Submission ID #: 61715
Scriptwriter Name: Bridget Colvin
Project Page Link: https://www.jove.com/account/file-uploader?src=18823123

Title: Assisted Selection of Biomarkers by Linear Discriminant Analysis Effect Size (LEfSe) in Microbiome Data

Authors and Affiliations: Fang Chang[1*], Shishi He[2*], and Chenyuan Dang[3]
*These authors contributed equally

[1]Marine Resources Research Centre, Tianjin Research Institute for Water Transport Engineering, M.O.T.
[2]Key Laboratory of Regional Energy Systems Optimization, Ministry of Education, College of Environmental Science and Technology, North China Electric Power University
[3]School of Environmental Science and Engineering, Huazhong University of Science and Technology

Corresponding Author:
Chenyuan Dang
dangchenyuan123@hotmail.com

Co-Authors:
changfang1999@163.com
heshishi@ncepu.edu.cn

Author Questionnaire

1. Microscopy: Does your protocol require the use of a dissecting or stereomicroscope for performing a complex dissection, microinjection technique, or similar? **N**

2. Software: Does the part of your protocol being filmed demonstrate software usage? **Y**

3. Interview statements: Considering the Covid-19-imposed mask-wearing and social distancing recommendations, which interview statement filming option is the most appropriate for your group? **Please select one**.

☒ Interview Statements are read by JoVE's voiceover talent.

4. Filming location: Will the filming need to take place in multiple locations (greater than walking distance)? **N**

Protocol Length
Number of Shots: **0**

Introduction

1. Introductory Interview Statements Voiceover Talent: please record Introduction statements

REQUIRED:

1.1. **JoVE Voiceover Talent**: The application of Linear Discriminant analysis Effect Size can solve the problem of finding good biomarkers with statistical differences between biologic groups **[1]**.

1.1.1. Use Figure 8

REQUIRED:

1.2. **JoVE Voiceover Talent**: Linear Discriminant analysis Effect Size provides a convenient method for identifying genomic biomarkers to characterize statistical differences between biologic groups **[1]**.

1.2.1. Use Figure 10

OPTIONAL:

1.3. **JoVE Voiceover Talent**: Be sure to take care with each step of the procedure, because the outcome of each preceding step may affect the subsequent step **[1]**.

1.3.1. Use 3.1.1.

Protocol

2. Linear Discriminant Analysis Effect Size (LEfSe) Native Analysis

2.1. After generating a linear discriminant analysis effect size input file **[1]**, run the commands as indicated to exclude the possibility of dependencies conflict and create a conda environment for linear discriminant analysis effect size **[2]**.

2.1.1. SCREEN: screenshot_1.mov: 00:00-00:55 *Video Editor: please speed up*
2.1.2. SCREEN: screenshot_1.mov:00:56-01:07

2.2. Use the commands as indicated to activate the created environment and to install linear discriminant analysis effect size with channel bioBakery **[1]**. To format data for linear discriminant analysis effect size, run the command to format the original file to the internal linear discriminant analysis effect size format **[2]**.

2.2.1. SCREEN: screenshot_1.mov: 02:35-03:28
2.2.2. SCREEN: screenshot_1.mov: 01:54-02:34 *Video Editor: please speed up*

2.3. To calculate the linear discriminant analysis effect size, run the command to perform a linear discriminant analysis and to generate the resulting data file **[1]**.

2.3.1. SCREEN: screenshot_1.mov: 02:25-03:28 *Video Editor: please speed up*

2.4. After the analysis, use the commands as indicated to plot the effect size of the biomarkers in a pdf file **[1]** and to draw the species tree to display the biomarkers in a cladogram **[2]**. To plot the differences of a single biomarker among different groups, use the command as indicated **[3]**.

2.4.1. SCREEN: screenshot_1.mov: 03:29-04:01 *Video Editor: please speed up*
2.4.2. SCREEN: screenshot_1.mov: 04:02-04:52
2.4.3. SCREEN: screenshot_1.mov: 04:53-07:03 *Video Editor: please speed up*

2.5. All of the features can also be drawn using the command if desired **[1]**.

2.5.1. SCREEN: screenshot_1.mov: 07:04-08:35

3. LEfSe Online Analysis (Galaxy)

3.1. For online LEfSe analysis using the galaxy server, navigate to the server. To upload the relevant files, click **Up** and **Choose local file** to select the files. Then select **Tabular** format and click **Start [1-TXT]**.

3.1.1. SCREEN: screenshot_2.mov: 00:00-00:34 *Video Editor: please speed up* **TEXT: http://huttenhower.sph.harvard.edu/galaxy**

3.2. To format the data for linear discriminant analysis effect size, click **LEfSe** (L-E-F-S-E) and **Format Data for LEfSe** select the specific rows for class, and click **Execute [1]**. To calculate the linear discriminant analysis effect size, click **LEfSe** and **LDA Effect Size**, select the parameter values according to the analysis requirements, and click **Execute [2]**.

3.2.1. SCREEN: screenshot_2.mov: 00:35-00:57 *Video Editor: please speed up*
3.2.2. SCREEN: screenshot_2.mov: 00:58-01:21 *Video Editor: please speed up*

3.3. To plot the linear discriminant analysis effect size results, click **LEfSe** and **Plot LEfSe Results** and click **Execute [1]**. To plot the cladogram, select the appropriate parameter values and click **Plot Cladogram** and **Execute [2]**.

3.3.1. SCREEN: screenshot_2.mov: 01:22-01:46 *Video Editor: please speed up*
3.3.2. SCREEN: screenshot_2.mov: 01:47-2:10 *Video Editor: please speed up*

3.4. To plot one feature, select the appropriate parameter values and click **Plot One Feature** and **Execute [1]**. To plot differential features, select the appropriate parameter values and click **Plot Differential Features** and **Execute** button **[2]**.

3.4.1. SCREEN: screenshot_2.mov: 02:11-02:40 *Video Editor: please speed up*
3.4.2. SCREEN: screenshot_2.mov: 02:41-03:16 *Video Editor: please speed up*

Protocol Script Questions

A. Which steps from the protocol are the most important for viewers to see?

n/a

B. What is the single most difficult aspect of this procedure and what do you do to ensure success?
2.4., 3.4.

Results

4. Results: Representative Microbiome LEfSe Characterization

4.1. Here the linear discrimination analysis scores of microbial communities with significant differences in each group determined analyzing the 16S ribosomal RNA gene sequences of three samples is shown **[1]**.

4.1.1. LAB MEDIA: Figure 8 *Video Editor: please sequentially emphasize blue, green, and red sections*

4.2. In this figure, the biomarkers with significant differences and species trees between different classification levels can be observed **[1]**.

4.2.1. LAB MEDIA: Figure 9

4.3. The circles radiating from the inside to the outside represent the classification levels from phylum to genus **[1]**, with the diameter of each species circle representing the level of abundance of each classification **[2]**.

4.3.1. LAB MEDIA: Figure 9 *Video Editor: please emphasize black circle lines form middle to outside of figure*
4.3.2. LAB MEDIA: Figure 9

4.4. The species with no significant differences appear in yellow **[1]** and the significantly different species biomarkers are colored to match the corresponding groups **[2]**.

4.4.1. LAB MEDIA: Figure 9 *Video Editor: please emphasize yellow circles*
4.4.2. LAB MEDIA: Figure 9 *Video Editor: please sequentially emphasize green circles and green shading*, *red circles and red shading, and blue circles and blue shading*

4.5. The corresponding species names of the biomarkers shown in the plot are listed here **[1]**.

4.5.1. LAB MEDIA: Figure 9 *Video Editor: please emphasize list of names*

4.6. Here a representative abundance barplot for one biomarker that shows differences among the different groups according to the linear discriminant analysis effect size results is shown **[1]**.

4.6.1. LAB MEDIA: Figure 10

4.7. The solid line represents the average relative abundance **[1]**, the dotted line represents the median relative abundance **[2]**, and each column represents the relative abundance of each sample in different groups **[3]**.

4.7.1. LAB MEDIA: Figure 10 *Video Editor: please emphasize solid lines*
4.7.2. LAB MEDIA: Figure 10 *Video Editor: please emphasize dotted lines*
4.7.3. LAB MEDIA: Figure 10 *Video Editor: please emphasize red data bars*

Conclusion

5. **Conclusion Interview Statements** Voiceover Talent: please record Conclusion statements

    5.1. **JoVE Voiceover Talent**: Principal Component Analysis can also be performed, as the dimensionality education is directly related to the data dimension and the projected coordinate system is orthogonal **[1]**.

        5.1.1. Use 2.1.1.

    5.2. **JoVE Voiceover Talent**: As the demand for high-dimensional data analysis increases, this method will aid in the exploration of biomarker features of interest **[1]**.

        5.2.1. Use Figure 9