

### Abstract

- High-throughput sequencing (HTS) datasets are compositional due to the arbitrary limit of reads imposed by the sequencing instrument<sup>1</sup>
- Command-line tools have a steep learning curve for biologists without scripting experience
- There is a need to reduce the learning curve of exploring HTS data and differential abundance analysis

### Objectives

- Facilitate workflow for exploring HTS data
- Develop graphical user interface for exploring HTS data using compositional methods<sup>2</sup>
- Ensure analysis is transparent and reproducible
- Provide resources for introducing biologists to exploratory analysis of their datasets

### Methods

- Developed using the **Shiny** package for interactive apps in the R programming language
- Compositional approach for transforming counts (log-ratio)
- Imports **ALDEx2** for differential abundance analysis

### Getting started

```
library(omicplotR)
omicplotr.run()
```

Choose Data

Browse... No file selected

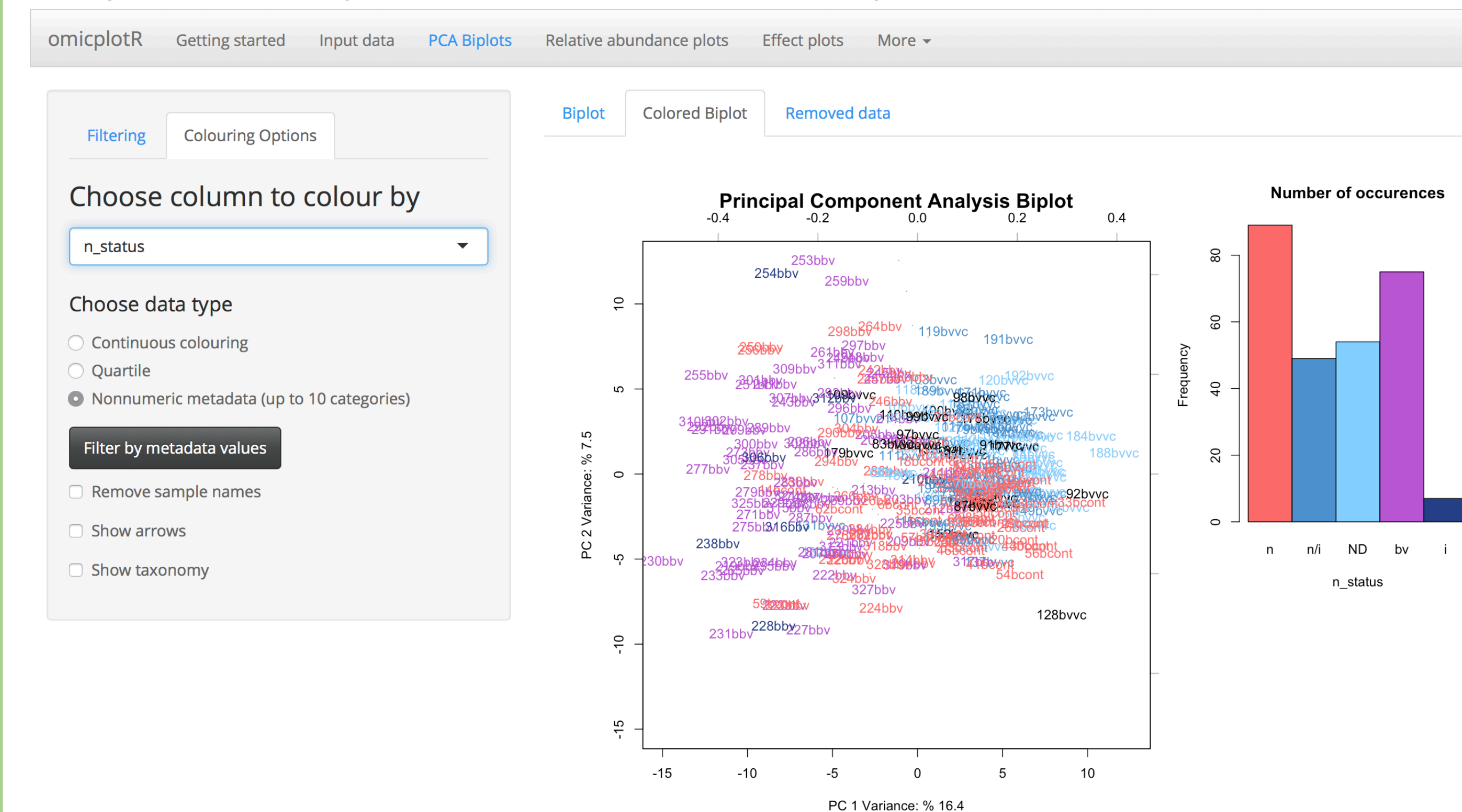
Example data and metadata:

#OTU ID	226bbv	35bcont	261bbv	21bcont	138bvc	taxonomy
0	2003	1814	654	19424	4419	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus;
1	4392	1081	269	2428	31495	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus;
2	886	6379	5286	1926	1840	Bacteria;Actinobacteria;Actinobacteria;Bifidobacteriales;Bifidobacteriaceae;
3	41	48	54	402	266	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus;
4	193	130	109	83	312	Bacteria;Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Megi;

#SampleID	person_id	time	study	probio	age	b_contra
2bcont	2	0	b_cont	NA	35	y
3bcont	3	0	b_cont	NA	44	n
4bcont	4	0	b_cont	NA	33	y
5bcont	5	0	b_cont	NA	33	y

### Graphical user interface

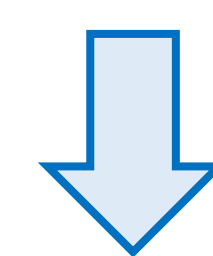
omicplotR: Visual exploration of omic datasets as compositions



### Exploratory analysis

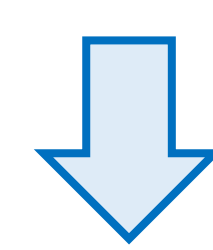
Principal components analysis biplot:

- Filter sparse features or samples
- Colour by metadata
- Filter samples by metadata (subgroup analysis)
- Scree plot
- Visualize removed features and samples



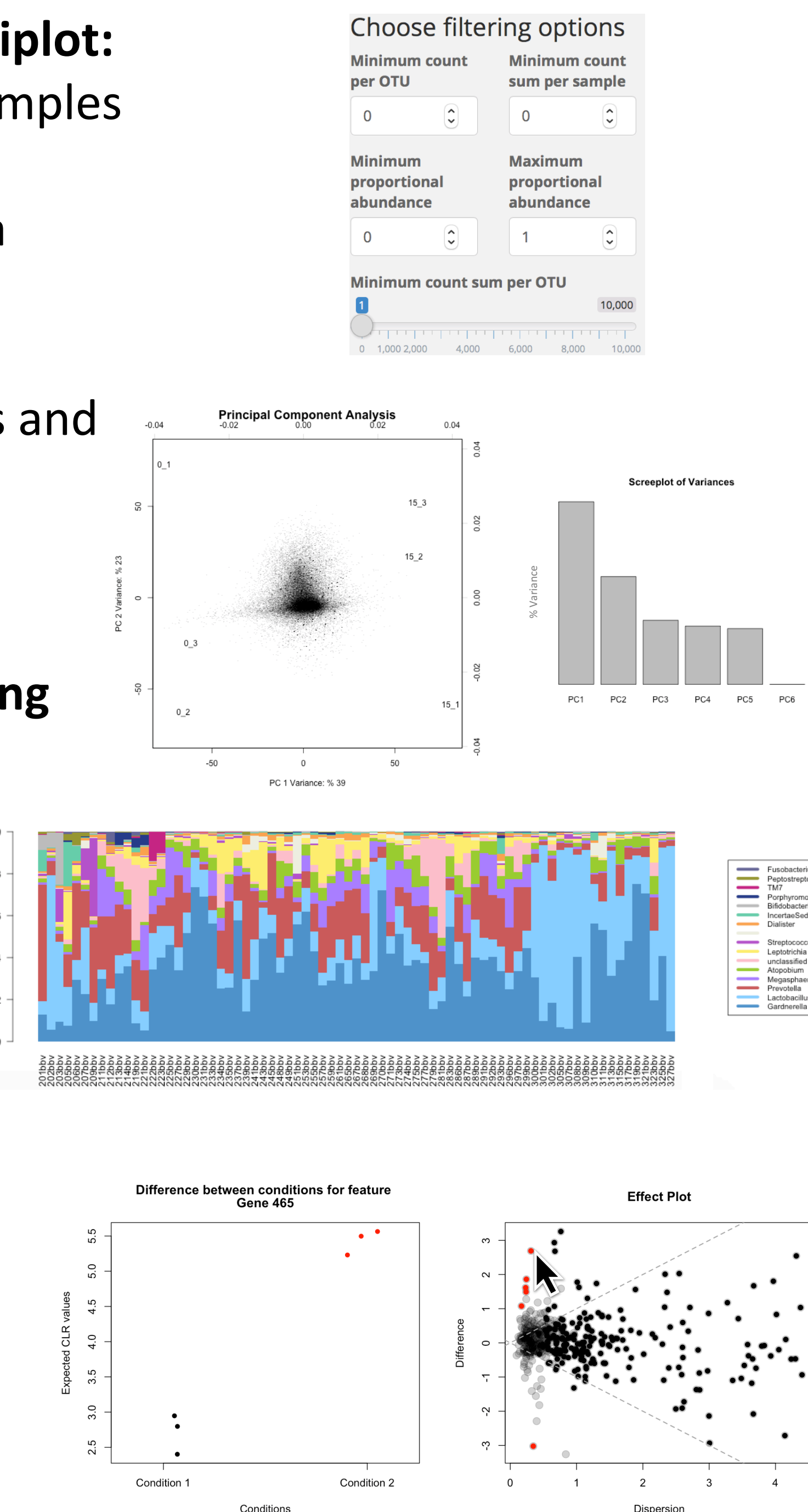
Relative abundance and clustering

- Clustering by dendrogram
- Relative abundance by taxonomic identifiers
- Interactive zoom



Differential abundance analysis

- ALDEx2
- Effect plot
- Bland-Altman plot
- Hover over point to visualize relative abundance of samples



### Development

Downloadable scripts to reproduce each plot

- Ensures reproducibility while allowing customization

Direct integration with EBI metagenomics

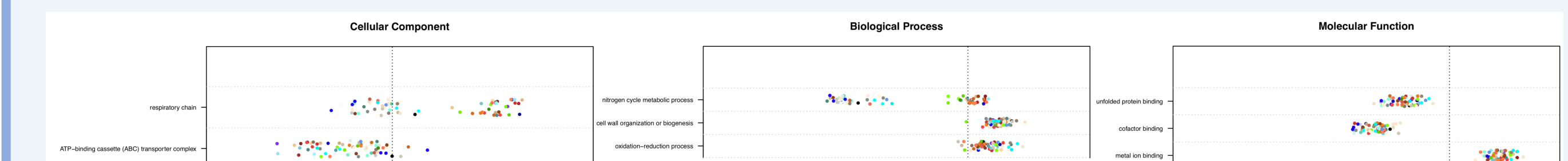
Download dataset from EBI

EBI Project ID  
Project ID

EBI Pipeline version  
Pipeline version (x.y)

Functional analysis using GO slim annotation

- Interactive charts to explore all samples



Tutorials for interpreting each plot are currently being developed for users unfamiliar with these visualization techniques

### Conclusions

- omicplotR facilitates exploratory analysis of HTS data for both new and experienced users
- Provides graphical user interface for exploring HTS data as compositions
- Maintains reproducibility while using **Shiny** app
- Will provide introductory tutorials for interpretation of plots in next release

### References

- Gloor, G.B. et al., 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in microbiology*, 8, p.2224.
- Fernandes, A.D. et al., 2013. ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. J. Parkinson, ed. *PLoS ONE*, 8(7), pp.e67019–15.

### Acknowledgements

Ontario Graduate Scholarship  
Bishnu Sanwal and Theodore Lo  
Graduate Endowment Fund

