














# EasyAmplicon: An easy-to-use, open-source, reproducible, and community-based pipeline for amplicon data analysis in microbiome research

Yong-Xin Liu<sup>1</sup>  | Lei Chen<sup>2</sup> | Tengfei Ma<sup>3</sup>  | Xiaofang Li<sup>4</sup>  |  
Maosheng Zheng<sup>5</sup>  | Xin Zhou<sup>6</sup>  | Liang Chen<sup>6</sup>  | Xubo Qian<sup>7</sup> |  
Jiao Xi<sup>8</sup>  | Hongye Lu<sup>9</sup>  | Huiluo Cao<sup>10</sup> | Xiaoya Ma<sup>11</sup>  | Bian Bian<sup>12</sup> |  
Pengfan Zhang<sup>13</sup> | Jiqui Wu<sup>14,15</sup>  | Ren-You Gan<sup>16</sup> | Baolei Jia<sup>17</sup> |  
Linyang Sun<sup>1</sup> | Zhicheng Ju<sup>1</sup>  | Yunyun Gao<sup>1</sup>  | Tao Wen<sup>18</sup>  | Tong Chen<sup>19</sup>

<sup>1</sup>Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, Guangdong, China

<sup>2</sup>Department of Vascular Surgery, Fu Xing Hospital, Capital Medical University, Beijing, China

<sup>3</sup>State Key Laboratory of Grassland Agro-ecosystems, Centre for Grassland Microbiome, College of Pastoral Agricultural Science and Technology, Lanzhou University, Lanzhou, Gansu, China

<sup>4</sup>Centre for Agricultural Resources Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Shijiazhuang, China

<sup>5</sup>College of Environmental Science and Engineering, North China Electric Power University, Beijing, China

<sup>6</sup>Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

<sup>7</sup>Department of Pediatrics, Affiliated Jinhua Hospital, Zhejiang University School of Medicine, Jinhua, Zhejiang, China

<sup>8</sup>College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi, China

<sup>9</sup>Key Laboratory of Oral Biomedical Research of Zhejiang Province, Cancer Center of Zhejiang University, Clinical Research Center for Oral Diseases of Zhejiang Province, School of Stomatology, Zhejiang University School of Medicine, Stomatology Hospital, Hangzhou, Zhejiang, China

<sup>10</sup>Department of Microbiology, University of Hong Kong, Hong Kong, China

<sup>11</sup>Center of Excellence in Fungal Research, Mae Fah Luang University, Chiang Rai, Thailand

<sup>12</sup>Graduate School of Frontier Sciences, University of Tokyo, Chiba, Japan

<sup>13</sup>Department of Plant-Microbe Interactions, Max Planck Institute for Plant Breeding Research, Cologne, Germany

<sup>14</sup>APC Microbiome Institute, University College Cork, Cork, Ireland

<sup>15</sup>Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

<sup>16</sup>Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A\*STAR), Singapore, Singapore

<sup>17</sup>Department of Life Science, Chung-Ang University, Seoul, Republic of Korea

<sup>18</sup>The Key Laboratory of Plant Immunity Jiangsu Provincial Key Lab for Organic Solid Waste Utilization Jiangsu Collaborative Innovation Center for Solid Organic Waste Resource Utilization, National Engineering Research Center for Organic-Based Fertilizers, Nanjing Agricultural University, Nanjing, China

<sup>19</sup>National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

Yong-Xin Liu, Lei Chen, and Tengfei Ma contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *iMeta* published by John Wiley & Sons Australia, Ltd on behalf of *iMeta* Science.

### Correspondence

Yong-Xin Liu, Shenzhen Branch,  
Guangdong Laboratory of Lingnan  
Modern Agriculture, Genome Analysis  
Laboratory of the Ministry of Agriculture  
and Rural Affairs, Agricultural Genomics  
Institute at Shenzhen, Chinese Academy  
of Agricultural Sciences, Shenzhen,  
Guangdong 518120, China.  
Email: [liuyongxin@caas.cn](mailto:liuyongxin@caas.cn)

Tao Wen, The Key Laboratory of Plant  
Immunity Jiangsu Provincial Key Lab for  
Organic Solid Waste Utilization Jiangsu  
Collaborative Innovation Center for Solid  
Organic Waste Resource Utilization,  
National Engineering Research Center for  
Organic-Based Fertilizers, Nanjing  
Agricultural University, Nanjing 210095,  
China.  
Email: [taowen@njau.edu.cn](mailto:taowen@njau.edu.cn)

Tong Chen, National Resource Center for  
Chinese Materia Medica, China Academy  
of Chinese Medical Sciences, Beijing  
100700, China.  
Email: [chent@nrc.ac.cn](mailto:chent@nrc.ac.cn)

### Funding information

Youth Innovation Promotion Association  
of the Chinese Academy of Sciences,  
Grant/Award Number: 2021092; National  
Natural Science Foundation of China,  
Grant/Award Number: 31772400

### Abstract

It is difficult for beginners to learn and use amplicon analysis software because there are so many software tools to choose from, and all of them need multiple steps of operation. Herein, we provide a cross-platform, open-source, and community-supported analysis pipeline EasyAmplicon. EasyAmplicon has most of the modules needed for an amplicon analysis, including data quality control, merging of paired-end reads, dereplication, clustering or denoising, chimera detection, generation of feature tables, taxonomic diversity analysis, compositional analysis, biomarker discovery, and publication-quality visualization. EasyAmplicon includes more than 30 cross-platform modules and R packages commonly used in the field. All steps of the pipeline are integrated into RStudio, which reduces learning costs, keeps the flexibility of the analysis process, and facilitates personalized analysis. The pipeline is maintained and updated by the authors and editors of WeChat official account “Meta-genome.” Our team will regularly release the latest tutorials both in Chinese and English, read the feedback from users, and provide help to them in the WeChat account and GitHub. The pipeline can be deployed on various platforms, and the installation time is less than half an hour. On an ordinary laptop, the whole analysis process for dozens of samples can be completed within 3 h. The pipeline is available at GitHub (<https://github.com/YongxinLiu/EasyAmplicon>) and Gitee (<https://gitee.com/YongxinLiu/EasyAmplicon>).

### KEYWORDS

amplicon; bioinformatics; microbiome; pipeline; visualization, metagenome

### Highlights

- EasyAmplicon is a user-friendly, cross-platform, and community-supported pipeline for amplicon data analysis.
- It has most of the modules for data processing and visualization in microbiome research.
- The pipeline is maintained and updated regularly, and we encourage users to contribute appropriate code.

## INTRODUCTION

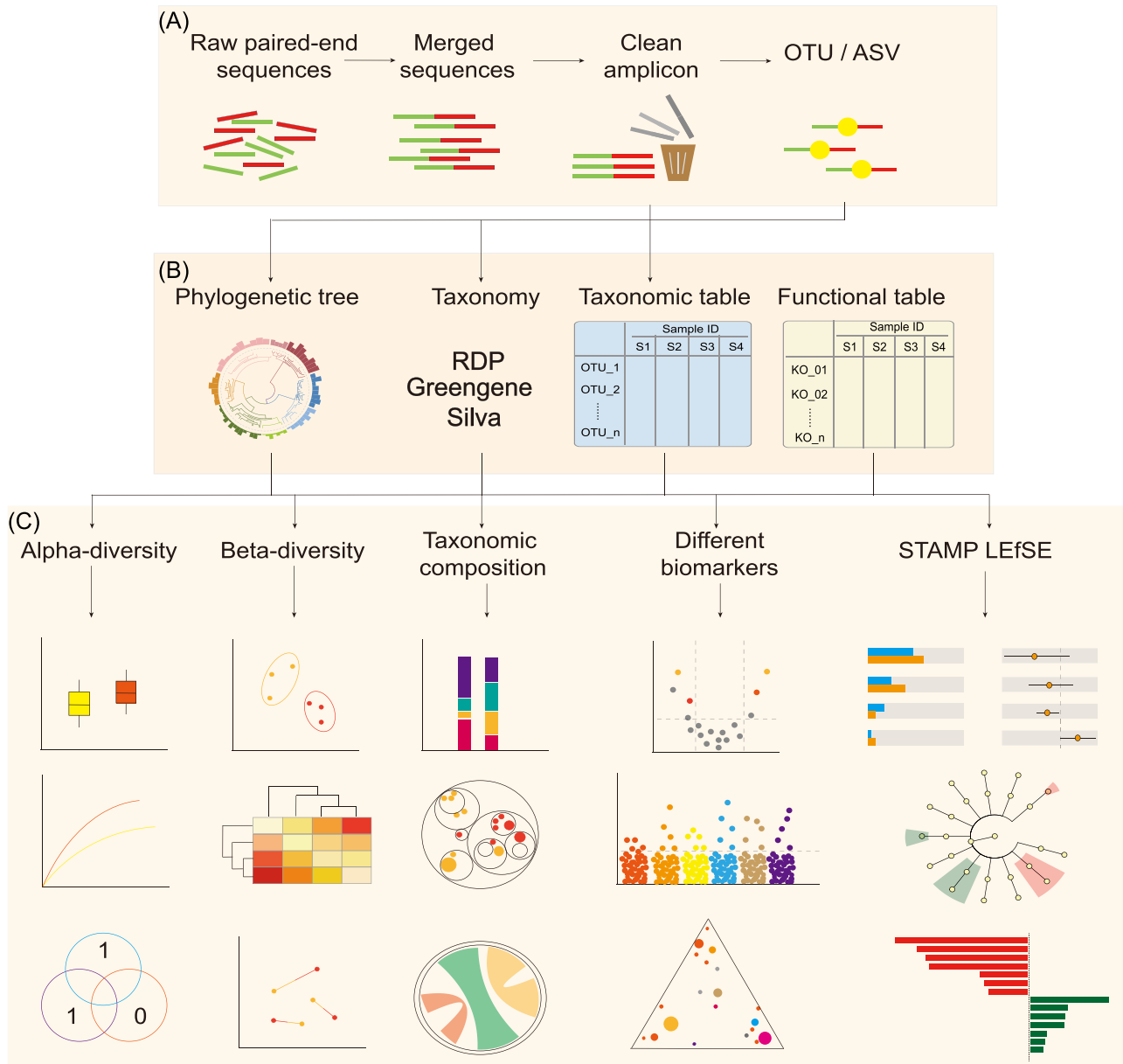
The rapid development of high-throughput sequencing technologies in the past 20 years has promoted an increasingly deeper exploration of the crucial roles of microbiome in humans [1–5], animals [6–8], plants [9–11], and the environment [12–14]. Most of them were driven by amplicon sequencing (such as 16S rDNA sequencing of bacteria or archaea, eukaryotic 18S rDNA or internal transcribed spacer, and nitrogen-fixing prokaryote's *nifH* gene) and had profiled the taxonomic

composition of the microbiome in various environments [15–17].

Wet laboratory operations of amplicon sequencing are now standardized, and most operations are implemented by specialized biotechnology companies or sequencing centers. However, bioinformatics analyses of amplicon data are still challenging, and the existence of overwhelming software, methods, and algorithms brings difficult choices for beginners. The popular amplicon analysis pipelines include mothur [18], USEARCH [19], and QIIME [20], all of which have been

cited over 10,000 times. However, they still have obvious shortcomings, such as a lack of downstream statistical analyses and visualization solutions, higher time costs, and being limited to specified operating systems. Some online analysis webservers are easy to use, such as Qiita [21], MGnify [22], and gcMeta [23], but they also have several limitations, such as slow upload speed, long waiting/running time, and few adjustable parameters, which make it impossible to conduct customized analyses [24, 25].

The lack of an easy-to-use and flexible amplicon analysis pipeline seriously restricts researchers to understand the data analysis process and hinders the development of this field. Therefore, we developed an easy-to-use, open-source, and cross-platform amplicon analysis pipeline—EasyAmplicon. It can be used in both command-line mode and interactive mode in RStudio. Currently, it provides more than 20 visualization styles and generates publication-quality figures easily. The open-source code could facilitate reproducible analysis



**FIGURE 1** Pipeline of EasyAmplicon for analyzing paired-end amplicon sequences. (A) Dimensionality reduction: processing raw sequencing reads into feature tables. (B) Analysis: providing phylogenetic analysis, taxonomic classification, functional prediction, and alpha- and beta-diversity calculations. (C) Statistics and visualization: generating publication-quality figures and performing statistical tests for biological interpretations. ASVs, amplicon sequence variants; OTU, operational taxonomic units.

and allow personalized modification. In addition, it also generates standard input for the most popular software, such as STAMP [26], LEfSe [27], PICRUSt 1 & 2 [28, 29], BugBase [30], FAPROTAX [31], ImageGP [32], and iTOL [33]. EasyAmplicon provides a free, reproducible, and personalized solution for amplicon analysis, which could be an amazing software tool for microbiome research.

## RESULTS

### Overview of EasyAmplicon pipeline

EasyAmplicon is an integrated pipeline for amplicon data analysis and visualization on a laptop or server, and it provides various tables and figures to explore underlying biological interpretations. This pipeline is easy to install on Windows, MacOS, and Linux systems. The installation method is described in detail in the Methods section or available at <https://github.com/YongxinLiu/EasyAmplicon>. For the test data, comprising 18 samples and 50,000 PE250 reads per sample, the complete analysis could be finished within about 3 h with a peak memory footprint of less than 4 gigabytes (CPU: 2 cores, 2.1 GHz).

The EasyAmplicon is an end-to-end pipeline. It starts with raw reads and ends with data tables and publication-quality figures (Figure 1). It mainly comprises three steps: dimensionality reduction, analysis, and visualization & statistics (Figure 1). All the related software is easy to install (Table 1), and we provide a batch download package to accelerate pipeline deployment.

### Running the pipeline in command line or R markdown mode

First, we open the pipeline file “pipeline.sh” using RStudio. After setting the working directory, the analysis process could be run step-by-step with only a mouse clicking the “Run” button. For users to conduct their own analysis, only the raw sequencing data and sample metadata are needed, and the following analysis would be processed by EasyAmplicon. If RStudio is not applicable, we can copy and paste the scripts into the pipeline.sh and run them in any Shell environment (such as a terminal in Linux/Mac either locally or remotely, or Git bash in Windows). All the related software and packages are listed in Table 1. All the figures are saved in PDF format by default, and some examples are shown in Figures 2 and 3.

To make statistics and visualization of microbiome data more personalized, users can open the “Tutorial.Rmd” document in RStudio and then modify the details of the figures, such as item order, color scheme, legend layout, and so on. It can even generate a publish-ready combo figure (Figures 2 and 3) and a reproducible HTML format report (Tutorial.html).

### Third-party software supporting

EasyAmplicon does not cover all the functions required for microbiome analysis. There are currently some mainstream and very distinctive microbiome analysis tools, such as STAMP [26], LEfSe [27], PICRUSt 1 & 2 [28, 29], BugBase [30], FAPROTAX [31], and iTOL [33].

**TABLE 1** Software and packages included in EasyAmplicon

| Software        | Function in the pipeline  | Website   |
|-----------------|---|---|
| Git for Windows | Provides Linux Shell like environment in Windows                                | <a href="http://gitforwindows.org/">http://gitforwindows.org/</a>                                       |
| R               | Statistical processing and data visualization                                   | <a href="https://www.r-project.org">https://www.r-project.org</a>                                       |
| RStudio         | Integrated development environment for R and Shell                              | <a href="https://posit.co/">https://posit.co/</a>   |
| VSEARCH         | A fast, free, and cross-platform pipeline for amplicon sequencing analysis [34] | <a href="https://github.com/torognes/vsearch">https://github.com/torognes/vsearch</a>                   |
| USEARCH         | Processes sequences and calculates alpha- and beta-diversities [19]             | <a href="http://www.drive5.com/usearch/">http://www.drive5.com/usearch/</a>                             |
| SeqKit          | Toolkit for FASTA or FASTQ file manipulation [35]                               | <a href="https://github.com/shenwei356/seqkit">https://github.com/shenwei356/seqkit</a>                 |
| ggplot2         | R package for data visualization  | <a href="https://github.com/tidyverse/ggplot2">https://github.com/tidyverse/ggplot2</a>                 |
| ggClusterNet    | R package for microbiome network visualization [36]                             | <a href="https://github.com/taowenmicro/ggClusterNet/">https://github.com/taowenmicro/ggClusterNet/</a> |
| vegan           | R package for alpha and beta diversity analysis                                 | <a href="https://cran.r-project.org/package=vegan">https://cran.r-project.org/package=vegan</a>         |
| ggraph          | Creates layout for tree map and circle packing chart                            | <a href="https://github.com/thomasps85/ggraph">https://github.com/thomasps85/ggraph</a>                 |
| circlize        | Circular visualization [37]   | <a href="https://github.com/jokergoo/circlize">https://github.com/jokergoo/circlize</a>                 |

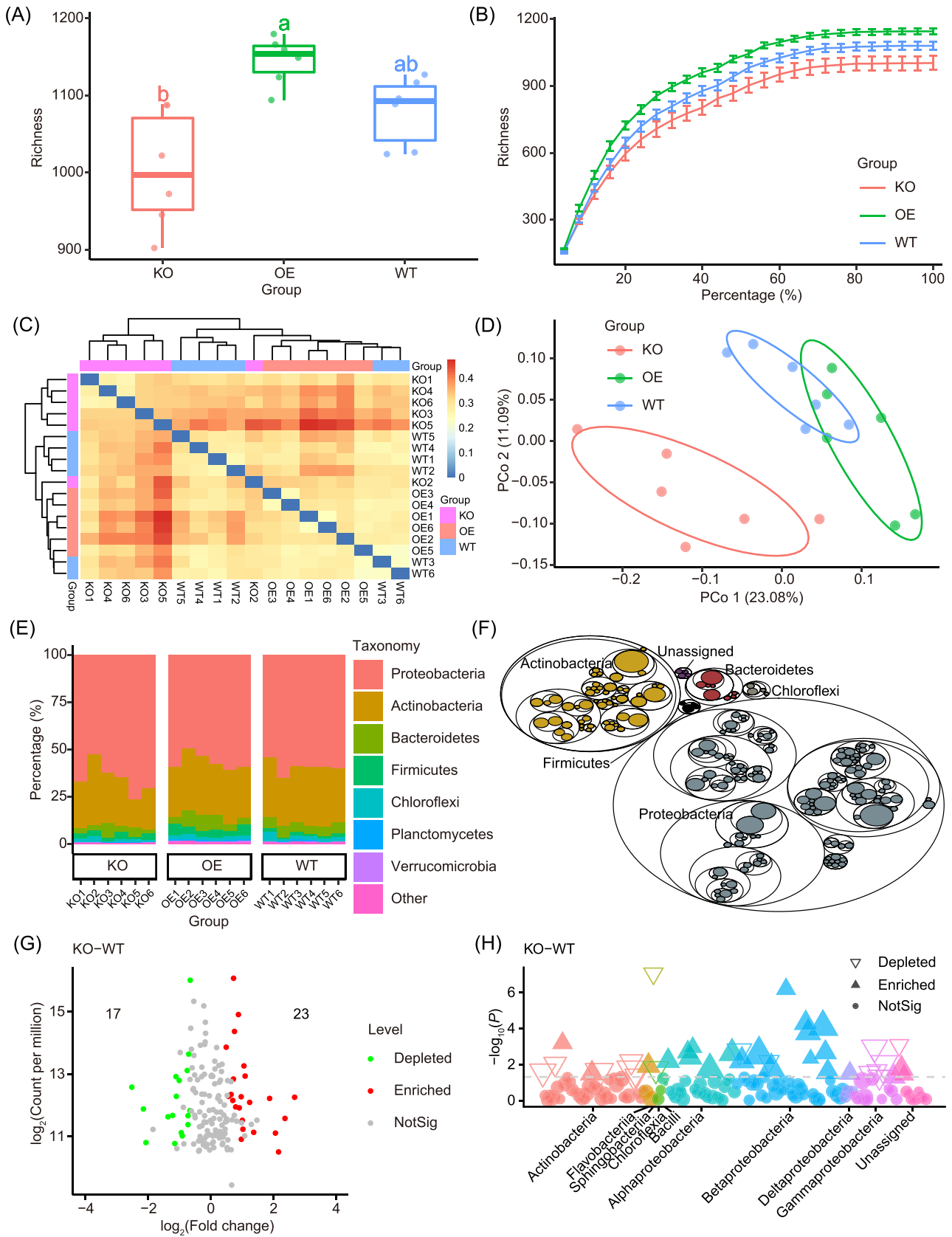


FIGURE 2 (See caption on next page)

However, some input files are difficult to prepare for users without bioinformatics backgrounds. In EasyAmplicon, a lot of scripts are used to prepare input for all the above software tools. The example visualizations, STAMP (Figure 4A), LEfSe (Figure 4B), BugBase (Figure 4C), and iTOL (Figure 4D) are shown. As for the most popular QIIME 2 pipeline, the intermediate files generated by EasyAmplicon can be imported into QIIME 2, and the output files from QIIME 2 can also be imported into EasyAmplicon for downstream analyses.

## Anticipated results

EasyAmplicon provides multiple visualization styles for amplicon data analysis. For alpha diversity (within-sample diversity), the boxplot is the best way to visualize the data and compare each group (Figure 2A), and the different letters represent significant differences ( $p < 0.05$ , ANOVA, Tukey HSD test). Rarefaction analysis reveals that the features reach the saturation stage with increasing sequencing depth, and lines and error bars represent the mean and standard error, respectively (Figure 2B). If you want to examine the unique or common features among samples or groups, the Venn diagram is a good way to show this pattern (Figure 3A). As for beta diversity, a heatmap based on Bray–Curtis dissimilarity would be a good visualization method. The colored grouping labels show how the samples cluster (Figure 2C).

## DISCUSSION

At present, for amplicon analysis, the most popular pipelines are QIIME [20] and QIIME 2 [15], which have been cited 54,900 times (Google Scholar, January 4, 2023). However, the two pipelines have some disadvantages that limit their use in microbiome analysis, such as a too-large installation package, no support for the Windows system, and a lack of publication-quality visualization. EasyAmplicon is trying to solve the above problems.

Currently, this is just the first version of EasyAmplicon. Runtime and memory usage depend on the data set size. The current version has been used by more than thousands of users and formally cited 36 times by the end of 2022 (searching “EasyAmplicon” in Google Scholar). The authors and core team of the WeChat account “Meta-genome” will update the pipeline in time. The scripts for correlation, network analysis [38, 39], random forest [40, 41], machine learning [42], deep learning [43], transfer learning [44], and source track [45–47] analyses are ongoing and will be included in the pipeline soon. A webserver version like MicrobiomeAnalyst [48, 49] will be set up in the future. More general command-line scripts and visualization styles are still in development, and they will be available in a new version of the pipeline. Anyone who is interested in this project is welcome to contribute scripts pertaining to analysis methods, visualization styles, and other issues mentioned in the GitHub repository.

## CONCLUSION

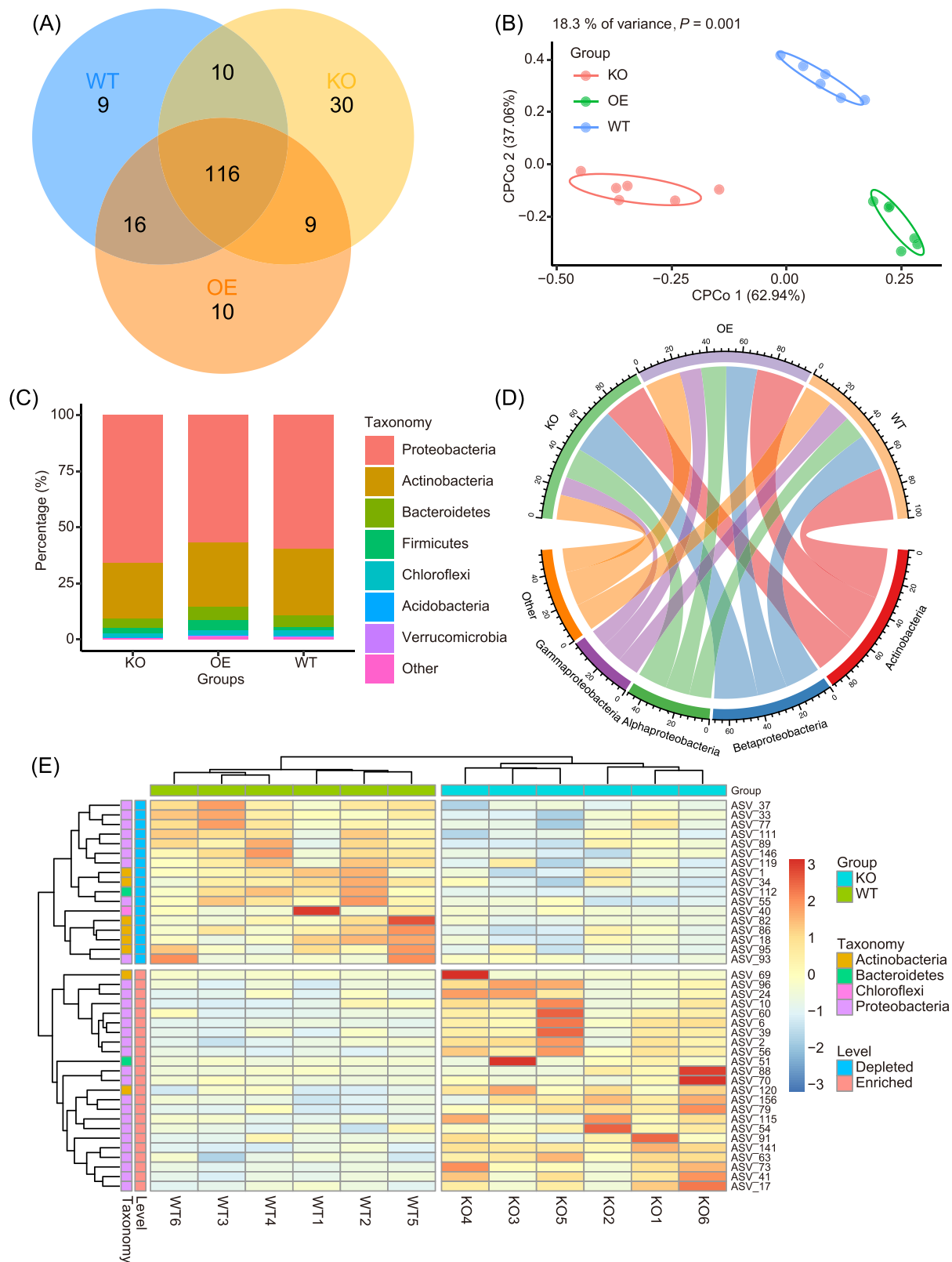
In summary, the EasyAmplicon pipeline provides an efficient, cross-platform framework for amplicon analysis. Additionally, more than 20 predefined analysis and visualization solutions are provided for the multi-dimensional exploration of your data and for generating publication-quality figures. Additionally, EasyAmplicon provides some utilities to be integrated with other widely used software for various needs.

## METHODS

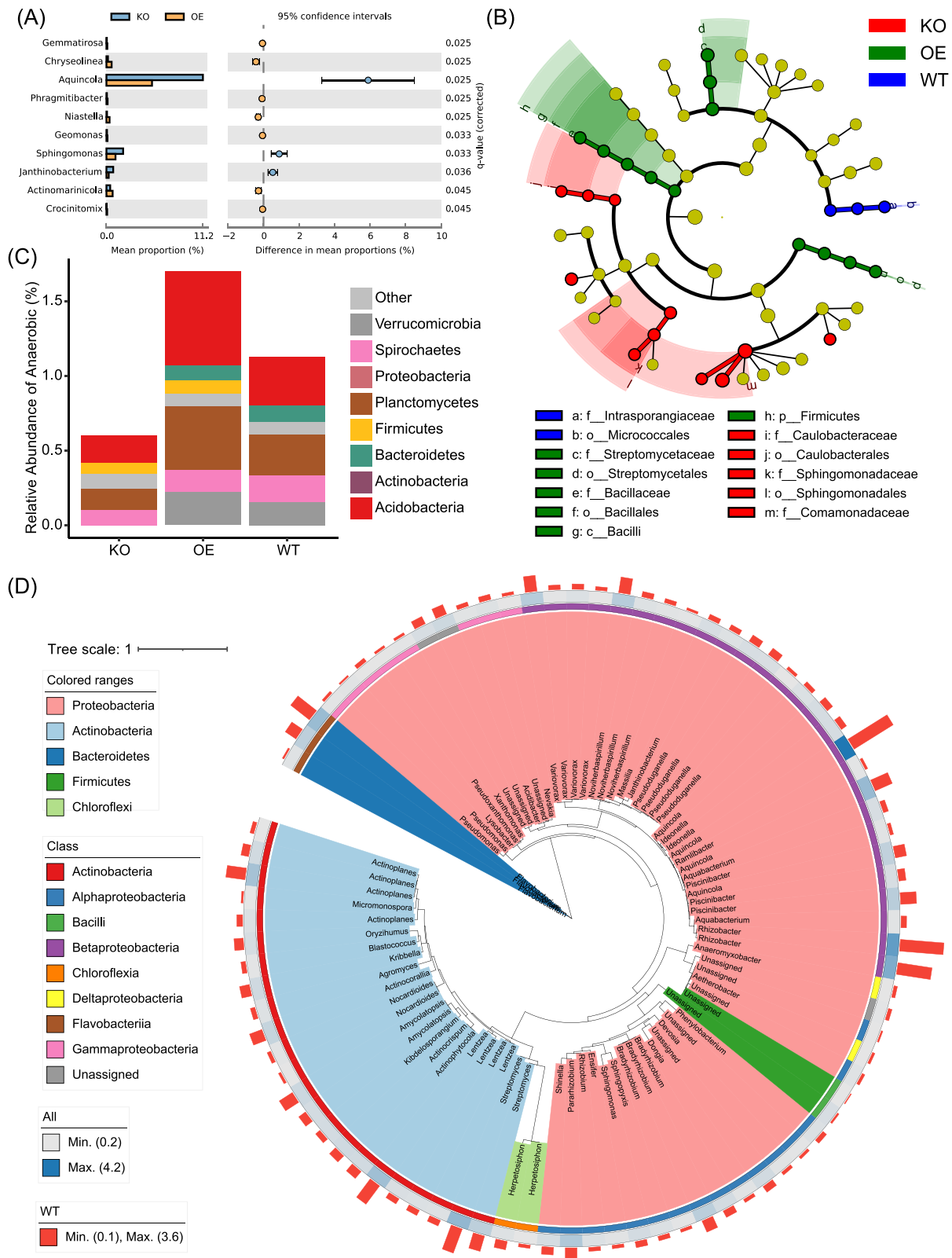
### Quick start of EasyAmplicon

EasyAmplicon is coded mainly in Shell bash and R language and could be run in command-line (terminal) mode or RStudio interactive mode. It is recommended to be deployed on a Windows system (with the installation of Git for Windows) and run in RStudio, especially for researchers

**FIGURE 2** Examples of publication-quality visualizations. (A) Boxplot showing alpha diversity in richness metrics among groups. Different letters indicate statistical significance among groups ( $p < 0.05$ , ANOVA, Tukey HSD). The horizontal bars within boxes represent medians. The tops and bottoms of the boxes represent the 75th and 25th percentiles, respectively. The upper and lower whiskers extend to data no more than 1.5× the interquartile range from the upper and lower edge of the box, respectively. (B) Rarefaction curve of richness shows that features reach saturation stage with increasing sequencing depth. Each vertical bar represents standard error. (C) Heatmap based on Bray–Curtis dissimilarity. (D) Principal coordinate analysis (PCoA) of Bray–Curtis dissimilarity. (E) Stacked bar plot of taxonomic composition in grouped samples at phylum level. (F) Tree map of taxonomic composition. (G) Volcano plot showing significantly differential abundance taxa between KO and WT groups. (H) Manhattan plot showing different features and related taxa between KO and WT groups. The numbers of replicated samples in this figure are as follows: in KO ( $n = 6$ ), OE ( $n = 6$ ), and WT ( $n = 6$ ). KO, knock-out; OE, overexpression; WT, wild-type.



**FIGURE 3** Supplementary examples of publication-quality visualizations to Figure 2. (A) Venn diagram showing common and unique ASVs (relative abundance >0.1%) among three groups. (B) Constrained principal coordinate analysis (CPCoA) of three groups. (C) Stacked plot of average relative abundance at phylum level of three groups. (D) Circle plot of average relative abundance at phylum level of three groups. (E) Heatmap showing significantly different ASVs between KO and WT groups (Wilcoxon test,  $p < 0.05$ ). ASVs, amplicon sequence variants; KO, knock-out; WT, wild-type.



**FIGURE 4** Visualizations generated by third-party software using the intermediate files of EasyAmplicon. (A) Extended error bar plot at genus level in WT and KO groups by STAMP. (B) Cladogram showing biomarkers in each group by LefSe. (C) Percentage of BugBase annotated anaerobic bacteria at the phylum level. (D) Phylogenetic tree of 86 ASVs (relative abundance > 0.2%). The tree background is colored by Phylum. The outer strip represents different classes. The heatmap represents the average relative abundance of all samples. The bar plot represents the relative abundance of the WT group. ASVs, amplicon sequence variants; KO, knock-out; WT, wild-type.



**TABLE 2** Summary of main visualization functions in EasyAmplicon

| Plot                    | Script                               | Description  | References |
|-------------------------|--------------------------------------|--|------------|
| Boxplot                 | alpha_boxplot.R                      | The plot shows data distribution in each group. Dots represent each sample, and labeled letters indicate statistical significance among groups | [10]       |
| Rarefaction curve       | alpha_rare_curve.R                   | Richness of rarefied samples or groups from 1% to 100%   | [10]       |
| Venn diagram            | sp_vennDiagram.sh                    | Visualizes common and unique elements among 2–5 groups   | [53, 54]   |
| Ordination scatter plot | beta_pcoa.R<br>beta_cpcoa.R          | Shows the results of dimensionality reduction  | [55]       |
| Heatmap                 | sp_pheatmap.sh<br>compare_heatmap.sh | Shows distance or similarity matrix and differential abundance features  | [50, 56]   |
| Stack plot              | tax_stackplot.R                      | Shows taxonomic and functional composition in each sample or group   | [57]       |
| Circular plot           | tax_circlize.R                       | Shows taxonomic composition  | [58]       |
| Circle packing chart    | tax_maptree.R                        | Shows the relations and relative abundance of taxonomic hierarchy  | [59]       |
| Volcano plot            | compare_volcano.R                    | The dots in the plot show abundance and fold changes between two groups  | [50]       |
| Manhattan plot          | compare_manhattan.sh                 | Shows taxonomy, abundance, and pattern between two groups  | [60]       |

without programming knowledge and skills. In addition, it also supports MacOS and Linux. To install it, please follow the instructions available at <https://github.com/YongxinLiu/EasyAmplicon>. Some dependent software and packages are listed in Table 1, and they are integrated for easy installation. The analysis process mainly includes three steps, as shown in Figure 1. To prove its practicability, we provide a demo data set that contains 18 samples belonging to three groups, and each sample is rarefied to 50,000 reads. This example data set is a subset of our previously published data (CRA001464) [50] (rarefied example data are deposited in GSA <https://ngdc.cncb.ac.cn/gsa/>, with accession ID: CRA002352).

## Dimensionality reduction (from sequences to tables)

The accepted input includes paired-end or single-end/merged sequences (fastq format), clean amplicons (fasta format), and even the intermediate files generated by other pipelines, as shown in Figure 1. Most amplicons are sequenced on the Illumina HiSeq. 2500 or NovaSeq. 6000 platform in paired-end 250 bp mode. Typically, the pipeline starts with paired-end reads in fastq format and merges them to get single-end sequences. Primers and barcodes are cut, and then low-quality reads are filtered out to get clean amplicons. These steps are performed mainly using vsearch [34] or USEARCH [19] (Figure 1A). The clean amplicons of 16S rDNA can be directly mapped to the reference database GreenGenes [51], and the closed-reference operational

taxonomic units (OTUs) table can be generated, which can be used as the input of PICRUST to predict the potential functions [28, 52], and can be used as the input of BugBase for phenotypic prediction [30]. Alternatively, clean amplicons are usually clustered into OTUs (97% similarity) or denoised into amplicon sequence variants (ASVs) in de novo mode. Finally, the clean amplicons will be mapped to the de novo identified OTUs/ASVs to generate a feature table. Representative sequences can be used to construct a phylogenetic tree and perform taxonomic annotation (Figure 1A).

## Analysis (from big tables to small tables)

Feature tables are milestone outputs of the dimensionality reduction step. We can use the feature tables and phylogenetic tree to calculate all kinds of alpha- and beta-diversity metrics. The feature tables with taxonomy annotations can be used to collapse into a specific taxonomic level and discover biomarkers at all taxonomic levels (Figure 1B). In addition, EasyAmplicon provides many glue scripts to generate input files for other widely used tools, such as QIIME 2 [15], STAMP [26], and LEfSe [27].

## Statistics and visualization (from tables to figures)

EasyAmplicon can generate visualizations for alpha diversity, beta diversity, taxonomic composition, and

biomarkers along with related statistical tables (Figure 1C and 2; Table 2), and these publication-quality graphs include a box-plot, scatter plot, stacked bar plot, and heatmap. In addition, the output of EasyAmplicon can be imported into STAMP [26], or LEfSe [27] for biomarker identification, and visualized in an extended error bar plot or Cladogram, respectively.

### AUTHOR CONTRIBUTIONS

Yong-Xin Liu, Tao Wen, Tong Chen, Lei Chen, and Tengfei Ma developed the pipeline, performed the analysis, and wrote the manuscript. Xiaofang Li, Maosheng Zheng, Xin Zhou, Liang Chen, Xubo Qian, Jiao Xi, Hongye Lu, Huiluo Cao, Xiaoya Ma, Bian Bian, Pengfan Zhang, Jiqu Wu, Ren-You Gan, Baolei Jia, Linyang Sun, Zhicheng Ju, and Yunyun Gao have tested the pipeline, suggested pipeline amendments, and revised the manuscript. Tao Wen, Tong Chen, and Tengfei Ma write the codes for several visualization styles. Tong Chen improved the pipeline and designed the framework of visualization. Xin Zhou contributed to the VSEARCH analysis pipeline document. Yong-Xin Liu, Tao Wen, Tong Chen, Xin Zhou, Xubo Qian, and Liang Chen contributed to the scripts and developed the user documentation. All authors have read the final manuscript and approved it for publication.

### ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (U21A20182), and the Youth Innovation Promotion Association CAS (2021092).

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### DATA AVAILABILITY STATEMENT

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive [61] in the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA002352) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. EasyAmplicon is freely available and implemented in Shell and R, and easy to install. The step-by-step protocols can be found at GitHub <https://github.com/YongxinLiu/EasyAmplicon> or Gitee <https://gitee.com/YongxinLiu/EasyAmplicon>. All the data of the figures can be downloaded in GitHub, or supplementary tables.

### ORCID

Yong-Xin Liu  <https://orcid.org/0000-0003-1832-9835>

Tengfei Ma  <https://orcid.org/0000-0003-1426-1060>

Xiaofang Li  <http://orcid.org/0000-0003-1554-4484>

Maosheng Zheng  <http://orcid.org/0000-0003-2633-7763>

Xin Zhou  <http://orcid.org/0000-0003-3143-2079>

Liang Chen  <http://orcid.org/0000-0002-5706-6559>

Jiao Xi  <http://orcid.org/0000-0001-5428-7391>

Hongye Lu  <http://orcid.org/0000-0003-4079-9681>

Xiaoya Ma  <http://orcid.org/0000-0002-2204-1992>

Jiqu Wu  <http://orcid.org/0000-0001-5078-7620>

Zhicheng Ju  <http://orcid.org/0000-0002-2562-5774>

Yunyun Gao  <http://orcid.org/0000-0003-3389-9385>

Tao Wen  <http://orcid.org/0000-0003-0102-0487>

### REFERENCES

- Proctor, Lita M., Heather H. Creasy, Jennifer M. Fettweis, Jason Lloyd-Price, Anup Mahurkar, Wenyu Zhou, Gregory A. Buck, et al. 2019. "The Integrative Human Microbiome Project." *Nature* 569: 641–48. <https://doi.org/10.1038/s41586-019-1238-8>
- Zuo, Tao, Xiaojian Wu, Weiping Wen, and Ping Lan. 2021. "Gut Microbiome Alterations in COVID-19." *Genomics, Proteomics & Bioinformatics* 19: 679–88. <https://doi.org/10.1016/j.gpb.2021.09.004>
- Chen, Lianmin, Daoming Wang, Sanzhima Garmaeva, Alexander Kurilshikov, Arnau Vich Vila, Ranko Gacesa, Trishla Sinha, et al. 2021. "The Long-Term Genetic Stability and Individual Specificity of the Human Gut Microbiome." *Cell* 184: 2302–2315. <https://doi.org/10.1016/j.cell.2021.03.024>
- Qin, Junjie, Ruiqiang Li, Jeroen Raes, Manimozhian Arumugam, Kristoffer Solvsten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, et al. 2010. "A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing." *Nature* 464: 59–65. <https://doi.org/10.1038/nature08821>
- Falony, Gwen, Marie Joossens, Sara Vieira-Silva, Jun Wang, Youssef Darzi, Karoline Faust, Alexander Kurilshikov, et al. 2016. "Population-Level Analysis of Gut Microbiome Variation." *Science* 352: 560–64. <https://doi.org/10.1126/science.aad3503>
- Stewart, Robert D., Marc D. Auffret, Amanda Warr, Alan W. Walker, Rainer Roehe, and Mick Watson. 2019. "Compendium of 4,941 Rumen Metagenome-Assembled Genomes for Rumen Microbiome Biology and Enzyme Discovery." *Nature Biotechnology* 37: 953–61. <https://doi.org/10.1038/s41587-019-0202-3>
- Huang, Peng, Yan Zhang, Kangpeng Xiao, Fan Jiang, Hengchao Wang, Dazhi Tang, Dan Liu, et al. 2018. "The Chicken Gut Metagenome and the Modulatory Effects of Plant-Derived Benzylisoquinoline Alkaloids." *Microbiome* 6: 211. <https://doi.org/10.1186/s40168-018-0590-5>
- Xiao, Liang, Jordi Estellé, Pia Kiillerich, Yuliaxis Ramayo-Caldas, Zhong kui Xia, Qiang Feng, Suisha Liang, et al. 2016. "A Reference Gene Catalogue of the Pig Gut Microbiome." *Nature Microbiology* 1: 16161. <https://doi.org/10.1038/nmicrobiol.2016.161>
- Liu, Yong-Xin, Yuan Qin, and Yang Bai. 2019. "Reductionist Synthetic Community Approaches in Root Microbiome Research." *Current Opinion in Microbiology* 49: 97–102. <https://doi.org/10.1016/j.mib.2019.10.010>

10. Zhang, Jingying, Yong-Xin Liu, Na Zhang, Bin Hu, Tao Jin, Haoran Xu, Yuan Qin, et al. 2019. "NRT1.1B is Associated with Root Microbiota Composition and Nitrogen Use in Field-Grown Rice." *Nature Biotechnology* 37: 676–84. <https://doi.org/10.1038/s41587-019-0104-4>
11. Xi, Jiao, Zanbo Ding, Tengqi Xu, Wenxing Qu, Yanzhi Xu, Yongqing Ma, Quanhong Xue, Yongxin Liu, and Yanbing Lin. 2022. "Maize Rotation Combined with *Streptomyces* Rochei D74 to Eliminate *Orobanche Cumana* Seed Bank in the Farmland." *Agronomy* 12: 3129. <https://doi.org/10.3390/agronomy12123129>
12. Nayfach, Stephen, Simon Roux, Rekha Seshadri, Daniel Udway, Neha Varghese, Frederik Schulz, Dongying Wu, et al. 2021. "A Genomic Catalog of Earth's Microbiomes." *Nature Biotechnology* 39: 499–509. <https://doi.org/10.1038/s41587-020-0718-6>
13. Liu, Yong-Xin, Tong Chen, Danyi Li, Jingyuan Fu, and Shuang-Jiang Liu. 2022. "iMeta: Integrated Meta-Omics for Biology and Environments." *iMeta* 1: e15. <https://doi.org/10.1002/imt2.15>
14. Li, Wenjun, Likun Wang, Xiaofang Li, Xin Zheng, Michael F. Cohen, and Yong-Xin Liu. 2022. "Sequence-Based Functional Metagenomics Reveals Novel Natural Diversity of Functioning CopA in Environmental Microbiomes." *Genomics, Proteomics & Bioinformatics*. <https://doi.org/10.1016/j.gpb.2022.08.006>
15. Bolyen, Evan, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, et al. 2019. "Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2." *Nature Biotechnology* 37: 852–57. <https://doi.org/10.1038/s41587-019-0209-9>
16. Zhang, Jingying, Yong-Xin Liu, Xiaoxuan Guo, Yuan Qin, Ruben Garrido-Oter, Paul Schulze-Lefert, and Yang Bai. 2021. "High-Throughput Cultivation and Identification of Bacteria from the Plant Root Microbiota." *Nature Protocols* 16: 988–1012. <https://doi.org/10.1038/s41596-020-00444-7>
17. Liu, Yong-Xin, Yuan Qin, Xiaoxuan Guo, and Yang Bai. 2019. "Methods and Applications for Microbiome Data Analysis." *Hereditas (Beijing)* 41: 845–826. <https://doi.org/10.16288/j.yczs.19-222>
18. Schloss, Patrick D., Sarah L. Westcott, Thomas Ryabin, Justine R. Hall, Martin Hartmann, Emily B. Hollister, Ryan A. Lesniewski, et al. 2009. "Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities." *Applied and Environmental Microbiology* 75: 7537–41. <https://doi.org/10.1128/aem.01541-09>
19. Edgar, Robert C. 2010. "Search and Clustering Orders of Magnitude Faster than BLAST." *Bioinformatics* 26: 2460–61. <https://doi.org/10.1093/bioinformatics/btq461>
20. Caporaso, J. Gregory, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D. Bushman, Elizabeth K. Costello, Noah Fierer, et al. 2010. "QIIME allows Analysis of High-Throughput Community Sequencing Data." *Nature Methods* 7: 335–36. <https://doi.org/10.1038/nmeth.f.303>
21. Gonzalez, Antonio, Jose A. Navas-Molina, Tomasz Kosciolk, Daniel McDonald, Yoshiki Vázquez-Baeza, Gail Ackermann, Jeff DeReus, et al. 2018. "Qiita: Rapid, Web-Enabled Microbiome Meta-Analysis." *Nature Methods* 15: 796–98. <https://doi.org/10.1038/s41592-018-0141-9>
22. Mitchell, Alex L., Alexandre Almeida, Martin Beracochea, Miguel Boland, Josephine Burgin, Guy Cochrane, Michael R. Crusoe, et al. 2020. "MGnify: The Microbiome Analysis Resource in 2020." *Nucleic Acids Research* 48: D570–D578. <https://doi.org/10.1093/nar/gkz1035>
23. Shi, Wenyu, Heyuan Qi, Qinglan Sun, Guomei Fan, Shuangjiang Liu, Jun Wang, Baoli Zhu, et al. 2019. "gcMeta: A Global Catalogue of Metagenomics Platform to Support the Archiving, Standardization and Analysis of Microbiome Data." *Nucleic Acids Research* 47: D637–D648. <https://doi.org/10.1093/nar/gky1008>
24. Liu, Yong-Xin, Yuan Qin, Tong Chen, Meiping Lu, Xubo Qian, Xiaoxuan Guo, and Yang Bai. 2021. "A Practical Guide to Amplicon and Metagenomic Analysis of Microbiome Data." *Protein & Cell* 12: 315–30. <https://doi.org/10.1007/s13238-020-00724-8>
25. Qian, Xu-Bo, Tong Chen, Yi-Ping Xu, Lei Chen, Fu-Xiang Sun, Mei-Ping Lu, and Yong-Xin Liu. 2020. "A Guide to Human Microbiome Research: Study Design, Sample Collection, and Bioinformatics Analysis." *Chinese Medical Journal* 133: 1844–55. <https://doi.org/10.1097/cm9.0000000000000871>
26. Parks, Donovan H., Gene W. Tyson, Philip Hugenholtz, and Robert G. Beiko. 2014. "STAMP: Statistical Analysis of Taxonomic and Functional Profiles." *Bioinformatics* 30: 3123–24. <https://doi.org/10.1093/bioinformatics/btu494>
27. Segata, Nicola, Jacques Izard, Levi Waldron, Dirk Gevers, Larisa Miropolsky, Wendy S. Garrett, and Curtis Huttenhower. 2011. "Metagenomic Biomarker Discovery and Explanation." *Genome Biology* 12: R60. <https://doi.org/10.1186/gb-2011-12-6-r60>
28. Langille, Morgan G. I., Jesse Zaneveld, J. Gregory Caporaso, Daniel McDonald, Dan Knights, Joshua A. Reyes, Jose C. Clemente, et al. 2013. "Predictive Functional Profiling of Microbial Communities Using 16S rRNA Marker Gene Sequences." *Nature Biotechnology* 31: 814–21. <https://doi.org/10.1038/nbt.2676>
29. Douglas, Gavin M., Vincent J. Maffei, Jesse R. Zaneveld, Svetlana N. Yurgel, James R. Brown, Christopher M. Taylor, Curtis Huttenhower, and Morgan G. I. Langille. 2020. "PICRUSt2 for Prediction of Metagenome Functions." *Nature Biotechnology* 38: 685–88. <https://doi.org/10.1038/s41587-020-0548-6>
30. Ward, Tonya, Jake Larson, Jeremy Meulemans, Ben Hillmann, Joshua Lynch, Dimitri Sidiropoulos, and John R. Spear, et al. 2017. "BugBase Predicts Organism-Level Microbiome Phenotypes." *bioRxiv*: 133462. <https://doi.org/10.1101/133462>
31. Louca, Stilianos, Laura Wegener Parfrey, and Michael Doebeli. 2016. "Decoupling Function and Taxonomy in the Global Ocean Microbiome." *Science* 353: 1272–77. <https://doi.org/10.1126/science.aaf4507>
32. Chen, Tong, Yong-Xin Liu, and Luqi Huang. 2022. "ImageGP: An Easy-To-Use Data Visualization Web Server for Scientific Researchers." *iMeta* 1: e5. <https://doi.org/10.1002/imt2.5>
33. Letunic, Ivica, and Peer Bork. 2021. "Interactive Tree Of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation." *Nucleic Acids Research* 49: W293–W296. <https://doi.org/10.1093/nar/gkab301>

34. Rognes, Torbjørn, Tomáš Flouri, Ben Nichols, Christopher Quince, and Frédéric Mahé. 2016. "VSEARCH: A Versatile Open Source Tool for Metagenomics." *PeerJ* 4: e2584. <https://doi.org/10.7717/peerj.2584>
35. Shen, Wei, Shuai Le, Yan Li, and Fuquan Hu. 2016. "SeqKit: A Cross-Platform and Ultrafast Toolkit for fasta/q File Manipulation." *PloS One* 11: e0163962. <https://doi.org/10.1371/journal.pone.0163962>
36. Wen, Tao, Penghao Xie, Shengdie Yang, Guoqing Niu, Xiaoyu Liu, Zhexu Ding, Chao Xue, et al. 2022. "ggClusterNet: An R Package for Microbiome Network Analysis and Modularity-Based Multiple Network Layouts." *iMeta* 1: e32. <https://doi.org/10.1002/imt2.32>
37. Gu, Zuguang, Lei Gu, Roland Eils, Matthias Schlesner, and Benedikt Brors. 2014. "Circlize Implements and Enhances Circular Visualization in R." *Bioinformatics* 30: 2811–12. <https://doi.org/10.1093/bioinformatics/btu393>
38. Lv, Zhiyao, Rui Dai, Haoran Xu, Yongxin Liu, Bo Bai, Ying Meng, Haiyan Li, et al. 2021. "The Rice Histone Methylation Regulates Hub Species of the Root Microbiota." *Journal of Genetics and Genomics* 48: 836–43. <https://doi.org/10.1016/j.jgg.2021.06.005>
39. Wang, Jinfeng, Jiayong Zheng, Wenyu Shi, Nan Du, Xiaomin Xu, Yanming Zhang, Peifeng Ji, et al. 2018. "Dysbiosis of Maternal and Neonatal Microbiota Associated with Gestational Diabetes Mellitus." *Gut* 67: 1614–25. <https://doi.org/10.1136/gutjnl-2018-315988>
40. Yuan, Jun, Tao Wen, He Zhang, Mengli Zhao, C. Ryan Penton, Linda S. Thomashow, and Qirong Shen. 2020. "Predicting Disease Occurrence with High Accuracy Based on Soil Macroecological Patterns of Fusarium Wilt." *The ISME Journal* 14: 2936–50. <https://doi.org/10.1038/s41396-020-0720-5>
41. Li, Xiaoqing, Jiayong Zheng, Xiuling Ma, Bing Zhang, Jinyang Zhang, Wenhuan Wang, Congcong Sun, et al. 2021. "The Oral Microbiome of Pregnant Women Facilitates Gestational Diabetes Discrimination." *Journal of Genetics and Genomics* 48: 32–39. <https://doi.org/10.1016/j.jgg.2020.11.006>
42. Qu, Kaiyang, Fei Guo, Xiangrong Liu, Yuan Lin, and Quan Zou. 2019. "Application of Machine Learning in Microbiology." *Frontiers in Microbiology* 10: 827. <https://doi.org/10.3389/fmicb.2019.00827>
43. Hernández Medina, Ricardo, Svetlana Kutuzova, Knud Nor Nielsen, Joachim Johansen, Lars Hestbjerg Hansen, Mads Nielsen, and Simon Rasmussen. 2022. "Machine Learning and Deep Learning Applications in Microbiome Research." *ISME Communications* 2: 98. <https://doi.org/10.1038/s43705-022-00182-9>
44. Wang, Nan, Mingyue Cheng, and Kang Ning. 2022. "Overcoming Regional Limitations: Transfer Learning for Cross-regional Microbial-Based Diagnosis of Diseases." *Gut* gutjnl-2022-328216. <https://doi.org/10.1136/gutjnl-2022-328216>
45. Wang, Jinfeng, Zhen Jia, Bing Zhang, Lei Peng, and Fangqing Zhao. 2020. "Tracing the Accumulation of In Vivo Human Oral Microbiota Elucidates Microbial Community Dynamics at the Gateway to the GI Tract." *Gut* 69: 1355–56. <https://doi.org/10.1136/gutjnl-2019-318977>
46. Knights, Dan, Justin Kuczynski, Emily S. Charlson, Jesse Zaneveld, Michael C. Mozer, Ronald G. Collman, Frederic D. Bushman, Rob Knight, and Scott T. Kelley. 2011. "Bayesian Community-Wide Culture-Independent Microbial Source Tracking." *Nature Methods* 8: 761–63. <https://doi.org/10.1038/nmeth.1650>
47. Shenhav, Liat, Mike Thompson, Tyler A. Joseph, Leah Briscoe, Ori Furman, David Bogumil, Itzhak Mizrahi, Itsik Pe'er, and Eran Halperin. 2019. "FEAST: Fast Expectation-Maximization for Microbial Source Tracking." *Nature Methods* 16: 627–32. <https://doi.org/10.1038/s41592-019-0431-x>
48. Dhariwal, Achal, Jasmine Chong, Salam Habib, Irah L. King, Luis B. Agellon, and Jianguo Xia. 2017. "MicrobiomeAnalyst: A Web-Based Tool for Comprehensive Statistical, Visual and Meta-Analysis of Microbiome Data." *Nucleic Acids Research* 45: W180–W188. <https://doi.org/10.1093/nar/gkx295>
49. Chong, Jasmine, Peng Liu, Guangyan Zhou, and Jianguo Xia. 2020. "Using MicrobiomeAnalyst for Comprehensive Statistical, Functional, and Meta-Analysis of Microbiome Data." *Nature Protocols* 15: 799–821. <https://doi.org/10.1038/s41596-019-0264-1>
50. Huang, Ancheng C., Ting Jiang, Yong-Xin Liu, Yue-Chen Bai, James Reed, Baoyuan Qu, Alain Goossens, et al. 2019. "A Specialized Metabolic Network Selectively Modulates *Arabidopsis* Root Microbiota." *Science* 364: eaau6389. <https://doi.org/10.1126/science.aau6389>
51. McDonald, Daniel, Morgan N. Price, Julia Goodrich, Eric P. Nawrocki, Todd Z. DeSantis, Alexander Probst, Gary L. Andersen, Rob Knight, and Philip Hugenholtz. 2011. "An Improved Greengenes Taxonomy With Explicit Ranks for Ecological and Evolutionary Analyses of Bacteria and Archaea." *The ISME Journal* 6: 610–18. <https://doi.org/10.1038/ismej.2011.139>
52. Zheng, Maosheng, Nan Zhou, Shufeng Liu, Chenyuan Dang, Yong-Xin Liu, Shishi He, Yijun Zhao, Wen Liu, and Xiangke Wang. 2019. "N<sub>2</sub>O and NO Emission from a Biological Aerated Filter Treating Coking Wastewater: Main Source and Microbial Community." *Journal of Cleaner Production* 213: 365–74. <https://doi.org/10.1016/j.jclepro.2018.12.182>
53. Chen, Tong, Haiyan Zhang, Yu Liu, Yong-Xin Liu, and Luqi Huang. 2021. "EVENn: Easy To Create Repeatable and Editable Venn Diagrams and Venn Networks Online." *Journal of Genetics and Genomics* 48: 863–66. <https://doi.org/10.1016/j.jgg.2021.07.007>
54. Gao, Chun-Hui, Guangchuang Yu, and Peng Cai. 2021. "ggVennDiagram: An Intuitive, Easy-to-Use, and Highly Customizable R Package to Generate Venn Diagram." *Frontiers in Genetics* 12: 706907. <https://doi.org/10.3389/fgene.2021.706907>
55. Chen, Qingwen, Ting Jiang, Yong-Xin Liu, Haili Liu, Tao Zhao, Zhixi Liu, Xiangchao Gan, et al. 2019. "Recently Duplicated Sesterterpene (C25) Gene Clusters in *Arabidopsis thaliana* Modulate Root Microbiota." *Science China Life Sciences* 62: 947–58. <https://doi.org/10.1007/s11427-019-9521-2>
56. Gu, Zuguang. 2022. "Complex Heatmap Visualization." *iMeta* 1: e43. <https://doi.org/10.1002/imt2.43>

57. Zhang, Jingying, Na Zhang, Yong-Xin Liu, Xiaoning Zhang, Bin Hu, Yuan Qin, Haoran Xu, et al. 2018. "Root Microbiota Shift in Rice Correlates with Resident Time in the Field and Developmental Stage." *Science China Life Sciences* 61: 613–21. <https://doi.org/10.1007/s11427-018-9284-4>
58. Xu, Jun, Jie-Ni Zhang, Bo-Hui Sun, Qing Liu, Juan Ma, Qian Zhang, Yong-Xin Liu, Ning Chen, and Feng Chen. 2022. "The Role of Genotype and Diet in Shaping Gut Microbiome in a Genetic Vitamin A Deficient Mouse Model." *Journal of Genetics and Genomics* 49: 155–64. <https://doi.org/10.1016/j.jgg.2021.08.015>
59. Carrión, Víctor J., Juan Perez-Jaramillo, Viviane Cordovez, Vittorio Tracanna, Mattias de Hollander, Daniel Ruiz-Buck, Lucas W. Mendes, et al. 2019. "Pathogen-Induced Activation of Disease-Suppressive Functions in the Endophytic Root Microbiome." *Science* 366: 606–12. <https://doi.org/10.1126/science.aaw9285>
60. Zgadzaj, Rafal, Ruben Garrido-Oter, Dorte Bodker Jensen, Anna Koprivova, Paul Schulze-Lefert, and Simona Radutoiu. 2016. "Root Nodule Symbiosis in Lotus Japonicus Drives the Establishment of Distinctive Rhizosphere, Root, and Nodule Bacterial Communities." *Proceedings of the National Academy of Sciences of the United States of America* 113: E7996–E8005. <https://doi.org/10.1073/pnas.1616564113>
61. Chen, Tingting, Xu Chen, Sisi Zhang, Junwei Zhu, Bixia Tang, Anke Wang, Lili Dong, et al. 2021. "The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types." *Genomics, Proteomics & Bioinformatics* 19: 578–83. <https://doi.org/10.1016/j.gpb.2021.08.001>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Liu, Yong-Xin, Lei Chen, Tengfei Ma, Xiaofang Li, Maosheng Zheng, Xin Zhou, Liang Chen, et al. 2023. "EasyAmplicon: An Easy-To-Use, Open-source, Reproducible, and Community-Based Pipeline for Amplicon Data Analysis in Microbiome Research." *iMeta*, 2, e83. <https://doi.org/10.1002/imt2.83>