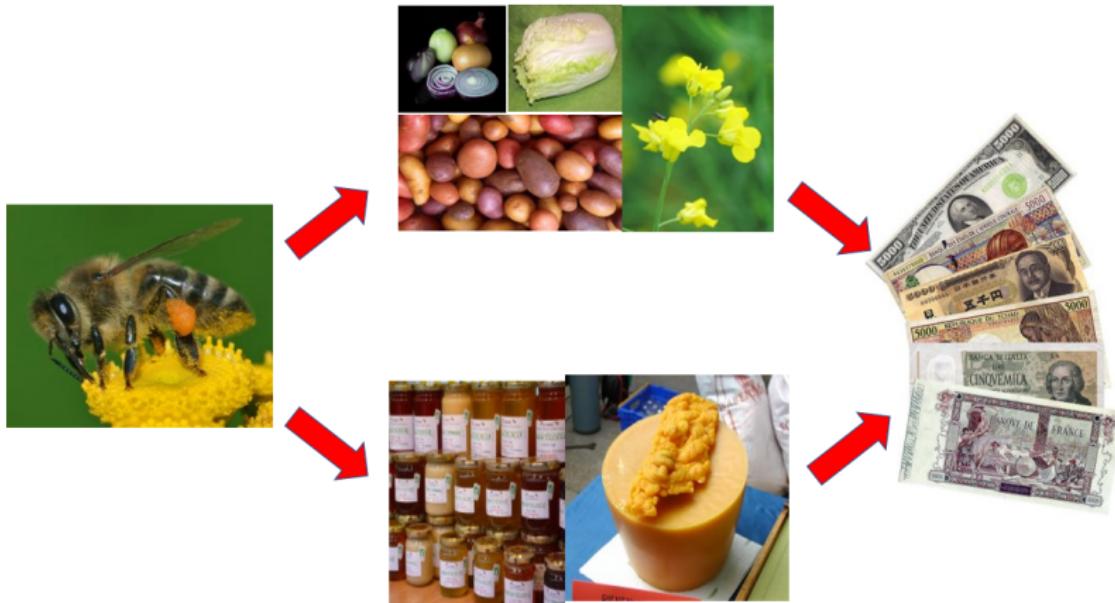


Optimize sequencing depth for shotgun metagenomics of pollination system by rarefaction, using a modular profiling pipeline

Cong Liu

2021.9

# Bee pollination provides economic benefits



# Factors challenging bee health



Invasive species



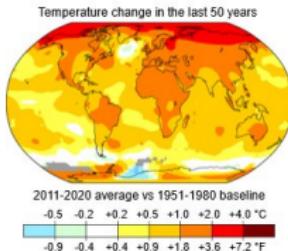
Parasites



Pesticides



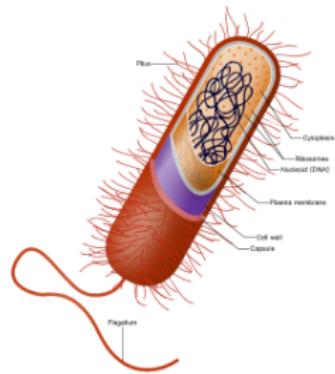
Habitat loss



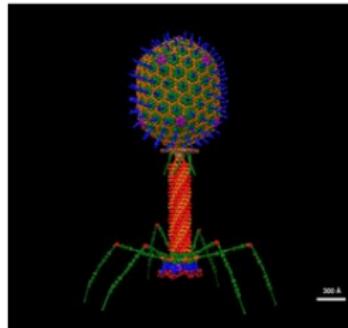
Climate change

# Diverse gut microbiome impacts bee health

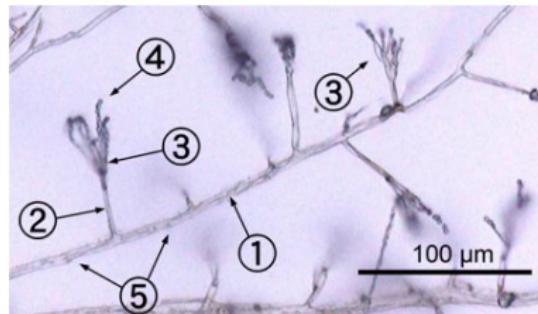
## Taxonomic diversity



Bacteria



Viruses



Fungi



Insects, plants, etc

# Diverse gut microbiome impacts bee health

Functional diversity



Bee microbiome



Food digestion



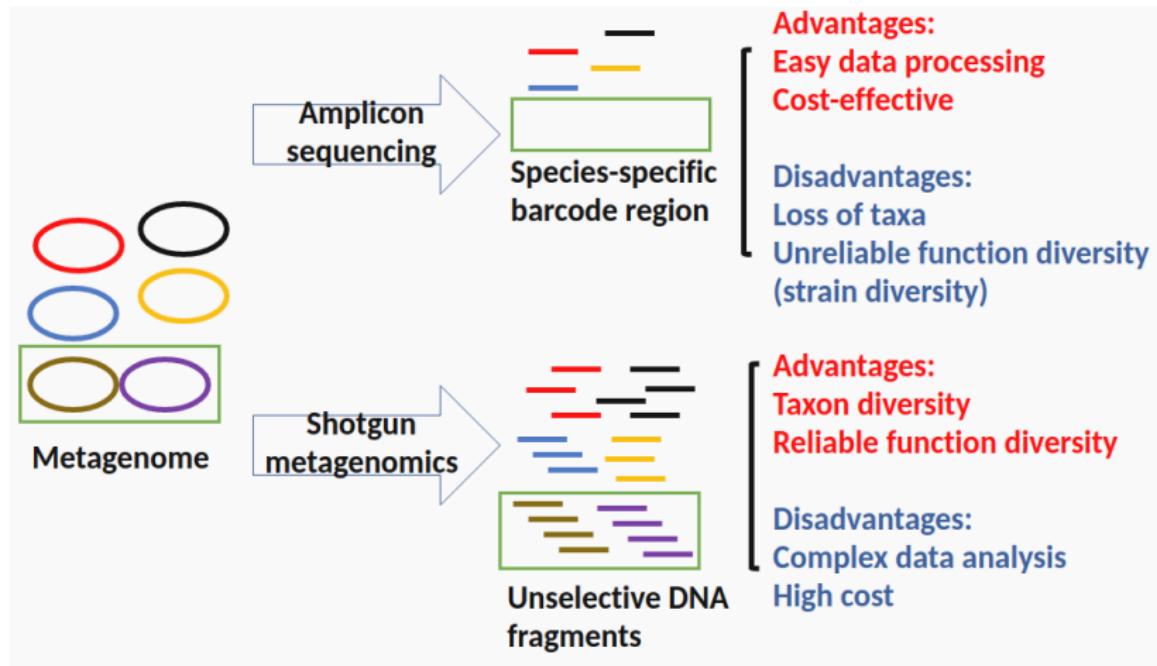
Parasite defence



Chemical detoxification

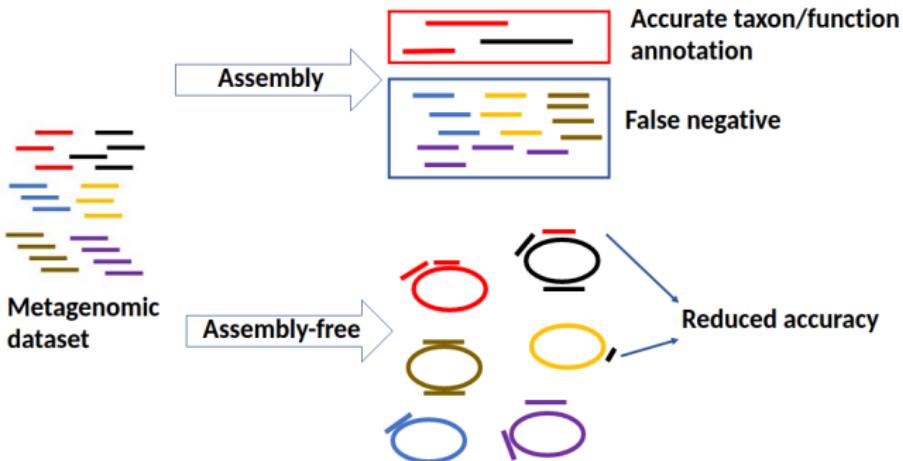
# High-throughput methods exploring taxon/function diversity of microbiome

## Amplicon sequencing and shotgun metagenomics



# Challenges of shotgun metagenomics

Complex data analysis and high cost



- ▶ Shotgun metagenomics:  $\sim 120\text{£}/(\text{sample} \times 6 \text{ Gbp})$
- ▶ Amplicon sequencing:  $\sim 20\text{£}/(\text{sample} \times 15 \text{ Mbp})$



# Aims

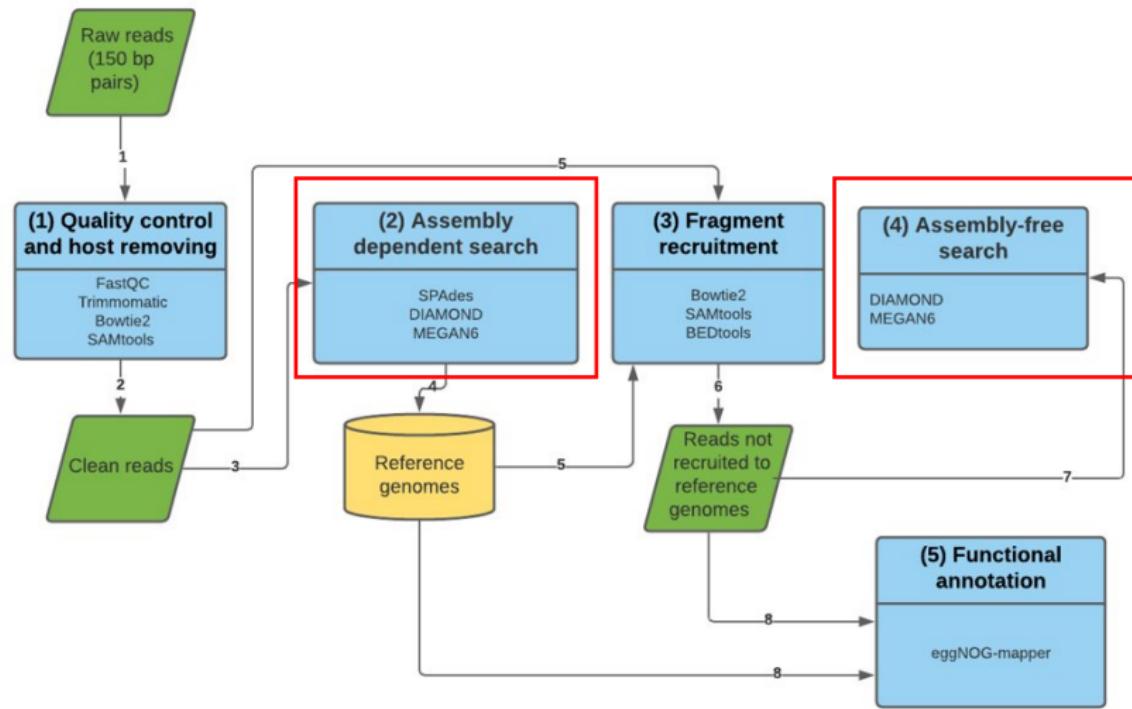
- ▶ Pipeline for metagenomic data analysis



- ▶ Optimizing sequencing depth



# Integrated pipeline for taxonomic/functional profiling of shotgun metagenomic data



# Samples and sequence data



*Apis mellifera*



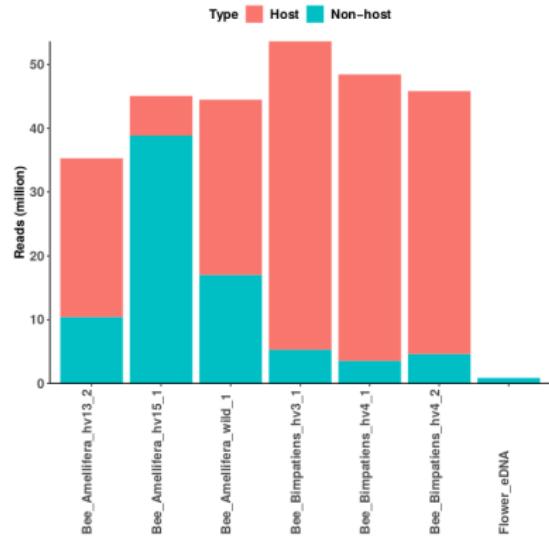
*Bombus impatiens*



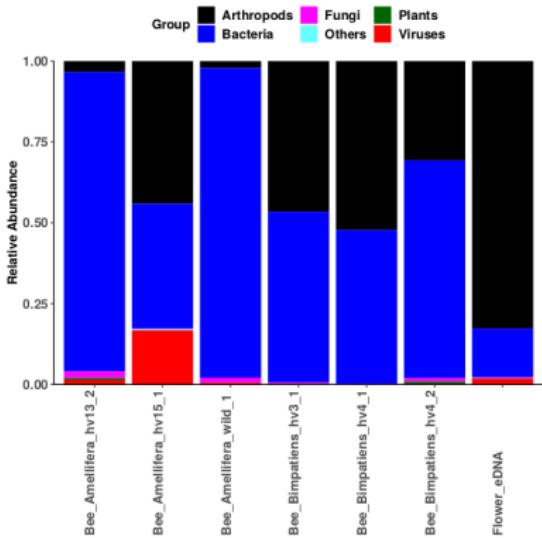
*Erigeron annuus*

# Analyze metagenomic data by integrated pipeline

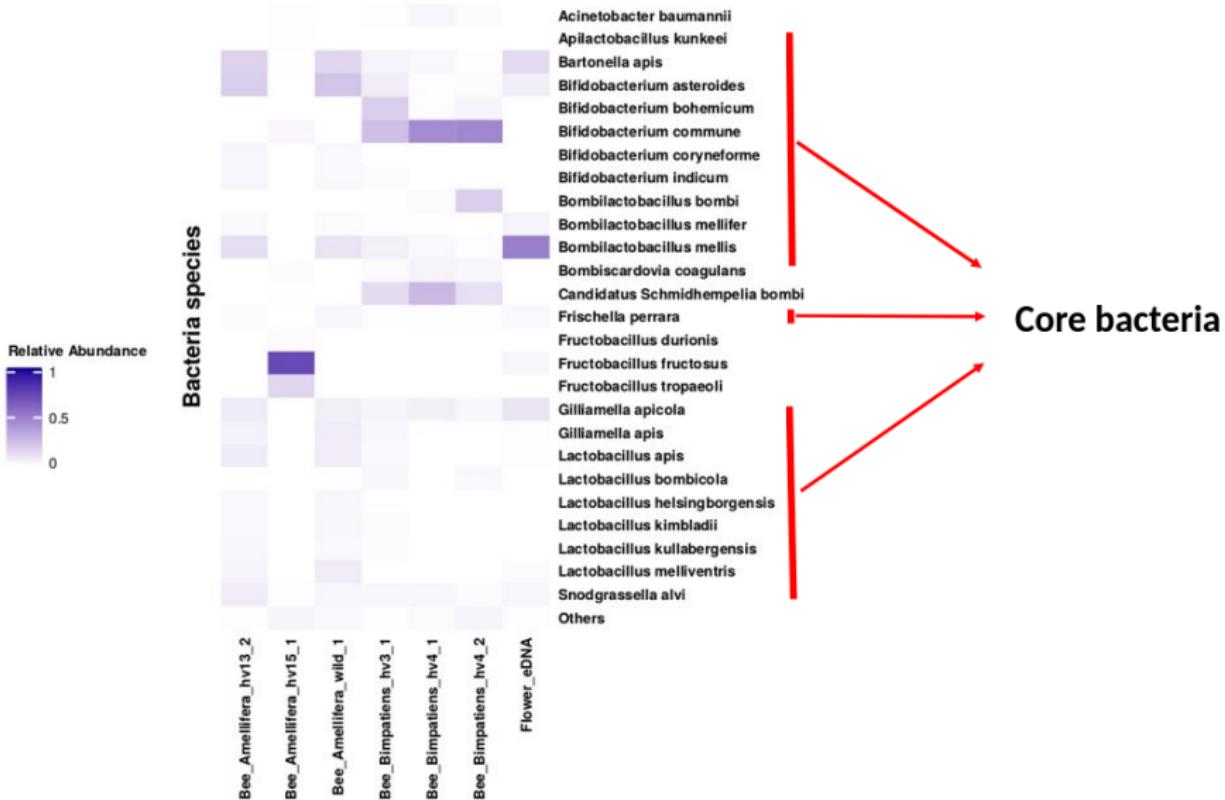
a



b



# Core bacteria indicate sample quality and representativity



# Multi-clade community revealed by integrated pipeline

## Arthropods



*Apis*

*Bombus*

## Plants



Rape

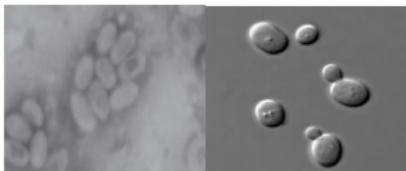
Chickpea



Radish

Sunflower

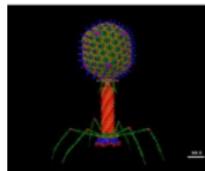
## Fungi



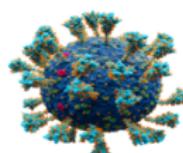
*Nosema*

Yeast

## Viruses

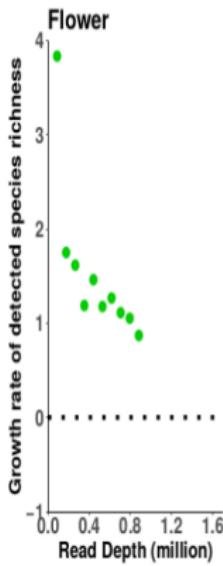
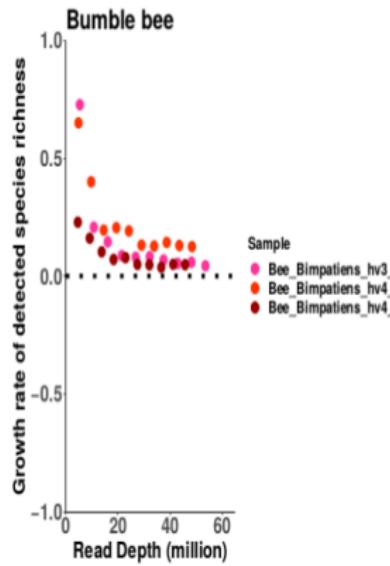
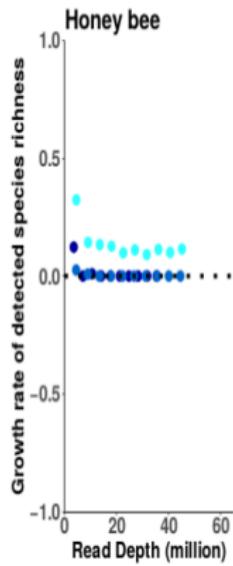


Phages



Arthropod viruses

# Integrated pipeline found species not represented by assembly



# Advantages of integrated pipeline

- ▶ **Identify more species**

Combination of assembly-free method helps solve species not represented by assembly.

- ▶ **Flexibility**

Modularity provides capability for incorporating of alternative tools.

- ▶ **Easy troubleshooting**

Output files generated by each module are recorded and can be inspected easily for troubleshooting.

# Aims

- ▶ Pipeline for metagenomic data analysis  
Integrated pipeline



- ▶ Optimizing sequencing depth



# Measure species/function diversity by Hill numbers

Indexes measuring species/function diversity:

Richness, Shannon index, Simpson index, **Hill numbers**, etc

Advantages of Hill numbers:

- ▶ Replication principle
- ▶ Modulate sensitivity to relative abundances via **order  $q$**
- ▶ Algebraic transformation to other diversity indexes

**Assemblage I**

$$p_1, p_2, p_3$$

**Assemblage II**

$$p_4, p_5, p_6, p_7$$

+



**Pooled assemblage**

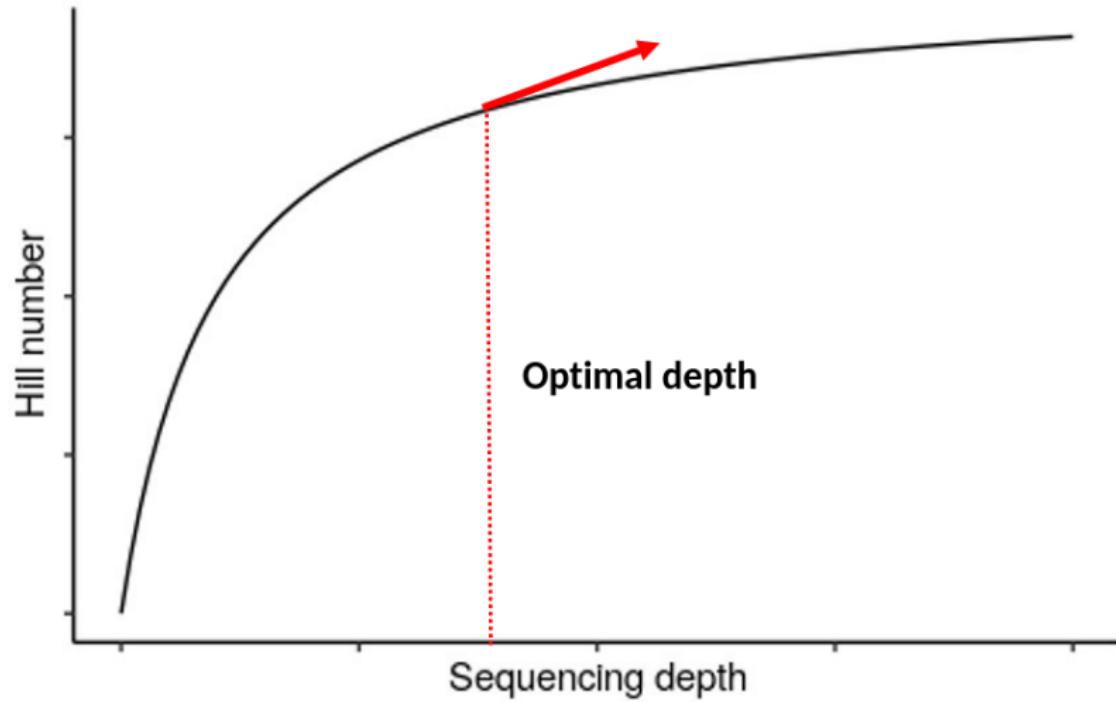
$$\frac{p_1}{2}, \frac{p_2}{2}, \frac{p_3}{2}$$
$$\frac{p_4}{2}, \frac{p_5}{2}, \frac{p_6}{2}, \frac{p_7}{2}$$

$${}^qD_1 = x$$

$${}^qD_2 = x$$

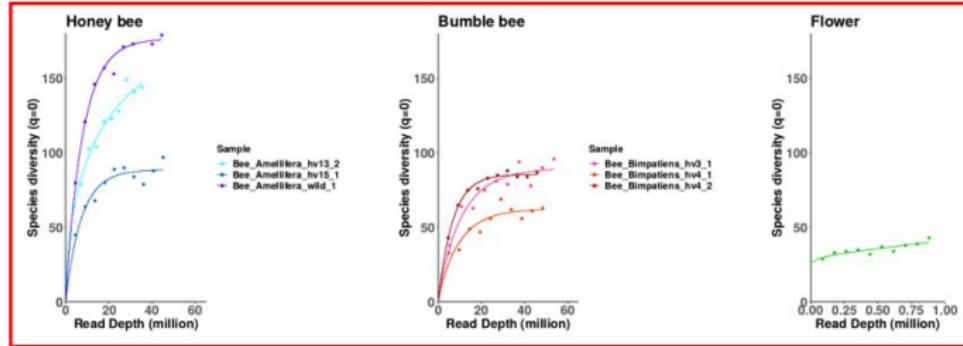
$${}^qD_{\text{pooled}} = 2x$$

## Optimizing sequencing depth according to slope of rarefaction curve

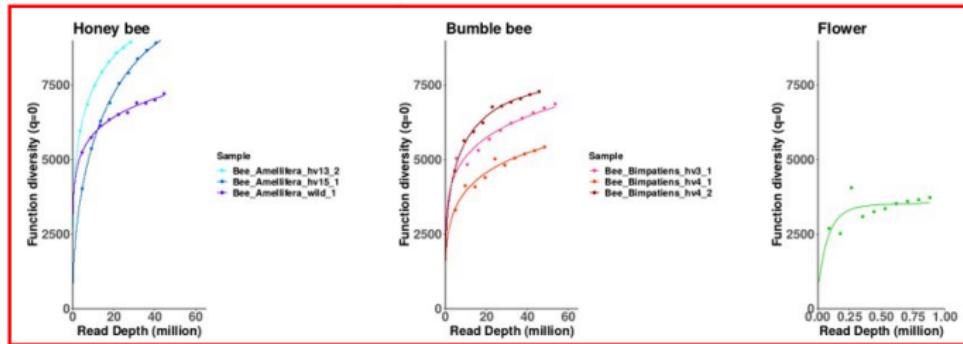


# Verify completeness of original datasets

Species



Function



# Optimal sequencing depth for species/function diversity estimation

Species diversity estimation:



Decreasing  
weight on  
rare species

12.10 Gbp

5.57 Gbp

5.24 Gbp

12.75 Gbp

12.10 Gbp

7.43 Gbp

*order = 0*

*order = 1*

*order = 2*

Function diversity estimation:

- ▶ All datasets were **incomplete** (final slope  $> 15$ )

# Sequencing depth can be optimized for species diversity estimation



Species diversity: 12.10 Gbp  
Function diversity: > 13.5 Gbp

12.75 Gbp  
> 15.9 Gbp

> 0.26 Gbp  
> 0.26 Gbp

# Summary

- ▶ Shotgun metagenomics: powerful in revealing species/function diversity of microbiomes
- ▶ Integrated pipeline: identifying more species, flexibility, easy troubleshooting
- ▶ Optimal depth for species identification: 12/12.9 Gbp for honey/bumble bees (~ 200£/sample)
- ▶ Shallower sequencing with reduced emphasis on rare species
- ▶ Deep sequencing for functional diversity
- ▶ Pilot studies for large scale metagenomic project