Biological Community Comparisons using PRIMER-E Software

1. SCOPE AND APPLICATION

1.1 This SOP describes analyses to compare taxa lists using PRIMER-E software. These analyses are conducted for Algal Taxonomy QC exercises described in SOP AB-14, but can be used for any investigation in which taxa lists (algae or macroinvertebrates) are compared.

2. SUMMARY OF METHOD

2.1 Bray-Curtis similarity indices are calculated to determine the percent similarity between two or more taxa lists and counts at the species- or genus-level. SIMPER analysis is conducted to determine the taxa that contributed most to differences between taxa lists. In comparisons of two or more samples and two or more analysts, ANOSIM analysis is conducted to determine whether taxa lists differed among samples or analysts and the sources of those differences.

3. APPARATUS AND EQUIPMENT

3.1 Primer-E software and documentation
Primer software and documentation may be obtained from Lab Data Support with Supervisor approval, provided there is sufficient license. This SOP is written for Version 5, but instructions are very similar for Version 6 and tips are given when commands differ.

3.2 Microsoft Excel software

3.3 Taxa lists

4. DATA ANALYSIS PROCEDURES

4.1 Import Excel QC data sheet.

4.1.1 Open the Primer-E program. Import the QC spreadsheet into Primer using File\Open. When the "Open" window appears, change "Files of type:" to "Excel files (*.xls)". Choose the file to open and either double-click on the file name or highlight it and click "Open."

4.1.2 In the "Excel Files properties" window make sure the "Data type" radio button highlighted is "Sample data." Leave the "Includes title" box checked. Click "OK." This creates a *.pri file.

Note: For PRIMER 6, select the worksheet to import from the drop-down menu, Data type●Sample data. Click Next>. Then, √ Title, √ Row labels, Orientation●Samples as columns, Data type●Abundance, Blank= ●Zero. Click Finish.

4.1.3 Perform Bray-Curtis similarity (section 4.2), SIMPER (section 4.3), add Factors (section 4.4), or perform ANOSIM (section 4.5) as needed.

4.2 Calculation of Bray-Curtis similarity for species- and genus-level taxa lists.

Bray-Curtis similarity is a measure of how similar data sets are to each other. Bray-Curtis similarity has a range of 0-100 and is affected by both taxa present and their relative abundance.

The OFFICIAL COPY is the on-line version. All other copies are considered unofficial and uncontrolled.

4.2.1 Choose Data\Similarity from the main menu. When the "Similarity" window opens, choose the "Samples" radio button under "Analyze between" and the "Bray-Curtis Similarity" radio button under "Measure". From the "Transformation" drop down list, choose "None". Do not check the "Standardize" check box. Click "OK."

Note: In PRIMER 6, click on Analyze\Resemblance, Analysis between●Samples, Measure●Bray-Curtis similarity, (no check in Add dummy variable). Click OK. When warned that no overall transformation or weighting has been applied, click OK.

Note: The output will appear as a triangular similarity matrix, with the separate analyses as row and column headers. Each inter-analysis comparison appears once in the matrix. For three samples, two replicate analyses, and three analysts, there will be 18 analyses in the data file and 306 similarity coefficients in the similarity matrix.

4.2.2 A results sheet and a similarity matrix will be generated. Examine the similarity matrix. For comparisons between two samples (initial demonstration of proficiency or ongoing QC), there will be only one similarity value and this value should be >70%. For comparisons between multiple samples and analysts (quarterly QC), intra-sample (same sample analyzed twice by one analyst) and inter-analyst (same sample analyzed by different analysts) similarities should be >70%. If similarities are >70%, the QC has passed. Record the similarity value in the original Excel file.

Note: A target similarity value of 70% was chosen because proficient analysts routinely exceed a Bray-Curtis value of 70% in inter-aliquot comparisons from the same sample.

4.2.3 If Bray-Curtis similarities are below 70%, repeat steps 4.2.1 and 4.2.2, but choose "Square root" from the "Transformation" drop down list. If similarities are >70%, the QC has passed with square root transformation. Record the similarity value in the original Excel file and note that square root transformation was required.

Note: In PRIMER 6, make the original taxa list the active worksheet. Click Analyze\Pretreatment\Transform(overall) and select from the drop down menu to perform the square-root transformation. Then follow steps in 4.2.1 to get to Bray-Curtis similarity.

4.2.4 If the 70% cutoff is not met in step 4.2.3, create an aggregation file that combines taxa in the sample to the generic level. The aggregation file may already have been created in step 4.1.3 of AB-14. If not, to create an aggregation file using Excel, open the original Excel spreadsheet containing the taxa and counts. Type "Species" over the column heading "Taxa." Insert a column after the "Species" column. In this new second column, type the header "Genus" and type in the genus name of each species in the first column. If the lowest possible identification was genus or higher (i.e. family, order, class, or phylum), the "Genus" column should contain the same entry as the "Species" column. Save this file as an Excel file with the same name as the original file, but add the suffix "_Agg_File" before the .xls extension.

Note: The goal of combining taxa to the generic level is to determine if analysts differ in level of identification (i.e. species vs. genus or genus vs. family).

4.2.5 Import the Aggregation file into Primer using menu option File\Open. When the "Open" window appears, change "Files of type:" to "Excel files (*.xls)." Choose the file to open and either double-click on the file name or highlight it and click "Open."

The OFFICIAL COPY is the on-line version. All other copies are considered unofficial and uncontrolled.

Note: Primer will automatically combine taxa into higher categories using an Aggregation file. An Aggregation file will contain * in the first column of the file.

4.2.6 In the "Excel files properties" window make sure the "Data type" radio button highlighted is "Aggregation data." Uncheck the "Includes title" box. Click "OK." When the "Aggregation Data Properties" window opens, it is not necessary to give this file a title.

Note: For PRIMER 6, select the worksheet to import from the drop down menu, Data type●Aggregation Data. Click Next>. Then, uncheck the Title box. Click Finish.

4.2.7 The associated data file should be open in Primer from the previous steps of this analysis. Make sure the original taxa file is the active window.

Note: The active window on the Primer desktop will have a dark blue line across the top of the file window.

4.2.8 Combine taxa to the generic level by choosing menu option Data\Aggregate. When the "AGGREGATE" window opens, make sure the file name that appears in the "Aggregation worksheet:" drop down list is the correct file name. It should match your data file name but have the "_Agg_File" suffix. For "Base column:" choose "Species," and for "Mapped column:" choose "Genus." Click "OK," and a new spreadsheet window will appear containing the taxa aggregated to the generic level.

Note: In PRIMER 6, click Tools\Aggregate. From the 3 drop down menus, select the _Agg. File, From level/species, To level/Genus. Click OK.

4.2.9 Using the new spreadsheet generated in Primer at the generic level, repeat calculation of Bray-Curtis similarity following steps 4.2.1 to 4.2.3. If the similarities are >70%, without or with transformation at the generic level, the QC has passed. When recording the similarities in the original Excel file, note that the QC passed at the generic level but not species level.

4.2.10 If the similarities do not pass the 70% cutoff at any level with square root transformation, perform the SIMPER test (section 4.3) to determine the sources of dissimilarity.

4.3 SIMPER test

The SIMPER routine determines what taxa contribute most to the differences between counts.

4.3.1 Make the original, un-aggregated *.pri file the active window in Primer.

Note: In PRIMER 6, you must add a Factor before proceeding (see section 4.4).

4.3.2 From the main menu, choose Analyze\SIMPER…. When the "SIMPER" window appears, do not check the "Standardize data" check box. Leave "Transformation" as "None" and "Cutoff percentage" as "90." Uncheck the "List only higher contributing species" check box. Choose "Sample" for "Factor name". Click "OK."

Note: In PRIMER 6, click on Analyze\SIMPER, Design●One way, Factor A: Analyst, Measure●Bray-Curtis similarity, (no check in List only higher contributing species), Cut-off percentage: 90. Click OK. When warned that no overall transformation or weighting has been applied, click OK.

4.3.3 A "Results" sheet will appear with the sample information and a list of taxa with dissimilarity values. Print this file immediately or copy the information to another file. If Primer is closed without printing this file, it will not be saved.

The OFFICIAL COPY is the on-line version. All other copies are considered unofficial and uncontrolled.

Note: Sometimes when this window appears, it looks as if it is blank, but it is just scrolled to the bottom of the page. Use the up arrow on the right side of the page to scroll up and see the data.

4.3.4 Using the taxa list and Average Dissimilarity values generated in the "Results" sheet, compare the differences between taxa lists. Those taxa contributing the highest percentage of the dissimilarity between samples (from the column labeled "Contrib%" on the "Results" sheet) should be examined to determine the source of the discrepancy (count, level of identification, etc.). All taxa contributing 5% or more of the total inter-sample dissimilarity should be examined.

Note: The output lists the taxa in order of the greatest differences between counts. The Bray-Curtis dissimilarity between counts is one minus the Bray-Curtis similarity between counts. Each taxon that is not identical in the samples contributes a portion of the inter-sample dissimilarity.

4.4 Adding Factors to dataset

Primer uses factors to identify groups of samples in a data set. Factors are added if there are more than two (2) taxa lists being compared and lists can be grouped by a factor such as analyst or sample. Factors are needed to run the ANOSIM analysis.

4.4.1 Make the original, un-aggregated *.pri file the active window in Primer.

4.4.2 From the main menu, choose Edit\Factors…. When the "Factors" window appears, choose Factors\Add…and type "Analyst" into the "Add factor named" box that appears. Click "OK." Enter the factor names in the column beneath "Analyst" using the initials of the analyst.

For Quarterly QC data, repeat the Factors\Add… step and type "Sample" for the second factor. Enter the factor names in the column beneath "Sample" using the sample codes created in Step 1 of this subsection (II). Click "OK." Click "OK" again.

Note: The Analyst factors should be the initials of each analyst. The Sample factors should be short codes like those established in Section II, Step 1 (e.g. A-MH-1). Using short abbreviations as factors facilitates graphical analyses of data.

Note: In PRIMER 6, right clicking over the taxa list also brings up the edit menu.

4.4.3 Save the data file by choosing File\Save as… from the main menu. Save the file with the same name and in the same location as the original Excel file generated in Section II, Step 1. When the "Save" window appears, choose "PRIMER data files (*.pri)" from the "Save as type…" drop down list.

Note: In PRIMER 6, File/Save Workspace As or File/Save Results As???? PRIMER Workspace files (*.pwk)"

4.5 ANOSIM test

ANOSIM is a non-parametric analysis of variance analogue using similarity coefficients and permutation tests of significance. It is used to compare taxa lists among samples and analysts in the Quarterly QC exercise (see SOP AB-14).

4.5.1 The similarity matrix with the most passing (>70%) Bray-Curtis similarity coefficients should be the active worksheet.

The OFFICIAL COPY is the on-line version. All other copies are considered unofficial and uncontrolled.

4.5.2 From the main menu, choose Analyze\ANOSIM. When the "ANOSIM" window appears, click the radio button "Two way crossed" under "Design." Choose "Sample" for "Factor A:" and "Analyst" for "Factor B." Do not choose anything for "Output", and leave "Max permutations" at 999. Click "OK."

Note: In PRIMER 6, click on Analyze\ANOSIM, Design●Two way crossed (with replicates), Factor A: Sample, Factor B: Analyst, Max permutations 999, √ pairwise tests to worksheet, √ Plot histogram, √ R values to file. Click OK.

4.5.3 A results sheet will appear. The section headed by "ANOSIM" contains the analytical results. Primer first lists the factors in the analysis and then reports their assignment for each sample identifier (the column heading in the data file).

Note: Sometimes when this window appears, it looks as if it is blank, but it is just scrolled to the bottom of the page. Use the up arrow on the right side of the page to scroll up and see the data.

4.5.4 The section headed "TESTS FOR DIFFERENCES BETWEEN Sample GROUPS" reports the overall significance of all comparisons between samples. "Sample" is the name of the Factor "Sample" when used in this context. Where three different samples were used in the exercise, the R value under the Global Test should be >0.90 with an associated probability level of <0.01. All samples should be clearly differentiated from each other (R.070, $p<0.05$) when three different samples are used in the exercise.

Note: R is a measure of the strength of the result and ranges from –1 to 1. A value of 1.0 means that samples are maximally different from each other while a value of 0 indicates the samples are identical. Negative values suggest that samples between groups are more similar to each other than within groups and are very unlikely to occur.

Note: Below the Global Test results is a table of R-values for the possible pair-wise comparisons between samples. In the Quarterly QC analysis, there are three pair-wise comparisons (Sample 1-2, Sample 2-3, and Sample 1-3). All samples should be clearly differentiated from each other ($R > 0.9$, $p<0.01$) when three different samples are used in the exercise.

Note: When two aliquots of the same sample are tested for the Quarterly QC, that pair of samples should have an R value < 0.5 and should not differ significantly from each other, but should differ significantly from the third sample. Check this in the pair-wise R-values comparisons section of the Results Log under the Global Test results.

4.5.5 The second section of the Results Log is headed "TESTS FOR DIFFERENCES BETWEEN Analyst GROUPS". The Global Test section will report a Global R. The R-value should be <0.5 and the associated probability should be >0.05. This result indicates that there are no significant differences among analysts under the most stringent test conditions.

Note: Below the Global Test results is a table of R-values for the possible pair-wise comparisons between analysts. In the Quarterly QC analysis there are three pair-wise comparisons (Analyst 1-2, Analyst 1-3, and Analyst 2-3).

4.5.6 If the results of the ANOSIM analysis match those detailed in steps 4.5.4 and 4.5.5, the QC exercise has passed. Record this information in the original Excel file.

The OFFICIAL COPY is the on-line version. All other copies are considered unofficial and uncontrolled.

Note: This standard demonstrates that samples may be clearly differentiated from each other and that results from different analysts cannot be statistically distinguished from each other.

4.5.7 If the results of the ANOSIM analysis do not match those detailed in steps 4.5.4 and 4.5.5, examine the similarity matrix used in the ANOSIM analysis. If the Bray-Curtis similarity coefficients were calculated on untransformed (raw unit count) data at either the lowest taxonomic or generic level, re-open the raw data file and calculate Bray-Curtis similarity choosing "Square root" for "Transformation" following Section I, Step 3.

4.5.8 Repeat the analysis in Section IV, Steps 1-5 on the square root-transformed similarity matrix. If the results are as described in Section IV, Steps 4-5, the statistical portion of the QC exercise has passed but note that square root transformation of the raw data was required for the ANOSIM analysis in the original Excel file.

4.5.9 If the statistical portion of the QC exercise did not pass, follow section 4.3 to generate a SIMPER file if one has not already been generated. Training will be targeted to remedy any problems identified in this exercise.

Note: Below the Global Test results section headed "TESTS FOR DIFFERENCES BETWEEN Analyst GROUPS" is a table of R-values for the possible pair-wise comparisons between analysts. In this analysis there are three pair-wise comparisons (Analysts 1-2, 2-3, and 1-3). Check these values to see if one analyst is significantly different from the other two or if all analysts differ from each other.

Appendix of Significant Changes:

03/17/2008 Created SOP AB-15-1.0 from portions of SOPs AB09, AB09_01, and AB09_03 that contained instructions for PRIMER analyses. Added updated commands for PRIMER v6.0.

04/15/2011 Added step 4.1 to import a file; re-numbered remaining sections. Added directions for PRIMER 6.

03/25/2015 Formatted for 508 compliance.

The OFFICIAL COPY is the on-line version. All other copies are considered unofficial and uncontrolled.