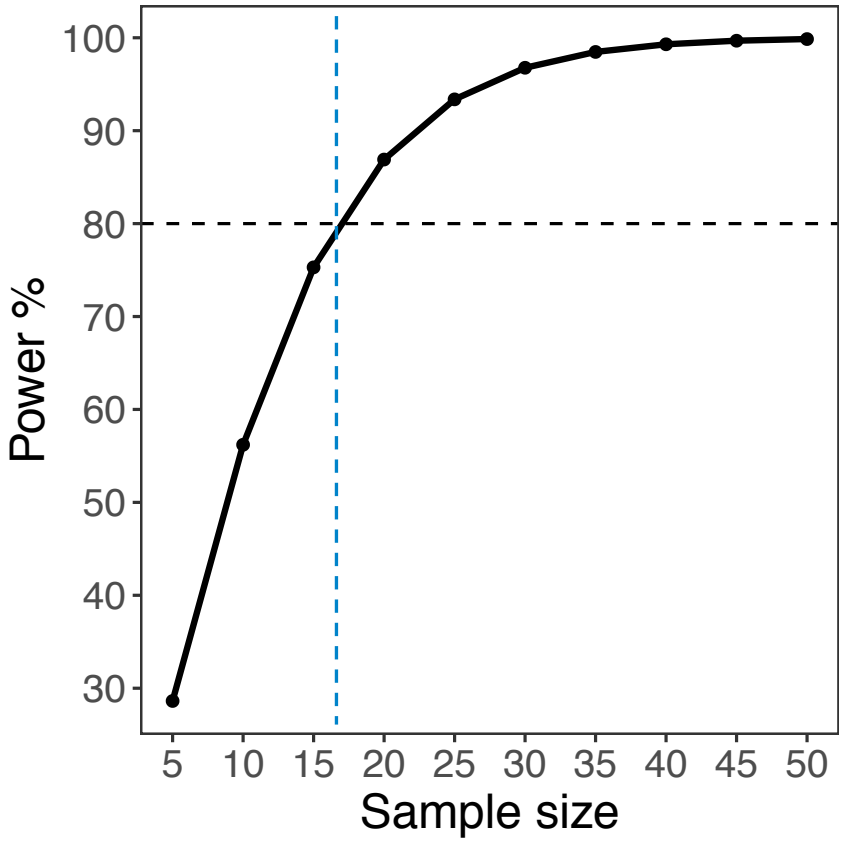


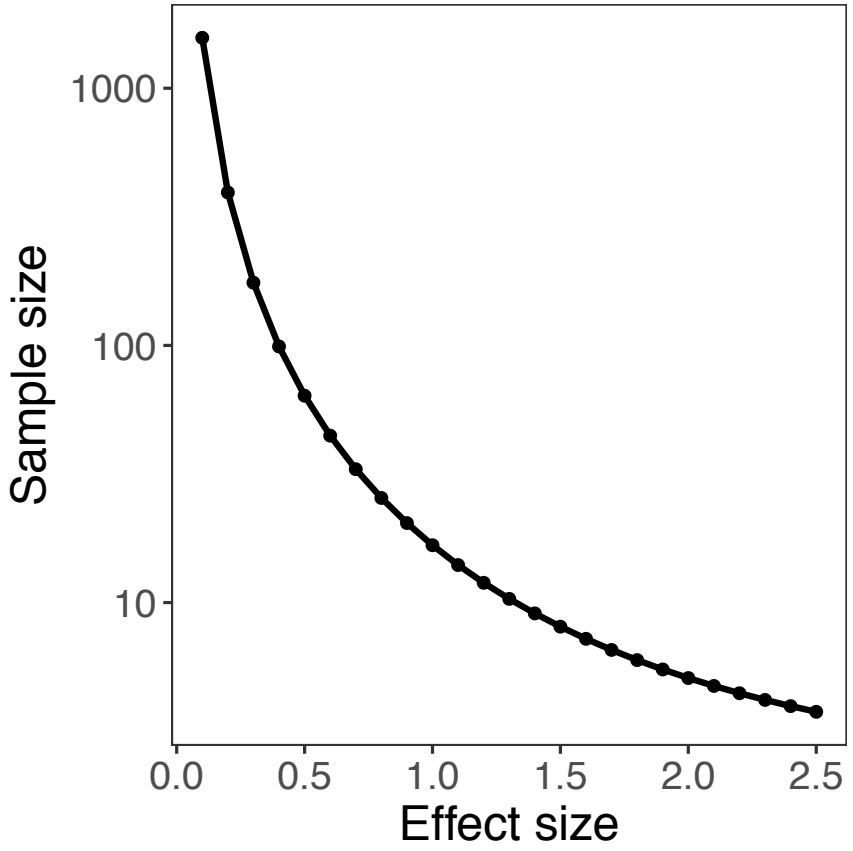
Example:

$Z_{0.025}$ = 1.96

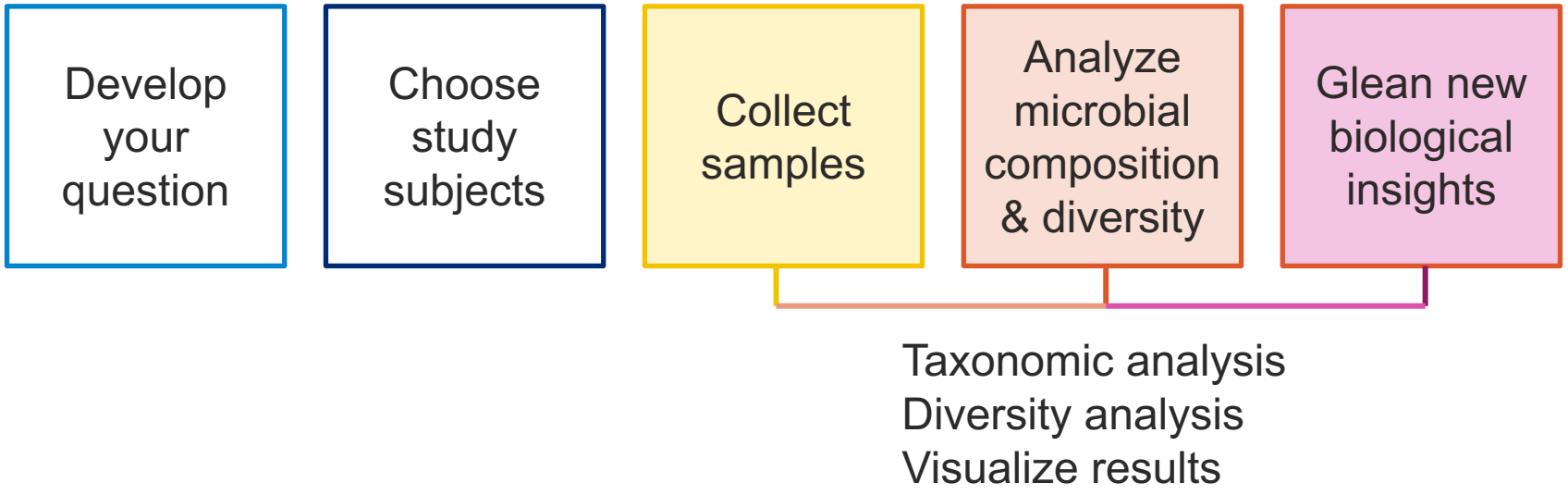
For data normally distributed around 0, only 2.5% of the values will be greater than 1.96

Study design



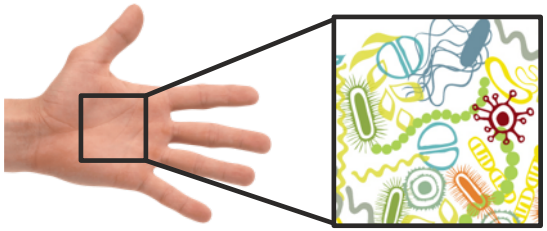


Study design



A workflow for 16S analysis

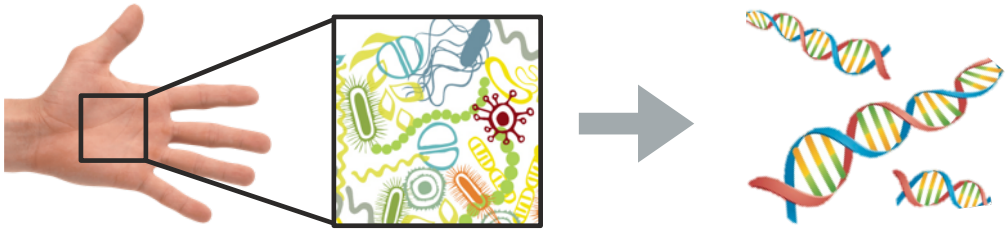
1 Collect sample
(e.g. skin swab)



A workflow for 16S analysis

1 Collect sample (e.g. skin swab)

2 Total DNA extraction



A workflow for 16S analysis

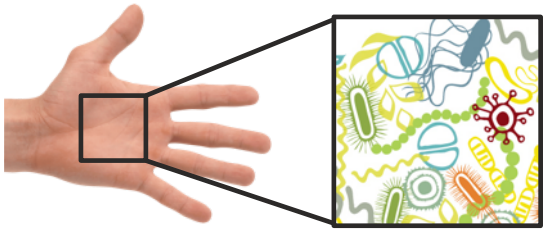
1 Collect sample (e.g. skin swab)

2 Total DNA extraction

3 Amplify variable region of 16S rRNA gene

A workflow for 16S analysis

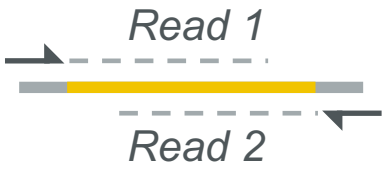
1 Collect sample (e.g. skin swab)



2 Total DNA extraction

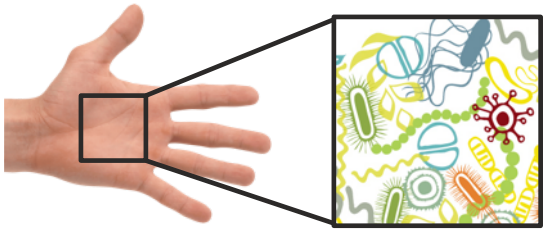
3 Amplify variable region of 16S rRNA gene

4 Sequence 16S rRNA genes



A workflow for 16S analysis

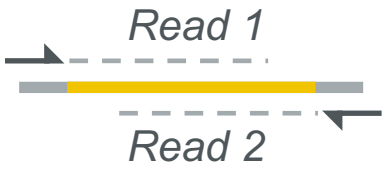
1 Collect sample (e.g. skin swab)



2 Total DNA extraction

3 Amplify variable region of 16S rRNA gene

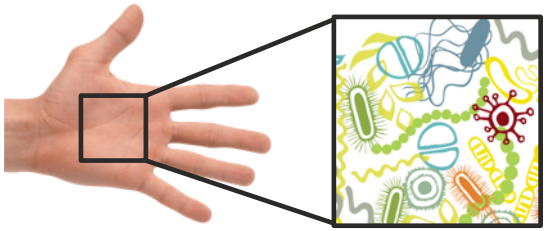
4 Sequence 16S rRNA genes



5 Processing, QC, and assembly of reads

A workflow for 16S analysis

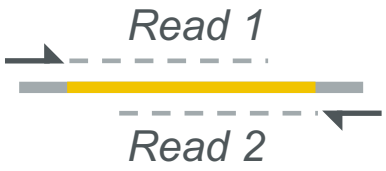
1 Collect sample (e.g. skin swab)



2 Total DNA extraction

3 Amplify variable region of 16S rRNA gene

4 Sequence 16S rRNA genes

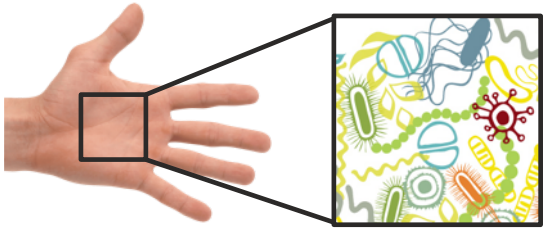


5 Processing, QC, and assembly of reads

6 Generate a representative sequence for each species and assign taxonomy

A workflow for 16S analysis

1 Collect sample (e.g. skin swab)



2 Total DNA extraction

3 Amplify variable region of 16S rRNA gene

4 Sequence 16S rRNA genes

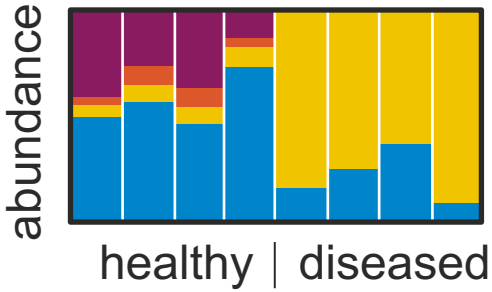
5 Processing, QC, and assembly of reads

6 Generate a representative sequence for each species and assign taxonomy

Typical output →

Who? How much?

relative abundance plot



diversity plot

A real world 16S application

Oh *et al. Genome Medicine* 2012, **4**:77
http://genomemedicine.com/content/4/10/77

RESEARCH Open Access

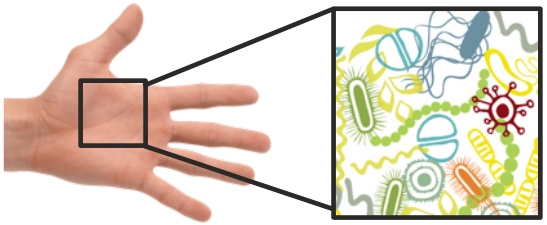
Shifts in human skin and nares microbiota of healthy children and adults

Julia Oh[1], Sean Conlan[1], Eric C Polley[2], Julia A Segre[1*†] and Heidi H Kong[3*†]

Question: *How does the skin microbiome change as children progress through puberty?*

Sample collection

1 Collect samples



- 28 healthy individuals (no current or prior chronic skin or medical conditions)
- Range in age from 2 – 40 years
- Puberty assessed by “Tanner staging”

“exclusion criteria” and metadata are key!

- No bathing, washing, antimicrobial treatment 7 days prior to sampling
- Swabs from nares (N), antecubital fossa (Af), volar forearm (Vf) and popliteal fossa (Pf)

DNA extraction

2 Total DNA extraction = cell lysis followed by DNA isolation

Cell lysis

Mechanical

Tissue lyzer II

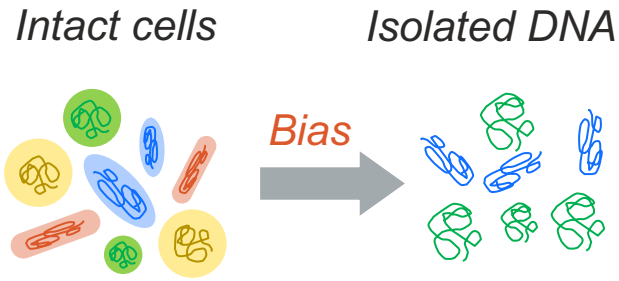
- Heating
- Freeze/thaw
- Bead-beating

Chemical/Enzymatic

- Detergents
  - Triton-X-100
- Cell wall disruption
  - Lysozyme
  - Mutanolysin
  - Lysostaphin
- Protein degradation
  - Proteinase K
- RNA digestion
  - RNAse A

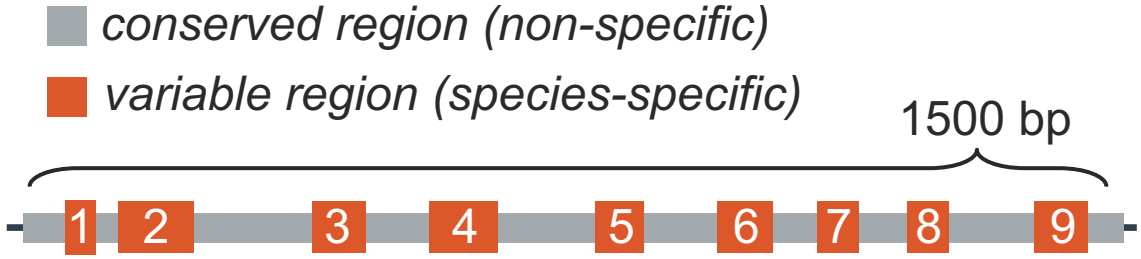
Bias in cell lysis can severely distort the measured composition

→ “mock community“



Amplify hypervariable regions of interest from 16S gene

16S rRNA contains 9 variable regions



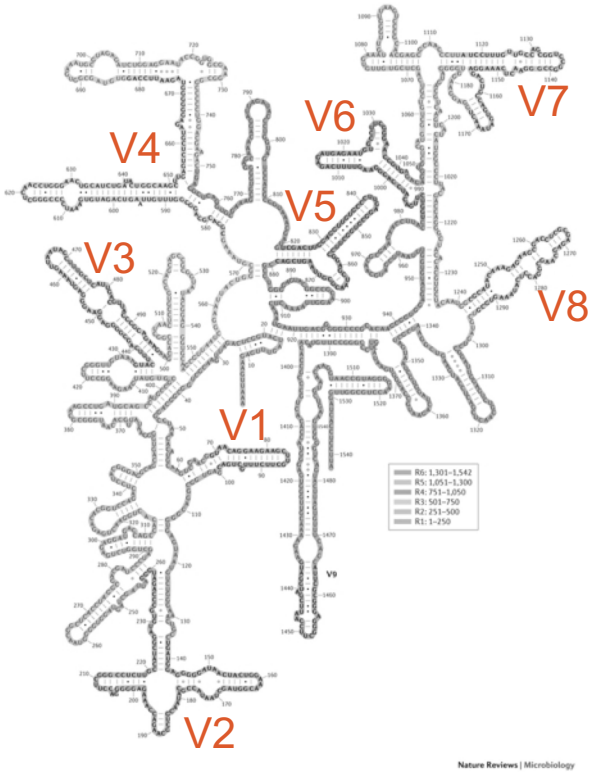
→ Conserved regions serve as universal primer binding sites

→ Variable regions can be used to distinguish organisms at various taxonomic levels

→ More distantly related species exhibit more divergent 16S RNA sequences

How much of the 16S gene do we need to sequence to distinguish between organisms?

16S rRNA secondary structure



Hypervariable 16S regions can be used for species identification

Example: We are studying a collection of sentences and want to differentiate them and determine how many of each type we have

1.	The	cow	jumps	over	the	moon
2.	The	cow	jumps	over	the	moon
3.	The	cow	jumps	under	the	sun
4.	The	cow	jumps	under	the	sun
5.	The	horse	jumps	over	the	moon
6.	The	horse	jumps	over	the	moon
7.	The	horse	jumps	over	the	sun
8.	The	horse	jumps	under	the	sun

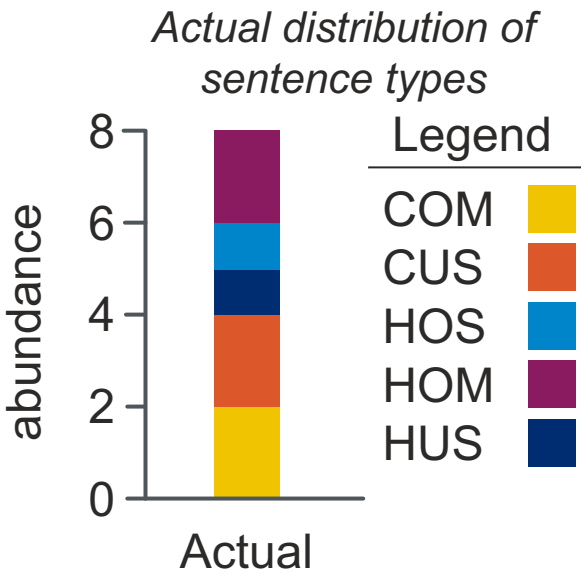
→ These sentences have some **conserved** (e.g. “the” “jump”) and some **variable** (e.g. (“over/under”) regions

→ How many variable regions do we need to consider to effectively catalog all 8 sentences

Hypervariable 16S regions can be used for species identification

Example: We are studying a collection of sentences and want to differentiate them and determine how many of each type we have

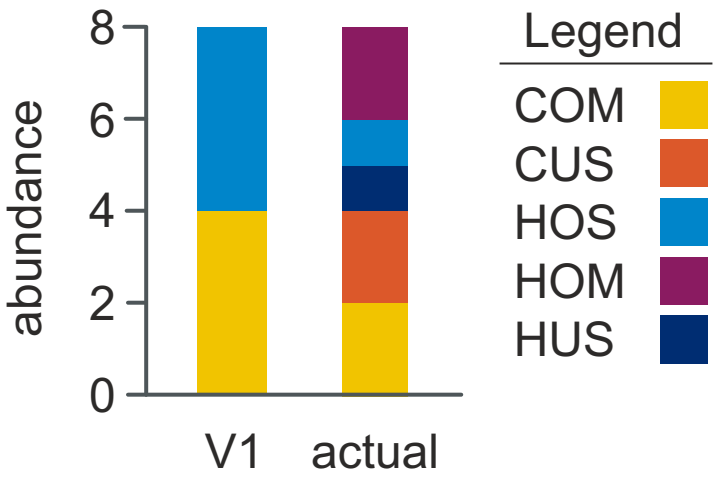
1.	The	cow	jumps	over	the moon
2.	The	cow	jumps	over	the moon
3.	The	cow	jumps	under	the sun
4.	The	cow	jumps	under	the sun
5.	The	horse	jumps	over	the moon
6.	The	horse	jumps	over	the moon
7.	The	horse	jumps	over	the sun
8.	The	horse	jumps	under	the sun



Hypervariable 16S regions can be used for species identification

	V1	V2	V3
1.	The cow	jumps over	the moon
2.	The cow	jumps over	the moon
3.	The cow	jumps under	the sun
4.	The cow	jumps under	the sun
5.	The horse	jumps over	the moon
6.	The horse	jumps over	the moon
7.	The horse	jumps over	the sun
8.	The horse	jumps under	the sun

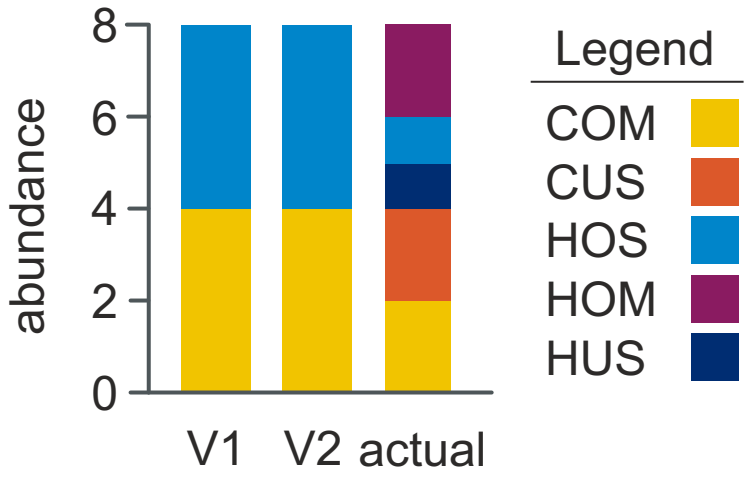
Regions considered	variants	# of variants
V1	cow, horse	2



Hypervariable 16S regions can be used for species identification

	V1		V2		V3
1.	The	cow	jumps	over	the moon
2.	The	cow	jumps	over	the moon
3.	The	cow	jumps	under	the sun
4.	The	cow	jumps	under	the sun
5.	The	horse	jumps	over	the moon
6.	The	horse	jumps	over	the moon
7.	The	horse	jumps	over	the sun
8.	The	horse	jumps	under	the sun

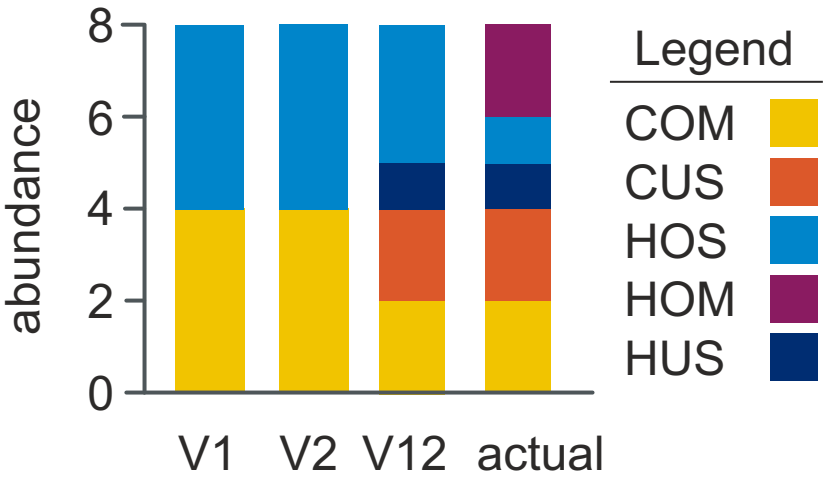
Regions considered	variants	# of variants
V1	cow, horse	2
V2	over, under	2



Hypervariable 16S regions can be used for species identification

	V1		V2		V3
1.	The	cow	jumps	over	the moon
2.	The	cow	jumps	over	the moon
3.	The	cow	jumps	under	the sun
4.	The	cow	jumps	under	the sun
5.	The	horse	jumps	over	the moon
6.	The	horse	jumps	over	the moon
7.	The	horse	jumps	over	the sun
8.	The	horse	jumps	under	the sun

Regions considered	variants	# of variants
V1	cow, horse	2
V2	over, under	2
V1 V2	cow-over horse-over cow-under horse-under	4

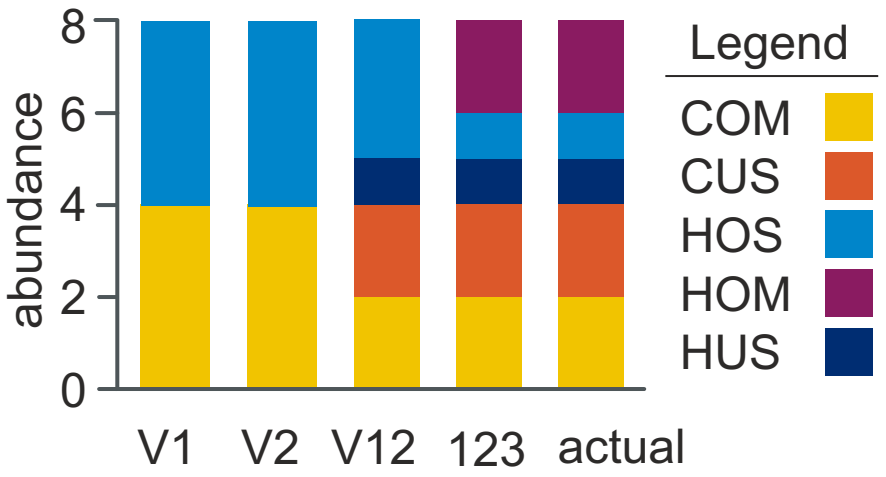


Hypervariable 16S regions can be used for species identification

		V1		V2		V3
1.	The	cow	jumps	over	the	moon
2.	The	cow	jumps	over	the	moon
3.	The	cow	jumps	under	the	sun
4.	The	cow	jumps	under	the	sun
5.	The	horse	jumps	over	the	moon
6.	The	horse	jumps	over	the	moon
7.	The	horse	jumps	over	the	sun
8.	The	horse	jumps	under	the	sun

The more variable regions considered, the higher the resolution

Regions considered	variants	# of variants
V1 V2 V3	cow-over-moon horse-over-moon horse-over-sun cow-under-sun horse-under-sun	5

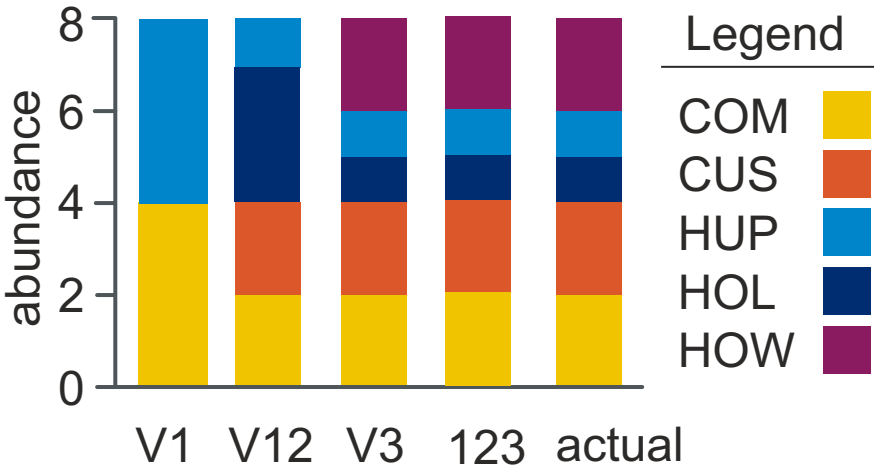


Hypervariable 16S regions can be used for species identification

	V1		V2		V3
1. The	cow	jumps	over	the	moon
2. The	cow	jumps	over	the	moon
3. The	cow	jumps	under	the	sun
4. The	cow	jumps	under	the	sun
5. The	horse	jumps	under	the	pond
6. The	horse	jumps	over	the	lake
7. The	horse	jumps	over	the	wood
8. The	horse	jumps	over	the	wood

But…the exact number of regions you need depends on composition

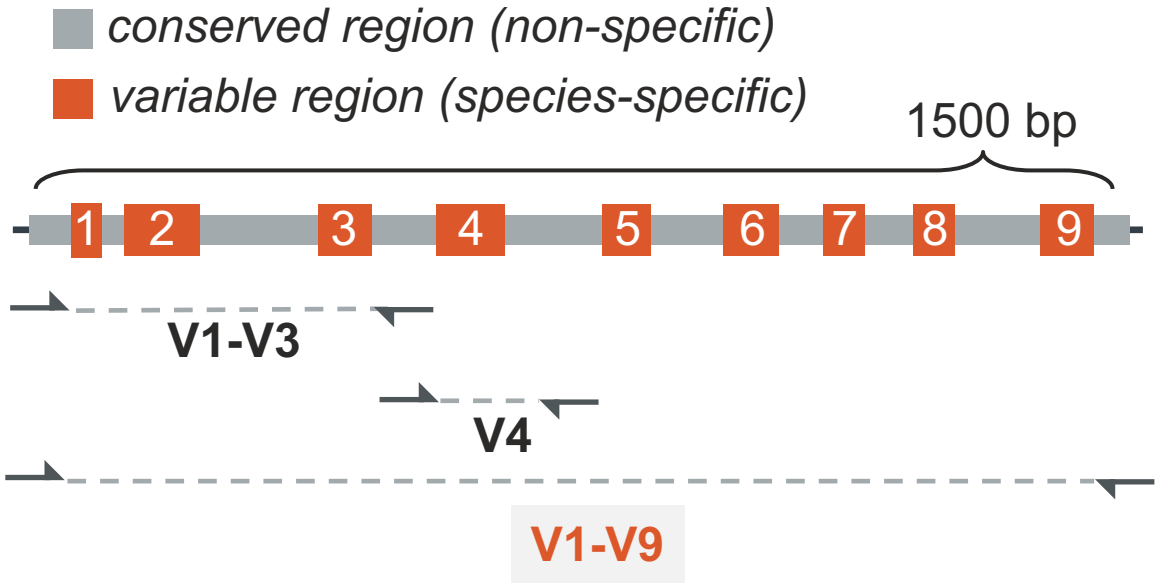
Regions considered	variants	# of variants
V1 V2 V3	cow-over-moon horse-over-moon horse-over-sun cow-under-sun horse-under-sun	5



Hypervariable 16S regions can be used for species identification

Phylogenetic resolution is a tradeoff between read length and read depth

→ Different read lengths can be achieved by different sequencing technologies



Length	Resolution
~480 bp	Genus, species
~255 bp	Family, genus
~1500 bp	Species, strain

16S Amplification occurs in 2 steps

→ We want to amplify copies of the 16S gene from our patient metagenomes so we have enough material for sequencing

1. PCR

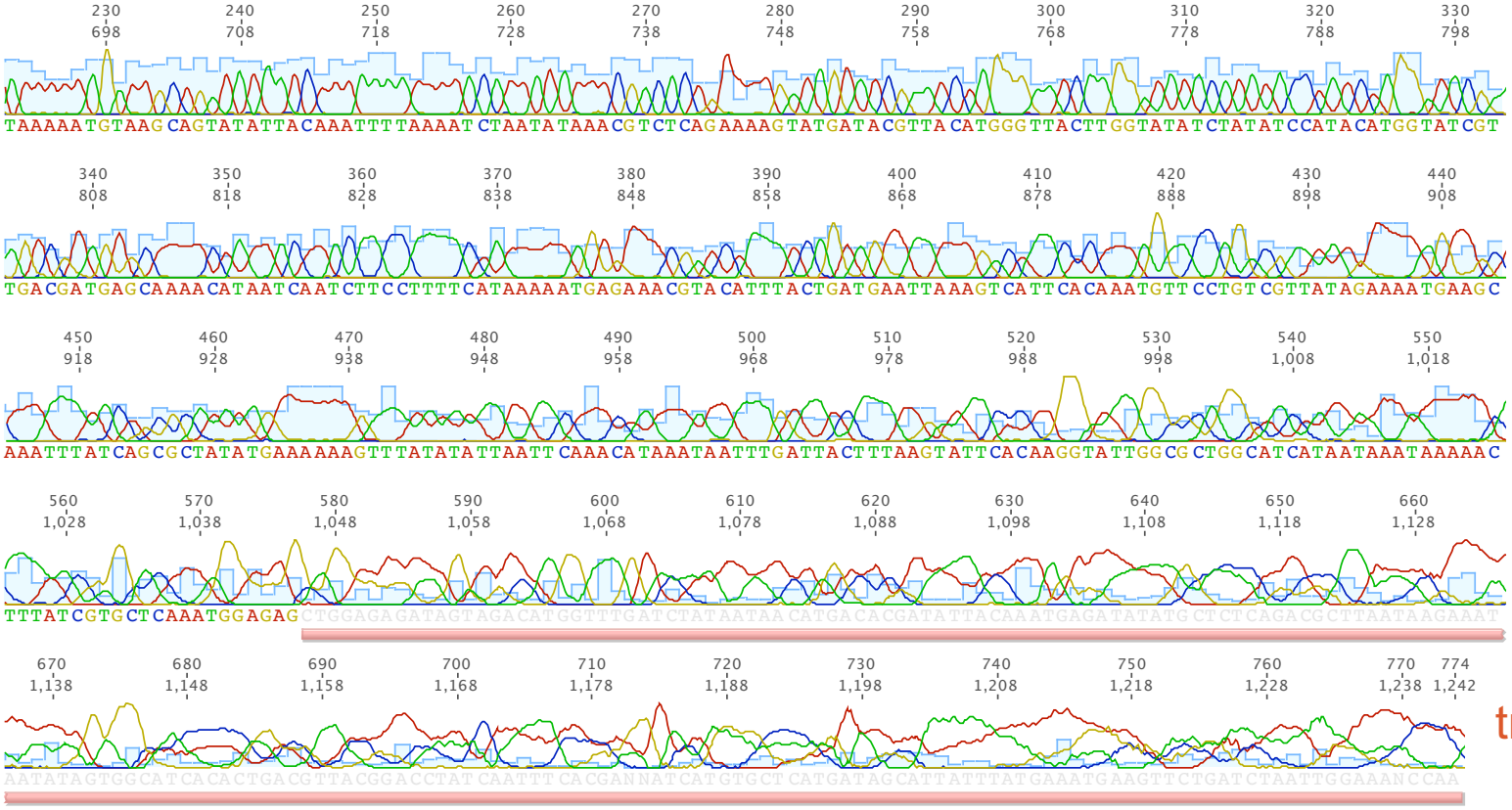
   very sensitive
   error prone

2. Clone into bacteria

   Huge amplification
   High fidelity
   Requires starting material

Sequence processing

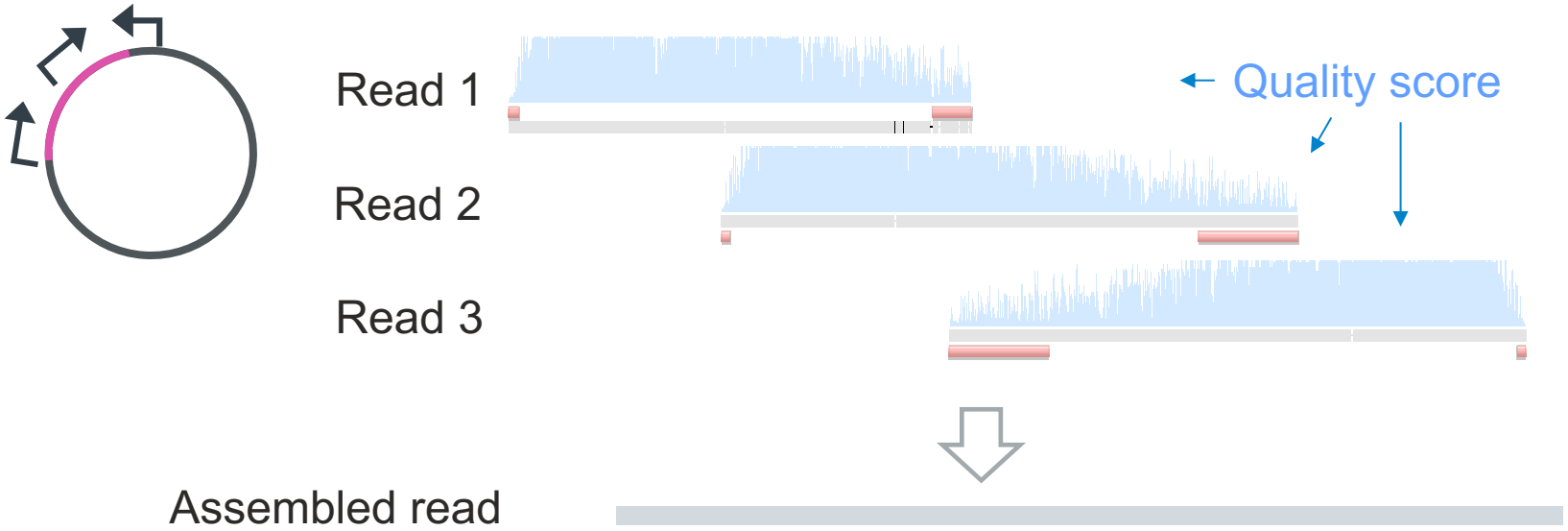
5 Processing and QC of reads: base calling & trimming



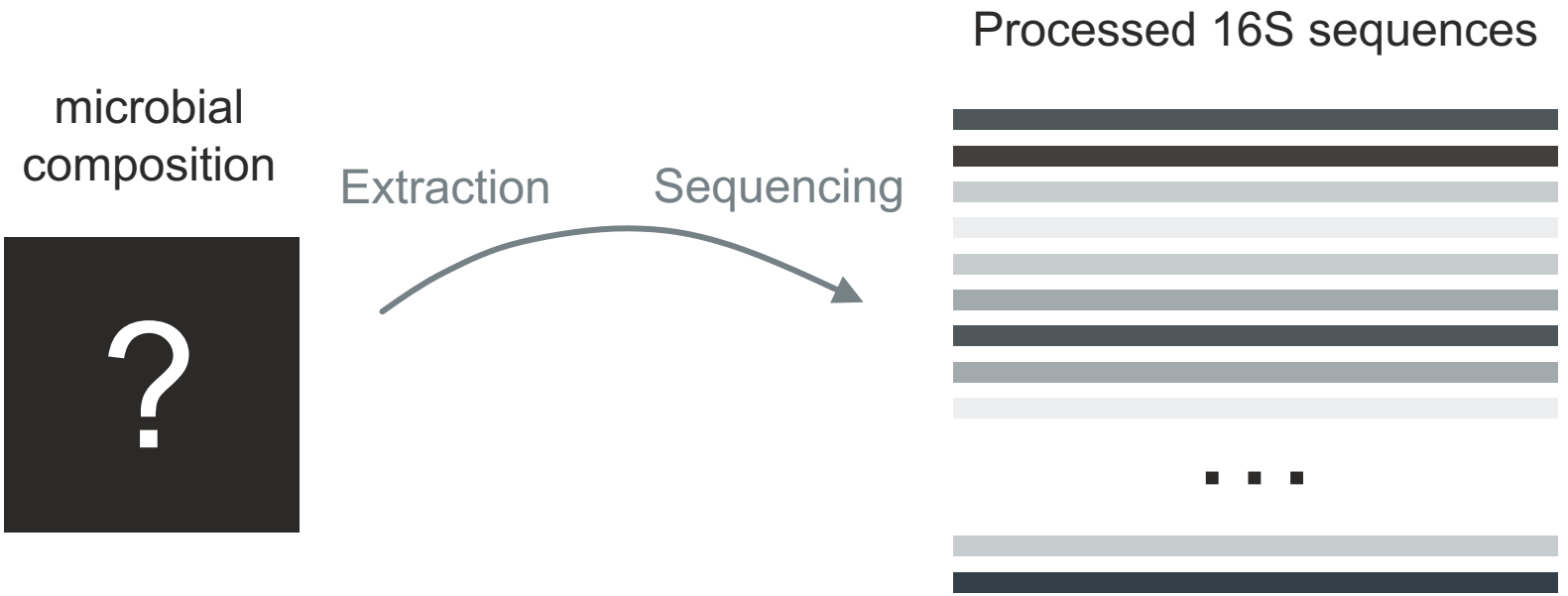
Remove (i) chimeras and (ii) sequences that align to the human genome (E<0.1)

Read assembly for each clone

There are three sequencing reads for each clone



Where do we want to go next?

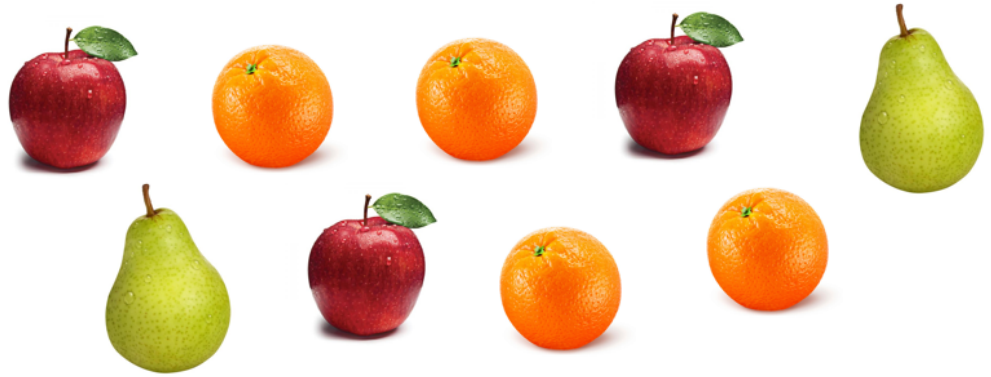


The challenge of metagenomics is that the sample is mixed!

→ Which 16S sequence came from which bacterium?
→ Does every unique 16S sequence come from a unique microbe?

It is trivial to catalog identical objects

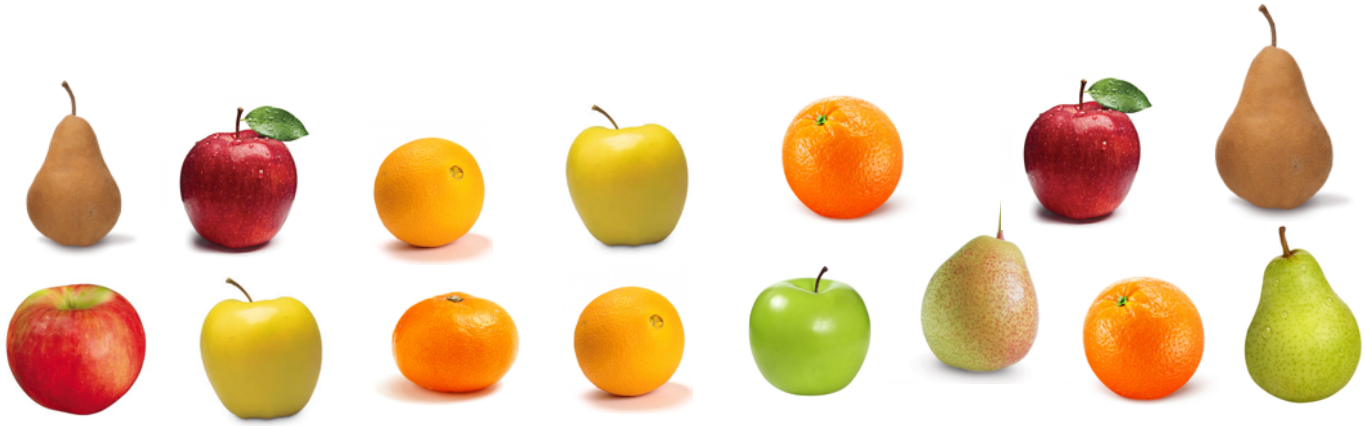
mixed sample



cataloged

Cataloging variable objects is hard

mixed sample



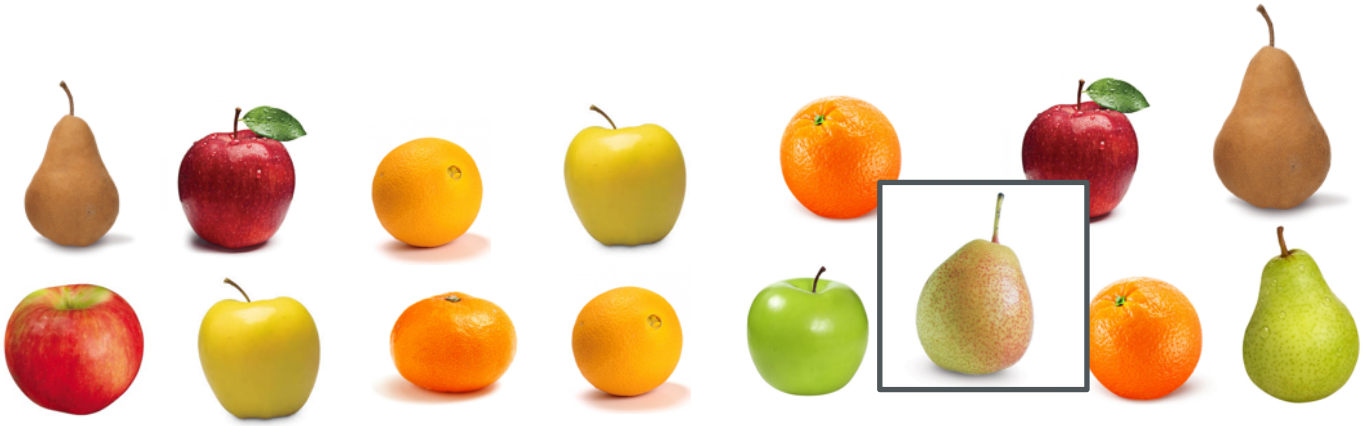
Cataloging variable objects is hard

mixed sample

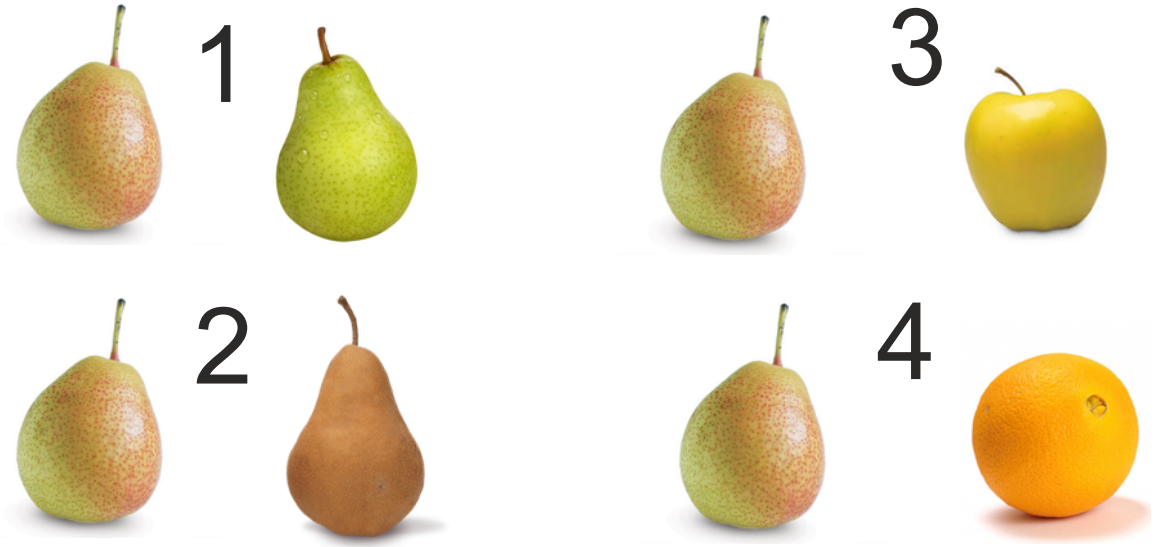
Intuitively we know that the starred pictures are all pears….but how would we approach this if we knew nothing about fruit?

We catalog variable objects by iterative pairwise comparison

mixed sample

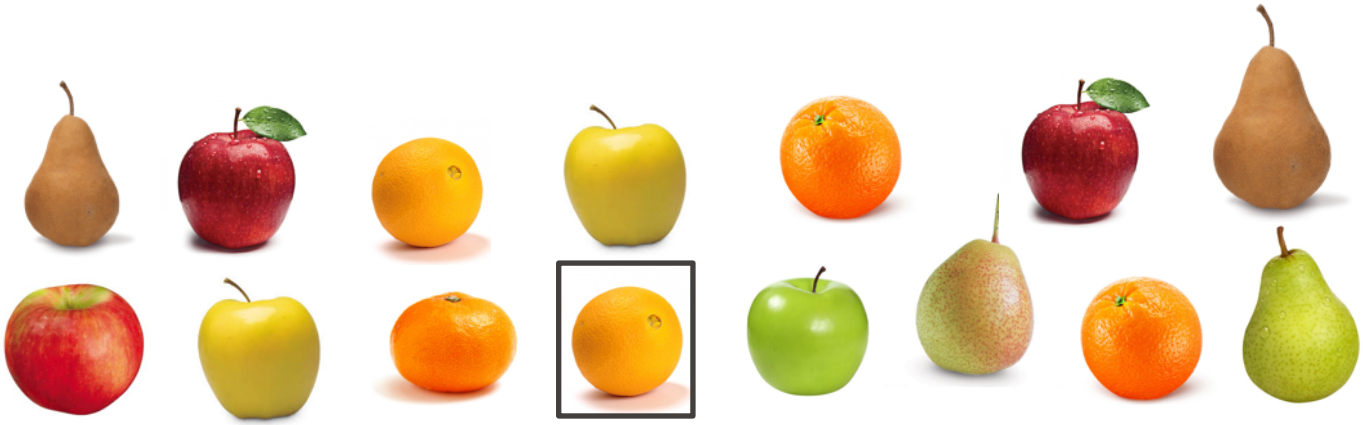


ranked pairwise comparisons

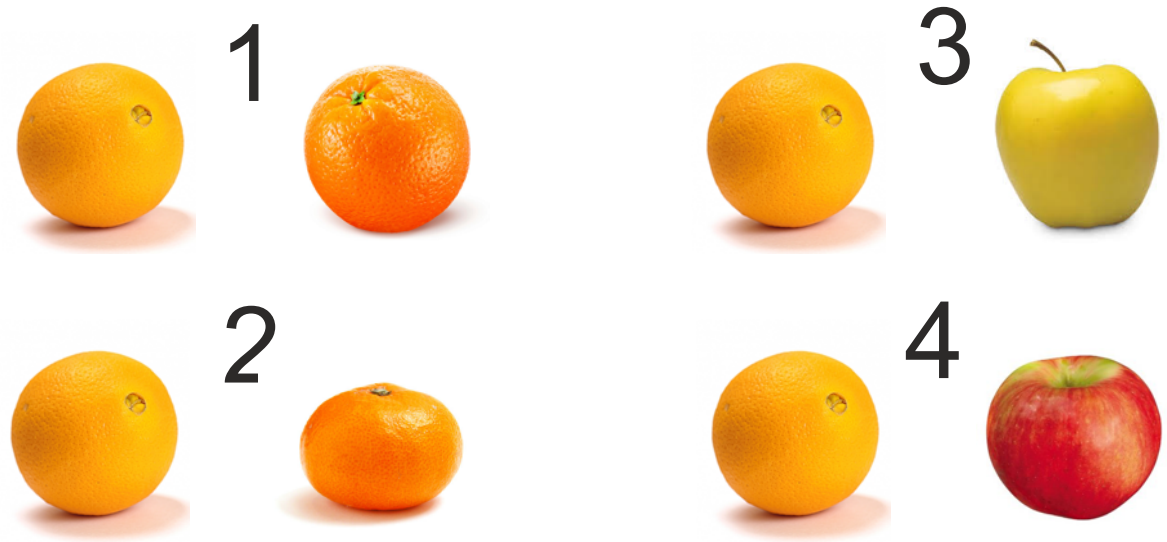


We catalog variable objects by iterative pairwise comparison

mixed sample



ranked pairwise comparisons



Clusters arise of similar items

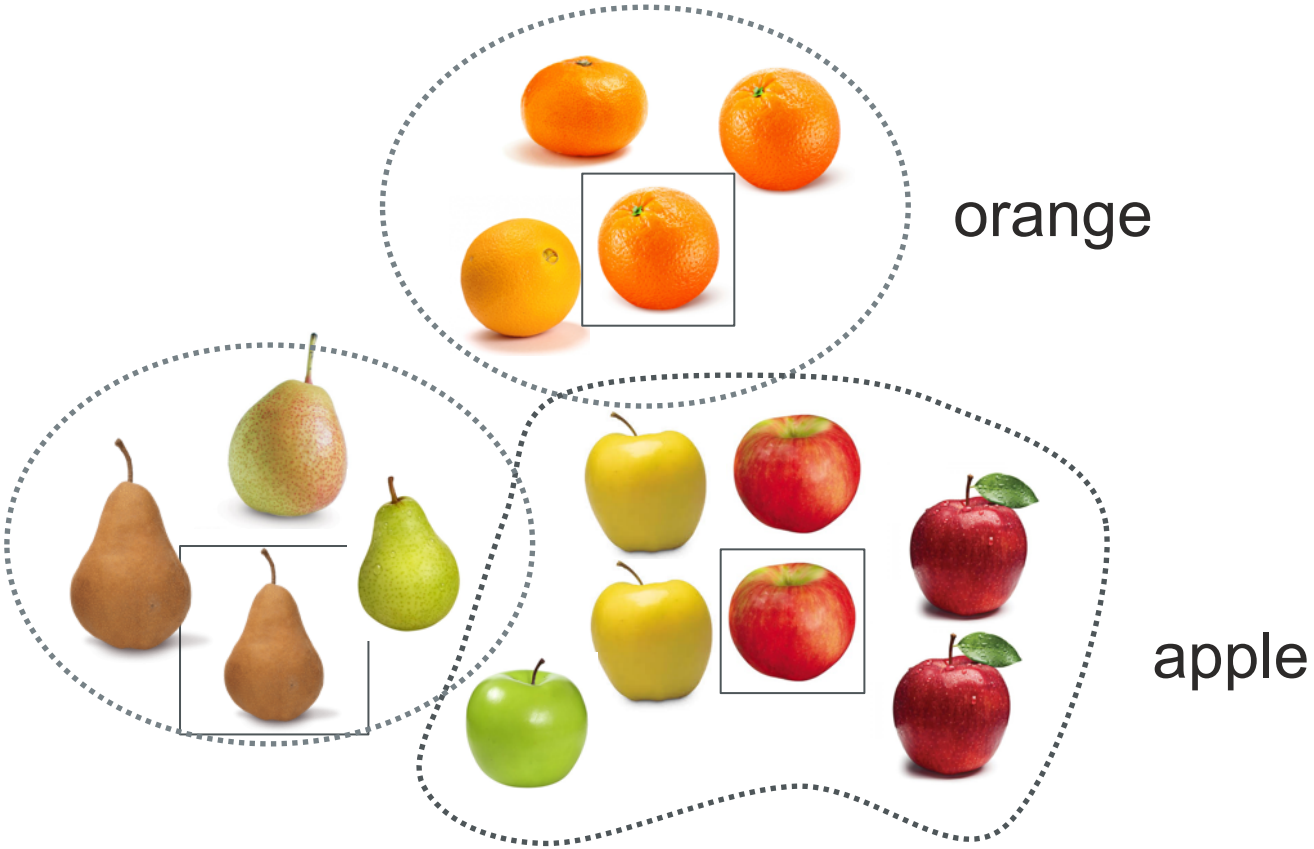
mixed sample

ranked pairwise comparisons

Similar fruits cluster together

mixed sample

pairwise comparison distances



Where do we want to go next?

Swab site composition

?

Extraction

Sequencing

Assembled and pooled 16S reads across samples

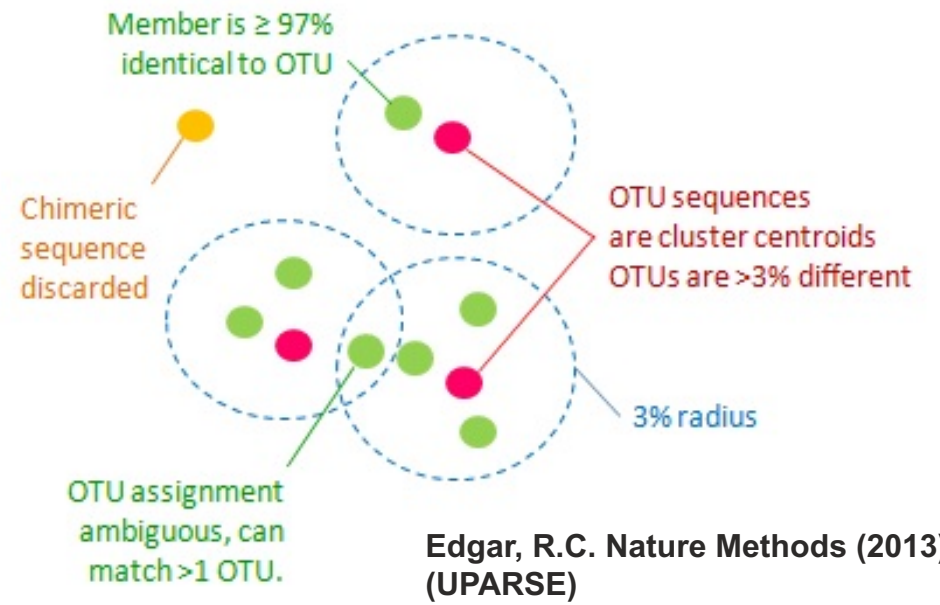
1. Identify unique sequences
2. Use pairwise comparison to cluster into operational taxonomic units (OTUs)
3. Count how many sequences match each OTU

Sequences to OTUs to OTU abundance

OTU sequences are representative sequences chosen for each OUT and are <97% similar compared to any other OTU sequence

~3% similarity = species

To generate a relative abundance table, count the number of 16S sequences matching each OUT sequence



Edgar, R.C. Nature Methods (2013) (UPARSE)

Samples

#OTU ID	F3D0	F3D141	F3D142
OTU_6	749	535	313
OTU_25	29	57	14
OTU_1	613	497	312
OTU_8	426	378	255
OTU_31	149	38	10
OTU_2	366	392	327

OTUs

Counts

How can we visualize this data?

3) What kind of biological insights can we derive?

Visualizing the OTU table: relative abundance plot

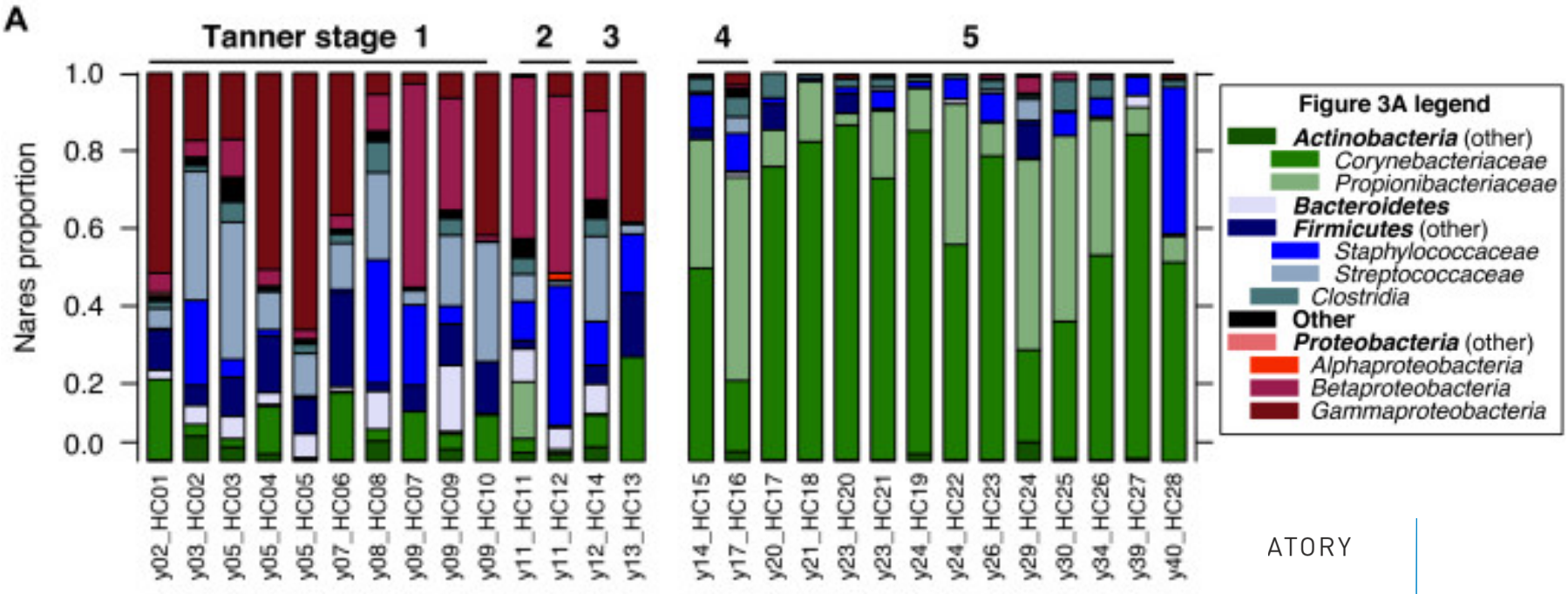
Samples

OTUs

#OTU ID	F3D0	F3D141	F3D142
OTU_6	749	535	313
OTU_25	29	57	14
OTU_1	613	497	312
OTU_8	426	378	255
OTU_31	149	38	10
OTU_2	366	392	327

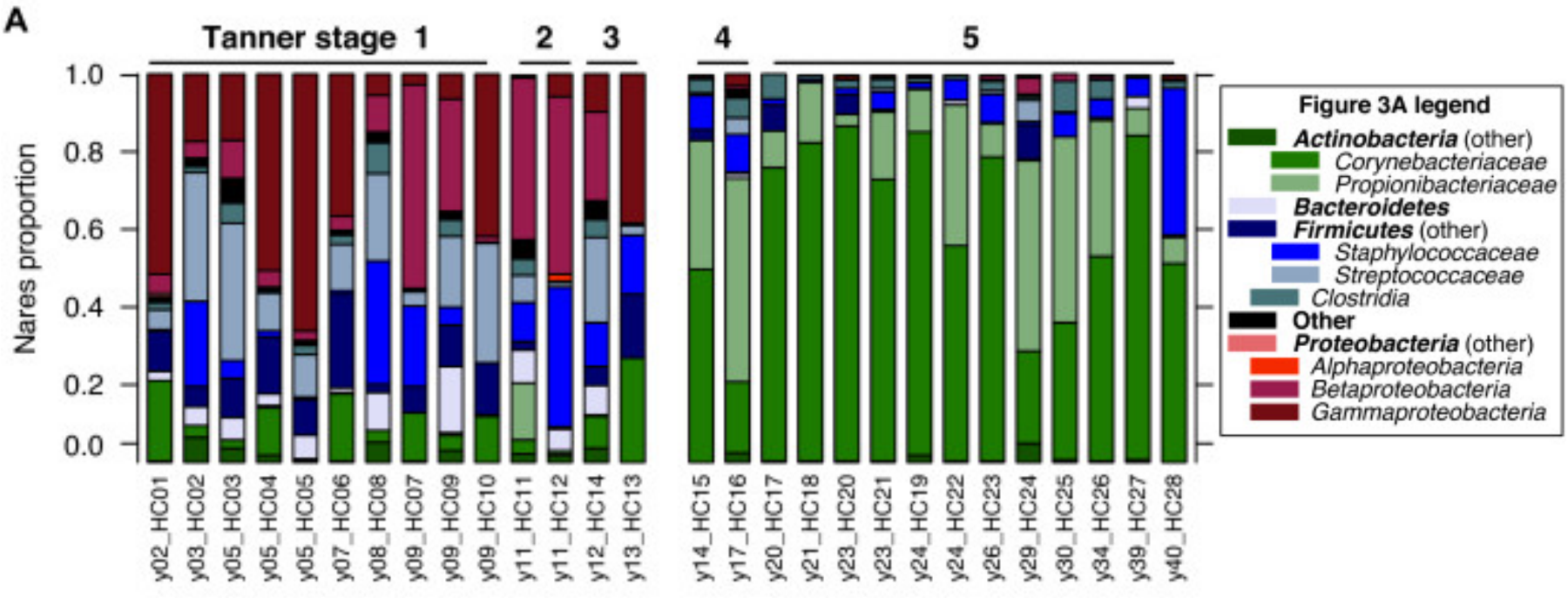
+

The Ribosomal Database Project (RDP) Classifier tool identifies OTU taxonomy from a reference database



ATORY

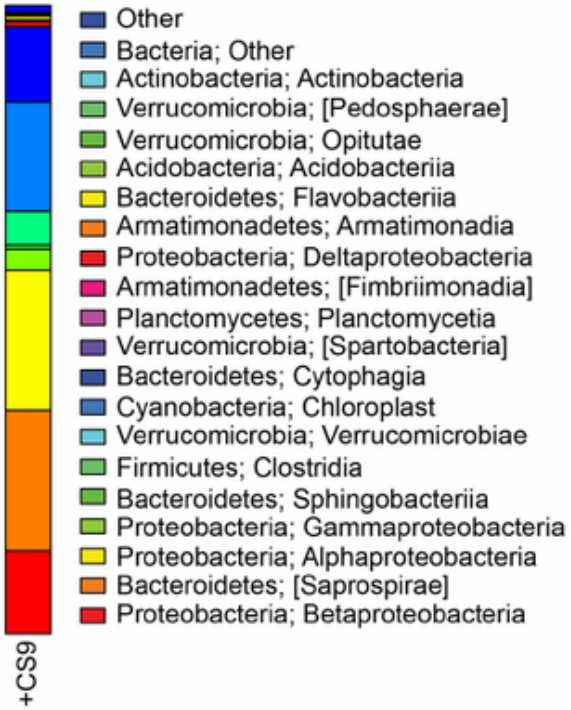
Relative abundance plot captures variation in abundance across samples



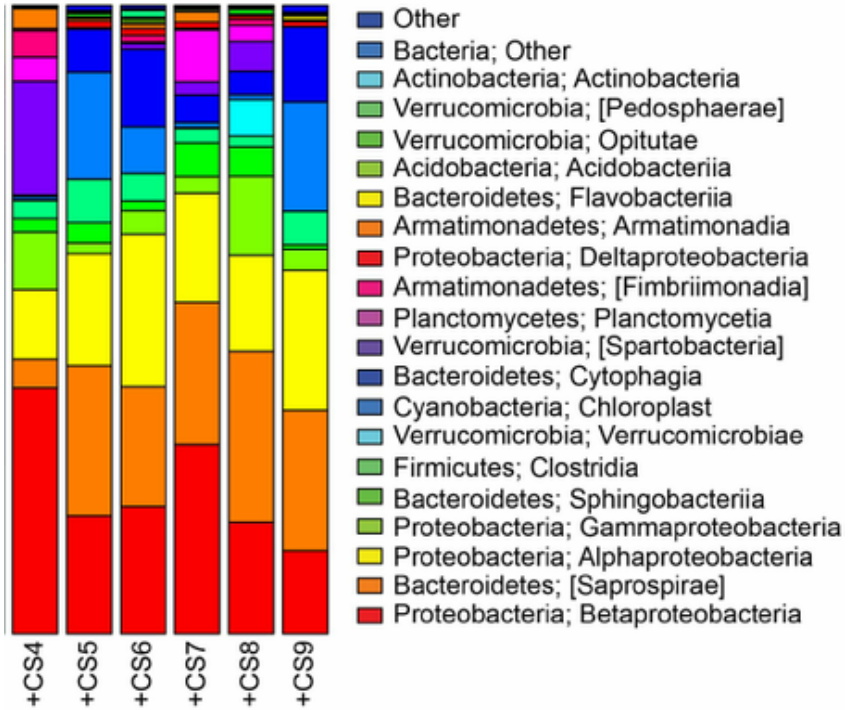
"we found that the microbial communities clustered into two distinct groups in which those of Tanner stages 1, 2, and 3 differed significantly from those of Tanner stages 4 and 5"

Visualizing diversity: alpha v beta

Alpha diversity: Complexity of community within a given sample



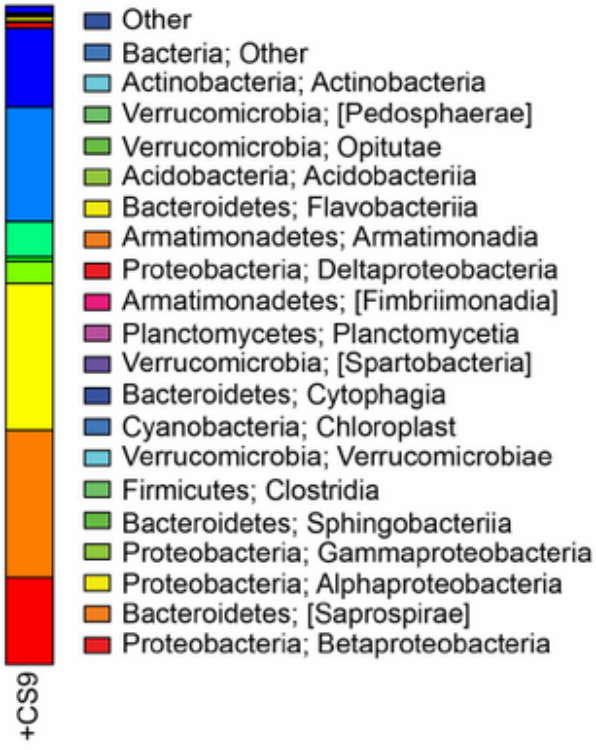
Beta diversity: Variation in microbial communities across samples



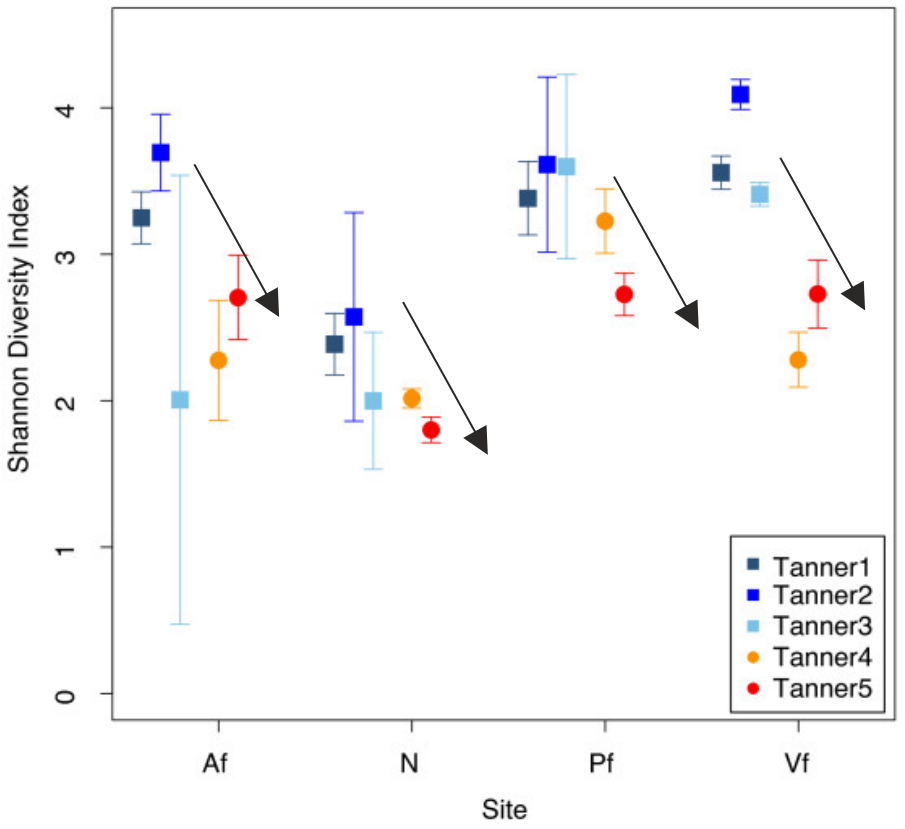
"To assess whether there is a shift in the microbial composition of children (Tanner stage 1) versus adults (Tanner stage 5), we calculated the variation between individuals within and between the two groups."

Visualizing diversity: alpha v beta

Alpha diversity: Complexity of community within a given sample



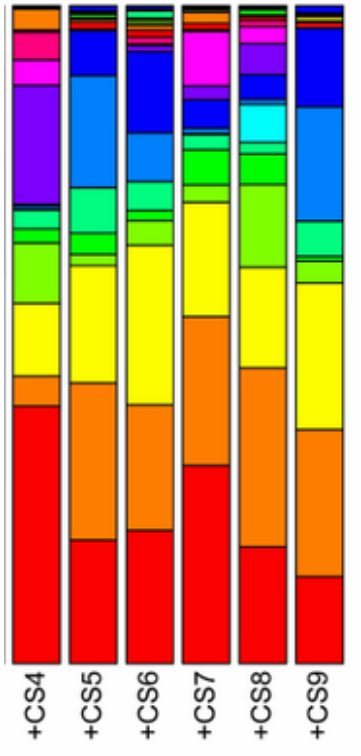
Many methods: Shannon, Simpson, Fisher, etc



"decreas[ing alpha diversity] with increasing sexual maturity suggests a stabilization and convergence of the nares microbiome in more mature individuals"

Visualizing diversity: alpha v beta

Beta diversity: Variation in microbial communities across samples



Bray-Curtis dissimilarity: quantifies how much compositional variation exists between pair of samples

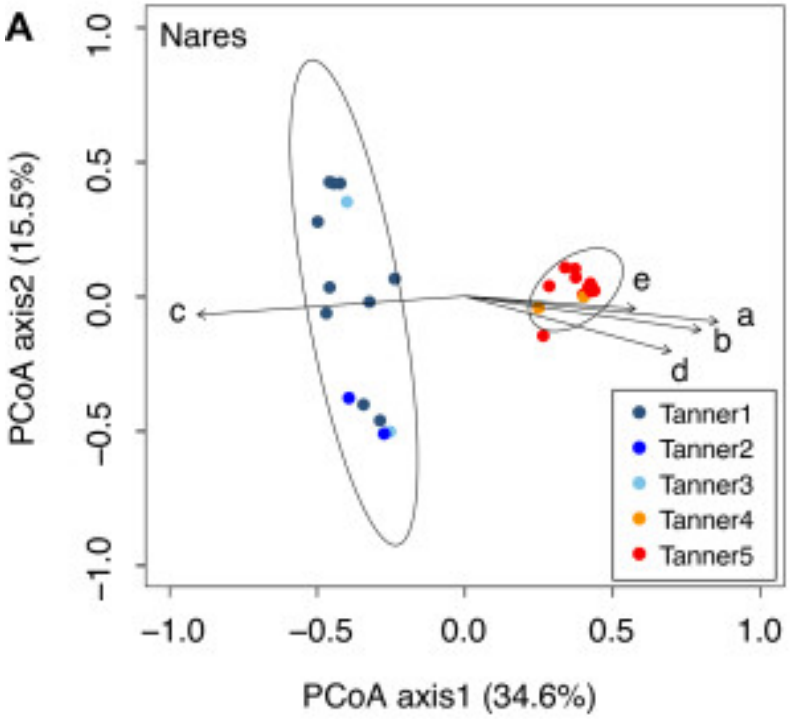
A Bray-Curtis = 0 → samples are identical.

A Bray-Curtis = 1 → no OTUs found in common.

→ Visualize by Principle Coordinates Analysis ordination plot

Visualizing diversity: alpha v beta

Beta diversity



"we observed that the microbial community memberships and structures of individuals of Tanner stages 2 and 3 trended towards more significantly resembling those of Tanner stage 1 than later stages…Physiologically, Tanner 4 individuals more closely resemble Tanner 5 individuals in terms of sexual maturity"

16S is powerful, but has limitations

- Phylogenetic classification is limited → Many OTUs are unclassified

	Phylum	Class	Order	Family	Genus
OTU_1	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	unclassified_Porphyromonadaceae
OTU_2	Firmicutes	Clostridia	Clostridiales	unclassified_Clostridiales	unclassified_Clostridiales
OTU_3	Firmicutes	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae	Clostridium_XVIII
OTU_4	Firmicutes	Clostridia	Clostridiales	unclassified_Clostridiales	unclassified_Clostridiales
OTU_5	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	unclassified_Porphyromonadaceae
OTU_6	Firmicutes	Clostridia	Clostridiales	Clostridiales_Incertae Sedis XIII	unclassified_Clostridiales_Incertae_Sedis_XIII

- 16S gene is only a single gene → biological insight is limited

Metagenomic shotgun sequencing looks at ALL the DNA, not just 16S

- Better phylogenetic classification → strain level identification (Marker gene analysis)

- All the gene content → functional insight (biosynthetic genes, resistance genes, etc)

Thank you

Julia Oh

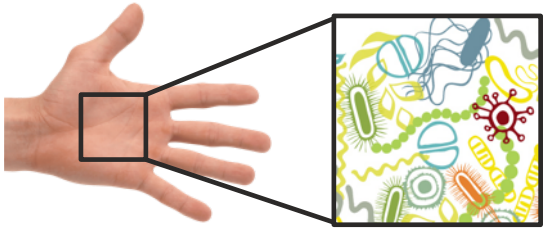
Jethro Johnson

Mark Adams

Synthesis: hands-on exercise

A workflow for 16S analysis

1 Collect sample (e.g. skin swab)



2 Total DNA extraction

3 Amplify variable region of 16S rRNA gene

4 Sequence 16S rRNA genes

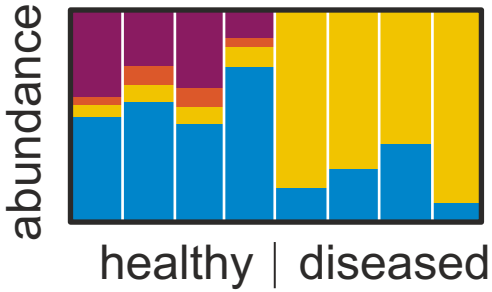
5 Processing, QC, and assembly of reads

6 Generate a representative sequence for each species and assign taxonomy

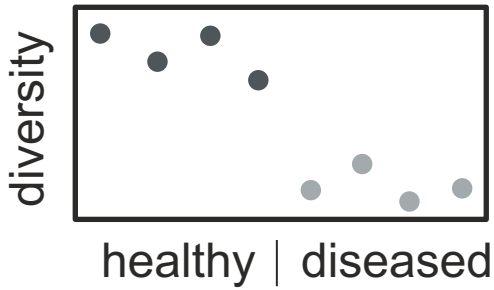
7

Who? How much?

relative abundance plot



diversity plot



A microbiome module by design

Conceptual goals for students

- why do we sequence the 16S gene?
- how do we decide which 16S sequences belong to which organism?
- what can relative abundance and diversity plots tell us?
- develop hypothesis and test it

Data science goals for students

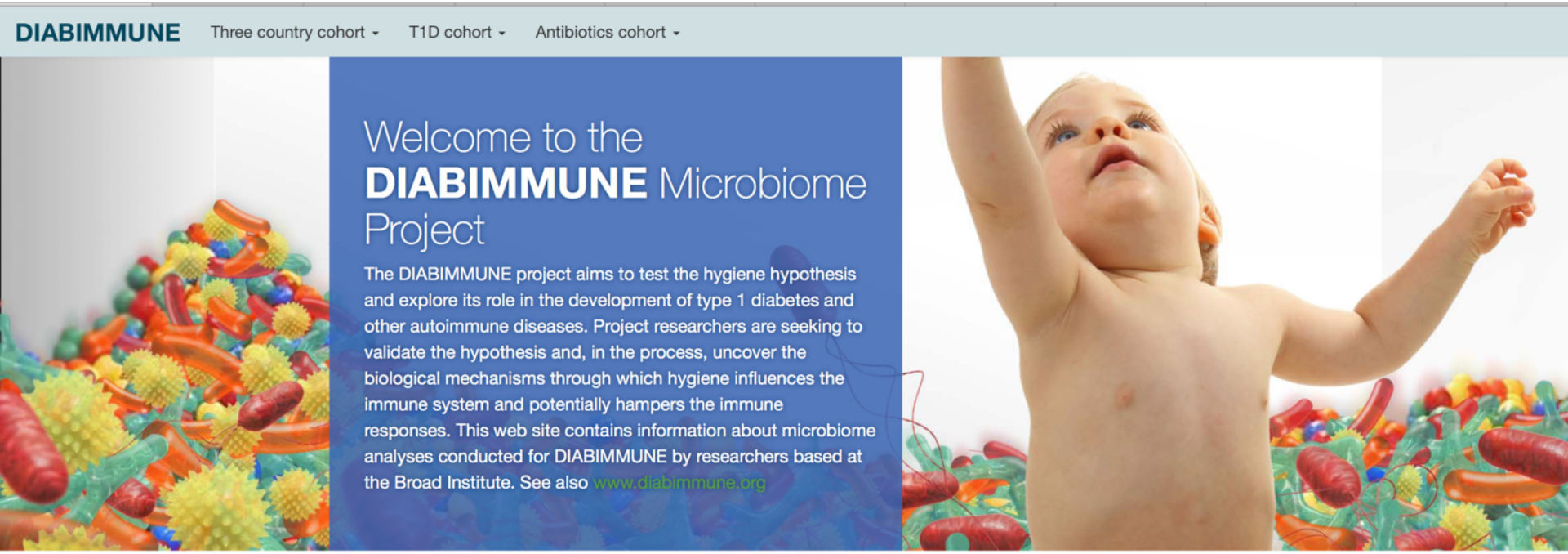
- file manipulation from the command line
- writing and executing ”for loops”
- running commands in terminal
- data visualization in R
- writing R notebooks

A microbiome module by design

Teaching implementation goals

- allows for individual student curiosity & open-endedness
- real-world medical relevance
- easily sourced raw data
- do not have to wait for code to run
- avoid painful coding “typos”
- avoid version/compatibility issues

The exercise: an overview



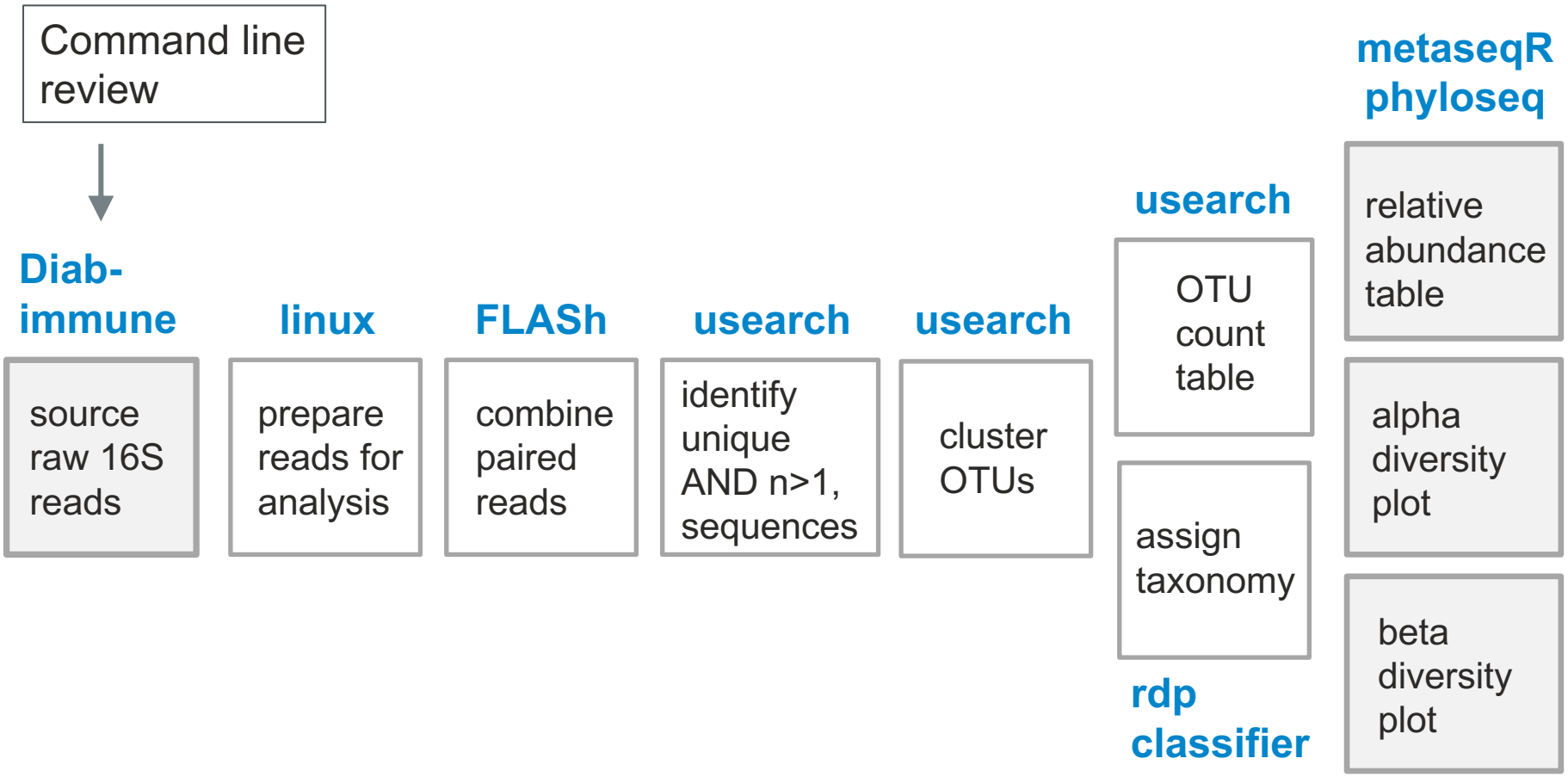
"What is the role of the microbiome and the "hygiene hypothesis" in the development of type 1 diabetes and other auto-immune diseases?"

https://pubs.broadinstitute.org/diabimmune

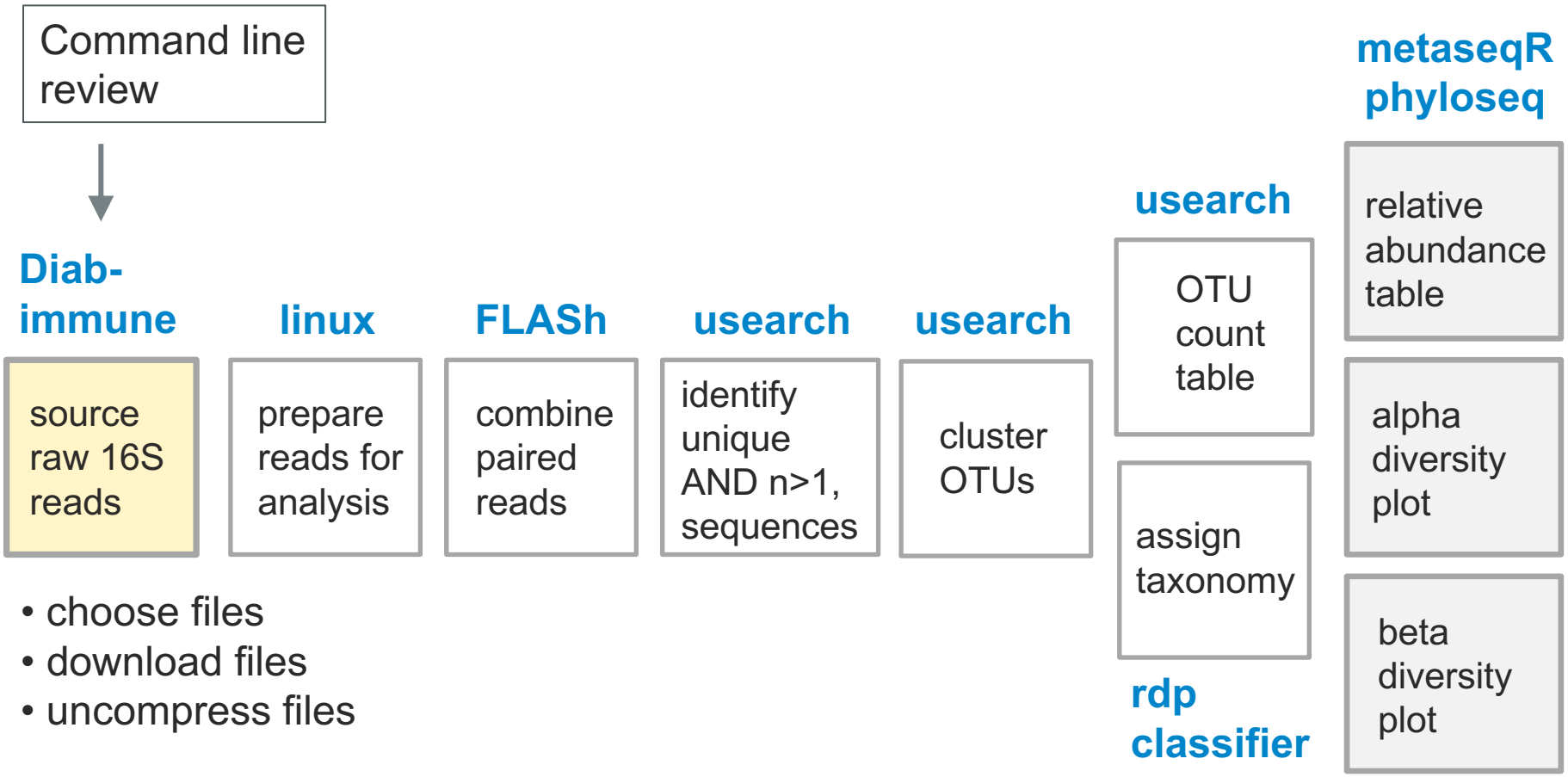
The exercise: an overview

“The goal of this cohort is to compare microbiome in infants who have developed type 1 diabetes (T1D) or serum autoantibodies (markers predicting the onset of T1D) with healthy controls in the same area.”

The exercise: an overview

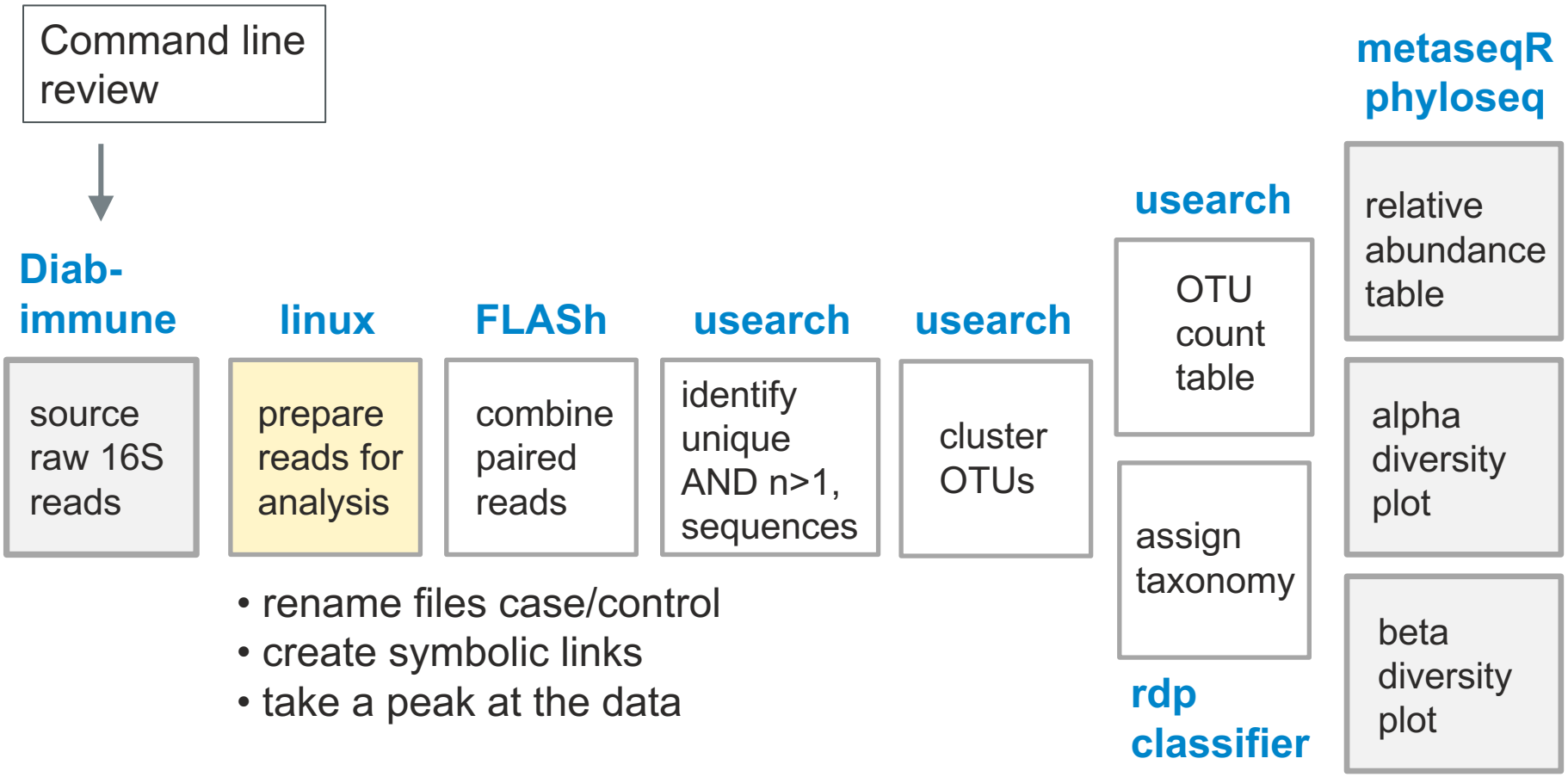


The exercise: an overview

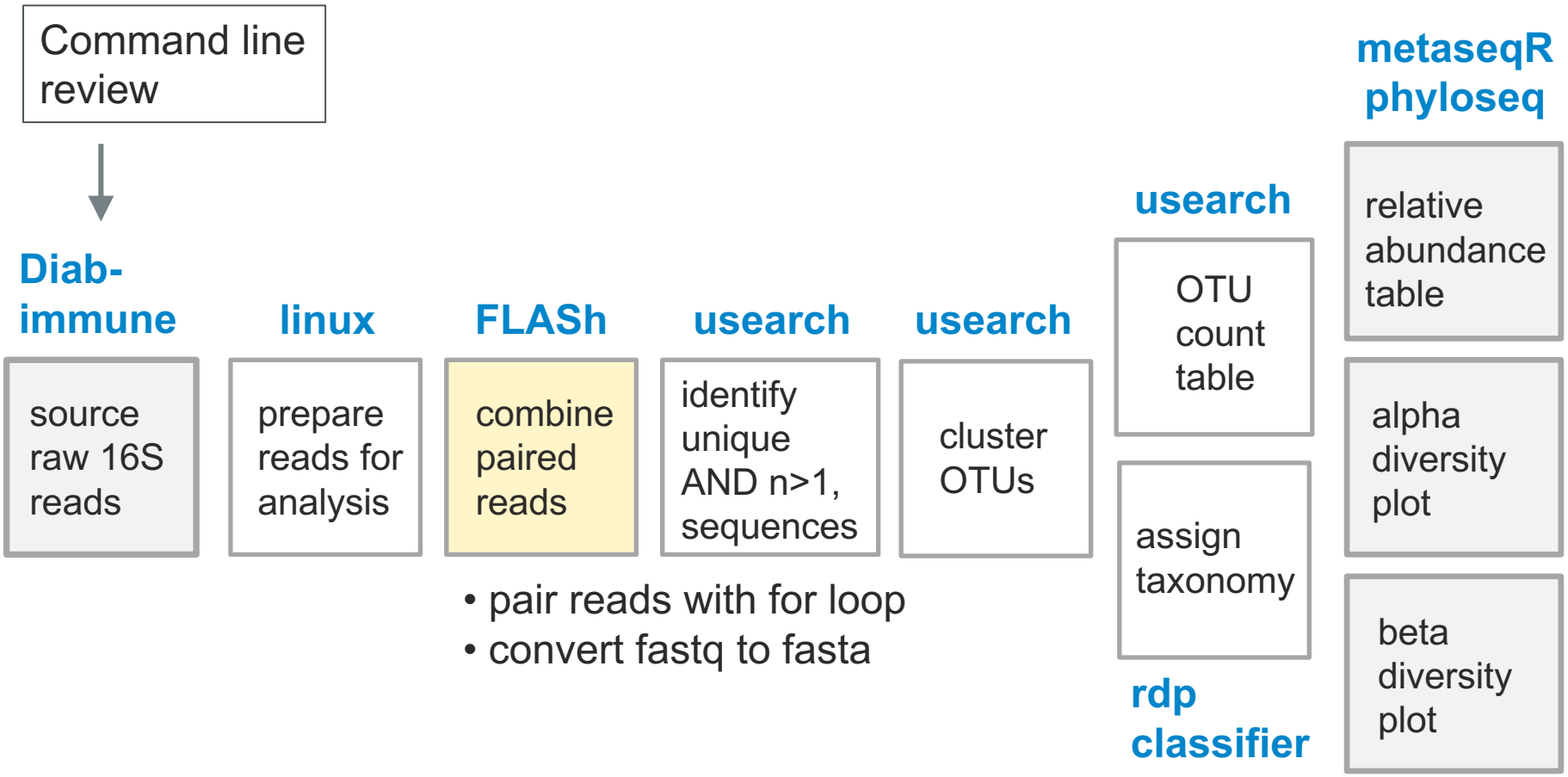


- choose files
- download files
- uncompress files

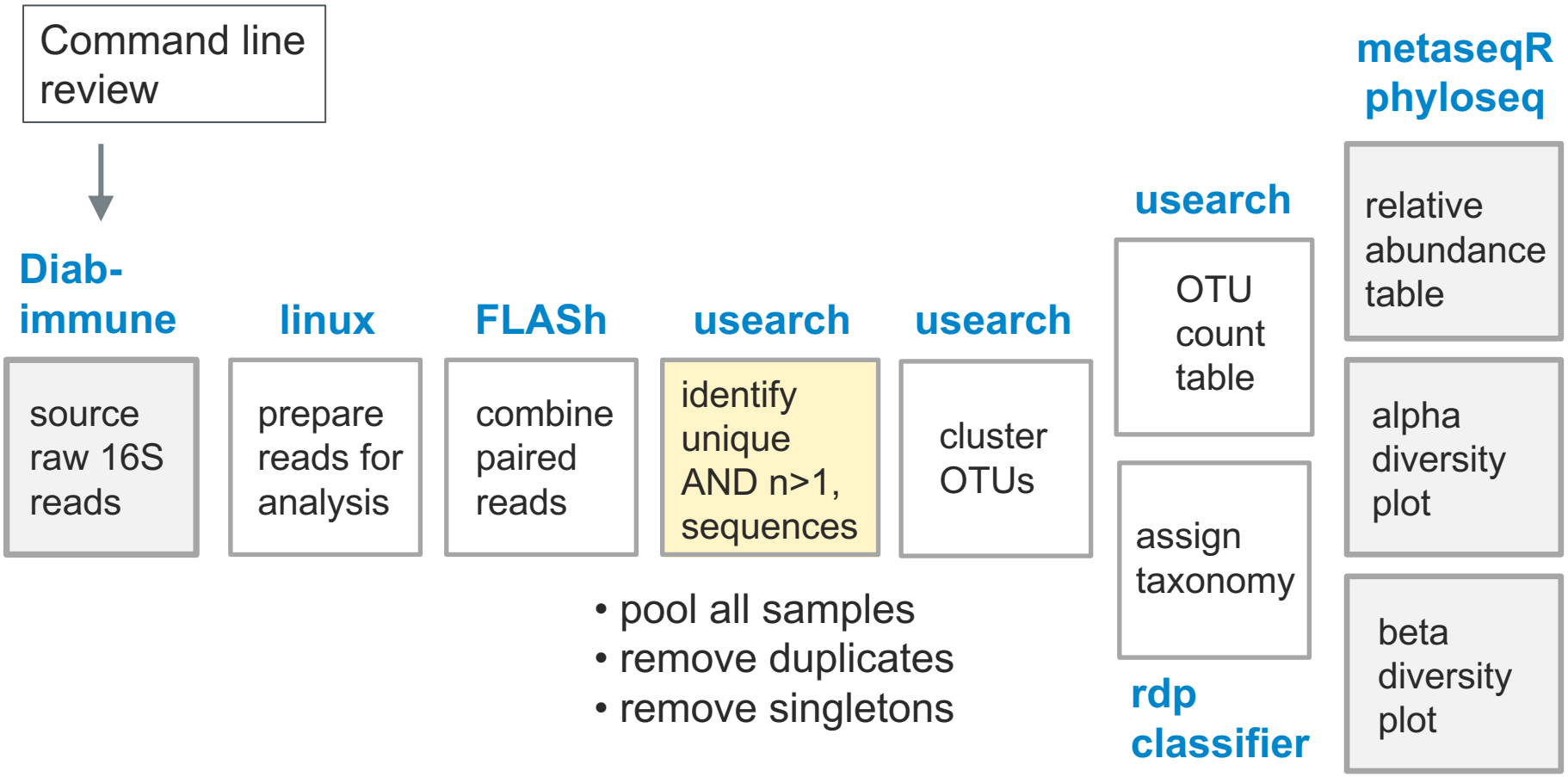
The exercise: an overview



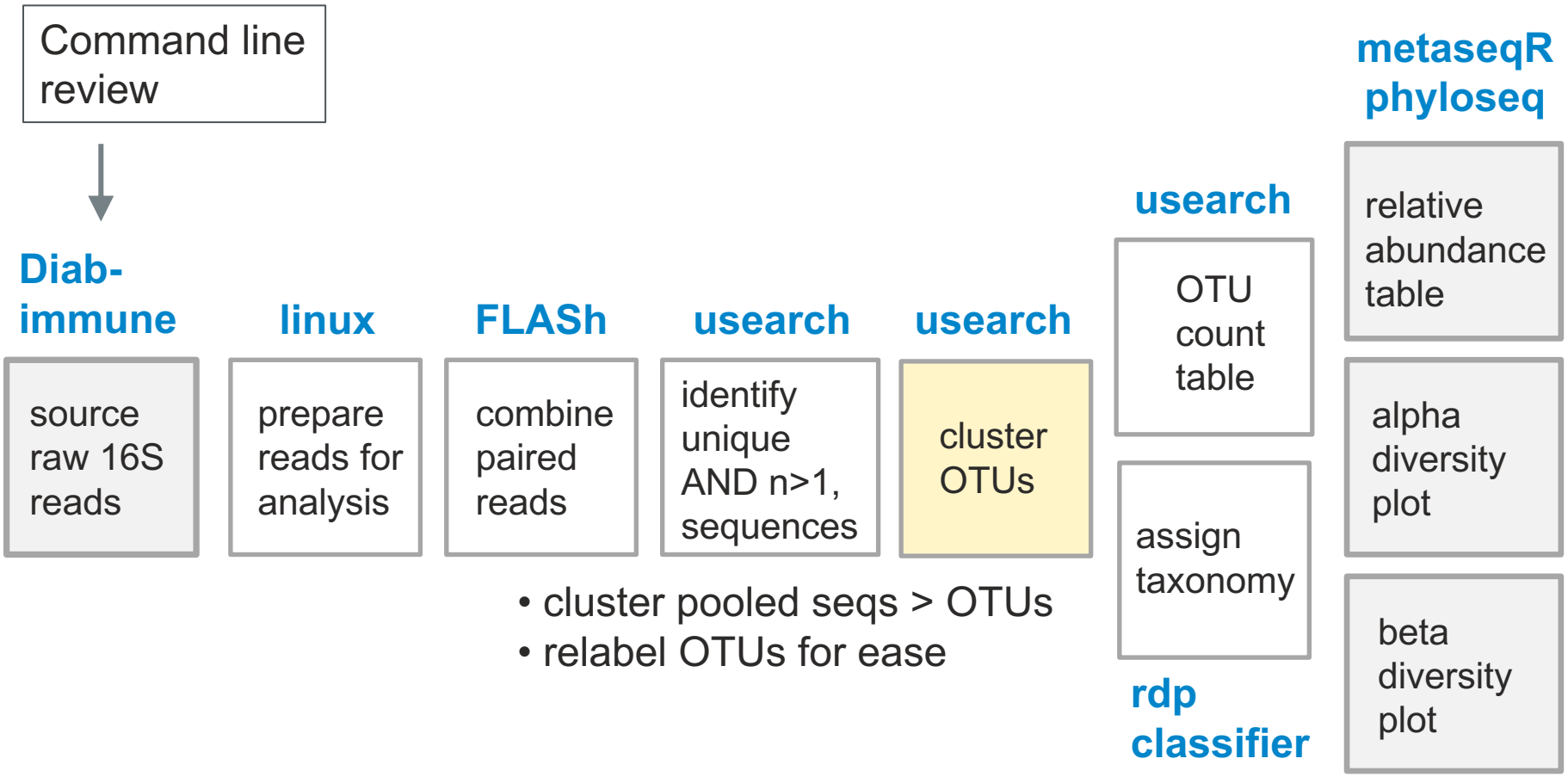
The exercise: an overview



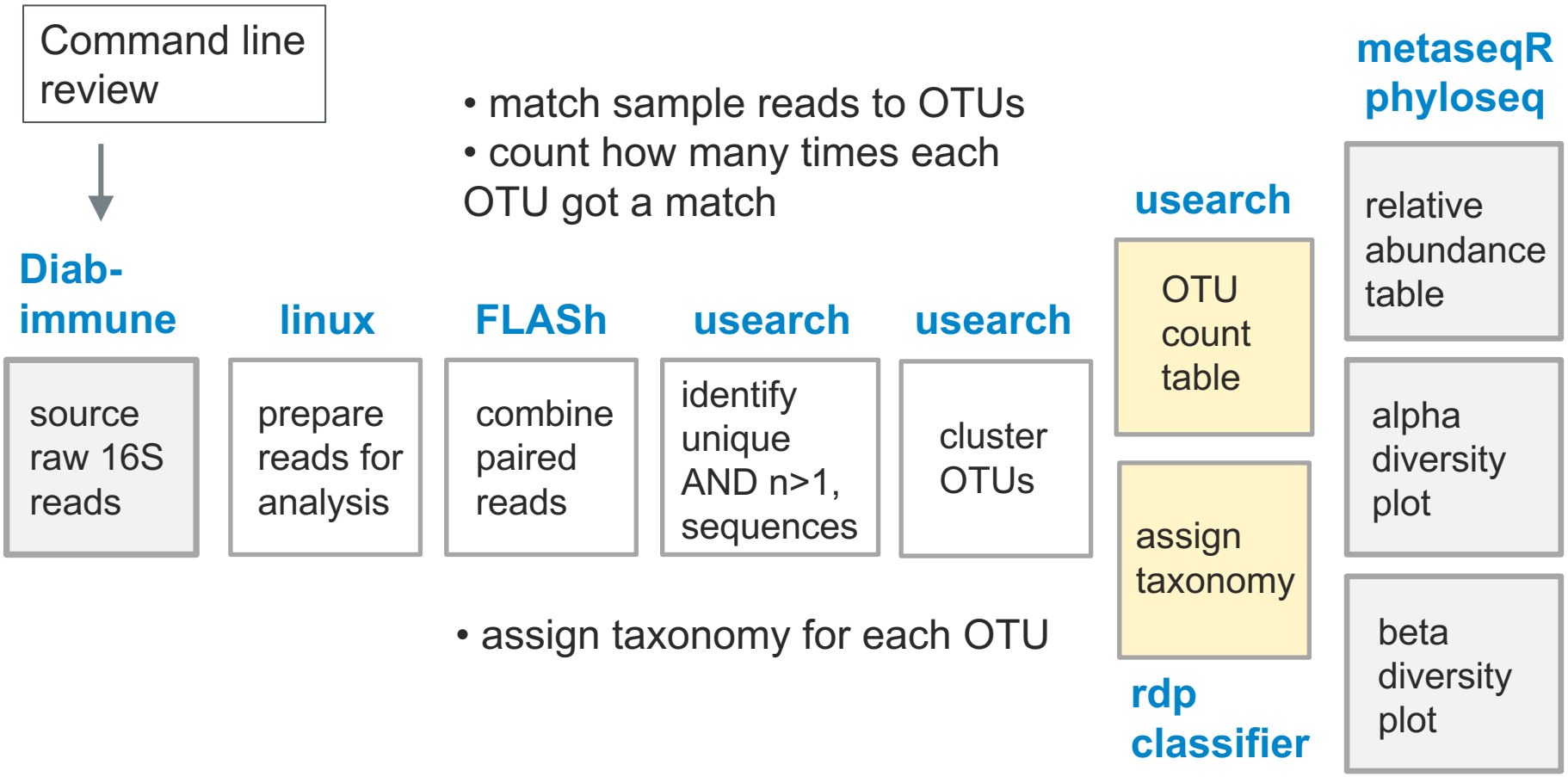
The exercise: an overview



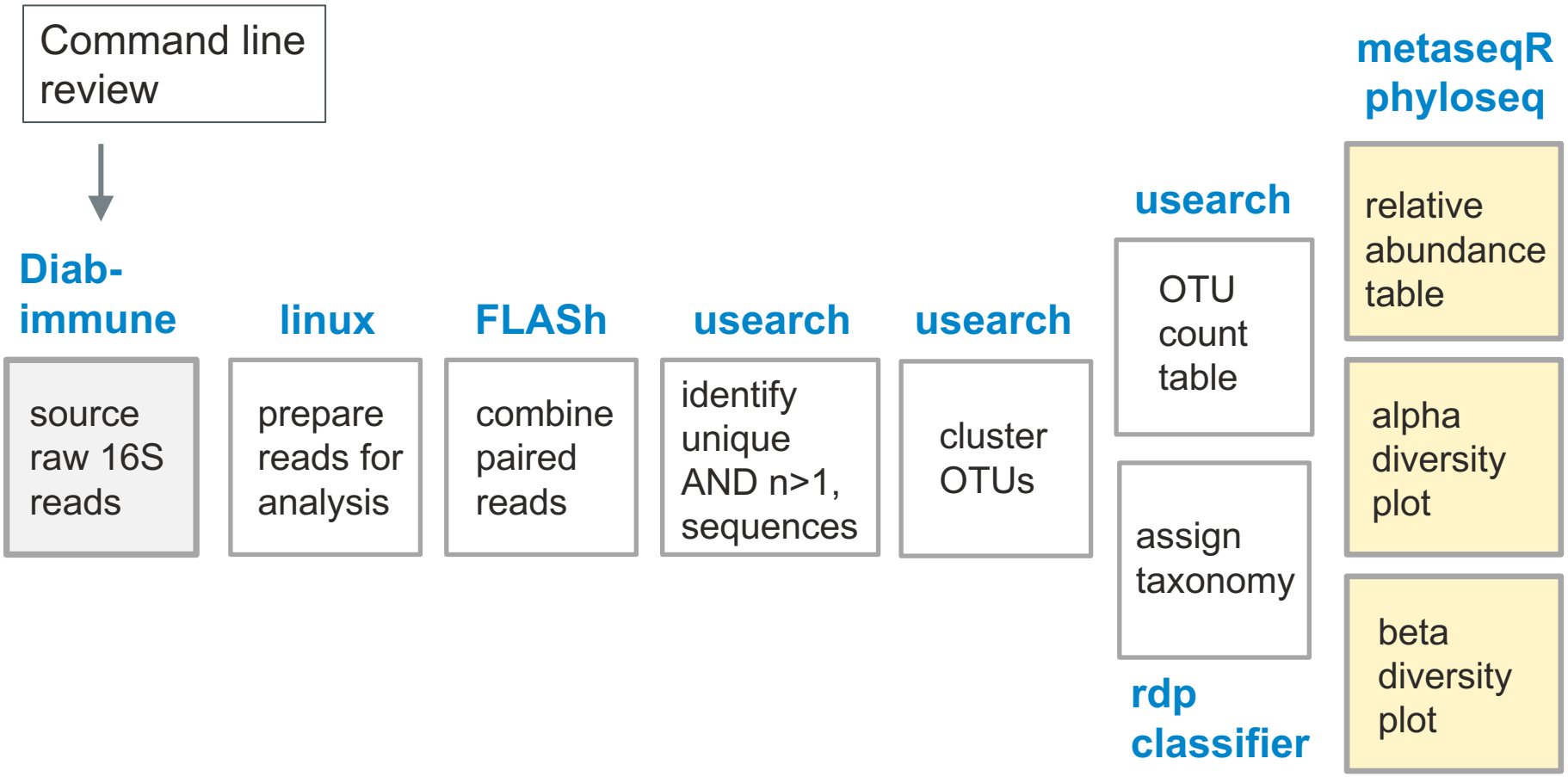
The exercise: an overview



The exercise: an overview



The exercise: an overview



Revisiting our goals

Conceptual goals for students

- why do we sequence the 16S gene?
- how do we decide which 16S sequences belong to which organism?
- what can relative abundance and diversity plots tell us?
- develop hypothesis and test it

Data science goals for students

- file manipulation from the command line
- writing and executing ”for loops”
- running commands in terminal
- data visualization in R
- writing R notebooks

Revisiting our goals

Teaching implementation goals

- allows for individual student curiosity & open-endedness
- real-world medical relevance
- easily sourced raw data
- do not have to wait for code to run
- avoid painful coding “typos”
- avoid version/compatibility issues