Microbial Community Analyses Using Colony Phenotype and Sequencing Databases

Instructor Notes

The Microbial Community Analysis Using Colony Phenotype Database and Microbial Community Analysis Using Colony-based Sequencing Database, may be used as two independent data sources for student to address questions about bean beetle microbiome community ecology. These analyses are not contingent on each other. The Colony Phenotypes were collected by undergraduate students who plated microbiome bacteria on media as described Microbiome Plating protocol and the Colony-based Sequencing is based on the Sanger sequences from picked colonies on which PCR was performed using 16S V4 primers as described in the Colony PCR protocol (protocols available at www.beanbeetles .org).

Materials

This is a computer-based activity so students will need individual computers with internet access. This study could be conducted by students working in pairs at one computer, but there is more to be gained by having students work individually while collaborating. Basic data manipulation requires Microsoft Excel. Data analyses may be conducted using Excel or the freely available RStudio program. Configuring computers prior to conducting the data manipulation and analyses is strongly suggested since it will minimize student frustration.

RStudio is an open access statistical analysis interface that requires the installation of the open access R program. R and RStudio may be downloaded from:
https://cran.r-project.org/
https://www.rstudio.com/products/rstudio/download/

Preparation for Conducting Community Ecology Analyses

We advise instructors to conduct the entire data manipulation and analysis that you want students to conduct prior to meeting with a class. That will ensure you are familiar with the potential problems students may have when conducting this study. The screen shots included in the Student Handout are from a WindowsOS computer so the screens that you and your students see will depend on the computer operating system and the version of Excel that you use. Microsoft Excel (or a similar spreadsheet program) is required to manipulate data prior to conducting the community ecology analyses. Once data are configured and organized, the data may be analyzed using either Excel or RStudio. RStudio will permit a more comprehensive analysis, but Excel will permit students to calculate the basic variables and indices necessary for community comparisons, as well as understand the underlying formulas used for calculating the indices.

The data manipulation required prior to data analysis could be performed by the instructor as a way to save time and have students just focus on the community ecology comparisons. This streamlining of the data manipulation may be appropriate depending on the specific learning goals for your course.

The basic community ecology variables: species (taxon) richness, Simpson Index (as well as the reciprocal and inverse Simpson), and Shannon-Weaver Index may be calculated using either Excel or RStudio. Calculating an index of community distance (or dissimilarity), such as the Bray Curtis Dissimilarity Index or the Morista-Horn Index of Dissimilarity, as well as plotting species accumulation curves, will be much more tractable using RStudio.

Choosing the dataset(s) that students evaluate, colony phenotype or colony-based sequence is a matter of preference and the time you want students to spend on these analyses. The two datasets are not simply different methods of defining taxa in a community analysis but also represent different ways of defining a community. Analyzing both types of data to address similar questions could lead to a very productive discussion with students on the limitations of each database and the analysis of the data within them. Each database could be used independently of the other and there is no necessity for one database analysis to precede the other. If students worked on one database, the approximate time required to perform the data manipulation and the analysis using either Excel or RStudio is one and one-half hours. Completing all the activities presented here would require two 3-hour laboratory periods.

This study is based on Blumer LS, Beck CW 2020. **Introducing community ecology and data skills with the bean beetle microbiome project**. Advances in Biology Laboratory Education 41.